Journal of Food Composition and Analysis 133 (2024) 106456

Contents lists available at ScienceDirect

Journal of Food Composition and Analysis

journal homepage: www.elsevier.com/locate/jfca

One shot-proteomic profiling, allergenicity assessment and varietal discrimination of lentils seeds by discovery high resolution tandem mass spectrometry

A. Lamonaca [a,b], E. De Angelis [a], L. Monaci [a], R. Pilolli [a,*]

[a] Istituto di Scienze delle Produzioni Alimentari, Consiglio Nazionale delle Ricerche, via G. Amendola 122/O, Bari 70126, Italy
[b] Dipartimento di Scienze del Suolo, della Pianta e degli Alimenti, Università degli Studi di Bari Aldo Moro, via G. Amendola 165/a, Bari 70126, Italy

ARTICLE INFO

Keywords:
Lentil
Protein profiling
Allergenicity assessment
Varietal comparison
Discovery proteomics

ABSTRACT

Lentils (*Lens culinaris*) represent important food staple replacing meat products in human diet with various applications of foods and feeds due to the high protein content. This work reported on the comparative proteomic profiling of four lentils commercial varieties Crimson, Eston, Laird and Black, providing novel knowledge on the differential expression of specific protein accessions with specific focus placed on the major protein families also investigating their allergenicity risk. The electrophoretic profile of lentil extracts confirmed that most of the proteomic profile was conserved across the investigated varieties, with only few differences highlighted for convicillin, vicilin and legumin subunits. A more in-depth analysis was carried out by one-shot discovery proteomics and provided the unequivocal identification and label-free quantification of 365 protein entries ranked into the main protein families. Statistically significant differences were disclosed for most of these protein groups and/or subgroups, cupins being the major contributors to the protein profile for all lentil samples. Noteworthy, the 7 S/11 S ratio describing seed nutritional quality highlighted clear differences among the four varieties with highest values reported for Eston and Laird samples. However, such higher nutritional quality was offset by a higher risk of allergenicity for sensitized individuals; indeed, the *in-silico* allergenicity prediction reported that most of the identified proteins presented either strong or weak evidence of immunogenicity with highest abundance in the Eston and Laird lentil samples. In perspective, the differential expression of specific protein accessions suggested the possibility to identify protein markers for varietal discrimination: a preliminary multivariate statistical analysis was accomplished to this aim.

1. Introduction

Pulses, defined as the edible seeds of legumes (Capurso et al., 2018), mainly include chickpea (*Cicer arietinum* L.), pigeonpea (*Cajanuscajan* L.), mungbean (*Vigna radiata* L. Wilczek), urdbean (*Vigna mungo* L. Hepper), lentil (*Lens culinaris* L.) and fieldpea (*Pisum sativum* L.). They represent an important component of a healthy diet due to their high content of proteins (20–35 %, on average), dietary fiber and carbohydrates (60 %), and to their low glycemic index (Marinangeli et al., 2017, Cavalluzzi et al. 2022).

India is the largest producer, consumer, and importer of pulses in the world contributing to 28.9 % of the global production (FAOSTAT, 2021). Pulses are important for the nutritional security of the Indian population with a vegetarian diet. Their own demand for pulses is currently growing at 2.8 % per annum, with an estimation of 27.5 million tonnes (mt) of pulses required by 2025 (Chauhan et al., 2016). Pulses can be grown on a wide range of soil and climatic conditions, and play important role in crop rotation, mixed and inter-cropping, maintaining soil fertility through nitrogen fixation, release of soil-bound phosphorus, and contribute significantly to the sustainability of food systems (Gan et al., 2015). There is also a growing interest towards legumes crops in Europe, that currently produces about 10 million tonnes of pulses for year (about 12 % of world production), occupying almost 5 million hectares for their production (FAOSTAT, 2021).

Lentil (*Lens culinaris*) is an important food legume with various uses as food and feed because of its protein-rich grains and straw. World lentil production has risen steadily from an average of 0.85 mt in 1961–5.6 mt in 2021. The two major geographical regions of lentil production are

* Corresponding author.
E-mail address: rosa.pilolli@ispa.cnr.it (R. Pilolli).

https://doi.org/10.1016/j.jfca.2024.106456
Received 23 April 2024; Received in revised form 14 June 2024; Accepted 17 June 2024
Available online 18 June 2024

Canada (28.4 %), and India (26.6 %). (FAOSTAT, 2021). The *Lens culinaris* Medik. cultivation history dates back the 7th century B.C. Initially, lentils were cultivated around the Mediterranean Basin, then cultivation extended to the Middle East, Ethiopia and the Indian subcontinent (Ladizinsky, 1979; Duran and De La Vega (2004)), thus generating a myriad of different native varieties characterized by a remarkable genetic variability. *Lens culinaris* Medik. showed a high adaptation to different environmental conditions and different landraces developed adaptive genetic complexes that were conserved through genetic linkage and selection (Erskine, 1997). Due to their genetic variability, nowadays, lentils come in various colours, including red, green, black, yellow, orange or brown depending on the cultivar and the composition of seed coats and cotyledons. The major commercial market classes of lentils are red (based on the cotyledon colour of dehulled seeds) and green (based on the seed coat colour) (Khazaei et al., 2019). Large green lentils (with yellow cotyledon) are primarily marketed in Europe, and parts of the Middle East and South America, mostly consumed as whole seeds (Muehlbauer et al., 2009).

Lentils are rich in proteins and their content ranges from 20 % to 36 % of the total seed weight (on dry matter) depending on the variety considered (Khazaei et al., 2019). The most important group of proteins are seed storage proteins, which represent approximately 80 % of the total seed proteins. These proteins are present in the seeds cotyledons and their primary function is to provide nutrients (nitrogen, carbon and sulphur) during the germination phase and growth/development of the plant. They are also involved in some plant defence mechanisms or antimicrobial activities (de Souza Cândido et al., 2011). Storage proteins can be classified into four groups based on their solubility in different solvents: globulins that are soluble in saline solutions (about 70 %), albumins soluble in water (about 16 %), glutenins soluble in acid solutions (about 11 %), prolamins soluble in ethanol (about 3 %) (Khazaei et al., 2019). Globulins can be divided into classes 7 S (vicilines and convicilines) and 11 S (legumines) depending on the sedimentation coefficient; vicilins have a molecular weight of approximately 20–82 kDa, legumines have a molecular weight of approximately 14–92 kDa. Albumins, glutelins and prolamins have a molecular weight of approximately 20–82, 17–46 and 17–64 kDa, respectively (Boye et al., 2010). In addition to storage proteins, a high number of metabolic proteins such as enzymes and structural proteins involved in the different metabolic/physiological processes of the seed have also been identified in lentil seeds (Khazaei et al., 2019).

Although the proteomic profile of lentils has been widely studied, only few information is available on differential protein expression upon varieties and/or cultivars (Scippa et al., 2010). Nevertheless, insight on this topic can be very interesting to feature specific dissimilarities among lentils varieties accounting for peculiar nutritional and/or anti-nutritional values. Moreover, since current dietary trends include an increased consumption of plant protein (Quintieri et al., 2023), lentils, among other plant proteins, were included in a “watch list” of concerning foods, prioritized for their allergenicity, growing consumption and sensitization occurrence in vulnerable population (World Health Organization, 2022).

In this frame, the present work focused on the differential characterization of the proteomic profile of dry mature seeds belonging to four *Lens culinaris* Medik. commercial varieties, namely Crimson, Eston, Laird and Black. The investigation was carried out on commercial samples with a common specified origin, Canada as main supplier for the Italian market of pulses. The set of samples shared not only the geographical origin but also the specific provider and year of production to avoid that climatic conditions and agronomic practices might affect the protein profiles likely differing only in the genotype. A particular focus was placed on the observed expression of main proteins and their *in-silico* allergenicity assessment. Typically, the overwhelming presence of storage proteins in mature seeds impairs the detection of proteins involved in different metabolic processes. In this investigation the recourse to an advanced proteomic approach based on untargeted high resolution mass spectrometry and bioinformatic analysis allowed to feature also proteins involved in disease/defence, metabolic and regulatory functions and provided novel knowledge on the differential expression of specific protein accessions.

2. Materials and methods

2.1. Chemicals and reagents

Salts, solvents (water, acetonitrile), and other reagents (formic acid, trizma base, urea, ammonium bicarbonate, dithiothreitol, iodoacetamide, acetic acid, chloridric acid) were purchased from Sigma–Aldrich (Milan, Italy) and VWR International PBI (Milan, Italy). Trypsin Gold Mass Spectrometry Grade was purchased from Promega (Milan, Italy). Disposable desalting cartridges PD-10 were purchased from Cytiva, GE Healthcare Life Sciences (Milan, Italy). Sep-Pak C18 solid phase extraction (SPE) columns (1 cc, 50 mg) were purchased from Waters (Milford, Massachusetts, USA).

2.2. Lentils samples

The study was performed on commercial varieties of dried lentil seeds (Lens culinaris Medik) purchased from the "Fulvio Corina" company (Taviano (LE), Italy), all produced in Canada, one of the main supplier countries for the Italian market of legume. The four varieties were: (1) Crimson, (2) Eston, (3) Laird and (4) Black. The lentil seeds were finely ground by a Retsch ZM 200 (Retsch, Haan, Germany) laboratory mill at 14 000 rpm, obtaining grounded samples with particle size $\leq$500 μm, thoroughly homogenized, and stored under vacuum at +4 ∘C till protein extraction.

2.3. Sample preparation protocol of protein from ground seed samples

Aliquots of protein from ground samples were extracted with 30 mL of denaturing buffer (200 mM Tris HCl pH 9.2, 8 M urea) with a solid: liquid ratio 1:10. After the buffer addition, the mixtures were shaken on a vortex for 2 min, mixed on an orbital shaker for 30 min and sonicated in a water bath for 15 min, at room temperature (Pilolli et al., 2024). The samples were then centrifuged for 10 min at 3500 g and the collected supernatants were manually filtered through 1.2 μm cellulose acetate syringe filters. The resulting filtrated solutions were purified by size exclusion chromatography (SEC; 5 kDa cut-off) on disposable cartridge. The columns were conditioned beforehand with three aliquots of water (4 mL each) followed by four aliquots of 50 mM ammonium bicarbonate buffer (AB, 4 mL each). The “Spin elution” protocol was carried out according to the producer instructions (Pilolli et al., 2024).

