Genomic Analysis of *Blomia tropicalis* Identifies Novel Allergens for Component-Resolved Diagnosis of Mite Allergy

Authors: Qing Xiong[1,2 †], Xiaoyu Liu[3 †], Angel Tsz-Yau Wan[1,2 †], Nat Malainual[4], Xiaojun Xiao[3], Hui Cao[3], Man-Fung Tang[5], Judy Kin-Wing Ng[1], Soo-Kyung Shin[1], Yang Yie Sio[6], Mingqiang Wang[1,7], Baoqing Sun[8], Ting-Fan Leung[5], Fook Tim Chew[6 *], Anchalee Tungtrongchitr[4 *], Stephen Kwok-Wing Tsui[1,2,9 *]

Affiliations:

1 School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong.
2 Hong Kong Bioinformatics Centre, The Chinese University of Hong Kong, Hong Kong.
3 Shenzhen Key Laboratory of Allergy and Immunology, School of Medicine, Shenzhen University, China.
4 Department of Parasitology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand.
5 Department of Paediatrics, The Chinese University of Hong Kong, Hong Kong.
6 Department of Biological Sciences, National University of Singapore, Singapore.
7 Stanford Cardiovascular Institute, Stanford University School of Medicine, Stanford, CA, 94305, USA.
8 State Key Laboratory of Respiratory Disease, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China.
9 Centre for Microbial Genomics and Proteomics, The Chinese University of Hong Kong, Hong Kong.

* Correspondence to Professor Stephen Kwok-Wing Tsui, E-mail: kwtsui@cuhk.edu.hk; Professor Anchalee Tungtrongchitr, E-mail: anchalee.tun@mahidol.ac.th; Professor Fook Tim Chew, E-mail: dbscft@nus.edu.sg.

† These authors contributed equally to this work.

We have no competing interest to disclose.

Authors' contributions

QX, XL and ATYW designed the experiments, analyzed the data, and wrote the manuscript. NM, XX, HC, MFT, JKWN, SKS, YYS and MW performed the experiments and analyzed the data. BS and TFL provided the resources. FTC, AT and SKWT supervised this study and revised the manuscript.

Abstract

Background

Blomia (*B.*) *tropicalis*, as an important species of house dust mites (HDMs), plays a critical role in allergic diseases in tropical populations, but its allergen components are less investigated than those of other HDMs. Multiple omics methods have largely improved the identification of mite allergens. Here, we sought to identify a comprehensive allergen profile of *B. tropicalis* and advance the allergen component-resolved diagnosis (CRD) of mite allergy.

Methods

Reference mite allergen sequences were searched in a high-quality genome of *B. tropicalis*. Comparative analysis was performed for important allergen groups. ELISA was used to assess the allergenicities of recombinant proteins of specific allergens.

Results

A complete allergen profile of *B. tropicalis* was revealed, including thirty-seven allergen groups (up to Blo t 42). In-depth comparative analysis not only determined the homology of major allergen groups 5 and 21 but also shed light on the emergence and divergence of chitin-binding allergens. The specific Blo t 12 was identified to be a chitin-binding protein originating from the chitinase of allergen group 15. Immunoassays of recombinant proteins verified three novel allergens and the ELISA results suggested geographical differences in the *B. tropicalis* sensitization rate.

Conclusions

The comprehensive allergen profile revealed in *B. tropicalis*, the comparative analysis of allergen groups and the immunoassay assessment of recombinant proteins largely expanded our knowledge to *B. tropicalis* allergens and could ultimately benefit the CRD of HDM allergy.

Key words

Blomia tropicalis; Chitin-binding allergens; Comparative analysis; Component-resolved diagnosis; Genomic analysis; House dust mites

Abbreviations

HDM, house dust mite; ELISA, enzyme-linked immunosorbent assay; NPC2, Niemann-Pick protein type C2; CRD, component-resolved diagnosis

Introduction

House dust mites (HDMs) are one of the major causes of human allergic diseases, including asthma, allergic rhinitis, and atopic dermatitis. The prevalence of HDM allergies highly varies among different climates, urbanization rates and hygiene situations (1). It was reported that the overall sensitization rate of HDMs in Asian countries can be up to 90% in atopic patients and was higher than that in Western countries (2). *B. tropicalis* was considered a storage mite but is now identified as an important species of HDM and causes a series of allergic diseases, especially in tropical regions such as Singapore (3). Compared with the major HDM species *Dermatophagoides* (*D.*) *farinae* and *D. pteronyssinus*, *B. tropicalis* has much fewer reported allergen groups (4-6). Because of the increasing sensitization rates to *B. tropicalis* over the last decades (7-10), comprehensive identification of its allergen components has become an urgent task when component-resolved diagnosis (CRD) has become an inevitable trend (5, 11).

In this study, genome-wide analysis identified a comprehensive putative allergen profile of *B. tropicalis*, including many novel groups, and cross-species comparative analysis revealed a range of allergen gene family variations. Moreover, three novel allergens were identified by ELISA experiments using serum samples from patients with sensitization to *B. tropicalis*. Collectively, this genome-wide analysis of *B. tropicalis* revealed a global picture of the allergen profile and provided insights into the CRD of mite allergy.

Methods

Genome and transcriptome data

The genome and transcriptome data of *B. tropicalis* have been reported (11) and deposited under in the NCBI database under BioProject accession PRJNA702011. The GenBank assembly accession is GCA_021650745.2 and the SRA accessions of transcriptome data are SRR13742047, SRR13742048, SRR13742049 and SRR13742050.

In silico identification of allergens

The protein sequences of allergens in astigmatic mites were downloaded from the WHO/IUIS allergen nomenclature database (4) (dated March 2022) as references. The reference allergens were searched in the annotated protein sequences of *B. tropicalis* genome (11) using BLASTP v2.9.0 (12) with the options '-evalue 1e-6 -max_hsps 1 -max_target_seqs 1 -outfmt 6'. The top matched proteins were considered as candidate allergens. Then, the transcriptome data were mapped to the genome using Hisat2 (13) and the format conversion was performed by SAMtools (14). The sorted and indexed bam file was visualized in the Integrative Genomics Viewer (IGV) (15) and used for the manual curation of the candidate allergens based on transcriptome data to confirm intron-exon split sites. The final *in silico* identified allergens were listed in Table 1 and the sequences have been deposited in NCBI GenBank database (accessions are OQ285922–OQ285989).

Quantification of gene expression level

To quantify the gene expression level of allergens, all the coding sequences (CDS) were collected and indexed using Salmon v0.12.0 (16) with the options '-i index --type quasi -k 31'. Then the transcriptome reads were mapped to the indexed CDS of all allergens using Salmon v0.12.0 (16) and the gene expression levels were represented by the transcript per million (TPM) values. The transcriptome data were obtained from pooled whole mite bodies and their NCBI SRA accessions SRR13742047, SRR13742048, SRR13742049 and SRR13742050 (Figure 1A). As for the gene expression of chitinases, the CDS of target genes were used as index. The bar chart and heatmaps were generated by the program GraphPad Prism v9.0.0.

