

Accepted Manuscript

Rapid and comprehensive discovery of unreported shellfish allergens using large-scale transcriptomic and proteomic resources

Roni Nugraha, MSc, Sandip D. Kamath, PhD, Elecia Johnston, PhD, Kyall R. Zenger, MSc(Hons), PhD, Jennifer M. Rolland, PhD, Robyn E. O'Hehir, FRACP, PhD, FRCP, FRCPATH, Andreas L. Lopata, PhD

PII: S0091-6749(17)32915-9

DOI: [10.1016/j.jaci.2017.11.028](https://doi.org/10.1016/j.jaci.2017.11.028)

Reference: YMAI 13175

To appear in: *Journal of Allergy and Clinical Immunology*

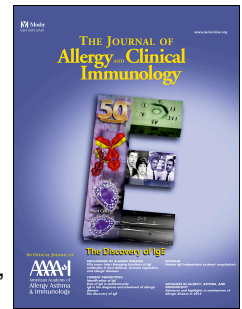
Received Date: 4 July 2017

Revised Date: 27 October 2017

Accepted Date: 1 November 2017

Please cite this article as: Nugraha R, Kamath SD, Johnston E, Zenger KR, Rolland JM, O'Hehir RE, Lopata AL, Rapid and comprehensive discovery of unreported shellfish allergens using large-scale transcriptomic and proteomic resources, *Journal of Allergy and Clinical Immunology* (2018), doi: 10.1016/j.jaci.2017.11.028.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Manuscript I.D. JACI-D-17-00913R2

Rapid and comprehensive discovery of unreported shellfish allergens using large-scale transcriptomic and proteomic resources

Roni Nugraha, MSc a,b,c,d, Sandip D. Kamath, PhD a,b,c, Elecia Johnston, PhD a,b,c, Kyall R. Zenger, MSc(Hons), PhD, Jennifer M. Rolland, PhD f, g, Robyn E. O'Hehir, FRACP, PhD, FRCP, FRCPath f,g, Andreas L. Lopata, PhD a,b,c *

a Centre for Biodiscovery and Molecular Development of Therapeutics, James Cook University, Townsville, Australia

b Australian Institute of Tropical Health and Medicine, James Cook University, Townsville, Australia

c College of Public Health, Medical and Veterinary Science, James Cook University, Townsville, Australia

d Department of Aquatic Product Technology, Bogor Agricultural University, Bogor, Indonesia

e Centre for Sustainable Tropical Fisheries and Aquaculture, College of Marine & Environmental Sciences, James Cook University, Townsville, Australia

f Department of Immunology and Pathology, Monash University, Melbourne, Australia

g Department of Allergy, Immunology and Respiratory Medicine, The Alfred Hospital and Monash University, Melbourne, Australia

* Corresponding Author:

Prof. Andreas L. Lopata

Molecular Allergy Research Laboratory

Centre for Biodiscovery and Molecular Development of Therapeutics

Building 47, James Cook Drive, Douglas Campus

James Cook University, TOWNSVILLE, QLD 4811, Australia

Telephone +61 (07) 47814563

Fax +61 (07) 47816078

Email: andreas.lopata@jcu.edu.au

35

36 **Capsule Summary**

37 A method for rapid large scale discovery of allergenic proteins, revealing molecular,
38 immunological and clinical variability, could considerably aid future management of
39 patients with shellfish allergy and development of new strategies for
40 immunotherapeutics.

41

42

43 **Keywords:** bioinformatics, food allergy, diagnostics, IgE cross-reactivity, proteomic,
44 shellfish.

45

To the Editor:

Discovery of allergens in various food and inhalant sources is central to our understanding of the molecular mechanisms of allergic reactions. Allergen characterization is the most important underlying factor for better patient management and the design and development of novel immunotherapeutics.¹

Of the 'Big eight' allergen food groups, shellfish presents a unique challenge in terms of allergen discovery due to the large number and diversity of consumed species, leading to heterogeneity of allergen structure and cross-reactivity among various sources. At present, 31 crustacean allergens have been officially registered in the WHO/IUIS Allergen Nomenclature Database as compared to 4 mollusk allergens due to pitfalls in current allergy discovery approaches. Co-sensitization of patients with crustacean and mollusk allergy is often described, however the current diagnostic approaches to manage these patients are not based on sufficient molecular knowledge of these shellfish allergens.

In this study, an innovative strategy was employed for the identification of novel cross-reactive allergenic proteins in the Pacific Oyster, using a combined approach of high-throughput screening of the genome and proteome along with comparative confirmation using protein family domains and allergen databases, identifying allergens with high probability. The study design (**Fig E1**), and **methods** are described in this article's **Online Repository**.

Initially, we utilized a bioinformatics approach for identifying potential allergens in the Pacific Oyster proteome. To select promising candidates, we first enumerated proteins from the Pacific Oyster genome-derived proteome (25,982 proteins). Subsequently allergen protein domains were detected by interrogating Hidden Markov Models (HMM) profiles using HMMER3.² The term 'allergen protein domain' is used to define protein families that include allergenic proteins. These domains were derived from the Pfam database version 29.0³ that accounted for only 273 of the known 16,295 protein domains (2%). We discovered that 2,504 Pacific Oyster proteins are associated with at least one of the known allergen protein domains. We next utilized the full-length protein sequence alignment to these 2,504 proteins using BLASTP program against 2,117 known allergens to identify proteins that are homologous to the known allergens. A minimum of 50% shared identity and an upper threshold of E = value of 10^{-7} ⁴ were used as display limits of alignment. Using

these two criteria, we identified 95 proteins that have significant sequence identity with known allergenic proteins. Of these, 22 proteins were categorized as “very likely allergenic” with an amino acid identity $\geq 70\%$ with known allergens and 73 proteins as “likely allergenic” with amino acid identities between 50% and 70% (**Fig. 1A.I** and **see Table E1 in this article’s Online Repository**).

