

Proteomic Analysis of Oil-Roasted Cashews Using a Customized Allergen-Focused Protein Database

Shimin Chen and Melanie L. Downs*

Cite This: *J. Proteome Res.* 2022, 21, 1694–1706

Read Online

ACCESS |



Metrics & More



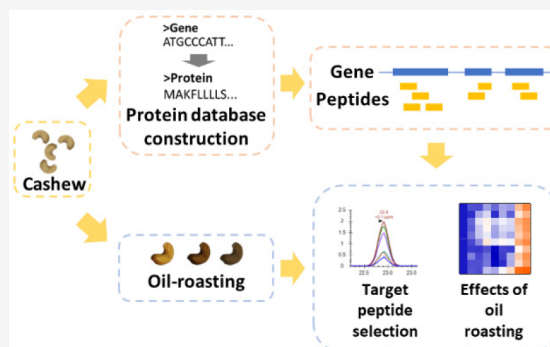
Article Recommendations



Supporting Information

ABSTRACT: Cashews are one of the most prevalent causes of tree nut allergies. However, the cashew proteome is far from complete, which limits the quality of peptide identification in mass spectrometric analyses. In this study, bioinformatics tools were utilized to construct a customized cashew protein database and improve sequence quality for proteins of interest, based on a publicly available cashew genome database. As a result, two additional isoforms for cashew 2S albumins and five other isoforms for cashew 11S proteins were identified, along with several other potential allergens. Using the optimized protein database, the protein profiles of cashew nuts subjected to different oil-roasting conditions (139 and 166 °C for 2–10 min) were analyzed using discovery LC-MS/MS analysis. The results showed that the cashew 2S protein is most heat-stable, followed by 11S and 7S proteins, though protein isoforms might be affected differently. Preliminary target peptide selection indicated that out of the 29 potential targets, 18 peptides were derived from the newly developed database. In the evaluation of thermal processing effects on cashew proteins, several Maillard reaction adducts were also identified. The cashew protein database developed in this study allows for comprehensive analyses of cashew proteins and development of high-quality allergen detection methods.

KEYWORDS: cashew, proteomics, proteogenomics, oil roasting, allergen



INTRODUCTION

Tree nuts are regulated as major allergenic foods in many jurisdictions, including the United States and the European Union.^{1,2} In the U.S., almonds, pecans, walnuts, pine nuts, cashew nuts, Brazil nuts, coconuts, and several others are considered tree nuts for the purposes of food allergen regulations.³ Approximately 0.2–2% of adults in the U.S. have tree nut allergies, of which 15–30% are allergic to cashew.⁴ Cashew trees (*Anacardium occidentale*) are in the same family (*Anacardiaceae*) as mango (*Mangifera indica*) and pistachio trees (*Pistacia vera*). Cashew-allergic patients are commonly found to have cross-reactivity with pistachio.^{5–7} It was suggested that cashew allergy may cause cardiovascular symptoms and anaphylactic reactions more frequently than peanut allergy, indicating a potentially higher level of severity of cashew allergies.^{8,9} Three major allergenic proteins have been identified in cashew nuts: Ana o 1 (7S vicilin), Ana o 2 (11S legumin), and Ana o 3 (2S albumin).^{10–12} The Ana o 2 proteins account for 50% of soluble proteins in raw cashews, followed by the Ana o 3 proteins, whereas Ana o 1 is the least abundant one among the three.¹³ Although oral immunotherapy for food allergies has been extensively researched in recent years, there have not been any therapies approved to treat cashew allergy.¹⁴ Therefore, avoidance of the offending foods is still required to prevent onset of allergic reactions in susceptible individuals.

Food allergen detection methods play critical roles in preventing allergen cross-contact and protecting allergic consumers. Enzyme-linked immunosorbent assays (ELISAs) are the most used technique for allergen detection and quantification. Despite the advantages of relatively fast detection and easy operation, ELISAs can be susceptible to matrix interferences and effects of food processing.^{15–17} One of the most relevant effects of food processing is changes in protein structure that can lead to differences in antibody recognition, resulting in incorrect results. Moreover, protein insolubility caused by thermal processing can be difficult to resolve due to limitations of extraction buffers.

In the past decade, many mass spectrometry (MS)-based methods have been developed for allergen detection and quantification. MS-based methods have shown advantages, such as less matrix interference and less susceptibility to processing effects, compared with ELISA methods. Most MS-based allergen detection methods rely on the detection of target

Received: February 16, 2022

Published: June 3, 2022



peptides that serve as surrogates for proteins from the allergenic source. Target peptides are often selected on the basis of a set of criteria predictive of performance, including abundance and susceptibility to modifications. However, the identification of quality target peptides, which is the key to MS-based methods, relies heavily on the completeness of protein sequence databases. While some allergenic foods such as soybean and peanut have been well investigated, the proteomes of many tree nuts remain incomplete. Currently, only 105 protein sequences for cashew are available on UniProt (accessed in June 2021), four of which are associated with allergenic proteins. Such a small protein database provides very limited options for target peptide selection, which may affect the sensitivity and accuracy of the MS method.

Researchers have explored alternatives for a surrogate protein database when the proteome is not available. Xiong et al. were able to select quality target peptides for English walnut based on translated genomic sequences before the complete proteome became available.¹⁸ In a subsequent study, Pirone-Davis et al.¹⁹ compared the peptide identification results of raw English walnut samples using five databases, including six-frame translation (6FT) of the genome, a tissue-specific transcriptome, and three proteomes produced using different bioinformatic pipelines. They found that all except for the 6FT method are efficient in constructing a protein database for peptide identifications.¹⁹ The 6FT method was less efficient due to the disproportional increase of database size and increase in the number of identified peptides.

Although a cashew proteome is not yet available, a prerelease version of the cashew genome has been published on Phytozome (<https://phytozome-next.jgi.doe.gov/>), the Plant Comparative Genomics portal of the US Department of Energy Joint Genome Institute. While the genome is incomplete, the protein sequences predicted from the database can provide much wider options for cashew proteome analysis, target peptide selection, and MS method development. In combination with *ab initio* gene prediction techniques, BLAST search against known allergens, and discovery MS analysis, amino acid sequences of proteins from commonly allergenic protein families can be identified. The customized protein database can provide more potential for target peptide selection, and thus may increase the quality of MS-based detection and quantification methods.

The objectives of this study were to identify protein sequences for known and potential cashew allergens, to improve sequence accuracy for proteins of interest by incorporating *ab initio* gene prediction methods, to analyze protein profiles in oil-roasted cashews using the newly developed database, and to select candidate target peptides for MS detection method development.

EXPERIMENTAL SECTION

Materials and Reagents

Ammonium bicarbonate (AmBic), Tris(hydroxymethyl) aminomethane (Tris), and iodoacetamide (IAA) were purchased from Sigma-Aldrich. Sequencing grade modified trypsin and chymotrypsin were purchased from Promega. Acetonitrile (ACN, ultrahigh-performance liquid chromatography–mass spectrometry, UHPLC–MS grade) and formic acid were from Fisher Scientific. Pierce 660 nm protein assay reagent, trifluoroacetic acid (TFA), dithiothreitol (DTT), and C18 spin columns were purchased from Thermo Fisher Scientific.

Urea and Mini-PROTEAN TGX stain-free gels (4–20%) were purchased from Bio-Rad.

Blanched raw cashew nuts (halves) were purchased from R U Nuts (Lincoln, NE). A portion of the raw cashew nuts was washed with distilled water five times and air-dried overnight to remove potential cross-contact from other tree nuts. The washed and unwashed cashew nuts were analyzed for peanuts, walnuts, almonds, and sesame by ELISA in a service laboratory (FARRP Confidential Analytical Testing Laboratory). The samples showed results of below the limit of quantification for all foods tested.

Sample Preparation for LC-MS/MS Discovery Analysis

Cashew nuts were ground using a freezer mill (SPEX 6850) and extracted at a 1:20 sample-to-buffer ratio (w/v, 0.07 g sample: 1.4 mL buffer) using 50 mM Tris-HCl, pH 8.6, with 6 M urea and 20 mM DTT. The samples were incubated in a 60 °C shaking water bath for 30 min, followed by centrifugation at 13 000g for 10 min. Soluble protein concentration of the extract supernatant was determined using the Pierce 660 nm protein assay. Sample extracts were analyzed on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) under reducing conditions using Tris-glycine gels. 10-fold diluted extracts (containing approximately 10 µg protein) were taken for insoluble tryptic digestion. The diluted extract (21 µL) was mixed with 50 mM AmBic (30 µL) and 100 mM DTT (3 µL) and incubated at 95 °C for 5 min. Upon cooling, 100 mM IAA (6 µL) was added to samples, which were then placed in the dark at room temperature for 20 min. After alkylation, samples were digested with trypsin at 1:100 protein-to-trypsin ratio (w/w) at 37 °C for 3 h, then added with the same amount of trypsin for overnight digestion (at least 12 h). Chymotryptic digestion was conducted on separate reduced and alkylated extract aliquots. The digestion conditions (time duration, protein-to-enzyme ratio) used for the chymotryptic samples were the same as the tryptic samples, with the only difference on incubation temperature (25 °C for chymotryptic samples).

