

ORIGINAL ARTICLE

Predicting *Blomia tropicalis* allergens using a multiomics approach

Jan Hubert^{1,2}  | Susanne Vrtala³  | Bruno Sopko¹  | Scot E. Dowd⁴  |
Qixin He⁵  | Pavel B. Klimov⁵  | Karel Harant^{6,7}  | Pavel Talacko⁶  |
Tomas Erban¹ 

¹Crop Research Institute, Prague, Czechia

²Department of Microbiology, Nutrition and Dietetics, Faculty of Agrobiological Sciences Prague, Prague, Czechia

³Department of Pathophysiology and Allergy Research, Center for Pathophysiology, Infectiology and Immunology, Medical University of Vienna, Vienna, Austria

⁴MR DNA (Molecular Research LP), Shallowater, Texas, USA

⁵Purdue University, Lilly Hall of Life Sciences, West Lafayette, Indiana, USA

⁶Proteomics Core Facility, Faculty of Science, Charles University, BIOCEV, Vestec, Czechia

⁷Institute for Environmental Studies, Faculty of Science, Charles University, Prague, Czechia

Correspondence

Jan Hubert, Crop Research Institute, Drnovska 507/73, CZ-161 06, Prague 6-Ruzyně, Czechia.
Email: hubert@vurv.cz

Funding information

Ministry of Science and Higher Education of the Russian Federation, Grant/Award Number: 193021X0012; Ministry of Education, Youth and Sports of the Czech Republic, Grant/Award Number: LUAUS23082

Abstract

Background: The domestic mite *Blomia tropicalis* is a major source of allergens in tropical and subtropical regions. Despite its great medical importance, the allergome of this mite has not been sufficiently studied. Only 14 allergen groups have been identified in *B. tropicalis* thus far, even though early radioimmuno-electrophoresis techniques (27 uncharacterized allergen complexes) and comparative data based on 40 allergen groups officially recognized by the World Health Organization (WHO)/IUIS in domestic astigmatid mites suggest the presence of a large set of additional allergens.

Methods: Here, we employ a multiomics approach to assess the allergome of *B. tropicalis* using genomic and transcriptomic sequence data and perform highly sensitive protein abundance quantification.

Findings: Among the 14 known allergen groups, we confirmed 13 (one WHO/IUIS allergen, Blo t 19, was not found) and identified 16 potentially novel allergens based on sequence similarity. These data indicate that *B. tropicalis* shares 27 known/deduced allergen groups with pyroglyphid house dust mites (genus *Dermatophagoides*). Among these groups, five allergen-encoding genes are highly expressed at the transcript level: Blo t 1, Blo t 5, Blo t 21 (known), Blo t 15, and Blo t 18 (predicted). However, at the protein level, a different set of most abundant allergens was found: Blo t 2, 10, 11, 20 and 21 (mite bodies) or Blo t 3, 4, 6 and predicted Blo t 13, 14 and 36 (mite feces).

Interpretation: We report the use of an integrated omics method to identify and predict an array of mite allergens and advanced, label-free proteomics to determine allergen protein abundance. Our research identifies a large set of novel putative allergens and shows that the expression levels of allergen-encoding genes may not be strictly correlated with the actual allergenic protein abundance in mite bodies.

KEYWORDS

enzyme, genome, IgE, label-free proteomics, mites, transcriptome

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. Clinical and Translational Allergy published by John Wiley & Sons Ltd on behalf of European Academy of Allergy and Clinical Immunology.

1 | INTRODUCTION

House dust mites and domestic mites cause IgE-mediated type I allergic conditions in an estimated 30% of the world's population.¹ In sensitized individuals, allergen exposure induces the cross-linking of high-affinity FcεRI receptor-bound IgE on effector cells, leading to the release of mediators that cause allergic symptoms.² In severe cases, mite allergies manifest as chronic diseases, such as allergic rhinitis, conjunctivitis, and bronchial asthma.³ Currently, 40 allergens of astigmatid mites are recognized by the Allergen Nomenclature Subcommittee of the World Health Organization (WHO) and the International Union of Immunological Societies (IUIS)⁴; however, allergens are not studied across all domestic astigmatid mite species to the same extent.

Blomia tropicalis is one of the most common and important house dust mite species inhabiting human houses in tropical or subtropical areas.^{5–9} Within its typical range, this mite species can be found in up to 96% of dust samples, with a 40% prevalence among all mite species.⁷ Two *B. tropicalis* allergens are considered major allergens, rBlo t 5 and rBlo t 21, which show IgE reactivity to 93% and 89% of sera from adult asthma patients, respectively.¹⁰ A recent study showed high IgE-binding frequencies for rBlo t 21 (43%) and rBlo t 2 (40%), whereas the other tested allergens were only recognized in the samples of a few patients (rBlo t 10, 6%; rBlo t 12, 3%).¹¹ Overall, 27 antigen-antibody precipitin complexes have been identified in *B. tropicalis* by crossed radioimmunoelectrophoresis.¹² However, out of 40 allergen groups known in domestic mites, only 14 allergen groups have been described in *B. tropicalis* thus far.^{4,13} Furthermore, 7 have been predicted based on their high similarity to identified allergens of other astigmatid mites.^{13,14} These data collectively indicate that *B. tropicalis* is likely to have a substantial number of unrecognized allergen groups.

One of the major challenges in molecular allergy research is the accurate prediction of the allergenic potential of proteins.¹⁵ Accurate diagnosis and targeted therapies for patients are critical for identifying allergenic mite species.¹⁴ Diagnosis is focused on single allergens of mites to determine the exact IgE sensitization pattern of a patient and distinguish between sensitization and cross-sensitization.¹⁴ Additional challenges associated with research linked to mite allergenicity are the lack of screening tools for quantifying all allergenic proteins in mite bodies and feces. Mite fecal pellets (FP) are arguably the most important allergy-inducing component, as they contain powerful digestive proteins.¹⁶ The feces accumulate over time in human-related environments and, because of their minute sizes, can easily become airborne, thus increasing the risk of contact with the skin and the lung epithelium when inhaled.¹⁶ Although only a small portion of the entire mite allergome may be present in mite feces,¹⁷ the quantification of these allergens is important. Mite feces may have distinct properties affecting sensitization and allergy induction in different ways in comparison to mites themselves. Using a proteomic approach, allergen abundance and diversity have been accurately estimated in several mite species, such as *Dermatophagoides farinae*,

Dermatophagoides pteronyssinus, and *Tyrophagus putrescentiae*^{18–20}; however, data on *B. tropicalis* are lacking. Earlier studies employing bottom-up gel-based proteomics relied on narrow subsets of allergen sequences available at those times.²¹

In this study, we generated both a high-quality de novo whole genome (long and short reads) and transcriptome of *B. tropicalis* and annotated them. Putative allergens were identified based on high sequence identities and structural similarities to known allergens⁴ and allergen-related proteins available in public databases. We also estimated the expression levels of known allergens/putative allergens in mite bodies and feces. An advanced proteome analysis using an Orbitrap Tribrid mass spectrometer and deduced allergen sequences was then performed to quantify allergic protein richness and abundance in mite bodies and feces.

2 | MATERIAL AND METHODS

2.1 | Sample of mites and feces

Laboratory cultures of *B. tropicalis* were obtained from the University of Cartagena, Colombia in 2008. Since then, the culture has been maintained under the rearing conditions of the Crop Research Institute, Prague, Czechia.²² The wheat-derived rearing diet powder contained flakes, wheat germ and Mauripan-dried yeast extract (AB Mauri, Balakong, Malaysia) (10:10:1 w/w). Mites were mass reared in tissue culture flasks on a house dust mite diet in a large population (approximately 4 weeks) when the diet was almost consumed. The detailed rearing and harvesting methodology is described elsewhere.²³ Live mites were collected from the plugs and surfaces of the flasks with a brush and placed in sterile microfuge tubes. "Pure" feces were obtained from the flasks after the mites and remnants of their bodies or eggs were removed and collected.¹⁹

The mite and fecal samples were weighed using a microbalance (Mettler-Toledo) to obtain samples with fresh weights in the range of 30–40 mg. For transcriptomics, mites were surface cleaned by placing them in 100% ethanol, followed by vortexing for 5 s and centrifugation at 13,000 × g for 1 min. The supernatant was replaced with a 1:10 bleach solution containing 5% sodium hypochlorite, and the samples were then mixed by vortexing for 5 s and centrifuged at 13,000 × g for 2 min. The last step was washing twice with dd H₂O to remove residual bleach. Surface sterilization was performed on ice. The provided rearing diet had the same weight as the feces. The rearing diet and feces were not cleaned and were stored directly in an ultracold freezer (–80°C). We collected five biological replicates of mites for RNA sequencing and 1 sample of mites for genomic sequencing. Each sample contained approximately 1000 mites.

For proteomic analyses, four sample types, each with three biological replicates, were prepared: (i) 1000 individually collected adult mites; (ii) pools of mites of different ages, including eggs; (iii) water extracts of feces; and (iv) detergent-buffer extracts of the remaining pellet of the water extract.

2.2 | Sample processing

Homogenization of mites was performed on ice. All samples were homogenized for 30 s in a glass tissue grinder (Kavalier glass, Prague, Czechia) in 500 μ L of lysis buffer from the kits. DNA was extracted from homogenates after overnight incubation with 20 μ L of proteinase K at 56°C using the QIAamp DNA Micro Kit (Qiagen, Hilden, Germany, cat. No. 56304) following the manufacturer's protocol for tissue samples. The concentration of the extracted DNA samples was quantified using the Qubit® dsDNA HS Assay Kit (Life Technologies), and the quality of the DNA was determined using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific). The average size of gDNA was determined on an E-Gel SizeSelect 2% agarose gel (Invitrogen) with a 1 kb ladder. The samples were sheared using a Covaris G-tube (Covaris Inc.). The average size of the sheared DNA was determined using an Agilent 2100 Bioanalyzer (Agilent Technologies).

RNA isolation was performed using a NucleoSpin RNA kit (catalog no. 740984.50; Macherey-Nagel, Duren, Germany) according to the manufacturer's instructions, with the following modifications: homogenized samples were centrifuged at 2000 $\times$ g for 3 s, and DNA was degraded by DNase I at 37°C according to the manufacturer's protocol (Riboclear plus, catalog no. 313-50; GeneAll, Lisbon, Portugal). RNA quality was evaluated using a NanoDrop (NanoDrop One; Thermo Scientific, Waltham, MA, USA) and an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), and RNA samples together with DNA samples were shipped on dry ice to the MrDNA laboratory (Shallowater, TX, USA) for downstream processing and sequencing.

For protein analyses, mite body samples (i and ii) were further homogenized in Potter-Elvehjem homogenizers with a Teflon pestle after the addition of 100 mM (1 mL/1000 mites) triethylammonium bicarbonate buffer (TEAB; Cat No. 17902, Fluka, Sigma-Aldrich, St. Louis, MO, USA) containing 2% w/v sodium deoxycholate (SDC; Cat No. 30970, BioXtra, Sigma-Aldrich). Water fractions of feces (iii) were obtained in tissue cell culture flasks as previously described.^{18–20} Briefly, 10 mL of ice-cold (4°C) Nanopure water (Thermo, Waltham, MA, USA) was added to the feces-coated flasks. The flasks were placed in a box with ice and mixed on a ProBlot Rocker (Labnet International, Woodbridge, NJ, USA) for 10 min laying on the side at 100 RPM, and the extract was transferred to 50 mL sterile flasks. Extraction of the second side of the feces-coated flask was performed with 10 mL of sterile water, and the extract was then added to the 50 mL flask. The third extraction in the flask was performed with 10 mL of water, which was thoroughly mixed within the flask. In total, 30 mL of the extract obtained from a flask was centrifuged at 10,000 $\times$ g for 15 min at 4°C (Thermo). The supernatant was divided into 3 equal portions, and portions of 1 mL were collected for total protein quantification. The contents were lyophilized in a PowerDry LL3000 lyophilizer (Thermo Fisher Scientific). The pellets of feces (iv) used for proteomic analysis were obtained by processing the pellet after water extraction. Three milliliters of 100 mM TEAB with 2% SDC was added to the pellet that remained from approximately 20 mL of the supernatant, and

the contents of the centrifuge tube were thoroughly mixed by vortexing plus 10 min of mixing on a rocker on ice. The procedure was repeated 3 times. The resulting sample was centrifuged at 10,000 $\times$ g for 15 min at 4°C (Thermo Fisher Scientific), and the supernatant was collected for further analysis. All samples were stored at –80°C until use.

2.3 | Sequencing and assemblage

For Illumina DNA sequencing, the libraries were prepared using a Nextera DNA Flex library preparation kit (Illumina) following the manufacturer's user guide. Fifty nanograms of DNA was used to prepare the libraries. The samples simultaneously underwent fragmentation and addition of adapter sequences. These adapters were utilized during limited-cycle PCR in which unique indices were added to the samples. The libraries were then pooled in equimolar ratios of 0.6 nM and subjected to paired-end sequencing for 500 cycles using a NovaSeq 6000 system (Illumina).

For PacBio sequencing, 250 ng of the sheared DNA was employed as the input for library preparation using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences). During library preparation, each sample underwent DNA damage and end repair as well as barcode adapter ligation. To reduce the number of smaller DNA fragments (<3 kb), an additional bead purification step was performed on the SMRTbell libraries using 40% diluted AMPure PB Beads. Each library was then sequenced using a 10-h movie time on a PacBio Sequel system (Pacific Biosciences). The SMRT Link Circular Consensus Sequencing workflow (SMRT Link v.9.0.0, CCS) was used to combine multiple subreads from the same molecule to generate a highly accurate consensus sequence. The sample metadata and the whole genome sequence were deposited in GenBank as part of project PRJNA625856.

The concentration of total RNA was determined using the Qubit® RNA Assay Kit (Life Technologies). Poly-A selection and library preparation were performed by using KAPA mRNA HyperPrep Kits (Roche) following the manufacturer's instructions. A total of 1000 ng RNA was used to prepare the library. Following library preparation, the final concentration of the library was measured using the Qubit® dsDNA HS Assay Kit (Life Technologies), and the average library size was determined using an Agilent 2100 Bioanalyzer (Agilent Technologies). The library was then diluted to 0.6 nM and subjected to paired-end sequencing for 500 cycles using the NovaSeq 6000 system (Illumina). The samples are deposited in GenBank as project PRJNA599071.

Illumina reads were trimmed in Trim Galore,²⁴ corrected in fastQC,²⁵ and aligned with the PacBio reads in hybrid SPADES v 3.14^{26,27} for DNA-based reads and rnaSPADES for RNA-based reads.²⁶ Transcriptome sequences were annotated by Prokka,²⁸ and predicted proteins were identified by KEGG analysis in Ghost-Koala.²⁹ Reads were mapped onto the assembled genome or transcriptome using Bowtie2^{30,31} and Minimap2³² for long sequences. Then, the assembled genomes were improved by Pilon (Walker et al.,

2014). The transcriptome is presented in Table S1. The GFF file from the genome and transcriptome was prepared by EXONERATE.³³ Expression was performed in CLC Worbench 22 (Qiagen, Venlo, Netherlands) using the total number of mapped reads as the expression value.

2.4 | Protein analyses

Samples for label-free mass spectrometry were processed and further analyzed using a nanoLC-MS/MS system employing an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific) as previously described.^{34,35} Mass spectrometry data were evaluated in MaxQuant version 2.2.0.0 using label-free quantification algorithms^{36,37} and the Andromeda search engine.³⁸ The key criteria used in data analyses were a false discovery rate of 0.01 for proteins and peptides, a minimum length of 7 amino acids, carbamidomethyl as a fixed modification, and variable modifications of N-terminal protein acetylation and methionine oxidation. These data were searched using a custom database with our transcriptome assemblies of *B. tropicalis*. In addition, we included a GenBank genomic sequence of *B. tropicalis* (NCBI: 16792seq; 06 Feb 2023) and sequences of possible associated contaminants and microorganisms: *Aspergillus* (UniProtKB: 5439 seq 07 Feb 2023), *Staphylococcus kloosii* (UniProtKB 4243seq: 07 Feb 2023), *Staphylococcus rev* (UniProtKB: 13547seq, 07 Feb 2023), *Staphylococcus xylosus* (UniProtKB: 12,883, 07 Feb 2023), and yeast (UniProtKB: 21667seq, 07 Feb 2023). The data were processed in Perseus version 2.0.7.0.³⁹

2.5 | Allergen identification

Sequences of allergens were downloaded from the WHO/IUIS database⁴ and subjected to protein BLAST searches against the GenBank database using BLASTp.⁴⁰ The most similar search results for *B. tropicalis* and other mites were added to extend the mite WHO/IUIS allergen database. Then, the predicted proteins from *B. tropicalis* were screened against the allergen database using HMMER⁴¹ on the Galaxy server. Predicted sequences (Supplementary material Table S2) with the highest scores were selected for analyses and compared to the proteome identification (see above).

In the next step, we performed structural alignment on the TCoFee server^{42,43} for each allergen group. For homologous sequence clusters, phylogenetic trees were inferred with PhyML.⁴⁴ The trees were visualized and edited in FigTree v.1.4.4,⁴⁵ and alignments were visualized in ESPrint.⁴⁶ Sequence identities were calculated in EMBOSS Water using the Smith–Waterman algorithm.⁴⁷ The protein structure, including the signal peptide, was predicted based on comparison to reference proteins in the EMBL database.⁴⁸

Trypsin identification was based on BLASTp, and we searched for homologous PDB structures. Then, JGLJHAHI_02083, JGLJHAHI_03207 and JGLJHAHI_04704 were aligned to the structures with the results of BLASTp searches using CLUSTALX 2.0.⁴⁹ The protein

alignment was then used as an input for MODELLER 10.2 software^{50,51} for 3D structure prediction. The validity of the resulting models was checked using PROCHECK software (G-score),⁵² and the model with the best valid score was chosen for further investigation. This model was optimized by running 10 steepest descent energy minimization steps followed by 10 conjugated gradients, 20 steepest descent, 10 conjugated gradients and finally 10 steepest descent steps (GROMOS 96 force field).⁵³ The validity of the final model was again checked by PROCHECK.

2.6 | Data analyses

Expression and protein heatmaps were created in R-4.2.2 software⁵⁴ using the Complex-Heatmap package.^{55,56} The Protein profiles of feces and mite bodies were compared using analysis of similarities (ANOSIM). Two distance matrices were chosen: the Jaccard index (presence/absence of proteins) and the Bray–Curtis index (protein hit intensity). To identify allergens/predicted allergens responsible for the differences between the profiles, we used similarity percentages (SIMPER). The calculations for all of these analyses were performed in PAST 4.⁵⁷

3 | RESULTS

3.1 | Genome, transcriptome and proteome of *B. tropicalis*

The *Blomia tropicalis* predicted genome (GenBank SUB13543831) included 1601 contigs of 61 Mb in length ($N_{50} = 223$ kB). According to BUSCO prediction based on the Arachnida set, the completeness was estimated to be 93.2%, with 2.6% duplicated genes. The annotated transcriptome had 7171 contigs, with a total length of 31 Mb ($N_{50} = 5996$ bp) and 18,164 predicted genes (Supplementary material Table S1). Its completeness was estimated at 95%, with 9.1% duplicated genes based on BUSCO with the Arachnida dataset. In contrast, a previously generated and annotated transcriptome included fewer predicted genes (16,590–14,899).^{58,59}

The proteome analyses identified 3714 protein hits in four protein sample types individually collected from adult mites; pools of mites of different ages, including eggs; water extracts of feces; and detergent-buffer extracts of the remaining pellets of the water extracts.

3.2 | Identification of known allergens and putative allergens in the mite transcriptome and proteome

We inferred a total of 31 allergen/putative allergen-encoding proteins (Table 1, Supplementary material Tables S2; Supplementary Figures S1–S50), among which 13 proteins were confirmed WHO/IUIS allergens (groups 1–8, 10–13, 21), 17 proteins represented 16

TABLE 1 Known and predicted allergen groups in *Blomia tropicalis* based on structural identity analyses (PHMMER) of transcriptomic data.

Allergen group	Domain	WHO/IUIS allergens	Transcriptome No	Others (GenBank)
1	SP, cathepsin pro-peptide inhibitor, Papain family cysteine protease	AAK58415 ⁶⁰ ; AAQ24541 ⁶¹	JGLJHAHI_03943	KAI2799130; ⁵⁹ WBV73454; ⁶² WBV73455 ⁶²
2	SP, ML_domain	AAQ73483 ⁶³ ; AAQ73482 ⁶³ ; AAQ73481 ⁶³	JGLJHAHI_04625	ABG76185 ^{64,65} ; ABG76186 ^{64,65} ; ABG76187 ^{64,65} ; ABG76188 ^{64,65} ; ABG76189 ^{64,65} ; ABG76190 ^{64,65} ; ABG76191 ^{64,65} ; ABG76192 ^{64,65} ; ABG76193 ^{64,65} ; KAI2804433 ⁵⁹
3	SP, trypsin	AAQ24542 ⁶⁶ AAM10779 ⁶⁷	JGLJHAHI_02083; JGLJHAHI_03207	AAM10779 ⁶⁷ ; KAH9390279 ⁵⁹ ; KAI2797715 ⁵⁹ ; KAI2802709 ⁵⁹ ; KAI2805554 ⁵⁹ ; WBV73464 ⁶²
4	SP, Alpha amylase, catalytic domain and alpha amylase, C-terminal 2all-beta domain	AAQ24543 ⁶⁸	JGLJHAHI_01112	KAI2795383 ⁵⁹ ; KAI2795737 ⁵⁹ ; KAI2801484 ⁵⁹ ; KAI2803944 ⁵⁹
5	SP, Blo_t_5 domain	AAD10850 ⁸	JGLJHAHI_14490	AAB49396 ⁶⁹ ; ABH06351 ⁷⁰ ; ABH06352 ⁷⁰ ; ABH06353 ⁷⁰ ; ABH06354 ⁷⁰ ; ABH06355 ⁷⁰ ; ABH06356 ⁷⁰ ; ABH06357 ⁷⁰ ; ABH06358 ⁷⁰ ; ABH06359 ⁷⁰ ; APU87554 ⁷¹ ; APU87555 ⁷¹ ; APU87556 ⁷¹ ; APU87557 ⁷¹ ; APU87558 ⁷¹ ; KAI2807952 ⁵⁹ ; WBV73467 ⁶²
6	SP, trypsin	AAQ24544 ⁷²	JGLJHAHI_04704	WBV73468 ⁶² ; KAI2808564 ⁵⁹
7	SP, Group 7 allergen	ASX95438 ⁷³	JGLJHAHI_05732; JGLJHAHI_09763	AAQ24545 ⁷⁴ ; KAI2808329 ⁵⁹ ; WBV73469 ⁶²
8	Glutathione S-transferase	ACV04860 ⁷⁵	JGLJHAHI_17878; JGLJHAHI_15972	AAP35069 ⁷⁶ ; KAI2802743 ⁵⁹ ; KAI2801258 ⁵⁹
10	Tropomyosin	ABU97466 ⁷⁷	JGLJHAHI_15939	KAI2797771 ⁵⁹
Group	Domain	Known allergens	Predicted	Supported sequences
11	Myosin tail 1	AAM83103 ^{107,108}	JGLJHAHI_05065	KAI2806507 ⁵⁹
12	SP, chitin binding Peritrophin-A	AAA78904 ⁸³	JGLJHAHI_10661	KAI2795557 ⁵⁹ WBV73481 ⁶²
13	Lipocalin/cytosolic fatty-acid binding protein family	AAC80579 ⁷⁸	JGLJHAHI_00046; JGLJHAHI_14091; JGLJHAHI_12788	AAP35071 ¹⁰⁹ ; KAI2806095 ⁵⁹ ; WBV73483 ⁶² ; WBV73484 ⁶²
14	SP, lipoprotein amino terminal region, vitellinogen, open beta-sheet		JGLJHAHI_08651	KAI2807817 ⁵⁹ ; WBV73485 ⁶²
15	SP, Glycosyl hydrolases family 18 and chitin binding Peritrophin-A domain		JGLJHAHI_17563	KAI2795555 ⁵⁹ ; WBV73486 ⁶²
16	Gelsolin repeat units		JGLJHAHI_10134	KAI2804260 ⁵⁹ ; WBV73487 ⁶²
18	SP, glycosyl hydrolase family 18, chitin binding Peritrophin-A		JGLJHAHI_08029	AAQ24549 ¹¹⁰
19	Nematode Antimicrobial peptide	AHG97583	No matches	No matches
20	ATP: Guanido phosphotransferase		JGLJHAHI_13643	KAI2800065 ⁵⁹ ; WBV73490 ⁶²
21	SP, mite allergen Blo t 5	AAX34047 ¹¹¹	JGLJHAHI_14491	ABH06346 ⁷⁰ ; ABH06347 ⁷⁰ ; ABH06348 ⁷⁰ ; ABH06350 ⁷⁰ ; WBV73492 ⁶²
24	Ubiquinol-cytochrome C reductase complex		JGLJHAHI_12583	KAI2804014 ⁵⁹ ; WBV73494 ⁶²
25	Triosephosphate isomerase		JGLJHAHI_16837	KAI2796809 ⁵⁹

(Continues)

TABLE 1 (Continued)

Group	Domain	Known allergens	Predicted	Supported sequences
26	Unknown structure		JGLJHAHI_10008	KAI2802972 ⁵⁹
28	Hsp70 protein		JGLJHAHI_14735; JGLJHAHI_17137	AAQ24552 ¹¹² ; KAI2799691 ⁵⁹ ; KAI2809076 ⁵⁹ ; WBV73501 ⁶²
29	Cyclophilin type peptidyl-prolyl cis-trans isomerase/CLD		JGLJHAHI_13666	KAI2796616 ⁵⁹
30	Ferritin-like domain		JGLJHAHI_17697	KAI2803899 ⁵⁹ ; WBV73503 ⁶²
32	Inorganic pyrophosphatase		JGLJHAHI_11580	KAI2804627 ⁵⁹ ; WBV73507 ⁶²
33	Tubulin		JGLJHAHI_03926; JGLJHAHI_05719	KAI2804271 ⁵⁹ ; WBV73508 ⁶²
34	EF-hand domain pairs		JGLJHAHI_15642	KAI2797062 ⁵⁹ ; WBV73517 ⁶²
35	Aldehyde dehydrogenase family		JGLJHAHI_01024	
36a	SP, peptides with C2 domain		JGLJHAHI_06496; JGLJHAHI_09342	KAI2809178 ⁵⁹ ; KAI2801606 ⁵⁹ ; WBV73513 ⁶²
36b	Profilin		JGLJHAHI_14665	AAQ24553 ¹¹³ ; KAI2803977 ⁵⁹ ; WBV73521 ⁶²
nd	60S acidic ribosomal protein P2		JGLJHAHI_08405	AAQ24548 ^{a,114} ; KAI2795631 ^{a,59}

Note: Confirmed allergens from the WHO/IUIS allergen database and the closest GenBank matches (annotated or unannotated) are given. Full analysis (allergen and nonallergen proteins) and alignments are given in Supplementary results and Supplementary figures 1–50, respectively.

Abbreviation: SP, signal peptide.

^aBased on proteome analysis AAQ24548 and KAI2795631 did not correspond to JGLJHAHI_08405.

allergen groups known in other mites (14–16, 18, 20, 24–26, 28–30, 32–35, 36) but not medically evaluated in *B. tropicalis*, and 1 protein (60S acidic ribosomal protein P2) represented a putative new allergen group²¹ (Table 1: nd, Supplementary material Tables S3–S32). The group 36 allergens incorrectly included two different structural groups (Table 1: group 36a,b). The groups of allergens (1, 2, 4, 5, 7, 14, 21) showed significant differences in protein identity based on mite species comparisons (Table 2). The identity of *B. tropicalis* allergens or putative allergens significantly differed from the remaining mite species in groups 1, 2, 3, 4, 5, 14, and 28.

For each of these allergens/putative allergens, protein alignments and phylogenetic trees (*B. tropicalis* and other mites) are presented in the Supplementary results (Supplementary Figures S1–S50). *Blomia tropicalis* showed the greatest number of WHO/IUIS allergens shared with pyroglyphid house dust mites *D. farinae* and *D. pteronyssinus* (34%) as opposed to the mold mite *T. putrescentiae*. Twenty-six allergens of *B. tropicalis* were shared with those of the pyroglyphid house dust mite genus *Dermatophagoides* (Figure 1).

The presence of Blo t 19, a WHO/IUIS allergen previously identified in *B. tropicalis*, could not be confirmed by our genome/transcriptome or GenBank data. Originally, the Blo t 19 (GenBank: AHG97583) protein was reported as a putative *B. tropicalis* protein with similarities to the nematode antimicrobial peptide. Our PHMMER search of this peptide did not return any mite matches; all of the closest matches came from nematodes: 81.7%–85.5% amino acid similarity: *Toxocara canis* KHN74506), *Halicephalobus* sp.

KAE9556153, and *Enterobius vermicularis* VDD96638). We therefore consider this previously reported mite allergen to be an artifact/contamination.

Group 36 contained two different allergens incorrectly assigned as proteins with the same function (see also Vrtala¹⁴). The Der p 36 (ATI08932) and Der f 36 (ATI08931) peptides showed a C2 domain (1–24 signal peptide, 90–208 C2 domain). The identity of the peptides was 77.6%, while Tyr p 36 had a profilin (alignment region 1–131) structure. Both peptides with C2 domains and profilin were present in *B. tropicalis*.

In this study, we found two Blo t 3 isoforms (Table 1; Supplementary material Figures S5–S6), but we did not identify an isoform connected to Blo t 3 AAQ24542 in the genome or transcriptome. However, we identified the AAQ24542 protein in the proteome^{58,59} (Table 1; Supplementary material Figures S5–S6). Isoform AAQ24542 was not detected in transcripts. PHMMER analyses showed that several allergen groups shared the same functional domains (Table 1). The allergens Blo t 3 and Blo t 6 both had a trypsin domain. The predicted 3D structures confirmed the trypsin-like core structure of JGLJHAHI_02083, JGLJHAHI_03207 and JGLJHAHI_04704 (Figure 2). Except for the trypsin-like core structure, the remaining sequences were classified into different groups. Blo t 3 JGLJHAHI_02083 exhibited similarity to factor IX of the blood clotting cascades. This protein was identified only in the body, not in excrement. Blo t3 JGLJHAHI_03207 was identified as excrement and was highly similar to trypsinogen. The protein structure indicated

TABLE 2 The comparison of sequence identity was based on PERMANOVA (Bray–Curtis, 1000 permutations), and the factors were species of mites or *Blomia tropicalis* versus other species. The similarity value levels are indicated by color.

Factor group	Species			<i>B. tropicalis</i> x other species		
	F	P	MIN ^{#1}	F	P	MIN ^{#2}
1	22.44	<0.001	30.4	33.66	<0.001	34.9
2	1291	<0.001	38.5	197	<0.001	86.6
3	1.204	0.100	45.000	3.58	0.004	49.6
4	84.22	0.005	62.9	49.85	0.001	92.7
5	132.2	0.008	42.4	243.1	0.009	78.1
6	0.00029	0.198	55.6	8.509	0.065	99.6
7	5.836	0.009	24.7	8.116	0.01	43.3
8	1.573	0.127	32.8	4.052	0.011	34.4
10	0.7165	0.689	93.3	1.006	0.4748	94
11	nd	nd	86.2	3.568	0.066	91.5
12	nd	nd	nd	nd	nd	91.2
13	0.674	0.992	51.9	0.8726	0.409	52.1
14	862.1	0.017	43	584.2	0.018	99.1
15	1696	0.052	61.4	250	0.1024	98.3
16	nd	nd	56.5	nd	nd	99.6
18	nd	nd	60.5	39.07	0.3235	99.6
20	0.4785	0.74	67.3	0.4892	0.5234	68.3
21	0.00029	0.022	38.5	1180	0.0216	98.4
24	1794	0.1	72	1794	0.1	99.2
25	36.64	0.203	35	64.99	0.105	35
26	0.0008	0.161	82.6	44.3	0.333	100
28	1.845	0.203	69.5	2.748	0.022	82.1
29	0.75	0.47	65	0.819	0.338	65
30	11.68	0.099	76.5	5.409	0.099	90.8
32	nd	nd	67.4	7.871	0.103	100
33	23.66	0.168	80.9	23.66	0.168	93.3
34	87.23	0.395	26.7 ^{#3}	1.35	0.407	99.3
35	nd	nd	73.3	nd	nd	nd
36a	2.3	0.203	43.6	2.3	0.203	52.3
36b	nd	nd	94.7	1.8	0.397	100

Note: The minimum sequence identity is provided and indicated by the color scale. #1 identity minimum from all sequences; #2 identity minimum for *Blomia tropicalis* sequences; #3 minimum caused by incorrectly included BAV90601.

Abbreviation: nd, not determined.

that trypsinogen is activated by the removal of the protein tail (Figure 2C). The final protein, Blo t 6 JGLJHAHI_04704, exhibited similarity to kallikrein-like serine protease. Both Blo t 5 and Blo t 21 shared the Blo t 5 domain, as confirmed by a 2D protein model

(Figure 3AB). The differences were caused by the protein tail of n Blo t 21 (Figure 3B).

3.3 | Expression of allergen-encoding genes and protein abundance

Our expression analysis identified six highly expressed allergen-encoding genes, Blo t 1, 5, 14, 15, 18, and 21 (Figure 4). However, a proteomic analysis identified only some of them as the most abundant, Blo t 14, 15, 18 (Figure 5). The proteomes of mite bodies and feces were significantly different in the presence/absence of the protein (ANOSIM Jaccard, $R = 0.509$, $P = 0.003$) and abundance (ANOSIM Bray–Curtis, $R = 0.677$, $P = 0.002$). The differences (SIMPER analysis) were due to the low abundance/absence of several proteins in mite feces: 60S acidic ribosomal protein P2 and Blo t 7, 8, 33, 34, which were represented in mite bodies. In our proteomic samples, alpha-amylase protein (Blo t 4) was highly abundant in both mite bodies and feces, indicating that this enzyme was present in the mite gut and then concentrated in feces. Blo t 20 (JGLJHAHI_13643) was absent in FP but showed intermediate abundance in water extracts from feces. Blo t 2 (JGLJHAHI_04625), Blo t 18 (JGLJHAHI_08029) and Blo t 29 (JGLJHAHI_136660) were more abundant in mite bodies than in feces.

In *B. tropicalis*, group 3 allergen (trypsin) had 3 isoforms (Figures S5 and S6), but only two of these isoforms were detected in feces at higher abundances than in whole mites (e.g., trypsinogen: JGLKHAHI_03207 and AAQ24542). Another isoform, Blo t 3 (IX factor of blood clotting cascade: JGLJHAMI_02083), was nearly absent in feces. Blo t 6 (kallikrein-like serine protease) presented a single isoform that was more abundant in feces than in mite whole bodies. Blo t 13 (lipocalin) showed 3 isoforms (Figure S19). One isoform (JGLKHAHI_14091) predominated in feces, while isoform JGLKHAHI_00046 had almost the same abundance in whole mite bodies and feces. Isoform JGLKHAHI_00101 was detected at low abundance in whole mite samples but not in feces.

