Genomic and Proteomic Characterisation of the European House Dust Mite, *Dermatophagoides pteronyssinus*

Rose Waldron BSc

Maynooth University
National University of Ireland Maynooth

Thesis submitted to Maynooth University

for the degree of Doctor of Philosophy

June 2019

Supervisor

Prof. Sean Doyle,

Department of Biology,

Maynooth University

Co- supervisor

Dr David Fitzpatrick,

Department of Biology,

Maynooth University

Co-supervisor

Dr Natasha Gordon,

Head of Research,

airmid healthgoup ltd

Head of Department

Prof. Paul Moynagh

Maynooth University

Table of Contents

List of Tables

List of Figures

Acknowledgements

I am profoundly grateful to my supervisor Professor Sean Doyle and co-supervisor's Dr Natasha Gordon and Dr David Fitzpatrick, for allowing me the opportunity to participate in this project, for their thoughtful mentorship, encouragement and enthusiasm which helped me to stay on-track.

I owe deep gratitude to Professor Sean Doyle, whose guidance and support carried me from the early days of my scientific career to completing this thesis, and through the many highs, lows and babies along the way. I cannot begin to express my thanks to Dr Natasha Gordon for sharing my weird passion for all things house dust mite related, and for being an advocate for me and our research throughout this PhD. I could not have completed this thesis nor the research it presents without the expert guidance from Dr David Fitzpatrick for which I am immensely grateful.

I would also like to thank the members of my supervisory committee, Dr Christine Griffin and Dr Paul Dowling, for their guidance with this project. I am thankful to the Irish Research Council who provided financial support for this research under the Enterprise Partnership Scheme (EPSPG/2014/64). I would also like to thank airmid healthgroup Ltd, Dublin, Ireland who co-funded this project. Mass spectrometry facilities were funded by a Science Foundation Ireland infrastructure award to SD (SFI 12/RI/2346(3)). I would like to take this opportunity to acknowledge the DJEI/DES/SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support.

Many thanks to all the postgraduate and post doctorate members of the Biotechnology Laboratory, past and present, for their advice, mentorship and encouragement (Luke O'Shaughnessy, Gráinne, Lara, Rebecca, Kevin, Stephen, Aimee, Eoin Mulvihill, Eoin O'Connor, Cindy). Special thanks to Jamie McGowan for his bioinformatic expertise, which contributed greatly to the research presented in this thesis. To the "Quad squad" and "Anyone for Nuggets" for the many laughs, lunches and coffee breaks. Thanks to my office buddies; Dearbhlaith, Sandra, Ciera, Betim, Felipe and Beverly for your friendship and many sanity-saving chats

Special thanks to all the airmid team past and present (Bruce, Maira, Jake, Vivienne, Tracy, Claire, Graeme, Siobhain, Shola, Naga, Dave, John Carrol, John Fallon, Maiklas, Angela, Sarah, Evelina & Mariarita) for their assistance throughout this project.

I would like to extend my sincere thanks to all the technical staff in the Biology department, for their help throughout my time in Maynooth University. Special thanks to Nick for fixing the AKTA and everything else I broke. Thanks to Caroline for her assistance with running MS samples and making every effort to find a slot to run my samples at short notice. I would like to take this opportunity to thank Dr Rebecca Owens for her technical guidance with Blast2GO and protein purification.

Heartfelt thanks to Dr Linan Yu for his friendship and encouragement throughout the course of this PhD, and for his molecular biology and protein purification pep-talks. Thanks to Elizabeth Smith, for her friendship, for babysitting Mason and pretending he was well behaved, and for the many polyacrylamide gels she made for me during my pregnancy. To my old friend Daragh Cuskelly for his expertise in all things molecular biology and beer related. I am especially grateful to my lab brother and sister, Nicola Moloney and Niall Conlon, for their friendship and unwavering support during those long years in the lab.

I am grateful to my siblings (Grace, Peter, Andrew, Kieran, Bernadette, Jacinta, Martin, Theresa, Joseph, Anna & Sharon) who have provided me encouragement and emotional support throughout my life, without which I would have never had the confidence to set out on this journey. I am also grateful to my extended family members and friends who have supported me along the way. To my two crazy boys, Mason and Myron, who I love more than I ever imagined possible. Thanks for keeping me grounded and always putting a smile on my face.

To my mums, Janet Waldron and Angela Stuart. Who sadly passed away before seeing me complete this thesis. You were with me in spirit through every setback and accomplishment, and continue to inspire and shape me long after you left this world. I aspire to be as good a mother, friend and mother in law as you both were.

I owe the greatest debt of gratitude to my husband and dear friend, Ben Waldron, for his compassion and understanding of the many turns my life has taken. For the unconditional support and love that he has given me from that fateful day we met on a bus in New Zealand. For the many sacrifices you made to support our little family so that I could complete this PhD, for all this and much more, I dedicate this thesis to him.

Declaration of Authorship

This thesis has not previously been submitted in whole or in part to this or any other University for any other degree. This thesis is the sole work of the author, with the exception of the bioinformatic genome assembly and phylogenetic analysis conducted by Dr David Fitzpatrick and Jamie McGowan.

_____

Rose Waldron, BSc

Publications and Presentations

Publications

Sipos, G., Prasanna, A.N., Walter, M.C., O'Connor, E., Balint, B., Krizsan, K., Kiss, B., Hess, J., Varga, T., Slot, J., Riley, R., Boka, B., Rigling, D., Barry, K., Lee, J., Mihaltcheva, S., LaButti, K., Lipzen, A., **Waldron, R**., Moloney, N.M., Sperisen, C., Kredics, L., Vagvolgyi, C., Patrignani, A., Fitzpatrick, D., Nagy, I., Doyle, S., Anderson, J.B., Grigoriev, I.V., Guldener, U., Munsterkotter, M., and Nagy, L.G. (2017) Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungi Armillaria. *Nat Ecol Evol* **1**: 1931-1941.

Waldron, R., McGowan, J., Gordon, N., McCarthy, C., Mitchell, E.B., Doyle, S., and Fitzpatrick, D.A. (2017) Draft Genome Sequence of Dermatophagoides pteronyssinus, the European House Dust Mite. *Genome Announcements* **5**: e00789-00717.

Waldron, R., McGowan, J., Gordon, N., McCarthy, C., Mitchell, E.B., and Fitzpatrick, D.A. (2019) Proteome and allergenome of the European house dust mite Dermatophagoides pteronyssinus. *PLOS ONE* **14**: e0216171.

Poster Presentations

Who is Who in house dust? Metaproteomics of House dust reveals major allergenic players and their metabolic activities within this poorly understood ecosystem. **Annual Biology Research Day 2018**. Maynooth University.

In Silico Genome Directed Investigation of the *Dermatophagoides pteronyssinus* Proteomic Response to Desiccation. **International Conference of Entomology 2017**. Florida, USA.

In Silico Genome Directed Investigation of the *Dermatophagoides pteronyssinus* Proteomic Response to Desiccation. **Annual Biology Research Day 2016**. Maynooth University. Awarded: Best Poster Presentation (runner up).

The first draft genome of *Dermatophagoides pteronyssinus* will allow studies of house dust mites to embrace the omics era. **Annual Biology Research Day 2015**. Maynooth University. Awarded: Best Poster Presentation (runner up).

Oral Presentations

The first draft genome of *Dermatophagoides pteronyssinus* will allow studies of house dust mites to embrace the omics era. **Biology Departmental Lunch Time Seminars**. Maynooth University. 2016

Characterization of the European house dust mite *Dermatophagoides pteronyssinus*. **Biology Departmental Lunch Time Seminars**. Maynooth University. 2015.

Abbreviations

a.a:	Amino Acids
AH:	Allergen Homolog
BP:	Bootstrap Support
EC:	Enzyme Code
E-Value:	BLAST Expected Value
FDR:	False Discovery Rate
GH:	Glycoside hydrolases
GH16:	Glycosyl hydrolase family 16
GO:	Gene ontology
GO BP:	Gene Ontology Biological Processes
GO CC:	Gene Ontology Cellular Component
GO MF:	Gene Ontology Molecular Function
HD:	House Dust
HDM:	House Dust Mites
HDMMM:	House Dust Mite Maximal Media
HGT:	Horizontal Gene Transfer
IPS:	InterPro Scan
LC-MS/MS:	Liquid Chromatography Mass Spectrometry Mass Spectrometry
LEAPdb:	Late Embryogenesis Abundant Protein database
LFQ:	Label Free Quantification
MB:	Mite Body
MES:	4-Morpholineethanesulfonic acid
MGA:	Mite Group Allergen
MGAH:	Mite Group Allergen Homolog
MM:	Minimal Media
NCBI:	National Centre for Biotechnology Information
PBS:	Phosphate Buffer Saline
PCR:	Polymerase Chain Reaction
PP:	Predicted Proteome
RBH:	Reciprocal Best Hit
RP:	Representative Proteome
RT:	Room Temperature
SDS-PAGE:	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SM:	Spent Culture Medium
TAE:	Tris-Acetate

Summary

House dust mites are major causative agents in the pathogenesis of allergy. Their proximity with human habitats, association with development of allergenic diseases, and resistance to physical and chemical control measures; make them some of the most medically important mites. Understanding of house dust mites has been hampered by a lack of genomic sequence data and limited to a discrete number of proteins. The work presented here is a detailed characterisation of the European house dust mite, *Dermatophagoides pteronyssinus* airmid strain, at the genomic and proteomic level. Genomic sequencing and assembly resulted in a high-quality assembly of 70.76 Mb in size with 96.86% coverage. A comprehensive bioinformatic and proteomic examination was conducted on the 12,530 predicted proteins, validating the expression of 4,002. A small group of *D. pteronyssinus* airmid proteins showed significant homology to known allergens from other species. A large scale comparative proteomic investigation of the mite body and spent growth medium allowed for: (i) qualitative assessment of allergen localisation and (ii) the identification of numerous enzymes that may be involved in key physiological activities. The characterisation of protein extracts from house dust also identified a substantial number of uncharacterised *D. pteronyssinus* proteins in addition to known and putative allergens. The genes encoding novel β-1,3 glucanases were identified within a trigene cluster in *D. pteronyssinus* airmid. Recombinant protein expression, biochemical and proteomic analysis revealed Glu1 and Glu2 to exhibit hydrolytic activity toward β-1,3 glucans and have increased expression in the mite body and excretome of *D. pteronyssinus* in response to yeast diet. Further proteomic and enzymatic analysis correlated glucanase activity in house dust with presence of Glu1 and Glu2. These findings provide evidence that active β-1,3 glucanases are expressed and excreted in the faeces of *D. pteronyssinus* in response to fungal diet, in both the laboratory and the wild-type environment.

Preface

This thesis is presented in a "Thesis by Published Works" format, comprising of two peer reviewed publications and one manuscript currently under peer review. A review of relevant literature is presented in Chapter 1 and concludes with a discussion of knowledge deficiencies and aims of the research presented. The first results chapter, Chapter 2, comprises of a short publication followed by supplementary materials, methods and results, describing the methodology used to successfully sequence the *D. pteronyssinus* airmid genome. Chapter 3 is a large publication, it extensively describes the predicted proteome, allergenome and proteome of laboratory-reared *D. pteronyssinus* airmid with reference to the proteome of wildtype *D. pteronyssinus*. Published versions of these manuscripts are available in appendix 3. Chapter 4 forms a submitted publication, describing the discovery and recombinant expression of a trigene cluster coding for putative β-1,3 glucanases that may have a role in the digestion of fungi. This thesis concludes with a general discussion, Chapter 5, in which the research presented in chapters 2 - 4 is summarised and discussed in the context of its contribution to the field of house dust mite biology and allergy.

Throughout this work a distinction will be made between "*D. pteronyssinus*" the species and "*D. pteronyssinus* airmid" the strain used in this study. *D. pteronyssinus* will be used in terms of a general discussion relating to the species and *D. pteronyssinus* airmid when describing proteins or genes identified from extracts derived directly from the airmid strain of *D. pteronyssinus*.

1.0 Introduction

1.1 House Dust Mite Phylogeny

The term house dust mite (HDM) typically refers to the most common house dust (HD) dwelling pyroglyphid mites associated with allergy: *Dermatophagoides pteronyssinus*, *Dermatophagoides farinae* and *Euroglyphus maynei*; however, other mites associated with household dusts are also known to cause allergy and are of significant medical and economical importance (Colloff, 2009). HDM belong to the Acari suborder Astigmata, containing more than 6,100 species from free-living and parasitic lineages. Their unusual evolutionary history has been used to demonstrate the reversible evolution of parasitism (Klimov & O'Connor, 2013). HDMs appear to have evolved from an avian parasite ancestor, which in turn had evolved from a free-living ancestor (Klimov & O'Connor, 2013). Some entomologists suggest this free-living ancestor was a mycophagous mite that originally inhabited soil, but transitioned into birds' nests, thus explaining the continued ability of several synanthropic mite species to digest fungi (O'Connor, 1979). Phylogenetic studies indicate that the closest parasitic relatives to *Dermatophagoides* spp. are bird parasites from the *Onychalges* spp. and the sheep scab mite *Psoroptes ovis* (Klimov & O'Connor, 2013). Morphological studies have placed *D. pteronyssinus* and *D. farinae* into the same genus, however recent phylogenetic studies have observed that *D. pteronyssinus* and *E. maynei* appear more closely related to each other than to *D. farinae* (Klimov *et al.*, 2015, Cui *et al.*, 2010).

The *D. pteronyssinus* life cycle, duration from egg to adult, is variable depending on the hygro-thermal conditions in which the mites are grown, ranging between 15 and 122 days (Arlian *et al.*, 1990). A typical life cycle for *D. pteronyssinus* is illustrated in Figure 1.1 A & B, showing the development of an egg to adult under controlled laboratory growth conditions of 23ºC and 75% relative humidity (RH). Larvae emerge from eggs after approximately 8 days, they have three pairs of legs and measure 170 µm in length. The other three life stages have four pairs of legs and can be most easily differentiated by their size. Upon reaching maturity, adults live for about one month, making the entire life span of *D. pteronyssinus* between 4 - 6 weeks (Colloff, 2009). Adult females are larger, measuring 350 µm in length and weighing 5.8 µg, compared to males which are typically

270 µm in length and weigh 3.48 µg (Arlian *et al.*, 1990, Colloff, 2009). The male to female sex ratio of pyroglyphids is typically 1:1, this is assumed to also be the case for *D. pteronyssinus* (Colloff, 2009).

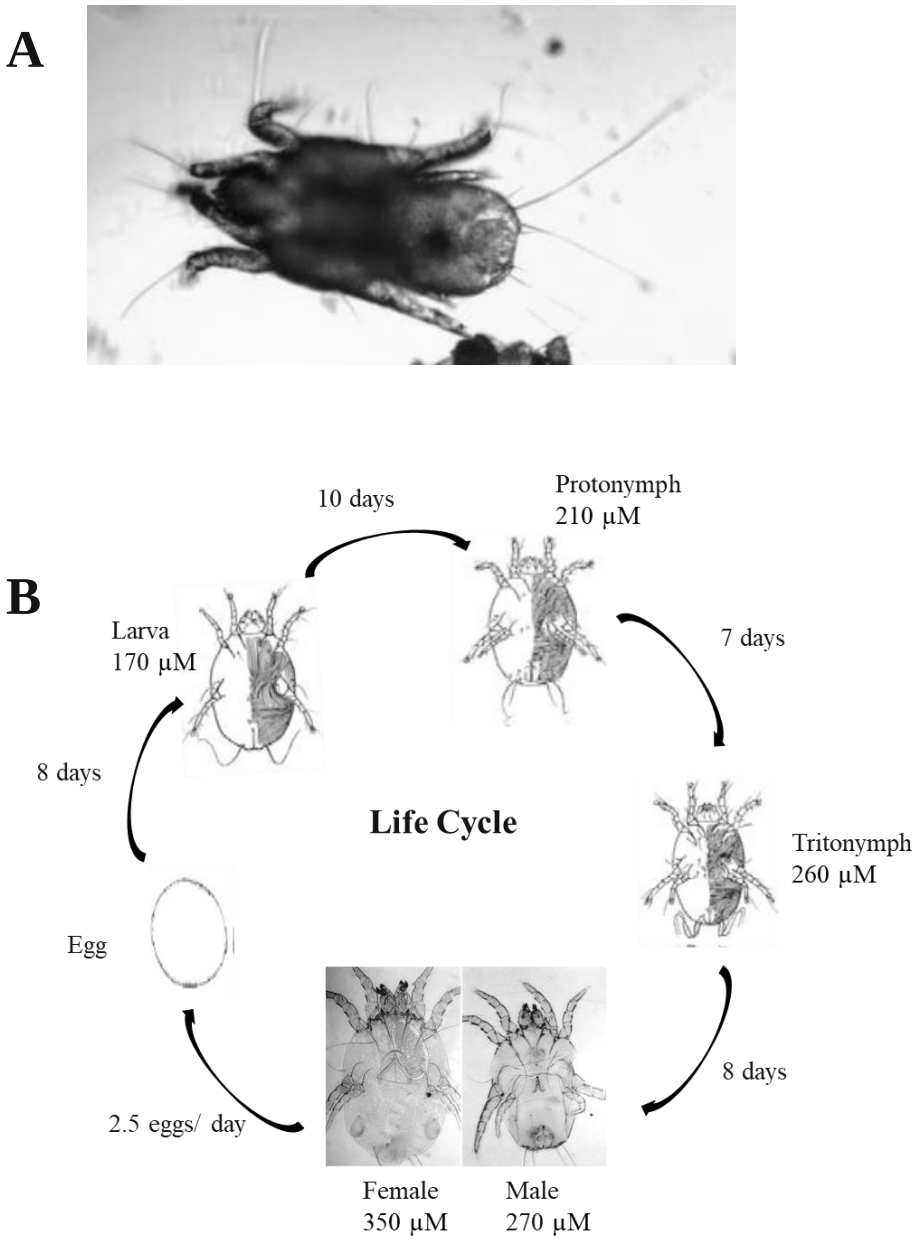


Figure 1.1 *D. pteronyssinus* physical characteristics and life cycle. A. Microscope image of *D. pteronyssinus* airmid with visible setae. B. Depiction of the five stages of *D. pteronyssinus* life cycle at 23ºC and 75% RH. Data: Arlian *et al.* (1990). Image: adapted from Colloff (2009).

1.2 Hygrothermal Requirements

The level of HDM contamination between residences varies enormously. The abiotic factors facilitating large populations in individual homes are not fully understood, however hygrothermal conditions are thought to be a significant factor (De Boer *et al.*, 1998, Arlian, 1992). Water availability in the form of RH is critical in dictating duration of the HDM life cycle, fecundity, survival, feeding and allergen production (Arlian *et al.*, 1998, Colloff, 1992).

HDMs are typically composed of 75% water by weight, they actively acquire moisture by extracting water from the air (Arlian, 1992). However, with a large surface area to volume ratio they also lose water rapidly when humidity is low. In the laboratory, to ensure optimal growth and avoid water loss, HDMs must be maintained above a set RH, however cultures exposed to sub-optimal RH have shown long-term survival (Ghazy & Suzuki, 2014, Arlian, 1992). The protonymph stage is highly resistant to desiccation, surviving several months at 0% RH (Arlian, 1992). It has been shown that short increases in RH can substantially promote survival. *D. pteronyssinus* held below 10% RH showed substantially higher rates of survival when transiently exposed to high humidity (90% RH; 1.5 hours), increasing exposure to three hours was sufficient for them to reproduce (deBoer & Kuller, 1997).

This is relevant for understanding HDM population growth within the fluctuating hygrothermal conditions of the home (Crowther *et al.*, 2006, Cunningham, 1998). Recent advances in house construction, insulation and central heating have resulted in warmer, drier homes, that may be less favourable to HDM colonisation (Hart & Whitehead, 1990, Crowther *et al.*, 2006). However, humidity within the home is not constant, as domestic activities such as cooking and bathing result in temporary humidity increases, therefore average indoor RH is a poor indicator of HDM survival or growth (De Boer *et al.*, 1998, Niven *et al.*, 1999). Although *D. pteronyssinus* can be killed by extended exposure to low RH, reduction of RH in the home does not always lead to a reduction in mite numbers or levels of allergen (Hyndman *et al.*, 2000, Niven *et al.*, 1999). It is evident from these studies that HDMs possess mechanisms to resist and tolerate periods of desiccation. The

ability of mites to withstand periods of environmental extremes is key to their continued persistence in the home (Colloff *et al.*, 1992, Calderón *et al.*, 2015).

Desiccation resistance and cold/freeze tolerance are intrinsically linked. Invertebrate studies have shown both extremes to elicit the same protective biochemical responses. Both stresses cause: changes to pH and osmolarity of cellular fluids, decreased volume of haemolymph and in extreme cases lead to cell shrinkage and ultimately death (Worland & Block, 2003). The mechanisms by which the animal responds to such stresses are equally similar, requiring osmo-protection, stabilisation of proteins and membrane structure. Prior exposure to desiccation has been shown to offer protection against cold and vice versa (Worland & Block, 2003, Sinclair *et al.*, 2013). This may explain why HDMs can survive sub-zero temperatures without prior acclimation. Clinicians frequently recommend freezing soft toys and pillows etc, at regular intervals to kill HDMs and reduce allergens (Arlian & Platts-Mills, 2001). However, freezing only reduces the number of live HDMs, with an average of 17 HDMs per toy surviving (Chang *et al.*, 2011).

Molecular and biochemical studies of desiccation tolerant organisms have revealed the Late Embryogenesis Abundant Proteins (LEAPs) play a key role in their survival. LEAPs were first identified and are best described in plants, however recent studies have identified LEA-like genes in nematodes and arthropods (Kikawada *et al.*, 2006, Solomon *et al.*, 2000, Browne *et al.*, 2004). Clear and consistent classification of LEAPs into classes is lacking; to date, eight sub-families have been described in the Pfam database, however inconsistencies still exist (Hunault & Jaspard, 2010).

Accurate *in silico* prediction of LEAPs in arthropods is hampered as only a handful of LEAPs have been described in arthropods (Hand *et al.*, 2011). Despite this, utilising known LEAPs to search for homologous proteins in other organisms may identify proteins involved in desiccation and freeze resistance (Du *et al.*, 2013).

1.3 HDM Allergy

Atopy is the tendency to produce IgE antibodies in response to ordinary exposure to allergens, it is heavily influenced by genetics, however defined genetic diagnostic criteria are lacking (Asher *et al.*, 2012). The atopic diseases most associated with HDM allergy are Atopic dermatitis (Eczema), Allergic Rhinitis and Asthma (Sanchez-Borges *et al.*, 2017). It is difficult to discuss one of these atopic diseases without drawing information from the others. Atopic individuals have a propensity towards allergenic diseases, frequently suffering from multiple diseases, this is referred to as the atopic march. The atopic march typically begins in infancy with Atopic Dermatitis (AD), then progressing to Allergic Rhinitis (AR) and Asthma in childhood (Bantz *et al.*, 2014). A study of 100 patients with AD, revealed that the majority (76%) also had Asthma and/or AR, while 39% had all three diseases (Banerjee *et al.*, 2015).

1.3.1 Atopic Dermatitis

Most AD patients have elevated IgE against HDM allergens, sensitisation accompanies early signs of AD but is not causative of AD (Sanchez-Borges *et al.*, 2017). AD is an inflammatory skin condition associated with skin barrier dysfunction, often as a result of a filaggrin mutation (Brown, 2016). Filaggrin is expressed in keratinocytes and functions to maintain skin barrier pH, hydration and antimicrobial protection. In particular, it inhibits HDM phospholipase A2. Without filaggrin, HDM phospholipase A2 generates antigenic neolipids, which activate T-cells causing inflammation (Jarrett *et al.*, 2016). AD skin lesions contain T-cells that recognise Der p 1, this proteolytic allergen has been demonstrated to also disrupt skin barrier function (Sanchez-Borges *et al.*, 2017). Moreover, higher numbers of HDMs are recovered from the skin scrapings of patients with AD compared to healthy individuals, indicating a transdermal route of allergen exposure (Teplitsky *et al.*, 2008). We spend approximately eight hours per day in bed, with our skin in direct contact with bedding materials. In HDM sensitised patients, particularly those with AD, bedding and clothing provide a means for HDMs to come into direct contact with the skin and cause transdermal sensitisation (Teplitsky *et al.*, 2008, Clarke *et al.*, 2015a, Tovey *et al.*, 1995).

Distinct patterns of HDM sensitisation are seen for patients who have AD symptoms only compared to those with respiratory symptoms only. Of the 10 *D. pteronyssinus* allergens examined, AD patients bound Der p 10, 11, 14 and 18 with higher frequency (2.5 – 10 times) than patients with respiratory symptoms only (Banerjee *et al.*, 2015). This study serves to highlight that the type of allergens a patient reacts to is altered by exposure route, with AD patients being transdermally exposed to different repertoire of allergens than patients with respiratory symptoms who are exposed via inhalation.

1.3.2 Asthma

Asthma is characterised by chronic airway inflammation, episodes of wheezing, breathlessness, chest tightness and coughing (Sanchez-Borges *et al.*, 2017). Ireland has the fourth highest prevalence of asthma in the world, estimated at 10% the population and recent statistics indicates the prevalence is increasing (NHQRS, 2017). The Asthma Society Ireland estimate that one person every week dies in Ireland as a result of asthma, in the UK this is as much as 1,200 per year (BLF, 2016). The economic burden of asthma is immense, accounting for up to 2% of total healthcare expenditure in developed countries (Asher *et al.*, 2012).

The mechanisms underlying the pathophysiology of asthma are complex, however it is now widely accepted that the airway epithelium plays a central role in orchestrating the physiological manifestations of asthma (Chanez & Bourdin, 2008). Asthmatic episodes are triggered by bronchoprovocation by common allergens, which are bound by IgE, triggering degranulation of mast cells. Immunomodulatory molecules released during degranulation results in airway narrowing, mucus secretion and infiltration of immune cells (Holgate, 2013). A key feature is the disruption of the bronchial epithelium, this allows the entry of the proinflammatory allergens that illicit local immune response and inflammation (Figure 1.2). Several major HDM allergens, including Der p 1, have been shown to directly disrupt the airway epithelium (Wang, 2013, Gregory & Lloyd, 2011). It is estimated that adults inhale 0.8 mg of dust every day and children as much as 2 mg (NIVM, 2008). This is clinically relevant if the dust is heavily HDM infested, as Der p 1 content in bronchoalveolar lavage fluid of asthmatic patients has been demonstrated to

correlate with concentrations in the environment (Ferguson & Broide, 1995). It is not surprising therefore that IgE hyperresponsiveness to HDM allergens is one of the most important asthma associated phenotypes and a risk factor for other allergenic diseases (Sanchez-Borges *et al.*, 2017). The prevalence of HDM sensitisation in asthmatics is enormous, estimated at 85% (Custovic *et al.*, 1996).

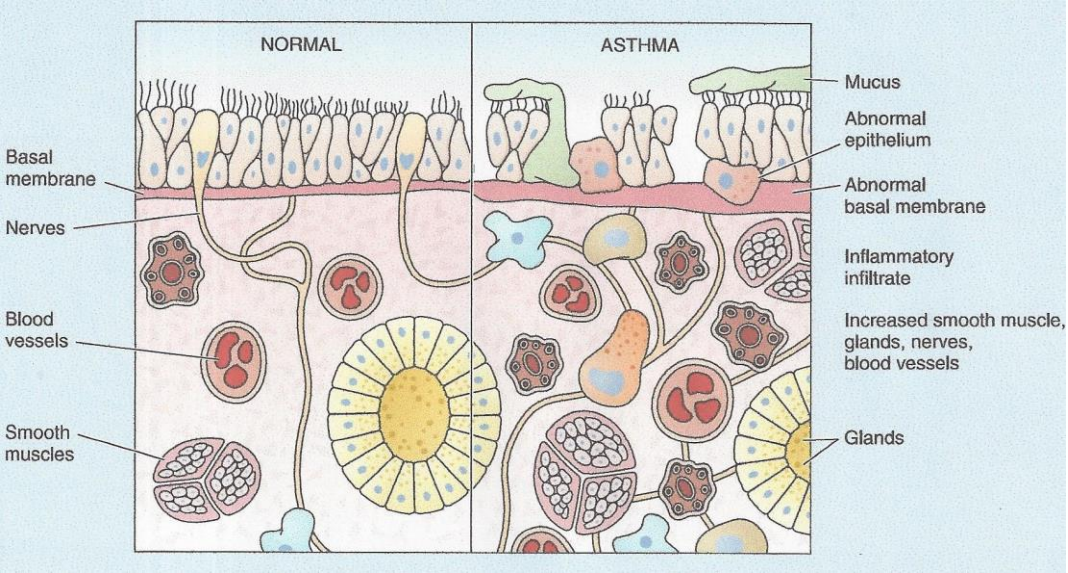


Figure 1.2 Illustration of endobronchial biopsy comparing normal physiology to that seen in asthma. Asthmatic bronchia depicts bronchial hypersecretion resulting in increased luminal mucus. Fragile epithelium showing loss of mechanical and biochemical barrier exposing the basement membrane which becomes enlarged and densified. Infiltration of immune cells (mast cells, eosinophils, neutrophils, lymphocytes and monocytes) into underlying tissues. Taken from Chanez & Burdin, 2008.

1.3.3 Allergic Rhinitis

Asthma and AR are common co-morbid diseases with overlapping symptoms, 42% of asthmatics also have AR (Steppuhn *et al.*, 2013). AR is associated with nasal mucosa inflammation resulting in symptoms of nasal congestion, sneezing, and itching of the eyes (Rowland-Seymour, 2010). As with other atopic diseases, HDM sensitisation is prevalent in patients with AR, estimated at 50-90% of AR patients, and of these 50-80% are polysensitised to HDM and non-HDM allergens (Sanchez-Borges *et al.*, 2017).

1.3.4 Allergen Exposure and Sensitisation

There is clear evidence for a bell-shaped dose response curve for sensitisation to HDM allergens, with high and low concentrations being protective (Sanchez-Borges *et al.*, 2017). Repeated exposure is associated with sensitisation and progression to asthma in susceptible individuals. However avoidance of allergens in infancy appears to only delay allergy onset rather than prevent it (Boner *et al.*, 2002). In 1989, The First International Workshop on Indoor Allergens and Asthma was held, here it was agreed by international experts that allergen levels of ≥2 μg Der p 1 per gram of dust are sufficient to induce sensitisation, while levels of ≥10 μg Der p 1 per gram are seen as a risk factor for the development of acute asthma (Platts-Mills et al., 1989).

1.3.5 Storage Mite Allergy

Aside from HDMs, other allergenic mites are found within household dust, these are typically referred to as storage mites due to their association with stored products (Colloff, 2009). Storage mites with listed allergens in WHO/USIS are as follows: *Acarus siro* (Aca s; *n* = 1), *Blomia tropicalis* (Blo t; *n* = 14), *Glycyphagus domesticus* (Gly d; *n* = 1), *Lepidoglyphus destructor* (Lep d; *n* = 5), *Tyrophagus putrescentiae* (Tyr p; *n* = 8) and *Chortoglyphus arcuatus* (Cho a; *n* = 1). These mites are significant elicitors of occupational allergy for farm, warehouse and granary workers, however in the home their role in allergy is poorly understood (Colloff, 2009). All the above-mentioned storage mites have been identified in household dusts, though in lower numbers compared to HDMs (Wraith *et al.*, 1979, Boquete *et al.*, 2006, Clarke *et al.*, 2016). While certainly present in HD, storage mites do not dominate this habitat, their allergenicity (excluding occupational illness) is not a result of their high abundance or allergen producing ability, but most likely a result of cross-reactivity with HDMs. A study examining co-sensitisation to HDM and storage mites found 88% of *D. pteronyssinus/D. farinae* allergic patients had skin prick positive results to one or more storage mite (*A. siro*, *L. destructor* & *T. putrescentiae*) and 73% were allergic to all three storage mites used in the study (Vidal *et al.*, 1997).

1.4 House Dust Mite Habitats

1.4.1 Geographical Distribution of House Dust Mites

Collectively, *D. pteronyssinus*, *D. farinae*, and *E. maynei* account for 90% of all HD dwelling Acari (Arlian *et al.*, 1992). HDMs are found in most homes globally, a few countries have a species bias based on their hygrothermal tolerances (Colloff, 2009). Australia, New Zealand and England tend toward *D. pteronyssinus* domination, while most of South Korea, U.S.A and Italy have *D. farinae* biased populations. Many countries have a regional species bias, but most exhibit mixed populations of HDM species (Thomas, 2010).

The first comprehensive analysis of mite species in Irish homes found that of the 19 different mite species identified, *D. pteronyssinus* was the principal mite present. It was found in 59% of all samples. *E. maynei* was identified in only 5% of Irish HD samples and only 1% of samples contained *D. farinae* (Clarke *et al*., 2016).

1.4.2 Microhabitats

HDMs are found in areas of the home that are used most frequently and where shed skin tends to accumulate, such as mattresses, carpets, pillows, bed clothes and soft toys (Nagakura *et al.*, 1996, Arlian, 1975, Arlian *et al.*, 1992, Mitchell *et al.*, 1969).

The type of bedding material used in the home may be a factor in determining the size of HDM populations. One study showed feather pillows to harbour fewer HDMs than synthetic ones, possibly due to the encasing of feather pillows restricting HDMs entering the filling (Rains *et al.*, 1999). Foam mattresses may harbour up to eight times more HDM faeces than spring mattresses (Schei *et al.*, 2002). Similar studies contradicted this, finding the converse or no difference (Mills *et al.*, 2002, van den Bemt *et al.*, 2006). These contradictory findings highlight the need for more studies which implement standardised methodologies to quantify HDM presence in the home and further the understanding of HDM habitat requirements.

1.5 House Dust Mite Digestion and Diet

1.5.1 House Dust Mite Digestive Physiology

HDMs ingest food using chelicerae, a pair of appendages that allow food to be pushed into the mouth. Food is then passed from the mouth, through the pharynx, where it is mixed with saliva before passing into the cuticle lined oesophagus which links the pharynx to the anterior midgut, the site of nutrient utilisation (Figure 1.3A). Here in the anterior midgut, cells bud off from the gut lumen and are thought to burst open releasing digestive enzymes (Figure 1.3B). The food then moves to the posterior midgut where it enters the peritrophic envelope, which is thought to be formed by components secreted by secretory cells of the anterior midgut. Here in the posterior midgut, the food bolus swells in mass and the sleeve-like peritrophic envelope is released from the gut epithelium and floats free (Figure 1.3A). More digestive enzymes enter the envelope and digest food; water and small digested fragments pass out of the peritrophic envelope, leaving behind undigested material. The peritrophic envelope then moves to the hindgut, where it contracts, changes colour and becomes the faecal particle that is then excreted by the mite into the surrounding dust. It is because of this faecal particle, that HDMs are prolific elicitors of allergy, these faecal particles are abundant in many potent digestive enzymes (Colloff, 2009).

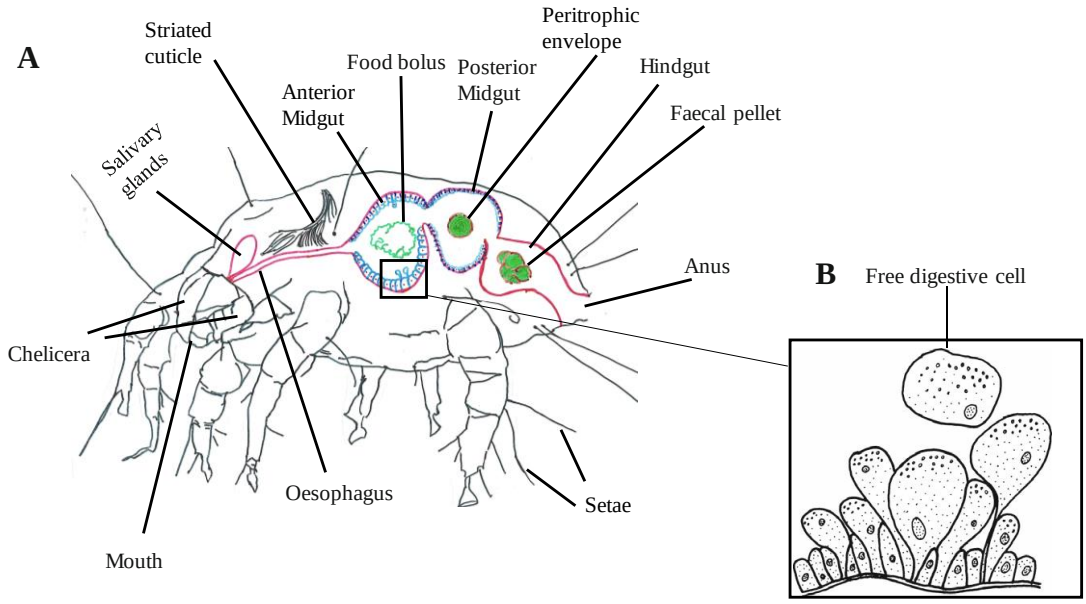


Figure 1.3 Illustrations of HDM digestive physiology. Adapted from Colloff, 2009. **A**. Illustration of *D. pteronyssnius* with simplified depiction of digestive system **B**. magnified view of budding digestive cells of the anterior midgut.

1.5.2 Composition of House Dust

HD is an amalgamation of particulate matter derived from both abiotic and biotic components present within the home and the homes' surroundings (Rintala *et al.*, 2012). This complex and highly variable matrix is composed of skin and hair from occupant and pets, fabric fibres, insect fragments, construction material, food particles, soil and plant material (Colloff, 2009). It is not an inert matrix, it provides ample nutrition to facilitate growth of diverse and highly variable fungal, bacterial and insect communities (Korpi *et al.*, 1997, Dannemiller *et al.*, 2017, Clarke *et al.*, 2016, Barberan *et al.*, 2015).

It is pragmatic to think of HD as a highly complex and variable ecosystem that is heavily influenced by the homes' occupants, unique features (heating, ventilation, age etc) and its geographical location (Barberan *et al.*, 2015, Wu & Lewis, 2013). Combined exposure to the numerous components mentioned above could have the potential to exacerbate allergy and contribute to the complex disease pathophysiology observed in atopic individuals (Gregory & Lloyd, 2011). How these allergenic components interact with each other in the dust ecosystem or act upon the immune system is largely unknown; research into this topic may explain why studies which focus on eliminating a single allergen source show limited success (Gøtzsche & Johansen, 2008). Moreover, the objective characterisation of allergenic components present in HD is lacking and was featured as one of the "unmet needs in mite allergy research" in the 2017 international consensus (ICON) report on the clinical consequences of mite hypersensitivity (Sanchez-Borges *et al.*, 2017).

1.5.3 House Dust Mite Diet

In the laboratory, *D. pteronyssinus* have been grown on diverse culture media including various combinations of wheat bran, wheat flour, dog food, rodent chow, ground porcine liver, dried egg powder, yeast, defatted skin scales and fish food (Andersen, 1991, Hubert *et al.*, 2016, Eraso *et al.*, 1997, Arlian & Morgan, 2015, Hart *et al.*, 2007). The ability of HDMs to grow on diverse nutrient sources within the laboratory indicates they can utilise similar nutrient sources in a wildtype setting. Moreover, their inability to move to new habitats in search of more favourable food would suggest it is advantageous for HDM to be trophic generalists, feeding on a variety of organic debris and microbes associated with their human proximity habitats (Colloff, 2009).

The diet of HDMs in the home is largely unknown and much of what is described in the literature dates back to observations made by early researchers (Colloff, 2009). In 1973, Van Bronswwijk examined the alimentary canal of 147 *D. pteronyssinus* specimens recovered from mattress and floor dust, observing pollen, fungal mycelia and spores, bacteria and plant fibres in the gut contents. Other similar studies have identified skin, cuticle fragments and insect fragments from HDM gut contents (Colloff, 2009).

There have been no laboratory studies to date that have extensively studied the mite carrying capacity of HD (Colloff, 2009). Given the high degree of HD composition variability, it is difficult to generate a model that is applicable to all homes. What little we know about the nutritional value of HD for HDMs survival, stems from a handful of studies (Colloff, 2009). It is evident that the protein content of HD is an important factor in determining the carrying capacity, as protein content below 110 mg/gram dust was found to be insufficient to HDM populations (Koren & Eckhardt, 1995).

The HD itself or as a consequence of being a microbial reservoir, is nutritionally sufficient to support large HDM numbers, some researchers report as many as 318 HDMs in just 100 mg of dust (Hart & Whitehead, 1990, Platts-Mills *et al.*, 1986). Epidemiological studies estimate that 100 mites per gram of house dust correspond with 2 µg of Der p 1, and 500 mites with 10 µg (Platts-Mills *et al.*, 1986). Based upon these estimations it is not difficult to understand how relatively few *D. pteronyssinus* can produce hazardous levels of allergens.