The total protein content of the protein extracts was quantified using a commercial kit for colorimetric assays (Quick Start™ Bradford protein assay, Bio-rad Laboratories) according to the producer instructions. Trypsin digestion was carried out on 500 μL of the SEC eluted fraction after proper denaturation (incubation for 15 min at 95 °C), reduction (addition of 50 μL of 500 mM dithiothreitol solution and incubation for 30 min at 60 °C), and alkylation steps (addition of 100 μL of 100 mM iodoacetamide solution and incubation for 30 min at 22 °C). The solutions were then diluted (20x) and digested with 5 μL of trypsin solution (1 μg/μL in 50 mM acetic acid) added to the samples (theoretical trypsin to protein ratio: 1/100) according to the Bradford assay values. The enzymatic hydrolysis was stopped after 16 h of incubation at 37°C by acidification with hydrochloric acid (7 μL, 6 M). The peptide pool was collected after centrifugation (13000 g; 10 min) and purified on Sep-Pak C18 SPE columns (1 cc, 50 mg) (Monaci et al., 2020). The columns were activated with ACN (3 mL) and conditioned using 0.1 % formic acid in water (3 mL). 500 μL of samples were loaded onto the column and washed with 0.1 % formic acid in water (3 mL). The peptides were eluted with acetonitrile/0.1 % formic acid in water mixture 80/20 (v/v) (1 mL). The eluted solutions were concentrated by evaporation under

nitrogen flow (N2; RT) up to dryness. The dried extracts were solubilized in 5 % acetonitrile in 0.1 % formic acid in water solution (100 μL), vortexed, and transferred into an injection vial and analyzed by UHPLC-MS/MS. Four independent samples for each lentil variety were analysed.

2.4. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was performed on 8–16 % polyacrylamide pre-cast gels Mini-PROTEAN TGX Any KDa, 10 well, 50 μL/well (Bio-Rad Laboratories, Segrate, MI, Italy) (Uasuf et al., 2020). 10 or 20 μg of protein extracts were mixed with Laemmli buffer (62.5 mM TrisHCl, pH 6.8, 25 % glycerol, 2 % SDS, 0.01 % bromophenol blue, 100 mM DTT) in a 1:1 proportion and then denatured for 5 min at 95 ∘C. The electrophoretic separation was performed in a running buffer (25 mM Tris, 192 mM glycine, 0.1 % SDS) at 80 V until the end. Gels were stained by a Coomassie Brilliant Blue G-250 solution. The bands were detected on a ChemiDoc™ Imaging System (Bio-Rad Laboratories, Segrate, MI, Italy). Precision Plus Protein™ All Blue Standards (10–250 kDa, Bio-Rad Laboratories, Segrate, MI, Italy) were used as protein molecular weight references.

2.5. Discovery HPLC-MS/MS analysis and software based identification

UHPLC-MS/MS analyses were performed on an Ultimate 3000 UHPLC system coupled to a hybrid quadrupole-Orbitrap™ mass spectrometer Q-Exactive Plus (Thermo Fisher Scientific, San Josè, USA). For the peptide chromatographic separation, a reversed phase Aeris™ 3.6 μm PEPTIDE XB-C18 column, 2.1 x 150 mm (Phenomenex, Torrance, California, USA) was used, at a flow rate of 250 μL/min. The elution gradient was as follows: solvent A 0.1 % formic acid in water (H_2O), solvent B 0.1 % formic acid in acetonitrile (ACN), multistep-gradient: 0–40 min linear 5–40 % B, 41 min step change to 50 % B, 41–51 min isocratic at 50 % B, 52 min step change to 90 % B, 52–67 min isocratic at 90 %, 68 min step change to 5 % B, 68–85 min isocratic at 5 % B. The column compartment was hold at 25 °C. 20 μL of each sample were injected for analysis.

Untargeted high resolution MS/MS analysis was performed by Full-MS/dd-MS^2 analisis mode, taking into account only positive ions, set up as follows (Pilolli et al., 2021): Full-MS: microscan 1, resolution 70 k, AGC target 1e6, maximum injection time 30 ms, scan range 200–2000 *m/z*; dd-MS^2 microscan 1, resolution 17.5 k, AGC target 1e5, maximum injection time 60 ms, loop count 5, isolation window 2.0 *m/z*, stepped collision energy 27, 30, minimum AGC target 5e2, charge exclusion unassigned 1, 4–8, >8, peptide match preferred, exclude isotopes on, dynamic exclusion 15 s. For the ionization of the compounds, the source parameters (HESI) used were the following: sheat gas flow rate 25, auxiliary gas flow rate 15, spray voltage 3.4 kV, Capillary temperature 320°C and S-lens RF level 55.

Raw data were processed by Proteome Discoverer™ version 3.0.1.27 (Thermo-Fisher-Scientific, Bremen, Germany), which is based on SequestHT algorithms for protein/peptide identification. The database used was downloaded from Uniprot (The UniProt Consortium (2023)) on 2 February 2023, including taxonomy *Fabales* ID 72025 and containing about 1143624 accessions. The general parameters set in the workflow for the identification of the peptides were the following: trypsin as cleavage enzyme, mass tolerance on the precursor and fragment ions 10 ppm and 0.05 Da, respectively, peptide length 6–144 amino acids (AA), dynamic modifications: methionine-oxidation, glutamine/asparagine-deamidation, N-terminal glutamine cyclization to pyroglutamate, N-terminal protein acetylation, and static modifications: cysteine-carbamidomethylation. Finally, to display only the proteins identified with greater reliability, the list of hits obtained was further filtered by applying the following criteria: at least 2 peptides for protein, at least 1 unique peptide for protein, score Sequest HT > 0, high confidence level for peptide identification (False Discovery Rate ≤ 1 %).

3. Results and discussion

3.1. Gel-based protein profiling of lentils varieties

General proteomic profile was characterized by 1-D electrophoresis. A comprehensive protein extraction was carried out under optimized conditions described in the paragraph 2.3. Briefly, a strong denaturing and reducing buffered solution was prepared and used for the extraction of each individual ground lentil sample. Total protein content was quantified in each extract using a commercial kit by two analytical replicates and two technical replicates (see Figure S1 of the supporting material). Notably, the average protein concentrations of the Eston, Laird lentils and the Crimson lentil extracts were found equal to 19.9 ±0.6 mg/mL, 19.9±0.8 mg/mL, 21.0±1.4 mg/mL, respectively, with no statistically significant difference. Noteworthy, for the Black lentil extract a higher protein content (25.8±1.1 mg/mL) was assayed resulting in an apparent higher yield of extraction, 105 % against approximately 80 % calculated for the other three varieties. In absence of reasonable explanations for such higher extraction yield involving only the Black lentil samples, and considering the strong dark colour obtained only for this protein extract (see Figure S1), we deemed that such result might be an artefact of the colorimetric assay due to intrinsic absorbance of the protein sample at the tested wavelength.

The protein profile of the lentil samples and potential differences across the varieties were first investigated by monodimensional electrophoresis (SDS-PAGE). Fig. 1 shows the electrophoretic profiles of the four extracts loaded with different absolute amounts of protein (10 and 20 μg of proteins for each lentil samples). As known from the literature, the bands at MW ~85 kDa was characteristic of the convicilin subunits, while the bands at MW ~50 kDa may correspond to vicilin subunits, which belong to a 7 S trimeric protein, a globulin, found in many pulses. The bands at 37 kDa and 20–25 kDa corresponded to legumin subunits, an 11 S globulin, which is a hexameric protein formed by ~60 kDa units, which consist in two subunits, one acid (~40 kDa) and one basic (~20 kDa) linked by disulfide bonding (Alonso-Miravalles et al., 2019).

A homogeneous profile was displayed for all samples, with only few differences observed which were marked with arrows in Fig. 1. As for bands ranging between 90 and 100 kDa (likely attributable to convicillin proteins), two clear lines were detected for Crimson and Black lentils against a single band detected for Eston and Laird lentils. All the samples exhibited a strong band at about 90 kDa, whereas a weak band at 95 kDa (marked with arrows in lanes 2, 3 and 8) was present only in the Crimson and Black lentil samples. Other differences could be observed in the MW range typically attributed to vicilin proteins, namely 40 and 60 kDa, by comparing the Eston lentil extract to the other sample analyzed. In fact, in all cases, a faint band at 60 kDa was detectable, and two more bands were highlighted at MW approximately 43 and 47 kDa, whereas the Eston lentil sample only reported a single band at 45 kDa (marked with arrows in lanes 4 and 5). Finally, some differences should be featured also in the low MW region (20–25 kDa) characteristic of the basic subunits of legumes. All samples exhibited two bands in this region with the strongest signal detectable around 18 kDa except for the black lentil (marked with an arrow in lane 7) which showed a more intense band at 20 kDa and a less intense one at 18 kDa.

The SDS-PAGE analysis of Crimson, Eston, Laird and Black lentil extracts confirmed that most of the proteomic profile was conserved across the investigated varieties but a differential protein expression was demonstrated for a few bands putatively ascribed to convicillin, vicilin and legumin subunits.

3.2. Discovery proteomics based on high resolution mass spectrometry for protein profile identification

Protein identification was carried out by LC-MS analysis using a

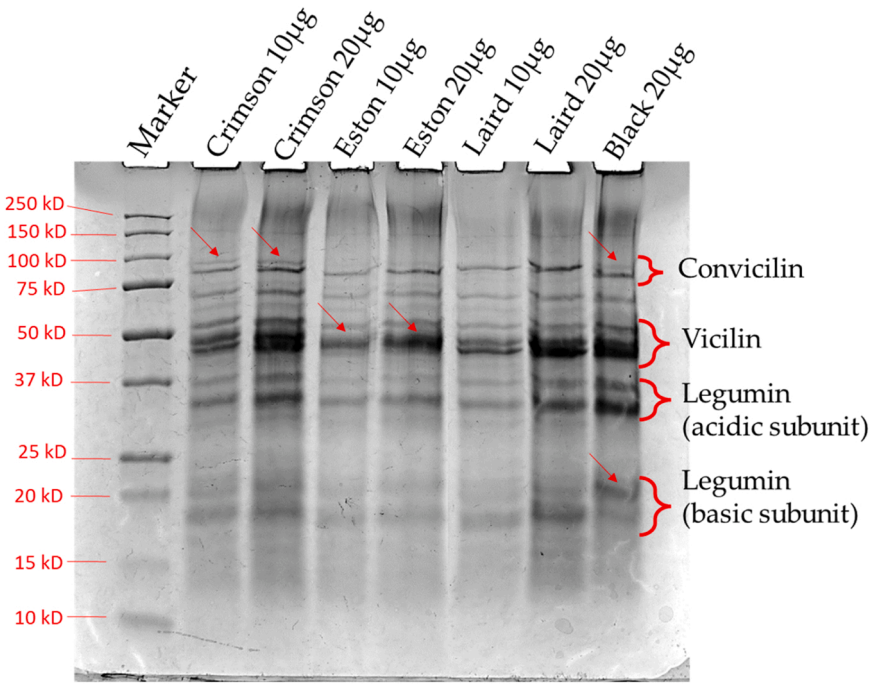

Fig. 1. Typical SDS-PAGE protein profile of four lentil varieties: Crimson, Eston, Laird, Black. Different lanes refer to different protein contents loaded: 10 and 20 µg.

typical *bottom-up* approach. The HR-MS/MS analysis was carried out in Full MS-data dependent acquisition mode, and the fragmentation spectra were processed via *Proteome Discoverer*™, a commercial software for protein identification via Sequest HT algorithm against a customized database of *Fabales* taxonomy. The software output was filtered to constrain the protein list to the most reliable identifications (see Section 2.5 for details). A total of 365 proteins were identified, most with a molecular weight less than 120 kDa and with 1–9 unique peptides identified. Most of the identified peptides (95,4 %) ranged between 400 and 1200 *m/z* and they were identified with a very high accuracy, indeed 97 % of the sequences resulted in a maximum shift of ±5 ppm compared to the theoretical value.