4 | DISCUSSION

Our multiomic analyses confirmed 13 known groups of *B. tropicalis* allergens (Table 1) and suggested that one previously identified WHO/IUIS allergen (Blo t 19) belonged to a nematode and not to *B. tropicalis* or any other mite. This last allergen may have been observed because of contamination by a nematode inhabiting the mite culture medium.³⁴ Based on the structural homology and sequence identity of known WHO/IUIS allergens from the pyroglyphid house dust mite *Dermatophagoides* sp. and the mold mite *T. putrescentiae*, we predicted 16 putative allergens in *B. tropicalis*, groups 14–16, 18, 20, 24–26, 28–30, 32–35, 36 (Table 1). A total of 26 known/predicted *B. tropicalis* allergens were found to be shared with *Dermatophagoides* and *Tyrophagus* mites. Interestingly, a similar number of candidate allergens (27) was identified previously using

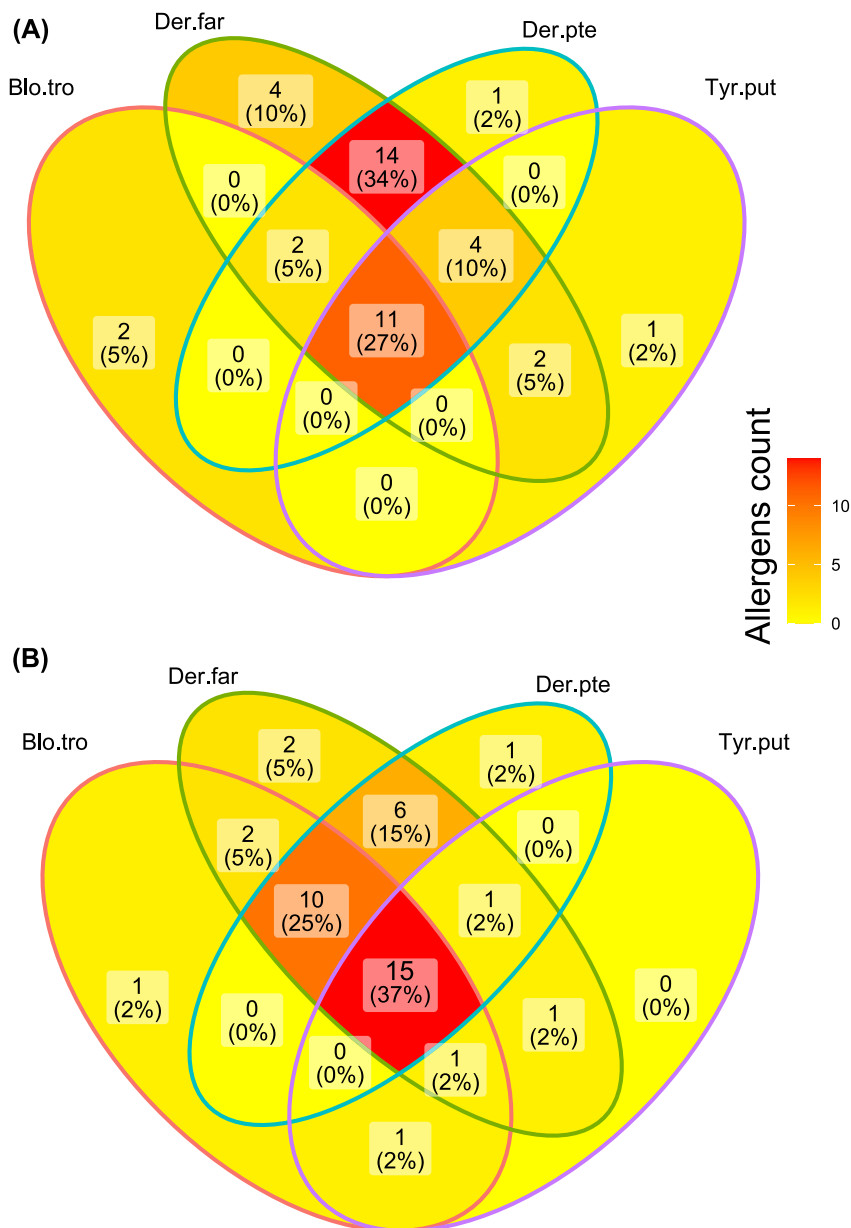


FIGURE 1 Venn diagrams of allergens only from the World Health Organization (WHO)/IUIS database (A) and WHO/IUIS allergens plus predicted allergens (B) in four species of mites: *Dermatophagoides farinae*, *Dermatophagoides pteronyssinus*, *Blomia tropicalis*, and *Tyrophagus putrescentiae*. The numbers of total and percentual numbers of allergens are showed. Legend: Blo.tro, *Blomia tropicalis*; Der.far, *Dermatophagoides farinae*; Der.pte, *Dermatophagoides pteronyssinus*; Tyr.put, *Tyrophagus putrescentiae*.

crossed radioimmuno-electrophoresis.¹² Many of these allergens/putative allergens (14) were also found in recently published genomic/transcriptomic⁵⁹ and proteomic analyses.²¹

Our study showed that Blo t 3, 8, 13, 33 36a (peptides with C2 domain) occurs in more isoforms; however, we did not confirm isoforms for Blo t 2. Blo t 2 allergen protein was abundant in mite bodies as well as FP (Figure 4). We found only a single isoform, which contrasts with the findings of Reginald et al.,⁶⁴ who reported nine Blo t 2 isoforms with 1–3 amino acid substitutions.⁶⁴ As these isoforms were obtained via RT-PCR/cloning and were not verified using direct DNA sequencing,⁶⁴ it is unknown whether these substitutions represent the natural variation of the protein, polymerase errors, or a mix of the two. However, the study by Reginald et al.⁶⁴ which use a high fidelity polymerase revealed that 8% of the examined Blo t 2 sequences showed the Blo t 2.0102 sequence (all with the same two

nucleotide substitutions). This cannot be a copy error and is consistent with the more extensive data for *Dermatophagoides* mites showing an evolutionary pattern of variation that differs in different regions and could be very important for the frequency of responses and for immunotherapy. A possible reason for not finding alleles is the use of a closed culture of mites.

A previous proteomic study revealed 3 isoforms with different protein profiles in *B. tropicalis* extracts.²¹ Lipocalin Blo t 13 is a minor allergen^{13,78} with 3 isoforms (our data). Some of these isoforms presented higher abundances in mite feces (Figure 4), as noted previously for *T. putrescentiae*.¹⁹ The study showed the existence of two isoforms of Blo t 1 with low similarity (35%) in *B. tropicalis*. We found evidence of two domains inside this protein: (i) a cathepsin propeptide inhibitor domain and (ii) Peptidase_C1. The differences were caused by the absence/presence of cathepsin propeptide inhibitors;

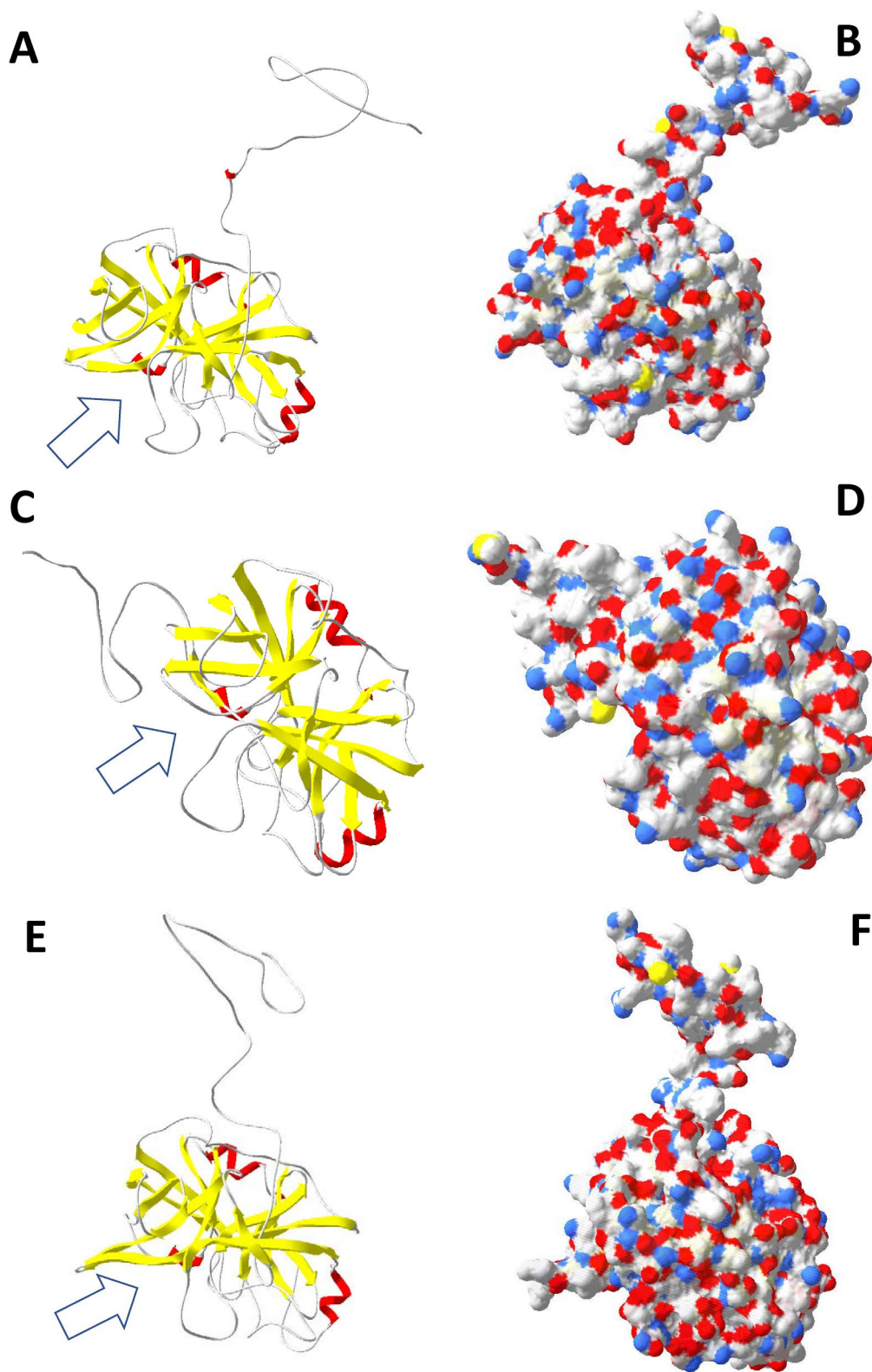


FIGURE 2 3D models of identified serine proteases factor IX of the blood clotting cascade (JGLJHAHI_02083) - (A and B); trypsinogen (JGLJHAHI_03207) - (C and D); kallikrein-like serine protease (JGLJHAHI_04704) - (E and F). Panels (A, C, and D) show the backbones. (B, D and F) show the surface. Positively charged surfaces are indicated in blue, and negatively charged surfaces are indicated in red. The active site is indicated by an arrow.

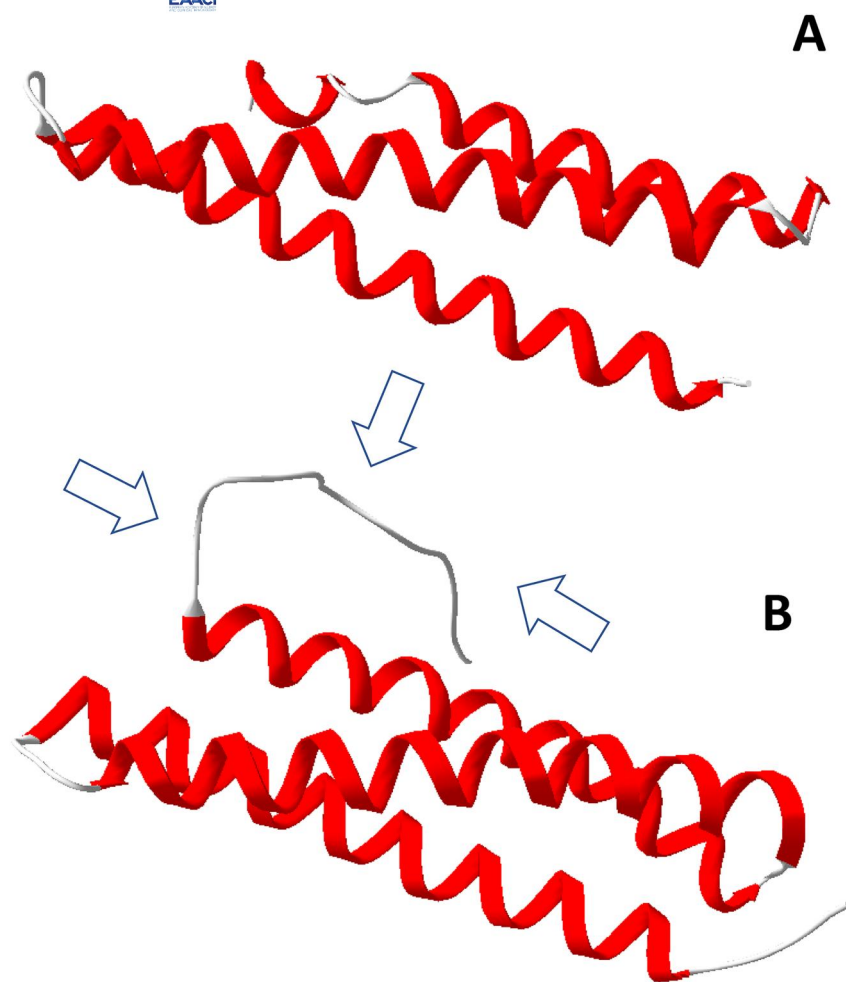


FIGURE 3 Predicted 2D model of allergen Blot 5 (A) and Blot 21 (B). Arrows point to the protein tail.

however, the biological functions and epitope binding of the isoforms should be solved in the future.

Blo t 14 allergen (vitellogenin)²¹ showed higher levels in both whole mite bodies and feces. In both *Dermatophagoides* mites, the allergen has been described as an apolipoprotein-like group 14 allergen, with predicted hydrophobicity and lipid-binding activity, inducing high levels of IgE responses and T-cell stimulation.^{79,80} This is an exoskeleton-bound protein with much higher abundance in females than in males of *D. farinae*.³⁴ In contrast to “exoskeleton-bound” proteins, vitellogenins are yolk protein precursors that are produced at extraovariole sites and taken up by the growing oocytes⁸¹; the presence of vitellogenin in feces is thus surprising, and the protein structure needs further attention. However, it has been suggested that it can bind several microbial ligands, such as LPS, lipoteichoic acid or β -glucan, due to the vitellogenin domain (as reviewed Jacquet⁸²).

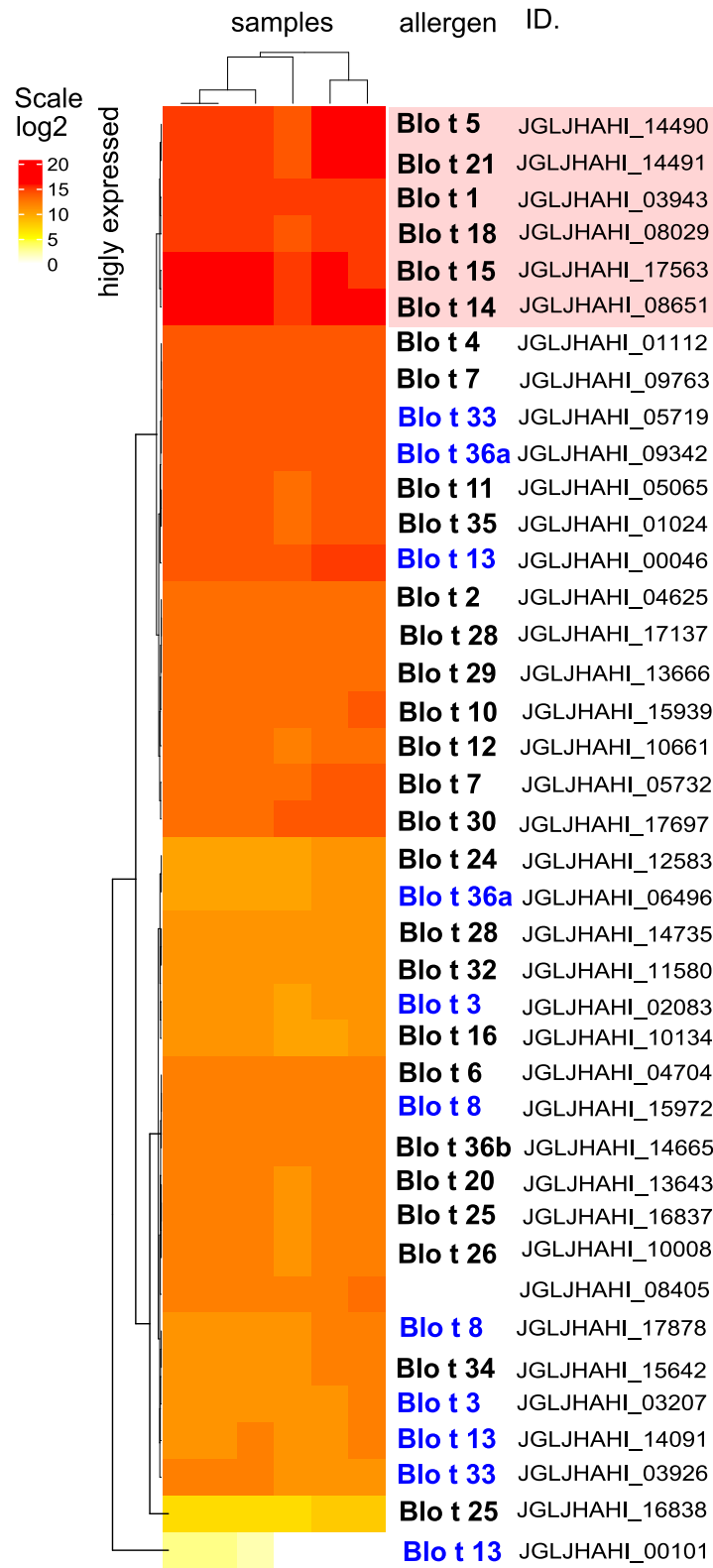
The chitin-binding peritrophin-A domain is the same domain identified exclusively in group 12 in *B. tropicalis*⁸³ and in group 23 (*D. farinae*^{84,85} and *D. pteronyssinus*⁸⁶) and group 37.⁴ We did not find any evidence of the presence of groups 23 and 37 in *B. tropicalis* based on our analyses. Similarly, as in *Dermatophagoides* mites,⁸⁷ Blo t 15 and Blo t 18 showed similar structures (i.e., signal peptide and glycosyl hydrolase family 18) (Blo t 15 position 33–386 and Blo t 18 position

29–355). In addition to the glycosyl hydrolase family, *Dermatophagoides* mites exhibit a chitin-binding domain. This was not found in *Blomia*. The previous observation showed that polyclonal anti-Der p 15 and Der p 18 mouse serum bound only one specific allergen, and no cross-reactivity was observed.⁸⁷ The situation in *B. tropicalis* is not known and needs further analysis.

Putative Blo t 20 allergen protein (arginine kinase⁸⁸ or ATP: guanido phosphotransferase (our PHMER analysis)) is relatively abundant in mite bodies but has low or zero abundance in feces. Der p 20 presented 40% IgE reactivity in nonhospitalized patients,⁸⁹ and Der f 20 showed 7% IgE reactivity to patient sera,^{90,91} indicating that in pyroglyphid mites, this allergen has moderate to low medical importance. However, arginine kinase is a major allergen in crustaceans that may cause asthma and fatal anaphylaxis in fishermen and processing plant workers.⁹² The Allergenic properties of putative Blo t 20 are unknown.

Protease allergens (Blo t 3 trypsin and Blo t 6 chymotrypsin) were the predominant proteins in mite feces. Similarly, group 3 and 6 proteins were highly abundant in the feces of the mold mite *T. putrescentiae*,^{18–20} but in *D. pteronyssinus*, only chymotrypsin was detected.^{17,18} Previous observations have indicated that in culture extracts of *B. tropicalis*, no trypsin or chymotrypsin enzymatic activity can be detected⁹³ or that these enzymes occur at low abundance.²¹

FIGURE 4 Expression of known and predicted allergen-encoding genes of *Blomia tropicalis*. Gene expression values were log2 transformed and visualized in a heatmap. The red box indicates allergens with the highest expression. In the allergen group names, blue indicates allergens identified or predicted allergen isoforms.



However, another study based on a design using inhibitors and specific substrates for enzymes identified trypsin, elastase, chymotrypsin, kallikrein, C3/C5 convertase, and mast cell protease.⁹⁴ Our results showed that the two isoforms of Blo t 3 differ in their protein structure and body localization: Blot 3 JGLJHAHI_02083 showing

similarity to factor IX of the blood clotting cascade is body localized, trypsinogen JGLJHAHI_03207 associated with feces. The last Blo t 3 (AAQ24542) isoform was obtained by the amplification of isolate RNA and confirmed by DNA sequencing.⁶⁶ The absence of the transcriptome and presence of the *B. tropicalis* proteome in this and

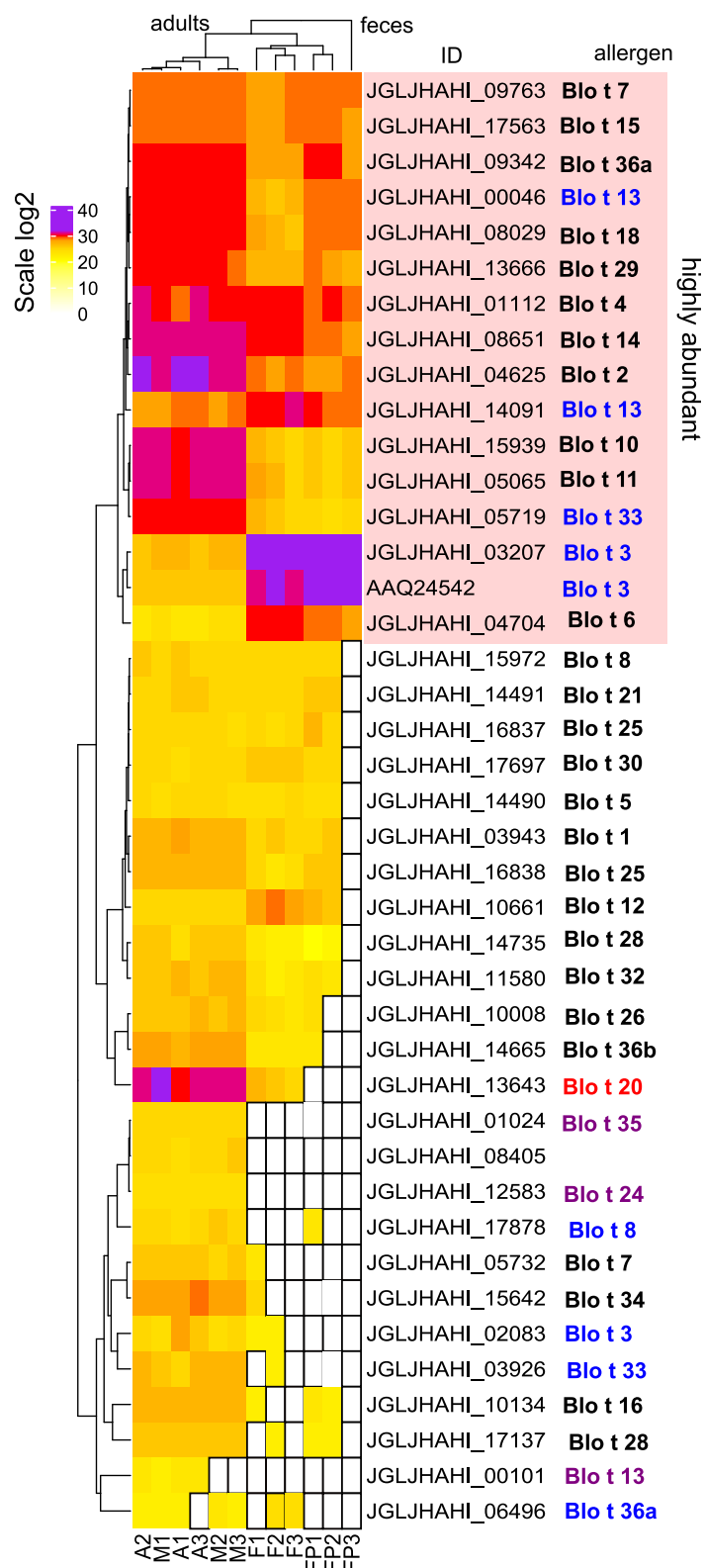


FIGURE 5 Abundance of allergenic proteins in the bodies and feces of *Blomia tropicalis* (A, adults, M, mixed stages, F, feces, FP, fecal pellets not dissolved in water). Data were log2 transformed. In the allergen group names, blue indicates that an allergen was identified with more isoforms, red indicates that an allergen was not present in fecal pellets (FP), and purple indicates allergens not present in feces (F).

previous studies²¹ indicated posttranslational modification. Surprisingly, Blo t 6 JGLJHAHI_04704 exhibited similarity to kallikrein-like serine protease, confirming kallikrein enzymatic activity of *B. tropicalis* extract.⁹⁴ These results can be reconciled as enzyme inactivation occurring in the mite bodies⁹⁵ or the expression of these enzymes

being affected due to the presence of the intracellular bacterial symbionts *Cardinium* and *Wolbachia*.³⁵

Blo t 5 and 21 have been reported as the major allergens of *B. tropicalis*.^{10,14} Our analyses showed that both proteins contained a signal peptide and shared a Blo t 5 domain. The majority (>75%) of

sensitized patients were cosensitized to both Blo t 5 and Blo t 21, which was expected based on their structural similarity.⁷⁰ Both of these allergens were found in the midgut and inside FP (Gao et al.,⁷⁰ as independently suggested by our gene expression and proteomic analyses.

Epidemiological studies provide further evidence of an empirical link between abundance and allergenicity.⁹⁶ The suggestion is that the allergens are more stable and more highly expressed than other proteins.⁹⁷ For example, the stability of Der p 1 is slightly less than the mean for DP proteins, but it is exceptionally highly transcribed. In contrast, Der p 23 is not as highly expressed but is extremely stable.⁹⁷ Based on our analyses, the comparison of allergens in mite bodies and feces enables the identification of stable allergens in terms of their resistance to degradation by mite digestive enzymes. The allergens that are present in feces are produced via the digestive tract.^{98,99} This means that these proteins are not degraded by digestive enzymes.⁹⁵ This is supported by studies showing low allergen degradation (Der f 1) after exposure in the environment.¹⁰⁰

In conclusion, we show that the medically important mite *B. tropicalis* has a large set of putative allergens ($n = 15$), as predicted by omics analyses (genome, transcriptome, proteome) and the use of carefully verified comparative data on known allergens. Although the clinical importance of these putative allergens should be verified in the future, a similar bioinformatics approach was successful in allergen discovery in the mold *T. putrescentiae*.¹⁰¹ It is therefore expected that a large portion of our predicted putative allergens will have medical significance. Our study also provides accurate quantification of immunogenic proteins in both mite whole bodies and FP, the two important mite-related components with distinct allergenic profiles. For this task, we used a shotgun proteomics approach with an advanced Tribrid Orbitrap Fusion spectrometer and contrasted our results with gene expression data. We showed that the expression levels of allergen-encoding genes may not be strictly correlated with actual allergenic protein abundance. Effective strategies for the identification and prediction of new allergens include high-throughput sequencing, proteomics, transcriptomics and genomics; however, the most accurate results are achieved when these technologies are integrated in a so-called *omics* approach.^{34,35,102–106} Moreover, label-free or other quantitative approaches enable the estimation of mite allergen abundance in proteome profiles.^{34,35}

AUTHOR CONTRIBUTIONS

Jan Hubert, Tomas Erban, Pavel B. Klimov, Qixin He, Bruno Sopko and Susanne Vrtala—scientific writing, Jan Hubert—detailed data evaluation and preparation of figures and tables, Karel Harant, Pavel Talacko, and Tomas Erban—proteomic analysis, Scot E. Dowd—genome and transcriptome sequencing, Bruno Sopko—protein analyses, Tomas Erban—experimental design; all authors commented on the final version of the draft.

ACKNOWLEDGMENTS

The authors thank Marta Nesvorna, Julie Chalupnikova, and Martin Markovic (all Crop Research Institute, Prague, Czechia) for technical

help. JH, BS, and TE were supported by project LUAUS23082 of the Ministry of Education, Youth and Sports of the Czech Republic. PBK was supported by the Ministry of Science and Higher Education of the Russian Federation within the framework of the Federal Scientific and Technical Program for the Development of Genetic Technologies for 2019–2027 (agreement N^o: 075-15-2021-1345, unique identifier RF 193021X0012).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the supplementary material of this article and Genbank.

ETHICAL STATEMENT

Not applicable.

ORCID

Jan Hubert  <https://orcid.org/0000-0003-0740-166X>

Susanne Vrtala  <https://orcid.org/0000-0003-4250-8243>

Bruno Sopko  <https://orcid.org/0000-0002-5580-1871>

Scot E. Dowd  <https://orcid.org/0000-0002-6296-1427>

Qixin He  <https://orcid.org/0000-0003-1696-8203>

Pavel B. Klimov  <https://orcid.org/0000-0002-9966-969X>

Karel Harant  <https://orcid.org/0000-0002-9654-5392>

Pavel Talacko  <https://orcid.org/0000-0002-0943-4564>

Tomas Erban  <https://orcid.org/0000-0003-1730-779X>

REFERENCES

- Aggarwal P, Senthilkumaran S. *Dust Mite Allergy*. StatPearls Publishing; 2022. Accessed 22 February 2023. <https://www.ncbi.nlm.nih.gov/books/NBK560718/>
- Bischoff S. Role of mast cells in allergic and non-allergic immune responses: comparison of human and murine data. *Nat Rev Immunol*. 2007;7(2):93–104. <https://doi.org/10.1038/nri2018>
- Amoretti M, Amsler C, Bonomi G, et al. Production and detection of cold antihydrogen atoms. *Nature*. 2002;419(6906):456–459. <https://doi.org/10.1038/nature01096>
- WHO/IUIS. *Allergen Nomenclature*. WHO/IUIS Allergen Nomenclature Sub-Committee; 2023. Accessed 22 February 2023. <http://www.allergen.org>
- Arlian LG, Morgan MS, Neal JS. Dust mite allergens: ecology and distribution. *Curr Allergy Asthma Rep*. 2002;2(5):401–411. <https://doi.org/10.1007/s11882-002-0074-2>
- Thomas WR, Hales BJ, Smith W. *Blomia tropicalis*: more than just another source of mite allergens. *Clin Exp Allergy*. 2003;33(4):416–418. <https://doi.org/10.1046/j.1365-2222.2003.01634.x>
- Fernandez-Caldas E, Puerta L, Mercado D, Lockey RF, Caraballo LR. Mite fauna, Der p I, Der f I and *Blomia tropicalis* allergen levels in a tropical environment. *Clin Exp Allergy*. 1993;23(4):292–297. <https://doi.org/10.1111/j.1365-2222.1993.tb00325.x>
- Arruda LK, Vailles LD, Platts-Mills TAE, et al. Sensitization to *Blomia tropicalis* in patients with asthma and identification of allergen Blo t 5. *Am J Respir Crit Care Med*. 1997;155(1):343–350. <https://doi.org/10.1164/ajrccm.155.1.9001334>

9. Huang H.-W, Lue K.-H, Wong R.-H, Sun H.-L, Sheu J.-N, Lu K.-H. Distribution of allergens in children with different atopic disorders in central Taiwan. *Acta Paediatr Taiwanica*. 2006;47(3):127-134.
10. Carvalho KA, de Melo-Neto OP, Magalhaes FB, et al. *Blomia tropicalis* Blo t 5 and Blo t 21 recombinant allergens might confer higher specificity to serodiagnostic assays than whole mite extract. *BMC Immunol*. 2013;14(1):11. <https://doi.org/10.1186/1471-2172-14-11>
11. Pauli G, Wurmser C, Roos A, et al. Frequent IgE recognition of *Blomia tropicalis* allergen molecules in asthmatic children and young adults in equatorial Africa. *Front Immunol*. 2023;14:1133935. <https://doi.org/10.3389/fimmu.2023.1133935>
12. Arlian LG, Vyszynski-Moher DL, Fernandez-Caldas E. Allergenicity of the mite, *Blomia tropicalis*. *J Allergy Clin Immunol*. 1993;91(5):1042-1050. [https://doi.org/10.1016/0091-6749\(93\)90218-5](https://doi.org/10.1016/0091-6749(93)90218-5)
13. da Silva ES, Asam C, Lackner P, et al. Allergens of *Blomia tropicalis*: an overview of recombinant molecules. *Int Arch Allergy Immunol*. 2017;172(4):203-214. <https://doi.org/10.1159/000464325>
14. Vrtala S. Allergens from house dust and storage mites. *Allergo J Int*. 2022;31(8):267-271. <https://doi.org/10.1007/s40629-022-00226-5>
15. Aalberse RC. Structural biology of allergens. *J Allergy Clin Immunol*. 2000;106(2):228-238. <https://doi.org/10.1067/mai.2000.108434>
16. Tovey ER, Chapman MD, Platts-Mills TAE. Mite faeces are a major source of house dust allergens. *Nature*. 1981;289(5798):592-593. <https://doi.org/10.1038/289592a0>
17. Thomas WR. Hierarchy and molecular properties of house dust mite allergens. *Allergol Int*. 2015;64(4):304-311. <https://doi.org/10.1016/j.alit.2015.05.004>
18. Erban T, Hubert J. Two-dimensional gel proteomic analysis of *Dermatophagoides farinae* feces. *Exp Appl Acarol*. 2015;65(1):73-87. <https://doi.org/10.1007/s10493-014-9848-1>
19. Erban T, Rybanska D, Harant K, Hortova B, Hubert J. Feces derived allergens of *Tyrophagus putrescentiae* reared on dried dog food and evidence of the strong nutritional interaction between the mite and *Bacillus cereus* producing protease bacillolysins and exochitinases. *Front Physiol*. 2016;7:53. <https://doi.org/10.3389/fphys.2016.00053>
20. Erban T, Harant K, Hubert J. Detailed two-dimensional gel proteomic mapping of the feces of the house dust mite *Dermatophagoides pteronyssinus* and comparison with *D. farinae*: reduced trypsin protease content in *D. pteronyssinus* and different isoforms. *J Proteonomics*. 2017;162:11-19. <https://doi.org/10.1016/j.jprot.2017.04.021>
21. da Silva ES, Ponte JCM, da Silva MB, et al. Proteomic analysis reveals allergen variability among breeds of the dust mite *Blomia tropicalis*. *Int Arch Allergy Immunol*. 2019;180(3):159-172. <https://doi.org/10.1159/000501964>
22. Hubert J, Nesvorna M, Kopecky J, Erban T, Klimov P. Population and culture age influence the microbiome profiles of house dust mites. *Microb Ecol*. 2019;77(4):1048-1066. <https://doi.org/10.1007/s00248-018-1294-x>
23. Erban T, Hubert J. Digestive function of lysozyme in synanthropic acarid mites enables utilization of bacteria as a food source. *Exp Appl Acarol*. 2008;44(3):199-212. <https://doi.org/10.1007/s10493-008-9138-x>
24. Krueger F. Trim Galore. *Babraham Bioinforma*. 2021. Accessed 22 February 2023. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
25. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. *Babraham Bioinformatics*; 2019. Accessed 22 February 2023. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
26. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455-477. <https://doi.org/10.1089/cmb.2012.0021>
27. Antipov D, Korobeynikov A, McLean JS, Pevzner PA. hybrid-SPAdes: an algorithm for hybrid assembly of short and long reads. *Bioinformatics*. 2016;32(7):1009-1015. <https://doi.org/10.1093/bioinformatics/btv688>
28. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068-2069. <https://doi.org/10.1093/bioinformatics/btu153>
29. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and meta-genome sequences. *J Mol Biol*. 2016;428(4):726-731. <https://doi.org/10.1016/j.jmb.2015.11.006>
30. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25. <https://doi.org/10.1186/gb-2009-10-3-r25>
31. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-359. <https://doi.org/10.1038/nmeth.1923>
32. Li H. Minimap2: Pairwise Alignment for Nucleotide Sequences, Version 5; 2018. *arXiv*, 1708.01492 [q-bio.GN]. Accessed 22 February 2023. <https://arxiv.org/abs/1708.01492>
33. Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinf*. 2005;6(1):31. <https://doi.org/10.1186/1471-2105-6-31>
34. Erban T, Klimov P, Talacko P, Harant K, Hubert J. Proteogenomics of the house dust mite, *Dermatophagoides farinae*: allergen repertoire, accurate allergen identification, isoforms, and sex-biased proteome differences. *J Proteonomics*. 2020;210:103535. <https://doi.org/10.1016/j.jprot.2019.103535>
35. Erban T, Klimov PB, Harant K, Talacko P, Nesvorna M, Hubert J. Label-free proteomic analysis reveals differentially expressed *Wolbachia* proteins in *Tyrophagus putrescentiae*: mite allergens and markers reflecting population-related proteome differences. *J Proteonomics*. 2021;249:104356. <https://doi.org/10.1016/j.jprot.2021.104356>
36. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008;26(12):1367-1372. <https://doi.org/10.1038/nbt.1511>
37. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics*. 2014;13(9):2513-2526. <https://doi.org/10.1074/mcp.M113.031591>
38. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res*. 2011;10(4):1794-1805. <https://doi.org/10.1021/pr101065j>
39. Tyanova S, Temu T, Sinitcyn P, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*. 2016;13(9):731-740. <https://doi.org/10.1038/nmeth.3901>
40. Johnson M, Zaretskaya I, Raytselis Y, Merezukh Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res*. 2008;36(Web server issue):W5-W9. <https://doi.org/10.1093/nar/gkn201>
41. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011;39(Web server issue):W29-W37. <https://doi.org/10.1093/nar/gkr367>
42. Notredame C, Higgins DG, Heringa J. T-coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol*. 2000;302(1):205-217. <https://doi.org/10.1006/jmbi.2000.4042>
43. Di Tommaso P, Moretti S, Xenarios I, et al. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids*