It is important to bear in mind that the composition of HD may alter allergen production, as laboratory studies show that different diets alter protein and allergen production (Avula-Poola *et al.*, 2012, Casset *et al.*, 2012, Vidal-Quist *et al.*, 2017). Studies utilising single allergens from a limited number of species, for measurement of HDM presence may not provide an accurate reflection of HDM contamination or allergen exposure. This issue was highlighted as an "unmet need in mite allergy research". Development of objective methods and high-throughput strategies to identify allergens and species present in HD is needed to better understand the pathogenesis of allergic diseases (Sanchez-Borges *et al.*, 2017).

1.5.4 Microbial Content of House Dust

Much like HDMs, fungi and bacteria flourish in warm humid environments. Traditionally, studies have used ergosterol, β-1,3 glucan and extracellular polysaccharides for quantification of fungal content in homes (Miller & Young, 1997, Douwes *et al.*, 1997, Karvonen *et al.*, 2014). Endotoxin is the most commonly used

biomarker for assessing bacterial content of HD (Taubel *et al.*, 2009). However, recent advancements in high-throughput DNA sequencing has enabled the detailed characterisation of microbiomes to the species level (Kong, 2011, Wu & Lewis, 2013). Such analysis has also been extended to cataloguing microbial species within homes.

1.5.5 Fungi Present in Homes

In HD from floors and mattresses the fungal species frequently identified by high-throughput DNA sequencing, to be of high abundance are: *Aspergillus* spp. *Cladosporium* spp., *Penicillium* spp. and *Alternaria* spp. (Kaarakainen *et al.*, 2009, Sousa *et al.*, 2014, Rintala *et al.*, 2012). The most comprehensive HD study to date included over 1,100 paired indoor-outdoor dust samples from across the continental US, identifying over 38,000 fungal taxa (Grantham *et al.*, 2015). Fungal taxa in dust samples was so geographically predictable they were a reliable means of pinpointing the sample location to a few hundred kilometres (Grantham *et al.*, 2015). Despite the biogeographical predictability of fungal taxa, the researchers were still able to identify subsets of taxa (1.1% of those identified) consistently found within the home, these included *Aspergillus*, *Penicillium*, *Alternaria* and *Fusarium* spp.

1.5.6 Mycophagy

There is no evidence to suggest that fungi are indigenous gut inhabitants of *D. pteronyssinus* (Hay *et al.*, 1992), therefore the presence of fungi in the HDM gut is due to ingestion either intentionally or unintentionally. If HDMs consume fungi as a nutritional source, mycophagy, they may have the ability to digest some species and not others. Naegele *et al.* (2013) observed that when offered a diet of six different molds, *D. farinae* preferably consumed *Alternaria alternata*, *Cladosporium sphaerospermum* and *Wallemia sebi* over other species. *Aspergillus* and *Penicillium spp.* spores have been isolated from the gut of laboratory-reared *D. pteronyssinus* and successfully cultured, indicating the spores were undamaged after digestion (VanBronswijk & Sinha, 1973). A recent study demonstrated that *D. pteronyssinus* is better suited to nutritionally exploit

fungi than *D. farinae*, and preferentially seek out *Aspergillus jensenii* and *Saccharomyces cerevisiae* to consume (Molva *et al.*, 2019).

The small size of HDMs make it logistically impossible to dissect them and distinguish between the gut lumen and fat body. Therefore presence of an enzyme in faeces is used as an indicator of a digestive function (Erban & Hubert, 2008). One avenue to explore the hypothesis of mycophagy is to examine the expression and localisation of the enzymes capable of digesting essential fungal components. In order to feed on fungi, a mite must possess the enzymatic arsenal to degrade the microbe's cell wall and digest the cell contents (Erban & Hubert, 2012). A typical fungal cell wall is a multilayer structure composed of a chitin layer located next to the cell membrane, linked to a β-1,3- and β-1,6-glucan layer via β-1,4-linkages, followed by an outermost layer of mannoproteins. β-glucans form 50-60% of the fungal cell wall with β-1,3 contributing 65-90% of the β-glucan content (Fesel & Zuccaro, 2016). Effective disruption of fungal cell walls requires the combined actions of β-1,3-1,6 glucanases, chitinases and proteases (Schiavone *et al.*, 2014). Arthropods utilise chitinases for cuticle remodelling, therefore, the presence of chitinases in itself does not indicate mycophagy; rather the location of the enzyme is key to its function. Glyco hydrolase activity has been observed in laboratory reared *D. pteronyssinus* (Martinez *et al.*, 1999). Such enzyme activity may be related to immune defence or digestive activities, the latter being indicated if the enzyme is present in faeces (Pauchet *et al.*, 2009, Erban & Hubert, 2008).

1.5.7 Other Digestive Activities

It is evident that HDMs possess the enzymatic arsenal to digest a wide variety of substances, as demonstrated by the diverse culture media used in laboratory settings to culture HDMs (Andersen, 1991, Hubert *et al.*, 2016, Eraso *et al.*, 1997, Arlian & Morgan, 2015, Hart *et al.*, 2007). Enzymes produced by HDMs are altered depending on the diet on which they are reared. For example, a diet high in protein increased production of allergens Der p 1 and Der p 4, as well as increasing elastase and chymotrypsin enzyme activities. Whereas a lipid rich diet (15.4%) increased Der p 14 allergen production,

trypsin, chymotrypsin, elastase, glutathione-S-transferase and endo-chitinase enzyme activities (Vidal-Quist *et al.*, 2017).

Spent culture medium (SM) and whole mite body extracts (MB) from *D. farinae* reared on a mixture of fish food and wheat, were demonstrated to have hydrolytic activity against wheat, potato, rice and corn starches, as well as amylopectin, maltase, dextrin and sucrase (Erban *et al.,* 2009). Digestive enzyme activities reported for *D. pteronyssinus* include 33 distinct digestive enzyme activities from HDMs reared on a variety of diets (Colloff, 2009). Casset *et al.* (2012) examined numerous commercially available diagnostic and therapeutic HDM extracts, finding many were lacking important allergens and some had several fold variations in Der p 1 to Der p 2 ratios (Casset *et al.*, 2012). These studies highlight the enormous variability in protein production by HDMs grown on different media. Characterising alterations in enzyme production can be difficult, relying on substrate assays or measurement of allergens for which there are ELISAs (Vidal-Quist *et al.*, 2017). Proteomics offers an alternative method to study enzymes and has the ability to quantify alterations in the production of hundreds of enzymes.

1.6 House Dust Mite Allergens

1.6.1 Allergen Databases and Allergen Nomenclature

Several public allergen databases are available which allow researchers to report the discovery of a new allergen. However, there are inconsistencies between these databases in the naming and official recognition of reported allergens.

The most extensive list of allergens/variants is held by Allergenome.org which lists 314 “mite” allergens/variants, 92 for *D. pteronyssinus* and 99 for *D. farinae*, the remaining are from various mite species (www.allergenome.org). Allergens included in this database are reported in scientific publications and reviewed by database administrators. Sequences for the allergen entries can be found in the protein sequence repository UniProt.

Allergenonline database, curated by University of Nebraska Lincoln, hosts full protein sequences for 2,129 allergen entries, of the 251 "mite" entries, 76 are for *D. pteronyssinus* and 88 for *D. farinae.* The allergens are listed, along with the protein sequence data and a brief description of allergenicity assessment (www.Allergenonline.org).

The official site for systematic allergen nomenclature and reporting is approved and maintained by the World Health Organisation and International Union of Immunological Societies Allergen Nomenclature Sub-committee (WHO/IUIS). This Database lists 92 mite allergens/variants, 23 allergens for *D. pteronyssinus* and 32 from *D. farinae* (Accessed: 30/04/2019). A key feature of inclusion in the WHO/IUIS database is that the allergen must satisfy strict criteria for biochemical and immunological characterisation. These criteria are set out by the WHO/IUIS material and are summarised in Table 1.1 (Chapman, 2008).

Table 1.1 Criteria for Inclusion in the WHO/IUIS Nomenclature

The molecular and structural properties should be clearly and unambiguously defined, including • Purification of the allergen protein to homogeneity. • Determination of molecular weight, pI, and carbohydrate composition. • Determination of nucleotide and/or amino acid sequence. • Production of monospecific or monoclonal antibodies to the allergen.
The importance of the allergen in causing IgE responses should be defined by; • Comparing the prevalence of serum IgE antibodies in large population(s) of allergic patients. Ideally, at least 50 or more patients should be tested. • Demonstrating biological activity, e.g., by skin testing or histamine release assay. • Investigating whether depletion of the allergen from an allergic extract (e.g., by immunoabsorption) reduces IgE-binding activity. • Demonstrating, where possible, that recombinant allergens have comparable IgE antibody-binding activity to the natural allergen.

Taken from Chapman (2008).

1.6.2 Allergen Families

The structure and function of a protein has important implications for allergenicity, most allergenic proteins are limited to just 2% of protein families (Radauer *et al.*, 2008). A comprehensive categorisation of allergens into allergen families (AllFams) was undertaken by Radauer *et al.* (2008) from which an online database was created. This AllFam database classifies allergens deposited into Allergenonline and WHO/IUIS

databases into AllFams based upon structure. The database gives details for 959 allergens belonging to 151 AllFams (Radauer, 2017).

1.6.3 Mite Group Allergens

The majority of mite allergens identified to date have been assigned to groups (Group 1-36) in accordance with their order of discovery (Thomas, 2015). The term "group" is used to describe structurally related allergens from different species within a genus or closely related genera, these typically exhibit 40 - 90% protein identity (Chapman, 2008). Much confusion still persists as to the identity of some more recently discovered allergens, with different research groups designating distinctly different proteins to the same mite allergen grouping (Rider *et al.*, 2017). To avoid any ambiguity, the work presented throughout this thesis will refer to allergen nomenclature used by the WHO/IUIS committee.

1.6.4 Allergen Isoforms and Isoallergens

Allergens frequently exhibit strong homology, and to aid in clarity and in understanding the relationships between allergens, the WHO/IUIS Allergen Nomenclature defined terms, Isoallergens and Isoforms will be used throughout this thesis. Isoallergens are homologous allergens (> 67% identity) from the same organism, that are similar in size and function to one another. Isoforms (variants) are a sub-category of Isoallergens, exhibiting greater than 90% identity, but typically differing by only a few amino acids. Isoforms are defined with numeral suffixes following the allergen group number (Pomés *et al.*, 2018). Several HDM allergens have geographical isoforms, observed to emerge in isolated HDM populations (Thomas, 2015). For example, the major *D. pteronyssinus* allergen Der p 1 is listed as having two Isoallergens and 22 Isoforms, these isoforms are designated Der p 1.0101 through Der p 1.0124 (WHO/IUIS).

1.6.5 Immuno-Reactivity and Biochemical Activity of Mite Group Allergens

Scientific publications detailing the immunological effects and pathways elicited by HDM allergens are innumerable and reviewing these in depth is beyond the scope of this thesis. Instead I will briefly discuss each *D. pteronyssinus* mite group allergen (MGA) in the context of its biochemical characteristics and briefly summarise its immunological importance. Where information is lacking for a *D. pteronyssinus* MGA, I will discuss the equivalent MGA present in *D. farinae.*

Immuno-reactivity to MGAs is hierarchical, with some MGA being considered major allergens, if they bind a large proportion of the anti-HDM IgE in patient serum. This relative immunoreactivity may stem from their abundance, aerobiology, stability in the environment or enzymatic potency (Thomas *et al.*, 2010, Wan *et al.*, 1999). A summary of MGAs is outlined in Table 1.2, representing the most up to date information in WHO/ISUS and supplemented with information regarding immunological hierarchy where known. Allergenicity for the purpose of this thesis is simplified to allow generalised insight into the importance of the MGA, thus allergenicity is described in two ways: (i) reactivity and (ii) potency.

Reactivity refers to the percentage of mite allergic patients that have significant reactions to a specific MGA. Further information regarding the various methodologies used (Skin prick test, Immunoblot, ELISA etc) in determining this reactivity can be retrieved from WHO/IUIS. IgE reactivity is an important consideration, as IgE mediated cellular reactions are triggered by trace amounts of allergen (Aalberse, 2000).

The potency of an MGA is discussed in terms of IgE titer, the amount of anti-HDM IgE directed against the specific MGA. This has significant implications for the severity of symptoms elicited upon allergen exposure (Rolinck-Werninghaus *et al.*, 2008).

1.6.6 Enzymatic Mite Group Allergens

Within HD a wide range of enzyme activities have been reported including trypsin, chymotrypsin, glucoamylase, chitinase, cellulase, lipase, elastase, lysozyme and alkaline phosphatase. Enzymatic activity is a common feature of allergens, of the 151 AllFams, 60 have known or predicted enzyme activities (Radauer, 2017). Of the 34 *D. pteronyssinus*/*D. farinae* MGAs 14 have known or predicted enzyme activity (Table 1.2).

1.6.7 Protease Allergens

Protease MGAs (Der p 1, 3, 6 and 9) are cystine/serine proteases with trypsin, chymotrypsin or collagenase activities (Table 2). The first identified and best described is Der p 1, a major allergen. Der p 1 has been demonstrated to break down epithelial barrier in skin and lungs, inactivate elastase inhibitors and stimulate innate immune cells to release large amounts of pro-inflammatory cytokines (Wang, 2013). In addition, the presence of Der p 1 can induce co-sensitisation to otherwise innocuous bystander molecules (Thomas, 2015). It is no wonder these potent immunostimulatory activities result in Der p 1 being a highly potent allergen, binding in excess of 75% of anti-*D. pteronyssinus* IgE (WHO/IUIS). Less studied are the serine proteases, Der p 3, 6 and 9. These minor potency allergens also disrupt epithelial tight junctions and induce production of proinflammatory cytokines (Gregory & Lloyd, 2011, Thomas, 2015).

Table 1.2 Biochemical Activity and Allergenicity of Mite Group Allergens

Allergen group	Allergen	Biochemical name	Biological activity	Isoforms	MW (KDa)	Allergenicity	Potency*
1	Der p 1	Cysteine protease	Proteolysis	24	24	>92%+	Major
2	Der p 2	NPC2 family	Lipid binding	15	15	>71%+	Major
3	Der p 3	Trypsin	Proteolysis	1	31	> 97%+	Minor
4	Der p 4	Alpha amylase	Glycoside hydrolase	1	60	>25%	Mid
5	Der p 5		Hydrophobic binding	2	14	>31%+	Mid
6	Der p 6	Chymotrypsin	Proteolysis	1	25	>41%+	Minor
7	Der p 7	Bactericidal permeability-increasing like protein	Lipid binding	1	26, 30 & 31	>37%+	Mid
8	Der p 8	Glutathione S-transferase		1	27	40%	Minor
9	Der p 9	Collagenolytic serine protease	Proteolysis	2	29	92%+	Minor
10	Der p 10	Tropomyosin	Structural Protein	1	36	5.60%	Minor
11	Der p 11	Paramyosin	Structural Protein	1	103	>18%	Minor
13	Der p 13	Cytosolic Fatty Acid Binding Protein	Lipid binding	1	15	>6%	Minor
14	Der p 14	Apolipophorin	Lipid binding protein	1	177	>65%[a]	unknown
15	Der p 15	Chitinase-like protein	Chitinase	2		70%[b]	Minor
16	Der f 16	Gelsolin/villin			53	47%	Minor
17	Der f 17	Calcium binding protein			53	35%	Minor
18	Der p 18	Chitin-binding protein	Glyco hydro family 18	1		63%[b]	Minor
20	Der p 20	Arginine kinase	ATP: guanido phosphotransferase	1		40%[c]	Minor
21	Der p 21	Group 5 Homologue		1		26%[d]	Mid

Allergen group	Allergen	Biochemical name	Biological activity	Isoforms	MW (KDa)	Allergenicity	Potency*
22	Der f 22	NPC2 family	Lipid binding	1			Unknown
23	Der p 23	Peritrophin-like protein domain	Chitin-binding	1	14	74%	Major
24	Der p 24	biquinol-cytochrome c reductase binding protein	ubiquinol-cytochrome c reductase complex component	1	13	>50%	Unknown
25	Der p 25	Triosphosphate isomerase	Carbohydrate biosynthesis	1		>46%	Unknown
26	Der f 26	Myosin alkali light chain	Structural Protein	1	18	>29%	Unknown
27	Der f 27	Serpin	protease inhibitor	1	48	>34%	Unknown
28	Der f 28	Heat Shock Protein		2	70	>68%	Unknown
29	Der f 29	Peptidyl-prolyl cis-trans isomerase (cyclophilin)	folding of proteins	1	16	>20%	Unknown
30	Der f 30	Ferritin	Iron Storage	1	16	>60%	Unknown
31	Der f 31	Cofilin	actin-binding protein	1	15	>30%	Unknown
32	Der f 32	Secreted inorganic pyrophosphatase		1	35	>15%	Unknown
33	Der f 33	alpha-tubulin	Structural Protein	1	52	>25%	Unknown
34	Der f 34	enamine/imine deaminase		1	16	68%	Unknown
35	Der f 35			1	14.4	51%	Unknown
36	Der p 36			1	23	>42%	Unknown
37	Der p 37	Peritrophic like protein domain	Predicted chitin binding	1	30	>19%	Unknown
38	Der p 38	Bacterial lytic enzyme		1	15	>45%	Unknown

Data curated from WHO/ISUS database correct as of 01/05/2019. Allergenicity: Percentage of Mite Allergic patients that have significant reactions to this allergen. MW: Determined by SDS-PAGE. a: Data for Der f14 as no data available for Der p 14. b: Data from (O'Neil *et al.*, 2006). c: Data from (Hales *et al.*, 2006) d: Data from (Weghofer *et al.*, 2008). * Taken from (Sanchez-Borges *et al.*, 2017). + Allergens with immunostimulatory adjuvant effects (Sanchez-Borges *et al.*, 2017).

1.6.8 Glycosyl hydrolase and Carbohydrate Binding Allergens

The Gene Ontology (GO) term, Carbohydrate metabolic process, is one of the top GO terms associated with the highest number of allergens in AllFam (Radauer *et al.*, 2008), therefore it is not surprising that several MGAs fall into this category. Carbohydrate degrading MGAs include the Glycosyl hydrolases (GH) α-amylase Der p 4 (GH13) and chitinase Der p 15 (GH18), the triosephosphate isomerase Der p 25 and the bacteriolytic enzyme Der p 38 (Table 1.2). Der p 38 is one of the most recently recognised allergens, added to WHO/ISUS database in the latest revision (01/01/2019). Der p 18 and Der p 23 are structural proteins that form part of the chitinous peritrophic membrane that surrounds the faecal pellet (Weghofer *et al.*, 2013, Bolognesi *et al.*, 2008, Resch *et al.*, 2016). These proteins appear to have similar function and localisation but exhibit vastly different allergenicity. Der p 18 is a minor allergen, whereas Der p 23 is a major allergen of similar importance to Der p 1 and Der p 2 (Sanchez-Borges *et al.*, 2017). Very little is known about newly identified Der p 37, it appears structurally similar to Der p 18 and Der p 23, containing a predicted peritrophic like protein domain (WHO/IUIS).

1.6.9 Other Enzymes

Other enzymatically active MGAs not discussed above are less studied but are members of well-defined allergen families. Der p 8 is among 13 glutathione-S-transferases listed in AllFam. A similar number of allergenic ATP: guanido phosphotransferase are listed alongside Der p 20. In the past Der f 29 has been reported both as a cyclophilin (An *et al.*, 2013) and profilin (Jiang *et al.*, 2015). It is officially recognized by WHO/IUIS as a cyclophilin, comprising a separate AllFam to that of profilin. Der f 32 an inorganic pyrophosphatase (PPase) is one of two PPases listed in AllFam, the other is from *Aedes aegypti* (Aed a 7). Der f 34 is the only enamine/imine deaminase listed in AllFam but has been shown to be cross-reactive with *A. fumigatus* (Radauer, 2017, ElRamlawy *et al.*, 2016).

1.6.10 Lipid-Binding Allergens

Lipid binding proteins are frequently identified as allergens, more than 50% of major allergens identified, have known or predicted lipid-binding activity (Thomas *et al.*, 2005). Of the officially recognised MGAs, five have lipid-binding activity, Der p 2, 7, 13, 14 and Der f 22 (Table 1.2). Der p 2 is considered a major *D. pteronyssinus* allergen, it has significant homology to MD-2, the LPS binding component of the TLR4 complex. The *D. farinae* homologue of Der p 2 (Der f 2), has been demonstrated to bind bacterial LPS and contains predicted TLR4 dimerization sites, suggesting it forms a TLR4-Der f 2-LPS complex (Ichikawa *et al.*, 2009). Der p 13 was biochemically characterised as having lipid-binding activity resulting in the activation of TLR signalling pathways (Satitsuksanoa *et al.*, 2016). The mechanisms of immune-stimulation for these lipid-binding allergens are not fully understood. In the case of Der p 7, the folded protein has similarities to lipopolysaccharide binding protein (LPB), it has been shown to bind bacterial derived lipid ligands which subsequently bind to TLRs on immune cells. The allergenicity of Der p 14 and Der f 22 are unknown, but their ability to bind lipid ligands is thought to be the causation of their immunostimulatory adjuvant effects (Gregory & Lloyd, 2011).

1.6.11 Structural Protein MGAs

In the muscles of multicellular organisms, contraction is enabled by the sliding of actin and myosin filaments over one another (Barton & Buckingham, 1985). The *D. pteronyssinus* homolog of Der f 26, a myosin alkali light chain, is likely complexed with Der p 11 (Paramyosin) a myosin binding protein that typically constitutes the core of myosin filaments (Sonobe *et al.*, 2016). Der p 10 Tropomyosin is an actin binding protein, while Der p 16 (Gelsolin) and Der f 31 (Cofilin) are actin remodelling proteins (Table 1.2). These proteins are found in close association with each other within the body of the HDM and therefore may have a similar route of allergen exposure. Der f 33 is an alpha-tubulin, a fundamental component of the cellular cytoskeleton (Nielsen *et al.*, 2010). Der f 33 may be found in the faeces, as the

peritrophic membrane is formed by cells budding off the midgut epithelium (Colloff, 2009).

1.6.12 Other MGA

MGAs 17, 27, 28 and 30 have been reported in *D. farinae*, however as yet these have not been directly identified as allergens in *D. pteronyssinus* (Table 1.2). Der f 17, a calcium binding protein, is a member of the broad AllFam group, EF-Hand Domain, which contains 74 recognised allergens and includes Myosin light chain allergens e.g. Der f 26. Serpin Der f 27, a putative serine protease inhibitor is listed in AllFam alongside only two other serpins from vastly different phyla, domestic chicken and wheat.

Heat shock proteins (HSPs) are common allergens and tend to be highly conserved and constituently expressed in all nucleated cells. They are thought to play an important role in chaperoning antigens and other immunologically important molecules between immune cells (Yusuf *et al.*, 2009). Three classes of HSPs are listed in AllFam HSP 20, HSP 70 (includes Der f 28) and HSP 90 (Table 1.2).

Ferritin is an important physiological protein in all arthropods, just as it is in nearly all other organisms, serving as the universal storage and transport protein for dietary iron (Whiten *et al.*, 2018). Der f 30 is the only allergenic ferratin officially recognised by WHO/ IUIS, its' importance as an allergen is yet to be determined (Table 1.2).

1.6.13 MGA with Unknown Function

MGA not discussed above are of unknown function. These include homologous allergens Der p 5 and Der p 21, and newly reported allergens Der f 35 and Der p 36. Despite having unknown function, Der p 5 and Der p 21 are well recognised allergens of equal importance to Der p 4 and Der p 7 (Table 1.2). They are unique to mites, and show moderate homology (21%) and cross-reactivity (Thomas, 2015). The protein

sequences for Der f 35 and Der p 36 have not been made publicly available, therefore further discussion of these allergens is not feasible at this point.

Though Der p 1, 2 and 23 represent the most prevalent allergens, understanding the diversity of allergens a patient is sensitised to has important clinical implications. Patients reactive to multiple HDM allergens show an increased risk for multi-allergic disease phenotypes (Kidon *et al.*, 2011). Moreover, different allergenic diseases show different patterns of allergen sensitisation (Banerjee *et al.*, 2015). Diagnosis of HDM allergy is typically performed with crude protein extracts, containing both allergenic and non-allergenic components in non-defined amounts. Researchers have examined numerous commercially available diagnostic and therapeutic HDM extracts, finding many were lacking important allergens and some had several fold variations in Der p 1 to Der p 2 ratios (Casset *et al.*, 2012). The 2017 international consensus (ICON) report on the clinical consequences of mite hypersensitivity outlines the benefits of using component specific diagnosis, which includes purified or recombinant allergens for the most important MGAs (groups 1, 2, 4, 5, 7, 10, 11, 14, 15, 18, 21 & 23) as a means of determining if a patient is genuinely sensitised to HDMs (Sanchez-Borges *et al.*, 2017).

1.6.14 Cross-reactivity

Patients sensitised to HDM also show sensitisation not only to other mites, but to food allergens such as snails, shrimp, mussels and fungi (Sidenius *et al.*, 2001, ElRamlawy *et al.*, 2016). As the structure and function of a protein has important implications for allergenicity, cross-reactivity is linked to protein family rather than allergen source (Hauser *et al.*, 2010, McKenna *et al.*, 2016, Radauer *et al.*, 2008). Most MGAs discussed so far have predicted biochemical functions that placed them in well-defined allergen families (Radauer *et al.*, 2008).

Identification of cross-reacting allergens is often the result of epidemiological studies, however homology may also highlight cross-reacting allergens (Sidenius *et al.*, 2001).

The FAO/WHO expert consultation report on Allergenicity of Foods Derived from Biotechnology, stipulates that a protein is to be considered potentially cross-reactive if it meets one or both of the following criteria; “exhibit more than 35 % identity in the amino acid sequence of the expressed protein”, “using a window of 80 amino acids and a suitable gap penalty” or “identity of 6 contiguous amino acids” with a known allergen (FAO/WHO, 2001).

The availability of large protein and allergen databases can allow researchers to predict biochemical functions of proteins and examine homology between allergens to establish patterns that may confer allergenicity. This technique can be applied to examining the proteome of newly sequenced species to give insight into potentially allergenic proteins.

1.7 Omics of HDMs

The recent emergence of sophisticated omic technologies has allowed for cost-effective global investigation of an organism at the genomic, transcriptomic and proteomic levels. Significant insight into the molecular toolkit an organism possesses can be gleaned from examination of the functional elements of its’ genome. The wealth of genomic and functional protein information available provides a rich resource for rapid assembly of *de novo* genome sequences, gene prediction and high accuracy of functional protein annotation (Chandramouli & Qian, 2009).

Bioinformatic prediction of protein-coding genes occasionally results in false positive gene calling, thus in the case of a newly sequenced genome, proteins must be identified to positively prove their existence (Stanke *et al.*, 2006, Prasad *et al.*, 2017). Proteomics offers a sensitive and high-throughput method for doing so, characterising the complete set of proteins expressed in a whole organism or a sub-section of that organism (Chandramouli & Qian, 2009). A strategic approach must be taken when conducting proteomic studies, as proteomic studies are intrinsically biased towards identification of high abundance proteins, that are favourable to tryptic digestion and

result in formation of optimal length peptides (0.5 – 3 kDa) for LC-MS/MS fragmentation (Fonslow *et al.*, 2013, Tsiatsiani & Heck, 2015). Some proteins are only expressed under particular conditions. Moreover, the sample preparation methods influence the subset of proteins that can be reliably identified through proteomics (Klont *et al.*, 2018). Comprehensive proteome characterisation studies require extensive use of multiple protein extraction methods and depletion of high abundance proteins (Klont *et al.*, 2018, Tsiatsiani & Heck, 2015, Fonslow *et al.*, 2013). Comparative proteomics, between specimens, allows for the measurement of organism-wide alterations in protein expression, independent of substrates or ELISAs (Hogrebe *et al.*, 2018). This can also allow us to identify proteins present in one sample and absent from another.

To date, proteomic studies of *D. pteronyssinus* have identified approximately 500 proteins utilising bottom-up proteomics (Bordas-Le Floch *et al.*, 2017) or 2D-electrophoresis coupled with tandem mass spectrometry of high abundance proteins (Erban *et al.*, 2017). These studies were limited to spectral searches against publicly available sequences (Erban *et al.*, 2017) or transcriptomic data (Bordas-Le Floch *et al.*, 2017). The many outstanding questions about *D. pteronyssinus* allergens and digestive enzymes may be answered by examination of the *D. pteronyssinus* airmid genome and proteome. Moreover, comparative proteomics of laboratory and wildtype *D. pteronyssinus* may identify the biochemical tools utilised by *D. pteronyssinus* in extracting nutrients from HD.

1.8 HDM Population Control and Allergen Avoidance Measures

Numerous HDM population control and allergen avoidance measures have been proposed as a means of improving clinical outcomes for sensitised individuals. These can be categorised into physical (heating, freezing, ventilation, washing, vacuuming, HDM barriers) or chemical controls (acaricides) and some were discussed earlier in this text. For the sensitised individual, eliminating HDMs from their home may not be enough to completely avoid exposure, as furniture in public spaces have been shown

to also harbour HDMs. HDMs have been found in soft furnishings on trains, buses, aeroplanes and private cars (Clarke *et al.*, 2015b, Colloff, 1987, Wickens *et al.*, 1997). In addition, clothing acts as a reservoir for HDMs, allowing them to be transported from an infested home and deposited into new habitats (Clarke *et al.*, 2015a). As we spend approximately 90% of our time indoors, complete allergen avoidance for the HDM allergic patients is almost impossible to achieve (Evans & McCoy, 1998).

A Cochrane review of commonly used physical, chemical and combined physical-chemical HDM control strategies has shown no clinical benefit or evidence that these control measures can reduce exposure to HDMs, their allergens or the severity/frequency of asthma symptoms (Gøtzsche & Johansen, 2008). A similar meta-analysis found moderate levels of symptom improvement in AD patients, however the authors concluded these improvements fell below what should be considered the basis for influencing clinical practice (Nankervis *et al.*, 2015). A common issue when assessing the efficacy of HDM population and allergen control measures is patient compliance. For example, regular and rigorous cleaning to remove HDM allergen is advised, however this cleaning must be conducted by someone other than the sensitised individual (Murray & Ferguson, 1983). Many of the proposed methods are expensive, complex or time consuming and therefore difficult to maintain for prolonged periods of time thus resulting in poor patient compliance (Nankervis *et al.*, 2015).

1.9 Knowledge Deficiencies

Although there are a vast number of publications examining the life cycle and allergens of *D. pteronyssinus*, there is a scarcity of in-depth molecular studies. Moreover, to date there have been few genomic studies of medically and economically important Acari. Thus, genomic sequencing and annotation of the *D. pteronyssinus* airmid genome would provide general knowledge of Acari and a necessary platform for the detailed molecular study of this organism. The fact that the predicted proteome of *D. pteronyssinus* is not publicly available, has hampered the investigation of HDM

proteomics. Indeed, no large-scale proteomic study of *D. pteronyssinus* has been undertaken of the mite body or of excreted proteins (excretome), but with a complete genome, the huge diversity of digestive enzymes and allergenic proteins can be determined. Furthermore, protocols and techniques for the implementation of large-scale proteomic studies in HDMs, and more particularly in *D. pteronyssinus*, have not been developed, and high-throughput technologies have not been utilised for investigation of this medically important mite. Development of objective methods and high-throughput strategies to identify allergens present in HD would contribute to the body of knowledge of the pathogenesis of allergic diseases and address one of the key "unmet needs in mite allergy research" highlighted in the International consensus (ICON) on: clinical consequences of mite hypersensitivity, a global problem (Sanchez-Borges *et al.*, 2017). Furthermore, identification of *D. pteronyssinus* airmid genes that may confer resistance to desiccation or are involved in key physiological processes, may allow for development of much needed secondary prevention strategies for mite-induced diseases, thus meeting another unmet need in mite allergy research (Sanchez-Borges *et al.*, 2017). Moreover, this research will provide the foundation upon which much needed easy-to-administer HDM reduction or avoidance measures can be developed and their efficacy measured.

1.10 Scope and Aims of this Project

Much of the existing knowledge of *D. pteronyssinus* stems from ecological studies of mite physiology, life cycle or immunological studies of its' allergens. Much insight into the factors that contribute to allergenicity or facilitate *D. pteronyssinus* persistence in the home, can be gained from examining *D. pteronyssinus* airmid on a genomic and proteomic level.

The objectives of the work presented in this thesis are (i) sequencing the genome of *D. pteronyssinus* airmid, (ii) the characterisation of *D. pteronyssinus* proteomes and (iii) functional analysis of a trigene cluster encoding putative glucanases. This will identify protein families, enzymes, allergens and predicted allergens encoded in the

genome of this medically important HDM. *In silico* examination of proteins with homology to known allergens should identify proteins putatively involved in cross-reactive immunological sensitisation. Development of methods for culture and extraction of protein from both the mite body and spent culture medium of *D. pteronyssinus* airmid will allow identification of proteins and allergens expressed and localised to either the mite body or excretome. Proteomic analysis of HD using *D. pteronyssinus* airmid predicted proteome will allow for the identification of numerous *D. pteronyssinus* proteins previously unknown to be present in HD and give insight into the wildtype pattern of protein/allergen expression. Proteomic analysis of proteins present in HD may represent a new testing methodology for characterising the allergenome of HD. Examining the repertoire of enzymes expressed by *D. pteronyssinus* airmid and their subsequent localisation will allow for insights into the digestive capabilities and diet of *D. pteronyssinus*. Phylogenetic analysis, recombinant expression and biochemical characterisation of novel glucanases may reveal their physiological role in *D. pteronyssinus* and other HDMs. To these ends the following approaches will be taken:

1. Culture and maintenance of *D. pteronyssinus* airmid in the laboratory. Genomic sequencing of *D. pteronyssinus* and construction of a genome assembly which will be annotated by bioinformatic analysis.

2. Phylogenetic analysis of *D. pteronyssinus* and close relatives using whole genome assemblies, proteo-genomic analysis of *D. pteronyssinus* airmid genome assembly with reference to other recently published *D. pteronyssinus* genomes. *In silico* examination of the presence of allergen homologs and predicted allergens in *D. pteronyssinus* airmid. Protein extraction from mite body, spent culture medium and HD, prior to separation and analysis by LC-MS/MS and interrogation of the predicted *D. pteronyssinus* airmid proteome database.

3. In depth phylogenetic, biochemical and proteomic analysis of novel *D. pteronyssinus* airmid β-1,3 glucanases by recombinant protein expression, enzyme

assay and analysis of enzyme expression by LC-MS/MS will allow for insight into the physiological role of these enzymes.

2.0 Draft Genome Sequence of *Dermatophagoides pteronyssinus*, the European HDM.

This paper has been published in the journal *Genome Announcements* (August 2017).

Citation

WALDRON, R., MCGOWAN, J., GORDON, N., MCCARTHY, C., MITCHELL, E. B., DOYLE, S. & FITZPATRICK, D. A. 2017. Draft Genome Sequence of Dermatophagoides pteronyssinus, the European House Dust Mite. *Genome Announcements,* 5**,** e00789-17.

Abstract

Dermatophagoides pteronyssinus is the European house dust mite and a major source of human allergens. Here, the first draft genome sequence of the mite is presented, as well as the *ab initio* gene prediction and functional analyses that will facilitate comparative genomic analyses with other mite species.

Introduction

Dermatophagoides pteronyssinus, the European House Dust Mite, belongs to the Acari suborder Astigmata, of which more than 6,100 species are described containing both free-living and parasitic lineages (Klimov & O'Connor, 2013). House dust mites (HDMs) live in close association with vertebrates and utilize powerful enzymes to digest organic debris that vertebrates leave behind. Many of these enzymes, secreted in the feces, are major sources of allergens and lead to sensitization in 15 to 20% of the population in industrialized countries, through activation of both innate and adaptive immune responses (Jacquet, 2013).

Materials and Methods

For genomic sequencing, *D. pteronyssinus* (airmid healthgroup Ltd., Ireland) was cultured for 28 days on house dust mite maximal medium (HDMMM) at 75% relative humidity and 25 °C. A multi-isolate sample of *D. pteronyssinus* was collected and separated from the culture medium by sieving, followed by saturated saline separation. HDMs were washed and subjected to 24 h starvation before sterilization using 70% ethanol; they were then washed and frozen in liquid nitrogen prior to DNA extraction, which was performed using a Promega genomic DNA purification kit and mouse-tail method. DNA was quantified using a Qubit dsDNA BR assay kit and examined for integrity by agarose gel electrophoresis. Four sequencing libraries—500 bp paired-end (PE), 2 kb mate-pair (MP), 5 kb MP, and 10 kb MP—were prepared for the Illumina HiSeq 2000, 2500, and 4000 platforms (BGI, China) with PE read sizes of 100 bp and MP read sizes of 49 bp. A total of 130,978,913 PE reads, 143,286,220 2 kb MP reads, 56,245,986 5 kb MP reads, and 29,806,232 10 kb MP reads were first trimmed for adapter and base call quality with Trimmomatic (Bolger *et al.*, 2014). Trimmed reads were then being used for *de novo* assembly in dipSPAdes version 1.0 (Safonova *et al.*, 2015), which resulted in 4,459 contigs with an N50 of 68,101 bp. Scaffolds were generated using SSPACE (Boetzer *et al.*, 2011), and gaps were closed using GapFiller (Boetzer & Pirovano, 2012).

Results

The final assembly resulted in 1,322 scaffolds, with an N50 value 450,436 bp, an L50 of 33 scaffolds, and a GC content of 30.93%. The largest scaffold was 3,593,316 bp in length. The genome size of *D. pteronyssinus* was estimated to be approximately 70.76 Mb with a total assembly gap length of 3.14%. Gene prediction employed AUGUSTUS version 3.1.0 (Stanke *et al.*, 2008), trained using the gene set of *Dermatophagoides farinae* (Chan *et al.*, 2015). Gene functions were annotated with Pfam domains (Finn *et al.*, 2016) using InterproScan version 5.3-46.0 (Jones *et al.*, 2014). CEGMA version 2.5 was used to identify the presence of core eukaryotic

protein-coding genes (Parra *et al.*, 2007). Secreted proteins were predicted using SignalP version 4.1 (Petersen *et al.*, 2011) and transmembrane helices were predicted through the TMHMM server version 2.0 (Krogh *et al.*, 2001). The *ab initio* gene prediction discovered 12,530 gene models, containing 48,371 exons in total. 419 of the 429 CEGMA eukaryotic core genes were identified, and full-length sequences for 39 known mite allergens were located, including the mite group allergens 1 to 11, 13 to 16, 18, and 20 to 33 (Chan *et al.*, 2015, Rider *et al.*, 2015). Functional annotation resulted in gene ontology terms for 5,622 genes and Pfam domains for 8,031 proteins; 1,619 proteins are predicted to have a signal peptide, and 3,610 contain a transmembrane domain.

Accession number(s). This whole-genome shotgun project was deposited at DDBJ/ENA/GenBank under the accession number MQNO00000000. The version described in this paper is the second version, MQNO02000000.

Acknowledgement

R.W. is an Irish Research Council Enterprise Partnership Scheme Scholar (EPSPG/ 2014/64).

References

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. (2011) Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**: 578-579.

Boetzer, M., and Pirovano, W. (2012) Toward almost closed genomes with GapFiller. *Genome Biol* **13**: R56.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.

Chan, T.F., Ji, K.M., Yim, A.K.Y., Liu, X.Y., Zhou, J.W., Li, R.Q., Yang, K.Y., Li, J., Li, M., Law, P.T.W., Wu, Y.L., Cai, Z.L., Qin, H., Bao, Y., Leung,

R.K.K., Ng, P.K.S., Zou, J., Zhong, X.J., Ran, P.X., Zhong, N.S., Liu, Z.G., and Tsui, S.K.W. (2015) The draft genome, transcriptome, and microbiome of Dermatophagoides farinae reveal a broad spectrum of dust mite allergens. *Journal of Allergy and Clinical Immunology* **135**: 539-548.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., and Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* **44**: D279-D285.

Jacquet, A. (2013) Innate immune responses in house dust mite allergy. *ISRN Allergy* **2013**: 735031.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., and Hunter, S. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236-1240.

Klimov, P.B., and O'Connor, B. (2013) Is Permanent Parasitism Reversible?-Critical Evidence from Early Evolution of House Dust Mites. *Systematic Biology* **62**: 411-423.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567-580.

Parra, G., Bradnam, K., and Korf, I. (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**: 1061-1067.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H., (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. In: Nat Methods. United States, pp. 785-786.

Rider, S.D., Morgan, M.S., and Arlian, L.G. (2015) Draft genome of the scabies mite. *Parasites & Vectors* **8**.

Safonova, Y., Bankevich, A., and Pevzner, P.A. (2015) dipSPAdes: Assembler for Highly Polymorphic Diploid Genomes. *J Comput Biol* **22**: 528-545.

Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**: 637-644.

2.1 Supplementary Materials and Methods

2.1.1 Materials

2.1.1.1 Saturated Saline Solution (36% (w/v))

Sodium chloride (360 g) was brought to 1 L with distilled water and mixed until dissolved.

