



A label-free shotgun proteomics analysis of macadamia nut

Johanna Rost, Sridevi Muralidharan*, N. Alice Lee*

ARC Training Centre for Advanced Technology in Food Manufacture, School of Chemical Engineering, University of New South Wales, Kensington, NSW 2052, Australia



ARTICLE INFO

Keywords:

Macadamia nut soluble proteome
Protein identification
LC-MS/MS
In-silico analysis
Potential allergen
Cross-reactivity
AllergenOnline

ABSTRACT

In this study, we present a systematic proteomic overview of macadamia nut using a label-free shotgun proteomic approach. We identified 947 proteins in 723 clusters and gene ontology analysis revealed proteins across 46 functional categories including carbohydrate metabolism (10%), protein metabolic processes (5%), amino acid metabolism (4%), transport (4%), stress response (3%), lipid metabolism (3%), protein folding (3%) and defense response (1.4%). The defense response proteins accounted for 24% of the total peptide abundance. The vicilin-like macadamia antimicrobial peptides 2–3 (MiAMP2) was the most abundant protein, followed by glyceraldehyde-3-phosphate dehydrogenase 3, 11S legumin-like protein, 2-phospho-D-glycerate hydrolase and heat shock 70 kDa protein among others. The cascading of amino acid and carbohydrate metabolic pathways in macadamia nut were constructed against reference maps from KEGG and proposed for the first time. Results were also indicative of useful protein candidates with possible allergenic potential and cross-reactivity in macadamia nut. The *in-silico* analysis revealed homology and linear epitope similarities to known allergens such as conglutin β allergen from lupin, Jug r2 vicilin allergens from walnut, Ara h3 11S globulin from peanut, small rubber particle protein Hev b3, hevein, enolase 2, HSP 70kDa Cla h4, Der f28 allergen, and methylglyoxalases. Label-free shotgun proteomics reveal valuable insights into the genetic and biological makeup of macadamia nut proteome and provide guidance on protein candidates with allergenic potential for further immunological investigation. Data are available via ProteomeXchange with identifier PXD015364.

1. Introduction

Macadamia nut is a native Australian tree nut and has gained increasing dietary popularity worldwide due to its beneficial properties for health. In Australia, commercial cultivars are derived from *Macadamia integrifolia* and hybrids of *M. integrifolia* and *M. tetraphylla*, where *M. integrifolia* was the main progenitor species used for the development of cultivars during initial domestication. Genetic diversity is currently lacking in macadamia (Nock et al., 2019), because the global macadamia population goes back to just three generations of wild cultivars arising from a narrow seed pool. Studies on characterizing the macadamia nut genome and proteome has only recently begun (Nock et al., 2016) and are notably under-represented in the literature compared to almonds (Li, Geng, Wang, Lu, & Ma, 2016), soybean (Xu et al., 2015), peanuts (Liu et al., 2018; Mamone et al., 2019), lupin (Cabello-Hurtado et al., 2016) and other tree nuts (Smeekens, Bagley, & Kulis, 2018). This paucity of genomic and proteomics studies in macadamia over other tree nuts limit macadamia breeding and crop improvement

efforts. Notably ultra-high throughput diversity array technology-based polymorphism marker investigations among a wide selection of macadamia root stocks and cultivars have successfully established genotype identities and their historical genetic differentiation across cultivars (Alam, Neal, O'Connor, Kilian, & Topp, 2018). More such research is needed to make breeding efficiency and improved genetic diversity feasible for macadamia breeders.

Macadamia nut is speculated to contain proteins similar to the dominant classes in tree nuts, such as the nutrient reservoir proteins or the storage proteins including the cupin and prolamin superfamily as well as the lipid transfer proteins and profilins, typically found in higher amounts in seeds for nourishment of the plant during germination and first growth (Mills & Shewry, 2008). Key protein characterization studies in macadamia nut include the macadamia antimicrobial peptides – 1 (MiAMP1) (Marcus, Goulter, Green, Harrison, & Manners, 1997; McManus et al., 1999), and 2 (MiAMP2) (Marcus, Green, Goulter, & Manners, 1999) and MiAMP2 was reported being a part of a 7S globulin. MiAMP proteins have been found to possess antimicrobial

Abbreviations: MiAMP1, Macadamia antimicrobial peptide 1; MiAMP2, Macadamia antimicrobial peptide 2; IEDB-AR, Immune Epitope Database Analysis Resource; NCBI, National Centre for Biotechnology Information; MS/MS, Tandem mass spectrometry; LTQ-FT, Linear ion Trap with a Fourier Transform ion cyclotron resonance mass spectrometer; BLAST, Basic Local Alignment Search Tool; FARRP, Food Allergy Research and Resource Program; GO, Gene Ontology

* Corresponding authors.

E-mail addresses: s.muralidharan@unsw.edu.au (S. Muralidharan), alice.lee@unsw.edu.au (N.A. Lee).

<https://doi.org/10.1016/j.foodres.2019.108838>

Received 28 May 2019; Received in revised form 12 November 2019; Accepted 18 November 2019

Available online 30 November 2019

0963-9969/© 2019 Elsevier Ltd. All rights reserved.

activities to a wide spectrum of bacterial and fungal phytopathogens. Additionally diacylglycerol acyltransferases (Arroyo-Caro, Manas-Fernandez, Alonso, & Garcia-Maroto, 2016), acyl-ACP thioesterases and desaturases (Moreno et al., 2011; Rodriguez, Sanchez-Garcia, Salas, Garces, & Martinez-Force, 2015) have been genes of interest for macadamia nut oil characterization in an effort to elucidate its lipid compositions and mechanism of oil accumulation. Several chloroplast protein coding genes in macadamia present significant number of SNPs as a result of evolution, however this high variability is in non-coding regions where gene function remained unaffected. They include NADPH oxidoreductase subunits 1, 4, and 5, DNA directed RNA polymerase subunits α and β , maturase k, 30 s ribosomal protein s16, ATP synthase α subunit and chlorophyll *a* apoproteins A1 and A2 (Nock et al., 2019). The authors regard positive selection of these genes may have conferred climatic adaptability in macadamia.

Further, macadamia nuts also pose a significant allergy risk to consumers who are hypersensitive to them. Compared to other tree nuts, the prevalence of macadamia allergy is still low. Several clinical cases on macadamia allergy have been reported with anaphylaxis reactions or severe symptoms after consuming macadamia nut (De Knop, Hagendorens, Bridts, Stevens, & Ebo, 2010; Ekbote, Hayman, & Bansal, 2010), but present little to no information on corresponding allergenic proteins. Compared to the depth of information on protein identifications and characterizations reported for other allergic tree nuts, macadamia nut still lack momentum (Geiselhart, Hoffmann-Sommergruber, & Bublin, 2018). So far low IgE cross-reactivity of a 17 kDa macadamia nut protein with hazelnut has been reported (Sutherland, O'Hehir, Czarny, & Suphioglu, 1999) and a 12 kDa protein from macadamia was found to show IgE reactivity and warrant further characterization (Herbst, Wahl, & Frosch, 2010). Although the gold standard of food allergy diagnosis is the DBPCFC, not all patients are able to undertake the food challenge or are willing to and this can only be performed in accredited clinics that have the expertise to handle anaphylactic emergency. Largely, the first step in any general practice is the skin prick test followed by a specific serum IgE test. Diagnosis by skin prick test and in vitro serum IgE tests of macadamia allergy is challenged by the lack of commercially available standardized macadamia nut protein extracts and clinicians tend to use whole nut or crushed nut for these tests. This is further complicated by the fact that most hybrid cultivars currently available share the same maternal lineage although the paternal partner of nuts is suspected to have been different (Nock et al., 2019).

These phylogenetic variations more likely translate from proteomic differences in the nut, and may in turn present different isoforms of allergens for macadamia that also require screening. Therefore it is necessary to characterize macadamia nut protein profile and identify proteins posing possible allergenic potential because macadamia nut products continue to be implicated in recent food-related deaths in Victoria, Australia. (McGinn, 2016). A recent study demonstrated screening of low-allergenic foods using the AllerCatPro tool (Maurer-Stroh et al., 2019), where the top abundant proteins resulting in matches to no known allergens, were considered as having low-allergenic potential (Krutz et al., 2019). As increasing number of novel plant-based products and foods become available in the commercial market, *in-silico* based screening of novel or new proteins with allergenic potential by amino acid and structural similarity in commodities previously not characterized for potential allergenic assessment, such as our paper, would become useful.

The primary focus of this paper is, therefore, providing an overview of the macadamia nut proteome by identifying and understanding of macadamia proteins and their function. For this, we used the label-free shotgun proteomics approach in this study. To understand the functional ontology of its proteins, the macadamia nut soluble proteome was processed into functional biological categories revealing varied protein and peptide distribution. The current challenge is lack of well-curated macadamia allergome that allows for the identification and

characterization of macadamia nut allergens. Therefore, as the secondary focus, the identified proteins were further processed using *in-silico* analyses to identify protein sequence similarities to known allergens to select potential allergen or cross-reactive candidates for future characterization.

2. Materials and methods

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD015364.

2.1. Extraction of soluble macadamia nut proteins

Macadamia samples (Macadamia Processing Company Ltd, Alphadale, NSW, Australia) represented a mixture of commercial cultivars (from *M. integrifolia* and hybrids of *M. integrifolia* and *M. tetraphylla*). Raw macadamia nuts were ground using a lab-grinder (IKA®-Werke GmbH & CO) and further finely ground with a mortar and pestle using liquid nitrogen. Defatting and protein extraction workflow employed in this study is typical for nuts and especially suited for allergen protein extraction (Walczyk, Smith, Tovey, & Roberts, 2017). Method used in our study was extensively optimised in-house to obtain maximum protein extraction. The ground macadamia nut was defatted thrice with a mixture of *n*-hexane and ethanol (4:1, v/v) with the sample to solvent ratio of 1:20 for 1 h. For the protein extraction, 0.3 g of defatted flour and 1 M sodium chloride (pH 8.4 adjusted with NaOH) were mixed in a sample to buffer ratio of 1:20 for the first extraction and 1:10 for the second extraction at 4 °C for 90 min on a rotary shaker. The sample was centrifuged at 13,000g at 4 °C for 20 min and the supernatant was filtered through 0.45 μ m syringe filters. The protein was precipitated with cold acetone overnight at -20 °C. The pellet was recovered by centrifuging at 17,000g for 5 min and washed three times with ice cold acetone. The acetone residue was removed by flushing nitrogen gas and the dried pellet was resuspended in 20 mM Tris-HCl buffer (pH 8.0). The protein content was determined by Bicinchoninic acid microplate assay (Sigma-Aldrich, USA) according to manufacturer's protocol and read at 562 nm using SPECTRAMax M2 Softmax Pro (Molecular Devices, LLC).