- Res. 2011;39(Web server issue):W13-W17. <https://doi.org/10.1093/nar/gkr245>
44. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59(3):307-321. <https://doi.org/10.1093/sysbio/syq010>
45. Rambaut A. FigTree, a Graphical Viewer of Phylogenetic Trees: 2014-07-09 – v1.4.2. Molecular Evolution, Phylogenetics and Epidemiology: Research, Software and Publications of Andrew Rambaut and Members of His Research Group; 2014. Accessed 22 February 2023. <http://tree.bio.ed.ac.uk/software/figtree/>
46. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. 2014;42(W1):W320-W324. <https://doi.org/10.1093/nar/gku316>
47. Madeira F, Pearce M, Tivey ARN, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res*. 2022;50(W1):W276-W279. <https://doi.org/10.1093/nar/gkac240>
48. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. *Nucleic Acids Res*. 2018;46(W1):W200-W204. <https://doi.org/10.1093/nar/gky448>
49. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and clustal X version 2.0. *Bioinformatics*. 2007;23(21):2947-2948. <https://doi.org/10.1093/bioinformatics/btm404>
50. Fiser A, Do RK, Sali A. Modeling of loops in protein structures. *Protein Sci*. 2000;9(9):1753-1773. <https://doi.org/10.1110/ps.9.9.1753>
51. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*. 2000;29(1):291-325. <https://doi.org/10.1146/annurev.biophys.29.1.291>
52. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*. 1993;26(2):283-291. <https://doi.org/10.1107/S0021889892009944>
53. Schuler LD, Daura X, van Gunsteren WF. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J Comput Chem*. 2001;22(11):1205-1218. <https://doi.org/10.1002/jcc.1078>
54. R Development Core Team. R-4.2.2 for Windows. The Comprehensive R Archive Network; 2022. Accessed 22 February 2023. <https://cran.r-project.org/bin/windows/base/>
55. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32(18):2847-2849. <https://doi.org/10.1093/bioinformatics/btw313>
56. Gu Z. Complex heatmap visualization. *iMeta*. 2022;1(3):e43. <https://doi.org/10.1002/imt.2.43>
57. Hammer O. *Past 4 – the Past of the Future*. Natural History Museum, University of Oslo; 2020. Accessed 22 February 2023. <https://www.nhm.uio.no/english/research/resources/past/>
58. Xiong Q, Wan T, Malainual N, Tsui SK-W. The genome analysis of *Blomia tropicalis* reveals a comprehensive allergen profile. *J Allergy Clin Immunol*. 2020;145(2 Suppl 1):AB82. AB82-AB82. <https://doi.org/10.1016/j.jaci.2019.12.682>
59. Xiong Q, Wan AT-Y, Liu X, et al. Comparative genomics reveals insights into the divergent evolution of astigmatic mites and household pest adaptations. *Mol Biol Evol*. 2022;39(5):msac097. <https://doi.org/10.1093/molbev/msac097>
60. Mora C, Flores I, Montealegre F, Diaz A. Cloning and expression of Blo t 1, a novel allergen from the dust mite *Blomia tropicalis*, homologous to cysteine proteases. *Clin Exp Allergy*. 2003;33(1):28-34. <https://doi.org/10.1046/j.1365-2222.2003.01480.x>
61. Chew FT, Wang W-L, Shang HS, Kuay KT, Lim SH, Lee BW. Identification and cloning of group 1 allergen from *Blomia tropicalis*; 2003. Unpublished. Accessed 22 February 2023. <https://www.ncbi.nlm.nih.gov/protein/AAQ24541>
62. Xiong Q, Liu X, Wan AT-Y, et al. Comprehensive allergen profile of *Blomia tropicalis* provides insights into the component-resolved diagnosis of mite allergy. *bioRxiv*. 2023. <https://doi.org/10.1101/2023.02.09.527948>. Accessed 22 February 2023.
63. Chew FT, Angus AC, Lim SH. Cloning and expression of group 2 allergen isoforms from the storage mites *Lepidoglyphus destructor* and *Blomia tropicalis*; 2003. Unpublished. Accessed 22 February 2023. <https://www.ncbi.nlm.nih.gov/protein/AAQ73483>
64. Reginald K, Pang SL, Chew FT. Blo t 2: group 2 allergen from the dust mite *Blomia tropicalis*. *Sci Rep*. 2019;9(1):12239. <https://doi.org/10.1038/s41598-019-48688-y>
65. Reginald K, Kuay KT, Chew FT. Cloning and characterization of a group 2 allergen from *Blomia tropicalis*; 2006. Unpublished. Accessed 22 February 2023. <https://www.ncbi.nlm.nih.gov/protein/ABG76185>
66. Cheong N, Yang L, Lee BW, Chua KY. Cloning of a group 3 allergen from *Blomia tropicalis* mites. *Allergy*. 2003;58(4):352-356. <https://doi.org/10.1034/j.1398-9995.2003.00033.x>
67. Flores I, Mora C, Rivera E, Donnelly R, Montealegre F. Cloning and molecular characterization of a cDNA from *Blomia tropicalis* homologous to dust mite group 3 allergens (trypsin-like proteases). *Int Arch Allergy Immunol*. 2003;130(1):12-16. <https://doi.org/10.1159/000068375>
68. Chew FT, Kuay KT, Shang HS, Wang W-L, Lim SH, Lee BW. Identification and cloning of group 4 allergen from *Blomia tropicalis*; 2003. Unpublished. Accessed 22 February 2023. <https://www.ncbi.nlm.nih.gov/protein/AAQ24543>
69. Caraballo L, Avjioglu A, Marrugo J, Puerta L, Marsh D. Cloning and expression of complementary DNA coding for an allergen with common antibody-binding specificities with three allergens of the house dust mite *Blomia tropicalis*. *J Allergy Clin Immunol*. 1996;98(3):573-579. [https://doi.org/10.1016/S0091-6749\(96\)70091-x](https://doi.org/10.1016/S0091-6749(96)70091-x)
70. Gao YF, Wang DY, 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):105-112. <https://doi.org/10.1016/j.jaci.2007.02.032>
71. Medina LR, Malainual N, Ramos JD. Genetic polymorphisms and allergenicity of Blo t 5 in a house dust mite allergic Filipino population. *Asian Pac J Allergy Immunol*. 2017;35(4):203-211. <https://doi.org/10.12932/AP0794>
72. Chew FT, Wang W-L, Shang HS, Kuay KT, Lim SH, Lee BW. Identification and cloning of group 6 allergen from *Blomia tropicalis*; 2003. Unpublished. Accessed 22 February 2023. <https://www.ncbi.nlm.nih.gov/protein/AAQ24544>
73. Jacquet APA. Blo t 7 precursor, partial [*Blomia tropicalis*]: direct submission. 2017. Unpublished. Accessed 22 February 2023. <https://www.ncbi.nlm.nih.gov/protein/ASX95438>
74. Chew FT, Shang HS. Identification and cloning of group 7 allergen from *Blomia tropicalis*. 2003. Unpublished. Accessed 22 February 2023. <https://www.ncbi.nlm.nih.gov/protein/AAQ24545.1>
75. Zakzuk J, Fernandez-Caldas E, Caraballo L. Characterization of a glutathione S-transferase from *Blomia tropicalis*; 2009. Unpublished. Accessed 22 February 2023. <https://www.ncbi.nlm.nih.gov/protein/ACV04860>
76. Chew FT, Wang W-L, Kuay KT, et al. Identification and cloning of group 8 allergen from *Blomia tropicalis* expressed sequence tags; 2003. Unpublished. Accessed 22 February 2023. <https://www.ncbi.nlm.nih.gov/protein/AAP35069>
77. Chew FT, Wang W-L, Kuay KT, Lim SH, Lee BW. Identification and cloning of group 10 allergen from *Blomia tropicalis* expressed sequence tags; 2007. Unpublished. Accessed 22 February 2023. <https://www.ncbi.nlm.nih.gov/protein/ABU97466>
78. Caraballo L, Puerta L, Jimenez S, 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):341-347. <https://doi.org/10.1159/000237478>

79. Epton MJ, Dilworth RJ, Smith W, Thomas WR. Sensitisation to the lipid-binding apolipoprotein allergen Der p 14 and the peptide Mag-1. *Int Arch Allergy Immunol.* 2001;124(1-3):57-60. <https://doi.org/10.1159/000053668>
80. El Ramlawy KG, Fujimura T, Aki T, et al. Prominent IgE-binding and cytokine-inducing capacities of a newly cloned N-terminal region of Der f 14, an apolipoprotein-like house dust mite allergen. *J Biochem.* 2018;163(1):51-60. <https://doi.org/10.1093/jb/mvx060>
81. Engelmann F. Insect vitellogenin: identification, biosynthesis, and role in vitellogenesis. *Adv Insect Physiol.* 1979;14:49-108. [https://doi.org/10.1016/S0065-2806\(08\)60051-X](https://doi.org/10.1016/S0065-2806(08)60051-X)
82. Jacquet A. The HDM allergen orchestra and its cysteine protease maestro: stimulators of kaleidoscopic innate immune responses. *Mol Immunol.* 2023;156:48-60. <https://doi.org/10.1016/j.molimm.2023.03.002>
83. Puerta L, Caraballo L, Fernandez-Caldas E, et al. Nucleotide sequence analysis of a complementary DNA coding for a *Blomia tropicalis* allergen. *J Allergy Clin Immunol.* 1996;98(5):932-937. [https://doi.org/10.1016/S0091-6749\(96\)80009-1](https://doi.org/10.1016/S0091-6749(96)80009-1)
84. Soh WT, Le Mignon M, Suratannon N, et al. The house dust mite major allergen Der p 23 displays O-glycan-independent IgE reactivities but no chitin-binding activity. *Int Arch Allergy Immunol.* 2015;168(3):150-160. <https://doi.org/10.1159/000442176>
85. He Y, Dou C, Su Y, et al. Identification of Der f 23 as a new major allergen of *Dermatophagoides farinae*. *Mol Med Rep.* 2019;20(2):1270-1278. <https://doi.org/10.3892/mmr.2019.10305>
86. Weghofer M, Grote M, Resch Y, 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):3059-3067. <https://doi.org/10.4049/jimmunol.1202288>
87. O'Neil SE, Heinrich TK, Hales BJ, et al. The chitinase allergens Der p 15 and Der p 18 from *Dermatophagoides pteronyssinus*. *Clin Exp Allergy.* 2006;36(6):831-839. <https://doi.org/10.1111/j.1365-2222.2006.02497.x>
88. Xing P, Yu H, Li M, et al. Characterization of arginine kinase, a novel allergen of *Dermatophagoides farinae* (Der f 20). *Am J Transl Res.* 2015;7(12):2815-2823.
89. Hales BJ, Martin AC, Pearce LJ, et al. IgE and IgG anti-house dust mite specificities in allergic disease. *J Allergy Clin Immunol.* 2006;118(2):361-367. <https://doi.org/10.1016/j.jaci.2006.04.001>
90. Jeong KY, Lee JY, Son M, et al. Profiles of IgE sensitization to Der f 1, Der f 2, Der f 6, Der f 8, Der f 10, and Der f 20 in Korean house dust mite allergy patients. *Allergy Asthma Immunol Res.* 2015;7(5):483-488. <https://doi.org/10.4168/aa.2015.7.5.483>
91. WHO/IUIS. Der p 20: allergen details. WHO/IUIS Allergen Nomenclature Sub-Committee; 2019. Accessed February 22, 2023. <http://www.allergen.org/viewallergen.php?aid=294>
92. Downs M, Johnson P, Zeece M. Insects and their connection to food allergy. In: Dossey AT, Morales-Ramos JA, Rojas MG, eds. *Insects as Sustainable Food Ingredients: Production, Processing and Food Applications.* Academic Press; 2016:255-272. <https://doi.org/10.1016/B978-0-12-802856-8.00009-0>
93. Cardona G, Guisantes J, Eraso E, Serna LA, Martinez J. Enzymatic analysis of *Blomia tropicalis* and *Blomia kulagini* (Acari: echimyopodidae) allergenic extracts obtained from different phases of culture growth. *Exp Appl Acarol.* 2006;39(3-4):281-288. <https://doi.org/10.1007/s10493-006-9009-2>
94. Montealegre F, Quinones C, Torres N, Goth K. Detection of serine proteases in extracts of the domestic mite *Blomia tropicalis*. *Exp Appl Acarol.* 2002;26(1-2):87-100. <https://doi.org/10.1023/a:1020931221953>
95. Vidal-Quist JC, Ortego F, Hernandez-Crespo P. Contribution of cysteine and serine proteases to proteolytic digestion in an allergy-eliciting house dust mite. *J Insect Physiol.* 2021;133:104285. <https://doi.org/10.1016/j.jinsphys.2021.104285>
96. Foo ACY, Mueller GA. Abundance and stability as common properties of allergens. *Front Allergy.* 2021;2:769728. <https://doi.org/10.3389/falgy.2021.769728>
97. Ogburn RN, Randall TA, Xu Y, et al. Are dust mite allergens more abundant and/or more stable than other *Dermatophagoides pteronyssinus* proteins? *J Allergy Clin Immunol.* 2017;139(3):1030-1032. <https://doi.org/10.1016/j.jaci.2016.08.016>
98. de Boer R, van der Hoeven WAD, Stapel SO. The decay of house dust mite allergens, Der p I and Der p II, under natural conditions. *Clin Exp Allergy.* 1995;25(8):765-770. <https://doi.org/10.1111/j.1365-2222.1995.tb00015.x>
99. Terra WR, Ferreira C. Insect digestive enzymes: properties, compartmentalization and function. *Comp Biochem Physiol B Comp Biochem.* 1994;109(1):1-62. [https://doi.org/10.1016/0305-0491\(94\)90141-4](https://doi.org/10.1016/0305-0491(94)90141-4)
100. Sidenius KE, Hallas TE, Stenderup J, Poulsen LK, Mosbech H. Decay of house-dust mite allergen Der f 1 at indoor climatic conditions. *Ann Allergy Asthma Immunol.* 2002;89(1):34-37. [https://doi.org/10.1016/S1081-1206\(10\)61908-4](https://doi.org/10.1016/S1081-1206(10)61908-4)
101. Zhou Y, Klimov PB, Gu X, et al. Chromosome-level genomic assembly and allergome inference reveal novel allergens in *Tyrophagus putrescentiae*. *Allergy.* 2023;78(6):1691-1695. <https://doi.org/10.1111/all.15656>
102. An S, Chen L, Long C, et al. *Dermatophagoides farinae* allergens diversity identification by proteomics. *Mol Cell Proteomics.* 2013;12(7):1818-1828. <https://doi.org/10.1074/mcp.M112.027136>
103. Chan T-F, Ji K-M, Yim AK-Y, 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):539-548. <https://doi.org/10.1016/j.jaci.2014.09.031>
104. Bordas-Le Floch V, Le Mignon M, Bussieres L, et al. A combined transcriptome and proteome analysis extends the allergome of house dust mite *Dermatophagoides* species. *PLoS ONE.* 2017;12(10):e0185830. <https://doi.org/10.1371/journal.pone.0185830>
105. Waldron R, McGowan J, Gordon N, McCarthy C, Mitchell EB, Fitzpatrick DA. Proteome and allergome of the European house dust mite *Dermatophagoides pteronyssinus*. *PLoS ONE.* 2019;14(5):e0216171. <https://doi.org/10.1371/journal.pone.0216171>
106. Hubert J, Nesvorna M, Klimov P, Dowd SE, Sopko B, Erban T. Differential allergen expression in three *Tyrophagus putrescentiae* strains inhabited by distinct microbiome. *Allergy.* 2019;74(12):2502-2507. <https://doi.org/10.1111/all.13921>
107. Ramos JD, Cheong N, Lee BW, Chua KY. cDNA cloning and expression of Blo t 11, the *Blomia tropicalis* allergen homologous to paramyosin. *Int Arch Allergy Immunol.* 2001;126(4):286-293. <https://doi.org/10.1159/000049525>
108. Ramos JDA, Teo ASM, Lee BW, Cheong N, Chua KY. DNA immunization for the production of monoclonal antibodies to Blo t 11, a paramyosin homolog from *Blomia tropicalis*. *Allergy.* 2004;59(5):539-547. <https://doi.org/10.1046/j.1398-9995.2003.00409.x>
109. Chew FT, Ong ST, Wang W-L, et al. Identification and cloning of group 13 allergen from *Blomia tropicalis* expressed sequence tags; 2003. Unpublished. Accessed 22 February 2023. <https://www.ncbi.nlm.nih.gov/protein/AAP35071.1>
110. Chew FT, Kuay KT, Shang HS, Wang W-L, Lim SH, Lee BW. Identification and cloning of chitinase homolog allergen from *Blomia tropicalis*; 2003. Unpublished. Accessed 22 February 2023. <https://www.ncbi.nlm.nih.gov/protein/AAQ24549>
111. Chew FT, Gao YF. Blo T 5.02 Allergen [*Blomia tropicalis*]: direct submission; 2004. Unpublished. Accessed 22 February 2023. <https://www.ncbi.nlm.nih.gov/protein/AAX34047>

112. Chew FT, Kuay KT, Shang HS, Wang W-L, Lim SH, Lee BW. Identification and cloning of Mag 29 allergen from *Blomia tropicalis*; 2003. Unpublished. Accessed 22 February 2023. <https://www.ncbi.nlm.nih.gov/protein/AAQ24552>
113. Chew FT, Wang W-L, Shang HS, Kuay KT, Lim SH, Lee BW. Identification and cloning of profilin homolog allergen from *Blomia tropicalis*; 2003. Unpublished. Accessed 22 February 2023. <https://www.ncbi.nlm.nih.gov/protein/AAQ24553>
114. Chew FT, Wang W-L, Shang HS, Kuay KT, Lim SH, Lee BW. Identification and cloning of Blo t Alt 6 homolog allergen from *Blomia tropicalis*; 2003. Unpublished. Accessed 22 February 2023. <https://www.ncbi.nlm.nih.gov/protein/AAQ24548>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hubert J, Vrtala S, Sopko B, et al. Predicting *Blomia tropicalis* allergens using a multiomics approach. *Clin Transl Allergy*. 2023;e12302. <https://doi.org/10.1002/clt2.12302>

Supplementary material

Title: Predicting *Blomia tropicalis* allergens using a multiomics approach

Authors: J. Hubert, S. Vrtala, B. Sopko, S. E. Dowd, Q. He, P. B. Klimov, K. Harant, P. Talacko, T. Erban

Journal: Clinical and Translational Allergy

Group 1 allergens showed structural homology to Peptidase_C1, similar to the other proteins classified in this group (Figure S1). Phmmer analyses of all protein sequences of *Blomia tropicalis* provided the following findings: (i) AAK58415, JGLKHAHI_00359 and JGLKHAHI_03943 are fragments that do not contain signal peptides; (ii) AAQ24541, WBV73454, KAI2799130 and WBV73455 contain a signal peptide, cathepsin propeptide inhibitor and Peptidase_C1; and (iii) KAI2799128 is a sequencing artifact containing a signal peptide, a triplicated cathepsin propeptide inhibitor and a Peptidase_C1 region.

The sequence identity within *Dermatophagoides* mites ranged between 80 and 99% (Table S1), and the *Tyrophagus putrescentiae* allergen (ABM53753) showed approximately 28% sequence identity to a *Dermatophagoides* protein. *Blomia* allergens (AAK58415, AAQ24541) shared approximately 35% identity with a *Dermatophagoides* protein, similar to the predicted *Blomia tropicalis* allergens. The identity of known Blot 1 (AAK58415 and AAQ24541) was 61.1%. These allergens belonged to different clusters located outside of the *Dermatophagoides* mite allergen and Tyr p 1 clusters (Figure S1).

Predicted *Blomia tropicalis* proteins JGLJHAHI_03943 and JGLJHAHI_00359 showed structural homology corresponding to other allergens from this group. JGLJHAHI_03943 showed 98.6% homology to AAQ24541 and 61% homology to AAK58415. The expression level of JGLJHAHI_00359 was close to 0, so this sequence was eliminated from the allergen list. The comparison of Blot 1 proteins showed significant numbers of conserved areas (Figure S2).

FIGURE S1 Comparison of group 1 allergens and predicted proteins. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was AAX34043 of *Suidasia medoanensis*.

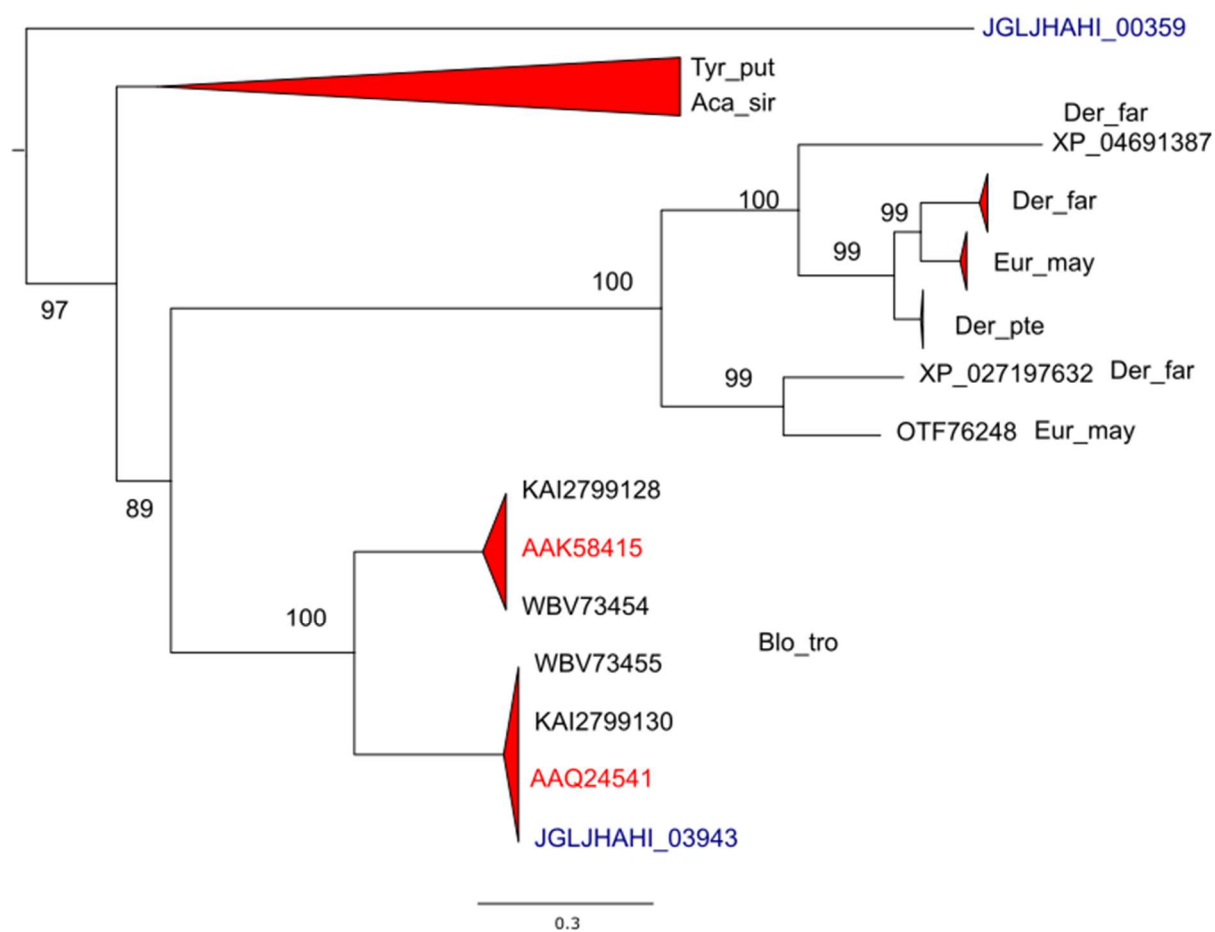


FIGURE S2 Alignment of Blo t 1 proteins.

```

AAK58415      .....
AAQ24541      .....
JGLKHAHI_00359 .....
JGLKHAHI_03943 .....
WBV73454      .....
KAI2799130    .....
WBV73455      .....
MKFLAALCALVAIGSCKPTREEIKTFEQFKKVF GKVYRNAEEEEARREHHFKEQ LKWWVEEH
MKFLAALCALVAIGSCKPTREEIKTFEQFKKVF GKVYRNAEEEEARREHHFKEQ LKWWVEEH
MKFLAALCALVAIGSCKPTREEIKTFEQFKKVF GKVYRNAEEEEARREHHFKEQ LKWWVEEH
MKFLAALCALVAIGSCKPTREEIKTFEQFKKVF GKVYRNAEEEEARREHHFKEQ LKWWVEEH

```

```

AAK58415      .....
AAQ24541      .....
JGLKHAHI_00359 .....
JGLKHAHI_03943 .....
WBV73454      .....
KAI2799130    .....
WBV73455      .....
NGIDGVEYAINNEYSDMSEQEFSLHLSGGGLNFTYMKMEAAAKEPLINTYGSIPANFDWRQK
MGLNKALYSEFPRTLQETEH..FEGN.....TINLPQFDWRQK
NGIDGVEYAINNEYSDMSEQEFSLHLSGGGLNFTYMKMEAAAKEPLINTYGSIPQFDWRQK
QGNKGAEYGINQYSDMSDEEFSQHNLN..GLNISQMMLEVPQAPVIDTYEQIPANFDWRQK
NGIDGVEYAINNEYSDMSEQEFSLHLSGGGLNFTYMKMEAAAKEPLINTYGSIPQFDWRQK
NGIDGVEYAINNEYSDMSEQEFSLHLSGGGLNFTYMKMEAAAKEPLINTYGSIPQFDWRQK

```

```

AAK58415      .....
AAQ24541      .....
JGLKHAHI_00359 .....
JGLKHAHI_03943 .....
WBV73454      .....
KAI2799130    .....
WBV73455      .....
THVNP I R N O G G C G S C W A F A A S S V A E T L Y A T H R H Q N I I L S E Q E L L D C T Y H L Y D P T Y K C H G C
A R L T R I R Q Q G S C G S C W A F A A A G V A E S L Y S I Q K Q Q S I E L S E Q E L V D C T Y N R Y D S S Y Q C N G C
G V I S E V K N O G M C G S C W A F S T I G N I E G V W A V K K N E T V S L S E Q E L V D C D . . S L D A . . . . . G C
A R L T R I R Q Q G E C G S C W A F A A A G V A E S L Y S I Q K Q Q S I E L S E Q E L V D C T Y N R Y D S S Y Q C N G C
THVNP I R N O G A C G S C W A F A A S S V A E T L Y A T H R H Q N I I L S E Q E L I D C T Y H L Y D P T Y K C H G C
A R L T R I R Q Q G A C G S C W A F A A A G V A E S L Y S I Q K Q Q S I E L S E Q E L V D C T Y N R Y D S S Y Q C N G C
A R L T R I R Q Q G A C G S C W A F A A A G V A E S L Y S I Q K Q Q S I E L S E Q E L V D C T Y N R Y D S S Y Q C N G C

```

```

AAK58415      .....
AAQ24541      .....
JGLKHAHI_00359 .....
JGLKHAHI_03943 .....
WBV73454      .....
KAI2799130    .....
WBV73455      .....
Q S G M S P E A F K Y M K Q K G L L E E S H Y P Y K M K L N Q C Q A N A R G T R Y H V S Y N S L R Y R A G D Q E I Q A
G S G Y S T E A F K Y M I R T G L V E E E N Y P Y N M R T Q W C N P D V E G Q R Y H V S G Y Q Q L R Y Q S S D E D V M Y
G G G L P T N A F K T I I R L G G I E S R D Y S Y D G V Q E K C R F N R T L S H V S I D S Y . . L E L P K N E T Y F Q N
G S G Y S T E A F K Y M I R T G L V E E K N Y P Y N M R T Q W C N P D V E G H R Y H V S G Y Q Q L R Y Q S S D E D V M Y
Q S G M S P E A F K Y M K Q K G L L E E S H Y P Y K M K L N Q C Q A N A R G T R Y H V S Y N S L R Y R A G D Q E I Q A
G S G Y S T E A F K Y M I R T G L V E E R N Y P Y N M R T Q W C D P D V E G Q R Y H V S G Y Q Q L R Y Q S S D E D V M Y
G S G Y S T E A F K Y M I R T G L V E E R N Y P Y N M R T Q W C D P D V E G Q R Y H V S G Y Q Q L R Y Q S S D E D V M Y

```

```

AAK58415      .....
AAQ24541      .....
JGLKHAHI_00359 .....
JGLKHAHI_03943 .....
WBV73454      .....
KAI2799130    .....
WBV73455      .....
A I M N H G P V V I Y I H G T E A H F R N L R K G I L R G A G Y N D A Q I D H A V I L V G W G T V Q G V D Y W I I R N S
T I Q Q H G P V V I Y M H G S N N Y F R N L G N G V L R G V A Y N D A Y T D H A V I L V G W G T V Q G V D Y W I I R N S
Y L F A N G P I S T I G L N A N A M Q F Y F G G I S H P P K F L C D P Q N L D H G V L I V G F G V G K T L P Y W I I K N S
T I Q Q H G P V V I Y M H G S N N Y F R N L G N G V L R G V A Y N D A Y T D H A V I L V G W G T V Q G V D Y W I I R N S
A I M N H G P V V I Y I H G T E A H F R N L R K G I L R G A G Y N D A Q I D H A V I L V G W G T V Q G V D Y W I I R N S
T I Q Q H G P V V I Y M H G S N N Y F R N L G N G V L R G V A Y N D A Y T D H A V I L V G W G T V Q G V D Y W I I R N S
T I Q Q H G P V V I Y M H G S N N Y F R N L G N G V L R G V A Y N D A Y T D H A V I L V G W G T V Q G V D Y W I I R N S

```

```

AAK58415      .....
AAQ24541      .....
JGLKHAHI_00359 .....
JGLKHAHI_03943 .....
WBV73454      .....
KAI2799130    .....
WBV73455      .....
W G T Q W G D A G Y G F V E R H H N S L G I N N Y P I Y A S L
W G T Q W G N G Y G Y V E R G H N S L G I N N F V T Y A T L
W G K G W G E H G Y Y R I Y R G D D S C G L T Q M A S S A V I
W G T Q W G N G Y G Y V E R G H N S L G I N N F V T Y A T L
W G T Q W G D A G Y G F V E R H H N S L G I N N Y P I Y A S L
W G T Q W G N G Y G Y V E R G H N S L G I N N Y V T Y A T L
W G T Q W G N G Y G Y V E R G H N S L G I N N Y V T Y A T L

```


All **group 2** allergens showed structural identity to the ML domain, i.e., Der p 2 (AAF86462) (signal peptide 1-15, ML domain 18–141). The sequence identity was high, ranging from 85 to 99% among *Dermatophagoides* allergens (Table S2) and among Blo t 2 (AAQ73481) and predicted proteins. However, for all other species, the sequence identity ranged from 35 to 50%. Predicted JGLJHAHI_04625 shared 87.3% to 95.1% identity with Blo t 2. All *B. tropicalis* proteins clustered outside the cluster of allergens from *Glycyphagus domesticus*, *Lepidoglyphus destructor* and *Tyrophagus putrescentiae*, and the cluster included *Dermatophagoides* mites and *Euroglyphus maynei* (Figure S3). The alignment of Blo t 2 showed that these proteins were highly similar (Figure S4).

FIGURE S3 Comparison of group 2 allergens. The description is provided in Fig S1. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was ABU97461 of *Aleuroglyphus ovatus*.

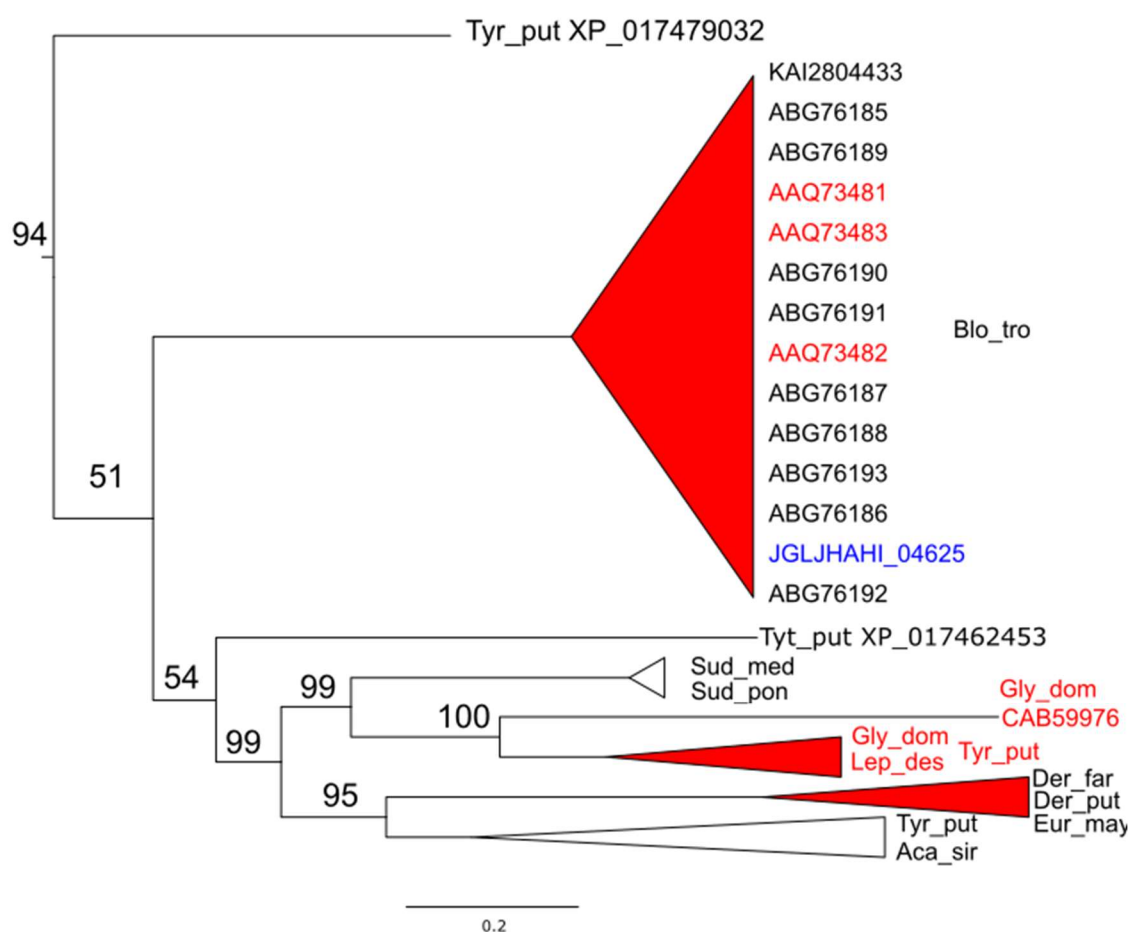


FIGURE S4 Alignment of Blo t 2.