The digested samples were desalted using Pierce C18 spin columns following the manufacturer's instructions. Samples were then dried in a vacuum evaporator and stored at –20 °C. Prior to injection, samples were resuspended in 40 µL 0.1% formic acid and 3% acetonitrile in HPLC-grade water. Each cashew sample was extracted in duplicates and analyzed in duplicate injections by LC-MS (*n* = 4).

LC-MS Settings: Discovery Analysis

Two experiments for discovery analysis were conducted. The first experiment was to generate proteomic evidence using raw cashew nuts digested using trypsin or chymotrypsin for cashew protein database development. The second experiment was to analyze change of protein profile in oil-roasted cashew nuts.

Raw cashew samples digested using trypsin were analyzed using Thermo Scientific Q Exactive HF Hybrid Quadrupole-Orbitrap Mass Spectrometer coupled with UltiMate 3000 RSLCnano system. The injection volume was 3 µL. The LC column was a Waters Acquity UPLC M-Class Peptide CSH C18 column (250 mm × 75 µm i.d.) with 1.7 µm particles (130 Å). The samples were eluted from 5 to 32% solvent B (80% ACN, 1% formic acid in water) in 96 min with 0.260 µL/min column flow rate. MS1 resolution was set at 120 000. Maximum injection time (IT) and automatic gain control (AGC) are 60 ms and 3 × 10⁶, respectively. MS2 resolution was set as 15 000. Maximum IT and AGC for MS2 were 250 ms and 1 × 10⁵, respectively. Top 15 precursors were selected for fragmentation

with 30 s dynamic exclusion. The isolation window was 1.6 m/z without offset. Intensity threshold was 1×10^5 . Charge states at 1, 6–8, >8, and unassigned were excluded.

Raw cashew samples digested using chymotrypsin and all cashew samples in the oil-roasting experiment were analyzed using a Thermo Scientific Q Exactive Plus Orbitrap Mass Spectrometer coupled with Dionex UltiMate 3000 RS ultrahigh performance liquid chromatography (UHPLC) system. Samples were loaded onto a Hypersil GOLD C18 Selectivity LC Columns (100 mm $\times$ 1 mm i.d.) with 1.9 μ m silica particles (175 Å). The column flow rate was maintained at 0.06 mL/min. Peptides were eluted with a linear gradient of 2–40% solvent B over 67 min, where mobile phase A consisted of 0.1% (v/v) formic acid in water and mobile phase B consisted of 0.1% (v/v) formic acid in ACN. MS1 resolution was set as 70 000. Maximum IT and AGC target for MS1 were set as 100 ms and 1×10^6 , respectively. Top 10 precursor ions were selected for fragmentation with 10 s dynamic exclusion. The isolation window was 2 m/z without offset. The intensity threshold was 1.5×10^4 . MS2 resolution was 35 000. Maximum IT and AGC for MS2 were 100 ms and 1×10^5 , respectively.

LC-MS Settings: Targeted Analysis

Oil-roasted cashew samples were prepared as mentioned above and injected into LC-MS for targeted analysis using Thermo Scientific Q Exactive Plus Orbitrap Mass Spectrometer coupled with an Dionex UltiMate 3000 RS UHPLC system. Peptides were eluted with a linear gradient of 2–25.8% solvent B over 42 min. A 5 min retention window was scheduled for target peptides. MS1 resolution was 17 500 followed by MS2 at 140 000 using Parallel Reaction Monitoring (PRM). ACG targets were 1×10^6 for MS1 and 2×10^5 for MS2. Maximum IT was 50 ms for MS1 and auto for MS2. Isolation window was 1.6 m/z without offset.

Data Analysis

All data generated from the discovery analyses were analyzed using Proteome Discoverer 2.3. Precursor mass tolerance and fragment mass tolerance were set to 10 ppm and 0.02 Da, respectively. False discovery rate (FDR) of 1% was applied to peptide spectrum matches, peptide, and protein levels. The maximum number of dynamic modifications allowed for each peptide was four. Dynamic modifications included oxidation (M) and carbamylation (peptide N-Terminus), and static modification included carbamidomethyl (C). Full tryptic cleavage with a maximum of two missed cleavage sites was used as digestion settings. For the chymotryptic digested samples, peptide settings were set as full chymotrypsin (FWY). Sequest HT was used as the search algorithm with Percolator used for peptide validation. For cashew protein database development, both tryptic and chymotryptic raw cashew samples were analyzed and combined in the results. For the analysis of oil-roasted cashew samples, the PDstudy files were imported into Skyline 19.0 to evaluate MS1 peak quality and to calculate the MS1 peak area. The Skyline file for MS1-based protein quantification of oil-roasted cashew samples has been uploaded to PanoramaWeb (<https://panoramaweb.org/CashewDB.url>).

For the data analysis of Maillard reaction products, the following dynamic modifications were included: Hex/+162.053 Da (K, R, peptide N-Terminus), Carboxymethyl/+58.005 Da (K, R, peptide N-Terminus), Carboxyethyl/+72.021 Da (K, R, peptide N-Terminus), HydroxymethylOP/+108.021 Da (K, R, peptide N-Terminus), Pyronyl-amine/+126.032 Da (K, R,

peptide N-Terminus). Trypsin (semi, C-terminus cleaved by trypsin) was used as the enzyme setting. To limit the database search time, only sequences for high confidence proteins identified in the peptide search using the regular settings on the same samples were included for the Maillard reaction product search.

The protein/peptide identification files have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD027543 and DOI: [10.6019/PXD027543](https://doi.org/10.6019/PXD027543).²⁰

Protein Database Development

Cashew genomic and protein sequence information was downloaded from JGI Phytozome V12.1. These sequence data were produced by the US Department of Energy Joint Genome Institute (<http://www.jgi.doe.gov/>) in collaboration with the user community. The downloaded files included protein sequences with all alternative splice variants, protein sequences from primary transcript only, genome assembly, and annotation information, separately. The details on genome analysis can be found at https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Aoccidentale_er. According to the information provided, the genome annotation was determined using homology-based gene prediction programs along with RNA-Seq evidence. The RNA-Seq results, however, were not available to download.

To construct a cashew protein database, the protein annotation information acquired from Phytozome was attached to the corresponding protein accessions using text concatenation in Microsoft Excel (not all accessions were annotated). All sequences were screened for duplication using DBToolKit.²¹ The accession number of the repeated sequence would be attached to the title of the first occurrence of the sequence and the repeated sequences were removed. AUGUSTUS^{22,23} gene prediction was performed on the Phytozome cashew genome on both forward and reverse strands, using *Arabidopsis thaliana* as the reference genome (see Table S1 for parameters used in AUGUSTUS prediction). Alternative splicing prediction was disabled in the initial prediction. The getAnnoFasta.pl script (<https://github.com/Gaius-Augustus/Augustus/blob/master/scripts/getAnnoFasta.pl>) was used to extract amino acid sequences from the results. The AUGUSTUS gene prediction generated 51 598 unique protein sequences. A second AUGUSTUS gene prediction was performed specifically on scaffold 8, a region where several 11S protein coding-genes were found, of the Phytozome cashew genome with alternative splicing enabled.

The final protein database contains 38 798 protein sequences, including all UniProt *Anacardium occidentale* sequences (denoted as “UniProt”, 96 sequences, accessed in January 2019), all Phytozome protein sequences predicted on the basis of primary transcripts (denoted as “Phytozome-primary”, 38 669 sequences), selected Phytozome protein sequences predicted on the basis of alternative transcripts (denoted as “Phytozome-selected secondary”, 21 sequences), and selected AUGUSTUS predicted protein sequences (denoted as “selected AUGUSTUS”, 12 sequences). The selection criteria for which sequence should be included in the final database are described in the Results and Discussion section.

Oil-Roasting of Cashew Nuts

Blanched raw cashew nut halves (100 g) were oil-roasted in 1 L of soybean oil using a kitchen-scale fryer at two different temperatures (139 and 166 °C) for different time durations (2,

