Lentil allergens identification and quantification: An update from omics perspective

Oumma Halima\(^a\), Fares Z. Najar\(^b\), Asfia Wahab\(^c\), Sanjeeewa Gamagedara\(^d\), Akibul Islam Chowdhury\(^e\), Steven B. Foster\(^f, g\), Nazma Shaheen\(^h, i\), Nagib Ahsan\(^f, g, *\)

\(^{a}\) Institute of Nutrition and Food Science, University of Dhaka, Dhaka, Bangladesh
\(^{b}\) High-Performance Computing Center (HPCC), Oklahoma State University, Stillwater, OK, USA
\(^{c}\) Department of Biotechnology, University of Oxford, York, UK
\(^{d}\) Department of Chemistry, University of Central Oklahoma, Edmond, OK, USA
\(^{e}\) Department of Food Technology and Nutrition Science, Noakhali Science and Technology University, Noakhali, Bangladesh
\(^{f}\) Department of Chemistry and Biochemistry, The University of Oklahoma, Norman, OK, USA
\(^{g}\) Mass Spectrometry, Proteomics and Metabolomics Core Facility, Stephenson Life Sciences Research Center, The University of Oklahoma, Norman, OK, USA

ARTICLE INFO

Keywords:
- In silico
- Legume
- Mass spectrometry
- Seed proteome

ABSTRACT

Among legumes, the lentil (Lens culinaris) is a major dietary component in many Mediterranean and Asian countries due to its high nutritional value, especially protein. However, allergic reactions triggered by lentil consumption have also been documented in many countries. Complete allergens profiling is critical for better management of lentil food allergies. Earlier studies suggested Len c 1, a 47 kDa vicilin, Len c 2, a seed-specific-biotinylated 66-kDa protein, and Len c 3, low molecular weight lipid transfer proteins (LTPs) were major allergenic proteins in lentils. Recently, mass-spectrometry-based proteomic platforms successfully identified proteins from lentil samples homologous to known plant allergens. Furthermore, in silico analysis using 337 protein sequences revealed lentil allergens that have not previously been identified as potential allergens in lentil. Herein, we discuss the feasibility of omics platforms utilized for lentil allergens profiling and quantification. In addition, we propose some future strategies that might be beneficial for profiling and development of precise assays for lentil allergens and could facilitate identification of the low allergen-containing lentil cultivars.

1. Introduction

Lentil (Lens culinaris) is a legume and a rich source of protein (22–26% dry weight; Longobardi et al., 2017). Lentil is also a good source of vitamins, dietary fiber, and minerals (Ibanez, Martinez, Sanchez, & Fernandez-Caldas, 2003; Longobardi et al., 2017); fat content and trans-fatty acids contents are lower than other legumes, such as chickpeas (Cicer arietinum), cowpeas (Vigna unguiculata), lupins (Lupinus albus), faba bean (Vicia faba), and pigeon bean (Cajanus cajan) (Ene-Obong & Carnovale, 1992; FAO, 2019; Urbano, Porres, Frias, & Vidal-Valverde, 2007). In 2019, around 5.7 million tons of lentils were produced around the world, wherein China, India, Nepal, Bangladesh, Canada, and USA (Fig. 1A and B) were the major lentil producers (FAO, 2019; Rawal & Navarro, 2019).

Although lentil is a common component of the diet in many countries, allergic reactions due to lentil consumption have been reported by many research groups (Leonardi et al., 2014; Sackesen et al., 2020; San Ireneo, Ibanez, Sanchez, Carnes, & Fernandez-Caldas, 2008). Strong IgE-mediated hypersensitivity and immune cross-reactivities have been reported in many populations, especially in Mediterranean and Asian countries (Crespo, Pascual, Burks, Helm, & Esteban, 1995). LTPs are well-known plant allergen proteins, have also been identified in lentils (Akkerdaas et al., 2012; Asero, 2011; E. Finkina et al., 2007; Shaheen et al., 2019). However, it is important to note that allergic reactions vary with race, age, and other factors (Dalal, Binson, Reifen, Ballin, & Somekh, 2002; Mahdavinia et al., 2021; Sasaki et al., 2020; Wollmann et al., 2015).

