Food Chemistry 348 (2021) 129110

Contents lists available at ScienceDirect

Food Chemistry

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

Protein extraction protocols for optimal proteome measurement and arginine kinase quantitation from cricket *Acheta domesticus* for food safety assessment

Utpal Bose [a,1], James A. Broadbent [a,1], Angéla Juhász [b], Shaymaviswanathan Karnaneedi [c], Elecia B. Johnston [c], Sally Stockwell [a], Keren Byrne [a], Vachiranee Limviphuvadh [d,e], Sebastian Maurer-Stroh [d,e,f], Andreas L. Lopata [c], Michelle L. Colgrave [a,b,*]

[a] CSIRO Agriculture and Food, 306 Carmody Rd, St Lucia, QLD 4067, Australia
[b] Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science, School of Science, Edith Cowan University, Joondalup, WA 6027, Australia
[c] Molecular Allergy Research Laboratory, Australian Institute of Tropical Health and Medicine, James Cook University, Townsville, QLD, Australia
[d] Biomolecular Function Discovery Division, Bioinformatics Institute, Agency for Science, Technology and Research, Singapore
[e] IFCS Programme, Singapore Institute for Food and Biotechnology Innovation, Agency for Science, Technology and Research, Singapore
[f] Department of Biological Sciences, National University of Singapore, Singapore

ARTICLE INFO

Keywords:
Cricket
Protein extraction
Protocol optimisation
Proteomics
LC-MRM-MS
Allergen
Cross-reactive
Arginine kinase

ABSTRACT

Insects have been consumed by people for millennia and have recently been proposed as a complementary, sustainable source of protein to feed the world's growing population. Insects and crustaceans both belong to the arthropod family. Crustacean (shellfish) allergies are common and potentially severe; hence, the cross-reactivity of the immune system with insect proteins is a potential health concern. Herein, LC-MS/MS was used to explore the proteome of whole, roasted whole and roasted powdered cricket products. Eight protein extraction protocols were compared using the total number of protein and distinct peptide identifications. Within these data, 20 putative allergens were identified, of which three were arginine kinase (AK) proteoforms. Subsequently, a multiple reaction monitoring MS assay was developed for the AK proteoforms and applied to a subset of extracts. This targeted assay demonstrated that allergen abundance/detectability varies according to the extraction method as well as the food processing method.

1. Introduction

Crickets (*Acheta domesticus*) possess a high quality nutritional profile that includes proteins (~45% of dry mass), fats (~26% of dry mass), vitamins and minerals (Montowska et al., 2019; Stull et al., 2018). The protein content and nutritional value varies substantially between cricket sub-species, developmental stages, source of collection (farm or open-field) and processing method (boiling, baking or roasting) (Kouřimská & Adámková, 2016; Yi et al., 2013). Furthermore, over the past five years several review papers have been published regarding bioactive peptides from insect protein hydrolysates with antihypertensive, antimicrobial, antioxidant and gut health properties (Jantzen da Silva Lucas et al., 2020; de Castro et al., 2018; Montowska et al., 2019). Although insect proteins have shown positive health benefits, their possible allergenic status remains a major concern. For instance, the protein arginine kinase (AK) has shown cross-allergenicity between *Macrobrachium* spp. (prawns or shrimp) and field cricket *Gryllus bimaculatus* (Srinroch et al., 2015). Likewise, IgE cross-reactivity is observed between tropomyosin from cricket and shrimp, as shown in human shrimp allergic sera (Hall et al., 2018; Kamemura et al., 2019; Ruethers

Abbreviations: LC-MS, Liquid chromatography-mass spectrometry; LC-MRM-MS, Liquid chromatography-multiple reaction monitoring-mass spectrometry; AK, Arginine kinase; SDS-PAGE, Sodium dodecyl sulphate-polyacrylamide gel electrophoresis; FASP, filter-aided sample preparation; DTT, dithiothreitol; HCA, Hierarchical clustering analysis; GO, Gene ontology; GRAVY, Grand average hydropathicity index; pI, isoelectric point; IDA, Information dependent acquisition.

* Corresponding author at: CSIRO Agriculture and Food, 306 Carmody Rd, St Lucia, QLD 4067, Australia.
E-mail address: michelle.colgrave@csiro.au (M.L. Colgrave).

[1] Authors contributed equally.

https://doi.org/10.1016/j.foodchem.2021.129110
Received 10 December 2020; Received in revised form 6 January 2021; Accepted 11 January 2021
Available online 19 January 2021
0308-8146/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

et al., 2018). As a result, research that aims to obtain new knowledge regarding the safety of dietary insect-based proteins is warranted.

Protein extraction is critical for proteome measurement where buffer systems will impact the representation of physicochemical properties and functional classes of the detected proteins (Bose et al., 2019a, 2019b; Ndiritu et al., 2017; Rodríguez-Rojas & Rolff, 2020). Insect tissues contain large amounts of lipids, fatty acids, vitamins and minerals that can interfere with protein extraction and downstream measurement (Belghit et al., 2019). The use of aqueous extraction followed by fat partitioning across five insect species has shown a species-specific difference in protein yield (Yi et al., 2013). Furthermore, *n*-hexane-based defatting followed by sonication has been shown to improve the yield of extractable protein by 35–94% in mealworm larvae, adult crickets and silkworm pupae (Choi et al., 2017). Likewise, sample processing that commences with defatting followed by alkaline buffer extraction and sonication increases the protein yield up to 57.5% and 55.2% for grasshopper and honeybee, respectively (Mishyna et al., 2019). Sodium phosphate buffer has also been used to extract allergen homologues from *Macrobrachium* spp. and the field cricket (*Gryllus bimaculatus*) (Srinroch et al., 2015). Therein the study revealed that the potent cross-activating allergen in *G. bimaculatus* and *Macrobrachium* spp. is arginine kinase. Until now, the majority of studies have focused on estimating crude protein content in insects using techniques such as protein combustion (Amarender et al., 2020) or protein-to-nitrogen conversion rate (Yi et al., 2013). Whereas protein detection and identification has been performed with sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) followed by mass spectrometry.

The advancement of genomics-guided MS-based proteomics techniques and data mining strategies have enabled the identification and quantitation of thousands of proteins from complex biological samples. Routine targeted and untargeted MS experiments have been used to measure these proteins in terms of their abundance, localisation, interactions and modifications to help interpret phenotypes (Bose et al., 2019). In all cases, protein extraction is critical. Herein, different buffer compositions that are commonly used to extract and analyse proteins from cricket (*A. domesticus*) were compared using bottom-up LC-MS/MS proteomics. Subsequently a liquid chromatography-multiple reaction monitoring-mass spectrometry (LC-MRM-MS) experiment was performed to examine the ability to efficiently and repeatably extract AK, which can be used as a marker of cricket presence in food products.

2. Materials and methods

2.1. Cricket samples

Whole cricket products were collected from the insect food company Grubs Up (Pinjarra, Australia) and roasted slowly at 150 °C for 4 h. The roasted whole cricket (150 °C for 4 h) and roasted cricket powder samples (150 °C for 4 h; *A. domesticus*) were collected from Grilo (Byron Bay, Australia). The Grubs Up samples will be termed as "whole cricket" throughout the manuscript. Samples were ground to powder using a mortar and pestle. The powder was passed through a sieve (commercial tea strainer) to collect a uniform flour for protein extraction.

2.2. Buffer compositions and protein extraction

Eight protein extraction protocols were used to solubilize proteins. Buffer compositions for the extraction protocols are as follows: Protocol 1 (*P*1): 100 mM Tris-HCl, 4% SDS, 50 mM dithiothreitol (DTT) (pH 7.6); Protocol 2 (*P*2): 200 mM Tris-HCl, 2 M urea, 50 mM DTT; Protocol 3 (*P*3): 20 mM Tris, 8 M urea, 2 M thiourea, 50 mM DTT; Protocol 4 (*P*4): 10 mM HEPES, 6 M urea, 2 M thiourea, 50 mM DTT (pH 8.0); Protocol 5 (*P5*): 50 mM Tris-HCl, 50 mM DTT, 100 mM KCl; Protocol 6 (*P*6): defatting with *n*-pentane followed by extraction with 100 mM Tris-HCl, 8 M urea, 50 mM DTT (pH 8.5); Protocol 7 (*P7*): 50 mM Tris-HCl, 1 M NaCl (pH 7.2); Protocol 8 (*P8*): 50 mM Tris-HCl, 6 M Urea, 1 M NaCl (pH 7.2). All buffers were freshly prepared on the experiment day.

The cricket flour samples were prepared as previously described (Fallahbaghery et al., 2017). Samples (n = 4 per buffer) were weighed into 1.5 mL micro-tubes; 200 μL (10 μL/mg) each of the afore-mentioned extraction buffers were added to respective samples followed by vortex mixing until a uniform solution was achieved. Samples were then agitated by sonication for 5 min. For *P*1, sample tubes were placed on a thermomixer (Eppendorf, Germany) at 600 rpm; 55 °C; 30 min. For the remaining protocols (*P*2-*P5*), sample tubes were placed on a thermomixer (600 rpm; room temperature; 30 min) before centrifugation for 15 min at 20,800 × *g*. Supernatants were used for subsequent processing.

For *P*6, flour samples were first defatted with 200 μL (w/v) of *n*-pentane prior to placing them on a thermomixer. In brief, ~20 mg aliquots of cricket flour were weighed into 1.5 mL micro-tubes and mixed with 200 μL of *n*-pentane. The tubes were placed on a rotator for 15 min, followed by centrifugation (5 min) at 20,800 × *g*. The supernatant was discarded, and the process was repeated two more times. The pellets were air dried and redissolved in 200 μL of extraction buffer.

In parallel, proteins were extracted using P7 buffer. Therein, sample tubes were incubated for 16 h at 4 °C in a tumbling motion followed by centrifugation at 10,000 × *g* for 30 min at 4 °C before collecting the supernatant. The pellet was immersed and homogenised in *P8* buffer and incubated, centrifuged, and supernatant collected as per *P7*. Extracts of *P7* and *P8* were filtered using glass fibre filters (Sartorius Stedim Biotech, Germany).

2.3. Protein concentration estimation by Bradford assay

Protein concentration estimation of the extracts was performed using a Bradford colorimetric assay. Samples were diluted 10x, 20x and 40x with water and standards were generated with BSA ranging from 7.8 μg/mL to 500 μg/mL. Sample dilutions and standards (10 μL) were added to 200 μL of Bradford Reagent (BioRad: 1:5 in H_2O), incubated for 5 min at RT and read at 595 nm.

2.4. Protein digestion

Protein extracts (100 μg) were transferred to 30 kDa MWCO filters (Merck Millipore, Bayswater, Vic). The protein on the filter was washed twice with UA buffer consisting of 8 M urea in 100 mM Tris-HCl (pH 8.5) with centrifugation for 15 min at 20,800 × *g*. Iodoacetamide (50 mM, 100 μL) in UA was added to the filters for cysteine alkylation with incubation in the dark for 20 min before centrifugation at 20,800 × *g* for 10 min. The buffer was exchanged with 100 mM ammonium bicarbonate (pH 8.0) by two consecutive wash/centrifugation cycles. The digestion enzyme, bovine trypsin (Sigma-Aldrich, NSW, Australia), was prepared as 250 μg/mL in 100 mM ammonium bicarbonate (pH 8.0) and 200 μL was added to each filter for overnight incubation at 37 °C. The filters were transferred to fresh collection tubes and centrifuged for 15 min at 20,800 × *g*. The filters were washed with 200 μL of 100 mM ammonium bicarbonate, and the peptides in the filtrates combined and dried by vacuum concentration.