2.1.1.2 75% Relative Humidity Solution

Saturated saline solution (Section 2.1.1.1) was added to NaCl to form a 50% (w/v slurry.

2.1.1.3 Wash Solution A

Sodium chloride (90 g) was brought to 1 L with distilled water and mixed until dissolved.

2.1.1.4 Wash Solution B

Sodium chloride (22.5 g) was brought to 1 L with distilled water and mixed until dissolved.

2.1.1.5 50 X Tris-Acetate Buffer (TAE)

Trizma base (242 g) was added to deionised water (700 ml) and mixed until dissolved. Glacial acetic acid (57.1 ml) and 0.5 M EDTA (100 ml) were added, mixed and pH adjusted (pH 8). The volume was adjusted to 1 L with distilled water. The solution was stored at room temperature.

2.1.1.6 1 X Tris Acetate Buffer (TAE)

50 X TAE (20 ml; Section 2.1.1.5) was added to distilled water (980 ml). The solution was stored at room temperature.

2.1.2 Methods

2.1.2.1 HDM Housing Unit

HDMs were grown within culture vessels consisting of a 250 ml conical flask plugged with cotton wool (Figure 2.1). Culture vessels were housed within a Secador desiccator cabinet 3.0 (Figure 2.1). Maintenance of constant humidity was facilitated by filling the desiccant reservoir of the Secador desiccator cabinet to a depth of 2-3 cm with 75% Relative Humidity Solution (Section 2.1.1.2). Temperature was maintained by placing Secador desiccator cabinet within an incubator set to 25 ^{o}C

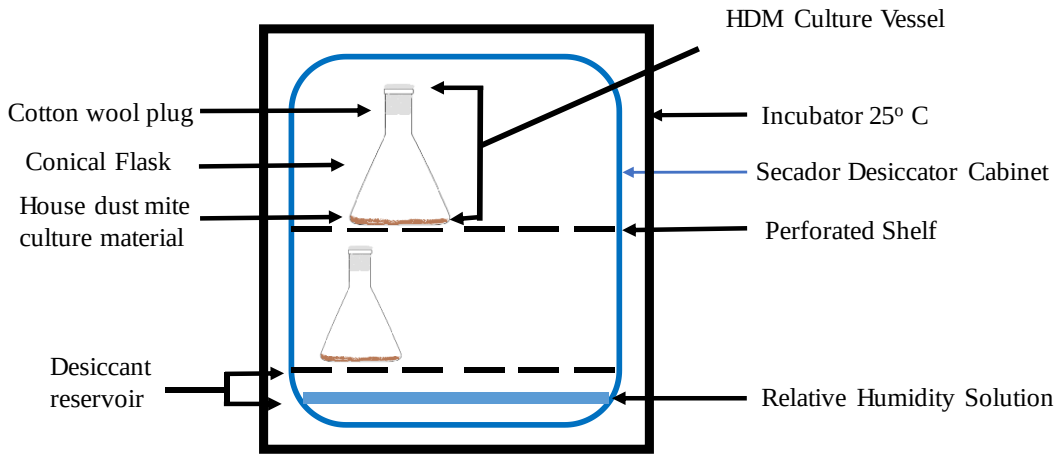


Figure 2.1: Schematic Representation of The HDM Housing Unit.

The housing unit consisting of Secador desiccator cabinet 3.0 placed within an incubator (25 ^{o}C). Relative humidity was maintained at 75% with use of relative humidity solution maintained at a depth of 2-3 cm. HDMs are cultured within a 250 ml conical flask and prevented from leaving the flask by a cotton wool plug.

(Figure 2.1). The Secador desiccator cabinet was cleaned with hot soapy water followed by 70% ethanol and the relative humidity solution was changed every three months to prevent microbial contamination.

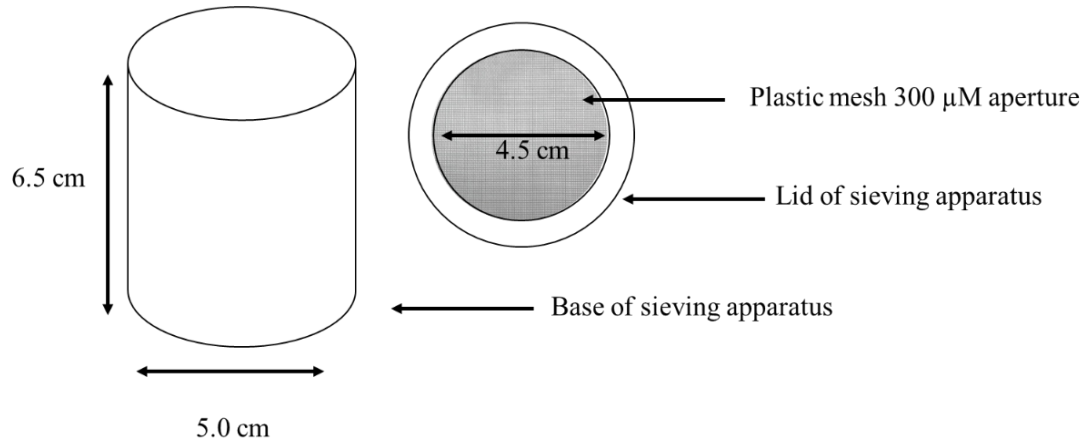


Figure 2.2: Sieving Apparatus.

A custom sieving apparatus constructed from 100 ml specimen pot (VWR) and 300 µM aperture plastic mesh was made. Sieving apparatus was used separate HDMs from culture media during culturing and harvesting.

2.1.2.2 Humidified HDMMM

HDMs are vulnerable to desiccation, so to avoid desiccation stress the culture medium was equilibrated to 75% RH by placing HDMMM (airmid healthgroup LtD, Dublin) into a weigh boat, and storing in a designated food desiccator with 75% RH solution for two weeks prior to use.

2.1.2.3 Sieving of HDMs

The sieving of HDMs was conducted in a laminar flow hood to minimise risks associated with allergen exposure to lab personnel. Culture media was placed in a custom-made sieving apparatus illustrated in Figure 2.2., lid of sieving apparatus was secured, and contents were sieved into a 1 L beaker by tapping the sieving apparatus against the side of the beaker. After a period of time no HDMMM passed through the mesh, the remaining culture contained approximately 90% HDMs and 10% culture media.

2.1.2.4 Culture of HDMs

D. pteronyssinus airmid were obtained from cultures housed for many years at airmid healthgroup LtD (Dublin, Ireland). The HDMs were sub-cultured every 28 days. Culturing was conducted in a laminar flow hood to minimise risks associated with allergen exposure to lab personnel. Sieved HDMs (1.25 g; Section 2.1.2.3) were placed into a culture vessel (Section 2.1.2.1) along with humidified HDMMM (3.75 g; Section 2.1.2.2), then the opening of the culture vessel was plugged with cotton wool. HDMs were incubated in housing unit (Section 2.1.2.1) for the 28 days before sub-culture.

2.1.2.5 Feeding HDMs

HDMs were fed 14 days post-culturing (Section 2.1.2.4) by adding humidified HDM culture media (2.5 g; Section 2.1.2.2).

2.1.2.6 Harvest and Washing of HDMs

Wash Solution A (50 ml; Section 2.1.1.3) was poured into a 500 ml separating funnel followed by sieved HDMs (3.5 g; Section 2.1.2.3) and topped up with an additional 450 ml of Wash Solution A. Contents were swirled gently to mix, then allowed to settle in the separating funnel until food debris sank to the bottom of the funnel and HDMs floated to the surface of the solution (approx. 2 min). Wash buffer A with food debris was drained from separating funnel into a waste beaker until only 50 ml remained, this ensured all the HDMs were retained. The washing procedure was then repeated a second time using 450 ml of Wash Solution B (Section 2.1.1.4) to remove remaining food debris. Wash Solution B was carefully removed from the separating funnel, ensuring all the HDMs were retained. Distilled water (450 ml) was added to the separating funnel and contents were swirled gently to mix. The entire contents of the separating funnel were passed through a Miracloth to separate the HDMs from the water. HDMs were washed by pouring additional distilled water (500 ml) through the Miracloth to remove residual wash buffer. HDMs were dried by placing Miracloth onto tissue paper. HDMs were used immediately for surface sterilisation (Section 2.1.2.8).

2.1.2.7 Starvation of HDMs

Washed HDMs (Section 2.1.2.6) were spread evenly onto a petri dish, using an inoculation loop to create a thin layer. Petri dish containing washed HDMs was placed in the housing unit (Section 2.6.2.1) for 24 h without culture media, allowing their digestive systems to empty of food.

2.1.2.8 Surface Sterilisation of HDMs

Following starvation for 24 h (Section 2.6.2.7), HDMs were collected from the petri dishes by the addition of cold 70% ethanol (5 ml). A pasture pipette was used to aspirate and dispense the HDM/ethanol mixture into a 50 ml Falcon tube. Additional

cold 70% ethanol was added until all HDMs were collected. The Falcon tube was filled to a final volume of 25 ml with cold 70% ethanol and incubated for 5 min at 4 ºC. Tubes were centrifuged (5 min, 4 ºC, 2000 x g) to pellet HDMs. Ethanol was removed and washing repeated twice more with 25 ml fresh 70% ethanol for a total of three washes. HDMs were then washed three times with sterile distilled water (45 ml) to remove ethanol. HDMs were aliquoted into sterile Eppendorf tubes (50 mg or 200 mg) used immediately for DNA extraction or snap-frozen in liquid nitrogen and stored at -80 ºC.

2.1.2.9 Isolation of Genomic DNA from *D. pteronyssinus* airmid

A multi-isolate sample of 28 day old *D. pteronyssinus* airmid cultures was prepared for genomic DNA extraction by separating mites from culture medium (Section 2.1.2.3). HDMs were washed (Section 2.1.2.6), subjected to 24 h starvation (Section 2.1.2.7) and surface-sterilised (Section 2.1.2.8). Following starvation, all steps were conducted in laminar flow hood to minimise risks of DNA contamination. DNA was extracted from 200 mg of ground mites (n = 8) using the Promega Wizard® Genomic DNA Purification Kit according to manufacturer's instructions for animal tissue (mouse tail method) with minor modifications. All buffers and reagents were supplied with the kit, except for RNase (R6513, Sigma) and proteinase K (P8450, Sigma). In brief, 120 µl of 0.5 M EDTA was added to 500 µl of chilled nuclei lysis buffer; 600 µl of EDTA-nuclei lysis buffer mixture was added to the ground HDM samples (200 mg; n = 8), followed by 17.5 µl proteinase K (20 mg/ml). Samples were inverted to mix then incubated at 65 ºC for 30 min. Samples were allowed to cool to room temperature, then 3 µl of RNase A solution (4 mg/ml) was added to the nuclear mixture and incubated for 20 min at 30 ºC. Once cool (RT), 200 µl protein precipitation solution was added to nuclear lysates. Samples were inverted to mix and incubated on ice (5 min). The protein pellet was collected by centrifugation (4 min; 13, 000 x g), supernatant (~ 600 µl) was removed to a fresh Eppendorf tube and protein pellet was discarded. Ice-cold isopropanol (600 µl) was added to supernatant and inverted to mix. Genomic DNA (gDNA) was collected by centrifugation (1 min; 13, 000 x g).

Supernatant was discarded and gDNA pellet was washed once more with 70% Ethanol. Following washing, gDNA pellet was allowed to air dry at 4 ºC for 1 h, then resuspended in 80 µl of nuclease free water (W4502, Sigma) and incubated on ice for a further 1 h. gDNA from eight replicates were pooled to obtain enough material for DNA sequencing. gDNA was centrifuged briefly to pellet insoluble gDNA then a 20 µl aliquot was removed for molecular analysis. The remaining DNA was stored at -70ºC until shipping to BGI China for sequencing. gDNA was quantified using a Qubit™ dsDNA BR Assay Kit (Life Technologies) and examined for integrity by agarose electrophoresis (Section 2.1.2.11 & 2.1.2.12). Presence of protein contaminants were assessed by reading the absorbance of the sample at 260 and 280 nm using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific).

2.1.2.10 Culture of Microbes from *D. pteronyssinus* at Various Stages of Preparation for Genomic DNA Extraction

Whole HDMs (50 mg) during various stages of genomic DNA preparation: harvesting (Section 2.1.2.6), surface sterilisation (Section 2.1.2.8) and ground HDM bodies, (Section 2.1.2.9) were examined for presence of culturable microbes. HDMs were washed with 1.5 ml sterile distilled water and vortexed briefly, then pelleted by centrifugation (3 min; 5,000 x g). Wash liquid was retained (100 µl) and used to inoculate agar plates in triplicate: A) Luria-Bertani agar (L3027, sigma), B) Malt extract agar (70145, Sigma), C) Sabouraud Dextrose Agar (CM0041, Oxoid), D) Difo^{TM}Czapek-DOX Broth (Cat No: 23381, Difco, Maryland, USA). Plates were incubated at 37 ºC for 5 days. All work was carried out using aseptic techniques in a laminar flow hood. Sterile distilled water was used as a negative control.

2.1.2.11 Agarose Gel Electrophoresis

Agarose Gel Electrophoresis was used to visualise PCR products, restriction digest reactions and examine genomic DNA integrity. Agarose gels were cast and run using Bio-Rad electrophoresis equipment. Agarose gels between 0.75 – 2 % (w/v) were

made by adding powdered agarose (Cat No: 551027, Invitrogen) to the appropriate volume of 1X TAE buffer (Section 2.1.1.6), in a 200 ml flask. The agarose was melted by heating in a microwave, with frequent mixing, until the agarose had dissolved into the TAE. Agarose-TAE mixture was allowed to cool (~50 ºC), then ethidium bromide (4 µl/100 ml) was added and swirled to mix. The molten gel was poured into the prepared mould. A gel comb was inserted, and the gel was allowed to cool on a level surface. Once set, the gel comb was removed, and the gel was placed into the gel electrophoresis tank, with the wells nearer to the negative (black) electrode. The 1X TAE Buffer (Section 2.1.1.6) was then poured into the tank to fully submerge the gel.

2.1.2.12 Loading and Running DNA Samples

gDNA samples were prepared by adding five volumes of gDNA samples to one volume of 6X Loading Dye (SMR0611, Thermo Scientific). Molecular weight markers were used to estimate gDNA fragment size; 1 kb DNA Ladder (SM1331, Thermo Scientific) and 100 bp GeneRuler (SM0321, Thermo Scientific), Lambda DNA marker (D2916, Sigma). Gels were electrophoresed at 60-120 V for 30-90 min and were visualised using a Syngene G:Box.

2.1.2.13 Polymerase Chain Reaction (PCR)

Polymerase chain reaction (PCR) was used to amplify four different Acari genes prior to sequencing or Restriction Fragment Length Polymorphism (RFLP). PCR was carried out using AccuTaq^{TM} LA DNA Polymerase (Sigma). Typical PCR reaction mixtures are shown in Table 2.1. All PCR reactions were carried out in 50 µl volumes using a G-Storm PCR (Roche) System. Primers used are shown on Table 2.2.

Table 2.1: PCR Reaction constituents

Reagent	**Volume**
Accu-Taq LA 10X Buffer	5 µl
dNTP Mixture (2.5 mM each)	1 µl
DMSO	1 µl
Molecular grade water	38.5 µl
Forward Primer	1 µl
Forward Primer	1 µl
Template DNA (50 ng/µl)	2 µl
Accu-Taq	0.5 µl

Table 2.2: Primers used to amplify HDM DNA.

Primer Name	Species Specificity	Sequence (5'- 3')
ITS2 mite universal-Forward	6933 Mites & Ticks	CGACTTTCGAACGCATATTGC
ITS2 Mite Universal-Reverse		GCTTAAATTCAGGGGGTAATCTCG
DM Group 2 Forward	*D. pteronyssinus*	CGAAGCCAACCAAAACAGA
DM Group 2 Reverse		TTCAGATTTTGGTGCAATTTTC
DM Group 3 Forward	*D. pteronyssinus*	CCCGGATCCATGATCATCTATAATAT
DM Group 3 Reverse		GAAGGAATTCTCACTGTGAACGTTTTG
Outer-Derf-F1	*D. farinae*	CAAGCGCTTGCCGTATCAATTCGGTTAACGTT
Outer-Derf-R1		ATGTTGCGAATTTGGTCGTCGGCATTGTTGT
Inner-Derf-F2		GGAATTGGATTTACGATCAC
Inner-Derf-R2		GTATGGATAGCTTCTTTCTTC
Outer-Derp-F1	*D. pteronyssinus*	CTGAAACTAACGCCTGCAGTATCAATGGAAAT
Outer-Derp-R1		TGTGCATTTGGTCGTCGGCATGATTGTTCCTA
Inner-Derp-F2		GAAATGCTCCAGCTGAAATC
Inner- Derp-R2		TCGATAGTAGCTTTCTTGGA

2.1.2.14 PCR amplification of Mite Group 2 Allergen Gene

PCR was used to amplify the mite group 2 allergen gene from *D. pteronyssinus* and control DNA (*D. farinae*) using primers (DM Group 2 Forward & DM Group 2 Reverse) shown in Table 2.2 and cycling parameters used as per Table 2.3.

Table 2.3: Mite Group 2 Allergen gene PCR cycling conditions.

Step	Temperature ºC	Time
Denaturation/ Activation	94	3 min
Denaturation	94	10 s
Annealing	54	20 s
Extension	68	30 s
Final Extension	68	10 min
Cycles	35	

2.1.2.15 PCR Amplification of Mite Group 3 Allergen Gene

PCR was used to amplify the mite group 3 allergen gene from *D. pteronyssinus* and control DNA (*D. farinae*) using primers (DM Group 3 Forward & DM Group 3 Reverse) shown on Table 2.2 and cycling parameters as per Table 2.4.

Table 2.4: Mite Group 3 Allergen gene PCR cycling conditions.

Step	Temperature ºC	Time
Denaturation/ Activation	94	3 min
Denaturation	94	30 s
Annealing	53	60 s
Extension	68	60 s
Final Extension	68	10 min
Cycles	35	

2.1.2.16 Nested PCR amplification of Mite Group 1 Allergen gene

Nested PCR was used to amplify mite group 1 allergen gene from *D. pteronyssinus* and control DNA (*D. farinae*) as described by Chan *et al.* (2015). Outer PCR was performed on *D. pteronyssinus* DNA (Section 2.1.2.9562.1.2.16) using primers (Outer-Derp-F1 & Outer-Derp-R1) and *D. farinae* DNA (Section 2.1.2.9) using primers (Outer-Derf-F1 and Outer-Derf-R1) shown in Table 2.2. PCR cycling parameters are outlined in Table 2.5. Following amplification, the PCR reaction was diluted 1:1000 and used as a template for the inner PCR amplification. *D. pteronyssinus* DNA was inner PCR amplified using primers (Inner-Derp-F2& Inner-Derp-R2) and *D. farinae* DNA using primers (Inner-Derf-F2 & Inner-Derf-R2); Table 2.2: Primers used to amplify HDM DNA, PCR cycling parameters; Table 2.5.

Table 2.5: Mite Group 1 Allergen gene Nested PCR cycling conditions.

Step	Temperature ºC	Time
Outer PCR		
Denaturation/ Activation	95	5 min
Denaturation	94	50 s
Annealing	68	50 s
Extension	68	50 s
Cycles	25	
Inner PCR		
Denaturation/ Activation	95	5 min
Denaturation	94	50 s
Annealing	55	50 s
Extension	68	50 s
Cycles	30	

2.1.2.17 PCR amplification and Restriction Fragment Length Polymorphism of mite ITS2 gene

Restriction Fragment Length Polymorphism (RFLP) analysis of mite ITS2 region was conducted as described by Wong *et al*. (2011) for *D. pteronyssinus* and mite control DNA (*D. farinae*). Amplification of ITS2 region was conducted using ITS2 mite universal-Forward and ITS2 Mite Universal-Reverse primers (Table 2.2) and

using PCR cycle conditions (Table 2.6). Restriction enzymes *Hinf*I (FD0804, Thermo Scientific) and *Alu*1 (FD0014, Thermo Scientific) were used for RFLP analysis of the ITS2 gene (Table 2.7). Reactions were carried out as per the manufacturer's instructions. Digestion reactions were visualised by agarose gel electrophoresis (Sections 2.1.2.11 & 2.1.2.12).

Table 2.6: Acari Mite ITS2 gene PCR cycling conditions.

Step	**Temperature ºC**	**Time**
Denaturation/ Activation	95	3 min
Denaturation	94	30 s
Annealing	59	30 s
Extension	68	50 s
Final Extension	68	10 min
Cycles	35	

Table 2.7: Restriction Pattern of Mite ITS2 Gene Reported by Wong et al. (2011).

Mite	PCR product size	*Hinf*I	*Alu*1
G. malaysiensis	260	190, 90	140,120
D. pteronyssinus	320	**320**	**210, 120**
B. tropicalis	325	200, 90	325
D. farina	330	**180, 110**	**220, 110**
A. ovatus	430	210, 120	270, 160
T. putrescentiae	490	250, 150	340, 150

2.1.2.18 PCR Purification

Purification of PCR products for sequencing was carried out using a QIAquick PCR Purification Kit (Qiagen, UK) as per manufacturer's instructions. PCR products were eluted by placing the column in sterile microcentrifuge tube and adding nuclease-free water (30 µl) to the column. The purified DNA was quantified using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific) and stored at -20 ºC, until required for use.

2.1.2.19 PCR Product Sequencing

Gene fragments were sequenced on a commercial basis (LGC Technologies, Belgium). Sequence data was searched against publicly available gene sequences for highest scoring match (BLASTn; NCBInr).

2.2 Supplementary Results

2.2.1 The Culturing of Microbes from *D. pteronyssinus* during gDNA Extraction Preparation

Liquid recovered from washing of HDMs (Section 2.1.2.10) was assessed for presence of viable microbes. After a 5-day incubation (37 ^{o}C) bacterial colonies were observed on Luria - Bertani Agar (LB agar) and Czapek-Dox agar plates from washed non-sterilised HDMs, HDMs post-sterilization were negative for bacterial growth (Table 2.8). Ground HDM bodies were also negative for bacterial growth. Fungal growth agars Malt Extract and Sabouraud Dextrose Agar, were negative for fungi (Table 2.8). These data indicated that both the surface and gut of the HDMs did not contain culturable bacteria or fungi following surface sterilisation, therefore reducing the risk of microbial contamination of extracted gDNA. Surface sterilisation of the HDMs prior to gDNA extraction minimised the amount of microbial DNA being isolated and sequenced along with HDM gDNA.

Table 2.8: Culture of Microbes from *D. pteronyssinus* at Various Stages of Preparation for Genomic DNA Extraction.

Growth Agar	HDM Material	Colony Forming Units (Positive/Negative)
Luria -Bertani	Washed Mites	Positive
Czapek-Dox	Washed Mites	Positive
Malt Extract	Washed Mites	Negative
Sabouraud Dextrose	Washed Mites	Negative
Luria -Bertani	Surface Sterilised Mites	Negative
Czapek-Dox	Surface Sterilised Mites	Negative
Malt Extract	Surface Sterilised Mites	Negative
Sabouraud Dextrose	Surface Sterilised Mites	Negative
Luria -Bertani	Ground Mites	Negative
Czapek-Dox	Ground Mites	Negative
Malt Extract	Ground Mites	Negative
Sabouraud Dextrose	Ground Mites	Negative
Luria -Bertani	Sterile Water	Negative
Czapek-Dox	Sterile Water	Negative
Malt Extract	Sterile Water	Negative
Sabouraud Dextrose	Sterile Water	Negative

2.2.2 Quantification and Quality Control of *D. pteronyssinus* genomic DNA prior to sequencing

D. pteronyssinus gDNA for genome sequencing (Section 2.1.2.9) was assessed for purity by A260/280 and A260/230 ratios (Table 2.9). gDNA integrity was assessed by Agarose Gel electrophoresis (Section 2.1.2.11). gDNA was observed to be of high molecular weight with minimal degradation and therefore of acceptable quality for sequencing.

Table 2.9: Quantification and Purity of *D. pteronyssinus* genomic DNA

DNA sample ID	A260/230	A260/280	*Concentration
RW_2	1.7	1.44	138 ng/µl

*Qubit BR DS DNA quantification Kit

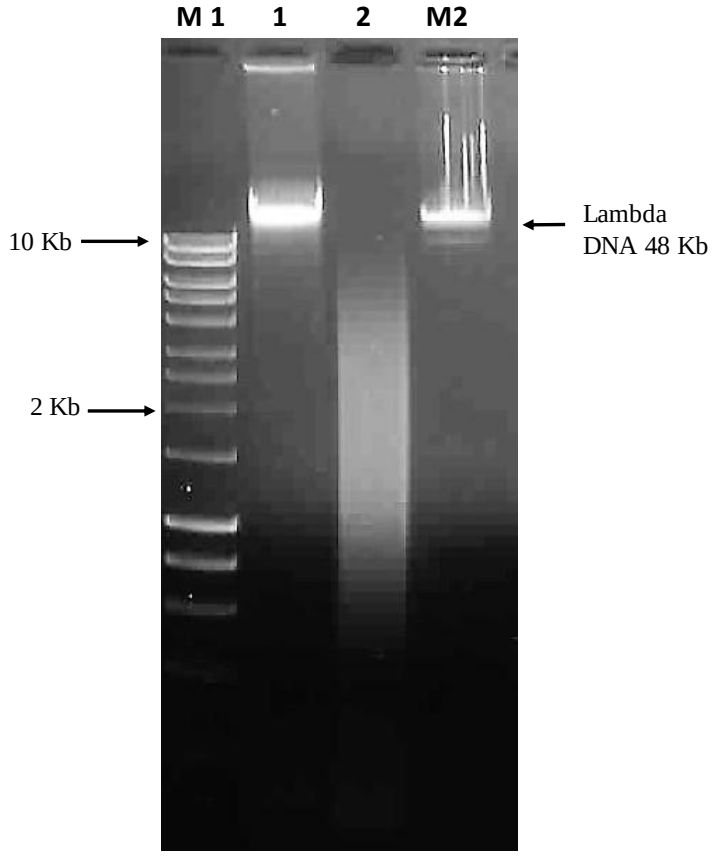


Figure 2.3: *D. pteronyssinus* genomic DNA Integrity

Purified extracted gDNA from HDM *D. pteronyssinus* analysed on a 0.75% agarose gel. Lane M1: Molecular weight marker. Lane 1: 400 ng gDNA. Lane 2: 400 ng degraded gDNA control. M2: Molecular weight marker. Large gDNA fragments were observed at >10 kb, with minimal degradation.

2.2.3 Species Purity of *D. pteronyssinus* genomic DNA.

2.2.3.1 Amplification and Sequencing of Mite Group Allergen 2 Gene

PCR was conducted on mite gDNA using dust mite group 2 allergen primers (Section 2.1.2.14). A single amplicon was observed for *D. pteronyssinus* and no amplicon was observed for control DNA (Figure 2.4A). Sequence data (139 bp) was searched in NCBI, analysis showed 99% identity (BLASTn; E-value; 8 e^{-64}) to *D. pteronyssinus* allergen 2 in NCBI database (JN222809.1 *Dermatophagoides pteronyssinus* Der p 2 variant 5 mRNA, partial cds) confirming presence of *D. pteronyssinus* gene in the gDNA (Figure 2.5). These genes were used to confirm presence of a single mite species in cultures of *D. pteronyssinus*.

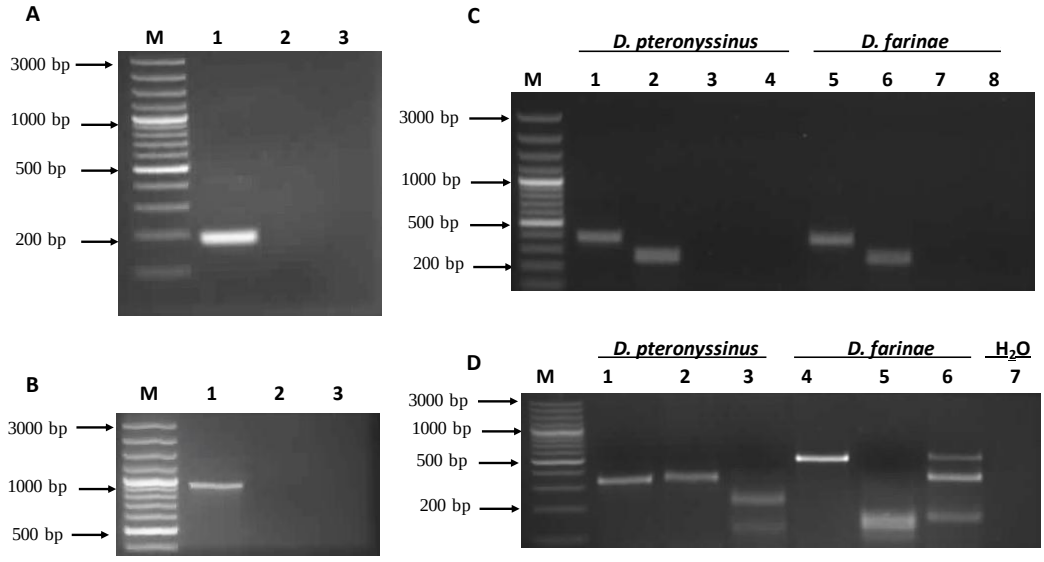


Figure 2.4: Molecular Assessment of Species Purity of gDNA from *D. pteronyssinus*.

(A) PCR amplification of Group 2 mite allergen gene. Lane M: Molecular weight marker ladder. Lane 1: *D. pteronyssinus airmid* DNA. Lane 2: Control DNA from *D. farinae*. Lane 3: Negative control with PCR master mix present but no DNA. **(B)** PCR amplification of Group 3 mite allergen gene. Lane M: Molecular weight marker ladder. Lane 1: *D. pteronyssinus airmid* DNA. Lane 2: Control DNA from *D. farinae.* Lane 3: Negative control with PCR master mix present but no DNA. **(C)** Nested PCR amplification of Group 1 mite allergen gene. Lane M: Molecular weight marker ladder. Lane 1-4: *D. pteronyssinus airmid* DNA. Lane 5-8: Control DNA from *D. farinae*. Lane 1: Outer-Derp. Lane 2: Inner-Derp. Lane 3: Outer-Derf. Lane 4: Inner-Derf. Lane 5: Outer-Derf. Lane 6: Inner-Derf. Lane 7. Outer-Derp. Lane 8: Inner-Derp. **(D)** RFLP of Dust Mite ITS2 gene. Lane M: Molecular weight marker ladder. Lane 1-3: *D. pteronyssinus airmid* DNA. Lane 4-6: Control DNA from *D. farinae*. Lane 1: ITS2 PCR product uncut. Lane 2. ITS2 PCR product cut with *Hinf*I. Lane 3. ITS2 PCR product Cut with *Alu*1. Lane 4: ITS2 PCR product uncut. lane 5: ITS2 PCR product cut with *Hinf*I. Lane 6. ITS2 PCR product cut with *Alu*1. Lane 7. PCR master mix present but no DNA.

```
Alignment of Sequence_1:  [JN222809.1 Dermatophagoides pteronyssinus Der p 2
variant 5 mRNA, partial cds] with  Sequence_2: [DM_GP2_PCR Product]

Seq_1  121   TTTTCGAAGCCAACCAAAACACAAAAACCGCTAAAATTGAAATCAAAGCCTCAATCGATG  180
              ##################################################||||##||||||
Seq_2  1     ----------------------------------------------------CCTC--TCGATG  10

Seq_1  181   GTTTAGAAGTTGATGTTCCCGGTATCGATCCAAATGCATGCCATTACATGAAATGCCCAT  240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  11    GTTTAGAAGTTGATGTTCCCGGTATCGATCCAAATGCATGCCATTACATGAAATGCCCAT  70

Seq_1  241   TGGTTAAAGGACAACAATATGATATTAAATATACATGGAATGTTCCGAAAATTGCACCAA  300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  71    TGGTTAAAGGACAACAATATGATATTAAATATACATGGAATGTTCCGAAAATTGCACCAA  130

Seq_1  301   AATCTGAAAATGTTGTCGTCACTGTTAAAGTTATGGGTGATGATGGTGTTTTGGCCTGTG  360
              |||||||||##################################################
Seq_2  131   AATCTGAAA----------------------------------------------------  139
```

Figure 2.5: Sequence alignment of mite group 2 allergen PCR product with highest scoring match gene in NCBInr.

2.2.3.2 Amplification and Sequencing of Mite Group Allergen 3 Gene

PCR was conducted on Mite gDNA using dust mite group 3 Allergen primers (Table 2.2). A single amplicon was observed for *D. pteronyssinus* and no amplicon occurred in control DNA (Figure 2.4B). Sequence data (219 bp) was searched in NCBI; BLASTn, the analysis showed 99% identity (BLAST; E-value; 3 e^{-82}) to *D. pteronyssinus* allergen 3 (EU092645.1 Dermatophagoides pteronyssinus Der p 3 allergen mRNA, complete cds) confirming presence of *D. pteronyssinus* (Figure 2.6).

```
Alignment of Sequence_1:  [DM_GP3_PCR Product] with  Sequence_2: [EU092645.1
Dermatophagoides pteronyssinus Der p 3 allergen mRNA, complete cds]

Seq_1  1     ------------------TTAAC-------------------------TTGGCTAATCCA  17
             ##################|||||############################|||||||||||
Seq_2  1     ATGATCATCTATAATATTTTAATTGTTTTATTATTGGCCATTAATACATTGGCTAATCCA  60

Seq_1  18    ATTCTACCAGCATCACCAAATGCAACTATTGTTGGTGGTGAAAAAGCATTAGCTGGTGAA  77
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  61    ATTCTACCAGCATCACCAAATGCAACTATTGTTGGTGGTGAAAAAGCATTAGCTGGTGAA  120

Seq_1  78    TGTCCATATCAGATTTCATTACAATCAAGTAGTCATTTTTGTGGTGGTACTATTCTTGAT  137
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  121   TGTCCATATCAGATTTCATTACAATCAAGTAGTCATTTTTGTGGTGGTACTATTCTTGAT  180

Seq_1  138   GAATATTGGATTTTAACAGCTGCACATTGTGTTGCCGGGTAAGTTTTGTTGTTGTTGAC  196
             ||||||||||||||||##########################################
Seq_2  181   GAATATTGGATTTTAA------------------------------------------  196
```

Figure 2.6: Sequence alignment of mite group 3 allergen PCR product with highest scoring gene match gene in NCBInr.

2.2.3.3 Nested PCR and Sequencing of Mite Group 1 Allergen Gene

A PCR was conducted on mite gDNA using outer primers (Table 2.2) for *D. pteronyssinus* and *D. farinae*. The PCR product was diluted (1:1000), then a second round of PCR was conducted using inner *D. pteronyssinus* and *D. farinae* primers **Table 2.2**). No amplicon was observed for PCR conducted on *D. pteronyssinus* DNA using *D. farinae* primers. Similarly, no amplicon was observed for PCR conducted on *D. farinae* DNA using *D. pteronyssinus* primers, indicating the high primer specificity for a single mite species and confirming the absence of *D. farinae* contamination in *D. pteronyssinus* genomic DNA. PCR reaction mixture (300 µl) was purified using Qiagen PCR purification Kit (Section 592.1.2.18), then sequenced (Section 2.1.2.19). Resultant sequence data (216 bp) was searched (NCBI; BLASTn) analysis showed 99% identity (BLAST; E-value; 9 e^{-105}) to *D. pteronyssinus* allergen 1 gene (KJ542096.1 *Dermatophagoides pteronyssinus* isolate RS68 cysteine proteinase-1 pro-enzyme gene, partial cds; Figure 2.7).

```
Alignment of Sequence_1:  [KJ542096.1 Dermatophagoides pteronyssinus isolate RS68
cysteine proteinase-1 preproenzyme gene, partial cds] with  Sequence_2:
[DM_GP1_InnerDerp]

Seq_1  707   CTTTCTTGGACGACACCATTATGTTGGATGTATTCAATACCACGTGGAATGGTATCACCA  648
             ################################||###|||||||||||#||||||||||||
Seq_2  1     --------------------------------TTTTCTACCACGTGG-ATGGTATCACCA  27

Seq_1  647   TGACAACCGTGTTGGGAAGCACAATCGACTAATTCTTGTTCAGCAAGATCCAATGATTGA  588
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  28    TGACAACCGTGTTGGGAAGCACAATCGACTAATTCTTGTTCAGCAAGATCCAATGATTGA  87

Seq_1  587   TTACGGTAAGCCAAATAAGCTGATTCAGTTGCGGCAACACCAGAGAAAGCCCAACATGAA  528
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  88    TTACGGTAAGCCAAATAAGCTGATTCAGTTGCGGCAACACCAGAGAAAGCCCAACATGAA  147

Seq_1  527   CCACAGCCTCCTTGCATACGAATGGGAGTGACAGTTCGCATTTGTCGCAAATCGATTTCA  468
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  148   CCACAGCCTCCTTGCATACGAATGGGAGTGACAGTTCGCATTTGTCGCAAATCGATTTCA  207

Seq_1  467   GCTGGAGCATTTCCATTGATACTGCAGGCGTTAGTTTCAGCATTCAAATCGAATTGAGTT  408
             ||||||||||||##################################################
Seq_2  208   GCTGGAGCATTT--------------------------------------------------  219
```

Figure 2.6: Sequence alignment of mite group 1 allergen PCR product with highest scoring gene match in NCBInr.

2.2.3.4 Sequencing and RFLP of Dust Mite ITS2 Gene

To confirm that the HDM cultures contained a single mite species (*D. pteronyssinus*) and were not contaminated with other HDM or storage mite species, PCR-RFLP of the ITS2 region was conducted as described by Wong *et al.* (2012), using primers (Table 2.2) specific for the ITS2 gene region of mites and ticks (Wong *et al.*, 2011). PCR reactions containing multiple amplicons would indicate more than one species was present. A single PCR product was observed from both *D. pteronyssinus* and *D. farinae* gDNA (Figure 2.4 D). ITS2 PCR products from *D. pteronyssinus* and mite control (*D. farinae*) were digested with *Hinf*I and *Alu*1 (Section 2.1.2.17; Figure 2.4D). *Hinf*I restriction enzyme (FD0804, Thermo Scientific) was selected as it cuts all mite ITS2 PCR amplicons except *D. pteronyssinus*, *Alu*1 (FD0014, Thermo Scientific) was

selected as restriction enzyme control as it cuts *D. pteronyssinus* ITS2 gene (Wong *et al.*, 2011). *D. pteronyssinus* ITS2 gene was uncut by *Hinf*I, but was cut by *Alu*1, resulting in fragments of approximately 210 bp and 120 bp. ITS2 PCR product from control DNA (*D. farinae*) was cut by *Hinf*I and *Alu*1 showing enzyme specificity and absence of star activity (Figure 2.4D). PCR reaction mixture was purified (Section 2.1.2.18) and sequenced (Section 2.1.2.19). Resultant sequence data (113 bp) was searched in NCBI; BLASTn and matched the *D. pteronyssinus* ITS2 gene (KY994137.1 *Dermatophagoides pteronyssinus* isolate 14B.3 5.8S ribosomal RNA gene, partial sequence; internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence; Figure 2.8) with 99% identity (BLAST; E-value 4 e^{-42}). Sequence data and RFLP of ITS2 gene from *D. pteronyssinus* gDNA indicated a pure culture containing only *D. pteronyssinus* mites.

```
Alignment of Sequence_1:  [DM_ITS2_PCR Product] with  Sequence_2: [KY994137.1
Dermatophagoides pteronyssinus isolate 14B.3 5.8S ribosomal RNA gene, partial
sequence; internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA
gene, partial sequence]

Seq_1  1     --A-----------------------------------GCTTCGTTTGTCTGAGCGTC  21
             ##|#####################################|||||||||||||||||||
Seq_2  1     CGACTTTCGAACGCATATTGCAGCCATTGGATAGCCGATGGCTTCGTTTGTCTGAGCGTC  60

Seq_1  22    GTTATCAAATTATGACCAAACAAAATGGGATAGATCTGTTTCGTCGTGCAAACGTCGTC-  80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||#
Seq_2  61    GTTATCAAATTATGACCAAACAAAATGGGATAGATCTGTTTCGTCGTGCAAACGTCGTCG  120

Seq_1  81    ----ATTTCCAAACATTTGATATGCTGACT----CGGTGAAAGTTG--------------  118
             ####|||||||||||||||||||||||||||#####|||||||####|##############
Seq_2  121   GATCATTTCCAAACATTTGATATGCTGACTTTTGTGGTGAAGAAGGCTTTGTTGCACATT  180
```

Figure 2.7: Sequence alignment of mite group ITS2 PCR product with highest scoring gene match in NCBInr.

2.3 Concluding Remarks

The work described in this chapter outlines the steps taken to grow *D. pteronyssinus* in the laboratory in sufficient numbers to successfully facilitate extraction of genomic DNA and allow for subsequent genome sequencing.