A label-free quantitation of the identified proteins was carried out via software based on summed abundances of unique and razor peptides. The intensities of precursor ions were used for automatic calculations and the abundances were normalized over the total peptide amounts for best comparability between samples. Such abundances were expressed as arbitrary units and were scientifically valuable for differential analysis but no direct conversion into absolute protein content can be carried out because the experimental design was not aimed at absolute protein quantitation. A box plot of the total protein abundance for each sample was built, as standardized approach to display the dataset based on a five-number summary: minimum, maximum, median, first and third quartiles (see Fig. 2). The total protein abundance was not different among the four varieties; this experimental evidence confirmed our previous hypothesis about the potential artefact in the black lentils absolute concentration estimated by the colorimetric assay. Noteworthy, heat-maps created over the full list of identified accessions disclosed substantial differences on individual proteins/protein groups (see Figure S2 of the supporting material).

In order to characterize such differential expression across varieties, proteins list was grouped based on the protein families (Pfam). Fig. 3 displayed the most abundant protein families (experimental relative abundance $\geq$ 0.5 %) highlighted for further discussion: cupins (Pf00190), late embryogenesis abundant (LEA) proteins (Pf02987, Pf03760, PF03168, PF03242, Pf04927), lypoxigenases (Pf00305, Pf01477), lectins (Pf00139), dehydrogenases (PF13561, PF13714, Pf00107, Pf08240, Pf00171, Pf00180, Pf00389, Pf01842, Pf02826,

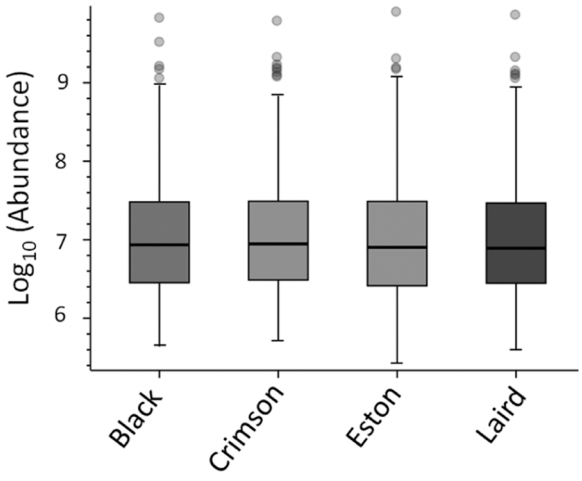

Fig. 2. Box plot of total protein abundance of four lentil varieties: Crimson, Eston, Laird, Black.

Pf19304, PF00984, PF03720, PF03721, Pf00044, Pf02800, Pf03358), redoxins (Pf00578, Pf10417, Pf08534, Pf00085, Pf13848, PF00462), albumins I (Pf08027, Pf16720), albumins 2 (Pf00045), elongation factors (Pf00009, Pf00679, Pf03144, Pf03764, Pf14492, Pf03143, Pf00647, Pf00043, Pf02798), peptidase (Pf00082, Pf12580, Pf21316, Pf21223, PF00883, PF02789, Pf14541, Pf14543, PF01432, PF19310) and heat shock proteins (Pf00183, Pf02518, Pf13589, Pf17830, Pf13414, Pf13432, Pf13181, Pf00011, Pf00012, Pf00931, Pf01582)

The cupins family clearly represented the major component of the protein profile for all lentil samples covering always more than the 73 % of the total protein content. In particular, the abundances varied from the 73.8±1.6 % of the Crimson lentil samples to the 78.6±1.5 % of the Black lentil samples, with quite similar and intermediate values for Eston and Laird lentil samples. The cupins consist in a diverse superfamily of proteins sharing a conserved β-barrel domain. This superfamily includes enzymes as well as non-enzymatic seed storage proteins, whose principal function appears to be nutrient reservoir as major nitrogen

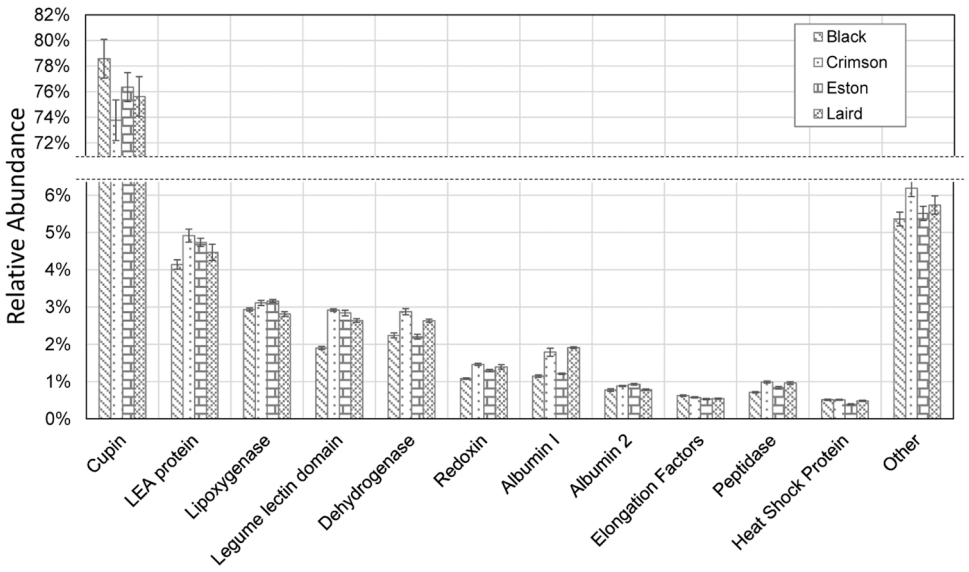

Fig. 3. Comparison across lentils varieties of the percent relative abundances experimentally calculated for specific protein families/groups.

source for the developing plant. Here, cupins related accessions were assigned to 7S-vicilin (55–64 %), 11S-legumin (24–38 %), 7S-convicilin (7–9 %), 7S-minor component (0.05–0.07 %) and 11S-conglutinin (0.004–0.007 %). The highest content of vicilins was reported for Eston and Laird lentil samples, 49.9±0.7 % and 49.4±1.0 % respectively, of the total protein, against the lowest content of Black and Crimson lentil samples equals to 43.7±0.7 % and 44.9±0.8 %, respectively, of the total protein. Differently, the legumins resulted overexpressed in the Black lentil samples 30.2±0.7 % of the total protein, against the 21.4±0.4 %, 21.8±0.4 % and 23.7±0.7 % of the Eston, Laird and Crimson lentil proteins respectively. The 7 S/11 S ratio is an important characteristic to describe seeds nutritional quality (Bourgeois et al., 2009, Khazaei et al., 2019) and was reported to be very high in lentil, close to three (Scippa et al., 2010), compared to other legumes. According to our data the 7 S/11 S ratios are very different among the four varieties, namely 1.65±0.07 for the Black lentils, 2.18±0.10 for Crimson lentils, 2.64±0.08 for Eston lentils, and 2.53±0.10 for Laird lentils samples. A higher nutritional quality is therefore envisaged for Eston and Laird lentils.

The LEA protein family varied from the 4.14±0.13 % of the Black lentil samples to the 4.92±0,17 % recorded for Crimson lentil samples. LEA proteins in plants have been found to accumulate to high levels during the last stage of seed formation when a natural desiccation of the seed tissues takes place and during periods of water deficit in vegetative organs. Different types of LEA proteins can be expressed at different stages of late embryogenesis in higher plant seed embryos and under conditions of dehydration stress, all classified into several subgroups and according to Bray and Dure (Hundertmark and Hincha, 2008). In all lentil samples, LEA protein entries belonging to group 1 (30–35 %), group 2 (14–16 %), group 3 (3–5 %), group 4 (78–80 %) and SMP-seed maturation protein- (20–21 %) have been identified and categorized among others. The LEA protein group 1 involved in the biological process of embryo development ending in seed dormancy and may play roles in the adaptive process to water deficit in higher plants. This group resulted slightly overexpressed in Crimson lentil sample (1.03±0.04 %) compared to the other three varieties. The LEA protein group 2, also referred to as dehydrins, are directly involved in the biological processes responding to water-stress (presence, absence or variation) and contribute to freezing stress tolerance in plants, likely due to their protective effect on membranes (Puhakainen et al., 2004). Such group was found equally expressed in Crimson, Eston and Laird lentil samples (0.45 % of the total proteins on average) with a lower expression for Black Lentil samples (0.36±0.01 %)). A similar trend was observed for the LEA protein group 3 and group 4. The function of these LEA proteins is not clear but may be involved in the re-establishment of desiccation tolerance in seeds. Comparable results were reported for the other LEA protein groups.

The lipoxygenase family is a class of iron-containing dioxygenases directly involved in biosynthesis and metabolism of fatty acid, lipid and oxylipin biosynthesis. They are common in plants where they may be involved in various aspects of plant physiology such as growth and development, pest resistance, and senescence or responses to wounding. The lipoxygenase relative abundance varied from the 2.81±0.06 % of the Laird lentils sample to the 3.16±0.04 % of the Eston lentil sample.