2.5. Proteome measurement

Digested peptides were resuspended in 100 μL of 1% formic acid before analysis by LC–MS/MS as previously described (Bose et al., 2019; Fallahbaghery et al., 2017). In brief, 5 μL of peptides were chromatographically separated and detected with an Ekspert nanoLC415 (Eksigent, Dublin, CA, USA) coupled to a TripleTOF 6600 MS (SCIEX, Redwood City, CA, USA). The peptides were desalted for 3 min on a polar C18 ProteCol trap column (Trajan; 3 μm Particle Size × 300 Å Pore Size, 10 mm × 300 μm ID) at a flow rate of 10 μL/min of 0.1% formic acid and separated on a ChromXP C18 (3 μm, 120 Å, 150 mm × 0.3 mm) column at a flow rate of 5 μL/min. The solvents used were: (A) aqueous

5% DMSO, 0.1% formic acid; and (B) aqueous 5% DMSO, 0.1% formic acid, 90% acetonitrile. A linear gradient from 3% to 25% solvent B over 68 min was employed followed by 25–35% B over 5 min, and then 35–80% over 2 min. Next, 3 min hold at 80% B, return to 3% B over 1 min, and 8 min of re-equilibration. The eluent from the HPLC was directly coupled to the Optiflow source of the TripleTOF 6600 MS. The ion spray voltage was set to 4500 V; the curtain gas was set to 138 kPa (30 psi), and the ion source gas 1 and 2 (GS1 and GS2) were set to 138 and 138 kPa (30 and 30 psi). The heated interface was set to 100 °C.

2.6. Database creation and protein identification

The cricket database was prepared by combining UniProt "Orthoptera" sequences (access date 06/04/2020) with genomes from NCBI *Locusta migratoria* whole genome sequencing (WGS), *Teleogryllus occipitalis* WGS, *Laupala kohalensis* reference genome, *Xenocatantops brachycerus* WGS, NCBI_04042020_*Gryllidea*_prot and IUIS reference allergens for crustaceans, house dust mite and cockroach. Contigs originating from WGS projects were used for open reading frame (ORF) predictions, using a minimum of 100 nucleotide ORF cut-off value. Predicted ORFs were translated and functionally annotated using Pfam domain and Blast2GO analyses before appending to the UniProt database. These protein sequences were merged with the common repository of adventitious proteins (cRAP) sequences as well as the iRT pseudo-protein sequence (2,600,927 sequences).

ProteinPilot v5.0.3 software (SCIEX) incorporating the Paragon and ProGroup algorithms (Shilov et al., 2007) was used for protein identification. The MS/MS data were searched against the abovementioned protein database. The search parameters were defined as iodoacetamide modified cysteine; trypsin as the digestion enzyme; instrument as 6600 TripleTOF; and, thorough search effort. The database search results from the combined cricket flour analyses were manually curated to yield protein identifications at a 1% global false discovery rate (FDR) determined by the SCIEX FDR tool (Tang et al., 2008). The SCIEX protein alignment template was used to unify protein identifications between the different samples. Peptide results were unified across the samples using a custom R script that filters for high quality distinct peptides meeting a 1% FDR threshold, fully tryptic, fixed carbamidomethyl (Cys), variable oxidation (Met) and *N*-terminal Gln to pyro-Glu modifications. To detect the arginine kinases (AK) within the cricket samples, a database search was performed that combined the data from all samples. The resulting output file was searched thoroughly to identify AKs that met a 1% global FDR threshold.

2.7. Physicochemical properties, multivariate and bioinformatics analysis

Based on protein identification frequency, the five top performing protein extraction protocols were selected for comparative analysis, physicochemical determination and functional annotation. Physicochemical parameters were determined for the protein sets detected from each of the extraction protocols for the three types of cricket samples. In this regard, protein sequence files in *.FASTA format were created for each set of proteins using the unified protein identifications determined by the SCIEX protein alignment template. These sequences were then processed using a custom Python script to determine Grand average hydropathicity index (GRAVY), aromaticity, isoelectric point (pI) and molecular weight for each protein in each set. Groups were compared using one-way ANOVA with Tukey's HSD test.

Hierarchical Clustering Analysis (HCA) was performed with Biovinci version 1.1.5 (BioTutoring Inc., San Diego, California, USA) and Morpheus software (Broad Institute, Cambridge MA, USA; https://software.broadinstitute.org/morpheus/). Gene Ontology (GO) and InterPro domain mapping were performed to obtain functional classification data using BLAST2GO software (Conesa et al., 2005). Insect proteins downloaded from UniProt (8,082,118 proteins; accessed on 05/06/2020) were used as a background for the BLAST2GO analysis.

2.8. Allergen protein mapping from cricket samples

The database search results comprising all cricket samples extracted with all extraction methods were searched against a database of Allergen families (AllFam, Medical University Vienna, Austria; http://www.meduniwien.ac.at/allfam/) to identify allergen domain-containing proteins. Allergen protein families with known allergy related Pfam domains characteristic of insect and crustacean species were retrieved from the AllFam database. Protein sequences from the World Health Organization and International Union of Immunological Societies (WHO/IUIS; http://www.allergen.org/) reference allergens were retrieved from the UniProt Knowledgebase. Pfam domain mapping of the identified cricket proteins was performed using hidden Markov models (HMMER3) in CLC Genomics Workbench 20.0.4. Protein sequences with Pfam domains characteristic of the collected allergen families were aligned to the reference protein sequences. Proteins with matching Pfam domain profiles were accepted as potential allergens and further investigated using sequence alignment. Proteins were also analysed using the AllerCatPro web tool to investigate their putative allergenicity (Maurer-Stroh et al., 2019). Pairwise comparisons of the identified protein sequences and reference allergens was performed using CLC Genomics Workbench 20.0.4. The percentage sequence identity between the identified proteins and the reference allergens were used to generate a heatmap using Morpheus software.

2.9. Relative quantitation of arginine kinase by MRM

Arginine kinase proteins were imported to Skyline software to identify suitable MRM transitions (MacLean et al., 2010). *In silico* digestion yielded 30 peptides representing three AKs. These were used initially to acquire MRM data from a pooled sample of all the cricket protein extracts. The results from these analyses were used to refine the transitions and schedule retention times. Peptides were selected that yielded intense peaks and were fully tryptic, with no variable modifications or missed cleavages.

3. Results

3.1. Extraction protocols and protein estimation

The efficiency of a suite of protein extraction buffers was examined by comparing the protein identification yields (based on unified protein accessions) and high-quality tryptic peptides resulting from LC-MS/MS analysis. To this end, two primary buffers (Tris-HCl and urea) were employed along with either SDS/thiourea/HEPES or NaCl additives. One protocol also included a defatting step prior to extraction with Tris-HCl/urea buffer. Total protein estimation indicated that urea and thiourea buffers extracted the maximum amount of protein followed by urea/thiourea/HEPES buffer across the three cricket samples (Table S1; top panel). The urea and thiourea buffer (*P*3) extracted ~ 60% more protein from whole cricket (~16.1 mg/mL) than urea/thiourea and HEPES buffer (*P*4; ~9.7 mg/mL). For the roasted whole cricket sample, urea and thiourea (21.1 mg/mL) yielded ~ 56% more protein than urea/thiourea/HEPES (*P*4; ~9.1 mg/mL) buffer. Likewise, urea and thiourea (*P*3; ~17 mg/mL) buffer yielded ~ 38% more protein than urea/thiourea/HEPES (*P*4; ~10 mg/mL) buffer for the powdered roasted samples. Although urea and thiourea (*P*3) yielded the maximum total protein concentration for all cricket samples, the result was variable (based on the coefficient of variation, CV). *P5* yielded a low protein concentration and was excluded from LC-MS/MS analysis.

3.2. Protein and peptide yields from discovery proteomics

The total number of proteins and peptides identified at 1% global FDR were determined for the three cricket samples extracted using the eight extraction protocols (Table S1; bottom panel). Extraction protocols

and buffer compositions were selected based on their application in previously published insect proteome-based experiments. The urea/thiourea/HEPES (*P*4) buffer yielded the maximum number of proteins (423) from the whole cricket sample. The urea/thiourea (*P*3) buffer yielded the maximum number of protein identifications (498) from the whole roasted sample. While notably, the SDS/Tris-HCl (*P*1) buffer yielded the maximum number of proteins (364) from the roasted powder sample. The use of a defatting step with Tris/urea buffer-based extraction (*P*6) did not provide a superior yield of protein or peptide identifications in comparison to other protocols; however, protocol *P*6 was more consistent in terms of peptide numbers and total protein yield across the protocols. Similar results were obtained for the number of detected peptides from the extraction buffers (Table S1; bottom panel). For instance, the maximum number of peptides were detected using the protocol *P*3 extraction buffer: whole (1,543), roasted whole (1,332) and roasted powder (1,259). Overall, the protocols associated with urea-buffers were able to yield more protein identifications than other buffers; Tris-HCl yielded the lowest number of identifications unless combined with SDS.

3.3. Comparison of five extraction protocols and three types of cricket samples

The use of unannotated genome-translated databases can give rise to protein redundancy. This redundancy can impede comparative proteomic analyses. To overcome this challenge all data were combined into a single search to map each individual search result to a minimum set of protein accessions. The distinct peptides from each search result were mapped against a non-redundant set of total distinct peptide identifications.

Five protein extraction protocols (*P*1-*P*4 and *P*6) were selected for comparison based on their yield of proteins and peptides (Table S1; top panel). The resulting matrices of aligned Paragon unused protein scores and sequence coverage were captured to provide measures of protein identification quality. Apex peptide intensity information was used to provide a measure of peptide response.

3.3.1. Whole cricket

Hierarchical clustering of the unused protein scores (a measure of the total, unique peptide evidence related to a given protein) was performed to investigate the relationships between the extraction protocols. The unsupervised analysis revealed that extraction protocol *P*2 (200 mM Tris-HCl, 2 M urea), which yielded the lowest number of protein identifications, separates from the four remaining protocols (Fig. 1A).The hierarchical clustering of apex peptide intensity also shows that *P*4 and *P*6 produced similar peptide responses (Fig. S1). Although Tris-HCl buffer yielded a small number of proteins, the addition of SDS increased the number of protein identifications by ~ 90% (Table S1; bottom panel). Both urea-based buffers, *i.e.* *P*3 (Urea + Thiourea) and *P*4 (Urea + Thiourea + HEPES) showed relatively close clustering. Buffer *P*1 (SDS/Tris-HCl) and the pre-extraction defatting followed by Tris/urea buffer solubilisation *P*6 clustered together showing a similar complement of detected proteins for both protocols. GRAVY score was calculated for proteins detected from whole cricket to measure how the extraction buffers affect the representation of protein hydrophobicity (Fig. 1B). Proteins detected from *P*2 were significantly ($p < 0.0001$) different than *P*1 and *P*4; whilst *P*3 was significantly higher than *P*1 ($p < 0.01$) and *P*4 ($p < 0.01$). All extraction methods yielded significant

