Identification of new allergens in *Periplaneta americana* by omics and immunological techniques for cockroach-related allergic diseases: abridged secondary publication

WNS Au, TF Leung, KWS Tsui *

KEY MESSAGES

1. A high-quality, chromosome-level genome of Periplaneta americana was assembled and well-annotated, with a size of 3.06 Gb and 29 939 predicted protein-coding genes.
2. In silico identification of putative allergens was performed in the high-quality P americana genome, revealing a complete allergen profile that includes 43 allergen groups.
3. Proteomic analysis using immunoassay and mass spectrometry methods detected seven novel allergen groups: Per a 14, Per a 15, Per a 16, Per a 17, Per a 18, Per a 19, and Per a 20.
4. A new isoallergen of tropomyosin (Per a 7.02) and multiple potential isoallergens of Per a 5 were identified through bioinformatics and proteomic techniques.

Hong Kong Med J 2025;31(Suppl 1):S38-40
HMRF project number: 07181266

[1] WNS Au, [2] TF Leung, [3] KWS Tsui

[1] *School of Life Sciences, The Chinese University of Hong Kong, Hong Kong SAR, China*
[2] *Department of Paediatrics, The Chinese University of Hong Kong, Hong Kong SAR, China*
[3] *School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong SAR, China*

* Principal applicant and corresponding author: kwtsui@cuhk.edu.hk

Introduction

The American cockroach (*Periplaneta americana*) is linked to allergic diseases. The degree of cockroach infestation is associated with the incidence of cockroach allergy in inner-city areas, leading to an increased risk of asthma onset and higher morbidity among asthmatic individuals.[1] The prevalences of cockroach allergy are 17% to 41% in the United States, 55% to 79% in Brazil, and 44% to 60% in Thailand. Cockroach allergens are detectable in their saliva, debris, secretions, and shed skin[2]; 23 allergens have been identified in the American cockroach and are divided into 13 groups (Per a 1-13). Despite increasing awareness of the negative impact and health implications of cockroach allergies, advances in diagnosis and treatment remain slow owing to the lack of a comprehensive cockroach allergen profile.

Methods

Samples of 8-month-old male and female adult *P americana* were raised. Paired-end sequencing of 150 bp gDNA was performed. A high-molecular-weight gDNA library was prepared and sequenced. Chicago and Hi-C libraries were constructed and sequenced. The backbone of the American cockroach genome was de novo assembled and polished. The final assembly was refined to chromosome level by scaffolding with Chicago and Hi-C data. Quality control was conducted.

The allergenic potential of annotated *P americana* proteins was evaluated based on amino acid sequence similarity with known allergens, mRNA expression levels, and the presence of T cell and B cell epitopes.

Skin prick tests were performed using a commercialised American cockroach extract and other common allergens. Individuals were asked to abstain from anti-allergy drugs for at least 7 days before testing. Histamine (2.5 mg/mL) served as a positive control; buffered saline solution constituted a negative control. Individuals with symptoms and positive results were recruited as cockroach-allergic patients, whereas healthy individuals with negative results were recruited as controls.

A crude protein extract of *P americana* was purified. Potential allergens in 43 gel spots were identified. The DNA sequences of eight potential allergens were retrieved from the genome and manually curated with support from transcriptomic reads. Start and stop codons, as well as exact exon-intron sites of each potential candidate, were identified to produce genes containing full-length sequences. To minimise mis-assembly, curated gene sequences were translated into amino acid sequences and searched against databases. Plasmid cloning, recombinant protein expression, and protein purification were performed. The allergenicity characteristics of the eight purified recombinant proteins were examined by enzyme-linked immunosorbent assay (ELISA).

