

Proteogenomics of the house dust mite, *Dermatophagoides farinae*: Allergen repertoire, accurate allergen identification, isoforms, and sex-biased proteome differences

Tomas Erban^{a,*}, Pavel Klimov^{b,c}, Pavel Talacko^d, Karel Harant^d, Jan Hubert^a

^a Crop Research Institute, Drnovska 507/73, Prague 6-Ruzyně, CZ-16106, Czechia

^b Department of Ecology and Evolutionary Biology, University of Michigan, 3600 Varsity Drive, Ann Arbor, MI 48109, USA

^c Institute of Biology, University of Tyumen, Pirogova 3, 625043 Tyumen, Russia

^d Proteomics Core Facility, Faculty of Science, Charles University, BIOCEV, Prumyslova 595, Vestec, CZ-25242, Czechia

ARTICLE INFO

Keywords:

Label-free proteomics
Serpins Der f 27
Ferritin Der f 30
Der f 28
Der f 9
Der f 38

ABSTRACT

The allergen repertoire of the house dust mite, *Dermatophagoides farinae*, is incomplete despite most mite allergens having been described in this species. Using proteogenomics, we aimed to compare proteins and allergens between sexes and provide a foundation for the identification of novel allergens. Overall, 6297 protein hits were identified, and 2899 and 886 were male- and female-specific, respectively. Removal of trace results narrowed the dataset to 3478 hits, including 275 and 157 male- and female-specific hits, respectively. All 34 WHO/IUIS-approved *D. farinae* allergens (omitting Der f 17) were identified, and we also identified homologs of the yet undescribed Der f 9 and 38. Der f 27/serpin exhibited the largest sex-dependent difference and was dominant in females. Using official protein sequences, Der f 11, 14, 23, 28 and 30 were identified with low success. However, identification success of Der f 11 and 14 was greatly increased by using longer/complete sequences. Der f 30 is characterized by the same tryptic digests as the more abundant Der f 30 (isoform) identified here. Der f 23 appears to be of low abundance in mite bodies. Der f 28.0101 and Der f 28.0201 were detected at low abundance and in trace amounts, respectively.

Significance: In this work, we performed a proteogenomic annotation of the house dust mite, *Dermatophagoides farinae*, which is the most important source of house dust allergens. The proteogenomic analysis performed here provides a foundation for not only understanding the biology of the mite but also the identification of novel allergens. This study generated a robust proteomic dataset for *D. farinae* and reviewed existing and candidate allergens in this species. We stress some pitfalls of high-throughput analyses, especially that improper headers of allergen protein records provided in databases can lead to confusion. Using partial sequences in proteomic identification and quantification can lead to low identification success (low signal intensity or MS/MS counts). Thus, we individually curated the protein sequences for proper identification and quantification. The discovered sex differences can be one factor affecting allergen/immunogen variations in mite extracts. Overall, this work provides a benchmark for accurate identification of mite immunogenic proteins using proteomics.

1. Introduction

Allergy to house dust mites (HDMs) affects millions of people worldwide, mostly asthma sufferers. In the latter category, 85% of people have asthma induced by HDMs [1,2]. According to the World Health Organization (WHO) and International Union of Immunological Societies (IUIS) allergen nomenclature subcommittee website (WHO/IUIS) [3], 101 (100 if omitting Der f 17) allergens from 10 astigmatid mite species (Acari: Astigmata or Acaridida) have been officially

identified and categorized into 39 biochemically distinct allergen groups. *Dermatophagoides farinae* Hughes, 1961 is the mite species with the largest number (34 if omitting Der f 17) of known allergens. However, one allergen group (Der f 17) appears to be represented by an incomplete record in the database; although it has been reported in the literature [4,5], the actual sequence has never been published (see WHO/IUIS [3]). Mass spectrometry-based proteomic (MSP) approach including targeted analysis employing multiple reaction monitoring (MRM) is a useful and beneficial strategy for quality control and

* Corresponding author at: Crop Research Institute, Proteomics and Metabolomics Laboratory, Drnovska 507/73, Prague 6-Ruzyně, CZ-16106, Czechia.
E-mail address: arachnid@centrum.cz (T. Erban).

<https://doi.org/10.1016/j.jprot.2019.103535>

Received 28 May 2019; Received in revised form 28 August 2019; Accepted 26 September 2019

Available online 17 October 2019

1874-3919/ © 2019 Elsevier B.V. All rights reserved.

standardization of allergenic extracts [6].

The MSP approach has been widely used in recent years for protein identification. For this purpose, whole-genomic sequence data greatly facilitate high-throughput proteomic analyses by providing large-scale datasets annotated with information on protein functions [7]. Another possibility is to use de novo transcriptomes to construct a database for proteomic identification [8]. For *D. farinae*, there are several proteomic studies involving allergen analysis using different sources for proteomic identification, i.e., publicly available sequences [9–14] or transcriptomes [15,16].

Currently, the National Center for Biotechnology Information (NCBI) has several hundred protein sequences of *D. farinae*, many of which are nonredundant (nr) and inferred from deduced protein sequences. Two draft genomes of *D. farinae* are also available. The first reported draft genome (size: 53.5 Mb) was sequenced using an Illumina platform (BioProject No. PRJNA174061) [17]. The second draft genome (size: 91.9 Mb) was assembled using both PacBio and Illumina data (BioProject No. PRJNA379991) [18]. These draft genomes can be used to construct databases for proteomic data evaluation. Conversely, the combination of proteomics and genomics, known as proteogenomics, can be used to facilitate de novo assemblies of genomes [19]. For proteogenomic annotations, multiple tissues of animals are commonly analyzed [19]; however, astigmatid mites, including HDMs and stored product mites (SPMs), are microscopic, making it virtually impossible to obtain specific mite tissues [20]. Furthermore, for accurate proteomic analysis, it is necessary to use high-quality samples that are free from contaminants, such as environmental particles of debris or fungal mycelia adhering to the mite cuticle or setae.

Extracts of HDMs are used for allergy diagnostics as well as for immunotherapy. For this reason, the HDM extracts must not contain impurities, and the allergen levels must be standardized. However, studies have shown that mite extracts may differ in their composition, which mainly depends on the method of cultivation and the method of sample processing prior to allergen extraction [21–25]. In addition, allergen composition changes over time with development of the mite culture [26–28]. In another important HDM species, *Dermatophagoides pteronyssinus* (Trouessart, 1898), expression of some allergens (as determined using qPCR) is influenced by mite developmental stage, mite sex and the environment [29]. As another example, a sex bias in allergen expression has been shown for the cat allergen Fel d 1, which is more abundant in male cats than in female cats [30]. Importantly, the sex ratio in *D. farinae* varies with culture development, and according to Hodgson [31], males become dominant in older cultures [31]. We therefore hypothesize potential sex-dependent differences in allergen production in mite cultures.

Here, we performed a proteogenomic annotation of *D. farinae*. We used separate samples of male and female mites for label-free quantitative proteomic analyses employing nanoliquid chromatography coupled with a state-of-the-art Orbitrap Fusion Tribrid mass spectrometer (nLC-MS/MS). We hypothesized that a sex-specific analysis would improve proteogenomic annotation and that the differences in expression levels between sexes would have implications for allergen/immunogen production. Finally, we provide a curated list of *D. farinae* allergens, give sequences for allergens yet to be described at the protein level that identify allergen-like proteins for future allergenicity evaluation, and point out inaccurate sequence headers in public databases.

2. Materials and methods

2.1. Mites

A culture of *D. farinae* was obtained from Prof. K. Solarz in Poland and since then has been maintained at the Crop Research Institute for ca. 15 years. This mite culture has been used by us in a number of

previous studies; for this work, culture maintenance was performed as described previously [12,32–34]. The culture was periodically checked for contamination by identifying a subsample of mites using morphology.

For genomic analyses, we used *D. farinae* from a pure laboratory culture maintained at the University of Michigan Museum of Zoology (BMOC 05-0812-001) started from specimens collected from a skeleton-cleaning culture of *Dermestes maculatus* (Coleoptera: Dermestidae) in Ann Arbor, MI, USA, in 2005. Fish flakes were used as food. Slide-mounted mites were identified using morphology, and the identification was confirmed by GenBank expressed sequence tag (EST) database sequences and multiple mitochondrial genes. The colony was used in several previous studies, including those that produced a complete mitochondrial genome [35,36] and whole-genomic and short RNA sequence data [18].

2.2. Collection of mites

Females and males of *D. farinae* were collected from healthy cultures (large proportion of live mites and low rate of mite death) and increasing population sizes. Mite cultures were maintained in tissue culture flasks [12,32–34]. Live mites that were similarly sized and free of food particles were collected individually from a glass Petri dish using a thin paintbrush with a sharp tip under a dissecting microscope (Carl Zeiss, Jena, Germany). Then, the mites were placed directly into a 2-mL Potter-Elvehjem tissue grinder with a Teflon pestle (Kartell, Labware division, Noviglio, Italy) kept in a box with dry ice. Each of the four biological replicates for male and female mite samples of *D. farinae* consisted of 1000 individual mites.

2.3. Proteomic analysis using label-free nLC-MS/MS analyses

To the Potter-Elvehjem homogenizers with 1000 *D. farinae* males or females, 1 mL of 100 mM 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) was added. Next, the samples were processed and analyzed using nLC-MS/MS as described elsewhere [37]. Each of the replicates per sex was analyzed twice; thus, a total of 16 consecutive nLC-MS/MS analyses were performed. The data were analyzed and quantified with label-free quantification (LFQ) algorithms in MaxQuant v1.6.3.3 [38] and the Andromeda search engine [39]. The false discovery rate (FDR) parameter was set to 1% for both proteins and peptides. The enzyme specificity of trypsin was set as C-terminal to Arg and Lys. MethylThio was the fixed modification, while N-terminal protein acetylation and methionine oxidation were variable modifications. The data were searched against a database created by us.

2.4. Search database creation

For the proteogenomic analysis, we performed a six-frame translation using MaxQuant v1.6.3.3. The draft genome available in NCBI under BioProject No. PRJNA379991 was used, and we selected a minimal length of 50 amino acids as an output. The resulting sequences were further annotated using Blast2GO 4 (Biobam, Valencia, Spain) against the NCBI Inr Arthropoda database available in the Blast2GO cloud in April 2018. The resulting dataset of annotated proteins was enriched for existing protein sequence records related to *D. farinae* taxonomy in NCBI Inr, which consisted of 315 sequences downloaded on 21 Oct 2018; we added the “gi” accession numbers to the sequences for easier comparison with previous studies. The database that we created is available for download along with the raw dataset used in our analyses. To detect contaminant sequences, we used the MaxQuant standard dataset and a dataset of 7918 sequences of yeast, *Saccharomyces*

cerevisiae, downloaded from UniProt on 30 Oct 2018. The latter database should be able to detect the level of food-derived contamination because yeasts constituted part of the mites' diet.

2.5. Overall proteomic data evaluation

The LFQ intensities from MaxQuant [38] were analyzed in Perseus v1.6.2.3 [40]. The dataset was log2-transformed. Potential contaminants, hits identified only by site and reverse hits were filtered out. The data were evaluated as the total number of 16 nLC-MS/MS runs since it was beneficial to see the positives in each of the analyses. The normal distribution of the data was inspected through histograms (Fig. S1). Protein hits that contained three or fewer valid values in the 16 nLC-MS/MS runs were filtered out, the data before and after this step were averaged via the median (of the valid values), and the data were evaluated using a two-sample *t*-test ($p < 0.05$). Both matrices were then arranged based on the difference between the averages of males and females, and the protein hits that were not detected in any of analyses in the males or females were colored.

During data evaluation, we worked with all 16 analyses and did not average the results of technical replicates since averaging analytical replicates within biological replicates does not mathematically affect the final result. We considered the qualitative results to be either truly qualitative or below the detection limit. Two heatmaps visualizing male vs. female proteome differences were generated. To produce the heatmaps, the missing values were replaced as follows: (a) missing values were replaced by the constant 0, or (b) missing values were replaced based on a normal distribution (width: 0.3; down shift: 1.7; total matrix). The second option was additionally normalized through the Z-score. Finally, the data were averaged via the mean in both cases, and the heatmaps were generated (Euclidian distance; 300 clusters; 10 iterations). The resulting heatmaps were adjusted to grayscale. To visualize the differentially expressed proteins between the sexes, we also performed two volcano plot analyses (*t*-test; number of randomizations: 250; FDR: 0.05; S0: 0.1) of data processed analogously to the heatmap data.

2.6. Identification of allergens, allergen isoforms, homologs, analogs, and proteins of allergen potency

We identified the official WHO/IUIS-approved mite allergens, isoforms, and homologs or analogs. Although there are additional databases of allergenic molecules, such as Allergome [41], we restricted our analysis to the *D. farinae* allergens approved by the WHO/IUIS Allergen Nomenclature Subcommittee [3]. The sequences that correspond to proteins of interest were found through individual examination of the list of protein hits. Identification of the protein hits corresponding to allergens in our dataset facilitated the inclusion protein sequences of NCBI in the search database. Every official allergen is listed in the WHO/IUIS database; therefore, we verified via individual inspection which results of our analysis exactly corresponded to the allergen Protein Access No. provided in the WHO/IUIS database. When only the Nucleotide Access No. was provided in the WHO/IUIS database, we found the protein sequence ID.