D. pteronyssinus were housed within a humidity and temperature-controlled chamber, grown on a defined culture medium, fed and sub-cultured according to a strict regimen ensuring optimal growth. A multi-isolate sample of *D. pteronyssinus* was starved and surface-sterilised prior to DNA extraction to minimise DNA contamination from bacteria and fungi present in the HDM cultures and from the culture media. A series of molecular tests was conducted whereby four mite genes were amplified from DNA extracted from *D. pteronyssinus* and sequenced, to ensure these genes were a match to

previously-reported gene sequences in *D. pteronyssinus* and exclude the possibility that there was other mite species present in the cultures.

gDNA was of good integrity allowing for construction of four sequencing libraries (BGI, China), one paired end (500 bp) and three mate-pair (2 kb, 5 kb and 10 kb). Utilising the Illumina HiSeq 2000, 2500, and 4000 platforms (BGI, China) generated paired-end read lengths of 100 bp (n = 130,978,913) and mate-pair reads of 49 bp in length (n = 43,286,220; 2 kb, (n = 56,245,986; 5 kb, (n = 29,806,232; 10 kb). Bioinformatic analysis and genome assembly resulted in an assembly consisting of 1,322 scaffolds and revealing the genome of *D. pteronyssinus* to be approximately 70.76 Mb in length and consisting of 12,530 genes.

The successful sequencing and assembly of *D. pteronyssinus* genome provided the genetic foundation upon which a detailed proteomic investigation of the organism was built, as described in subsequent thesis chapters.

3.0 Proteome and Allergenome of the European house dust mite *Dermatophagoides pteronyssinus*

This paper has been published in the journal *PLOS ONE* (May 2019).

Citation

WALDRON, R., MCGOWAN, J., GORDON, N., MCCARTHY, C., MITCHELL, E. B. & FITZPATRICK, D. A. 2019. Proteome and allergenome of the European house dust mite Dermatophagoides pteronyssinus. *PLOS ONE*, 14**,** e0216171.

Abstract

The European house dust mite *Dermatophagoides pteronyssinus* is of significant medical importance as it is a major elicitor of allergic illnesses. In this analysis we have undertaken comprehensive bioinformatic and proteomic examination of *Dermatophagoides pteronyssinus* airmid, identified 12,530 predicted proteins and validated the expression of 4,002 proteins. Examination of homology between predicted proteins and allergens from other species revealed as much as 2.6 % of the *D. pteronyssinus* airmid proteins may cause an allergenic response. Many of the potential allergens have evidence for expression (n = 259) and excretion (n = 161) making them interesting targets for future allergen studies. Comparative proteomic analysis of mite body and spent growth medium facilitated qualitative assessment of mite group allergen localisation. Protein extracts from house dust contain a substantial number of uncharacterised *D. pteronyssinus* proteins in addition to known and putative allergens. Novel *D. pteronyssinus* proteins were identified to be highly abundant both in house dust

and laboratory cultures and included numerous carbohydrate active enzymes that may be involved in cuticle remodelling, bacteriophagy or mycophagy. These data may have clinical applications in the development of allergen-specific immunotherapy that mimic natural exposure. Using a phylogenomic approach utilising a supermatrix and supertree methodologies we also show that *D. pteronyssinus* is more closely related to *Euroglyphus maynei* than *Dermatophagoides farinae*.

Keywords: House dust mites, allergens, proteomics, house dust, enzymes, asthma.

Introduction

House dust mites (HDM) are the most prevalent source of indoor allergens worldwide, with 1-2% of the total population experiencing an allergic response in their presence [1]**.** HDM allergens are major causative agents in the pathogenesis of asthma, allergic rhinitis and atopic dermatitis [2]. Protease allergens disrupt the epithelial barrier and activate immune cells resulting in the production of large amounts of proinflammatory cytokines [3, 4]. Sero-dominant allergens; Der p 1and Der p 2 account for 50 – 60% of IgE reactivity in individuals tested [5]. Allergenic protein families represent only 2% of all protein families. Allergenicity and cross-reactivity is linked to the allergen family rather than allergen source [6]. Use of publicly available allergen databases to query newly sequenced genomes for the presence of potentially allergenic or cross-reactive proteins has enormous potential in identifying new allergens.

HDM allergens are either located within the mite body or in faecal particles. Current knowledge of mite allergen localisation is limited [7] and may be improved by employing comparative proteomics to study the mite body and spent culture media proteins of *D. pteronyssinus*. Proteins present in HDM faeces are of particular importance as faecal particles are inhaled deep into the lungs, due to their small size [4, 8, 9]. Very little is known about which HDM components are present in house dust or inhalable air, and

assessment is limited to allergens for which there are ELISAs [10]. Therefore, it is our belief that characterising *D. pteronyssinus* proteins present in house dust could potentially yield much needed insights into allergens present in house dust.

Previous studies have shown that approximately 50% of all European homes contain HDM [11]. Therefore, evidenced based biocontrol strategies are needed to curtail HDM populations in homes. A Cochrane review of commonly used physical, chemical and combined physical chemical HDM control strategies has shown no clinical benefit or evidence that these control measures can reduce exposure to HDMs, their allergens or the severity/frequency of asthma symptoms [12]. Reducing humidity within the home has been proposed as a means of constraining HDM populations and limiting allergen production [13]. However, transient exposure to moist air allows for long term survival and reproduction indicating HDM may employ mechanisms to resist desiccation [14]. Therefore, genomic and proteomic characterisation of *D. pteronyssinus* has the potential to reveal biochemical pathways that could be exploited in future biocontrol strategies. These "Omic" approaches have accelerated digestive enzyme discovery [15], enabling *in silico* prediction of biochemical activities coupled with measurement of gene or protein expression. Potent enzymes are excreted by HDM into their surroundings as a by-product of their digestive processes, therefore the presence of putative enzymes in faeces is a strong indicator of a digestive function [16]. Surveying the predicted proteome of *D. pteronyssinus airmid* and subsequent proteomic examination of enzyme expression and localisation could identify new enzymes utilised in nutrient acquisition.

Here we describe the proteome of the European HDM *D. pteronyssinus* using a strain of mite housed at *airmid* healthgroup ltd. We have analysed the proteome in an attempt to elucidate the phylogenetic relationships between different species of HDMs and determined the localisation of allergenic components and enzymes involved in nutrient acquisition. The predicted proteome provides the basis to further understand the reported cross-reactivity between HDM and phylogenetically distinct species. We have examined and report the *D. pteronyssinus* airmid predicted proteome, mite body proteome and excretome with reference to a wild-type proteome as a means of (i) identifying potentially

allergenic proteins, (ii) inferring localisation of allergenic and potentially allergenic molecules, and (iii) identifying proteins involved in key physiological processes.

Materials and methods

Genomic data

To construct and compare the phylogenetic relationships between the Acari, 12 genomes were downloaded from the NCBI database for use in the analyses (S1 Table). Two Arachnid outgroups namely the Arizona bark scorpion (*Centruroides sculpturatus*) and the American house spider (*Parasteatoda tepidariorum*) were also downloaded. Assembly completeness of each genome was assessed using BUSCO v3 (Benchmarking Universal Single-Copy Ortholog) [17] with the Arthropoda dataset. Comparative analysis of *D. pteronyssinus* genome assemblies were conducted against previously published *D. pteronyssinus* strains [18, 19].

Phylogenetic analysis

Phylogenomic analysis of *D. pteronyssinus* was undertaken with reference to 11 other species (subclass Acari), consisting of six Parasitiformes and five Acariformes (S1 Table). Supermatrix and supertree phylogenomic methods were employed to infer the evolutionary relationships. Suitable phylogenetic markers were selected by locating single copy orthologs (from BUSCO analysis above) in the Acari and outgroup genomes. Single copy orthologs that were ubiquitously present (n = 111) were aligned using MUSCLE [20]. Individual gene families alignments were subsequently concatenated to yield a supermatrix 77,878 amino acids in length. This supermatrix was used to reconstruct a maximum likelihood phylogenomic species tree using RAxML [21] utilising the LG+G+I+F model as selected by ProtTest [22], branch supports were determined using 100 bootstrap replicates.

Single copy ortholog families, present in at least four species (n = 2,796), were identified and individually aligned using MUSCLE. Subsequent phylogenies were generated using FastTree [23]. A supertree was constructed using the matrix representation with

parsimony (MRP) method implemented in Clann [24] using the 2,796 gene trees as input with 100 bootstrap replicates. The resultant phylogeny was visualised and annotated using the Interactive Tree of Life (iTOL) [25].

Proteo-genomic analysis of assembly completeness

Proteo-genomic software Peppy [26] was used to generate peptide databases from three *D. pteronyssinus* assemblies [18, 19, 27]. LC-MS/MS spectra derived from proteomic experimentation on *D. pteronyssinus* airmid were searched against the six-frame translated genomes (maximum FDR 0.01; precursor tolerance 2000; fragment tolerance 300; digestion rules – cleavage acid R & K, missed cleavages 1; static mods -mod C 57.021464). For completeness and comparative purposes, the spectra were searched against the corresponding translated predicted protein-coding genes for each assembly. Output files were filtered to locate unique peptides and corresponding genomic locations.

Annotation of predicted proteome

Annotation of predicted proteins was achieved using BLAST2GO Version 5.0 [28] to sequentially search SwissProt (Downloaded; 14/01/2018) then NCBInr (Downloaded; 30/08/2017) database. Gene Ontology (GO) terms were assigned to predicted proteins (GO cut off 55, GO weight 5, E value hit filter E^{-06} and default computational evidence codes). InterPro Scan was used to identify; families, domains, sites and repeats in predicted proteins (CCD, HAMAP, HMMPanther, HMMpfam, FPrintscan, BLASTPromDom) and performed secretion peptide prediction using SignalP ver4.0 in parallel. Mapping feature facilitated mapping of GO terms to enzyme codes.

***D. pteronyssinus* specific proteins**

D. pteronyssinus predicted proteins without BLASTp homology (n = 3,906) to proteins in NCBInr/SwissProt were searched (tBLASTn; E value ≤ E^{-05}) against Acari genomes for presence of homologs in closely related species (S1 Table). Predicted proteins without significant homology were considered *D. pteronyssinus*-specific proteins (n = 1,848) and were then searched against other the other two available *D. pteronyssinus* genome

assemblies [18, 19]. Proteins without homology to predicted proteins in these assemblies were considered *D. pteronyssinus* airmid-specific proteins.

Identification of LEA-like proteins

LEA homologs were identified by performing BLASTp (E-value ≤ $1E^{-05}$) searches [29] against the Late Embryogenesis Abundant Proteins Database (LEAPdb) [30].

Mite group allergen orthologs

The majority of mite allergens identified to date have been assigned to groups (Group 1-33) in accordance with their order of discovery [7]. Chan *et al* (2015) reported a further seven non-chronological allergens for *D. farinae [31]*. Mite group allergen (MGA) orthologs were identified by performing BLASTp searches of query FASTA files containing MGA from *D. pteronyssinus* & *D. farinae* (when no other sequences were available for *D. pteronyssinus*) against the Acari genomes utilised in this study (S1 Table). Allergen orthologs had to satisfy the following criteria; have reciprocal best BALST hits (RBH) with an allergen (E-value ≤ $1E^{-05}$) and a minimum alignment length of 80 amino acids with an identity ≥ 35% in accordance with FAO/WHO guidelines [32]. Allergens that had a RBH but did not meet the alignment and identity criteria were considered RBH homologs, these were included in the sequence similarity network and visualised using Gephi [33]. Each protein was represented by a node, two proteins were connected by an undirected edge if they were homologous (BLASTp ≥ $1E^{-05}$).

Allergens and predicted allergens

D. pteronyssinus airmid MGA and non-chronological allergens were located by performing a local BLASTp (E-value ≤ 1E-03) search of the *D. pteronyssinus* predicted proteome against the FASTA sequence file containing the MGAs located above. The match with the smallest E-value was chosen as designated MGA. These sequences were annotated as "Der p1 Allergen" or Der f 22 like-allergen etc. for *D. farinae* based BLAST hits. Subsequent BLASTp hits were considered MGA homologs and were assessed for potential cross-reactivity with MGA in accordance with FAO/WHO guidelines [32].

Sequence similarity between known allergens (Uniprot "allergenome") and *D. pteronyssinus* airmid predicted proteins was assessed in accordance with FAO/WHO guidelines [32].

D. pteronyssinus airmid culture

D. pteronyssinus airmid were obtained from cultures housed at airmid healthgroup ltd (Dublin, Ireland) and maintained on diet composed of dried porcine liver and yeast, house dust mite maximal media (HDMMM, airmid healthgroup ltd, Ireland) at 75% relative humidity and 25^{o}C.

Harvest of mites and spent culture medium

Mites from replicate cultures (n = 5) were separated from spent culture medium (SM) by sieving, and saline floatation method [16], washed with distilled water and surface-sterilised by submersion in 70% ethanol (3 min) followed by washing with sterile distilled water. An average of 10.8 mites were present per mg of spent medium. Mites were then snap frozen in liquid nitrogen, and lyophilised. SM (n = 5) was divided into aliquots (200 mg) and stored at -70 °C prior to protein extraction.

Protein extraction

Lyophilised mite bodies (MB) were ground to a fine powder. Proteins were extracted from MB (25 mg) by addition of 500 µl lysis buffer and quantified according to methods described by Owens *et al* (2015) [34]. Proteins from spent culture medium (SM, n = 5) were extracted by addition of glass beads (50 mg; 0.5 mm BioSpec Products) and 1000 µl lysis buffer [34], followed by bead-beating (30 Hz; 5 min; MM300, Retsch®). MB (n = 2) and SM lysates (n = 2) were pooled for gel filtration (one biological replicate). Protein lysates for shotgun proteomic analysis (n = 4) were normalised (MB; 0.4 mg/ml, SM 0.3 mg/ml), then prepared for digestion according to Owens *et al* (2015) [34].

Culture media contaminants

Proteins deriving from HDMMM (200 mg) were extracted (n = 1) and prepared for shotgun proteomic analysis as per methods for spent growth culture medium.

House dust protein extracts

Der p 1 positive (0.2 – 16.94 µg Der p 1 per gram house dust) house dust protein extracts (n = 21) were provided by airmid healthgroup ltd (Dublin, Ireland).

Gel filtration chromatography

Gel filtration chromatography was carried out using an ÄKTA purifier coupled with a Superdex 200 10/300 GL column (GE Healthcare, Germany), equilibrated in PBS. Filtered specimens (0.22 µm), were injected (500 µl) and separated (flow rate 0.4 ml/min) with absorbance monitored at 215, 254 and 280 nm. Fractions were collected (between ~ 6 and 32 ml) and stored at -20 °C until further analysis.

Proteolytic digestion of protein extracts for proteomic analysis

Specimens for proteomic analysis were prepared for LC-MS/MS as described by Owens *et al* (2015) [34].

Nano-flow liquid chromatography electro-spray ionization tandem mass spectrometry (LC-MS/MS) analysis

Peptide mixtures were analysed using a Thermo Fisher Q-Exactive mass spectrometer coupled to a Dionex RSLC nano for LC-MS/MS analysis. LC gradients operated from 3–40% acetonitrile over 40 min, with data collection using a Top15 method for MS/MS scans [35].

Representative proteome

LC-MS/MS spectra obtained from proteomic analysis of *D. pteronyssinus airmid* were randomised into 5 groups (14-20 files each). Spectra were searched using Sequest HT

engine within Proteome Discoverer (Version 1.4) against *D. pteronyssinus* airmid predicted proteome (peptide filters; set to medium peptide confidence and protein filters; set to two peptides per protein). Protein molecular weight and pI for predicted proteome and representative proteome were calculated using JVirGel [36].

MaxQuant and perseus data analysis

Protein identification and label free quantitative (LFQ) analysis was conducted using MaxQuant (Version 1.6.1.0; http://maxquant.org/), statistical analysis of MaxQuant output data was performed by Perseus (Version 1.6.2.2) as described in O'Keeffe *et al* (2014) [37].

Culture media contaminants database

As it was not possible to fully remove culture media from specimens prior to proteomic analysis, a custom contaminates database was generated. This allowed for exclusion of protein identifications deriving from culture media (HDMMM) which contained porcine liver and baker's yeast. Spectra obtained from LC-MS/MS of HDMMM were interrogated (MaxQuant and Perseus) against combined proteomic database of *Sus scrofa* and *Saccharomyces cerevisiae*, resulting in the identification of 2,135 proteins (min. 1 peptide). These proteins were added to MaxQuant contaminants database to generate a custom contaminants database (n = 2,380 contaminants).

Data analysis of *D. pteronyssinus* proteomes

Spectra obtained from LC-MS/MS were interrogated (MaxQuant and Perseus) using either standard contaminant (HD samples) or custom contaminant (MB & SM samples) databases. Proteins were considered present when a minimum of two peptides (1 unique) for each parent protein was observed. Proteins meeting the following criteria were included in the analysis; (i) identified in two of the four non-fractionated whole protein extracts or (ii) identified in one chromatographic fraction.

Qualitative assessment of mite group allergen localisation

A qualitative assessment of localisation of mite group allergens to MB or SM was conducted. An allergen was considered present in MB/SM proteome if it was; (i) absent from one dataset or (ii) found at higher LFQ intensity and ms/ms count.

Results and discussion

Phylogenomic assessment of *D. pteronyssinus* reveals closest relative to be *Euroglyphus maynei*

Supermatrix and supertree phylogenomic methods were employed to infer the evolutionary relationships between the Acari species that have genome data available (S1 Table). Both supertree and supermatrix methodologies generated phylogenies with identical topologies and similarly high levels of bootstrap support (BP) for the monophyly of the Parasitiformes and Acariformes superorders (Fig 1A, 100% BP). Within the Parasitiformes superorder, Ixodida and Mesotigmata are monophyletic (Fig 1A, 100% BP). Within the Acariformes superorder, Trombidiformes and Sarcoptiformes are also found to be monophyletic (100% BP) although only a single representative of the Trombidiformes (*Tetranychus urticae*) is represented in our dataset (Fig 1A).

To date, phylogenetic studies of Acari (mites and ticks) have been restricted to multi-locus studies utilising a small number of genes, due to the absence of full genome sequences [38-40]. Our phylogeny infers a strongly supported (87% and 100% BP for supermatrix and supertree methods, respectively) sister group relationship between *D. pteronyssinus* and *Euroglyphus maynei* to the exclusion of *D. farinae*. This confirms previous studies which utilised only two and six genes but observed the same phylogenetic relationship between *D. pteronyssinus* and *E. maynei* [38, 41]. Therefore, in molecular evolutionary terms, *D. pteronyssinus* and *E. maynei* are more closely related to one another, even though *D. pteronyssinus* and *D. farinae* are currently classified within the same genus. Furthermore, the phylogeny confirms the paraphyly of the *Dermatophagoides* genus as previously reported [41].

Comparison of *D. pteronyssinus* airmid predicted proteome with other *D. pteronyssinus* assemblies

D. pteronyssinus airmid genome assembly [27] completeness was assessed by proteo-genomics. Peppy software [26] facilitated the mapping of 615,150 spectra (28,001 non-redundant) to the predicted proteome and 402,998 (21,505 non-redundant) to the assembly. Peptides that spanned intron-exon junctions identified in the predicted proteome were not mapped to the assembly. Of the 21,505 peptides mapped to the genome assembly, 96.2 % were also identified in predicted proteins (n = 20,683). Several peptides (n = 65) were located adjacent to predicted genes indicating that 65 gene models may need to be extended. The predicted proteome of *D. pteronyssinus* airmid [27] incorporates 8.3 million amino acids, this is significantly more than the Liu *et al*. (2018) and Randall *et al*. (2018) assemblies, which have 6.7 million and 5.9 million amino acids respectively. Moreover, despite having fewer predicted proteins than the other two available *D. pteronyssinus* proteomes (i.e. 12,530 versus 15,846 proteins [19] and 19,368 [18]), *D. pteronyssinus* airmid has on average longer protein coding genes (557aa vs. 425aa & 304aa respectively). These results indicate that differences in gene calling methodologies are most likely responsible for the differences in the number of protein coding genes.

Proteogenomic comparison of the predicted proteomes reveals the highest number of proteins were identified using the Waldron *et al* (2017) proteome (n = 4,581), followed by Randall *et al* (2018) with 4,416, then Liu *et al* (2018) with only 3,408 proteins identified. Therefore, unsurprisingly proteogenomic analysis of the available *D. pteronyssinus* genome assemblies against our protein samples which are derived from *D. pteronyssinus* airmid reveals the Waldron *et al* (2017) assembly and predicted proteome to be the most appropriate for our analyses, as higher numbers of protein identifications are uncovered relative to the other current assemblies [18, 19, 27].

Mite group allergen orthologs in Arachnidia

The presence of common MGA in the genomes of Arachnida species (S1 Table) was investigated, identifying multiple putative cross-reactive MGA orthologs in mite species. Most MGA had numerous orthologs distributed across all species (S1 Data) with presence closely linked to phylogeny. To help visualise the abundance of MGAs in the different Arachnida species a homology network was generated. Our results show that *D. farinae*, *D. pteronyssinus*, and *E. maynei* contained at least one MGA ortholog for all groups investigated, with the exception of Groups 23 & 24 in *E. maynei*. Group 7 and 14 allergens are only located in the Sarcoptiformes subset of Acariformes. Homologous allergens, Der p 5 and Der p 21, were present in the closely related Acariformes *D. pteronyssinus*, *D. farinae*, *E. maynei* and *Psoroptes ovis*, but absent from the other species (Fig 1B, S1 Data). Group 23 allergens are specific to *D. farinae* and *D. pteronyssinus*. More distantly related species from the Parasitiformes superorder were either missing orthologs of particular allergens such as Der p 4, 5 & 21 or contained RBH homologs only, Der p 1, 2, 22, 23 & 27 for example, (S1 Data). Serine proteases (groups 3, 6 & 9) appear to be expanded in some species, with a minimum of 10 homologs in *D. pteronyssinus* to a maximum of 28 in *ixodes scapularis* (Fig 1B, S2 Data).

Annotation of *D. pteronyssinus airmid* predicted proteome

Predicted proteome annotations

Multi-database Blast2GO workflow enabled annotation of 96% of the predicted proteome (*n* = 11,996, S3 Data). Gene Ontology (GO) terms were assigned to 68.2% of proteins (*n* = 8,546, Fig 2A). InterPro Scan (IPS) assigned IPS annotations to 95.5% of predicted proteins (*n* = 11,971), 6, 874 with IPS GO terms and IPS IDs to a further 3,804 proteins (Fig 2B). SignalP4.0 identified eukaryotic secretion signals in 10.3% of predicted proteins (*n* = 1,293, Fig 2B). Enzyme codes (EC) were assigned to 21.5% of the predicted proteins (*n* = 2,689, Fig 2C) with hydrolases (EC:3.0) representing the largest enzyme category (*n* = 1,244). Putative peptidase activity was identified in 377 predicted proteins comprising almost 3% of the total predicted proteome. Peptidase EC

(EC:3.4) were assigned to 275 peptidases, the remaining peptidases (*n* = 102) were identified by GO annotations. Enzymes have a propensity to cause allergy [42]. The potent peptidase activity of Der p 1 has been shown to disrupt numerous immune system processes [43] and it is thought lesser studied peptidases may have a similar effect [42]. Enzymes, particularly those with predicted secretion peptides, should be considered in the context of patient exposure as they are more likely to be excreted into house dust and therefore may augment the immune response.

D. pteronyssinus specific proteins

Predicted proteins without BLASTp alignments (*n* = 3,906) to proteins in NCBI/Swissprot were searched (E-value $\leq 1E^{-05}$) against closely related species (S1 Table), 2,054 had homologs in one or more species. The remaining 1,848 uncharacterised proteins (S4 Data) represented *D. pteronyssinus*-specific proteins [44]. Of these, 1,475 (S4 Data), were specific to *D. pteronyssinus* airmid strain as they were not found in the other *D. pteronyssinus* assemblies [18, 19]. These data suggest that 88.3% of identified proteins are core protein coding genes as they are found in all 3 *D. pteronyssinus* assemblies, with the remaining 11.7% being strain specific. Some uncharacterised *D. pteronyssinus* airmid-specific proteins may represent adaptations while others are a likely consequence of genetic drift occurring in isolated populations [45]. Bacterial pan-genomic studies estimate that strain-specific genes range from 5% to 35% per genome [46]. These strain-specific accessory genes are generally under relaxed mutational pressure, accumulating mutations more frequently than those of the core genome [47]. Further validation of the expression of these strain specific proteins is necessary to determine if they are functional proteins facilitating strain specific adaptations. It is worth highlighting that within the *D. pteronyssinus* representative proteome (discussed in more detail later) were 172 *D. pteronyssinus*-specific proteins, 23 of which had homologs in other *D. pteronyssinus* assemblies [18, 19], while 149 were only found in *D. pteronyssinus* airmid. The putative functions of these strain specific proteins are unknown, their role in strain specific adaptions may be discovered through further proteomic investigation.

LEA-like *D. pteronyssinus airmid* predicted proteins

HDM lose water readily through evaporation when the critical equilibrium humidity falls below optimum levels [48, 49]. Studies of biochemical mechanisms to resist desiccation have revealed late embryogenesis abundant proteins (LEAPs) play a key role in plant, insect and nematode desiccation survival [50-52]. Our analysis revealed 18 *D. pteronyssinus* airmid predicted proteins (S2 Table) to have significant homology with reported LEA proteins [30]. Gusev *et al* (2014) used a similar bioinformatic approach to identify 27 LEA-like proteins in the anhydrobiotic sleeping chromatid, *Polypedilum vanderplanki* [53]. This anhydrobiotic organism can tolerate extreme water loss of 97% by entering a state of suspended animation. The presence of similar proteins in *D. pteronyssinus* may explain the ability of mites in the protonymph developmental stage being entirely resistant to desiccation [49]. Although *D. pteronyssinus* can be killed by extended exposure to sub-critical equilibrium humidity, reduction of humidity in the home does not lead to a reduction in mite numbers or levels of allergen [54], as HDM return to a normal metabolic and reproductive state following short periods of optimal humidity. Furthermore, mattresses when occupied provide ample humidity to ensure survival of HDM in low humidity homes [14].

Expression of *D. pteronyssinus* airmid proteins exhibiting LEA-like proteins (n = 7) was validated by proteomics. All LEA-like proteins were found at low intensity (~LQF intensities of 1E+08) except for DERPT_G12026 and DERPT_G404 (LFQ Intensities > 7.8E+10). These two highly abundant LEA-like proteins are expressed under optimal non-desiccating laboratory growth conditions and were identified in both mite body and spent media. The ability of *D. pteronyssinus* airmid to utilise LEA-like proteins under normal laboratory conditions or in response to desiccation/cold may yield information that could be exploited for biocontrol strategies. The role of LEA-like proteins in relation to *D. pteronyssinus* is yet to be determined, however our data highlights them as potential players in desiccation resistance and hence as interesting biocontrol targets.

Allergens and predicted allergens

We searched the *D. pteronyssinus* airmid proteome for the presence of 37 previously reported mite allergens [7, 31]. MGA were identified for *D. pteronyssinus* airmid, we found full protein sequences (*n* = 37) for all reported MGA; Groups 1-11, 13-16, 18, 20-33 and seven non-chronological allergens (81.5-100 % identity, Table 1). Subsequent BLAST hits (E-value $\leq 1E^{-03}$) were considered MGA homologs (*n* = 233, Table 1, S5 Data). Der p 1-like cysteine proteases were represented by 31 homologs (20.9 – 63.7 % identity), several were found in clusters of 2-3 adjacent protein coding genes (*n* = 13). MGA homologs with high sequence similarity to the query MGA (> 67% identity) met criteria for being considered isoallergens [55]. Isoallergens were identified for eight different MGAs (Table 1). Der p 28 has five isoallergens (69.4 – 85.7% identity) and two were identified for Der p 29 (75.7 - 88.1% identity). One-third of MGA homologous proteins met the criteria for potential allergenicity [32] and were annotated as MGA Homologs (MGAH) (*n* = 93, S5 Data). In addition, many *D. pteronyssinus airmid* predicted proteins (*n* = 326, S6 Data) also exhibited significant similarity [32] to allergens from other species, suggesting they may be cross-species allergens. Most of these allergen homologs (AH) had multiple high scoring alignments with non *D. pteronyssinus* allergens (*n* = 991, Range:1 - 204, Median: 4). Significant overlap was seen between allergenic (MGA) and predicted allergenic (MGAH & AH) proteins with many being observed in more than one category, illustrated in Fig 2D. The structure and function of a protein has important implications for allergenicity, most allergenic proteins are limited to just 2% of protein families, with cross-reactivity linked to protein family rather than allergen source [6, 56, 57]. Most potential allergens highlighted in this study had predicted biochemical functions that placed them in well defined allergen families [6]. For example, predicted enolases DERPT_G12026 and DERPT_G4831 have high levels of sequence similarity (67 – 87%) with enolases from up to 19 phylogenetically distinct species (S6 Data). Enolase has long been recognised as a major cross-reacting allergen in plants, fungi, fish, and arthropods [58]. Moreover, the presence of at least one putative cross-reactive enolase ortholog [32] in all 12 Acari and two Arachnid outgroups (S2 Data) highlights the importance of this pan-allergen protein family. Several cyclophilins

(*n* = 12) were annotated as putative cross-reactive proteins, of note Der f 29 like allergen (DERPT_G9923) exhibited sequence homology (52 – 83 % identity) with cyclophilins from 12 different species (S6 Data). Cross-reactive cyclophilins from HDM, mouse, humans and fungi are well reported in the literature [59-61]. HDM and fungi are frequently co-present in HD [62] with A*lternaria* and *Aspergillus spp.* being the most common source of mould allergens [63]. *D. pteronyssinus airmid* predicted proteins exhibited homology to *Aspergillus fumigatus* (*n* = 19) and A*lternaria* a*lternata* (*n* = 6) allergens. Homology between HDM and fungal proteins may play a role in reported fungal exacerbation of HDM-induced asthmatic symptoms [64, 65]. Excretion of putative allergenic proteins into HD *via* faecal particles would implicate a route of exposure and therefore has significant implications for allergy. Eukaryotic secretion peptides were predicted in 20.7% of allergenic/potentially allergenic proteins. Even if these putative allergens were unable to induce immune responses in their own right, co-presence with other immune modulators may be involved in bystander sensitization. Der p 1 accumulates in HD, levels exceeding 2 µg/g of dust are considered hazardous to the occupants [1]. In addition to being a potent activator of the immune system, Der p 1 has an adjuvant effect, enhancing IgE production against bystander molecules that may be present in HD [66]. Therefore, any protein accumulating in HD should be considered in the context of being a bystander allergen candidate.

Proteomic characterisation of *D. pteronyssinus airmid*

Representative proteome

Analysis of LC-MS/MS spectra obtained from proteomic analysis of *D. pteronyssinus airmid* resulted in the high confidence identification of 3,931 proteins (S7 Data). This representative proteome comprised 31.4% of the predicted proteome of *D. pteronyssinus airmid* with experimental evidence for expression. Protein molecular weight and pI were widely distributed in the representative proteome (Range: 5.25- 3086.2 kDa, pI 3.26 – 12.85) and similar to that of the predicted proteome (Range: 3.94 - 3086.2 kDa, pI 2.63 – 13.27), confirming the protein extraction methods were optimal for the characterisation of *D. pteronyssinus airmid* proteome (Fig 3A). Establishing a representative proteome,

that reflects the methodological limitations of protein extraction and identification is essential for subsequent enrichment analyses. This background proteome is defined by Bessarabova *et al* (2012) as "the complete set of proteins known to be expressed in an organ/tissue/body liquid/cell line of sample origin" [67]. To date, proteomic investigation of *D. pteronyssinus* has extended to a few discrete studies [68-71]. Laboratory HDM populations have been shown to have different reproduction rates to wild-type strains [72] and isolated populations to give rise to geographical allergen variants [7]. Experimental examination of uncharacterised *D. pteronyssinus airmid*-specific proteins, particularly those with evidence for expression, may provide useful insights into the genes evolving within isolated populations. Our representative proteome may be expanded by employing alternative protein extraction methods, use of trypsin alternative or multi-protease digestion and depletion of high abundance proteins [73-75].

Wildtype proteome of *D. pteronyssinus*

Extending the relevance of *D. pteronyssinus* proteomics to the environment is essential, as very little is known about which HDM components are present in inhalable air [10]. Proteomic analysis of 21 Der p 1 positive HD samples revealed the presence of 150 *D. pteronyssinus* proteins (S8 Data), the ten most abundant are given in Table 2, with sequence coverage ranging from 9.6 – 73.3%. Here, it can be seen that allergens Der p 1, 2, 5, 14 and Der f 6 like allergen are amongst the most abundant *D. pteronyssinus* proteins in HD. Der p 1 and Der p 2 are considered major allergens while Der p 5, 6 and 14 are mid-tier IgE binders, all were amongst the ten most abundant proteins in HD [7]. Five non-allergenic proteins were also highly abundant, suggesting that sensitisation and IgE binding is a result of the unique properties of the protein rather than its abundance in house dust. Allergenic/potentially allergenic molecules accounted for almost 51 % of all proteins identified in HD (n = 76). The predicted functions of many of these proteins place them into well-established allergen families, therefore their presence in HD and sequence similarity to known allergens make them interesting targets for future studies seeking to identify new allergens. Comparative analysis between allergenic and non-allergenic components of HD may reveal epitopes or structural characteristics common

to inhalant allergens [10]. Our data illustrates the utility of high sensitivity protein MS as a novel way to identify HDM products in the wild-type environment and has significant implications for the development of immunotherapies that mimic natural exposure. Previously, researchers have examined numerous commercially available diagnostic and therapeutic HDM extracts, finding many were lacking important allergens and some had several fold variations in Der p 1 to Der p 2 ratios [76]. For example, the 2017 international consensus (ICON) report on the clinical consequences of mite hypersensitivity states that the "development of objective methods to assess allergen exposure and environmental control outcomes" are unmet and in need in mite allergy research [1]. Furthermore, the European medicines agencies guideline on the clinical development of products for specific immunotherapy for the treatment of allergic diseases, states that for seasonal allergies "it is mandatory to document the exposure to the relevant allergens" and "it is recommended to document the exposure level for the individual patient especially for the evaluation periods to evaluate the variation of indoor allergens" [77]. Generating diagnostic and therapeutic HDM extracts with allergen content and ratios that mimic natural exposure is of great importance. Characterisation of the factors affecting allergen repertoire and accumulation in different home microhabitats may give rise to much needed strategies for reducing allergen exposure for sensitised individuals [78].

Proteome of laboratory-reared *D. pteronyssinus airmid*

D. pteronyssinus airmid whole protein lysates (WPL) were analysed directly (shotgun) and subjected to fractionation by size exclusion chromatography resulting in identification of 1,948 MB and SM proteins by high sensitivity protein mass spectrometry. Protein identification was confirmed by detection of at least two peptides per protein, and percentage sequence coverage ranged from 0.2 to 99.4 % across the entire protein dataset. Gel filtration fractionation of MB extracts and SM (Fig 3A1-A3) led to the unique identification of 248 and 105 proteins (18% of total identified proteins) from chromatographic fractions. Of the proteins identified, 1,076 proteins (58.2%) were uniquely found in MB extracts (S9 Data), while 99 proteins (0.8%) were solely identified

as secreted proteins (S10 Data). Faecal rich SM was obtained by sieving to remove large mites, therefore smaller mites remained. In a previous proteomic study of *D. pteronyssinus* faeces, it was suggested that their method of faeces sample collection is beneficial over the sieving technique, which unavoidably contains mite bodies and growth media [69]. While culture media-derived proteins (contaminants) will also be detected in proteomic analysis, it is possible to differentiate true dust mite proteins from contaminants by use of a culture media contaminants database as demonstrated in our study. Several functional terms assigned to proteins were differentially represented in MB (S9A-C Data) and SM (S10A-C Data) proteomes compared to the RP.

Several GO terms were differentially represented (Over-represented: n = 161, Under-represented: n = 6, S9A Data), of note, GO Cellular Component terms cytosol (Fisher's P = 5.3E-49, GO:0005829) and mitochondrion (Fisher's P = 8.5E-33, GO:0005739) were the most highly over-represented terms. This finding supports our proteomic strategy, showing significant enrichment of GO terms associated with intracellular activities. EC were highly represented in the MB proteome, assigned to 44.6% of proteins identified (n = 824). Only two enzyme names were found to be differentially represented, acting on peptide bonds (Fisher's P = 1.28E-05) was over-represented and transferring phosphorous-containing groups (Fisher's P = 1.03E-04) underrepresented (S9B Data). NAD(P)-binding domain superfamily (IPR036291) was the most highly over-represented IPS ID (Fishers P = 1.63E-12) of 25, with mobidb-lite IPS ID (Fishers P = 5.58E-17) for Intrinsic disorder protein sequences, representing the most significantly underrepresented of 30 IPS IDs (S9C Data).

All MGA apart from Der p 28 were found in the MB, six were among the most abundant proteins identified (S3 Table). In the absence of Der p 28, nine Der p 28 homologs were identified including four isoallergens (Table 1, S9 Data). MGA were also found to be amongst the most abundant proteins in the SM, including Der p 2 and Der p 14 (S4 Table). Allergenic/potentially allergenic molecules were highly represented in both MB and SM proteomes accounting for 11.4% and 18.5 % of all proteins identified. Homologs of Der p 1 (n = 17), Der p 2 (n = 2) and Der p 23 (n = 1) were identified in the SM. Given

that allergens Der p 1and Der p 2) account for up to 60% of IgE reactivity in HDM sensitised individuals [5], and Der p 23 sensitivity is seen in 79% of HDM allergic patients [79] examination of cross-reactivity between these MGA and excreted homologs is warranted.

MB and SM proteomes were abundant in enzymes, which accounted for 44.6 % and 52.4% of all proteins identified respectively (S9&10 Data), some of these enzymes may be involved in digestion. HDM have long been associated with feeding on shed skin present in HD. While they have been observed to eat skin, the poor nutritional value of keratin makes it unlikely to be a primary food source. Rather, HDM are trophic generalists, they feed on organic debris associated with their proximity to humans [16]. In the laboratory, *D. pteronyssinus* have been grown on diverse culture media including various combinations of wheat bran, wheat flour, dog food, rodent chow, ground porcine liver, dried egg powder, defatted skin scales, and fish food. Most research groups use dried yeast to supplement diets and improve mite population growth [72, 80-83]. HDM have been observed to feed on bacteria and fungi in laboratory experiments [84, 85]. Whether HDM consume bacteria or fungi in a wildtype setting as a nutrient source needs further experimental evidence. Expression and excretion of bacterial and fungal degrading enzymes may indicate a role in digestion [85].

Numerous glycoside hydrolases (EC:3.2.1) were identified in the predicted proteome (*n* = 57) and proteomic datasets (RP, MB & SM) summarised in Table 3. Each glycoside hydrolase enzyme sub-family were represented by at least one proteomic identification. Two predicted lysozymes (EC:3.2.1.17) and one 14.5 kDa bacteriolytic enzyme, Der p38 (DERPT_G10989) were identified with evidence for high expression, these enzymes may be responsible for bacteriolytic activity in HDM extracts [86]. Excreted proteins with predicted activities against major components of fungal cell walls were identified and include; chitinases (*n* = 5), four glycoside hydrolase family 16 members with putative β-1,3 glucanase activity, β-mannosidase (*n* = 1), α-mannosidases (*n* = 4), chitosanase (*n* = 1) and α-N-Acetyl hexosaminidaseine (*n* = 2). Carbohydrate metabolism GO terms were among the 318 over-represented GO terms (Fisher's exact test < 0.05 FDR) in the

excretome and included carbon utilization, hydrolase activity hydrolyzing O-glycosyl compounds, chitin metabolic process, chitin binding and starch metabolic process (S10A Data). Of the putative enzymes identified in the SM proteome (*n* = 457), enrichment analysis showed 35 to be over-represented, many of which related to carbohydrate digestion; Alpha-glucosidase, Chitinase and Alpha-mannosidase (S10B Data). The presence of a predicted secretion signal was a strong indicator of excretion as 22.5% of all proteins identified in the SM proteome contained predicted secretion signals. Moreover, secretion signal peptides were the most highly over-represented IPS ID (Fisher's *P* = 1.08E-26) of 61 (S10C Data).

The expression of the numerous carbohydrate active enzymes listed above provides compelling evidence to support observations of *D. pteronyssinus* feeding on fungi and bacteria and demonstrate that they possess the necessary enzymes to utilise bacteria and fungi as a nutrient source [85]. Feeding on bacteria or fungi within wild-type microhabitats may alter allergen repertoire between homes, as diet has been demonstrated to alter allergen production in laboratory HDM cultures [87]. This new insight compounds the necessity for characterising factors affecting HDM allergen production within the home.

