

Blueprint for the House Dust Mite

Wayne R. Thomas

Telethon Kids Institute, University of Western Australia, West Perth, WA, Australia

Why would you want to sequence the house dust mite genome? Randall et al. [1] help explain this in an analysis of their assembly of the genomic sequence of *Dermatophagoides pteronyssinus*. The assembly, also reported by Waldron et al. [2], in fact follows those for *D. farinae* [3] and the closely related *Euroglyphus maynei* [4], but for the first time attention is directed to the issues that are important for allergy research, going beyond being a tool for the verification of newly discovered allergens. The completeness of the coverage of a genome assembly is a critical starting point for all genomic projects so considerable attention is given to the sequencing metrics, which showed high indicators of success. The genome length of 52.5 Mb encoding a predicted 19,368 proteins was similar to that found for *D. farinae* (53.5 Mb) and scabies (56.2 Mb) [5], although somewhat different to the 70.7 Mb found for *E. maynei*, indicating that further resolution is required. Despite the gross morphological difference of *E. maynei* from the 2 *Dermatophagoides* species, the sequences of the allergens of *E. maynei* have shown only the same disparity from *Dermatophagoides* sp. as that found between *D. farinae* and *D. pteronyssinus*. Therefore, if the difference is confirmed by continued analysis, it would be of major evolutionary significance. As expounded by Rider et al. [4], the original appraisals of completeness of the genome assemblies of the Acari were low but these were based on evolutionary expectations for the presence of different types of genes found in other arthropods and in particular mainly in insects. Now that a number of di-

verse acarid genome assemblies have been obtained for comparison it has been possible to appreciate their consistency and their completeness.

The *D. pteronyssinus* genomic sequence was found to contain all of the sequences for allergens registered in the WHO/IUIS allergen nomenclature subcommittee database as well as the homologues of all of the sequences entered into the database for *D. farinae* except for the EF-hand calcium-binding protein Der f 17, for which there is no recorded molecular information. It was similarly left unassigned by Chan et al. [3] for *D. farinae*. The homologues include those of the newly described Der f 24–34 allergens that were mostly identified in proteomic screens in China and have yet to be assessed for their IgE binding by gravimetric estimates or by titrations compared to the serodominant group 1 and 2 allergens. The ability to make recombinant *D. pteronyssinus* homologues of these components from the information in the genome not only opens up these investigations to studies in the many more regions where people are substantively or almost exclusively sensitized by *D. pteronyssinus*, but also to the possibility that investigators in these regions are more interested in making authenticable assessments. The T-cell responses of house dust mite-allergic subjects from Europe and America made to peptides representing predicted epitopes of many of these proteins showed that they did not induce substantive T-cell responses compared to the responses induced by the high IgE-binding components authenticated earlier with quantitative analyses [6],

so there are questions to be resolved especially on the size and consistency of the responses and possible environmental influences and cross-reactivities.

The dominant allergens of the grass [7, 8], ragweed [9], and birch [10] pollens belong to gene families that might be an important determinant for their prevalent and high immune reactivity. The production of proteins from several related genes would not only allow more protein to be produced, and perhaps in different circumstances, but also facilitate binding to a wide range of the antigen-presenting major histocompatibility molecules. The house dust mite genomes show unequivocally that the serodominant group 1 and 2 allergens are produced from single genes, although the genes for Der p 1 and Der f 1 were both closely linked to another cysteine protease gene coding for a protein of about 50% amino acid identity [1, 3]. This is obviously a prime candidate for further study. The recently discovered Der f 35, shown by quantitative analysis to have high IgE binding [11], is interestingly an ML-domain protein like the group 2 allergens but, in keeping with its overall disparate sequence, has little cross-reactivity with its group 2 counterpart and is not genetically linked for coordinated expression. The group 23 peritrophin-like components can now be classed as serodominant because of their high prevalence of IgE binding [12–14], and from their chitin-binding sequence motifs were shown to be members of large families. *D. pteronyssinus* had a family of 48 chitin-binding domain proteins with the Der p 23 gene linked to genes coding for 5 of these proteins, and there were 47 such chitin-binding domain homologues for *D. farinae*, although only 3 were linked to that for Der f 23 [1]. As recorded in GenBank, there have been a large number of allelic variants reported for Der f 23, with 1 showing a large insertion of a possible oligosylation (threonine, serine, proline-rich) domain and these might induce different specificities of IgE antibodies with different affinities to the group 23 proteins. Gene families thus might have some importance in the response to the group 23 allergens, which might be significant in the variation of titers reported in the sera of subjects from different geographical regions [14]. They can now be definitively studied with information from the genome.