To date, numerous efforts have been made to identify and quantify lentil allergens. A search on the NCBI website using the keywords “legumes”, “allergen/allergy”, and “proteomics” revealed a total of 3663...
research articles published in the last two decades, wherein lentil allergens (Supplementary Table 1, Fig. 1C) and proteome-based research (Supplementary Table 1, Fig. 1D) are steadily in the bottom compared to other legume species. However, a word cloud of all research papers published on lentils showed that allergen protein identification, characterization, and allergic symptoms from lentil samples were key interests (Fig. 1E). These results showed there is a lack in lentil research, particularly identification and profiling of allergens, which might be related to the incomplete protein sequences. It is important to note that precise protein sequencing, profiling, and quantification of plant allergens is challenging for non-model species using mass-spectrometry due to the availability of a publicly accessible protein sequence database (Carpentier et al., 2008). Standard methods, such as immunoblot assay for identification and quantifications of lentil allergens have been reported (Kumar et al., 2010; Lopez-Torrejon et al., 2003; Fasquel et al., 1999; Shaheen et al., 2019). Enzyme linked-immunosorbent assay (ELISA) has also been widely used for allergens research. However, this method depends on recognition of food allergens by monoclonal or polyclonal antibodies (Vidova & Spacil, 2017). These conventional methods have been successfully applied for the detection of particular allergen and quantification in various biological samples (Koeberl Clarke, & Lopata, 2014). However, there are some limitations with these antibody-based methodologies (Koeberl Clarke, & Lopata, 2014), which might be overcome with peptide-based microarray assay (Sackesen et al., 2020; Vereda et al., 2010).

Proteomic-based platforms combined with liquid chromatography and coupled with tandem mass spectrometry (LC-MS/MS) have gained importance because of their accuracy, multiplexity, identification, and quantification of allergens from biological samples including lentil (Lopez-Pedrouso, Lorenzo, Gagaoua, & Franco, 2020; Scippa et al., 2010). Thus, successful application of mass spectrometry-based proteomics platforms could expand the identification and quantification of lentil allergens and pave the way to future studies. Furthermore, multiple bioinformatics-based tools have been developed that could successfully predict allergens and/or allergen-like proteins from available protein sequences (Garino, Gossin, & Arlorio, 2016; Gupta et al., 2013; Hall & Liceaga, 2021; Maurer-Stroh et al., 2019).

In this review, we discuss the status of lentil allergens and allergen-
like proteins as well as the methods utilized for identification and quantification. Additionally, we performed in silico analysis using 337 lentil protein sequences, which provided a list of potential allergens. We also considered potential research strategies that would help to elucidate and validate detailed information for the lentil allergome as well as help to develop strategies for lentil allergy management through modifications in food systems.

2. Known lentil allergens or allergen-like proteins

According to the plant allergen database (https://www.allergome.org; https://www.allergenonline.org), a total of ten lentil allergen proteins, namely Len c, Len c 1, Len c 1.0101, Len c 1.0102, Len c 1.0103, Len c 2, Len c 2.0101, Len c 3, Len c 3.0101, and Len c Agglutin in were identified and verified (Akkerdaas et al., 2012; Sánchez-Monge et al., 2000). A recent update (February 2021) of the online allergen database (https://www.allergenonline.org/) also reports three allergen protein groups from lentil along with 951 sequences. However, it is important to note that different databases may be highly redundant, and/or missing many of the recently identified allergens. The majority of known allergens in lentil seeds belong to the cupulin, tryp alpha amylo, and lectin legB family (Punta et al., 2012). However, recent studies identified a number of allergens or allergen-like proteins from lentil samples (Table 1).

Based on the available information, lentil allergens can be grouped into four major protein families, namely convicilin, vicilin, LTP, and lectin. As expected, phylogenetic analysis with the same protein group homologs from different members of the Fabaceae family reveals close evolutionary relationships (Supplementary Fig. 1). This result is in agreement with earlier studies that showed plant allergen homologs shared highly conserved sequences (Arora, Kumar, Singh, & Singh, 2020; Verma, Kumar, Das, & Dwivedi, 2013).