	1	10	20	30	40
KAI2804433	MEKFFICLALLVSYAAAGDVKFTD	CAHGEVTSLDLS	GCSGDHCT	THKG	
ABG76185	MEKFFICLALLVSYAAAGDVKFTD	CAHGEVTSLDLS	GCSGDHCT	THKG	
ABG76189	MEKFFICLALLVSYAAAGDVKFTD	CAHGEVTSLDLS	GCSGDHCT	THKG	
AAQ73481	MEKFFICLALLVSSRRQRC	QI.YRSCTW	EVTSLDLS	GCSGDHCT	THKG
AAQ73483	MEKFFICLALLP	FWRA	AAGDVKFTD	CAHGEVTSLDLS	GCSGDHCT
ABG76190	MEKFFICLALLVSYAAAGDVKFTD	CAHGEVTSLDLS	GCSGDHCT	THKG	
ABG76191	MEKFFICLALLVSYAAAGDVKFTD	CAHGEVTSLDLS	GCSGDHCT	THKG	
AAQ73482	MEKFFICLALLVSPAAGDVKFTA	CAHGEVTSLDLS	GCSGDHCT	THKG	
ABG76187	MEKFFICLALLVSYAAAGDVKFTD	CAHGEVTSLDLS	GCSGDHCT	THKG	
ABG76188	MEKFFICLALLVSYAAAGDVKFTD	CAHGEVTSLDLS	GCSGDHCT	THKG	
ABG76193	MEKFFICLALLVSYAAAGDVKFTD	CAHGEVTSLDLS	GCSGDHCT	THKG	
ABG76186	MEKFFICLALLVSYAAAGDVKFTD	CAHGEVTSLDLS	GCSGDHCT	THKG	
JGLJHAHI_04625	MEKFFICLALLVSYAAAGDVKFTD	CAHGEVTSLDLS	GCSGDHCT	THKG	
ABG76192	MEKFFICLALLVSYATAGDVKFTD	CAHGEVTSLDLS	RCSGDHCT	THKG	

	50	60	70	80	90	100
KAI2804433	KSFTLKTFFIANQDSEKLEIK	ISAN	MNG	IEVPVPGVDKD	GCKHTTCPLKKGQKYELD	Y
ABG76185	KSFTLKTFFIANQDSEKLEIK	ISAT	MNG	IEVPVPGVDKD	GCKHTTCPLKKGQKYELD	Y
ABG76189	KSFTLKTFFIANQDSEKLEIK	ISAIMNN	IEVPVPGVDKD	GCKHTTCPLKKGQKYELD	Y	
AAQ73481	KSFTLKTFFIANQDSEKLEIK	ISAIMNN	IEVPVPGVDKD	GCKHTTCPLKKGQKYELD	Y	
AAQ73483	KSFTLKTFFIANQDSEKLEIK	ISAIMNN	IEVPVPGVDKD	GCKHTTCPLKKGQKYELD	Y	
ABG76190	KSFTLKTFFIANQDSEKLEIK	ISAIMNN	IEVPVPGVDKD	GCKHTTCPLKKGQKYELD	Y	
ABG76191	KSFTLKTFFIANQDSEKLEIK	ISAIMNN	IEVPVPGVDKD	GCKHTTCPLKKGQKYELD	Y	
AAQ73482	KSFTLKTFFIANQDSEKLEIK	ISAIMNN	IEVPVPGVDKD	GCKHTTCPLKKGQKYELD	Y	
ABG76187	KSFTLKTFFIANQDSEKLEIK	ISAIMNN	IEVPVPGVDKD	GCKHTTCPLKKGQKYELD	Y	
ABG76188	KSFTLKTFFIANQDSEKLEIK	ISAIMNN	IEVPVPGVDKD	GCKHTTCPLKKGQKYELD	Y	
ABG76193	KSFTLKTFFIANQDSEKLEIK	ISAIMNN	IEVPVPGVDKD	GCKHTTCPLKKGQKYELD	Y	
ABG76186	KSFTLKTFFIANQDSEKLEIK	ISAT	MNG	IEVPVPGVDKD	GCKHTTCPLKKGQKYELD	Y
JGLJHAHI_04625	KSFTLKTFFIANQDSEKLEIK	ISAIMNG	IEVPVPGVDKD	GCKHTTCPLKKGQKYELD	Y	
ABG76192	KSFTLKTFFIANQDSEKLEIK	ISAT	MNG	IEVPVPGVDKD	GCKHTTCPLKKGQKYELD	Y

	110	120	130
KAI2804433	T.....LKT...	VT	TASLVGDHGVVACGKVNT
ABG76185	SLIIPTILPNLKT	VT	TASLVGDHGVVACGKVNT
ABG76189	SLIIPTILPNLKT	VT	TASLVGDHGVVACGKVNT
AAQ73481	SLIIPTILPNLKT	VT	TASLVGDHGVVACGKVNT
AAQ73483	SLIIPTILPNLKT	VT	TASLVGDHGVVACGKVNT
ABG76190	SLIIPTILPNLKT	VT	TASLVGDHGVVACGKVNT
ABG76191	SLIIPTILPNLKT	VT	TASLVGDHGVVACGKVNT
AAQ73482	SLIIPTILPNLKT	VT	TASLVGDHGVVACGKVNT
ABG76187	SLIIPTILPNLKT	VT	TASLVGDHGVVACGKVNT
ABG76188	SLIIPTILPNLKT	VT	TASLVGDHGVVACGKVNT
ABG76193	SLIIPTILPNLKT	VT	TASLVGDHGVVACGKVNT
ABG76186	SLIIPTVLP...NVK	VT	TASLVGDHGVVACGKVNT
JGLJHAHI_04625	SLIIPTVLP...NVK	VT	TASLVGDHGVVACGKVNT
ABG76192	SLIIPTVLP...NVK	VT	TASLVGDHGVVACGKVNT

Group 3 allergens show structural homology to trypsin, e.g., Der f 3 (BAA09920) signal peptide pos 1-16 and trypsin pos 28-253. The sequence identity of Der p 3 (AAA19973) and Der f 3 (BAA09920) was 80.1%; however, their identity to the remaining allergens ranged from 45 to 60% (Table S3). Blo t 3 allergens AAM10779 and AAQ24542 presented 98.9% identity. The sequence identities of JGLJHAHI_03207 and JGLJHAHI_02083 vs. Blo t 3 (AAM10779) were 62.0% and 49.5%, respectively. Predicted sequences of *Blomia* formed two separate clusters supported by other known sequences from GenBank. These two clusters were outside the Tyr p3 and Blo t 3 clusters (Figure S5). Alignment of Blo t 3 showed many conserved regions (Figure S6). As the third isoform, we identified the AAQ24542 (A1KXI1) protein in the proteome analysis. However, we did not find evidence of AAM10779 (Q8I916) in the proteome analyses.

FIGURE S5 Phylogenetic tree of group 3 allergen isoforms of domestic mites. Red indicates known allergen proteins, blue indicates predicted proteins of *Blomia tropicalis* (this work). The outgroup was *Acarus siro* ABL09311 (not shown).

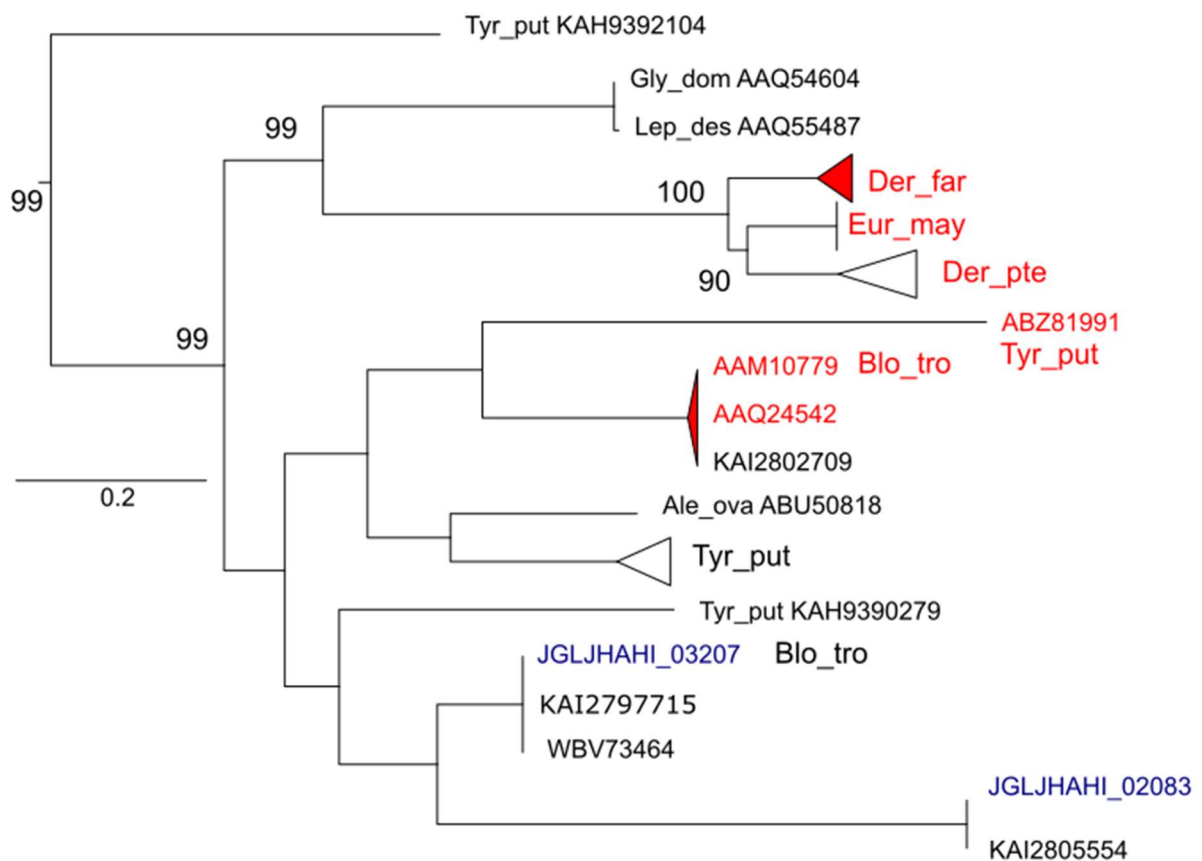
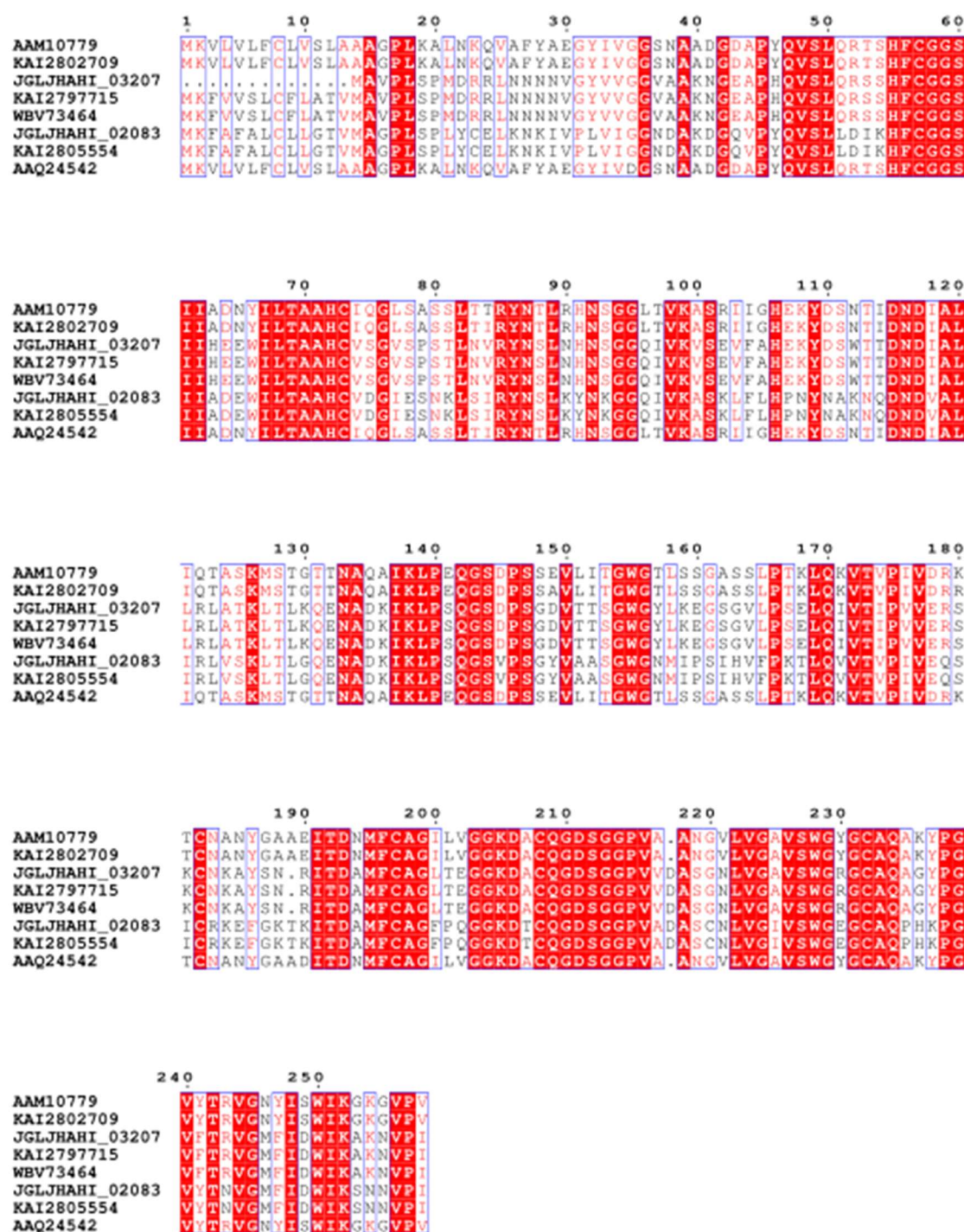


FIGURE S6 Alignment of Blo t 3 sequences



Group 4 allergens showed structural homology to alpha amylase, e.g., Der f 4 (AHX03180) contained signal peptide 1-23, alpha amylase, catalytic domain 62-327 and alpha amylase, and a C-terminal all-beta domain (429-520). Der f 4 (AHX0318) and Der p 4 (AAD38942) showed 88.2% identity. Blo t 4 (AAQ24543) showed 69.9% sequence identity to Der p 4 (AAD38942), 73.2% identity to Tyr p 4 (ABM537540) and 65.1% identity to Der f 4 (AHX03180). Predicted JGLJHAHI_01112 showed 99% identity to Blo t 4 (Table S4). The Blo t 4 alignment indicated highly similar proteins; however, the KAI2801484 sequence contained an incorrectly predicted start and end of the sequence (Figure S7).

FIGURE S7 Alignment of Blo t 4 sequences.

```

JGLJHAHI_01112 .....
AAQ24543 .....
KAI2795383 .....
KAI2795737 .....
KAI2801484 TKNVSFGLVGKKQDASYSAFLEFRKLRTVYDPTKYNNVLKCKKNKPQEAINDKENGLIGCLR
KAI2803944 .....

```

```

          1
JGLJHAHI_01112 .....MIVEL.....
AAQ24543 .....MIVEL.....
KAI2795383 .....M.....
KAI2795737 .....MIVEL.....
KAI2801484 LEPGGAFIRCIFDLGDSNKSPPFNEIGYFFENVYQKGMWGAFRYLSLPKNETVDTQFAYRG
KAI2803944 .....MIVEL.....

```

```

JGLJHAHI_01112 .....
AAQ24543 .....
KAI2795383 .....
KAI2795737 .....
KAI2801484 DLSSFESEHKYFDKIRKADQIEGLLCDVVCSSLNFKMLATGRIQPGPESVGEFSGRRRDN
KAI2803944 .....

```

```

          10          20          30          40          50
JGLJHAHI_01112 .....IVFSLALTIPYQTLASSPYSDPHEQHNRKVITHLMQWKFVDIADECERFLGPGY
AAQ24543 .....IVFSLALTIPYQTLASSPYSDPHEQHNRKVITHLMQWKFVDIADECERFLGPGY
KAI2795383 .....IVFSLALTIPYQTLASSPYSDPHEQHNRKVITHLMQWKFVDIADECERFLGPGY
KAI2795737 .....IVFSLALTIPYQTLASSPYSNPHQHNRKVITHLMQWKFVDIADECERFLGPGY
KAI2801484 GKRVMMYIVFSLALTIPYQTLASSPYSDPHEQHNRKVITHLMQWKFVDIADECERFLGPGY
KAI2803944 .....IVFSLALTIPYQTLASSPYSDPHEQHNRKVITHLMQWKFVDIADECERFLGPGY.

```

```

          60          70          80          90          100          110
JGLJHAHI_01112 YGGVQVSPVNEHASLDRHPWYELYQPVSYRIVSRSGTESQFRDMVHRCNKAGVRIYVDVV
AAQ24543 YGGVQVSPVNEHASLDRHPWYELYQPVSYRIVSRSGTESQFRDMVHRCNKAGVRIYVDVV
KAI2795383 .....PWYELYQPVSYRIVSRSGTESQFRDMVHRCNKAGVRIYVDVV
KAI2795737 YGGVQVSPVNEHASLDRHPWYELYQPVSYRIVSRSGTESQFRDMVHRCNKAGVRIYVDVV
KAI2801484 YGGVQVSPVNEHASLDRHPWYELYQPVSYRIVSRSGTESQFRDMVHRCNKAGVRIYVDVV
KAI2803944 .....VNEHASLDRHPWYELYQPVSYRIVSRSGTESQFRDMVHRCNKAGVRIYVDVV

```

```

          120          130          140          150          160          170
JGLJHAHI_01112 LNHMTGPQSGVGIDGTHYDGNSMQYPGIPFGPNDFHGHESCPTSNLDIQNYDDPTQARNC
AAQ24543 LNHMTGPQSGVGIDGTHYDGNSMQYPGIPFGPNDFHGHESCPTSNLDIQNYDDPTQARNC
KAI2795383 LNHMTGPQSGVGIDGTHYD.....DIQNYDDPTQARNC
KAI2795737 LNHMTGPQSGVGIDGTHYDGNSMQYPGIPFGPNDFHGHESCPTSNLDIQNYDDPTQARNC
KAI2801484 LNHMTGPQSGVGIDGTHYDGNSMQYPGIPFGPNDFHGHESCPTSNLDIQNYDDPTQARNC
KAI2803944 LNHMTGPQSGVGIDGTHYDGNSMQYPGIPFGPNDFHGHESCPTSNLDIQNYDDPTQARNC

```

```

          180          190          200          210          220          230
JGLJHAHI_01112 RLSGLRDLKQSADYVRTKQADFLNHLIDIGVACFRFDASKHMWPCDLQAIYSKLHHLNEK
AAQ24543 RLSGLRDLKQSADYVRTKQADFLNHLIDIGVACFRFDASKHMWPCDLQAIYSKLHHLNEK
KAI2795383 RLSGLRDLKQSADYVRTKQADFLNHLIDIGVACFRFDASKHMWPCDLQAIYSKLHHLNEK
KAI2795737 RLSGLRDLKQSADYVRTKQADFLNHLIDIGVACFRFDASKHMWPCDLQAIYSKLHHLNEK
KAI2801484 RLSGLRDLKQSADYVRTKQADFLNHLIDIGVACFRFDASKHMWPCDLQAIYSKLHHLNEK
KAI2803944 RLSGLRDLKQSADYVRTKQADFLNHLIDIGVACFRFDASKHMWPCDLQAIYSKLHHLNEK

```

```

          240          250          260          270          280          290
JGLJHAHI_01112 YFPSNSNPFYIYHEVIYSNNNAISISDYTKLGRSIEFHYYHELCNVIRGNNKLKWLHNFGQ
AAQ24543 YFPSNSNPFYIYHEVIYSNNNAISISDYTKLGRSIEFHYYHELCNVIRGNNKLKWLHNFGQ
KAI2795383 YFPSNSNPFYIYHEVIYSNNNAISISDYTKLGRSIEFHYYHELCNVIRGNNKLKWLHNFGQ
KAI2795737 YFPSNSNPFYIYHEVIYSNNNAISISDYTKLGRSIEFHYYHELCNVIRGNNKLKWLHNFGQ
KAI2801484 YFPSNSNPFYIYHEVIYSNNNAISISDYTKLGRSIEFHYYHELCNVIRGNNKLKWLHNFGQ
KAI2803944 YFPSNSNPFYIYHEVIYSNNNAISISDYTKLGRSIEFHYYHELCNVIRGNNKLKWLHNFGQ

```

FIGURE S7 continuation.

	300	310	320	330	340	350
JGLJHAHI_01112	PWGMVFNDDALIMVDSHDLQRCGHTGQLGLNINYPESRLLKVATAFMLAWPYGIPRVMSSY					
AAQ24543	PWGMVFNDDALIMVDSHDLQRCGHTGQLGLNINYPESRLLKVATAFMLAWPYGIPRVMSSY					
KAI2795383	PWGMVFNDDALIMVDSHDLQRCGHTGQLGLNINYPESRLLKVATAFMLAWPYGIPRVMSSY					
KAI2795737	PWGMVFNDDALIMVDSHDLQRCGHTGQLGLNINYPESRLLKVATAFMLAWPYGIPRVMSSY					
KAI2801484	PWGMVFNDDALIMVDSHDLQRCGHTGQLGLNINYPESRLLKVATAFMLAWPYGIPRVMSSY					
KAI2803944	PWGMVFNDDALIMVDSHDLQRCGHTGQLGLNINYPESRLLKVATAFMLAWPYGIPRVMSSY					

	360	370	380	390	400	410
JGLJHAHI_01112	RWNQKIVNCKDENDWIGPPTDGSGSILSVKPNSDLTCNQEWICEHRWQIYSMVQFRNTA					
AAQ24543	RWNQKIVNCKDENDWIGPPTDGSGSILSVKPNSDLTCNQEWICEHRWQIYSMVQFRNTA					
KAI2795383	RWNQKIVNCKDENDWIGPPTDGSGSILSVKPNSDLTCNQEWICEHRWQIYSMVQFRNTA					
KAI2795737	RWNQKIVNCKDENDWIGPPTDGSGSILSVKPNSDLTCNQEWICEHRWQIYSMVQFRNTA					
KAI2801484	RWNQKIVNCKDENDWIGPPTDGSGSILSVKPNSDLTCNQEWICEHRWQIYSMVQFRNTA					
KAI2803944	RWNQKIVNCKDENDWIGPPTDGSGSILSVKPNSDLTCNQEWICEHRWQIYSMVQFRNTA					

	420	430	440	450	460	470
JGLJHAHI_01112	GDEPVKNWWDNGDYQIAFSRGAKTFIAINLQNGQHLKQNLHTGLPSGNYCDLVTGNV					
AAQ24543	GDEPVKNWWDNGDYQIAFSRGAKTFIAINLQNGQHLKQNLHTGLPSGNYCDLVTGNV					
KAI2795383	GDEPVKNWWDNGDYQIAFSRGAKTFIAINLQNGQHLKQNLHTGLPSGNYCDLVTGNV					
KAI2795737	GDEPVKNWWDNGDYQIAFSRGAKTFIAINLQNGQHLKQNLHTGLPSGNYCDLVTGNV					
KAI2801484	GDEPVKNWWDNGDYQIAFSRGAKTFIAINLQNGQHLKQNLHTGLPSGNYCDLVTGNV					
KAI2803944	GDEPVKNWWDNGDYQIAFSRGAKTFIAINLQNGQHLKQNLHTGLPSGNYCDLVTGNV					

	480	490	500	510
JGLJHAHI_01112	KCSGQMVHVDGSGHAMISIGANADDPFIALSVDARV			
AAQ24543	KCSGQMVHVDGSGHAMISIGANADDPFIALSVDARV			
KAI2795383	KCSGQMVHVDGSGHAMISIGANADDPFIALSVDARV			
KAI2795737	KCSGQMVHVDGSGHAMISIGANADDPFIALSVDARV			
KAI2801484	KCSGQMVHVDGSGHAMISIGANADDPFIALSVDARV			
KAI2803944	KCSGQMVHVDGSGHAMISIGANADDPFIALSVDARV			

JGLJHAHI_01112	
AAQ24543	
KAI2795383	
KAI2795737	
KAI2801484	SAAGGVGQAAIRICQNKNCIEFVTVGNNKEKREFIKREFGIPDENIFNSRDISFERLIKER
KAI2803944	

JGLJHAHI_01112	
AAQ24543	
KAI2795383	
KAI2795737	
KAI2801484	TKSGVDIVLNSLAEEKLQASLRCLGLNGRFLFLEIKGYDMQMNKQIGMFAFLQNISFHGVC
KAI2803944	

JGLJHAHI_01112	
AAQ24543	
KAI2795383	
KAI2795737	
KAI2801484	LDSYFREGFDSPLKRFMRKMTLELLKDGIANGVVRPIQRTSYAWNQVEEAFRYMSTGKIG
KAI2803944	

JGLJHAHI_01112	
AAQ24543	
KAI2795383	
KAI2795737	
KAI2801484	KVIKMRDEEKEKITIAPLMVQAIRRTWFNPSKSYIIVGGLGGFGLLELLYWMVMRGAKK
KAI2803944	

FIGURE S7 continuation

JGLJHAHI_01112
AAQ24543
KAI2795383
KAI2795737
KAI2801484 F I L S S R N G F K N S Y Q K L Y F K R L K E L S N Y V S I F D N D V M V T T S N V T T V D G C E R L I E E A I E K G P
KAI2803944

JGLJHAHI_01112
AAQ24543
KAI2795383
KAI2795737
KAI2801484 I G G I F N L A L V L H D A Y L E N Q T V E T F K D V C N P K L T G L S N L D K I S R T K C P E W N I S G R G N A G Q V
KAI2803944

JGLJHAHI_01112
AAQ24543
KAI2795383
KAI2795737
KAI2801484 N Y G Y A N S A M E R I C E M R R F K G L P G H A I Q W G P I G D V G V V A D Q L V G E D D T K E D L L K L L G G I S L
KAI2803944

JGLJHAHI_01112
AAQ24543
KAI2795383
KAI2795737
KAI2801484 Q R I S S C L E V L D R F L Q I P I P A L S S V I K S N L E D S S K K S D D D I L D Q L C A N V D K R N R D E T I G E A
KAI2803944

JGLJHAHI_01112
AAQ24543
KAI2795383
KAI2795737
KAI2801484 G L D S M T V V E I Q Q R L E R D Y D V S F S V N D V K R I T I G E L K D F R D G K R E A L K Q F S Q D L K Q A R M A T
KAI2803944

JGLJHAHI_01112
AAQ24543
KAI2795383
KAI2795737
KAI2801484 I K F E I A T E P Y T I V N Q T G K L D N A S K P I E F L P P I E G I Y E S L K E L I S E L A K L R P V I T L N W V R K
KAI2803944

JGLJHAHI_01112
AAQ24543
KAI2795383
KAI2795737
KAI2801484 M S E L K T V K K V A E Y F R G L M L K F E P K G D H E I I G H S F G A I V G I C R K A V P V K T L I A M D P M D V S N
KAI2803944

JGLJHAHI_01112
AAQ24543
KAI2795383
KAI2795737
KAI2801484 F K E E Y D R N E K V E L V L S Y L R N F L S Q R L V E K L R K E A L S I K T E A E C I Q R I T D A L K S H G G K T V V
KAI2803944

FIGURE S7 continuation

```
JGLJHAHI_01112 .....  
AAQ24543 .....  
KAI2795383 .....  
KAI2795737 .....  
KAI2801484 GKDIDEI IKDSFERAELVLKYQQKNIRK LKMF GREMTQKTMQKLLLHKITVNVTVLKRLN  
KAI2803944 .....
```

```
JGLJHAHI_01112 .....  
AAQ24543 .....  
KAI2795383 .....  
KAI2795737 .....  
KAI2801484 QESDMDKLNNILLTSYGINRQDFVGRFDV FQVIGDEEEYLTPFI IETIKTKLEM  
KAI2803944 .....
```

Group 5 was described as the Blo t-5 domain group, i.e., Der f 5 (ABO84970) contained a signal peptide (pos 1-19) and Mite allergen Blo t 5 (pos 15-132). The identity of Blo t 5 (AAD10850) to Der f 5 (ABO84970) was 46.2%, Der p 5 (CAA35692) was 43.3% and Lep d 5 (CAB62212) was 48.2%. The predicted *Blomia tropicalis* protein JGLJHAHI_14490 showed 100% identity to Blo t 5 (AAD10850). The AAB49396 sequence presented lower similarity to *Blomia tropicalis* allergens and predicted allergens. This was caused by the missing first part of the protein (Figure S9).

FIGURE S8 Comparison of group 5 allergens. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was AAX34051 of *Suidasia medianensis*.

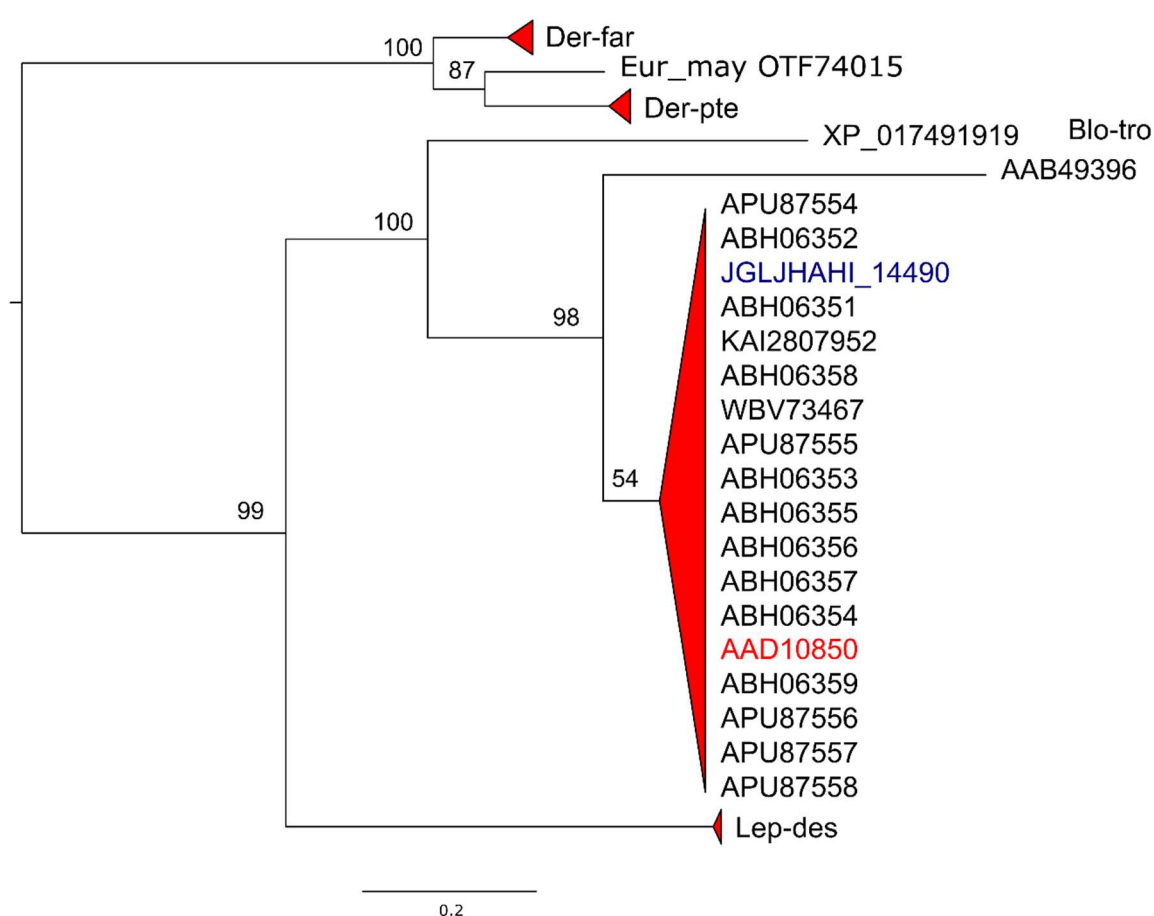
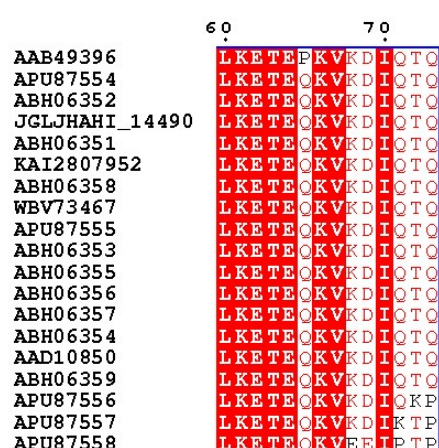
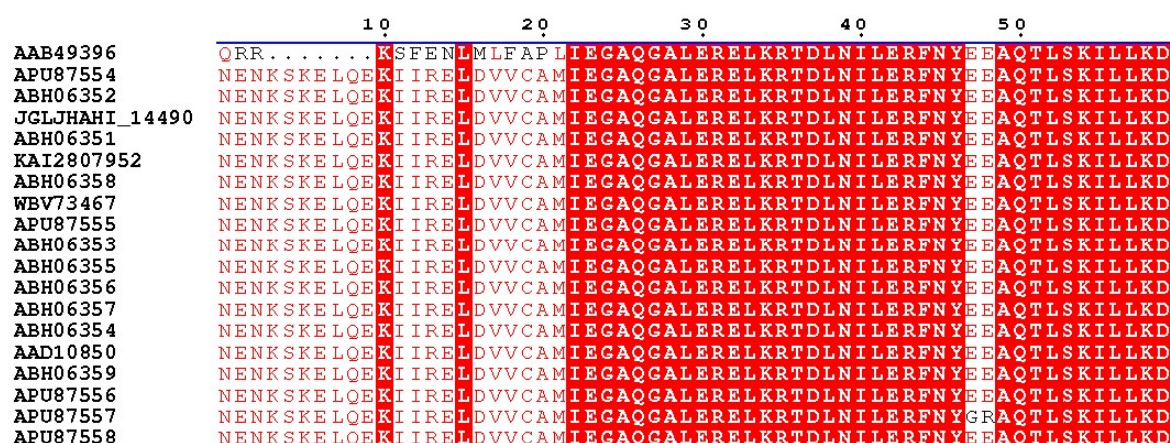
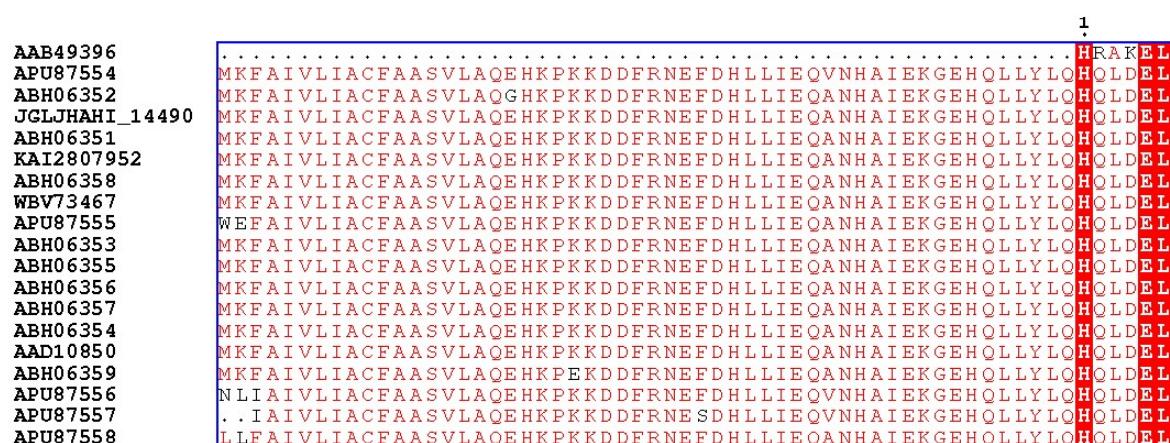
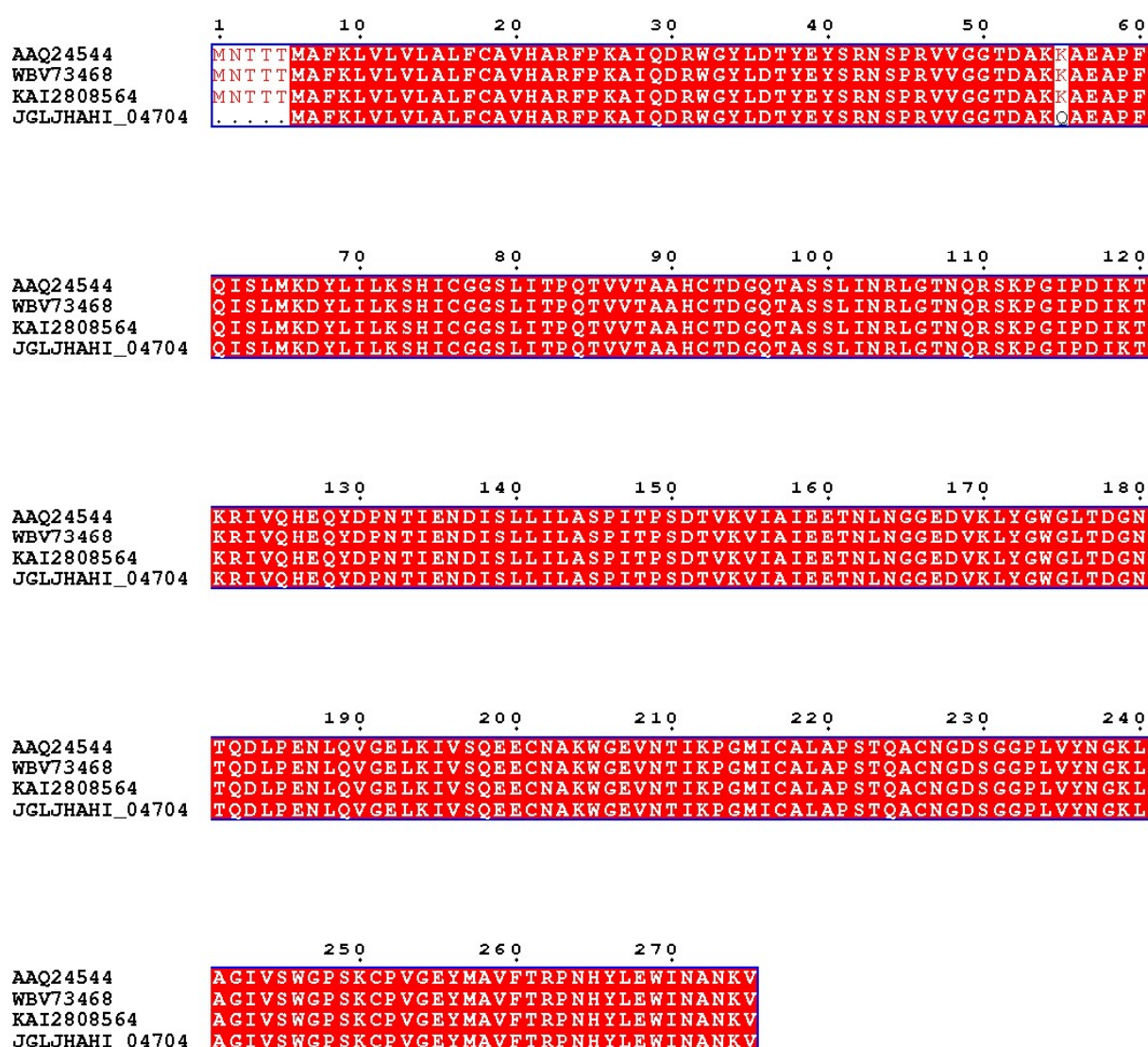


FIGURE S9 Alignment of Blo t 5.