Of the 4 mid-tier allergens known to make important contributions to the overall size of IgE responses of most patients, there is only 1 copy of the genes for the group 7 components and there are no further members of the group 5 and 21 allergens that are both homologous to and genetically closely linked to each other. Groups 5 and 21 are interesting from an evolutionary perspective because

the mature proteins show a low (26%) degree of sequence identity with each other, but they each have about 40% identity to the important group 5 allergen of *Blomia tropicalis*. Der p 21, but not Der p 5, is a significant source of IgE cross-reactivity between the species [15]. The other allergen components recognized as showing mid-tier IgE binding are the group 4 amylases. As found and highlighted by Randall et al. [1], and as reported in abstract form from the genomic study by Mills et al. [16], *D. pteronyssinus*, like many organisms, has a second adjacent amylase gene. Its sequence, however, as found in the *D. pteronyssinus* genome and by Mills et al. [16], has a truncated leader sequence and a large degree of sequence disparity from Der p 4 at the C-terminal that includes the absence of 4 cysteines for disulphide bonds that are highly conserved in the amylases of many species. The genomic sequence for this gene in *D. farinae* [3] has the truncated leader sequence but retains the conserved cysteines, indicating a region of divergence between the species and a greater possibility that it is expressed by the mite and is stable. Given that, as a mid-tier allergen, Der p 4 might be an important constituent of formulations for immunotherapy and diagnosis, the product of the second gene with its 70% sequence identity warrants further investigation.

The amylases are also important because scabies-infected and previously exposed subjects have been found to produce high titers of IgE binding to Der p 4 and to Der p 20 in the absence of IgE antibodies to the serodominant house dust mite allergens. These components might accordingly be able to identify confounding reactions to skin prick tests and IgE binding due to the previous and current scabies infections, which are highly prevalent in regions of low income and amongst immigrants and refugees [17]. The Der p 20 arginine kinase is well known for its abundance in mite bodies and its propensity for IgE cross-reactivities because of its highly conserved primary structure. The Der p 4 amylase cross-reactivity, however, has become a conundrum because the several genome sequences of scabies [5, 18] have shown the notable absence of an amylase gene and no genes for proteins that would be expected to be highly cross-reactive. The induction of anti-amylase antibodies by scabies has also been found following the experimental infection of pigs [19], so perhaps the response could be to an amylase of a commensal microorganism. Since the reactivity in pigs could be detected with bacterial and house fly amylases, it might also be due to interactions with 1 of the organisms of the skin lesions. Another possibility is a cross-reactivity with autoantibodies found after scabies infestations [20]. What-

ever the cause, this is a good example of the definitive information and questions that the genome sequence has produced. Indeed, investigators had been unsuccessfully scouring expressed sequence tag data for scabies for amylase transcripts.

In further bounty for the aficionado, it is noted that there was only 1 gene for the large lipid-binding protein Der p 14 and no closely related homologues. IgE binding to Der p 14 and a small group of other apparently minor IgE-binding proteins has been associated with an additional propensity of house dust mite-allergic children to develop asthma [21], and it has also been found to be one of the most stimulatory allergens as assessed by the T-cell epitope study by Oseroff et al. [6]. Antibodies produced to recombinant Der p 14 have not only revealed the presence of a Der p 14-like molecule in house dust mite eggs, which would be expected to be a vitellogenin known to belong to this type of protein, but also a protein in the bodies of male mites that correspond to the related lipid transporter apolipoprotein [22]. The information revealing a single gene shows that 1 molecule is performing both functions, as also described for the crab, and that is all that is needed for the recombinant formulations.