Furthermore, a comprehensive analysis of the expression of allergen coding genes, across different developmental stages of the oyster and different tissues was conducted based on currently available transcriptomic data from the oyster genome project.⁵ As illustrated in **Fig. 1A.II** and **Fig E2 in this article’s Online Repository**, potential allergens were differentially expressed across developmental stages and various tissues of the Pacific Oyster. For example both isoforms of tropomyosin are highly expressed in the adult stage, specifically in the adductor muscle and mantle tissues. In contrast, heat shock protein HSP 68 is expressed at a higher level in early developmental stages as compared to the adult oyster.

The next step involving allergenomics included the identification of IgE-binding proteins using serum of five mollusk-sensitized patients. These patients reported clinical reactivity to various shellfish species including, oyster, mussel, prawn and lobster as well as reactivity to fish and house dust-mite (**see Table E2 in this article’s Online Repository**). Protein extracts from raw and heated Pacific Oyster were separated by 2D electrophoresis, and visualized using Coomassie Blue or IgE antibody binding analyzed by immunoblotting (**Fig. 1B**). Twenty-two IgE-reactive spots were excised from the resolved raw oyster extract and five spots from the heated extract (**see Fig E3 in this article’s Online Repository**), followed by in-gel tryptic digestion and mass spectrometric analysis. We identified a total of 332 and 26 proteins in the IgE-reactive spots of the raw and heated extracts, respectively. Some proteins were discovered in several different spots, due to polymorphisms (**source data is available via ProteomeXchange with identifier PXD006624**).

The composition of water-soluble proteins in the raw and heated oyster extracts were identified using shotgun mass spectrometry. We identified 1,097 proteins in the raw extract and 133 proteins in the heated extract. The comparative analysis of relative protein abundance, using the emPAI scores, indicated that sarcoplasmic calcium binding protein is the most abundant water-soluble protein in both extracts, followed by arginine kinase and fatty acid binding protein in the raw extract. Other known heat-stable shellfish allergens identified in the heated extracts are

tropomyosin and myosin light chain (**Fig. 1C**). Most of allergens that have been characterized in commonly consumed shellfish species have been demonstrated to be heat stable.⁶

The comparative evaluation of *in silico* bioinformatics analysis with the allergenomic and proteomic data generated, confirms that 24 proteins in the raw extract and four in the heated extract were identified using all three methods (**Fig. 1D** and **Table 1**). These allergenic proteins have a high probability to elicit immunological cross-reactivity in oyster-allergic patients to other organisms containing the same or highly similar allergens. For example, triosephosphate isomerase and 78 kDa glucose regulated protein share 76% and 65% amino acid identity with homologous dust-mite allergens respectively, potentially responsible for clinical mollusk-mite cross-reactivity, previously reported during mite immunotherapy.⁷ Furthermore oyster enolases share 61-71% amino acid identity with enolase from fish, latex, grass pollen and fungi. However, clinical reactivity due to enolase of oyster or these other allergen sources has not yet been reported. The importance of early identification of sensitizing allergens has been corroborated by several clinical studies, including the recently described chicken-fish syndrome.⁸ Similarly, Chan *et al.*⁹ conducted the first genomic analysis of house dust-mites (HDM; *Dermatophagoides farinae*) and developed a combined genomic-transcriptomic-proteomic approach to elucidate previously unreported HDM allergens, identifying 12 new allergens also present in non-mite species.

In summary, our innovative methodological approach utilizing biochemical and computational tools in addition to antibody reactivity was successful in identifying 24 previously unreported allergens from over 25,000 proteins of the Pacific Oyster. This is the first study to demonstrate the presence of multiple novel allergens in any shellfish species. Some of these are common to very different allergen sources incorporating animal, including fish and mites, as well as plant allergens from pollen, latex and fungi. Importantly we demonstrated that these allergenic proteins identified are reactive to shellfish allergic patients' IgE antibodies under *in vitro* conditions. The rapid and comprehensive analysis of unreported allergenic proteins fills a major gap in the current management of patients at high-risk of concurrent reactivity to diverse allergen sources.

Authors:

Roni Nugraha, MSc^{a,b,c,d},
Sandip D. Kamath, PhD^{a,b,c},
Elecia Johnston, PhD^{a,b,c},
Kyll R. Zenger, MSc(Hons), PhD^e,
Jennifer M. Rolland, PhD^{f,g},
Robyn E. O'Hehir, FRACP, PhD, FRCP, FRCPATH^{f,g},
Andreas L. Lopata, PhD^{a,b,c,*}

ACKNOWLEDGEMENTS

A.L.L. received financial support from the Australian Research Council (ARC). S.D.K is holder of the NHMRC Peter Doherty Fellowship and supported by the Allergy and Immunology Foundation of Australasia (AIFA), and R.N. received a James Cook University Postgraduate Research Scholarship.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests

The mass spectrometry proteomics data are available via ProteomeXchange with the dataset identifier PXD006624 and 10.6019/PXD006624.

REFERENCES

1. Yu W, Freeland DMH, Nadeau KC. Food allergy: immune mechanisms, diagnosis and immunotherapy. *Nat Rev Immunol* 2016; 16:751-65.
2. Eddy SR. Accelerated Profile HMM Searches. *PLoS Comput Biol* 2011; 7:e1002195.
3. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 2016; 44:D279-85.
4. Goodman RE, Ebisawa M, Ferreira F, Sampson HA, van Ree R, Vieths S, et al. AllergenOnline: A peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. *Mol Nutr Food Res* 2016; 60:1183-98.
5. Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, et al. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 2012; 490:49-54.
6. Kamath SD, Rahman AM, Voskamp A, Komoda T, Rolland JM, O'Hehir RE, et al. Effect of heat processing on antibody reactivity to allergen variants and fragments of black tiger prawn: A comprehensive allergenomic approach. *Mol Nutr Food Res* 2014; 58:1144-55.
7. Tsapis M, Chabane H, Teychene AM, Laurent M, Roland PN, de Pontual L. Fatal anaphylaxis after snail ingestion in a child after 3 years of house dust mite immunotherapy. *Rev Fr Allergol* 2013; 53:436-8.
8. Kuehn A, Codreanu-Morel F, Lehnert-Weber C, Doyen V, Gomez-Andre SA, Bienvenu F, et al. Cross-reactivity to fish and chicken meat - a new clinical syndrome. *Allergy* 2016; 71:1772-81.
9. Chan TF, Ji KM, Yim AK, Liu XY, Zhou JW, Li RQ, et al. The draft genome, transcriptome, and microbiome of *Dermatophagoides farinae* reveal a broad spectrum of dust mite allergens. *J Allergy Clin Immunol* 2015; 135:539-48.