2.2. SDS-PAGE

A sample protein (50 μ g) was mixed with 2 $\times$ Laemmli sample buffer (Bio-Rad, USA) containing 3% DTT, heated for 5 min at 95 °C, and then centrifuged at 10,000g for 3 min. Protein samples were subjected to gel electrophoresis on 15% Tris-glycine gel using a mini-PROTEAN Tetra cell system (Bio-Rad, USA). The gel images were acquired using the ChemiDoc™ MP System equipped with the Image Lab™ Software (Bio-Rad, USA). All experiments were conducted in biological triplicates of raw macadamia nuts sourced from a batch of commercial macadamia nuts.

2.3. Sample preparation for shotgun proteomics

Peptide samples for shotgun proteomics were prepared and digested in similar detail to the methodology previously published for plant proteins (Neilson et al., 2014). Proteins from each sample separated and visualized on the SDS-PAGE gel were further sliced and diced into 16 fractions. The stained gel pieces were first equilibrated with 50 μ L of 100 mM NH_4HCO_3 to adjust the pH to 8 and then thoroughly destained by washing with 200 μ L 50% acetonitrile/50 mM NH_4HCO_3 . The clear gel pieces were dehydrated with 100% acetonitrile and protein fractions were then reduced with 50 μ L of 10 mM DTT in 100 mM NH_4HCO_3 at 37 °C for 30 min. They were alkylated with 50 μ L of 55 mM iodoacetamide in 100 mM NH_4HCO_3 at room temperature for 30 min in the dark. The gel pieces were washed once with 100 μ L of

100 mM NH_4HCO_3 for 5 min, twice with 200 μL 50% acetonitrile/50 mM NH_4HCO_3 for 5 min each and then dehydrated again with 100% acetonitrile. Following this, the protein fractions were subjected to in-gel tryptic digestion with trypsin (12.5 ng/ μL sequencing grade, Promega Inc USA) incubated for 30 min at 4 °C initially, and then topped up with 50 μL of 50 mM NH_4HCO_3 and incubated at 37 °C overnight. After digestion, peptides from each protein fraction were extracted by adding 50 μL of 50% acetonitrile/2% formic acid to the gel pieces, vortexed and then incubated for 30 min. This extraction was repeated twice and the corresponding fractions were pooled. The peptide digests were dried with a vacuum centrifuge and reconstituted into 5 μL of 0.1% formic acid prior to mass spectrometry.

2.4. Nano-LC-MS/MS

An LTQ-FT Ultra mass spectrometer (Thermo Scientific, USA) coupled with Ultimate 3000 nano-LC and Dionex autosampler (Thermo Scientific, USA) was used to analyze the macadamia nut peptide digests. The samples were loaded onto a micro C18 pre-column (500 $\mu\text{m} \times 2$ mm, Michrom Bioresources Inc, USA) with buffer A (98% water, 2% acetonitrile, 0.1% formic acid) at 10 $\mu\text{L min}^{-1}$. After a 4 min wash, the pre-column was switched into line using Valco 10 port valve with a fritless nanocolumn (75 $\mu\text{m i.d.} \times 10$ cm) containing reverse phase C18 media (3 μm , 200 Å, Michrom Bioresources Inc, USA). Peptides were eluted using a linear gradient of buffer A (98% water, 2% acetic acid, 0.1% formic acid) to buffer B (98% acetonitrile, 2% water, 0.1% formic acid) at 250 $\mu\text{L min}^{-1}$ over 30 min. The column tip was positioned ~0.5 cm from the heated capillary ($T = 200$ °C) of the LTQ-FT, at 1800 V applied to a low volume tee junction. The instrument was operated in data-dependent acquisition mode, with positive ions generated by nanospray ionization. Automatic peak recognition, dynamic exclusion and MS/MS of the top 10 most intense ions at collision induced dissociation was performed where survey scan was acquired in m/z range 195–2000.

2.5. Database search for macadamia protein identification

All MS/MS spectra were analyzed using Mascot (Matrix Science) and X! Tandem (the Global Proteome Machine; thegpm.org version CYCLONE) and searched against the NCBI non-redundant protein database (selected for *Viridiplantae*, 1,530,227 entries) with trypsin as the proteolytic enzyme. Mascot and X! Tandem searches were performed with a fragment ion mass tolerance of 0.40 Da and a parent ion tolerance of 2.0 Da. Carbamidomethylation of cysteine was accounted as complete modification and oxidation of methionine was accounted as variable modifications. Scaffold (version Scaffold_4.5.3, Proteome Software Inc.) was used to validate MS/MS based peptide and protein identifications. Protein Prophet algorithm (Nesvizhskii, Keller, Kolker, & Aebersold, 2003) and Peptide Prophet algorithm (Keller, Nesvizhskii, Kolker, & Aebersold, 2002) were used to filter protein and peptide identifications at $\geq 80\%$ and 90% threshold probabilities respectively, such that FDRs estimations were less than 1%, with at least 1 peptide as the minimum threshold for protein identifications. Proteins sharing significant peptide evidence were grouped into clusters based on shared peptide evidence. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Reproducible proteins were analyzed for total peptide counts, unique peptide counts, sequence matches, and percentage coverage. Proteins were annotated with Gene Ontology (GO) in terms of biological process information at levels 3 and 4 in the hierarchy of GO annotation sourced from geneontology.org, NCBI, UniProt, InterPro and KEGG databases collectively. For cataloguing the overall protein profile of macadamia nut, 947 protein identifications that satisfy the protein and peptide FDR thresholds in at least one replicate were filtered and binned into 46 functional categories. Identifiers with no protein annotation information were subjected to

protein sequence alignment using Basic Local Alignment Search Tool (BLAST) against all proteins database at NCBI to retrieve annotation information for matches with $> 90\%$ identity and expect value < 0.05 .

2.6. Database search for macadamia allergen matching

The AllergenOnline database (version 16) (Goodman et al., 2016) provided by Food Allergy Research and Resource Program (FARRP) was used to compare macadamia nut proteins identified with the NCBI nr shotgun proteomics identifications, against protein sequences of publicly known allergens. The allergen-specific database contained 1956 protein sequences which belonged to 778 protein groups. Database searches were carried out with expect value cut off ≤ 0.05 and used BLOSUM50 as the scoring matrix. Only the protein identifications with a minimum of two spectrum counts in at least two of the three biological replicates were used for the screening. Macadamia proteins that matched to known allergenic proteins were filtered based on their % shared peptide identity (Jamakhani, Lele, & Rekadwad, 2018; Krutz et al., 2019).

2.7. Database search for the epitope matching

Immune Epitope Database Analysis Resource (IEDB-AR; <http://www.iedb.org/>) was used to BLAST search the above-identified protein sequences of the allergens to match with the FARRP database for epitopes. Due to the laborious nature of this search, this was limited to linear epitopes matches only. Substring BLAST searches were conducted for positive assays, T-Cell, B-cell and MHC ligand assays with a linear structure of the epitopes from *Homo sapiens* and were restricted to allergic diseases. Linear epitopes reveal the allergenicity of a protein significantly as they are less affected by processing and digestion conditions (Sathe, Teuber, & Roux, 2005).

3. Results and discussion

3.1. Overview of macadamia protein profile based on shotgun proteomics

Macadamia nut soluble proteins presented a distinctive profile upon electrophoresis indicative of proteins between the range 75–5 kDa (Fig. 1). Qualitative and quantitative analysis of proteins and peptides based on spectral counting were performed and protein distributions were studied to profile the macadamia nut proteome. The mass spectrometric shotgun analysis identified 1079 proteins based on their tryptic peptide evidence for macadamia nut as listed in Table S1 in the [Supplementary Material](#). Of the 1079 identified proteins, a total of 947 proteins in 723 clusters met 0.9% protein false discovery rate and 0.2% peptide false discovery rate and corresponded to 80% protein and 93% peptide confidence threshold respectively. The observed proteins correspond to the merged total peptide counts of 16 peptide fractions across biological triplicates. Studying the proteomics distribution revealed vital clues about both the genetic and biological makeup of macadamia and helps to establish corroboration for its actual phenotype.

3.2. Qualitative and quantitative distribution of proteins in macadamia

The qualitative distribution pattern of the identified proteins in terms of the number of proteins present in each functional category (Fig. S1 in the [Supplementary Material](#)) as well as their quantitative distribution based on relative protein abundance in terms of peptide counts using spectral counting approach within each functional category were studied (Fig. S2 in the [Supplementary Material](#)). For this analysis, we assigned protein functional categorization to each protein on the basis of GO annotation in terms of biological process (levels 3–4) information, and classified into 46 functional categories.

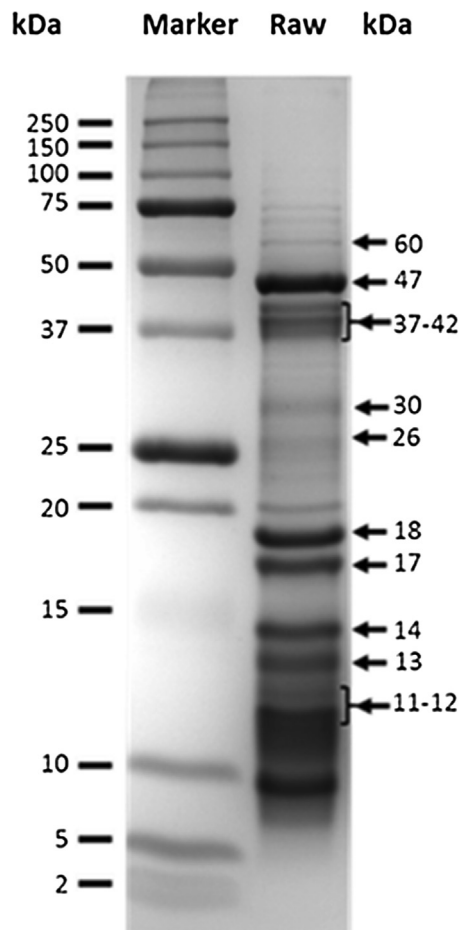


Fig. 1. Soluble proteins in raw macadamia nut extract, 15% SDS-PAGE.