Group 6 allergens showed structural homology to trypsin, i.e., Der f 6 (AAF28423) signal peptide 1-19 and trypsin 50-274. Der f 6 (AAF28423) and Blo t 6 (AAQ24544) showed 55.6% identity. The predicted *Blomia tropicalis* protein JGLJHAHI_04704 showed 99.6% identity to Blo t 6 (AAQ24544) (Table S6), as shown in the Blo t 6 alignment (Figure S10).

FIGURE S10 Alignment of Blo t 6 allergens



Group 7 allergens contained a protein with the Grp 7 allergen domain, i.e., Der f 7 (AAB35977) signal peptide 1-17 and group 7 allergen 28-207. The sequence identity between Der f 7 (AAB35977) and Der p 7 (AAA80264) was 85.9%, while Der f 7 (AAB35977) and Blo t 7 (ASX95438) showed 28.4% identity. Blo t 7 (ASX95438) and Tyr p 7 (ABM53750) showed 50.3% identity (Table S7). Predicted *Blomia tropicalis* proteins JGLJHAHI_09763 and JGLJHAHI_05732 showed 100 and 45% identity to Blo t 7 (ASX95438). JGLJHAHI_09763 was included in a cluster with the Blo t 7 protein (ASX95438) (Figure S11). JGLJHAHI_05732 formed a cluster located outside of the Blo t 7 proteins. Both predicted *Blomia tropicalis* sequences showed 4 identical highly conserved areas (Figure S12).

FIGURE S11 Comparison of group 7 allergens. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was KPM06925 of *Sarcoptes scabiei*.

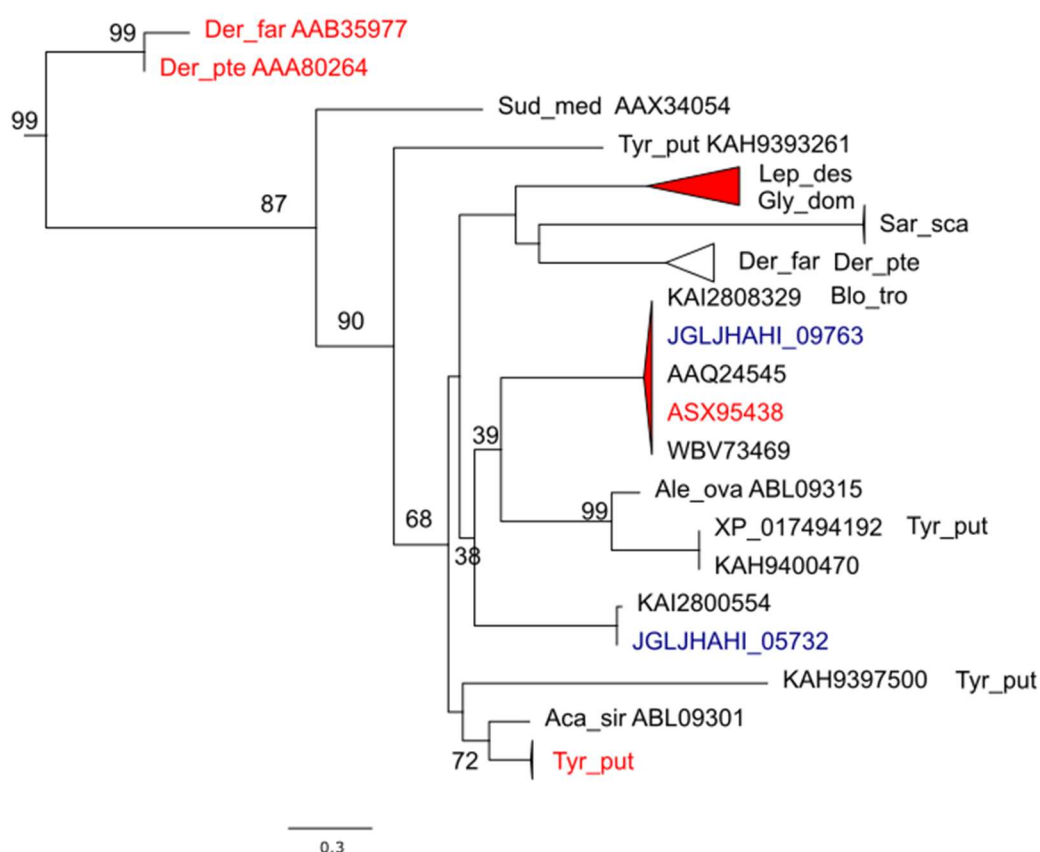


FIGURE S12 Alignment of Blo t 7 allergens.

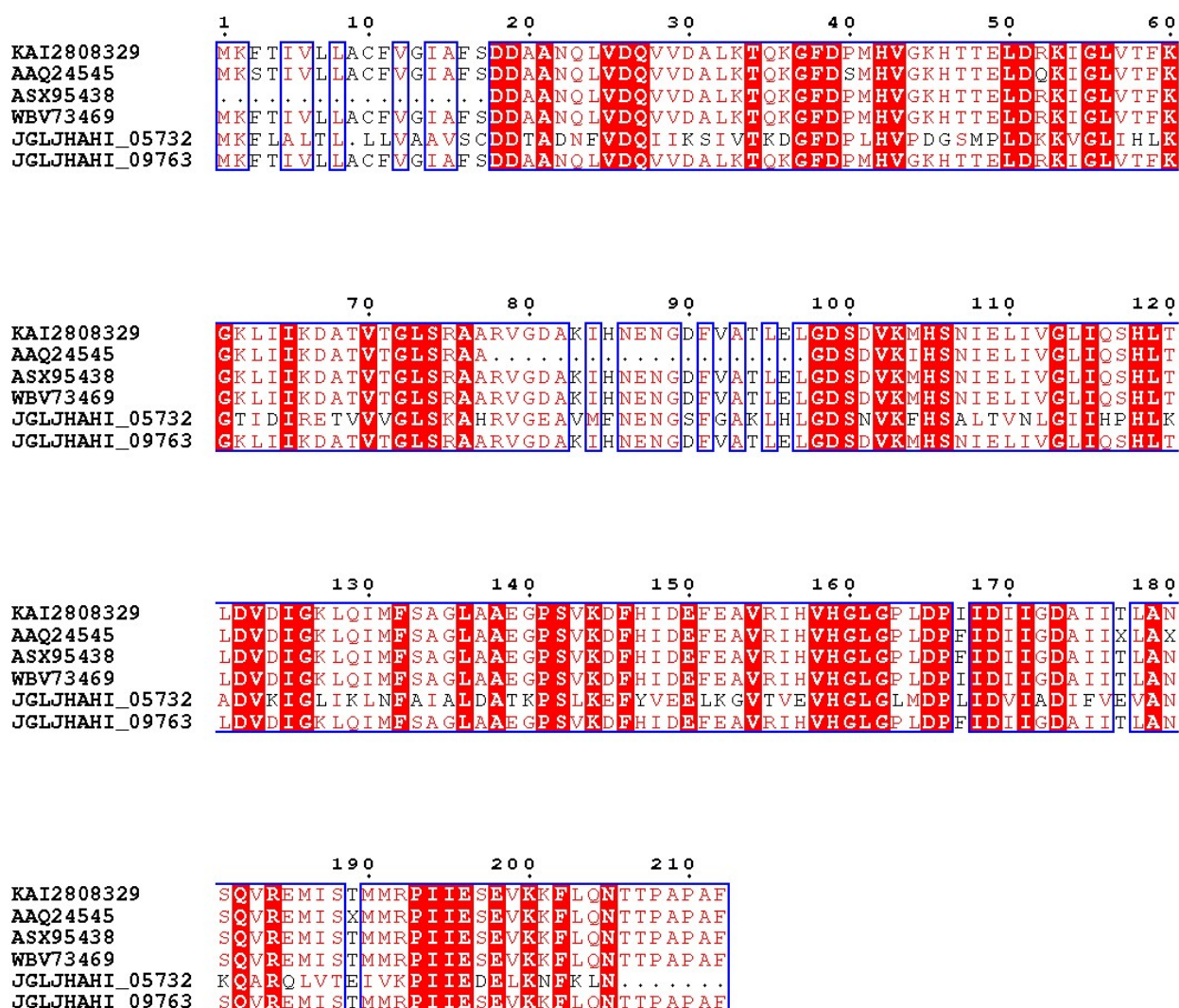


FIGURE S14 Alignment of Blo t 8 allergens.

	1	10	20	30	40	50	60
KAI2801258	MAPK	LCYWDIRGL	GEPIRMML	KYLN	IEFEEVRY	GFGDDKAESFPS	RDEWLAEKFNLCMDF
KAI2802743	...	KIGYWDIRGF	FAEPIRMML	KHLN	IEFEEKRY	GFGNESEASFPH	RDEWLAEKFNLCGLDF
ACV04860	MAPK	IGYWDVRCF	FAEPIRMML	KHLN	IEFEETRY	GFGNDSEESLPN	RDEWLAEKFTLCLEF
AAP35069	MAPK	IGYWDIRGF	FAEPIRMML	KHLN	IEFEETRY	GFGNDSEESFPN	RDEWLAEKFTLCFEF
JGLJHAHI_17878	KTIT	LCYWNIQNA	QPIRLLL	AYTK	TNYVERN	YNS.....	EMDRAEWLKEKFDLGLDF
JGLJHAHI_15972	...	ILCYWDIRGL	FAEPIRMML	AYGE	IAYEDKRY	T.....	DRDVWLAEKFTLCGLDF

	70	80	90	100	110	120					
KAI2801258	PNLPY	LI	EGDLKMT	QSVS	ILKHLAR	ANGLVVTTE	PAQTELEMF	EAI	VLDI	LHI	LVFTIYS
KAI2802743	PNLPY	LI	DGDFKMT	ESVA	ILKRRLAR	ANGMIATTE	PALSHSEMF	EAM	IIDI	RNR	LSHVNYI
ACV04860	PNLPY	LF	DGDFKMT	ESVA	ILKRRLAR	ANGMIATTE	PALSYSEMI	EAM	IIDI	RNR	LINVVYA
AAP35069	PNLPY	LF	DGDFKMT	ESVA	ILKRRLAR	ANGMIATTE	PALSYSEMI	EAM	VIDI	RNR	LVYVIYA
JGLJHAHI_17878	PNLPY	YMEGD	LRIS	QSLT	ILRYLAD	KHDLA	GSNEEEK	IRIDL	CEQLR	DYR	QFIRL.A.
JGLJHAHI_15972	PNLPY	FF	DGDVKLS	QTLA	ILRYLAR	KLNLD	GADCKEK	VTISLV	EQQIV	DLNMAM	GRIED.

	130	140	150	160	170		
KAI2801258	D...TAEDFEKAMETVRFM	LAS	SLKQLDE	WFDKNGCK	WLGRDQ	LT	YVDFMAYEYLDWYRV
KAI2802743	DNKGTAEEFDKKLGAIRER	LQTT	LEQMES	FFTKHGNQ	WIAGDK	LT	YVDFMAYEYLDWYRV
ACV04860	ENSGTPEEFQKLADLRER	LET	SLGQLEA	FFQKHGSQ	WVAGDK	LT	YVDFLAYEYLDWYRV
AAP35069	ENSGTPEEFQKLADLRER	LET	SLGQLEA	FFQKHGSQ	WVAGDK	LT	YVDFLAYEYLDWYRV
JGLJHAHI_17878	YLSQ	LPEKLR	SLSV	FLD..RRP	FFAGNS	IS	YVDFMAYEYIDHF..Y
JGLJHAHI_15972	YLKA	LPDQLK	LVSN	ELG..EKP	FAAGAN	LS	YVDFVMYEYLLKKVSV

	180	190	200	210	220	230				
KAI2801258	FAKSKP	IFEK	FPNV	SA	YMKRF	EE	LPNLKEYI	IASDQHRLADC	VSPFVR	I GHRWAKE
KAI2802743	FAKATP	IFKK	LPKI	SE	YMKRF	EE	LPNLSKEYI	IASDAHRLASC	VSPFART	I GHRWAKE
ACV04860	FVKSTP	IFEK	FAKV	SD	YMKRF	EE	LPNLSKEYI	IARDEHRSASC	LSPFAR	I GHRWAKE
AAP35069	FVKSTP	IFEK	FAKV	SD	YMKRF	EE	LPNLSKEYI	IASDEHRSASC	LSPFAR	I GHRWAKE
JGLJHAHI_17878	..LYPE	LF	GKLSN	LKQ	FLRRLES	LPTIREY	QKKI	IENSSA		
JGLJHAHI_15972	..LVP	..GS	LDNL	LAK	EVGR	IES	LPKVAQYI			

FIGURE S15 Alignment of Blo t 10.

KAI2797771
 ABU97466
 JGLJHAHI_15939

1 10 20 30 40 50 60
 MEAIKKKKMQAMKLEKDNAIDRAETAEQKSRDANLRAEKSEEEVRLQKKIQQTENELDQV
 MEAIKKKKMQAMKLEKDNAIDRAETAEQKSRDANLRAEKSEEEVRLQKKIQQTENELDQV
 MEAIKKKKMQAMKLEKDNAIDRAETAEQKSRDANLRAEKSEEEVRLQKKIQQTENELDQV

70 80 90 100 110 120
 QESLTQANTKLEEKEKSLQTAEGDVAALNRRIQLIEEDLERSEERLKVATAKLEEASHSA
 QESLTQANTKLEEKEKSLQTAEGDVAALNRRIQLIEEDLERSEERLKVATAKLEEASHSA
 QESLTQANTKLEEKEKSLQTAEGDVAALNRRIQLIEEDLERSEERLKVATAKLEEASHSA

130 140 150 160 170 180
 DESERM RKMLEHRSITDEERMDGLESQ LKEARMM AEDADRKYDEVARKLAMVEADLERAE
 DESERM RKMLEHRSITDEERMDGLESQ LKEARMM AEDADRKYDEVARKLAMVEADLERAE
 DESERM RKMLEHRSITDEERMDGLESQ LKEARMM AEDADRKYDEVARKLAMVEADLERAE

190 200 210 220 230 240
 ERAETGETKIVELEELRVVGN NLKSLEVSEEKAQQREEAYEQQIRM MTGKLKEAEARAE
 ERAETGETKIVELEELRVVGN NLKSLEVSEEKAQQREEAYEQQIRM MTGKLKEAEARAE
 ERAETGETKIVELEELRVVGN NLKSLEVSEEKAQQREEAYEQQIRM MTGKLKEAEARAE

250 260 270 280
 FAERSVOKLQKEVDRLEDELVHEKEKEYKSISDEL DQTF AELTGY
 FAERSVOKLQKEVDRLEDELVHEKEKEYKSISDEL DQTF AELTGY
 FAERSVOKLQKEVDRLEDELVHEKEKEYKSISDEL DQTF AELTGY

Group 11 included proteins with structural homology to Myosin tail 1. The allergens and predicted allergens were very similar among species, with the lowest sequence identity of 86% (Table S10). The predicted *Blomia tropicalis* protein JGLJHAHI_05065 showed 99.7% identity to Blo t 10 (AAM83103).

FIGURE S16 Alignment of Blo t 11.

AAM83103	1	10	20	30	40	50	60
KAI2806507	MAARSAKMYQSSRAGHGGDISIEYCTDLGALTRLEDKIRLLSEDLSESEELRQRVEREK						
JGLJHAHI_05065	MAARSAKMYQSSRAGHGGDISIEYCTDLGALTRLEDKIRLLSEDLSESEELRQRVEREK						
AAM83103	70	80	90	100	110	120	
KAI2806507	SDITVQLMNLTERLEETEGSSESVTENMKKRDSELAKLRLLEDVHMESEETAHHLRQKH						
JGLJHAHI_05065	SDITVQLMNLTERLEEAEGSSESVTENMKKRDSELAKLRLLEDVHMESEETAHHLRQKH						
AAM83103	130	140	150	160	170	180	
KAI2806507	QAAIQEMQDQLDQVQKAKNKSDEKQKFOAEVVFELLAQVETANKDKLVAQKTVEKLEYTV						
JGLJHAHI_05065	QAAIQEMQDQLDQVQKAKNKSDEKQKFOAEVVFELLAQVETANKDKLVAQKTVEKLEYTV						
AAM83103	190	200	210	220	230	240	
KAI2806507	HELNKIEEINRTVVEVTAHRQRLSQENSELIKEVHEYKISLDNANHLCQIAQQLEDTR						
JGLJHAHI_05065	HELNKIEEINRTVVEVTAHRQRLSQENSELIKEVHEYKISLDNANHLCQIAQQLEDTR						
AAM83103	250	260	270	280	290	300	
KAI2806507	HRLEDEERKRSSLENHAHTLEVELESLKVQLEEESEARLELERQLTKANGDAASWKSKEYE						
JGLJHAHI_05065	HRLEDEERKRSSLENHAHTLEVELESLKVQLEEESEARLELERQLTKANGDAASWKSKEYE						
AAM83103	310	320	330	340	350	360	
KAI2806507	AELQAHVDEVEELRRKMAQKISEYEQLLEALLNKCSALEKQKARLQSEVEVLIMDLEKAT						
JGLJHAHI_05065	AELQAHVDEVEELRRKMAQKISEYEQLLEALLNKCSALEKQKARLQSEVEVLIMDLEKAT						
AAM83103	370	380	390	400	410	420	
KAI2806507	AHAQALEKRVVSQLEKINLDLKSKEEVSMLEQTQKDLRVKIADLQRLQHEYEKLRDQKE						
JGLJHAHI_05065	AHAQALEKRVVSQLEKINLDLKSKEEVSMLEQTQKDLRVKIADLQRLQHEYEKLRDQKE						
AAM83103	430	440	450	460	470	480	
KAI2806507	ALARENKKLADDLAEAKSQLNDAHRRHEQETIEIKRLENEREELAAAYKEAETLRKQEEA						
JGLJHAHI_05065	ALARENKKLADDLAEAKSQLNDAHRRHEQETIEIKRLENEREELAAAYKEAETLRKQEEA						
AAM83103	490	500	510	520	530	540	
KAI2806507	KNQRLTAELAQTRHDYEKRLAQKEEIEALRKQYQIEIEQLNMRLAEAEAKLKTEVARLK						
JGLJHAHI_05065	KNQRLTAELAQTRHDYEKRLAQKEEIEALRKQYQIEIEQLNMRLAEAEAKLKTEVARLK						
AAM83103	550	560	570	580	590	600	
KAI2806507	KKYQAQITAELESLDAANKANIDLQKTIKKQALQITGLQAHYDEVHRLQQAQVDQLGVTQ						
JGLJHAHI_05065	KKYQAQITAELESLDAANKANIDLQKTIKKQALQITEGLQAHYDEVHRLQQAQVDQLGVTQ						
AAM83103	610	620	630	640	650	660	
KAI2806507	RRCQALTAEEEMRVNLEQALRAKRAAEQMHEEAVVRVNELTITINVNLASAKSKLETEFS						
JGLJHAHI_05065	RRCQALTAEEEMRVNLEQALRAKRAAEQMHEEAVVRVNELTITINVNLASAKSKLETEFS						

FIGURE S16 continuation

	670	680	690	700	710	720
AAM83103	ALQNDYDEVHKELRISDERVQKLTIEVKSTKDLESETERVTKLETIKKSLETEVRNLQI					
KAI2806507	ALQNDYDEVHKELRISDERVQKLTIEVKSTKDLESETERVTKLETIKKSLETEVRNLQI					
JGLJHAHI_05065	ALQNDYDEVHKELRISDERVQKLTIEVKSTKDLESETERVTKLETIKKSLETEVRNLQI					

	730	740	750	760	770	780
AAM83103	RIEEVEANALAGGKRVIAKLESRI RDVEIEVEEERRRHAETEKMLRKKDHRVKELLLQNE					
KAI2806507	RIEEVEANALAGGKRVIAKLESRI RDVEIEVEEERRRHAETEKMLRKKDHRVKELLLQNE					
JGLJHAHI_05065	RIEEVEANALAGGKRVIAKLESRI RDVEIEVEEERRRHAETEKMLRKKDHRVKELLLQNE					

	790	800	810	820	830	840
AAM83103	EDHKQIQLLQEMSDKLNKVKVYKROMQEQEGMSQQNLTRVRRFQRELEAAEDRADQAES					
KAI2806507	EDHKQIQLLQEMSDKLNKVKVYKROMQEQEGMSQQNLTRVRRFQRELEAAEDRADQAES					
JGLJHAHI_05065	EDHKQIQLLQEMSDKLNKVKVYKROMQEQEGMSQQNLTRVRRFQRELEAAEDRADQAES					

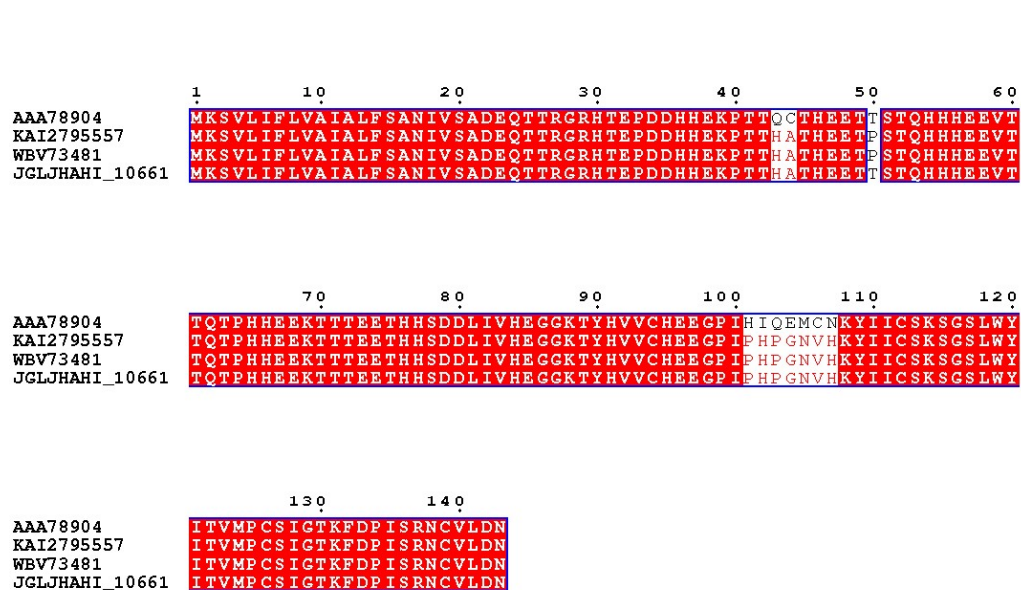
	850	860	870
AAM83103	NLSFIRAKHRSWVTTSQVPGGTRQVFVTEQESSNF		
KAI2806507	NLSFIRAKHRSWVTTSQVPGGTRQVFVTEQESSNF		
JGLJHAHI_05065	NLSFIRAKHRSWVTTSQVPGGTRQVFVTEQESSNF		

Only **Blo t 12** was identified in this group of allergens. Blo t 12 (AAA78904) contained signal peptide 1-21 and Chitin binding Peritrophin-A domain 95-144. The predicted *Blomia tropicalis* protein JGLJHAHI_10661 shared 91.2% identity with Blo t 12 (AAA78904) (Table S11). The comparison of allergens showed that Blo t 12 clustered with *Blomia tropicalis* proteins and *L. destructor* proteins outside of *Tyrophagus putrescentiae* proteins (Figure S17). The alignment of Blo t 12 showed a high degree of conservation (Figure S18).

FIGURE S17 Comparison of group 12 allergens. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence UXW65971 from *Tyrophagus putrescentiae*.



FIGURE S18 Alignment of Blo t 12 allergens.



Group 13 allergens showed structural homology to lipocalin; i.e., Der f 13 (AAP35078) contained the lipocalin/cytosolic fatty-acid binding protein family 5-131. The signal peptide was missing. Der f 13 (AAP35078) and Der p 13 (ADK92390) showed 95.4% sequence identity. Der f 13 (AAP35078) and Blo t 13 (AAC80579) showed 82.0% sequence identity (Table S12). Aca s 13 (ABL09307) and Blo t 13 (AAC80579) showed 65.4% sequence identity. The following peptides were predicted from the *Blomia tropicalis* transcriptome: JGLJHAHI _00046, JGLJHAHI _00101, JGLJHAHI _14091 and JGLJHAHI _12788 and showed structural homology to lipocalin. JGLJHAHI _00101 showed a low level of expression and was suggested to be an artifact. The proteins were grouped into different clusters (Figure S6). JGLJHAHI _14091 formed a separate cluster from known mite allergens, although it showed 56.3% identity to Blo t 13 (AAC80579). JGLJHAHI _12788 formed a separate cluster near Aca s 13 (ABL09307). The protein shared 61.7% identity with Blo t 13 (AAC80579) and 84.6% identity with Aca s 13 (ABL09307). JGLJHAHI _00046 showed the highest identity (91.5%) to Blo t 13 (AAC80579) and belonged to the same cluster (Figure S19). In the proteome analyses, JGLJHAHI _00046 and JGLJHAHI _12788 were identified as the same protein. The Blo t 13 alignment indicated high structural homology among sequences (Figure S20).

FIGURE S19 Comparison of group 13 allergens. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was KAF7493768 from *Sarcoptes scabiei*.

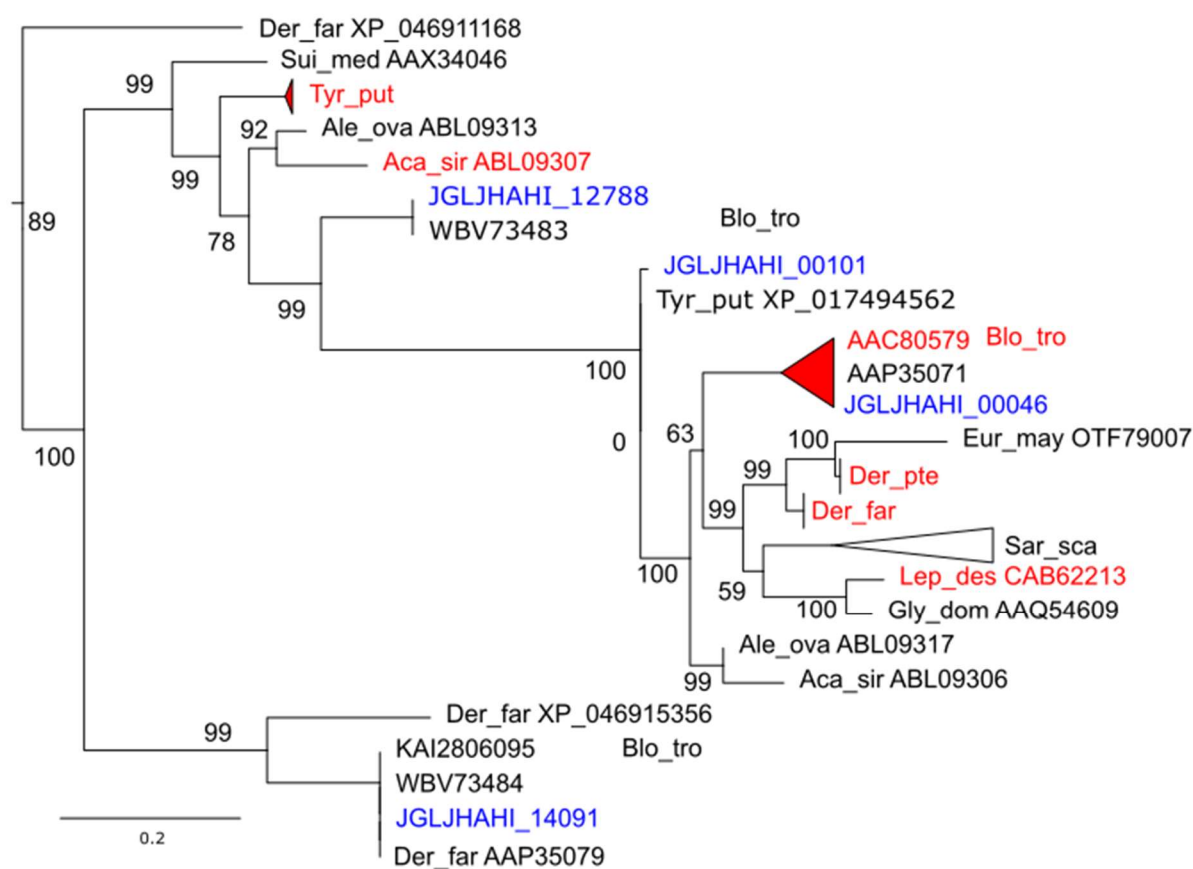


FIGURE S20 Alignment of Blo t 13.

	1	10	20	30	40	50	60
JGLJHAHI_14091	MV	VSFV	CKYQLES	SENFDE	FLKELGVN	FILRLNAKT	SKPTIEITLDGDTYTIKTTITLKT
JGLJHAHI_00046	.MP	IECKYXXXXXX	FLDELGVG	FMVKTA	AKT	LKPTLEV	VDVQGDYTVFVRSLS
JGLJHAHI_12788	MV	QLNGSYKLDK	SDNFDAFLKELGLN	FVTRNLAKT	ASPTVEVS	VDGDN	YTIKSTSTFKNT
AAC80579	.MP	IECKYKLEK	SDNFDAFLKELGVG	FMVKTA	AKT	LKPTLEV	VDVQGDYTVFVRSLS
AAP35071	.MP	IECKYKLEK	SDNFDAFLKELGVG	FMVKTA	AKT	LKPTLEV	VDVQGDYTVFVRSLS
WBV73484	MV	VSFV	CKYQLES	SENFDE	FLKELGVN	FILRLNAKT	SKPTIEITLDGDTYTIKTTITLKT
WBV73483	MV	QLNGSYKLDK	SDNFDAFLKELGLN	FVTRNLAKT	ASPTVEVS	VDGDN	YTIKSTSTFKNT
KAI2806095	MV	VSFV	CKYQLES	SENFDE	FLKELGVN	FILRLNAKT	SKPTIEITLDGDTYTIKTTITLKT

	70	80	90	100	110
JGLJHAHI_14091	VITFKIGEEFEE	SRMDGKT	VKTVITQEGD	KLIVQVQGDKEVK	IVREFTEHLLTTIC
JGLJHAHI_00046	EIKFKLGEEFEE	DRADGKRVK	TVVNKEGDNKE	IQTOYQGDKEVK	IVRDEQGGDVVVTATVG
JGLJHAHI_12788	EVKFKIGEEFEE	DRADGKRVK	TVVNKEGDNKE	IQTOYQGDKEVK	IVRDEQGGDVVVTATVG
AAC80579	EIKFKLGEEFEE	DRADGKRVK	TVVNKEGDNKE	IQTOYQGDKEVK	IVRDEQGGDVVVTATVG
AAP35071	EIKFKLGEEFEE	DRADGKRVK	TVVNKEGDNKE	IQTOYQGDKEVK	IVRDEQGGDVVVTATVG
WBV73484	VITFKIGEEFEE	SRMDGKT	VKTVITQEGD	KLIVQVQGDKEVK	IVREFTEHLLTTIC
WBV73483	EVKFKIGEEFEE	DRADGKRVK	TVVNKEGDNKE	IQTOYQGDKEVK	IVRDEQGGDVVVTATVG
KAI2806095	VITFKIGEEFEE	SRMDGKT	VKTVITQEGD	KLIVQVQGDKEVK	IVREFTEHLLTTIC

	120	130
JGLJHAHI_14091	EITSTRVYKRV	
JGLJHAHI_00046	DVTSVRTYKRI	
JGLJHAHI_12788	GVSSVREYKRL	
AAC80579	DVTSVRTYKRI	
AAP35071	DVTSVRTYKRI	
WBV73484	EITSTR...V	
WBV73483	GVSSVREYKRL	
KAI2806095	EITSTRVYKRV	

Group 14 included 3 known allergens: Der p 14 (AAM21322) was vitellogenin, i.e., SP (1-41), lipoprotein amino terminal region (59-591), vitellinogen, and open beta-sheet (638–851). The same structure was found for Eur m 1 (AAF14270), a predicted *Blomia tropicalis* protein (JGLJHAHI_08651) and the KAI2807817 protein. However, this structure was not recognized in allergen Der f 1 (BAA04558). Although the *Blomia tropicalis* proteins belonged to the same cluster (Figure S21), two proteins (AAM10780, ABU97467) did not have remarkable structures, similar to Der f 1 (BAA04558), because of an insufficient length of some proteins (Figure S22). The identity of Der p 14 (AAM21322) and Eur m 1 (AAF14270) was 88.9%. The identity of Der p 14 (AAM21322) and (JGLJHAHI_08651) was 46.4% (Table S13).

FIGURE S21 Comparison of group 14 allergens. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was XP_017492868 from *Tyrophagus putrescentiae*.