198 Table 1: Proteins identified across all three methods with their matched allergens source and routes of sensitization. Proteins
 199 identified in both raw and heated extract are shaded. The proteins are sorted based on amino acid sequence identity with their
 200 homologous allergens in descending order.

No	Entry	Protein name	Homologous allergen in IUIS database*	Amino Acid identity (%)	Overlap (AA)	E-value	Organism	Route of sensitization	Source
1	B7XC66	Tropomyosin	Hel as 1	75.7	284	4.00E-133	<i>Helix aspersa</i> (Brown garden snail)	Ingestion	Animal
2	K1PCV6	Triosephosphate isomerase	Cra c 8	74.03	77	2.00E-41	<i>Crangon crangon</i> (North sea shrimp)	Ingestion	Animal
3	K1PJ59	Triosephosphate isomerase	Der f 25	73.37	169	6.00E-92	<i>Dermatophagoides farinae</i> (House dust mite)	Inhalation	Animal
4	K1QX37	Enolase	Thu a 2	72.83	357	0	<i>Thunnus albacares</i> (Yellowfin tuna)	Ingestion	Animal
5	K1Q350	Glyceraldehyde-3-phosphate dehydrogenase	Tri a 34	71.21	330	7.00E-173	<i>Triticum aestivum</i> (Wheat)	Ingestion	Plant
6	Q75W49	78kDa glucose regulated protein	Cor a 10	71.04	618	0	<i>Corylus avellana</i> (European hazelnut)	Inhalation	Plant
7	K1QTC1	Paramyosin [†]	-	68.59	799	0	<i>Haliotis discus discus</i> (Abalone)	Ingestion	Animal
8	K1R8R6	Fructose-bisphosphate aldolase	Sal s 3	67.31	364	4.00E-177	<i>Salmo salar</i> (Atlantic salmon)	Ingestion	Animal

9	K1RTQ6	Fructose-bisphosphate aldolase	Sal s 3	66.67	363	1.00E-178	<i>Salmo salar</i> (Atlantic salmon)	Ingestion	Animal
10	K1PNQ5	Heat shock protein HSP 90-alpha 1	Asp f 12	66.26	412	0	<i>Aspergillus fumigatus</i>	Inhalation	Fungi
11	K1QNV6	Tropomyosin	Tod p 1	65.31	271	3.00E-84	<i>Todarodes pacificus</i> (Squid)	Ingestion	Animal
12	K1R266	Retinal dehydrogenase 1	Tyr p 35	61	472	1.00E-170	<i>Tyrophagus putrescentiae</i> (Storage mite)	Inhalation	Animal
13	K1QNT7	Aldehyde dehydrogenase, mitochondrial	Tyr p 35	60	482	1.00E-180	<i>Tyrophagus putrescentiae</i> (Storage mite)	Inhalation	Animal
14	K1QVK0	Transaldolase	Fus p 4	59.2	326	1.00E-125	<i>Fusarium proliferatum</i>	Inhalation	Fungi
15	K1QVG5	Retinal dehydrogenase 1	Tyr p 35	59	474	1.00E-169	<i>Tyrophagus putrescentiae</i> (Storage mite)	Inhalation	Fungi
16	K1PLF9	Arginine kinase	Bomb m 1	59.13	345	3.00E-147	<i>Bombyx mori</i> (Silk moth)	Ingestion	Animal
17	K1Q3F4	Inorganic pyrophosphatase	Der f 32	58.24	261	1.00E-113	<i>Dermatophagoides farinae</i> (House dust mite)	Inhalation	Animal
18	K1Q9Z4	Aldehyde dehydrogenase	Tyr p 35	56	195	8.00E-65	<i>Tyrophagus putrescentiae</i> (Storage mite)	Inhalation	Animal
19	K1P9D0	Stress-70 protein,	Pen c 19	55.22	431	3.00E-163	<i>Penicillium citrinum</i>	Inhalation	Fungi

mitochondrial									
20	K1QX26	Endoplasmin	Asp f 12	54.04	198	9.00E-61	<i>Aspergillus fumigatus</i>	Inhalation	Fungi
21	K1Q7T5	Protein disulfide-isomerase	Alt a 4	52.27	44	3.00E-11	<i>Alternaria alternata</i>	Inhalation	Fungi
22	K1Q5P7	Peptidyl-prolyl cis-trans isomerase	Cat r 1	52.17	161	5.00E-53	<i>Catharanthus roseus</i> (Madagascar periwinkle)	Inhalation	Plant
23	K1R4Z3	Malate dehydrogenase, mitochondrial	Mala f 4	51.43	280	1.00E-91	<i>Malassezia furfur</i>	Inhalation	Fungi
24	K1Q6X5	Protein disulfide-isomerase	Alt a 4	50	68	1.00E-15	<i>Alternaria alternata</i>	Inhalation	Fungi

