Allergy Asthma Immunol Res. 2025 Jan;17(1):94-110
https://doi.org/10.4168/aair.2025.17.1.94
pISSN 2092-7355·eISSN 2092-7363

Allergy, Asthma & Immunology Research AAIR

Original Article

Check for updates

Genome-Wide Identification and Comparative Analysis of Allergens in *Procambarus clarkii*

Judy Kin Wing NG,[1†] **Soo Kyung Shin**,[1†] **Xiaojun Xiao**,[2†] **Qing Xiong**,[1,3] **Hui Cao**,[2] **Ruyi Yuan**,[2] **Baoqing Sun**,[4] **Xiaoyu Liu**,[2*] **Stephen Kwok-Wing Tsui** [1,3,5*]

[1]School of Biomedical Sciences, The Chinese University of Hong Kong, Sha Tin, Hong Kong
[2]Shenzhen Key Laboratory of Allergy and Immunology, School of Medicine, Shenzhen University, Shenzhen, China
[3]Hong Kong Bioinformatics Centre, The Chinese University of Hong Kong, Sha Tin, Hong Kong
[4]State Key Laboratory of Respiratory Disease, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China
[5]Centre for Microbial Genomics and Proteomics, The Chinese University of Hong Kong, Sha Tin, Hong Kong

OPEN ACCESS

Received: Mar 16, 2024
Revised: Aug 5, 2024
Accepted: Aug 7, 2024
Published online: Oct 24, 2024

Correspondence to
Stephen Kwok-Wing Tsui, PhD
School of Biomedical Sciences, Hong Kong Bioinformatics Centre, Centre for Microbial Genomics and Proteomics, The Chinese University of Hong Kong, 32 Tai Po Road, Ma Liu Shui, Sha Tin, Hong Kong.
Tel: +852-3943 6381
Fax: +852-2603-5123
Email: kwtsui@cuhk.edu.hk

Xiaoyu Liu, PhD
Shenzhen Key Laboratory of Allergy and Immunology, School of Medicine, Shenzhen University, Ke Yuan Lu, Nanshan, Shenzhen 518054, China.
Tel: +86-755-8667 1907
Fax: +86-755-8667, 1906
Email: lxy0901@szu.edu.cn

[†]Judy Kin Wing NG, Soo Kyung Shin, and Xiaojun Xiao contributed equally to this work.

ABSTRACT

Purpose: Crustacean shellfish is one of the eight most common food allergens, and crayfish is a highly valued shellfish species for consumption in China. However, the detailed allergen profile of crayfish remains unknown, with only four allergen groups reported in the WHO/IUIS allergen nomenclature database. In this study we aimed to identify novel allergens based on the *Procambarus clarkii* genome and to reveal its allergen profile for developing better diagnostic tools and treatments.
Methods: We assembled the crayfish genome using both long-read and short-read sequencing data and identified putative allergens using the BLAST algorithm based on sequence homology. We employed bioinformatics tools to investigate the expression levels, gene structure, and synteny of these putative allergens. We also applied indirect enzyme-linked immunosorbent assay by using patients' sera to determine allergenicity and utilized proteomic methods to identify novel allergens.
Results: We identified a total of 11 putative allergen groups, including all isoforms or homologs for each allergen group based on the genome and three putative allergens by using 2-dimensional (2D) mass spectrometry. We identified 2 novel allergens, pPro c 3.0301 and pPro c 6.0201, with immunoglobulin E reactivity of 33.3% and 20%, respectively.
Conclusions: By providing a comprehensive understanding of the complete allergen profile, our study presents a foundation for comprehending *P. clarkii*-associated allergy. The knowledge could facilitate the implementation of a components-resolved diagnostic test and preventive immunotherapy based on molecular allergens for crayfish allergy.

Keywords: Crayfish allergy hypersensitivity; seafood; tropomyosin; genome; allergens; shellfish

INTRODUCTION

Shellfish is one of the eight major food allergens, affecting over 2% of the general population and the sensitization usually persist in adult.[1,2] Shellfish allergy is a global issue, especially

ORCID iDs

Judy Kin Wing NG

https://orcid.org/0000-0003-4228-0349

Disclosure

There are no financial or other issues that might lead to conflict of interest.

Availability of Data

The genomes have been deposited under National Center for Biotechnology Information (NCBI) with BioProject accession number PRJNA977390 and CNGB Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb) with project accession number CNP0004517.

in Asia. Previous studies have reported the prevalence of shellfish allergy is 1.3% in United States and up to 7.7% in Asian countries.[3,4] Notably, shellfish has also been found to be a major cause of food-induced anaphylaxis which could be life-threatening.[5-7] Among various shellfish species, edible shellfish usually refers to crustacean and mollusk belonging the phylum of Arthropoda and Mollusca respectively. Crustaceans are more allergenic than mollusks, with most patients sensitized to prawn and crab.[3,8] Moreover, tropomyosin has been identified as the major allergen in 20 commonly consumed shellfish and is highly cross-reactive within arthropods.[9]

With the advancement of freshwater aquaculture, the demand for and supply of freshwater shellfish have increased significantly in recent years. Red swamp crayfish (*Procambarus clarkii*) is a famous seafood and one of the most widely cultivated species in China.[10] However, despite the increasing production and consumption rate, much less is known about its allergen composition compared to that of prawn (*Penaeus monodon*), as reported in the WHO/IUIS nomenclature database. Currently, four allergen groups have been reported including tropomyosin, arginine kinase, myosin light chain 1, and triosephosphate isomerase.[11,12] A case report revealed an incident of non-occupational protein contact dermatitis after ingestion of crayfish, demonstrating that *P. clarkii* is associated with allergic disorders and has a serious impact on people with food allergies.[13]

To cope with the increased allergic reactions caused by crayfish, an in-depth investigation of its allergen profile is required. Multi-omics approaches in previous studies on cockroach, house dust mite, and storage mite have resolved the complex allergen profiles and identified novel allergens for component-resolved diagnosis.[14-17] In this study, we aimed to provide a foundation for translational research, such as producing recombinant allergens for crayfish allergy diagnosis, and shed light on the development of tailored treatment for *P. clarkii*-sensitive patients. The insights gained from the study will not be limited to *P. clarkii* only but will be applicable to other types of shellfish which usually show high cross-reactivity.

MATERIALS AND METHODS

Genomic DNA extraction

Qiagen Puregene Core Kit A was used for extracting genomic DNA of *P. clarkii*, following the protocol "DNA Purification from Tissue Using the Gentra Puregene Tissue Kit." Then ethanol precipitation was performed, followed by DNA purification with Wizard® SV Genomic DNA Purification System (Promega) according to the manufacturer's instructions. Genomic DNA concentration was measured by using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA) and Qubit Fluorometer (Thermo Fisher Scientific Inc.), and the integrity was determined by using gel electrophoresis in 0.5% agarose gel. Then genomic DNA was sent to Berry Genomics (Beijing, China) to generate PacBio long-read sequencing reads.

Genome assembly and annotation

PacBio reads were used for *de novo* genome assembly with Flye v2.8.2.[18] Then scaffolding was performed by SSPACE-BASIC-2.0 and SSPACE-LongRead.[19] Next, gap filling was performed by using GapFiller with paired-end short reads. Then the gapfilled genome was polished by using Pilon, and the scaffolds with length less than 5,000 bp were discarded to get the final genome assembly.[20] The completeness of the genome assembly was assessed by using BUSCO

https://doi.org/10.4168/aair.2025.17.1.94

v3.1.0.[21] The final genome assembly was subjected to repeat masking with RepeatMasker v4.1.0.[22] The repeat-masked genome was used in the Maker pipeline with protein homology evidence from closely related species.[23] The completeness of annotated proteins was also assessed by using BUSCO v3.1.0 with arthropoda_odb9 in protein mode. The genomes have been deposited in the National Center for Biotechnology Information (NCBI) with BioProject accession number PRJNA977390 and in the CNGB Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb) with project accession number CNP0004517.

Tissue RNA extraction and sequencing

The crayfish tail muscle, hepatopancreas, antennal gland, ovary, and gut tissues were kept in TRIzol for RNA extraction by using the PureLink RNA Mini Kit (Thermo Fisher Scientific Inc.) according to the manufacturer's instructions. RNA concentration of the samples was measured by using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific Inc.) and the Qubit Fluorometer (Thermo Fisher Scientific Inc.). Two replicates from each tissue type were sent to Groken Bioscience for RNA sequencing using Illumina HiSeq 2500.