Among the lentil allergens, Len c 1 (Len c 1.01), the earliest discovery as a major allergen, is a 50 kDa γ-vicilin chain from the cupulin superfamily (López-Torrejón et al., 2003). Vicilin, also known as 7S globulins, are among the major storage proteins of legumes (Finkina et al., 2017). A total of three genetic isoforms namely Len c 1.0101, Len c 1.0102, and Len c 1.0103 have been identified. Len c 1.02 is a 12- to 16-kd protein of beta-subunit of lentil vicilin protein that may be produced by the posttranslational modification of the precursor Len c 1.01 (López-Torrejón et al., 2003). On the other hand, Len c 2 protein is an IgE-binding component of a 66 kDa seed-specific biotinylated protein. It is isolated from boiled lentils (Bouakkadia et al., 2015; Sánchez-Monge et al., 2000). Len c 3, a lipid transfer protein (LTPs) is a small molecular weight protein (around 9 kDa) that belongs to the superfamily “prolamins.” A subfamily of six lentil LTP isomers designated as Lc- LTP-6 has been identified and verified as immunologically potential allergens using immunoblot analysis and mass spectrometry-based platforms (Akkerdaas et al., 2012; Asero, 2011; Bogdanov et al., 2015; E. I. Finkina et al., 2020; Shaheen et al., 2019). LTPs are considered as one of the main plant allergens that are highly cross-reactive. They can easily bind to different types of lipid molecules such as fatty acids, phospholipids, sterols, galactolipids, and lignins (Shenkarev et al., 2017). However, LTP showed a lower binding affinity to IgE proteins isolated from sera of patients with lentil allergy by heating and digestion processes (E. I. Finkina et al., 2020). Besides Len c, agglutinin (lectin) has been reported in several studies (Bouakkadia et al., 2015). In our recent proteomic-based discovery study, we have identified a total of 44 allergens and allergen-like proteins from lentil, including Len c 1, Len c 2, Len c 3, vicilin, convicilin, legumin, lectins along with their isomers (Shaheen et al., 2019). Label-free quantitative analysis of these allergen proteins revealed that Len c 1, Len c 2, and vicilin are the most abundant allergens, whereas convicilin and Len c 3 are less abundant (Shaheen et al., 2019).

### Table 1

| Name of the allergen/allergen-like proteins | Identification Method | Quant Method | References |
|------------------------------------------|-----------------------|-------------|------------|
| Albumin 2, Allergen Len c 1.0101 partial, Allergen Len c 1.0102 partial, Beta-lathyrin 2 partial, Chain D lentil lectin, Convicilin partial, Convicilin Precursor, CVC partial, Endochitinase, Lectin, Legumin, Legumin A, Legumin A precursor, Legumin A2, Legumin B, Legumin K, Legumin type B alpha chain precursor, LTP 2, LTP 6, Pollen specific peptin methyl ester, Provicilin, Seed albumin 2 partial, Storage protein, Vicilin, Vicilin 14 kDa component, Vicilin 47 K, Vicilin partial, Vicilin precursor, Vicilin type C partial, Len c 3 | Immunoblot, MALDI-TOF | – | (Akkerdaas et al., 2012) |
| Len c 1, Len c 2 | Immunoblot | – | (Sánchez-Monge et al., 2000) |
| Len c 1.0101, Len c 1.0102, Len c 3 | Lymphoblot, ELISA | – | (López-Torrejón et al., 2003) |
| Len c 1 | Peptide microarray immunoblotting | – | (Vereda et al., 2010) |
| Len c 1 | Peptide microarray immunoblotting | – | (Sackesen et al., 2020) |
| Len c 1 | IgE | – | (Armentia et al., 2006) |
| LTP1, LTP2, LTP3, LTP4, LTP5, LTP6, LTP7, LTP8, LTP-1, LTP-3 | Biochemical assay | – | (E. Finkina et al., 2007) |
| Len c 1, Len c 2, Len c 3, Len c Agglutinin | Immunoblotting | – | (Bogdanov et al., 2015) |
| Provicilin, Vicilin, 14 kDa component, F-box protein, ATP-dependent Clp protease, Pyruvate phosphate dikinase regulatory protein 1, Glutathione S-transferase | Immunogold labeling | – | (E. I. Finkina et al., 2020) |
| Len c 3 | Binding Assay | LC-MS/MS | (Bautista-Exposito et al., 2018) |
| Convicilin, Len c 1.0101, Len c 1.0102, Lectin | MALDI-TOF, LC-ESI-LIT-MS/MS | – | (Sciappa et al., 2016) |
3.1. PCR-based methodologies