The lectins family showed a wider variability ranging from 1.91 ±0.04 % calculated for Black lentil samples to 2.92±0.04 % and 2.84 ±0.07 % calculated for Crimson and Eston lentil samples, respectively. Noteworthy, lectins serve as defence mechanism for plants being involved into response to biotic stresses: as carbohydrate-binding proteins they cause agglutination of targeted cells or precipitation of glycoconjugates and polysaccharides, all processes associated with the plant response to pathogens, but they are also involved in seed germination/conservation (Scippa et al., 2010). Lectins can pose challenges for human diet as antinutritional factor (ANF). Indeed, active lectins might interact with minerals (calcium, iron, phosphorus, zinc) and interfere with their absorption and utilization in the body affecting the overall health. (Joehnke et al., 2021); however, such effect can be modulated by food process. In particular, thermal processing can reduce/eliminate heat-labile ANFs, such as lectins and protease inhibitors (Zhou et al. (2023)).

The dehydrogenase, redoxin, albumin I and peptidase protein groups showed a very similar trend with a higher abundance reported for Crimson and Laird lentil samples than Eston and Black lentil samples. The dehydrogenase group included alcohol dehydrogenases (Pf00107, Pf08240), aldehyde/glyceraldehyde dehydrogenase (Pf00171, Pf02800) and malate dehydrogenase (Pf00056), among others; all of them share an oxidoreductase function into different metabolic processes. The redoxins, also serving as oxidoreductase enzymes, were classified as thioredoxins (small enzymes involved in protein disulfide redox reaction). This protein group has great relevance for the antioxidant biological function as member of the peroxiredoxin superfamily which protects cells against membrane oxidation through glutathione (GSH)-dependent reduction of phospholipid hydroperoxides to the corresponding alcohols (Choi et al., 1998). The albumin I protein

detected here as a fragment (insecticidal lentil peptide), belongs to the seed storage proteins and, as such, boasts nutrient reservoir activity; moreover, it also provides a toxic activity as hormone-like peptide, that stimulates kinase activity upon binding specific membrane receptor (Yamazaki et al., 2003). Among the peptidase family, the aspartic- type endopeptidase was the main component, belonging to the peptidase A1 (pepsin) family. All the identified accessions were variously involved in the protein metabolic processes.

The albumin 2 group, here populated by three accessions, are generally classified as storage proteins but can also play important roles in germination and stress defense functions. The elongation factors identified are all involved in protein synthesis. The heat shock protein (HSP) family includes subgroups of protein synthesized by the plant as response to thermal or other environmental stresses. Such response is the most highly conserved genetic system known, existing in every organism, from archaebacteria to eubacteria, from plants to animals (Lindquist and Craig, 1988). Proteins encoded by the hsp70 and hsp90 gene families are synthesized in response to elevated temperatures, and in these lentil samples represented about the 47–58 % and 38–43 %, respectively, of the total HSP. The third class of HSP (about 4–10 %) with an average molecular weight of 20 Kd, known as the hsp20 proteins, likely acts as chaperones protecting other proteins against heat-induced denaturation and aggregation.

Among the minor protein components summed up in Fig. 3 into the 'other' category it deserved to be mentioned ferritin and Bowman-Birk protease inhibitors (BBIs) both involved in the defense mechanism: the ferritin was proved to protect against oxidative damage by limiting levels of reactive oxygen species and facilitating seed germination and the BBIs activate defense mechanism of the seed against insect midgut proteases. Since serine proteases play a pivotal role in the development and pathogenesis of cancer, Dengue fever, inflammatory and allergic disorders, some evidence of involvement of such BBIs in the prevention of these diseases have been previously reported. As such, BBIs boasted a positive contribution to the nutritional value of lentil seeds (Qi et al., 2005) The highest abundance for BBIs was disclosed for Laird lentils samples (0.175±0.004 %) and the lowest (0.132±0.009 %) for Eston lentils samples, with intermediate value for Crimson and Black varieties.

3.2.1. *In-silico allergenicity assessment*

Moving forward the general overview provided in the previous section, an in-depth analysis was carried out on allergenicity assessment of the detected proteins and on their differential expression among varieties. In the last years lentils gained relevance for growing consumption and sensitization in vulnerable population, up to requiring the inclusion in a "watch list" of concerning foods, prioritized for the allergenicity potential (World Health Organization, 2022). Allergic reactions due to lentil consumption have been reported in many countries (Sackesen et al., 2020), especially in Mediterranean and Asian countries (Crespo et al., 1995). However, few information is available on identification and profiling of allergens mainly due to the incomplete sequencing of lentils proteome.

The untargeted LC-HR-MS/MS analysis combined with bioinformatics tools provides high accuracy in the identification/quantification of allergens as well as in the prediction of allergenicity potential (Halima et al., 2022). Based on this, here we applied an in-silico analysis of the identified protein lists to predict immunogenic potential of each variety by alignment and comparison of the identified sequences with known allergens. In particular, the web server *AllerCatPro* 2.0 was used to predict the similarity between input proteins using both their amino acid sequences and the predicted 3D structures towards the most comprehensive datasets of reliable proteins associated with allergenicity (Maurer-Stroh et al., 2019; Nguyen et al., 2022). The web server applies the guidelines provided by FAO/WHO and European Food Safety Authority (EFSA) to assess whether a query protein is potentially allergenic, namely it seeks for sequence identity higher than 35 % on a frame of 80 amino acids and for exact matches of sequences from 6 to 8 amino acids with an allergenic protein. Furthermore, query proteins are also browsed for identification of aligned 3D surface residues greater than 93 %. The overall accuracy of *AllerCatPro* is 84 % compared with other current methods which range from 51 % to 73 %. (Maurer-Stroh et al., 2019).

A curated list of lentil proteins with high and low immunogenic potential was obtained thanks to such *in-silico* homology search against known allergenic sequences. A total of 132 out of 365 identified lentil proteins were categorized with strong evidence of allergenicity and a total of 69 out of 365 proteins were categorized with weak evidence of allergenicity (see Table S1 of the supplementary material). Abundances of proteins accessions with 'strong', 'weak' and no evidence of allergenicity were summed-up into specific protein rank. The proteins with strong evidence represented the most abundant category in all varieties accounted for about 67–71 % of the total protein abundance. The proteins with weak evidence of allergenicity accounted for about 19–23 % of the total protein abundance, the remainder presented no evidence of allergenic potential according to the current knowledge. The most abundant identified proteins belonged to the family of storage proteins, whose strong allergenic potential is well established. Such results confirmed the importance to include lentils in the watch list for new prioritized allergenic foods.

Multiple comparisons of variety-correlated mean values were performed by a Tukey's post hoc ANOVA test. Mean that were not significantly different were marked with equal labels (a, b, c) in relevant plots (see Fig. 4). The results of the Tukey's test showed for proteins with strong evidence, that the Eston lentils have the highest abundance, significantly different from the abundance calculated in the Crimson and Black samples and equal only to the Laird sample. As for proteins with weak evidence, the Black lentils reported the highest abundance significantly different from all other samples.

Hierarchical Cluster Analysis (HCA) of abundance data collected for strong and weak allergens was carried to visualize similarity and grouping of specific protein accessions and varieties. Abundances were scaled before clustering to avoid overweight of the most abundant proteins on the overall analysis. The heat-maps were used for visual clarity with a color-coded HCA (see Fig. 5) and dendrograms on the left and on the top of the heat maps, displayed the cluster nodes of HCA. The size of the dendrograms disclosed the similarity between two cluster nodes, namely closer nodes represent more similar protein accessions or sample groups (i.e. lentils varieties). Specifically, the dendrograms above the heat map, presented the similarity among lentils varieties suggesting three clusters for both strong allergens (Fig. 5, left panel) and weak allergens (Fig. 5, right panel), even if distances between cluster were quite small in agreement with the results of Tukey's test.

As for strong allergens (Fig. 5, left panel), Laird lentils group was close to Crimson lentils group in cluster 1, then this cluster was close to Black lentils group (cluster 2), and finally cluster 2 was linked to Eston lentils group. As for weak allergens (Fig. 5, right panel), Laird lentils group was close to Crimson lentils group in cluster 1, Black lentils group was close to Eston lentils group (cluster 2), and finally cluster 1 and 2 were linked together in cluster 3. On the right side of each heat-map, the top50 most abundant accessions were highlighted with a red square, and a full list of them was reported in Table 1. Main contributors to this list belonged to cupin family, classified as strong or weak evidence of allergenicity depending on the specific accessions (see Table 1). Focusing on proteins with strong allergenicity potential, 7S-vicilins and 11 S legumins resulted the most abundant: vicilins resulted overexpressed in Eston and Laird lentil samples, about 43–44 % of the total proteins against about 38 % of the total protein reported for Black and Crimson samples, whereas the 11 S legumin resulted overexpressed in Black lentil samples, about 14 % against 10–12 % of the other three varieties. The vicilins accessions Q84UI1, A0A9D4WHZ2, P13918, Q84UI0, Q702P0, A0A9D4X1R0, A0A1S2XQ88, D3VND7, Q41677, Q2HW16, A0A2Z6P9Y2, P08438 featured both 100 % identity with 3D epitopes and 88.8–100 % identity over a linear 80 aa window with Pis s

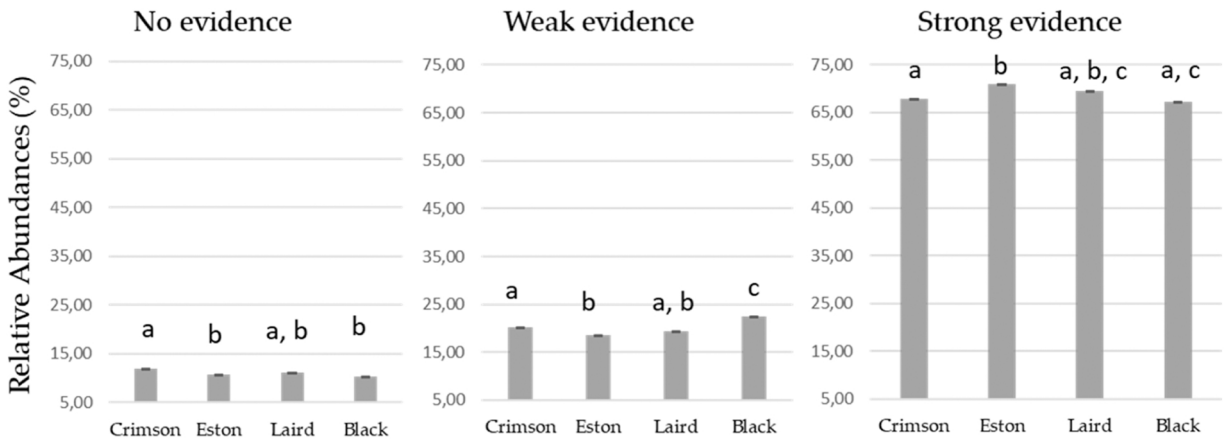

Fig. 4. Relative abundance of proteins with no-evidence (left panel), weak evidence (middle panel) and strong evidence (right panel) of allergenicity potential proteins identified in Crimson, Eston, Laird, Black lentils using AllerCatPro. The results of a Tukey statistical test for multiple mean comparisons (n = 4) are also reported as labels (a, b, c); equal labels highlight mean values that are not significantly different.