In silico identification of allergens

The putative allergens in *P. clarkii* were identified by inferring homology from known allergens of closely related species in the order Decapoda. The respective gene sequences were extracted from the genome by TBLASTN search using the reported protein sequences that available on the webpage of WHO/IUIS allergen nomenclature database http://allergen.org/ as the query. An E-value cutoff of 1e-6 was applied in the algorithm. The sequence homology of the putative allergen to the reported allergen was identified by BLASTP algorithm and to determine the percentage identity. The genomic data was visualized through Integrative Genomics Viewer (IGV) for observing alternative splicing in selected genes.[24]

Transcriptome analysis and expression profile

Transcriptome data was processed using both reference-mapping and *de novo* assembly methods with Hisat2 v2.0.4 and Trinity v2.8.4, respectively.[25,26] The output from Hisat2 was converted to FASTA format by using StringTie v1.3.0 and gffread v0.12.1.[27,28] The resulting transcriptome data was then combined and prepared for gene annotation. To analyze allergen expressions in various organs, pair-end clean reads from each organ were processed using Salmon v0.13.0 and the expression level of each gene was measured in transcripts per million (TPM).[29]

Cloning and indirect enzyme-linked immunosorbent assay (ELISA) of recombinant protein

The putative allergen coding DNA sequences (CDSs) were obtained from the genome and the synthesized CDSs were inserted into the pET-30a+ vector for expression in TOP10 *Escherichia coli*. The recombinant proteins were expressed by Sangon Biotech (Shanghai, China). To test the immunoglobulin (Ig)E reactivity of the recombinant proteins, an indirect ELISA was performed by using serum samples from 30 crayfish-reactivity patients and 5 healthy controls. Purified proteins (5 μg/mL) were coated onto a 96-well microtiter plate in a sodium bicarbonate coating buffer (100 mM, pH 9.6) and incubated at 37°C for 3 hours. The plate was washed three times with 0.05% Tween-20/PBS (PBST) and blocked with 8% fetal bovine serum (FBS, Gibco, Thermo Fisher Scientific Inc.) in PBS at room temperature for 2 hours. Serum samples were diluted 1:5 with blocking buffer and 50 μL of diluted serum was added to each well overnight at 4°C. After washing, horseradish peroxidase conjugated anti-human IgE antibodies (Thermo Fisher Scientific Inc.) were added at a 1:1,000 dilution and incubated at

https://doi.org/10.4168/aair.2025.17.1.94

room temperature for 1 hour. The plate was then washed five times with PBST and incubated with TMB-ELISA substrate (Abcam, Cambridge, UK) for color development. The reaction was stopped by adding 0.1M sulfuric acid and the absorbance was measured at 450 nm using a microplate reader (Bio-Rad, Hercules, CA, USA).

Protein extraction and 2D gel electrophoresis please refer to **Supplementary Data S1**.

Ethics approval

This study was approved by the Ethics Committee of the First Affiliated Hospital of Guangzhou Medical University with ethics number GYFYY-2022-135.

RESULTS

Genome-wide identification of putative allergen in P. clarkii

The assembly and annotation of the *P. clarkii* genome represents a crucial step in advancing our understanding of this species. It enables us to investigate allergen genes at a multi-omics scale. We successfully generated a complete genome of *P. clarkii*, measuring 3.2 Gb in size, with 93.4% BUSCO completeness (**Supplementary Fig. S1**). To identify allergen genes, we utilized the reported *P. clarkii* sequence and other decapod allergens from the WHO/IUIS allergen nomenclature database as reference genes. We identified eleven distinct allergen groups based on the genome sequence (**Table 1**). A collection of transcript and protein sequences of the listed putative allergens were retrieved from the genome by manual curation, and the data were subjected to downstream analysis.

To assign allergen and isoallergen IDs to the putative allergen groups and individual isoallergens, we utilized a standardized format as presented in the WHO/IUIS database. Specifically, the IDs for putative allergens were denoted by the letter "p" followed by the species abbreviation and the corresponding group number. Under group 1, 2, 5, and 8, the first isoallergen was considered the reported *P. clarkii* allergen, exhibiting the highest sequence identity. Putative allergens in other allergen groups were classified based on their sequence identity of over 30% to the reference genes from closely related species.[17] By analyzing the sequence homology to the reported allergens, we successfully identified all homologs and isoforms within each allergen group, along with their corresponding gene loci within the genome. Remarkably, we observed the presence of multiple homologs in the majority of the allergen groups, with groups 6 and 7 showing extensive tandem duplication. Notably, despite only identifying one gene locus in group 1, we discovered a striking combination of 9 exons which could potentially generate 192 isoforms within the tropomyosin gene, assuming no exon skipping occurred. This finding prompted us to conduct further in-depth investigations to gain a clearer understanding of the allergen profile in *P. clarkii*.

Comparative analysis of tropomyosin in crustaceans

To explore the diversification of tropomyosin in decapods, we identified the tropomyosin gene in seven crustacean species and examined their gene structures (**Fig. 1**). In *P. clarkii*, we observed alternative splicing events at exons 2, 4, 5, 7, and 9, accompanied by a lengthy intron spanning over 30 kb between exons 2 and 3 (**Fig. 1A**). A similar gene structure was noted in the genomes of other decapods, such as prawn (*P. monodon*), shrimp (*Macrobrachium nipponense*), and American lobster (*Homarus americanus*). However, *Panulirus ornatus* exhibited the deletion of the second exon, resulting in a gene with only eight exons. This configuration

https://doi.org/10.4168/aair.2025.17.1.94

Table 1. Overview of all in silico predicted allergen genes in *P. clarkii*

Allergen ID	Isoallergen ID	Gene locus	Biochemical function	No. of exons	No. of amino acid	Homolog identity
Pro c 1	Pro c 1.0101*	CF_00019402-RA	Tropomyosin	9	284	ACN87223.1 Pro c 1 (98.6%)
Pro c 2	Pro c 2.0101*	CF_00036538-RA	Arginine kinase	2	357	H6VGI2 Pro c 2 (100%)
	pPro c 2.0201	CF_00111069-RA	Arginine kinase	2	363	H6VGI2 Pro c 2 (54.8%)
pPro c 3	pPro c 3.0101	CF_00066465-RA	Myosin, light chain 2	4	178	ADV17342.1 Pen m 3 (93.2%)
	pPro c 3.0201	CF_00013960-RA	Myosin, light chain 2	4	213	ADV17342.1 Pen m 3 (68.3%)
	pPro c 3.0301	CF_00002567-RA	Myosin, light chain 2	4	175	ADV17342.1 Pen m 3 (65.2%)
	pPro c 3.0401	CF_00010548-RA	Myosin, light chain 2	4	172	ADV17342.1 Pen m 3 (37.1%)
pPro c 4	pPro c 4.0101	CF_00052896-RA	Sarcoplasmic calcium-binding protein	5	193	ACR43475.1 Pen m 4 (80.8%)
	pPro c 4.0102	CF_00052896-RA	Sarcoplasmic calcium-binding protein	5	193	ACR43475.1 Pen m 4 (83.9%)
	pPro c 4.0103	CF_00052896-RA	Sarcoplasmic calcium-binding protein	5	193	ACR43475.1 Pen m 4 (82.4%)
Pro c 5	Pro c 5.0101*	CF_00017644-RA	Myosin light chain 1	5	153	AFP95338.1 Pro c 5 (100%)
pPro c 6	pPro c 6.0101	CF_00028022-RA	Troponin C	6	150	ADV17344.1 Pen m 6 (83.3%)
	pPro c 6.0201	CF_00028023-RA	Troponin C	6	150	ADV17344.1 Pen m 6 (71.3%)
	pPro c 6.0301	CF_00028020-RA	Troponin C	6	155	ADV17344.1 Pen m 6 (75.3%)
	pPro c 6.0401	CF_00028025-RA	Troponin C	6	150	ADV17344.1 Pen m 6 (82.0%)
	pPro c 6.0501	CF_00028024-RA	Troponin C	6	150	ADV17344.1 Pen m 6 (83.3%)
	pPro c 6.0601	CF_00028021-RA	Troponin C	6	150	ADV17344.1 Pen m 6 (81.3%)
	pPro c 6.0701	CF_00010560-RA	Troponin C	6	155	ADV17344.1 Pen m 6 (59.8%)
	pPro c 6.0801	CF_00045349-RA	Troponin C	6	149	ADV17344.1 Pen m 6 (45.9%)
	pPro c 6.0901	CF_00072387-RA	Troponin C	6	140	ADV17344.1 Pen m 6 (50.3%)
	pPro c 6.1001	CF_00017925-RA	Troponin C	6	152	ADV17344.1 Pen m 6 (51.0%)
	pPro c 6.1101	CF_00045350-RA	Troponin C	6	152	ADV17344.1 Pen m 6 (45.4%)
pPro c 7	pPro c 7.0101	CF_00078879-RA	Hemocyanin	2	493	AEB77775.1 Pen m 7 (72.7%)
	pPro c 7.0201	CF_00078879-RA	Hemocyanin	2	493	AEB77775.1 Pen m 7 (72.2%)
	pPro c 7.0301	CF_00078880-RA	Hemocyanin	2	567	AEB77775.1 Pen m 7 (74.7%)
	pPro c 7.0401	CF_00019522-RA	Hemocyanin	2	570	AEB77775.1 Pen m 7 (69.0%)
	pPro c 7.0501	CF_00019523-RA	Hemocyanin	2	597	AEB77775.1 Pen m 7 (68.9%)
	pPro c 7.0601	CF_00019524-RA	Hemocyanin	2	545	AEB77775.1 Pen m 7 (68.0%)
	pPro c 7.0701	CF_00019525-RA	Hemocyanin	2	496	AEB77775.1 Pen m 7 (67.5%)