Table 1. Results Summary of BLASTP Search against AllergenOnline V18B

query ID	top match NCBI protein ID (AllergenOnline V18B)	sequence identity of the top match	average sequence identity of all matches	number of matches (sequence identity $\geq 35\%$)
Anaoc.0001s2422.1.p	gil24473800 gil77l2s albumin [<i>Anacardium occidentale</i>] (Q8H2B8 ^a)	100.0	50.8	9
Anaoc.0001s2421.1.p	gil24473800 gil77l2s albumin [<i>Anacardium occidentale</i>]	83.3	49.8	9
Anaoc.0020s0683.1.p	gil110349081 gil1270lPis v 1 allergen 2S albumin [<i>Pistacia vera</i>]	74.0	48.8	9
Anaoc.0002s1062.1.p	gil25991543 gil76lallergen Ana o 2 [<i>Anacardium occidentale</i>] (Q8GZP6)	100.0	47.0	8
Anaoc.0008s0618.1.p	gil110349083 gil1093lPis v 2.0101 allergen11S globulin precursor [<i>Pistacia vera</i>]	85.2	45.7	42
Anaoc.0008s0618.2.p	gil110349083 gil1093lPis v 2.0101 allergen11S globulin precursor [<i>Pistacia vera</i>]	84.0	46.6	42
Anaoc.0008s0621.1.p	gil110349083 gil1093lPis v 2.0101 allergen11S globulin precursor [<i>Pistacia vera</i>]	83.6	44.8	8
Anaoc.0603s0005.1.p	gil110349083 gil1093lPis v 2.0101 allergen11S globulin precursor [<i>Pistacia vera</i>]	78.0	45.0	10
Anaoc.0004s0038.1.p	gil110349083 gil1093lPis v 2.0101 allergen11S globulin precursor [<i>Pistacia vera</i>]	77.1	45.0	10
Anaoc.0008s0624.1.p	gil110349083 gil1093lPis v 2.0101 allergen11S globulin precursor [<i>Pistacia vera</i>]	63.1	44.1	28
Anaoc.0006s0989.1.p	gil25991543 gil76lallergen Ana o 2 [<i>Anacardium occidentale</i>]	54.5	39.4	6
Anaoc.0005s1212.1.p	gil29839254 gil531l13S globulin seed storage protein 1 precursor [<i>Fagopyrum esculentum</i>]	36.3	35.9	2
Anaoc.0476s0001.1.p	gil29839254 gil531l13S globulin seed storage protein 1 precursor [<i>Fagopyrum esculentum</i>]	36.3	36.3	1
Anaoc.0006s0058.1.p	gil21914823 gil75lvcilin-like protein [<i>Anacardium occidentale</i>] (Q8LSL5)	100.0	57.0	9
Anaoc.0006s0058.1.p	gil2166498 gil75lvcilin-like protein [<i>Anacardium occidentale</i>] (Q8LSL6)	99.8	51.7	8
Anaoc.0008s0426.1.p	gil42414627 gil257lvcilin [<i>Pisum sativum</i>]	42.1	39.4	4
Anaoc.0228s0005.2.p	gil42414627 gil257lvcilin [<i>Pisum sativum</i>]	35.5	35.5	1
Anaoc.0228s0005.3.p	gil42414627 gil257lvcilin [<i>Pisum sativum</i>]	35.5	35.5	1
Anaoc.0010s1033.1.p	gil187766749 gil567lGly m Bd 28K allergen [<i>Glycine max</i>]	58.9	58.2	5
Anaoc.0005s1273.1.p	gil187766749 gil567lGly m Bd 28K allergen [<i>Glycine max</i>]	58.3	56.7	5
Anaoc.0228s0005.1.p	gil187766747 gil567lGly m Bd 28K allergen [<i>Glycine max</i>]	58.2	58.2	1
Anaoc.0017s0279.1.p	gil187766749 gil567lGly m Bd 28K allergen [<i>Glycine max</i>]	36.7	36.5	4
Anaoc.0758s0003.1.p	gil122218540 gil2284lOleoin 2 [<i>Arachis hypogaea</i>]	73.4	56.1	9
Anaoc.0018s0441.1.p	gil29170509 gil389loleosin [<i>Corylus avellana</i>]	71.7	65.6	6
Anaoc.0013s0411.1.p	gil5381321 gil1238l15 kDa oleosin [<i>Sesamum indicum</i>]	58.3	48.1	9
Anaoc.0004s1797.1.p	gil10834827 gil1893loleosin [<i>Sesamum indicum</i>]	50.9	45.3	7

^aQ8H2B8, Q8GZP6, Q8LSL5, and Q8LSL6 are UniProt accession numbers for the same sequence in NCBI.

4, 8, and 10 min). The temperature fluctuation, measured by a digital thermometer, was ± 5 °C over the course of roasting. Approximately 25 g of cashew halves were sampled at each time point. All cashew halves were placed on a paper towel to remove excess oil. Samples were then ground and stored at -20 °C. Cashew samples were prepared for LC-MS/MS analysis following extraction, reduction, alkylation, in-solution tryptic digestion, and C18 cleanup as described above. Each oil-roasted cashew sample was extracted in duplicate and analyzed in duplicate injections by LC-MS ($n = 4$).

RESULTS AND DISCUSSION

Peptide Identification Using the Protein Database Available from Phytozome

Raw cashew nut samples were analyzed using LC-MS/MS in data dependent acquisition mode. The peptide identification was initially performed using a database that contains UniProt sequences (96 sequences) and Phytozome-primary sequences

(38 669 sequences). A total of 6595 high confidence peptides were identified in raw cashew samples (FDR < 1%, identified in at least 2 out of 4 replicates, tryptic digestion). Upon adding protein sequences from Phytozome alternative transcripts (28 319 sequences), the number of high confidence peptides that can be exclusively attributed to the alternative transcripts only increased by 22, which was not a substantial increase given the number of additional protein sequences in the database. Therefore, to shorten the run time of peptide searches and avoid false-positive issues in the statistical analyses of the search algorithm, only the protein sequences that have at least 1 unique and high confidence peptide identified in the Phytozome alternative transcripts were included in future database searches (Phytozome-selected secondary, 21 sequences).

Potential Allergen Identification and Classification Using AllergenOnline

Although the annotation information for many protein sequences is available via Phytozome, the annotation is mostly

Table 2. Comparison of Gene Prediction Methods

cashew protein group	temporary protein ID	transcript name	source of gene prediction ^c	genomic location (strand) ^d	number of unique peptides ^e	sequence coverage (%)
Ana o 1 (7S)	Ana o 1.1 ^a	Anaoc.1125s0001.1	P	contig_1125:4,696–6,964(–)	59	82
		Anaoc.0006s0058.1	P	scaffold_6:4,091,886–4,094,154(–)		
Ana o 2 (11S)	Ana o 2.1	Anaoc.0002s1062.1	P; AG	scaffold_2:14,428,739–14,431,163(–)	63	89
	Ana o 2.2 ^b	Anaoc.0004s0038.1	P	scaffold_4:234,132–235,889(–)	51	80
		Anaoc.0603s0005.1	P	contig_603:32,000–33,756(–)	11	69
	Ana o 2.3	g19481.t1	AG	scaffold_8:3,581,945–3,583,745(–)		
	Ana o 2.4	g600.t3	AG-alt	scaffold_8:3,567,552–3,569,367(+)	10	52
	Ana o 2.5	g602.t3	AG-alt	scaffold_8:3,576,579–3,578,454(+)	10	62
	Ana o 2.6	Anaoc.0008s0624.1	P	scaffold_8:3,595,033–3,596,788(–)	31	58
Ana o 3 (2S)	Ana o 3.1	Anaoc.0001s2422.1	P; AG	scaffold_1:26,642,130–26,643,032(–)	14	79
	Ana o 3.2	Anaoc.0020s0683.1	P; AG	scaffold_20:9,974,954–9,975,606(–)	12	65
	Ana o 3.3	Anaoc.0001s2421.1	P; AG	scaffold_1:26,640,627–26,641,712(–)	11	64
Cupin family protein	Ana o 4.1	Anaoc.0005s1273.1	P	scaffold_5:13,846,350–13,848,657(+)	5	49
	Ana o 4.2	Anaoc.0010s1033.1	P	scaffold_10:12,433,535–12,436,104(+)	13	37
	Ana o 4.3	Anaoc.0017s0279.1	P; AG	scaffold_17:1,742,852–1,747,444(–)	33	53
	Ana o 4.4	Anaoc.0228s0005.1	P	scaffold_228:23,267–25,313(+)	2	51
Oleosins	Ana o 5.1	Anaoc.0004s1797.1	P; AG	scaffold_4:17,984,031–17,985,076(+)	11	44
	Ana o 5.2	g14981.t1	AG	scaffold_6:9,939,182–9,940,201(+)	12	54
	Ana o 5.3	Anaoc.0013s0411.1	P; AG	scaffold_13:2,799,872–2,800,315(+)	7	46
	Ana o 5.4	g32004.t1	AG	scaffold_14:10,700,811–10,701,316(+)	12	46
	Ana o 5.5	Anaoc.0018s0441.1	P; AG	scaffold_18:3,510,835–3,512,387(+)	3	19
	Ana o 5.6 ^a	Anaoc.0010s0356.1	P	scaffold_10:4,241,394–4,242,735(+)	4	34
		Anaoc.0758s0003.1	P	contig_758:80,258–81,599(+)		

^aAnaoc.1125s0001.1 and Anaoc.0006s0058.1 have identical sequences. Anaoc.0010s0356.1 and Anaoc.0758s0003.1 also share identical sequences.

