

Proteomic characterization of low molecular weight allergens and putative allergen proteins in lentil (*Lens culinaris*) cultivars of Bangladesh

Nazma Shaheen^{a,*}, Oumma Halima^{a,1}, Kazi Turjaun Akhter^a, Noshin Nuzhat^b,
R. Shyama Prasad Rao^c, Rashaun S. Wilson^d, Nagib Ahsan^{b,e,*}

^a Institute of Nutrition and Food Science, University of Dhaka, Dhaka 1000, Bangladesh

^b Center for Cancer Research Development, Proteomics Core Facility, Rhode Island Hospital, Providence, RI 02903, USA

^c Biostatistics and Bioinformatics Division, Yenepoya Research Center, Yenepoya University, Mangalore 575018, India

^d Keck MS & Proteomics Resource, Yale University, New Haven, CT 06511, USA

^e Division of Biology and Medicine, Brown University, Providence, RI 02903, USA

ARTICLE INFO

Keywords:

Allergens

Lentil

Lipid transfer proteins

MRM

Proteome

ABSTRACT

Here, we present the proteome profiling of low-molecular weight (< 50 kDa) proteins of seven different lentil cultivars developed by Bangladesh Agricultural Research Institute. A total of 2873 peptides corresponding to 180 unique proteins were identified wherein > 24% of them were described lentil allergens. Comparative relative quantitation showed differences in protein abundance of major allergen proteins such as Len c 1.0101, Len c 1.0102, and lipid transfer proteins (LTPs), indicating qualitative and quantitative variations in allergen proteins in lentil cultivars. In this report, for the first-time, the amino acid sequence of LTPs in lentil has been confirmed by high resolution mass spectrometry. In addition, ideal peptides of Len c 1.0101, Len c 1.0102, and LTPs allergens were further determined with multiple reaction monitoring (MRM) analysis. Therefore, this data could provide a great resource for further development of targeted, proteomics-based assays for quantification of lentil allergens.

1. Introduction

Lentil (*Lens culinaris*) is one of the most consumed pulses in Asian countries, and is becoming increasingly popular in developed countries due to its high nutritional composition and protein content (Mudryj, Yu, & Aukema, 2014). Lentils are listed among the healthiest foods and considered as the fourth most important pulse crop in the world (Longobardi et al., 2017). As an alternative source to animal protein, lentils are a pivotal component of the predominantly rice-based diet in Bangladesh. It has been estimated that approximately 3.1% of children in Bangladesh suffer from severe, acute malnutrition wherein a ready-to-use therapeutic food made with rice and lentil is well-accepted by the children, thus potentially saving children's lives through the supplementation of essential amino acids (Choudhury et al., 2018). In addition, recent studies have shown that a lentil-based diet is comparable to an animal protein-based diet in terms of absorption of many essential nutrients in malnourished children (Hossain, Islam, Wahed, Khatun, & Kabir, 2009; Podder et al., 2018). In our earlier study, we demonstrate that lentils, if combined with rice or wheat flour can help meet the

indispensable amino acids requirements of the Bangladeshi population (Shaheen, Islam, Munmun, Mohiduzzaman, & Longvah, 2016). Conversely, lentils are among the main plant foods causing allergic reactions in pediatric patients in many Asian and European countries (Akkerdaas et al., 2012; Asero, 2011; Cuadrado et al., 2009; López-Torrejón et al., 2003; Sánchez-Monge et al., 2000; Yavuz et al., 2013).

Although IgE-mediated hypersensitivity reactions of lentils have been reported in many studies (Cuadrado et al., 2009; Ibáñez Sandín et al., 1999; Manso et al., 2010; Martínez San Ireneo, Ibáñez Sandín, & Fernández-Caldas, 2000; Pascual et al., 1999), deep sequencing of lentil allergens or putative allergen proteins at their amino acid level is very limited and poorly annotated compared to other legumes such as peanut and soybean. To date, only partial peptide sequencing of a handful of allergen proteins such as Len c 1.0101, a 47-kDa vicilin from the cupin superfamily, Len c 2.0101, a seed-specific biotinylated protein, and Len c 3.0101, which is a lipid transfer protein (LTP) from the prolamin superfamily, along with several isoforms have been identified and verified as lentil allergens (Akkerdaas et al., 2012; Asero, 2011; Bogdanov et al., 2015; Finkina et al., 2007, 2012).

* Corresponding authors at: Center for Cancer Research Development, Proteomics Core Facility, Rhode Island Hospital, Providence, RI 02903, USA (N. Ahsan).
E-mail addresses: nazmashaheen@du.ac.bd (N. Shaheen), nagib_ahsan@brown.edu, nagib.ahsan@lifespan.org (N. Ahsan).

¹ Current address: Noakhali Science and Technology University, Noakhali 3814, Bangladesh.

Lentil crop area and production have been simultaneously decreasing in Bangladesh over the last two decades, which significantly affects the per capita pulse consumption. In an effort to become self-sufficient in lentil production, Bangladesh Agricultural Research Institute (BARI) has been making a continuous effort to develop high-yielding, environmentally sustainable lentil varieties through introduction of new germplasm, particularly from the International Center for Agricultural Research in the Dry Areas (ICARDA). However, it is important to note that cultivar-specific and other agricultural practices mediate the heterogeneity of allergen proteins' distribution and quantification, which has been reported in many seeds (Chen, Yang, & Downs, 2019; Geng et al., 2017; McClain et al., 2018; Stevenson et al., 2012; Stevenson, McClain, & Thelen, 2015). Although, there is no comparative quantitative analysis that has been carried out on lentil allergens, cultivar-specific, differential proteome expression has been demonstrated by two-dimensional gel electrophoresis (2DE)-based proteomic approaches (Iallicco et al., 2012; Scippa et al., 2010). Furthermore, a recent metabolomic analysis of different lentil cultivars showed significant alteration in many nutritional components (Farg et al., 2018), indicating lentil allergens could also show genotype-specific heterogeneity. Interestingly, cultivar-specific skin reactivity and *in vitro* IgE binding potencies have been demonstrated in legumes (Codina, Arduoso, Lockey, Crisci, & Medina, 2003), suggesting that plant allergenic potential varies with genotype.

The aim of this work is to deep sequence and quantifies allergens and putative allergen proteins in lentil cultivars using a label-free proteomic approach. Specifically, we focused on low molecular weight proteins (< 50 kDa), as most of the lentil allergens are verified in immunoblot analysis. To the best of our knowledge, the dataset generated here is the largest proteome sequencing of lentils. Furthermore, we listed the ideal tryptic peptides for all lentil allergens that could be very useful in further quantitative analysis by using single/multiple reaction monitoring (SRM/MRM) based targeted proteomic approaches.

2. Materials and methods

2.1. Lentil cultivars and seed collection

Seeds of lentil cultivars were collected from BARI, which were grown in its on-farm, research-based plot. Each of these cultivars have very different genetic backgrounds with distinct characteristics such as origin, protein content, seed weight, etc. Detailed information regarding these cultivars is listed in [Supplemental Table 1](#).

2.2. Protein extraction from lentil seeds and SDS-PAGE analysis

De-husked dried seeds were obtained from BARI. For each cultivar, a mixture of seeds (100 mg) obtained from field-grown plants were ground using liquid nitrogen. Total seed protein was extracted from each sample under denaturing conditions using a modified phenol partitioning protocol (Stevenson et al., 2015) and precipitated with three different precipitation buffers: (1) 0.1 M ammonium acetate in methanol, (2) ice cold 75% acetone, and (3) ice cold 10% trichloroacetic acid (TCA), 20 mM dithiothreitol (DTT) in acetone. Precipitated protein from three different precipitation buffers was resuspended in dissolution lysis buffer (8 M urea, 50 mM Tris-HCl, pH 8.0) and quantified using the BCA Protein Kit (Thermo Fisher Scientific, Houston, TX) using bovine serum albumin (BSA) as a standard. Protein lysate was diluted 4-fold with 20 mM HEPES buffer, pH 8.0 prior to protein concentration measurement and trypsin digestion. Gel electrophoresis was performed under denaturing conditions in 13% polyacrylamide gels using 20 mA per gel. Equal amounts of protein (25 µg) were loaded in each lane. Gels were stained with colloidal Coomassie blue stain under standard conditions.