Because the major mite allergens include various proteases and proteins with chitinase/chitin-binding domains, we focused on these proteins as potential new immunogens/allergens. In addition, we selected hits corresponding to vitellogenin because of its domain similarity to Der f 14 and because vitellogenin is an official allergen in other organisms (i.e., Api m 12 and Ves v 6). Selected results were then inspected individually and verified using BLASTP [42] and the conserved domains (CCDs) [43,44]. For data evaluation, we also used the WHO/IUIS allergen database [3], with which we individually verified allergen sequences and related references. Selected proteins were aligned using Clustal Omega v1.2.4 [45] and visualized using Jalview v2.10.5 [46].

3. Results

3.1. Overall proteogenomic data evaluation

Our proteogenomic analysis produced 6297 protein hits for both sexes of *D. farinae* (Table S1); 2899 and 886 hits were specific to males and females, respectively. Reducing the matrix to at least three positive results of the 16 total nLC-MS/MS runs narrowed the dataset to 3478 hits, and the numbers of male- and female-specific hits were reduced to 275 and 157, respectively (Table S2). Thus, we consider the 2819 proteins excluded by the threshold of at least three positives to be of very low abundance or trace amounts in the proteome. Most of the trace protein hits were detected in only females. Protein hits that could be identified using the current records of NCBI nr-protein (containing a "gb" identifier) are highlighted in Table S2. The two heatmaps (Fig. 1AB) and volcano plots (Fig. 2AB; Tables S3 and S4) created using two different algorithms illustrate the large proteome dissimilarity between sexes. The volcano plot in Fig. 2A nicely separates the relatively high-intensity protein hits that were detected in only one sex. An inspection of these groups of proteins specific to males or females in Table S3 revealed that all the hits with a $-\log(p\text{-value}) > 20$ (see also Fig. 2A) were detected in all eight nLC-MS/MS runs in only one of the sexes. Furthermore, all of these proteins were newly identified. The data created from the Z-score-subtracted dataset (Figs. 1B and 2B) provide different outputs because the data being rounded to accommodate the processing method, that is, the key qualitative differences are not as easily identifiable in the chart.

3.2. Identified allergens of *D. farinae* and homologs of allergens

All 34 *D. farinae* allergens currently listed in the WHO/IUIS database (and for which sequences are available) were identified (Table 1; Tables S2 and S5). Furthermore, we identified proteins homologous to allergens described in other mites, especially *D. pteronyssinus*, but not yet identified in *D. farinae*. Thus, we provide evidence at the protein level for as yet undescribed allergens Der f 9 and 38 (Table 1: entries marked by an exclamation mark). However, we were not successful in identifying Der f 12 and 19; finally, Der f 17 could not be identified because the corresponding sequence was not available (Table 1: Der f 12, 17 and 19 are marked by a question mark). Table 1 provides a quantitative comparison of allergens (WHO/IUIS approved or predicted by us) between the sexes, while Table S5 summarizes the selected proteins in more detail, including actual sequences, individual BLASTP searches and notes from our analyses. Table 2 provides simplified information for additional proteins investigated in Table S5.

3.3. The serpin allergen Der f 27 exhibited a large difference between sexes, and additional serpins were identified

Among the official WHO/IUIS allergens, the largest quantitative difference between sexes was found for Der f 27 (serpin; serpin SMSB3a variant). Der f 27 was more abundant in females, and the log2 fold change was 4.95, one of the largest differences observed. We found two isoforms of Der f 27, which were also serpin SMSB3a variants, and both were detected in all samples. Interestingly, one isoform (scaffold138|size173686_True144872; query coverage of 89% and identity of 32.65% to Der f 27) was more abundant in males (log2 fold change of -1.69), unlike Der f 27. The third protein hit of serpin (scaffold1|size1740722_False_942723 serpin SMSB3a variant; query coverage of 94% and identity of 28.93% to Der f 27) was also significantly more abundant in males, but the log2 fold change was only -0.54 . Finally, we identified a serpin B10-like protein (scaffold13|size626939_True484469). While the overall intensity and MS/MS counts indicated a relatively high abundance of the three SMSB3a variant serpins, serpin B10-like appeared to be of low abundance.

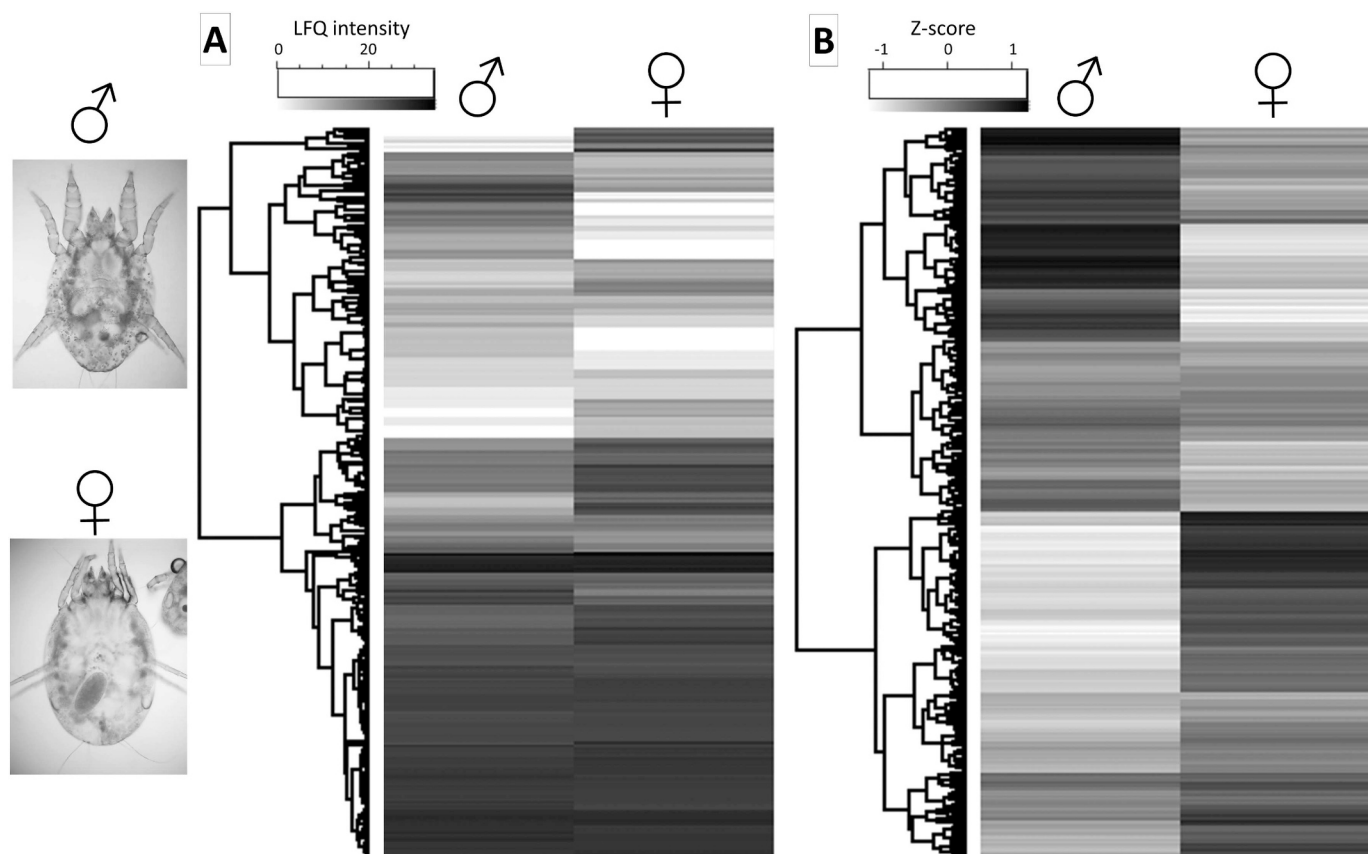


Fig. 1. Two heatmaps demonstrating the large proteome difference between *D. farinae* males and females. A) A resulting heatmap after the missing values were replaced by the constant “0” and a mean was calculated. The heatmap clearly shows the missing values or qualitative differences (in white). B) A heatmap was generated after the missing values were replaced according to a normal distribution and the data were normalized through Z-score. This heatmap appropriately represents the data comparison for proteins that were steadily detected among the samples, but the qualitative result is less visible.

3.4. Low identification success for some allergens when using the WHO/IUIS database

The allergens Der f 11, 14, 23, 28, and 30 were detected with low success (low intensity and MS/MS counts) using sequences provided in the WHO/IUIS database (Table 1, Table S5). Therefore, these allergens appeared to be of low abundance. To verify this result, we further searched hits similar to Der f 11, 14, 23, 28, and 30 in our dataset (Table S2). Our investigation revealed alternative protein hits to the official WHO/IUIS Der f 11 and Der f 14 protein records: AAK39511 and BAA04558, respectively. In particular, Der f 11 (AAK39511, 692 aa) provided only 42 MS/MS counts in the entire analysis, while the results related to the longer/complete paramyosin sequence (AIO08864; 876 aa) provided a much larger number of MS/MS counts (2784). We observed a similar situation for the Der f 14 allergen (Table 1 and Table S2). While the official WHO/IUIS Der f 14 entry (BAA04558; 349 aa) provided only eight MS/MS counts and was identified in only one of the 16 nLC-MS/MS runs, the longer sequence of apolipoprotein (BBD75204; 1666 aa) had much higher identification success, resulting in a large number of MS/MS counts (4275), and the identification was successful in all 16 nLC-MS/MS runs. Note that in the shorter WHO/IUIS sequence of Der f 11, there is one missing amino acid compared to the longer sequence, while the shorter WHO/IUIS sequence of the Der f 14 allergen is identical to the corresponding part of the sequence record BBD75204.

For Der f 23, we did not find alternative protein hits. Thus, based on this result, we may consider Der f 23 a low-abundance allergen in mite bodies.

The protein hit corresponding to the official WHO/IUIS ferritin/Der f 30 sequence (AGC56219) provided identification in only one analysis in males with a total MS/MS count of only 17. However, in our dataset, we found additional matches representing two ferritin isoforms: (i) scaffold39|size422454_False_409955 and scaffold211|size71058_False_20376 of the same protein sequence, which we denoted for use in this study as Der f 30 (isoform), and (ii) a short ferritin isoform, scaffold24|size505992_False_78980, which we denoted Der f 30-like. The identification success for Der f 30 (isoform) and Der f 30-like was substantially higher than that for the ‘official’ Der f 30. Moreover, we found that Der f 30 was similar to Der f 30 (isoform), with a 98% query coverage and an 83% identity between them, while Der f 30-like was a short form, probably a light chain, sharing a sequence coverage of 78% and an 83% identity with Der f 30. Importantly, we found that Der f 30 and Der f 30 (isoform) produce analogous tryptic digests (Fig. 3). Thus, we hypothesized that these isoforms could be confused in studies through MSP identification because the sequence of Der f 30 (isoform) was not available prior to this study. We checked previously published studies that identified ferritin/Der f 30 using an MSP approach. In our previous 2D-E-MS/MS study focusing on the proteome/allergome of mite feces, we provided MS/MS-identified peptides [12], and therefore, we were able to recover the MS/MS-identified peptides LADYITNLER and TVNQALLDLHAVATK that matched ferritin/Der f 30. Next, in a study by Choopong et al. [11], ferritin/Der f 30 was identified by the peptides LADYITNLER, TVNQALLDLHAVATK and IVLDDIHKPQQQD-WSSGLEAMR (see Table S3 in [11]). The two former peptides were the same in both studies. Furthermore, in a study by An et al. [9], two peptides, namely, NDAQFADFIETHYLTEQVEAIK and

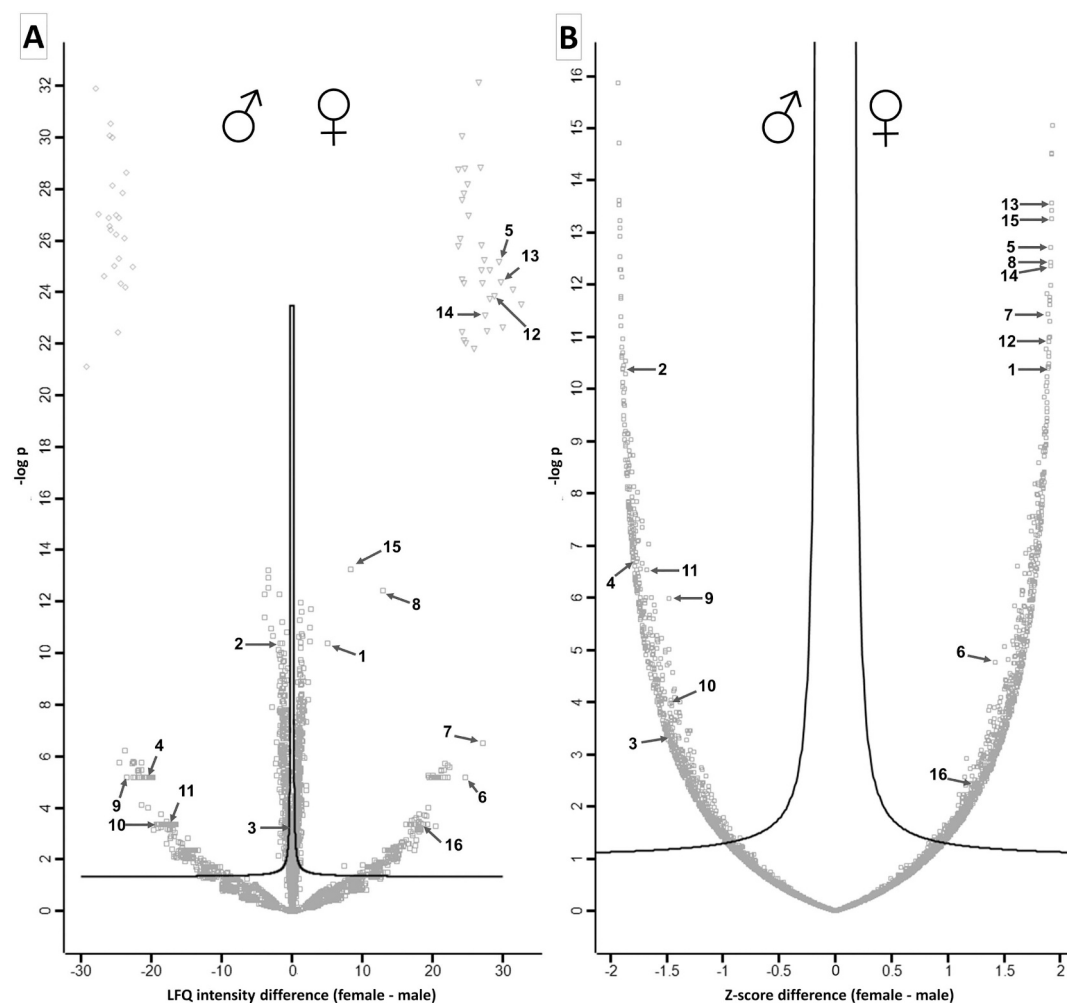


Fig. 2. Volcano plots showing proteome differences between *D. farinae* males and females when A) the missing values were replaced by zeroes or B) the dataset was normalized through Z-scores. The first plot clearly identifies protein hits identified in one sex at a relatively high abundance and not detected in the other sex. Moreover, the allergen that showed the largest quantitative difference (Der f 27/serpin – marked as 1) is shown. Furthermore, we highlight the other serpin isoforms, cathepsins L, serine protease 27, vitellogenins and apolipophorins. Data for Fig. 2A and B are given in Tables S3 and S4, respectively.