Randall et al. [1] noted that their assembly for *D. pteronyssinus* revealed the full sequence of the gene of the Der p 4 amylase allergen. To elaborate, the mRNA sequence in the GenBank lacked the leader and the short proenzyme sequences, which can be critical for recombinant protein production, and was provided from the genome. The production of recombinant Der p 4 and other amylases in *Escherichia coli* hosts does not produce an enzymatically active protein, but it could be produced as an active enzyme by the yeast *Pichia pastoris*, although with a low yield curtailing the investigations planned for this allergen. From the experience reported for other amylases this can be done with a better yield from the proenzyme [23] and, as discussed above, would be a priority for molecular diagnosis. The genome assembly also provided similar information lacking from the Der p 14 cDNA sequence. The 1650 amino acid allergen is highly susceptible to proteolysis as a natural protein and, to date, has only been represented by recombinant peptides in IgE assays, meaning an effective recombinant expression system for the whole molecule is sorely needed. Knowledge of the introns can also help the production of recombinant proteins with constructs with introns being expressed at higher levels in mammalian and insect cells and yeast than intron-less constructs, with a 23-fold improvement being reported [24]. This could not only be important for allergens with a poor quantity and quality,

such as Der p 11 and 14, but also for all other allergens for industrial-level production in eukaryotic systems. From these examples, the assembly will contain a wealth of information to be mined for recombinant expression not only by microorganisms, but also from transgenic hosts.

From the products being licensed for immunotherapy, it appears that the practice of using house dust mite extracts in lieu of the development of standardized formulations of allergen components will be supported for many years. Such extracts vary enormously even in the concentrations of the long-recognized major components Der p 1 and 2, and often lack other IgE-binding proteins [25]. High-resolution and accurate mass spectrometry is finding considerable application in measuring the protein constituents of these complex mixtures [26]. The quantitative technology depends on a comparison of the amount of a marker peptide from the constituent being measured compared with a reference peptide synthesized with the same sequence but with heavy amino acid isotopes. The comprehensive sequences now provided in the genome will allow a full audit of different extracts which, when compared to their efficacy for diagnosis and therapy, could provide telling information for the requirements for different components. Mass spectrometry linked to a database of all the protein sequences encoded in the genome can also be a discovery tool to ascertain the distribution of house dust mite products in the environment, especially in inhalable air. Little is known about which house dust mite products are inhaled. It is thought, although with little evidence, that there could be a large contribution from proteins distributed in the feces compared to proteins from decaying body components, but there might also be novel sources such as eggs. Comparisons of the inhaled house dust mite components that do and do not induce IgE antibodies could be very informative for the quest to find characteristics that are linked to allergenicity or to nonallergenic immunoregulatory pathways. The comparisons would extend from concentration and stability [27] through to the propensity to interact with the different elements of the innate immune system [28]. Currently, there is almost no information on proteins that are in inhalable air besides that gleaned for Der p 1 by monoclonal antibody assays, but this has been highly instructive in showing that nasal filters and personal monitors can collect enough protein to measure and to provide baseline information on Der p 1 for a reference [29]. From measuring T-cell responses to peptides representing the epitopes of allergens and other proteins represented in the mite transcriptome, there is evidence that quite a few nonallergenic proteins induce T-cell re-

sponses in humans [6]. Therefore, these can now be studied comprehensively and be extended by studying responses to proteins known to be in the inhalable air at known doses.

The genomic sequences that have become available for several different mite species and their evolutionary relationships provide a valuable resource for understanding many aspects of mite biology and for understanding their growth and development, as well as for research into biological control. Randall et al. [1] now provide the first analysis of the genomic organization of the house dust mite as it affects our knowledge of allergens. The advances have come in tandem with other big data on the transcriptome [3, 6, 30] and proteome [30], noting that both these depend on growth conditions and can now be referenced to the genome data. Although not so up-scalable, the implementation and refinement of megapool peptide techniques for comparing T-cell responses to large assemblies of house dust mite proteins provides a powerful tool with which to interrogate the interactions of mite

proteins with the human immune system [6], as does the systematic production of recombinant proteins for serology. The time for comprehensive investigations of allergic and nonallergic responses to house dust mite proteins at the molecular level has now well and truly arrived. It carries not only the promise of gaining new insights from a new perspective, but also for them to be underpinned by precise and reproducible data obtained with known amounts of defined reagents. The genomes of storage and related mites are, however, yet to come, especially of *B. tropicalis*, which not only causes a high prevalence of allergy in highly populous tropical and subtropical countries, but does so by inducing responses to a different profile of allergenic components [31].

Disclosure Statement

Wayne Thomas is an inventor with house dust mite allergen patents assigned to the Telethon Kids Institutes and receives royalty payments from their licensing of the patents.