1 and Len c 1 allergens, thus confirming their strong allergenic potential. Similarly, the legumin accessions Q41676, A0A2K3N205, Q41702, Q03971, A0A396H636 featured both a 94.7–100 % identity with 3D epitopes and 83.8–91.2 % identity over a linear 80 aa window with Cic a 6, Gly m 6, Pis s 1 allergens (see Table 1 for details). In addition, the entries Q702P0 (vicilin) and A0A396H636 (legumin) also presented the 100 % identity with gluten-like Q-repeats.

All the other identified protein accessions with strong allergenicity potential presented lower abundances ranging between 0.2 % and 6 % of the total proteins (see Table 1 for details), including 7 S convicilins (Q9M3X8, B0BCK5), albumin-2 (A0A9D5BII4), peroxiredoxin (A0A9D4ZWX1), oleosin (A0A9D4Y0S6), non-specific lipid-transfer protein 2 (A0AT29), LEA proteins group 4 (A0A9D4X2N1, A0A9D5A5D8, A0A9D4X1H4), lectins (E3UFD6, P04122), and dehydrogenases (A0A9D4XPH4, A0A2K3MMN2). All of them presented a strong evidence of allergenicity meeting at least one identity criterium (3D epitope and/or linear 80 aa window), which correlated them specifically to the following known allergens: to Pis s 2, Pis s albumin, Hor v 32, Ara h 10, Len c 3, Glu m 7/ Cic a 1,Len c Agglutinin/Lat oc Agglutinin, and an unclassified allergen from Sesamum indicum, respectively.

Protein accessions with weak evidence of allergenicity were also listed in Table 1 together with all information collected in terms of similarity with known allergenic proteins. Again, the cupin family represented the prevalent component, 11 S legumins being the most abundant proteins (9–14 %). Such protein group including Q41703, A0A9D4XTC5, A0A9D4W4Q7, P16078, P05190 accessions, resulted significantly overexpressed in Black lentil samples (14.0±0.3 %) compared to the other three varieties (9.1–9.8 %). All of them displayed a quite high similarity in terms of linear 80 aa window (73.8–78.8 %) with Gly m 6, but a 3D epitope identity lower than the prescribed 93 % cut-off. Second most abundant protein group reported as weak allergenicity evidence was the 7 S vicilins (O49927, A0A2K3L0Z1, A0A9D5BCZ9) with a certain variability among varieties comprised between the highest value recorded for Crimson lentils (5.92±0.07 %) and the lowest value recorded for Black lentils (4.65±0.09 %). In this case, the allergenicity potential was accounted for similarity with Jug r 6, Cor a 11, and Coc n 1, respectively.

3.3. *Protein profile analysis for perspective varietal discrimination*

The detailed proteomic investigation carried out proved that despite the overall conserved profile characterizing the lentil samples, peculiar differences which are statistically significant on specific protein accessions can be spotted among the four varieties. This experimental evidence suggested the perspective to exploit such information to convey an analytical approach for lentil sample discrimination. Seeking this purpose, a differential analysis supported by volcano plots and unsupervised multivariate statistical analysis was carried out to select a subset of variables fitting the scope.

In Figure S3 the volcano plots for all combinations of paired sample groups were reported. Such plots are used commonly in omics experiments to quickly identify changes between two conditions/groups in large data sets. The statistically significant changes can be clearly visualised by applying meaningful cut-offs on ratios and p-values. In this experiment, an absolute value for the fold change higher than 1 with a p-value $\leq$0.05 were applied for the selection of potential discriminant variables. A total of 35 entries characterized by 211 peptides were highlighted as up or down regulated in at least one ratio (see Table S2). In terms of relative abundance, such potential discriminant proteins represented a minor component of the protein profile in all lentil samples: about 3 % of the total proteins identified in the Crimson and Eston lentil samples and about 6 % of the total proteins identified in the Laird and Black lentil samples.

A Principal Component Analysis (PCA) was applied to visualize the multidimensional data and potential correlation between characterizing proteins. The first five PC explained 97 % of the total variance (PC1 48.7 %, PC2 32.5 %, PC3 15.9 %, PC4 1.2 %, PC5 0.6 %) and interestingly the score plot (Fig. 6a) disclosed a clear grouping of the lentil samples according to the variety factor. Fig. 6b displayed the loadings plot with the relationship between the selected variables. Noteworthy, all the selected protein accessions resulted indicative for at least one PC (see Table S2 of the Supporting Information) with coefficient values for the loading plot higher than 0.1 or lower than −0.1, being as such discriminative among varieties. Both Black and Eston lentils showed positive scores on PC1 (averaging around 1.7 and 5.5, respectively) against negative scores on PC1 for Crimson and Laird lentils (averaging around −1.9 and −5.4, respectively). According to this discrimination over PC1, 9 accessions presenting indicative positive coefficients on PC1 correlated to characterize Black and Eston lentils against 15 accessions with negative coefficients on PC1 correlating to characterize the Crimson and Laird lentils. Differently, PC2 contributed mainly to discriminate the Black lentil samples (scores averaging around −5.3), from the Crimson, Eston, and Laird lentils (scores averaging around 2.5, 2.7, and 0.2, respectively) with only 6 entries indicative for Black lentils and 18 variables correlating to describe the other three varieties. Full details about loading coefficient can be found in Table S2 of the supporting information.

Even if in its infancy, such preliminary analysis confirmed the potential to exploit specific protein accessions identified and quantified by one-shot discovery proteomics to discriminate commercial lentils

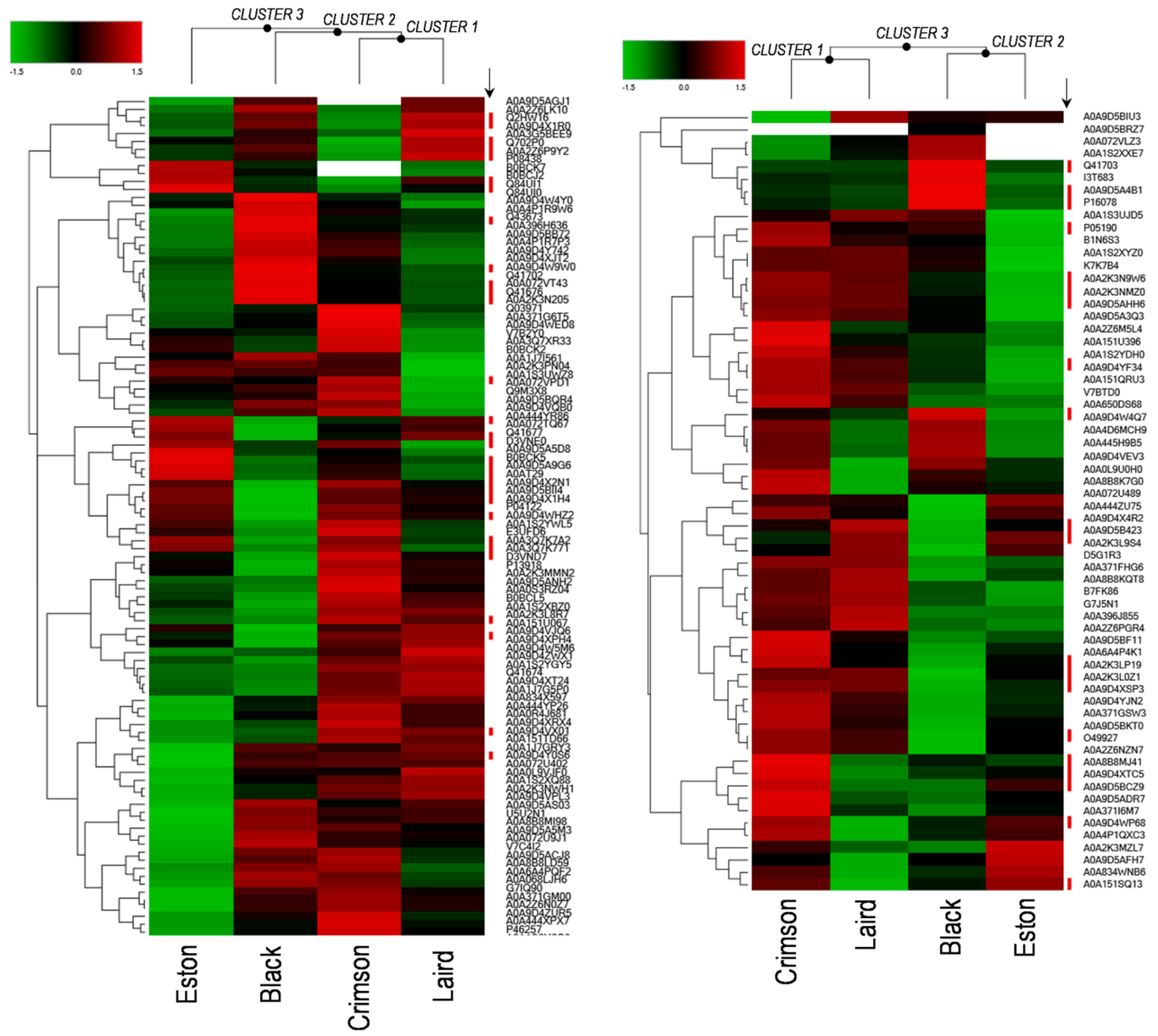

Fig. 5. Hierarchical Cluster Analysis (HCA) and heat maps of protein accessions with strong evidence of allergenicity (left panel) and weak evidence of allergenicity (right panel). HCA was carried out with Euclidean distance function and complete linkage method for abundances values which were scaled before clustering. The red squares pointed with the arrow on the right side of each heatmap displayed the top50 most abundant accessions with any predicted allergenicity potential.

varieties. Although further developments of this perspective are required on a larger number of samples to confirm the observed trends, the significance of these results is very high compared to previous reported investigation due to the viability of the one-shot analytical approach applied. Indeed, other authors in the past have attempted to exploit protein abundance/expression for varietal discrimination. Scippa et al. in 2010 investigated the potential of proteomics as a tool in phylogenetic studies, testing the ability to identify specific markers of different plant landraces on lentil populations belonging to a local ecotype (Capracotta) and by analysing five commercial varieties. The multivariate statistical analyses carried out on 122 variably expressed protein spots excised by 2D-electrophoretic maps showed that 24 entries were essential for population discrimination, thus suggesting them as landrace markers (Scippa et al. 2010). More recently, Halima et al. in 2022 have reviewed the feasibility of omics platforms for lentil allergens profiling and quantification also proposing strategies that might be used for profiling and assays development for lentil allergens facilitating identification of the low allergen-containing lentil cultivars. To the best of our knowledge, in this investigation it was proved for the first time the feasibility of a one-shot discovery proteomic analysis to provide a complete proteomic profiling of four commercial lentil varieties, their allergenicity assessment by sequence alignment and the identification of potential varietal markers for perspective application in authenticity studies.