1) gi|685432794|gb|AIO08851.1| Der f 27 allergen; 2) scaffold138|size173686_True_144872 serpin SMSB3a variant; 3) scaffold1|size1740722_False_942723 serpin SMSB3a variant; 4) scaffold33|size445179_True_225812 serine protease 27; 5) scaffold90|size237115_True_117857 vitellogenin 1; 6) scaffold90|size237115_True_117574 vitellogenin-like protein; 7) scaffold90|size237115_True_122408 vitellogenin, putative; 8) scaffold90|size237115_True_118681 vitellogenin-like protein; 9) scaffold368|size33359_True_18656 cathepsin L1-like; 10) scaffold195|size88217_True_14968 cathepsin L1-like; 11) scaffold89|size238932_True_235616 cathepsin L1-like; 12) scaffold168|size125973_False_51957 Group 14 mite allergen-like protein (Apolipophorin-like); 13) scaffold40|size417961_False_203038 Apolipophorin-like protein; 14) scaffold40|size417961_False_203757 Apolipophorin-like protein; 15) scaffold168|size125973_False_51293 Apolipophorin-like protein; 16) scaffold38|size423637_True_361700 putative bifunctional chitinase/lysozyme.

MAANPESTTKTSRVR (N-terminus), matched ferritin/Der f 30 (see Supplement 1 Fig. S2H in An et al. [9]). Together, we found that the four peptides of the three abovementioned studies (LADYITNLER, TVNQALLDLHAVATK, IVLDDIHKPQQDWSSGLEAMR and NDAQFADFIETHYLTEQVEAIK) are identical in both Der f 30 and Der f 30 (isoform) (Fig. 3); therefore, these peptides, resulting from trypsin digestion, identify the two isoforms with the same probability. Note that the N-terminal sequence MAANPESTTKTSRVR highlighted in An et al. [9] predicts trypsin digestion at three cleavage sites (10, 13 and 15; as analyzed in [47]), and the authors [9] reportedly performed sequencing using Edman degradation. Thus, the N-terminal peptide sequence of Der f 30 should not be included in the comparison of peptides for MSP identification. Overall, this result shows that the two ferritin isoforms with a mutual query coverage of 98% and an identity of 83% (Fig. 3) could be mistaken in the MSP identification; however, results of this study show that Der f 30 (isoform) is considerably more abundant in mite bodies. The third, shorter ferritin variant (Der f 30-like) identified in all our analyses contained different peptides for identification

(Fig. 3).

Finally, the Der f 28 allergen isoform Der f 28.0101 (AGC56218.1) was identified in only three analyses in both females and males, with similar abundances in both sexes. In addition, in the dataset in which trace results were excluded, we were unable to find the second WHO/IUIS Der f 28 isoform, Der f 28.0201 (AIO08848.1) (Table S2); however, we found it in the full dataset with a positive result in only one out of 16 performed analyses (Table S1). Thus, both Der f 28 isoforms appear at a low abundance, with the latter isoform appearing only in trace amounts in mite bodies. Furthermore, the WHO/IUIS provides a comment on these two isoforms posted on 5 Nov 2014: “The overall identity match of Der f 28.0101 and Der f 28.0201 is ~ 57 %, which is beyond the normal range for isoallergens in WHO/IUIS, however, the first 3/4 of the sequences are of much higher identity, warranting the designation”. Der f 28 is a heat shock protein 70, and in our dataset, we found five protein hits that we denoted here as Der f 28-like. All of them had a higher identification success than the official WHO/IUIS sequence for Der f 28. All of the sequences corresponding to Der f 28-like appear to

Table 1
Thirty-nine allergen groups currently known in astigmatid mites and their curated identification by a proteogenomic approach in *D. farinae*.

Intensity	MS/MS count	DF _f	DF _m	DE _f - DF _m	T-test	Allergen designation	WHO/IUIS	Representative protein	Note	Conserved domains (CCDs)
3E+10	557	30.89	30.11	0.78	+	Der f 1	Yes	BAC53948	This study identifies Der f 5 using proteomics.	Peptidase_C1A; Inhibitor_I29
3E+10	209	30.63	30.04	0.59	+	Der f 2	Yes	BA01239		Der-p2 like; ML s.f.
2E+09	185	26.85	26.52	0.32	+	Der f 3	Yes	BA00920		Tryp_SPC
6E+09	273	28.11	27.83	0.28	+	Der f 4	Yes	AHX03180		AmyAc_bac_euk_AmyA; Amy_C
2E+10	263	30.24	29.22	1.02	+	Der f 5	Yes	AB084970		Blo-t-5 s.f.
1E+09	97	25.92	25.46	0.46	+	Der f 6	Yes	AAT28423	This study identifies Der f 9 at the protein level.	Tryp_SPC
3E+10	383	30.60	29.84	0.76	+	Der f 7	Yes	AAB35977		Grip7 allergen s.f.
2E+10	358	30.28	29.69	0.59	+	Der f 8	Yes	AGC56215		GST_C family s.f.; Thioredoxin like s.f.
8E+08	74	25.5	25.2	0.25	+	Der f 9 (!)	No	AAP35067		Tryp_SPC
2E+11	789	32.01	33.28	-1.27	+	Der f 10	Yes	BA04557		Tropomyosin
9E+08	42	25.27	26.18	-0.91	+	Der f 11	Yes	AAK39511	Low identification success due to partial sequence! Longer sequence increased the identification success above. Only Blo t 12 has been described and similar protein was identified in <i>Lepidoglypus destructor</i> .	SMC_N s.f.
3E+11	2784	32.42	33.72	-1.30	+	Der f 11	No	AIO08864		SMC_N s.f.
						Der f 12 (?)	No			Blo t 12 (gb AAA78904) domain: CBM_14
2E+10	218	30.16	29.61	0.55	+	Der f 13	Yes	AAP35078		Lipocalin s.f.
2E+07	8	NaN	21.23	NaN	+	Der f 14	Yes	BA04558		none
3E+11	4275	33.95	33.85	0.10	+	Der f 14	No	BB075204	Low identification success due to partial sequence! Longer sequence increased the identification success above.	LPD_N (Vitellogenin_N s.f.); DUF1943 s.f.
8E+09	368	29.10	28.18	0.92	+	Der f 15	Yes	AAD52672		GH18_chitinase-like s.f.; ChtBD2
3E+10	471	29.82	30.76	-0.95	+	Der f 16	Yes	AAM64112		gelsolin_S1 like; gelsolin_S3 like; 2 ×
										ADF_gelsolin s.f.
										? EF-hand calcium-binding?
6E+09	290	28.63	27.77	0.86	+	Der f 17 (?)	No		Sequence absent! Empty allergen group of the mites? Only Blo t 12 has been described. Nematode-like peptide.	GH18_chitinase-like s.f.; CBM_14
						Der f 18	Yes	AAM19082		Blo t 19 (gb AHG97583): Antimicrobial25 s.f.
						Der f 19 (?)	No			arginine_kinase like
										arginine_kinase like
										Blo-t-5s.f.
2E+11	1105	32.00	33.41	-1.41	+	Der f 20	Yes	AIO08850	Der f 20.0101 (gb AIO08850) and Der f 20.0201 (gb ABU97470) isoforms have an identity of only 83.99%, different protein hits.	ML s.f.
7E+10	666	31.54	31.60	-0.06	+	Der f 20	Yes	ABU97470		ChtBD2
3E+10	326	30.60	29.81	0.79	+	Der f 21	Yes	AHC94806		UCR_14kD
9E+09	99	29.17	28.33	0.85	+	Der f 22	Yes	ABG35122		PTZ00333
1E+08	17	23.00	24.03	-1.04	+	Der f 23	Yes	ALU66112		EFH_PEF s.f.
6E+09	263	27.70	28.08	-0.39	+	Der f 24	Yes	AGI78542	Low abundance in mite bodies. Der f 25.0101 and Der f 25.0201 (gb AIO08860) identity of 87.85%. Renamed from Der f 30, but the header is still not changed!	SERPIN
2E+10	308	29.64	29.90	-0.25	+	Der f 25	Yes	AGC56216		HSP70 s.f.
1E+11	741	31.15	32.33	-1.18	+	Der f 26	Yes	AIO08852		cyclophilin_ABH like
1E+10	421	30.40	25.45	4.95	+	Der f 27	Yes	AIO08851		Euk_Ferritin
2E+09	23	26.92	26.97	-0.05	+	Der f 28	Yes	AGC56218		Euk_Ferritin
1E+10	174	29.16	28.98	0.18	+	Der f 29	Yes	AAP35065	Number of potential isoforms with a higher identification success! Renamed from Der f 261. The header is Der f Mal f 6 allergen. Low-abundance ferritin isoform. An isoform of AGC56219 with a higher identification success.	ADF_cofilin like; ADF_gelsolin s.f.
4E+08	17	NaN	24.44	NaN	+	Der f 30	Yes	AGC56219		pyrophosphatase s.f.
5E+09	132	27.09	27.91	-0.82	+	Der f 30	No	*		PTZ00335
						(isoform)				TIGR00004
3E+10	243	29.66	30.47	-0.81	+	Der f 31	Yes	AIO08870		ML s.f.
5E+09	292	28.05	27.81	0.24	+	Der f 32	Yes	AIO08849	(continued on next page)	
2E+09	191	27.09	26.33	0.76	+	Der f 33	Yes	AIO08861		
2E+09	109	25.41	25.05	0.36	+	Der f 34	Yes	BAV90601		
4E+09	75	27.94	26.98	0.96	+	Der f 35	Yes	BAX34757		

Table 1 (continued)

Intensity	MS/MS count	DF_f	DF_m	DF_f - DF_m	T-test	Allergen designation	WHO/IUIS	Representative protein	Note	Conserved domains (CCDs)
1E+10	224	29.35	28.90	0.45	+	Der f 36	Yes	AT108931		C2
1E+09	106	25.7	25.1	0.61	+	Der f 37	No	QBF67839**	Our sequence has 98.36% identity to QBF67839.	CBM_14; CBM_14
3E+09	77	27.2	27	0.23		Der f 38 (I)	No	***	Der p 38 described, we provide the sequence for Der f 38.	Spr s.f.
						Der f 39	Yes	QBF67841****	We found two hits with similar sequences.	PT200184 s.f.

Details on the protein IDs and FASTA headers of each allergen are given in Tables S1, S2, and S5. For accurate identification of the Der f 11 and 14 allergens, sequences longer (complete) than those (partial) provided in the official WHO/IUIS allergen database should be used. We also suggest that the Der f 30 allergen can actually be represented by the most abundant isoform that we identified here (while previous works linked a low-abundance isoform to this allergen). Our study provides a sequence for the yet undescribed allergen Der f 38. DF_m - *D. farinae* male; DF_f - *D. farinae* female; DF_f - DF_m - the log2 fold change difference between male and female LFQ intensities; s.f. - superfamily; + - a T-test found significant differences between male and female allergen abundances; (I) - evidence at the protein level using *D. farinae* allergens reported in this study; (?) - a protein sequence for this potential allergen group was not found (this allergen group is most likely absent in *D. farinae*). The protein sequences of these putative allergens can be found in the database used for the search (see data availability section): * - scaffold39|size422454,False_409955 ferritin or scaffold211|size71058,False_20376 ferritin; ** - scaffold105|size208323,False_49412 flocculation protein FLO11; *** - scaffold89|size238932,True_214417 peptidoglycan endopeptidase RipB-like; **** - scaffold88|size239338,True_146363 tropinin C and scaffold488|size239338,True_146033 tropinin C (see Table 2). The corresponding DNA sequences to the above proteins can be found in the following genomic scaffolds (GenBank BioProject No. PRJNA379991).

be of lower molecular weight than Der f 28. Two of the isoforms were denoted as Mag 29 allergens [48,49]; however, our protein sequences derived from the draft genome annotation are identical but longer than those entries (see Table S5, i.e., P39674.1 MAG29_DERFA vs scaffold96|size232027,True_55359; AAP35066.1 Der f Mag 29 allergen vs scaffold96|size232027,True_52373).