2.3. In-gel protein digestion and mass spectrometry analysis

Prior to protein digestion, gel lanes for each sample were cut from 50 kDa to the bottom of the stain, diced into approximately 1 mm cubes with a scalpel, and transferred into an Eppendorf tube. Tryptic digestion was carried out according to a previously-established procedure (Stevenson et al., 2015). Tryptic peptides were desalted using C18 Sep-Pak plus cartridges (Waters, Milford, MA) and were lyophilized for 48 h to dryness. The dried, eluted peptides were reconstituted in 20 µl of buffer A (0.1 M acetic acid), and 5 µl was injected for each analysis.

The LC-MS/MS was performed on a fully automated proteomic technology platform that includes an Agilent 1200 Series Quaternary HPLC system (Agilent Technologies, Santa Clara, CA) connected to a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific, Waltham, MA). The LC-MS/MS set up was used as described earlier (Ahsan, Belmont, Chen, Clifton, & Salomon, 2017). Briefly, the peptides were separated through a linear reversed-phase 90 min gradient from 0% to 40% buffer B (0.1 M acetic acid in acetonitrile) at a flow rate of 3 µl/min through a 3 µm × 20 cm C18 column. The electrospray voltage of 2.0 kV was applied in a split flow configuration, and spectra were collected using a top-9 data-dependent method. Survey full scan MS spectra (*m/z* 400–1800) were acquired at a resolution of 70,000 with an automatic gain control target value of 3×10^6 ions or a maximum ion injection time of 200 ms. The peptide fragmentation was performed via higher-energy collision dissociation (HCD) with the collision energy set at 28. The MS/MS spectra were acquired at a resolution of 17,500, with a targeted value of 2×10^4 ions or a maximum integration time of 200 ms. The ion selection abundance threshold was set at 8.0×10^2 with charge state exclusion of unassigned and $z = 1$, or 6–8 ions and dynamic exclusion time of 30 s.

2.4. Database search and relative quantitation

Due to lack of complete genome sequence of the lentil, MS RAW files were searched against the Fabaceae databases (TaxID163743, NCBI; downloaded 13/02/2018) annotated over 4888 proteins using the Sequest (XCorr Only) algorithm within Proteome Discoverer v 2.1.1.21. The Sequest (Thermo Fisher Scientific, San Jose, CA) database search was performed with the following parameters: trypsin enzyme cleavage specificity, 2 possible missed cleavages, 10 ppm mass tolerance for precursor ions, 0.020 Da mass tolerance for fragment ions. Search parameters permitted variable modification of methionine oxidation (+15.9949 Da) and static modification of carbamidomethylation (+57.0215 Da) on cysteine. The resulting peptide spectrum matches (PSMs) were imported into Scaffold software v. 4.8.9 (Proteome Software Inc., Portland, OR) for further filtering. Peptide and protein identifications were accepted at a threshold of 95% and 99% probability, respectively, which was determined by the Peptide Prophet algorithm. The minimum number of identified peptides/protein was set to 1. Furthermore, a BLAST search was performed to annotate peptides identified in hypothetical or unknown proteins.

2.5. Functional annotation and classification

Identified proteins were annotated and functionally classified based on gene ontology (GO) terms using Blast2GO v 5.2.5. Alignments were obtained using the BLASTP program against the NCBI nr database using default parameters and 1×10^{-3} for expected value with no taxonomy filter and Blast Description Annotator enabled. The word size was limited to 3 and the HSP Length Cutoff was set to 33. Mapping and annotation steps were also performed using BLAST2GO default values. Mapping was performed using the latest database version and Goa Version 2018.12. The Annotation CutOff and GO Weight was set to 55 and 5, respectively. The E-Value-Hit-Filter was set at 1×10^{-6} and the Hit Filter was set at 500. GO combined charts related to biological process and molecular function were represented at the tertiary level.

2.6. Multiple reaction monitoring (MRM) based LC–MS/MS analysis

Targeted peptides coupled with an internal standard peptide LVN-ELTEFAK corresponding to BSA with their doubly-charged states resulted a total of 146 product ions. These product ions were subjected to MRM analysis by a triple quadrupole mass spectrometer (TSQ Quantiva, Thermo Fisher Scientific, Germany) equipped with a nano UHPLC (EASY-nLC™ 1200, Thermo Fisher Scientific, Germany). For the majority of the peptides, the y-ions were selected for analysis. The selected peptides corresponded to three known allergen proteins namely Len c 1.0101, Len c 1.0102 and LTPs. The MRM method was developed with Skyline software version 4.1, which is an open source tool for mass spectrometry analysis. See [Supplemental Table 2](#) for peptide details, including the precursor and product ions used for MRM method development. The parameters used for the quadrupole mass spectrometer instrument are as follows: positive ion spray voltage, 2200 kV; ion transfer tube temperature, 350 °C; cycle time, 0.5 s; chromatographic peak width, 3.0 s; collision pressure, 1.5 mTorr; and Q1 and Q3 resolutions, 0.2 and 0.7 (FWHM), respectively.

Trypsin-digested samples (5 µl) containing the desalted peptides were chromatographically separated using reverse-phase chromatography with acidified ACN (0.1 M acetic acid in acetonitrile) and water (0.1 M acetic acid in LC-MS grade water) mobile phases. Separation of peptides was achieved with a 30 min linear gradient from 5 to 95% over 22 min of ACN using a flow rate of 300 nL/min through a 3 µm (75 µm × 15 cm) C18 analytical column. MS RAW files were further processed and analyzed by Skyline software (Ver. 4.1).

3. Results

3.1. Low molecular weight proteome of lentil

Among the three precipitation buffers tested for evaluating the efficiency of the protein extraction method, buffer 1 and 3 both yielded relatively higher amounts of protein comparative to buffer 2 (ice cold 75% acetone) (data not shown). Furthermore, the low molecular bands displayed higher intensity and visibility in buffer 1 compared to the other two precipitation buffers ([Supplementary Fig. 1](#)). Therefore, in this study, phenol extraction followed by precipitation with 0.1 M ammonium acetate in methanol (buffer 1) was used for downstream LC–MS/MS analysis for all seven samples.

A total of 2873 peptides were identified [possibly due to charge and post translational modification (PTM) variants] corresponding to 782 unique peptide sequences from 180 proteins. A detailed list of the identified peptides and proteins is provided in the [Supplementary dataset](#) ([Supplementary Tables 3 and 4](#)). The expression pattern and overall protein abundance (peptide spectrum match, PSM) of proteins in seven lentil samples are shown in [Fig. 1A](#) and [B](#), respectively. It is important to note that most of the high-abundant proteins identified in this study are known allergens (such as Len c 1.0101 and Len c 1.0102) and/or putative allergen proteins. Overall, the allergens or putative allergen proteins vary widely in their abundance. A total of 44 allergens or putative allergen proteins (including various isoforms) were identified, which constitutes 24.4% of the total identified proteins ([Fig. 1C](#)). In addition, allergens comprised of more than 86% of total protein abundance ([Fig. 1D](#)) indicating that a major portion of the low molecular weight proteins are allergens or putative allergen proteins.

3.2. Peptide sequencing and relative quantitation of allergens and putative allergen proteins in lentil cultivars

A heat map and clustering of 44 identified allergens and/or putative allergen proteins is displayed in [Fig. 2A](#). Len c 1.0101 and Len c 1.0102 are two of the most abundant allergens, whereas LTP 2, and LTP 6 are some of the least abundant allergens. Expression profiles of selected allergens across seven lentil cultivars are shown in [Fig. 2B](#). As indicated,

some known allergens such as allergen Len c 1.0101, Len c 1.0102, and vicilin are highly expressed and present in nearly equal amounts in all seven lentil cultivars, whereas other allergens such as provicilin, legumin A2, and LTPs are more variable across samples. In general, low abundant allergens are more variable across samples, and the coefficient of variation (CV) ranges from 4.8 to 264.6%.