^bAnaoc.0004s0038.1 and Anaoc.0603s0005.1 are 97% identical, but no unique peptides can be found in Anaoc.0603s0005.1. ^cP = Phytozome; AG = AUGUSTUS prediction for primary transcript only; AG-alt = AUGUSTUS prediction for alternative splicing variants in scaffold 8. ^dGenomic locations in JGI cashew genome v0.9 (Phytozome genome ID: 449). ^eAll peptides were identified with high confidence (FDR < 1%) in at least two out of four replicates in either the tryptic or chymotryptic digestion group. Data were collected from discovery analysis on raw cashew samples, combining tryptic and chymotryptic results. Peptides with different modifications were counted as one peptide. See the details of the peptides in Table S4.

total of 187 high confidence peptides (FDR < 1%, found in at least 2 out of 4 samples) were unique to the AUGUSTUS predicted sequences. Most of the proteins from which these peptides derived were functional proteins, such as heat shock protein and aspartyl protease family protein. Two proteins (g14981.t1 and g32004.t1) shared high degrees of sequence similarity (40–60%) to oleosins in other plants like peanuts, hazelnuts, and sesame. Twelve unique peptides were found in each of these proteins, resulting in at least 46% protein coverage (Table 2). The low protein coverage could be a result of several factors: the size of the protein, the distribution of enzymatic cleavage sites, and the hydrophobic nature of the central region, which renders many peptides derived from oleosins unsuitable for MS detection. Given the analytical difficulties, in silico tryptic

digestion of the protein indicated that most of the tryptic peptides (10 out of 11) that have a peptide length between 6 and 25 aa have been identified in the discovery MS analysis.

To compare the gene prediction accuracy for the allergens of interest, a proteomics approach was implemented. The protein coding-genes predicted by AUGUSTUS were aligned with ones predicted by Phytozome for manual inspection of gene structure. Raw cashew proteins were digested using either trypsin or chymotrypsin, followed by discovery MS. Chymotrypsin was used to help with the identification of exon–exon spanning peptides because tryptic cleavage sites often coincide with exon–ending and exon–exon junction residues.³⁵ High quality peptides (FDR < 1%, found in at least two of four replicates) identified in the database search were then mapped

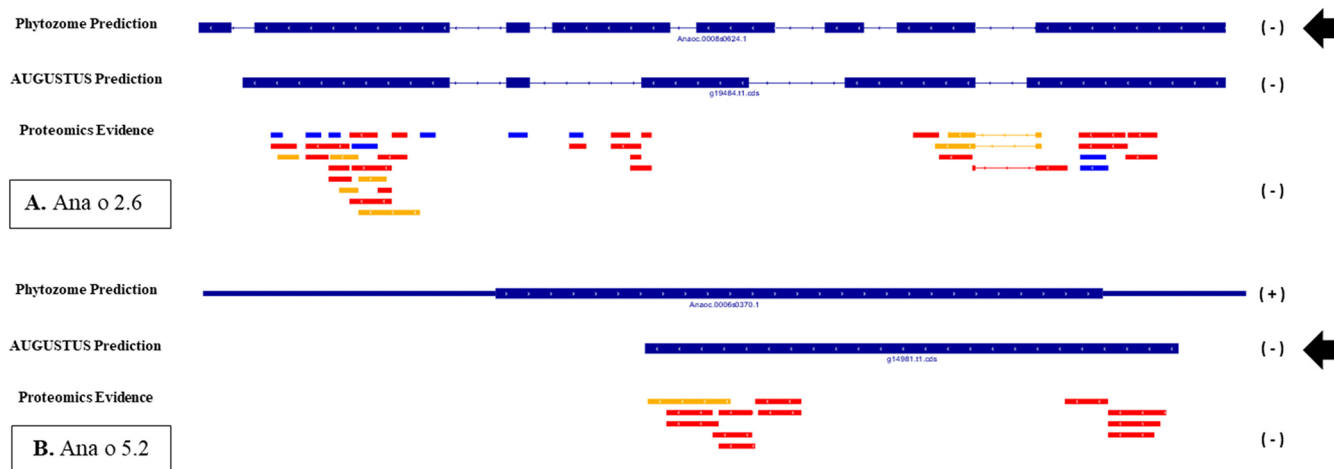


Figure 2. Comparison of gene prediction results in genomic region scaffold_8:3,595,033–3,596,788 (Ana o 2.6) and scaffold_6:9,939,182–9,940,201 (Ana o 5.2). In the proteomics evidence section, red indicates unique tryptic peptide, yellow indicates unique chymotryptic peptide, and blue indicates nonunique peptides that were found in the protein. The correct prediction is indicated by the black arrow. The $\pm$ sign indicates whether the prediction was based on the forward or reverse strand. Gene prediction comparison for other proteins can be found in Figure S4.

to the genome using BLAST and a genome location conversion tool³⁶ to assist in the assessment of gene prediction accuracy. A manual inspection was conducted using the proteomic evidence provided by the discovery MS analysis. Several factors were considered in the assessment: the number of peptides identified in the transcripts predicted by the two sources, the uniqueness of the identified peptides, and the biochemical properties of proteins (length of signal peptides, starting position of the mature protein, and amino acid pattern typical to the protein family). The assessment was performed according to the two following rules: (1) if all peptides identified in the region were shared between the two predictions, and there was no other evidence to support either prediction, the Phytozome prediction was selected by default, and (2) when there is at least one peptide unique to either prediction and none in the other prediction, the prediction with unique peptides was selected. Examples of the assessment are demonstrated in Figure 2. The comparison results are summarized in Table 2. A temporary protein ID is given for each distinguishable protein sequence (at least one unique peptide can be found). However, these temporary IDs are only for simplified protein reference in this study and not any indications of the WHO/IUIS allergen nomenclature. Except Ana o 1.1, 2.1, and 3.1, all other protein sequences listed in Table 2 were newly identified.

As suggested in Table 2, among the 22 selected protein sequences, nine were contributed by Phytozome, five by AUGUSTUS, and eight shared by both. This result suggested that Phytozome had predicted accurate results in most cases, and AUGUSTUS could improve the prediction in some cases (Figure 2). As illustrated in Figure 2A, several unique peptides can be mapped to the Phytozome prediction (the fifth exon from the right) but none to the AUGUSTUS prediction. Moreover, three exon–exon spanning peptides between the first and second exons were identified, of which two are chymotryptic peptides. In the case of Ana o 5.2 (Figure 2B), AUGUSTUS predicted a gene on the reverse strand while Phytozome predicted one on the forward strand. The AUGUSTUS prediction was considered correct because several unique peptides were identified for the AUGUSTUS protein sequence but none for the Phytozome sequence. A more complicated case lies within Ana o 2.4 and Ana o 2.5, which correspond to

Anaoc.0008s0618.2 and Anaoc.0008s0618.1 in the Phytozome predictions. These two proteins were predicted as a result of alternative splicing by Phytozome. However, it was predicted that Anaoc.0008s0618.1 contains an intron that is 8000 nucleotides long, which is much larger compared with those in other 11S protein-coding genes (<200 nucleotides). Therefore, a more thorough gene prediction was performed using AUGUSTUS with the inclusion of all possible transcripts on scaffold 8, a region hosting several 11S protein-coding genes. Through the second prediction using AUGUSTUS, g600.t3 and g602.t3 were considered to be the correct predictions as each of those contained a number of peptides (nine for g600.t3 and five for g602.t3) that cannot be explained by the Phytozome sequences. Moreover, both g600.t3 and g602.t3 have reasonable intron length similar to other 11S proteins. The two proteins also had different coding genes instead of being the result of alternative splicing suggested by Phytozome. The customized cashew protein database can be found in the Supporting Information S2 and the sequences for cashew major seed proteins in the Supporting Information S3. Figure S4 shows the comparison of gene predictions and proteomic evidence mapped to the genome. The number of unique peptides identified in each protein and the sequence coverage results are listed in Table 2. Compared with using the Phytozome-primary and UniProt database, the finalized database detected 168 additional high confidence peptides (FDR < 1%, detected in at least 2 out of 4 replicates, tryptic and chymotryptic raw cashew digests) with 143 peptides derived from cashew allergens or potential allergens.

Effects of Oil-Roasting on Cashew Nuts

The protein database developed in this study has provided opportunities for MS analyses of cashew proteins affected by food processing methods. Oil roasting is a common food processing method for tree nuts. The processing operation is similar to frying, where the food is submerged in an oil bath for a certain amount of time to reach preferred taste and texture properties. Temperatures used for oil roasting are generally between 120 and 180 °C for 2–10 min.^{37,38} Although the effects of dry roasting on tree nuts have been widely studied,^{39,40} research on the effects of oil roasting on tree nuts is rare. In this study, cashew nuts were submerged in an oil bath and oil-roasted



Figure 3. Appearance of oil-roasted cashew nuts prepared under different roasting conditions.

at 139 and 166 °C for 2, 4, 8, and 10 min, separately. The physical appearance of the oil-roasted cashews can be found in Figure 3. The colors of cashew nuts roasted at 139 °C for 10 min and ones roasted at 166 °C for 2 min were visually similar to those of commercially available snack nuts. When protein from oil-roasted cashew nuts was extracted (50 mM Tris-HCl, pH 8.6, with 6 M urea and 20 mM DTT), the soluble protein content decreased substantially with increasing roasting time and temperature (Figure 4). These results were expected because

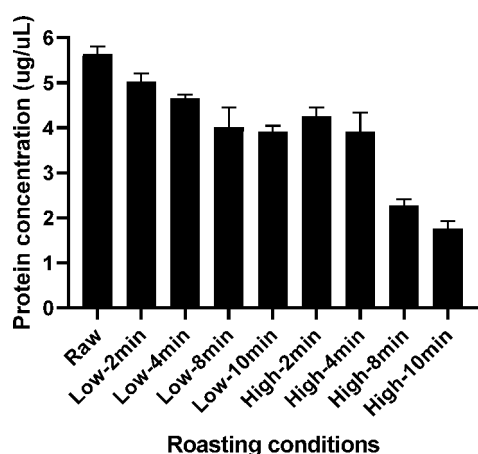


Figure 4. Soluble protein concentration of the oil-roasted cashew extracts [mean $\pm$ standard deviation (SD); $n = 4$, two extractions measured in duplicate].

heat processing induces protein aggregation, thus reducing protein solubility. The degree of decrease in protein solubility was greater in high-temperature roasting compared with the low temperature condition for the same time duration.