DNA-based methods such as polymerase chain reaction (PCR), transcriptomic, microarray assay, and single nucleotide polymorphisms (SNPs) have been used for the identification and quantification of allergenic epitopes (Besler, 2001; Monaci & Visconti, 2010; Schmitt, Nesbit, Hurlbut, Cheng, & Malecki, 2010). Identification of allergens using a gene-based platform is an indirect method but fast and reliable (Wang et al., 2011). SNPs are also used for identifying the mutations in genes (Shaheen et al., 2019; Soo, Walczyk, & Smith, 2012). The microarray-based genetic approach has not only been used to detect the legume epitope proteins, but also determine and monitor the disease profile of allergens and to establish multi-allergen tests (Flinterman et al., 2008; Harwaneg et al., 2003; Shreffler, Lencer, Bardina, & Sampson, 2005). Additionally, electrophoresis after the PCR assay is proposed by several authors to evaluate the amplicons (Sena-Torralba, Pallás-Tamarit, Morais, & Maqueira, 2020).

3.2. Conventional Immunology-based approach

The immunological-based approach is the basic technique for searching novel and known allergen proteins from biological samples (González-Buitrago et al., 2007). This platform is often combined with different technologies with one common thread using antibodies that recognize allergen proteins or part of these proteins (epitopes). The combination of immunochemical identification and gel electrophoresis provides a wide range of information about the level of interrogated protein, molecular weight of the protein, and its dispensation between the cellular fragments (Litovchick, 2020). Most of the known lentil allergen proteins such as Len c 1 (López-Torrejón et al., 2003), Len c 2 (Sánchez-Monge et al., 2004), and LTPs (Akkerdaas et al., 2012) have been successfully identified by the immunoblot-based approach. Currently, numerous commercial allergen strips are available as a low-cost test, which provides rapid results, ideal for quick diagnosis of many known allergen proteins (Hamid, Elfedawy, Mohamed, & Mosaad, 2009).

Similarly, ELISA is one of the most popular immunological methods where proteins are directly detected from the allergenic source (Immer & Lacorn, 2015). A recent study successfully utilized the ELISA-based method to demonstrate the allergenicity and allergen compositions of 53 Chinese peanut cultivars (Wu et al., 2016). However, the efficacy of multiple ELISA kits to detect allergen proteins from a wide range of lupines and legumes showed variable sensitivity among the samples, which relates to the cross-reactivity of the sample species (Koeberl, Clarke, & Lopata, 2014). A number of factors can determine the performance of the ELISA methods to detect food allergens (Abbott, 2010), wherein the major challenge of this method is the selection of an ideal buffer for the extraction of proteins from the food samples (Parker, 2015). Nevertheless, the application of polyclonal and/or monoclonal antibodies for detection, identification, and quantification of precise allergen proteins is still divisive (Ascoli & Agerler, 2018).

3.3. Peptide microarray-based immunoassay

Peptide-based microarrays are similar to protein microarrays, but they have been constructed using thousands of pre-synthesized peptides on a microarray surface (Li et al., 2021). Compared to proteins, synthesized peptides are more stable, easy to produce and modify, and more importantly, less expensive. Recently, peptide-based microarrays have been successfully applied for detection and mapping of the IgE epitopes of several food allergens (Flinterman et al., 2008; Han et al., 2016; Kühne et al., 2015; Martínez-Botas & de la Hoz, 2010) including lentilis (Sackesen et al., 2020; Vereda et al., 2010). Using lentil allergic patients’ serum samples, for the first time, Vereda and coworkers identified several IgE-binding sequential epitopes of Len c 1 which are in the C-terminal region (Vereda et al., 2010). Similarly, a recent work by Sackesen and coworkers also used microarray-based immunoassays to detect antibodies such as IgE and IgG4 binding to Len c 1 epitopes in the patients’ sera. The study also revealed that IgE and IgG4 binding to epitopes were significantly higher in the reactive patient than the tolerant one. However, IgE epitopes binding to Len c 1 were not detectable for all lentil allergic patients as it was not the only allergen in lentils (Sackesen et al., 2020). The peptide-based microarray technique required a low amount of human serum sample and showed higher sensitivity compared to the conventional methods. Therefore, it could be adopted as a cost-effective screening method for the identification of other lentil allergen epitopes (Kühne et al., 2015; Li et al., 2021).

3.4. Mass spectrometric-based proteomic methodologies

Recently, liquid chromatography-tandem mass spectrometry (LC-MS/MS)-based proteomic platforms have been successfully utilized and are considered a reliable method to identify and quantify plant allergen proteins (Ahsan, Rao, Gruppuso, Ramratnam, & Salomon, 2016; Gavage et al., 2020; López-Pedrouso et al., 2020; Nakamura & Teshima, 2013; Righetti & Bocchetti, 2016; Stevenson, Houston, & Thelen, 2010; The- len, 2009). In this section, we discuss LC-MS/MS-based proteomics platforms that have been used to identify and quantify lentil allergen proteins.

Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) is a well-established and widely used mass spectrometry-based method in food chemistry. MALDI-TOF has been shown to be suitable for the identification of known food contamination and allergen proteins composition during industrial production (Calvano, Bianco, Losito, & Cataldi, 2021). Earlier studies showed successful identification of known lentil allergen proteins such as Len c1 and Len c3 using immunoblot followed by MALDI-TOF analysis (Akkerdaas et al., 2012; Scippa et al., 2010). However, recent advancement of LC-MS/MS-based methods mainly known as shotgun proteomics or discovery proteomics can be essential for this purpose. This platform provides highly accurate identification and quantification of many known and unknown plant allergens and allergen-like proteins with their posttranslational modified forms (Rahman et al., 2021; Root, Muralidharan, & Lee, 2020; Spiric et al., 2017). Only a handful of gel-based and/or shotgun proteomic analyses have been performed to catalogue the total proteome and allergen proteins from lentil seed samples. Thus far, only a couple hundred lentil seed proteins have been profiled using the gel-based proteomic approach (Akkerdaas et al., 2012; Bouakkadia et al., 2015; Scippa et al., 2010; Shaheen et al., 2019). Most of the earlier proteomic studies successfully identified major lentil allergen proteins such as Len c1, Len c2, Len c3, provicil, vicil, convicilin, and lectin (Akkerdaas et al., 2012; Bouakkadia et al., 2015; Scippa et al., 2010). However, for the first time, in a recent study, we have identified and sequenced the isoforms of LTPs, legumins, and many other allergen-like proteins such as beta-lathyrin 2, albumin 2, endochitinase, pollen-specific pectin methylester from lentil seed using nano-LC systems coupled with high-resolution mass spectrometry (Shaheen et al., 2019). Taken together, these earlier studies suggest that optimization of protein extraction and sample preparation methodologies coupled with an advanced LC-MS/MS pipeline could significantly improve the sequencing depth of lentil seed proteome, thus providing identification and confirmation of many novel allergen proteins.

Targeted proteomics can be performed using technical methodologies including selected reaction monitoring (SRM), multiple reaction monitoring (MRM) and parallel reaction monitoring (PRM), with the latter being one of the new concepts and application of mass spectrometry-based proteomics invented in the last decade (Gillett et al., 2012; Picotti & Aebersold, 2012; Picotti, Bodenmiller, Mueller, Doman, & Aebersold, 2009; Picotti et al., 2010). Among targeted proteomic platforms, absolute quantification (AQUA)-MBR is considered one of the most reliable for accurate protein quantitation from any biological sample (Ahsan, Wilson, & Thelen, 2018; Picotti et al., 2009).
advantage of using the targeted proteomic platform is the specific identification, and quantification of a target set of peptides and/or proteins in an MS analysis with high sensitivity, reproducibility, and quantitative accuracy. Therefore, the application of targeted proteomic pipelines has rapidly increased in the last five years particularly in the field of food industry (Ahsan et al., 2016; Gavage et al., 2020). Targeted proteomic-based assays has been successfully utilized for identification and quantification of many well-known legume allergen proteins (Gavage et al., 2020; Stevenson et al., 2010). Recently, we have established the ideal peptide list of three lentil allergen proteins namely Len C1.0101, Len c 1.0102, and LTP2/4/5/6 for MRM-based assay (Shaheen et al., 2019). One of the main criteria of developing a successful MRM-based assay depends on the reproducibility of the targeted peptides in the actual biological sample (Ahsan et al., 2016). Due to the lack of a large-scale discovery proteomics dataset, the lentil seed proteome data is still fragmentary. Therefore, a great deal of discovery proteomic analysis remains to be done to develop an efficient and precise MRM-based quantitation assay for lentil allergome.

4. In silico analysis for potential lentil allergens

In silico or computational methods are widely used to determine or predict physiochemical properties such as structures and domains of a protein. The probability of a protein being allergenic can also be calculated in silico by using different algorithms and comparing the structures that are known to have allergenic properties (Garino et al., 2016; Maurer-Stroh et al., 2019). Although, in silico methodologies are incapable of distinguishing between the sensitization and elicitation phases of allergies, the FAO/WHO and the European Food Safety Authority (EFSA) guidelines state that a query protein is potentially allergenic if it has a sequence identity of >35% identical over a frame of 80 amino acids or matched exactly with 6 to 8 amino acids with an allergen (Maurer-Stroh et al., 2019).