Cloning and expression of recombinant proteins

Protein and CDS sequences of specific allergens proteins of *B. tropicalis* were obtained after manual curation. The codon-optimized CDS sequences were synthesized and subcloned into plasmid vector (pET-28a+ for rBlo t 5, 18, 21, 23 and 26, pET-32a+ for rBlo t 12, 24, 25 and 31), and then transformed in *E. coli* cells, which was performed by Sangon Biotech (Shanghai, China). The *E. coli* cells were induced by isopropyl β-D-1-thiogalactopyranoside (final concentration: 1 mM) for protein expression and incubated at 16°C for 16 hours or 37°C for 4 hours. Proteins were extracted after cell ultrasonication, and 12% (w/v) SDS-PAGE analysis were performed to determine the protein solubility. The recombinant proteins were finally purified to over 90% by Ni^{2+} affinity chromatography using a fast protein liquid chromatography system. Four recombinant proteins, rBlo t 5, 12, 24 and 31, were soluble and purified from supernatant after centrifuge. The other five recombinant proteins were initially insoluble, then were denatured and purified from inclusion body pellets, and finally refolded in dialysis.

Immunoassay of specific allergens

Allergen-specific IgE levels > 0.35 kUA/L were considered positive, which was further graded into classes 1 to 6 according to manufacturer's instructions. For the ELISA experiments, serum samples were collected in Prince of Wales Hospital (Hong Kong) and obtained from 26 patients who presented with positive ImmunoCAP results (at least class 3) towards *B. tropicalis* (d201), and from 17 healthy individuals without mite sensitization as negative controls. The patients have at least one type of allergic disease in asthma, allergic rhinitis, and atopic dermatitis.

The allergenicity of the recombinant proteins was assessed by the enzyme-linked immunosorbent assay (ELISA). In the indirect sandwich ELISA, a 96-well microtiter plate was coated with 100 ul (2 ug/ml) of recombinant proteins per well in sodium bicarbonate coating buffer (0.1 mol/L, pH 9.5). The plate was washed with 0.05% (v/v) Tween-20 in PBS washing buffer for 4 times and then was blocked with 375 ul 3% (w/v) skim milk in PBST buffer for 30 minutes at 37□. Serum samples were diluted 1:4 in PBST containing 1% (w/v) BSA and 50 ul was added to each blocked well for 1 hr at room temperature. After washing, IgE antibodies were detected using biotinylated goat anti-human

IgE (Vector, Burlingame, CA, USA) and streptavidin peroxidase (Sigma-Aldrich, St. Louis, MO, USA). The wells were then washed again with PBST buffer and incubated with 100 ul 3,3',5,5'-Tetramethylbenzidine (Kirkegaard and Perry Laboratories, Gaithersburg, MD, USA) for 15 mins at room temperature. The reactions were stopped with the addition of 2M sulphuric acid (50 ul/well). The plates were then placed in the Benchmark Plus Microplate Spectrophotometer (BioRad, Hercules, CA, USA) and absorbance at 450 nm was recorded, where a high absorbance is indicative of a high concentration. The ELISA results were visualized and statistically analyzed in the program GraphPad Prism v9.0.0.

Comparative analysis of gene families

All proteins of the six astigmatic mites, *D. farinae*, *D. pteronyssinus*, *Psoroptes* (*P.*) *ovis*, *Sarcoptes* (*S.*) *scabiei*, *B. tropicalis* and *Tyrophagus* (*T.*) *putrescentiae*, were searched by BLASTP v2.9.0 (12) with reference proteins at E-value cutoff of 1E-6. To collect all the genes in a family, the corresponding genes of astigmatic mite in Swiss-Prot database (17) (dated October 2021) or the WHO/IUIS allergen nomenclature database (4) (dated March 2022) were used for searching in the annotated protein sequences. After manual curation based on transcriptome data to confirm intron-exon split sites and filtering out annotated proteins with shorter than 50% average length of the gene family, all proteins of the six astigmatic mites identified in the target gene families were aligned and drawn into a phylogenetic tree. Sequence alignment was performed by CLUSTAL W (18) and MUSCLE (19) in MEGA v11.0.13 (20), and all phylogenetic trees were constructed by MEGA v11.0.13 (20) with maximum likelihood (ML) algorithm (21) in the Jones-Taylor-Thornton (JTT) model (22), 80% site coverage and 100 bootstrap replicates, and then edited by Interactive Tree of Life (iTOL) (23).

If two genes were located adjacently on genome and no other gene was located between them, they were considered as tandemly arrayed genes. If two genes were separated by no more than 10 genes, they were considered as proximally arrayed. Frequent tandemly arrayed genes were identified in gene families and connected with curve lines in phylogenetic trees edited by the online tool iTOL (23) with validated gene synteny information.

Data availability

The genome and sequencing data of *B. tropicalis* are deposited in NCBI database under BioProject accession PRJNA702011. The *in silico* identified allergen sequences were uploaded in NCBI GenBank database under accessions OQ285922–OQ285989.

Ethics approval

This study was approved by the institutional review board of Prince of Wales Hospital (CREC Ref. No.: 2018.663) for using the patient sera in ELISA experiments, and subjects and/or their parents have provided informed consent to participate.

Results

Genome-wide analysis of allergens

As an important HDM, *B. tropicalis* has only fourteen groups of allergens reported in the WHO/IUIS Allergen Nomenclature database (4). To fully understand the allergens of *B. tropicalis*, a genome-wide analysis was performed using the reported mite allergens as references to search in the high-quality genome (11). Thirty-seven allergen groups (up to Blo t 42) were identified and covered a comprehensive putative allergen profile of *B. tropicalis* (Table 1), including a wide range of homologs (or isoallergens). The reported group 19 allergen of *B. tropicalis*, Blo t 19 (GenBank accession: AHG97583), could not be found in the genome. In our naming rules, we used the first and last two digits after the decimal point to differentiate the different genes and isoforms, respectively. Multiple isoforms originating from the same gene locus were not included in this study, such as those caused by single-nucleotide polymorphism or alternative splicing. In our identification, group 2 and 12 allergens shared 93.8% and 92.4% identities with the reported allergens, respectively (Table 1). All other reported allergens in the database could be matched in our identified sequences with over 97% identity (Table 1). All these discrepancies could be explained by intraspecific differences.

The group 13 allergen, Blo t 13.0101, and the two major allergens of *B. tropicalis*, Blo t 5.0101 and 21.0101 (7, 24, 25), had significantly higher gene expression levels (Figure 1A). The high gene expression of the group 13 allergen was similarly found in the storage mite *T. putrescentiae* (26). The relatively high expression levels of the group 5 and 21 allergens may explain why they are major allergens of *B. tropicalis* but not the other mites. In the canonical HDMs *D. farinae* and *D. pteronyssinus*, we could not observe similar gene expression features (27).

Group 1 allergens of mites are characterized as cysteine proteases and are considered major allergens of HDMs but do not include *B. tropicalis*. Five homologs of group 1 allergens were identified in *B. tropicalis*, namely Blo t 1.0101, 1.0201, 1.0301, 1.0401 and 1.0501. Blo t 1.0101 and 1.0201 shared high similarities with the two reported Blo t 1 sequences (Table 1). All five homologs were proximally arrayed in the genome of *B. tropicalis* (Figure S1) and possibly originated from the duplication of one ancestral gene. Similarly, four homologs were observed in the group 2 allergen, in

which Blo t 2.0201 and 2.0301 were tandemly arrayed (Figure S1). The best-matched group 2 allergen, Blo t 2.0101 (gene locus: BT_003416.02), shared only 93.8% identity with the reported Blo t 2.0101 and the difference was mainly caused by a mismatch region on the N-terminus (Figure S2). Along with groups 22 and 35 allergens of mites (including Der f 22, Der f 35 and Der p 22(28)), group 2 allergens belong to the Niemann-Pick protein type C2 (NPC2) family and have been well analyzed (26).