201

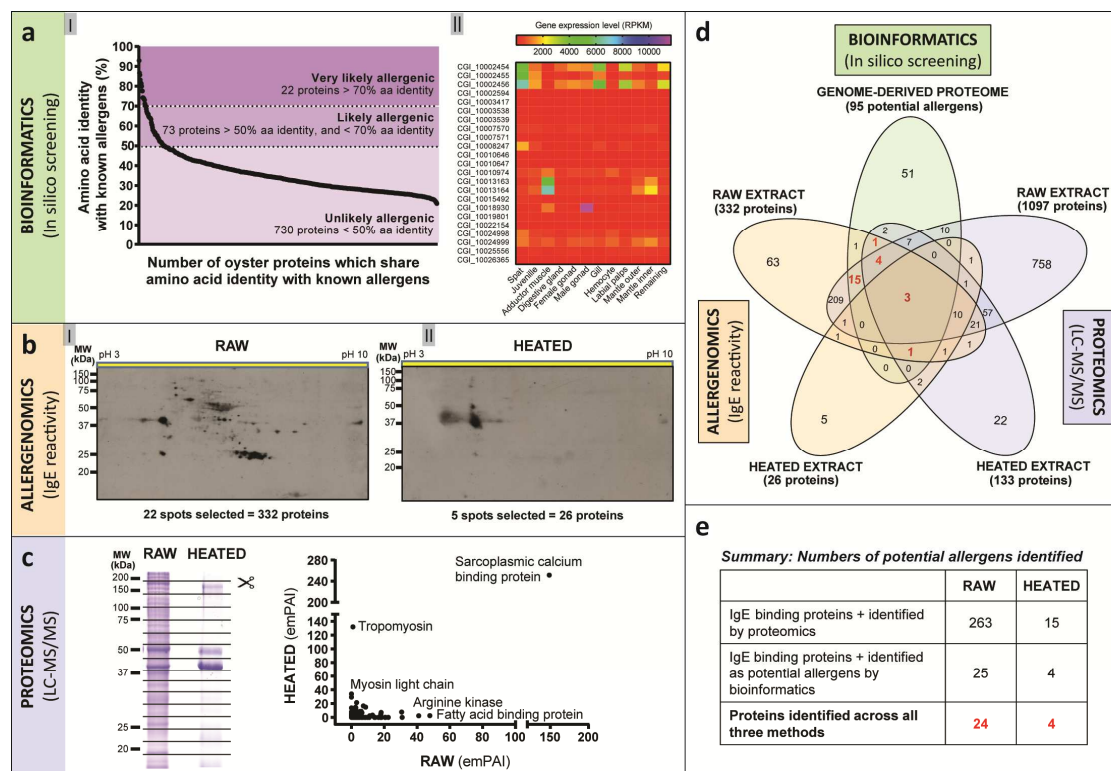
202 * Only homologous allergen with the highest identity has been displayed

203 ‡ Paramyosin currently not registered in IUIS database has over 50% identity. The alignment was determined against allergens

204 from AllergenOnline database

Figure Legends

Fig 1. A combined approach demonstrates the increased sensitivity of identifying allergens in low abundance. **(A.I)** *In silico* analysis identified 22 proteins as “very likely allergenic” (aa identity $\geq 70\%$) and 73 proteins as “likely allergenic” (aa identity $\geq 50\%$, and $\leq 70\%$). **(A.II)** The expression levels of the “very likely allergenic” proteins were different across development stages and organs of the Pacific Oyster. **(B)** Multiple spots in the raw **(I)** and heated extract **(II)** were recognized to bind serum IgE. The IgE-reactive spots in the raw and heated extract contained 332 and 26 proteins, respectively. **(C)** Shotgun proteomics analysis of the extracts indicates higher relative abundance of IgE binding proteins in the heated extract. **(D)** Venn diagram displaying proteins shared by the bioinformatics, the allergenomic and the proteomic approach. Numbers of proteins shared by the three methods are coloured in red. **(E)** Summary of numbers of potential allergens shared by different methods. Twenty-four proteins in the raw extract and four proteins in the heated extract were identified across all three methods.



Supplementary Methods (for Online Repository)***In silico identification of potential allergens from Pacific Oyster***

To identify potential allergens of Pacific Oyster, full-length sequence alignment of oyster proteins were carried out against a repertoire of known allergens. For this purpose, two datasets were assembled. The first dataset contained a FASTA file of 25,982 genome-derived proteins of the Pacific Oyster⁶ collected from the Uniprot database (Proteome ID UP000005408, last modification October 9, 2016). The second dataset contained 2117 allergen sequences compiled from two main allergen databases: the World Health Organization and International Union of Immunological Societies (WHO/IUIS) Allergen Nomenclature (<http://www.allergen.org/>)⁹ and the Food Allergy Research and Resource Program (FARRP) (Version 16, <http://www.allergenonline.org/>)⁴. Genbank accession IDs of all allergenic proteins were collected from these databases and the IDs uploaded in the Batch Entrez menu on the NCBI website to obtain the sequence of the protein and remove duplicate proteins. Prior to the sequence alignment, we filtered the oyster proteome by the Pfam domains containing allergens. The latest distribution of protein families from the allergen dataset were defined by running the *hmmsearch* program² against the Pfam database (version 29.0³). The BLASTP program was employed to align the Pacific Oyster proteins and the repertoire of known allergens using a cut-off E-value of 10^{-7} and sequence identity >50%.

Gene expression analysis of potential allergens in the Pacific Oyster

The expression levels of potential allergen genes were analysed from the available RNA sequencing data from the oyster genome project⁶. The expression profiles were analysed from two developmental stages: spat and juvenile, and from ten adult

organs, including the adductor muscle, the digestive gland, the female gonad, the male gonad, the gill, the hemocyte, the labial palp, the outer mantle, the inner mantle, and the remaining tissue. The expression levels were calculated by RPKM (Reads Per Kilobase of transcript per Million mapped reads).