(continued to the next page)

https://doi.org/10.4168/aair.2025.17.1.94

Table 1. (Continued) Overview of all in silico predicted allergen genes in *P. clarkii*

Allergen ID	Isoallergen ID	Gene locus	Biochemical function	No. of exons	No. of amino acid	Homolog identity
	pPro c 7.0801	CF_00019526-RA	Hemocyanin	2	496	AEB77775.1 Pen m 7 (66.3%)
	pPro c 7.0901	CF_00019529-RA	Hemocyanin	2	499	AEB77775.1 Pen m 7 (65.7%)
	pPro c 7.1001	CF_00048326-RA	Hemocyanin	2	492	AEB77775.1 Pen m 7 (74.5%)
	pPro c 7.1101	CF_00048327-RA	Hemocyanin	2	566	AEB77775.1 Pen m 7 (76.0%)
	pPro c 7.1201	CF_00048328-RA	Hemocyanin	2	492	AEB77775.1 Pen m 7 (74.5%)
	pPro c 7.1301	CF_00048329-RA	Hemocyanin	2	492	AEB77775.1 Pen m 7 (74.9%)
	pPro c 7.1401	CF_00048330-RA	Hemocyanin	2	492	AEB77775.1 Pen m 7 (71.7%)
	pPro c 7.1501	CF_00048331-RA	Hemocyanin	2	567	AEB77775.1 Pen m 7 (74.3%)
	pPro c 7.1601	CF_00014449-RA	Hemocyanin	2	600	AEB77775.1 Pen m 7 (59.3%)
	pPro c 7.1701	CF_00014450-RA	Hemocyanin	2	573	AEB77775.1 Pen m 7 (61.3%)
	pPro c 7.1801	CF_00014451-RA	Hemocyanin	2	573	AEB77775.1 Pen m 7 (61.0%)
	pPro c 7.1901	CF_00014452-RA	Hemocyanin	2	573	AEB77775.1 Pen m 7 (61.4%)
	pPro c 7.2001	CF_00004013-RA	Hemocyanin	2	661	AEB77775.1 Pen m 7 (57.8%)
	pPro c 7.2101	CF_00004013-RA	Hemocyanin	2	660	AEB77775.1 Pen m 7 (57.5%)
	pPro c 7.2201	CF_00004013-RA	Hemocyanin	2	660	AEB77775.1 Pen m 7 (57.9%)
	pPro c 7.2301	CF_00004013-RA	Hemocyanin	2	660	AEB77775.1 Pen m 7 (57.3%)
	pPro c 7.2401	CF_00004013-RA	Hemocyanin	2	660	AEB77775.1 Pen m 7 (57.5%)
	pPro c 7.2501	CF_00004013-RA	Hemocyanin	2	660	AEB77775.1 Pen m 7 (57.5%)
Pro c 8	Pro c 8.0101*	CF_00024098-RA	Triosephosphate isomerase	4	248	AEB54655.1 Pro c 8 (100%)
	pPro c 8.0201	CF_00008401-RA	Triosephosphate isomerase	4	270	AEB54655.1 Pro c 8 (64.1%)
	pPro c 8.0301	CF_00008403-RA	Triosephosphate isomerase	4	269	AEB54655.1 Pro c 8 (64.1%)
	pPro c 8.0401	CF_00008402-RA	Triosephosphate isomerase	4	292	AEB54655.1 Pro c 8 (61.0%)
pPro c 9	pPro c 9.0101	CF_00080058-RA	Filamin C	17	977	QFI57017.1 Scy p 9 (91.5%)
	pPro c 9.0201	CF_00024447-RA	Filamin C	10	660	QFI57017.1 Scy p 9 (31.15%)
	pPro c 9.0301	CF_00001639-RA	Filamin C	28	1,809	QFI57017.1 Scy p 9 (26.7%)
pPro c 10	pPro c 10.0101	CF_00015035-RA	Troponin I	11	225	P05547 Pon l 7 (86.5%)
pPro c 13	pPro c 13.0101	CF_00031936-RA	Fatty Acid Binding Protein	4	136	AEP84100.1 Pen m 13 (82.4%)
	pPro c 13.0201	CF_00009656-RA	Fatty Acid Binding Protein	4	93	AEP84100.1 Pen m 13 (75.0%)
	pPro c 13.0301	CF_00036676-RA	Fatty Acid Binding Protein	1	93	AEP84100.1 Pen m 13 (52.2%)

*Reported allergens in *P. clarkia*.

https://doi.org/10.4168/aair.2025.17.1.94

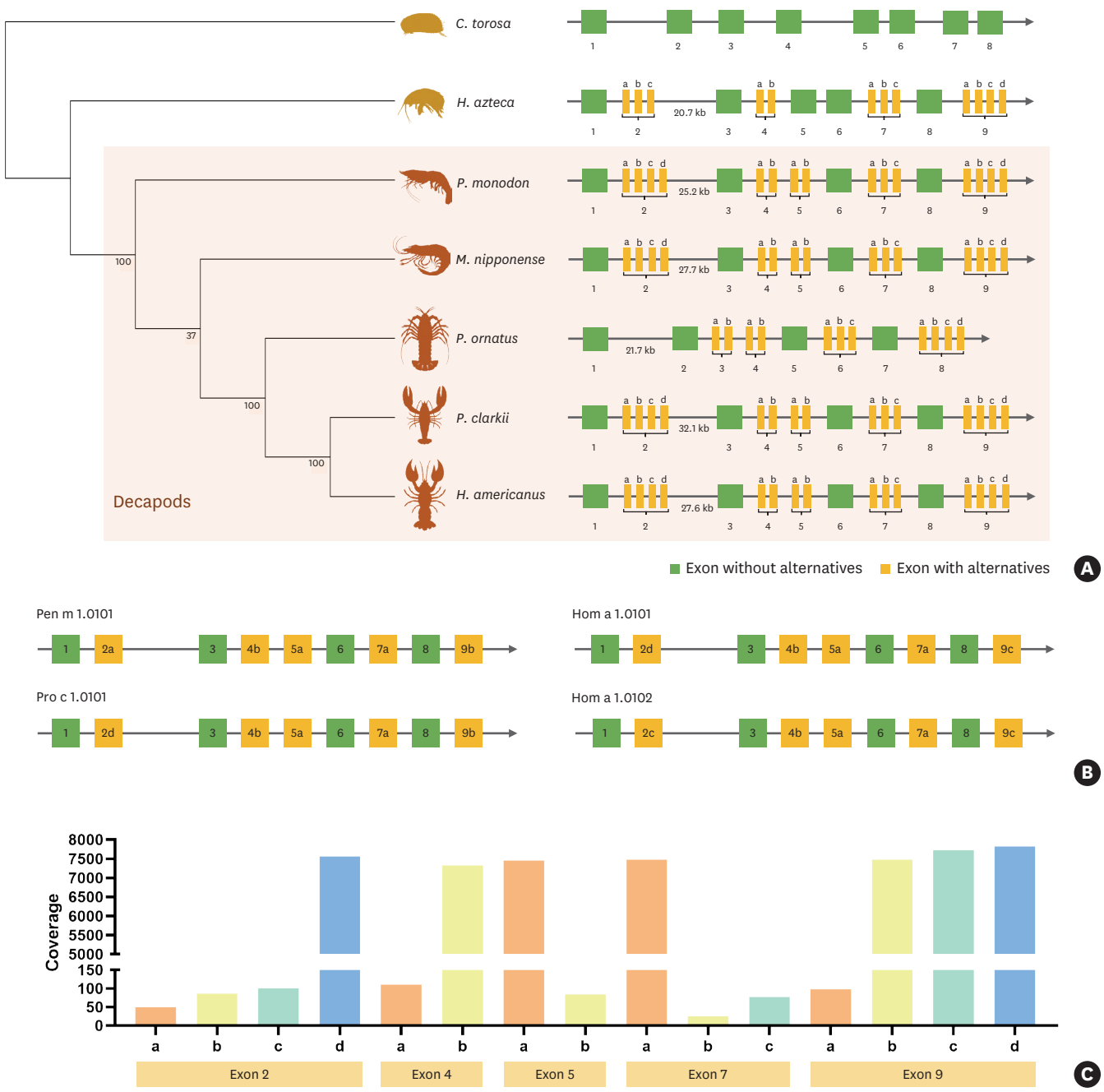


Fig. 1. Comparative analysis of tropomyosin in Decapoda. (A) The tree was constructed based on COX1 gene sequences of seven species by MEGA-X. All species selected are in the subphylum Crustacea, with 5 species of *P. monodon*, *M. nipponense*, *P. ornatus*, *P. clarkii*, and *H. americanus* are in the order of Decapoda. The gene structure of tropomyosin in each species is illustrated in which exons are presented as colored blocks, and the arrow indicate the transcription direction. The silhouettes of organisms are not drawn in scale. (B) Gene structures of tropomyosin in *H. americanus* (Hom a 1.0101, Hom a 1.0102), *P. monodon* (Pen m 1.0101) and *P. clarkii* (Pro c 1.0101). The exon order is labelled in numbers and choice of exon with alternatives are denoted by letter. (C) Bar chart illustrate the average alignment coverage of each exon sequence with alternative splicing in *P. clarkii*. The coverage is generated by transcriptome data mapping to genome sequence and correlated to the proportion of sequence transcribed. The coverage level of exon 9b, 9c and 9d cannot be differentiated due to partially overlapping of sequences.