Group 3, 6 and 9 allergens are all members of the serine protease family (5), in which massive tandem gene duplications of *B. tropicalis* occurred (11). According to the adapted phylogenetic tree (11), the three serine protease allergens were located in distinct clusters (Figure S3). Group 3 allergens were close to the species-specific inactive serine proteases of *B. tropicalis* and *S. scabiei*, while group 6 and 9 allergens had no close homolog from the two parasitic mites *P. ovis* and *S. scabiei* (Figure S3).

At least two highly identical genes encoding alpha amylase, were identified as single-exon genes, and shared over 99% identity with the reported group 4 allergen, Blo t 4 (GenBank accession: AAQ24543). Because these two homologs of Blo t 4 are located on different scaffolds in different gene syntenies, we determined that they were generated via a recent transposable duplication. Initially, four homologous genes of Blo t 4 were identified in *B. tropicalis* (11). However, two low-quality genes were discarded because one gene was located on a short scaffold encoding only one gene, while the other gene on a short scaffold shared highly conserved synteny with BT_008084.02 and was considered a heterozygous gene.

For the major allergen groups 5 and 21 of *B. tropicalis* (7, 24, 25), we noticed that Blo t 5.0101 and 21.0101 are tandemly arrayed homologs that share 41.9% identity. To further explore the divergence of this gene family (G5/21), we collected all the genes of six mites and performed phylogenetic analysis (Figure 1B). Two major allergens, groups 5 and 21 of *B. tropicalis*, were compared with those of the other five astigmatic mites (11) (Figure 1B). In *D. farinae*, *D. pteronyssinus* and *P. ovis*, the two tandemly arrayed G5/21 genes were located on opposite strands and shared identical gene

synteny (Figure 1C), which supports the close phylogenetic relationship among the three species (11). For the skin-burrowing mite *S. scabiei*, the absence of the G5/21 gene may be caused by complete gene decay. In two canonical storage mites, *B. tropicalis* and *T. putrescentiae*, the gene syntenies surrounding G5/21 genes are totally different from those in other species (Figure 1C). Unlike *D. farinae*, *D. pteronyssinus* and *P. ovis*, the two G5/21 genes of *B. tropicalis* are located on the same strand. For *T. putrescentiae*, four tandemly arrayed G5/21 genes, including a dimer gene TP_014182.02, were identified (26). Both group 5 and 21 allergens were highly expressed in *B. tropicalis* (Figure 1A) but expressed at low levels in *T. putrescentiae* (26). We suggest that group 5 and 21 allergens of mites are homologous genes that originated from an ancestral gene.

Only one homolog of the group 7 allergen was identified, while up to five tandemly arrayed homologs of group 8 allergens were found in *B. tropicalis* (Table 1, Figure S1). Group 8 allergens of mites were characterized as glutathione-S-transferase (5), of which the gene family variation has been well explored (26). Two tropomyosins and one paramyosin were identified in group 10 and 11 allergens, respectively (Table 1). Notably, the group 12 allergen, a protein containing a chitin-binding domain that has not been reported in other HDMs, could be identified in *B. tropicalis* (gene locus: BT_016677.01) but presented only 92.4% identity to that reported for Blo t 12.0101 (GenBank accession: AAA78904, Table 1). In group 13 allergens, three homologs were found, including two tandemly arrayed genes (Table 1, Figure S1).

Moreover, a wide range of newly identified allergen groups were identified in *B. tropicalis* (Table 1). All these newly predicted allergens of *B. tropicalis* exhibit high similarities (>45% identity) to those of *D. pteronyssinus* (prefix: Der p) or *D. farinae* (prefix: Der f), and no homolog of the group 27 allergen was identified. Along with group 2 allergens, group 22 and 35 allergens are homologs in the NPC2 family (26) and were not identified in *B. tropicalis*.

Comparative analysis of chitin-binding allergens

Chitinases and other chitin-binding domain containing proteins have been reported to play important roles in human allergic disease (29, 30). Groups 12, 23 and 37 allergens of mites contain chitin-

binding domains, while groups 15 and 18 allergens are characterized as chitinases (5).

Group 12 allergen was only reported in *B. tropicalis* (31) in the WHO/IUIS allergen nomenclature database. The reported Blo t 12 (GenBank accession: AAA78904) was from mites collected in Colombia, while the *B. tropicalis* mite in this study originated from Singapore (NCBI BioSample: SAMN17935020). Therefore, it is reasonable that our identified Blo t 12 is nearly identical (only one different amino acid) to that from mites collected in Singapore (32) but shares only 92.4% identity with the reported Blo t 12. In sequence alignment, we observed that the major difference was the consequence of a regional frame-shift mutation (block 2, Figure S4). To explore the genomic specificity of the group 12 allergen, gene synteny analysis was performed among six astigmatic mites (Figure 2A). The group 12 allergen gene of *B. tropicalis*, Blo t 12, was located adjacently with a chitinase gene that shared as high as 59.8% identity in protein sequence with the group 15 allergen of *D. pteronyssinus* and identified as Blo t 15 (Table 1). Therefore, Blo t 12 and 15 were tandemly arrayed and shared high conservation on their C-terminal sequences (chitin-binding domain). (Figure S5). Group 23 was reported in *D. farinae* and *D. pteronyssinus* (33, 34). We identified a homolog of group 23 allergen in the annotated genome of *B. tropicalis*, Blo t 23, that shares 40.0% identity with Der p 23 (GenBank accession: ACB46292). Despite the low similarity, the chitin-binding domains of Blo t 23 and Der p 23 were highly conserved (Figure S6).

Group 15 and 18 allergens of mites are both identified as chitinases. All the chitinase genes of the six astigmatic mites were divided into eleven clusters in the phylogenetic tree (Figure 2B). Tandem gene duplications were observed in Clusters 10 and 11(Figure 2B). Clusters 2 and 3 were horizontally transferred genes, as previously reported (11). The chitinases in the two distinct Clusters 1 and 10 were identified as group 18 and 15 allergens of mites, respectively (Figure 2B). Interestingly, both group 15 and 18 allergens decayed in the genome of *S. scabiei* (Figure 2B). The gene synteny alignment supported the gene decay of the group 15 allergen in *S. scabiei* (Figure 2A). Both group 15 and 18 allergens are highly conserved among astigmatic mites in protein sequences (Figures S7 and S8). In gene expression analysis, we observed the consistently high expression of group 15 and 18 allergens in astigmatic mites (Figure S9). In particular, the chitinases in Cluster 11 were highly

expressed in two parasitic mites, *P. ovis* and *S. scabiei* (Figure S9).

Immunoassay of recombinant allergen proteins

Although our genome-wide analysis revealed a comprehensive allergen profile of *B. tropicalis* (Table 1), serological evidence is still needed to confirm the allergenicity of many novel allergens. To verify and evaluate the allergenicity of the proposed novel allergens, their recombinant proteins were expressed and subjected to ELISA experiment using patient serum samples. Serum samples were collected in Hong Kong from allergic patients for testing and healthy individuals as controls (Table S1). When at least class 3 allergy (over 3.5 kUA/L of IgE) was considered positive, 58% (29/50) of those tested positive for HDM (Der p 1, *D. pteronyssinus*) were identified to be positive for *B. tropicalis*, while only one sample (BC0699 out of 20 samples) was negative for HDM (class 0) but positive for *B. tropicalis* (class 4). The sensitization rate of those serum samples was apparently lower than that in Singapore where the prevalence of *B. tropicalis* could be up to 97.5% in allergic children (35).