Protein content in macadamia is under 10% of nut composition when compared to that in almonds (21%) (Li et al., 2016) or peanuts (30%) (Koppelman et al., 2016), but macadamia was found to have a similar functional protein profile. Carbohydrate metabolic proteins were the largest annotated class (10%) in terms of the distribution of total number of proteins in macadamia nut, followed by translation (8%), protein metabolic processes (5%), transport (4%), amino acid metabolism (4%), and 3% each by lipid metabolic processes, protein folding and stress response proteins among others. Proteins belonging to oxidation-reduction processes, metabolic process, proteolysis, transcription, secondary metabolite metabolism, and seed storage proteins were about 2% each of total identified proteins. Defense response proteins and the 30-remaining biological process functional categories each represented $\leq 1.5\%$ of the total number of proteins.

Conversely, the distribution pattern of protein abundance was not the same as the observed protein counts. Greatest protein abundances identified in macadamia nut were the defense response proteins quantified at 24% of the total abundance by peptide counts (Fig. S2 in the Supplementary Material). The second most abundant category was the carbohydrate metabolic proteins (15%), followed by seed storage proteins (8%), stress response (5%), translation (4%), lipid (3%) and protein metabolic processes (4%). Proteins belonging to functional categories namely iron homeostasis, oxidation-reduction processes, amino acid metabolism, signal transduction, seed development, DNA metabolism, secondary metabolite metabolic proteins had similar distribution patterns in terms of protein count but were less in terms of abundance. Given the fat content of the nut is between 70 and 80%, only 3% of the total proteins identified were adequately annotated as lipid metabolic proteins. Lipids are considered to be a major source of energy and nutrient resource in macadamia. Membrane proteins and

proteins with lipid anchors remain largely unknown in macadamia and are likely to contribute to the lipid storage and transport of molecules and proteins across the large membrane structures of the cell. In almonds, about 22% of identified proteins belonged to the membrane components of the cell (Li et al., 2016).

3.3. Defense response proteins

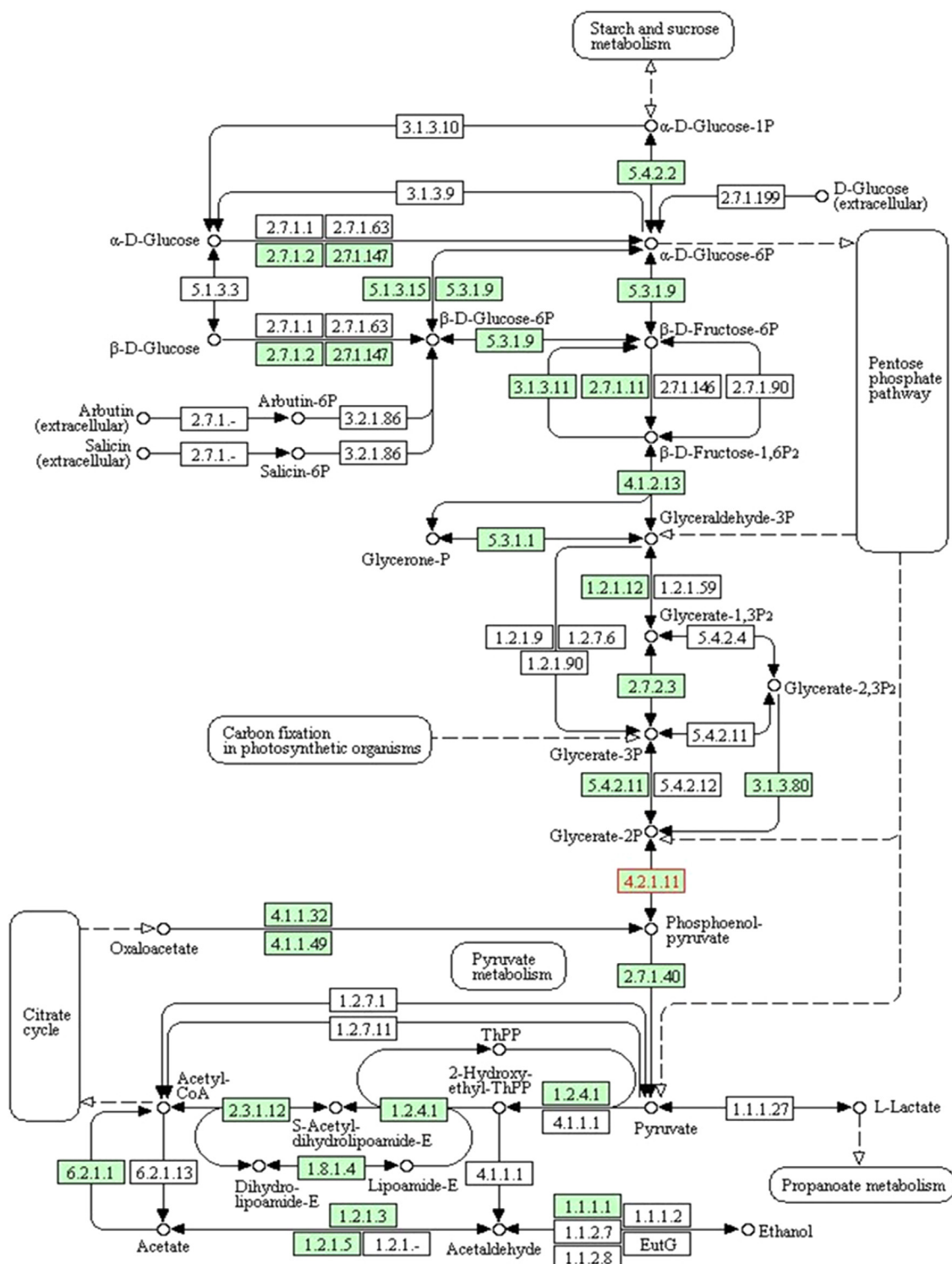
The defense response proteins accounted for 24% of protein abundance. Of this, 94% was chiefly contributed by MiAMP2 vicilin-like antimicrobial peptides 2–3 and our study is the first to recognize this protein as the most abundant protein in macadamia. Their greatest abundances seen here is evidence at the peptide level and could be directly correlated to the biological and genetic features of macadamia, vital for seed germination and tree development and were also indicative of the nut's vital phenotypic traits, and allergenic content (discussed below). Genomic analysis also showed more gene families in plant defense and pathogen recognition (Nock et al., 2016) in *M. integrifolia*. Since macadamia is naturally adapted for rain forest habitat, our proteomic results are in agreement with previous genomics findings, where increased vicilin protein abundance could be an evolutionary change suited to protect against its pests and diseases. Further recent findings point that the three seed storage MiAMP2b-d peptides in macadamia appear to be interstitial peptides that are buried within the N-terminal region of a larger preprovicilin, similar to the preproalbumin of sunflower, which was noted to split into seed storage albumin and sunflower trypsin inhibitor 1 protein (Zhang et al., 2019).

3.4. Seed storage proteins

Proteins related to seed storage were just fewer than 3% of proteins but represented 8% in terms of total protein abundance. Seed storage proteins could be broadly classified as globulins (vicilins, legumins, α , β conglutins, glycinin), albumins (2S albumins, δ -conglutinin), prolamins (gliadins, hordeins, zein, avenin), and glutelins (Geiselhart et al., 2018). Significant seed storage protein matches were identified among macadamia nut proteins and include 11S legumin-like protein, which upon BLASTp analysis revealed $> 50\%$ similarity to 11S globulin protein from sesame, tobacco, potato, soybean, and capsicum, α , β and δ conglutins, which were highly similar to lupin conglutins, glycinin similar to that found in peanuts (Rabjohn et al., 1999), cruciferin, basic 7S globulin of soybean, glutelin and the allergen 19 kDa Ory s globulin protein of rice, as well as 11S storage globulin from *Coffea arabica*. In addition to their roles in nutrient reservoir activity, they also show significant allergenic potential in many tree nuts, peanuts, other oil seeds, and legume species (Liu, Burks, & Green, 2015; Mennini, Dahdah, Mazzina, & Fiocchi, 2016). On a biological process functional annotation perspective for the macadamia proteome, it seems under-representing that the MiAMP2 vicilin-like proteins were currently assigned just for their defense response role in UniProt. Further gene ontology would benefit to also assign them to seed storage proteins because the MiAMP2 for instance shares $> 80\%$ query coverage and around 50% identity to seed storage vicilins in other species (as discussed below).

3.5. Carbohydrate metabolism

Of the 74 carbohydrate metabolic proteins, the glyceraldehyde-3-phosphate dehydrogenase predominantly involved in glycolysis/gluconeogenesis and in carbon fixation for photosynthetic plants as part of the reductive pentose phosphate cycle for energy metabolism was notable. We also identified 2-phospho-D-glycerate hydrolase or enolase, pyruvate kinase, pyruvate dehydrogenase dihydrolipoamide acetyltransferase, malate synthase, which would feed the phosphoenolpyruvate pathway generating cellular energy. Phosphoglycerate kinase, glucose-6-phosphate isomerase, phosphoglycerate mutase, fructose-



bisphosphate aldolase, alcohol dehydrogenase, fructokinase, triose-phosphate isomerase were also identified in macadamia nut, being part of glycolysis/gluconeogenesis (Fig. 2). It has been reported that over-expression of fructose bisphosphate aldolase might impart salt stress adaptation, and this could particularly be useful for macadamia adapting to grow in excess salt conditions (Konishi, Kitano, & Komatsu, 2005). Proteins namely class I chitinase, involved in aminoglycoside and

nucleosugar metabolism, α - and β -galactosidases involved in galactose metabolism, glucan endo-1,3- β -glucosidase involved in starch and sucrose metabolism, isocitrate lyase involved in glyoxalate metabolism, β -mannosidase and xylose isomerase involved in fructose and mannose metabolism were also present among others grouped within carbohydrate metabolism. We also identified malate dehydrogenase involved in the tricarboxylic acid cycle and known for its role in providing energy.

especially for fruit ripening (Centeno et al., 2011; Voet, Voet, & Pratt, 2016).