Preparation of protein extracts

Fresh oyster specimens were purchased from the local market and stored at -20 °C prior to use. Protein extracts were prepared according to method of Kamath et al.¹⁰ with slight modification. The meat of the oyster was finely cut and homogenized in phosphate-buffered saline (PBS, pH 7.2) for 10 min, using an Ultra turrax homogenizer (IKA, Staufen, Germany). After gentle shaking at 4 °C for 3 h and centrifugation at 20,000 *g* for 20 min, supernatants were clarified through a glass fibre filter, followed by filtration through a 0.45 µm membrane filter (Sartorius AG, Goettingen, Germany) and stored at -80 °C before further use. To produce heated protein extracts, the meat was heated in PBS at 95-100 °C for 20 min instead of heating the raw extract, to mimic the way consumers are exposed to food allergens. The meat was removed after cooling, and the proteins were extracted using the same method as described above. The total protein content of the extracts was quantified using the bicinchoninic acid (BCA) assay (Thermo Scientific™, USA) following the manufacturer's instructions. A pre-diluted set of bovine serum albumin was used as protein standards (Pierce™).

Identification of total protein in the extracts

Total protein of the Pacific Oyster raw and heated extracts was identified after trypsin in-gel digestion. Briefly, 20 µg of the extracts were loaded onto a 12% polyacrylamide gel and run at 170 V for 1 h. The gels were cut into pieces and

washed with 25 mM ammonium bicarbonate (AMBIC). After being dried using a vacuum dryer, the gels were reduced by 20 mM dithiothreitol (DTT) at 65°C for 1 h and alkylated by 50 mM iodoacetamide for 40 min at 37°C in the dark. Gel pieces were washed and dried using a SpeedVac. Dried gel pieces were rehydrated with 20 ng/μl of trypsin dissolved in 40 mM ammonium bicarbonate and 10% acetonitrile for 1 h at room temperature and subsequently incubated overnight at 37°C. The digested proteins were acidified using 0.1% formic acid and the peptides were concentrated on a SpeedVac and subjected to Liquid chromatography tandem-mass spectrometry (LC-MS/MS) analysis.

Patient selection

Five subjects with a convincing clinical history of allergic reactivity to shellfish and one non-atopic subject were recruited from The Alfred Hospital Allergy Clinic, Melbourne, Victoria, Australia. Skin prick testing and oral challenge with mollusc extracts were not conducted routinely in these patients, in keeping with the clinicians' preference for safer serum specific allergen IgE testing in adult patients due to comorbidities, together with clinical history of reactions on exposure. Ethics approval for this study was granted by James Cook University's Ethics Committee (Project number H4313) in collaboration with The Alfred Hospital (Project number 192/07) and Monash University's Ethics Committees (MUHREC CF08/0225).

Identification of IgE-reactive proteins using 2D-gel electrophoresis and immunoblotting

The proteins in the oyster extracts were separated by 2D-gel electrophoresis. Proteins were first lyophilized and resuspended in a 8 M urea, 2% CHAPS, 50 mM DTT and 0.2%(w/v) Biolyte 3/10 ampholytes buffer. The extract was subjected to

isoelectric focusing using a 3–10 NL pH range 12% ReadyStrip™ IPG Strips (Bio-Rad, USA), as per the manufacturer's instructions. Briefly, 185 µl of rehydration buffer containing 200 µg of raw extract or 100 µg of heated extract was loaded on to the IPG tray, and a strip was gently placed side down on to the sample and left to incubate overnight at room temperature. Isoelectric focusing was conducted using a PROTEAN IEF cell (Bio-Rad, USA) with a maximum current of 50 µA/strip. After focusing, the IPG strips were equilibrated with equilibration buffer (ES) 1 and ES 2. The ES 1 contained 6 M urea, 2% SDS, 20% glycerol, 0.375 M Tris-HCl pH 8.8 and 2% (w/v) DTT, and the ES 2 contained the same solution as ES 1 except it contained 4% iodoacetamide instead of DTT. The strip was then washed with SDS-PAGE running buffer and laid on top of a 12% polyacrylamide gel. The gels were run at 170 V until bromophenol blue dye reached the bottom of each gel. Gels were either stained with Coomassie Brilliant Blue R-250 or the separated proteins were transferred to polyvinylidene difluoride (PVDF) membrane. Protein transfer was performed using the Semi-dry TransBlot Apparatus (BioRad, USA). After blocking with 5% (w/v) skim milk powder in phosphate buffered saline with 0.05% Tween (PBS-T), the membrane was incubated with a serum pool from five shellfish-allergic patients at a 1:20 dilution overnight at 4°C with shaking. The membrane was subsequently incubated with 1:10.000 dilution of rabbit anti-human IgE (Dako, France) followed by 1:10.000 dilution of horseradish peroxidase-conjugated goat anti-rabbit antibodies (Promega, USA). Specific IgE binding was detected by chemiluminescence and exposed to photographic film (GE Healthcare Biosciences, USA) to visualize the antibody-binding protein spots. Serum from a non-atopic donor was used as a negative control. IgE reactive spots were annotated using the

proteome map and corresponding bands were cut, tryptic digested and analysed using mass spectrometry.