In subsequent ELISA experiments, 26 serum samples were collected from allergic patients who tested positive for *B. tropicalis* (BT-positive). Nine recombinant proteins of *B. tropicalis* were expressed and purified (Figure S10), including three reported allergens (rBlo t 5, 21 and 12) and six novel proposed allergens (rBlo t 18, 23, 24, 25, 26 and 31). The absorbance values at 450 nm of recombinant proteins represented the IgE levels of specific allergens (Table S2, Figure 3 and S11). The major allergens rBlo t 5 and 21 were used as positive controls (Figure 3A and B). According to the allergen-specific IgE levels in BT-allergic sera, rBlo t 18, 23 and 26 were verified as novel allergens of *B. tropicalis* (Figure 3C-E), while rBlo t 24, 25 and 31 were not confirmed to be novel allergens, because of their low IgE levels (Figure S11). Unexpectedly, rBlo t 12 presented significantly higher IgE levels in BT-allergic sera, but the low absorbance values (all below 0.15) could not well confirm its allergenicity (Figure S11A), which was similarly observed in rBlo t 24 (Figure S11B). The IgE levels of two major allergens, rBlo t 5 and 21 (Figure 3A and B), were highly dispersed with high maximum values (above 1.0) and standard deviations (SDs), which affected the significance levels (*P* values). Similar features were observed in three novel allergens, rBlo t 18, 23

and 26 (Figure 3C-E). Intriguingly, regression analysis revealed good correlations among the IgE levels of rBlo t 5, 21, 23 and 26, but not rBlo t 18, and suggested the cosensitization of these allergens (Figure S12).

For the sensitization rates of specific allergens (Figure 3F), unexpectedly, rBlo t 26 and 18 presented high positive percentages of 88.46% and 73.08%, respectively, while two major allergens rBlo t 5 and 21 demonstrated only 53.85%, and the chitin-binding allergen rBlo t 23 showed low sensitization rates 38.46%. The prevalence observed in this batch of Hong Kong sera was totally different from that found in Singapore samples, which showed the highest positivity to the two major allergens Blo t 5 and 21 (7). When we regarded absorbance values equal to or higher than 0.2 as high IgE and allergy levels, two major allergens, rBlo t 5 and 21, demonstrated higher percentages of 38.46% and 30.77% (Figure 3F).

Discussion

With the advent of omics era, a series of genome-based approaches have been applied to the study of allergens and mite allergies (11, 27, 36-41). Based on the high-quality genome of the dust mite *B. tropicalis*, our genome-wide analysis identified a comprehensive putative allergen profile of thirty-seven allergen groups (up to Blo t 42), which covered a wide range of allergen homologs. The specific allergen Blo t 12 has been well confirmed, while Blo t 19 was considered nonexistent in the genome. Considering group 19 allergen could not been identified in *D. farinae*, we speculated that the allergen Blo t 19 was falsely reported (42). Two cysteine proteases of *B. tropicalis* have been identified as isoallergens in group 1. In the allergen profile, many more homologs of *B. tropicalis* allergens were identified and proposed to be allergens (Table 1), including up to five homologs of group 1 allergens. Consistent with a previous report (11), tandem gene duplication could cause gene family expansion and result in multiple homologs (Figure S1), such as five tandemly arrayed homologous genes of Blo t 8 (glutathione S-transferase). Transcriptomic analysis revealed interesting gene expression levels of the *B. tropicalis* allergens. Unexpectedly, the Blo t 13 homolog, Blo t 13.0101, presented the highest expression level (Figure 1A), which was similarly observed in the storage mite *T. putrescentiae* (26). After Blo t 13.0101, two major allergens of *B. tropicalis*, Blo t 5 and 21, were the two most highly expressed allergens (Figure 1A), which may be related to why they were the major allergens. Blo t 13 was not identified to be major allergen (43), possibly because it could only be found in mite bodies rather than fecal pellets (44). The transcriptome data and expression levels are highly variable with geographical location, mite collection and culture conditions, especially when Blo t 2 was identified to be the most abundant allergen in a previous report (45). Therefore, multiple sampling is necessary in future allergen abundance analysis of mites.

To better understand the complicated allergen profile, more comparative analysis was performed for important allergen groups. A wide range of tandem duplications were found in the expansion and evolution of many allergen groups. In particular, the two major allergens of *B. tropicalis*, Blo t 5 and 21, were identified to be tandemly arrayed homologs. In the gene family of group 5 and 21 allergens, two tandemly arrayed homologs were identified in other astigmatic mites, including *D. farinae*, *D. pteronyssinus* and *P. ovis*. Four homologs were found in the storage mite *T. putrescentiae*, while none

was observed in the skin-borrowing mite *S. scabiei* (Figure 1C). Not only the specific Blo t 12 but also group 15, 18 and 23 allergens of mites contain chitin-binding domains. Gene synteny and sequence alignment suggested that the group 12 allergens were specific to canonical storage mites (including *B. tropicalis* and *T. putrescentiae*) and derived from the group 15 allergen (chitinase). Comparative analysis of chitinases further revealed the divergence of group 15 and 18 allergens that are especially allergenic in dogs. Despite the absence of aligned gene synteny, Blo t 23 was identified with a conserved chitin-binding domain shared with Der p 23. Such a wide range of chitin-binding allergens would be an interesting topic of mite allergy and deserves further exploration.

In the ELISA test of recombinant proteins (Figure 3), allergenicity was assessed for five specific allergens of *B. tropicalis*, including three novel allergens, Blo t 18, 23 and 26. In particular, Blo t 18 and 26 presented not only high positivity but also a relatively high proportion of high allergy levels, which suggests that they are important novel allergens of *B. tropicalis*. The high proportion of high allergy levels to rBlo t 5 and 21 indicated the importance of these two major allergens. The sensitization rates of specific allergens in this batch of Hong Kong sera demonstrated totally distinct features from those observed in Singapore samples with obviously high sensitization rates of Blo t 5 and 21 (7), which indicated the geographical difference in *B. tropicalis* allergy. Considering that nearly all the samples (BT-positive) were also positive for *D. pteronyssinus* (Table S1), we cannot exclude the influence of cross-reactivity in the ELISA results. A major limitation of this study is that only nine recombinant proteins were expressed.

In summary, this genome-wide analysis shed light on the allergen profile in *B. tropicalis*, the comparative analysis revealed insights into the divergence of allergen gene families, and the immunoassay experiments of recombinant proteins provided serological evidence for novel allergens. Taken together, this comprehensive analysis of *B. tropicalis* paves the way to the CRD of mite allergy.