3.5. Troponin C aka Der f 39

We found the two following entries corresponding to troponin C (Der f 39.0101; nucleotide: MK419032): scaffold88|size239338,True_146363 and scaffold88|size239338,True_146033 (Table 2 and Table S5). Both protein hits were found to be present at higher abundances in males than in females (log2 fold change males-females was 1.22 and 1.59, respectively). The BLASTP [42] analysis showed that both these sequences had parts identical to the deduced protein sequence of Der f 39 (QBF67841.1); however, the sequences can be considered different and are both shorter than QBF67841.1.

3.6. Allergen-like proteins and other selected important proteins

We identified a number of allergen-like proteins that represent allergen isoforms or potential novel immunogens (Table 2 and Table S5). Furthermore, we selected protein hits with potential allergenic properties by domain architecture and predicted enzymatic activity, i.e., proteases, chitinases/chitin-binding proteins or those with similar action, such as lysozyme, and apolipophorin and vitellogenin/vitellogenin-like proteins. These proteins, along with their BLAST hits, identification properties, and protein IDs, are listed in Table S5.

We detected qualitative and/or quantitative sex-dependent differences in other proteins with allergenic/immunogenic potential (Table 2). For example, among proteases, three cathepsin Ls and serine protease 27 (Fig. 2) were detected only in males. Furthermore, putative bifunctional chitinase/lysozyme (Fig. 2) was detected in females only. A substantial difference was found for vitellogenins and apolipophorins, whose isoforms qualitatively or quantitatively dominated in females (Fig. 2).

3.7. Identification of misleading sequence headers of proteins

We found that some proteins are denoted as “allergen” in NCBI despite not being proven *D. farinae* allergens (Table 3). Furthermore, we identified FASTA headers of allergen sequences officially provided in the WHO/IUIS database (linked to NCBI) that do not exactly identify allergens of *D. farinae* (Table 4).

4. Discussion

We made publicly available the raw data that obtained here by 16 consecutive nLC-MS/MS runs. These data were evaluated using the LFQ MaxQuant algorithms [38] through our constructed database of the Blast2GO [50] annotated six-frame translated *D. farinae* draft genome [18], which we enriched for the current species-specific NCBI nr sequences. Using this approach, we identified a very large number (6297) of protein hits in both sexes and, consequently, were able to easily recognize the previously and newly identified proteins. By contrast, a recent study that claimed to extend the *D. farinae* allergome identified 326 proteins using nLC-MS/MS and a transcriptome database [16]. Thus, our work represents a very substantial improvement, providing a large protein catalog for *D. farinae*.

A separate proteomic analysis of males and females facilitated proteogenomic annotation, illustrating the number of proteins detected in only males or females. In addition, the separate analyses of males and females enabled comparison of the abundances of proteins, including allergens, between the sexes. The large differences between sexes are illustrated in the heatmaps and volcano plots, both created from the

Table 2
Selected allergen-like proteins and other proteins with potential allergenic properties. These protein hits, along with their blast hits, identification properties, and full list of protein IDs, are listed in Table S5.

Intensity	MS/MS count	DF_f	DF_m	DF_f-DF_m	T-test	Designation	FASTA headers	Conserved domains (CDs)
3E+08	11	NaN	24.18	NaN		Der f 1-like	scaffold9 size717327_False_545116 Group 1 mite allergen-like protein (Cysteine protease)	none
4E+09	76	24.41	27.69	-3.29	+	Der f 1-like	scaffold9 size717327_False_543914 Group 1 mite allergen-like protein (Cysteine protease)	Peptidase C1 s.f.
5E+09	75	24.84	28.24	-3.40	+	Der f 1-like	scaffold368 size33359_True_18241 Group 1 mite allergen-like protein (Cysteine protease);scaffold195 size88217_True_9479 Group 1 mite allergen-like protein (Cysteine protease);scaffold103 size218184_False_4443 Group 1 mite allergen-like protein (Cysteine protease)	Peptidase C1 s.f.
1E+09	19	23.04	26.21	-3.17		Der f 1-like	scaffold9 size717327_False_544650 Group 1 mite allergen-like protein (Cysteine protease)	none
3E+09	76	25.52	27.19	-1.68	+	Der f 2-like	scaffold421 size522413_False_25461 group 2 allergen Sui p 2	ML s.f.
8E+09	234	29.05	28.29	0.76	+	Der f 7-like	scaffold4431 size30458_True_3073 mite allergen Lep d 7-like;scaffold196 size87070_False_8408 mite allergen Lep d 7-like;scaffold149 size199602_False_196431 mite allergen Lep d 7-like;gi 37958151 gb AAP35070.1 Der f 7 allergen [Dermatophagoides farinae]	Gnp7_allergen s.f.
3E+09	105	27.17	26.90	0.27	+	Der f 7-like	scaffold4827 size22995_True_3908 Der f 7 allergen;scaffold195 size88217_True_31809 Der f 7 allergen	Gnp7_allergen s.f.
1E+10	267	29.45	28.34	1.11	+	Der f 8-like	gi 685432826 gb AO08867.1 Der f 8 allergen [Dermatophagoides farinae];gi 37958171 gb AAP35080.1 Der f 8 allergen, partial [Dermatophagoides farinae];scaffold91 size236664_True_211140 Der f 8 allergen;scaffold1082 size18999_False_15702 Der f 8 allergen	GST_C_family s.f.; GST_N_Mu (Thioredoxin_like s.f.)
2E+09	105	25.87	26.31	-0.44	+	Der f 8-like	scaffold146 size174271_True_55950 Group 8 mite allergen-like protein (mu class glutathione transferase)	GST_C_family s.f.; GST_N_Mu (Thioredoxin_like s.f.)
4E+09	150	26.98	27.61	-0.63	+	Der f 13-like	gi 37958159 gb AAP35074.1 Der f 13 allergen [Dermatophagoides farinae];scaffold923 size21705_True_7474 Group 13 mite allergen-like protein (Lipocalin)	Lipocalin s.f.
1E+09	69	26.05	25.00	1.05	+	Der f 13-like	scaffold21 size522413_False_177253 Group 13 mite allergen-like protein (Lipocalin)	none
2E+10	303	29.22	28.93	0.30		Der f 14-like (isoform)	scaffold218 size67746_False_45175 Group 14 mite allergen-like protein (Apolipophorin-like)	none
3E+09	124	28.83	NaN			Der f 14-like	scaffold168 size125973_False_51957 Group 14 mite allergen-like protein (Apolipophorin-like)	Vitellogenin_N s.f.
1E+10	158	28.88	27.98	0.90		Der f 14-like	scaffold218 size67746_False_47827 Group 14 mite allergen-like protein (Apolipophorin-like)	Vitellogenin_N; LPD N
6E+09	529	29.87	NaN			Der f 14-like	scaffold440 size417961_False_203038 Apolipophorin-like protein	Vitellogenin_N s.f.; DUF1943
2E+09	320	28.26	NaN			Der f 14-like	scaffold440 size417961_False_198455 Apolipophorin-like protein	none
1E+09	64	27.65	NaN			Der f 14-like	scaffold168 size125973_False_51293 Apolipophorin-like protein	Vitellogenin_N s.f.
3E+10	661	32.15	23.62	8.53	+	Der f 14-like	scaffold138 size173686_True_144872 serpin SMSB3a variant	Vitellogenin_N s.f.; DUF1943
6E+09	253	26.52	28.22	-1.69	+	Serpin (Der f 27-like)	scaffold1 size1740722_False_942723 serpin SMSB3a variant	SERPIN
3E+10	460	29.79	30.33	-0.54	+	Serpin (Der f 27-like)	scaffold496 size232027_True_55359 heat shock protein 70;gi 729970 sp P39674.1 MAG29 DERA	SERPIN
7E+09	188	28.36	28.05	0.31	+	Der f 28-like (MAG29)	RecName: Full = Allergen MAG29;gi 666007 dbj BAA04556.1 Mag29, partial [Dermatophagoides farinae]	HSP70 s.f.
1E+09	103	26.04	25.70	0.33		Der f 28-like (MAG29)	scaffold96 size232027_True_52373 heat shock protein 70;gi 37958143 gb AAP35066.1 Der f Mag 29 allergen [Dermatophagoides farinae]	HSP70 s.f.
5E+08	97	25.11	24.33	0.78	+	Der f 28-like	scaffold499 size20630_False_7035 hypoxia up-regulated protein 1-like;scaffold1227 size16440_False_15711 hypoxia up-regulated protein 1-like;scaffold118 size237414_False_1084	HSP70 s.f.
1E+09	32	26.25	25.94	0.31		Der f 28-like	scaffold59 size318664_True_154978 heat shock protein 70 kDa	HSP70 s.f.
4E+07	24	22.56	21.45	1.11		Der f 28-like	scaffold499 size20630_False_4899 Sar s 28 (heat shock protein 70-like protein 7);scaffold350 size34216_False_31600 Sar s 28 (heat shock protein 70-like protein 7);scaffold220 size67183_True_15627 Sar s 28 (heat shock protein 70-like protein 7)	none
5E+09	132	27.09	27.91	-0.82	+	Der f 30 (isoform)	scaffold39 size42454_False_409955 ferritin;scaffold211 size71058_False_20376 ferritin	Euk_Ferritin
2E+09	94	26.52	25.53	0.99	+	Der f 30-like	scaffold424 size505992_False_78980 ferritin	Euk_Ferritin
2E+08	25	22.87	23.45	-0.58		Der f 33-like	scaffold21 size2522413_True_182318 Der p 36 allergen	PTZ00335
2E+08	42	22.95	23.38	-0.42		Der f 36-like	scaffold488 size239338_True_146033 troponin C	none
7E+09	107	26.87	28.46	-1.59	+	Troponin C/Der f 39?!	scaffold488 size239338_True_146033 troponin C	PTZ00184 s.f.
2E+10	215	28.64	29.87	-1.22	+	Troponin C/Der f 39?!	gi 37958173 gb AAP35081.1 Der f Alt a 10 allergen [Dermatophagoides farinae]	PTZ00184 s.f.
3E+10	515	30.50	30.55	-0.05		Der f Alt a1 allergen	gi 13774176 gb AAK38775.1 cysteine proteinase, partial [Dermatophagoides farinae]	ALDH_F1AB_F2_RALDH1
6E+08	37	24.03	25.04	-1.00	+	Cathepsin F-like	scaffold368 size33359_True_18656 cathepsin L1-like;scaffold195 size88217_True_9894 cathepsin L1-like;scaffold103 size218184_False_4030 cathepsin L1-like	Peptidase C1 s.f.
2E+09	27	NaN	26.70	NaN		Cathepsin L	scaffold195 size88217_True_14968 cathepsin L1-like;scaffold4368 size33359_True_22006 cathepsin L1-like	Peptidase C1 s.f.
6E+08	40	NaN	25.47	NaN		Cathepsin L	scaffold489 size238932_True_235616 cathepsin L1-like;scaffold44 size389383_True_9516 cathepsin L1-like	Peptidase C1A
2E+08	30	NaN	23.50	NaN		Cathepsin L	scaffold159 size132448_True_85030 cathepsin B	Peptidase C1A_CathepsinB
6E+09	227	28.24	27.82	0.41	+	Cathepsin B		

(continued on next page)

Table 2 (continued)