Noteworthy, the proteomic profile built in this investigation makes up a snapshot of dry mature seeds, and it is not strictly and directly representative of consumed food. Typically, lentils are either used cooked as whole or dehulled seeds for direct consumption or processed into derived formulations, such as flour, protein isolate, starch, and fibre, which can be used as ingredients in diverse food applications. Different preprocessing (e.g., dehulling and milling) and processing (e.g., cooking, fermentation, soaking, and germination) methods have been featured to modify the nutritional profile of lentil samples, also minimizing or inactivating the amount of ANFs (Dhull et al. (2023)) and

Table 1
List of the top 50 most abundant proteins identified and quantified and prioritized for a predicted strong (S) or weak (A) allergenic potential according to similarity check with known protein allergens.

Accession	Description	Pfam IDs	Predicted allergenicity by similarity				Relative Protein abundance			
			S/W	Protein allergen	% identity					
					linear 80 aa window	3D epitope	Black	Crimson	Eston	Laird
Q84UI1	Allergen Len c 1.0101 (Fragment)	Pf00190–7 S vicilin	S	Len c 1	100.0	100.0	22,13 ± 0,19 %	20,6 ± 0,2 %	25,7 ± 0,3 %	24,3 ± 0,4 %
A0A9D4WHZ2	Cupin type-1 domain-containing protein	Pf00190–7 S vicilin	S	Pis s 1	96.2	100.0	3,77 ± 0,06 %	4,87 ± 0,11 %	4,88 ± 0,08 %	4,63 ± 0,10 %
P13918	Vicilin	Pf00190–7 S vicilin	S	Pis s 1	100.0	100.0	2,64 ± 0,03 %	5,13 ± 0,10 %	4,86 ± 0,07 %	2,88 ± 0,02 %
Q84UI0	Allergen Len c 1.0102 (Fragment)	Pf00190–7 S vicilin	S	Len c 1	100.0	100.0	2,22 ± 0,14 %	1933 ± 0006 %	3,03 ± 0,05 %	2,32 ± 0,02 %
Q702P0	Vicilin (Fragment)	Pf00190–7 S vicilin	S	Pis s 1	100.0	100.0	1,79 ± 0,04 %	1,36 ± 0,08 %	1,67 ± 0,05 %	2,01 ± 0,05 %
A0A9D4X1R0	Cupin type-1 domain-containing protein	Pf00190–7 S vicilin	S	Pis s 1	95.0	100.0	1,76 ± 0,04 %	0,0684 ± 0,0014 %	0415 ± 0013 %	2,10 ± 0,10 %
A0A1S2XQ88	Vicilin-like	Pf00190–7 S vicilin	S	Pis s 1	90.0	100.0	1,58 ± 0,06 %	1,70 ± 0,03 %	0523 ± 0008 %	1,78 ± 0,05 %
D3VND7	Vicilin 47k	Pf00190–7 S vicilin	S	Pis s 1	100.0	100.0	1,05 ± 0,03 %	1,63 ± 0,04 %	1,64 ± 0,04 %	1,26 ± 0,03 %
Q41677	Vicilin	Pf00190–7 S vicilin	S	Pis s 1	92.5	100.0	0,62 ± 0,03 %	0,72 ± 0,04 %	0,84 ± 0,04 %	0,79 ± 0,02 %
Q2HW16	Cupin, RmlC-type	Pf00190–7 S vicilin	S	Pis s 1	91.2	100.0	0,53 ± 0,02 %	0236 ± 0011 %	0280 ± 0010 %	0,64 ± 0,04 %
A0A2Z6P9Y2	Cupin type-1 domain-containing protein	Pf00190–7 S vicilin	S	Pis s 1	88.8	100.0	0,27 ± 0,02 %	0045 ± 0011 %	0139 ± 0005 %	0333 ± 0013 %
P08438	Vicilin	Pf00190–7 S vicilin	S	Pis s 1	91.2	100.0	0213 ± 0002 %	0123 ± 0005 %	0165 ± 0003 %	0274 ± 0004 %
Q9M3X8	Convicilin (Fragment)	Pf00190–7 S convicilin	S	Pis s 2	95.0	100.0	4,91 ± 0,04 %	5,55 ± 0,06 %	5,02 ± 0,04 %	4,34 ± 0,03 %
B0BCK5	Convicilin (Fragment)	Pf00190–7 S convicilin	S	Pis s 2	98.8	100.0	0249 ± 0008 %	0,29 ± 0,02 %	0292 ± 0012 %	0228 ± 0012 %
Q41676	Legumin A	Pf00190–11 S legumin	S	Cic a 6	86.2	94.7	5,28 ± 0,07 %	4,51 ± 0,11 %	3,95 ± 0,05 %	4,08 ± 0,10 %
A0A2K3N205	Legumin A	Pf00190–11 S legumin	S	Cic a 6	87.5	100.0	4,85 ± 0,14 %	4,12 ± 0,14 %	3,67 ± 0,05 %	3,73 ± 0,06 %
Q41702	Legumin A	Pf00190–11 S legumin	S	Cic a 6	85.0	100.0	1,96 ± 0,02 %	1,62 ± 0,06 %	1,46 ± 0,05 %	1,48 ± 0,03 %
Q03971	N-terminal incomplete legumin A1 pre-pro-polypeptide (Fragment)	Pf00190–11 S legumin	S	Cic a 6	85.0	100.0	1,42 ± 0,07 %	1,22 ± 0,02 %	1,08 ± 0,03 %	1,10 ± 0,05 %
A0A396H636	Putative 11-S seed storage protein, plant	Pf00190–11 S legumin	S	Gly m 6	83.8	100.0	0,58 ± 0,04 %	0453 ± 0012 %	0,38 ± 0,03 %	0433 ± 0013 %
A0A9D5BII4	Albumin-2	PF00045	S	Pis s Albumin	100.0	100.0	0,68 ± 0,03 %	0772 ± 0007 %	0,83 ± 0,02 %	0690 ± 0013 %
A0A9D4ZWX1	Peroxiredoxin	PF10417, PF00578	S	Hor v 32	75.0	93.3	0503 ± 0007 %	0765 ± 0011 %	0633 ± 0005 %	0,78 ± 0,03 %
A0A9D4Y0S6	Oleosin	PF01277	S	Ara h 10	83.8	-	0222 ± 0010 %	0268 ± 0009 %	0214 ± 0010 %	0,26 ± 0,02 %
A0AT29	Non-specific lipid-transfer protein 2	Pf00234	S	Len c 3	100.0	100.0	0404 ± 0003 %	0458 ± 0007 %	0518 ± 0008 %	0,42 ± 0,02 %
A0A9D4X2N1	Seed biotin-containing protein sbp65	Pf02987 - LEA group 4	S	Gly m 7	68.8	-	0,81 ± 0,02 %	0,95 ± 0,02 %	1071 ± 0014 %	0,81 ± 0,03 %
A0A9D5A5D8	Uncharacterized protein	Pf02987 - LEA group 4	S	Cic a 1	67.5	-	0,39 ± 0,02 %	0454 ± 0013 %	0493 ± 0004 %	0,49 ± 0,02 %
A0A9D4X1H4	Seed biotin-containing protein SBP65	Pf02987 - LEA group 4	S	Gly m 7	72.5	-	0365 ± 0013 %	0468 ± 0011 %	0445 ± 0015 %	0,42 ± 0,02 %
E3UFD6	Lectin	Pf00139	S	Len c Agglutinin	100.0	100.0	1,49 ± 0,03 %	2,31 ± 0,02 %	2,22 ± 0,07 %	2,08 ± 0,04 %
P04122	Lectin beta-1 and beta-2 chains	Pf00139	S	Lat oc Agglutinin	100.0	100.0	0409 ± 0006 %	0,61 ± 0,02 %	0618 ± 0006 %	0,56 ± 0,02 %
A0A9D4XPH4	NADPH-dependent aldehyde reductase 1	PF13561	S	Unknown	87.5	100.0	0436 ± 0014 %	0,67 ± 0,02 %	0463 ± 0014 %	0598 ± 0011 %
A0A2K3MMN2	Glucose and ribitol dehydrogenase-like protein	Pf13561	S	Unknown	87.5	100.0	0257 ± 0009 %	0337 ± 0011 %	0289 ± 0006 %	0,2962 ± 0,0012 %
O49927	p54 protein	Pf00190–7 S vicilin	W	Jug r 6	57.5	76.9	3,08 ± 0,05 %	3,97 ± 0,03 %	3,59 ± 0,04 %	3,76 ± 0,04 %
A0A2K3L0Z1	Vicilin-like protein antimicrobial peptides 2–1-like protein	Pf00190–7 S vicilin	W	Cor a 11	67.5	84.6	0,83 ± 0,03 %	1,04 ± 0,03 %	0,96 ± 0,02 %	1,04 ± 0,02 %

(*continued on next page*)

Table 1 (*continued*)