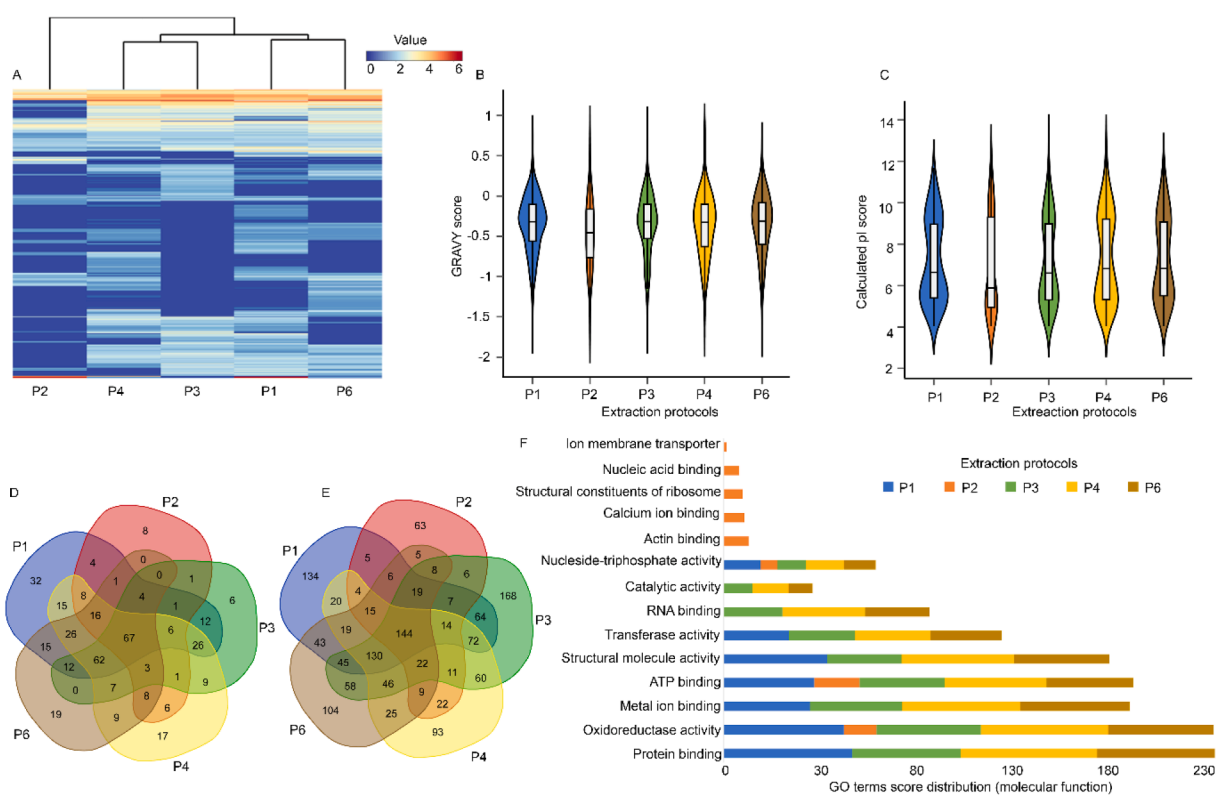

Fig. 1. Summary of protein and distinct peptide identifications (1% FDR) from whole crickets extracted with five extraction protocols. (A) Heatmap and HCA showing the unused protein scores of unified proteins detected following five extraction protocols: (*P*1) SDS/Tris-HCl; (*P*2) 200 mM Tris, 2 M Urea; (*P*3) Urea + Thiourea; (*P*4) Urea + Thiourea + HEPES; (*P*6) Defatted with *n*-pentane followed by Tris-HCl, 8 M urea solubilisation. Violin plots comparing calculated GRAVY score (B) and pI (C) for proteins detected from whole cricket samples. The y-axis indicates the values for GRAVY scores (positive GRAVY values indicate hydrophobic; negative values indicate hydrophilic) and pI. The median (black line) and boxes spanning the interquartile range (IQR) range are shown. Each whisker extends to the most extreme value no>1.5 × IQR from the box. Venn diagrams show the representation of proteins (D) and peptides (E) detected from whole cricket sample using the top five protocols. (F) Representative GO terms (molecular function) obtained from whole cricket proteins.

differences in terms of pI (Fig. 1C), except when comparing *P*1 to *P*4 or *P*4 to *P*6. Calculated aromaticity (Fig. S2A) and molecular weight (Fig. S2B) also show extraction protocol-dependent differences.

Using the SCIEX protein alignment template, 401 unique proteins were detected at 1% FDR from a combined database search of the five extraction methods (Fig. 1D). Overall, *P*1 yielded the maximum number of proteins (307; ~77%) followed by *P*4 (286; ~71%). In concordance with the Bradford protein concentration estimation, *P*2 yielded the lowest number of proteins (134; ~33%). Overall, 67 proteins (~17%) were commonly extracted by the five extraction methods. The maximum number of unique proteins was 32 (~8%) detected from *P*1; whilst *P*3 extracted the lowest number of unique proteins (6; ~2%).

Peptides detected at 1% FDR were aligned across the five extraction protocols to understand their representation within each extraction method. In total, 1,441 distinct peptides were detected from the whole cricket sample where 144 peptides (~10%) were commonly identified by all five protocols (Fig. 1D). As with the protein identification findings, peptides were also uniquely detected within each extraction method. For instance, extraction protocol *P*3 yielded the highest number of unique peptides (168; ~12%) followed by *P*1 (134; ~9%). As per the protein identification yield analysis, *P*2 also shows the lowest detection rate for unique peptides (63; 4.4%) (Fig. 1E).

Functional classification was performed for each of the unified protein identification sets (Fig. 1F). Protein binding, oxidoreductase activity, metal ion binding, ATP binding and structural molecule activities were the top five molecular functions resulting from all extraction methods, except *P*2. This method uniquely identified proteins associated with actin binding (14%), calcium ion binding (12%), structural constituents of the ribosome (11%), nucleic acid binding (9%) and ion transmembrane transporter activities (2%).

3.3.2. Roasted whole cricket

The top five extraction protocols were compared (*P*1-*P*4 and *P*6) in terms of their efficiency to extract proteins from roasted whole cricket samples. Hierarchical clustering of unused protein scores revealed that *P*4 and *P*6 were most closely related (Fig. 2A), while *P*4 separates from *P*1 and *P*2 and formed a separate cluster. This was replicated in the peptide intensity data (Fig. S3). Hydrophobicity was variable across the extraction methods (Fig. 2B), with differences observed between *P*1 and *P*3 ($p < 0.0002$), *P*3 and *P*2 ($p < 0.0001$) and *P*4 and *P*3 ($p < 0.008$). Analysis of the distribution of pI showed only two pairs of extraction protocols (*P*1 and *P*3; *P*4 and *P*6) were not different (Fig. 2C). Extraction protocol-dependent differences were also observed for aromaticity (Fig. S4A) and molecular weight (Fig. S4B).

In total, 401 unique proteins were detected following database search and protein accession unification across the five extraction protocols for roasted whole cricket, of which 91 (~23%) proteins were commonly extracted by the five extraction buffers (Fig. 2D). In contrast to the whole cricket sample, *P*3 extracted the highest number of proteins (333; ~83%) in comparison to proteins detected by *P*1 (190; ~47%) for the roasted whole cricket. However, *P*2 led to the lowest number of detected proteins (160; ~40%) for whole cricket and whole roasted cricket samples (Fig. 1D and Fig. 2D). All five extraction methods have shown their capacity to extract unique proteins. For instance, *P*3 extracted the maximum number of unique proteins (50; ~12%) for the roasted cricket samples. *P*1, *P*2 and *P*4 each extracted ~ 2% of the unique proteins from the total list of detected proteins. The comparison of peptide yield between extraction protocols showed that 1,155 distinct peptides were detected, where ~ 10% of these peptides were commonly identified with all five extraction protocols (Fig. 2E). The *P*3 extraction led to the highest frequency of unique peptides (~21%), whilst *P*1 led to the lowest number of unique peptides (~3%).

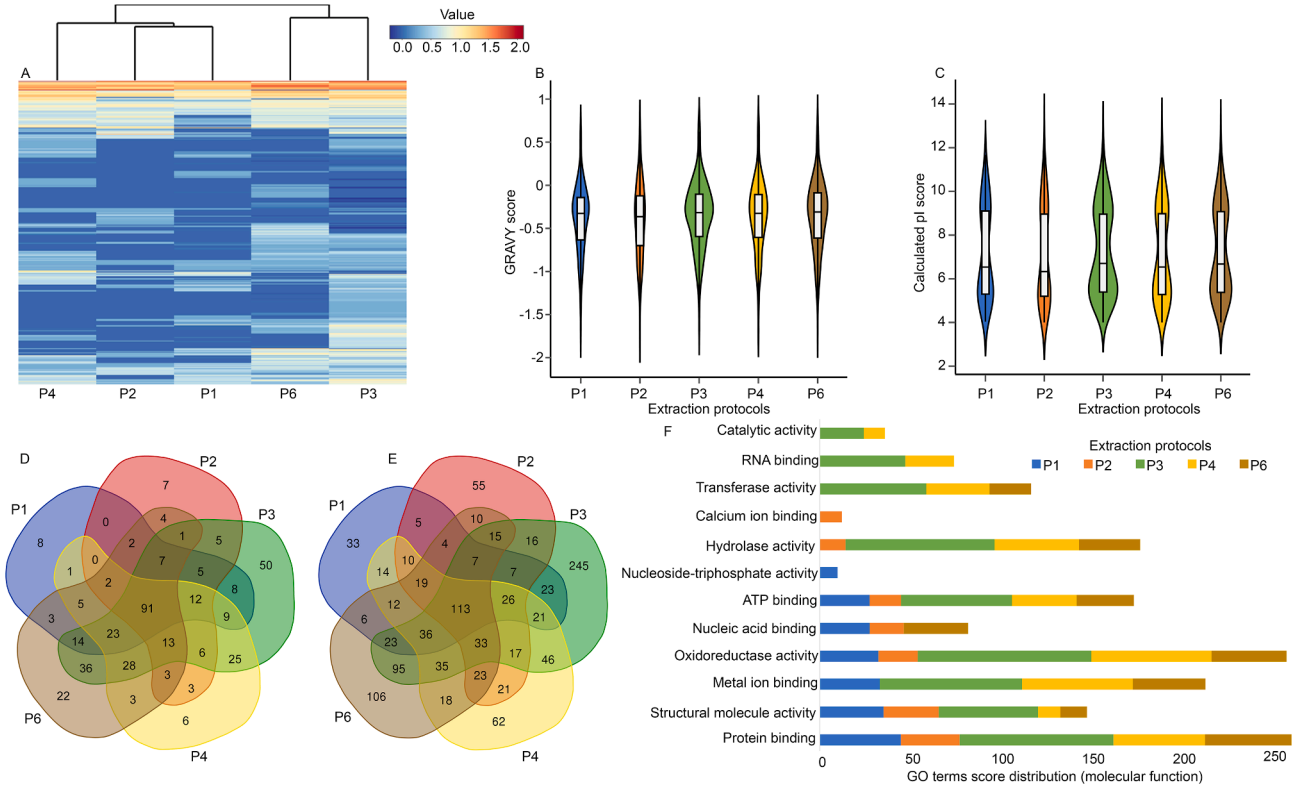

Fig. 2. Summary of protein and distinct peptide identifications (1% FDR) from roasted whole cricket. (A) Heatmap and hierarchical clustering showing the relationship of unused protein scores for aligned proteins detected following five extraction protocols: (*P*1) SDS/Tris-HCl; (*P*2) 200 mM Tris, 2 M Urea; (*P*3) Urea + Thiourea; (*P*4) Urea + Thiourea + HEPES; (*P*6) Defatted with *n*-pentane followed by Tris-HCl, 8 M urea solubilisation. Violin plots comparing GRAVY score (B) and pI (C) for proteins detected from whole cricket samples. Venn diagrams show the representation of proteins (D) and peptides (E) detected from roasted whole cricket samples. (F) Representative GO terms (molecular function) obtained for roasted whole cricket proteins.

GO analysis was performed on individual sets of proteins across the protocols (Fig. 2F). The top-three molecular functions associated with the extracted proteins across all protocols were: protein binding, structural molecule binding and metal ion binding. Notably, >40% of proteins resulting from *P*3 were associated with protein binding, structural molecule binding, metal ion binding and oxidoreductase activities in comparison to the four remaining protocols. Furthermore, calcium ion binding activity was only found with *P*2 while nucleoside-triphosphate activity was found with *P*1. The proteins with nucleoside-triphosphate activities are a family of membrane proteins. Anionic denaturing detergents such as SDS can be efficient in disrupting membranes and denaturing higher order structures by breaking protein–protein and protein-lipid interactions and hence can solubilize membrane proteins allowing access to proteolytic enzymes.

3.3.3. Roasted cricket powder

The impact of extraction protocols on protein detection was investigated for roasted cricket powder. Hierarchical clustering of the aligned unused protein scores for the protein extraction methods indicate that *P*1 and *P*3 can extract a similar suite of proteins, which differ from those extracted by *P*2 (Fig. 3A). Peptide intensities also show distinct separation based on extraction protocol for the roasted cricket powder samples (Fig. S5). GRAVY scores for the extraction methods also revealed that the proteins extracted by *P*2 were different from those extracted by *P*1 ($p < 0.0007$), *P*3 ($p < 0.006$) and *P*4 ($p < 0.04$) (Fig. 3B). Likewise, the pI values for the extraction protocols show that comparison of the protocols *P*1 to *P*3 ($p < 0.36$) and *P*4 to *P*6 ($p < 0.99$) were not different whereas all other protocols showed significantly different distributions to one another (Fig. 3C). Proteins extracted with the protocols also show differences in terms of their aromaticity (Fig. S6A) and molecular weight (Fig. S6B).