3.3. Functional analysis of the lentil low molecular proteins

In order to investigate the overall role of the lentil low molecular weight proteome, the identified proteins were assigned GO annotations using Blast2GO analysis. As expected, > 80% of the proteins are associated with different metabolic processes including cellular metabolic (20%), nitrogen compound metabolic (18%), and primary metabolic (16%) processes, whereas only 17% of proteins are involved with biosynthetic processes ([Supplementary Fig. 2A](#)). Similarly, > 55% of the identified proteins were cytoplasmic, followed by intercellular non-membrane bound organelle (20%), membrane bound organelle (13%) and ribosome (10%) ([Supplementary Fig. 2B](#)). In terms of molecular functions, over 40% of the proteins are involved in ion binding, whereas proteins involved in oxidoreductase, transferase, hydrolase, and lyase activities are equally abundant ([Supplementary Fig. 2C](#)).

3.4. Comparative analysis of the lentil proteome with other studies

A comparison of the proteome profile between the current study and two previously published reports showed a higher number of proteins were identified in the current study compared to the earlier published reports. More importantly, approximately 50% of the proteins identified in previous studies overlapped with the current study ([Fig. 3](#)), suggesting a reasonable consistency of the lentil seed proteome. It is important to note that earlier studies conducted with a 2DE proteomics approach characterized both low and high molecular weight proteins (208–7.5 kDa) of the lentil seed proteome whereas only low molecular weight proteins (< 50 kDa) were examined in the current study. As expected, a large number of unique proteins were identified, including known allergens such as LTPs. The presence of LTPs in lentils was confirmed previously by immunoblot analysis ([Akkerdaas et al., 2012; Asero, 2011; Bogdanov et al., 2015; Finkina et al., 2007, 2012](#)), however two peptides corresponding to LTP2 and LTP6 were identified by mass spectrometry for the first time, to the best of our knowledge. In this study, we have found that LTPs are less abundant than other allergen proteins ([Fig. 2](#)). Therefore, it is not surprising why the earlier studies were unable to sequence LTPs, as both used the 2DE gel-based approach. Among the overlapped proteins, most are known allergens such as Len c 1.0101, Len c 1.0102, vicilin, lectin, convicilin, etc. ([Fig. 3](#)).

3.5. Lentil LTP homology and amino acid sequence assessments

In order to develop a multiple monitoring reaction (MRM) based assay for quantitation of lentil LTP isoforms, amino acid sequences of all six known lentil LTP isoforms were grouped by their phylogenetic relationship ([Fig. 4A](#)). A close relationship was observed among the four isoforms LTP2, 4, 5, and 6, suggesting these isoforms originate from the same clade. A pair-wise sequence alignment with LTP1 indicated that the sequence identity ranged from 65 to 74% among the LTPs ([Fig. 4B](#)). Based on the amino acid sequence and alignment, LTP4 is likely a splice variant of LTP6, as there was no N-terminal start codon for LTP4. As shown, the peptide sequence (LLAAANTTPD), identified by discovery proteomics, corresponded to all 4 of these LTPs ([Fig. 4C](#)), suggesting these isoforms or one of these LTP could be the most abundant. Furthermore, feasibility of using an MRM assay for validating the presence of all LTP isoforms was assessed using sequence alignment followed by critiquing the possible tryptic peptides. As expected, the number of unique tryptic peptides corresponding to

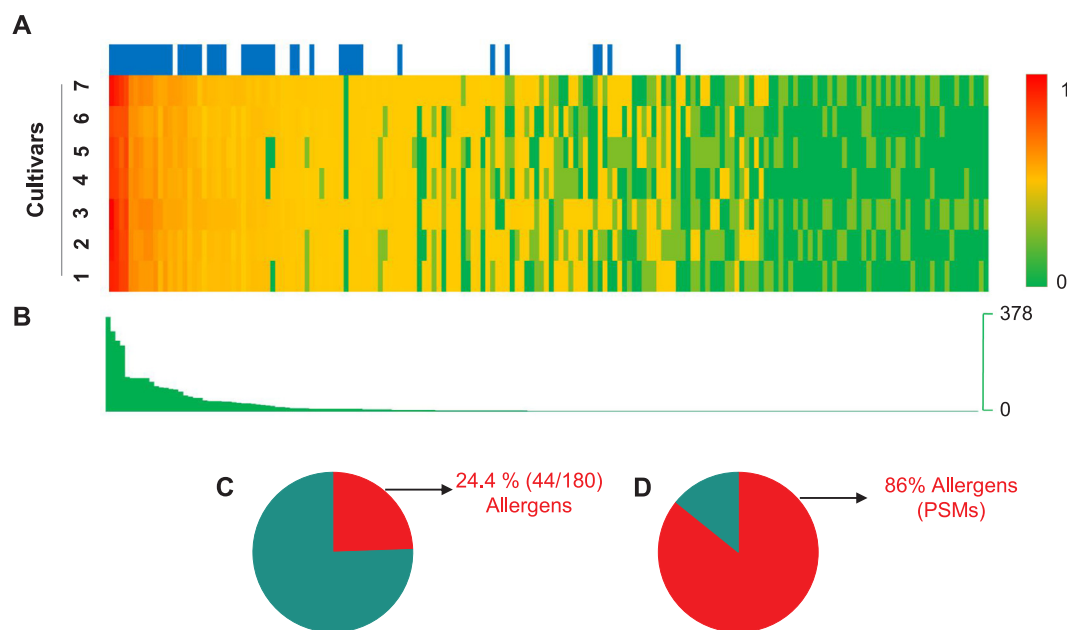


Fig. 1. (A) Heat map of < 50 kDa protein abundances from seven lentil samples (1–7). Of the 180 proteins (including different isoforms) identified, at least 44 are known/probable allergens (indicated by blue rectangles at the top of the heat map). Scale bar indicates normalized PSM to the most abundant protein. (B) Graph shows the maximum PSM per protein (among seven samples), arranged in descending order. (C) Allergens formed 24.4% of the identified proteins (44/180), but (D) it may be noted that most of the high abundant < 50 kDa proteins identified in this study are known/probable allergens which make up more than 86% of abundance by PSMs.

individual LTP is very low, and the unique peptides ranged from 1 to 6. The highest numbers of unique peptides were demonstrated in LTP1 whereas LTP2 and LTP6 have only a single unique peptide. Interestingly, no unique tryptic peptide was identified corresponding to LTP4 (Fig. 4B). These peptides along with other identified peptides corresponding to Len c 1.0101 and Len c 1.0102 allergen protein isoforms were selected for further MRM analysis (Fig. 4D).

3.6. Peptide sequence validation of lentil allergen proteins by MRM analysis

A total of 17 peptides were subjected to MRM analysis (Fig. 5). Among these peptides, the control Len c allergen peptides displayed significantly higher signals compared to the LTPs peptides (Fig. 5), which is also consistent with our discovery data, indicating LTPs are less abundant in the samples. As shown, all transition ions for a selected unique peptide corresponding to Len c allergen proteins are co-eluted under the same retention time with high signal. Therefore, these peptides are ideal for further isoform-specific MRM-based quantitative analysis of Len c allergens (Fig. 5). Among the ten LTP peptides selected for MRM analysis only one peptide “LLAAANTTPDR” corresponding to LTPs 2, 4, 5, and 6 displayed high signal with ideal characteristics including but not limited to identification of multiple transitions. This suggests one or more of these LTPs could be more abundant than other LTP isoforms (Fig. 5). Unique peptides corresponding to other LTP isoforms also displayed acceptable signals but only for one or two transitions (Supplementary Fig. 3). Therefore, these peptides were not considered for a positive identification for further quantitative analysis of LTP allergen proteins.