3.6. Cell redox homeostasis

Proteins involved in cell redox homeostasis are classified as those proteins involved in maintaining the redox homeostasis environment within the cell, whereas oxidation-reduction processes are defined as metabolic processes resulting in the removal or addition of electrons which may have a subsequent addition or removal of protons. Although both terms are sometimes co-occurring in literature, for gene ontology annotations, classifications were primarily based on UniProt GO entries and protein subsets were classified based on their pre-defined annotations. Proteins belonging to cell redox homeostasis represented 1.4% of the total count of proteins and those belonging to oxidation-reduction processes were just over 2.4% of proteins.

In terms of protein abundance, proteins of both categories accounted to about 1% of protein content each. Some important cell redox homeostasis proteins identified in macadamia included peroxiredoxin 1 and 2, thioredoxin H3, manganese superoxide dismutase, catalase, disulfide isomerases and dihydrolipoyl dehydrogenase 2 among others. Among those involved in oxidation-reduction processes, the important proteins identified included NADPH dehydrogenase FQR1, ascorbate oxidase, quinone oxidoreductase-like protein, cytochrome P₄₅₀ isoform 1, betaine-aldehyde dehydrogenase, and methylmalonate-semialdehyde dehydrogenase.

3.7. Lipid metabolic proteins

Macadamia kernels contain over 80% nutritional lipid content, of which palmitoleic acids in addition to oleic acids contribute to 80% of monounsaturated fatty acids present in macadamia (Arroyo-Caro et al., 2016). Of the few lipid metabolic proteins identified with protein annotation, notable ones include enoyl-[acyl-carrier-protein] reductase, aspartic proteinase, β -hydroxyacyl-ACP dehydrase 1, preproliferin C, S-adenosyl L-methionine dependent thiol methyltransferase 2, neutral/alkaline non-lysosomal ceramidase, pyruvate kinase 1, phospholipase D, pyruvate dehydrogenase E1 component subunit β -2, glycerophosphoryl diester phosphodiesterase, ceramidase, cytosolic acetoacetyl-coenzyme A thiolase, and esterase D. Some of these enzymes may be contributing to the predominant monounsaturated fatty acid composition in macadamia unlike the polyunsaturated fatty acid rich oils.

3.8. Stress response proteins

We identified various forms of heat shock proteins, including a cold shock protein 2, desiccation PCC13-like protein, annexins and few uncharacterized proteins belonging to this category. Heat shock proteins (70 kDa and 18 kDa) are key stress response proteins that allow plants not only to adapt during various biotic and abiotic stresses such as drought, salinity, temperature or fungal infections but are also important for fruit development (Faurobert et al., 2007) and repairing damaged proteins. We did not find any matches to dehydrins in macadamia unlike previously reported in almonds (Koppelman et al., 2016). Interestingly, a desiccation PCC13-like protein was identified in macadamia indicating a possible role for imparting desiccation tolerance during growing stages of this nut which could be targeted for selective breeding in the future. Cyclophilins are known to be involved in many different cell processes including protein folding and signal transduction, and having them present in increased abundance is indicative of a possible role in regulating water stress or impart water-deficit tolerance in macadamia, as seen previously in the studies on sorghum (Hajheidari et al., 2005), peanuts and chickpeas (Kottapalli et al., 2009).

3.9. Amino acid metabolism

Amino acid metabolic proteins identified in macadamia represented 4% of the total protein count and 2.5% of the protein abundance relatively. Of those identified, we mapped several crucial enzymes such as fumarylacetoacetase known for the formation of fumarate used for ATP energy generation, and adenosylhomocysteinase involved in the metabolism of S-adenosyl-L-homocysteine to L-homocysteine. Phenylalanine, tyrosine, and tryptophan synthesis initiate when phosphoenolpyruvate and erythro-phosphate converts to corresponding intermediates in the shikimate pathway. Following this, enzymatic conversions of intermediates such as prephenate to arogenate and chorismate to tyrosine form the final conversion steps (Galili, Amir, & Fernie, 2016). Cysteine synthase isoforms X1 and X2 were also identified, which convert serine to cysteine in the biosynthesis pathway. Serine hydroxymethyltransferase, which converted serine to glycine in the cell, cystathionine beta-synthase which help to form cysteine from homocysteine, 3-isopropylmalate dehydrogenase aiding in leucine synthesis, and chorismate synthase were also identified. Succinate synthesis from succinate semialdehyde substrate was also reflected by the presence of succinate dehydrogenase activity protein. Several of these proteins are also intermediates within carbohydrate metabolic pathways such as glycolysis, the tricarboxylic acid cycle, and pentose phosphate pathway (Pratelli & Pilot, 2014).

We also mapped D-3-phosphoglycerate dehydrogenase in the conversion of 3-phospho-D-glycerate to phosphoserine, glutamine synthetase for the reaction glutamate to glutamine, imidazole glycerol-phosphate dehydratase and histidinol dehydrogenase, which catalyze immediate steps leading to biosynthesis of histidine, and glutathione S-transferase towards glutathione formation. Glutamate dehydrogenase, known for catalyzing the conversion of glutamate to 2-oxoglutarate, and aminoacylase-1, known for the conversion of glutamate to ornithine forming proline and arginine amino acids (Galili et al., 2016), were also found in macadamia. The roles of many of these enzymes in plants have been well characterized and recently reviewed in literature. Collectively identification of these enzymes confirms evidence to the biosynthesis of a total of 13 amino acids, many of which are final enzymatic conversions and those reaction steps are mapped into the biosynthesis pathway outlined in Fig. 3.

3.10. Protein metabolism, folding, translation, and transcription

Proteins involved in protein metabolic process accounted for 5% of all identified proteins. Elongation factor 1 α , protein-L-isoaspartate O-methyltransferases, several forms of endopeptidases, proteases and proteasome subunits were identified as part of this group. The TCP domain class transcription factor, peptidyl-prolyl *cis-trans* isomerase 1, chaperonin 60 α -subunit, heat shock protein 70, endoplasmic, and calreticulin-2 identified are known for their roles in the formation of disulfide bonds, folding and assembly of new polypeptides, and cellular translocation of precursor proteins. Peptidyl-prolyl *cis-trans* isomerase 1 has been found to initiate cytokines accumulation and help prolong eosinophil survival, involved in allergic asthma (Nechama et al., 2018). Several proteins involved in protein synthesis and translation machinery, which are also critical for seed development, were identified to represent 4% of the relative protein abundance in macadamia. They included 40 S ribosomal proteins S3a, S6, S8, SA, S12, S13-2, S15c, S15d, S18, S20, S23, S24, S29, as well as 60 S ribosomal proteins L6, L9, L11-1, L12, L14, L18a, L19-3, L23a, L24, L34, L35, and L37 were among others.

3.11. Cell cycle, growth and cell wall organization

Only 1% of those identified were involved in cell cycle, cell growth, chitin catabolic process, cold acclimation, and plant cell wall organization. Of these, PEARL1 4 protein has been known to be involved in the

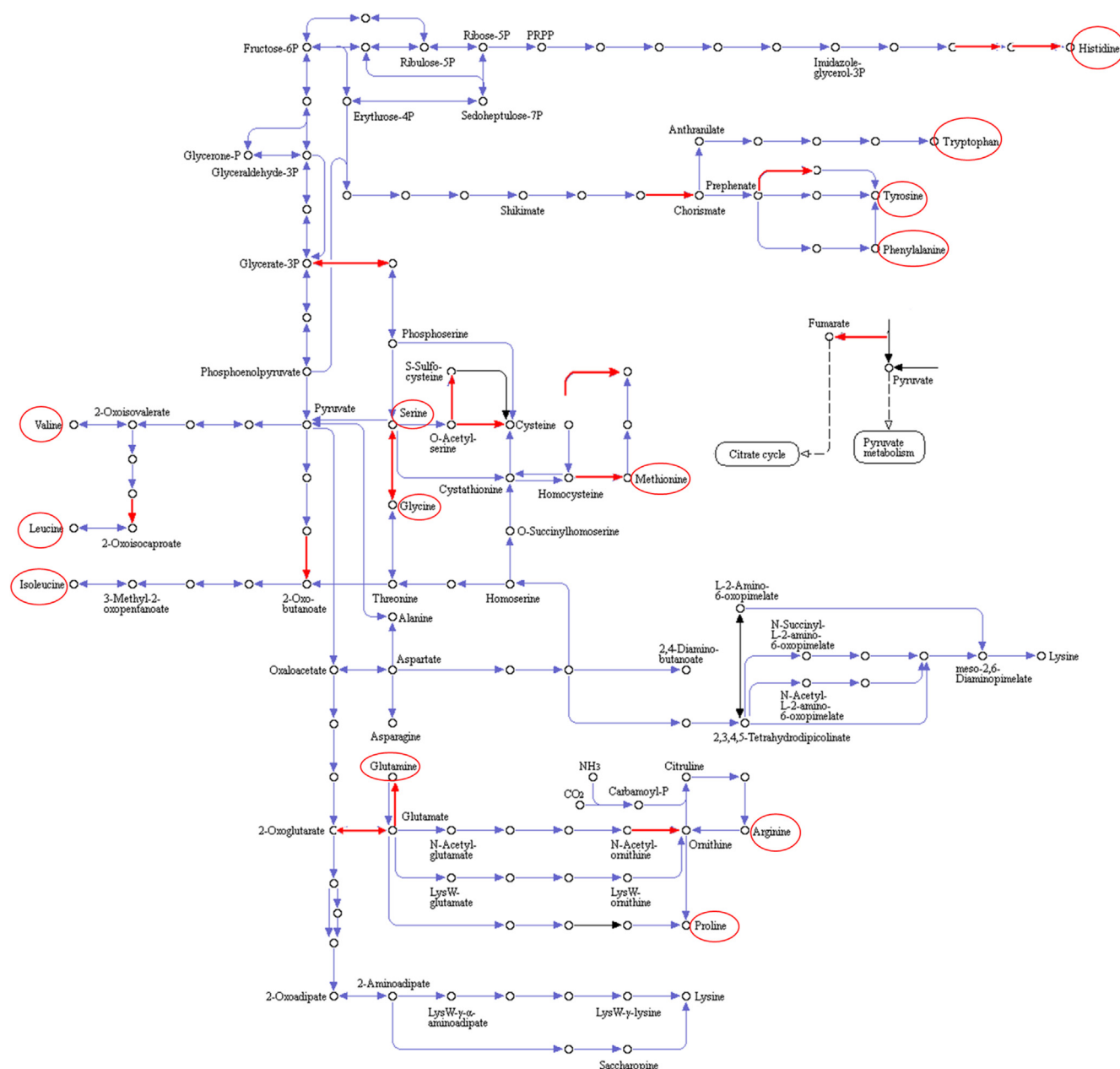


Fig. 3. Proposed cascading of the amino acid metabolic pathway in macadamia. Enzymatic conversions regulated by the proteins identified in this functional category and their corresponding amino acids are mapped into this pathway, highlighted in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

regulation of cell cycle, arabinogalactan protein 5 involved in cell growth, several forms of chitinases and pathogenesis-related protein 3-2 known for their roles in chitin and cell wall macromolecule catabolic processes, expansins and pectin esterases for cell wall organization (Faurobert et al., 2007; Huang, Zhang, Jiang, & Yu, 2016). Among other functional categories, we also noted those involved in DNA metabolic process, repair or replication and together also account for 1% of all proteins both in terms of protein count and protein abundance. Notable ones of these are nucleoside diphosphate kinases, integrase, hypothetical protein M569, ubiquitin receptor RAD 23, ubiquitin-conjugating enzyme E2 variant 1D among others. These are known for their distinct roles in nucleoside biosynthesis, DNA integration, mismatch repairs, nucleotide excision repairs, post-replication repairs and regulation of DNA replication.