In combination, the processing protocols extracted 361 unique proteins from roasted cricket powder samples where 59 (~16%) proteins were commonly extracted by the five protocols (Fig. 3D). In comparison to the whole cricket and roasted whole cricket, the roasted cricket powder yielded ~ 11% less proteins. *P*1 yielded the highest number of protein identifications 250 (62%) followed by *P*3 (59%). All five extraction protocols extracted a unique set of proteins and peptides from roasted cricket powder, as evidenced previously for the two other product types (Fig. 3D-E). Four protocols (*P*1, *P*3, *P*4 and *P*6) co-detected 74 (~7%) peptides whilst the two protocols that yielded the highest protein identifications (*P*1 and *P*3) commonly identified 82 unique peptides (7.5%), while *P*1 was shown to uniquely detected 162 (~15%) peptides and *P*3 yielded an additional 110 (~10%) peptides.

All protocols yielded proteins that classified under oxidoreductase activity, ATP binding and transferase activity molecular functions (Fig. 3F). Although protocol *P*2 yielded the lowest number of proteins, it shows the unique ability to extract proteins with specific molecular functions, including calcium ion binding, cycloskeletal protein binding, protein dimerization activity and ion transmembrane transporter activity. The molecular function structural molecule activity was only found in extracts following defatting (*P*6).

3.4. Identification of putative allergens from cricket samples

A database search was performed using the combined data acquired for all eight extraction methods to identify putative allergens from the cricket samples. Additionally, Pfam domain profiles of the proteins detected at 1% global FDR from the combined database search for all cricket samples were compared to the WHO/IUIS insect and crustaceans reference allergen protein sequences collected from the AllFam database. In total, 52 potential cricket allergens representing 15 allergen

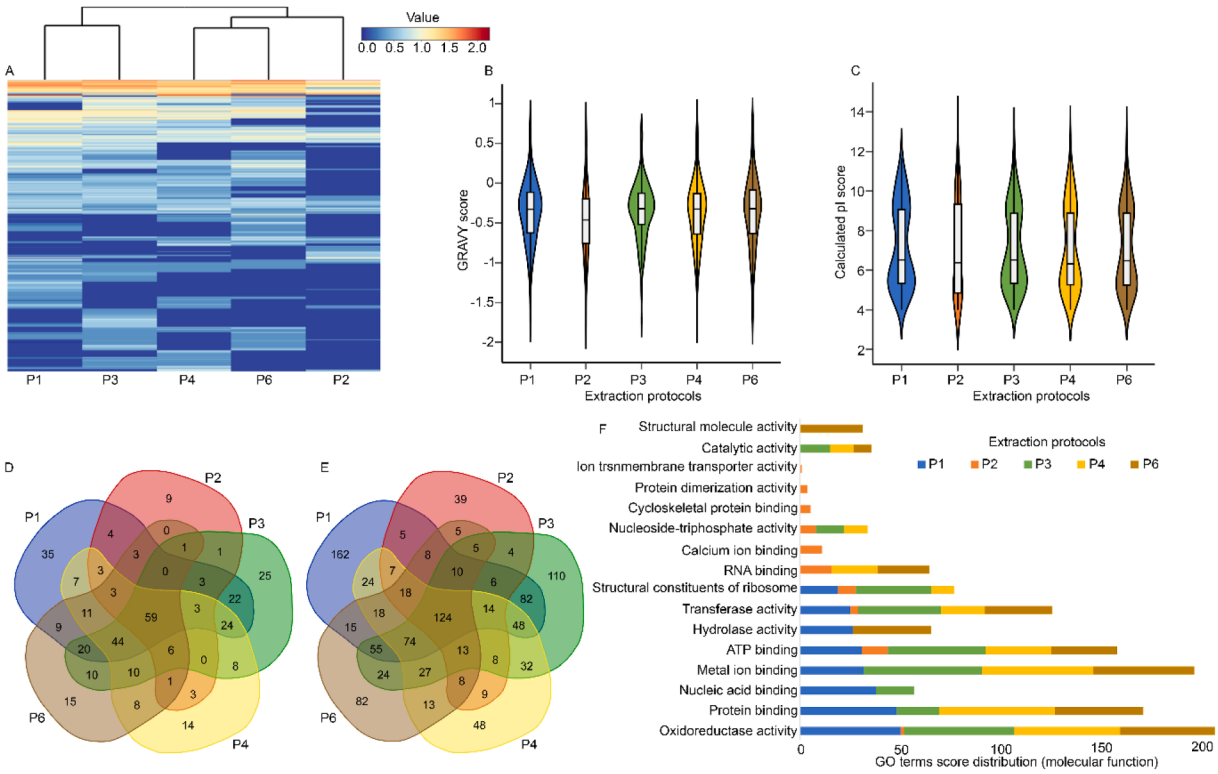

Fig. 3. Summary of protein and distinct peptide identifications (1% FDR) from roasted whole cricket powder. (A) Heatmap and hierarchical clustering based on unused protein scores of aligned proteins detected following five extraction protocols: (*P*1) SDS/Tris-HCl; (*P*2) 200 mM Tris, 2 M Urea; (*P*3) Urea + Thiourea; (*P*4) Urea + Thiourea + HEPES; (*P*6) Defatted with *n*-pentane followed by Tris-HCl, 8 M urea solubilisation. Violin plots comparing GRAVY score (B) and pI (C) for proteins detected from whole roast cricket powder samples. Venn diagrams show the representation of proteins (D) and peptides (E) identified from roasted whole cricket powder sample. (F) Representative GO terms (molecular function) obtained from roasted whole cricket powder proteins.

families were detected of which AF007: EF hand family proteins (15 sequences), AF054: tropomyosins (7 sequences), and AF002: heat shock protein Hsp70 (7 sequences) were the most abundant (Fig. 4, Table S2). Both tropomyosins and EF hand proteins are known allergens in crustaceans, insects and mites, while AF002 allergens have only been detected in house mites. Additionally, cricket proteins showing close homology to crustacean allergens were also detected in the AF049: ATP: guanido phosphotranferases (arginine kinases), AF146: troponin I and T and AF032: triosphophate isomerase allergen families. Sequence homology and conserved Pfam domains between cricket proteins and reference allergens indicate that the putative allergens from cricket samples are closely related to the known allergens retrieved from the AllFam database.

To further refine the lists of putative allergens from cricket, *in-silico* prediction was performed using AllerCatPro software. To this end, 20 putative allergens were identified from the comprehensive *in-silico* prediction and combined database search. Notably, the percent identity of linear 80 amino acid (aa) stretches, percentage identity of 3D epitope and allergen evidence columns show the allergen family and Pfam domain similarities between AllFam registered allergens and putative allergens detected in the present study (Fig. 4). Overall, these analyses provide evidence that 80% of the putative cricket allergens detected by LC-MS have high allergenic potential; however, future IgE binding assays would be required in order to validate their immunological reactivity.

3.5. Sample-processing influence on proteome content

To assess the influence of product preparation (whole; roasted; roasted-powdered) on proteome composition, a comparative analysis was performed on the protein and peptide identification resulting from *P*3 extraction. In total, 388 unique proteins were detected, where the whole, roasted whole and roasted powder cricket samples yielded 217, 333 and 236 proteins, respectively (Fig. S7A). The protein level comparisons revealed that the three food processing methods commonly shared 145 (37.4%) proteins (Fig. S7). Notably, the roasted cricket

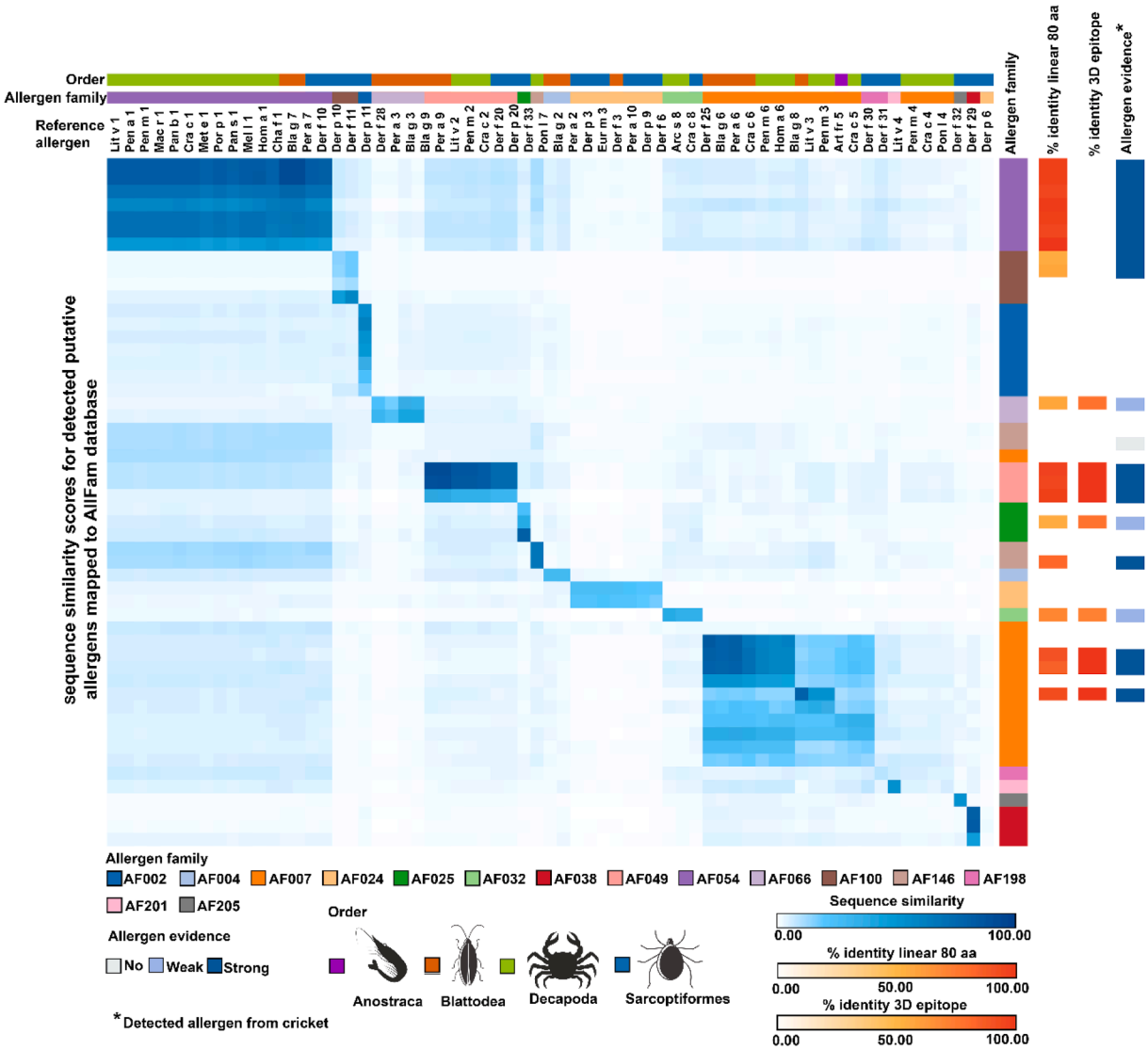

Fig. 4. The diversity of allergen proteins from cricket and crustaceans. The heatmap shows sequence similarity scores for the 73 reference allergens (columns) versus identified cricket proteins (rows). The reference allergens comprise both ingestion and inhalation-related insect and crustacean allergen families from the AllFam database. Above the heatmap, the reference allergens are classified according to insect order (taxonomy) and AF allergen families (AllFam database) using the IUIS allergen nomenclature system. To the right of the heatmap, the putative allergens detected by LC-MS are classified according to allergen family. The putative allergens from this analysis were then searched using AllerCatPro software. The resulting percent identity for both linear 80 aa sequences and percent identity of 3D epitopes are shown. The last column indicates the allergen evidence defined as no, weak or strong potential.

sample yielded the maximum 87 (22.4%) unique proteins followed by 31 (8%) proteins from whole cricket samples. As expected, the two roasting processes shared a high proportion (80.4%) of proteins. Unlike the protein level comparisons, the peptide level analysis revealed that the whole cricket sample had the maximum unique peptides 329 (25.1%) followed by roasted cricket samples 217 (16.5%) (Fig. S7B). The roasted powder samples yielded the lowest frequency of protein and peptide identifications.