Intensity	MS/MS count	DF_f	DF_m	DF_f-DF_m	T-test	Designation	FASTA headers	Conserved domains (CCDs)
4E+09	115	27.36	27.25	0.10		Cathepsin L	scaffold12 size1149099,False,470038,cathepsin L1-like	Peptidase_C1 s.f.
4E+09	170	28.23	27.18	1.04	+	Cathepsin L	scaffold523 size28000,False,22097,cathepsin L,scaffold21 size522413,True,343997,cathepsin L	Peptidase_C1; Inhibitor_129
4E+09	111	28.07	26.84	1.23	+	Cathepsin C	scaffold97 size230796,False,129670,cathepsin C	CathepsinC exc; Propeptide_C1
4E+09	170	28.38	26.97	1.42	+	Dipeptidyl peptidase 1	scaffold97 size230796,False,129670,cathepsin C	Peptidase_C1 s.f.
7E+09	398	28.47	28.11	0.36	+	Dipeptidyl peptidase 3	scaffold21 size522413,True,255033,Dipeptidyl peptidase 3	Peptidase_M_49
3E+09	175	27.47	26.68	0.79	+	Cathepsin B	scaffold159 size132448,True,82012,cathepsin B	Peptidase_C1A, CathepsinB
2E+09	101	26.99	26.22	0.77	+	Cathepsin L	scaffold481 size44186,False,35098,cathepsin L,scaffold17 size560895,False,230222,cathepsin L	Peptidase_C1
2E+09	146	26.12	26.18	-0.06		Cathepsin B	scaffold138 size173686,False,78452,cathepsin B	Peptidase_C1;
5E+08	44	25.52	22.51	3.02	+	Cathepsin L	scaffold481 size44186,False,35734,cathepsin L,scaffold17 size560895,False,230858,cathepsin L	Peptidase_C1A, CathepsinB
9E+08	114	26.65	22.44	4.21	+	Cathepsin L	scaffold59 size318664,False,147964,cathepsin L	Peptidase_C1A; COG4870 s.f.
3E+09	216	27.86	26.63	1.23	+	Cysteine protease	scaffold743 size24137,False,20071,digestive cysteine proteinase 2-like;scaffold8 size760192,True,602658	Peptidase_C1A; Inhibitor_129
7E+07	19	22.80	22.51	0.29		Cysteine protease	digestive cysteine proteinase 2-like	Peptidase_C1 s.f.
2E+08	42	NaN	23.69	NaN		Serine protease	scaffold43 size445179,True,225812,serine protease 27	Tryp_SpC
2E+09	153	26.63	25.86	0.77	+	Serine protease	scaffold39 size422454,False,258151,putative serine protease K12H4.7;scaffold47 size29079,False,26296	Peptidase_S28 (Abhydrolase s.f.)
1E+09	93	26.44	25.41	1.03	+	Serine protease	putative serine protease K12H4.7	Peptidase_S28 (Abhydrolase s.f.)
5E+08	41	24.80	24.67	0.12		Serine carboxypeptidase	scaffold478 size29079,False,25211,putative serine protease K12H4.7;scaffold39 size422454,False,257069	Peptidase_S10 s.f.
1E+10	309	28.91	29.04	-0.13		Aspartic protease	venom serine carboxypeptidase;scaffold1047 size19947,False,10024,venom serine carboxypeptidase	Pepsin_retropepsin_like s.f.
1E+09	73	25.81	25.38	0.43	+	Aspartic protease	scaffold35 size434843,False,346669,lysosomal aspartic protease-like	Pepsin_retropepsin_like s.f.
6E+09	202	28.42	27.65	0.77	+	chitin binding protein	scaffold35 size434843,False,347008,lysosomal aspartic protease-like	CE4_CDA_like_2
2E+08	22	24.27	NaN	NaN		Chitinase/lysozyme	scaffold682 size25061,True,2883,peritrophic membrane chitin binding protein, putative	GH18, chitinase-like s.f.
1E+09	64	26.29	25.21	1.09	+	Chitinase	putative;scaffold134 size177260,False,18931,peritrophic membrane chitin binding protein, putative	GH18, chitinase-like s.f.
2E+09	67	26.67	25.74	0.93	+	Chitinase	scaffold38 size423637,True,361700,putative bifunctional chitinase/lysozyme	GH18, chitinase-like s.f.
2E+09	158	26.74	25.95	0.79	+	Chitinase	scaffold614 size26335,False,24253,cathepsin L,scaffold10 size739037,False,130434	GH18, chitinase-like s.f.
3E+08	43	23.81	23.74	0.07		Chitinase/lysozyme	chitinase-3-like protein 1	GH18, chitinase-like s.f.; ChBD2
1E+08	28	22.54	22.62	-0.08		Chitinase	scaffold10 size739037,False,129991,cathepsin L,scaffold10 size739037,False,130917	GH18, chitinase-like s.f.; ChBD2
9E+07	15	22.55	23.45	-0.90		Der f 15-like	chitinase-3-like protein 1	(CBM_14 s.f.)
1E+09	51	26.18	25.59	0.59	+	Lysozyme	scaffold481 size248248,True,176749,putative bifunctional chitinase/lysozyme	GH18, chitinase-like s.f.; ChBD2
5E+09	183	29.65	NaN	NaN		Vitellogenin	scaffold261 size46442,True,9686,putative lysozyme;scaffold886 size22254,False,17973,putative	lysozyme-like s.f.
2E+09	34	28.13	NaN	NaN		Vitellogenin	lysozyme;scaffold84 size245478,False,9378,putative lysozyme	Vitellogenin_N s.f.
8E+09	133	30.19	23.68	6.51		Vitellogenin	scaffold490 size237115,True,117857,cathepsin L	Vitellogenin_N s.f.
4E+10	766	32.57	19.63	12.94	+	Vitellogenin	scaffold490 size237115,True,122408,cathepsin L,putative	VWD s.f.
						Vitellogenin	scaffold490 size237115,True,118681,cathepsin L,putative	Vitellogenin_N s.f.; DUF1943 s.f.

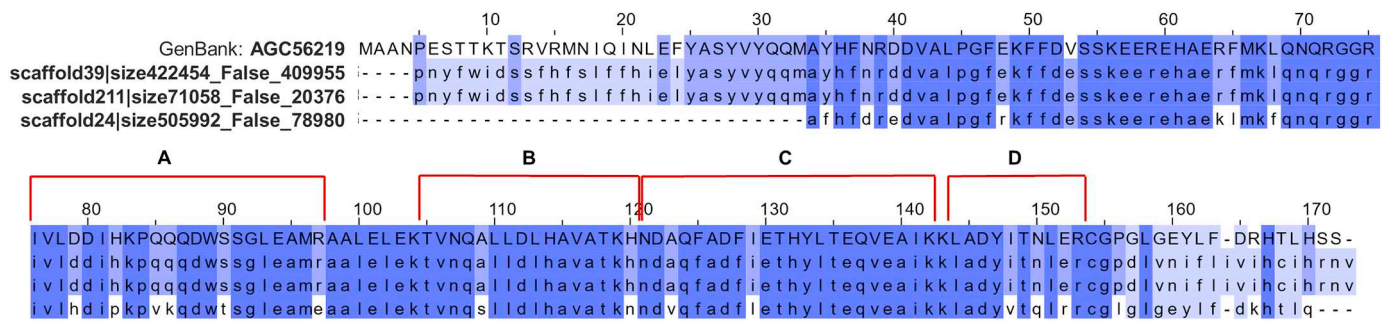


Fig. 3. The alignment of ferritins. The official WHO/IUIS sequence of Der f 30 (AGC26219), the Der f 30 (isoform) (scaffold39|size422454_False_409955 and scaffold211|size71058_False_20376), and a short ferritin variant (Der f 30-like) (scaffold24|size505992_False_78980) are aligned. The alignment clearly shows that peptides (A–D) are the same in Der f 30 and Der f 30 (isoform) discovered here; therefore, the same peptides can identify different isoforms at different abundances. The third, shorter ferritin variant (Der f 30-like) identified in all our analyses contains different peptides for identification. A–D – the four peptides that identified ferritin using MSP in previous studies [9,11,12].

dataset normalized through Z-scores or by replacement of the missing values by zeroes. It appears that the two methods of data evaluation have some advantages for the data of this study, and the replacement of the missing values by zeroes appears to be beneficial when qualitative differences are observed (see [51,52]).

We reviewed and verified the identification of all WHO/IUIS-approved allergens, the sequences of which are linked to the NCBI repository. We show, that one of the most questionable aspects appears to be the presence of partial/incomplete sequences in databases. Use of these incomplete sequences for data evaluation can negatively affect the accuracy of proteomic identification and especially quantification. In addition, some entries contain misleading headers, such as the word “allergen” provided in the header despite the entry not being a WHO/IUIS-approved allergen. In addition, in some headers, the allergens are named differently than the official allergen. Thus, it would be useful if the headers in the publicly available database are renamed according to the official allergen. Such a change would prevent these misleading entries, which provide incorrect allergen identification, from being used when using the publicly available database as the source for data evaluation. Currently, for proper identification, a manually curated database should be used, or the data need to be individually inspected.

We also completed the allergen list of *D. farinae* by including the yet-uncharacterized allergens with sequences homologous to confirmed allergens of other mites; these allergens include Der f 9 and 38. In addition, our comprehensive proteomic analysis enabled the identification of new candidate immunogenic proteins and/or allergen isoforms in *D. farinae*. We found significant differences in the expression of allergens between the sexes and this result has potentially practical significance as a factor affecting the differences in the allergen levels observed in commercially produced allergen extracts [21–28,53].

We demonstrated that the MSP approach, consistent with the

previous suggestion [6], is useful for the standardization of allergenic extracts. The challenge for the future is to apply our results to a targeted analytical method, such as selected monitoring reaction (SRM) analysis, to detect proteins of interest. However, it is crucial to use correct/appropriate sequences to construct a reference database for these methods.

4.1. The importance of biological material purity and sample analysis

Dermatophagoides farinae, similar to other domestic mite species, is microscopic [20]; therefore, obtaining pure representative material is difficult. In previous studies utilizing proteomic analyses, mites were mostly obtained by flotation (salt solution) combined with sieving [15,16], or the procedure of mite harvesting was not described in detail [9–11]. Transcriptomic and genomic sequencing also involved mites collected by flotation in salt solutions [16–18], although in the latter study, this was the case only for the PacBio data, while Illumina sequencing was based on hand-picked female mites [18]. From a methodological standpoint, requirements for DNA data collection are not as strict as those for proteomics. Previously, we showed that it is possible to collect highly pure live *D. farinae* mites by contactless processing [12,54]. In this study, we collected pools of 1000 individually collected, verified and sex-sorted individuals per biological replicate, and we ensured that each of the mites included in the analysis was alive, free of contaminants and of a specific size at the time of collection. The separate analyses of males and females enabled us to compare the proteome and allergome between the mite sexes.

4.2. Completion of the list of *D. farinae* allergens

The WHO/IUIS database currently lists 39 allergen groups present

Table 3

Misleading allergen FASTA headers. Using these sequences can lead to incorrect allergen identification or give a false impression that the allergen was identified correctly, although its sequence actually does not represent a proven allergen.

Official in WHO/IUIS	Comment in WHO/IUIS	Confusing FASTA header	Reason
No	No	gb AAP35070.1 Der f 7 allergen [<i>Dermatophagoides farinae</i>]	Only 18.22% identity with Der f 7 (gb AAB35977.1)
No	No	gb AAP35074.1 Der f 13 allergen [<i>Dermatophagoides farinae</i>]	Only 50.78% identity with Der f 13 (gb AAP35078.1)
No	No	gb AAU34048.1 Der f 5.02 allergen [<i>Dermatophagoides farinae</i>]	The sequence relates to Der f 21 allergen.
No	No	sp L7UZ85.1 ACTN_DERFA RecName: Full = Alpha-actinin; AltName: Full = F-actin cross-linking protein; AltName: Allergen = Der f 24	The sequence does not characterize Der f 24. However, α -actinin was tested for allergy [84].
No	No	gb AIO08856.1 Der f 26 allergen [<i>Dermatophagoides farinae</i>]	The sequence relates to the Der f 29 allergen.
Yes	No	gb AAP35065.1 Der f Mal f 6 allergen [<i>Dermatophagoides farinae</i>]	Der f 29 allergen (100% identity to above).
No	Yes (Der f 29)	gb AIO08866.1 Der f 29 allergen [<i>Dermatophagoides farinae</i>]	It is not Der f 29, the sequence relates to profilin.
Yes	Yes (Der f 26)	gb AIO08852.1 Der f 30 allergen [<i>Dermatophagoides farinae</i>]	The sequence relates to the Der f 26 allergen.
Yes	No	dbj BAX34757.1 group II allergen [<i>Dermatophagoides farinae</i>]	It is the Der f 35 allergen.
No	No	gb AAP35081.1 Der f Alt a 10 allergen [<i>Dermatophagoides farinae</i>]	Not a proven allergen.

Table 4

FASTA headers of allergen sequences officially provided in the WHO/IUIS database that do not exactly identify allergens of *D. farinae*. For the majority of sequences, allergen names are not provided in these FASTA headers. The allergens Der f 26, 29 and 35 were misidentified in the FASTA headers (see also Table 3).

Allergen	FASTA header
Der f 8	gi 442565870 gb AGC56215.1 glutathione S-transferase, partial [<i>Dermatophagoides farinae</i>]
Der f 10	gi 1359436 dbj BAA04557.1 Mag44, partial [<i>Dermatophagoides farinae</i>]
Der f 11	gi 13785807 gb AAK39511.1 paramyosin-like allergen, partial [<i>Dermatophagoides farinae</i>]
Der f 14	gi 1545803 dbj BAA04558.1 Mag3, partial [<i>Dermatophagoides farinae</i>]
Der f 15	gi 5815436 gb AAD52672.1 AF178772_1 98kDa HDM allergen [<i>Dermatophagoides farinae</i>]
Der f 20.0201	gi 156938897 gb ABU97470.1 arginine kinase [<i>Dermatophagoides farinae</i>]
Der f 22	gi 109629736 gb ABG35122.1 allergen [<i>Dermatophagoides farinae</i>]
Der f 24	gi 477541860 gb AGI78542.1 ubiquinol-cytochrome c reductase binding protein-like protein [<i>Dermatophagoides farinae</i>]
Der f 25.0101	gi 442565872 gb AGC56216.1 triose-phosphate isomerase [<i>Dermatophagoides farinae</i>]
Der f 26	gi 685432796 gb AIO08852.1 Der f 30 allergen [<i>Dermatophagoides farinae</i>]
Der f 28.0101	gi 442565876 gb AGC56218.1 heat shock protein 70 [<i>Dermatophagoides farinae</i>]
Der f 29	gi 685432804 gb AIO08856.1 Der f 26 allergen [<i>Dermatophagoides farinae</i>]
Der f 30	gi 442565878 gb AGC56219.1 ferritin [<i>Dermatophagoides farinae</i>]
Der f 34	gi 1098871171 dbj BAV90601.1 enamine/imine deaminase [<i>Dermatophagoides farinae</i>]
Der f 35	gi 1187443130 dbj BAX34757.1 group II allergen [<i>Dermatophagoides farinae</i>]

in astigmatid mites; however, Der f 9, 12, 19, and 38 have not yet been described/approved [3]. Furthermore, the Der f 17 allergen, a putative EF-hand calcium-binding protein, was cited in the original submissions as an allergen [4,5], but the corresponding sequence has never been deposited in a public database (see also a note on the WHO/IUIS site [3]). Therefore, we referred to allergen group 17 of the mite as an ‘empty’ group, although our dataset included EF-hand calcium-binding proteins, one of which may be that referred to as Der f 17 [4,5]. Thus, our proteomics approach successfully identified 34 allergens, and we found reliable matches for the yet undescribed allergens Der f 9 and 38. However, we could not find protein hits corresponding to the group 12 and 19 allergens described in only *B. tropicalis*. Thus, we suggest that these two groups of allergens are not expressed in *Dermatophagoides*, a conclusion that is also supported by the absence of these allergen groups in *D. pteronyssinus*. Blo t 12 (AAA78904.1), including isoallergens, is similar to a number of the mite allergens produced only as recombinants [55–57]. A protein similar to Blo t 12 was also found in *Lepidoglyphus destructor* (AAQ55550.1; query coverage 100% and identity 93.06% to AAA78904.1). Therefore, a possible explanation for the fact that we did not find a homolog to Blo t 12 in *D. farinae* is that *B. tropicalis* is phylogenetically distant from *Dermatophagoides*. Proteomic identification is an important piece of evidence supporting the existence of a protein at the protein level. Thus, it should also be of interest that by using proteogenomics, we identified *D. farinae* proteins with some similarity to the Blo t 12 domain in terms of the domain architecture. In the case of Blo t 19, the domain architecture (AHG97583, Antimicrobial25 s.f. domain) [43,44] indicates its high similarity to antimicrobial peptides found in nematodes. Thus, in future research, the origin of Blo t 19 should be verified.