3.12. *In silico* analysis of proteins with potential allergenicity and cross-reactivity

Allergen protein characterization to ascertain allergenic activity warrants testing sample protein extract with clinical studies such as SPT and/or Nasal Provocation Test, 1DE or 2DE coupled IgE-based immunoblot for evaluating allergen protein reactivity followed by MS/MS for *in-silico* assignment (Cetereisi et al., 2018). In the absence of IgE serum pool or when the available serum pool is small, as it is the case for macadamia allergic patients, our approach involving shotgun proteomics coupled with AllergenOnline and IEDB analyses could provide an alternative for studying potential allergenic proteins and is applicable for screening novel allergens.

Of the total 1079 proteins detected in for macadamia, several matched to known allergens in the AllergenOnline database. Based on the percentage of shared identity with the known allergens, the

allergenic protein matches were categorized into three classes: (1) $\geq 70\%$ identity, (2) $\geq 50\%$ but $\leq 69\%$ identity, and (3) $\leq 49\%$ identity with a high overall abundance (total spectral count ≥ 25). The percentage of shared identity is related to the peptide sequence coverage and the possible cross-reactivity. Where protein matches with $\geq 70\%$ identity are more likely to show similar folding and cross-reactivity with known allergens, those with $\leq 50\%$ identity are less likely to show cross-reactivity (Aalberse, Akkerdaas, & van Ree, 2001). A similar study on *in-silico* allergen assignments for potential pollen allergens using Structural Database of Allergenic Proteins tool and AllergenOnline found 27 out of 249 identified proteins as possible allergens and includes heat shock protein, enolase, pollen Ole e 1 allergen, thioredoxin and profilins as putative allergens (Wang et al., 2018). In comparison, our approach expands the scope for screening large datasets.

3.13. Proteins matching allergens with $\geq 70\%$ identity

Macadamia proteins that matched with $\geq 70\%$ identity were grouped into 23 protein clusters which comprised of 55 known allergenic proteins across different species (Table S2 in the [Supplementary Material](#) and summarised in Table 1). Of all the proteins with $\geq 70\%$ identity, we found three proteins showing 100% identity with known allergens, conglutin β -Lup an 1.0101 and peanut allergens Ara h 3 arachin and Ara h 3 glycinin isoforms. The significant sequence coverage shared between these matches further signifies the presence of an 11S globulin in macadamia similar to peanuts. The most abundant proteins are confirmed to be seed storage proteins belonging to 11S globulins and 7S vicilins, sharing homologies between 44 and 77% with sesame and walnut allergens. Interestingly, we did not find any direct matches to cross-species 2S albumin proteins in this study.

Matches to enolase-2 and its isoform 1 from *Hevea brasiliensi*, which are rubber tree latex allergens contained 90–95% identity at e-values of 1.4×10^{-155} showing up to 51% peptide sequence coverage. Enolases are homodimeric enzymes that are present in all tissues and organisms capable of glycolysis or fermentation and are known to be cross-reactive allergens in certain molds such as *Cladysporium herbarum* and *Alternaria alternata* (Pomes, Chapman, & Wunschmann, 2016). The likelihood for cross-reactivity of latex allergens with macadamia is high due to this high percentage of shared identity and coverage.

The peptidyl-prolyl *cis-trans* isomerase 1-like protein identified in macadamia matched a birch pollen allergen, Bet v 7, with 90% identity and 12% sequence coverage (Table 3), indicating the possible cross-reactivity between macadamia nut and birch pollen. This correlates with clinical cases that showed positive specific IgE but are non-responders of macadamia nut basophil activation tests demonstrating serum IgE cross-reactivity to major birch pollen allergens Bet v1, v7 and its homologues from hazelnut (cor a1), and apple (mal d1) and profilins (De Knop et al., 2010). Results reveal cross-reactivity links between birch pollen, hazelnut and macadamia nut; and together with epitope analysis (further discussed below) mainly point to lipid transfer proteins, latex proteins, Cor a1 and vicilin which are known to resist proteolytic digestion leading to severe allergic reactions, and is in agreement with patterns reviewed for other tree nuts and peanuts (Smeekens et al., 2018). Further to this, the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) protein cluster accounted for 0.22% of all the spectra, showing a significant abundance in macadamia. We found GAPDH that matched a GAPDH gliadin in wheat, known as Tri a 34.0101, with 80–89% identity and 41% sequence coverage. The 70 kDa heat shock protein in macadamia nut shared around 71% identity and 30% sequence coverage with the heat shock 70 kDa protein from the common mold, *Cladysporium herbarum*, the airborne spore allergen Cla h 4, also this protein matched to Der f 28, a dust mite allergen, with 72% identity and 25% coverage. The Hsp70 protein cluster accounted for 0.2% of the detected spectra and was one of the most abundant proteins in macadamia. The 70 kDa heat shock proteins

in macadamia are most likely to be cross-reactive with fungi and house dust mite.

Other proteins that matched allergens with high percent identity included seed maturation-like protein from sesame (*Sesame indicum*), the European white birch allergen Bet v 7, an endochitinase Pers a 1 from avocado, the European chestnut class I chitinase isoform 2, the banana allergen Mus a 2.0101 and the cyclophilin protein from carrot. Notably, despite its low abundance in macadamia, lactoylglutathione lyase (glyoxalases) also matched lactylglutathione lyase from rice with 80% identity and 30% sequence coverage.

3.14. Proteins matching allergens with $\geq 50\%$ and $\leq 70\%$ identity

Of those proteins with percentage identity $\geq 50\%$ and $\leq 70\%$, an 11S globulin (legumin-like) protein from sesame was identified to be the second most abundant protein in macadamia and represented 0.32% of the total spectra, but with only 4 unique peptides (Table S3 in the [Supplementary Material](#) and summarized in Table 2). The overall coverage of these peptides was only 5% of the sequence, suggesting that this highly abundant 11S globulin protein in macadamia may have a significantly different protein sequence compared to other 11S globulins. The 11S legumin-like protein matched to the minor allergen from sesame, Ses i 6.0101 which was previously noted for its cross-reactive nature with walnut allergens (Beyer, Grishina, Bardina, & Sampson, 2007; Wallowitz, Chen, Tzen, & Teuber, 2007). The 11S globulins belong to the cupin superfamily and are seed storage proteins known for their high allergenicity in many tree nuts such as anacardein (Ana o 2) in cashew, amandin (Pru du 6) in almond, Cor a 9 in hazelnut and Ara h 3 in peanuts amongst others.

Further, a glucose/ribitol dehydrogenase homolog identified in macadamia matched to a seed maturation-like protein, which is a minor allergen from sesame (Beyer, Bardina, Grishina, & Sampson, 2002) with 69% identity and 7% sequence coverage. A protein similar to conglutin α from lupin shared 52% identity and 18% sequence coverage with glycinin, a soybean allergen. Also, we found malate dehydrogenase showing 54% sequence homology with a major fungal allergen of *Malassezia furfur* with a high sequence coverage (26%).

3.15. Proteins matching allergens with identity $< 50\%$

Proteins matching allergens with identity $< 50\%$ include the three vicilin-like antimicrobial peptides (MiAMP2) 2–1, 2–2 and 2–3, which are the most abundant among all proteins in the soluble macadamia nut protein extract accounting for 1.7% of the total spectra (Table S4 in the [Supplementary Material](#) and summarized in Table 3). As our sample was sourced from a commercial mix of various cultivars, there may be different isoforms of the MiAMP2 present in the mix sample. The MiAMP2 vary in lengths, characterized by their anti-microbial property, are part of the larger vicilin precursor protein, which originates from a hydrophilic region of the N-terminus (Marcus et al., 1999). We found that the three peptides of MiAMP2 showed significant homologies to the vicilin-like protein precursor, Jug r 2.0101, which is an allergen found in walnut (*Juglans regia*). The MiAMP2-3 protein showed 45% identity to Jug r 2, and shared a 49% sequence homology, indicating that it could be a potential major allergen in macadamia. Findings from this group also include a match to Cor a1, a major allergen in hazelnut and shares 8 epitopes in parts. This is because it is also highly likely that these proteins in macadamia have considerable sequence polymorphism due to their taxonomical and genetic variation from other nut species. This possible cross-reactivity between hazelnut and macadamia was also previously suggested (Sutherland et al., 1999).

In addition, the percentage of coverage determined would depend on the total length of the peptide and the length of the query protein sequence and could result in a greater coverage for shorter proteins or incomplete transcripts. Low scoring for a protein match could also be due to incomplete tryptic digestion or a low number of peptide copies.

Table 1
Proteins identified from soluble macadamia nut protein extract and matching allergens from FARRP database with identity $\geq 70\%$, coverage $\geq 5\%$ and > 1 unique peptide, sorted by abundance and species of matching allergen, in brackets: number of similar proteins in cluster, AA: amino acids, shading: proteins matching to allergens in the same species.