potentially generated 48 isoforms due to the various combinations of exons. In *Hyalella azteca*, an amphipod closely related to decapods, the tropomyosin gene also comprised four exons with alternative splicing, leading to the formation of 72 potential isoforms. In contrast, *Cyprideis torosa*, a more distantly related decapod, did not display such complex

https://doi.org/10.4168/aair.2025.17.1.94

gene structures or alternative splicing. Notably, all the species examined possessed a single tropomyosin gene within their entire genomes, regardless of genome size. Rather than gene duplication giving rise to homologs, the subfunctionalization of tropomyosin was achieved through increased gene complexity and the generation of isoforms. These findings suggest that the diversification of tropomyosin through alternative splicing emerged later in crustaceans and primarily exists within decapods.

Among the various tropomyosin isoforms, our investigation focused on identifying the isoform corresponding to the reported allergenic form in the WHO/IUIS database. We illustrated the exon combination of Pen m 1.0101, Pro c 1.0101, Hom a 1.0101, and Hom a 1.0102 (**Fig. 1B**). Interestingly, we observed similarities in the choice of exons among these four allergens from different species. Specifically, the allergen genes consisted of exon 4b, 5a, and 7a, while the selection of exon 2 and 9 varied across species. To delve deeper, we conducted sequence alignment to explore the amino acid similarity in exons 4, 5, and 7. Remarkably, we found that the amino acid sequences encoded by same exons in the three species were either identical or highly similar (**Supplementary Fig. S2**). For instance, we observed that the amino acid sequence encoded by exon 4a in *P. clarkii* showed a higher degree of similarity to the corresponding exon 4a in two other species, rather than to exon 4b within *P. clarkii* itself. This suggests that the exon combination of each isoform could potentially impact the allergenicity of tropomyosin. Furthermore, we observed that Pro c 1.0101, characterized by the exon combination 2d-4b-5a-7a-9b, exhibited the highest exon coverage (**Fig. 1C**). The exon coverage analysis indicates that the expression of Pro c 1.0101 is most abundant in muscle tissues.

In addition, we conducted multiple sequence alignments to investigate the sequence similarity among the aforementioned reported allergic tropomyosins (**Fig. 2**). Although the epitope of *P. clarkii* tropomyosin has yet to be experimentally validated, we could infer potential epitope regions based on the high sequence similarities observed with other crustacean tropomyosins.[9] We compared the sequence identities in these regions across different species with the IgE-binding epitope of tropomyosin reported in a freshwater prawn, *Exopalaemon modestus* (Exo m 1.0101).[30] The IgE-binding epitopes, which exhibited a high degree of conservation, were found in regions translated by exon 2, 3, 4, 6, 8, and 9 among the species analyzed. Conversely, sequence variations were observed at the N-terminal of the protein, specifically in regions translated by exon 1.

Gene synteny of putative group 6 and 7 allergens

Group 6 and 7 allergens in decapods are represented by troponin C and hemocyanin, respectively. In our investigation, we also explored the putative allergen genes related to these proteins within the *P. clarkii* genome. Troponin C, a calcium-binding protein with four EF-hand motifs, is found in striated muscle tissue and has been reported as an allergen in mites, cockroaches, and prawns.[31-33] Troponin C genes in *P. clarkii* were identified by sequence homology with reference to Pen m 6 (GenBank accession: ADV17344.1), and we explored all the homologs by manual curation. There were 11 troponin C homologs of which pPro c 6.0101 and pPro c 6.0501 shared the highest sequence identity (83.3%) with Pen m 6 (**Table 1**). Interestingly, 6 out of the 7 troponin C genes were found to be clustered in the genome. We illustrated the synteny of these tandem arrayed genes based on their locus (**Fig. 3A**). Notably, pPro c 6.0101 and pPro c 6.0501 were tandem arrayed genes, suggesting a possible formation through gene duplication. Furthermore, the genes within the cluster exhibited a higher similarity to Pen m 6 compared to other troponin C homologs.

https://doi.org/10.4168/aair.2025.17.1.94

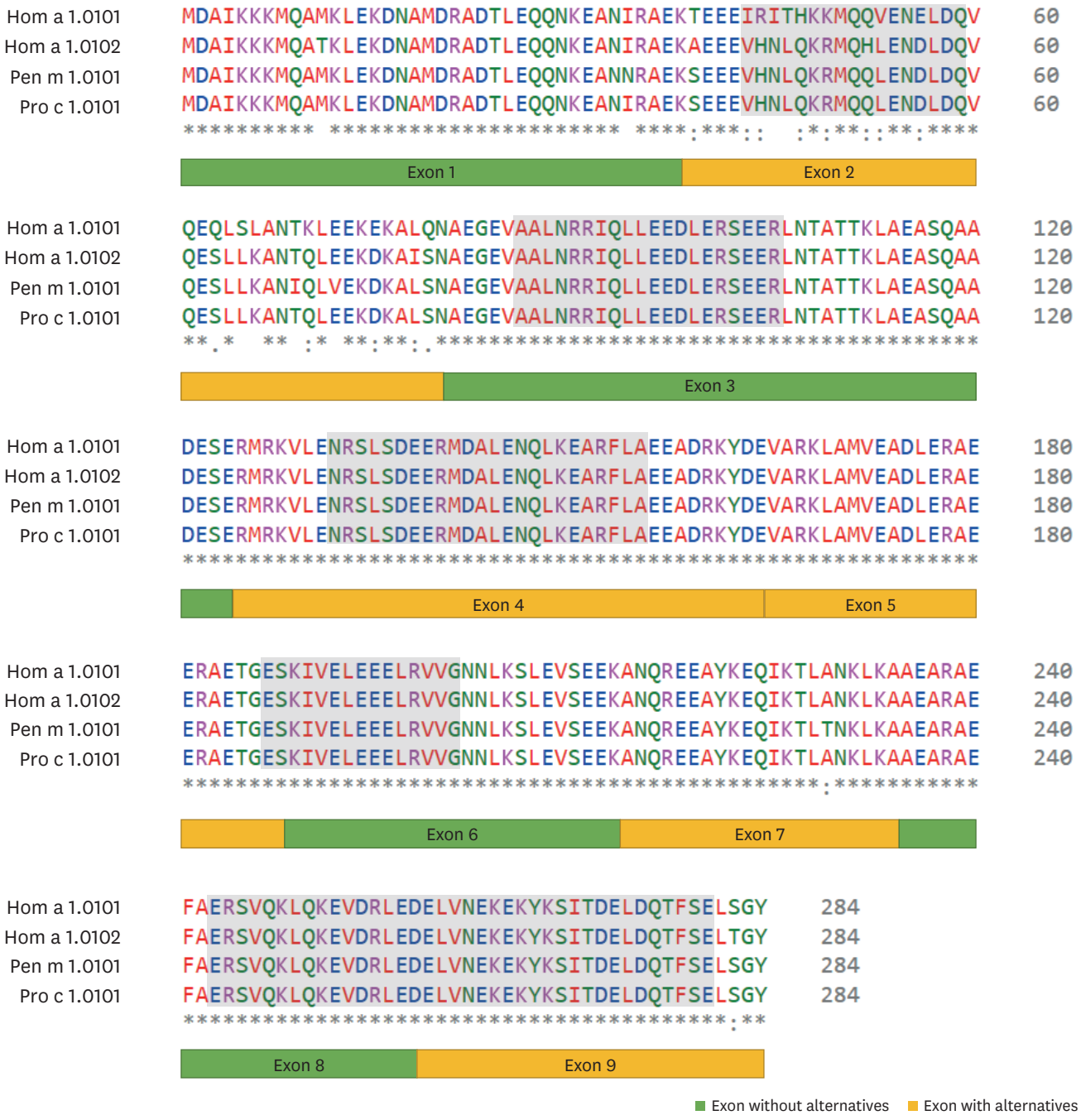
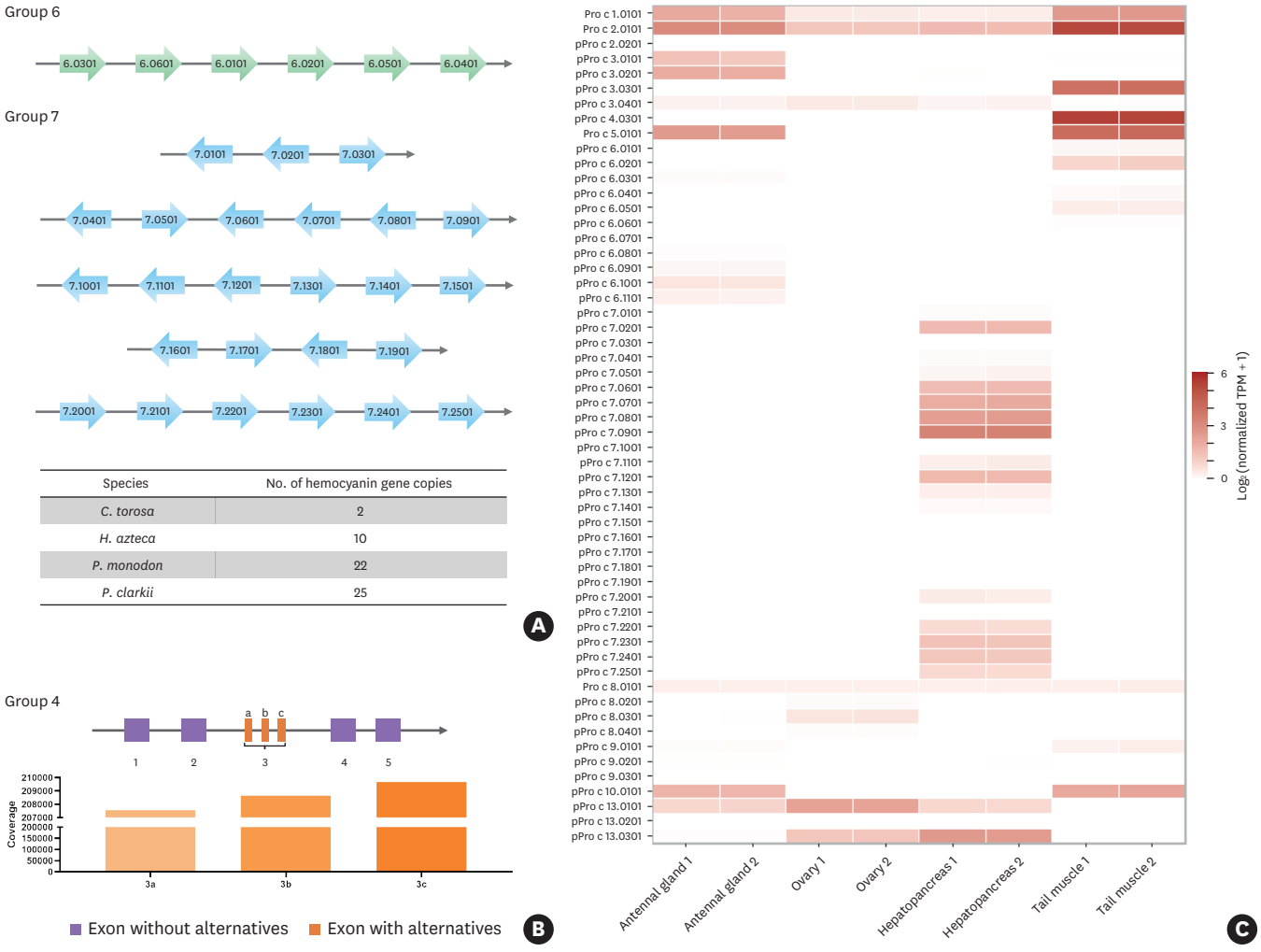