Here, we provided evidence for Der f 38 because we found sequences homologous to the corresponding approved allergen Der p 38 that was recently deposited in the WHO/IUIS database [3]. However, it will be necessary to confirm the immunogenicity of Der f 38. Compared to Der f 38, the sequences related to Der f 9 were available in NCBI at the time of our analysis. It is a similar situation for Der f 5, which was recently added to the WHO/IUIS list of allergens [3]. It is quite surprising that Der f 5 and 9 have not earlier or yet, respectively, been confirmed/approved as allergens since their homologs in *D. pteronyssinus*, i.e., Der p 5 and 9, respectively, were described a long time ago [58–60]. Note that Der f 5 was claimed by Le Mao et al. [61] according to the location of spots in 2D-E. Here, we found the following: (i) both Der f 5 and 9 were present in males and females, (ii) Der f 5 was more abundant than Der f 9, and (iii) there were no significant sex differences in Der f 9 abundance, while Der f 5 was more abundant in females. The sequence record AAP35067.1/2 for Der f 9 was provided by a research group [62] that also provided several sequences for Der f

5 (i.e., ABO84970.1, ABO84974.1, and ABO84975.1). Another research group provided the Der f 5 allergen sequence AIO08857.1. Furthermore, Bordas-Le Floch et al. [16] recently retrieved the partial counterpart sequences that matched Der f 5 (ATI08933.1) and Der f 9 (ATI08934.1) (see Table S3 in [16]). However, we found that in addition to the sequences that correctly denoted Der f 5, NCBI included the record AAX34048.1 Der f 5.02 allergen [*Dermatophagoides farinae*], which in fact relates to the group 21 allergen. Thus, the header of the Der f 5.02 allergen is misleading and should be renamed accordingly to avoid confusion in the future. We suggest that this pitfall could be caused by the fact that both groups 5 and 21 have the Blo-t-5 domain. Thus, multiple sequences that characterize the same protein/allergen exist in NCBI, but some of them can lead to incorrect identification (see also the last section of the Discussion).

In our previous study, we suggested that the group 9 allergen constitutes an important difference between *D. farinae* and *D. pteronyssinus* fecal proteomes because we failed to detect it in *D. farinae*, while *D. pteronyssinus* Der p 9 was detected in the acidic isoelectric point (pI) of 2D-E [34]. Subsequently, Cui et al. [63] produced a recombinant Der f 9 and tested it for allergenicity [63]. Thus, considering all of these results together, it is important that we detected/confirmed Der f 9 at the protein level here, although it seems that this protease allergen is less important from a quantitative point of view than its homolog Der p 9 in *D. pteronyssinus* [34].

4.3. New candidate allergens and other important proteins of *D. farinae*

The 34 WHO/IUIS currently approved allergens, together with the likely allergens Der f 9 and 38, make *D. farinae* a very diverse source of allergens; however, there are a number of additional proteins of potential immunogenic importance in this species. Therefore, we selected proteins with primary structures similar to those of known allergens, denoted here as allergen-like proteins. For a subset of these proteins, one can at least suggest that they are allergen isoforms. We then selected proteins that are candidate allergens based on their function, domain architecture and/or enzyme activity. In general, for each allergen or a newly considered allergen compound, it is important to consider its abundance in the environment, especially whether the protein is defecated/excreted and thereby quantitatively dominates other proteins present only in the mite body [12,34,64]. However, the abundance of the allergen in the mite body is also important, especially if mite bodies are used for antigen preparation.

In general, one of the most potent groups of proteins provoking allergies are mite proteases, which can have a dual allergen function. Proteases such as Der f 1 can affect epidermal barrier structure and function, stimulate protease-activated receptor 2 (PAR-2) [65–69] and/

or inactivate PAR-1 [69]. Although the very important protease allergen groups 1, 3, 6 and 9 are already known, additional proteases can also affect both the epidermal barrier and PAR-1/2. Here, we identified a number of additional serine and cysteine proteases that may be potent allergens directly or at least due to their mechanism of action. Furthermore, the synergic or cumulative effect of the wide spectra of proteases can contribute to allergy manifestation.

The next group of proteins that can generally be considered to provoke allergies are chitinases or proteins with a similar mode of action. Diverse protozoan and metazoan pathogens use chitinases through infection and thereby induce diseases in humans and animals [70]. In HDMs, there are two recognized chitinase-like allergen groups, but the allergenic properties of groups 15 and 18 have been confirmed for dogs only [71,72]. Both Der f 15 and Der f 18 share the GH18_chitinase-like s.f. domain and have a second ChtBD2 and CBM_14 domain, respectively. Moreover, among the HDM allergens, group 23 allergens have a ChtBD2 domain, and a more recently described group 37 allergen has two CBM_14 domains. The CBM_14 (Chitin-binding Peritrophin-A) domain also characterizes the Blo t 12 allergen; therefore, among the proteins of *Dermatophagoides* with a similar domain architecture, one can also find a Blo t 12 homolog/analog. Overall, proteins with the chitin-binding domain and/or chitinase activity can be considered potential allergen analogs to those that have already been described. Finally, we previously identified exochitinases of the bacterium *Bacillus cereus* in fecal extracts from the mite *Tyrophagus putrescentiae* [64]. Thus, the bacterial exochitinases, if present within the mites, can contribute to the allergy directly by their enzyme activity [64], which is in principle similar to the action of the proteases.

Another member of the canonical mite allergens is Der f 14, which is an apolipoprotein. Der f 14 has the LPD_N (Vitellogenin_N s.f.) and DUF1943 s.f. domains. Furthermore, vitellogenin with the Vitellogenin_N, DUF1943, and VWD s.f. domains is a proven allergen, specifically, Api m 12 and Ves v 6 of *Apis mellifera* and *Vespula vulgaris*, respectively [73]. Thus, proteins with apolipoprotein/vitellogenin-like domain architecture are of potential immunogenic interest in mites. Finally, the results showed that an abundance of vitellogenins exhibit a large difference between the sexes, dominating in females of *D. farinae*.

Another group of allergens of importance that we focus on here are serpins. Serpins are highly conserved proteins with the principal function of protease inhibitor activity [74]. Their immunogenic potential was first characterized in *Schistosoma haematobium* [75], long before the Der f 27 allergen was characterized in *D. farinae* [76]. Interestingly, the sex-dependent difference in Der f 27 expression found in our study mirrored that observed for a serpin in the tick *Amblyomma americanum* [74]. Overall, our results suggest that Der f 27 may have a higher allergy potential than previously suggested. However, we identified two other serpin isoforms similar to Der f 27 as SMSB3a variants, but they shared low similarity with Der f 27. Both of these isoforms were more abundant in males, but the difference was not as large as that for Der f 27. Furthermore, a serpin B10-like protein was identified in *D. farinae*, and it also may be of interest.

4.4. Explanation of the low identification success for some allergens

The abundance of an immunogen in the environment affects its allergy potential [77]. We observed a low identification success for the allergens Der f 11, 14, 23, 28 and 30 when the official WHO/IUIS sequences were used for identification. Therefore, we investigated whether these allergens are truly at a low abundance in mite bodies or whether there is some other explanation for the low identification success. Our investigation provided support that Der f 23 is a low-abundance allergen in *D. farinae* mite bodies, which can be explained by it being localized to the peritrophic membrane. In contrast, the mite allergen group 23 is excreted with feces, as found for Der p 23 in *D. pteronyssinus* [78]. The low abundance of Der f 23 in the mite bodies determined here by label-free proteomics is consistent with that

reported for its homolog Der p 23 in *D. pteronyssinus* [79]. Moreover, for the two allergens, Der f 11 and Der f 14, we found that the identification success was higher when a sequence longer than the partial official WHO/IUIS sequence was used to identify the protein hit. Thus, partial sequences inaccurately characterize the allergen and should be replaced in the database, at least for the use in MSP identification.

A more complicated investigation was necessary for the Der f 30 allergen (i.e., ferritin), for which we found that *D. farinae* expresses more isoforms, which were reported here as three separate protein hits. We revealed that the isoform denoted here as Der f 30 (isoform) provides the same tryptic digests as the official WHO/IUIS sequence of Der f 30 (AGC56219.1). Furthermore, the results of our investigation suggest that the ferritin/Der f 30 identified in earlier analyses employing MSP [9,11,12,14] may actually have been Der f 30 (isoform), the sequence of which was not known/available at the time. This suggestion is based on the fact that according to our results, Der f 30 appears to be a very-low-abundance protein/allergen in mite bodies.

Finally, our results showed that Der f 28.0101 (AGC56218.1) and Der f 28.0201 (AIO08848.1) are of low abundance, and the latter Der f 28 isoform we found only in the trace dataset. Der f 28 is a heat shock protein 70, and this group of proteins is usually very abundant in organisms; however, the low abundance of the two WHO/IUIS approved isoforms of Der f 28 allergen [3] decreases their importance from a quantitative point of view. In our dataset, we found a variety of Der f 28-like protein hits, including those denoted as Mag 29 allergens [48,49], and they were identified with higher identification success despite being characterized by shorter sequences than Der f 28.0101 and Der f 28.0201. Future studies with detailed sequence identification analyses will be necessary to fully explain the quantitative ambiguity of Der f 28.

4.5. Sex-dependent allergen expression and its practical significance

There was a large difference between the male and female proteomes. Furthermore, there were series of protein hits detected in only males or females. Therefore, it is important to carefully consider the significance of this difference, and the threshold for assessing substantial importance can be approximately ± 1 on a log2 fold scale. The key qualitative differences were revealed in the volcano plot analysis, especially when missing values were replaced by zeroes. Sex-related differences were also observed at the quantitative level (protein was detected in both sexes). There were sex-dependent LFQ differences in various allergens, reaching an approximate difference of ± 1 on a log2 fold scale; however, there was one allergen, Der f 27 (serpin; see [76]), whose difference between females and males was the most substantial (approximately 5 on the log2 fold scale). Moreover, the overall intensities or MS/MS counts suggest that Der f 27 is a highly abundant protein in mite bodies, especially in females. Other potential immunogens of *D. farinae* discussed above exhibited important differences between sexes.

Considering the fact that we demonstrated sex-biased differences in allergens and that variation in the sex ratio with culture age was previously observed [31], the allergen levels in differently aged cultures should also impact the appropriate proportion. Thus, the differences in allergens between males and females can at least partially explain the observed variations in the composition of (commercial) mite allergen extracts [21–28,53]. However, there are additional factors that can greatly influence the production of proteins/allergens. Therefore, it is important to consider the role of the mite strain and associated bacterial microbiome [80] and cultivation method, including differences in the diet used [53,81,82]. Further work will be needed to demonstrate the level at which the changes in the sex ratio can affect the allergen composition in allergen extracts.

The sex differences in proteomes could also be notably influenced by proteins from the unlaidd eggs of females. To prove this suggestion and show its significance, it will be necessary in the future to compare

the proteomes of eggs and mature females with unlaidd eggs. In the future, the impact of the composition of different life stages in the mite culture on the total proteome and allergome of the culture should also be studied. Although our analysis can be considered comprehensive, some proteins and allergens might be missed or underestimated in the analysis because of the restriction to adult males and females.

4.6. Misleading sequence headers can lead to inaccuracies in MSP allergen identification

Recent high-throughput “omic” methods have become very effective tools in the analysis of complex data and biomarker discovery; however, such complex analyses contain pitfalls that can result in inaccurate data interpretation. In the case of MSP, researchers usually work with large databases used to search the MS data. Furthermore, many sequences provided in public databases are not verified or are redundant. Thus, for proper interpretation, in some cases, it is appropriate to individually verify the sequences and their corresponding headers. For example, in our previous study [83], automatic evaluation wrongly identified the VP3 structural protein of deformed wing virus because the sequence in fact related to a different part of the virus [83]. Here, we identified a set of headers from NCBI entries that can easily lead to improper identification of allergens because they contain the word “allergen”, which is misleading for unproven allergens. Thus, it is important to be aware of these entries.

The sequence records of *D. farinae* that we found to be misleading need to be properly considered until they are renamed or the sequence containing the word “allergen” is verified as a true allergen approved by WHO/IUIS [3]. We stress that among the astigmatid mites, the allergen groups 5 and 21 can easily be confused due to incorrect sequence designation. For example, confusion is probable due to the sequence AAX34048.1 of *D. farinae* having the header Der f 5.02 allergen, although the sequence was actually identified as Der f 21. Furthermore, the official WHO/IUIS sequence record for Blo t 21 (AAX34047.1) has the header Blo t 5.02 allergen [*Blomia tropicalis*]. Similarly, the entry BAX34757.1 has the header group II allergen [*Dermatophagoides farinae*], but the entry in fact characterizes the allergen Der f 35.