Scaffold search with NCBI nr		FARRP Allergen Database						Allergen sequence coverage					
Protein name (no of total similar proteins)	GI Accession Number	Species	Average total spectrum count	Average unique spectrum count	GI Accession number	Allergen description	Allergen	Species	Common name	% identity	E-value	No of AA in allergen	% coverage by sample
RecName: Full = Enolase (5)	1169534	<i>Ricinus communis</i>	55	23	14423687	Enolase 2 (2-phosphoglycerate dehydratase 2)	Unassigned	<i>Hevea brasiliensis</i>	Para rubber tree	95	6.60E-177	445	51
Enolase (4)	34597330	<i>Brassica rapa</i> subsp. <i>oleifera</i>	42	17	9581744	Enolase, isoform 1	Hev b 9	<i>Hevea brasiliensis</i>	Para rubber tree	90	4.60E-158	445	38
Manganese superoxide dismutase	406870049	<i>Elaeis guineensis</i>	2	2	348137	Superoxide dismutase (manganese)	Hev b 10.0101	<i>Hevea brasiliensis</i>	Para rubber tree	79	5.00E-87	223	5
Unnamed protein (BLAST search 99% identity predicted: glyceraldehyde-3-phosphate dehydrogenase)	312282755	<i>Eutrema halophilum</i>	51	17	253783729	Glyceraldehyde-3-phosphate dehydrogenase	Tri a 34.0101	<i>Triticum aestivum</i>	Wheat	87	2.10E-107	337	41
Triosephosphate isomerase family protein (3)	118481564	<i>Populus trichocarpa</i>	3	2	11124572	Triosephosphate isomerase	Tri a 31.0101	<i>Triticum aestivum</i>	Wheat	79	3.10E-58	253	10
Heat shock protein 70 (6)	340784765	<i>Paeonia lactiflora</i>	40	18	729764	heat shock 70 kDa protein	Cla h 4	<i>Cladosporium herbarum</i>	Aero fungi very common	71	2.80E-180	643	30
Heat shock protein, putative (4)	255543357	<i>Ricinus communis</i>	24	10	685432788	Der f 28 allergen	Der f 28.0201	<i>Dermatophagoides farinae</i>	House dust mite	72	3.30E-173	654	25
Predicted: glucose and ribitol dehydrogenase-like	470133153	<i>Pragaria vesca</i> ssp. <i>vesca</i>	23	3	171853012	Seed maturation-like protein precursor	Unassigned minor allergen	<i>Sesamum indicum</i>	Sesame	77	9.50E-93	345	10
Conglutin beta (2)	169950562	<i>Lupinus angustifolius</i>	19	8	169950562	Conglutin beta	Lup an 1.0101	<i>Lupinus angustifolius</i>	Blue lupin	100	5.00E-157	611	22
Pathogenesis-related protein 3-2, class 1 chitinase (3)	345546656	<i>Lolium perenne</i>	14	2	3201547	Endochitinase	Pers a 1	<i>Persea americana</i>	Avocado	81	5.80E-48	326	6
Chitinase	259479248	<i>Diospyros kaki</i>	12	1	307159110	Class I chitinase isoform 2	Unassigned	<i>Castanea sativa</i>	European chestnut	80	7.40E-29	298	5
Hypothetical protein, region of lactoylglutathione lyase (6)	2213425	<i>Citrus × paradisi</i>	10	5	84029333	Lactoylglutathione lyase	Unassigned	<i>Oryza sativa</i>	Rice	80	1.40E-101	291	30
Peptidyl-prolyl cis-trans isomerase 1-like	255628137	<i>Glycine max</i>	6	4	21886603	(Methylglyoxalase) Peptidylprolyl isomerase (cyclophilin)	Bet v 7	<i>Betula pendula</i>	European white birch	90	2.80E-64	173	12
Arachin Ara h3 isoform	199732457	<i>Arachis hypogaea</i>	6	5	199732457	Arachin 11 s globulin	Ara h 3 isoform	<i>Arachis hypogaea</i>	Peanut	100	4.10E-170	530	13
Glycinin	3703107	<i>Arachis hypogaea</i>	5	4	3703107	Glycinin (cupin)	Ara h 3.0101	<i>Arachis hypogaea</i>	Peanut	100	3.30E-163	507	11
Predicted: peptidyl-prolyl cis-trans isomerase-like	357138513	<i>Brachypodium distachyon</i>	4	3	373939374	Cyclophilin	Unassigned	<i>Daucus carota</i>	Carrot	86	1.30E-62	171	15
Putative cyclophilin	146233381	<i>Pinus taeda</i>	4	2	65985152	Chain A, Cat R 1	Unassigned	<i>Catharanthus roseus</i>	Madagascar periwinkle	88	2.30E-63	178	23

Table 2
Proteins identified from soluble macadamia nut protein extract and matching allergens from FARPP database with identity $\leq 70\%$ and $\geq 50\%$, average total peptide count > 5 , sorted by abundance and species of matching allergen, in brackets: number of similar proteins in cluster, AA: amino acids, shading: proteins matching to allergens in the same species.

Scaffold search with NCBI nr				FARRP Allergen Database						Allergen sequence coverage			
Protein name (no of total similar proteins)	GI Accession Number	Species	Average total spectrum count	Average unique spectrum count	GI Accession number	Allergen description	Allergen	Species	Common name	% identity	E-value	No of AA in allergen	% coverage by sample
Legumin-like protein	1296435	<i>Asarum europaeum</i>	146	4	5,381,325	11S globulin precursor	Ses i 6.0101	<i>Sesamum indicum</i>	Sesame	52	1.10E-72	459	5
Glucose and ribitol dehydrogenase homolog	326495458	<i>Hordeum vulgare subsp. vulgare</i>	14	2	171,853,012	Seed maturation-like protein	Unassigned minor allergen	<i>Sesamum indicum</i>	Sesame	69	8.10E-75	345	7
Conglutin alpha 2 (2)	328684561	<i>Lupinus angustifolius</i>	22	4	18,641	precursor Glycinin	Unassigned	<i>Glycine max</i>	Soy	52	4.90E-36	562	18
Hsp70 (4)	148910700	<i>Picea sitchensis</i>	17	5	685,432,788	Der f 28 allergen	Der f 28.0201	<i>Dermatophagoides farinae</i>	House dust mite	63	3.30E-148	654	8
Ferritin-3	388510730	<i>Lotus japonicus</i>	12	5	442,565,878	Ferritin	Der f 30.0101	<i>Dermatophagoides farinae</i>	House dust mite	50	3.30E-26	171	13
Fructose-bisphosphate aldolase (6)	10800924	<i>Persea americana</i>	13	3	291,195,949	Aldolase A	Unassigned	<i>Thunnus albacares</i>	Yellowfin tuna	61	2.40E-96	364	9
Heat shock 70 kDa protein (3)	123588	<i>Daucus carota</i>	10	5	729,764	Heat shock 70 kDa protein	Cla h 4	<i>Cladosporium herbarum</i>	Aeroo fungi very common	66	9.00E-149	643	14
Ferritin 1	289546505	<i>Lupinus luteus</i>	10	3	15,072,346	ferritin heavy chain-like protein	Unassigned	<i>Dermatophagoides pteronyssinus</i>	House dust mite	51	1.40E-31	180	12
Malate dehydrogenase, chloroplastic (2)	462404918	<i>Prunus persica</i>	8	7	4,587,985	Major allergenic protein Mal f4	Mala f 4	<i>Malassezia furfur</i>	Skin fungi	54	2.50E-59	342	26

Table 3
Proteins identified from soluble macadamia nut protein extract and matching allergens from FARRP database with identity < 50% and > 25%, sorted by abundance and species of matching allergen, in brackets: number of similar proteins in the cluster, AA: amino acids, shading: proteins matching to allergens in the same species.

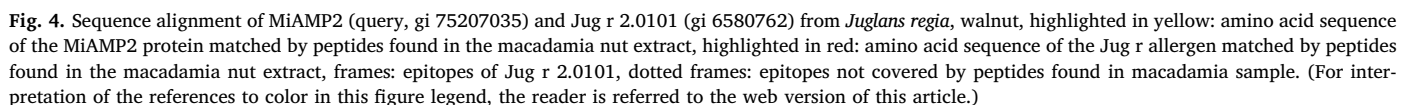
Protein	FARRP Allergen Database						Amino acid coverage		
	GI Accession Number	Species	Average total spectrum count	Average unique spectrum count	GI Accession number	Allergen	Allergen	Species	Common name
Vicilin-like antimicrobial peptides 2-3 (MIAMP2)	75207035	<i>Macadamia integrifolia</i>	681	72	6,580,762	Vicilin-like precursor	Jug r 2.0101	<i>Juglans regia</i>	Walnut
Vicilin-like antimicrobial peptides 2-2 (MIAMP2)	75266171	<i>Macadamia integrifolia</i>	541	62	6,580,762	Vicilin-like precursor	Jug r 2.0101	<i>Juglans regia</i>	Walnut
Vicilin-like antimicrobial peptides 2-1 (MIAMP2)	75207036	<i>Macadamia integrifolia</i>	503	60	6,580,762	Vicilin-like precursor	Jug r 2.0101	<i>Juglans regia</i>	Walnut
Phosphoglycerate kinase (2)	357451631	<i>Medicago truncatula</i>	34	13	5,726,304	Major allergen	Cor a 1.0401	<i>Corylus avellana</i>	European hazelnut
Cytosolic phosphoglycerate kinase family protein	224109062	<i>Populus trichocarpa</i>	28	10	22,690	Major allergen, aero	Cor a 1.0102	<i>Corylus avellana</i>	European hazelnut

The original amino acid sequence might be altered after translation leading to modifications that might enhance or decrease the allergenicity of the final protein product (Goodman et al., 2016). Naturally, the surface structure created by the final folding of the protein would be most relevant for the interaction with IgE antibodies (Aalberse et al., 2001) as well as for the true allergenicity to elicit a clinically relevant reaction. However, if the proteins under investigation and the known allergens share sufficient primary structure it is even more likely that their final folding and their function are also very similar (Hileman et al., 2002). It is vital to bear in mind that cross-reactive proteins can have a similar fold (3D structure), but the reverse may not apply, meaning not all similar folded proteins may be cross-reactive (Negi & Braun, 2017). Although cross-reactivities between different tree nuts are not necessarily correlated with the botanical taxonomy, 35% patients with a confirmed allergy (with specific IgE) to one tree nut could experience cross-reactivity with another tree nut. This threshold is higher for certain combination such as walnut-pecan, almond-hazelnut, cashew-pistachio (Sicherer & Sampson, 2014).

3.16. In silico analysis of IgE-binding epitopes

To evaluate potential cross-reactivity based on shared IgE-epitopes, we incorporated airborne allergens in addition to food allergens when screening matches to linear epitopes using the IEDB-AR and BLAST. We were able to successfully match macadamia protein candidates that may be allergens and identify possible cross-reactivities among existing allergens in foods (Tables S5–S7 in the Supplementary Material).