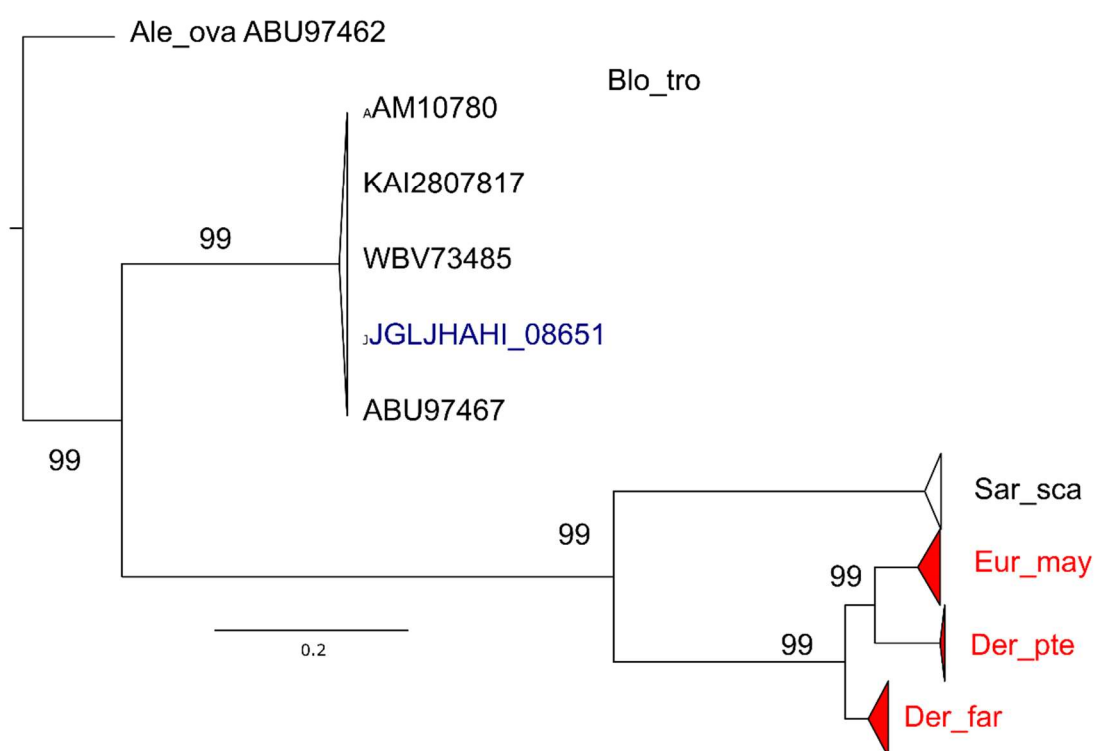


FIGURE S22 Alignment of predicted Blo t 14 proteins

```

AAM10780      . . . . .
KAI2807817    MSNTITAGNCLVRQIQMRTMRSSVKVVVFLLAALLGVSYAQKCTNICPRSLPTRIVPKV
WBV73485      MSNTITAGNCLVRQIQMRTMRSSVKVVVFLLAALLGVSYAQKCTNICPRSLPTRIVPKV
JGLJHAHI_08651 MSNTITAGNCLVRQIQMRTMRSSVKVVVFLLAALLGVSYAQKCTNICPRSLPTRIVPKV
ABU97467      . . . . .

AAM10780      . . . . .
KAI2807817    ESTYVYSLEAKTILTPRDTQKVTIKADAEVSIQSVCEASLHLKNVVIEGVPNGAELATQV
WBV73485      ESTYVYSLEAKTILTPRDTQKVTIKADAEVSIQSVCEASLHLKNVVIEGVPNGAELATQV
JGLJHAHI_08651 ESTYVYSLEAKTILTPRDTQKVTIKADAEVSIQSVCEASLHLKNVVIEGVPNGAELATQV
ABU97467      . . . . .

AAM10780      . . . . .
KAI2807817    AANPIGFYFNGKVRGVC SVKNEEEWIVNIKKAIVSALQLSGPGDEGSAEIDEHDFS GKC
WBV73485      AANPIGFYFNGKVRGVC SVKNEEEWIVNIKKAIVSALQLSGPGDEGSAEIDEHDFS GKC
JGLJHAHI_08651 AANPIGFYFNGKVRGVC SVKNEEEWIVNIKKAIVSALQLSGPGDEGSAEIDEHDFS GKC
ABU97467      . . . . .

AAM10780      . . . . .
KAI2807817    LTKFRTLETTANGAVIEKKKDLNKCTERRVDLRQTPDQLLGQVKELIRMYAHPMDSQLTC
WBV73485      LTKFRTLETTANGAVIEKKKDLNKCTERRVDLRQTPDQLLGQVKELIRMYAHPMDSQLTC
JGLJHAHI_08651 LTKFRTLETTANGAVIEKKKDLNKCTERRVDLRQTPDQLLGQVKELIRMYAHPMDSQLTC
ABU97467      . . . . .

AAM10780      . . . . .
KAI2807817    KLTVNDKVIKNVNCEETHTLIHRTEKPIHISNLQNLKENKNGVSTELADKELDNQEIFV
WBV73485      KLTVNDKVIKNVNCEETHTLIHRTEKPIHISNLQNLKENKNGVSTELADKELDNQEIFV
JGLJHAHI_08651 KLTVNDKVIKNVNCEETHTLIHRTEKPIHISNLQNLKENKNGVSTELADKELDNQEIFV
ABU97467      . . . . .

AAM10780      . . . . .
KAI2807817    SYDSQHKHAATEANVVEVLKELCPAISGSQASIEASTITAKLVGLLKHLPVDAANSVDEA
WBV73485      SYDSQHKHAATEANVVEVLKELCPAISGSQASIEASTITAKLVGLLKHLPVDAANSVDEA
JGLJHAHI_08651 SYDSQHKHAATEANVVEVLKELCPAISGSQASIEASTITAKLVGLLKHLPVDAANSVDEA
ABU97467      . . . . .

AAM10780      . . . . .
KAI2807817    VKSGNICPEAQKRLRSYFLDSCAMAASDSAIFVKAHDNGELSLIRSTAGFTVIAIKAA
WBV73485      VKSGNICPEAQKRLRSYFLDSCAMAASDSAIFVKAHDNGELSLIRSTAGFTVIAIKAA
JGLJHAHI_08651 VKSGNICPEAQKRLRSYFLDSCAMAASDSAIFVKAHDNGELSLIRSTAGFTVIAIKAA
ABU97467      . . . . .

AAM10780      . . . . .
KAI2807817    PNEETIRSLPMVESEKSTRP LLLGLSVLTRYCEKTKDCETNAAVQEARDAIMTRIGKT
WBV73485      PNEETIRSLPMVESEKSTRP LLLGLSVLTRYCEKTKDCETNAAVQEARDAIMTRIGKT
JGLJHAHI_08651 PNEETIRSLPMVESEKSTRP LLLGLSVLTRYCEKTKDCETNAAVQEARDAIMTRIGKT
ABU97467      . . . . .

AAM10780      . . . . .
KAI2807817    TDP LLKITLIKALENLNVNKNANENMIKMLDEIINSPEAGDAARLAAVKALPNEASINDQ
WBV73485      TDP LLKITLIKALENLNVNKNANENMIKMLDEIINSPEAGDAARLAAVKALPNEASINDQ
JGLJHAHI_08651 TDP LLKITLIKALENLNVNKNANENMIKMLDEIINSPEAGDAARLAAVKALPNEASINDQ
ABU97467      . . . . .

```

FIGURE S22 continuation

AAM10780	
KAI2807817	LKSIVLKESNPNEIRVAAFQKLAANKGLKDANELMNVQEHCVKNYVLT YMKNLKNNAKNPV
WBV73485	LKSIVLKESNPNEIRVAAFQKLAANKGLKDANELMNVQEHCVKNYVLT YMKNLKNNAKNPV
JGLJHAHI_08651	LKSIVLKESNPNEIRVAAFQKLAANKGLKDANELMNVQEHCVKNYVLT YMKNLKNNAKNPV
ABU97467	
AAM10780	
KAI2807817	RRSLLPENFELKEEPTKFGITKNLAYEYGPVSVDDIDIVYPEGNVTR SITARVSRVVDDK
WBV73485	RRSLLPENFELKEEPTKFGITKNLAYEYGPVSVDDIDIVYPEGNVTR SITARVSRVVDDK
JGLJHAHI_08651	RRSLLPENFELKEEPTKFGITKNLAYEYGPVSVDDIDIVYPEGNVTR SITARVSRVVDDK
ABU97467	
AAM10780	
KAI2807817	IVELGEFSVRQDGLDKQVGNLLNVVSRKISPSQFGFELIKHLSDFVAKLRKKGHTHDHLQ
WBV73485	IVELGEFSVRQDGLDKQVGNLLNVVSRKISPSQFGFELIKHLSDFVAKLRKKGHTHDHLQ
JGLJHAHI_08651	IVELGEFSVRQDGLDKQVGNLLNVVSRKISPSQFGFELIKHLSDFVAKLRKKGHTHDHLQ
ABU97467	
AAM10780	
KAI2807817	LYGKINGKNVYFSDVFEDLHHSIDRIKATVEKLI DEKKFDRTVSGLFDSRVVVPTISGI
WBV73485	LYGKINGKNVYFSDVFEDLHHSIDRIKATVEKLI DEKKFDRTVSGLFDSRVVVPTISGI
JGLJHAHI_08651	LYGKINGKNVYFSDVFEDLHHSIDRIKATVEKLI DEKKFDRTVSGLFDSRVVVPTISGI
ABU97467	
AAM10780	
KAI2807817	PLIYKLSDNALIQSNGEII SNGNKKHITWNHSEFITGLHASAKLKLKDQKLGVDYDAKLAF
WBV73485	PLIYKLSDNALIQSNGEII SNGNKKHITWNHSEFITGLHASAKLKLKDQKLGVDYDAKLAF
JGLJHAHI_08651	PLIYKLSDNALIQSNGEII SNGNKKHITWNHSEFITGLHASAKLKLKDQKLGVDYDAKLAF
ABU97467	
AAM10780	
KAI2807817	TPRFDMETERKDKGVSLMLNIRDEKHTILKFKQSLKEIKSTGPESQSENELEAEPRSDNC
WBV73485	TPRFDMETERKDKGVSLMLNIRDEKHTILKFKQSLKEIKSTGPESQSENELEAEPRSDNC
JGLJHAHI_08651	TPRFDMETERKDKGVSLMLNIRDEKHTILKFKQSLKEIKSTGPESQSENELEAEPRSDNC
ABU97467	
AAM10780	
KAI2807817	FNPLILNYCRKSSQVKGLKMHNF EYYVSKPDPALTSVKIDLDLEAPDATTKKFSAK
WBV73485	FNPLILNYCRKSSQVKGLKMHNF EYYVSKPDPALTSVKIDLDLEAPDATTKKFSAK
JGLJHAHI_08651	FNPLILNYCRKSSQVKGLKMHNF EYYVSKPDPALTSVKIDLDLEAPDATTKKFSAK
ABU97467	
AAM10780	
KAI2807817	GTANSHEHLAQLEVTPKDPAPKEVKLLKSPENSEFETKFALNLDA SRNLDIKLNLPHLID
WBV73485	GTANSHEHLAQLEVTPKDPAPKEVKLLKSPENSEFETKFALNLDA SRNLDIKLNLPHLID
JGLJHAHI_08651	GTANSHEHLAQLEVTPKDPAPKEVKLLKSPENSEFETKFALNLDA SRNLDIKLNLPHLID
ABU97467	GTANSHEHLAQLEVTPKDPAPKEVKLLKSPENSEFETKFALNLDA SRNLDIKLNLPHLID
AAM10780	
KAI2807817	IHLKNQFEKDKETNLMKNDMVLEYKFPNDETVHTLKT HHHELGYNLKRNGKDKVANFDFKS
WBV73485	IHLKNQFEKDKETNLMKNDMVLEYKFPNDETVHTLKT HHHELGYNLKRNGKDKVANFDFKS
JGLJHAHI_08651	IHLKNQFEKDKETNLMKNDMVLEYKFPNDETVHTLKT HHHELGYNLKRNGKDKVANFDFKS
ABU97467	IHLKNQFEKDKETNLMKNDMVLEYKFPNDETVHTLKT HHHELGYNLKRNGKDKVANFDFKS

FIGURE S22 continuation

AAM10780	
KAI2807817	KFESSRRPFLNHRSLVQFKYRPYKLOELVLEFGYGEKMDNVYKFSRISKIDVQEFKPFKM
WBV73485	KFESSRRPFLNHRSLVQFKYRPYKLOELVLEFGYGEKMDNVYKFSRISKIDVQEFKPFKM
JGLJHAHI_08651	KFESSRRPFLNHRSLVQFKYRPYKLOELVLEFGYGEKMDNVYKFSRISKIDVQEFKPFKM
ABU97467	KFESSRRPFLNHRSLVQFKYRPYKLOELVLEFGYGEKMDNVYKFSRISKIDVQEFKPFKM
AAM10780	
KAI2807817	NSETDLNIVATDFDVDYEIKADNRVLNDRGNALEFDMNLKGKDRSKRAAENGNGQIEIGKI
WBV73485	NSETDLNIVATDFDVDYEIKADNRVLNDRGNALEFDMNLKGKDRSKRAAENGNGQIEIGKI
JGLJHAHI_08651	NSETDLNIVATDFDVDYEIKADNRVLNDRGNALEFDMNLKGKDRSKRAAENGNGQIEIGKI
ABU97467	NSETDLNIVATDFDVDYEIKADNRVLNDRGNALEFDMNLKGKDRSKRAAENGNGQIEIGKI
AAM10780	
KAI2807817	KYRNKGSAVDSKLEASLKGVGHDFAWNSELKQVEPQKYEKGKITIQTEKDKKIFITHKEEI
WBV73485	KYRNKGSAVDSKLEASLKGVGHDFAWNSELKQVEPQKYEKGKITIQTEKDKKIFITHKEEI
JGLJHAHI_08651	KYRNKGSAVDSKLEASLKGVGHDFAWNSELKQVEPQKYEKGKITIQTEKDKKIFITHKEEI
ABU97467	KYRNKGSAVDSKLEASLKGVGHDFAWNSELKQVEPQKYEKGKITIQTEKDKKIFITHKEEI
AAM10780	
KAI2807817	SKPTKEIHEKSEADISYSYKPDKKTVMYMEAKKQGTAYILKGEAKKDGTVIFSNDINFESS
WBV73485	SKPTKEIHEKSEADISYSYKPDKKTVMYMEAKKQGTAYILKGEAKKDGTVIFSNDINFESS
JGLJHAHI_08651	SKPTKEIHEKSEADISYSYKPDKKTVMYMEAKKQGTAYILKGEAKKDGTVIFSNDINFESS
ABU97467	SKPTKEIHEKSEADISYSYKPDKKTVMYMEAKKQGTAYILKGEAKKDGTVIFSNDINFESS
AAM10780	110
KAI2807817	NGNLKALIKRDTRSVDLNVDNVFRPREATLLFKIKDREYNIKMDREPFKYITLKVDGNEN
WBV73485	NGNLKALIKRDTRSVDLNVDNVFRPREATLLFKIKDREYNIKMDREPFKYINLKVDGNEN
JGLJHAHI_08651	NGNLKALIKRDTRSVDLNVDNVFRPREATLLFKIKDREYNIKMDREPFKYINLKVDGNEN
ABU97467	NGNLKALIKRDTRSVDLNVDNVFRPREATLLFKIKDREYNIKMDREPFKYINLKVDGNEN
AAM10780	203040506070
KAI2807817	ALIKNGKAHLSIMDPTTLNLVTKANSNVDFSMDLFASINKKVALKIDSPKYNFFHDGDI
WBV73485	ALIKNGKAHLSIMDPTTLNLVTKANSNVDFSMDLFASINKKIALKIDSPKYNFFHDGDI
JGLJHAHI_08651	ALIKNGKAHLSIMDPTTLNLVTKANSNVDFSMDLFASINKKIALKIDSPKYNFFHDGDI
ABU97467	ALIKNGKAHLSIMDPTTLNLVTKANSNVDFSMDLFASIRHQIALKIDSPKYNFFHDGDI
AAM10780	8090100110120130
KAI2807817	LSIVNRRLLWKSITKKDDREYKFNADIARKGSMISLTKVTPDRTSSVQYSRNGEKIEVNI
WBV73485	LSIVNRRLLWKSITKKDDREYKFNADIARKGSMISLTKVTPDRTSSVQYSRNGEKIEVNI
JGLJHAHI_08651	LSIVNRRLLWKSITKKDDREYKFNADIARKGSMISLTKVTPDRTSSVQYSRNGEKIEVNI
ABU97467	LSIVNRRLLWKSITKKDDREYKFNADIARKGSMISLTKVTPDRTSSVQYSRNGEKIEVNI
AAM10780	140150160170180190
KAI2807817	DTEYLEGKVEGDRFSGKIVLKNKQNDYELESTYKRENGRLVIESVNGKNKMEAVFSRKE
WBV73485	DTEYLEGKVEGDRFSGKIVLKNKQNDYELESTYKRENGRLVIESVNGKNKMEAVFSRKE
JGLJHAHI_08651	DTEYLEGKVEGDRFSGKIVLKNKQNDYELESTYKRENGRLVIESVNGKNKMEAVFSRKE
ABU97467	DTEYLEGKVEGDRFSGKIVLKNKQNDYELESTYKRENGRLVIESVNGKNKMEAVFSRKE
AAM10780	200210220230240250
KAI2807817	PSKPVLETPNTKAKIDMDLTAPVKTFKLDFDNPRYQKKIDASMEEPKSKPKYSSYSNQKNE
WBV73485	PSKPVLETPNTKAKIDMDLTAPVKTFKLDFDNPRYQKKIDASMEEPKSKPKYSSYSNQKNE
JGLJHAHI_08651	PSKPVLETPNTKAKIDMDLTAPVKTFKLDFDNPRYQKKIDASMEEPKSKPKYSSYSNQKNE
ABU97467	PSKPVLETPNTKAKIDMDLTAPVKTFKLDFDNPRYQKKIDASMEEPKSKPKYSSYSNQKNE

FIGURE S22 continuation

	260	270	280	290	300	310
AAM10780	KKERKIEIDGVHMKEL	NVDIDFPDFKFKVKQPESSKKVEFSYTFNNYTETEEYDFDPHKA				
KAI2807817	KKERKIEIDGVHMKEL	NVDIDFPDFKFKVKQPESSKKVEFSYTFNNYTETEEYDFDPHKA				
WBV73485	KKERKIEIDGVHMKEL	NVDIDFPDFKFKVKQPESSKKVEFSYTFNNYTETEEYDFDPHKA				
JGLJHAHI_08651	KKERKIEIDGVHMKEL	NVDIDFPDFKFKVKQPESSKKVEFSYTFNNYTETEEYDFDPHKA				
ABU97467	KKERKIEIDGVHMKEL	NVDIDFPDFKFKVKQPESSKKVEFSYTFNNYTETEEYDFDPHKA				

	320	330
AAM10780	YLVNWFVNALRQYVQTFVVQN	
KAI2807817	YLVNWFVNALRQYVQTFVVQN	
WBV73485	YLVNWFVNALRQYVQTFVVQN	
JGLJHAHI_08651	YLVNWFVNALRQYVQTFVVQN	
ABU97467	YLVNWFVNALRQYVQTFVVQN	

Group 15 showed known allergens in *Dermatophagoides* mites. The structural domains were Glycosyl hydrolase family 18 and Chitin binding Peritrophin-A domain, i.e., Der f 15 (AAD52672) signal peptide 1-19, Glycosyl hydrolase family 18, 34-381 and Chitin binding peritrophin-A domain 507-554. The identities of Der f 15 (AAD52672) and Der p 15 (AAY84564, AAY84565) were 90 and 88%, respectively. Predicted *Blomia* protein JGLJHAHI_17563 showed the same structural function, but the chitin-binding peritrophin-A domain was missing. The *Blomia tropicalis* proteins presented lower similarity to *Dermatophagoides* mites by approximately 62% (Table S14). The predicted JGLJHAHI_17563 protein showed 65% sequence identity to Der f 15 (AAD52672) (Figure S23). All predicted Blo t 15 proteins were almost identical (Figure S22).

FIGURE S23 Comparison of group 15 allergens. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was UXW65971 from *Tyrophagus putrescentiae*.

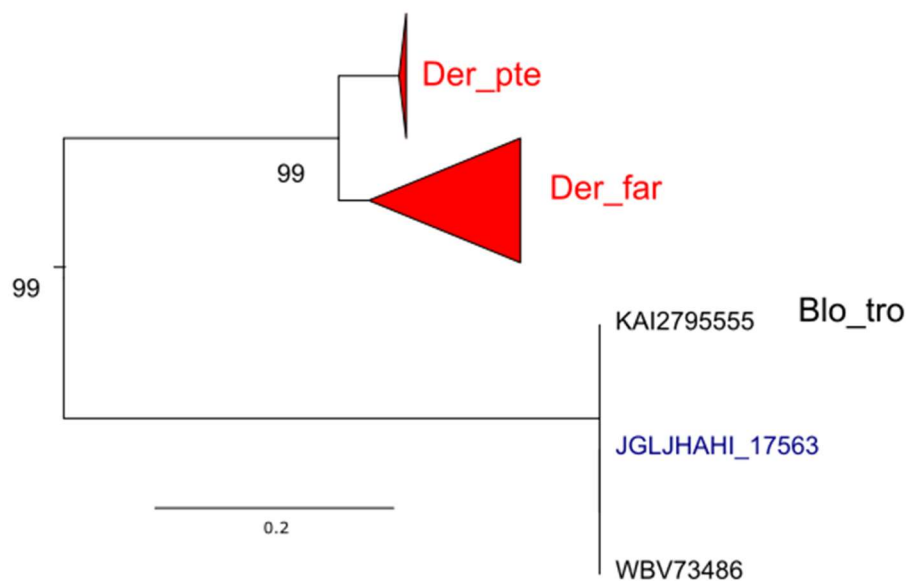


FIGURE S24 Alignment of predicted Blo t 15 proteins.

[illegible]

Der f 16 (AAM64112) is a known allergen from **group 16**, and its protein structure contains gelsolin repeat units, i.e., Gelsolin repeats 25-121, 160-231, 296-367 and 396-473. No signal peptide was found. The same structure showed the predicted *Blomia tropicalis* protein JGLJHAHI_10134 with 57% sequence identity to Der f 16 (AAM64112) (Table S15). The proteins from *B. tropicalis* and *Tyrophagus putrescentiae* formed a separate cluster from those from *Dermatophagoides* mites (Figure S25). The alignment of predicted Blo t 16 proteins showed that JGLJHAHI_10134 differed by 2/484 amino acids from the other two *Blomia tropicalis* proteins (Figure S26).

FIGURE S25 Comparison of group 16 allergens. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was KAI7694635 from *Sarcoptes scabiei*.

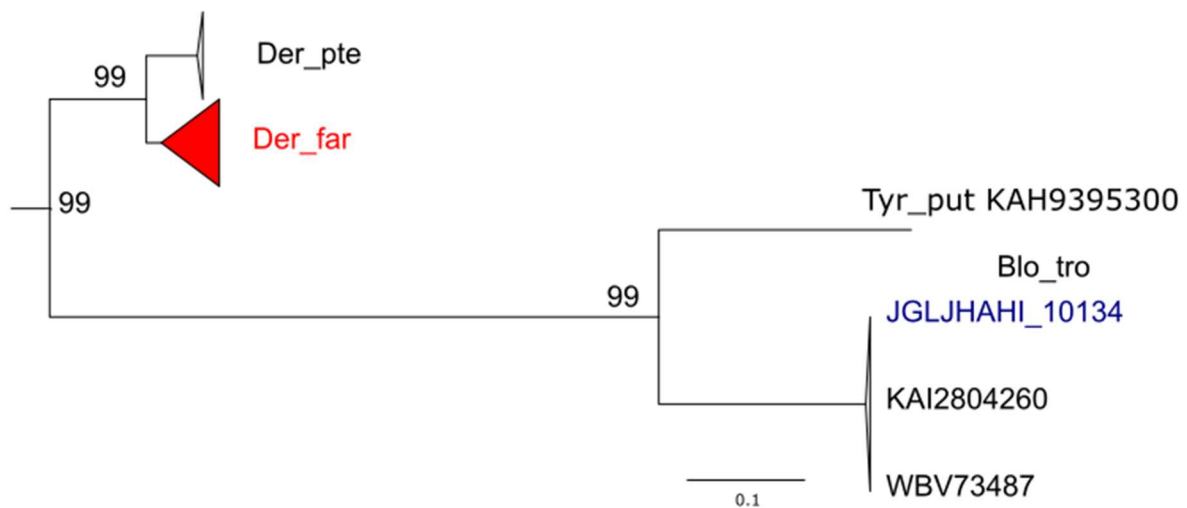
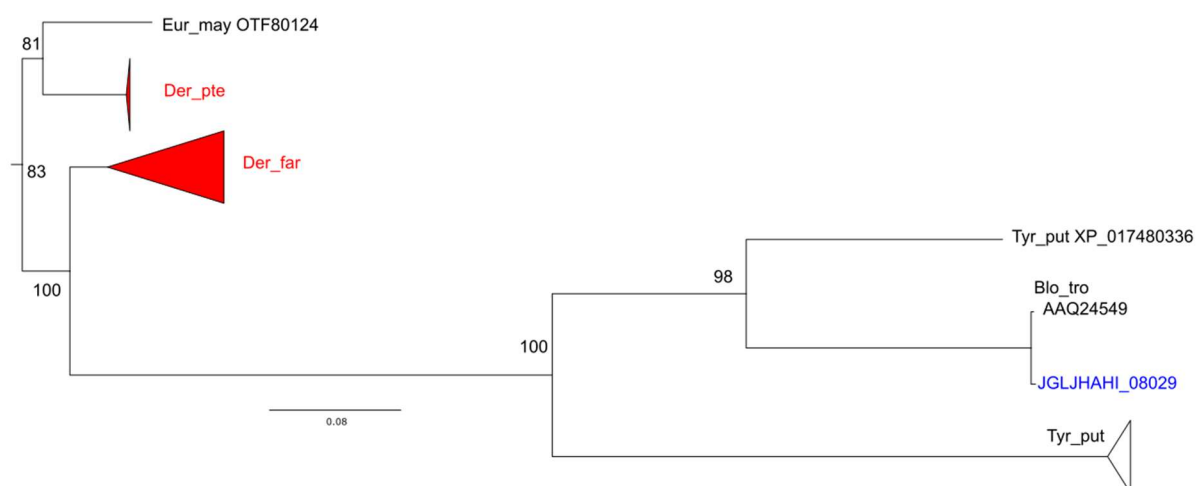


FIGURE S26 Predicted Blo t 16 protein alignment.

	1	10	20	30	40	50	60
JGLJHAHI_10134	MASNAPHD	VFDVVPIGSSFFFIWVRKEFNLI	PVPKEEY	GKFKGDSYVVICCTEYRTGG			
KAI2804260	MASNAPHD	VFDVVPIGSSFFFIWVRKEFNLI	PVPKEEY	GKFKGDSYVVICCTEYRTGG			
WBV73487	MASNAPHD	VFDVVPIGSSFFFIWVRKEFNLI	PVPKEEY	GKFKGDSYVVICCTEYRTGG			
	70	80	90	100	110	120	
JGLJHAHI_10134	HSNMPIEPLKTGGIGMIDIHFWIGSNSSQDES	AVAAIKTVELDDFLGGTPVQHREVEGHE					
KAI2804260	HSNMPIEPLKTGGIGMIDIHFWIGSNSSQDES	AVAAIKTVELDDFLGGTPVQHREVEGHE					
WBV73487	HSNMPIEPLKTGGIGMIDIHFWIGSNSSQDES	AVAAIKTVELDDFLGGTPVQHREVEGHE					
	130	140	150	160	170	180	
JGLJHAHI_10134	SKQFLSYFRNGIRLLNGCFESGFHKVDKELHPS	LYQVKGKKRPVMHELNEIKWSSINNND					
KAI2804260	SKQFLSYFRNGIRLLNGCFESGFHKVDKELHPS	LYQVKGKKRPVMHELNEIKWSSINNND					
WBV73487	SKQFLSYFRNGIRLLNGCFESGFHKVDKELHPS	LYQVKGKKRPVMHELNEIKWSSINNND					
	190	200	210	220	230	240	
JGLJHAHI_10134	VLLLVVPKYIFVWVGKSA	NHFERLSAINIANELKNELSRFHLS	IVIVDDGQEAQQLSSNE				
KAI2804260	VLLLVVPKYIFVWVGKSA	NHFERLSAINIANELKNELSRFHLS	IVIVDDGQEAQQLSSNE				
WBV73487	VLLLVVPKYIFVWVGKSA	NHFERLSAINIANELKNELSRFHLS	IVIVDDGQEAQQLSSNE				
	250	260	270	280	290	300	
JGLJHAHI_10134	NTEFNRYLPLESKDSQIKPA	DEFDWTSEKFEKAERKFVHLYHCKEFDDKILIKSVKDG	P				
KAI2804260	NTEFNRYLPLESKDSQIKPA	DEFDWTSEKFEKAERKFVHLYHCKEFDDKILIKSVKDG	P				
WBV73487	NTEFNRYLPLESKDSQIKPA	DEFDWTSEKFEKAERKFVHLYHCKEFDDKILIKSVKDG	P				
	310	320	330	340	350	360	
JGLJHAHI_10134	LTRSDLDSDNSFIVDNGANGI	WVWVGKDASTKERSSGLRYAMELIKDEYPAKTEVTKVI					
KAI2804260	LTRSDLDSDNSFIVDNGANGI	WVWVGKDASTKERSSGLRYAMELIKDEYPAKTEVTKVI					
WBV73487	LTRSDLDSDNSFIVDNGANGI	WVWVGKDASTKERSSGLRYAMELIKDEYPAKTEVTKVI					
	370	380	390	400	410	420	
JGLJHAHI_10134	EGGETIEFRSLFKQWS	AVKGFSA	GKQARLFNLT	KKCKFAQI	INFEKHDLEEDNVMILDV		
KAI2804260	EGGETIEFRSLFKQWS	AVKGFSA	GKQARLFNLT	KKCKFAQI	INFEKHDLEEDNVMILDV		
WBV73487	EGGETIEFRSLFKQWS	AVKGFSA	GKQARLFNLT	KKCKFAQI	INFEKHDLEEDNVMILDV		
	430	440	450	460	470	480	
JGLJHAHI_10134	GEKLFVWLCSQLQEKL	ADESKVGKIAEFY	LKKDKSGRQLEPAKII	YVKQGS	SENDEFKSYF		
KAI2804260	GEKLFVWLCSQLQEKL	ADESKVGKIAEFY	LKKDKSGRQLEPAKII	YVKQGS	SENDEFKSYF		
WBV73487	GEKLFVWLCSQLQEKL	ADESKVGKIAEFY	LKKDKSGRQLEPAKII	YVKQGS	SENDEFKSYF		
JGLJHAHI_10134	KSWN						
KAI2804260	KSWN						
WBV73487	KSWN						

Group 18 contained allergens of *Dermatophagoides* mites, including Der f 18 (AAM19082) and Der p 18 (AAY84563), which belonged to glycosyl hydrolase family 18, i.e., Der p 18 (AAY84563) signal peptide 1-16, Glycosyl hydrolase family 18 30-356 and Chitin binding peritrophin-A domain 407-462. Der p 18 and Der F 18 differed in that Der p 18 contained both glycosyl hydrolase family 18 and the chitin binding peritrophin-A domain, which was not recognized in Der f 18. The sequence identity of those proteins was 89% (Table S16). Predicted *Blomia tropicalis* protein JGLJHAHI_08029 contained glycosyl hydrolase family 18 but did not bind the peritrophin-A domain, which caused the formation of different clusters among these proteins (Figure S27). The identities were 61.1% for Der f 18 (AAM19082) and 60.7% for Der p 18 (AAY84563), and the identity between *B. tropicalis* proteins AAQ24549 and JGLJHAHI_08029 was 100%.

FIGURE S27 Comparison of group 18 allergens. The description is provided in Fig S1. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was OTF77271 from *Euroglyphus maynei*.



Group 20 contained identified allergens with structural homology to ATP:guanido phosphotransferase, i.e., Der f 20 (AIO08850) ATP:guanido phosphotransferase, N-terminal domain 21-87 and ATP:guanido phosphotransferase, C-terminal catalytic domain 147-356. No signal peptide was found. The *Dermatophagoides farinae* allergens Der f 20 (AIO08850 and ABU97470) showed 84.0% sequence identity (Table S17). *Tyrophagus putrescentiae* Tyr p 20 (QOI58528 and UXW65972) peptides shared 71% sequence identity. Tyr p 20 formed two separate clusters. Two predicted *Blomia tropicalis* proteins, JGLJHAHI_13643 and JGLJHAHI_04702, clustered into Tyr p 20 clusters (Figure S28). The expression level of JGLJHAHI_04702 was low, and no match was found in the proteome. JGLJHAHI_13643 showed 88.5% sequence identity to Tyr p 20 (QOI58528). The alignment of predicted proteins from Blo t 20 showed that KAI2800065 and WBV73490 were identical proteins, while JGLJHAHI_13643 showed differences (Figure S29).

FIGURE S28 Comparison of group 20 allergens. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was KAF7493571 from *Sarcoptes scabiei*.

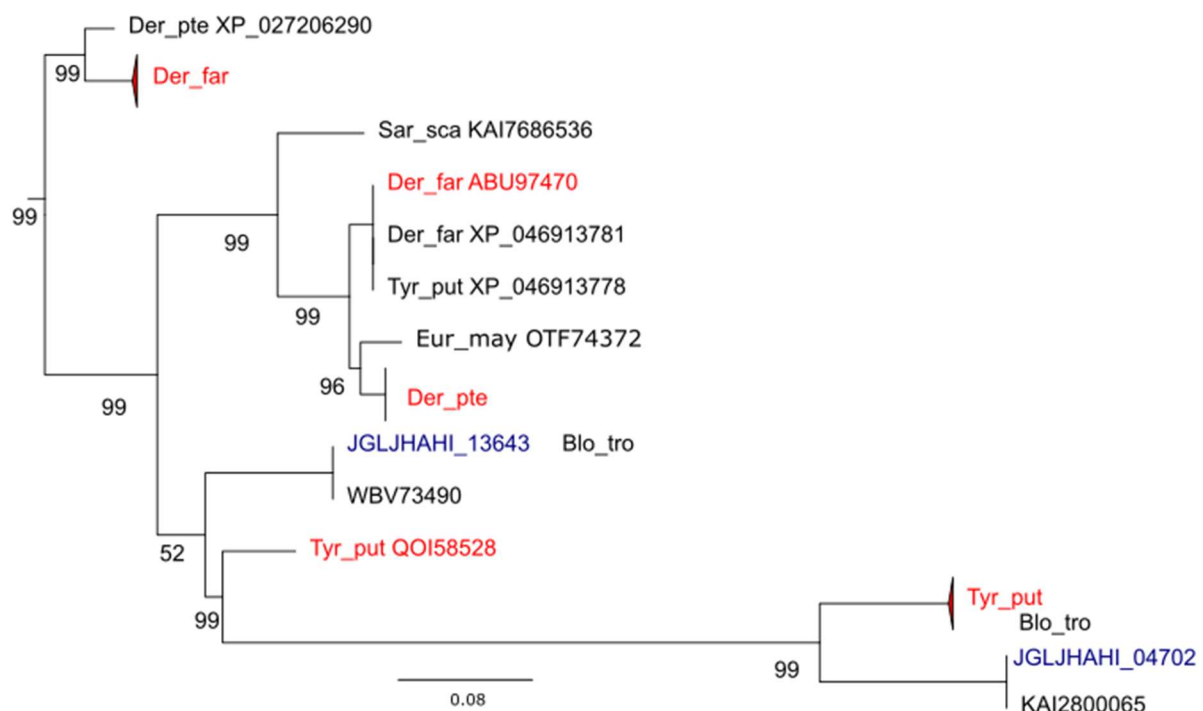
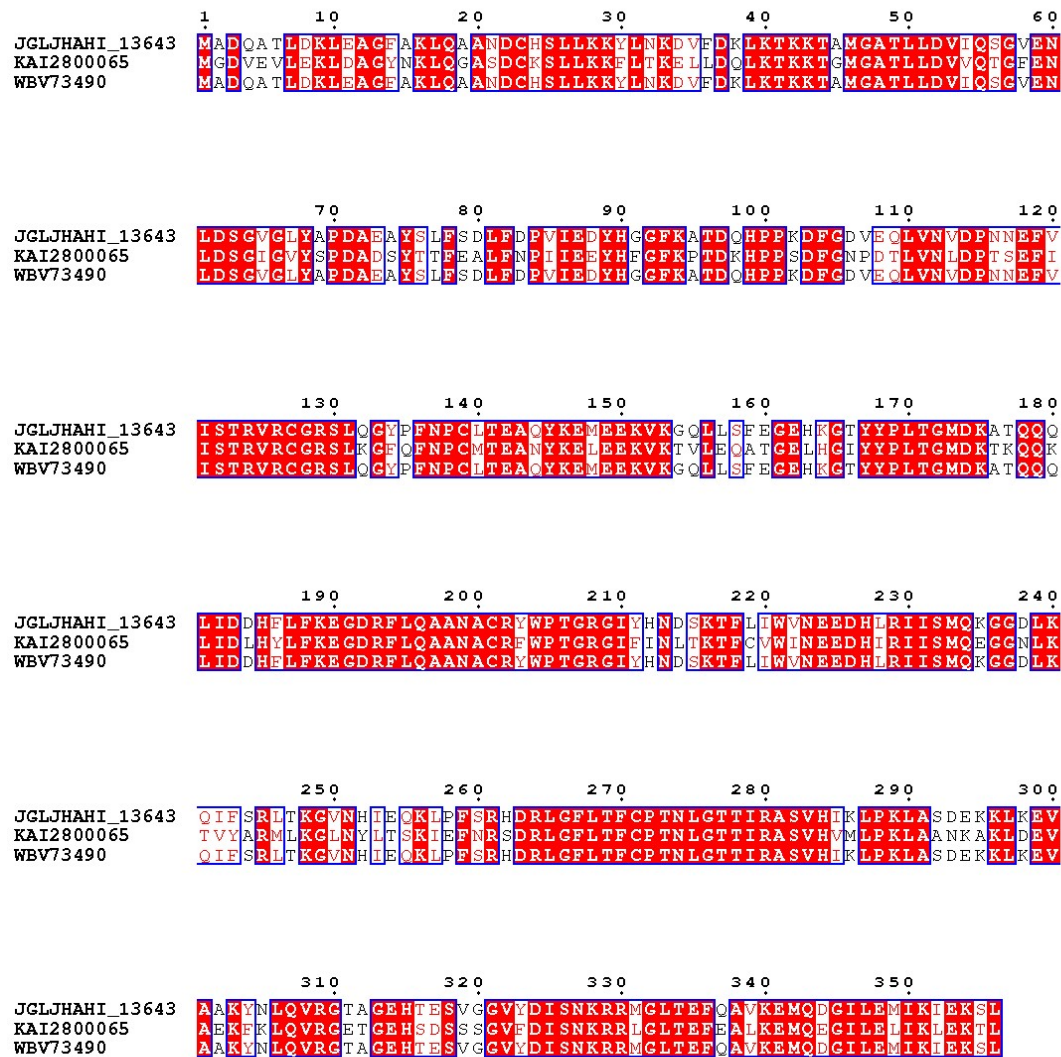


FIGURE S29 Alignment of predicted Blo t 20.