The process of chitin synthesis and remodelling is an integral part of the growth and development of all arthropods. Chitin remodelling enzymes include chitinase, β-N-Acetylhexosaminidase and the highly conserved chitin synthase, a key enzyme in the insect biosynthetic pathway [88]. Proteomic profiling of *D. pteronyssinus airmid* facilitated identification of eight predicted chitinases (EC:3.2.1.14, five β-N-Acetylhexosaminidases (EC:3.21.52) and two chitin synthases (IPR004835) (Table 3) putatively involved in chitin remodelling. These enzymes represent important biocontrol targets, as chitin is absent from vertebrates, dysregulation of these enzymes could provide a much needed method of curtailing HDM populations in the homes of sensitised individuals [12, 88].

Mite Group Allergen Localisation

All proteins are synthesised in the MB prior to excretion, however excreted proteins are likely to accumulate in growth medium and HD. Data regarding localisation of MGA are limited [7]. Localisation is often linked to the degree of protein allergenicity, identifying sites of MGA accumulation may reveal trends of exposure that can be applied to assessing new allergens. In our analysis, the majority of MGA were detected in MB and SM proteomes, the relative amounts in each proteome was used to infer localisation. Proteomic assessment of localisation showed sero-dominant allergens Der p 1, Der p 2 and Der p 23 to accumulate in SM (Table 4), with Der p 1 and Der p 2 being the two most abundant proteins (S4 Table). Der p 23 has previously been reported to be found only in low quantities in SM relative to Der p 2 [79], we observed the same trend, more Der p 2 ms spectra (n = 486) were detected than for Der p23 (n = 90). Allergens Der p 3, Der f 6 like allergen, Der p 9, Der p 15 and Der p 28 were also found to accumulate in the SM (Table 4). Localisation of Der p 3 to SM is consistent with previous observations for Der f 3 [89], the serine peptidases Der p 3, Der f 6 like allergen and Der p 9 were all among the top 10 most abundant SM proteins (S4 Table). Despite a different methodological approach, another study also found Der p 1, Der p 2, Der p 6 and Der p 15 to be major proteins in *D. pteronyssinus* faeces and SM [69].

Analysis of localisation must not be restricted to laboratory-based studies and should include environmental reference samples where possible as demonstrated for allergens Der p 5 and Der p 21. Initial proteomic assessment found Der p 5 and Der p 21 were not excreted under laboratory conditions as they were absent from SM (S10 Data). However, proteomic analysis of HD revealed Der p 5 to be among the top 10 most abundant proteins (Table 2) and Der p 21 the 63rd most abundant, both were identified in at least 16 of the Der p 1 positive dust samples (S8 Data). The location of Der p 5 and Der p 21 in laboratory cultures, present in MB and absent from SM, would indicate they are not excreted, however, their presence in HD shows that there must be other factors that contribute to accumulation of non-excreted allergens in the home. These allergens may accumulate in HD as the mite bodies begin to degrade. The accumulation of dead mites

in laboratory cultures is avoided by regular sub-culturing of mites. Cross-referencing SM with MB associated proteins as demonstrated in Table 4, can infer localisation and can allow for novel insights into modes of accumulation when compared to wild-type dust samples. Our work has demonstrated the utility of high sensitivity mass spectrometry in characterising the complex proteomes of *D. pteronyssinus*. Laboratory cultures show parallels in protein expression with wild-type samples, as most proteins were identified in two or more of the proteomes (Fig 3C) with only four proteins uniquely identified in HD. Additional research is required to characterise the various microhabitats of HDM and factors affecting allergen presence in the home.

Conclusion

Here we performed a comprehensive bioinformatic and proteomic examination of *D. pteronyssinus* airmid describing the expression of 4,002 proteins (S11 Data) and identified 332 potential allergens. High sensitivity mass spectrometry allowed for the description of novel *D. pteronyssinus* components in HD and facilitated qualitative assessment of MGA localisation. This research has expanded the knowledge of proteins utilised by *D. pteronyssinus* for key physiological processes and will form the basis for further research into biocontrol strategies for the medically important HDM.

Acknowledgements

RW is a recipient of an Irish Research Council Enterprise Partnership Scheme Scholar (EPSPG/2014/64). CGPM is funded by an Irish Research Council Government of Ireland Postgraduate Scholarship (Grant No. GOIPG/2015/2242). JM is funded by a postgraduate scholarship from the Irish Research Council, Government of Ireland (grant number GOIPG/2016/1112). Mass spectrometry facilities were funded by Science Foundation Ireland (SFI 12/RI/2346(3)). We acknowledge the DJEI/DES/SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support. Continuous access to laboratory facilities and consumables funding, as well

as expert project guidance and proteomic expertise, was provided by Professor Sean Doyle of Maynooth University. This work was co-funded by airmid healthgroup Ltd, Dublin, Ireland. The authors would like to thank Maire Fox and Shola Ibitoye for their ELISA work on house dust extracts.

References

1. Sanchez-Borges M, Fernandez-Caldas E, Thomas WR, Chapman MD, Lee BW, Caraballo L, et al. International consensus (ICON) on: clinical consequences of mite hypersensitivity, a global problem. World Allergy Organ J. 2017;10(1):14. Epub 2017/04/30. doi: 10.1186/s40413-017-0145-4. PubMed PMID: 28451053; PubMed Central PMCID: PMCPMC5394630.

2. Oseroff C, Christensen LH, Westernberg L, Pham J, Lane J, Paul S, et al. 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(4):577-92. Epub 2016/09/30. doi: 10.1111/cea.12829. PubMed PMID: 27684489; PubMed Central PMCID: PMCPMC5374056.

3. Wan H, Winton HL, Soeller C, Tovey ER, Gruenert DC, Thompson PJ, et al. Der p 1 facilitates transepithelial allergen delivery by disruption of tight junctions. J Clin Invest. 1999;104(1):123-33. Epub 1999/07/07. doi: 10.1172/jci5844. PubMed PMID: 10393706; PubMed Central PMCID: PMCPMC408401.

4. Jacquet A. Innate immune responses in house dust mite allergy. ISRN Allergy. 2013;2013:735031. Epub 2013/06/01. doi: 10.1155/2013/735031. PubMed PMID: 23724247; PubMed Central PMCID: PMCPMC3658386.

5. Hales BJ, Martin AC, Pearce LJ, Laing IA, Hayden CM, Goldblatt J, et al. IgE and IgG anti–house dust mite specificities in allergic disease. Journal of Allergy and Clinical Immunology. 2006;118(2):361-7. doi: https://doi.org/10.1016/j.jaci.2006.04.001.

6. Radauer C, Bublin M, Wagner S, Mari A, Breiteneder H. Allergens are distributed into few protein families and possess a restricted number of biochemical functions. Journal of Allergy and Clinical Immunology. 2008;121(4):847-52.e7. doi: http://dx.doi.org/10.1016/j.jaci.2008.01.025.

7. Thomas WR. Hierarchy and molecular properties of house dust mite allergens. Allergology International. 2015;64(4):304-11. doi: http://dx.doi.org/10.1016/j.alit.2015.05.004.

8. Platts-Mills TA, Heymann PW, Longbottom JL, Wilkins SR. Airborne allergens associated with asthma: particle sizes carrying dust mite and rat allergens measured with a cascade impactor. J Allergy Clin Immunol. 1986;77(6):850-7. Epub 1986/06/01. PubMed PMID: 3711552.

9. Tovey ER, Chapman MD, Platts-Mills TA. Mite faeces are a major source of house dust allergens. Nature. 1981;289(5798):592-3. Epub 1981/02/12. PubMed PMID: 7464922.

10. Thomas WR. Blueprint for the House Dust Mite. Int Arch Allergy Immunol. 175. Switzerland2018. p. 121-5.

11. Zock JP, Heinrich J, Jarvis D, Verlato G, Norback D, Plana E, et al. Distribution and determinants of house dust mite allergens in Europe: the European Community Respiratory Health Survey II. J Allergy Clin Immunol. 2006;118(3):682-90. Epub 2006/09/05. doi: 10.1016/j.jaci.2006.04.060. PubMed PMID: 16950288.

12. Gøtzsche P, Johansen H. House dust mite control measures for asthma. Cochrane Database of Systematic Reviews [Internet]. 2008; (2). Available from: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD001187.pub3/epdf.

13. Arlian LG, Neal JS, Morgan MS, Vyszenski-Moher DL, Rapp CM, Alexander AK. Reducing relative humidity is a practical way to control dust mites and their allergens in homes in temperate climates. Journal of Allergy and Clinical Immunology. 2001;107(1):99-104. doi: 10.1067/mai.2001.112119. PubMed PMID: WOS:000166533300018.

14. deBoer R, Kuller K. Mattresses as a winter refuge for house-dust mite populations. Allergy. 1997;52(3):299-305. PubMed PMID: WOS:A1997WV34300006.

15. Ravee R, Mohd Salleh FI, Goh H-H. Discovery of digestive enzymes in carnivorous plants with focus on proteases. PeerJ. 2018;6:e4914. doi: 10.7717/peerj.4914.

16. Colloff MJ. Dust Mites. Springer, Van Godewijckstraat 30

3311 GX Dordrecht The Netherlands: CSIRO Publishing; 2009.

17. Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 2017. Epub 2017/12/09. doi: 10.1093/molbev/msx319. PubMed PMID: 29220515; PubMed Central PMCID: PMCPMC5850278.

18. Randall TA, Mullikin JC, Mueller GA. The Draft Genome Assembly of Dermatophagoides pteronyssinus Supports Identification of Novel Allergen Isoforms in Dermatophagoides Species. Int Arch Allergy Immunol. 2018;175(3):136-46. Epub 2018/01/11. doi: 10.1159/000481989. PubMed PMID: 29320781; PubMed Central PMCID: PMCPMC5847439.

19. Liu XY, Yang KY, Wang MQ, Kwok JS, Zeng X, Yang Z, et al. High-quality assembly of Dermatophagoides pteronyssinus genome and transcriptome reveals a wide range of novel allergens. J Allergy Clin Immunol. 2018;141(6):2268-71.e8. Epub 2018/01/07. doi: 10.1016/j.jaci.2017.11.038. PubMed PMID: 29305317.

20. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792-7. Epub 2004/03/23. doi: 10.1093/nar/gkh340. PubMed PMID: 15034147; PubMed Central PMCID: PMCPMC390337.

21. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3. Epub 2014/01/24. doi: 10.1093/bioinformatics/btu033. PubMed PMID: 24451623; PubMed Central PMCID: PMCPMC3998144.

22. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics. 2011;27(8):1164-5. Epub 2011/02/22. doi: 10.1093/bioinformatics/btr088. PubMed PMID: 21335321; PubMed Central PMCID: PMCPMC5215816.

23. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5(3):e9490. Epub 2010/03/13. doi: 10.1371/journal.pone.0009490. PubMed PMID: 20224823; PubMed Central PMCID: PMCPMC2835736.

24. Creevey CJ, McInerney JO. Clann: investigating phylogenetic information through supertree analyses. Bioinformatics. 2005;21(3):390-2. Epub 2004/09/18. doi: 10.1093/bioinformatics/bti020. PubMed PMID: 15374874.

25. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 2007;23(1):127-8. Epub 2006/10/20. doi: 10.1093/bioinformatics/btl529. PubMed PMID: 17050570.

26. Risk BA, Spitzer WJ, Giddings MC. Peppy: proteogenomic search software. J Proteome Res. 2013;12(6):3019-25. Epub 2013/04/26. doi: 10.1021/pr400208w. PubMed PMID: 23614390; PubMed Central PMCID: PMCPMC4117251.

27. Waldron R, McGowan J, Gordon N, McCarthy C, Mitchell EB, Doyle S, et al. Draft Genome Sequence of Dermatophagoides pteronyssinus, the European House Dust Mite. Genome Announcements. 2017;5(32):e00789-17. doi: 10.1128/genomeA.00789-17. PubMed PMID: PMC5552995.

28. Conesa A, Götz S. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. International Journal of Plant Genomics. 2008;2008:619832. doi: 10.1155/2008/619832. PubMed PMID: PMC2375974.

29. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25(17):3389-402. Epub 1997/09/01. PubMed PMID: 9254694; PubMed Central PMCID: PMCPMC146917.

30. Hunault G, Jaspard E. LEAPdb: a database for the late embryogenesis abundant proteins. BMC Genomics. 2010;11:221-. doi: 10.1186/1471-2164-11-221. PubMed PMID: PMC2858754.

31. Chan TF, Ji KM, Yim AKY, Liu XY, Zhou JW, Li RQ, et al. The draft genome, transcriptome, and microbiome of Dermatophagoides farinae reveal a broad spectrum of dust mite allergens. Journal of Allergy and Clinical Immunology. 2015;135(2):539-48. doi: 10.1016/j.jaci.2014.09.031. PubMed PMID: WOS:000349372300028.

32. FAO/WHO. Joint FAO/WHO Expert Consultation on Allergenicity of Foods Derived from Biotechnology. In: Food and Agriculture Organization of the United Nations WHO, editor. 2001.

33. Bastian M, Heymann S, Jacomy M. Gephi: An Open Source Software for Exploring and Manipulating Networks. 2009.

34. Owens RA, O'Keeffe G, Smith EB, Dolan SK, Hammel S, Sheridan KJ, et al. Interplay between Gliotoxin Resistance, Secretion, and the Methyl/Methionine Cycle in Aspergillus fumigatus. Eukaryotic cell. 2015;14(9):941-57. Epub 2015/08/27. doi: 10.1128/EC.00055-15. PubMed PMID: 26150413.

35. Sipos G, Prasanna AN, Walter MC, O'Connor E, Balint B, Krizsan K, et al. Genome expansion and lineage-specific genetic innovations in the forest pathogenic

fungi Armillaria. Nat Ecol Evol. 2017;1(12):1931-41. Epub 2017/11/01. doi: 10.1038/s41559-017-0347-8. PubMed PMID: 29085064.

36. Hiller K, Schobert M, Hundertmark C, Jahn D, Münch R. JVirGel: calculation of virtual two-dimensional protein gels. Nucleic Acids Research. 2003;31(13):3862-5. PubMed PMID: PMC168943.

37. O'Keeffe G, Hammel S, Owens RA, Keane TM, Fitzpatrick DA, Jones GW, et al. RNA-seq reveals the pan-transcriptomic impact of attenuating the gliotoxin self-protection mechanism in Aspergillus fumigatus. BMC Genomics. 2014;15(1):894. doi: 10.1186/1471-2164-15-894.

38. Cui Y, Gao C, Zhou Y, Zhou P, Peng M, Lin Y, et al. Phylogenetic analysis of house dust mites. Central European Journal of Medicine. 2010;5(1):69-74. doi: 10.2478/s11536-009-0106-6.

39. Klimov PB, O'Connor B. Is Permanent Parasitism Reversible?-Critical Evidence from Early Evolution of House Dust Mites. Systematic Biology. 2013;62(3):411-23. doi: 10.1093/sysbio/syt008. PubMed PMID: WOS:000318000600005.

40. Dermauw W, Leeuwen T, Vanholme B, Tirry L. The complete mitochondrial genome of the house dust mite Dermatophagoides pteronyssinus (Trouessart): a novel gene arrangement among arthropods. BMC Genomics. 2009;10(1):1-20. doi: 10.1186/1471-2164-10-107.

41. Klimov PB, Bochkov AV, O'Connor BM. Phylogenetic position of the house dust mite subfamily Guatemalichinae (Acariformes: Pyroglyphidae) based on integrated molecular and morphological analyses and different measures of support. Cladistics. 2015;32(3):261-75. doi: 10.1111/cla.12126.

42. Huby RDJ, Dearman RJ, Kimber I. Why Are Some Proteins Allergens? Toxicological Sciences. 2000;55(2):235-46. doi: 10.1093/toxsci/55.2.235.

43. Reithofer M, Jahn-Schmid B. Allergens with Protease Activity from House Dust Mites. International journal of molecular sciences. 2017;18(7):1368. doi: 10.3390/ijms18071368. PubMed PMID: 28653989.

44. Muzzi A, Masignani V, Rappuoli R. The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterials. Drug Discov Today. 2007;12(11-12):429-39. Epub 2007/05/29. doi: 10.1016/j.drudis.2007.04.008. PubMed PMID: 17532526.

45. Tutar Y. Pseudogenes. Comparative and functional genomics. 2012;2012:424526-. Epub 2012/05/07. doi: 10.1155/2012/424526. PubMed PMID: 22611337.

46. Jordan IK, Makarova KS, Spouge JL, Wolf YI, Koonin EV. Lineage-specific gene expansions in bacterial and archaeal genomes. Genome Res. 2001;11(4):555-65. Epub 2001/04/03. doi: 10.1101/gr.166001. PubMed PMID: 11282971; PubMed Central PMCID: PMCPMC311027.

47. Guimarães LC, Florczak-Wyspianska J, de Jesus LB, Viana MVC, Silva A, Ramos RTJ, et al. Inside the Pan-genome - Methods and Software Overview. Current genomics. 2015;16(4):245-52. Epub 2015/08/. doi: 10.2174/1389202916666150423002311. PubMed PMID: 27006628.

48. Ghazy NA, Suzuki T. Desiccation tolerance in diapausing spider mites Tetranychus urticae and T. kanzawai (Acari: Tetranychidae). Experimental and Applied Acarology. 2014;63(1):49-55. doi: 10.1007/s10493-013-9760-0.

49. Arlian LG. WATER-BALANCE AND HUMIDITY REQUIREMENTS OF HOUSE DUST MITES. Experimental & Applied Acarology. 1992;16(1-2):15-35. doi: 10.1007/bf01201490. PubMed PMID: WOS:A1992KH47700004.

50. Kikawada T, Nakahara Y, Kanamori Y, Iwata KI, Watanabe M, McGee B, et al. Dehydration-induced expression of LEA proteins in an anhydrobiotic chironomid. Biochemical and Biophysical Research Communications. 2006;348(1):56-61. doi: 10.1016/j.bbrc.2006.07.003. PubMed PMID: WOS:000239796100007.

51. Solomon A, Salomon R, Paperna I, Glazer I. Desiccation stress of entomopathogenic nematodes induces the accumulation of a novel heat-stable protein. Parasitology. 2000;121:409-16. doi: 10.1017/s0031182099006563. PubMed PMID: WOS:000090071900009.

52. Browne JA, Dolan KM, Tyson T, Goyal K, Tunnacliffe A, Burnell AM. Dehydration-Specific Induction of Hydrophilic Protein Genes in the Anhydrobiotic Nematode Aphelenchus avenae. Eukaryotic Cell. 2004;3(4):966-75. doi: 10.1128/EC.3.4.966-975.2004. PubMed PMID: PMC500876.

53. Gusev O, Suetsugu Y, Cornette R, Kawashima T, Logacheva MD, Kondrashov AS, et al. Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge. Nature Communications. 2014;5:4784. doi: 10.1038/ncomms5784

https://www.nature.com/articles/ncomms5784#supplementary-information.

54. Hyndman SJ, Vickers LM, Htut T, Maunder JW, Peock A, Higenbottam TW. A randomized trial of dehumidification in the control of house dust mite. Clin Exp Allergy. 2000;30(8):1172-80. Epub 2000/08/10. PubMed PMID: 10931126.

55. Radauer C, Nandy A, Ferreira F, Goodman RE, Larsen JN, Lidholm J, et al. Update of the WHO/IUIS Allergen Nomenclature Database based on analysis of allergen sequences. Allergy. 2014;69(4):413-9. doi: 10.1111/all.12348.

56. Hauser M, Roulias A, Ferreira F, Egger M. Panallergens and their impact on the allergic patient. Allergy, Asthma, and Clinical Immunology : Official Journal of the Canadian Society of Allergy and Clinical Immunology. 2010;6(1):1-. doi: 10.1186/1710-1492-6-1. PubMed PMID: PMC2830198.

57. McKenna OE, Asam C, Araujo GR, Roulias A, Goulart LR, Ferreira F. How relevant is panallergen sensitization in the development of allergies? Pediatric Allergy and Immunology. 2016;27(6):560-8. doi: 10.1111/pai.12589. PubMed PMID: PMC5006871.

58. Fæste CK, Jonscher KR, Dooper MMWB, Egge-Jacobsen W, Moen A, Daschner A, et al. Characterisation of potential novel allergens in the fish parasite Anisakis simplex. EuPA open proteomics. 2014;4:140-55. doi: 10.1016/j.euprot.2014.06.006. PubMed PMID: PMC4840464.

59. Flückiger S, Fijten H, Whitley P, Blaser K, Crameri R. Cyclophilins, a new family of cross-reactive allergens. European Journal of Immunology. 2001;32(1):10-7. doi: 10.1002/1521-4141(200201)32:1<10::AID-IMMU10>3.0.CO;2-I.

60. Glaser Andreas G, Limacher A, Flückiger S, Scheynius A, Scapozza L, Crameri R. Analysis of the cross-reactivity and of the 1.5 Å crystal structure of the Malassezia sympodialis Mala s 6 allergen, a member of the cyclophilin pan-allergen family. Biochemical Journal. 2006;396(Pt 1):41-9. doi: 10.1042/BJ20051708. PubMed PMID: PMC1449988.

61. Joshi SS, Gan HY, Chew FT. Cross Comparison between the IgE Binding Profiles of Cyclophilins from Human, Mouse, Dust Mites and Fungi. Journal of Allergy and Clinical Immunology. 2006;117(2):S120. doi: 10.1016/j.jaci.2005.12.481.

62. Norbäck D, Markowicz P, Cai G-H, Hashim Z, Ali F, Zheng Y-W, et al. Endotoxin, Ergosterol, Fungal DNA and Allergens in Dust from Schools in Johor Bahru, Malaysia- Associations with Asthma and Respiratory Infections in Pupils. PLOS ONE. 2014;9(2):e88303. doi: 10.1371/journal.pone.0088303.

63. Żukiewicz-Sobczak WA. The role of fungi in allergic diseases. Advances in Dermatology and Allergology/Postępy Dermatologii I Alergologii. 2013;30(1):42-5. doi: 10.5114/pdia.2013.33377. PubMed PMID: PMC3834689.

64. Zhang Z, Biagini Myers JM, Brandt EB, Ryan PH, Lindsey M, Mintz-Cole RA, et al. beta-Glucan exacerbates allergic asthma independent of fungal sensitization and promotes steroid-resistant TH2/TH17 responses. J Allergy Clin Immunol. 2017;139(1):54-65.e8. Epub 2016/05/26. doi: 10.1016/j.jaci.2016.02.031. PubMed PMID: 27221135; PubMed Central PMCID: PMCPMC5073040.

65. Snelgrove RJ, Gregory LG, Peiró T, Akthar S, Campbell GA, Walker SA, et al. Alternaria-derived serine protease activity drives IL-33-mediated asthma exacerbations. The Journal of allergy and clinical immunology. 2014;134(3):583-92.e6. doi: 10.1016/j.jaci.2014.02.002. PubMed PMID: 24636086.

66. Gough L, Sewell HF, Shakib F. The proteolytic activity of the major dust mite allergen Der p 1 enhances the IgE antibody response to a bystander antigen. Clin Exp Allergy. 2001;31(10):1594-8. Epub 2001/10/27. PubMed PMID: 11678860.

67. Bessarabova M, Ishkin A, JeBailey L, Nikolskaya T, Nikolsky Y. Knowledge-based analysis of proteomics data. BMC bioinformatics. 2012;13 Suppl 16(Suppl 16):S13-S. doi: 10.1186/1471-2105-13-S16-S13. PubMed PMID: 23176192.

68. Bordas-Le Floch V, Le Mignon M, Bussieres L, Jain K, Martelet A, Baron-Bodo V, et al. A combined transcriptome and proteome analysis extends the allergome of house dust mite Dermatophagoides species. PLoS One. 2017;12(10):e0185830. Epub 2017/10/06. doi: 10.1371/journal.pone.0185830. PubMed PMID: 28982170; PubMed Central PMCID: PMCPMC5628879.

69. Erban T, Harant K, Hubert J. Detailed two-dimensional gel proteomic mapping of the feces of the house dust mite Dermatophagoides pteronyssinus and comparison with D. farinae: Reduced trypsin protease content in D. pteronyssinus and different isoforms. J Proteomics. 2017;162:11-9. Epub 2017/04/27. doi: 10.1016/j.jprot.2017.04.021. PubMed PMID: 28442447.

70. Choopong J, Reamtone O, Sookrung N, Seesuay W, Indrawattana N, Sakolvaree Y, et al. Proteome, Allergenome, and Novel Allergens of House Dust Mite, Dermatophagoides farinae. Journal of Proteome Research. 2016;15(2):422-30. doi: 10.1021/acs.jproteome.5b00663. PubMed PMID: WOS:000369771700007.

71. An S, Chen LL, Long CB, Liu XY, Xu XM, Lu XR, et al. Dermatophagoides farinae Allergens Diversity Identification by Proteomics. Molecular & Cellular

Proteomics. 2013;12(7):1818-28. doi: 10.1074/mcp.M112.027136. PubMed PMID: WOS:000321559000006.

72. Hart BJ, Crowther D, Wilkinson T, Biddulph P, Ucci M, Pretlove S, et al. Reproduction and development of laboratory and wild house dust mites (Acari : Pyroglyphidae) and their relationship to the natural dust ecosystem. Journal of Medical Entomology. 2007;44(4):568-74. doi: 10.1603/0022-2585(2007)44[568:radola]2.0.co;2. PubMed PMID: WOS:000247748400003.

73. Klont F, Bras L, Wolters JC, Ongay S, Bischoff R, Halmos GB, et al. Assessment of Sample Preparation Bias in Mass Spectrometry-Based Proteomics. Anal Chem. 2018;90(8):5405-13. Epub 2018/04/03. doi: 10.1021/acs.analchem.8b00600. PubMed PMID: 29608294; PubMed Central PMCID: PMCPMC5906755.

74. Tsiatsiani L, Heck AJ. Proteomics beyond trypsin. Febs j. 2015;282(14):2612-26. Epub 2015/04/01. doi: 10.1111/febs.13287. PubMed PMID: 25823410.

75. Fonslow BR, Stein BD, Webb KJ, Xu T, Choi J, Park SK, et al. Digestion and depletion of abundant proteins improves proteomic coverage. Nature methods. 2013;10(1):54-6. doi: 10.1038/nmeth.2250. PubMed PMID: PMC3531578.

76. Casset A, Mari A, Purohit A, Resch Y, Weghofer M, Ferrara R, et al. Varying allergen composition and content affects the in vivo allergenic activity of commercial Dermatophagoides pteronyssinus extracts. Int Arch Allergy Immunol. 2012;159(3):253-62. Epub 2012/06/23. doi: 10.1159/000337654. PubMed PMID: 22722650; PubMed Central PMCID: PMCPMC4594775.

77. Guideline on the Clinical Development of Products for Specific Immunotherapy for the Treatment of Allergic Diseases, (2008).

78. Vidal-Quist JC, Ortego F, Lombardero M, Castanera P, Hernandez-Crespo P. Allergen expression in the European house dust mite Dermatophagoides pteronyssinus throughout development and response to environmental conditions. Med Vet Entomol. 2015;29(2):137-46. doi: 10.1111/mve.12102. PubMed PMID: 25565338.

79. Weghofer M, Grote M, Resch Y, Casset A, Kneidinger M, Kopec J, et al. Identification of Der p 23, a peritrophin-like protein, as a new major Dermatophagoides pteronyssinus allergen associated with the peritrophic matrix of mite fecal pellets. J Immunol. 2013;190(7):3059-67. Epub 2013/03/06. doi: 10.4049/jimmunol.1202288. PubMed PMID: 23460742; PubMed Central PMCID: PMCPMC4582595.

80. Andersen A. Nutritional value of yeast for Dermatophagoides pteronyssinus (Acari: Epidermoptidae) and the antigenic and allergenic composition of extracts during extended culturing. J Med Entomol. 1991;28(4):487-91. Epub 1991/07/01. PubMed PMID: 1941906.

81. Hubert J, Kopecky J, Sagova-Mareckova M, Nesvorna M, Zurek L, Erban T. Assessment of Bacterial Communities in Thirteen Species of Laboratory-Cultured Domestic Mites (Acari: Acaridida). J Econ Entomol. 2016;109(4):1887-96. Epub 2016/04/29. doi: 10.1093/jee/tow089. PubMed PMID: 27122496.

82. Eraso E, Guisantes JA, Martinez J, Saenz-de-Santamaria M, Martinez A, Palacios R, et al. Kinetics of allergen expression in cultures of house dust mites, Dermatophagoides pteronyssinus and D. farinae (Acari: Pyroglyphidae). J Med Entomol. 1997;34(6):684-9. Epub 1998/01/24. PubMed PMID: 9439123.

83. Arlian LG, Morgan MS. Reproductive biology of Euroglyphus maynei with comparisons to Dermatophagoides farinae and D. pteronyssinus. Exp Appl Acarol. 2015;66(1):1-9. Epub 2015/01/31. doi: 10.1007/s10493-015-9882-7. PubMed PMID: 25633266.

84. Naegele A, Reboux G, Scherer E, Roussel S, Millon L. Fungal food choices of Dermatophagoides farinae affect indoor fungi selection and dispersal. International Journal of Environmental Health Research. 2013;23(2):91-5. doi: 10.1080/09603123.2012.699029. PubMed PMID: WOS:000312440500001.

85. Erban T, Hubert J. Digestive function of lysozyme in synanthropic acaridid mites enables utilization of bacteria as a food source. Exp Appl Acarol. 2008;44(3):199-212. Epub 2008/03/22. doi: 10.1007/s10493-008-9138-x. PubMed PMID: 18357505.

86. Tang VH, Chang BJ, Srinivasan A, Mathaba LT, Harnett GB, Stewart GA. Skin-associated Bacillus, staphylococcal and micrococcal species from the house dust mite, Dermatophagoides pteronyssinus and bacteriolytic enzymes. Experimental and Applied Acarology. 2013;61(4):431-47. doi: 10.1007/s10493-013-9712-8. PubMed PMID: WOS:000326398200004.

87. Avula-Poola S, Morgan MS, Arlian LG. Diet Influences Growth Rates and Allergen and Endotoxin Contents of Cultured Dermatophagoides farinae and Dermatophagoides pteronyssinus House Dust Mites. International Archives of Allergy and Immunology. 2012;159(3):226-34. doi: 10.1159/000336026.

88. Merzendorfer H, Zimoch L. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. Journal of Experimental Biology. 2003;206(24):4393.

89. Zhan Z-k, Ji K-m, Liu X-y, Liu Z-g, Li M, Chen J-j, et al. Monoclonal antibodies against recombinant Der f 3 reveal localization of Der f 3 in the gut and faecal pellets of Dermatophagoides farinae. Experimental and Applied Acarology. 2010;52(1):63-71. doi: 10.1007/s10493-010-9349-9.

Figures

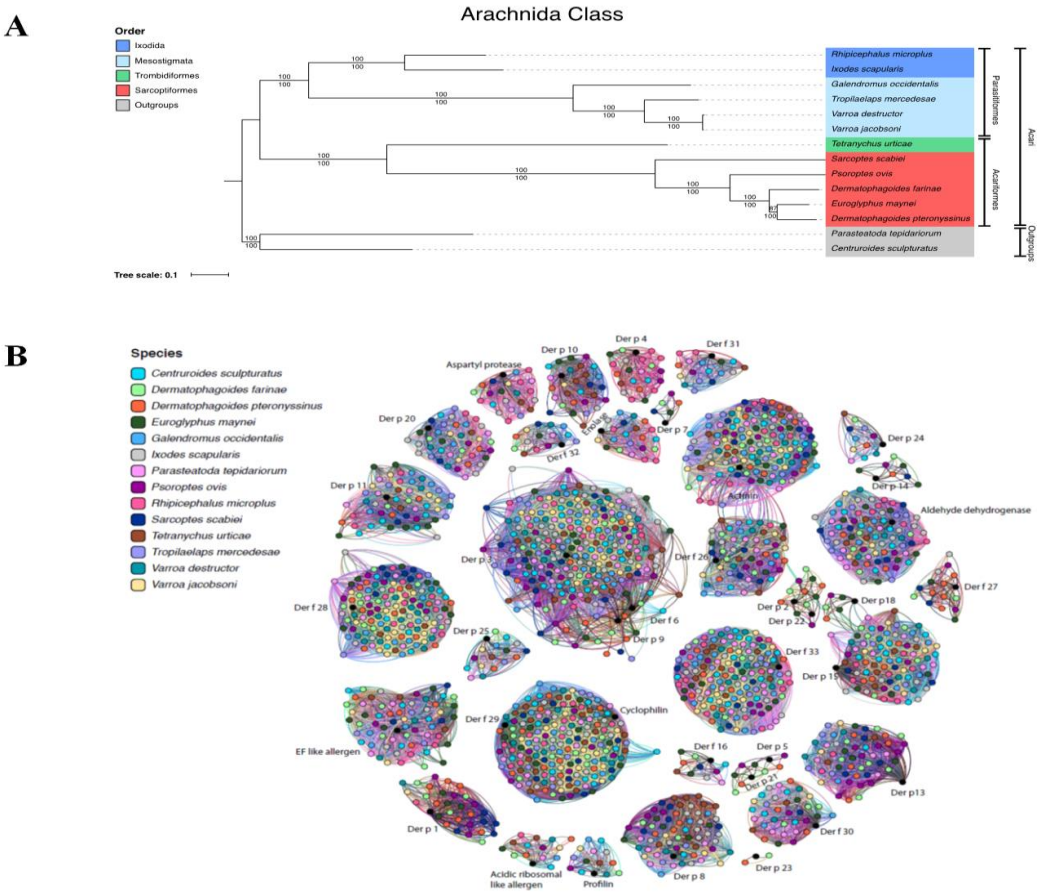


Figure 1. Arachnida Phylogenomic Species Tree and Network of Mite Group Allergen Orthologs in Genomes of twelve Acari and two Arachnids.

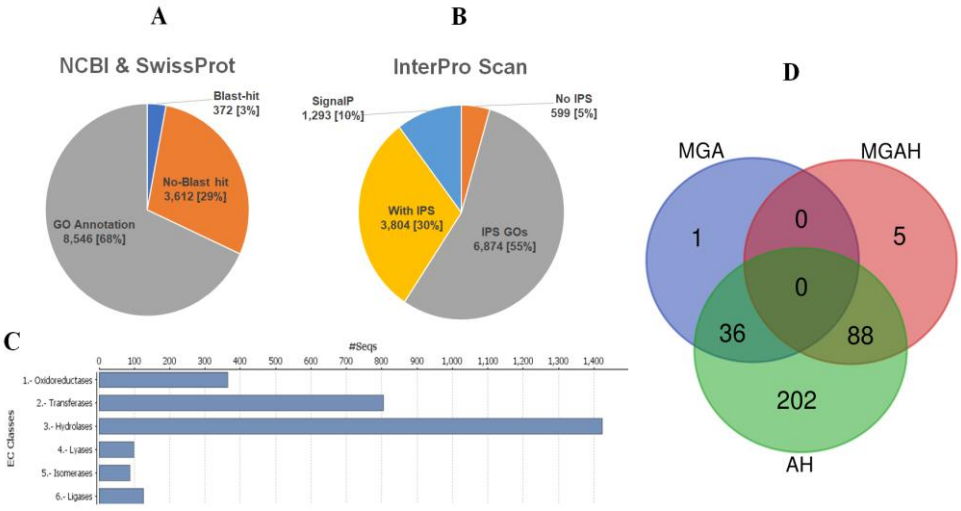


Figure 2. **Annotation of *D. pteronyssinus airmid* Predicted Proteome**.

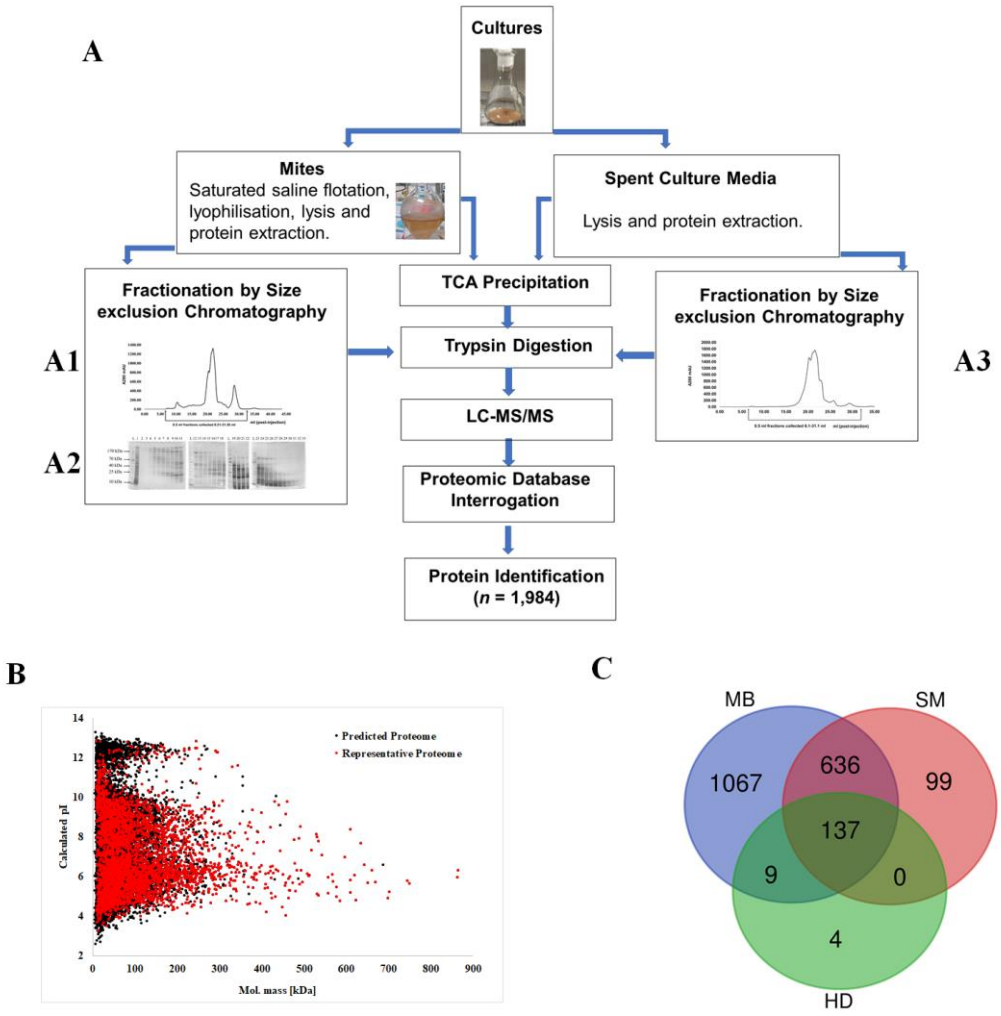


Figure 3. Proteomic Characterisation of *D. pteronyssinus airmid*

Figure legends

Figure 1. Arachnida Phylogenomic Species Tree and Network of Mite Group Allergen Orthologs in Genomes of twelve Acari and two Arachinds

A. Phylogenomic species tree of 11 Acari species inferred using maximum-likelihood supermatrix (111 ortholog families, 77,878 aa aligned, LG+G+I+F model) and supertree (2,796 ortholog families) methods. Bootstrap support values are given for each branch, with values above branches corresponding to the supermatrix phylogeny and values below corresponding to the supertree phylogeny. Both supermatrix and supertree methods infer a strongly supported (87% and 100% bootstrap support respectively) sister group relationship between *D. pteronyssinus* and *E. maynei* to the exclusion of *D. farinae*. Therefore, in evolutionary terms *D. pteronyssinus* and *E. maynei* are more closely related to one another even though *D. pteronyssinus* and *D. farinae* belong to the same genus. **B**. Network map depicting abundance of mite group allergen homologs in each species. Homologous allergen group together forming an interconnected network distinct from non-homologous allergens i.e 2 & 22, groups 3, 6 & 9, groups 5 & 21, and groups 15 & 18.

Figure 2. **Annotation of *D. pteronyssinus airmid* Predicted Proteome**.