Epitopes matching to allergens from the rubber tree, peanuts, walnut, hazelnut, the fungi *Cladosporium herbarum*, rice, timothy grass, birch pollen, carrot, and Madagascar periwinkle were established for macadamia proteins. The highest number of epitope matches originated from the timothy grass (*Phleum pratense*) allergens. We were able to match one epitope from the house dust mite allergen, Der f 28.0201 (VEIANDQGNRTTPS), one rubber tree allergen from Hev b 3 (AEEV-EEERLK) and one of Ara h 3 isoforms (FYSNAPQEIFIQQGR) with 100% sequence coverage. In addition, 10 epitopes of arachin Ara h 3 isoforms and 6 epitopes of glycinin Ara h 3.0101 were partly covered by the peptides found in the macadamia sample. We identified two macadamia proteins MiAMP2 and phosphoglycerate kinase whose peptides matched to parts of many epitopes from walnut allergen Jug r 2 (up to 15 of 17 known epitopes of Jug r 2.0101) and Cor a 1 hazelnut allergen (11 of 24 known epitopes of Cor a 1.0401) respectively, along with significant peptide counts. Although most of their epitopes were only partly matched, the three antimicrobial peptides of MiAMP2 in macadamia shared close to 50% identity with Jug r 2 (Fig. 4). Recent findings on pollen allergen studies using a similar approach revealed 27 proteins as putative allergens (Wang et al., 2018). Of these Cla h4 (EAVAYGAAYQAAILS), is also an allergen match for macadamia and resulted with an epitope hit, but Hev b9 and Tri a25 matched the AllergenOnline search as potential pollen allergen but resulted no epitope hit. Hazelnut extract among other nuts was shown to weakly inhibit IgE-binding to macadamia proteins, which suggest that proteins of these two nuts may share similar IgE epitopes (Sutherland et al., 1999). This led to the speculation that macadamia may contain 2S albumins and legumin seed storage proteins (Teuber & Peterson, 2010). From the studies based on the IgG reactivity, macadamia appears to be partially cross-reactive with peanut and hazelnut (Kiening et al., 2005), but not with Brazil nut (Sharma, Roux, & Sathe, 2009) or walnut (Doi et al., 2008). The epitope matches identified between macadamia and the known allergens may not provide sufficient evidence to reliably predict the clinical significance (such as cross-reactivity) (Ito, Morishita, Ohshima, Sakamoto, & Tanaka, 2005), but the shared epitopes are an indication of possible cross-reactivity that warrants further investigation. Supplementary Material Tables A–J provides all data from all proteins searched against the FAARP and IEDB resources and includes all epitope assignments.



Label-free shotgun proteomics approach is well suited to provide valuable insight into the macadamia nut protein repertoire at the peptide level by identifying 947 proteins in 723 protein clusters. This study adds value to existing literature with information on protein and peptide abundances, quantitative distribution and ontology analysis on macadamia proteins extending beyond the antimicrobial MiAMP proteins and fatty acid genes that were predominantly studied. The most abundant proteins belong to 11S globulins and 7S vicilins, sharing

homologies with sesame and walnut allergens. Enolase-2 and its isoform 1 rubber tree latex allergens and the peptidyl-prolyl *cis-trans* isomerase 1 birch pollen allergen, Bet v 7, matched with 90% identity in macadamia. Many known allergens also presented shared epitope evidence. With increased demand for macadamia nut containing pre-packaged foods the prevalence of macadamia nut allergy is expected to rise, and characterization of its allergenic proteins warrants more attention. We demonstrate how shotgun proteomics data coupled with *in-silico* allergen analysis can indicate useful candidate proteins with allergenic and cross-reactivity potential (in clinical diagnosis and food

analysis) for subsequent IgE characterization studies of first-hand under-characterized allergenic foods, such as macadamia nut. Integration of genomic data (currently underway (Nock et al., 2019)) with the proteomic data will be indispensable for future molecular characterization of macadamia including allergens.

CRedit authorship contribution statement

Johanna Rost: Formal analysis, Investigation, Writing -original draft. **Sridevi Muralidharan:** Data curation, Formal analysis, Investigation, Methodology, Supervision, Writing - original draft. **N. Alice Lee:** Conceptualization, Project administration, Funding Acquisition, Resources, Supervision, Writing - review editing.

Acknowledgements

We express our gratitude towards the Australian Research Council (Australia, ITTC Project number IC130100021) for funding this project, postdoctoral fellowship for Sridevi Muralidharan and Ph.D. scholarship for Johanna Rost. We thank the Bioanalytical Mass Spectrometry Facility within the Mark Wainwright Analytical Centre of the UNSW. This work was undertaken using the infrastructure provided by the NSW Government co-investment in the National Collaborative Research Infrastructure Scheme (NCRIS) providing subsidized access to this facility, which is gratefully acknowledged.

Declaration of Competing Interest

The authors have declared no conflicts of interest.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.foodres.2019.108838>.

References

- Aalberse, R. C., Akkerdaas, J., & van Ree, R. (2001). Cross-reactivity of IgE antibodies to allergens. *Allergy*, 56(6), 478–490.
- Alam, M., Neal, J., O'Connor, K., Kilian, A., & Topp, B. (2018). Ultra-high-throughput DArTseq-based silicoDART and SNP markers for genomic studies in macadamia. *PLoS One*, 13(8), e0203465.
- Arroyo-Caro, J. M., Manas-Fernandez, A., Alonso, D. L., & Garcia-Maroto, F. (2016). Type I diacylglycerol acyltransferase (MtDGAT1) from macadamia tetraphylla: cloning, characterization, and impact of its heterologous expression on triacylglycerol composition in yeast. *Journal of Agricultural and Food Chemistry*, 64(1), 277–285.
- Beyer, K., Bardina, L., Grishina, G., & Sampson, H. A. (2002). Identification of sesame seed allergens by 2-dimensional proteomics and Edman sequencing: Seed storage proteins as common food allergens. *Journal of Allergy and Clinical Immunology*, 110(1), 154–159.
- Beyer, K., Grishina, G., Bardina, L., & Sampson, H. A. (2007). Identification of 2 new sesame seed allergens: Ses i 6 and Ses i 7. *Journal of Allergy and Clinical Immunology*, 119(6), 1554–1556.
- Cabello-Hurtado, F., Keller, J., Ley, J., Sanchez-Lucas, R., Jorrián-Novo, J. V., & Ainouche, A. (2016). Proteomics for exploiting diversity of lupin seed storage proteins and their use as nutraceuticals for health and welfare. *Journal of Proteomics*, 143, 57–68.
- Centeno, D. C., Osorio, S., Nunes-Nesi, A., Bertolo, A. L., Carneiro, R. T., Araújo, W. L., ... Fernie, A. R. (2011). Malate plays a crucial role in starch metabolism, ripening, and soluble solid content of tomato fruit and affects postharvest softening. *Plant Cell*, 23(1), 162–184.
- Cetereisi, D., Karlioglu, N., Gelincik, A., Michelland, S., Onay-Ucar, E., Ertek, B., ... Arda, N. (2018). Proteomic Identification of Allergenic Proteins of Morus alba L. Pollenexacerbation. *Asian Pacific Journal of Allergy and Immunology*.
- De Knop, K. J., Hagendorens, M. M., Bridts, C. H., Stevens, W. J., & Ebo, D. G. (2010). Macadamia nut allergy: 2 case reports and a review of the literature. *Acta Clinica Belgica*, 65(2), 129–132.
- Doi, H., Touhata, Y., Shibata, H., Sakai, S., Urisu, A., Akiyama, H., & Teshima, R. (2008). Reliable enzyme-linked immunosorbent assay for the determination of walnut proteins in processed foods. *Journal of Agricultural and Food Chemistry*, 56(17), 7625–7630.
- Ekbote, A., Hayman, G., & Bansal, A. (2010). Macadamia nut allergy: Potentially misleading specific IgE results. *Allergy*, 65(10), 1345.
- Faurobert, M., Mihr, C., Bertin, N., Pawlowski, T., Negroni, L., Sommerer, N., & Causse, M. (2007). Major proteome variations associated with cherry tomato pericarp development and ripening. *Plant Physiology*, 143(3), 1327–1346.
- Galili, G., Amir, R., & Fernie, A. R. (2016). The regulation of essential amino acid synthesis and accumulation in plants. *Annual Review of Plant Biology*, 67, 153–178.
- Geiselhart, S., Hoffmann-Sommergruber, K., & Bublin, M. (2018). Tree nut allergens. *Molecular Immunology*, 100, 71–81.
- Goodman, R. E., Ebisawa, M., Ferreira, F., Sampson, H. A., van Ree, R., Vieths, S., ... Taylor, S. L. (2016). AllergenOnline: A peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. *Molecular Nutrition & Food Research*, 60(5), 1183–1198.
- Hajheidari, M., Abdollahian-Noghabi, M., Askari, H., Heidari, M., Sadeghian, S. Y., Ober, E. S., & Salekdeh, G. H. (2005). Proteome analysis of sugar beet leaves under drought stress. *Proteomics*, 5(4), 950–960.
- Herbst, R. A., Wahl, R., & Frosch, P. J. (2010). Specific IgE reactivity and identification of potential allergens in macadamia allergy. *Journal of the European Academy of Dermatology and Venereology*, 24(11), 1361–1363.
- Hileman, R. E., Silvanovich, A., Goodman, R. E., Rice, E. A., Holleschak, G., Astwood, J. D., & Heffle, S. L. (2002). Bioinformatic methods for allergenicity assessment using a comprehensive allergen database. *International Archives of Allergy and Immunology*, 128(4), 280–291.
- Huang, J. S., Zhang, Y. X., Jiang, L., & Yu, Z. F. (2016). Comparative proteomics analysis of differential proteins in response to 6-benzylaminopurine treatment in Pteridium aquilinum senescence. *Postharvest Biology and Technology*, 116, 66–74.
- Ito, K., Morishita, M., Ohshima, M., Sakamoto, T., & Tanaka, A. (2005). Cross-reactive carbohydrate determinant contributes to the false positive ige antibody to peanut. *Allergy International*, 54(3), 387–392.
- Jamakhani, M., Lele, S. S., & Rekadwad, B. (2018). In silico assessment data of allergenicity and cross-reactivity of NP24 epitopes from Solanum lycopersicum (Tomato) fruit. *Data Brief*, 21, 660–674.
- Keller, A., Nesvizhskii, A. I., Kolker, E., & Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Analytical Chemistry*, 74(20), 5383–5392.
- Kiening, M., Niessner, R., Drs, E., Baumgartner, S., Krska, R., Bremer, M., ... Weller, M. G. (2005). Sandwich immunoassays for the determination of peanut and hazelnut traces in foods. *Journal of Agricultural and Food Chemistry*, 53(9), 3321–3327.
- Konishi, H., Kitano, H., & Komatsu, S. (2005). Identification of rice root proteins regulated by gibberellin using proteome analysis. *Plant Cell and Environment*, 28(3), 328–339.
- Koppelman, S. J., Jayasena, S., Luyckx, D., Schepens, E., Apostolovic, D., de Jong, G. A., ... Maleki, S. (2016). Allergenicity attributes of different peanut market types. *Food and Chemical Toxicology*, 91, 82–90.
- Kottapalli, K. R., Rakwal, R., Shibato, J., Burrow, G., Tissue, D., Burke, J., ... Payton, P. (2009). Physiology and proteomics of the water-deficit stress response in three contrasting peanut genotypes. *Plant, Cell & Environment*, 32(4), 380–407.
- Krut, N. L., Winget, J., Ryan, C. A., Wimalasena, R., Maurer-Stroh, S., Dearman, R. J., ... Gerberick, G. F. (2019). Proteomic and bioinformatic analyses for the identification of proteins with low allergenic potential for hazard assessment. *Toxicological Sciences*, 170(1), 210–222.
- Li, S., Geng, F., Wang, P., Lu, J., & Ma, M. (2016). Proteome analysis of the almond kernel (Prunus dulcis). *Journal of the Science of Food and Agriculture*, 96(10), 3351–3357.
- Liu, H., Li, H., Gu, J., Deng, L., Ren, L., Hong, Y., ... Liang, X. (2018). Identification of the candidate proteins related to oleic acid accumulation during peanut (Arachis hypogaea L.) seed development through comparative proteome analysis. *International Journal of Molecular Sciences*, 19, 4.
- Liu, M., Burks, A. W., & Green, T. D. (2015). Tree nut allergy: Risk factors for development, mitigation of reaction risk and current efforts in desensitization. *Expert Review of Clinical Immunology*, 11(5), 673–679.
- Mamone, G., Di Stasio, L., De Caro, S., Picariello, G., Nicolai, M. A., & Ferranti, P. (2019). Comprehensive analysis of the peanut allergome combining 2-DE gel-based and gel-free proteomics. *Food Research International (Ottawa, Ont.)*, 116, 1059–1065.
- Marcus, J. P., Goulter, K. C., Green, J. L., Harrison, S. J., & Manners, J. M. (1997). Purification, characterisation and cDNA cloning of an antimicrobial peptide from Macadamia integrifolia. *European Journal of Biochemistry*, 244(3), 743–749.
- Marcus, J. P., Green, J. L., Goulter, K. C., & Manners, J. M. (1999). A family of antimicrobial peptides is produced by processing of a 7S globulin protein in Macadamia integrifolia kernels. *The Plant Journal*, 19(6), 699–710.
- Maurer-Stroh, S., Krut, N. L., Kern, P. S., Gunalan, V., Nguyen, M. N., Limvipuvadh, V., ... Gerberick, G. F. (2019). AllerCatPro-prediction of protein allergenicity potential from the protein sequence. *Bioinformatics*, 35(17), 3020–3027.
- McGinn, C. (2016). Coroner calls for better food labelling after teen Jack Irvine died days after eating a Subway cookie. *Monash leader*, Vol. 2016.
- McManus, A. M., Nielsen, K. J., Marcus, J. P., Harrison, S. J., Green, J. L., Manners, J. M., & Craik, D. J. (1999). MiAMP1, a novel protein from Macadamia integrifolia adopts a Greek key beta-barrel fold unique amongst plant antimicrobial proteins. *Journal of Molecular Biology*, 293(3), 629–638.
- Mennini, M., Dahdah, L., Mazzina, O., & Fiocchi, A. (2016). Lupin and other potentially cross-reactive allergens in peanut allergy. *Current Allergy and Asthma Reports*, 16(12), 84.
- Mills, E. N. C., & Shewry, P. R. (2008). *Plant food allergens*. Blackwell Publishing Ltd.
- Moreno, Y., Gros, P. P., Tam, M., Segura, M., Valanparambil, R., Geary, T. G., & Stevenson, M. M. (2011). Proteomic analysis of excretory-secretory products of Heligmosomoides polygyrus assessed with next-generation sequencing transcriptomic information. *PLoS Neglected Tropical Diseases*, 5(10), e1370.
- Nechama, M., Kwon, J., Wei, S., Kyi, A. T., Welner, R. S., Ben-Dov, I. Z., ... Lu, K. P. (2018). The IL-33-PIN1-IRAK-M axis is critical for type 2 immunity in IL-33-induced allergic airway inflammation. *Nature Communications*, 9(1), 1603.
- Negi, S. S., & Braun, W. (2017). Cross-React: A new structural bioinformatics method for