4. Discussion

Human nutrition is enormously complex, and a “healthy diet” may vary widely depending on an individual’s genetic make-up, environment, and cultural context. Legumes are becoming a focus for sustainable production and consumption in achieving sustainable development according to goal 3 “Good Health” in many Asian countries. Among the legumes, lentils are vital components in diversification of Bangladesh’s

predominantly rice-based crop system as well as its consumption pattern. In addition, lentils have frequently been used supplemented to address the existing malnutrition of under-five children (Desrochers & Brauer, 2001). To mitigate the huge demand of lentils in the dietary system in Bangladesh, priority has been given to collection and evaluation of local and exotic lentil cultivars, as well as development of high yielding and environmental sustainable cultivars (Choudhury et al., 2018; Hossain et al., 2009; Podder et al., 2018). To the best of our knowledge, no large-scale, systematic analysis has been conducted and/or reported on the allergenic responses of the Bangladeshi population in lentils. However, allergic symptoms and/or IgE cross-reactivity in response to lentil proteins have been demonstrated in a larger, worldwide population (Akpınar, 2018; Asero, 2011; Cuadrado et al., 2009; Ibanez, Martinez, Sanchez, & Fernandez-Caldas, 2003; Kumar et al., 2010; Vitaliti et al., 2012; Yavuz et al., 2013). Interestingly, most of the IgE cross-reactivity that has been confirmed so far is attributed to the low molecular weight proteins of lentils, particularly LTPs (Akkerdaas et al., 2012; Asero, 2011; Bogdanov et al., 2015; Cuadrado et al., 2009; Finkina et al., 2007, 2012; Gizatullina et al., 2013; Ibanez et al., 2003; Manso et al., 2010).

Therefore, determining the accurate protein sequence of these known allergen proteins will provide a better platform to develop alternative mass spectrometry-based quantitation methods of lentil allergens. Mass spectrometry-based proteomic approaches have been successfully employed not only to determine and quantify the food allergens (Ahsan, Rao, Gruppiso, Ramratnam, & Salomon, 2016; Chen et al., 2019; Planque et al., 2019) but also to demonstrate the conformational changes of many allergen proteins (Yang, Liu, & Wang, 2019). Among the mass spectrometry-based methods, a targeted proteomic approach using the absolute quantification (AQUA)-MRM platform has been considered a reliable, highly sensitive, and specific method for quantification of allergens or putative allergen proteins from many plant species (Ahsan, Wilson, & Thelen, 2018; McClain et al., 2018; Stevenson et al., 2012, 2015). However, it is important to note that peptide sequencing of the corresponding target allergen protein is the principal step to develop an AQUA-MRM-based method for allergen quantification (Ahsan et al., 2018).

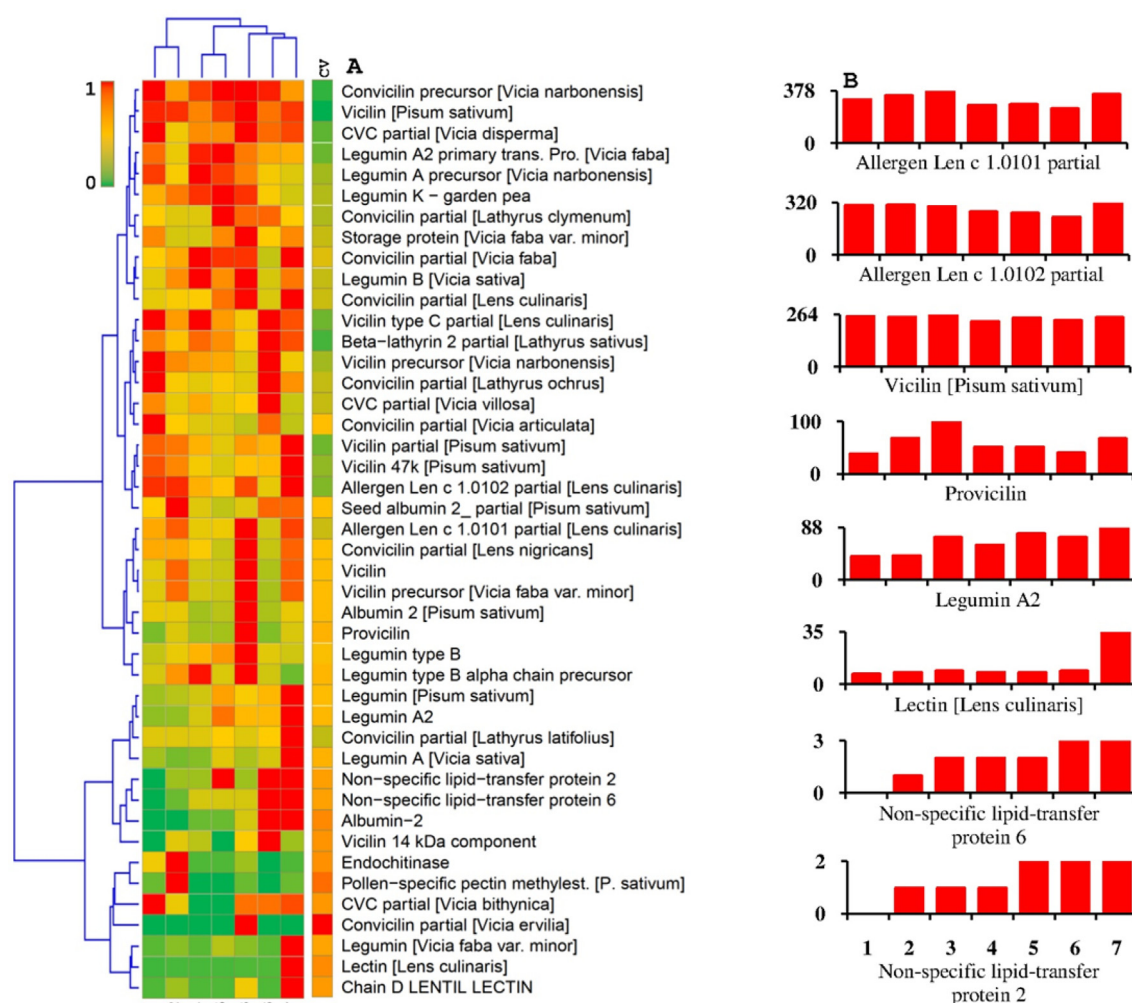


Fig. 2. (A) Heat map and clustering of 44 allergen isoforms identified in seven lentil samples. Many high-abundant allergens are less variable across samples (as indicated by lower coefficient of variation [CV], which ranged from 4.8 to 264.6%). (B) PSM expression profiles of selected allergens are shown. Allergen isoforms such as allergen Len c 1.0101, Len c 1.0102, vicilins, and legumins are some of the most abundant, whereas convicilins, and lipid-transfer proteins are the least abundant.

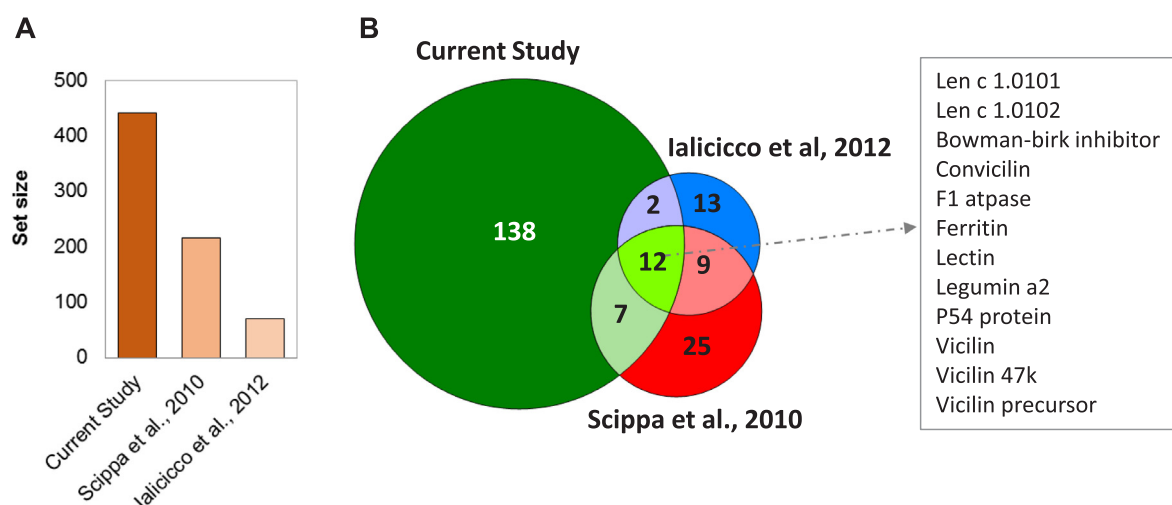
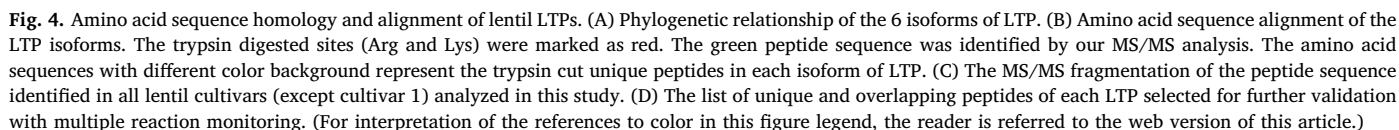