The results yielded from the discovery proteomics analysis were searched manually to identify the influence of extraction buffer on the enrichment of allergen proteins. Table S3 summarises the putative allergens detected across the seven extraction buffers and product types. As an example, the cross-reactive allergen arginine kinase was detected as 3 proteoforms with buffer *P*4 from whole cricket (Table S3). While the extraction buffer *P*3 and *P*6 extracted the highest number of arginine kinases (3 proteoforms) from roasted cricket sample. Altogether, the results from this comparative study reveals variation of proteome and allergen content that are not only dependent on the extraction buffer but also the product processing techniques (Table S3).

3.6. Extraction protocol-dependent arginine kinase extraction and quantitation from cricket samples by LC-MRM-MS

Arginine kinase has shown cross-allergenicity between

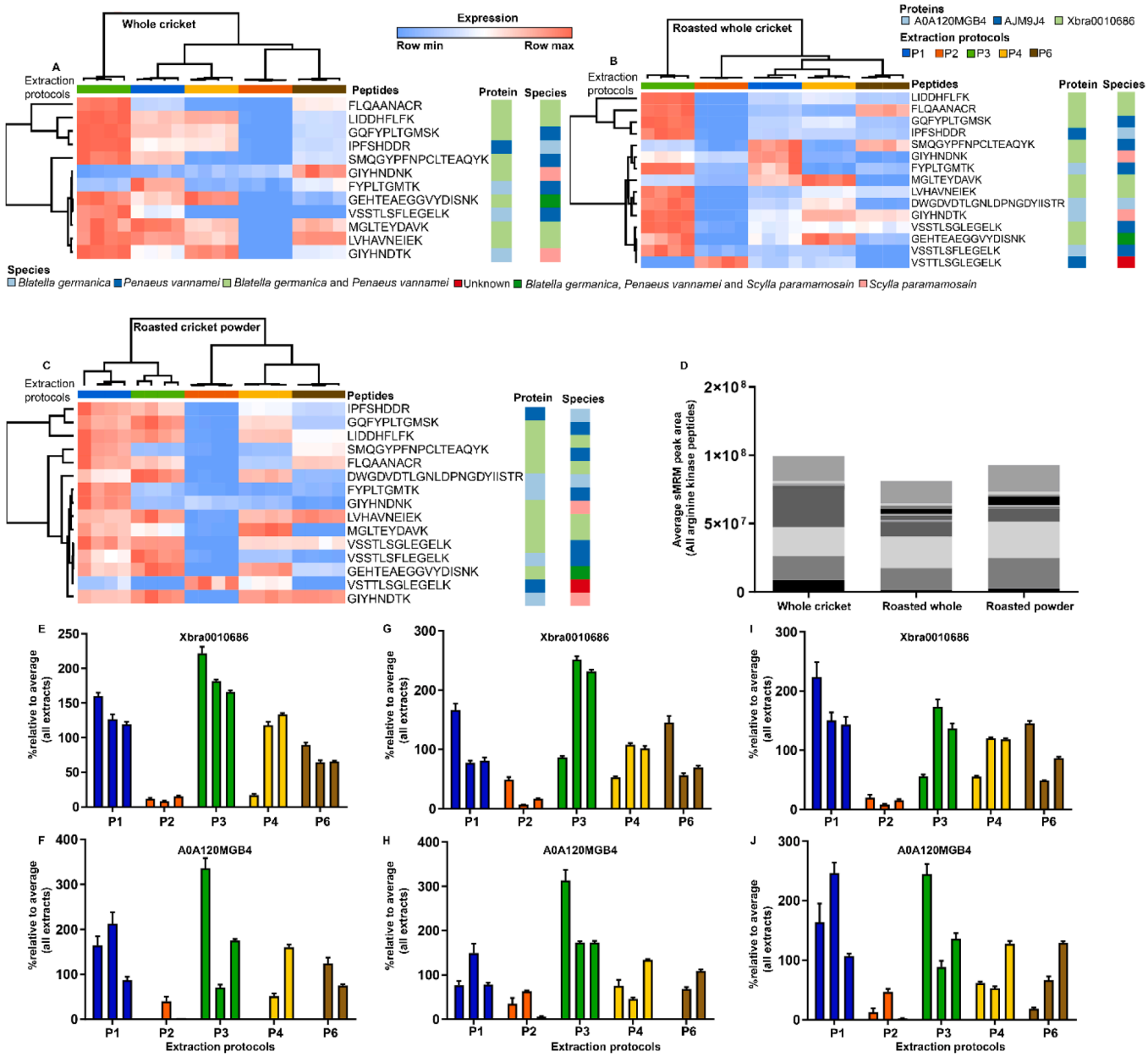

Fig. 5. Quantitation of arginine kinases by LC-MRM-MS across three cricket products. Heatmap and HCA showing the relative abundances of arginine kinase peptides and the relationships of the extraction protocols for whole (A); whole roasted (B); and whole roasted powdered (C) products. The colour in each cell of the heatmap depicts the relative peak area of each peptide from row min (low) to row max (high). The protein column on the heatmap indicates the protein origin of each peptide. The species column shows matches with > 80% sequence identity to the cricket AK peptides for allergens from three cross-reactive species. Three peptides were excluded from the heatmap (A) as they were not detectable by MRM. (D) Summed MRM peak area for monitored AK peptides from *P*3. Relative quantitation of three exemplar peptides from two cricket AKs (Xbra0010686 and A0A120MGB4). Graphs show the relative percentage of MRM peak area (compared to the average of all extraction protocols; n = 4 replicates); whole cricket (E-F); whole roasted cricket (G-H); roasted cricket powder (I-J).

Macrobrachium spp. (prawns or shrimp) and field cricket *Gryllus bimaculatus* (Srinroch et al., 2015). An LC-MRM-MS method to detect arginine kinase and explore relative quantitation was developed. In this case arginine kinase (AK) was used as an exemplar of a cross-allergenic protein to assess the influence of extraction protocols on allergen yield/detectability. In total, three AK proteoforms were detected from the master database search and used to build the LC-MRM-MS method. To confirm the potential for cross-allergenicity for the detected AKs in the present study, all sequences were examined using the AllerCatPro tool (Maurer-Stroh et al., 2019). Notably, all three AKs have shown strong evidence as a cross-allergen based on > 95% identity (linear 80 amino acid window) and 100% identity (3D epitope structure). LC-MRM-MS data was acquired for 15 AK-derived peptides from across these three proteoforms (Table S2; Fig. S8A); their relative abundances were determined and compared across the five extraction protocols and product types. The whole and whole roasted cricket samples showed a similar pattern of MRM response, while the whole roasted powdered sample showed a different response to the buffers as seen in the structure of the sample hierarchical clustering (Fig. 5A-C). The species column shows organisms with > 80% BLAST matches for cricket AK peptides compared to allergen peptides from three allergenic species (Fig. 5A-C). Of interest, searching the cricket AK peptides against the Immune Epitope Database (IEDB) revealed sequence homology to allergen proteins from shrimp, cockroaches and mud crab.

The AK peptides were noted to have the highest yield from *P*3 for both whole cricket and roasted whole cricket (Fig. 5A-B). For roasted cricket powder, *P*1 extracted the highest abundance of peptides followed by *P*3. To assess the proportion of AKs that were extracted using each protocol, the average abundance for all peptides was calculated (considering the five methods and four replicate extracts) from across the three cricket samples. Each individual peptide measurement was compared to the average and converted to a percentage, with all data presented as Violin plot (wherein the centre line equates to the median AK content). As an example, examining whole cricket extracted with five protocols shows a median value of *P*1 (119%), *P*3 (166%), *P*4 (76%) and *P*6 (71%) (Fig. S8B). Notably, all peptides were not detectable in all samples and variation in the abundance for each of the peptides was observed across samples. Overall, *P*3 resulted in the highest response for whole and roasted whole cricket while *P*1 led to the highest response for roasted powder samples, followed by *P*3.

The peak areas were summed for the *P*3 samples to compare how sample processing affects AK detection across the three products (Fig. 5D). In this regard, whole cricket was shown to yield the highest amount of detectable AK per μg of total protein (Fig. 5D).

The technical variation for sample preparation and analysis was typically < 15%, while the variation in the peptide response ranged from 0 to 90% across the three product types. Three selected peptides, from AK proteins Xbra0010686 and A0A120MGB4, were detected with 1.4- and 1.25-fold higher response by *P*3 than *P*1, respectively, from whole cricket (Fig. 5E-F). Likewise, the same panel of peptides were extracted 1.75- and 2.17-fold higher with *P*3 in comparison to *P*1 from the roasted whole cricket sample, respectively (Fig. 5G-H). In contrast, *P*1 yielded 1.41- and 1.1-fold higher AK response than *P*3 from roasted cricket powder for the same peptides, respectively (Fig. 5I-J). Taken together, the results highlight the importance of extraction protocol optimization for specific sample matrices.

4. Discussion

Protein extraction from insect tissues can be challenging due to the presence of significant amount of non-protein components. To overcome these challenges, a number of extraction protocols have been used in order to maximise the protein yield from processed or unprocessed insect samples and correlate protein concentration, identifications and/or their quantitative measurements with the extraction protocol (Choi et al., 2017; Hall et al., 2018; Srinroch et al., 2015; Yi et al., 2013). In this study, eight protein extraction buffer compositions were examined in terms of their capacity to yield protein and distinct peptide identifications, as well as broadly representing physicochemical and functional properties. The capacity of these buffers to yield specific signal for putative allergens was also investigated using a novel LC-MRM method. The initial proteome-scale study revealed that the urea-based buffers (*P*3) yielded the maximum number of proteins for whole cricket and roasted whole cricket (Figs. 1-2). The number of detected proteins were higher for roasted cricket powder when extracted with SDS buffer (*P*1, Fig. 3). To extract the proteins from insects, in-gel and in-solution-based sample preparation methods have been used with various solvents. The study presented herein provides evidence that the inclusion of a defatting step (*P*6) led to a consistent yield of peptide identifications and total protein concentration (Table S1; top panel).

The most notable differences between the protocols used in this study resulted from the strengths of the extraction buffers. Urea and thiourea are neutral chaotropic agents commonly used to solubilise and denature proteins by disrupting hydrogen bonds and hydrophobic interactions between and within proteins. The results herein indicate that these chaotropic agents help to denature proteins for a more efficient solubilisation in comparison to Tris-HCl-based buffer; as observed previously for *Hermetia illucens*, *Tenebrio molitor*, *Alphitobius diaperinus* and *Acheta domesticus* (Belghit et al., 2019). The exclusion of DTT from Tris-HCl buffer (*P7*) may also explain the detection of a comparatively small number of proteins as the addition of DTT will increase the reducing capacity of the buffer and allow for enhanced protein unfolding. In protocol 1 (*P*1), the inclusion of SDS (anionic detergent) led to a higher protein identification yield than other protocols tested (Table S1; bottom panel), likely due to the superior solubilisation power of the detergent. Although SDS is often used with urea to aid in dissolution of proteins, subsequent clean-up steps are required to separate the protein(s) from these compounds (Rodríguez-Rojas & Rolff, 2020). However, studies have reported that removing SDS can be difficult, and they can negatively impact the proteolytic enzymes and results in ion suppression (Bose et al., 2019; Fic et al., 2010). Thus, urea can be used as an alternative to SDS for protein extraction. Taken together, the use of urea-based extraction processes was found to be rapid, simple and readily implemented without the need for additional steps such as defatting.

Food processing is a critical determinant in terms of identification of proteins and peptides from various insect species (Belghit et al., 2019). The results from this study point to the existence of product processing-dependent variation within the proteome that could be exploited during food processing to prepare potentially less-allergenic products for consumers (Fig. S7; Table S3). The physicochemical profiling revealed that each buffer extracts unique sets of proteins which have different hydrophobicity, pI, aromaticity and molecular weight (Figs. 1-3). This information can be useful when designing immunological assays or to develop ELISA assays for epitope quantitation.