References

- Randall TA, Mullikin J, Mueller GA: The draft genome assembly of *Dermatophagoides pteronyssinus* supports identification of novel allergen isoforms in *Dermatophagoides* species. *Int Arch Allergy Immunol* 2018, DOI: 10.1159/000481989.
- Waldron R, McGowan J, Gordon N, McCarthy C, Mitchell EB, Doyle S, Fitzpatrick DA: Draft genome sequence of *Dermatophagoides pteronyssinus*, the European house dust mite. *Genome Announc* 2017;5:e00789-17.
- Chan TF, Ji KM, Yim AK, Liu XY, Zhou JW, Li RQ, Yang KY, Li J, Li M, Law PT, Wu YL, Cai ZL, Qin H, Bao Y, Leung RK, Ng PK, Zou J, Zhong XJ, Ran PX, Zhong NS, Liu ZG, Tsui SK: The draft genome, transcriptome, and microbiome of *Dermatophagoides farinae* reveal a broad spectrum of dust mite allergens. *J Allergy Clin Immunol* 2015;135:539-548.
- Rider SD, Morgan MS, Arlian LG: Allergen homologs in the *Euroglyphus maynei* draft genome. *PLoS One* 2017;12:e0183535.
- Rider SD, Morgan MS, Arlian LG: Draft genome of the scabies mite. *Parasit Vectors* 2015;8:585.
- Oseroff C, Christensen LH, Westernberg L, Pham J, Lane J, Paul S, Greenbaum J, Stranzl T, Lund G, Hoof I, Holm J, Wurtzen PA, Meno KH, Frazier A, Schulten V, Andersen PS, Peters B, Sette A: Immunoproteomic analysis of house dust mite antigens reveals distinct classes of dominant T cell antigens according to function and serological reactivity. *Clin Exp Allergy* 2017;47:577-592.
- Andersson K, Lidholm J: Characteristics and immunobiology of grass pollen allergens. *Int Arch Allergy Immunol* 2003;130:87-107.
- Sharma A, Sharma N, Bhalla P, Singh M: Comparative and evolutionary analysis of grass pollen allergens using *Brachypodium distachyon* as a model system. *PLoS One* 2017;12:e0169686.
- Wolf M, Twaroch TE, Huber S, Reithofer M, Steiner M, Aglas L, Hauser M, Aloisi I, Asam C, Hofer H, Parigiani MA, Ebner C, Bohle B, Briza P, Neubauer A, Stolz F, Jahn-Schmid B, Wallner M, Ferreira F: Amb a 1 isoforms: unequal siblings with distinct immunological features. *Allergy* 2017;72:1874-1882.
- Spiric J, Engin AM, Karas M, Reuter A: Quality control of biomedical allergen products – highly complex isoallergen composition challenges standard MS database search and requires manual data analyses. *PLoS One* 2015;10:e0142404.
- Fujimura T, Aki T, Isobe T, Matsuoka A, Hayashi T, Ono K, Kawamoto S: Der f 35: an MD-2-like house dust mite allergen that cross-reacts with Der f 2 and Pso o 2. *Allergy* 2017;72:1728-1736.
- Weghofer M, Grote M, Resch Y, Casset A, Kneidinger M, Kopec J, Thomas WR, Fernández-Caldas E, Kabisch M, Ferrara R, Mari A, Purohit A, Pauli G, Horak F, Keller W, Valent P, Valenta R, Vrtala S: Identification of Der p 23, a peritrophin-like protein, as a new major *Dermatophagoides pteronyssinus* allergen associated with the peritrophic matrix of mite fecal pellets. *J Immunol* 2013;190:3059-3067.
- Resch Y, Michel S, Kabisch M, Lupinek C, Valenta R, Vrtala S: Different IgE recognition of mite allergen components in asthmatic and nonasthmatic children. *J Allergy Clin Immunol* 2015;136:1083-1091.
- Batard T, Baron-Bodo V, Martelet A, Le Mignon M, Lemoine P, Jain K, Mariano S, Horiot S, Chabre H, Harwanegg C, Marquette CA, Corgier BP, Soh WT, Satitsuksanoa P, Jacquet A, Chew FT, Nony E, Moingeon P: Patterns of IgE sensitization in house dust mite-allergic patients: implications for allergen immunotherapy. *Allergy* 2016;71:220-229.
- Tan KW, Ong TC, Gao YF, Tiong YS, Wong KN, Chew FT, Mok YK: NMR structure and IgE epitopes of Blo t 21, a major dust mite allergen from *Blomia tropicalis*. *J Biol Chem* 2012;287:34776-34785.
- Mills KL, Thomas WR, Smith W: Characterization of the group 4 allergens of the house dust mite. *J Allergy Clin Immunol* 2002;109:S180.
- Walton SF, Slender A, Pizutto S, Mounsey K, Opresecu F, Thomas WR, Hales BJ, Currie BJ: Analysis of IgE binding patterns to house dust mite allergens in scabies endemic communities: insights for both diseases. *Clin Exp Allergy* 2015;45:1868-1872.
- Mofiz E, Holt DC, Seemann T, Currie BJ, Fischer K, Papenfuss AT: Genomic resources and draft assemblies of the human and porcine varieties of scabies mites, *Sarcoptes scabiei* var. *hominis* and var. *swis*. *GigaScience* 2016;5:23.