Mass Spectrometry Analysis

The LC-MS/MS was carried out on an LTQ Orbitrap Elite (Thermo Scientific) with a nanoESI interface in conjunction with an Ultimate 3000 RSLC nanoHPLC (Dionex Ultimate 3000) at the Bio21 Institute, Melbourne. The LC system was equipped with an Acclaim Pepmap nano-trap column (Dionex-C18, 100 Å, 75 µm x 2 cm) and an Acclaim Pepmap RSLC analytical column (Dionex-C18, 100 Å, 75 µm x 50 cm). The tryptic peptides were injected into the enrichment column at an isocratic flow of 5 µL/min of 3% v/v CH₃CN containing 0.1% v/v formic acid for 5 min before the enrichment column was switched in-line with the analytical column. The eluents were 0.1% v/v formic acid (solvent A) and 100% v/v CH₃CN in 0.1% v/v formic acid (solvent B). The flow gradient was (i) 0-5 min at 3% B, (ii) 5-25 min, 3-25% B (iii) 25-27 min, 25-40% B (iv) 27-29 min, 40-80% B (v) 29-31 min at 80% B (vi) 31-32 min, 80-3% B and (viii) 32-38 min at 3% B. The LTQ Orbitrap Elite spectrometer was operated in the data-dependent mode with nanoESI spray voltage of 1.8 kV, a capillary temperature of 250°C and S-lens RF value of 55%. All spectra were acquired in positive mode with full scan MS spectra from m/z 300-1650 in the FT mode at 240,000 resolution. Automated gain control was set to a target value of 1.0⁻⁶ and a lock mass of 445.120025 was used. The top 20 most intense precursors were subjected to rapid collision induced dissociation (rCID) with a normalized collision energy of 30 and activation q of 0.25. A dynamic exclusion of 30 seconds was applied for repeated precursors.

Protein Identification

All MS/MS files were analysed using Mascot v2.4 against the in-house database of the oyster proteome downloaded from the UniProt, supplemented with the common Repository of Adventitious Proteins sequences. Search parameters were as follows: precursor mass tolerance of 200 ppm, fragment mass tolerance of 0.6 Da (CID). Carbamidomethyl (C) was set as a fixed modification and oxidation (M) and deamidated (NQ) were set as variable modifications. Trypsin with a maximum of 3 missed cleavages was used as the cleavage enzyme. Scaffold (version Scaffold_4.7.3, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Peptide Prophet algorithm¹¹ with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm¹². Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters.

Accession numbers

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE¹³ partner repository with the dataset identifier PXD006624 and 10.6019/PXD006624

Supplementary References

2. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol 2011; 7:e1002195
3. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 2016; 44:D279-85.

4. Goodman RE, Ebisawa M, Ferreira F, Sampson HA, van Ree R, Vieths S, et al. AllergenOnline: A peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. *Mol Nutr Food Res* 2016; 60:1183-98.
6. Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, et al. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 2012; 490:49-54.
9. Radauer, C. et al. Update of the WHO/IUIS Allergen Nomenclature Database based on analysis of allergen sequences. *Allergy* 2014;**69**, 413-9.
10. Kamath, S.D., Abdel Rahman, A.M., Komoda, T. & Lopata, A.L. Impact of heat processing on the detection of the major shellfish allergen tropomyosin in crustaceans and molluscs using specific monoclonal antibodies. *Food. Chem.* 2013; **141**, 4031-9.
11. Keller, A., Nesvizhskii, A.I., Kolker, E. & Aebersold, R. Empirical Statistical Model To Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Anal Chem.* 2002; **74**, 5383-5392.
12. Nesvizhskii, A.I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem.* 2003; **75**, 4646-58.
13. Vizcaíno JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* 2016; 44:D447-D56.

Table E1. Identified potential allergens of Pacific Oyster from the *in silico* analysis.

Protein	Gene name	Best matched allergen	Amino acid identity (%)
Very likely allergenic (amino acid identity $\geq 70\%$)			
Tropomyosin	CGI_10013163	tropomyosin [<i>Crassostrea gigas</i>]	92.94
Tropomyosin	CGI_10013164	tropomyosin, partial [<i>Crassostrea virginica</i>]	86.67
Tubulin alpha chain	CGI_10002456	Der f 33 allergen [Mite]	85.71
Tubulin alpha-1C chain	CGI_10002455	Der f 33 allergen [Mite]	83.2
Tubulin alpha-3 chain	CGI_10018930	Der f 33 allergen [Mite]	82.43
Tubulin alpha-1C chain	CGI_10024998	Der f 33 allergen [Mite]	81.8
Tubulin alpha-1C chain	CGI_10007570	Der f 33 allergen [Mite]	81.53
Tubulin alpha-1C chain	CGI_10024999	Der f 33 allergen [Mite]	81.35
78 kDa glucose-regulated protein	CGI_10015492	Aed a 8 [Mosquito]	81
Tubulin alpha-1C chain	CGI_10002454	Der f 33 allergen [Mite]	80
Tubulin alpha-1C chain	CGI_10008247	Der f 33 allergen [Mite]	77.7
Tubulin alpha-1A chain	CGI_10007571	Der f 33 allergen [Mite]	77.63
Fructose-bisphosphate aldolase	CGI_10019801	Thu a 3 [Tuna]	74.29
Heat shock protein 70 B2	CGI_10010646	Der f 28 allergen [Mite]	74.27
Heat shock protein 70 B2	CGI_10010647	Der f 28 allergen [Mite]	74.27
Triosephosphate isomerase	CGI_10003538	triosephosphate isomerase [shrimp]	74.03
Heat shock protein 68	CGI_10002594	Der f 28 allergen [Mite]	73.96
Triosephosphate isomerase	CGI_10003539	Der f 25 allergen [Mite]	73.37
Heat shock protein 70 B2	CGI_10003417	Der f 28 allergen [Mite]	73.13
Enolase	CGI_10022154	enolase [Tunas]	72.83
Fructose-bisphosphate aldolase	CGI_10025556	Thu a 3 [Tuna]	72.22
Glyceraldehyde-3-phosphate dehydrogenase	CGI_10010974	Glyceraldehyde-3-phosphate dehydrogenase [Wheat]	71.21
Peptidyl-prolyl cis-trans isomerase E (PPIase E)	CGI_10026365	Der f 6 [Mite]	70.19
Likely allergenic (amino acid identity $\geq 50\%$, but $< 70\%$)			
Ferritin	CGI_10016317	Ferritin [Mite]	69.92
Paramyosin	CGI_10001653	Paramyosin [<i>Abalone</i>]	68.59
Arginine kinase	CGI_10021480	Arginine kinase [<i>Octopus</i>]	66.96
Arginine kinase	CGI_10021483	Arginine kinase [<i>Octopus</i>]	66.67
Heat shock protein HSP 90- α 1	CGI_10017621	Asp f 12 [Fungus]	66.26
Ferritin	CGI_10021660	Ferritin [Mite]	66.06
Cytochrome c	CGI_10012574	Cur I 3 [Fungus]	66.02
Superoxide dismutase [Cu-Zn]	CGI_10017958	Ole e 3 [Olive tree]	66