One of the principal goals of analytical chemistry is to deliver specific quantitative assays that can be easily applied to industrial or clinical needs, such as allergen quantitation in traditional and novel food matrices. In this regard, previous studies have reported the cross-reactivity of shrimp-allergic patients (Kamemura et al., 2019) to cricket proteins (*Gryllus bimaculatus*), necessitating the need to specifically measure these new allergens in any relevant consumer product. As a result, AK was selected as an exemplar target protein in the current study due to its putative status as a novel allergen. Relative quantitation of AKs can assist in the development of product processing methods that reduce the allergenic content for consumers (Figs. 4 and 5). While the method developed herein demonstrates a requirement for optimization for the detection and quantitation of AK in cricket samples and potentially in food products (Fig. 5), it should be noted that the use of denaturants during extraction may preclude such extracts being useful for functional or activity testing. Insect-derived food products offer promise as an alternative high-protein source to meet global food demand for the growing population. According to EU guidelines, insect-based food

products are regarded as novel foods (Belluco et al., 2013). Therefore, allergenic risk assessment should be performed through comparing any newly identified proteins with known allergens and immune cross-reactivity analysis prior to accepting insect products as a food ingredient for commercial use (de Gier & Verhoeckx, 2018; Mazzucchelli et al., 2018). This current study provides a framework for bioinformatics analysis, in-silico mapping and quantitation strategies for other cross-reactive immunogenic AK peptides from various insect species (Fig. 5). Moreover, this study provides future research opportunities to develop LC-MRM-MS methods for further cross-reactive proteins (Fig. 4) such as tropomyosin (Kamemura et al., 2019) for industrial and food safety applications.

5. Conclusion

In conclusion, the results highlight that urea-based extraction buffer systems coupled with FASP can extract complex proteomes with diverse physicochemical properties. Furthermore, the inclusion of a defatting step prior to protein extraction did not provide superior protein and peptide identification yields, although it was shown to provide consistency in protein yield. No single method of protein extraction can be universally applied to all types of samples; however, the selection of extraction method will impact the overall representation of the cricket proteome as well as the detectability of specific cricket allergen proteins. This study has shown that extraction buffer and sample processing affects the detectability of an exemplar cricket allergen arginine kinase; future study will focus on developing quantitative methods for the remaining allergens detected herein. Taken together, specific or complementary buffer systems should be used to achieve the greatest proteome coverage or lowest limit of quantification, depending on the application.

6. Credit author statement

This work was carried out in collaboration between all authors from four research institutions: Commonwealth Scientific and Industrial Research Organisation (CSIRO; Australia), Edith Cowan University (ECU; Australia), James Cook University (JCU; Australia) and Agency for Science, Technology and Research (A*STAR; Singapore). M.C., A.L.L. and S.M.S. acquired research funding and conceived the project idea. U. B., S.K., E.B.J. and S.S. prepared the samples. U.B. J.B., K.B. and M.C. designed the LC-MS data acquisition pipeline; developed the acquisition methods and acquired the discovery and targeted proteomics data. A.J. prepared the database and performed the *in silico* allergen mapping and subsequent bioinformatics analysis; V.L. performed the allergen mapping for the putatively detected allergens using AllerCatPro. All authors reviewed and approved the manuscript.

Declaration of Competing Interest

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

Acknowledgements

The project was funded by the 1st Singapore-Australia Bilateral Program on Innovations in Food for Precision Health, 2019 (Grant number: SG-AUS2019_191D4) encompassing a partnership between The Commonwealth Scientific and Industrial Research Organisation (Australia), James Cook University (Australia) and Agency for Science, Technology and Research (A*STAR, Singapore). We also would like to thank Paula Pownall (Grubs Up, Australia) for supplying the whole cricket sample and the company Grilo for supplying roasted whole cricket and roasted cricket powder samples.

Data Availability

Raw and processed LC-MS data files have been uploaded to https://doi.org/10.25919/rb15-em18.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2021.129110.

References

Amarender, R. V., Bhargava, K., Dossey, A. T., & Gamagedara, S. (2020). Lipid and protein extraction from edible insects–Crickets (Gryllidae). *LWT, 125*, 109222. https://doi.org/10.1016/j.lwt.2020.109222.

Belghit, I., Lock, E.-J., Fumière, O., Lecrenier, M.-C., Renard, P., Dieu, M., … Rasinger, J. D. (2019). Species-specific discrimination of insect meals for aquafeeds by direct comparison of tandem mass spectra. *Animals, 9*(5), 222. https://doi.org/10.3390/ani9050222.

Belluco, S., Losasso, C., Maggioletti, M., Alonzi, C. C., Paoletti, M. G., & Ricci, A. (2013). Edible insects in a food safety and nutritional perspective: A critical review. *Comprehensive Reviews in Food Science and Food Safety, 12*(3), 296–313. https://doi.org/10.1111/crf3.2013.12.issue-310.1111/1541-4337.12014.

Bose, U., Broadbent, J., Byrne, K., Hasan, S., Howit, C. A., & Colgrave, M. L. (2019). Optimisation of protein extraction for in-depth profiling of the cereal grain proteome. *Journal of Proteomics, 197*, 23–33. https://doi.org/10.1016/j.jprot.2019.02.009.

Bose, U., Byrne, K., Howit, C., & Colgrave, M. L. (2019). Targeted proteomics to monitor the extraction efficiency and levels of barley α-amylase trypsin inhibitors that are implicated in non-coeliac gluten sensitivity. *Journal of Chromatography A, 1600*, 55–64. https://doi.org/10.1016/j.chroma.2019.04.043.

Choi, B. D., Wong, N. A., & Auh, J.-H. (2017). Defatting and sonication enhances protein extraction from edible insects. *Food Science of Animal Resources, 37*, 955. https://doi.org/10.5851/kosfa.2017.37.6.955.

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., & Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics, 21*(18), 3674–3676. https://doi.org/10.1093/bioinformatics/bti610.

Jantzen da Silva Lucas, A., Menegon de Oliveira, L., da Rocha, M., & Prentice, C. (2020). Edible insects: An alternative of nutritional, functional and bioactive Compounds. *Food Chemistry, 311*, 126022. https://doi.org/10.1016/j.foodchem.2019.126022.

de Castro, R. J. S., Ohara, A., dos Santos Aguilar, J. G., & Domingues, M. A. F. (2018). Nutritional, functional and biological properties of insect proteins: Processes for obtaining, consumption and future challenges. *Trends in Food Science and Technology, 76*, 82–89. https://doi.org/10.1016/j.tifs.2018.04.006.

de Gier, S., & Verhoeckx, K. (2018). Insect (food) allergy and allergens. *Molecular Immunology, 100*, 82–106. https://doi.org/10.1016/j.molimm.2018.03.015.

Fallahbaghery, A., Zou, W., Byrne, K., Howitt, C. A., & Colgrave, M. L. (2017). Comparison of gluten extraction protocols assessed by LC-MS/MS analysis. *Journal of Agricultural and Food Chemistry, 65*(13), 2857–2866. https://doi.org/10.1021/acs.jafc.7b00063.

Fic, E., Kedracka-Krok, S., Jankowska, U., Pirog, A., & Dziedzicka-Wasylewska, M. (2010). Comparison of protein precipitation methods for various rat brain structures prior to proteomic analysis. *Electrophoresis, 31*(21), 3573–3579. https://doi.org/10.1002/elps.v31:2110.1002/elps.201000197.

Hall, F., Johnson, P. E., & Liceaga, A. (2018). Effect of enzymatic hydrolysis on bioactive properties and allergenicity of cricket (*Gryllodes sigillatus*) protein. *Food Chemistry, 262*, 39–47. https://doi.org/10.1016/j.foodchem.2018.04.058.

Kamemura, N., Sugimoto, M., Tamehiro, N., Adachi, R., Tomonari, S., Watanabe, T., & Mito, T. (2019). Cross-allergenicity of crustacean and the edible insect *Gryllus bimaculatus* in patients with shrimp allergy. *Molecular Immunology, 106*, 127–134. https://doi.org/10.1016/j.molimm.2018.12.015.

Kouřimská, L., & Adámková, A. (2016). Nutritional and sensory quality of edible insects. *NFS Journal, 4*, 22–26. https://doi.org/10.1016/j.nfs.2016.07.001.

B. MacLean D.M. Tomazela N. Shulman M. Chambers G.L. Finney B. Frewen … M.J. MacCoss Skyline: An open source document editor for creating and analyzing targeted proteomics experiments 26 7 2010 2010 966 968 10.1093/bioinformatics/btq054.

S. Maurer-Stroh N.L. Krutz P.S. Kern V. Gunalan M.N. Nguyen V. Limviphuvadh … A. Valencia AllerCatPro—prediction of protein allergenicity potential from the protein sequence 35 17 2019 2019 3020 3027 10.1093/bioinformatics/btz029.

Mazzucchelli, G., Holzhauser, T., Cirkovic Velickovic, T., Diaz-Perales, A., Molina, E., Roncada, P., … Hoffmann-Sommergruber, K. (2018). Current (food) allergenic risk assessment: Is it fit for novel foods? Status quo and identification of gaps. *Molecular Nutrition & Food Research, 62*(1), 1700278. https://doi.org/10.1002/mnfr.201700278.

Mishyna, M., Martinez, J.-J.-I., Chen, J., & Benjamin, O. (2019). Extraction, characterization and functional properties of soluble proteins from edible grasshopper (*Schistocerca gregaria*) and honey bee (*Apis mellifera*). *Food Research International, 116*, 697–706. https://doi.org/10.1016/j.foodres.2018.08.098.

Montowska, M., Kowalczewski, P.Ł., Rybicka, I., & Fornal, E. (2019). Nutritional value, protein and peptide composition of edible cricket powders. *Food Chemistry, 289*, 130–138. https://doi.org/10.1016/j.foodchem.2019.03.062.

Ndiritu, A. K., Kinyuru, J. N., Kenji, G. M., & Gichuhi, P. N. (2017). Extraction technique influences the physico-chemical characteristics and functional properties of edible crickets (*Acheta domesticus*) protein concentrate. *Journal of Food Measurement and Characterization, 11*(4), 2013–2021. https://doi.org/10.1007/s11694-017-9584-4.

Rodríguez-Rojas, A., & Rolff, J. (2020). Preparation of insect protein samples for label-free proteomic quantification by LC-Mass spectrometry. In T., G., Sandrelli, & F. (Eds.), *Immunity in Insects. Springer Protocols Handbooks* (pp. 53–61). New York, NY: Humana.

Ruethers, T., Taki, A. C., Johnston, E. B., Nugraha, R., Le, T. T. K., Kalic, T., … Lopata, A. L. (2018). Seafood allergy: A comprehensive review of fish and shellfish allergens. *Molecular Immunology, 100*, 28–57. https://doi.org/10.1016/j.molimm.2018.04.008.

Shilov, I. V., Seymour, S. L., Patel, A. A., Loboda, A., Tang, W. H., Keating, S. P., … Schaeffer, D. A. (2007). The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Molecular and Cellular Proteomics, 6*(9), 1638–1655. https://doi.org/10.1074/mcp.T600050-MCP200.

Srinroch, C., Srisomsap, C., Chokchaichamnankit, D., Punyarit, P., & Phiriyangkul, P. (2015). Identification of novel allergen in edible insect, *Gryllus bimaculatus* and its cross-reactivity with *Macrobrachium* spp. allergens. *Food Chemistry, 184*, 160–166. https://doi.org/10.1016/j.foodchem.2015.03.094.

Stull, V. J., Finer, E., Bergmans, R. S., Febvre, H. P., Longhurst, C., Manter, D. K., … Weir, T. L. (2018). Impact of edible cricket consumption on gut microbiota in healthy adults, a double-blind, randomized crossover trial. *Scientific Reports, 8*(1). https://doi.org/10.1038/s41598-018-29032-2.

Tang, W. H., Shilov, I. V., & Seymour, S. L. (2008). Nonlinear fitting method for determining local false discovery rates from decoy database searches. *Journal of Proteome Research, 7*(9), 3661–3667. https://doi.org/10.1021/pr070492f.

Yi, L., Lakemond, C. M. M., Sagis, L. M. C., Eisner-Schadler, V., van Huis, A., & van Boekel, M. A. J. S. (2013). Extraction and characterisation of protein fractions from five insect species. *Food Chemistry, 141*(4), 3341–3348. https://doi.org/10.1016/j.foodchem.2013.05.115.