Fig. 2. Protein sequence alignment of tropomyosin in 3 Decapods. The protein sequences of reported allergenic tropomyosin were obtained from WHO/IUIS nomenclature database. Tropomyosin from *H. americanus* (Hom a 1.0101, Hom a 1.0102), *P. monodon* (Pen m 1.0101) and *P. clarkii* (Pro c 1.0101) were aligned by Clustal Omega, and the corresponding exons are marked in parallel to the amino acid sequences. The respective regions of immunoglobulin E epitopes in *E. modestus* (Exo m 1.0101) are shaded in grey and located in regions translated by exon 2, 3, 4, 6, 8 and 9. The sequences are highly conserved, with variations among species occurring at the N-terminal of tropomyosin.

Hemocyanin is a copper containing protein that is responsible for oxygen transport in invertebrates.[34] It has previously been identified as a group 7 allergen in seafood, such as shrimp, squid, and crab.[35-37] We identified 25 homologs of hemocyanin in *P. clarkia,* and these genes were organized into 5 clusters (**Fig. 3A**). By comparing the gene copy number across different crustaceans, we found that *C. torosa* had 2 copies of hemocyanin, while *H. azteca*

https://doi.org/10.4168/aair.2025.17.1.94



Species	No. of hemocyanin gene copies
C. torosa	2
H. azteca	10
P. monodon	22
P. clarkii	25

Fig. 3. Allergen gene expression levels and tandem gene synteny alignment. (A) Gene synteny of tandemly arrayed allergen genes. The position of each gene is illustrated in colored arrows, and the direction of gene transcription is presented by the direction of arrows. All the distances between neighboring genes are below 5 Kb. (B) Alignment coverage of each exon with alternative splicing in group 4 allergen. The abundance of sequences is reflected by the reads coverage and the bar chart illustrate the average per-base coverage of the whole exon sequence. (C) The gene expression levels of all identified putative allergens in *P. clarkia* (**Table 1**) in 4 with 2 sets of transcriptome data. The expression level was calculated by Log2 (normalized TPM + 1) and TPM was normalized by the TPM level of GAPDH.
TPM, transcripts per million.

had 10 copies. Interestingly, we observed a significant expansion of hemocyanin genes in *P. monodon* and *P. clarkii*, with a duplication in the gene number compared to *H. azteca*. This result suggests that gene expansion through tandem duplication may contribute to the emergence of allergenic hemocyanin in decapods.

Expression profile of curated allergens

We investigated the gene expression levels of all putative allergen groups in 4 different tissues, including tail muscle, hepatopancreas, ovary, and antennal gland (**Fig. 3C**). For group 1, the isoform with highest exon coverage in muscle was selected to calculate the expression profile (**Figs. 1C** and **3B**). Pro c 1.0101 was also found to be expressed in other tissues. Arginine kinase (group 2) has previously been identified as a major allergen in shellfish and *P. clarkii*.[11] In the genome, we identified 2 homologs of arginine kinase with Pro c 2.0101 showing expression in all tissues, particularly in muscle. Group 3 contained 4 homologs, but muscle

https://doi.org/10.4168/aair.2025.17.1.94

tissue exclusively utilized pPro c 3.0301, with the other homologs not exppressed. This results also unveiled the subfunctionalization and tissue specificity of homologs in terms of their expression levels. While in group 4, 1 of 3 isoforms with highest exon coverage was also selected, and as expected, gly pPro c 4.0103 was highly expressed in tail muscle. Additionally, known allergens in *P. clarkii* included myosin light chain 1 (group 5) and triosephosphate isomerase (group 8).[38,39] Although only 1 gene was found for group 5, its expression was notably high in muscle and antennal gland. In contrast, group 8 consisted of 4 homologs, but their expression levels were relatively low in those tissues.

Immunoassay of recombinant putative allergens

Based on the previously described expression profiles, we selected the homolog or isoform with the highest expression in muscle for allergenicity testing. The coding gene sequences retrieved from the genome were used to clone the recombinant putative allergens. Allergenicity was assessed based on IgE reactivity using sera from patients (**Table 2**). Serum samples from patients with a positive reaction to prawn were collected and screened using total protein extracts from *P. clarkii*. The 30 serum samples that tested positive for *P. clarkii* crude protein (**Supplementary Table S1**, **Fig. 4F**) were utilized to test the recombinant allergens.