Accession	Description	Pfam IDs	Predicted allergenicity by similarity				Relative Protein abundance			
			S/W	Protein allergen	% identity					
					linear 80 aa window	3D epitope	Black	Crimson	Eston	Laird
A0A9D5BCZ9	Cupin type-1 domain-containing protein	Pf00190–7 S vicilin	W	Coc n 1	46.2	60.0	0739 ± 0007 %	0,90 ± 0,02 %	0826 ± 0014 %	0,74 ± 0,02 %
A0A2K3L9S4	Convicilin-like protein (Fragment)	Pf00190–7 S convicilin	W	Pis s 2	80.0	88.2	0,38 ± 0,02 %	0,51 ± 0,06 %	0619 ± 0007 %	0,64 ± 0,14 %
A0A9D5A4B1	Cupin type-1 domain-containing protein	Pf00190–7 S convicilin	W	Gly m 6	78.8	80.0	0358 ± 0012 %	0179 ± 0011 %	0144 ± 0005 %	0159 ± 0004 %
A0A2K3N9W6	Convicilin (Fragment)	Pf00190–7 S convicilin	W	Cic a 1	79.7	92.9	0195 ± 0004 %	0,22 ± 0,02 %	0175 ± 0005 %	0213 ± 0002 %
Q41703	Legumin B	Pf00190–11 S legumin	W	Gly m 6	77.5	86.7	11,04 ± 0,19 %	6,93 ± 0,13 %	6,73 ± 0,06 %	6,92 ± 0,04 %
A0A9D4XTC5	Cupin type-1 domain-containing protein	Pf00190–11 S legumin	W	11 S globulin-like	73.8	88.9	1,38 ± 0,04 %	1,62 ± 0,04 %	1,42 ± 0,03 %	1,32 ± 0,03 %
A0A9D4W4Q7	Cupin type-1 domain-containing protein	Pf00190–11 S legumin	W	Gly m 6	78.8	80.0	0,81 ± 0,04 %	0,66 ± 0,02 %	0476 ± 0009 %	0567 ± 0006 %
P16078	Legumin type B (Fragment)	Pf00190–11 S legumin	W	Gly m 6	77.5	80.0	0733 ± 0012 %	0,50 ± 0,04 %	0428 ± 0012 %	0,47 ± 0,02 %
P05190	Legumin type B	Pf00190–11 S legumin	W	Gly m 6	76.2	86.7	0069 ± 0004 %	0074 ± 0009 %	0059 ± 0003 %	0,07 ± 0,02 %
A0A2K3LP19	Glucose and ribitol dehydrogenase-like protein (Fragment)	Pf13561	W	Unknown	81.2	90.0	0,65 ± 0,02 %	0,80 ± 0,02 %	0,72 ± 0,02 %	0725 ± 0008 %
A0A9D4XSP3	Peptidase A1 domain-containing protein	Pf14541, Pf14543	W	Lup an gamma-conglutin	75.0	92.0	0,65 ± 0,02 %	0,92 ± 0,03 %	0,77 ± 0,03 %	0,90 ± 0,03 %
A0A9D4WP68	Protein disulfide-isomerase	Pf00085, Pf13848	W	Alt a 4 homolog	60.0	53.8	0149 ± 0002 %	0169 ± 0004 %	0160 ± 0003 %	0135 ± 0005 %
A0A151SQ13	Protein disulfide-isomerase	Pf00085, Pf13848	W	Alt a 4 homolog	60.0	53.3	0092 ± 0002 %	0101 ± 0002 %	0102 ± 0002 %	0080 ± 0005 %
A0A2K3NMZ0	Alcohol dehydrogenase	Pf00107, Pf08240	W	Cand a 1	36.2	53.3	0334 ± 0012 %	0417 ± 0014 %	0251 ± 0008 %	0397 ± 0006 %
A0A9D5AHH6	Alcohol dehydrogenase	Pf00107, Pf08240	W	Cand a 1	36.2	53.3	0083 ± 0002 %	0103 ± 0005 %	0053 ± 0003 %	0100 ± 0003 %
A0A9D5B423	Ferritin	Pf00210	W	Pro c 21kD	66.2	84.6	0132 ± 0004 %	0187 ± 0004 %	0,18 ± 0003 %	0219 ± 0009 %
A0A9D4YF34	Nucleoside diphosphate kinase	Pf00334	W	Gad m NDKB	65.0	69.2	0126 ± 0004 %	0143 ± 0002 %	0118 ± 0004 %	0136 ± 0005 %
A0A8B8MJ41	DEAD-box ATP-dependent RNA helicase 3, chloroplastic-like	Pf00270, Pf00271, Pf00400	W	Bla g RACK1	25.0	50.0	0,0587 ± 0,0013 %	0065 ± 0002 %	0,0577 ± 0,0006 %	0057 ± 0002 %

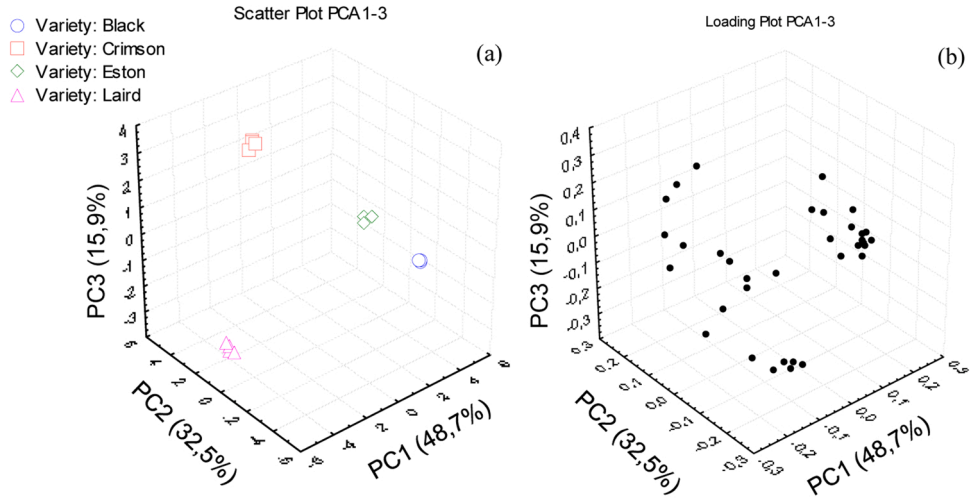

Fig. 6. Graphical visualization of principal component analysis. a) Score plots and b) Loadings plots of the four lentil varieties (Crimson, Eston, Laird, Black) using only the discriminating proteins.

promoting the use of lentil-based products without any negative effects on human health. In some instances, data reported on different processing techniques (dehulling, splitting, milling, cooking, extrusion, germination or sprouting, and fermentation) showed conflicting effects (i.e., increasing or decreasing patterns) on the composition and nutrient profile of lentils and this could be attributed to the specific lentil variety

used in the test and/or to the testing methodology itself (Dhull et al. (2023)). Thermal processing can effectively degrade heat-labile anti-nutritional factors, such as protease inhibitors and lectins (Zhou et al. (2023)), but it will not be effective on several heat-stable allergenic proteins (e.g., Len c 1, Len c 2) mainly belonging to the storage protein family, which might, then, keep its immunogenic potential also in cooked samples and still pose risk for the health of allergic consumers (Shaheen et al., 2019).

4. Conclusions

The present work provided a differential characterization of the proteomic profile of four lentils commercial varieties Crimson, Eston, Laird and Black, by one-shot discovery proteomics with particular focus on the observed expression of main proteins and their in-silico allergenicity assessment. The investigation was carried out by means of conventional electrophoresis and advanced proteomic approach based on untargeted high resolution tandem mass spectrometry (HR-MS/MS) and bioinformatic analysis for protein identification and allergenicity prediction. The electrophoretic profile of lentil extracts confirmed that most of the proteomic profile was conserved across the investigated varieties, with only few differences highlighted for convicillin, vicilin and legumin subunits. The protein identification and label-free quantification by HR-MS/MS analysis provided a list of 365 entries ranked into main protein families prioritized basing on their experimental relative abundance (>0.5 %): cupins, late embryogenesis abundant proteins, lypoxigenases, lectins, dehydrogenases, redoxins, albumins I, albumins 2, elongation factors, peptidase, and heat shock proteins. Statistically significant differences were disclosed for most of these protein groups and/or subgroups: for example as for the cupins family, representing the major component of the protein profile for all lentil samples, an overexpression of vicilins in Eston and Laird lentil samples and of legumins in the Black lentil samples were calculated compared to the other varieties. The 7 S/11 S ratio describing seed nutritional quality disclosed clear differences among the four varieties with highest values reported for Eston and Laird samples. Moreover, ANFs such as lectins also showed a wide variability with the lowest expression calculated in Black lentil samples, although such differences might not be relevant for the direct consumption of cooked seeds because as heat labile proteins, will potentially be deactivated during the boiling step.

The *in-silico* allergenicity assessment pointed out that most of the identified proteins features either strong (132 out of 365) or weak (69 out of 365) evidence of immunogenicity confirmed the importance of including this legume in the watch-list of new prioritized allergens. The total abundance of proteins with either strong or weak evidence of allergenicity presented a certain variability across varieties with highest values reported for Eston and Black lentil samples, respectively. The hierarchical cluster analysis of the predicted allergenic accessions disclosed a first level of similarity between Laird and Crimson lentils, linked on a higher level to Black and Eston lentil samples, the latter being featured for the highest risk of allergenicity. Among the most relevant proteins with strong allergenic potential a key role was played by cupins family, notably, by the 7S-vicilins (about 43–44 % of total proteins), and 11 S legumins (about 14 % of total proteins) confirming the trend already discussed.

Finally, preliminary results obtained by unsupervised multivariate statistical analysis of the data collected by HR-MS/MS suggested the feasibility to select protein markers for lentil varieties discrimination, the significance of this results is very high compared to previous reported investigation due to the viability of the one-shot analytical approach applied.

Funding

This research received no external funding.

Declaration of Competing Interest

None

Data Availability

All data collected in this investigation have been included in the Manuscript and in the Supplementary materials

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jfca.2024.106456.

References

Alonso-Miravalles, L., Jeske, S., Bez, J., Detzel, A., Busch, M., Krueger, M., Arendt, E.K., 2019. Membrane filtration and isoelectric precipitation technological approaches for the preparation of novel, functional and sustainable protein isolate from lentils. Eur. Food Res. Technol. *245*, 1855–1869. https://doi.org/10.1007/s00217-019-03296-y.

Bourgeois, M., Jacquin, F., Savois, V., Sommerer, N., Labas, V., Henry, C., Burstin, J., 2009. Dissecting the proteome of pea mature seeds reveals the phenotypic plasticity of seed protein composition. Proteomics *9*, 254–271.

Boye, J., Zare, F., Pletch, A., 2010. Pulse proteins: praocessing, characterization, functional properties and applications in food and feed. Food Res. Int. 43, 414–431.

Capurso, A., Crepaldi, G., Capurso, C., 2018. Legumes and pulses. Benefits of the Mediterranean Diet in the Elderly Patient, Practical Issues in Geriatrics. Springer, Cham, pp. 285–324. https://doi.org/10.1007/978-3-319-78084-9_12.