Protein extraction protocols for optimal proteome measurement and arginine kinase quantitation from cricket *Acheta domesticus* for food safety assessment

Utpal Bose[1‡], James A. Broadbent[1‡], Angéla Juhász[2], Shaymaviswanathan Karnaneedi[3], Elecia B. Johnston[3], Sally Stockwell[1], Keren Byrne[1], Vachiranee Limviphuvadh[4,5], Sebastian Maurer-Stroh[4,5,6], Andreas L. Lopata[3], Michelle L. Colgrave[1,2*]

[1]CSIRO Agriculture and Food, 306 Carmody Rd, St Lucia QLD 4067, Australia; [2]Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science, School of Science, Edith Cowan University, Joondalup WA 6027, Australia; [3]Molecular Allergy Research Laboratory, Australian Institute of Tropical Health and Medicine, James Cook University, Townsville, QLD, Australia; [4]Biomolecular Function Discovery Division, Bioinformatics Institute, Agency for Science, Technology and Research, Singapore. [5]IFCS Programme, Singapore Institute for Food and Biotechnology Innovation, Agency for Science, Technology and Research, Singapore. [6]Department of Biological Sciences, National University of Singapore, Singapore.

[*]Corresponding author: Michelle L. Colgrave, CSIRO Agriculture and Food, 306 Carmody Rd, St Lucia QLD 4067, Australia; Phone: +61 7 3214 2697; Fax: +61 7 3214 2900; Email: michelle.colgrave@csiro.au

[‡]Authors contributed equally

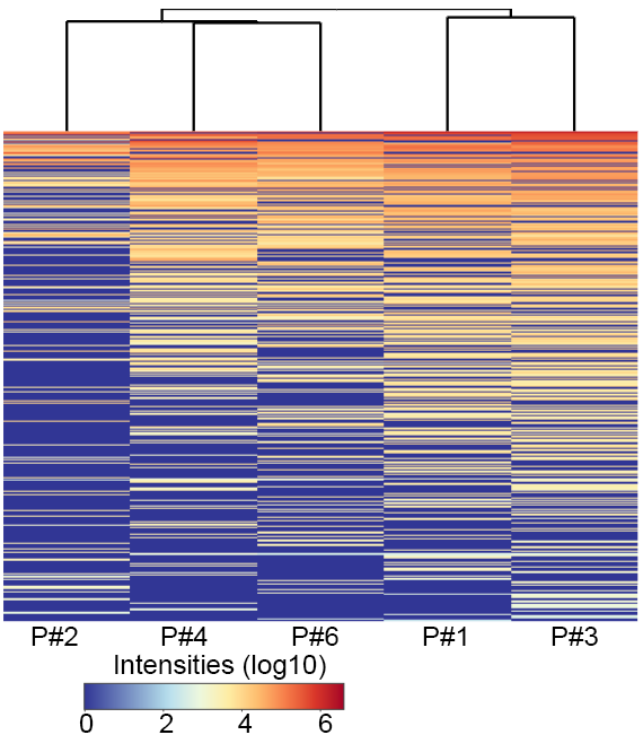

Figure S1: Heatmap plot and hierarchical clustering showing peptide peak apex intensities across the five extraction protocols applied to whole cricket samples.

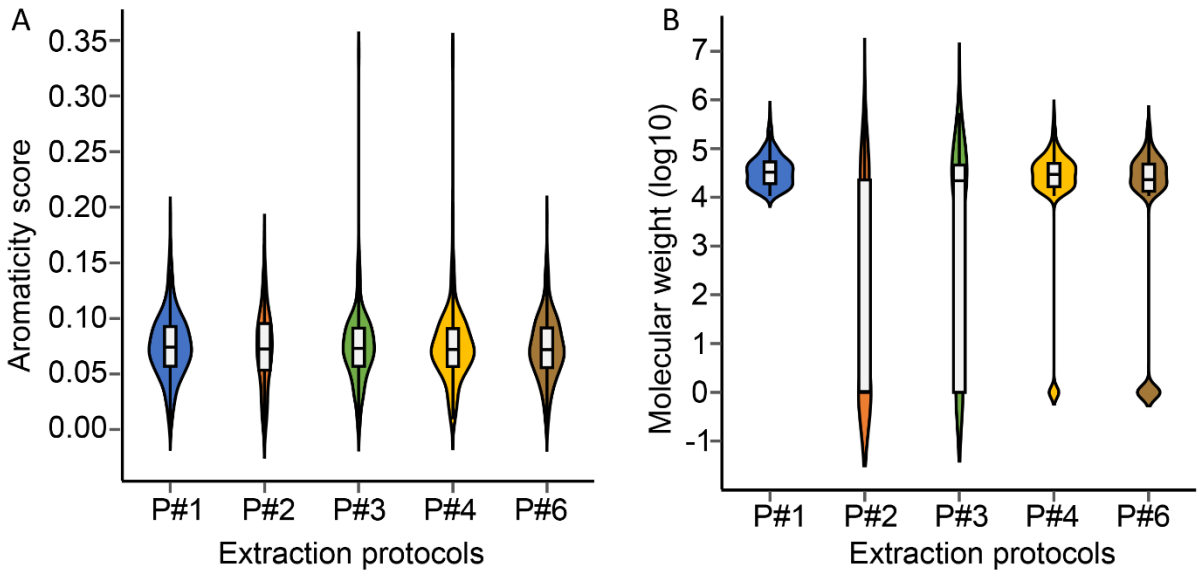

Figure S2: Five extraction protocols yield different physicochemical properties for whole cricket samples: (**A**) aromaticity score; and (**B**) molecular weight. The median (black line) and boxes spanning the interquartile range (IQR) range are shown. Each whisker extends to the most extreme value no more than 1.5 × IQR from the box. The wide kernel density indicates that there is a high density of points; where it is narrow, a low density of points. The tight distribution for P#1 indicates the proteins detected from this extraction buffer shows mostly similar distribution and the violin is short. While the violin plot with short and narrow distribution for instance P#3 and P#4 indicates the calculated aromaticity and molecular weight scores are widely dispersed.

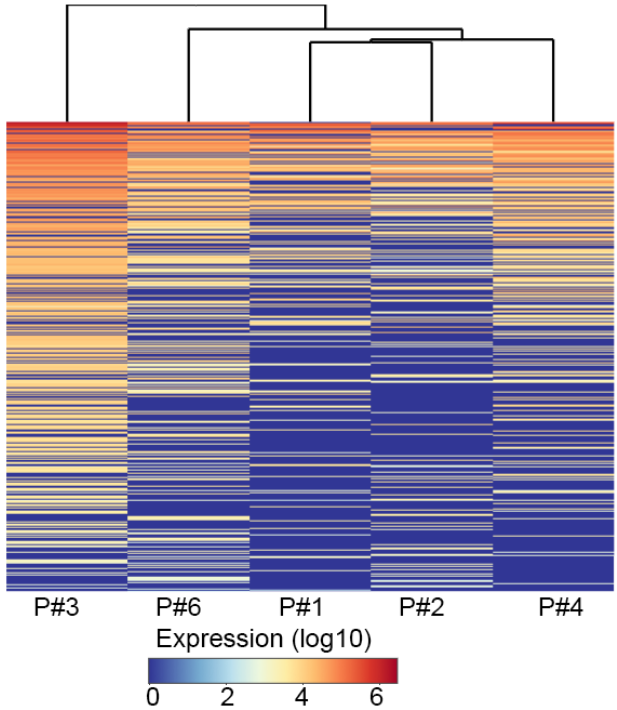

Figure S3: Heatmap plot and hierarchical clustering showing the peptide peak apex intensities across the five extraction protocols applied on roasted whole cricket sample.

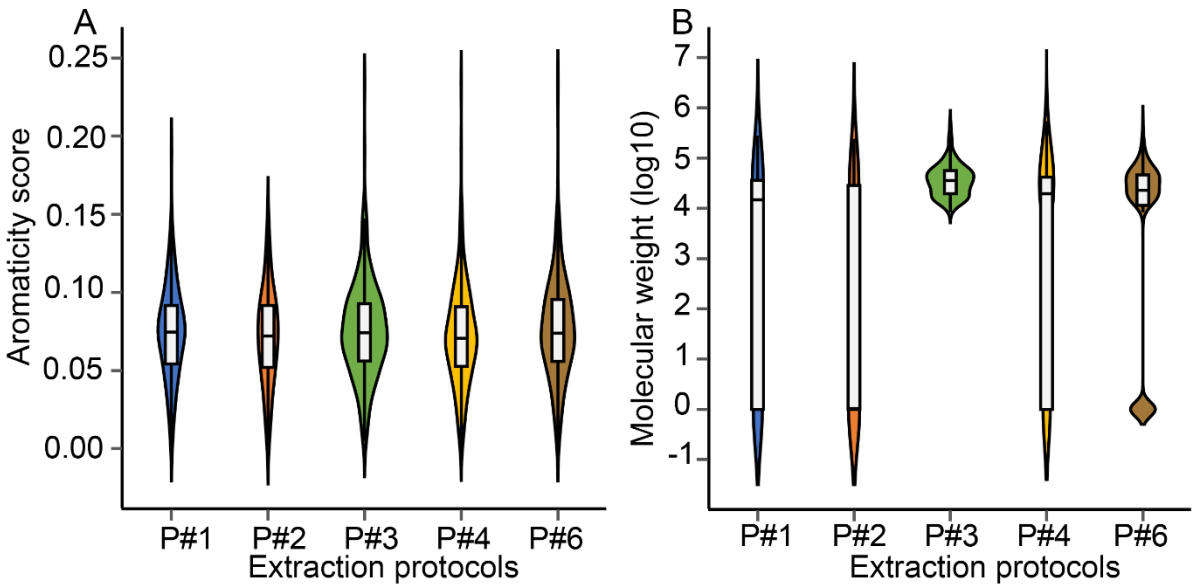

Figure S4: Five extraction protocols yield different physicochemical properties for roasted whole cricket samples: (**A**) aromaticity score; and (**B**) molecular weight. The median (black line) and boxes spanning the interquartile range (IQR) range are shown. Each whisker extends to the most extreme value no more than 1.5 × IQR from the box. Explanation of data distribution and kernel density for protocol-specific detected proteins and their corresponding aromaticity score and molecular weight was explained in **Fig S2**.

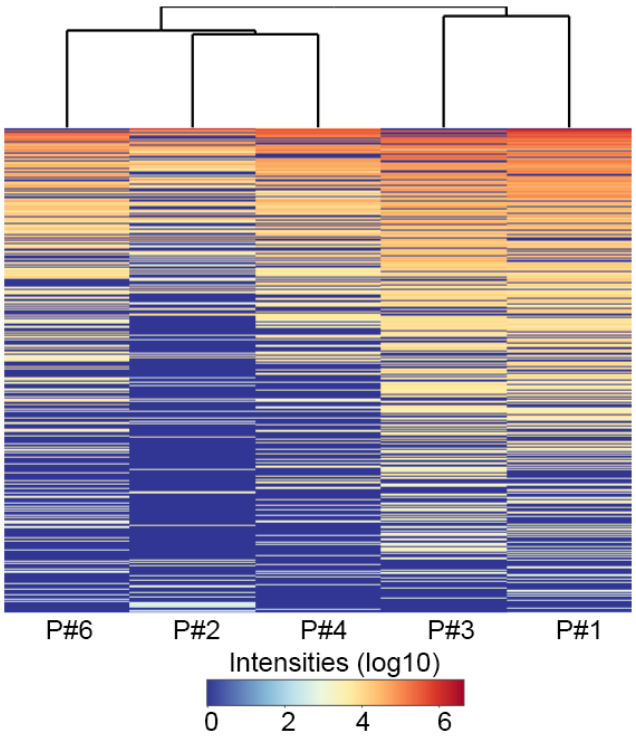

Figure S5: Heatmap plot and hierarchical clustering showing the peptide peak apex intensities across the five extraction protocols applied on roasted cricket powder sample.