Results

To de novo assemble the *P americana* genome, 129.3 Gb of long reads and 139.9 Gb of short reads were generated, with coverages of 42.3X and 45.8X, respectively. A draft genome (3.3 Gb) was de novo assembled; Chicago and Hi-C sequencing were used to improve chromosome-level scaffolding. The final assembly resulted in a smaller genome (3.06 Gb) with an N50 of 150.67 Mb, where the largest 48 scaffolds represent 94.6% completeness of the *P americana* genome. Among those scaffolds, the 25 largest exceeded the length threshold of 1 Mb, suggesting successful assembly of 25 chromosome-level scaffolds.

In silico prediction of putative allergens, based on amino acid sequence similarity relative to known allergens, identified 182 potentially allergenic hits in the *P americana* proteome. These protein hits were filtered according to % identity (threshold >50%), expression level (average transcripts per million >10), and the presence of potential epitopes, yielding a *P americana* allergome that comprised 135 components. Among these components, most previously reported *P americana* allergens across 13 groups were retrieved; the corresponding hits showed high % identities (90.74%-100.00%). No identical hits were observed for Per a 11.0101 or Per a 12.0101; however, highly conserved sequences (76.69% and 82.8%, respectively) were identified, suggesting the existence of potential isoallergens in groups 11 and 12. Moreover, groups 4 (lipocalin) and 8 (myosin) matched Pam_039016-T1 and Pam_063052-T1, with high identities of 98.91% and 99.52%, respectively.

112 of the 135 potential candidates were evaluated for allergenic potential. Most showed high similarity to allergens of arthropod origin such as those from cockroach, mosquito, and several mite species. Additionally, 20 and 19 Per a allergen candidates matched published Bla g allergens and Der f allergens with high similarity, respectively, indicating potential cross-reactivity between cockroach and mite allergens. Smaller numbers of the 112 candidates displayed similarity with allergens from fungi and other eukaryotes. Approximately 10% of the candidates showed some similarity with allergens from plants or sea animals, implying cross-reactivity between cockroach allergens and allergens from plant foods or seafood. Along with isoallergens from known groups, 30 novel allergen groups were predicted including cofilin, paramyosin, alpha-tubulin, cyclophilin, heat shock protein, porin 3, aldehyde dehydrogenase, ribosomal protein, enolase, and aldolase. Many newly predicted allergens belonged to three multigene families: glutathione-S transferases, alpha-tubulin, and cyclophilin. The high diversification in group 5 (glutathione-S transferases) emphasised that the 67% similarity cut-off for isoallergen assignment was arbitrary and only intended to serve as a guide. Two group 5 allergens, Per a 5.0101 and Per a 5.0102, shared only 18% identity and 33% similarity with the first identified glutathione-S transferases allergen Bla g 5 (O18598) in *Blattella germanica*, suggesting that criteria for determining isoallergens should be flexible when studying less conserved groups.

To consolidate our predictions, conventional methods for novel allergen identification (two-dimensional gel electrophoresis, immunoblotting, and mass spectrometry) were performed using *P americana* protein extracts and serum samples from allergy-positive patients. The well-annotated proteome served as a customised protein database that replaced the use of redundant public databases in identifying mass spectrometry hits, allowing 43 of 45 spots to be classified by allergen database searches. By integrating mass spectrometry results with in silico predictions, we selected several targets for experimental validation. The allergenicity of novel allergens: enolase (Per a 14), cytochrome C (Per a 15), cofilin (Per a 16), alpha-tubulin (Per a 17), cyclophilin (Per a 18), porin3 (Per a 19), and peroxiredoxin-6 (Per a 20), as well as a novel isoallergen in the tropomyosin group (Per a 7.02) were evaluated in allergy-positive patients using the ELISA. Exception for cytochrome C (Per a 15), the immunoglobulin E (IgE) reactivities of the other novel allergens were significantly higher in allergy-positive patients. The IgE sensitisation rates for these allergens exceeded 50%, suggesting that they can serve as novel major allergens in *P americana*. Cytochrome C (Per a 15) exhibited an IgE reactivity of 30% in allergy-positive patients.