On the WHO/IUIS site, comments are provided for some allergen sequences. In particular, a comment had been posted on 06 Nov 2014 for the Der f 26 allergen stating that “the sequence entries should be amended”, but we found that the correction has not been made in NCBI. Furthermore, for Der f 29, the WHO/IUIS notes that “NCBI and UniProt show “Der f 29” as profilin, but that entry (AIO08866.1) was not submitted to IUIS and is illegitimate”. In fact, the heading of AIO08866.1 is Der f 29 allergen, but it is not in that allergen group, which actually refers to peptidyl-prolyl cis-trans isomerase (cyclophilin) rather than profilin. It is also of importance that the entry L7UZ85.1, which is the putative mite allergen α -actinin [84], does not characterize Der f 24 as incorrectly indicated in the header. Thus, if in the future α -actinin is approved by WHO/IUIS as an allergen, it will be necessary place this protein in another group of allergens.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jprot.2019.103535>.

Authors' contributions

TE devised the study, evaluated the data in detail, and wrote the main manuscript. PT and KH performed the nLC-MS/MS analysis. PK and JH contributed with writing. All authors have read and approved the manuscript.

Funding

This study was supported by grant No. 17-12068S from the Czech Science Foundation (GACR) (<https://gacr.cz/>) and by grant No. RO0418 from the Ministry of Agriculture of the Czech Republic (<http://>

eagri.cz/). PBK was supported by the Russian Science Foundation (grant No. 19-14-00004) and the US National Science Foundation (grant No. NSF DEB-0613769).

Availability of data and materials

The accession numbers for the raw nLC-MS/MS runs reported in this paper are MassIVE MSV000083744 (doi: [10.25345/C59G9J](https://doi.org/10.25345/C59G9J)) and PXD013714. Furthermore, we provide for download the entire “combined/txt” folder of MaxQuant (v1.6.3.3) data processing and the database used for the search.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgements

We thank Julie Chalupnikova and Martin Markovic for valuable help. We also thank Barry OConnor for critical reading of the earlier draft of this manuscript. We thank Bruno Sopko for discussion of the manuscript revision.

References

- [1] T.A.E. Platts-Mills, J.A. Woodfolk, Allergens and their role in the allergic immune response, *Immunol. Rev.* 242 (1) (2011) 51–68.
- [2] T.A.E. Platts-Mills, A.L. de Weck, Dust mite allergens and asthma—a worldwide problem, *J. Allergy Clin. Immunol.* 83 (2) (1989) 416–427.
- [3] WHO/IUIS, Allergen Nomenclature, WHO/IUIS Allergen Nomenclature Subcommittee, 2019, <http://www.allergen.org>, Accessed date: 28 August 2019.
- [4] S. Kawamoto, T. Aki, M. Yamashita, A. Tategaki, T. Fujimura, S. Tsuboi, T. Katsutani, O. Suzuki, S. Shigeta, Y. Murooka, K. Ono, Toward elucidating the full spectrum of mite allergens — state of the art, *J. Biosci. Bioeng.* 94 (4) (2002) 285–298.
- [5] A. Tategaki, S. Kawamoto, T. Aki, T. Jyo, O. Suzuki, S. Shigeta, K. Ono, Newly described house dust mite allergens, *Allergy Clin. Immunol. Int. Suppl.* 1 (2000) 74–76.
- [6] J. Spirc, A. Reuter, R.L. Rabin, Mass spectrometry to complement standardization of house dust mite and other complex allergenic extracts, *Clin. Exp. Allergy* 47 (5) (2017) 604–617.
- [7] M. Tyers, M. Mann, From genomics to proteomics, *Nature* 422 (6928) (2003) 193–197.
- [8] T. Erban, K. Harant, J. Chalupnikova, F. Kocourek, J. Stara, Beyond the survival and death of the deltamethrin-threatened pollen beetle *Meligethes aeneus*: an in-depth proteomic study employing a transcriptome database, *J. Proteomics* 150 (2017) 281–289.
- [9] S. An, L. Chen, C. Long, X. Liu, X. Xu, X. Lu, M. Rong, Z. Liu, R. Lai, *Dermatophagoides farinae* allergens diversity identification by proteomics, *Mol. Cell. Proteomics* 12 (7) (2013) 1818–1828.
- [10] T. Batard, A. Hrabina, X.Z. Bi, H. Chabre, P. Lemoine, M.-N. Couret, D. Faccenda, B. Villet, P. Harzic, F. Andre, S.Y. Goh, C. Andre, F.T. Chew, P. Moingeon, Production and proteomic characterization of pharmaceutical-grade *Dermatophagoides pteronyssinus* and *Dermatophagoides farinae* extracts for allergy vaccines, *Int. Arch. Allergy Immunol.* 140 (4) (2006) 295–305.
- [11] J. Choopong, O. Reamtong, N. Sookrung, W. Seesuan, N. Indrawattana, Y. Sakolvaree, W. Chaicumpa, A. Tungtrongchitr, Proteome, allergenome, and novel allergens of house dust mite, *Dermatophagoides farinae*, *J. Proteome Res.* 15 (2) (2016) 422–430.
- [12] T. Erban, J. Hubert, Two-dimensional gel proteomic analysis of *Dermatophagoides farinae* feces, *Exp. Appl. Acarol.* 65 (1) (2015) 73–87.
- [13] T. Erban, Purification of tropomyosin, paramyosin, actin, tubulin, troponin and kinases for chemoproteomics and its application to different scientific fields, *PLoS One* 6 (8) (2011) e22860.
- [14] T. Erban, R. Klubal, Non-protease native allergens partially purified from bodies of eight domestic mites using p-aminobenzamidine ligand, *Allergol. Immunopathol. (Madr.)* 46 (3) (2018) 218–225.
- [15] T. Batard, V. Baron-Bodo, A. Martelet, M. Le Mignon, P. Lemoine, K. Jain, S. Mariano, S. Horiot, H. Chabre, C. Harwanegg, C.A. Marquette, B.P. Corgier, W.T. Soh, P. Satitsuksanoa, A. Jacquet, F.T. Chew, E. Nony, P. Moingeon, Patterns of IgE sensitization in house dust mite-allergic patients: implications for allergen immunotherapy, *Allergy* 71 (2) (2016) 220–229.
- [16] V. Borda-Le Floch, M. Le Mignon, L. Bussieres, K. Jain, A. Martelet, V. Baron-Bodo, E. Nony, L. Mascarell, P. Moingeon, A combined transcriptome and proteome analysis extends the allergome of house dust mite *Dermatophagoides* species, *PLoS One* 12 (10) (2017) e0185830.
- [17] T.-F. Chan, K.-M. Ji, A.K.-Y. Yim, X.-Y. Liu, J.-W. Zhou, R.-Q. Li, K.-Y. Yang, J. Li, M. Li, P.T.-W. Law, Y.-L. Wu, Z.-L. Cai, H. Qin, Y. Bao, R.K.-K. Leung, P.K.-S. Ng,