Fig. 3. Comparative lentil proteome analysis. (A) Comparison of the current study with literature-derived lentil proteome data. The publications are cited in the main text. The set size indicates the total number of proteins identified from each dataset. (B) Venn diagram analysis of unique proteins (by name in each experiment) shows the unique and overlapped proteins identified in the current study with earlier two lentil proteomic studies.



One of the best examples could be the LTPs in lentils which were identified and confirmed as potential allergens by immunoblot analysis (Akkerdaas et al., 2012; Asero, 2011; Bogdanov et al., 2015; Finkina et al., 2007, 2012; Gizatullina et al., 2013). The protein database contained amino acid sequences of six lentil LTP isoforms, three of which have been validated by immunoblot analysis (Bogdanov et al., 2015; Gizatullina et al., 2013). However, the peptide sequencing was not validated by any proteomic experiments. So far, to the best of our knowledge, no LTPs were sequenced in the earlier two discovery-based comparative proteomic experiments on lentil cultivars (Iallicco et al., 2012; Scippa et al., 2010). LTPs are of interest because of their wide range of substrate-binding specificity and are considered one of the

In this study, for the first time, we have successfully identified multiple peptides corresponding to the isoforms of LTP2 and LTP6. Similar to other plant LTPs, the amino acid sequences of lentil LTPs are highly conserved (Fig. 4A and B). Interestingly, peptide LLAAANTPDR which was identified in this study in all samples (Fig. 4C and D) and is present among four LTP isoforms (LTP2, LTP4, LTP5 and LTP6), supporting the phylogenetic data that these four LTPs originated from the same clade. Collectively, these results suggest that LTP2, LTP4, LTP5, and/or LTP6 are significantly more abundant than LTP1 and LTP3, at least for these lentil samples. In support of this hypothesis, as an abundant LTP in lentil seed, LTP2 has been identified and characterized by other biochemical assays (Gizatullina et al., 2013). However, it is important to note that in this study overall LTP expression levels were less abundant than other low molecular weight allergen proteins identified in lentil seed. Therefore, it is likely that LTPs could not be detectable or could show very minimal differential changes in 2DE based proteomics platforms (Iallicco et al., 2012; Scippa et al., 2010). To the best of our knowledge, in this study, for the first time major lentil allergen proteins Len c 1.0101, Len c 1.0102 and LTPs were verified by a MRM proteomic platform. Thus, amino acid sequence validation of LTPs coupled with two other known allergen peptides in the lentil proteome, Len c 1.0101 and Len c 1.0102 by a targeted proteomic

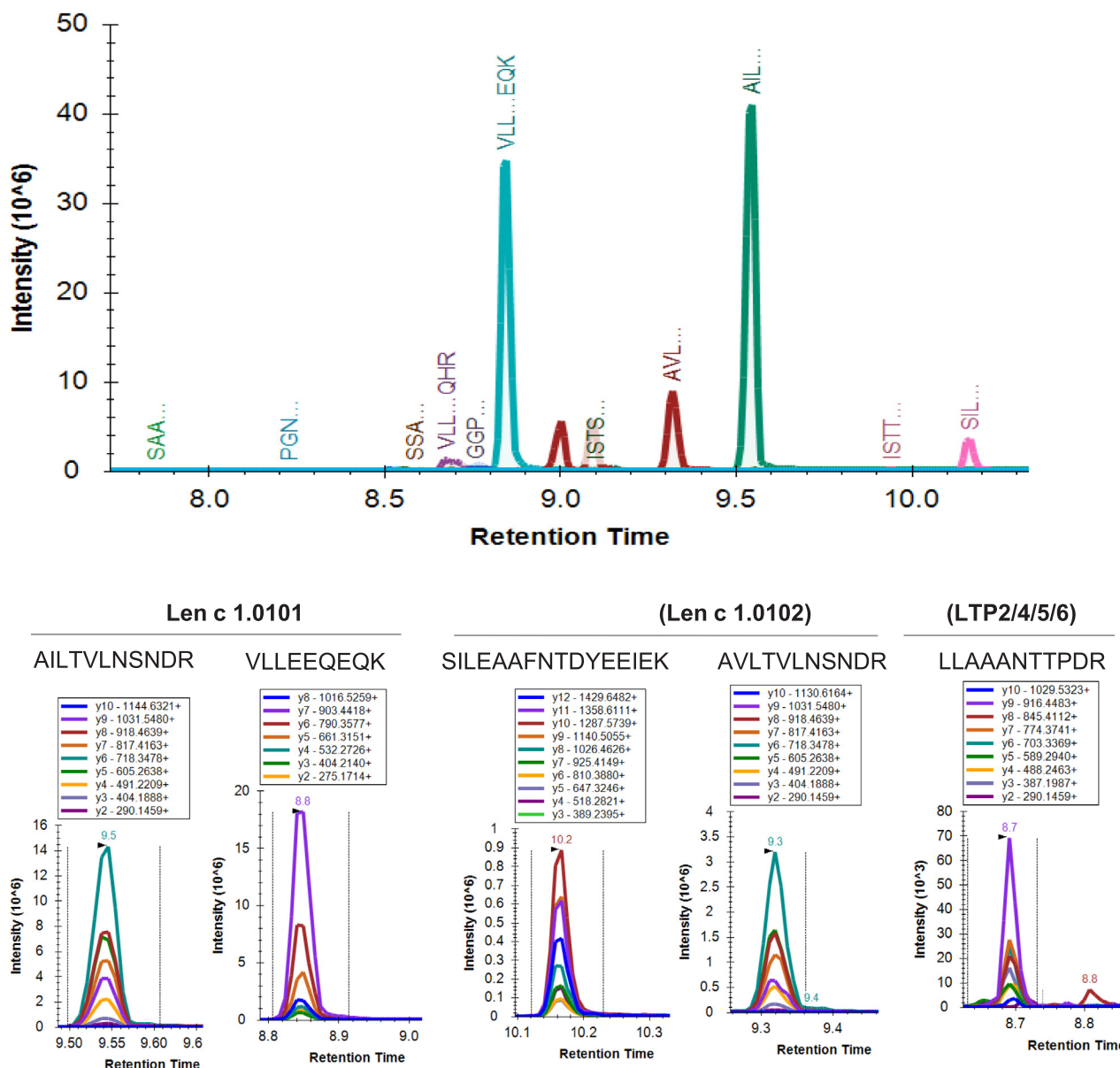


Fig. 5. Selection of ideal lentil allergen peptides for targeted proteomics platform using multiple reaction monitoring (MRM). Trypsin digested peptides (5 μ l) were analyzed by a triple quadrupole mass spectrometer using a MRM platform. The peaks represent the chromatogram of successful identification of some peptides visualized by Skyline software. The bottom panel shows the ideal MRM for some of the targeted peptides of Len c 1.0101, Len c 1.0102, and LTPs. As shown, the transition ions (different color peaks) of these peptides precursor ions displayed very ideal characteristics in terms of high signal and all the transitions are eluted under the same retention time. Therefore, these peptides are considered as positive identifications by MRM analysis in lentil samples.

approach provides a great platform for selecting ideal peptides for development of an absolutely quantitative proteomic assay.