Peptidyl-prolyl cis-trans isomerase 6	CGI_10022249	Cyclophilin [<i>Carrot</i>]	65.96
Ferritin	CGI_10027591	Ferritin [<i>Mite</i>]	65.58
78 kDa glucose-regulated protein	CGI_10008834	Der f 28 [<i>Mite</i>]	65.3
78 kDa glucose-regulated protein	CGI_10027395	Der f 28 [<i>Mite</i>]	65.25
Arginine kinase	CGI_10024056	Arginine kinase [<i>Octopus</i>]	64.65
Inorganic pyrophosphatase	CGI_10027722	Der f 32 [<i>Mite</i>]	64.24
Fructose-bisphosphate aldolase	CGI_10000078	Sal s 3 [<i>Salmon</i>]	64.12
60S ribosomal protein L3	CGI_10010529	Asp f 23 [<i>Fungus</i>]	63.96
Arginine kinase	CGI_10021482	Arginine kinase [<i>Octopus</i>]	63.83
Heat shock protein 68	CGI_10002823	Der f 28 [<i>Mite</i>]	63.34
40 kDa peptidyl-prolyl cis-trans isomerase	CGI_10015504	Cyclophilin [<i>Carrot</i>]	62.72
Peptidyl-prolyl cis-trans isomerase	CGI_10013880	Cyclophilin [<i>Carrot</i>]	61.76
Plasma kallikrein	CGI_10016607	Der f 3 [<i>Mite</i>]	61.7
60S ribosomal protein L3	CGI_10012282	Asp f 23 [<i>Fungus</i>]	60.76
Arginine kinase	CGI_10021481	Arginine kinase [<i>Octopus</i>]	60.6
Tubulin alpha chain	CGI_10018903	Der f 33 [<i>Mite</i>]	60.49
78 kDa glucose-regulated protein	CGI_10018425	Der f 28 [<i>Mite</i>]	59.9
Eukaryotic translation initiation factor 3 subunit I (Fragment)	CGI_10025943	For t 1 [<i>Midges</i>]	59.88
Calcium-binding atopy-related autoantigen 1	CGI_10026057	Hom s 4 [<i>Human</i>]	59.3
Transaldolase	CGI_10002311	Fus p 4 [<i>Fungus</i>]	59.2
Peptidyl-prolyl cis-trans isomerase B	CGI_10023851	Cyclophilin [<i>Carrot</i>]	58.7
Peptidyl-prolyl cis-trans isomerase	CGI_10023850	Asp f 27 [<i>Fungus</i>]	58.33
Aldehyde dehydrogenase, mitochondrial	CGI_10012671	Cla h 10 [<i>Fungus</i>]	58.02
Thioredoxin	CGI_10021611	Mala s 13 [<i>Yeast</i>]	58
Peptidyl-prolyl cis-trans isomerase	CGI_10024975	Cyclophilin [<i>Carrot</i>]	57.63
Thioredoxin domain-containing protein 5	CGI_10009327	Alt a 4 [<i>Fungus</i>]	57.45
60S acidic ribosomal protein P1	CGI_10009326	Alt a 12 [<i>Fungus</i>]	57.14
NK-tumor recognition protein	CGI_10007438	Asp f 27 [<i>Fungus</i>]	56.8
Thaumatococcus-like protein 1a	CGI_10012508	Pathogenesis related protein 5 [<i>Apple</i>]	55.62
Superoxide dismutase	CGI_10017307	Pis v 4 [<i>Pistachio</i>]	55.61