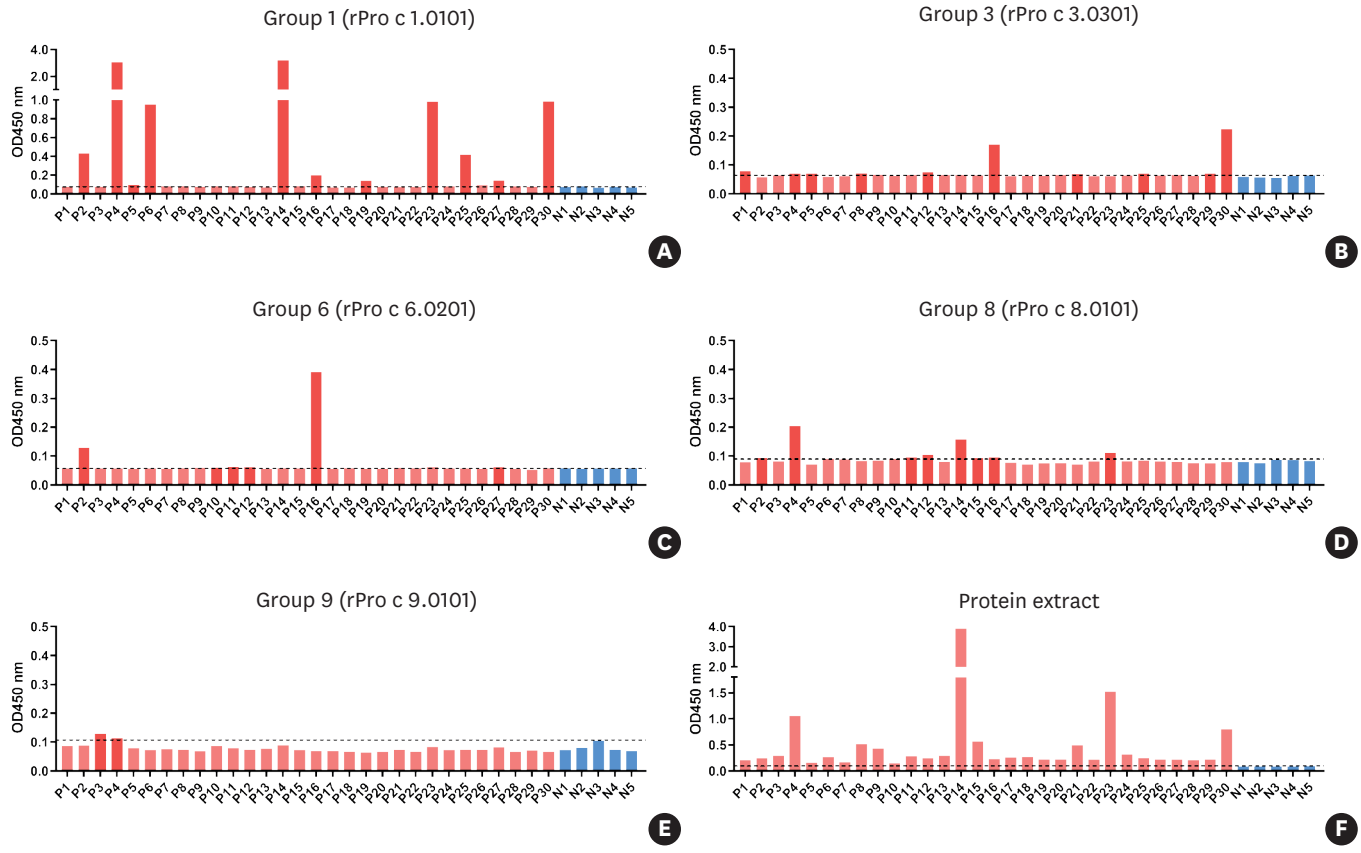
We confirmed the allergenicity of group 1, 3, 6, 8 and 9 through indirect ELISA (**Fig. 4A-E**). Our results were consistent with previous findings, indicating that tropomyosin is the major allergen in *P. clarkii.* Among the tested groups, rPro c 1.0101 produced the strongest positive signals and highest IgE reactivity (36.7%). While most of the *P. clarkii* allergy patients tested positive for group 1, our results also revealed that patients could exhibit allergic reactions to different proteins within *P. clarkii*. Notably, P16 exhibited high IgE reactivity to group 3 (**Fig. 4B**) and group 6 (**Fig. 4C**) rather than group 1. While at least 5 patients are needed to be enlisted in the IUIS allergen database, only 2 out of 30 individuals (6.67%) were positive to rPro c 9.010. However, P3 only showed a positive reaction to group 9 (**Fig. 4E**).

Table 2. MALDI-TOF mass spectrometry results of IgE-bound *P. clarkii* proteins

ID on 2D gel	Gene ID	Accession	Allergen	Protein	Score	MW (kDa)	pI	No. of peptides	Coverage (%)
1	CF_00036538-RA	XP_045591110.1	Pro c 2.0101*	Arginine kinase	1,322	40.4	6.19	71	70
	CF_00019402-RA	ACN87223.1	Pro c 1.0101*	Tropomyosin	537	32.7	4.73	46	64
		XP_045617636.1	pPro c 1.0102	Tropomyosin isoform X19	517	32.8	4.74	47	69
2	CF_00056759-RA	XP_045608233.1	-	Nesprin-1-like	1,759	71.0	5.10	85	69
	CF_00063257-RA	XP_045626220.1	-	COP1-interactive protein 1-like	917	68.3	5.13	33	39
	CF_00046766-RA	AXR98453.1	-	heat shock protein 70 kDa cognate 3, partial	578	72.4	5.07	28	37
	CF_00004013-RA	XP_045601340.1	-	hemocyanin A chain-like	531	75.6	5.48	51	36
3	CF_00074725-RA	XP_045608729.1	-	ATP synthase subunit beta	2,218	60.1	5.29	70	69
	CF_00009041-RA	XP_045612182.1	-	Actin-3, muscle-specific-like	775	42.1	5.24	32	55
	CF_00002725-RA	XP_045616694.1	-	Phenoloxidase-activating factor 2-like	285	47.2	5.00	12	23
4	CF_00019405-RA	XP_045617639.1	pPro c 1.0103	Tropomyosin isoform X22	4,106	32.8	4.74	227	97
	CF_00019405-RA	XP_045617635.1	pPro c 1.0104	Tropomyosin isoform X18	3,486	32.8	4.72	231	96
	CF_00024116-RA	BAC77082.1	-	Glyceraldehyde-3-phosphate dehydrogenase	230	35.9	6.54	16	49
	CF_00006191-RA	QPM92670.1	-	Malate dehydrogenase, NAD-dependent	114	36.5	6.02	7	23
5	CF_00009027-RA	XP_045612187.1	-	Actin-3	380	42.1	5.24	22	49
6	CF_00036538-RA	XP_045591110.1	Pro c 2.0101*	Arginine kinase	4,409	40.4	6.19	195	79
	CF_00022008-RA	XP_045614404.1	-	Fructose-bisphosphate aldolase-like	1,838	39.6	6.61	59	85
	CF_00058899-RA	XP_045607792.1	-	Aldo-keto reductase family 1 member B1-like	305	39.8	6.52	22	41
	CF_00024116-RA	BAC77082.1	-	Glyceraldehyde-3-phosphate dehydrogenase	947	36.0	6.54	36	76

*Reported allergens in *P. clarkii*.

https://doi.org/10.4168/aair.2025.17.1.94



	No. of positive	Allergenicity (%)
Group 1	11	36.7
Group 3	10	33.3
Group 6	6	20.0
Group 8	8	26.7
Group 9	2	6.67

Crayfish positive
Positive to recombinant protein
Negative control

Fig. 4. ELISA results for recombinant proteins and crude protein extract. The allergenicity of recombinant proteins and whole protein extract of *P. clarkii* was tested by ELISA using sera samples from 30 allergy patients (**Supplementary Table S1**) and 5 healthy individuals as negative controls. The dotted lines indicate the cutoff value, mean+ 2*standard deviation of the negative controls. The allergenicity of recombinant proteins was as follows: (A) rPro c 1.0101 (36.7%), (B) rPro c 3.0301 (33.3%), (C) rPro c 6.0201 (20%), (D) rPro c 8.0101 (26.7%) and (E) rPro c 9.0101 (6.67%). Also, (F) 30/30 (100%) of the patients' serum samples showed positive responses to the crude protein extract.
ELISA, enzyme-linked immunosorbent assay.

Proteomic identification of *P. clarkii* allergens

In addition to our *in silico* prediction of putative allergens in *P. clarkii*, we have identified and characterized IgE-binding proteins in crayfish muscle protein extract by using pooled patients' sera. The proteins were separated by 2D gel electrophoresis and 6 protein spots (numbered 1-6) were targeted for analysis (**Supplementary Fig. S3**). After sequencing the peptides by MALDI-TOF mass spectrometry and identifying the protein identities by amino acid sequence mapping (**Table 2**), we discovered 3 putative allergens that had not been previously reported in decapods. These groups included the nesprin-1-like protein in spot 2, the ATP synthase subunit beta in spot 3, and actin-3 in spot 5. Nesprin-1-like is a nuclear envelope protein in muscle cells that may be an invertebrate orthologue of nesprin-1 in *P. clarkii*.[40] ATP synthase subunit beta is the catalytic site of ATP synthase in mitochondria, and actin-3 is a highly conserved, abundant protein in eukaryotes.[41,42] Additional analyses are needed to determine the IgE reactivities of the proteins.

https://doi.org/10.4168/aair.2025.17.1.94

DISCUSSION

The genomic approach allows for the accurate identification of protein homologs and isoforms within a single species. Instead of relying solely on mass spectrometry to determine peptide sequences, a genome-based allergen search can provide a more comprehensive picture of the allergen profile. In this study, we added value on *P. clarkii* allergen composition and diversification of those homologs and isoforms. The multi-omic approach allows us to accurately predict allergens by sequence homology, and their gene and protein sequences can be retrieved. We identified 14 allergen groups (including putative allergens) in *P. clarkii*, and the allergenicity of two novel allergens (group 3 and 6) was confirmed by indirect ELISA using patients' sera. We also evaluated the gene expression level along with the transcriptome data, demonstrating alternative usage and subfunctionalization of allergen genes based on their expression profiles in various tissues. However, this study is limited by its reliance on the WHO/IUIS allergen database, which may not include all the most up-to-date allergen information. Although the proteomic results identified novel putative allergens, further verification is needed to test their actual allergenicity.