- 19 Sandeman RM, Zalunardo M, Cargill CF: Method for detection of mite infection. World Intellectual Property Organization Patent WO2001011365. 2001.
- 20 Toet HM, Fischer K, Mounsey KE, Sandeman RM: Autoantibodies to iron-binding proteins in pigs infested with *Sarcoptes scabiei*. *Vet Parasitol* 2014;205:263–270.
- 21 Posa D, Perna S, Resch Y, Lupinek C, Panetta V, Hofmaier S, Rohrbach A, Hatzler L, Grabenhenrich L, Tsilochristou O, Chen KW, Bauer CP, Hoffman U, Forster J, Zepp F, Schuster A, Wahn U, Keil T, Lau S, Vrtala S, Valenta R, Matricardi PM: Evolution and predictive value of IgE responses toward a comprehensive panel of house dust mite allergens during the first 2 decades of life. *J Allergy Clin Immunol* 2017;139:541–549.
- 22 Epton MJ, Malainual N, Smith W, Thomas WR: Vitellinogenin-apolipoprotein like allergen Der p 14 is a major specificity in house dust mite sensitisation. *J Allergy Clin Immunol* 2001;107:S107.
- 23 Kato S, Ishibashi M, Tatsuda D, Tokunaga H, Tokunaga M: Efficient expression, purification and characterization of mouse salivary alpha-amylase secreted from methylotrophic yeast, *Pichia pastoris*. *Yeast* 2001;18:643–655.
- 24 Hasannia S, Lotfi AS, Mahboudi F, Rezaii A, Rahbarizadeh F, Mohsenifar A: Elevated expression of human alpha-1 antitrypsin mediated by yeast intron in *Pichia pastoris*. *Bio-technol Lett* 2006;28:1545–1550.
- 25 Casset A, Mari A, Purohit A, Weghofer M, Ferrara R, Thomas WR, Zaffiro A, Chen KW, de Blay F, Valenta R, Vrtala S: Varying allergen composition and content affects in vivo allergenic activity of commercial *Dermatophagoides pteronyssinus* extracts. *Int Arch Allergy Immunol* 2012;159:253–226.
- 26 Seppala U, Daulay C, Robinson S, Hornshaw M, Larsen JN, Ipsen H: Absolute quantification of allergens from complex mixtures: a new sensitive tool for standardization of allergen extracts for specific immunotherapy. *J Proteome Res* 2011;10:2113–2122.
- 27 Ogburn RN, Randall TA, Xu Y, Roberts JH, Mebrahtu B, Karnuta JM, Rider SD, Kissling GE, London RE: Are dust mite allergens more abundant and/or more stable than other *Dermatophagoides pteronyssinus* proteins? *J Allergy Clin Immunol* 2017;139:1030–1032.
- 28 Thomas WR: Innate affairs of allergens. *Clin Exp Allergy* 2013;43:152–63.
- 29 Tovey ER, Liu-Brennan D, Garden FL, Oliver BG, Perzanowski MS, Marks GB: Time-based measurement of personal mite allergen bio-aerosol exposure over 24 hour periods. *PLoS One* 2016;11:e0153414.
- 30 Bordas-Le Floch V, Le Mignon M, Bussieres L, Jain K, Martelet A, Baron-Bodo V, Nony E, Mascarell L, Moingeon P: A combined transcriptome and proteome analysis extends the allergome of house dust mite *Dermatophagoides* species. *PLoS One* 2017;12:e0185830.
- 31 Santos da Silva E, Asam C, Lackner P, Hofer H, Wallner M, Silva Pinheiro C, Alcântara-Neves NM, Ferreira F: Allergens of *Blomia tropicalis*: an overview of recombinant molecules. *Int Arch Allergy Immunol* 2017;172:203–214.