References

1. Calderon MA, Linneberg A, Kleine-Tebbe J, De Blay F, Hernandez Fernandez de Rojas D, Virchow JC, et al. Respiratory allergy caused by house dust mites: What do we really know? J Allergy Clin Immunol. 2015;136(1):38-48.
2. Tham EH, Lee AJ, Bever HV. Aeroallergen sensitization and allergic disease phenotypes in Asia. Asian Pac J Allergy Immunol. 2016;34(3):181-9.
3. Guilleminault L, Viala-Gastan C. *Blomia tropicalis*: A house dust mite in the tropics. Rev Mal Respir. 2017;34(8):791-801.
4. Pomés A, Davies JM, Gadermaier G, Hilger C, Holzhauser T, Lidholm J, et al. WHO/IUIS Allergen Nomenclature: Providing a common language. Mol Immunol. 2018;100:3-13.
5. Thomas WR. Hierarchy and molecular properties of house dust mite allergens. Allergol Int. 2015;64(4):304-11.
6. Xiong Q, Wan ATY, Tsui SK. A Mini-review of the Genomes and Allergens of Mites and Ticks. Curr Protein Pept Sci. 2020;21(2):114-23.
7. Kidon MI, Chiang WC, Liew WK, Ong TC, Tiong YS, Wong KN, et al. Mite component-specific IgE repertoire and phenotypes of allergic disease in childhood: the tropical perspective. Pediatr Allergy Immunol. 2011;22(2):202-10.
8. Juliá-Serdá G, Cabrera-Navarro P, Acosta-Fernández O, Martín-Pérez P, García-Bello MA, Antó-Boqué J. Prevalence of sensitization to *Blomia tropicalis* among young adults in a temperate climate. J Asthma. 2012;49(4):349-54.
9. Zakzuk J, Acevedo N, Cifuentes L, Bornacelly A, Sánchez J, Ahumada V, et al. Early life IgE responses in children living in the tropics: a prospective analysis. Pediatr Allergy Immunol. 2013;24(8):788-97.
10. Santos da Silva E, Asam C, Lackner P, Hofer H, Wallner M, Silva Pinheiro C, et al. Allergens of *Blomia tropicalis*: An Overview of Recombinant Molecules. Int Arch Allergy Immunol. 2017;172(4):203-14.
11. Xiong Q, Wan AT-Y, Liu X, Fung CS-H, Xiao X, Malainual N, et al. Comparative Genomics Reveals Insights into the Divergent Evolution of Astigmatic Mites and Household Pest Adaptations. Mol Biol Evol. 2022;39(5).
12. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10(1):421.
13. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357-60.
14. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
15. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6.
16. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14(4):417-9.
17. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 2000;28(1):45-8.
18. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22(22):4673-80.
19. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic

Acids Res. 2004;32(5):1792-7.
20. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol. 2021;38(7):3022-7.
21. Dhar A, Minin VN. Maximum Likelihood Phylogenetic Inference. In: Kliman RM, editor. Encyclopedia of Evolutionary Biology. Oxford: Academic Press; 2016. p. 499-506.
22. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Bioinformatics. 1992;8(3):275-82.
23. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293-W6.
24. Gao YF, Wang de Y, Ong TC, Tay SL, Yap KH, Chew FT. Identification and characterization of a novel allergen from *Blomia tropicalis*: Blo t 21. J Allergy Clin Immunol. 2007;120(1):105-12.
25. Kuo IC, Cheong N, Trakultivakorn M, Lee BW, Chua KY. An extensive study of human IgE cross-reactivity of Blo t 5 and Der p 5. J Allergy Clin Immunol. 2003;111(3):603-9.
26. Wan AT-Y, Xiong Q, Xiao X, Ao KF-K, Jang SW, Wong BS-H, et al. Multi-Omic Analysis of *Tyrophagus putrescentiae* Reveals Insights into the Allergen Complexity of Storage Mites. bioRxiv. 2022:2022.10.21.513148.
27. Liu X-Y, Yang KY, Wang M-Q, Kwok JS-L, Zeng X, Yang Z, et al. High-quality assembly of *Dermatophagoides pteronyssinus* genome and transcriptome reveals a wide range of novel allergens. J Allergy Clin Immunol. 2018;141(6):2268-71.e8.
28. Zhou Y, Li L, Yu Z, Gu X, Pan R, Li Q, et al. *Dermatophagoides pteronyssinus* allergen Der p 22: Cloning, expression, IgE□binding in asthmatic children, and immunogenicity. Pediatr Allergy Immunol. 2022;33(8):e13835.
29. Lee CG. Chitin, chitinases and chitinase-like proteins in allergic inflammation and tissue remodeling. Yonsei Med J. 2009;50(1):22-30.
30. Elias JA, Homer RJ, Hamid Q, Lee CG. Chitinases and chitinase-like proteins in T(H)2 inflammation and asthma. J Allergy Clin Immunol. 2005;116(3):497-500.
31. Puerta L, Caraballo L, Fernández-Caldas E, Avjioglu A, Marsh DG, Lockey RF, et al. Nucleotide sequence analysis of a complementary DNA coding for a *Blomia tropicalis* allergen. J Allergy Clin Immunol. 1996;98(5 Pt 1):932-7.
32. Zakzuk J, Jiménez S, Cheong N, Puerta L, Lee BW, Chua KY, et al. Immunological characterization of a Blo t 12 isoallergen: identification of immunoglobulin E epitopes. Clin Exp Allergy. 2009;39(4):608-16.
33. Weghofer M, Grote M, Resch Y, Casset A, Kneidinger M, Kopec J, et al. Identification of Der p 23, a peritrophin-like protein, as a new major *Dermatophagoides pteronyssinus* allergen associated with the peritrophic matrix of mite fecal pellets. J Immunol. 2013;190(7):3059-67.
34. He Y, Dou C, Su Y, Chen J, Zhang Z, Zhao Z, et al. Identification of Der f 23 as a new major allergen of *Dermatophagoides farinae*. Mol Med Rep. 2019;20(2):1270-8.
35. Chew FT, Lim SH, Goh DY, Lee BW. Sensitization to local dust-mite fauna in Singapore. Allergy. 1999;54(11):1150-9.
36. Stewart GA. Studies of house dust mites can now fully embrace the "-omics" era. J Allergy Clin Immunol. 2015;135(2):549-50.
37. Chan T-F, Ji K-M, Yim AK-Y, Liu X-Y, Zhou J-W, Li R-Q, 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(2):539-48.
38. Yang Z, Wan AT, Liu XY, Xiong Q, Liu ZG, Tsui SK. Identification, functional annotation, and stability

analysis of miRNA in *Dermatophagoides pteronyssinus*. Allergy. 2020;75(5):1237-40.
39. 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. 2022.
40. Zhou Y, Klimov PB, Gu X, Yu Z, Cui X, Li Q, et al. Chromosome-level genomic assembly and allergome inference reveal novel allergens in *Tyrophagus putrescentiae*. Allergy. 2023.
41. Lin J, Huang N, Li J, Liu X, Xiong Q, Hu C, et al. Cross-reactive antibodies against dust mite-derived enolase induce neutrophilic airway inflammation. Eur Respir J. 2021;57(1).
42. Thomas WR, Smith WA, Hales BJ, Mills KL, O'Brien RM. Characterization and immunobiology of house dust mite allergens. Int Arch Allergy Immunol. 2002;129(1):1-18.
43. Caraballo L, Puerta L, Jimenez S, Martinez B, Mercado D, Avjiouglu A, et al. Cloning and IgE binding of a recombinant allergen from the mite *Blomia tropicalis*, homologous with fatty acid-binding proteins. Int Arch Allergy Immunol. 1997;112(4):341-7.
44. Satitsuksanoa P, Kennedy M, Gilis D, Le Mignon M, Suratannon N, Soh WT, et al. The minor house dust mite allergen Der p 13 is a fatty acid-binding protein and an activator of a TLR2-mediated innate immune response. Allergy. 2016;71(10):1425-34.
45. Santos da Silva E, Marques Ponte JC, Barbosa da Silva M, Silva Pinheiro C, Carvalho Pacheco LG, Ferreira F, et al. Proteomic Analysis Reveals Allergen Variability among Breeds of the Dust Mite *Blomia tropicalis*. Int Arch Allergy Immunol. 2019;180(3):159-72.