Discussion

This study integrated genomic and transcriptomic analyses with proteomic techniques, including two-dimensional gel electrophoresis, mass spectrometry, and ELISA, to identify novel allergens in *Blattodea* species. The gene set annotated from our high-quality genome served as a customised protein database, which replaced the redundant public databases, in identifying mass spectrometry hits, allowing construction of a broader and more precise allergen spectrum. In total, 135 allergenic hits, comprising 23 previously published allergens and 112 in silico predicted allergens, were identified in the American cockroach. The allergenic features of multiple novel allergens or isoallergens (eg, alpha-tubulin, cyclophilin, cofilin, enolase, porin3, peroxiredoxin-6, and tropomyosin) were confirmed by mass spectrometry and ELISA.

The diagnosis of cockroach allergy primarily relies on skin prick test results or specific IgE sensitivity to cockroach allergens. A panel of purified recombinant cockroach allergens was previously established to facilitate component-resolved

diagnosis, which resolves the IgE specificity of each component and characterises patient sensitisation profiles. However, it is challenging to achieve diagnostic sensitivity comparable to that of natural cockroach extracts because of the incomplete allergen profile available, lack of immunodominant allergens, variability in allergen content across cockroach extracts, and heterogeneous patterns of IgE sensitivity among different patient groups. The incorporation of established allergen panels with our newly identified allergens (especially those belonging to novel groups) could potentially enhance IgE sensitivity in diagnostic tests, thereby improving the efficacy and accuracy of species-specific component-resolved diagnosis in certain populations. The absence of a standardised cockroach extract hinders routine clinical implementation of immunotherapy. Newly identified allergens with high IgE sensitisation rates (eg, enolase, alpha-tubulin, cyclophilin, cofilin, and tropomyosin) should be combined with current cockroach extracts to facilitate the standardisation of cockroach allergen extracts, enable risk stratification, and improve the management of individuals with complex sensitisation profiles.

The presence of contaminants, such as endotoxins and microbial proteins, can trigger unwanted immunological responses during allergen immunotherapy. In the future, the limitations of allergen extracts may be addressed by using purified allergens standardised for purity and biological activity. Pre-clinical trials involving Per a 9 and Per a 10 have been conducted. We anticipate that the novel cockroach allergens identified in the present study will become candidates for a purified allergen mixture in next-generation allergen immunotherapy, potentially reducing hypersensitive reactions triggered by *P americana*.

Funding

This study was supported by the Health and Medical Research Fund, Health Bureau, Hong Kong SAR Government (#07181266). The full report is available from the Health and Medical Research Fund website (https://rfs2.healthbureau.gov.hk).

Disclosure

The results of this research have been previously published in:

1. Wang L, Xiong Q, Saelim N, et al. Genome assembly and annotation of Periplaneta americana reveal a comprehensive cockroach allergen profile. Allergy 2023;78:1088-103.
2. Xiong Q, Wan AT, Liu X, et al. Comparative genomics reveals insights into the divergent evolution of astigmatic mites and household pest adaptations. Mol Biol Evol 2022;39:msac097.
3. Shi L, Xiong Q, Ao FK, et al. Comparative analysis of cysteine proteases reveals gene family evolution of the group 1 allergens in astigmatic mites. Clin Transl Allergy 2023;13:e12324.
4. Xiong Q, Liu X, Wan AT, et al. Genomic analysis reveals novel allergens of Blomia tropicalis. Allergol Int 2024;73:340-4.
5. Wan AT, Xiong Q, Xiao X, et al. Multi-omic analysis of *Tyrophagus putrescentiae* reveals insights into the allergen complexity of storage mites. Clin Exp Allergy 2024;54:77-81.

References

1. Do DC, Zhao Y, Gao P. Cockroach allergen exposure and risk of asthma. Allergy 2016;71:463-74.
2. Eggleston PA, Arruda LK. Ecology and elimination of cockroaches and allergens in the home. J Allergy Clin Immunol 2001;107(3 Suppl):S422-S429.