- J. Zou, X.-J. Zhong, P.-X. Ran, N.-S. Zhong, Z.-G. Liu, S.K.-W. Tsui, The draft genome, transcriptome, and microbiome of *Dermatophagoides farinae* reveal a broad spectrum of dust mite allergens, *J. Allergy Clin. Immunol.* 135 (2) (2015) 539–548.
- [18] M. Mondal, P. Klimov, A.S. Flynt, Rewired RNAi-mediated genome surveillance in house dust mites, *PLoS Genet.* 14 (1) (2018) e1007183.
- [19] T.S. Prasad, A.K. Mohanty, M. Kumar, S.K. Sreenivasamurthy, G. Dey, R.S. Nirujogi, S.M. Pinto, A.K. Madugundu, A.H. Patil, J. Advani, S.S. Manda, M.K. Gupta, S.B. Dwivedi, D.S. Kelkar, B. Hall, X. Jiang, A. Peery, P. Rajagopalan, S.D. Yelamanchi, H.S. Solanki, R. Raja, G.J. Sathe, S. Chavan, R. Verma, K.M. Patel, A.P. Jain, N. Syed, K.K. Datta, A.A. Khan, M. Dammali, S. Jayaram, A. Radhakrishnan, C.J. Mitchell, C.-H. Na, N. Kumar, P. Sannis, I.V. Sharakhov, C. Wang, H. Gowda, Z. Tu, A. Kumar, A. Pandey, Integrating transcriptomic and proteomic data for accurate assembly and annotation of genomes, *Genome Res.* 27 (1) (2017) 133–144.
- [20] T. Erban, J. Hubert, Determination of pH in regions of the midguts of acarid mites, *J. Insect Sci.* 10 (2010) 42.
- [21] F. Moreno Benitez, M. Espinazo Romeu, A. Letren Camacho, S. Mas, F.J. Garcia-Cozar, A.I. Tabar, Variation in allergen content in sublingual allergen immunotherapy with house dust mites, *Allergy* 70 (11) (2015) 1413–1420.
- [22] S. Post, M.C. Nawijn, T.L. Hackett, M. Baranowska, R. Gras, A.J.M. van Oosterhout, I.H. Heijink, The composition of house dust mite is critical for mucosal barrier dysfunction and allergic sensitization, *Thorax* 67 (6) (2012) 488–495.
- [23] H. Henmar, S.M.T. Frisenette, K. Grosch, K. Nielsen, G. Smith, S. Sonderkaer, J.N. Larsen, Fractionation of source materials leads to a high reproducibility of the SQ house dust mite SLIT-tablets, *Int. Arch. Allergy Immunol.* 169 (1) (2016) 23–32.
- [24] K.Y. Jeong, S.-Y. Choi, J.-H. Lee, I.-Y. Lee, T.-S. Yong, J.-S. Lee, C.-S. Hong, J.-W. Park, Standardization of house dust mite extracts in Korea, *Allergy, Asthma Immunol. Res.* 4 (6) (2012) 346–350.
- [25] J. Carnes, V. Iraola, S.H. Cho, R.E. Esch, Mite allergen extracts and clinical practice, *Ann. Allergy Asthma Immunol.* 118 (3) (2017) 249–256.
- [26] E. Eraso, J.A. Guisantes, J. Martinez, M. Saenz-de-Santamaria, A. Martinez, R. Palacios, R. Cisterna, Kinetics of allergen expression in cultures of house dust mites, *Dermatophagoides pteronyssinus* and *D. farinae* (Acari: Pyroglyphidae), *J. Med. Entomol.* 34 (6) (1997) 684–689.
- [27] E. Eraso, J. Martinez, P. Garcia-Ortega, A. Martinez, R. Palacios, R. Cisterna, J.A. Guisantes, Influence of mite growth culture phases on the biological standardization of allergenic extracts, *J. Investig. Allergol. Clin. Immunol.* 8 (4) (1998) 201–206.
- [28] E. Eraso, J. Martinez, A. Martinez, R. Palacios, J.A. Guisantes, Quality parameters for the production of mite extracts, *Allergol. Immunopathol. (Madr.)* 25 (3) (1997) 113–117.
- [29] J.C. Vidal-Quist, F. Ortego, M. Lombardero, P. Castanera, P. Hernandez-Crespo, Allergen expression in the European house dust mite *Dermatophagoides pteronyssinus* throughout development and response to environmental conditions, *Med. Vet. Entomol.* 29 (2) (2015) 137–146.
- [30] J. Jalil-Colome, A.D. de Andrade, J. Birnbaum, D. Casanova, J.-L. Mege, A. Lanteaume, D. Charpin, D. Vervolte, Sex difference in Fel d 1 allergen production, *J. Allergy Clin. Immunol.* 98 (1) (1996) 165–168.
- [31] R.K. Hodgson, Sex ratio and sex determination in the American house dust mite, *Dermatophagoides farinae*, *Ann. Entomol. Soc. Am.* 69 (6) (1976) 1085–1086.
- [32] T. Erban, J. Hubert, Digestive function of lysozyme in synanthropic acarid mites enables utilization of bacteria as a food source, *Exp. Appl. Acarol.* 44 (3) (2008) 199–212.
- [33] J. Hubert, J. Kopecky, M. Sagova-Mareckova, M. Nesvorna, L. Zurek, T. Erban, Assessment of bacterial communities in thirteen species of laboratory-cultured domestic mites (Acari: Acarididae), *J. Econ. Entomol.* 109 (4) (2016) 1887–1896.
- [34] T. Erban, K. Harant, J. Hubert, 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. Proteomics* 162 (2017) 11–19.
- [35] P.B. Klimov, B.M. O'Connor, Improved tRNA prediction in the American house dust mite reveals widespread occurrence of extremely short minimal tRNAs in acariform mites, *BMC Genomics* 10 (2009) 598.
- [36] P.B. Klimov, L.L. Knowles, Repeated parallel evolution of minimal tRNAs revealed from detailed comparative analysis, *J. Hered.* 102 (3) (2011) 283–293.
- [37] T. Erban, B. Sopko, P. Talacko, K. Harant, K. Kadlikova, T. Halesova, K. Riddellova, A. Pekas, Chronic exposure of bumblebees to neonicotinoid imidacloprid suppresses the entire mevalonate pathway and fatty acid synthesis, *J. Proteomics* 196 (2019) 69–80.
- [38] J. Cox, M.Y. Hein, C.A. Luber, I. Paron, N. Nagaraj, M. Mann, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ, *Mol. Cell. Proteomics* 13 (9) (2014) 2513–2526.
- [39] J. Cox, N. Neuhauser, A. Michalski, R.A. Scheltema, J.V. Olsen, M. Mann, Andromeda: a peptide search engine integrated into the MaxQuant environment, *J. Proteome Res.* 10 (4) (2011) 1794–1805.
- [40] J. Cox, M. Mann, 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data, *BMC Bioinformatics* 13 (Suppl. 16) (2012) S12.
- [41] A. Mari, D. Riccioli, The allergome web site - a database of allergenic molecules. Aim, structure, and data of a web-based resource, *J. Allergy Clin. Immunol.* 113 (2) (2004) S301.
- [42] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, *J. Mol. Biol.* 215 (3) (1990) 403–410.
- [43] A. Marchler-Bauer, Y. Bo, L. Han, J. He, C.J. Lanczycki, S. Lu, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwatz, D.I. Hurwitz, F. Lu, G.H. Marchler, J.S. Song, N. Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng, L.Y. Geer, S.H. Bryant, CDD/SPARCLE: functional classification of proteins via subfamily domain architectures, *Nucleic Acids Res.* 45 (D1) (2017) D200–D203.
- [44] L.Y. Geer, M. Domrachev, D.J. Lipman, S.H. Bryant, CDART: protein homology by domain architecture, *Genome Res.* 12 (10) (2002) 1619–1623.
- [45] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol. Syst. Biol.* 7 (2011) 539.
- [46] A.M. Waterhouse, J.B. Procter, D.M.A. Martin, M. Clamp, G.J. Barton, Jalview Version 2—a multiple sequence alignment editor and analysis workbench, *Bioinformatics* 25 (9) (2009) 1189–1191.
- [47] SIB, PeptideCutter, ExPASy - Swiss Institute of Bioinformatics (SIB) Bioinformatics Resource Portal, https://web.expasy.org/peptide_cutter/, (2019), Accessed date: 6 February 2019.
- [48] J.-x. Sun, L.-l. Yu, Y. Zhou, Y.-h. Bian, L. Yang, N. Wang, Y.-b. Cui, Cloning, expression and analysis of Der f Mag 29 allergen of *Dermatophagoides farinae*, *Chinese J. Parasitol. Parasit. Dis.* 31 (6) (2013) 480–482 in Chinese with English abstract.
- [49] T. Aki, A. Fujikawa, T. Wada, T. Jyo, S. Shigeta, Y. Murooka, S. Oka, K. Ono, Cloning and expression of cDNA coding for a new allergen from the house dust mite, *Dermatophagoides farinae*: homology with human heat shock cognate proteins in the heat shock protein 70 family, *J. Biochem.* 115 (3) (1994) 435–440.
- [50] A. Conesa, S. Gotz, J.M. Garcia-Gomez, J. Terol, M. Talon, M. Robles, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, *Bioinformatics* 21 (18) (2005) 3674–3676.
- [51] T. Erban, J. Zitek, M. Bodrinova, P. Talacko, M. Bartos, J. Hrabak, Comprehensive proteomic analysis of exoproteins expressed by ERIC I, II, III and IV *Paenibacillus larvae* genotypes reveals a wide range of virulence factors, *Virulence* 10 (1) (2019) 363–375.
- [52] T. Erban, E. Shcherbachenko, P. Talacko, K. Harant, The unique protein composition of honey revealed by comprehensive proteomic analysis: allergens, venom-like proteins, antibacterial properties, royal jelly proteins, serine proteases, and their inhibitors, *J. Nat. Prod.* 82 (5) (2019) 1217–1226.
- [53] A. Casset, A. Mari, A. Purohit, Y. Resch, M. Weghofer, R. Ferrara, W.R. Thomas, C. Alessandri, K.-W. Chen, F. de Blay, R. Valenta, S. Vrtala, Varying allergen composition and content affects the in vivo allergenic activity of commercial *Dermatophagoides pteronyssinus* extracts, *Int. Arch. Allergy Immunol.* 159 (3) (2012) 253–262.
- [54] T. Erban, R. Klubal, Zpusob pripravy exkrementovych antigenu domacich roztocu: patent cislo 304351 [Process for preparing excrement antigens of domestic mites: patent: registration number: CZ304351], Industrial Property Office, Prague, 2014 in Czech.
- [55] E. Santos da Silva, C. Asam, P. Lackner, H. Hofer, M. Wallner, C. Silva Pinheiro, N.M. Alcantara-Neves, F. Ferreira, Allergens of *Blomia tropicalis*: an overview of recombinant molecules, *Int. Arch. Allergy Immunol.* 172 (4) (2017) 203–214.
- [56] L. Puerta, L. Caraballo, E. Fernandez-Caldas, A. Avjioglu, D.G. Marsh, R.F. Lockey, M.L. Dao, Nucleotide sequence analysis of a complementary DNA coding for a *Blomia tropicalis* allergen, *J. Allergy Clin. Immunol.* 98 (5) (1996) 932–937.
- [57] J. Zakzuk, S. Jimenez, N. Cheong, L. Puerta, B.W. Lee, K.Y. Chua, L. Caraballo, Immunological characterization of a Blo t 12 isoallergen: identification of immunoglobulin E epitopes, *Clin. Exp. Allergy* 39 (4) (2009) 608–616.
- [58] K.-L. Lin, K.-H. Hsieh, W.R. Thomas, B.-L. Chiang, K.-Y. Chua, Characterization of Der p V allergen, cDNA analysis, and IgE-mediated reactivity to the recombinant protein, *J. Allergy Clin. Immunol.* 94 (6) (1994) 989–996.
- [59] A. Simpson, R. Green, A. Custovic, A. Woodcock, L.K. Arruda, M.D. Chapman, Skin test reactivity to natural and recombinant *Blomia* and *Dermatophagoides* spp. allergens among mite allergic patients in the UK, *Allergy* 58 (1) (2003) 53–56.
- [60] C. King, R.J. Simpson, R.L. Moritz, G.E. Reed, P.J. Thompson, G.A. Stewart, The isolation and characterization of a novel collagenolytic serine protease allergen (Der p 9) from the dust mite *Dermatophagoides pteronyssinus*, *J. Allergy Clin. Immunol.* 98 (4) (1996) 739–747.
- [61] J. Le Mao, C.E. Mayer, G. Peltre, F.X. Desvaux, B. David, A. Weyer, H. Senechal, Mapping of *Dermatophagoides farinae* mite allergens by two-dimensional immunoblotting, *J. Allergy Clin. Immunol.* 102 (4) (1998) 631–636.
- [62] H.S. Shang, S.T. Ong, X.Z. Bi, L.C. Tsai, S.H. Lim, F.T. Chew, Cloning and immunological characterization of complete group 9 allergens from 3 major dust mites, *Blomia tropicalis*, *Dermatophagoides farinae* and *Dermatophagoides pteronyssinus*, *J. Allergy Clin. Immunol.* 111 (2) (2003) S324.
- [63] Y. Cui, F. Teng, L. Yu, Y. Zhou, C. Zhang, L. Yang, *Dermatophagoides farinae* allergen Der f 9: cloning, expression, purification, characterization and IgE-binding in children with atopic asthma, *Pediatr. Pulmonol.* 52 (3) (2017) 282–292.
- [64] T. Erban, D. Rybanska, K. Harant, B. Hortova, J. Hubert, 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 exo-chitinases, *Front. Physiol.* 7 (2016) 53.
- [65] S. Kondo, H. Helin, M. Shichijo, K.B. Bacon, Cockroach allergen extract stimulates protease-activated receptor-2 (PAR-2) expressed in mouse lung fibroblast, *Inflamm. Res.* 53 (9) (2004) 489–496.
- [66] S.K. Jeong, H.J. Kim, J.-K. Youm, S.K. Ahn, E.H. Choi, M.H. Sohn, K.-E. Kim, J.H. Hong, D.M. Shin, S.H. Lee, Mite and cockroach allergens activate protease-activated receptor 2 and delay epidermal permeability barrier recovery, *J. Invest. Dermatol.* 128 (8) (2008) 1930–1939.
- [67] T. Roelandt, C. Heughebaert, J.-P. Hachem, Proteolytically active allergens cause barrier breakdown, *J. Invest. Dermatol.* 128 (8) (2008) 1878–1880.
- [68] H. Wan, H.L. Winton, C. Soeller, E.R. Tovey, D.C. Gruenert, P.J. Thompson, G.A. Stewart, G.W. Taylor, D.R. Garrod, M.B. Cannell, C. Robinson, Der p 1 facilitates transepithelial allergen delivery by disruption of tight junctions, *J. Clin.*

- Invest. 104 (1) (1999) 123–133.
- [69] N. Asokanathan, P.T. Graham, D.J. Stewart, A.J. Bakker, K.A. Eidne, P.J. Thompson, G.A. Stewart, House dust mite allergens induce proinflammatory cytokines from respiratory epithelial cells: the cysteine protease allergen, Der p 1, activates protease-activated receptor (PAR)-2 and inactivates PAR-1, *J. Immunol.* 169 (8) (2002) 4572–4578.
- [70] R. Hamid, M.A. Khan, M. Ahmad, M.M. Ahmad, M.Z. Abidin, J. Musarrat, S. Javed, Chitinases: an update, *J. Pharm. BioAllied Sci.* 5 (1) (2013) 21–29.
- [71] C. McCall, S. Hunter, K. Stedman, E. Weber, A. Hillier, C. Bozic, B. Rivoire, T. Olivry, Characterization and cloning of a major high molecular weight house dust mite allergen (Der f 15) for dogs, *Vet. Immunol. Immunopathol.* 78 (3–4) (2001) 231–247.
- [72] E. Weber, S. Hunter, K. Stedman, S. Dreitz, T. Olivry, A. Hillier, C. McCall, Identification, characterization, and cloning of a complementary DNA encoding a 60-kd house dust mite allergen (Der f 18) for human beings and dogs, *J. Allergy Clin. Immunol.* 112 (1) (2003) 79–86.
- [73] S. Blank, H. Seismann, M. McIntyre, M. Ollert, S. Wolf, F.I. Bantleon, E. Spillner, Vitellogenins are new high molecular weight components and allergens (Api m 12 and Ves v 6) of *Apis mellifera* and *Vespula vulgaris* venom, *PLoS One* 8 (4) (2013) e62009.
- [74] L. Porter, Z. Radulovic, T. Kim, G.R.C. Braz, I. Da Silva Vaz Jr., A. Mulenga, Bioinformatic analyses of male and female *Amblyomma americanum* tick expressed serine protease inhibitors (serpins), *Ticks Tick Borne Dis.* 6 (1) (2015) 16–30.
- [75] Z. Li, C.L. King, J.O. Ogundipe, L.S. Licata, R.E. Blanton, Preferential recognition by human IgE and IgG4 of a species-specific *Schistosoma haematobium* serine protease inhibitor, *J. Infect. Dis.* 171 (2) (1995) 416–422.
- [76] J. Lin, M. Li, Y. Liu, C. Jiang, Y. Wu, Y. Wang, A. Gao, Z. Liu, P. Yang, X. Liu, Expression, purification and characterization of Der f 27, a new allergen from *Dermatophagoides farinae*, *Am. J. Transl. Res.* 7 (7) (2015) 1260–1270.
- [77] W.R. Thomas, Hierarchy and molecular properties of house dust mite allergens, *Allergol. Int.* 64 (4) (2015) 304–311.
- [78] M. Weghofer, M. Grote, Y. Resch, A. Casset, M. Kneidinger, J. Kopec, W.R. Thomas, E. Fernandez-Caldas, M. Kabesch, R. Ferrara, A. Mari, A. Purohit, G. Pauli, F. Horak, W. Keller, P. Valent, R. Valenta, S. Vrtala, 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.* 190 (7) (2013) 3059–3067.
- [79] G.A. Mueller, T.A. Randall, J. Glesner, L.C. Pedersen, L. Perera, L.L. Edwards, E.F. DeRose, M.D. Chapman, R.E. London, A. Pomes, Serological, genomic and structural analyses of the major mite allergen Der p 23, *Clin. Exp. Allergy* 46 (2) (2016) 365–376.
- [80] J. Hubert, M. Nesvorna, P. Klimov, S.E. Dowd, B. Sopko, T. Erban, Differential Allergen Expression in Three *Tyrophagus putrescentiae* Strains Inhabited by Distinct Microbiome, *Allergy* (in Press), (2019), <https://doi.org/10.1111/all.13921>.
- [81] A. Andersen, Nutritional value of yeast for *Dermatophagoides pteronyssinus* (Acari: Epidermoptidae) and the antigenic and allergenic composition of extracts during extended culturing, *J. Med. Entomol.* 28 (4) (1991) 487–491.
- [82] S. Avula-Poola, M.S. Morgan, L.G. Arlian, Diet influences growth rates and allergen and endotoxin contents of cultured *Dermatophagoides farinae* and *Dermatophagoides pteronyssinus* house dust mites, *Int. Arch. Allergy Immunol.* 159 (3) (2012) 226–234.
- [83] T. Erban, K. Harant, M. Hubalek, P. Vitamvas, M. Kamler, P. Poltronieri, J. Tyl, M. Markovic, D. Titera, In-depth proteomic analysis of *Varroa destructor*: detection of DWV-complex, ABPV, VdMLV and honeybee proteins in the mite, *Sci. Rep.* 5 (2015) 13907.
- [84] S. An, C. Shen, X. Liu, L. Chen, X. Xu, M. Rong, Z. Liu, R. Lai, Alpha-actinin is a new type of house dust mite allergen, *PLoS One* 8 (12) (2013) e81377.

Fig. S1 Histograms after the 16 nLC-MS/MS runs. **A)** before and **B)** after the matrix was filtered to proteins that contained at least three valid values in total.