The SDS-PAGE protein profile of oil-roasted cashews showed that protein aggregation was induced with increased heating, while some of the proteins were still soluble. The bands at around 20–25 and 7 kDa remained visible after roasting at 166 °C for 10 min, while the intensity of bands at 50 and 17 kDa decreased significantly at high temperature and long roasting time (Figure 5). On the basis of the molecular weight of major cashew proteins reported in previous studies,^{10–12} the bands between 20 and 25 kDa and 7 kDa likely corresponded to Ana o 2 (11S legumins) and to Ana o 3 (2S albumin), respectively. The 50 kDa band is likely to be the Ana o 1 proteins (7S proteins). Other studies on thermal processing of cashews indicated that Ana o 1 is heat-labile while Ana o 2 and 3 are more stable to heat treatment, which is consistent with the SDS-PAGE profile observed in the current study.⁴⁰

The effects of oil-roasting on major cashew proteins were further analyzed using LC-MS/MS. In order to calculate the MS1 abundance of major cashew proteins (2S, 7S, and 11S), the summed MS1 peak area of the top three most abundant peptides was used for the calculation (Figure 6). Only peptides that are

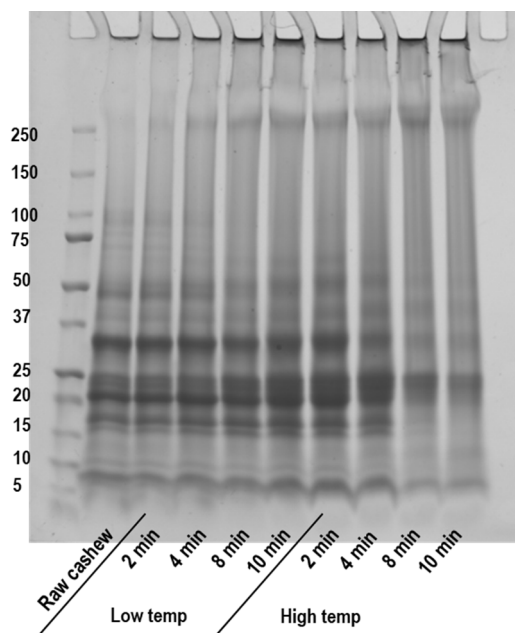


Figure 5. SDS-PAGE of the oil-roasted cashew samples (reducing condition, equal volume loading).

unique to each isoform and were consistently detected in all roasting conditions were included. The results indicated that among the three major cashew protein groups, 11S proteins and 2S proteins account for approximately 63 and 35% of total abundance in raw cashews, respectively (Figure 6A). In samples that have been roasted for 10 min at 166 °C, 11S proteins and 2S proteins represent 57% and 43% of the total abundance, respectively. The relative abundance of the 7S protein changed from 3% (Raw) to 0.1% (High-10m). The results also suggested that the decrease in protein abundance of all 2S proteins was less substantial compared with 11S proteins when roasted at the lower temperature (Figure 6B). The protein with the least reduction in abundance after oil-roasting is Ana o 3.1, retaining 31% abundance in the most processed condition compared with raw, followed by Ana o 2.2 (Figure 6B).

Target Peptide Selection for Oil-Roasted Cashew Nuts

The customized cashew protein database provides many more options for target peptide selection for MS detection method development. A set of predetermined criteria was used for target peptide selection. Three criteria were quality-related, including high confidence peptide identification (FDR < 1%), low mass error (± 5 ppm), and consistent peak quality (isotope dot product values > 0.8 in 34 out of 36 samples) throughout all oil-roasting conditions applied to cashews. Two criteria were related to suitability as target peptides: specificity to cashew (determined by BLASTp search against NCBI protein database) and free of common modifications (oxidation and carbamylation). A list of peptides that passed all the criteria was further

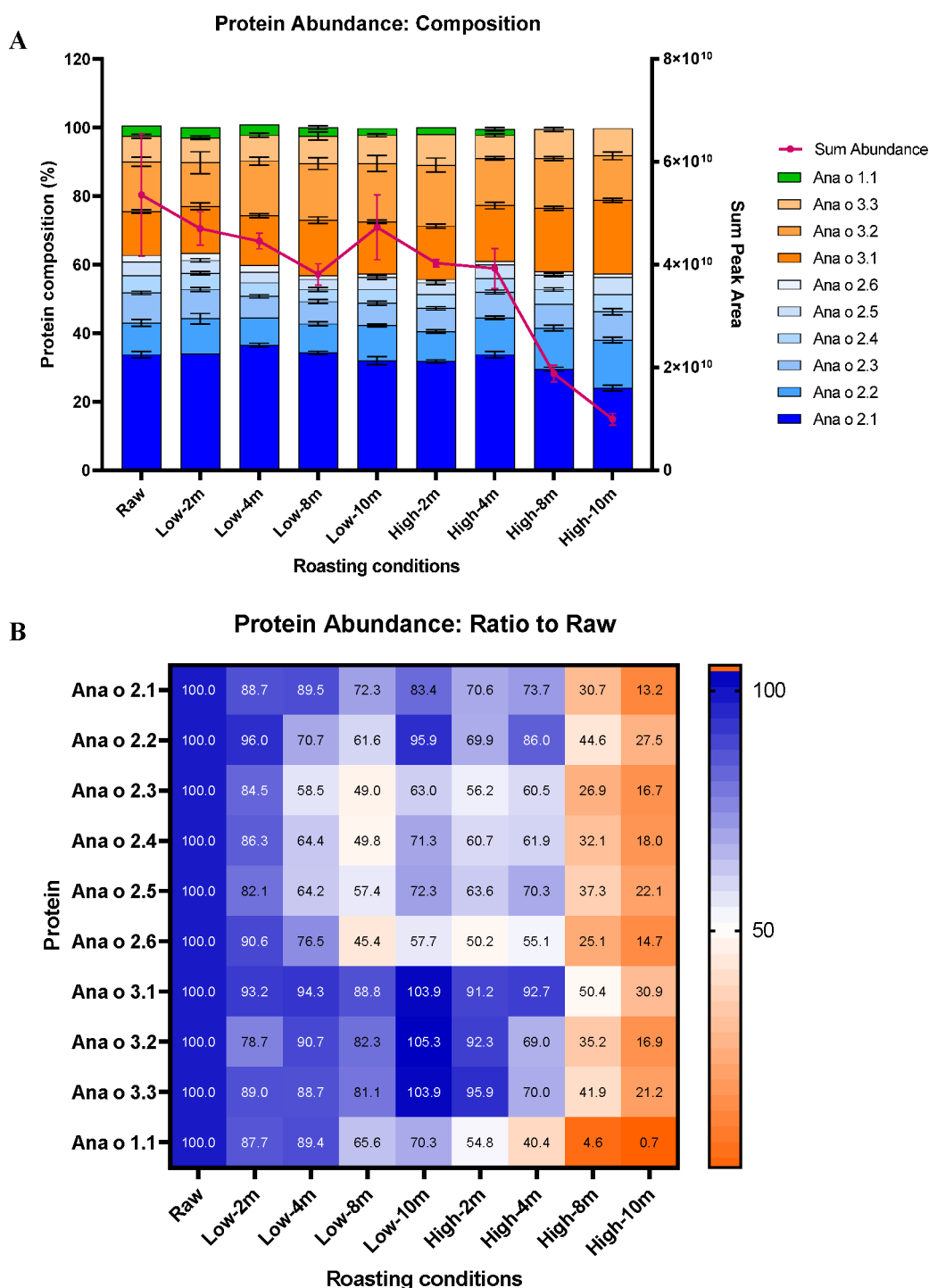


Figure 6. Protein abundance calculated using the sum of the top 3 most abundant peptides (only unique and fully tryptic peptides were considered). See protein ID information in Table 2. (A) Change of composition of cashew major protein with roasting conditions. Percentage composition was calculated by normalizing the value to the sum of protein abundances (red line). Data points represent mean $\pm$ SD ($n = 4$). (B) Mean protein abundance normalized to the counterpart in raw cashews. The mean of the normalized abundance is given on the figure. The Skyline file for MS1 peak abundance calculation can be found at <https://panoramaweb.org/CashewDB.url>.

examined. All modified and miscleaved versions identified for these peptides were investigated. If the total abundance of these versions of the peptide was greater than 10% of that of the unmodified peptide, the peptide was excluded. The modifications investigated in this phase of selection included carbamylation [peptide N-terminus], oxidation [M], and several modifications derived from Maillard reactions (hexose-adduct, *N*- ϵ -carboxymethyl-amine, carboxyethyl-amine, pyrroline, and

pyrroline-amine). The details of these modifications are discussed in the next section.