Group 21 allergens showed structural homology to the mite allergen Blo t 5, i.e., Der f 21 (AHC94806) 1-17 signal peptide and mite allergen Blo t 5 17-135. The identity of Blo t 21 (AAX34047) to Der f 21 (AHC94806) was 39%, and that to Der p 21 (ABC73706) was 39.4% (Table S18). Predicted JGLJHAHI_14491 showed 100% identity to Blo t 21 (AAX34047). The identity of Blo t 5 (AAD10850) and Blo t 21 (AAX34047) was 41.5%. The alignment of Blo t 21 showed 4/130 different amino acids (Figure S30).

FIGURE S30 Alignment of Blo t 21.

	1	10	20	30	40	50	60
JGLJHAHI_14491	MKFI	IALAALI	IAVACALPVSNDNFRHEFDHMI	VNT	ATQRFHEIEKFL	LLHITHEVDDLEK	T
ABH06346	MKFI	IALAALI	IAVACALPVSNDNFRHEFDHMI	VNT	ATQRFHEIEKFL	LLHITHEVDDLEK	T
ABH06347	MKFI	IALAALI	IAVACALPVSNDNFRHEFDHMI	VNT	ATQRFHEIEKFL	LLHITHEVDDLEK	T
ABH06348	MKFI	IALAALI	IAVACALPVSNDNFRHEFDHMI	VNT	ATQRFHEIEKFL	LLHITHEVDDLEK	T
ABH06350	MKFI	IALAALI	IAVACALPVSNDNFRHEFDHMI	VNT	ATQRFHEIEKFL	LLHITHEVDDLEK	T
AAX34047	MKFI	IALAALI	IAVACALPVSNDNFRHEFDHMI	VNT	ATQRFHEIEKFL	LLHITHEVDDLEK	T
WBV73492	MKFI	IALAALI	IAVACALPVSNDNFRHEFDHMI	VNT	ATQRFHEIEKFL	LLHITHEVDDLEK	T

	70	80	90	100	110	120
JGLJHAHI_14491	GNKDEKARLL	RELTVSEAFIEGSRGYFQRELKRTDLDLLEKFNFEAALATGDLLLKDLKA				
ABH06346	GNKDEKARLL	RELTVSEAFIEGSRGYFQRELKRTDLDLLEKFNFEAALATGDLLLKDLKA				
ABH06347	GNKDEKARLL	RELTVSEAFIEGSRGYFQRELKRTDLDLLEKFNFEAALATGDLLLKDLKA				
ABH06348	GNKDEKARLL	RELTVSEAFIEGSRGYFQRELKRTDLDLLEKFNFEAALATGDLLLKDLKA				
ABH06350	GNKDEKARLL	RELTVSEAFIEGSRGYFQRELKRTDLDLLEKFNFEAALATGDLLLKDLKA				
AAX34047	GNKDEKARLL	RELTVSEAFIEGSRGYFQRELKRTDLDLLEKFNFEAALATGDLLLKDLKA				
WBV73492	GNKDEKARLL	RELTVSEAFIEGSRGYFQRELKRTDLDLLEKFNFEAALATGDLLLKDLKA				

JGLJHAHI_14491	LQKRVQDSE
ABH06346	LQKRVQDSE
ABH06347	LQKRVQDSE
ABH06348	LQKRVQDSE
ABH06350	LQKRVQDSE
AAX34047	LQKRVQDSE
WBV73492	LQKRVQDSE

Group 24 allergens from *Dermatophagoides* mites (Der f 24 (AGI78542) and Der p 24 (ALA65345)) showed structural homology to the ubiquinol-cytochrome C reductase complex 14 kD subunit, i.e., Der f 24 (AGI78542), Ubiquinol-cytochrome C reductase complex 14 kD subunit 7-109, and no signal peptide. The sequence identity was 96.6%. Predicted *Blomia tropicalis* protein JGLJHAHI_12583 showed the same structure as *Dermatophagoides* allergens and presented identities of 72% and 75% to these allergens (Table S19). The sequences of *Blomia tropicalis* formed a separate cluster from the *Dermatophagoides* allergens (Figure S31). The alignment of predicted Blo t 24 showed 100% identity to *Blomia tropicalis* proteins (WBV73494 and KAI2804014), and JGLJHAHI_12583 differed by 1/119 amino acids (Figure S32).

FIGURE S31 Comparison of group 24 allergens. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was KAI7692480 from *Sarcoptes scabiei*

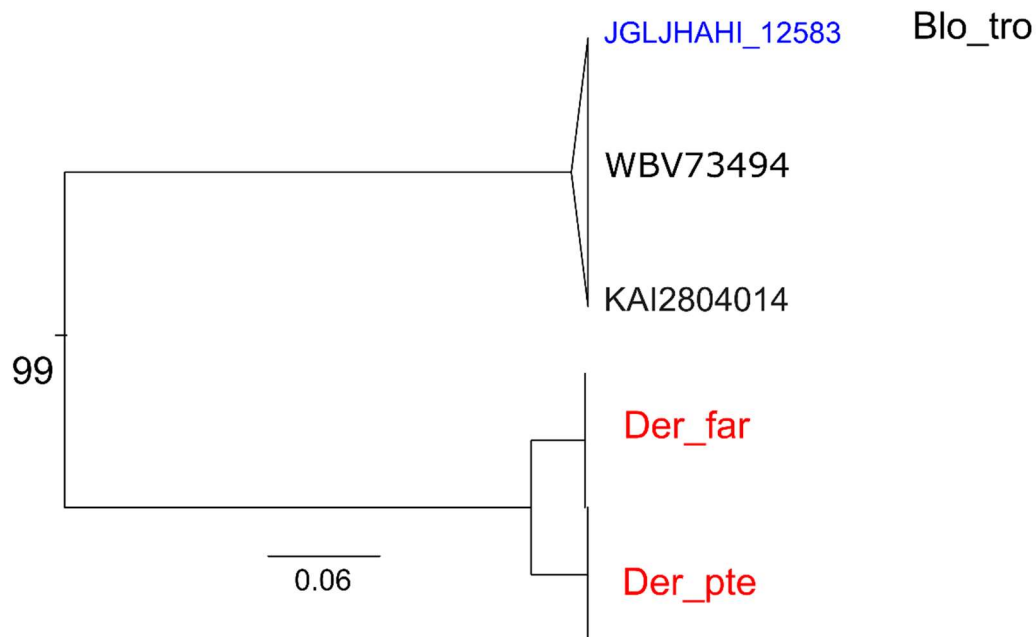


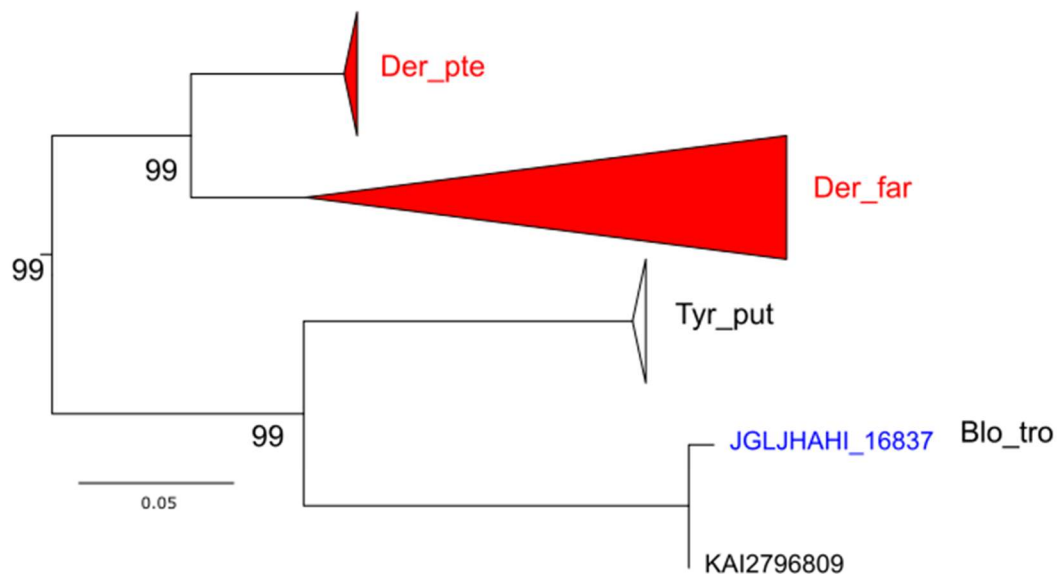
FIGURE S32 Alignment of Blo t 24.

	1	10	20	30	40	50	60
JGLJHAHI_12583	MVFLTRTLRVFSQPAFKKMYFGMQGYNKFGLYRDDLYDYTDPVHLEALRRLPPDVYDQHC						
WBV73494	MVFLTRTLRVFNQPAFKKMYFGMQGYNKFGLYRDDLYDYTDPVHLEALRRLPPDVYDQHC						
KAI2804014	MVFLTRTLRVFNQPAFKKMYFGMQGYNKFGLYRDDLYDYTDPVHLEALRRLPPDVYDQHC						

	70	80	90	100	110
JGLJHAHI_12583	FRVVRASQLEFTKQFLPQEEWPSYEEDQTKGRFLQPYIEEVQKEKLEKEEWVSFLSKD				
WBV73494	FRVVRASQLEFTKQFLPQEEWPSYEEDQTKGRFLQPYIEEVQKEKLEKEEWVSFLSKD				
KAI2804014	FRVVRASQLEFTKQFLPQEEWPSYEEDQTKGRFLQPYIEEVQKEKLEKEEWVSFLSKD				

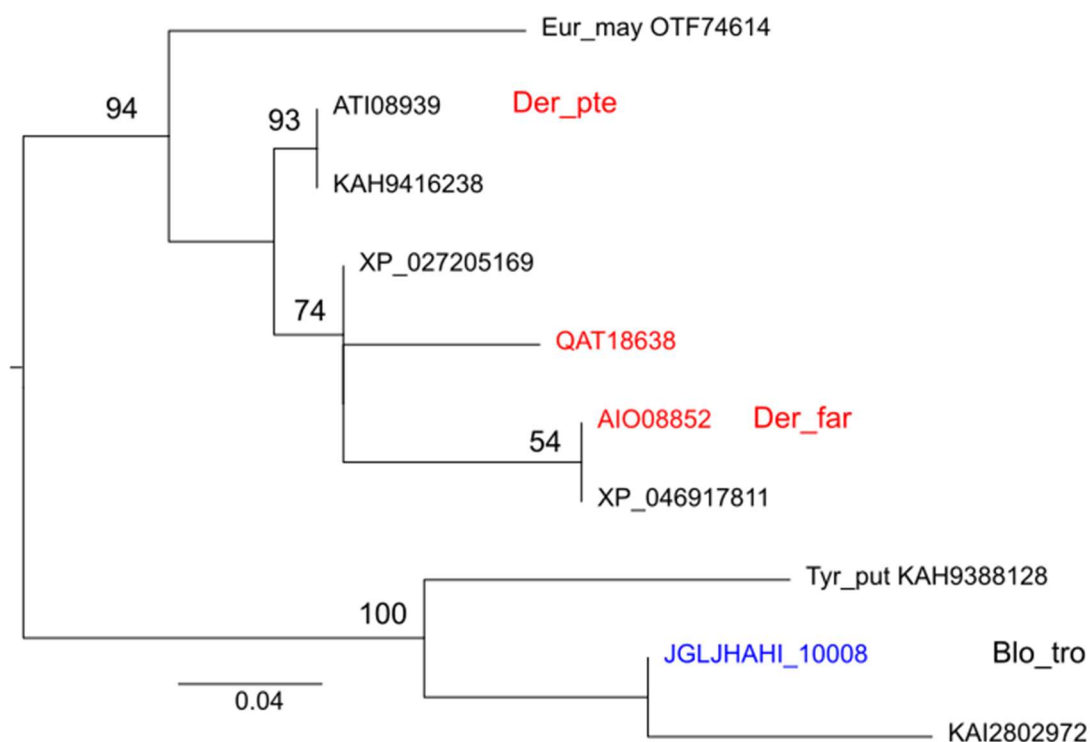
The structure of allergens from **group 25** corresponded to triosephosphate isomerase, i.e., Der f 25 (AGC56216) Triosephosphate isomerase 5-246, with no signal protein. Der f 25 allergens (AGC56216 and AIO08860) showed 87.9% sequence identity. The identities to Der p 25 (QAT186370) were 85 and 91.9%. Predicted *Blomia tropicalis* protein JGLJHAHI_16837 showed a different structural function, i.e., AIG2-like family. The similarity of JGLJHAHI_16837 to *Dermatophagoides* allergens was 77.6 to 81.20% (Table S20). *Blomia tropicalis* proteins formed a separate cluster outside of *Dermatophagoides* allergens together with *Tyrophagus putrescentiae* proteins (Figure S33). The most similar protein was KAI2796809, which shared structural homology with both the AIG2-like family (62-147) and triosephosphate isomerase (177-418). The identity of *Blomia tropicalis* proteins JGLJHAHI_10008 and KAI2796809 was 99.6%. The analogous situation was that the *Tyrophagus putrescentiae* protein KAH9409787 showed structural homology to the AIG2-like family (77-165) and triosephosphate isomerase (199-440), but the AIG2-like family domain was missing in XP_017478126. This suggested that the absence of the AIG2-like family domain or triosephosphate isomerase was an artifact due to incomplete sequences.

FIGURE S33 Comparison of group 25 allergens. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was KAF7491894 from *Sarcoptes scabiei*.



Dermatophagoides allergens **Der p 26** (QAT18638) and Der f 26 (AIO08852) did not show any known structure. Their identity was 93.9%. Predicted *Blomia tropicalis* protein JGLJHAHI_10008 shared 82.7 and 85.2% identity with these allergens (Table S21). The *B. tropicalis* proteins (JGLJHAHI_10008 and KAI2802972) shared 100% identity, and together with *Tyrophagus putrescentiae* proteins, they formed a sister cluster to *Dermatophagoides* (Figure S34).

FIGURE S34 Comparison of group 26 allergens. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was KAF7490318 from *Sarcoptes scabiei*.



Group 28 was formed by *Dermatophagoides* allergens Der p 28 (QAT18639) and Der f 28 (AGC56218, AIO08848). The allergen structure corresponded to the Hsp70 protein, i.e., Der f 28 (AGC56218) Hsp70 protein, 6-597 and 605-631 signal protein. Der f 28 allergens showed 69.8% identity; however, AGC56218 was similar to an *Aspergillus fumigatus* (AAS58470) protein (identity= 78.7%). The identity between QAT18639 and AIO08848 was much higher, at 95.6%. This was the reason for excluding AGC56218 from the comparison. Two predicted *Blomia tropicalis* proteins, Hsp70 protein JGLJHAHI_17137 and JGLJHAHI_14735, shared 81% identity with each other and 78 and 79.4% identity with AIO08848 (Table S22). These sequences formed two separate clusters outside of house dust mite allergens (Figure S35). The alignment of predicted Blo t 28 allergens is shown in Figure S36.

FIGURE S35 Comparison of group 28 allergens. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was AOD75395 from *Tyrophagus putrescentiae*.

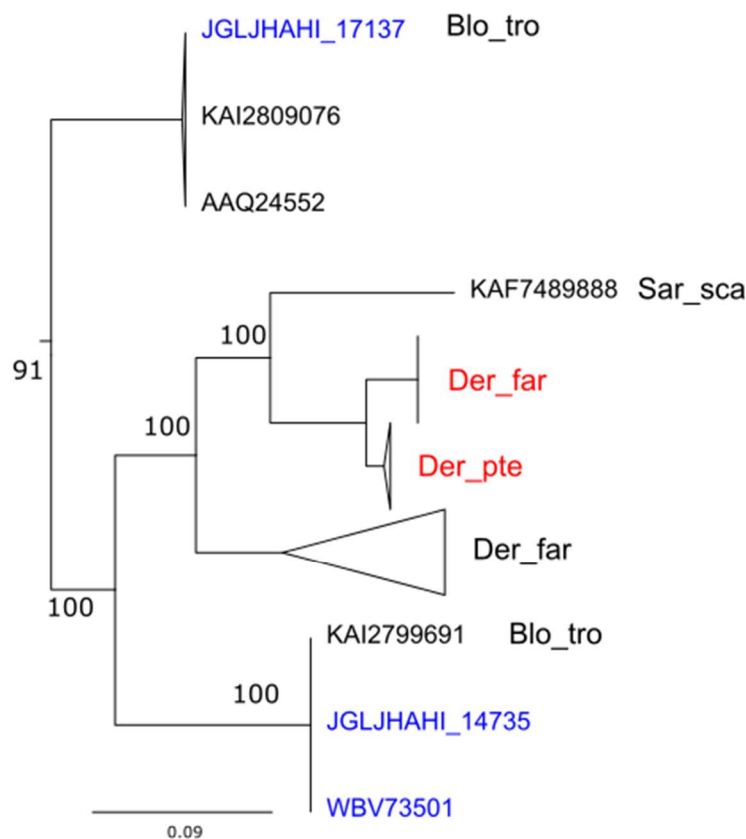


FIGURE S36 Alignment of predicted Blo t 28.

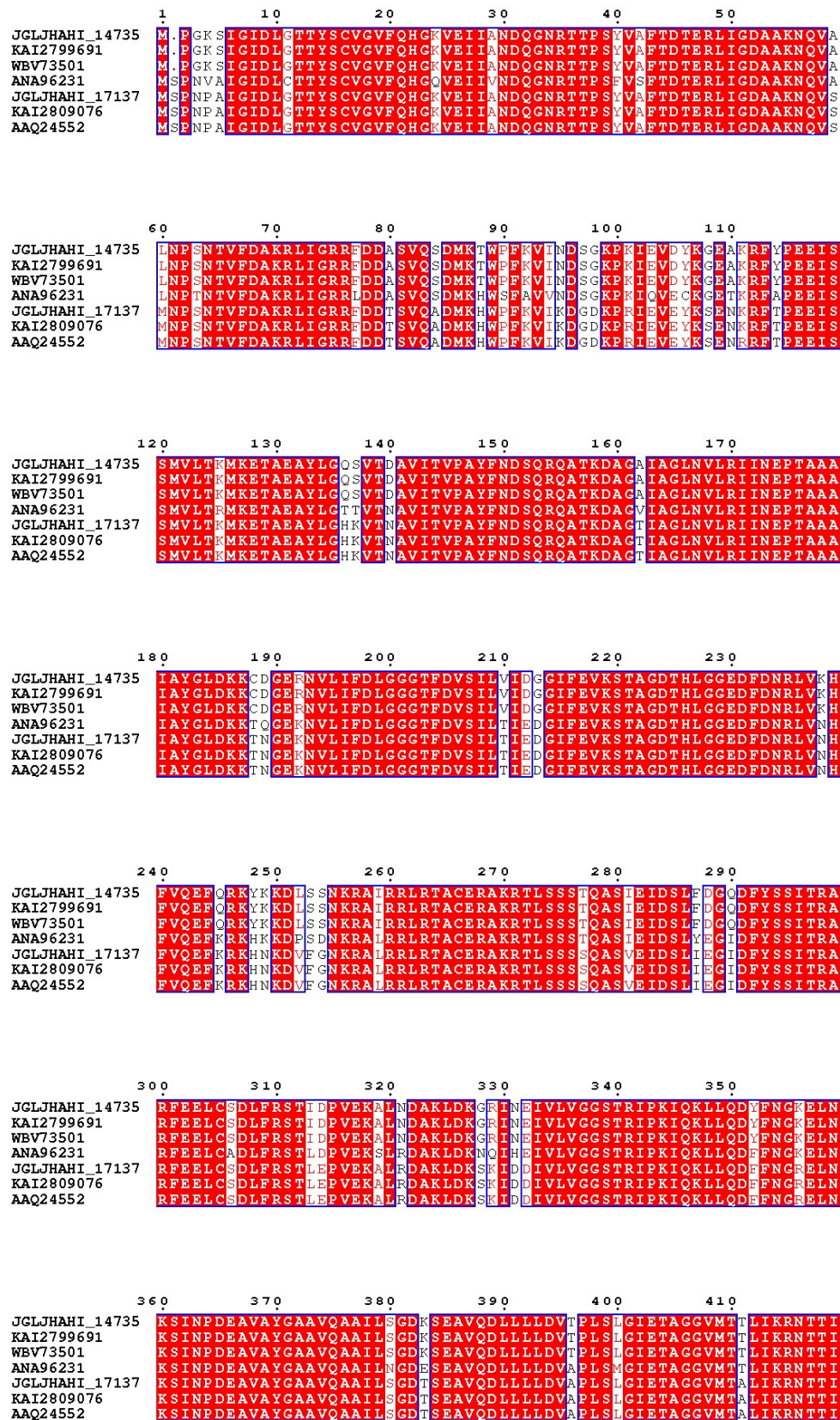


FIGURE S36 continuation

	420	430	440	450	460	470
JGLJHAHI_14735	PTKQTQI	FTTYSNQP	CVTVQ	VYEGERT	MTKDNLLG	QFTLS
KAI2799691	PTKQTQI	FTTYSNQP	CVTVQ	VYEGERT	MTKDNLLG	QFTLS
WBV73501	PTKQTQI	FTTYSNQP	CVTVQ	VYEGERT	MTKDNLLG	QFTLS
ANA96231	PTKQTQI	FTTYSNQP	CVTVQ	VYEGERT	MTKDNLLG	QFTLS
JGLJHAHI_17137	PTKQTQT	FTTYSNQP	CVNLIQ	VYEGERT	MTKDNLLG	QFTLS
KAI2809076	PTKQTQT	FTTYSNQP	CVNLIQ	VYEGERT	MTKDNLLG	QFTLS
AAQ24552	PTKQTQT	FTTYSNQP	CVNLIQ	VYEGERT	MTKDNLLG	QFTLS

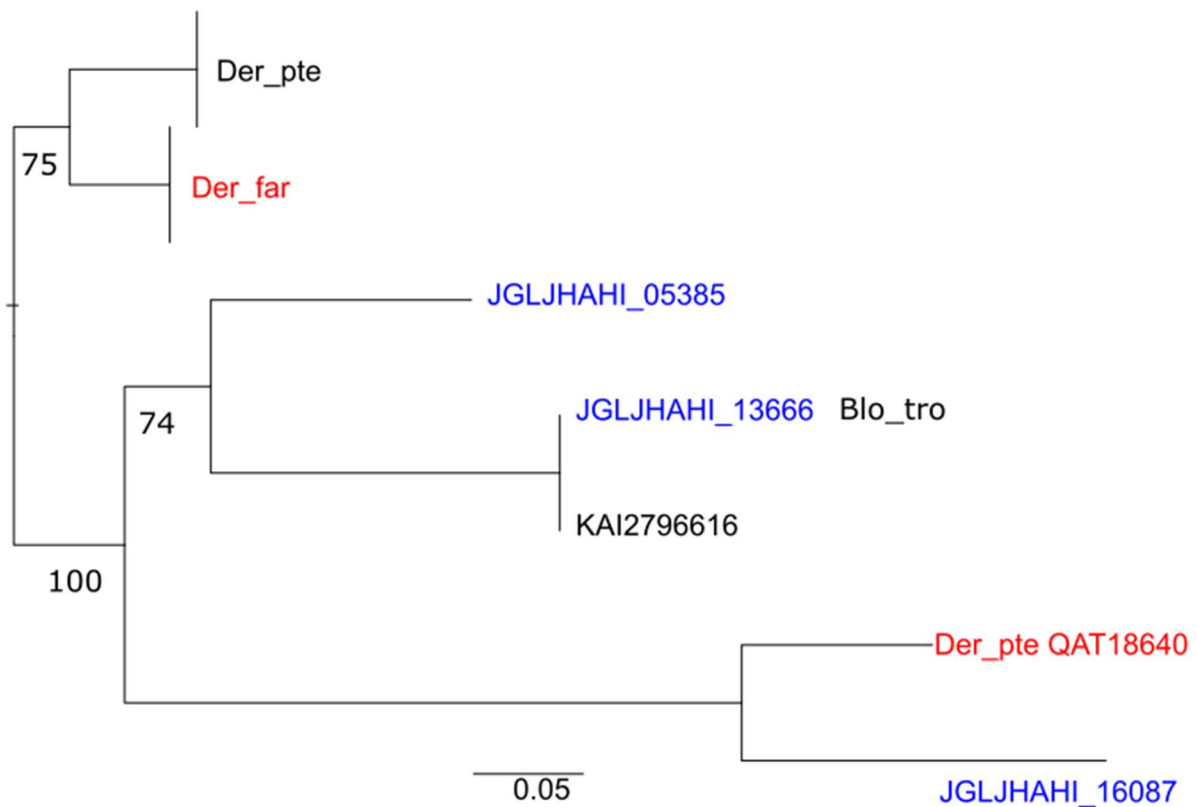
	480	490	500	510	520	530
JGLJHAHI_14735	DANGILH	VSAQDKST	GKQNKITI	TNDKGRL	SKADIEK	MVRDAE
KAI2799691	DANGILH	VSAQDKST	GKQNKITI	TNDKGRL	SKADIEK	MVRDAE
WBV73501	DANGILH	VSAQDKST	GKQNKITI	TNDKGRL	SKADIEK	MVRDAE
ANA96231	DANGILH	VSAQDKST	GKQNKITI	TNDKGRL	SKADIEK	MVRDAE
JGLJHAHI_17137	DANGILNV	SAVDKST	GKQNKITI	TNDKGRL	TKEDIER	MVKEAE
KAI2809076	DANGILNV	SAVDKST	GKQNKITI	TNDKGRL	TKEDIER	MVKEAE
AAQ24552	DANGILNV	SAVDKST	GKQNKITI	TNDKGRL	TKEDIER	MVKEAE

	540	550	560	570	580	590
JGLJHAHI_14735	QLESYAYS	IKNTIEDE	ALKSKISDS	DRKNILE	KINATIA	WLDNNQ
KAI2799691	QLESYAYS	IKNTIEDE	ALKSKISDS	DRKNILE	KINATIA	WLDNNQ
WBV73501	QLESYAYS	IKNTIEDE	ALKSKISDS	DRKNILE	KINATIA	WLDNNQ
ANA96231	QLESYAYS	IKNTIEDE	ALKSKISDS	DRKNILE	KINATIA	WLDNNQ
JGLJHAHI_17137	ALESYAF	TIKSTLDD	ETVKT	KISEDR	DRKIND	KIADTMK
KAI2809076	ALESYAF	TIKSTLDD	ETVKT	KISEDR	DRKIND	KIADTMK
AAQ24552	ALESYAF	TIKSTLDD	ETVKT	KISEDR	DRKIND	KIADTMK

	600	610	620	630
JGLJHAHI_14735	TF	CRPIMTKLY	SGGGRAN	PSAG
KAI2799691	TF	CRPIMTKLY	SGGGRAN	PSAG
WBV73501	TF	CRPIMTKLY	SGGGRAN	PSAG
ANA96231	GICNP	ITTKLYQQA	CGAEMPGG	MPGMPGFP
JGLJHAHI_17137	AVCNP	IMS	SKLHQAA	CGQOMP
KAI2809076	AVCNP	IMS	SKLHQAA	CGQOMP
AAQ24552	AVCNP	IMS	SKLHQAA	CGQOMP

Group 29 included *Dermatophagoides* mite allergens Der f 29 (AAP35065) and Der p 29 (QAT18640). The allergens showed structural homology to cyclophilin-type peptidyl-prolyl cis-trans isomerase/CLD, i.e., Der f 29 (AAP35065) cyclophilin-type peptidyl-prolyl cis-trans isomerase/CLD, 7-163, no signal protein. Their sequence identity was 76.2% (Table S23). The predicted Blo t 29 allergens showed different positions in the clusters, i.e., Der f 29 (AAP35065) formed a separate cluster with *Dermatophagoides pteronyssinus* proteins. Der p 29 formed a separate cluster with JGLJHAHI_16087. The last cluster was formed by the remaining *Blomia tropicalis* proteins (JGLJHAHI_05385, JGLJHAHI_13666 and KAI2796616) (Figure S37). Although four predicted *Blomia tropicalis* proteins showed similarity and structural homology to these allergens, JGLJHAHI_13666 was confirmed only by proteomic analyses. The sequence identities were 83.5% to Der f 29 (AAP35065), 60.8% to Der p 29 (QAT18640) and 100% to a *Blomia tropicalis* protein (KAI2796616).

FIGURE S37 Comparison of group 29 allergens. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was ASL05640 from *Psoroptes ovis*.



Group 30 included *Dermatophagoides* mite allergens Der f 30 (AGC56219) and Der p 30 (QAT18641), with 82.5% sequence identity and structural homology to the ferritin-like domain, i.e., Der f 30 (AGC56219) Ferritin-like domain 14–154, with no signal peptide. Three peptides of *Blomia tropicalis* with the same structural homology were predicted, but only JGLJHAHI_17697 was confirmed by proteomic analyses. The sequence identity was 74.7% for Der f 30 (AGC56219) and 85.4% for Der p 30 (QAT18641) (Table S24). The *B. tropicalis* proteins clustered together with *Tyrophagus putrescentiae* proteins and formed a separate cluster from *Dermatophagoides* allergens (Figure S38). The predicted Blo t 30 alignment showed differences in 10/170 amino acids (Figure S39).

FIGURE S38 Comparison of group 30 allergens. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was KAI2803900 from *Blomia tropicalis*.

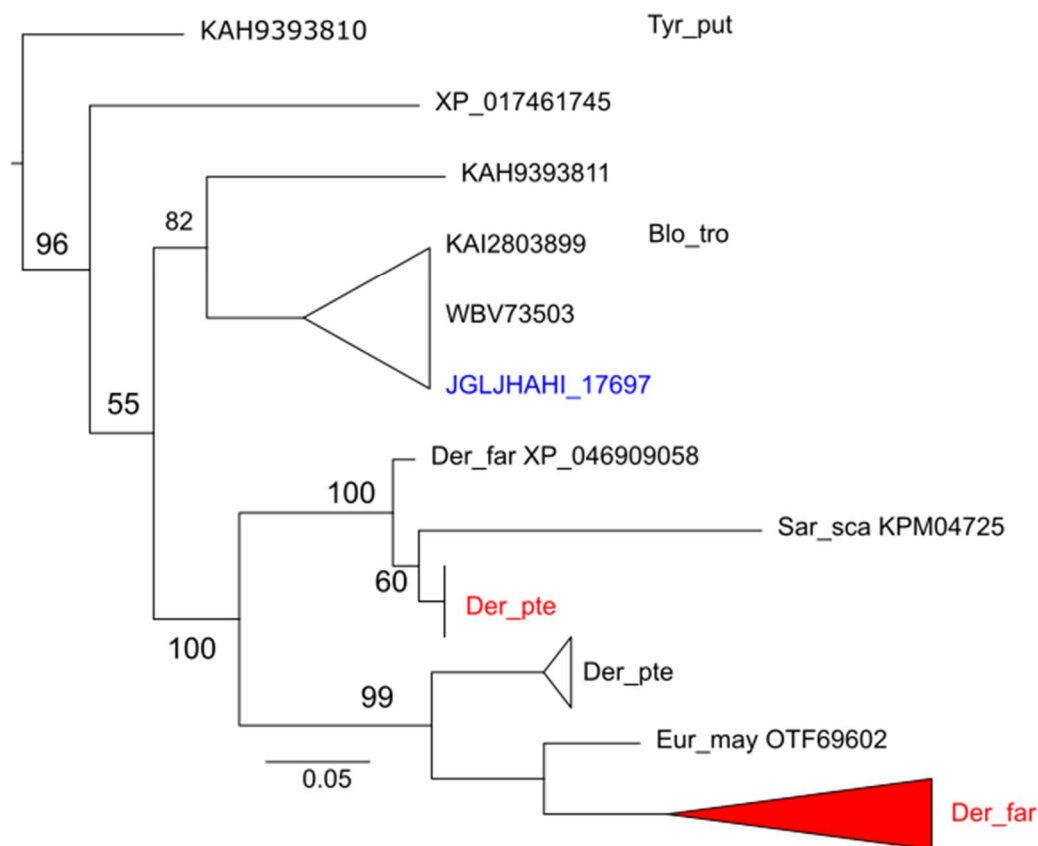


FIGURE S39 Predicted Blo t 30 alignment.

	1	10	20	30	40	50
KAI2803899	MADVSKVRMNFHKECEAGVNKQINLELYASLSTLI...MHLPGFEKFFRENSKEEREH					
WBV73503	MADVSKVRMNFHKECEAGVNKQINLELYASYVYQQLNRDDVALPGFEKFFRENSKEEREH					
J_17697	MADVSKVRMNFHKECEAGVNKQINLELYASYVYQQLNRDDVALPGFEKFFRENSKEEREH					

	60	70	80	90	100	110
KAI2803899	AEKLMKFQNERGGRIILHDVPKPIKQDWSSGLEAMEAAELEKTVNQSLDLHGVA					
WBV73503	AEKLMKFQNERGGRIILHDVPKPIKQDWSSGLEAMEAAELEKTVNQSLDLHGVA					
J_17697	AEKLMKFQNERGGRIILHDVPKPIKQDWSSGLEAMEAAELEKTVNQSLDLHGVA					

	120	130	140	150	160
KAI2803899	DPQFADFLETHFLTEQVEAIKLLADYVTNLRRVGSGLGEYLFDKHTLSS				
WBV73503	DPQFADFLETHFLTEQVEAIKLLADYVTNLRRVGSGLGEYLFDKHTLSS				
J_17697	DPQFADFLETHFLTEQVEAIKLLADYVTNLRRVGSGLGEYLFDKHTLSS				

Group 32 included known allergens with structural homology to inorganic pyrophosphatase, i.e., Der f 32 (AIO08849) inorganic pyrophosphatase,51-233., with no signal peptide. It included *Dermatophagoides* mite Der p 32 (QAT18643) and Der f 32 (AIO08849), with 90.5% identity to each other. *Tyrophagus putrescentiae* Tyr p 32 (UXW65973) presented 68.5% and 70.0% identity to *Dermatophagoides* allergens (Table S25). Predicted *Blomia tropicalis* protein JGLJHAHI_11580 showed sequence identities of 59.5 and 67.7% to *Dermatophagoides* allergens and 83.6% to Tyr p 32 (UXW65973). The *B. tropicalis* proteins formed a sister cluster to *T. putrescentiae* proteins outside the *Dermatophagoides* allergen cluster (Figure S40). Blo t 32 showed high sequence identity (Figure S41).