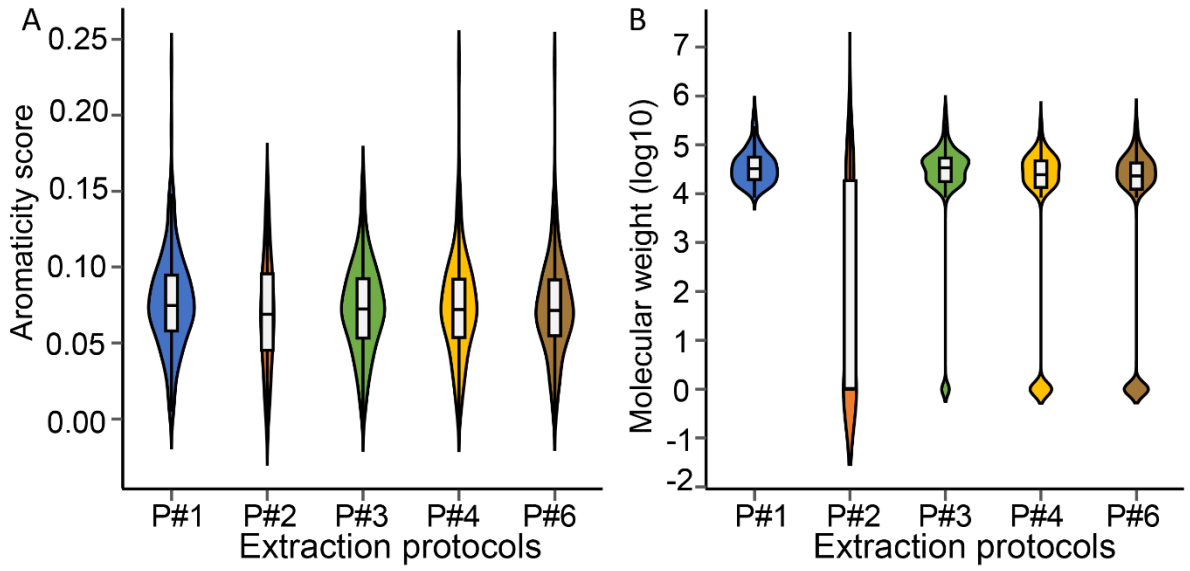

Figure S6: Five extraction protocols yield different physicochemical properties for roasted cricket powder samples: (**A**) aromaticity score (**B**) molecular weight. The median (black line) and boxes spanning the interquartile range (IQR) range are shown. Each whisker extends to the most extreme value no more than 1.5 × IQR from the box. Explanation of data distribution and kernel density for protocol-specific detected proteins and their corresponding aromaticity score and molecular weight was explained in **Fig S2**.

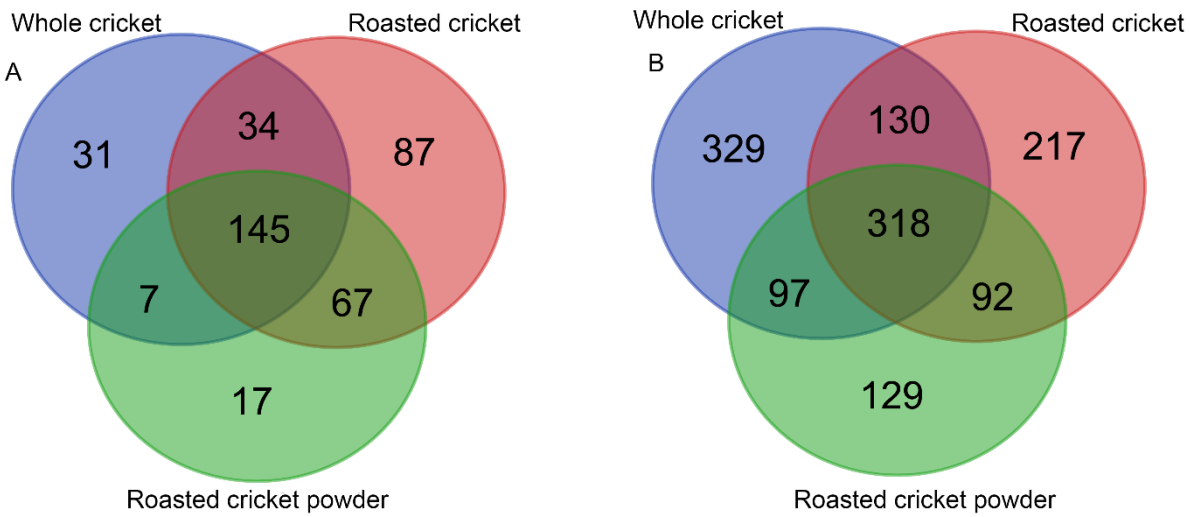

Figure S7: Comparison of proteins (**A**) and peptides (**B**) extracted with (P3) across three types of cricket samples.

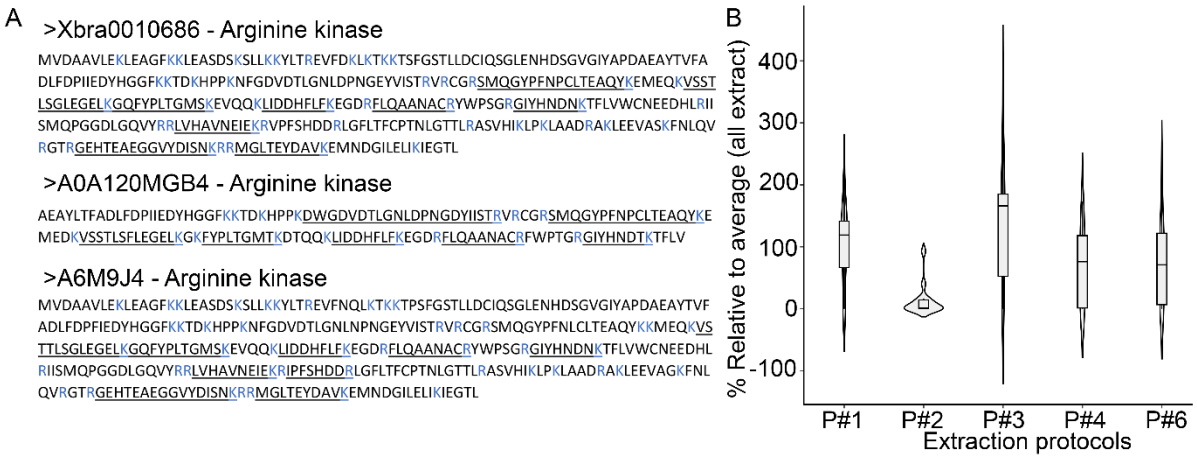

Figure S8: Analysis of arginine kinase (AK) protein isoforms by LC-MRM-MS examining whole cricket samples. (**A**) AK proteins identified from cricket samples. Underlined peptides were monitored by LC-MRM-MS from each protein. (**B**) Violin plots show the extraction protocol-dependent significant differences for monitored peptides from AK.

Table S1. Protein estimation by Bradford analysis (or BCA analysis for protocols *P7-P8*) and summary of proteins (and peptides) identified following eight extraction protocols. For Bradford assay, data shown as mean ± S.D. (CV) where n=4 replicate analyses. In each case, 100 µg of protein was loaded onto each 30 kDa filter for LC-MS/MS. The number of proteins identified at global 1% FDR (and distinct peptides identified with local 95% confidence) are shown. P=Protocol number; NA = Not acquired.

Sample type (source)	**P1 (SDS + Tris-HCl)**	**P2 (Tris-HCl + 2M Urea)**	**P3 (Urea + Thiourea + Tris-Hcl)**	**P4 (Urea + Thiourea + HEPES)**	**P5 (Tris-HCl 50 mM)**	**P6 (Defat + 8M Urea)**	**P7 (50 mM Tris + 1M NaCl)**	**P8 (50 mM Tris-HCl + 1M NaCl + 6M Urea)**
Bradford assay								
Whole [Grubsup]	7.68 ± 0.99 (12.9%)	0.91 ± 0.27 (30.3%)	16.1 ± 6.3 (39.2%)	9.7 ± 1.3 (13.1%)	0.18 ± 0.12 (62.3%)	6.96±1.30 (18.7%)	5.95±0.02 (0.34%)	8.36±0.05 (0.60%)
Roasted whole [Grilo]	6.79 ± 0.61 (8.9%)	0.77 ± 0.09 (11.6%)	21.1 ± 9.3 (44.1%)	9.1 ± 2.7 (29.3%)	0.14 ± 0.04 (29.7%)	4.47±0.84 (18.9%)	7.33±0.01 (1.4%)	9.14±0.01 (0.11%)
Roasted powder [Grilo]	7.36 ± 0.50 (6.8%)	1.38 ± 0.14 (9.9%)	17.0 ± 4.7 (27.4%)	10.5 ± 0.6 (5.4%)	0.33 ± 0.21 (63.3%)	3.51±0.33 (9.5%)	7.33±0.01 (0.14%)	11.10±0.06 (0.54%)
LC-MS/MS identification								
Whole [Grubsup]	369 (1175)	156 (736)	230 (1543)	423 (1275)	N/A	345 (1237)	35 (186)	89 (399)
Roasted whole [Grilo]	242 (709)	210 (792)	498 (1332)	300 (965)	N/A	351 (1123)	33 (272)	55 (344)
Roasted powder [Grilo]	364 (1226)	116 (678)	300 (1259)	259 (1011)	N/A	261 (1101)	38 (209)	113 (468)

Table S2: AK peptides monitored by LC-MRM-MS.

Peptide	**Protein accession**	**Matching allergenic species**	**No. matching allergens (>80% BLAST match)**
SMQGYPFNPCLTEAQYK	Xbra0010686	*Penaeus vannamei* (White Pacific shrimp)	7
VSSTLSGLEGELK	Xbra0010686	*Penaeus vannamei* (White Pacific shrimp)	1
GQFYPLTGMSK	Xbra0010686	*Penaeus vannamei* (White Pacific shrimp)	4
LIDDHFLFK	Xbra0010686	*Penaeus vannamei* (White Pacific shrimp) *Blattella germanica* (German cockroach)	6
FLQAANACR	Xbra0010686	*Penaeus vannamei* (White Pacific shrimp) *Blattella germanica* (German cockroach)	2
GIYHNDNK	Xbra0010686	*Scylla paramamosain* (green mud crab)	2
LVHAVNEIEK	Xbra0010686	*Penaeus vannamei* (White Pacific shrimp) *Blattella germanica* (German cockroach)	3
GEHTEAEGGVYDISNK	Xbra0010686	*Penaeus vannamei* (White Pacific shrimp) *Blattella germanica* (German cockroach) *Scylla paramamosain* (green mud crab)	4
MGLTEYDAVK	Xbra0010686	*Penaeus vannamei* (White Pacific shrimp) *Blattella germanica* (German cockroach)	4
DWGDVDTLGNLDPNGDYIISTR	A0A120MGB4	*Blattella germanica* (German cockroach)	2
VSSTLSFLEGELK	A0A120MGB4	*Penaeus vannamei* (White Pacific shrimp)	1
FYPLTGMTK	A0A120MGB4	*Penaeus vannamei* (White Pacific shrimp)	2
GIYHNDTK	A0A120MGB4	*Scylla paramamosain* (green mud crab)	1
VSTTLSGLEGELK	A6M9J4	Unknown	0
IPFSHDDR	A6M9J4	*Blattella germanica* (German cockroach)	2

Table S3: Identification of protocol-specific enrichment of allergen proteins from three types of cricket samples.

Whole Cricket							
Protein	**P1**	**P2**	**P3**	**P4**	**P6**	**P7**	**P8**
Myosin	4	2	1	5	2	3	3
Tropomyosin	3	6	4	5	4	2	3
Troponin	4	3	4	5	4	1	4
Arginine kinase	2	1	2	3	2	0	1
Hexamerin	0	1	0	0	1	0	1
Triosephosphate isomerase	1	0	1	1	1	0	0
Tubulin alpha chain	1	1	1	1	1	0	0
Roasted whole cricket							
Myosin	4	3	4	4	3	2	2
Tropomyosin	3	5	6	5	3	2	3
Troponin	4	4	4	4	3	1	3
Arginine kinase	2	1	2	2	2	0	0
Hexamerin	1	1	1	1	1	1	0
Triosephosphate isomerase	0	0	0	0	0	0	0
Tubulin alpha chain	2	2	2	1	2	0	0
Roasted cricket powder							
Myosin	6	3	4	5	3	2	4
Tropomyosin	5	5	5	5	6	2	5
Troponin	4	5	4	4	5	2	5
Arginine kinase	2	1	3	2	3	0	0
Hexamerin	1	0	1	1	1	0	0
Triosephosphate isomerase	0	0	0	1	1	0	0
Tubulin alpha chain	1	1	2	1	1	0	0