Acknowledgments

We would like to thank the staff members in the Siriraj Dust Mite Center for Services and Research, Siriraj Hospital, Bangkok, Thailand, for the mite culturing work.

This work was made possible by grants:

1. General Research Fund from Research Grants Council of Hong Kong (Reference numbers: 464710, 475113, 14119219, 14119420, 14175617)

2. Health and Medical Research Fund from Food and Health Bureau of Hong Kong (Reference numbers: 06171016, 07181266)

3. Continuation project of Joint Research Fund for Overseas Chinese Scholars and Scholars in Hong Kong and Macao Young Scholars (Reference number: 31729002)

4. National Natural Science Foundation of China (Reference numbers: 81971514, 82073950)

5. Shenzhen Science and Technology Plan Project (Reference number: GJHZ20190822095605512, SGDX20201103095609027)

Figure legends

Figure 1. Genomic and transcriptomic analyses of *B. tropicalis* allergens

(A) Expression level of all the allergens genes identified in the *B. tropicalis* genome (Table 1). The gene expression level was represented by transcript per million (TPM) values. In the bar chart, error bar means standard deviation (SD). The coding sequences of all the allergen genes were used as index for read mapping of four transcriptome data, namely SRR13742048, SRR13742047, SRR13742050 and SRR13742049 (from top to bottom in the heatmap). (B) Phylogenetic analysis of the gene family of group 5 and 21 allergens in six astigmatic mites. All the allergens from the WHO/IUIS Allergen Nomenclature database were highlighted in red. (C) Gene synteny alignment of group 5 and 21 allergens in six astigmatic mites. A dimer gene of *T. putrescentiae* (TP_014182.02) was highlighted in red. G5/21, homologs of group 5 and 21 allergens. The turnover symbol means reverse complement.

Figure 2. Comparative analysis of chitin-binding allergens

(A) Gene synteny alignment of the group 12 allergen loci in six astigmatic mites. The group 12 allergen homologs of *B. tropicalis* and *T. putrescentiae* and the chitinase genes were highlighted in red and yellow boxes, respectively. The turnover symbol means reverse complement. None of group 12 allergen homolog or chitinase was identified in the locus of *S. scabiei*. (B) Phylogenetic analysis of chitinase genes in six astigmatic mites. Eleven clusters were classified, including the Cluster 1 and 10 identified as the group 18 and 15 allergens, respectively. Four reported allergens in the WHO/IUIS Allergen Nomenclature database were marked by black triangles. Tandemly arrayed genes and proximally arrayed genes (separated by no more than 10 genes) were connected by curved solid lines and dotted lines, respectively.

Figure 3. ELISA results of recombinant allergen proteins of *B. tropicalis*

The IgE antibody binding to recombinant proteins of *B. tropicalis* was evaluated by ELISA using serum samples from 26 patients positive for *B. tropicalis* (BT-allergic sera) for testing and 17 healthy individuals as control sera (Table S1). The absorbance values at 450 nm of recombinant proteins were recorded to represent the IgE levels of specific allergens. The lines in dot plots were set as the

mean with a standard deviation (SD) and the dotted lines indicated the cutoff value, mean + 2*SD of the healthy controls. The allergenicity was evaluated for five recombinant proteins as follows: (A) rBlo t 5, (B) rBlo t 21, (C) rBlo t 18, (D) rBlo t 23, (E) rBlo t 26. Unpaired two-tailed t-test P-values indicate statistical significance (all below 0.05). (F) Positive percentage of five recombinant proteins for the allergens of *B. tropicalis*. The absorbance values equal to or higher than 0.2 were regarded as high IgE and allergy levels and their percentages were highlighted in columns.

523 **Table 1. Summary of all allergen genes identified in the *B. tropicalis* genome**

Allergen ID	Isoallergen ID	Gene locus	Biochemical function	No. of exons	No. of amino acids	TPM value (Avg±Stdev)	Homologue (Identity)
Blo t 1	Blo t 1.0101*	BT_013135.03	Cysteine protease	4	331	18744±2261	AAK58415 Blo t 1.0101 (98.2%)
	Blo t 1.0201*	BT_013137.01	Cysteine protease	4	333	24102±3762	AAQ24541 Blo t 1.0201 (98.8%)
	Blo t 1.0301	BT_013135.02	Cysteine protease	4	333	757±582	AAQ24541 Blo t 1.0201 (63.8%)
	Blo t 1.0401	BT_013135.04	Cysteine protease	2	330	509±158	AAQ24541 Blo t 1.0201 (39.0%)
	Blo t 1.0501	BT_013144.01	Cysteine protease	3	369	650.575±127	AAQ24541 Blo t 1.0201 (32.9%)
Blo t 2	Blo t 2.0101*	BT_003416.02	ML-domain protein	2	142	42029±16214	AAQ73483 Blo t 2.0101 (93.8%)
	Blo t 2.0201	BT_007270.01	ML-domain protein	2	144	40370±13402	AAQ73483 Blo t 2.0101 (50.0%)
	Blo t 2.0301	BT_007271.01	ML-domain protein	3	145	372±57	AAQ73483 Blo t 2.0101 (42.5%)
	Blo t 2.0401	BT_007142.01	ML-domain protein	2	151	5289±838	CAI05848 Blo t 2.0101 (41.9%)
Blo t 3	Blo t 3.0101*	BT_006831.01	Trypsin	2	266	10277±880	AAM10779 Blo t 3.0101 (98.9%)
	Blo t 3.0201	BT_015438.01	Trypsin	2	269	2075±267	AAM10779 Blo t 3.0101 (55.6%)
Blo t 4	Blo t 4.0101*	BT_008084.02	Alpha-amylase	1	515	5328±1077	AAQ24543 Blo t 4.0101 (99.2%)
	Blo t 4.0201*	BT_011038.02	Alpha-amylase	1	515	5328±1077	AAQ24543 Blo t 4.0101 (99.0%)
Blo t 5	Blo t 5.0101*	BT_005893.01	Unknown	2	134	111590±10602	AAD10850 Blo t 5.0101 (100%)
Blo t 6	Blo t 6.0101*	BT_006509.01	Chymotrypsin	3	275	5203±695	AAQ24544 Blo t 6.0101 (100%)
Blo t 7	Blo t 7.0101*	BT_006271.01	Bactericidal/permeability-increasing protein	2	212	24757±7825	ASX95438.1 Blo t 7.0101 (99.5%)
Blo t 8	Blo t 8.0101*	BT_006863.04	Glutathione S-transferase	2	236	7738±5410	ACV04860 Blo t 8.0101 (97.0%)