Annotation of *D. pteronyssinus airmid* proteins conducted using Blast2GO to search SwissProt, NCBI and InterPro databases for homology to known proteins and functional domains. **A**. GO annotations were assigned to 68.2% of the predicted proteome (*n* = 8,546). A small number of proteins had Blast hits with proteins in NCBInr or SwissProt databases but were not assigned a GO term (*n* = 372). The remaining proteins (*n* = 3,612) were not homologous with proteins in these databases. Most of the predicted proteins were annotated with more than one GO category (Biological Processes, Molecular function, Cellular component) with median number of assigned GO terms being 9 (Range: 0-164). **B.** InterPro scan feature of Blast2GO enabled assignment of InterPro IDs to 95.5% of predicted proteins (*n* = 11,971). InterPro GO IDs were assigned to 55 % (*n* = 6, 874) of the predicted proteome and predicted eukaryotic secretion peptides to 10.3%

of the predicted proteome (*n* = 1,293). A further 3, 804 proteins were assigned InterPro Scan (IPS) IDs. Multi-database Blast2GO workflow enabled 96% of the predicted proteome to be assigned some form of annotation (*n* = 11,996). A small number of proteins had BLAST hits with proteins in NCBInr/SwissProt databases but were not assigned a GO term (*n* = 372). Several predicted proteins (*n* = 294) were assigned GO annotations but BLASTp hit alignments did not meet the required threshold of significance. The remaining proteins (*n* = 3,612) were not homologous with proteins in these databases **C**. Enzyme code classes assigned to *D. pteronyssinus airmid* predicted proteins (*n* = 2,689), representing 21.5% of *D. pteronyssinus airmid* predicted proteome. Hydrolyses (EC:3.0) formed the largest enzyme category (*n* = 1,244) and Isomerases (EC:5.0) the smallest (*n* = 87). **D.**Venn diagram depicting overlapping categorisation of allergenic and predicted allergenic proteins. *D. pteronyssinus airmid* predicted proteins were assigned into unambigous allergen groupings based upon BlastP homology to known allergens. We found full predicted proteins for all reported MGA (Groups 1-11, 13-16, 18, 20-33) and Seven non-chronalogical allergens (81.5 - 100 % identity). Subsequent blast hits were considered MGAH. Predicted proteins with potential cross-reactivity (FAO/WHO, 2001) with allergens from other species (Uniprot "allergenome") were annotated as Allergen homolog (AH). Many Allergenic/potentially allergenic proteins were present in multiple allergen analyses, in total 332 allergenic/potentially allergenic proteins were identified.

Figure 3. **Proteomic Characterisation of *D. pteronyssinus airmid***

A. Proteomic Strategy for characterisation of *D. pteronyssinus airmid*. Flow diagram depicting proteomic strategy for characterisation of *D. pteronyssinus airmid* excretome and mite body proteome utilising whole sample analysis and sample fractionation analysis. **A1**. Gel filtration chromatography of *D. pteronyssinus airmid* whole body homogenate. Whole body homogenate protein extract (4.5 mg; 500 µl injection) separated by size exclusion chromatography (Superdex 200 10/300 GL; 0.4 ml/min flow rate). 0.5 ml fractions collected between 6.21 and 32.26 ml post-injection. **A2.** SDS-

PAGE analysis of *D. pteronyssinus airmid* whole body homogenate fractions. SDS-PAGE (4-20%) with silver staining of whole body homogenate protein extract (Lane 1; 10 µg) from *D. pteronyssinus airmid* and fractions from gel filtration (6.2 -22.2 ml post-injection) of same (Lanes 2-13, 17 µl loaded). SDS-PAGE visualisation of fractions (0.5ml) depicting fractionation of complex protein mixture into reduced complexity extracts containing proteins of approximately similar sizes. Fractions, 7.2 – 32.2 ml, post injection (*n* = 50) were pooled to give 1 -1.5 ml aliquots (*n* = 21),100 µl was processed for LC-MS/MS. **A3.** Gel filtration chromatography of *D. pteronyssinus airmid* spent growth medium (SM). Protein extract of *s*pent growth media (0.2 mg; 500 µl injection) separated by size exclusion chromatography (Superdex 200 10/300 GL; 0.4 ml/min flow rate). 0.5 ml fractions collected between 6.1 and 32.1 ml post-injection for SDS-PAGE analysis. Proteins in gel filtration fractions were acidic and therefore unsuitable for analysis by SDS-PAGE. Therefore, fractions were pooled and analysed according the methods used for mite body extract fractions. Fractions, 7.1 – 32.1 ml, post injection (*n* = 50) were pooled to give 1 -1.5 ml aliquots (*n* = 21), 100 µl was processed for LC-MS/MS. **B.** Alignment of theoretical molecular mass and pI from *D. pteronyssinus airmid* PP *(n* = 12, 530) and representative proteome (*n* = 3,931) Software: http://www.juirgel.de Hiller et al., (2006). Proteins > 900 kDa excluded for graphing purposes (DERPT_G11449, DERPT_G9606, DERPT_G4775, DERPT_G8007,3088 kDa, 2021 kDa, 1122 kDa & 1001 kDa respectively). Predicted proteins with strong proteomic evidence for expression (2+ medium confidence peptides) accounted for 31.4% of the predicted proteome (*n* = 3, 931). Calculated molecular mass and pI (Range: 5.25- 3086.2 kDa, pI 3.26 – 12.85) of representative proteome was approximate to that of the predicted proteome (Range: 3.94 - 3086.2 kDa, pI 2.63 – 13.27). Median protein size in the *D. pteronyssinus airmid* PP was 54.25 kDa. The largest predicted protein coding gene was DERPT_G11449 at 79,548 bp and encoded a paladin (Titin) protein homolog (26,516 a.a, 3086.2 kDa, pI 4.93). The smallest predicted protein was DERPT_G12515 (36 a.a, 3.94 kDa, pI 6.04) of unknown function, encoded by a 106 bp gene. The median protein pI was calculated to be 7.42, the protein with lowest predicted pI of 2.63 was DERPT_G12367 (50 a.a, 5.45 kDa), DERPT_G11425 had the highest

predicted pI at 13.27 (67 a.a, 6.94 kDa). **C.** Venn Diagram depicting overlap between Mite Body (MB), Spent Culture Medium (SM) and House Dust (HD) Proteomes, totalling 1,952 proteins identified.

Tables

Table 1. Mite Group Allergens (MGA), Isoallergens and MGA Homologs.

Group[A] Allergen	Query Sequence ID[B]	Designated MGA Sequence ID[C]	Description[D]	Biochemical Function	% Identities	E-value	Isoallergens	No. Homologs[E]
1	P08176	DERPT_G1283	Der p 1	Cysteine protease 6	99.7	0		31
2	Q1H8P8	DERPT_G8792	Der p 2	Lipid binding	97.9	2.0E-104		6
3	P39675	DERPT_G8859	Der p 3	Trypsin 2	98.1	0		41
4	Q9Y197	DERPT_G360	Der p 4	a-Amylase 3	100	0	DERPT_G359	2
5	P14004	DERPT_G7260	Der p 5	Structural protein 2	99.2	1.1E-91		2
6	DEFA_160240	DERPT_G9187	Der f 6 like Allergen	Serine proteases	83.8	5.4E-140		39
7	P49273	DERPT_G8042	Der p 7	Unknown	99.5	8.3E-156		4
8	Q2YFE5	DERPT_G10534	Der p 8	Glutathione transferase 2	99.5	3.7E-166	DERPT_G10535	5
9	Q7Z163	DERPT_G7100	Der p 9	Serine protease	98.5	0		35
10	O18416	DERPT_G8047	Der p 10	Tropomyosin 5	99.3	0		5
11	Q6Y2F9	DERPT_G8381	Der p 11	Paramyosin 11	99.2	0		6
13	E0A8N8	DERPT_G8964	Der p 13	Fatty acid binding 2	100	3.3E-91		8
14	Q8N0N0	DERPT_G9820	Der p 14	Vitellogenin: egg yolk storage 6	99.6	0		3
15	Q4JK69	DERPT_G3350	Der p 15	Chitinase	99.3	0		13
16	A0A291KZA1	DERPT_G5615	Der f 16 like Allergen	Gelsolin: actin binding	99.8	0		2
18	Q4JK71	DERPT_G784	Der p 18	Chitinase	96.3	0		11
20	B2ZSY4	DERPT_G59	Der p 20	Arginine kinase	100	0	DERPT_G10997	3
21	Q2L7C5	DERPT_G7259	Der p 21	Structural protein 2	100	6.1E-98		2

Group[A] Allergen	Query Sequence ID[B]	Designated MGA Sequence ID[C]	Description[D]	Biochemical Function	% Identities	E-value	Isoallergens	No. Homologs[E]
22	A0A291KZA0	DERPT_G5267	Der f 22 like Allergen	MD-2–related lipid recognition	100	1.7E-106		7
23	L7N6F8	DERPT_G11207	Der p 23	Chitin-binding domain type 2	100	4.9E-65		14
24	A0A0K2GUJ4	DERPT_G5941	Der p 24 Allergen	Ubiquinol-cytochrome c reductase	100	3.1E-86		1
25	A0A291KYZ7	DERPT_G6990	Der f 25 like Allergen	Triosephosphate isomerase	100	0		2
26	A0A088SAG5	DERPT_G10519	Der f 26 like Allergen	Myosin light chain	92.4	7.6E-99		9
27	A0A291KZ97	DERPT_G8127	Der f 27 like Allergen	Serpin	82	0	DERPT_G8125 DERPT_G8126 DERPT_G8127	14
28	A0A291KZD8	DERPT_G6942	Der f 28 like Allergen	Heat shock protein 70	99.5	0	DERPT_G4999 DERPT_G585 DERPT_G7298 DERPT_G7750 DERPT_G9299	16
29	A1KXG2	DERPT_G9923	Der f 29 like Allergen	Cyclophilin	83.2	9.7E-111	DERPT_G1407 DERPT_G9923	14
30	Q962I7	DERPT_G8780	Der f 30 like Allergen Ferritin	Ferritin	100	5.7E-136	DERPT_G4895 DERPT_G9802	4
31	A0A088SAY1	DERPT_G5053	Der f 31 like Allergen	Cofilin	100	1.7E-108		1

Group[A] Allergen	Query Sequence ID[B]	Designated MGA Sequence ID[C]	Description[D]	Biochemical Function	% Identities	E-value	Isoallergens	No. Homologs[E]
32	A0A291KZC9	DERPT_G5605	Der f 32 like Allergen	Pyrophosphatase	100	3.0E-174		1
33	A0A2L0EBJ4	DERPT_G7263	Der f 33 like Allergen	a-tubulin	100	0	DERPT_G10901 DERPT_G8894	12
Non-Chronological Allergens	DEFA_097280	DERPT_G5569	Der f Acidic ribosomal like Allergen	Alpha Actin	81.6	1.3E-37		18
	L7UZ85	DERPT_G9110	Der f Actinin like Allergen	Aldehyde dehydrogenase	88.7	2.8E-141		11
	DEFA_095270	DERPT_G5156	Der f Aldehyde dehydrogenase like Allergen	Acidic Ribosomal Protein	88.8	0		3
	X4ZE83	DERPT_G12026	Der f Enolase like Allergen	Elongation Factor	93.8	0		15
	DEFA_072440	DERPT_G5247	Der f EF Elongation Factor like Allergen	Enolase	98.5	0		2
	DEFA_098190	DERPT_G10356	Der f Eukaryotic aspartyl protease like Allergen	Eukaryotic aspartyl proteas	86.2	0		1
	DEFA_029120	DERPT_G5894	Der f Profilin like allergen	Profilin	96.2	3.6E-93		1

Allergen Group[A], Chronological and non-chronological allergens described for mites. Query Sequence ID[B], UniProt Accession or *D. farinae* Accession (Chan *et al*., 2015). Sequence ID[C], *D. pteronyssinus airmid* protein sequence ID. Description[D], mite group allergen name.

Table 2. Top 10 Most Abundant *D. pteronyssinus* Proteins Identified in House Dust Protein Extracts

Sequence IDs[A]	Description[B]	N: Samples[C]	N: Unique peptides[D]	Sequence coverage [%]	Mol. Mass [kDa]	N: MS/MS count[E]	LFQ Intensity[F]
DERPT_G8792	Der p 2 Allergen; Proposed Mon-Allergen, Allergen Homolog (Der f2, Eur m2, Der s2)	21	7	56.2	15.9	178	4.3E+10
DERPT_G1283	Der p 1 Allergen, Proposed Sten-Allergen, Allergen Homolog (Der f1, Der m1, Pso o1)	21	11	18.8	96.0	129	1.9E+10
DERPT_G212	Polyubiquitin	21	9	18	287.2	133	1.5E+10
DERPT_G10213	Glutamate receptor kainate 3	17	12	17.7	99.7	97	5.1E+09
DERPT_G9820	Der p14 Allergen; Proposed Sten-Allergen, Allergen Homolog (Sar s14, Eur m14, Der f14)	17	51	36.2	191.4	212	4.9E+09
DERPT_G7260	Der p5 Allergen; Allergen Homolog (Der f5)	17	8	21.5	36.2	70	4.5E+09
DERPT_G4157	Der p36 Allergen	20	7	64.7	24.5	73	3.4E+09
DERPT_G114	Proposed Sten-Allergen, Allergen Homolog (ole e5, Sola 1 SOD)	13	7	74.3	15.7	60	3.4E+09
DERPT_G9187	Der f6 like allergen; Allergen Homolog (Blo t6)	18	11	9.6	228.1	122	3.2E+09
DERPT_G9697	Sucrase- intestinal, alpha-1,4-glucosidase activity	18	16	16.6	202.1	119	2.8E+09

Sequence ID[A], *D. pteronyssinus airmid* protein sequence ID. Description[B], annotations assigned by Blast2GO. N: Samples[C], number of extracted dust samples in which specified protein was identified. N: Unique peptides[D], number of unique (not present in any other protein sequence in the predicted proteome) peptides identified by LC-MS/MS for specified protein. N: MS/MS count[E], sum of peptides selected for ms/ms analysis. LFQ Intensity[F], label free quantification intensity. (Software: Maxquant version 1.6.2.10, Perseus version 1.6.2.2)

Table 3. Carbohydrate Degrading Enzymes.

		Proteomic Evidence			
Carbohydrate active enzyme	**EC No/ InterPro ID**	**PP[A]**	**RP[B]**	**MB[C]**	**SM[D]**
Glycosidases	3.2	63	44	32	30
Glycosyl hydrolases	3.2.1	57	42	30	29
a-amylase	3.2.1.1	2	1	1	1
Chitinase	3.2.1.14	10	8	3	5
Lysozyme	3.2.1.17	3	1	2	1
Alpha-Glucosidase	3.2.1.20	4	4	4	4
Alpha-Galactosidase	3.2.1.22	3	2	2	2
Beta-Galactosidase	3.2.1.23	3	2	2	2
Alpha-Mannosidase	3.2.1.24	11	8	5	5
Beta-Mannosidase	3.2.1.25	1	1	1	1
Trehalase	3.2.1.28	2	2	0	0
Beta-Glucuronidase	3.2.1.31	1	1	1	1
amylo-alpha-1,6-glucosidase	3.2.1.33	2	1	1	0
Hyaluronoglucosaminidase	3.2.1.35	2	0	1	0
Glucosylceramidase	3.2.1.45	2	2	2	2
Alpha-L-Fucosidase	3.2.1.51	3	2	2	2
Alpha-*N*-Acetyl hexosaminidase	3.2.1.52	5	5	2	2
mannosyl-glycoprotein endo-β-N-acetylglucosaminidase	3.2.1.96	1	1	0	0
Chitosanase	3.2.1.132	2	1	1	1
Glysoside Hydrolase family 16	IPR000757	6	4	3	14
Glyco Hydrolase family 18 (W/O EC)[F]	IPR001223	9	8	4	4

PP[A], Predicted Proteome. RP[B], set of all proteins identified in the Representative proteome. MB[C], set of all proteins identified in the mite body. SM[D], set of all proteins identified in the Excretome. HD[E], set of all *D. pteronyssinus* proteins identified in house dust. Glyco Hydrolase family 18 (W/O EC)[F], proteins annotated as Glyco Hydrolase family 18 but lacking enzyme code annotation.

Table 4. Proteomic Evaluation of MGA Localization in *D. pteronyssinus airmid*

MGA[A]	Sequence ID[B]	N: Mol. mass [kDa][C]	N: Sequence coverage [%][D]	N: Score[E]	N: MS/MS count[F]	N: LFQ Intensity[G]	Localisation[H]
1	DERPT_G1283	96.0	41.3	323.31	1511	4.9E+11	**SM**
			32.2	323.31	966	2.7E+11	
2	DERPT_G8792	15.9	78.8	323.31	486	5.2E+11	**SM**
			76.7	323.31	419	6.7E+11	
3	DERPT_G8859	28.1	67.4	323.31	510	4.4E+11	**SM**
			66.7	323.31	250	5.1E+10	
4	DERPT_G360	60.3	75.5	323.31	370	1.5E+11	**MB**
			76.9	323.31	410	1.8E+11	
5	DERPT_G7260	36.2	25.5	323.31	188	2.7E+10	**MB**
6	DERPT_G9187	228.1	14.4	323.31	468	2.6E+11	**SM**
			17.9	323.31	192	2.4E+10	
7	DERPT_G8042	28.5	48.6	256.04	105	1.7E+10	**MB**
			53.7	323.31	273	6E+10	
8	DERPT_G10534	25.6	57.1	248.79	65	5.3E+09	**MB**
			97.3	323.31	165	3.8E+10	
9	DERPT_G7100	29.4	87.2	323.31	440	2.9E+11	**SM**
			85.3	323.31	162	2.2E+10	
10	DERPT_G8047	33.0	77.1	323.31	171	3.9E+09	**MB**
			81.7	323.31	923	5.4E+11	
11	DERPT_G8381	184.0	48.9	323.31	1822	4.7E+11	**MB**
13	DERPT_G8964	15.0	71.8	222.22	66	4E+09	**MB**
			90.8	323.31	387	2E+11	
14	DERPT_G9820	191.4	78.5	323.31	1388	1.6E+11	**MB**
			88.8	323.31	4729	2.2E+12	

MGA[A]	Sequence ID[B]	N: Mol. mass [kDa][C]	N: Sequence coverage [%][D]	N: Score[E]	N: MS/MS count[F]	N: LFQ Intensity[G]	Localisation[H]
15	DERPT_G3350	183.4	27.1	323.31	706	3E+11	**SM**
			26.3	323.31	398	1.1E+11	
16	DERPT_G5615	55.4	62.1	323.31	127	6.1E+09	**MB**
			92.1	323.31	718	1.5E+11	
18	DERPT_G784	63.1	48	323.31	267	4.7E+10	**SM**
			45.3	323.31	218	4.6E+10	
20	DERPT_G59	43.8	72.4	323.31	196	2.6E+10	**MB**
			85.1	323.31	727	3.7E+11	
21	DERPT_G7259	16.5	14.3	10.742	1	1.2E+07	**MB**
			70.7	138.58	117	2.3E+10	
22	DERPT_G5267	15.6	51	135.52	58	1.9E+10	**MB**
			65	135.32	62	2.1E+10	
23	DERPT_G11207	10.3	44.4	323.31	90	6.6E+10	**SM**
			34.4	32.386	26	3.6E+09	
24	DERPT_G5941	16.6	49.3	36.632	37	1.4E+09	**MB**
25	DERPT_G6990	30.0	76.4	323.31	157	2.4E+10	**MB**
			80	323.31	293	1E+11	
26	DERPT_G10519	133.7	10.8	138.1	134	3.1E+10	**MB**
27	DERPT_G8127	43.3	37.8	247.65	19	3.3E+09	**MB**
			37.8	189.99	54	2E+10	
28	DERPT_G6942	72.5	23	33.474	22	2E+09	**SM**
29	DERPT_G9923	40.8	19.7	172.32	66	2.2E+10	**MB**
			27.7	155.96	122	4.8E+10	
30	DERPT_G8780	20.8	86.1	323.31	153	3E+10	**MB**
			99.4	323.31	572	6E+11	

MGA[A]	Sequence ID[B]	N: Mol. mass [kDa][C]	N: Sequence coverage [%][D]	N: Score[E]	N: MS/MS count[F]	N: LFQ Intensity[G]	Localisation[H]
31	DERPT_G5053	16.8	77	251.82	57	2.1E+09	**MB**
			89.2	275.91	154	3.9E+10	
32	DERPT_G5605	166.9	33.4	323.31	237	2E+10	**MB**
			37.3	323.31	286	5.2E+10	
33	DERPT_G7263	51.4	38.3	87.829	21	7.6E+08	**MB**
Der f acidic ribosomal like allergen	DERPT_G5569	11.5	45.1	49.195	8	1.7E+08	**MB**
			45.1	71.984	30	3E+09	
Der f Actinin like allergen	DERPT_G9110	149.6	20	219.08	45	3.8E+08	**MB**
			62.1	323.31	841	1.9E+11	
Der f aldehyde dehydrogenase like allergen	DERPT_G5156	54.1	94.1	323.31	303	2.7E+10	**MB**
			95.9	323.31	840	4.3E+11	
Der f EF elongation factor like Allergen	DERPT_G5247	94.7	70.1	323.31	297	2.6E+10	**MB**
			78.2	323.31	765	2.4E+11	
Der f enolase like Allergen	DERPT_G12026	47.4	78.8	323.31	261	4E+10	**MB**
			84.3	323.31	597	2.4E+11	
Der f eukaryotic aspartyl protease like allergen	DERPT_G10356	42.9	45.6	172.3	38	1.2E+09	**MB**
			62.2	166.75	137	1.8E+10	
Der f profilin like allergen	DERPT_G5894	14.3	73.8	111.29	30	2.4E+09	**MB**
			86.2	99.86	67	1.2E+10	

Highlighted in in grey: Mite Body (MB) proteome Data. In white: Excretome proteome data (SM). MGA[A], Mite Group Allergen. Sequence ID[B], *D. pteronyssinus airmid* protein sequence ID. Description[B], annotations assigned by Blast2GO. N: Mol. mass [kDa][C], calculated Molecular mass of protein in kDa.. N: Sequence coverage [%][D], percentage of total protein sequence for which peptides were identified. N: Score[E], protein score, cumulative score of individual peptide mass spectra's. N: MS/MS count[F], sum of peptides selected for ms/ms analysis. LFQ Intensity[G], label free quantification intensity. (Software: Maxquant version 1.6.2.10, Perseus version 1.6.2.2). Localisation[H], qualitative assessment of allergen localisation to MB or SM based upon MS/MS count and LFQ intensity.

Supporting Information

S1 Table: Genome Assemblies Utilised for Phylogenetic Analysis.

S2 Table: LEA Homologs in *D. pteronyssinus airmid*

S3 Table: Top 10 Most Abundant Proteins Identified in *D. pteronyssinus airmid* Mite Body

S4 Table: Top 10 Most Abundant Proteins Identified in *D. pteronyssinus airmid* Spent Culture Medium.

Appendix 1. Electronic Files (Excel)

S1_Data: Allergen Orthologs Potentially cross-reactive orthologs of known mite group allergens (Der p1-11, 13-16, 18, 20-33 or non-chronalogical allergen sequences available in UniProt for *D. pteronyssinus* and *D. farinae* ,where *D. pteronyssinus* sequences were unavailable) identified by BLAST searches in 12 acari and 2 arachnid out-grouping. To be considered an allergen ortholog, proteins had to have reciprocal best hits with an allergen (E-value ≤1E-05), an alignment length of at least 80 amino acids and share at least 35% identity. We also identified allergen homologs by removing the criteria of being a reciprocal best hit.

S2_Data: Number of potentially cross-reactive orthologs of known MGA (Der p1-11, 13-16, 18, 20-33 or non-chronalogical allergen sequences available in UniProt for *D. pteronyssinus* and *D. farinae* ,where *D. pteronyssinus* sequences were unavailable) identified by BLAST searches in 12 acari and two arachnid outgrouping. To be considered an allergen ortholog, proteins had to have reciprocal best hits with an allergen (E-value ≤1E-05), an alignment length of at least 80 amino acids and share at least 35% identity.

S3_Data: *D. pteronyssinus* and *D. pteronyssinus airmid* specific genes. *D. pteronyssinus* specific genes (*n* = 1,850) are absent from other acari but present in all

one or more *D. pteronyssinus* genome assemblies (SI Table 1). *D. pteronyssinus airmid* specific genes (*n* = 1,475) were found only in *D. pteronyssinus airmid* genome assembly.

S4_Data. *D. pteronyssinus airmid* predicted proteome (*n* = 12,530 sequences) annotated by blast homology to sequences in NCBInr, SwissProt and InterPro databases utilising Blast2GO.

S5_Data: BLAST alignment results for *D. pteronyssinus airmid* predicted proteins (E-value ≤1E-03) to MGA sequences available in UniProt for *D. pteronyssinus* and *D. farinae* (where *D. pteronyssinus* sequences were unavailable).

S6_Data: List of all *D. pteronyssinus airmid* protein sequences with known immunomodulatory effects or predicted ,"allergenic molecules (*n* = 332)", based upon BLASThomology to known allergens (E-value ≤1E-05), with an alignment length of at least 80 amino acids and share at least 35% identity. This list is a consolidation of protein sequences designated as"Mite Group Allergens and Non-chronological allergens" (MGA, *n* = 37), "Mite group allergen homologs" (MGAH, *n* = 93), "Allergen Homologs" (AH, *n* = 326). Allergen Homolog BLASTp alignment data.

S7_Data: List of *D. pteronyssinus airmid* proteins (*n* = 3,931) identified by meta-proteomic analysis of 88Gb of spectra (99 LC-MS/MS files) representing the proteome available for analysis utilising extraction and analysis techniques employed during this study (Representative Proteome).

S8_Data: Proteomic results of all *D. pteronyssinus airmid* proteins (*n* = 150) identified from house dust protein extracts by LC-MS/MS.

S9_Data: **A)** Proteomic results of all *D. pteronyssinus airmid* proteins (*n* = 1,849) identified from the mite body through combined proteomic strategy of whole extract protein analysis and gel filtration protein extract fractionation, followed by LC-MS/MS. **B)** List of the most specific GO terms (*n* = 167) differentially represented

between the representative proteome and mite body proteome (<0.05 FDR) illustrating sample enrichment for proteins associated with mite body processes. **C)** List of enzyme names (*n* = 2) differentially represented between the representative proteome and mite body proteome (<0.05 FDR) illustrating sample enrichment for enzymes associated with mite body processes. **D)** List of InterPro IDs (*n* = 57) differentially represented between the representative proteome and mite body proteome (<0.05 FDR) illustrating sample enrichment of proteins with functional domains associated with mite body processes.

S10_Data: **A)** Proteomic results of all *D. pteronyssinus airmid* proteins (*n* = 873) identified from the spent culture media (Excretome) through combined proteomic strategy of whole extract protein analysis and gel filtration protein extract fractionation, followed by LC-MS/MS. **B)** List of the most specific GO terms *(n* = 320) differentially represented between the representative proteome and Excretome (<0.05 FDR) illustrating selective extraction of proteins associated with the Excretome. **C**) List of enzyme names (*n* = 35) differentially represented between the representative proteome and Excretome (<0.05 FDR) illustrating selective extraction of certain groups of enzymes associated with the Excretome. **D)** List of InterPro IDs (*n* = 69) differentially represented between the representative proteome and Excretome (<0.05 FDR) illustrating selective extraction of proteins with functional domains associated with the excretome.

S11_Data: List of all *D. pteronyssinus airmid* proteins identified in this study by proteomics (*n* = 4, 002).

S1 Table. Genome data utilised in our assessment of genomic completeness and phylogenetic analysis

Species	**Genome Reference**	**Suborder**
Centruroides sculpturatus		Arachnid outgroup
Parasteatoda tepidariorum		Arachnid outgroup
Dermatophagoides farina	[2]	Acariformes
Euroglyphus maynei	[3]	Acariformes
Sarcoptes scabiei	[4]	Acariformes
Tetranychus urticae	[5]	Acariformes
Psoroptes ovis	[6]	Acariformes
Ixodes scapularis	[7]	Parasitiforme
Galendromus occidentalis	[8]	Parasitiforme
Rhipicephalus microplus	[9]	Parasitiforme
Tropilaelaps mercedesae	[10]	Parasitiforme
Varroa destructor	[11]	Parasitiforme
Varroa jacobsoni		Parasitiforme

S2 Table. LEA Homologs in *D. pteronyssinus airmid* assigned based upon single best highest scoring Blast hit with known LEA protein.

Query Sequence ID[A]	Sequence ID[B]	Description[C]	LEA Class (Best hit)[D]	% Identities	E-value
AMR19109	DERPT_G248	UDP-glucose 4-epimerase	99	22.508	8.95E-07
JAC82215	DERPT_G290	Phospholipase ABHD3	99	29.551	7.23E-54
NP_001256174	DERPT_G404	Laminin subunit gamma-1	6	21.64	8.74E-09
AMR19109	DERPT_G997	Natural resistance-associated Macrophage 1 Short	99	25.904	9.20E-09
JAC82215	DERPT_G3026	Phospholipase ABHD3	99	28.198	1.43E-49
AMR19109	DERPT_G3148	GDP-mannose 4,6-dehydratase	99	24.671	4.28E-13
XP_006488730	DERPT_G4173	F-box LRR-repeat 20	6	25.824	3.02E-07
NP_851129	DERPT_G4823	LEA homology	6	29.688	1.41E-08
KCW52581	DERPT_G4831	Allergen Homolog (Hev b9, Sac Enolase, Alt a 11) Enolase	6	55.708	8.0E-152
AMR19109	DERPT_G6082	UDP-glucose 4-epimerase	99	26.036	3.51E-14
BAE92616	DERPT_G6261	Flagellar attachment zone 1	6	22.145	6.06E-09
AMR19109	DERPT_G7527	3 beta-hydroxysteroid dehydrogenase Delta 5--	99	26.829	6.14E-08
LEA76_BRANA	DERPT_G8349	Zonadhesin Flags: Precursor	6	21.053	2.03E-06
JAC82215	DERPT_G8791	Phospholipase ABHD3	99	28.909	1.27E-43
AMR19109	DERPT_G10968	UDP-glucuronic acid decarboxylase 1	99	24.924	4.37E-19
AMR19109	DERPT_G11799	Fatty acyl- reductase 1	99	25.595	6.23E-06
KCW52581	DERPT_G12026	Der f enolase like Allergen; Allergen homolog (Bla g Enolase, Sal s 2, Bos D Enolase), Proposed Pan-Allergen: Enolase	6	56.308	3.6E-173
BAE92616	DERPT_G12094	Liver stage antigen	6	23.642	8.14E-08

Query Sequence ID[A], from LEAP database (http://forge.info.univ-angers.fr/~gh/Leadb/index.php). Sequence ID[B], *D. pteronyssinus airmid* protein sequence ID. Description[C], annotations assigned by Blast2GO. LEA Class (Best hit)[D] , single best Blast alignment.

S3 Table. Top 10 Highest Abundant Protein Identified in *D. pteronyssinus airmid* Mite Body

Sequence IDs[A]	Description[B]	N: WE[C]	N: Fractions[D]	N: Peptides[E]	N: Unique peptides[F]	Sequence coverage [%]	Mol. Mass [kDa]	N: MS/MS count[G]	LFQ Intensity[H]
DERPT_G9820	Der p14 Allergen; Proposed Sten-Allergen, Allergen Homolog (Sar s14, Eur m14, Der f14)	4	20	261	261	88.8	191.4	4729	2.2E+12
DERPT_G8046	Myosin-7	4	20	297	297	76.2	249.0	3554	1.1E+12
DERPT_G11558	ATP synthase subunit	4	20	47	47	79.6	56.5	1101	7.6E+11
DERPT_G8792	Der p 2 Allergen; Proposed Mon-Allergen, Allergen Homolog (Der f2, Eur m2, Der s2)	4	20	20	20	76.7	15.9	419	6.7E+11
DERPT_G8780	Der f30 like allergen: Ferritin heavy chain	4	20	27	27	99.4	20.8	572	6.0E+11
DERPT_G8047	Der p10 Allergen; Proposed Pan-Allergen, Allergen Homolog (Der p10, Aca s10, Pso o10, Sar s10, Blo t10)	4	20	51	51	81.7	33.0	923	5.4E+11
DERPT_G749	Allergen Homolog (Sal s3, Thua3): Fructose-bisphosphate aldolase	4	20	42	42	85.9	39.4	667	4.9E+11
DERPT_G8381	Der p11 Allergen; Proposed Sten-Allergen, Allergen Homolog (Ani s2, Blo t11, Der f11, Der p11)	4	20	173	173	48.9	184.0	1822	4.7E+11
DERPT_G5156	Der f aldehyde dehydrogenase like allergen; Allergen homolog (Tyr p 35, Cla h3)	4	20	69	68	95.9	54.1	840	4.3E+11
DERPT_G3392	Allergen homolog (Tri a34, Asp vGAPDH): Glyceraldehyde-3-phosphate dehydrogenase	4	20	53	52	98.5	35.7	674	3.9E+11

Sequence ID[A], *D. pteronyssinus airmid* protein sequence ID. Description[B], annotations assigned by Blast2GO. N: WE[C], number of whole protein extract replicates in which specified protein was identified. N: Fractions[D], number of gel filtration fractions in which specified protein was identified. N: WE[D], number of whole protein extract replicates in which specified protein was identified. N: Peptides[E], number of peptides identified by LC-MS/MS for specified protein. N: Unique peptides[F], number of unique (not present in any other protein sequence in the predicted proteome) peptides identified by LC-MS/MS for specified protein. N: MS/MS count[G], sum of peptides selected for ms/ms analysis. LFQ Intensity[H], label free quantification intensity. (Software: Maxquant version 1.6.2.10, Perseus version 1.6.2.2).

S4 Table. Top 10 Most Abundant Proteins Identified in Spent Culture Medium of *D. pteronyssinus airmid*

Sequence IDs[A]	Description[B]	N: WE[C]	N: Fractions[D]	N: Peptides[E]	N: Unique peptides[F]	Sequence coverage [%]	Mol. Mass [kDa]	N: MS/MS count[G]	LFQ Intensity[H]
DERPT_G8792	Der p 2 Allergen; Proposed Mon-Allergen, Allergen Homolog (Der f2, Eur m2, Der s2)	4	21	17	17	78.8	15.9	486	5.2E+11
DERPT_G1283	Der p 1 Allergen, Proposed Sten-Allergen, Allergen Homolog (Der f1, Der m1, Pso o1)	4	21	35	35	41.3	96.0	1511	4.9E+11
DERPT_G8859	Der p3 Allergen, Proposed Sten-Allergen, Allergen Homolog (Der p3, Der f3, Eur m3, Blo t3, Tyr p3)	4	21	30	30	67.4	28.1	510	4.4E+11
DERPT_G9697	Sucrase- intestinal, alpha-1,4-glucosidase activity	4	21	106	102	61.7	202.1	1288	3.4E+11
DERPT_G3350	Der p15 Allergen	4	15	52	52	27.1	183.4	706	3.0E+11
DERPT_G7100	Der p9 Allergen	4	21	24	24	87.2	29.4	440	2.9E+11
DERPT_G9187	Der f6 like allergen; Allergen Homolog (Blo t6)	4	21	29	29	14.4	228.1	468	2.6E+11
DERPT_G10989	kDa bacteriolytic enzyme	4	21	10	10	37	30.9	302	1.6E+11
DERPT_G9149	Lysosomal alpha-mannosidase	4	11	89	89	74.5	118.3	875	1.6E+11
DERPT_G9820	Der p14 Allergen; Proposed Sten-Allergen, Allergen Homolog (Sar s14, Eur m14, Der f14)	4	20	183	183	78.5	191.4	1388	1.6E+11

Sequence ID[A], *D. pteronyssinus airmid* protein sequence ID. Description[B], annotations assigned by Blast2GO. N: WE[C], number of whole protein extract replicates in which specified protein was identified. N: Fractions[D], number of gel filtration fractions in which specified protein was identified N: Peptides[E], number of peptides identified by LC-MS/MS for specified protein. N: Unique peptides[F], number of unique (not present in any other protein sequence in the predicted proteome) peptides identified by LC-MS/MS for specified protein. N: MS/MS count[G], sum of peptides selected for ms/ms analysis. LFQ Intensity[H], label free quantification intensity. (Software: Maxquant version 1.6.2.10, Perseus version 1.6.2.2)

4.0 Characterisation of three Novel β-1,3 glucanases from the Medically Important House Dust Mite *Dermatophagoides pteronyssinus* (airmid).

This manuscript has been submitted for peer review.

Rose Waldron,[a,b] Jamie McGowan,[a,c] Natasha Gordon,[b] E. Bruce Mitchell,[b] David A. Fitzpatrick[a,c] and Sean Doyle [a]*.

[a]Department of Biology, Maynooth University, Maynooth, Co. Kildare, Ireland.
[b]Airmid Healthgroup Ltd., Trinity Enterprise Campus, Dublin, Ireland.
[c]Human Health Research Institute, Maynooth University, Maynooth, Co. Kildare, Ireland.

* Corresponding author:

Professor Sean Doyle, Department of Biology, Maynooth University, Maynooth, Co. Kildare, Ireland.

Email: sean.doyle@mu.ie; Tel : +353-1-7083858

ABSTRACT

The European house dust mite, *Dermatophagoides pteronyssinus* is a major source of airborne allergens worldwide and is found in half of European homes. Interactions between microbes and house dust mites (HDM) are considered important factors that allow them to persist in the home. Laboratory studies indicate the European HDM, *D. pteronyssinus* is a mycophagous mite, capable of utilising a variety of fungi for nutrients, however specific mycolytic digestive enzymes are unknown. Our previous work identified a number of putative glycosyl hydrolases present in the predicted proteome of *D. pteronyssinus* airmid and validated the expression of 42 of these. Of note, three GH16 proteins with predicted β-1,3 glucanase activity were found to be consistently present in the mite body and excretome. Here, we performed an extensive bioinformatic, proteomic and biochemical study to characterize three-novel β-1,3 glucanases from this medically important house dust mite. The genes encoding novel β-1,3 glucanases designated *Glu1*, *Glu2* and *Glu3* were identified in *D. pteronyssinus* airmid, each exhibited more than 59% amino acid identity to one another. These enzymes are encoded by *Glu* genes present in a tri-gene cluster and protein homologs are found in other acari. The patchy phyletic distribution of Glu proteins means their evolutionary history remains elusive, however horizontal gene transfer cannot be completely excluded. Recombinant Glu1 and Glu2 exhibit hydrolytic activity toward laminarin, pachyman and barley glucan. Excreted β-1,3 glucanase activity was increased in response to *D. pteronyssinus* airmid feeding on baker's yeast. Active β-1,3 glucanases are expressed and excreted in the faeces of *D. pteronyssinus* airmid indicating they are digestive enzymes capable of breaking down β-1,3 glucans of fungi present in house dust.

KEYWORDS

β-1,3 glucanase, enzyme, House Dust Mite, Yeast, *D. pteronyssinus*, quantitative proteomics, phylogenetics.

HIGHLIGHTS

- β-1,3 glucanases found in HDMs and other Acari species, specifically those within the Acariformes but not Parasitiformes orders.
- Mycophagous HDM, *D. pteronyssinus*, increases expression and excretion of β-1,3 glucanases in response to *S. cerevisiae* diet.
- β-1,3 glucanases excreted by *D. pteronyssinus* are found in house dust.

ABBREVIATIONS

HDM	House Dust Mite
HGT	Horizontal Gene Transfer
MES	4-Morpholineethanesulfonic acid
MM	Minimal Media
HDMMM	HDM Maximal Media
PBS	Phosphate Buffered Saline
SM	Spent Culture Medium
MB	Mite Body

INTRODUCTION

House dust mites (HDMs) appear to have evolved from an avian parasite ancestor, which in turn may have evolved from a mycophagous free-living ancestor (Klimov and O'Connor, 2013; O'Connor, 1979). Close relatives of the most common house dust-dwelling acariforme mites; *Dermatophagoides pteronyssinus*, *Dermatophagoides farinae* and *Euroglyphus maynei*, exhibit facultative mycophagy (O'Connor, 1979). The stored product mite, *Tyrophagous puterscentiae* is frequently found to contaminate laboratory fungal cultures, feeding on hyphae and spores of dermatophytes, molds and yeasts (Duek et al., 2001). Recent studies have demonstrated *D. pteronyssinus* is better suited to nutritionally exploit fungi than *D. farinae*, and is capable of using yeasts and the filamentous fungi, *Aspergillus* and *Penicillium spp.* as food sources (Molva et al., 2019).

In order to feed on fungi, *D. pteronyssinus* must possess the enzymatic arsenal to degrade the microbe's cell wall and digest the cell contents (Erban and Hubert, 2012). A typical fungal cell wall is a multilayer structure composed of a chitin linked to a β-1,3- and β-1,6-glucan layer, followed by an outermost layer of mannoproteins. β-glucans form 50-60% of the fungal cell wall with β-1,3 contributing 65-90% of the β-glucan content (Fesel and Zuccaro, 2016). Enzymes from the Glyco Hydrolase family 16 (GH16) exhibit hydrolytic activity against a variety of polysaccharides, including β-1,3 glucans commonly found in fungal cell walls (Alvarez et al., 2015). Enzymes capable of hydrolysing β-1,3 glucans exhibit endo- β-1,3 glucanase (EC 3.2.1.39), endo- β-1,3-1,4 glucanase (EC 3.2.1.6) or exo- β-1,3 glucanase activity (EC 3.2.1.58) (Song et al., 2010).

In previous work, we identified six putative GH16 proteins with predicted β-1,3 glucanase activity in *D. pteronyssinus* airmid, four were validated as expressed (Waldron et al., 2019). Moreover, previous studies have demonstrated β -glucosidase activity to be present in protein extracts from *D. pteronyssinus* (Martinez et al., 1999). This enzyme activity may be utilised by *D. pteronyssinus* for immune defence and/or digestive activities (Erban and Hubert, 2008; Pauchet et al., 2009).