Overall, 29 peptides were selected as preliminary target peptides for an MS detection method. The sensitivity of these candidate peptides was also investigated. The reconstituted, raw cashew digest was serially diluted to different levels. The on-column protein loadings (calculated based on soluble protein assay results) of the serial dilutions were 1146, 114.6, 22.9, 11.5,

1.1, and 0.1 ng protein, respectively. Nineteen peptides were detected at or below the 11.5 ng level (Table S3). Three of those peptides can be detected with high peak quality (low mass error, high fragment ion dot product value, and symmetrical peak shape) in the 0.1 ng sample (Figure 7). The 19 peptides were

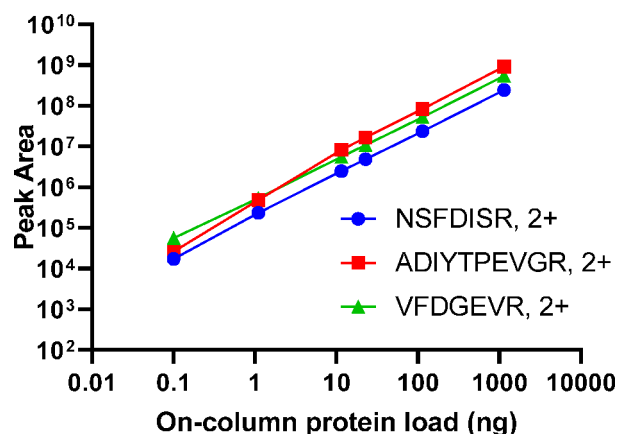


Figure 7. Serial dilution results of the three most sensitive peptides. Individual values are plotted ($n = 4$, two extractions were analyzed in duplicate by MS).

taken for further examination for suitability as target peptides. Among the 19 highly sensitive peptides, 13 peptides were from newly identified protein sequences (Table S3).

The protein database developed in this study reveals 17 additional amino acid sequences for cashew major proteins and potentially allergenic proteins. These additional protein sequences had contributed 13 potential target peptides (Table S3) for cashew detection using MS. Although some peptides (WLQLSVEK, ADIYTPEVGR, and VFDGEVR) have been identified as target peptides in previous studies based on a limited number of cashew protein sequences,^{41–44} the extra protein sequences can provide additional flexibility for target peptide identification. As demonstrated in the oil-roasting experiment, certain proteins/peptides are more susceptible to the food processing effects and thus may not be ideal for the MS analysis of oil-roasted cashew nuts. Moreover, as demonstrated by Xiong et al.,¹⁸ a comprehensive protein database enables the selection of target peptides with enhanced differentiation

between closely related species. This is critical because cashews are closely related to pistachios (*Pistacia vera*), another important food allergen. Overall, the new sequences provide the ability for quality target peptide selection and comprehensive cashew allergen analyses.

Identification of Maillard Reaction Products

Maillard reactions, also known as nonenzymatic browning reactions, are a set of complicated chemical reactions that occur between reducing sugars and primary amine groups, predominantly upon heating.^{45,46} The initial products of Maillard reactions are the result of condensation of primary amines and reducing sugars, including the formation Schiff bases, also known as early glycation products. As Schiff bases degrade, advanced glycation end-products are formed, such as carboxymethyl-amine, carboxyethyl-amine, and pyrraline. Finally, a set of colored compounds that contribute to the browning is formed. Many factors can affect Maillard reactions, such as temperature, pH, water activity, and presence of metal ions.⁴⁶ The main implication of Maillard reaction product (MRP) formation in this study is that covalent modifications alter the peptide mass-to-charge ratio, resulting in a loss of target peptide signals. More broadly speaking, however, the analysis of MRPs may also be useful in understanding the impact of heating on cashew allergenicity.

In this study, several MRPs were investigated, including hexose-adducts (early glycation products), *N*- ϵ -carboxymethyl-amine, carboxyethyl-amine, pyrraline (advanced glycation endproducts), and pyronyl-amine (color products).⁴⁷ Cross-linked products were not investigated due to complications in both the instrument and data analysis components. The criteria to confirm the presence of MRPs include the following: (1) at least one fragment ion containing the modification site should be identified; (2) the peptide should be identified with high confidence determined by Sequest HT in at least 2 out of 4 replicates.

By including the modification settings in the peptide search, several MRPs were identified. Interestingly, KFHLAGNPK, a peptide derived from a cashew 11S protein (Ana o 2.1), was observed with several types of Maillard reaction modifications (see spectra matches in Figure S2). For the modifications of KFHLAGNPK that have peak quality suitable for MS1 peak area calculation, the change of peptide abundance over the course of roasting time and temperature is shown in Figure 8A. The

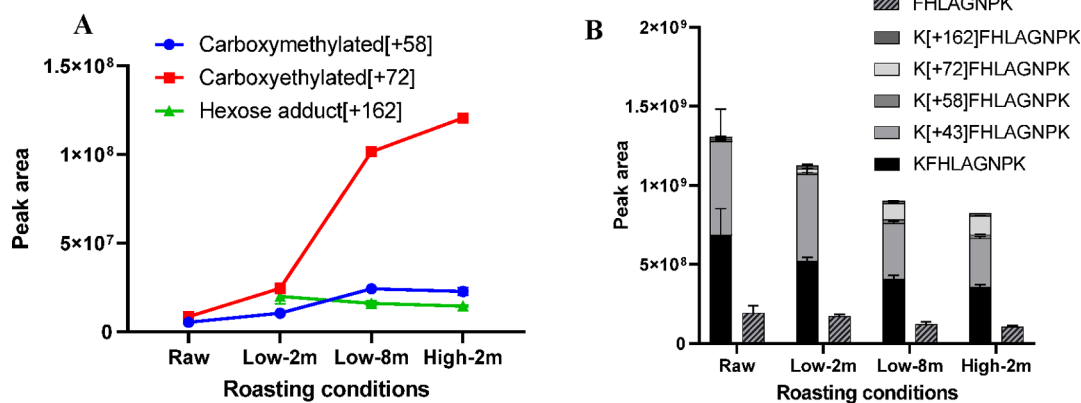


Figure 8. Maillard reaction modifications on peptide KFHLAGNPK. (A) Peptide MS1 abundance of different modifications on KFHLAGNPK (11S peptide) after different roasting treatments. Only peptides that have high MS1 peak quality (peak shape, idotp, etc.) by manual examination were plotted (mean $\pm$ SD, $n = 4$). (B) MS1 abundance of different modifications and missed cleavages for peptide FHLAGNPK.

presence of MRPs in raw cashew could be a result of the preprocessing and/or storage conditions of cashew nuts. Cashew nuts are naturally surrounded by a thick shell that contains cashew nut shell liquid, a toxic liquid that makes the removal of the shell by hand unsafe. Therefore, removal of cashew nut shell liquid by heating (steam roasting, open-pan roasting, etc.) is a common practice in cashew preprocessing.^{48,49} Such processes have likely initiated the Maillard reactions, resulting in the formation of MRPs. The change in MS1 abundance of the modified peptide may indicate the progress of Maillard reactions, with a higher abundance of peptides that have carboxymethyl-amine and carboxyethyl-amine modification with the increase of roasting time and temperature (Figure 8A).

As stated earlier, another purpose of identifying MRPs was to distinguish the peptides that are less affected by Maillard reaction modifications. Figure 8B demonstrates that for some peptides, the peptide abundance of the modified forms can be much higher than the unmodified and properly cleaved form. In the case of peptide FHLGNPK, KFHLGNPK (the missed cleavage form) and the carbamylated form K[+43]FHLGNPK have much higher MS1 abundances compared with the unmodified and fully tryptic-cleaved form. In this case, although peptide FHLGNPK had robust performance (high idotp values, high abundance, etc.) across different roasting conditions and remained with high detectability in the extensively roasted samples, the peptide was excluded from the potential target peptide list.

CONCLUSIONS

In this study, a protein database for cashews (*Anacardium occidentale*) was developed using genomic and bioinformatic tools, and the accuracy of protein sequences of interest was validated using proteomic evidence. Gene predictions produced by an evidence-based approach (Phytozome) and an ab initio approach (AUGUSTUS) were investigated. Although the ab initio gene prediction approach used in this study is relatively simple (without training for prediction parameters), it was sufficient for the analysis focusing on seed storage protein in cashews and target peptide selection. The number of high confidence cashew peptides (FDR < 1%, found in at least 2 out of 4 replicates, tryptic digestion) identified when using the AUGUSTUS database alone was similar to those using a combined database of UniProt and all Phytozome sequences (6578 vs 6595 peptides, see details in Table S2). This result indicated that even a protein database generated using a simple gene prediction may be able to provide a solution for MS discovery analysis when a reference proteome is not available, which is the case for many food commodities. However, because ab initio gene prediction tools use mathematical characteristics of a reference genome to discern genes in the genome of interest, the quality of the prediction is subject to the similarity between the genome of interest and the reference genome. Therefore, further research is needed to assess how robust this approach may be for the analysis of other foods. Despite the challenges associated with the gene prediction without training, the results in this study showed that the newly developed cashew protein database is important for target peptide selection for the MS-based cashew allergen detection method. The database can also facilitate proteomic analyses of cashew nut. Compared with de novo sequencing, which may be advantageous to target peptide selection when a complete proteome is not available, the approach presented in this study can provide protein-level

information, enabling the analysis of protein and protein isoforms.