Tropomyosin, a double-stranded α-helical actin-associated protein, is found in the cytoskeleton and muscles.[43] Approximately 60% of shellfish allergy patients exhibit sensitization to tropomyosin. This protein displays highly conserved sequences across animals and encompasses a wide range of isoforms.[2,43] In this study, tropomyosin was the major allergen in *P. clarkii* with allergenicity of 36.7%. Our alignment results highlight the conserved sequences corresponding to the reported epitopes, which could potentially contribute to IgE cross-reactivity among allergenic tropomyosins in various species. Based on the genome sequence and annotation, the gene structure of tropomyosin is composed of 9 exons, of which 5 exons composed of alternative splicing. Similar gene structures were observed in other decapods including prawn and American lobster, but not in ostracod (*C. torosa*). This may imply that the development of striated muscle in crustacean involved diversification and subfunctionalization of tropomyosin. Instead of generating new gene copies by duplication, deriving isoforms could be a more efficient way to form over hundreds of gene products within one gene locus. Even though gene duplication and alternative splicing could be interchangeable, such strategy could allow more combinations of spliced forms of gene with conserved function and structure.[44] Interestingly, the second intron of tropomyosin gene in decapods and amphipods were over 20 kb, and those sequences may involve binding of post-transcriptional regulatory elements or spliceosome that affected splice-site recognition of pre-mRNA.[45]

In addition to tropomyosin, arginine kinase (group 2), myosin light chain 1 (group 5), triosephosphate isomerase (group 8), which have been previously reported as allergen in WHO/IUIS allergen nomenclature database, we provided details of homologs including their gene locus. Zhang *et al.*[38] purified myosin light chain glycoprotein from *P. clarkii* muscle, and they found 2 isoforms of myosin light chain (named MLC1 and MLC2) with their epitopes. Based on the genome, we provided additional information on the number of homologs of myosin light chain 1 and myosin light chain 2. Referring to the expression level in muscle, 2 forms of myosin light chain identified by Zhang *et al.* could possibly be pPro c 3.0301 and Pro c 5.0101. We also confirmed the IgE reactivity of Pro c 3.0301 (myosin light chain 1) using recombinant proteins, and the allergenicity was 33.3%. Moreover, Yang *et al.*[39] reported the cross-reactivity of triosephosphate isomerase binding IgE with filamin C (group 9) by using surface plasmon resonance. We also confirmed the allergenicity of Proc 8.0101 and Pro c 9.0101 in *P. clarkii*, but

https://doi.org/10.4168/aair.2025.17.1.94

the pattern of IgE reactivity of individual patients were different among those groups. These findings suggest that each patient may have allergies to different components within the same food item. Furthermore, it is worth noting that some patients tested positive for total proteins in crayfish but not for individual proteins. This may be attributed to the cumulative effects of different allergens or the presence of undiscovered allergens. Our results also highlight the importance of identifying and characterizing allergenic components in food items to improve diagnosis and treatment of food allergies. Most importantly, we identified troponin C as a novel allergen in group 6 in *P. clarkii,* and one patients showed high IgE reactivity towards Pro c 6.0201 and Pro c 3.0301 instead of Pro c 1.0101.

Component-resolved diagnosis using molecular allergens enables identification of sensitization profiles and the introduction of immunotherapies and appropriate disease management targeting the causative allergens.[46] Currently, the ImmunoCAP system is available for recombinant fish parvalbumin (rCyp c 1 and rGad c 1) and prawn tropomyosin (rPen m 1).[47] An exhaustive understanding of crayfish allergens is essential to facilitate allergy diagnosis, treatment, and prevention. The identification of multiple isoforms or homologs for major allergens, such as tropomyosin and arginine kinase, suggests that a single recombinant allergen may not be sufficient to capture the diverse sensitization patterns. Thus, the allergen profile provides a basis for tailored crayfish allergy diagnostics. Panels of purified recombinant crayfish allergens could enable comprehensive patient sensitization assessment. Using these purified allergens, functional assays like the basophil activation test could offer insights into allergen biological activity and clinical relevance. Evaluating their ability to induce basophil degranulation could help prioritize the most clinically significant allergens for future diagnostics and immunotherapies.

In this study, we provided a well-defined allergen composition in *P. clarkii*, and a component-resolved diagnostic test could be developed for individual patients with crayfish allergy. We also provided a detailed expression profile for patients to avoid having tissue with high levels of allergens. Genomic information provides complete data on an organism, and recombinant allergens can be produced efficiently using gene sequences retrieved from the genome. However, our findings may not fully capture the diversity of sensitization patterns across the crayfish-allergic population due to limited sample size. Future studies could expand the patient cohort to include a larger and more diverse population of crayfish-allergic individuals. To gain a more comprehensive understanding, it would be beneficial to compare the IgE reactivity and investigate the cross-reactivity among popular crustaceans using inhibition ELISA. Additionally, detailed studies on epitope mapping would provide deeper insights to differentiate the inter-species variation of IgE reactivity. Future research could focus on developing a component-resolved diagnostic test specific to *P. clarkii* allergens, as well as exploring the feasibility of using recombinant allergens for immunotherapy. Additionally, similar studies could be conducted on other common allergenic species to improve our understanding of their allergen profile and develop tailored diagnostic and therapeutic approaches.

Our study made significant progress in understanding the allergen profile of *P. clarkii*, a species of crayfish. We identified 2 previously unknown allergens, namely pPro c 3.0301 and pPro c 6.0201, and have discovered a total of 11 putative allergen groups, which include all isoforms or homologs for each allergen group based on the genome. Additionally, we identified 3 putative allergens through 2D mass spectrometry. These findings provided a comprehensive understanding of the complete allergen profile of *P. clarkii*, which is crucial for developing components-resolved diagnostic tests and preventive immunotherapy based on molecular

https://doi.org/10.4168/aair.2025.17.1.94

allergens. By identifying the specific allergens responsible for triggering an allergic reaction, healthcare professionals can provide more targeted and effective treatment options for patients.

SUPPLEMENTARY MATERIALS

Supplementary Data S1
Supplementary Materials and Methods

Supplementary Table S1
Serum samples information

Supplementary Table S2
Enzyme-linked immunosorbent assay results of recombinant putative allergen

Supplementary Fig. S1
Genome construction of crayfish *P. clarkia*. The genomic DNA was extracted from the muscle of *P. clarkii*, and the genome was constructed by hybrid assembly approach. (A) Image of female *P. clarkii* used for genomic DNA extraction. (B) Genome statistic of the assembled and annotated genome. The genome size of *P. clarkii* was 3.2 Gb with completeness of 93.4%.

Supplementary Fig. S2
Protein sequence alignment of exon 4, 5 and 7 in tropomyosin. The protein sequences were obtained from the genome of *H. americanus*, *P. monodon* and *P. clarkii*. The respective amino acid sequences encoded by (A) exon 4, (B) exon 5 and (C) exon 6 were extracted and aligned by Clustal Omega. The label of sequence was given by the species name followed by the exon number and the alphabet denoted the alternatives in each exon as mentioned in **Fig. 1**.

Supplementary Fig. S3
2D gel electrophoresis of *P. clarkii* proteins. (A) SDS-PAGE image of Coomassie blue stained 2D gel. The result showed the successful separation of *P. clarkii* proteins. (B) Western blot image of 2D separated protein gel. The results of immunoblotting showed that there were 6 protein spots (circled in red) in the total protein of *P. clarkii* bound by the specific immunoglobulin E in the patients' sera (**Supplementary Table S2**). The corresponding protein spots was denoted by the same number in the 2D gel image.