GH16 proteins represent interesting targets for further study as active β-1,3 glucanases have been reported only in a handful of arthropods including Lepidoptera, Collembola and Diptera (Bragatto et al., 2010; Moraes et al., 2014). Moreover, they have been reported to be lost from Chelicerates (scorpions, mites, spiders and ticks) (Jiggins and Palmer, 2015). The unexpected discovery of an endo- β-1,3 glucanase in the Antarctic springtail, *Cryptopygus antarcticus* was explained by horizontal gene transfer (HGT) (Song et al., 2010). Thus, the presence of β-1,3 glucanases in the genome of *D. pteronyssinus* may also point to acquisition by HGT, as HGT has been seen previously in this species (Tang et al., 2017).

In the present study, we conducted an extensive bioinformatic and biochemical investigation of three recently discovered GH16 proteins with predicted β-1,3 glucanase activity to (i) examine the evolutionary history of predicted β-1,3 glucanases present in the *D. pteronyssinus* (ii) conduct functional characterisation of predicted β-1,3 glucanases and (iii) examine if *D. pteronyssinus* utilise β-1,3 glucanases to digest the yeast, *Saccharomyces cerevisiae*.

MATERIALS AND METHODS

Bioinformatic Analysis.

The evolutionary history of the three *Glu* genes relative to closely related species was investigated by taking the three corresponding Glu protein sequences and using each as a query sequence in a BLASTp (Altschul et al., 1997) database search (e-value cutoff of $1e^{-10}$) against a local protein database containing eleven other Acari species. Six of these species belong to the Parasitiformes order while the remaining five belong to the Acariformes order (Table 1). Homologous sequences were retrieved and aligned using MUSCLE (Edgar, 2004) using the default settings. The resultant alignment was used to reconstruct a maximum likelihood tree using RAxML (Stamatakis, 2014) utilising the LG+G+I+F model as selected by ProtTest (Darriba et al., 2011), branch supports were determined using 100 bootstrap replicates.

A broader evolutionary analysis was also undertaken by performing a BLASTp database search with an e-value cutoff of $1e^{-10}$ of the three Glu proteins against a dataset representative of fully sequenced prokaryotic and eukaryotic species. This dataset was composed of over 8 million protein sequences from 1,698 genomes sampled from all three domains of life which had been used in previous interdomain evolutionary analysis (McCarthy and Fitzpatrick, 2016). Homologous sequences were retrieved and aligned using MUSCLE using the default settings. The resultant alignment was used to reconstruct a maximum likelihood tree using FastTree2 (Price et al., 2010) utilising the LG model and local supports values were also determined.

Protein Extraction.

Culture and protein extraction from *D. pteronyssinus* airmid mite body (MB) and spent culture medium (SM) were conducted as described in Waldron et al. (2019). HDM were cultured on a diet of dried porcine liver (MM) or a mixture of porcine liver and baker's yeast (HDMMM).

Purification of β-1,3 glucanase from *D. pteronyssinus* airmid.

Gel filtration chromatography was carried out using an ÄKTA Purifier coupled with a Superdex 200 10/300 GL gel filtration column (GE Healthcare, Germany), equilibrated in Phosphate Buffered Saline (PBS). MB protein extracts were filtered (0.22 µm), injected (500 µl) and separated (flow rate 0.4 ml/min) with absorbance monitored at 215, 254 and 280 nm. Fractions were collected (between ~ 8 and 26 ml) and assayed for β-1,3-glucanase activity using the AZCL-Pachyman assay (described below). Fractions with detectable activity (16 – 22 ml post-injection) were pooled, buffer exchanged (50 mM 4-Morpholineethanesulfonic acid (MES), pH 5.5) and concentrated using centrifugal filters (Amicon; 3 kDa MWCO). The resulting protein solutions were further fractionated by cation exchange chromatography, and the β-1,3 glucanases bound the column (HiTrap SP Xl) and were eluted with sodium chloride gradient (60 - 90 mM NaCl in 50 mM MES pH 5.5).

Expression Analysis by RT-PCR.

RNA was extracted from *D. pteronyssinus* airmid using methods for nucleic acid extraction in Waldron et al. (2017) using the Purelink RNA mini kit (Invitrogen) and cDNA synthesised (qScript; Quantabio). Expression of *Glu1*, *Glu2* and *Glu3* genes were assessed by RT-PCR of cDNA with reference to control gene *Der p1*, using primers listed in Table 2.

Recombinant Protein Expression.

D. pteronyssinus airmid *Glu1*, *Glu2* and *Glu3* genes were commercially synthesised (IDT; USA) to allow for signal peptide removal, incorporation of terminal *Bgl*II and *Xho*I sites and codon optimisation for expression in *Escherichia coli*. Gene fragments were cloned into the pEX-N-GST Precision Shuttle vector (OriGene), recombinantly expressed in and purified from *E. coli* according to methods described by Dolan et al. (2014). The identity of purified recombinant proteins was confirmed by performing in-gel proteolytic digestion followed by LC-MS analysis, as described by Shevchenko et al. (2006).

Proteolytic Digestion and Nano-flow Liquid Chromatography Electro-Spray Ionization Tandem Mass Spectrometry (LC-MS/MS) Analysis of Protein Extracts.

Protein extracts from *D. pteronyssinus* airmid were prepared and analysed by LC-MS/MS as described by Waldron *et al*. (2019). Protein identification and label free quantitative (LFQ) analysis was conducted using MaxQuant (Version 1.6.1.0; http://maxquant.org/), statistical analysis of MaxQuant output data was performed by Perseus (Version 1.6.2.2) as described in O'Keeffe et al. (2014).

House Dust Protein Extract Analysis.

House dust protein extracts used in this study were provided by airmid healthgroup ltd (Dublin, Ireland). House dust protein extracts (100 µl) were brought to a final concentration of 1 M urea, then digested using methods described in Waldron et al.

(2019). Statistical analysis of the correlation between β-1,3 glucanase activity (mU/ml), Der p 1 (µg/g) and LFQ intensity of Glu1, Glu2, Glu3 was performed using Graphpad (PRISM) statistical software, Spearman's correlation with 95% confidence interval.

Laminarin Assay.

β-1,3 glucanase activity was measured using laminarin substrate (β-1,3-1,6 glucan; L9634; Sigma). Protein extracts (50 µl; 0.2 mg/ml), recombinant proteins (50 µl; 0.125 mg/ml), β-1,3 glucanase enzyme standard (67138; Sigma; 50 µl; 5 mU) and house dust protein extracts (50 µl) were individually incubated with laminarin substrate as follows: (200 µl; 0.25% (w/v) laminarin; 50 mM sodium acetate pH 6.0; 2.5 h; 37 ^{o}C). Assays were terminated by heating samples (90 ^{o}C; 5 min).

Barley Glucan Assay.

β-1,3-1,4 glucanase activity was measured using low viscosity barley glucan (Megazyme, Ireland) using methods described for the laminarin assay.

Glucose Measurement Assay.

Glucose measurement was performed according to Bethke and Busse (2008) with minor modifications. Sample/standard were diluted (1:10 to 1:80; 50 mM sodium phosphate pH 7.4), then added (50 µl) to each microplate well. Enzyme-Ampliflu® Red mixture (50 µl) was added and microplate was incubated in the dark (RT; 30 min). Absorbance (560 nm) was measured (BioTek Instruments, Inc., USA). Enzyme mix sufficient for one plate was prepared by combining 50 µl Ampliflu® Red reagent stock solution (Sigma 90101; 10 mM in DMSO), 100 µl horseradish peroxidase stock solution (Sigma P8375; 10 U/ml in 50 mM sodium phosphate pH 7.4), and 100 µl glucose oxidase stock solution (Sigma G7141; 100 U/ml in 50 mM sodium phosphate pH 7.4) with 4.75 ml of 50 mM sodium phosphate pH 7.4. All stock solutions were frozen as aliquots at −20°C prior to use. A glucose standard curve (0 - 1.8 µg). enabled the glucose content of samples to be calculated.

Glucanase Quantification.

β-1,3 glucanase activity was quantified by performing laminarin assay followed by glucose measurement assay. Endogenous glucose in *D. pteronyssinus* airmid protein extracts was quantified using glucose measurement assay and subtracted from total glucose measured following laminarin assay. Relative specific β-1,3 glucanase activity was calculated with reference to β-1,3 glucanase enzyme standard. One unit of β-1,3 glucanase was defined as the amount of enzyme that liberated the equal amount of glucose to that of the β-1,3 glucanase standard under the conditions described above.

AZCL-Pachyman Assay.

Endo- β-1,3 glucanase activity was measured using AZCL-Pachyman (Megazyme, Ireland). Briefly, recombinant proteins (200 µl; 0.125 mg/ml) were incubated with the substrate (800 µl; 0.25% (w/v) AZCL-Pachyman; 50 mM sodium acetate pH 6; 2.5 h; 37 ^{o}C), centrifuged (10,000 *g*; 5 min) and supernatant recovered. Samples were tested in triplicate; the mean absorbance was recorded (600 nm).

AZCL-HE-Cellulose Assay.

Cellulase activity (β-1,4 glucanase) was measured using AZCL-HE-Cellulose (Megazyme, Ireland) using methods described for AZCL-Pachyman assay.

Biochemical Characterisation of Recombinant Glucanases.

Optimal pH and temperature of recombinant proteins was assessed by performing laminarin assays across pH range (pH 4, 5, 6, 7, 8, 9) and temperatures (4^{o}C, 25^{o}C, 30^{o}C, 37^{o}C, 50^{o}C). Enzyme mixtures were neutralised (pH 7 - 8) and glucose concentration measured.

RESULTS

Bioinformatic Analysis of *D. pteronyssinus* airmid Glucanases.

Three GH16 genes with putative β-1,3 glucanase or β-1,3-1,4 glucanase activity were located in the *D. pteronyssinus* airmid genome assembly and designated *Glu1* (DERPT_G3105), *Glu2* (DERPT_ G3104) and *Glu3* (DERPT_ G3106) in order of their discovery. The genes had open reading frame lengths of 816-828 bp, were located within the same contig forming a tri-gene cluster, designated the *Glu* cluster, with intergenic distances of 241 bp between *Glu2* and *Glu1*, and 282 bp between *Glu1* and *Glu3* (Figure 1A). The genes translated to Glu proteins that exhibited a high degree of amino acid (a.a) sequence similarity to one another ranging between 59% and 69% identity

The Glu proteins of 272 - 275 a.a in length, each contained a signal sequence of 18-19 a.a. The Glu proteins are of a similar size (31.7-32.1 kDa) and contain a glycoside hydrolase family 16 (GH16) catalytic domain with catalytic residues (Figure 1B). Proteomic analysis of *D. pteronyssinus* airmid excreted proteins present in SM resulted in the identification of all three β-1,3 glucanases with sequence coverage of 25.8 – 77.9% (Figure 1C).

Blastp search of other Acari genomes (Table 1) identified three additional Glu homologs in *D. pteronyssinus* airmid while five in total were found in *D. farinae* and four in *E. maynei*. Homologs were also located in *Sarcoptes scabiei* (two) and *Psoroptes ovis* (two) but absent from the Parasitiforme order. A maximum likelihood phylogeny was constructed which revealed that Glu1, Glu2 and Glu3 are grouped in a single clade with strong (86%) bootstrap support (BP). Glu3 is grouped with orthologs from *D. farinae* and *E. maynei* (Figure 1F, 99% BP). *D. pteronyssinus* Glu1 and Glu2 are grouped in a single clade with 87% BP (Figure 2). Based on our phylogeny, *D. farinae* and *E. maynei* have an ortholog of Glu1 but not of Glu2 (Figure 2).

Blastp searched against a local proteome database located 299 homologs, primarily located in bacterial species but also select fungal species, two animals (*Ciona intestinalis*

and *Daphnia pulex*), oomycetes such as *Saprolegnia* and *Aphanomyces* species as well as a number of diatoms including *Ectocarpus* and *Aurococcus* species. The resultant maximum likelihood phylogeny shows the majority of eukaryote homologs are located in a single highly supported clade (S1 Figure 78% BP). Three of the Acariformes homologs (*D. farinae*, *E. maynei* and *D. pteronyssinus* airmid) share sister group relationships with bacterial homologs.

Enzyme Purification and Identification.

Identification of β-1,3 glucanase active native protein (Glu1) was facilitated by a two-step purification workflow (Figure 3A-D), the β-1,3 glucanase active fraction was visualized by SDS-PAGE (Figure 3E) and bands excised for identification by LC-MS/MS (Figure 3F). High confidence identification of putative β-1,3 (4) glucanase (DERPT_G3105) of approximately 25 kDa in size was achieved with 70.6% coverage and identification of 19 peptides. Purification resulted in the recovery of 20 µg of glucanase enriched extract with specific activity 670 mU/mg from 1.4 mg of starting material with 106 mU/mg specific activity, summarised in Table 3.

Cloning and Expression.

Total RNA was extracted from *D. pteronyssinus* airmid and converted to cDNA. RT-PCR analysis was performed using intron containing control gene, *Der p1*. Absence of genomic DNA was confirmed by the smaller amplicon size of the *Der p1* cDNA amplicon relative to the corresponding genomic amplicon (Figure 4A). *Glu1*, *Glu2* and *Glu3* amplicons from cDNA and genomic DNA were of identical size, confirming absence of introns. Amplicons derived from cDNA confirmed expression of all three genes in the *Glu* gene cluster (Figure 4A). Commercially synthesised *Glu1*, *Glu2* and *Glu3* in expression vector pEX-N-GST (Figure 4B) were expressed in *E. coli* with an N-terminal GST tag. Recombinant proteins were purified by affinity chromatography as determined by SDS-PAGE and western blot analysis (Figure 4C & D). SDS-PAGE gel bands corresponding to recombinant proteins were excised (Figure 4C), proteins were

digested, and peptides were identified by LC-MS/MS (Figure 4E) confirming expression and successful purification of the three recombinant proteins.

Substrate Specificity and Relative Specific activity of Glu1 and Glu2.

Four substrates were used to test the ability of GST-rGlu1, GST-rGlu2 and GST-rGlu3 to hydrolyse β-1,3-1,6 glucans, β-1,3 glucans, β-1,3-1,4 glucans and β-1,4 glucans. Activity for GST-rGlu1 and GST-rGlu2 was highest on laminarin (β-1,3-1,6 glucan) resulting in 37% and 8% of substrate hydrolysis, respectively, compared to only 7% and 2% on barley glucan (β-1,3-1,4 glucan) (Figure 5B). GST-rGlu2 exhibited approximately 20% of the hydrolytic activity of GST-rGlu1 on laminarin and barley glucan, and 33% on AZCL-Pachyman (β-1,3 glucans) (Figure 5C). GST-rGlu1 showed highest relative specific activity on laminarin substrate with 743.5 mU/mg and GST-rGlu2 with 158.3 mU/mg (Figure 5D). GST-rGlu1 and GST-rGlu2 were unable to hydrolyse AZCL-HE-cellulose (β-1,4 glucan) and GST-rGlu3 was found to be inactive on all substrates (data not shown).

Effects of Temperature and pH on Enzyme Activity.

An evaluation of temperature on GST-rGlu1 and GST-rGlu2 hydrolytic activity against laminarin showed an increase in activity as temperatures increased peaking at 37 ^{o}C for GST-rGlu1 and 50 ^{o}C for GST-rGlu2 (Figure 5E). The effect of pH on hydrolysis of laminarin was also determined, both GST-rGlu1 and GST-rGlu2 showed activity in the pH range 4.0 – 8.0. Maximum hydrolytic activity was observed at pH 5.0 for GST-rGlu1 and pH 4.0 for GST-rGlu2 (Figure 5F).

Internal and Excreted HDM Glucanase Activity and Abundance Increased on Yeast-based Diet.

Protein extracts from MB (n = 5) and SM (n = 5) of *D. pteronyssinus* airmid cultured on growth medium, with (HMDMM) and without yeast (MM), were assayed for β-1,3 glucanase activity. Extracts were incubated with laminarin substrate, released glucose was quantified by subtracting endogenous glucose present in protein extracts (approx. 6-

28% of total glucose). Specific β-1,3 glucanase activity in MB of *D. pteronyssinus* airmid fed yeast was higher than in HDMs cultured without yeast ($P = 0.0088$), this difference was also consistent in the SM ($P < 0.0001$). There was no significant difference in β-1,3 glucanase activity found between MB and SM of HDMs cultured without yeast ($P = 0.167$) however, a significant difference was seen when HDMs were fed yeast ($P = 0.0083$) (Figure 5A). Comparative proteomic analysis of SM from *D. pteronyssinus* airmid grown with and without yeast revealed all three β-1,3 glucanases increased in abundance (20- 39 %) in the SM of HDMs fed yeast (Table 4, S1 Data).

Glucanase Activity in House Dust Correlates with Glu1/2 Presence.

Protein extracts from HD positive for Der p 1 (0.2 – 16.94 µg Der p 1 per gram house dust) contained β-1,3 glucanase activity (Range; 2.5 - 370 mU/ml), activity levels correlated with Der p 1 concentration (Figure 6A). Subsequent proteomic analysis identified Glu1 or Glu2 to be present in many of the HD extracts ($n = 16$), nine extracts were found to contain both enzymes. LFQ intensity of both Glu1 and Glu2 correlated significantly (Spearman's correlation) with relative specific β-1,3 glucanase activity (Figure 6B & C).

DISCUSSION

HDM are trophic generalists feeding on pollen, bacteria, plant fibres, fungal mycelia and spores associated with their human proximity habitats (Colloff, 2009; van Bronswijk, 1973). β-1,3 glucan content of house dust ranges between 15 -79,000 µg per gram and may be an important part of the HDM diet in the home (Brooks et al., 2013; Maheswaran et al., 2014). β-glycosidase activity has been reported previously in enzymatic studies of *D. pteronyssinus* however no specific proteins were linked to the activity seen (Martinez et al., 1999). Studying the expression, localisation and biochemical activity of HDM enzymes can allow for insight into their physiological function (Erban and Hubert, 2012).

To date, biochemical characterisation of *D. pteronyssinus* proteins have primarily focused on allergenic proteins (Bordas-Le Floch et al., 2017). In the present study, we

conducted an extensive bioinformatic and biochemical investigation of three recently discovered GH16 proteins to gain insight into their role in *D. pteronyssinus* digestion.

Glu genes coding for these Glu proteins lacked introns and were found adjacent to one another, forming a tri-gene cluster designated the *Glu* cluster. A further three homologs for Glu proteins were located in the genome of *D. pteronyssinus* airmid outside of the *Glu* cluster. Molecular phylogeny of these Glu proteins confirms their close relatedness with the several Acariformes β-1,3 glucanases, with homologs found in other HDM species (*D. farinae* & *E. maynei*) and closely related parasitic mites (*P. ovis* & *S. scabei*). Glucanase genes were expanded in HDMs, which contained between four and six homologs, compared to the parasitic Acariformes which contained only two copies. Interestingly, none of the six species from the Parasitiforme order contained homologs, suggesting glucanase activity is a recent adaption in Acariformes. Interestingly, our phylogeny demonstrates that *D. pteronyssinus* has an additional Glu (Glu2) homolog, not seen in the other two HDMs (Figure 2). Based on its genomic location and high sequence identity, it appears to be a tandem duplication of Glu1. We are confident this is not an assembly/annotation error as this ortholog is present in a separate *D. pteronyssinus* assembly (Randall et al., 2018)). Moreover, the expression of all three *Glu* genes was confirmed by RT-PCR and their translation into protein confirmed by proteomic analysis, both in the course of this study and in previous work (Waldron et al., 2019). Recent studies have demonstrated *D. pteronyssinus* is better suited to nutritionally exploit fungi than *D. farinae*, and is capable of using yeasts and filamentous fungi as a food source (Molva et al., 2019). The presence of an additional Glu protein in *D. pteronyssinus* may allow it to exploit different nutrients than *D. farinae*, namely fungi, thus explaining why the presence of one species does not influence the numbers found of the other (Zock et al., 2006). Moreover, *D. farinae* exhibit a higher rate of population growth on bacteria-enriched diets than *D. pteronyssinus* (Erban and Hubert, 2008). The two mites may be utilising different components of house dust for food, with *D. farinae* better adapted to the exploitation of bacteria and *D. pteronyssinus* adapted for the exploitation of fungi.

To further investigate the evolutionary history of the three *Glu* genes, we searched their protein sequences against taxonomically diverse proteomes including prokaryotes, plants, fungi, heterokonts and animals (McCarthy and Fitzpatrick, 2016) and reconstructed a phylogenetic tree. The majority of eukaryote homologs including Glu1, 2 & 3 are located in a single highly-supported clade (S.Figure, 78% BP). Interestingly, three separate Glu Acariformes homologs (*D. farinae*, *E. maynei* and *D. pteronyssinus* airmid) share a sister group relationship with bacterial homologs (S.Figure). Some of the inconsistencies in the phylogenetic placement of eukaryote and prokaryote homologs may be due to HGT (Fitzpatrick, 2012) or other phylogenetic artefacts, such as long branch attraction (Felsenstein, 1978). To confidently infer potential HGT, a more comprehensive phylogenetic analysis is required. Such an analysis is beyond the scope of this work. However, we can confidently say that the distribution of Glu homologs amongst the eukaryotes in our proteome database is patchy at best. We are cognisant our inferences may be influenced by taxon sampling issue as our local proteome database consists of a large proportion of bacterial sequences (56%). To further investigate this we searched all three Glu proteins against the non-redundant database of GenBank (Sayers et al., 2019). As with our local proteome database search results the vast majority (~80%) of homologs located in GenBank are bacterial in origin. A number of additional eukaryote homologs from species not in our local database were located including homologs in the snowberry fruit fly (*Rhagoletis zephyria*), springtails (*Folsomia candida*, *Cryptopygus antarcticus* and *Orchesella cincta*) and, of special noteworthiness, the mycophagous amoeba *Planoprotostelium fungivorum*. Therefore, additional database search of the Glu proteins against GenBank confirms that the distribution of these proteins in eukaryotes is patchy.

Fractionation of *D. pteronyssinus* airmid MB protein lysates via a combination of gel filtration and cation exchange chromatography resulted in a semi-pure protein preparation which retained β-1,3 glucanase activity. SDS-PAGE of this fraction and subsequent high-sensitivity mass spectrometric analysis of a band corresponding to 25 kDa revealed the presence of a putative β-1,3-1,4 glucanase (DERPT_G3105; Glu1). Thus, confirming the source of β-1,3 glucanase activity seen in *D. pteronyssinus* airmid

protein extracts was from the mites themselves and not microbes present in the cultures. A typical purification with 1.4 mg of crude protein extract resulted in the recovery of 20 µg of β-1,3 glucanase enriched extract with 670 mU/mg specific activity.

The low yield of purified native β-1,3 glucanase necessitated the generation of recombinant proteins to allow for full biochemical characterisation. Recombinant proteins were purified by affinity chromatography, however solubility was poor with soluble enzyme accounting for < 0.5 % of the total protein of crude cell-free extract. Enzyme activity analysis of purified recombinant GST-rGlu1 and GST-rGlu2 confirmed that they are β-1-3 glucanases, with highest hydrolytic activity seen against laminarin substrate and moderate activity on AZCL-Pachyman and barley glucan. Relative specific β-1-3 glucanase activity was highest for GST-rGlu1 at 743.5 mU/mg, while GST-rGlu2 had relative specific activity of 158.3 mU/mg. None of the three recombinant proteins exhibited β-1-4 glucanase activity against AZCL-HE-Cellulose. This biochemical analysis has confirmed Glu1 and Glu2 to be laminarinases (EC 3.2.1.6) capable of hydrolysing β-1,3-glucans and mixed β-1,3-1,4/ β-1,3-1,6 glucans. Peak enzyme activity was observed for GST-rGlu1 at a pH 5 and temperature of 37 ^{o}C. GST-rGlu2 showed optimal activity at lower pH (pH 4) and higher temperature (50 ^{o}C).

Recombinantly expressed GST-rGlu3 lacked hydrolytic activity against all substrates tested. This maybe a result of *E. coli* as a poor choice of expression system resulting in the production of an inactive enzyme. The increased excretion of Glu3 in the SM of *D. pteronyssinus* airmid fed a diet containing yeast is a strong indicator that it is a digestive enzyme, expression of rGlu3 in an alternative expression system may yield an active protein in future studies.

In laboratory HDM cultures, fungi and bacteria are constituently present, their growth is kept to a minimum by limiting humidity and regular sub-culturing of HDMs (Colloff, 2009; Molva et al., 2018). Laminarinase activity was observed in the MB and SM of *D. pteronyssinus* airmid fed MM lacking yeast, indicating a basal level of expression and excretion in response to microbes naturally present in their environment (Molva et al., 2018). During basal expression, there was no significant difference in laminarinase

activity between the MB and SM. However, when fed heat-killed yeast, laminarinase activity was increased 2.6-fold internally in the MB and 5.9-fold externally in SM. Comparative proteomic analysis supported these assay results, identifying Glu1, Glu2 and Glu3 to be increased in abundance (20 – 39%) in the SM from HDMs fed on yeast (Table 4, S1 Data). Other studies have demonstrated that *D. pteronyssinus* is capable of fully digesting *S. cerevisiae* (Molva et al., 2018), and here we have demonstrated that they increase the production of three β-1-3 glucanase in response to feeding on this yeast. Moreover, a number of other enzymes were identified as uniquely present in the SM of yeast fed mites, including three alpha-mannosidase (DERPT_G10145, DERPT_G9692 & DERPT_G222, S2 Data). Mannoproteins form the outermost layer of the yeast cell wall and 40% of the carbohydrate content (Aguilar-Uscanga and Francois, 2003; Fesel and Zuccaro, 2016). Laminarinases are frequently used to remove mannoproteins from *S. cerevisiae* and disrupt the cell wall (Van Rinsum et al., 1991). It is possible that Glu1, Glu2 and Glu3 remove this mannoprotein layer which is then further hydrolysed by these mannosidases.

Protein extracts from HD contained notable laminarinase activity, correlating significantly ($P = 0.0001$) with *D. pteronyssinus* biomarker Der p 1. Moreover, proteomic analysis facilitated the identification of Glu1 and Glu2, and correlation between laminarinase activity in HD and LFQ intensity of Glu1 and Glu2 (Figure 6B & C). In house dust from floors and mattresses the fungal species frequently identified by high-throughput DNA sequencing, to be of high abundance are; *Aspergillus* spp, *Cladosporium* spp., *Penicillium spp.* and *Alternaria* spp. (Kaarakainen et al., 2009; Rintala et al., 2012; Sousa et al., 2014). *D. pteronyssinus* has been demonstrated to consume *Aspergillus* spp, and *Penicillium spp.* in the laboratory (Molva *et al.,* 2019). Therefore, it is highly probable they also consume these fungi in the home, and may utilise Glu enzymes to digest these fungi as they do for *S. cerevisiae*, thus explaining high presence of Glu1 and Glu2 in house dust. Our previous proteomic characterisation of HD and SM identified distinct parallels in the types of proteins identified in both of these conditions (Waldron et al., 2019). This data further demonstrates the utility of laboratory-based models for understanding the wildtype proteome of *D. pteronyssinus*.

Overall, our data infers that Glu1, Glu2 and Glu3 are digestive enzymes involved in HDM digestion of fungi. *D. pteronyssinus* contains an additional Glu protein, Glu2, which may confer a specialisation for mycophagy over other HDMs. Both Glu1 and Glu2 exhibit hydrolytic activity against β-1,3 glucans a major structural component of fungal cell walls and were increased in abundance in the SM of *D. pteronyssinus* airmid fed a diet containing *S. cerevisiae*. Basal expression and excretion of Glu1 and Glu2, in *D. pteronyssinus* airmid reared on a diet free of yeast, indicate they are produced in response to common microbes co-present in laboratory cultures. The identification of β-1,3 glucanase activity, and positive identification of Glu1 and Glu2 in HD provides compelling evidence that these are important enzymes utilised by *D. pteronyssinus* in both the laboratory and wildtype setting, for the digestion of fungi.

ACKNOWLEDGMENTS

RW is a recipient of an Irish Research Council Enterprise Partnership Scheme Scholar (EPSPG/2014/64). JM is funded by a postgraduate scholarship from the Irish Research Council, Government of Ireland (GOIPG/2016/1112). Mass spectrometry facilities were funded by a Science Foundation Ireland infrastructure award to SD (SFI 12/RI/2346(3)). We acknowledge the DJEI/DES/SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support. This work was co-funded by airmid healthgroup td, Dublin, Ireland.

REFERENCES

Aguilar-Uscanga, B., Francois, J.M., 2003. A study of the yeast cell wall composition and structure in response to growth conditions and mode of cultivation. Lett Appl Microbiol 37, 268-274.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389-3402.

Alvarez, T.M., Liberato, M.V., Cairo, J.P.L.F., Paixão, D.A.A., Campos, B.M., Ferreira, M.R., Almeida, R.F., Pereira, I.O., Bernardes, A., Ematsu, G.C.G., Chinaglia, M., Polikarpov, I., de Oliveira Neto, M., Squina, F.M., 2015. A Novel Member of GH16 Family Derived from Sugarcane Soil Metagenome. Applied Biochemistry and Biotechnology 177, 304-317.

Barrero, R.A., Guerrero, F.D., Black, M., McCooke, J., Chapman, B., Schilkey, F., Perez de Leon, A.A., Miller, R.J., Bruns, S., Dobry, J., Mikhaylenko, G., Stormo, K., Bell, C., Tao, Q., Bogden, R., Moolhuijzen, P.M., Hunter, A., Bellgard, M.I., 2017. Gene-enriched draft genome of the cattle tick Rhipicephalus microplus: assembly by the hybrid Pacific Biosciences/Illumina approach enabled analysis of the highly repetitive genome. Int J Parasitol 47, 569-583.

Bethke, P.C., Busse, J.C., 2008. Validation of a Simple, Colorimetric, Microplate Assay using Amplex Red for the Determination of Glucose and Sucrose in Potato Tubers and other Vegetables. American Journal of Potato Research 85, 414.

Bordas-Le Floch, V., Le Mignon, M., Bussieres, L., Jain, K., Martelet, A., Baron-Bodo, V., Nony, E., Mascarell, L., Moingeon, P., 2017. A combined transcriptome and proteome analysis extends the allergome of house dust mite Dermatophagoides species. PLoS One 12, e0185830.

Bragatto, I., Genta, F.A., Ribeiro, A.F., Terra, W.R., Ferreira, C., 2010. Characterization of a β-1,3-glucanase active in the alkaline midgut of Spodoptera frugiperda larvae and its relation to β-glucan-binding proteins. Insect Biochemistry and Molecular Biology 40, 861-872.

Brooks, C.R., Siebers, R., Crane, J., Noss, I., Wouters, I.M., Sander, I., Raulf-Heimsoth, M., Thorne, P.S., Metwali, N., Douwes, J., 2013. Measurement of beta-(1,3)-glucan in

household dust samples using Limulus amebocyte assay and enzyme immunoassays: an inter-laboratory comparison. Environ Sci Process Impacts 15, 405-411.

Burgess, S.T.G., Bartley, K., Marr, E.J., Wright, H.W., Weaver, R.J., Prickett, J.C., Hughes, M., Haldenby, S., Thi Le, P., Rombauts, S., Van Leeuwen, T., Van de Peer, Y., Nisbet, A.J., 2018. Draft Genome Assembly of the Sheep Scab Mite, Psoroptes ovis. Genome Announcements 6.

Chan, T.F., Ji, K.M., Yim, A.K.Y., Liu, X.Y., Zhou, J.W., Li, R.Q., Yang, K.Y., Li, J., Li, M., Law, P.T.W., Wu, Y.L., Cai, Z.L., Qin, H., Bao, Y., Leung, R.K.K., Ng, P.K.S., Zou, J., Zhong, X.J., Ran, P.X., Zhong, N.S., Liu, Z.G., Tsui, S.K.W., 2015. The draft genome, transcriptome, and microbiome of Dermatophagoides farinae reveal a broad spectrum of dust mite allergens. Journal of Allergy and Clinical Immunology 135, 539-548.

Colloff, M.J., 2009. Dust Mites. CSIRO Publishing, Springer, Van Godewijckstraat 303311 GX Dordrecht The Netherlands.

Cornman, S.R., Schatz, M.C., Johnston, S.J., Chen, Y.P., Pettis, J., Hunt, G., Bourgeois, L., Elsik, C., Anderson, D., Grozinger, C.M., Evans, J.D., 2010. Genomic survey of the ectoparasitic mite Varroa destructor, a major pest of the honey bee Apis mellifera. Bmc Genomics 11.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27, 1164-1165.

Dolan, Stephen K., Owens, Rebecca A., O'Keeffe, G., Hammel, S., Fitzpatrick, David A., Jones, Gary W., Doyle, S., 2014. Regulation of Nonribosomal Peptide Synthesis: bis-Thiomethylation Attenuates Gliotoxin Biosynthesis in Aspergillus fumigatus. Chemistry & Biology 21, 999-1012.

Dong, X., Armstrong, S.D., Xia, D., Makepeace, B.L., Darby, A.C., Kadowaki, T., 2017. Draft genome of the honey bee ectoparasitic mite, Tropilaelaps mercedesae, is shaped by the parasitic life history. Gigascience 6, 1-17.

Duek, L., Kaufman, G., Palevsky, E., Berdicevsky, I., 2001. Mites in fungal cultures. Mycoses 44, 390-394.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32, 1792-1797.

Erban, T., Hubert, J., 2008. Digestive function of lysozyme in synanthropic acaridid mites enables utilization of bacteria as a food source. Exp Appl Acarol 44, 199-212.

Erban, T., Hubert, J., 2012. Digestive Physiology of Synanthrophic Mites (Acari: Acaridida). SOAJ Entomological Studies, pp. 1-32.

Felsenstein, J., 1978. Cases in which Parsimony or Compatibility Methods Will be Positively Misleading. Systematic Zoology 27, 401-410.

Fesel, P.H., Zuccaro, A., 2016. β-glucan: Crucial component of the fungal cell wall and elusive MAMP in plants. Fungal Genetics and Biology 90, 53-60.

Fitzpatrick, D.A., 2012. Horizontal gene transfer in fungi. FEMS Microbiol Lett 329, 1-8.

Grbić, M., Van Leeuwen, T., Clark, R.M., Rombauts, S., Rouzé, P., Grbić, V., Osborne, E.J., Dermauw, W., Thi Ngoc, P.C., Ortego, F., Hernández-Crespo, P., Diaz, I., Martinez, M., Navajas, M., Sucena, É., Magalhães, S., Nagy, L., Pace, R.M., Djuranović, S., Smagghe, G., Iga, M., Christiaens, O., Veenstra, J.A., Ewer, J., Villalobos, R.M., Hutter, J.L., Hudson, S.D., Velez, M., Yi, S.V., Zeng, J., Pires-daSilva, A., Roch, F., Cazaux, M., Navarro, M., Zhurov, V., Acevedo, G., Bjelica, A., Fawcett, J.A., Bonnet, E., Martens, C., Baele, G., Wissler, L., Sanchez-Rodriguez, A., Tirry, L., Blais, C., Demeestere, K., Henz, S.R., Gregory, T.R., Mathieu, J., Verdon, L., Farinelli, L.,

Schmutz, J., Lindquist, E., Feyereisen, R., Van de Peer, Y., 2011. The genome of Tetranychus urticae reveals herbivorous pest adaptations. Nature 479, 487.

Gulia-Nuss, M., Nuss, A.B., Meyer, J.M., Sonenshine, D.E., Roe, R.M., Waterhouse, R.M., Sattelle, D.B., de la Fuente, J., Ribeiro, J.M., Megy, K., Thimmapuram, J., Miller, J.R., Walenz, B.P., Koren, S., Hostetler, J.B., Thiagarajan, M., Joardar, V.S., Hannick, L.I., Bidwell, S., Hammond, M.P., Young, S., Zeng, Q., Abrudan, J.L., Almeida, F.C., Ayllón, N., Bhide, K., Bissinger, B.W., Bonzon-Kulichenko, E., Buckingham, S.D., Caffrey, D.R., Caimano, M.J., Croset, V., Driscoll, T., Gilbert, D., Gillespie, J.J., Giraldo-Calderón, G.I., Grabowski, J.M., Jiang, D., Khalil, S.M.S., Kim, D., Kocan, K.M., Koči, J., Kuhn, R.J., Kurtti, T.J., Lees, K., Lang, E.G., Kennedy, R.C., Kwon, H., Perera, R., Qi, Y., Radolf, J.D., Sakamoto, J.M., Sánchez-Gracia, A., Severo, M.S., Silverman, N., Šimo, L., Tojo, M., Tornador, C., Van Zee, J.P., Vázquez, J., Vieira, F.G., Villar, M., Wespiser, A.R., Yang, Y., Zhu, J., Arensburger, P., Pietrantonio, P.V., Barker, S.C., Shao, R., Zdobnov, E.M., Hauser, F., Grimmelikhuijzen, C.J.P., Park, Y., Rozas, J., Benton, R., Pedra, J.H.F., Nelson, D.R., Unger, M.F., Tubio, J.M.C., Tu, Z., Robertson, H.M., Shumway, M., Sutton, G., Wortman, J.R., Lawson, D., Wikel, S.K., Nene, V.M., Fraser, C.M., Collins, F.H., Birren, B., Nelson, K.E., Caler, E., Hill, C.A., 2016. Genomic insights into the Ixodes scapularis tick vector of Lyme disease. Nature Communications 7, 10507.

Jiggins, F.M., Palmer, W.J., 2015. Comparative Genomics Reveals the Origins and Diversity of Arthropod Immune Systems. Molecular Biology and Evolution 32, 2111-2129.

Kaarakainen, P., Rintala, H., Vepsalainen, A., Hyvarinen, A., Nevalainen, A., Meklin, T., 2009. Microbial content of house dust samples determined with qPCR. Science of the Total Environment 407, 4673-4680.

Klimov, P.B., O'Connor, B., 2013. Is Permanent Parasitism Reversible?-Critical Evidence from Early Evolution of House Dust Mites. Systematic Biology 62, 411-423.

Maheswaran, D., Zeng, Y., Chan-Yeung, M., Scott, J., Osornio-Vargas, A., Becker, A.B., Kozyrskyj, A.L., 2014. Exposure to Beta-(1,3)-D-glucan in house dust at age 7-10 is associated with airway hyperresponsiveness and atopic asthma by age 11-14. PloS one 9, e98878-e98878.

Martinez, J., Eraso, E., Palacios, R., Guisantes, J.A., 1999. Enzymatic analyses of house dust mite extracts from Dermatophagoides pteronyssinus and Dermatophagoides farinae (Acari: Pyroglyphidae) during different phases of culture growth. J Med Entomol 36, 370-375.

McCarthy, C.G., Fitzpatrick, D.A., 2016. Systematic Search for Evidence of Interdomain Horizontal Gene Transfer from Prokaryotes to Oomycete Lineages. mSphere 1.

Molva, V., Bostlova, M., Nesvorna, M., Hubert, J., 2018. Do the microorganisms from laboratory culture spent growth medium affect house dust mite fitness and microbiome composition? Insect Sci.

Molva, V., Nesvorna, M., Hubert, J., 2019. Feeding Interactions Between Microorganisms and the House Dust Mites Dermatophagoides pteronyssinus and Dermatophagoides farinae (Astigmata: Pyroglyphidae). Journal of Medical Entomology.

Moraes, C.d.S., Diaz-Albiter, H.M., Faria, M.d.V., Sant'Anna, M.R.V., Dillon, R.J., Genta, F.A., 2014. Expression pattern of glycoside hydrolase genes in Lutzomyia longipalpis reveals key enzymes involved in larval digestion. Frontiers in physiology 5, 276-276.

O'Connor, B.M., 1979. EVOLUTIONARY ORIGINS OF ASTIGMATID MITES INHABITING STORED PRODUCTS, in: Rodriguez, J.G. (Ed.), Recent Advances in Acarology. Academic Press, pp. 273-278.

O'Keeffe, G., Hammel, S., Owens, R.A., Keane, T.M., Fitzpatrick, D.A., Jones, G.W., Doyle, S., 2014. RNA-seq reveals the pan-transcriptomic impact of attenuating the gliotoxin self-protection mechanism in Aspergillus fumigatus. BMC Genomics 15, 894.

Pauchet, Y., Freitak, D., Heidel-Fischer, H.M., Heckel, D.G., Vogel, H., 2009. Immunity or digestion: glucanase activity in a glucan-binding protein family from Lepidoptera. J Biol Chem 284, 2214-2224.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One 5, e9490.

Randall, T.A., Mullikin, J.C., Mueller, G.A., 2018. The Draft Genome Assembly of Dermatophagoides pteronyssinus Supports Identification of Novel Allergen Isoforms in Dermatophagoides Species. Int Arch Allergy Immunol 175, 136-146.

Rider, S.D., Jr., Morgan, M.S., Arlian, L.G., 2017. Allergen homologs in the Euroglyphus maynei draft genome. PLOS ONE 12, e0183535.