In summary, relative quantitative analysis among the cultivars showed differences in abundance of the allergen proteins indicating cultivar specificity for lentil allergen proteins. In addition, validation of some lentil allergen proteins by MRM analysis offers ideal peptides for further targeted, quantitative analysis. Therefore, this dataset could be used as a valuable tool further of a development MS-based assay for quantification of lentil allergens. Although, peptide sequencing of lentil low molecular-weight seed proteins with high resolution mass spectrometry provides successful identification of many known allergens and putative allergen proteins, we were unable to detect and/or sequence unique peptides for all known six lentil LTP isoforms. This limitation could be due to use of a gel-based proteomics platform coupled with the limited lentil variety that was screened. Genetic

variation of LTP genes has recently been demonstrated in lentil germ-plasms (Basheer-Salimia, Camilli, Scacchi, Noli, & Awad, 2015). Thus, shotgun proteomics coupled to proteogenomic analysis with an available transcriptome dataset would be a better choice for protein sequencing and accurate protein annotation not only for all known allergens proteins but also other expressed proteins in lentil seed.

Declaration of Competing Interest

None.

Acknowledgements

This work was partly supported by the Higher Education Quality Enhancement Project (HEQEP), University Grants Commission,

Bangladesh (CP-3270) and partially by Grant for Advanced Research in Education (GARE), Bangladesh Bureau of Educational Information & Statistics (BANBEIS), Ministry of Education, Government of the People's Republic of Bangladesh. We are very thankful to Drs Rakhohori Sarkar, Aftab Uddin and Rezaul Karim for their generous help prior sample collection and processing steps.

Authors contribution

NA and NS designed the experiment. OH and KTA performed tissue processing and protein extraction methods. NA performed mass spectrometry, and RSW performed database searching and quantitative analysis. NA, SPR, NN and RSW implemented bioinformatics analysis. NA performed the mining and functional analysis of the data, and designed research. NA and NS, wrote the manuscript; RSW and SRP edited and contributed to the writing. All authors discussed the results and commented on the manuscript.

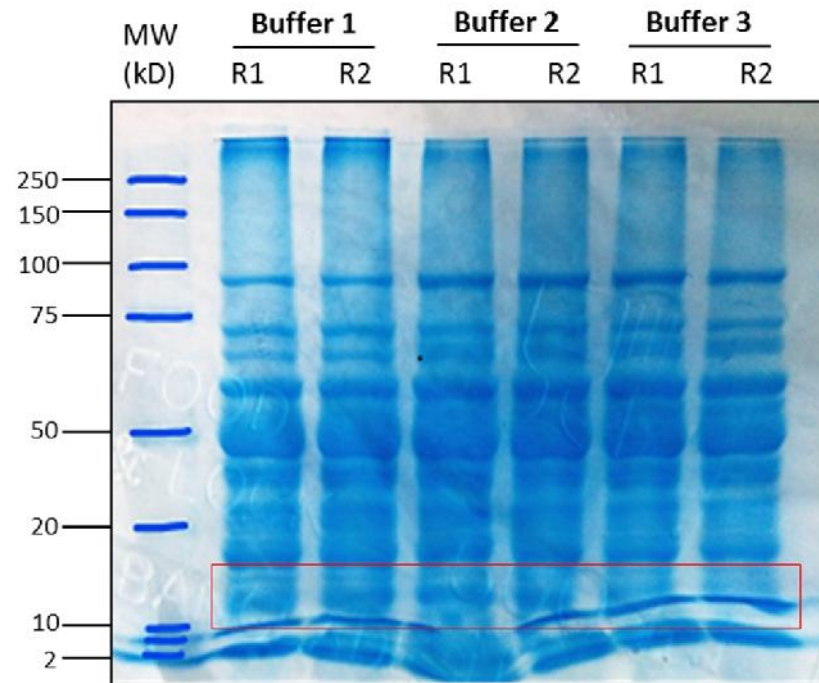
Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.foodchem.2019.06.003>.