	Blo t 8.0201	BT_006863.05	Glutathione S-transferase	2	236	1463±414	ACV04860 Blo t 8.0101 (90.3%)
	Blo t 8.0301	BT_006865.01	Glutathione S-transferase	2	236	5687±1686	ACV04860 Blo t 8.0101 (80.9%)
	Blo t 8.0401	BT_006863.03	Glutathione S-transferase	2	236	452±256	ACV04860 Blo t 8.0101 (45.5%)
	Blo t 8.0501	BT_006863.02	Glutathione S-transferase	2	237	133±25	ACV04860 Blo t 8.0101 (43.5%)
Blo t 9	Blo t 9.0101	BT_003514.02	Serine protease	2	272	7454±1549	AAP57077 Der p 9 (57.5%)
	Blo t 9.0201	BT_016173.01	Serine protease	3	274	368±156	AAP57077 Der p 9 (53.8%)
	Blo t 9.0301	BT_001003.01	Serine protease	2	288	173±58	AAP57077 Der p 9 (49.8%)
Blo t 10	Blo t 10.0101*	BT_015321.01	Tropomyosin	9	284	12444±1051	ABU97466 Blo t 10.0101 (99.6%)
	Blo t 10.0201	BT_006475.01	Tropomyosin	9	353	2500±863	ABU97466 Blo t 10.0101 (59.1%)
Blo t 11	Blo t 11.0101*	BT_008464.01	Paramyosin	10	875	3119±379	AAM83103 Blo t 11.0101 (99.7%)
Blo t 12	Blo t 12.0101*	BT_016677.01	Chitin-binding domain containing protein	2	143	19241±2339	AAA78904 Blo t 12.0101 (92.4%)
Blo t 13	Blo t 13.0101*	BT_008106.02	Fatty acid-binding protein	4	130	136948±13042	AAC80579 Blo t 13.0101 (99.2%)
	Blo t 13.0201	BT_008106.03	Fatty acid-binding protein	2	131	32933±8181	AAC80579 Blo t 13.0101 (60.8%)
	Blo t 13.0301	BT_005106.01	Fatty acid-binding protein	3	131	16078±7775	AAC80579 Blo t 13.0101 (56.3%)
Blo t 14	Blo t 14.0101	BT_005757.01	Apolipophorin	6	1700	9095±2545	AF373221_1 Der p 14 (46.3%)
Blo t 15	Blo t 15.0101	BT_016675.01	Chitinase	6	575	8653±2836	AAY84565 Der p 15 (59.8%)
Blo t 16	Blo t 16.0101	BT_003239.01	Gelsolin/villin	8	484	767±284	AAM64112 Der f 16 (57.0%)
Blo t 18	Blo t 18.0101	BT_010400.01	Chitinase	4	461	12893±7857	AAM19082 Der f 18 (60.9%)

Blo t 20	Blo t 20.0101	BT_008259.02	Arginine kinase	3	357	9153±1347	ABU97470 Der f 20 (86.0%)
	Blo t 20.0201	BT_008259.03	Arginine kinase	4	356	3159±955	ABU97470 Der f 20 (83.1%)
	Blo t 20.0301	BT_012279.01	Arginine kinase	2	358	134±10	ABU97470 Der f 20 (68.2%)
Blo t 21	Blo t 21.0101*	BT_005892.01	Unknown	2	129	105276±23261	AAX34047 Blo t 21.0101 (99.2%)
Blo t 23	Blo t 23.0101	BT_016256.01	Peritrophin-like protein	1	113	4870±2707	ACB46292 Der p 23 (40.0%)
Blo t 24	Blo t 24.0101	BT_008155.01	UQCRB protein	3	118	9069±1578	AGI78542 Der f 24 (72.9%)
Blo t 25	Blo t 25.0101	BT_013681.01	Triosphosphate isomerase	6	419	6001±1630	AIO08860 Der f 25 (80.9%)
Blo t 26	Blo t 26.0101	BT_007095.02	Myosin light-chain	6	158	26093±5323	AIO08852 Der f 26 (82.7%)
Blo t 28	Blo t 28.0101	BT_002686.01	Heat shock protein 70	1	650	1895±644	AIO08848 Der f 28 (80.5%)
	Blo t 28.0201	BT_001717.01	Heat shock protein 70	1	647	1310±121	AIO08848 Der f 28 (80.2%)
	Blo t 28.0301	BT_000218.01	Heat shock protein 70	1	650	16851±1947	AIO08848 Der f 28 (79.4%)
	Blo t 28.0401	BT_003376.01	Heat shock protein 70	1	658	10802±517	AIO08848 Der f 28 (79.2%)
	Blo t 28.0501	BT_011597.01	Heat shock protein 70	1	635	23005±14241	AIO08848 Der f 28 (78.0%)
Blo t 29	Blo t 29.0101	BT_014409.01	Cyclophilin	2	254	12008±686	AAP35065 Der f 29 (83.5%)
Blo t 30	Blo t 30.0101	BT_008038.02	Ferritin	3	174	27150±883	AGC56219 Der f 30 (74.3%)
	Blo t 30.0201	BT_008039.02	Ferritin	2	276	15119±16669	AGC56219 Der f 30 (66.1%)
	Blo t 30.0301	BT_008039.03	Ferritin	2	173	2487±566	AGC56219 Der f 30 (75.5%)
Blo t 31	Blo t 31.0101	BT_016210.01	Cofilin	4	147	16453±4806	AIO08870 Der f 31 (80.1%)
Blo t 32	Blo t 32.0101	BT_003614.02	Inorganic pyrophosphatase	2	296	1544±335	AIO08849 Der f 32 (67.7%)

Blo t 33	Blo t 33.0101	BT_003250.01	Alpha-tubulin	1	450	9708±578	AIO08861 Der f 33 (81.3%)
	Blo t 33.0201	BT_000606.02	Alpha-tubulin	3	450	290±76	AIO08861 Der f 33 (82.0%)
	Blo t 33.0301	BT_005979.01	Alpha-tubulin	1	448	2357±496	AIO08861 Der f 33 (81.1%)
	Blo t 33.0401	BT_002245.02	Alpha-tubulin	1	449	12065±172	AIO08861 Der f 33 (79.4%)
Blo t 34	Blo t 34.0101	BT_004919.01	Rid-like protein	1	136	8805±900	BAV90601 Der f 34 (58.4%)
Blo t 36	Blo t 36.0101	BT_000325.01	Unknown	3	221	13971±3133	ATI08931 Der f 36 (53.4%)
Blo t 37	Blo t 37.0101	BT_008575.02	Chitin binding protein	2	190	9952±2395	QBF67839 Der f 37 (50.9%)
	Blo t 37.0201	BT_008575.03	Chitin binding protein	2	186	2977±1032	QBF67839 Der f 37 (37.1%)
Blo t 38	Blo t 38.0101	BT_009281.01	Bacteriolytic enzyme	1	153	5043±2717	QHQ72282 Der f 38 (56.2%)
Blo t 39	Blo t 39.0101	BT_015114.01	Troponin C	5	153	7377±2066	QBF67841 Der f 39 (95.4%)
Blo t 40	Blo t 40.0101	BT_000785.01	Thioredoxin-like protein	3	105	49617±18539	UJQ69787 Der f 40 (71.7%)
Blo t 41	Blo t 41.0101	BT_002783.03	Aldehyde dehydrogenase	3	483	1206±352	AOD75396.1 Tyr p 35 (85.4%)
	Blo t 41.0201	BT_002783.02	Aldehyde dehydrogenase	3	481	8888±2370	AOD75396.1 Tyr p 35 (73.3%)
Blo t 42	Blo t 42.0101	BT_008117.01	Profilin	3	130	16035±11211	AOD75399 Tyr p 36 (94.7%)

* These genes were suggested as the identical allergen genes by the high identity to those reported in the WHO/IUIS Allergen Nomenclature database. All other genes are putative allergen genes.