The proteomic analysis of oil-roasted cashew demonstrated that cashew 2S albumins and 11S legumins are more heat stable compared to 7S vicilin, and thus can serve as quality targets for both MS-based and antibody-based methods, although there might be discrepancies among different isoforms. The criteria used for target peptide selection took various modifications into consideration so that the selected peptides might be less susceptible to modifications caused by Maillard reactions. The selected target peptides can be used for method development for the detection of cashew proteins that have gone through oil roasting.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jproteome.2c00095>.

Supporting Information S1: Figure S1 and S2: Peptide spectra match for peptides containing the 11S protein cleavage sites, 2S protein extra-cysteine, or Maillard reaction induced modifications; Figure S3: Sequence alignment of all 11S protein isoforms; Figure S4: Mapping of peptides identified in raw cashew nuts to the Phytozome cashew genome; Table S1: Parameters used in AUGUSTUS gene prediction; Table S2: Number of high confidence peptides identified when searching against different databases; Table S3: List of potential target peptides (PDF)

Table S4: List of unique peptides identified in cashew 7S, 11S, 2S, oleosins, and cupin proteins (XLSX)

Supporting Information S2: Customized cashew protein sequence database developed in this study (TXT)

Supporting Information S3: Protein sequences for cashew 7S, 11S, 2S, oleosins, and cupin proteins (TXT)

AUTHOR INFORMATION

Corresponding Author

Melanie L. Downs – Food Allergy Research and Resource Program, Department of Food Science and Technology, University of Nebraska–Lincoln, Lincoln, Nebraska 68588, United States; orcid.org/0000-0002-3753-3658; Phone: 402-472-5423; Email: mdowns2@unl.edu

Author

Shimin Chen – Food Allergy Research and Resource Program, Department of Food Science and Technology, University of Nebraska–Lincoln, Lincoln, Nebraska 68588, United States; orcid.org/0000-0001-9267-558X

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jproteome.2c00095>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Cashew genomic sequences and other corresponding information were used with permission and provided by the Joint Genome Institute. These sequence data were produced by the US Department of Energy Joint Genome Institute (<http://www.jgi.doe.gov/>) in collaboration with the user community. We also

thank Holland Computing Center at the University of Nebraska–Lincoln for providing supercomputing resources for bioinformatic data processing. We also thank Mohamed Abdelmoteleb for providing initial guidance for the bioinformatics analysis. Research funding was provided by the Food Allergy Research and Resource Program (FARRP) at the University of Nebraska, a food industry-sponsored consortium of over 100 food processing companies and their suppliers. This research is also part of a collaboration between FARRP and Thermo Fisher Scientific.

REFERENCES

- (1) United States Congress. *Food Allergen Labeling and Consumer Protection Act of 2004, Public Law 108–282*; United States, 2004.
- (2) European Union. Regulation (EU) No 1169/2011. *Off. J. Eur. Union* **2011**, 304, 18–63.
- (3) US Food and Drug Administration. Guidance for Industry: Questions and Answers Regarding Food Allergens (Edition 4), <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-industry-questions-and-answers-regarding-food-allergens-edition-4> (accessed on June 16, 2021).
- (4) McWilliam, V.; Koplin, J.; Lodge, C.; Tang, M.; Dharmage, S.; Allen, K. The Prevalence of Tree Nut Allergy: A Systematic Review. *Curr. Allergy Asthma Rep.* **2015**, 15 (9), 555.
- (5) Van Der Valk, J. P. M.; Dubois, A. E. J.; Gerth Van Wijk, R.; Wichers, H. J.; De Jong, N. W. Systematic Review on Cashew Nut Allergy. *Allergy Eur. J. Allergy Clin. Immunol.* **2014**, 69 (6), 692–698.
- (6) Willison, L. N.; Tawde, P.; Robotham, J. M.; Penney, R. M.; Teuber, S. S.; Sathe, S. K.; Roux, K. H. Pistachio Vicilin, Pis v 3, Is Immunoglobulin E-Reactive and Cross-React with the Homologous Cashew Allergen, Ana o 1. *Clin. Exp. Allergy* **2008**, 38 (7), 1229–1238.
- (7) Savvatanos, S.; Konstantinopoulos, A. P.; Borga, A.; Stavroulakis, G.; Lidholm, J.; Borres, M. P.; Manousakis, E.; Papadopoulos, N. G. Sensitization to Cashew Nut 2S Albumin, Ana o 3, Is Highly Predictive of Cashew and Pistachio Allergy in Greek Children. *J. Allergy Clin. Immunol.* **2015**, 136 (1), 192–194.
- (8) Clark, A. T.; Anagnostou, K.; Ewan, P. W. Cashew Nut Causes More Severe Reactions than Peanut: Case-Matched Comparison in 141 Children. *Eur. J. Allergy Clin. Immunol.* **2007**, 62 (8), 913–916.
- (9) Grigg, A.; Hanson, C.; Davis, C. M. Cashew Allergy Compared to Peanut Allergy in a US Tertiary Care Center. *Pediatr. Asthma. Allergy Immunol.* **2009**, 22 (3), 101–104.
- (10) Wang, F.; Robotham, J. M.; Teuber, S. S.; Tawde, P.; Sathe, S. K.; Roux, K. H. Ana o 1, a Cashew (*Anacardium Occidentale*) Allergen of the Vicilin Seed Storage Protein Family. *J. Allergy Clin. Immunol.* **2002**, 110 (1), 160–166.
- (11) Wang, F.; Robotham, J. M.; Teuber, S. S.; Sathe, S. K.; Roux, K. H. Ana o 2, a Major Cashew (*Anacardium Occidentale* L.) Nut Allergen of the Legumin Family. *Int. Arch. Allergy Immunol.* **2003**, 132 (1), 27–39.
- (12) Robotham, J. M.; Wang, F.; Seamon, V.; Teuber, S. S.; Sathe, S. K.; Sampson, H. A.; Beyer, K.; Seavy, M.; Roux, K. H. Ana o 3, an Important Cashew Nut (*Anacardium Occidentale* L.) Allergen of the 2S Albumin Family. *J. Allergy Clin. Immunol.* **2005**, 115 (6), 1284–1290.
- (13) Teuber, S. S.; Sathe, S. K.; Peterson, W. R.; Roux, K. H. Characterization of the Soluble Allergenic Proteins of Cashew Nut (*Anacardium Occidentale* L.). *J. Agric. Food Chem.* **2002**, 50 (22), 6543–6549.
- (14) Kim, E. H.; Burks, A. W. Food Allergy Immunotherapy: Oral Immunotherapy and Epicutaneous Immunotherapy. *Allergy* **2020**, 75 (6), 1337–1346.
- (15) Downs, M. L.; Taylor, S. L. Effects of Thermal Processing on the Enzyme-Linked Immunosorbent Assay (ELISA) Detection of Milk Residues in a Model Food Matrix. *J. Agric. Food Chem.* **2010**, 58 (18), 10085–10091.
- (16) Jablonski, J. E.; Fu, T.-J. J.; Jackson, L. S.; Gendel, S. M. Summary for Policymakers. In *Climate Change 2013—The Physical Science Basis*; Intergovernmental Panel on Climate Change, Ed.; Cambridge University Press: Cambridge, 2010; Vol. 93, pp 1–30.
- (17) Cucu, T.; Platteau, C.; Taverniers, I.; Devreese, B.; De Loose, M.; De Meulenaer, B. Effect of Partial Hydrolysis on the Hazelnut and Soybean Protein Detectability by ELISA. *Food Control* **2013**, 30 (2), 497–503.
- (18) Xiong, W.; McFarland, M. A.; Pirone, C.; Parker, C. H. Selection of Tree Nut Allergen Peptide Markers: A Need for Improved Protein Sequence Databases. *J. AOAC Int.* **2019**, 102 (5), 1263–1270.
- (19) Pirone-Davies, C.; McFarland, M. A.; Parker, C. H.; Adachi, Y.; Croley, T. R. The Utility of Genomic and Transcriptomic Data in the Construction of Proxy Protein Sequence Databases for Unsequenced Tree Nuts. *Biology (Basel)* **2020**, 9 (5), 104.
- (20) Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. The PRIDE Database and Related Tools and Resources in 2019: Improving Support for Quantification Data. *Nucleic Acids Res.* **2019**, 47 (D1), D442–D450.
- (21) Martens, L.; Vandekerckhove, J.; Gevaert, K. DBToolkit: Processing Protein Databases for Peptide-Centric Proteomics. *Bioinformatics* **2005**, 21 (17), 3584–3585.
- (22) Stanke, M.; Steinkamp, R.; Waack, S.; Morgenstern, B. AUGUSTUS: A Web Server for Gene Finding in Eukaryotes. *Nucleic Acids Res.* **2004**, 32 (Web Server), W309–W312.
- (23) Stanke, M.; Keller, O.; Gunduz, I.; Hayes, A.; Waack, S.; Morgenstern, B. AUGUSTUS: Ab Initio Prediction of Alternative Transcripts. *Nucleic Acids Res.* **2006**, 34 (Web Server), W435–W439.
- (24) Goodman, R. E.; Ebisawa, M.; Ferreira, F.; Sampson, H. A.; van Ree, R.; Vieths, S.; Baumert, J. L.; Bohle, B.; Lalithambika, S.; Wise, J.; et al. AllergenOnline: A Peer-Reviewed, Curated Allergen Database to Assess Novel Food Proteins for Potential Cross-Reactivity. *Mol. Nutr. Food Res.* **2016**, 60 (5), 1183–1198.
- (25) Codex Alimentarius. *Guideline for the Conduct of Food Safety Assessment of Foods Derived from the Recombinant-DNA Plants*; Codex Alimentarius, 2003.
- (26) Reitsma, M.; Bastiaan-Net, S.; Sforza, S.; Van Der Valk, J. P. M.; Van Gerth Van Wijk, R.; Savelkoul, H. F. J.; De Jong, N. W.; Wichers, H. J. Purification and Characterization of *Anacardium Occidentale* (Cashew) Allergens Ana o 1, Ana o 2, and Ana o 3. *J. Agric. Food Chem.* **2016**, 64 (5), 1191–1201.
- (27) Adachi, M.; Kanamori, J.; Masuda, T.; Yagasaki, K.; Kitamura, K.; Mikami, B.; Utsumi, S. Crystal Structure of Soybean 11S Globulin: Glycinin A3B4 Homohexamer. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, 100 (12), 7395–7400.
- (28) Wallowitz, M.; Peterson, W. R.; Uratsu, S.; Comstock, S. S.; Dandekar, A. M.; Teuber, S. S. Jug r 4, a Legumin Group Food Allergen from Walnut (*Juglans Regia* Cv. Chandler). *J. Agric. Food Chem.* **2006**, 54 (21), 8369–8375.
- (29) Huang, A. H. C. C. Oleosins and Oil Bodies in Seeds and Other Organs. *Plant Physiol.* **1996**, 110 (4), 1055–1061.
- (30) Lin, L. J.; Liao, P. C.; Yang, H. H.; Tzen, J. T. C. Determination and Analyses of the N-Termini of Oil-Body Proteins, Steroleosin, Caleosin and Oleosin. *Plant Physiol. Biochem.* **2005**, 43 (8), 770–776.
- (31) Tsuji, H.; Bando, N.; Hiemori, M.; Yamanishi, R.; Kimoto, M.; Nishikawa, K.; Ogawa, T. Purification and Characterization of Soybean Allergen Gly m Bd 28K. *Biosci. Biotechnol. Biochem.* **1997**, 61 (6), 942–947.
- (32) Hiemori, M.; Ito, H.; Kimoto, M.; Yamashita, H.; Nishizawa, K.; Maruyama, N.; Utsumi, S.; Tsuji, H. Identification of the 23-KDa Peptide Derived from the Precursor of Gly m Bd 28K, a Major Soybean Allergen, as a New Allergen. *Biochim. Biophys. Acta - Gen. Subj.* **2004**, 1675 (1–3), 174–183.
- (33) Xiang, P.; Haas, E. J.; Zeece, M. G.; Markwell, J.; Sarath, G. C-Terminal 23 KDa Polypeptide of Soybean Gly m Bd 28 K Is a Potential Allergen. *Planta* **2004**, 220 (1), 56–63.
- (34) Stanke, M.; Schöffmann, O.; Morgenstern, B.; Waack, S. Gene Prediction in Eukaryotes with a Generalized Hidden Markov Model That Uses Hints from External Sources. *BMC Bioinformatics* **2006**, 7, 1–11.