REFERENCES

1. Loh W, Tang ML. The epidemiology of food allergy in the global context. Int J Environ Res Public Health 2018;15:2043. PUBMED | CROSSREF
2. Lopata AL, Kleine-Tebbe J, Kamath SD. Allergens and molecular diagnostics of shellfish allergy: Part 22 of the Series Molecular Allergology. Allergo J Int 2016;25:210-8. PUBMED | CROSSREF
3. Wang HT, Warren CM, Gupta RS, Davis CM. Prevalence and characteristics of shellfish allergy in the pediatric population of the United States. J Allergy Clin Immunol Pract 2020;8:1359-1370.e2. PUBMED | CROSSREF
4. Lee AJ, Thalayasingam M, Lee BW. Food allergy in Asia: how does it compare? Asia Pac Allergy 2013;3:3-14. PUBMED | CROSSREF
5. Smit DV, Cameron PA, Rainer TH. Anaphylaxis presentations to an emergency department in Hong Kong: incidence and predictors of biphasic reactions. J Emerg Med 2005;28:381-8. PUBMED | CROSSREF

6. Thong BY, Cheng YK, Leong KP, Tang CY, Chng HH. Anaphylaxis in adults referred to a clinical immunology/allergy centre in Singapore. Singapore Med J 2005;46:529-34. PUBMED
7. Lertnawapan R, Maek-a-nantawat W. Anaphylaxis and biphasic phase in Thailand: 4-year observation. Allergol Int 2011;60:283-9. PUBMED | CROSSREF
8. Lee AJ, Gerez I, Shek LP, Lee BW. Shellfish allergy--an Asia-Pacific perspective. Asian Pac J Allergy Immunol 2012;30:3-10. PUBMED
9. Faber MA, Pascal M, El Kharbouchi O, Sabato V, Hagendorens MM, Decuyper II, et al. Shellfish allergens: tropomyosin and beyond. Allergy 2017;72:842-8. PUBMED | CROSSREF
10. Hu F, Zhong H, Wu C, Wang S, Guo Z, Tao M, et al. Development of fisheries in China. Reproduction and Breeding. 2021;1:64-79. CROSSREF
11. Chen HL, Mao HY, Cao MJ, Cai QF, Su WJ, Zhang YX, et al. Purification, physicochemical and immunological characterization of arginine kinase, an allergen of crayfish (*Procambarus clarkii*). Food Chem Toxicol 2013;62:475-84. PUBMED | CROSSREF
12. Yang Y, Yan HF, Zhang YX, Chen HL, Cao MJ, Li MS, et al. Expression and epitope identification of myosin light chain isoform 1, an allergen in *Procambarus clarkii*. Food Chem 2020;317:126422. PUBMED | CROSSREF
13. Perez-Carral C, Garcia-Abujeta JL, Vidal C. Non-occupational protein contact dermatitis due to crayfish. Contact Dermat 2001;44:50-1. PUBMED | CROSSREF
14. Wang L, Xiong Q, Saelim N, Wang L, Nong W, Wan AT, et al. Genome assembly and annotation of *Periplaneta americana* reveal a comprehensive cockroach allergen profile. Allergy 2023;78:1088-103. PUBMED | CROSSREF
15. Xiong Q, Liu X, Wan AT, Malainual N, Xiao X, Cao H, et al. Genomic analysis reveals novel allergens of *Blomia tropicalis*. Allergol Int 2024;73:340-4. PUBMED | CROSSREF
16. Wan AT, Xiong Q, Xiao X, Ao KF, Jang SW, Wong BS, et al. Multi-omic analysis of *Tyrophagus putrescentiae* reveals insights into the allergen complexity of storage mites. Clin Exp Allergy 2024;54:77-81. PUBMED | CROSSREF
17. Pearson WR. An introduction to sequence similarity ("homology") searching. Curr Protoc Bioinformatics 2013;Chapter 3:3.1.1-7. PUBMED | CROSSREF
18. Freire B, Ladra S, Parama JR. Memory-efficient assembly using Flye. IEEE/ACM Trans Comput Biol Bioinformatics 2022;19:3564-77. PUBMED | CROSSREF
19. Boetzer M, Pirovano W. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinformatics 2014;15:211. PUBMED | CROSSREF
20. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 2014;9:e112963. PUBMED | CROSSREF
21. Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. Methods Mol Biol 2019;1962:227-45. PUBMED | CROSSREF
22. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics 2009;Chapter 4:4.10.1-14. PUBMED | CROSSREF
23. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 2011;12:491. PUBMED | CROSSREF
24. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol 2011;29:24-6. PUBMED | CROSSREF
25. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 2015;12:357-60. PUBMED | CROSSREF
26. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 2011;29:644-52. PUBMED | CROSSREF
27. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 2015;33:290-5. PUBMED | CROSSREF
28. Pertea G, Pertea M. GFF utilities: GffRead and GffCompare. F1000 Res 2020;9:304. PUBMED | CROSSREF
29. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 2017;14:417-9. PUBMED | CROSSREF
30. Zhang Z, Li XM, Xiao H, Nowak-Wegrzyn A, Zhou P. IgE-binding epitope mapping of tropomyosin allergen (Exo m 1) from Exopalaemon modestus, the freshwater Siberian prawn. Food Chem 2020;309:125603. PUBMED | CROSSREF

31. Jeong KY, Kim CR, Un S, Yi MH, Lee IY, Park JW, et al. Allergenicity of recombinant troponin C from *Tyrophagus putrescentiae*. Int Arch Allergy Immunol 2010;151:207-13. PUBMED | CROSSREF
32. Hindley J, Wünschmann S, Satinover SM, Woodfolk JA, Chew FT, Chapman MD, et al. Bla g 6: a troponin C allergen from Blattella germanica with IgE binding calcium dependence. J Allergy Clin Immunol 2006;117:1389-95. PUBMED | CROSSREF
33. Kalyanasundaram A, Santiago TC. Identification and characterization of new allergen troponin C (Pen m 6.0101) from Indian black tiger shrimp Penaeus monodon. Eur Food Res Technol 2015;240:509-15. CROSSREF
34. Coates CJ, Costa-Paiva EM. Multifunctional roles of hemocyanins. In: Hoeger U, Harris JR, editors. Vertebrate and invertebrate respiratory proteins, lipoproteins and other body fluid proteins. Cham: Springer International Publishing; 2020. 233-50.
35. Piboonpocanun S, Jirapongsananuruk O, Tipayanon T, Boonchoo S, Goodman RE. Identification of hemocyanin as a novel non-cross-reactive allergen from the giant freshwater shrimp *Macrobrachium rosenbergii*. Mol Nutr Food Res 2011;55:1492-8. PUBMED | CROSSREF
36. Zhang Y, Deng Y, Zhao Y. Structure-based modelling of hemocyanin allergenicity in squid and its response to high hydrostatic pressure. Sci Rep 2017;7:40021. PUBMED | CROSSREF
37. Lu J, Luan H, Wang C, Zhang L, Shi W, Xu S, et al. Molecular and allergenic properties of natural hemocyanin from Chinese mitten crab (*Eriocheir sinensis*). Food Chem 2023;424:136422. PUBMED | CROSSREF
38. Zhang YX, Chen HL, Maleki SJ, Cao MJ, Zhang LJ, Su WJ, et al. Purification, characterization, and analysis of the allergenic properties of myosin light chain in *Procambarus clarkii*. J Agric Food Chem 2015;63:6271-82. PUBMED | CROSSREF
39. Yang Y, Zhang YX, Liu M, Maleki SJ, Zhang ML, Liu QM, et al. Triosephosphate Isomerase and filamin C share common epitopes as novel allergens of *Procambarus clarkii*. J Agric Food Chem 2017;65:950-63. PUBMED | CROSSREF
40. Zhang J, Felder A, Liu Y, Guo LT, Lange S, Dalton ND, et al. Nesprin 1 is critical for nuclear positioning and anchorage. Hum Mol Genet 2010;19:329-41. PUBMED | CROSSREF
41. Jonckheere AI, Smeitink JA, Rodenburg RJ. Mitochondrial ATP synthase: architecture, function and pathology. J Inherit Metab Dis 2012;35:211-25. PUBMED | CROSSREF
42. Dominguez R, Holmes KC. Actin structure and function. Annu Rev Biophys 2011;40:169-86. PUBMED | CROSSREF
43. Schevzov G, Whittaker SP, Fath T, Lin JJ, Gunning PW. Tropomyosin isoforms and reagents. Bioarchitecture 2011;1:135-64. PUBMED | CROSSREF
44. Lambert MJ, Cochran WO, Wilde BM, Olsen KG, Cooper CD. Evidence for widespread subfunctionalization of splice forms in vertebrate genomes. Genome Res 2015;25:624-32. PUBMED | CROSSREF
45. Fox-Walsh KL, Dou Y, Lam BJ, Hung SP, Baldi PF, Hertel KJ. The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc Natl Acad Sci U S A 2005;102:16176-81. PUBMED | CROSSREF
46. Incorvaia C, Rapetti A, Aliani M, Castagneto C, Corso N, Landi M, et al. Food allergy as defined by component resolved diagnosis. Recent Pat Inflamm Allergy Drug Discov 2014;8:59-73. PUBMED | CROSSREF
47. Tong WS, Yuen AW, Wai CY, Leung NY, Chu KH, Leung PS. Diagnosis of fish and shellfish allergies. J Asthma Allergy 2018;11:247-60. PUBMED | CROSSREF