Cavalluzzi, M.M., Lamonaca, A., Rotondo, N.P., Miniero, D.V., Muraglia, M., Gabriele, P., Corbo, F., De Palma, A., Budriesi, R., De Angelis, E., Monaci, L., Lentini, G., 2022. Microwave-Assisted Extraction of Bioactive Compounds from Lentil Wastes: Antioxidant Activity Evaluation and Metabolomic Characterization. Molecules *27* (21), 7471. https://doi.org/10.3390/molecules27217471.

Chauhan, J.S., Singh, B.B., Gupta, S., 2016. Enhancing pulses production in India through improving seed and variety replacement rates. Indian J. Genet. Plant Breed. *76* (04), 410–419. https://doi.org/10.5958/0975-6906.2016.00060.2.

Choi, H.J., Kang, S.W., Yang, C.H., Rhee, S.G., Ryu, S.E., 1998. Crystal structure of a novel human peroxidase enzyme at 2.0 Å resolution. Nat. Struct. Biol. *5* (5), 400–406. https://doi.org/10.1038/nsb0598-400.

Crespo, J.F., Pascual, C., Burks, A.W., Helm, R.M., Esteban, M.M., 1995. Frequency of food allergy in a pediatric population from Spain. Pediatr. Allergy Immunol. *6* (1), 39–43. https://doi.org/10.1111/j.1399-3038.1995.tb00256.x.

Dhull, S.B., Kinabo, J., Uebersax, M.A., 2023. Nutrient profile and effect of processing methods on the composition and functional properties of lentils (*Lens culinaris* Medik): A review. Legume Sci. 5 (1), e156 https://doi.org/10.1002/leg3.156.

Duran, Y., De La Vega, M.P., 2004. Assessment of genetic variation and species relationships in a collection of Lens using RAPD and ISSR [Inter-simple sequence repeats. Span. J. Agric. Res. *2* (4), 538–544. https://doi.org/10.5424/sjar/2004024-110.

Erskine, W., 1997. Lessons for breeders from land races of lentil. Euphytica *93*, 107–112. https://doi.org/10.1023/A:1002939704321.

FAOSTAT. (2021). Retrieved from ⟨https://www.fao.org/faostat/en/#data/QCL⟩. Accessed April 12, 2024.

Gan, Y., Hamel, C., O'Donovan, J.T., Cutforth, H., Zentner, R.P., Campbell, C.A., Niu, Y., Poppy, L., 2015. Diversifying crop rotations with pulses enhances system productivity. Sci. Rep. *5*, 14625 https://doi.org/10.1038/srep14625.

Halima, O., Najar, F.Z., Wahab, A., Gamagedara, S., Chowdhury, A.I., Foster, S.B., Ahsan, N., 2022. Lentil allergens identification and quantification: An update from omics perspective. Food Chem.: Mol. Sci. *4*, 100109 https://doi.org/10.1016/j.fochms.2022.100109.

Hundertmark, M., Hincha, D.K., 2008. LEA (late embryogenesis abundant) proteins and their encoding genes in Arabidopsis thaliana. BMC Genom. *9*, 1–22. https://doi.org/10.1186/1471-2164-9-118.

Joehnke, M.S., Jeske, S., Ispiryan, L., Zannini, E., Arendt, E.K., Bez, J., Petersen, I.L., 2021. Nutritional and anti-nutritional properties of lentil (Lens culinaris) protein isolates prepared by pilot-scale processing. Food Chem.: X *9*, 100112. https://doi.org/10.1016/j.fochx.2020.100112.

Khazaei, H., Subedi, M., Nickerson, M., Martínez-Villaluenga, C., Frias, J., Vandenberg, A., 2019. Seed protein of lentils: Current status, progress, and food applications. Foods *8* (9), 391. https://doi.org/10.3390/foods8090391.

Ladizinsky, G., 1979. The origin of lentil and its wild genepool. Euphytica *28*, 179–187. https://doi.org/10.1007/BF00029189.

Lindquist, S., Craig, E.A., 1988. The heat-shock proteins. Annu. Rev. Genet. *22*, 631–677. https://doi.org/10.1146/annurev.ge.22.120188.003215.

Marinangeli, C.P., Curran, J., Barr, S.I., Slavin, J., Puri, S., Swaminathan, S., Patterson, C. A., 2017. Enhancing nutrition with pulses: defining a recommended serving size for adults. Nutr. Rev. *75* (12), 990–1006. https://doi.org/10.1093/nutrit/nux058.

Maurer-Stroh, S., Krutz, N.L., Kern, P.S., Gunalan, V., Nguyen, M.N., Limviphuvadh, V., Gerberick, G.F., 2019. AllerCatPro—prediction of protein allergenicity potential from the protein sequence. Bioinformatics *35* (17), 3020–3027. https://doi.org/10.1093/bioinformatics/btz029.

Monaci, L., De Angelis, E., Guagnano, R., Ganci, A.P., Garaguso, I., Fiocchi, A., Pilolli, R., 2020. Validation of a MS Based Proteomics Method for Milk and Egg Quantification in Cookies at the Lowest VITAL Levels: An Alternative to the Use of Precautionary Labeling. Foods *9* (10), 1489. https://doi.org/10.3390/foods9101489.

Muehlbauer, F.J., Mihov, M., Vandenberg, A., Tullu, A., Materne, M., 2009. Improvement in developed countries. In: Erskine, W., Muehlbauer, F.J., Sarker, A., Sharma, B. (Eds.), The Lentil: Botany, Production and Uses. CAB International, Oxford, UK, pp. 137–154.

Nguyen, M.N., Krutz, N.L., Limviphuvadh, V., Lopata, A.L., Gerberick, G.F., Maurer-Stroh, S., 2022. AllerCatPro 2.0: a web server for predicting protein allergenicity potential. Nucleic Acids Res. *50* (W1), W36–W43. https://doi.org/10.1093/nar/gkac446.

Pilolli, R., Lamonaca, A., Nitride, C., De Angelis, E., van Poucke, C., Gillard, N., Huet, A. C., De Loose, M., Henrottin, J., Mills, E.C.N., Monaci, L., 2024. In-house validation of an LC-MS method for the multiplexed quantitative determination of total allergenic food in chocolate. Anal. Bioanal. Chem. *416* (3), 809–825. https://doi.org/10.1007/s00216-023-04894-2.

Pilolli, R., Van Poucke, C., De Angelis, E., Nitride, C., de Loose, M., Gillard, N., Huet, A. C., Tranquet, O., Larré, C., Adel-Patient, K., Bernard, H., Mills, E.N.C., Monaci, L., 2021. Discovery based high resolution MS/MS analysis for selection of allergen markers in chocolate and broth powder matrices. Food Chem. *343*, 128533 https://doi.org/10.1016/j.foodchem.2020.128533.

Puhakainen, T., Hess, M.W., Mäkelä, P., Svensson, J., Heino, P., Palva, E.T., 2004. Overexpression of multiple dehydrin genes enhances tolerance to freezing stress in Arabidopsis. Plant Mol. Biol. *54*, 743–753. https://doi.org/10.1023/B:PLAN.0000040903.66496.a4.

Qi, R.F., Song, Z.W., Chi, C.W., 2005. Structural features and molecular evolution of bowman-birk protease inhibitors and their potential application. Acta Biochim. Et. Biophys. Sin. 37 (5), 283–292. https://doi.org/10.1111/j.1745-7270.2005.00048.x.

Quintieri, L., Nitride, C., De Angelis, E., Lamonaca, A., Pilolli, R., Russo, F., Monaci, L., 2023. Alternative protein sources and novel foods: benefits, food applications and safety issues. Nutrients *15* (6), 1509. https://doi.org/10.3390/nu15061509.

Sackesen, C., Erman, B., Gimenez, G., Grishina, G., Yilmaz, O., Yavuz, S.T., Sampson, H. A., 2020. IgE and IgG4 binding to lentil epitopes in children with red and green lentil allergy. Pediatr. Allergy Immunol. *31* (2), 158–166. https://doi.org/10.1111/pai.13136.

Scippa, G.S., Rocco, M., Ialicicco, M., Trupiano, D., Viscosi, V., Di Michele, M., Arena, S., Chiatante, D., Scaloni, A., 2010. The proteome of lentil (Lens culinaris Medik.) seeds: discriminating between landraces. Electrophoresis *31*, 497–506 https://doi.org:10.1002/elps.200900459.

Shaheen, N., Halima, O., Akhter, K.T., Nuzhat, N., Rao, R.S.P., Wilson, R.S., Ahsan, N., 2019. Proteomic characterization of low molecular weight allergens and putative allergen proteins in lentil (Lens culinaris) cultivars of Bangladesh. Food Chem. 297, 124936 https://doi.org/10.1016/j.foodchem.2019.06.003.

de Souza Cândido, E., Pinto, M.F.S., Pelegrini, P.B., Lima, T.B., Silva, O.N., Pogue, R., Franco, O.L., 2011. Plant storage proteins with antimicrobial activity: novel insights into plant defense mechanisms. FASEB J. *25* (10), 3290–3305. https://doi.org/10.1096/fj.11-184291.

The UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase. Retrieved from ⟨https://www.uniprot.org/⟩. Accessed February 2, 2023.

Uasuf, C.G., De Angelis, E., Guagnano, R., Pilolli, R., D'Anna, C., Villalta, D., Brusca, I., Monaci, L., 2020. Emerging allergens in Goji berry superfruit: the identification of new IgE binding proteins towards allergic patients' Sera. Biomolecules *10* (5), 689 https://doi.org.10.3390/biom10050689.

World Health Organization, 2022. Risk Assessment of Food Allergens. Part 1-Review and validation of Codex Alimentarius priority allergen list through risk assessment: meeting report (Rome). Food Saf. Qual. Ser. No 14. https://doi.org/10.4060/cb9070en.

Yamazaki, T., Takaoka, M., Katoh, E., Hanada, K., Sakita, M., Sakata, K., Hirano, H., 2003. A possible physiological function and the tertiary structure of a 4-kDa peptide in legumes. Eur. J. Biochem. *270* (6), 1269–1276. https://doi.org/10.1046/j.1432-1033.2003.03489.x.

Zhou, J., Li, M., Bai, Q., de Souza, T.S.P., Barrow, C., Dunshea, F., Suleria, H.A.R., 2023. Effects of different processing methods on pulses phytochemicals: an overview. Food Rev. Int. *40* (4), 1138–1195. https://doi.org/10.1080/87559129.2023.2212041.