- predicting allergen cross-reactivity. *Bioinformatics*, 33(7), 1014–1020.
- Neilson, K. A., George, I. S., Emery, S. J., Muralidharan, S., Mirzaei, M., & Haynes, P. A. (2014). Analysis of rice proteins using SDS-PAGE shotgun proteomics. In J. V. Jorrión, S. Komatsu, W. Weckwerth, & S. Wienkoop (Eds.). *Plant proteomics: Methods and protocols* (pp. 289–302). Totowa, NJ: Humana Press.
- Nesvizhskii, A. I., Keller, A., Kolker, E., & Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Analytical Chemistry*, 75(17), 4646–4658.
- Nock, C. J., Baten, A., Barkla, B. J., Furtado, A., Henry, R. J., & King, G. J. (2016). Genome and transcriptome sequencing characterises the gene space of *Macadamia integrifolia* (Proteaceae). *BMC Genomics*, 17(1), 937.
- Nock, C. J., Hardner, C. M., Montenegro, J. D., Ahmad Termizi, A. A., Hayashi, S., Playford, J., ... Batley, J. (2019). Wild origins of macadamia domestication identified through intraspecific chloroplast genome sequencing. *Frontiers in Plant Science*, 10, 334.
- Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., ... Vizcaino, J. A. (2019). The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Research*, 47(D1), D442–D450.
- Pomes, A., Chapman, M. D., & Wunshmann, S. (2016). Indoor allergens and allergic respiratory disease. *Current Allergy and Asthma Reports*, 16(6), 43.
- Pratelli, R., & Pilot, G. (2014). Regulation of amino acid metabolic enzymes and transporters in plants. *J Exp Bot*, 65(19), 5535–5556.
- Rabjohn, P., Helm, E. M., Stanley, J. S., West, C. M., Sampson, H. A., Burks, A. W., & Bannon, G. A. (1999). Molecular cloning and epitope analysis of the peanut allergen Ara h 3. *Journal of Clinical Investigation*, 103(4), 535–542.
- Rodríguez, M. F., Sanchez-Garcia, A., Salas, J. J., Garces, R., & Martinez-Force, E. (2015). Characterization of soluble acyl-ACP desaturases from *Camelina sativa*, *Macadamia tetraphylla* and *Dolichandra unguis-cati*. *Journal of Plant Physiology*, 178, 35–42.
- Sathe, S. K., Teuber, S. S., & Roux, K. H. (2005). Effects of food processing on the stability of food allergens. *Biotechnology Advances*, 23(6), 423–429.
- Sharma, G. M., Roux, K. H., & Sathe, S. K. (2009). A sensitive and robust competitive enzyme-linked immunosorbent assay for Brazil nut (*Bertholletia excelsa* L.) detection. *Journal of Agricultural and Food Chemistry*, 57(2), 769–776.
- Sicherer, S. H., & Sampson, H. A. (2014). Food allergy: Epidemiology, pathogenesis, diagnosis, and treatment. *Journal of Allergy and Clinical Immunology*, 133(2), 291–307 quiz 308.
- Smeeckens, J. M., Bagley, K., & Kulis, M. (2018). Tree nut allergies: Allergen homology, cross-reactivity, and implications for therapy. *Clinical and Experimental Allergy*, 48(7), 762–772.
- Sutherland, M. F., O'Hehir, R. E., Czarny, D., & Suphioglu, C. (1999). Macadamia nut anaphylaxis: Demonstration of specific IgE reactivity and partial cross-reactivity with hazelnut. *Journal of Allergy and Clinical Immunology*, 104(4 Pt 1), 889–890.
- Teuber, S. S., & Peterson, W. R. (2010). IgE-binding Proteins in Macadamia Nut Cross-react with Walnut. *Journal of Allergy and Clinical Immunology*, 125(2) Ab222–Ab222.
- Voet, D., Voet, J. G., & Pratt, C. W. (2016). (5 ed.). Hoboken, NJ: Wiley.
- Walczyk, N. E., Smith, P. M. C., Tovey, E. R., & Roberts, T. H. (2017). Peanut protein extraction conditions strongly influence yield of allergens Ara h 1 and 2 and sensitivity of immunoassays. *Food Chemistry*, 221, 335–344.
- Wallowitz, M. L., Chen, R. J., Tzen, J. T., & Teuber, S. S. (2007). Ses i 6, the sesame 11S globulin, can activate basophils and shows cross-reactivity with walnut in vitro. *Clinical and Experimental Allergy*, 37(6), 929–938.
- Wang, L., Zhang, X., Zhang, J., Fan, W., Lu, M., & Hu, J. (2018). Proteomic Analysis and Identification of Possible Allergenic Proteins in Mature Pollen of *Populus tomentosa*. *International Journal of Molecular Sciences*, 19(1), 250.
- Xu, X. P., Liu, H., Tian, L., Dong, X. B., Shen, S. H., & Qu le, Q. (2015). Integrated and comparative proteomics of high-oil and high-protein soybean seeds. *Food Chemistry*, 172, 105–116.
- Zhang, J., Payne, C. D., Pouvreau, B., Schaefer, H., Fisher, M. F., Taylor, N. L., ... Mylne, J. S. (2019). An ancient peptide family buried within vicilin precursors. *ACS Chemical Biology*, 14(5), 979–993.