Although, in silico analysis has been successfully employed to determine the potential allergenicity of many plant proteins (Garino et al., 2016; Kulkarni, Ananthanarayan, & Raman, 2013), to the best of our knowledge, no large-scale in silico analysis has been conducted using the lentil protein database. Therefore, it is imperative to investigate a genome/proteome-based search for potential allergen-like proteins. While the complete lentil genome sequence is not publicly available, we extracted a total of 337 lentil protein sequences from the UniProt database (Taxonomic ID: 3864). These protein sequences were further subjected to AllerCatPro, an online-based pipeline to determine the allergenicity potential (Maurer-Stroh et al., 2019). Out of 337 lentil proteins analyzed, a total of 22 and 19 proteins were categorized as strong and weak allergens, respectively (Fig. 2, Supplementary Table 3). As expected, well-characterized lentil allergens such as LTPs and Len c1 were identified with strong evidence (Fig. 2A). Two pathogenesis-related (PR) proteins, namely PR4 and PR-10b, and a disease-resistance response protein 49 (DRR49) were also identified as
potential new allergens with strong evidence (Fig. 2A), while PR-10a was identified with weak evidence (Fig. 2B). Among the PR proteins, PR-10b and DRR49 showed > 75% sequence similarity, >94% identity with 3D epitope, and matched with 250 known allergens (Supplementary Table 2). A phylogenetic analysis of lentil PR proteins with 250 allergens hits further revealed close evolutionary relationships of PR10b and DRR49 with soybean (Glycine max), lupin (Lupinus luteus), mung bean (Vigna radiata), and peanut (Arachis hypogaea) PR proteins (Fig. 3A). Among the 250 allergens, soybean Gly m 4 was the top hit against these PR-10b and DRR49 proteins (Supplementary Table 3).

Amino acid sequence analysis of PR10b and DRR49 with other soybean, lupin, mung bean, and peanut PR proteins showed the epitope binding residues and region are highly conserved (Fig. 3B). It is also interesting to find out that the epitopes containing tryptic peptides of lentil PR10b and DRR49 are slightly different, suggesting a successful MRM-based targeted proteomic assay could be developed for these two potential lentil allergens (Ahsan et al., 2016).

Plant PR proteins are a known class I allergen showing strong IgE cross-reactivity (Arora et al., 2020). It is important to note that, PR proteins have never been claimed or identified as potential allergens for lentils. Nevertheless, the allergenic activity of potential allergens identified by in silico analyses need to be verified using further experimental strategies. Taken together, these results suggest that in silico analysis of the complete lentil proteome would be a potential strategy to discover many putative allergens which could help to assess the complete list of lentil allergens. Thus, food safety regulations could be evaluated more efficiently for lentils.

5. Future omics-based strategies for identification and quantification of lentil allergens

Minimal changes of allergen protein levels in food can trigger allergic reaction, a potential life-threatening condition for allergic patients (Ballmer-Weber et al., 2015). Early allergen detection and quantification methodologies improve our understanding of well-known lentil allergens. However, these conventional methodologies are unable to predict, detect, and quantify the isoform-specific allergens, multiple types of post-translational modifications in allergens, and/or less characterized allergens (Koeberl, Clarke, & Lopata, 2014). More importantly, conventional methodologies are incapable of being performed as a multiplexed system (Ahsan et al., 2016). Thus, there is great demand for developing multiplexed methodologies to identify and quantify the allergen proteins in the fresh food industry. Herein, we propose some future strategies that could be beneficial for the profiling and quantification of lentil allergome.

1. In silico analysis using the complete lentil genome-derived amino acid sequences would be significantly beneficial for predicting the overall lentil allergome.