References

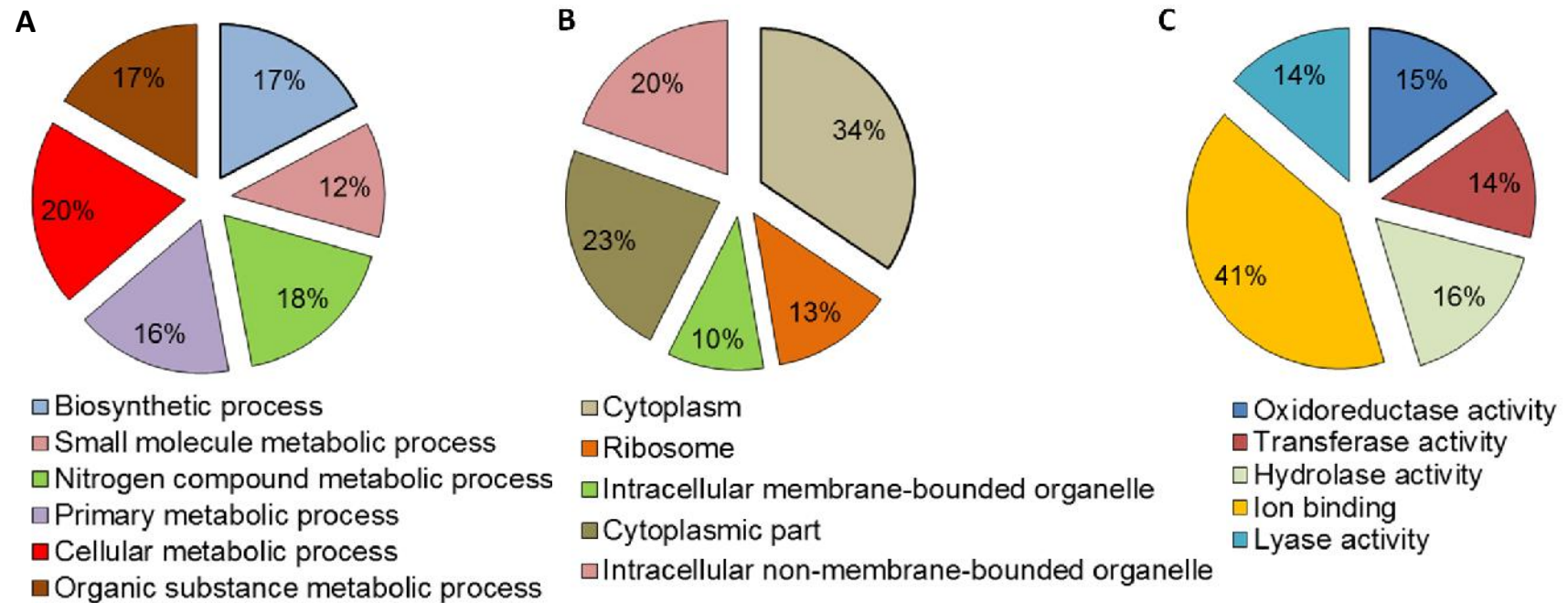
- Ahsan, N., Belmont, J., Chen, Z., Clifton, J. G., & Salomon, A. R. (2017). Highly reproducible improved label-free quantitative analysis of cellular phosphoproteome by optimization of LC-MS/MS gradient and analytical column construction. *Journal of Proteomics*, 165, 69–74.
- Ahsan, N., Rao, R. S. P., Gruppiso, P. A., Ramratnam, B., & Salomon, A. R. (2016). Targeted proteomics: Current status and future perspectives for quantification of food allergens. *Journal of Proteomics*, 143, 15–23.
- Ahsan, N., Wilson, R. S., & Thelen, J. J. (2018). Absolute quantitation of plant proteins. *Current Protocols in Plant Biology*, 3, 1–13.
- Akpınar, O. (2018). Urticaria and anaphylaxis in a child after skin contact of lentil soup. *Indian Journal of Pediatrics*, 85, 575.
- Akkerdaas, J., Finkina, E. I., Balandin, S. V., Santos Magadán, S., Knulst, A., Fernandez-Rivas, M., ... Ovchinnikova, T. V. (2012). Lentil (*Lens culinaris*) lipid transfer protein Len c 3: A novel legume allergen. *International Archives of Allergy and Immunology*, 157, 51–57.
- Asero, R. (2011). Lipid transfer protein cross-reactivity assessed in vivo and in vitro in the office: Pros and cons. *The Journal of Investigational Allergology and Clinical Immunology*, 21, 129–136.
- Basheer-Salimia, R., Camilli, B., Scacchi, S., Noli, E., & Awad, M. (2015). Assessment of genetic diversity in lentils (*Lens culinaris* Medik.) based on SNPs. *Genetics and Molecular Research*, 14, 5870–5878.
- Bogdanov, I. V., Finkina, E. I., Balandin, S. V., Melnikova, D. N., Stukacheva, E. A., & Ovchinnikova, T. V. (2015). Structural and functional characterization of recombinant isoforms of the lentil lipid transfer protein. *Acta Naturae*, 7, 65–73.
- Chen, S., Yang, C. T., & Downs, M. L. (2019). Detection of six commercially processed soy ingredients in an incurred food matrix using parallel reaction monitoring. *Journal of Proteome Research*, 18, 995–1005.
- Choudhury, N., Ahmed, T., Hossain, M. I., Islam, M. M., Sarker, S. A., Zeilani, M., & Clemens, J. D. (2018). Ready-to-use therapeutic food made from locally available food ingredients is well accepted by children having severe acute malnutrition in Bangladesh. *Food and Nutrition Bulletin*, 39, 116–126.
- Codina, R., Arduoso, L., Lockey, R. F., Crisci, C., & Medina, I. (2003). Allergenicity of varieties of soybean. *Allergy*, 58, 1293–1298.
- Cuadrado, C., Cabanillas, B., Pedrosa, M. M., Varela, A., Guillamón, E., Muzquiz, M., ... Burbano, C. (2009). Influence of thermal processing on IgE reactivity to lentil and chickpea proteins. *Molecular Nutrition & Food Research*, 53, 1462–1468.
- Desrochers, N., & Brauer, P. (2001). Legume promotion in counseling: An email survey of dietitians. *Canadian Journal of Dietetic Practice and Research*, 62, 193–198.
- Farag, M. A., Khatatab, A. R., Ehrlich, A., Kropf, M., Heiss, A. G., & Wessjohann, L. A. (2018). Gas chromatography/mass spectrometry-based metabolite profiling of nutrients and antinutrients in eight *Lens* and *Lupinus* seeds (Fabaceae). *Journal of Agricultural and Food Chemistry*, 66, 4267–4280.
- Finkina, E. I., Akkerdaas, J., Balandin, S. V., Santos Magadán, S., Knulst, A., Fernandez-Rivas, M., ... Ovchinnikova, T. V. (2012). Lentil (*Lens culinaris*) lipid transfer protein Len c 3: A novel legume allergen. *International Archives of Allergy and Immunology*, 157, 51–57.
- Finkina, E. I., Balandin, S. V., Serebryakova, M. V., Potapenko, N. A., Tagaev, A. A., & Ovchinnikova, T. V. (2007). Purification and primary structure of novel lipid transfer proteins from germinated lentil (*Lens culinaris*) seeds. *Biochemistry (Moscow)*, 72, 430–438.
- Geng, T., Stojšin, D., Liu, K., Schaalje, B., Postin, C., Ward, J., ... Glenn, K. (2017). Natural variability of allergen levels in conventional soybeans: Assessing variation across North and South America from five production years. *Journal of Agricultural and Food Chemistry*, 65, 463–472.
- Gizatullina, A. K., Finkina, E. I., Mineev, K. S., Melnikova, D. N., Bogdanov, I. V., Telezinskaya, I. N., ... Ovchinnikova, T. V. (2013). Recombinant production and solution structure of lipid transfer protein from lentil *Lens culinaris*. *Biochemical and Biophysical Research Communications*, 439, 427–432.
- Hossain, M. I., Islam, M. M., Wahed, M. A., Khatun, M., & Kabir, I. (2009). Lentil-based high protein diet is comparable to animal-based diet in respect to nitrogen absorption and nitrogen balance in malnourished children recovering from shigellosis. *Asia Pacific Journal of Clinical Nutrition*, 18, 8–14.
- Iallicco, M., Viscosi, V., Arena, S., Scaloni, A., Trupiano, D., Rocco, M., ... Scippa, G. S. (2012). *Lens culinaris* Medik. seed proteome: Analysis to identify landrace markers. *Plant Science*, 197, 1–9.
- Ibanez, M. D., Martinez, M., Sanchez, J. J., & Fernandez-Caldas, E. (2003). Legume cross-reactivity. *Allergologia et immunopathologia*, 31, 151–161.
- Ibáñez Sandín, D., Martínez San Ireneo, M., Marañón Lizana, F., Fernández-Caldas, E., Alonso Lebrero, E., & Laso Borrego, T. (1999). Specific IgE determinations to crude and boiled lentil (*Lens culinaris*) extracts in lentil-sensitive children and controls. *Allergy*, 54, 1209–1214.
- Kumar, R., Kumari, D., Srivastava, P., Khare, V., Fakhr, H., Arora, N., ... Singh, B. P. (2010). Identification of IgE-mediated food allergy and allergens in older children and adults with asthma and allergic rhinitis. *The Indian Journal of Chest Diseases & Allied Sciences*, 52, 217–224.
- Longobardi, F., Innamorato, V., Di Gioia, A., Ventrella, A., Lippolis, V., Logrieco, A. F., ... Agostiano, A. (2017). Geographical origin discrimination of lentils (*Lens culinaris* Medik.) using 1H NMR fingerprinting and multivariate statistical analyses. *Food Chemistry*, 237, 743–748.
- López-Torrejón, G., Salcedo, G., Martín-Esteban, M., Díaz-Perales, A., Pascual, C. Y., & Sánchez-Monge, R. (2003). Len c 1, a major allergen and vicilin from lentil seeds: Protein isolation and cDNA cloning. *Journal of Allergy and Clinical Immunology*, 112, 1208–1215.
- Manso, L., Pastor, C., Pérez-Gordo, M., Cases, B., Sastre, J., & Cuesta-Herranz, J. (2010). Cross-reactivity between coconut and lentil related to a 7S globulin and an 11S globulin. *Allergy*, 65, 1487–1488.
- Martínez San Ireneo, M., Ibáñez Sandín, M. D., & Fernández-Caldas, E. (2000). Hypersensitivity to members of the botanical order Fabales (legumes). *Journal of Investigational Allergology and Clinical Immunology*, 10, 187–199.
- McClain, S., Stevenson, S. E., Brownie, C., Herouet-Guicheney, C., Herman, R. A., Ladics, G. S., ... Thelen, J. J. (2018). Variation in seed allergen content from three varieties of soybean cultivated in nine different locations in Iowa, Illinois, and Indiana. *Frontiers in Plant Science*, 9, 1025.
- Mudryj, A. N., Yu, N., & Aukema, H. M. (2014). Nutritional and health benefits of pulses. *Applied Physiology, Nutrition, and Metabolism*, 39, 1197–1204.
- Pascual, C. Y., Fernandez-Crespo, J., Sanchez-Pastor, S., Padial, M. A., Diaz-Pena, J. M., Martin-Muñoz, F., & Martin-Esteban, M. (1999). Allergy to lentils in Mediterranean pediatric patients. *Journal of Allergy and Clinical Immunology*, 103, 154–158.
- Planque, M., Arnould, T., Delahaut, P., Renard, P., Dieu, M., & Gillard, N. (2019). Development of a strategy for the quantification of food allergens in several food products by mass spectrometry in a routine laboratory. *Food Chemistry*, 274, 35–45.
- Podder, R., Khan, S. M., Tar'an, B., Tyler, R. T., Henry, C. J., Jalal, C., ... Vandenberg, A. (2018). Sensory acceptability of iron-fortified red lentil (*Lens culinaris* Medik.) dal. *Journal of Food Science*, 83, 804–813.
- Sánchez-Monge, R., Pascual, C. Y., Díaz-Perales, A., Fernández-Crespo, J., Martín-Esteban, M., & Salcedo, G. (2000). Isolation and characterization of relevant allergens from boiled lentils. *Journal of Allergy and Clinical Immunology*, 106, 955–961.
- Scippa, G. S., Rocco, M., Iallicco, M., Trupiano, D., Viscosi, V., Di Michele, M., ... Scaloni, A. (2010). The proteome of lentil (*Lens culinaris* Medik.) seeds: Discriminating between landraces. *Electrophoresis*, 31, 497–506.
- Shaheen, N., Islam, S., Munmun, S., Mohiduzzaman, M., & Longvah, T. (2016). Amino acid profiles and digestible indispensable amino acid scores of proteins from the prioritized key foods in Bangladesh. *Food Chemistry*, 213, 83–89.
- Stevenson, S. E., McClain, S., & Thelen, J. J. (2015). Development of an isoform-specific tandem mass spectrometry assay for absolute quantitation of maize lipid transfer proteins. *Journal of Agricultural and Food Chemistry*, 63, 821–828.
- Stevenson, S. E., Woods, C. A., Hong, B., Kong, X., Thelen, J. J., & Ladics, G. S. (2012). Environmental effects on allergen levels in commercially grown non-genetically modified soybeans: Assessing variation across North America. *Frontiers in Plant Science*, 3, 196.
- Vitaliti, G., Morselli, I., Di Stefano, V., Lanzafame, A., La Rosa, M., & Leonardi, S. (2012). Urticaria and anaphylaxis in a child after inhalation of lentil vapours: A case report and literature review. *Italian Journal of Pediatrics*, 38, 71.
- Yang, Y., Liu, G., & Wang, H. (2019). Investigation of the mechanism of conformational alteration in ovalbumin as induced by glycation with different monoses through conventional spectrometry and liquid chromatography high-resolution mass spectrometry. *Journal of Agricultural and Food Chemistry*, 67, 3096–3105.
- Yavuz, S. T., Sahiner, U. M., Buyuktiryaki, B., Tuncer, A., Yilmaz, E. A., Cavkaytar, O., ... Sackesen, C. (2013). Role of specific IgE in predicting the clinical course of lentil allergy in children. *Pediatric Allergy and Immunology*, 24, 382–388.