- (35) Wang, X.; Codreanu, S. G.; Wen, B.; Li, K.; Chambers, M. C.; Liebler, D. C.; Zhang, B. Detection of Proteome Diversity Resulted from Alternative Splicing Is Limited by Trypsin Cleavage Specificity. *Mol. Cell. Proteomics* **2018**, *17* (3), 422–430.
- (36) Mills, D. B.; Francis, W. R.; Vargas, S.; Larsen, M.; Elemans, C. P. H.; Canfield, D. E.; Wörheide, G. The Last Common Ancestor of Animals Lacked the HIF Pathway and Respired in Low-Oxygen Environments. *eLife* **2018**, *7*, 1–17.
- (37) Shi, X.; Davis, J. P.; Xia, Z.; Sandeep, K. P.; Sanders, T. H.; Dean, L. O. Characterization of Peanuts after Dry Roasting, Oil Roasting, and Blister Frying. *LWT - Food Sci. Technol.* **2017**, *75*, 520–528.
- (38) Perren, R.; Escher, F. E. Impact of Roasting on Nut Quality. In *Improving the Safety and Quality of Nuts*; Elsevier, 2013; pp 173–197.
- (39) Mattison, C. P.; Grimm, C. C.; Li, Y.; Chial, H. J.; McCaslin, D. R.; Chung, S.-Y. Y.; Bren-Mattison, Y.; Wasserman, R. L. Identification and Characterization of Ana o 3 Modifications on Arginine-111 Residue in Heated Cashew Nuts. *J. Agric. Food Chem.* **2017**, *65* (2), 411–420.
- (40) Mattison, C. P.; Bren-Mattison, Y.; Vant-Hull, B.; Vargas, A. M.; Wasserman, R. L.; Grimm, C. C. Heat-Induced Alterations in Cashew Allergen Solubility and IgE Binding. *Toxicol. Reports* **2016**, *3*, 244–251.
- (41) Korte, R.; Lepski, S.; Brockmeyer, J. Comprehensive Peptide Marker Identification for the Detection of Multiple Nut Allergens Using a Non-Targeted LC–HRMS Multi-Method. *Anal. Bioanal. Chem.* **2016**, *408* (12), 3059–3069.
- (42) Sealey-Voyksner, J.; Zweigenbaum, J.; Voyksner, R. Discovery of Highly Conserved Unique Peanut and Tree Nut Peptides by LC–MS/MS for Multi-Allergen Detection. *Food Chem.* **2016**, *194*, 201–211.
- (43) Planque, M.; Arnould, T.; Delahaut, P.; Renard, P.; Dieu, M.; Gillard, N. Development of a Strategy for the Quantification of Food Allergens in Several Food Products by Mass Spectrometry in a Routine Laboratory. *Food Chem.* **2019**, *274*, 35–45.
- (44) Luparelli, A.; Losito, I.; De Angelis, E.; Pilolli, R.; Lambertini, F.; Monaci, L. Tree Nuts and Peanuts as a Source of Beneficial Compounds and a Threat for Allergic Consumers: Overview on Methods for Their Detection in Complex Food Products. *Foods* **2022**, *11* (5), 728.
- (45) Martins, S. I. F. S.; Jongen, W. M. F.; van Boekel, M. A. J. S. A Review of Maillard Reaction in Food and Implications to Kinetic Modelling. *Trends Food Sci. Technol.* **2000**, *11*, 364–373.
- (46) Lund, M. N.; Ray, C. A. Control of Maillard Reactions in Foods: Strategies and Chemical Mechanisms. *J. Agric. Food Chem.* **2017**, *65* (23), 4537–4552.
- (47) Downs, M. L.; Baumert, J. L.; Taylor, S. L.; Mills, E. N. C. Mass Spectrometric Analysis of Allergens in Roasted Walnuts. *J. Proteomics* **2016**, *142*, 62–69.
- (48) Mattison, C. P.; Malveira Cavalcante, J.; Izabel Gallão, M.; Sousa de Brito, E. Effects of Industrial Cashew Nut Processing on Anacardic Acid Content and Allergen Recognition by IgE. *Food Chem.* **2018**, *240*, 370–376.
- (49) Trox, J.; Vadivel, V.; Vetter, W.; Stuetz, W.; Scherbaum, V.; Gola, U.; Nohr, D.; Biesalski, H. K. Bioactive Compounds in Cashew Nut (*Anacardium occidentale* L.) Kernels: Effect of Different Shelling Methods. *J. Agric. Food Chem.* **2010**, *58* (9), 5341–5346.

Recommended by ACS

How (Not) to Generate a Highly Predictive Biomarker Panel Using Machine Learning

Heather Desaire.

AUGUST 25, 2022
JOURNAL OF PROTEOME RESEARCH

READ 

Variation in Milk Proteins Across Lactation in *Pongo pygmaeus* and *Gorilla gorilla*

Timothy P. Cleland and Michael L. Power

OCTOBER 24, 2022
JOURNAL OF PROTEOME RESEARCH

READ 

Novel Deep Eutectic Solvent-Based Protein Extraction Method for Pottery Residues and Archeological Implications

Manasij Pal Chowdhury, Michael Buckley, *et al.*

OCTOBER 21, 2022
JOURNAL OF PROTEOME RESEARCH

READ 

Human Plasma Proteome During Normal Pregnancy

Adi L. Tarca, Nardhy Gomez-Lopez, *et al.*

SEPTEMBER 26, 2022
JOURNAL OF PROTEOME RESEARCH

READ 

Get More Suggestions >