Peptidyl-prolyl cis-trans isomerase	CGI_10006179	Asp f 11 [<i>Fungus</i>]	55.56
Peptidylprolyl isomerase domain and WD repeat-containing protein 1	CGI_10011521	Mala s 6 [<i>Yeast</i>]	55.47
Stress-70 protein, mitochondrial	CGI_10016162	Pen c 19 [<i>Fungus</i>]	55.22
Peptidyl-prolyl cis-trans isomerase-like 6	CGI_10024382	Cat r 1 [<i>Periwinkle</i>]	55.22
U4/U6.U5 tri-snRNP-associated protein 1	CGI_10021218	Hom s 1 [<i>Human</i>]	54.96
Alpha-amylase	CGI_10022190	Bla g 11 [<i>Cockroach</i>]	54.37
Alpha-amylase	CGI_10022189	Bla g 11 [<i>Cockroach</i>]	54.18
Endoplasmin	CGI_10025730	Asp f 12 [<i>Fungus</i>]	54.04
Calmodulin	CGI_10006247	Amb a 9 [<i>Ragweed</i>]	53.7
Calmodulin	CGI_10014525	B1 protein allergen [<i>Bermuda grass</i>]	53.57
Aldehyde dehydrogenase	CGI_10021688	Cla h 10 [<i>Fungus</i>]	52.82
Retinal dehydrogenase 1	CGI_10026868	Cla h 10 [<i>Fungus</i>]	52.33
Protein disulfide-isomerase	CGI_10011652	Alt a 4 [<i>Fungus</i>]	52.27
Collagen alpha-3(VI) chain	CGI_10015798	collagen alpha [<i>Bovine</i>]	52.27
Peptidyl-prolyl cis-trans isomerase B	CGI_10027458	Cat r 1 [<i>Periwinkle</i>]	52.17
78 kDa glucose-regulated protein	CGI_10011272	Der f 28 [<i>Mite</i>]	52.1
Thioredoxin	CGI_10003765	Asp f 28 [<i>Fungus</i>]	51.85
Uncharacterized protein	CGI_10016401	Der f 3 [<i>Mite</i>]	51.85
Transcription elongation factor 1-like protein	CGI_10026699	Tri a 45 [<i>Wheat</i>]	51.81
Alpha-amylase	CGI_10023778	Bla g 11 [<i>Cockroach</i>]	51.75
Malate dehydrogenase, mitochondrial	CGI_10015004	Mala f 4 [<i>Yeast</i>]	51.43
Calmodulin	CGI_10022491	Tyr p 24 [<i>Mite</i>]	51.32
Retinal dehydrogenase 1	CGI_10026867	Cla h 10 [<i>Fungus</i>]	51.27
Peptidyl-prolyl cis-trans isomerase	CGI_10005876	Mala s 6 [<i>Yeast</i>]	51.16
Calmodulin	CGI_10011301	Par j 4 [<i>Weed</i>]	50.94
Calmodulin	CGI_10006481	Syr v 3 [<i>Common lilac</i>]	50.85
Alpha-amylase	CGI_10023781	Bla g 11 [<i>Cockroach</i>]	50.51
Elongation factor 1-beta	CGI_10021397	Pen c 22 [<i>Fungus</i>]	50.44
Calmodulin	CGI_10002924	B1 protein allergen [<i>Bermuda grass</i>]	50
Calmodulin-like protein 12	CGI_10004114	putative Cup a 4 allergen [<i>Cypress</i>]	50
Calmodulin	CGI_10011293	polcalcin [<i>Artemisia vulgaris</i>]	50
Supervillin	CGI_10014153	villin 2 [<i>Tobacco</i>]	50
Calmodulin	CGI_10017056	Bla g 6 [<i>Cockroach</i>]	50
Protein disulfide-isomerase	CGI_10026048	Alt a 4 [<i>Fungus</i>]	50

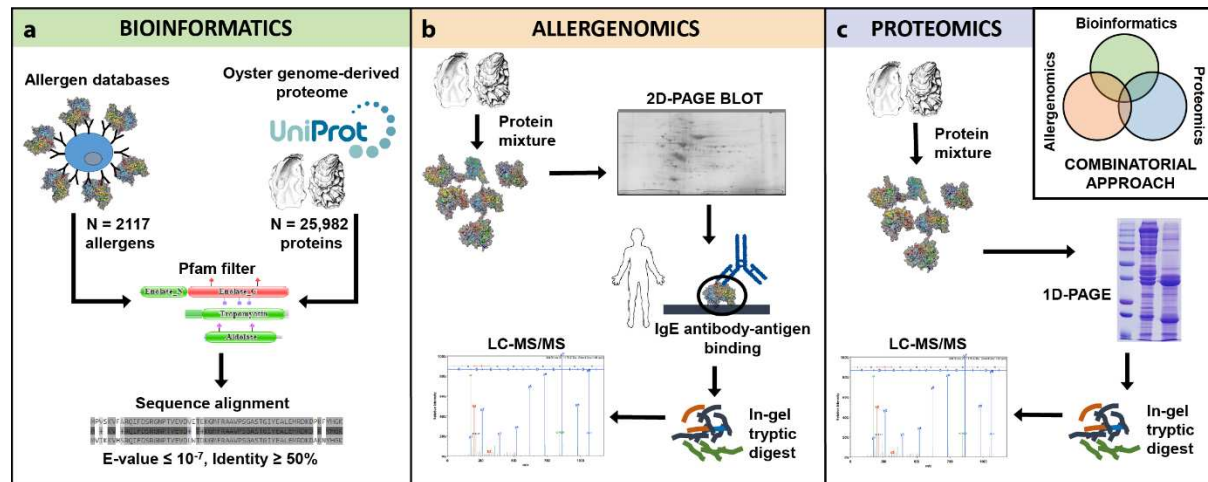
Table E2. Demographics of patients recruited for this study. Note: NT=Not Tested.

Subject	Sex	Age	Total IgE (kU/L)	Specific IgE (ImmunoCAP) kU/L			Skin Prick Test			Other Allergies
				Oyster	Shrimp	HDM	Oyster	Shrimp	HDM	
1	M	50	976	2.04	9.03	13.60	NT	NT	12 mm	Tuna, Cod
2	F	28	461	0.11	0.36	54.8	NT	NT	6 mm	NT
3	M	43	194	NT	1.41	0.35	3 mm	10 mm	10 mm	NT
4	F	38	28	3.75	9.82	2.66	NT	NT	0 mm	NT
5	M	38	183	1.04	6.84	31.70	NT	NT	NT	NT

Fig E1. Workflow for identification of the potential cross-reactive allergens from Pacific Oyster (*C. gigas*) by a combination of bioinformatics, allergenomic and proteomic approaches. (A) *In silico* identification based on sequence alignment was conducted to determine Pacific Oyster proteins which have amino acid sequence identity >50% of the known allergens and an E-value <10⁻⁷. (B) 2D-immunoblotting against a serum pool from 6 mollusc allergic patients was performed and the IgE-reactive proteins were identified using LC-MS/MS. (C) Shotgun proteomics was used to identify total proteins in the raw and heated extracts of the Pacific Oyster.

Fig E2. Expression profiles of genes encoding 'likely allergens' across different development stages and different tissues of the Pacific Oyster. The transcriptome expression data were obtained from GigaDB database under dataset number 100030. Gene expression levels were measured by RPKM (Reads Per Kilobase per Million mapped reads).

Fig E3. Coomassie-stained 2D-PAGE of the proteins from the raw and heated extracts of the Pacific Oyster. Spot numbers indicated on the gel correspond to IgE-reactive spots (Supplementary Table 2 & Supplementary Table 3) that were subjected to LC/MS mass spectrometry analysis.



Supplementary Figure 1