2. Developing a mass-spectrometry based allergen detection and quantification assay is a key step in profiling the complete allergome (allergen and allergen likes proteins) of lentil seed. It is important to note that, earlier lentil proteomic studies (Akkerdaas et al., 2012; Bouakkadia et al., 2015; Scippa et al., 2010; Shaheen et al., 2019) using gel-based proteomics platforms identified a total of 200 proteins. Surely this number is a small portion of the lentil seed proteome. However, complementary proteomic platforms such as a combination of gel-based and gel-free proteomic platforms coupled with various protein extraction methodologies offers extensive sequencing depth of seed proteome and allergome (Bose et al., 2019; Natarajan, Xu, Caperna, & Garrett, 2005; Romero-Rodríguez, Maldonado-Alconada, Valledor, & Jorrín-Novo, 2014). We have demonstrated that gel-free shotgun proteomics analysis of lentil seed samples significantly increased the sequencing depth of the lentil seed proteome (unpublished). While it has already been demonstrated that the choice of protein extraction buffer enriches different protein functional classes from seed samples, systematic optimization of protein extraction is needed for in-depth profiling of the lentil seed proteome.

3. A comprehensive proteome dataset combined with a thorough search against all annotated genome is necessary to identify the complete translational product (proteome) in non-model plant species (Al-Mohanna, Bokros, Ahsan, Popescu, 2020). Thus, a proteogenomics approach would provide a precise annotation of the lentil allergen proteins.
4. Systemic proteomic screening of thousands of lentil cultivars may identify lentil cultivars with low allergenicity.

5. Most of the plant allergen proteins are seed storage and defense-related proteins, therefore, the effect of geographical variation and cultural practices on lentil allergome should also be investigated.

6. Processing steps could significantly influence the allergenicity of many legume seed proteins including lentils (Beyer et al., 2001; Cabanillas, Jappe, & Novak, 2018; Cuadrado et al., 2009; Mondoulet et al., 2005). Therefore, analyses of complete lentil allergome profiles in response to cooking methods (boiling, frying etc.) would be a beneficial strategy to minimize the allergic reaction to the sensitive population.

6. Conclusion

The in silico analysis revealed that the complete profiling of lentil allergome is still in the juvenile stage due to the lack of a publicly available genome sequence. Precise detection, quantification, and meaningful labeling of lentil allergen proteins could significantly mitigate the allergenic reaction of these sensitive populations and thereby, could offer better consumer safety. Thus, various omics platforms such as in silico analysis, peptide-based microarray, and mass spectrometry-based proteomics could significantly contribute to the identification and quantification of potential lentil allergens.

Author contributions

Conceptualization, N.A; data analysis and bioinformatics, N.A, F.Z. N, A.W, O.H; writing, review, and editing; N.A, O.H, N.S, S.G, A.I.C, F.Z. All authors have read and agreed to the published version of the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

N.A., and S.B.F. gratefully acknowledge the initial funding support from the OU VPRF Office for the establishment of the Proteomics Core Facility. N.S and O.H. profoundly grateful to the HEQEP project of University Grants Commission (UGC), Bangladesh (CP-3270) and Advanced Research in Education (GARE), Bangladesh Bureau of Educational Information & Statistics (BANBIES), Ministry of Education, Government of the People’s Republic of Bangladesh for supporting the establishment of laboratory facilities for the analysis of food allergens in INFS.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochms.2022.100109.

References

Abbott, M. (2010). Validation procedures for quantitative food allergen ELISA methods: Community guidance and best practices. Journal of AOAC International, 93(3), 442–450.

Ahsan, N., Wilson, R. S., & Thelen, J. J. (2018). Absolute quantitation of plant proteins. Current Protocols in Plant Biology, 9(1), 1–13.

Akkerdaas, J., Finkina, E., Balandin, S., Magdans, S. S., Knutst, A., Fernandez-Rivas, M., & Ovchinnikova, T. (2012). Lentil (Lens culinaris) lipid transfer protein Len c 3: A novel legume allergen. International Archives of Allergy and Immunology, 157(1), 51–57.
Wang, W., Han, J., Wu, Y., Yuan, F., Chen, Y., & Ge, Y. (2011). Simultaneous detection of eight food allergens using optical thin film biosensor chips. *Journal of Agricultural and Food Chemistry, 59*(13), 6889–6894.

Wollmann, E., Hamsten, C., Sibanda, E., Ochome, M., Focke-Tejkl, M., Asarnoj, A., … Lupinek, C. (2015). Natural clinical tolerance to peanut in African patients is caused by poor allergenic activity of peanut IgE. *Allergy, 70*(6), 638–652.

Wu, Z., Zhou, N., Xiong, F., Li, X., Yang, A., Tong, P., … Chen, H. (2016). Allergen composition analysis and allergenicity assessment of Chinese peanut cultivars. *Food Chemistry, 196*, 459-465.