Supplementary Figure 1



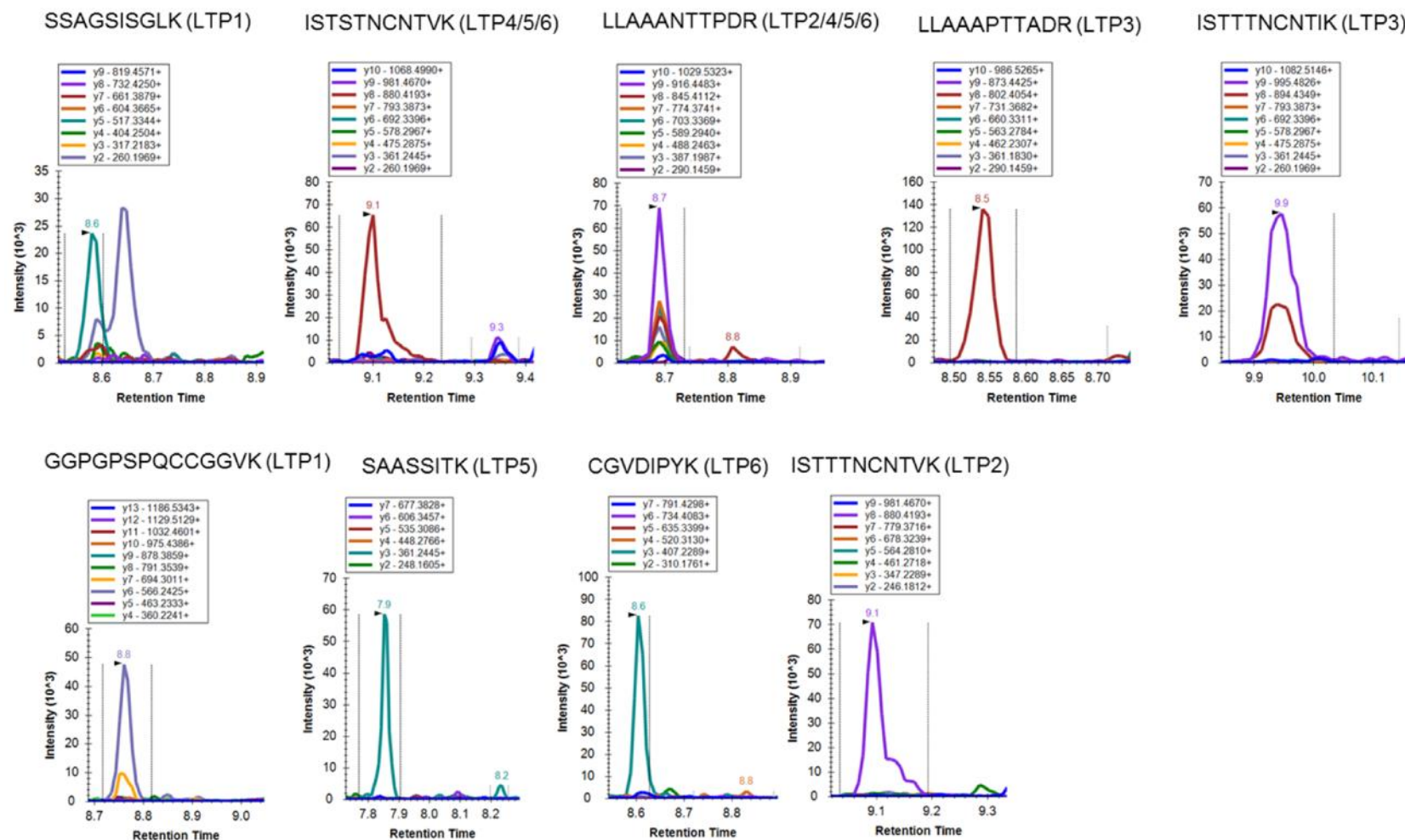
Supplementary Figure 1. SDS-PAGE analysis of lentil proteins. Protein were extracted by a phenol based method described in the materials and method section followed by precipitation with three buffers. Buffer 1, 0.1 M ammonium acetate in methanol; buffer 2, ice cold 75% acetone; and buffer 3, ice cold 10% TCA, 20 mM dithiothreitol (DTT) in acetone. For each buffer condition, two replication of 25 μ g of proteins were loaded in each well.

Supplementary Figure 2



Supplementary Figure 2. GO analysis of the lentil low molecular weight proteome. Identified proteins were categorized according to their gene ontology using Blast2GO program for their (A) biological process, (B) cellular component, and (C) molecular function.

Supplementary Figure 3



Supplementary Figure 3. MRM analysis of unique and overlapped LTPs peptides. Trypsin digested peptides (2 ug) were analyzed by a triple quadrupole mass spectrometer using a MRM platform. The peaks represent the chromatogram of successful identification of some peptides visualized by Skyline software. The transition ions represents by different color peaks in each chromatogram of peptide.

Supplementary Table 1. Genetic background and features of seven lentil cultivars.

Cultivar name	Germplasm	1000 seed weight (g)	Yield (kg/hectare)	Protein content (%)	Crop duration (days)	Year of release	Tolerant /resistance
BARI-Mosur 1	Local	15-16	1700-1800	26-28	105-110	1991	NA*
BARI-Mosur 2	ICARDA	12-13	1500-1700	27-29	105-110	1993	NA
BARI-Mosur 3	Cross between 2 local	22-25	1500-1700	24-26	100-105	1996	Rust and stemphylium
BARI-Mosur4	Crossing material collected from ICARDA and local	18-20	1600-1700	24-26	105-110	1996	Rust and stemphylium
BARI-Mosur 5	Local	19-20	2000-2200	24-26	110-115	2006	Rust, foot rot, aphid
BARI-Mosur 6	ICARDA	22-23	2200-2300	24-26	105-110	2006	Rust, foot rot, aphid
BARI-Mosur 7	ICARDA hybrid	23-25	1800-2300	24-25	110-115	2013	Disease tolerant

*NA: not available/known