Supplementary data

Tables S1–2

Figures S1–12

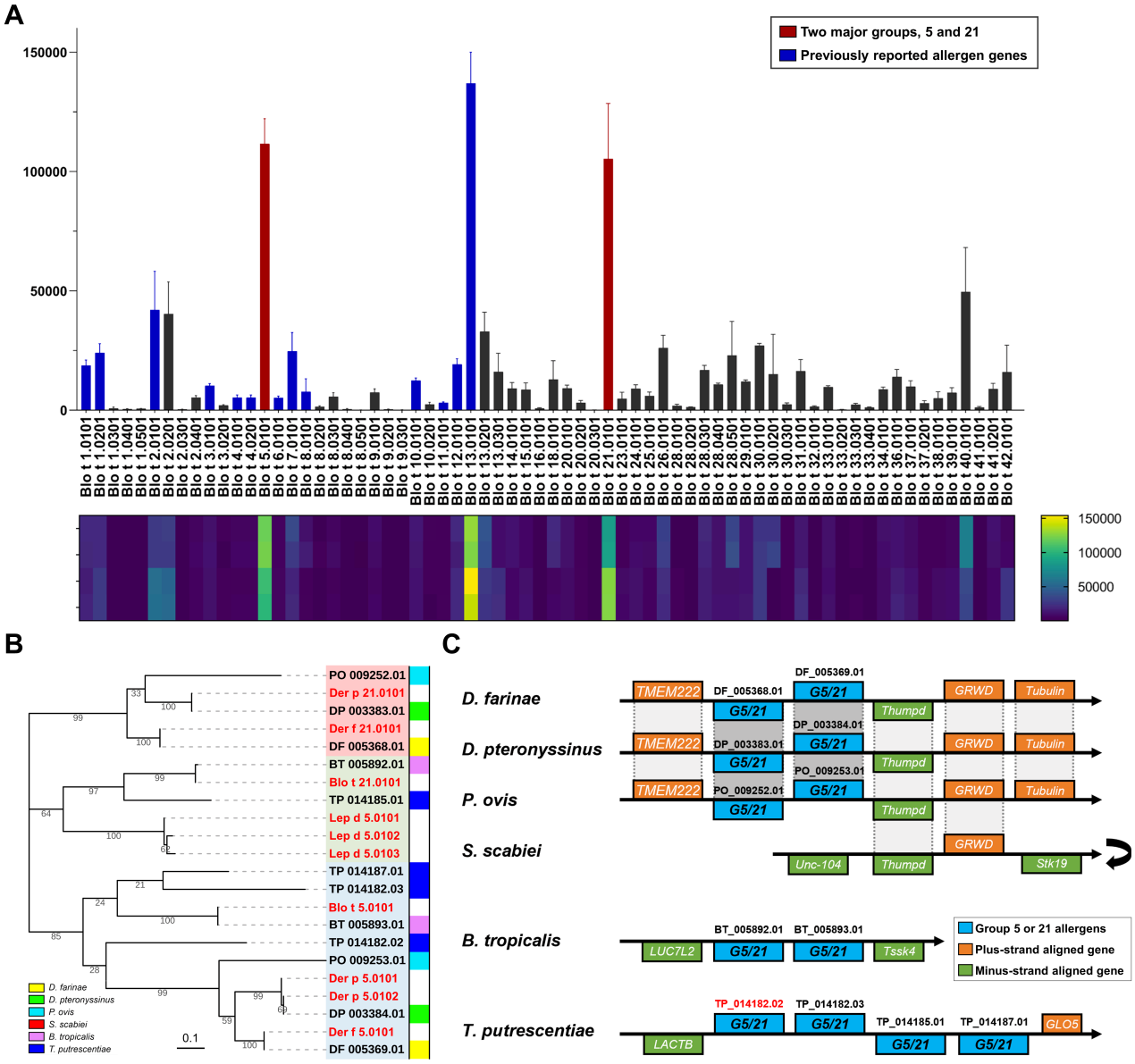

Figure 1_Xiong et al.

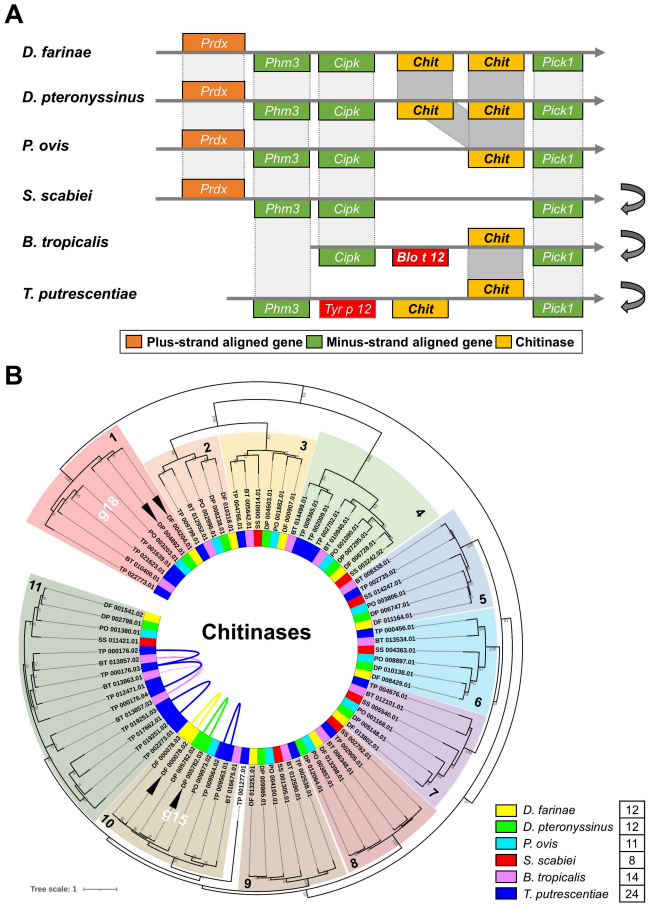

Figure 2_Xiong et al.

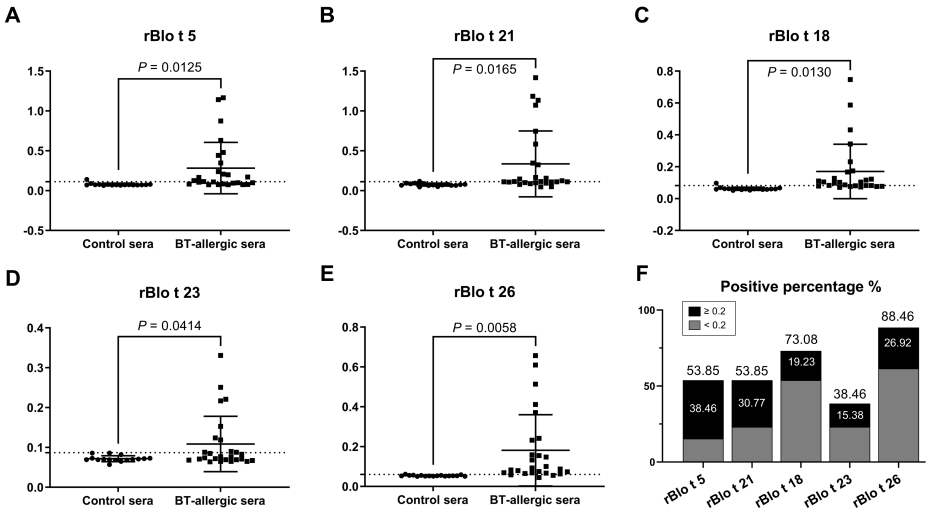

Figure 3_Xiong et al.