FIGURE S40 Comparison of group 32 allergens. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was KAF7488884 from *Sarcoptes scabiei*.

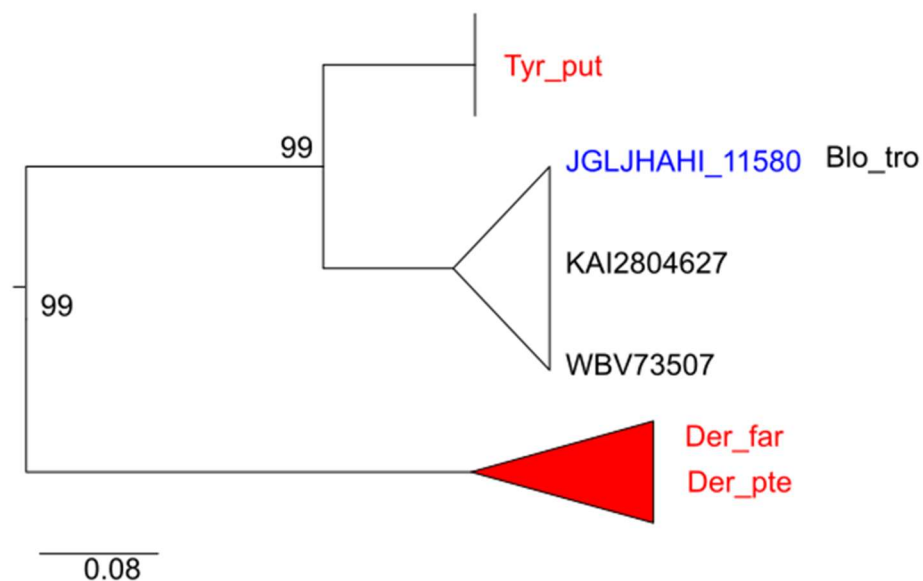


FIGURE S41 Alignment of predicted Blo 32 proteins.

	1	10	20	30	40	50	60
JGLJHAHI_11580	SIFVIKNRQLLLKHLATQTCTNHVHFVTNHSRLQLNTYCKPRKITFTYNHCPTRQLGS						
KAI2804627							
WBV73507							

	70	80	90	100	110	120
JGLJHAHI_11580	TCTNL MIGPSLTLSHSSVLCTSLIVSHRNLTKMSSNVSYTVDERGASNSLDYRVYFKDAT					
KAI2804627	 MIGPSLTLSHSSVLCTSLIVSHRNLTKMSSNVSYTVDERGASNSLDYRVYFKDAT					
WBV73507	 MIGPSLTLSHSSVLCTSLIVSHRNLTKMSSNVSYTVDERGASNSLDYRVYFKDAT					

	130	140	150	160	170	180
JGLJHAHI_11580	GQPISPWHDIPLFSDQSKKIYNMVVEIPRWSNAKMEIATAEPLNPIKQDVKKGALRYVRN					
KAI2804627	GQPISPWHDIPLFSDQSKKIYNMVVEIPRWSNAKMEIATAEPLNPIKQDVKKGALRYVRN					
WBV73507	GQPISPWHDIPLFSDQSKKIYNMVVEIPRWSNAKMEIATAEPLNPIKQDVKKGALRYVRN					

	190	200	210	220	230	240
JGLJHAHI_11580	FFPHKGYIWNYGAFPQTWEDPTHVDPSTHAKGDNDPIDVLEIGQRVANRGEVLQVKILGT					
KAI2804627	FFPHKGYIWNYGAFPQTWEDPTHVDPSTHAKGDNDPIDVLEIGQRVANRGEVLQVKILGT					
WBV73507	FFPHKGYIWNYGAFPQTWEDPTHVDPSTHAKGDNDPIDVLEIGQRVANRGEVLQVKILGT					

	250	260	270	280	290	300
JGLJHAHI_11580	VALIDEGETDWKIIITINVDDPLAEQLNDIDDVEKFMPGFLRATIEWFKIYKIPDGKPA NE					
KAI2804627	VALIDEGETDWKIIITINVDDPLAEQLNDIDDVEKFMPGFLRATIEWFKIYKIPDGKPA NE					
WBV73507	VALIDEGETDWKIIITINVDDPLAEQLNDIDDVEKFMPGFLRATIEWFKIYKIPDGKPA NE					

	310	320	330	340	350	360
JGLJHAHI_11580	FAFNGEAKNREFAEKVVEETANFWKALVGKSPNESKLDLKN TKLSNDYSVKVEDSKSIVE					
KAI2804627	FAFNGEAKNREFAEKVVEETANFWKALVGKSPNESKLDLKN TKLSNDYSVKVEDSKSIVE					
WBV73507	FAFNGEAKNREFAEKVVEETANFWKALVGKSPNESKLDLKN TKLSNDYSVKVEDSKSIVE					

	370	380
JGLJHAHI_11580	AKAVTSSLDKVP IADQSAIDKWYYIKLN	
KAI2804627	AKAVTSSLDKVP IADQSAIDKWYYIKLN	
WBV73507	AKAVTSSLDKVP IADQSAIDKWYYIKLN	

Only **Der f 33** (AIO08861) was the only identified allergen in group 33. The protein was composed of a Tubulin/FtsZ family domain, GTPase domain (alignment region 3-220) and Tubulin C-terminal domain (alignment region 270-399). Three proteins of *Blomia tropicalis* (JGLJHAHI_03926, JGLJHAHI_05719, JGLJHAHI_06710) showed the same structural identity, but the expression of JGLJHAHI_06710 was low and was not confirmed by proteomic analyses. The sequence identity of JGLJHAHI_03926 and JGLJHAHI_05719 was 93.3%, and the sequences showed 81 and 82% identity to Der f 33 (AIO08861), respectively (Table S26). The *B. tropicalis* protein sequences formed a separate cluster with the allergens of *D. farinae* together with the proteins of *Dermatophagoides pteronyssinus*, *Lepidoglyphus destructor*, *Sarcoptes scabiei* and *Tyrophagus putrescentiae* (Figure S42). However, all *Blomia tropicalis* proteins showed a high level of identity in the alignment (Figure S43).

FIGURE S42 Comparison of group 33 allergens. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was XP_017481387 from *Tyrophagus putrescentiae*.

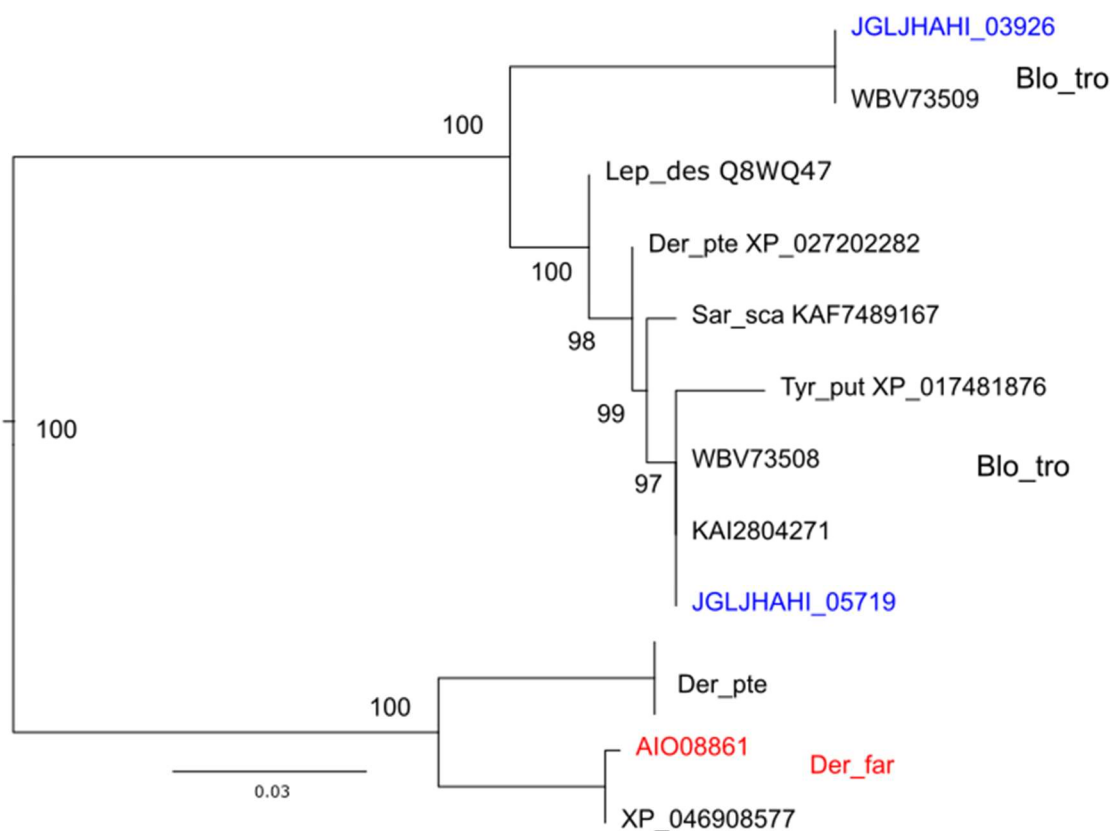


FIGURE S43 Alignment of predicted Blo t 33 allergens

	1	10	20	30	40	50	60
WBV73508	MRECI	SVHV	GQAGVQ	IGNACWELYCLEHGIQPDGQMP	SDKT	CT	GDDSFNTFFSETGSGK
KAI2804271	MRECI	SVHV	GQAGVQ	IGNACWELYCLEHGIQPDGQMP	SDKT	CT	GDDSFNTFFSETGSGK
J_03926	MRECI	SVHV	GQAGVQ	IGNACWELYCLEHGIQPDGQMP	SDKT	CT	GDDSFNTFFSETGSGK
J_05719	MRECI	SVHV	GQAGVQ	IGNACWELYCLEHGIQPDGQMP	SDKT	CT	GDDSFNTFFSETGSGK
	70	80	90	100	110	120	
WBV73508	HVPRAVY	VDLEP	TVVDEV	RTGTYRQLFHPEQLITGKEDAANNYARGHYTIGKEIVD	VVLD		
KAI2804271	HVPRAVY	VDLEP	TVVDEV	RTGTYRQLFHPEQLITGKEDAANNYARGHYTIGKEIVD	VVLD		
J_03926	HVPRAVY	VDLEP	TVVDEV	RTGTYRQLFHPEQLITGKEDAANNYARGHYTIGKEIVD	VVLD		
J_05719	HVPRAVY	VDLEP	TVVDEV	RTGTYRQLFHPEQLITGKEDAANNYARGHYTIGKEIVD	VVLD		
	130	140	150	160	170	180	
WBV73508	RIRKLS	DQCT	GLQGFLIFHSFGGCTGSGFTSL	LMERLSVDYGKKS	KLEFAV	YPAPQVSTA	
KAI2804271	RIRKLS	DQCT	GLQGFLIFHSFGGCTGSGFTSL	LMERLSVDYGKKS	KLEFAV	YPAPQVSTA	
J_03926	RIRKLS	DQCT	GLQGFLIFHSFGGCTGSGFTSL	LMERLSVDYGKKS	KLEFAV	YPAPQVSTA	
J_05719	RIRKLS	DQCT	GLQGFLIFHSFGGCTGSGFTSL	LMERLSVDYGKKS	KLEFAV	YPAPQVSTA	
	190	200	210	220	230	240	
WBV73508	VVEPYNS	ILTHTT	LEHSDCAF	MVDNEAIYDICRRNLDIERPTYTNLNRLIGQIVSSITA			
KAI2804271	VVEPYNS	ILTHTT	LEHSDCAF	MVDNEAIYDICRRNLDIERPTYTNLNRLIGQIVSSITA			
J_03926	VVEPYNS	ILTHTT	LEHSDCAF	MVDNEAIYDICRRNLDIERPTYTNLNRLIGQIVSSITA			
J_05719	VVEPYNS	ILTHTT	LEHSDCAF	MVDNEAIYDICRRNLDIERPTYTNLNRLIGQIVSSITA			
	250	260	270	280	290	300	
WBV73508	SLRFDGALNVDL	TEFQTNLVPYPRIHFP	LV	TYAPVISAEKAYHEQLTVAEITNT	CFEPAN		
KAI2804271	SLRFDGALNVDL	TEFQTNLVPYPRIHFP	LV	TYAPVISAEKAYHEQLTVAEITNT	CFEPAN		
J_03926	SLRFDGALNVDL	TEFQTNLVPYPRIHFP	LV	TYAPVISAEKAYHEQLTVAEITNT	CFEPAN		
J_05719	SLRFDGALNVDL	TEFQTNLVPYPRIHFP	LV	TYAPVISAEKAYHEQLTVAEITNT	CFEPAN		
	310	320	330	340	350	360	
WBV73508	QMVKCDPRHGKYMA	C	CLLYRGDVVPKDVNA	AAIA	GIKTKRSIQFVDWCPTGFKVG	INYP	
KAI2804271	QMVKCDPRHGKYMA	C	CLLYRGDVVPKDVNA	AAIA	GIKTKRSIQFVDWCPTGFKVG	INYP	
J_03926	QMVKCDPRHGKYMA	C	CLLYRGDVVPKDVNA	AAIA	GIKTKRSIQFVDWCPTGFKVG	INYP	
J_05719	QMVKCDPRHGKYMA	C	CLLYRGDVVPKDVNA	AAIA	GIKTKRSIQFVDWCPTGFKVG	INYP	
	370	380	390	400	410	420	
WBV73508	TVVPGGDLAKVQRAVC	MLSNTTAIAEAWARLDH	KFDLMYAKRAFVHWYV	VGCMEEGEFSE			
KAI2804271	TVVPGGDLAKVQRAVC	MLSNTTAIAEAWARLDH	KFDLMYAKRAFVHWYV	VGCMEEGEFSE			
J_03926	TVVPGGDLAKVQRAVC	MLSNTTAIAEAWARLDH	KFDLMYAKRAFVHWYV	VGCMEEGEFSE			
J_05719	TVVPGGDLAKVQRAVC	MLSNTTAIAEAWARLDH	KFDLMYAKRAFVHWYV	VGCMEEGEFSE			
	430	440	450				
WBV73508	AREDLAALEKDYE	EVGLD	STEADD	TAGEEF			
KAI2804271	AREDLAALEKDYE	EVGLD	STEADD	TAGEEF			
J_03926	AREDLAALEKDYE	EVGLD	STEADD	TAGEEF			
J_05719	AREDLAALEKDYE	EVGLD	STEADD	TAGEEF			

Group 34 was formed by two allergens, Der f 34 (BAV90601) and Tyr p 34 (ACL36923). These two proteins presented low (20.9%) sequence identity. A phmmer search of Der f 34 (BAV90601) revealed endoribonuclease L-PSP (alignment region 12-126) with similarity only to bacterial proteins. Der f 34 (BAV90601) was eliminated from the search in this group. Tyr p 34 (ACL36923) contained two EF-hand domain pairs (alignment regions 14–74 and 88–151). *Blomia tropicalis* protein JGLJHAHI_15642 showed 100% identity to Tyr p 34 (ACL36923) (Table S27), and proteins with the same structure were found in *Dermatophagoides* mites, *S. scabei*, and *Psoroptes ovis* (Figure S44). Tyr p 34 (ACL36923), the predicted JGLJHAHI_15642 protein and the *B. tropicalis* protein WBV73517 differed from the remaining compared proteins in the second EF-hand domain. However, all aligned proteins from *B. tropicalis* showed variability at the beginning of the protein, and the second part of the protein was homogenous (Figure S45).

FIGURE S44 The clustering of Tyr p 34 (ACL36923) and the selected protein of mites and predicted *Blomia tropicalis* JGLJHAHI_15642. Red indicates the identified allergen proteins, and blue indicates the predicted proteins of *B. tropicalis*. The outgroup sequence was KAI7697693 from *Sarcoptes scabiei*.

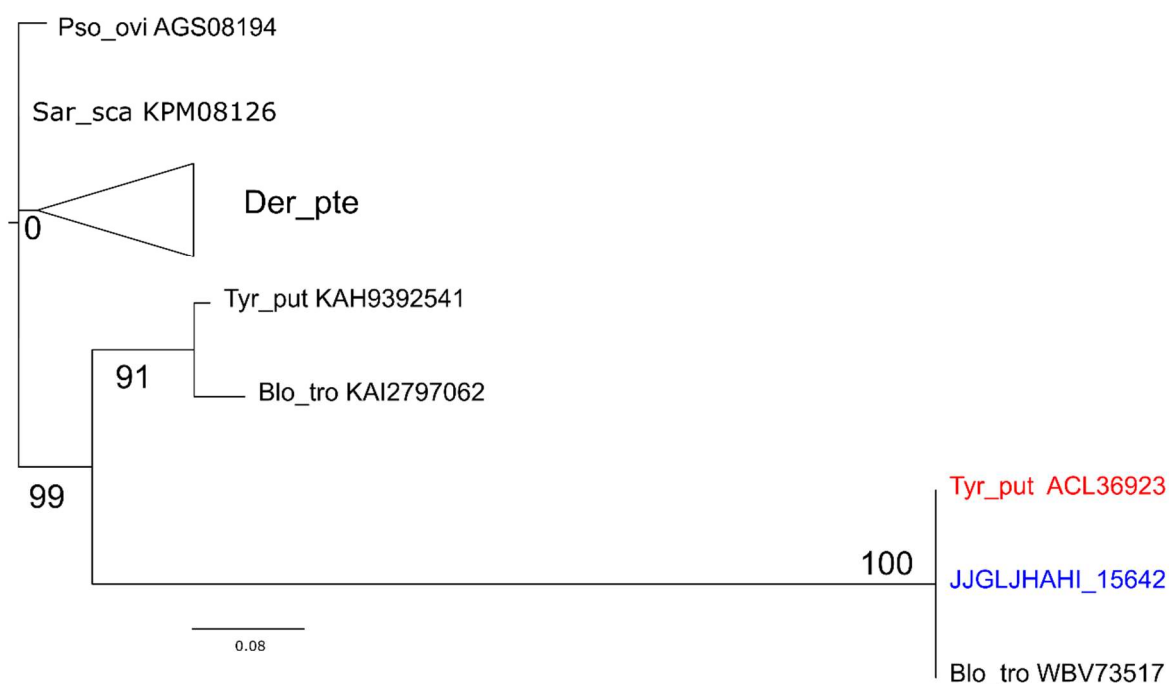
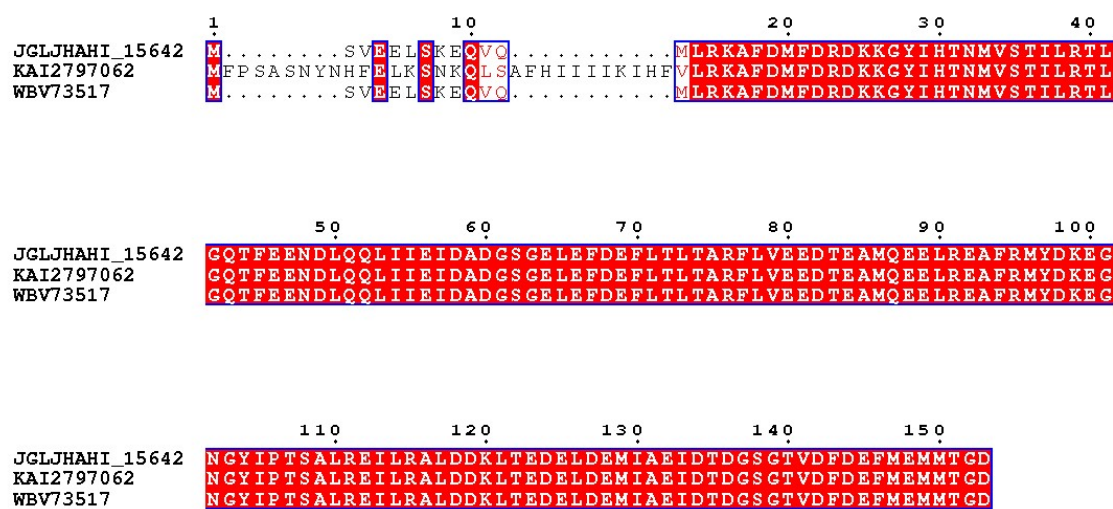
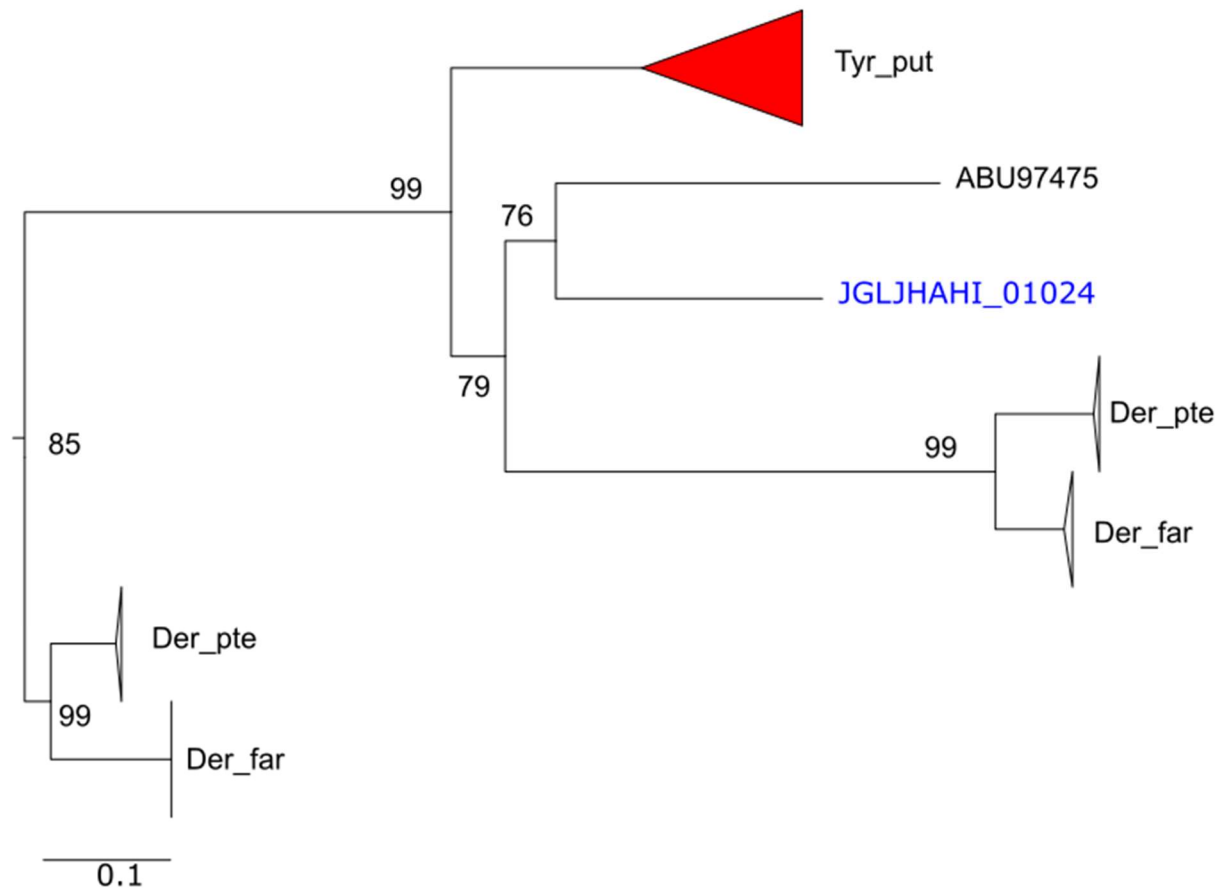


FIGURE S45 Alignment of predicted Blot t 34 proteins.



Group 35 was formed by two identified allergens, Der f 35 (BAX34757) and Tyr p 35 (AOD75396). The sequence identity was 24.4%. Der f 35 (BAX34757) showed structural homology to the ML domain, i.e., the 1-17 signal peptide 19-141 ML domain. This was analogous to the group 2 allergens, i.e., the identity to Der f 2 (CAI05849) was 40.3%. For this reason, BAX34757 was eliminated from further comparisons. The Tyr p 35 (AOD75396) structure corresponded to the aldehyde dehydrogenase family (14-478). Two *Blomia tropicalis* sequences were predicted (JGLJHAHI_01024 and JGLJHAHI_07968), with the same structure as Tyr p 35 (AOD75396), but JGLJHAHI_07968 was not confirmed by proteomic analyses. The identity of Tyr p 35 (AOD75396) and JGLJHAHI_01024 was 73.3% (Table S28). The JGLJHAHI_01024 sequence clustered as a sister group to Tyr p 35 allergens (Figure S46).

FIGURE S46 Comparison of group 35 allergens. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was OTF72067 from *Euroglyphus maynei*.



Group 36 contained two different allergens that were incorrectly assigned here: the Der p 36 (AT108932) and Der f 36 (AT108931) peptides, with a C2 domain (i.e., AT108931: 1-24 signal peptide, 90-208 CZ domain). The identity of the peptides was 77.6%. Two predicted *Blomia tropicalis* proteins (JGLJHAHI_06496 and JGLJHAHI_09342) with the same structure were found.

The identity of the proteins was 52.3%. JGLJHAHI_06496 showed 45.5% identity to Der p 36 (AT108932) and 44.1% identity to Der f 36 (AT108931). JGLJHAHI_09342 showed 52.6% identity to Der p 36 (AT108932) and 51.9% identity to Der f 36 (AT108931) (Table S29). The *B. tropicalis* proteins formed two separate clusters outside of *Dermatophagoides* allergens (Figure S47). However, the alignment of predicted *B. tropicalis* allergens showed high levels of heterogeneity (Figure S48).

FIGURE S47 Comparison of group 36a allergens (SP, peptides with C2 domain). Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was OTF76064 from *Euroglyphus maynei*.

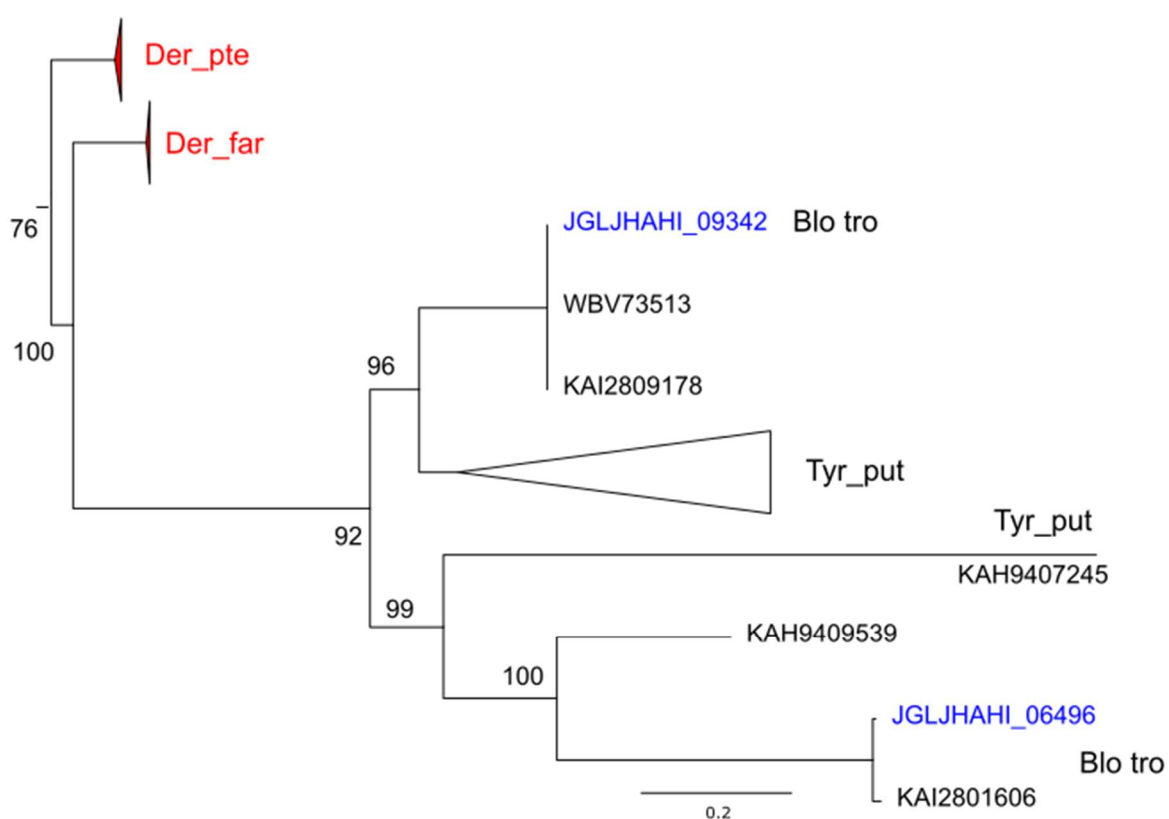


FIGURE S48 Alignment of the predicted group 36a allergens (SP, peptides with C2 domain) from *Blomia tropicalis*.

	1	10	20	30	40	50	60
JGLJHAHI_09342	MKEIHSIVLVAF	CWASVQADQLAEAK	CRSLHHVDVDP	STSYNSQC	LLT	CVIHGSANYHE	
JGLJHAHI_06496	MKS.QIILLICF	VATCTKASTEAEEQ	CRTIHHTDVDAE	TRYFNEK	CLLNCNVH	GKIWSHF	
WBV73513	MKEIHSIVLVAF	CWASVQADQLAEAK	CRSLHHVDVDP	STSYNSQC	LLT	CVIHGSANYHE	
KAI2809178	MKEIHSIVLVAF	CWASVQADQLAEAK	CRSLHHVDVDP	STSYNSQC	LLT	CVIHGSANYHE	

	70	80	90	100	110	120
JGLJHAHI_09342	MNEGLPCPLASSG	ICRRGQCVQQSAL	GHIDIEMT	SASLSAAYNAYAT	VCIONASVPM	MTLP
JGLJHAHI_06496	LNENGPCGPGYGM	ICNDGSCIKAAHI	GHIDIEIK	SASFRAVNAYAM	VSIFNSTNE	MTLP
WBV73513	MNEGLPCPLASSG	ICRRGQCVQQSAL	GHIDIEMT	SASLSAAYNAYAT	VCIONASVPM	MTLP
KAI2809178	MNEGLPCPLASSG	ICRRGQCVQQSAL	GHIDIEMT	SASLSAAYNAYAT	VCIONASVPM	MTLP

	130	140	150	160	170	180
JGLJHAHI_09342	IQDRSGCTSCST	SVITSTYPRWNTV	CAGSGKNLFVSE	SRVTFEVWDNH	GAGAANKI	FLGG
JGLJHAHI_06496	IQDRS	SATTC	KTSVITNDYPHWN	SICTGSDALRFVE	QSRVIFEFVDN	FENG.D.RQTF
WBV73513	IQDRSGCTSCST	SVITSTYPRWNTV	CAGSGKNLFVSE	SRVTFEVWDNH	GAGAANKI	FLGG
KAI2809178	IQDRSGCTSCST	SVITSTYPRWNTV	CAGSGKNLFVSE	SRVTFEVWDNH	GAGAANKI	FLGG

	190	200	210	220
JGLJHAHI_09342	ASLTIPQLLSHND	NHNPLHLALT	GGPKPGQINAR	VTWTPRN
JGLJHAHI_06496	ASLTIPQLLSHND	NKQLHLALS	GGLIASQINCRIT	WTPI.
WBV73513	ASLTIPQLLSHND	NHNPLHLALT	GGPKPGQINAR	VTWTPRN
KAI2809178	ASLTIPQLLSHND	NHNPLHLALT	GGPKPGQINAR	VTWTPRN

Tyr p 36 had profilin (alignment region 1-131) domains and showed 94.7% identity to the predicted *Blomia tropicalis* JGLJHAHI_14665 protein (Table S30). JGLJHAHI_14665 formed a sister cluster to Tyr p 36 (Figure S49). The *B. tropicalis* proteins showed 100% identity (Figure S50).

FIGURE S49 Comparison of group 36b (profilin) allergens. Red indicates the identified allergen proteins, and blue indicates predicted proteins of *Blomia tropicalis*. The outgroup sequence was AAX34044 from *Suidasia medianensis*.

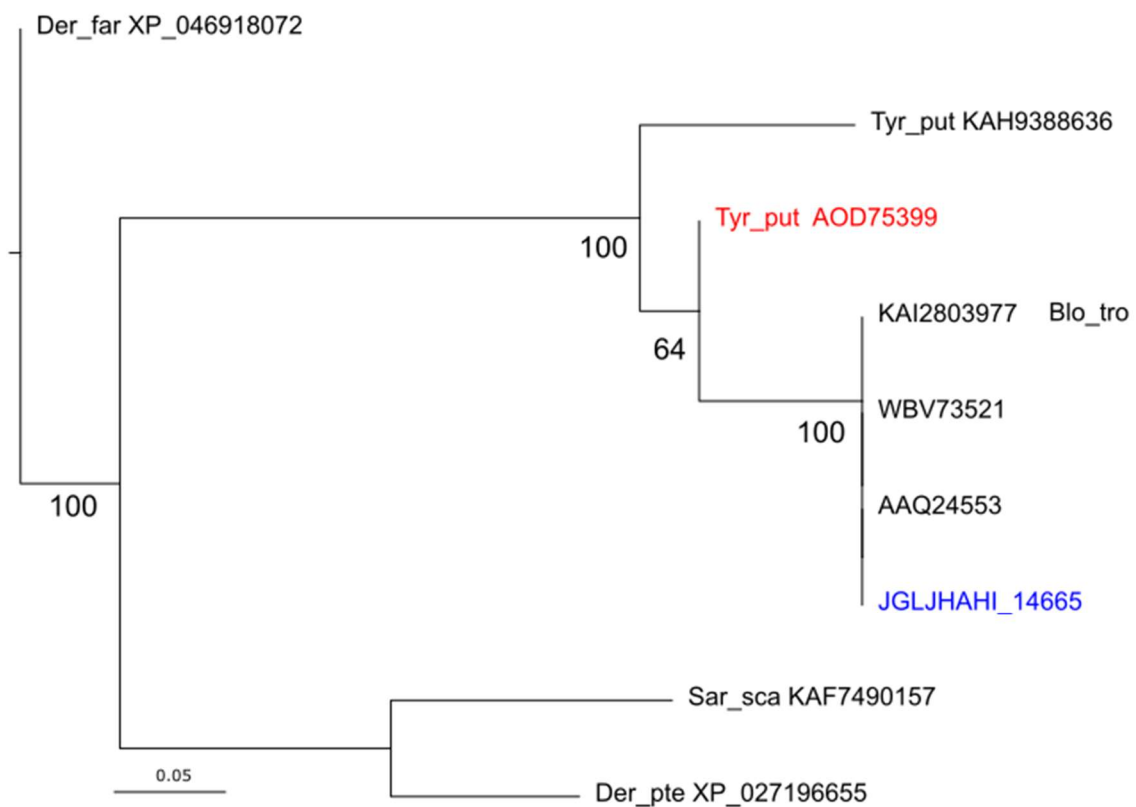


FIGURE S50 Alignment of predicted group 36b (profilin) allergens of *Blomia tropicalis*.



Comments on the analyses: We used GenBank sequences assigned as *Rhagoletis zephyria* (GCF_001687245.1). Our previous analyses showed that this culture was contaminated by the stored product mite *Tyrophagus putrescentiae*.⁹¹ Therefore, we refer to the *R. zephyria* sequences as coming from *T. putrescentiae* in this analysis.

Supplementary references

91. Hubert J, Nesvorna M, Klimov P, Dowd SE, Sopko B, Erban T. Differential allergen expression in three *Tyrophagus putrescentiae* strains inhabited by distinct microbiome. *Allergy*. 2019;74(12):2502–2507. <https://doi.org/10.1111/all.13921>