Rider, S.D., Morgan, M.S., Arlian, L.G., 2015. Draft genome of the scabies mite. Parasites & Vectors 8.

Rintala, H., Pitkaranta, M., Taubel, M., 2012. Microbial communities associated with house dust. Adv Appl Microbiol 78, 75-120.

Sayers, E.W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K.D., Karsch-Mizrachi, I., 2019. GenBank. Nucleic Acids Res 47, D94-d99.

Schwager, E.E., Sharma, P.P., Clarke, T., Leite, D.J., Wierschin, T., Pechmann, M., Akiyama-Oda, Y., Esposito, L., Bechsgaard, J., Bilde, T., Buffry, A.D., Chao, H., Dinh, H., Doddapaneni, H., Dugan, S., Eibner, C., Extavour, C.G., Funch, P., Garb, J., Gonzalez, L.B., Gonzalez, V.L., Griffiths-Jones, S., Han, Y., Hayashi, C., Hilbrant, M., Hughes, D.S.T., Janssen, R., Lee, S.L., Maeso, I., Murali, S.C., Muzny, D.M., Nunes da Fonseca, R., Paese, C.L.B., Qu, J., Ronshaugen, M., Schomburg, C., Schonauer, A., Stollewerk, A., Torres-Oliva, M., Turetzek, N., Vanthournout, B., Werren, J.H., Wolff, C., Worley, K.C., Bucher, G., Gibbs, R.A., Coddington, J., Oda, H., Stanke, M., Ayoub, N.A., Prpic, N.M., Flot, J.F., Posnien, N., Richards, S., McGregor, A.P., 2017. The house

spider genome reveals an ancient whole-genome duplication during arachnid evolution. BMC Biol 15, 62.

Shevchenko, A., Tomas, H., Havli, J., Olsen, J.V., Mann, M., 2006. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nature Protocols 1, 2856-2860.

Song, J.M., Nam, K., Sun, Y.-U., Kang, M.H., Kim, C.-G., Kwon, S.-T., Lee, J., Lee, Y.-H., 2010. Molecular and biochemical characterizations of a novel arthropod endo-β-1,3-glucanase from the Antarctic springtail, Cryptopygus antarcticus, horizontally acquired from bacteria. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 155, 403-412.

Sousa, A.C., Almeida, J.R., Pereira, C.C., Ramiro Pastorinho, M., Pereira, A.M., Nogueira, A.J., Taborda-Barata, L., Teixeira, J.P., Correia, A.C., Alves, A., 2014. Characterization of fungal communities in house dust samples collected from central Portugal-a preliminary survey. J Toxicol Environ Health A 77, 972-982.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312-1313.

Tang, V.H., Stewart, G.A., Chang, B.J., 2017. Dermatophagoides pteronyssinus lytFM encoding an NlpC/P60 endopeptidase is also present in mite-associated bacteria that express LytFM variants. FEBS Open Bio 7, 1267-1280.

van Bronswijk, J.E.M.H., 1973. Dermatophagoides Pteronyssinus (Trouessart, 1897) in Mattress and Floor Dust in a Temperate Climate (Acari: Pyroglyphidae). Journal of Medical Entomology 10, 63-70.

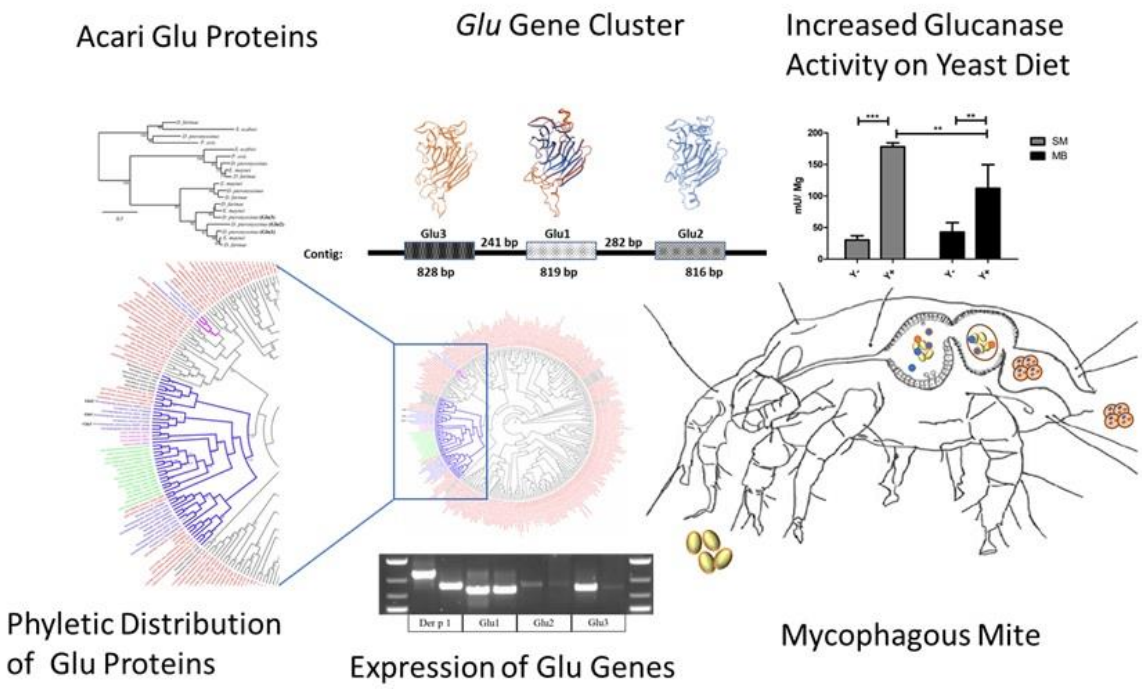
Van Rinsum, J., Klis, F.M., van den Ende, H., 1991. Cell wall glucomannoproteins of Saccharomyces cerevisiae mnn9. Yeast 7, 717-726.

Waldron, R., McGowan, J., Gordon, N., McCarthy, C., Mitchell, E.B., Doyle, S., Fitzpatrick, D.A., 2017. Draft Genome Sequence of Dermatophagoides pteronyssinus, the European House Dust Mite. Genome Announcements 5, e00789-00717.

Waldron, R., McGowan, J., Gordon, N., McCarthy, C., Mitchell, E.B., Fitzpatrick, D.A., 2019. Proteome and allergenome of the European house dust mite Dermatophagoides pteronyssinus. PLOS ONE 14, e0216171.

Zock, J.P., Heinrich, J., Jarvis, D., Verlato, G., Norback, D., Plana, E., Sunyer, J., Chinn, S., Olivieri, M., Soon, A., Villani, S., Ponzio, M., Dahlman-Hoglund, A., Svanes, C., Luczynska, C., 2006. Distribution and determinants of house dust mite allergens in Europe: the European Community Respiratory Health Survey II. J Allergy Clin Immunol 118, 682-690.

Figures



Graphical Abstract

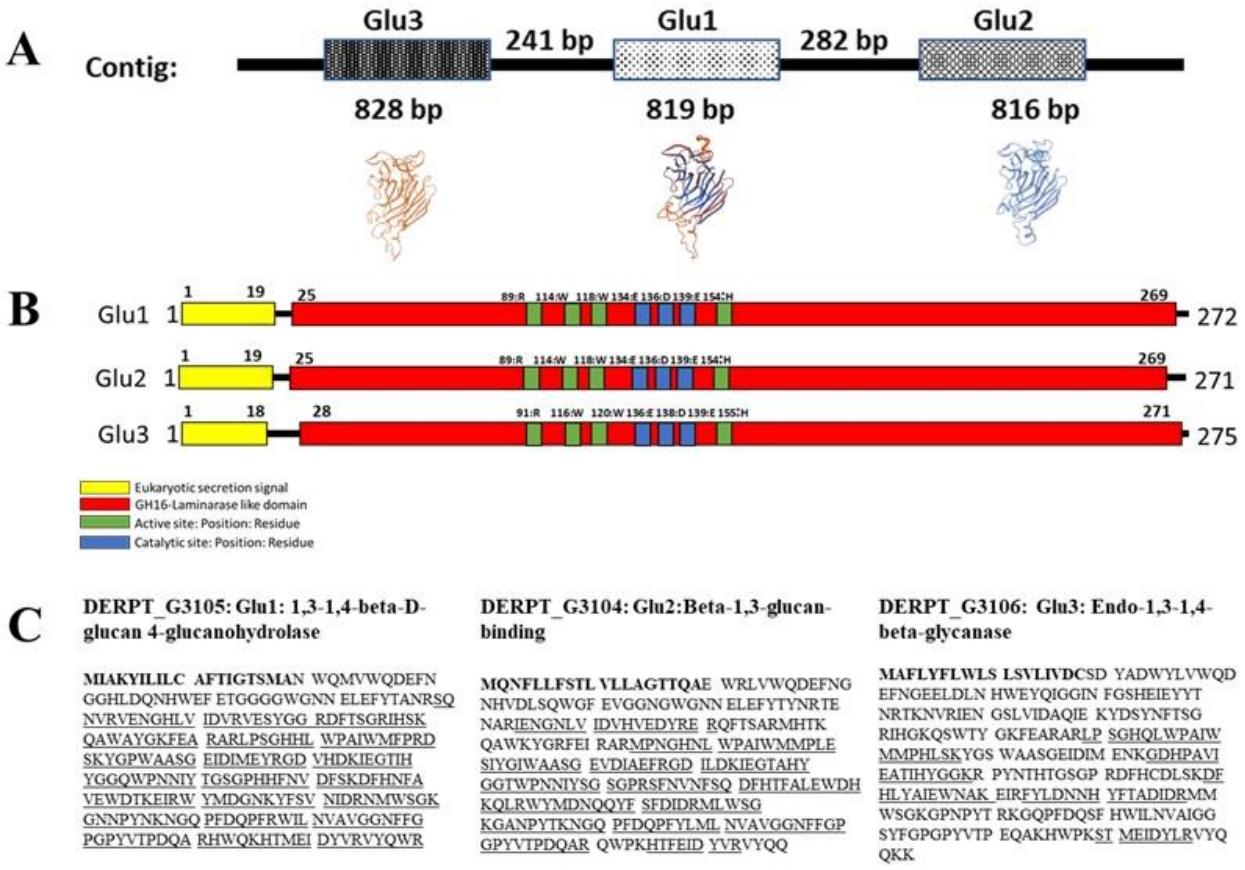

A
Contig:
Glu3
241 bp
Glu1
282 bp
Glu2
828 bp
819 bp
816 bp
B
Glu1
Glu2
Glu3
Eukaryotic secretion signal
GH16-Laminarase like domain
Active site: Position: Residue
Catalytic site: Position: Residue
C
DERPT_G3105: Glu1: 1,3-1,4-beta-D-glucan 4-glucanohydrolase
DERPT_G3104: Glu2:Beta-1,3-glucan-binding
DERPT_G3106: Glu3: Endo-1,3-1,4-beta-glycanase

Figure 1.

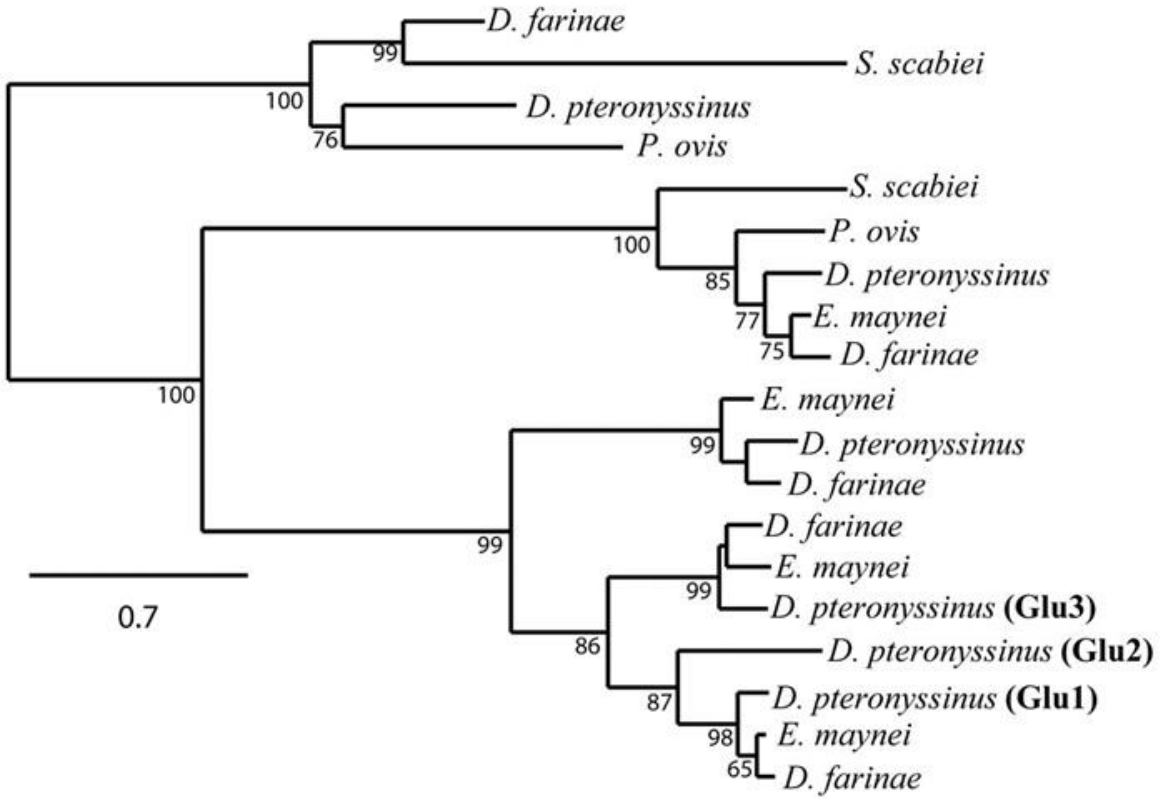

D. farinae
S. scabiei
D. pteronyssinus
P. ovis
S. scabiei
P. ovis
D. pteronyssinus
E. maynei
D. farinae
E. maynei
D. pteronyssinus
D. farinae
D. farinae
E. maynei
D. pteronyssinus (Glu3)
D. pteronyssinus (Glu2)
D. pteronyssinus (Glu1)
E. maynei
D. farinae
99
100
76
100
85
77
75
100
99
99
99
86
87
98
65
0.7

Figure 2.

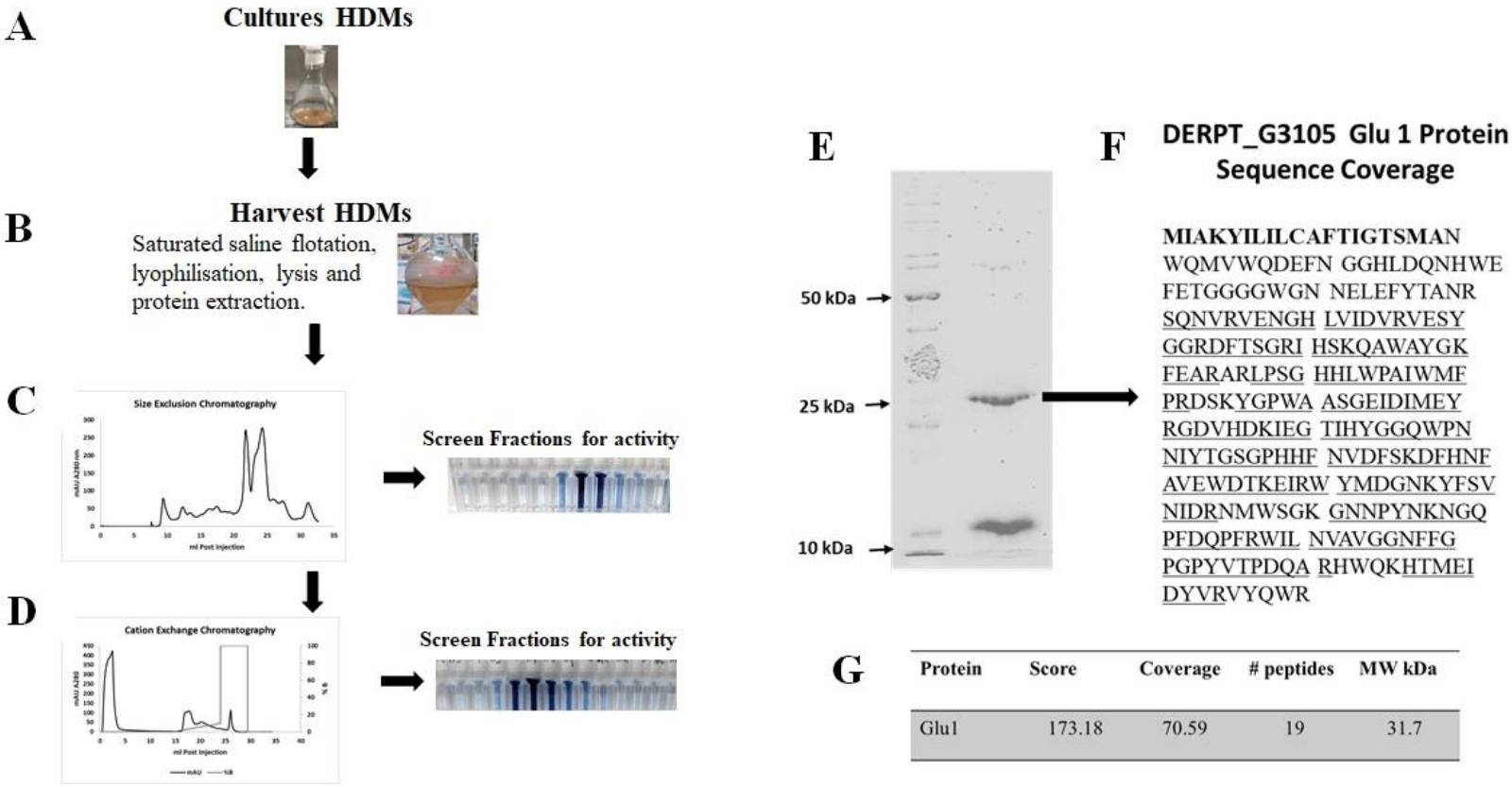
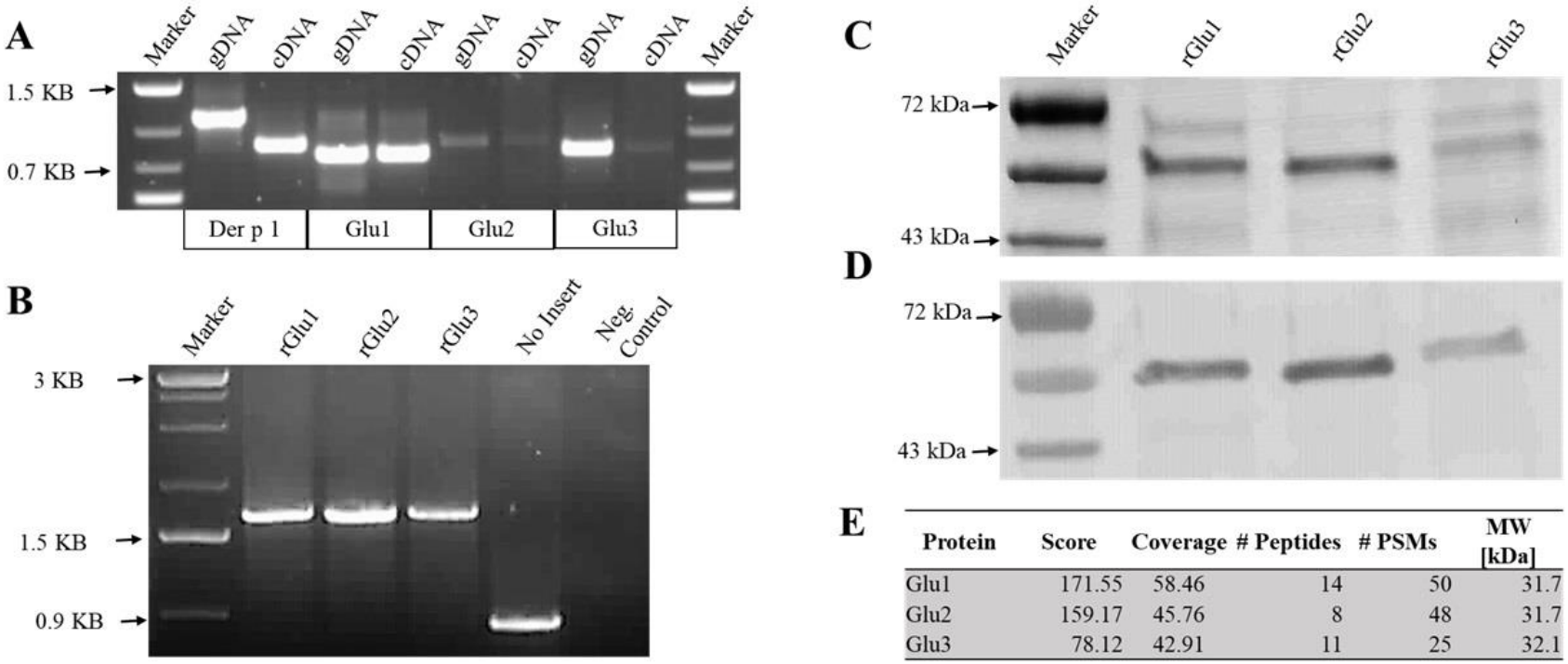

A
Cultures HDMs
B
Harvest HDMs
Saturated saline flotation, lyophilisation, lysis and protein extraction.
C
Size Exclusion Chromatography
Screen Fractions for activity
D
Cation Exchange Chromatography
Screen Fractions for activity
E
50 kDa
25 kDa
10 kDa
F
DERPT_G3105 Glu 1 Protein Sequence Coverage
MIAKYILILCAFTIGTSMAN
WQMVWQDEFN GGHLDQNHWE
FETGGGGWGN NELEFYTANR
SQNVRVENGH LVIDVRVESY
GGRDFTSGRI HSKQAWAYGK
FEARARLPSG HHLWPAIWMF
PRDSKYGPWA ASGEIDIMEY
RGDVHDKIEG TIHYGGQWPN
NIYTGSGPHHF NVDFSKDFHNF
AVEWDTKEIRW YMDGNKYFSV
NIDRNMWSGK GNNPYNKNGQ
PFDQPFRWIL NVAVGGNFFG
PGPYVTPDQA RHWQKHTMEI
DYVRVYQWR
G
Protein | Score | Coverage | # peptides | MW kDa
Glu1 | 173.18 | 70.59 | 19 | 31.7

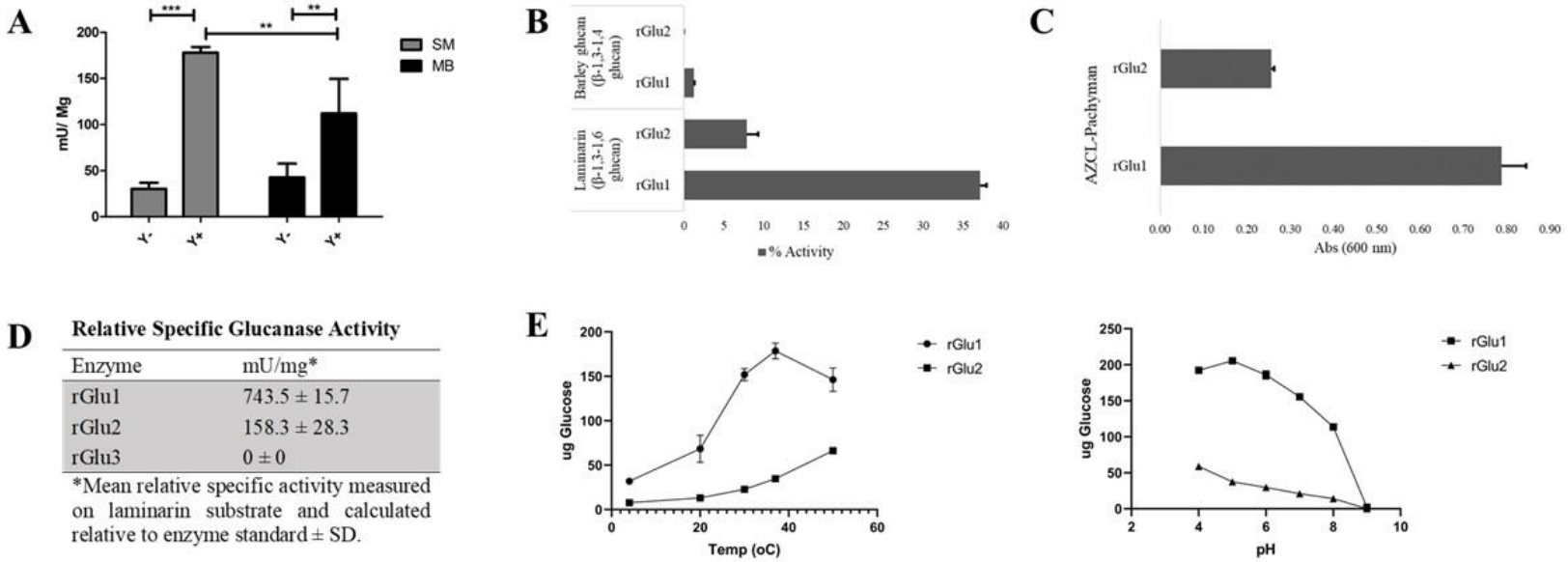
Figure 3.



E

Protein	Score	Coverage	# Peptides	# PSMs	MW [kDa]
Glu1	171.55	58.46	14	50	31.7
Glu2	159.17	45.76	8	48	31.7
Glu3	78.12	42.91	11	25	32.1

Figure 4.



Enzyme	mU/mg*
rGlu1	743.5 ± 15.7
rGlu2	158.3 ± 28.3
rGlu3	0 ± 0

*Mean relative specific activity measured on laminarin substrate and calculated relative to enzyme standard ± SD.

Figure 5.

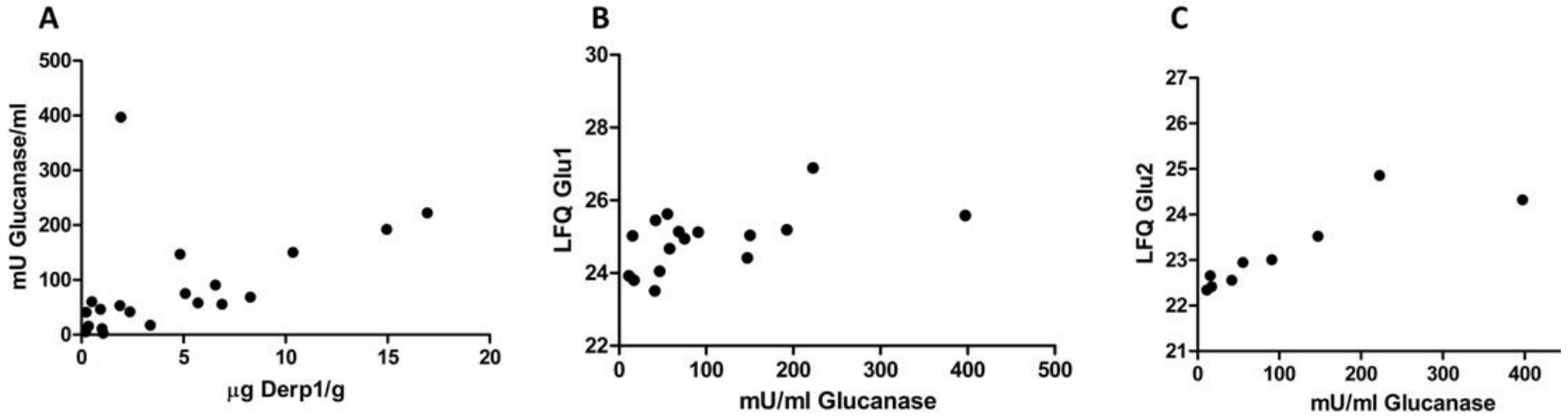
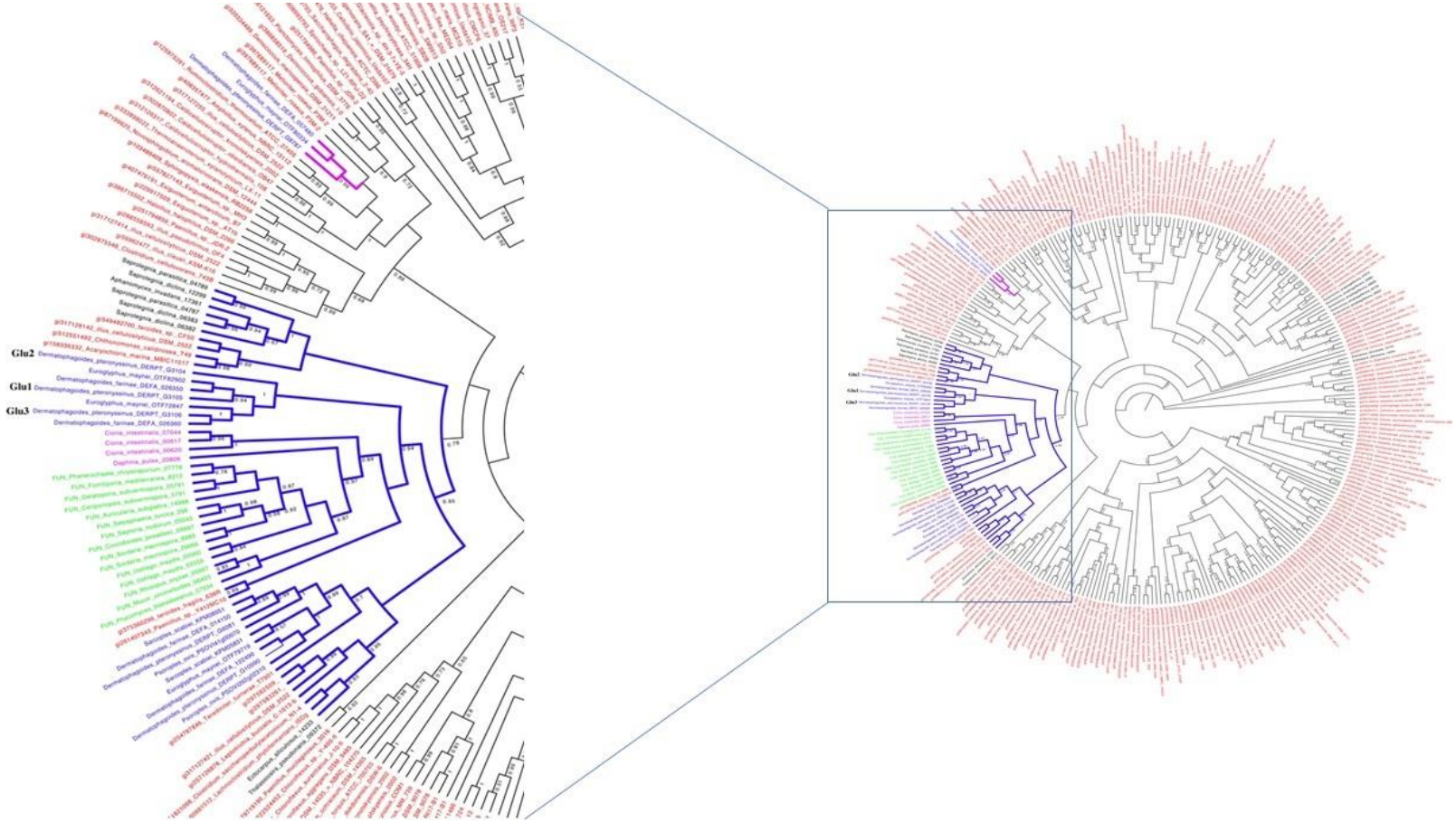

A
mU Glucanase/ml
500
400
300
200
100
0
0
5
10
15
20
µg Derp1/g
B
LFQ Glu1
30
28
26
24
22
0
100
200
300
400
500
mU/ml Glucanase
C
LFQ Glu2
27
26
25
24
23
22
21
0
100
200
300
400
mU/ml Glucanase

Figure 6.



S1 Figure

Figure Legends

Figure 1. Bioinformatic Analysis of *D. pteronyssinus* Glucanase Trigene Cluster Containing Glu1, Glu2 & Glu3. A. Illustration of *Glu* gene cluster depicting gene length, inter gene distance between *Glu1*, *Glu2* and *Glu3* and predicted 3D protein structure (swissmodel.expasy.org). **B**. Map of protein functional domains in Glu1, Glu2 and Glu3 predicted using InterproScan. **C**. Protein sequences showing LC-MS/MS sequence coverage (underlined).

Figure 2. Maximum Likelihood Phylogeny of Glu Homologs in Fully Sequenced Acari Proteomes. Bootstrap support values are shown at nodes. All homologs come from the Acariformes order, no homologs were located in members of the Parasitiforme order.

Figure 3. Workflow Depicting Purification and Subsequent Identification of Novel Glucanase, Glu1 (DERPT_G3105) from *D. pteronyssinus* airmid Protein Extracts. Flow diagram depicting proteomic strategy for the identification of the glucanases from *D. pteronyssinus* airmid. **A.** *D. pteronyssinus* airmid were cultured for 28 days, then **B**. Separated from culture medium using the saturated saline flotation method and proteins from whole body homogenate were extracted. **C**. Gel filtration chromatography of *D. pteronyssinus* airmid whole body homogenate. Protein extract (2.75 mg; 500 µl injection) separated by size exclusion chromatography (Superdex 200 10/300 GL; 0.4 ml/min flow rate). Fractions (2 ml) were collected between 8 ml and 26 ml post-injection and assayed for glucanase activity. Positive fractions (16 – 22 ml post-injection) were pooled and further separated by **D**. Cation exchange chromatography. Glucanases bound the column (HiTrap SP Xl) and were eluted in 60-90 mM NaCl. **E**. SDS-PAGE analysis of glucanase active semi-purified native protein elution **F.** LC-MS/MS of SDS-PAGE band at approximately 25 kDa identified a putative β-1,3-1,4 glucanase designated Glu1, predicted eukaryotic secretion signal highlighted in bold. **G.** High confidence identification of *D. pteronyssinus* Glu1 protein by LC-MS/MS with 70.6% sequence coverage.

Figure 4. Cloning and Recombinant Protein Expression of *D. pteronyssinus* Glucanase Trigene Cluster Containing *Glu1*, *Glu2* & *Glu3*. **A**. RT-PCR of Glu1, Glu2 & Glu3 from genomic DNA (gDNA) and cDNA confirming gene expression and absence of introns. *Der p1* gene containing three introns revealed gDNA amplicon of ~1135 bp and cDNA amplicon of ~867 bp. **B**. PCR confirmation of insertion of *E. coli* codon-optimised r*Glu1*, *rGlu2* and *rGlu3* gene sequences into pEX-N-GST expression vector. Amplicons containing insert were seen at ~1700 bp and amplicons from empty plasmid were seen at ~900 bp. **C.** SDS-PAGE of affinity purified GST-rGlu1, GST-rGlu2 and GST-rGlu3 and **D.** Western blot of same using an anti-GST antibody. **E**. LC-MS/MS of GST-rGlu1, GST-rGlu2 and GST-rGlu3 excised from SDS-PAGE gel (Figure 3C) confirming protein expression and successful purification.

Figure 5. Biochemical Characterisation of *D. pteronyssinus* airmid Glucanases. **A**. Relative specific glucanase activity (mU/mg protein) in *D. pteronyssinus* airmid protein extracts from mites grown with and without yeast. Protein extracts from Spent Culture Medium (SM) and Mite Body (MB) of *D. pteronyssinus* grown on HDMMM containing Yeast (Y+) and MM without yeast (Y-). Values represent mean of five replicates. *** Unpaired T-test $P \leq 0.001$. ** Unpaired T-test $P \leq 0.01$. **B.** Substrate specific activity (Percentage of substrate converted to glucose) of GST-rGlu1 and GST-rGlu2 on laminarin and Barley glucan. **C.** Activity of GST-rGlu1 and GST-rGlu2 on AZCL-Pachyman substrate. **D.** Relative specific β-1,3 glucanase activity (mU/mg) of GST-rGlu1 an GST-rGlu2. **E.** Activity of GST-rGlu1 and GST-rGlu2 activity in the temperature range 4 ^{o}C to 50 ^{o}C, shows optimal temperature for GST-rGlu1 to be 37 ^{o}C and GST-rGlu2 to be ≥ 50 ^{o}C. **F.** Activity of GST-rGlu1 and GST-rGlu2 in the pH range 4 to 9 showing optimal pH of GST-rGlu1 to be pH 5 and GST-rGlu2 to be ≤ pH4. B-E: Values: mean values from triplicate sampling. Error bars: SD.

Figure 6. β-1,3 glucanase Activity in House Dust Correlates with LFQ Intensity of Glu1 and Glu2. Proteomic analysis of Der p 1 positive house dust protein extracts. **A**. A positive correlation between β-1,3 glucanase activity (mU/ml) and Der p 1 (µg/gram dust) content of house dust protein extracts. **B.** & **C.** Correlation between β-1,3 glucanase

activity (mU/ml) and LFQ intensity of Glu 1 (n = 16) and Glu 2 (n = 9), respectively, identified in house dust protein extracts. *Spearman's Correlation $p \leq 0.05$ ** Spearman's Correlation $p \leq 0.005$ *** Spearman's Correlation $p \leq 0.0001$.

S1 Figure. Maximum Likelihood Phylogeny of Glu Homologs from Diverse Taxa. Bootstrap support values are shown at nodes. Bacterial, Fungal, Acariformes and other animals have their names highlighted with red, green, blue and purple text, respectively. Glu homologs are grouped in a single clade (highlighted with blue coloured branches) that also contains the majority of eukaryote homologs. Three additional Acariformes homologs (highlighted with purple branches) are grouped amongst bacterial species.

Tables

Table 1: Genome Assemblies Utilised for Phylogenetic Analysis.

Species	Genome Reference	Suborder
Centruroides sculpturatus		Arachnid outgroup
Parasteatoda tepidariorum		Arachnid outgroup
Dermatophagoides farinae	(Chan et al., 2015)	Acariformes
Euroglyphus maynei	(Rider et al., 2017)	Acariformes
Sarcoptes scabiei	(Rider et al., 2015)	Acariformes
Tetranychus urticae	(Grbić et al., 2011)	Acariformes
Psoroptes ovis	(Burgess et al., 2018)	Acariformes
Ixodes scapularis	(Schwager et al., 2017)	Parasitiforme
Galendromus occidentalis	(Gulia-Nuss et al., 2016)	Parasitiforme
Rhipicephalus microplus	(Barrero et al., 2017)	Parasitiforme
Tropilaelaps mercedesae	(Dong et al., 2017)	Parasitiforme
Varroa destructor	(Cornman et al., 2010)	Parasitiforme
Varroa jacobsoni		Parasitiforme

Genomes available for 11 species from the subclass Acari, six of these genomes belong to the Parasitiformes superorder while the remaining five belong to the Acariformes superorder. The evolutionary history of three *Glu* genes relative to closely related species was investigated by taking the three corresponding Glu protein sequences and using each as a query sequence in a BLASTp 1 (Altschul et al., 1997) database search (e-value cutoff of $1e^{-10}$) against a local protein database containing the eleven other Acari species listed.

Table 2. List of Primers.

Primer Name	**Annealing Temperature (ºC)**	**Sequence (5'- 3')**
Derp1_F	59	TCGTCCATCATCGATCAAAA
Derp1_R		TCGATGTTGGCAGCAAAATA
Glu1_Nat_F	63	ATGGCCAATTGGCAGATGGTC
Glu1_Nat_R		TTATCGCCATTGATAAACACGAACAT
Glu3_Nat_F	51	ATGGCTTTTCTCTACTTCC
Glu3_Nat_R		TTATTTTTTTTGTTGATAAACAC
Glu2_Nat_F	50	ATGCAAAATTTTCTTTTGTTT
Glu2_Nat_R		TTATTGTTGATAAACACGGAC
Glu3_Nat_F	51	ATGGCTTTTCTCTACTTCC
Glu3_Nat_R		TTATTTTTTTTGTTGATAAACAC
pEX-N-GST_F	56	AACGTATTGAAGCTATCCCAC
pEX-N-GST_R		TTCTACCATCGACACCACCA

Primers used for the amplification of *Glu1*, *Glu2* and *Glu3* and *Der p 1* genes from gDNA and cDNA. pEX-N-GST primers designed to span the vector multiple cloning site.

Table 3. Purification of β-1-3 glucanase Active Fraction from *D. pteronyssinus* airmid Protein Extracts.

Purification Step	**Volume (ml)**	**Total Protein (mg)**	**Total Activity (mU)**	**Specific Activity (mU/mg)**	**Yield (%)**	**Purification Factor**
Crude enzyme	0.5	1.4	148	106.07	100	1
Gel Filtration	0.4	0.48	221	460.01	146	4.3
Cation Exchange	0.1	0.02	13.4	670.45	8.8	6.3

Crude enzyme (106 mU/mg) was separated by gel filtration chromatography, β-1-3 glucanase active fractions (16 to 22 ml post-injection) were pooled, concentrated and buffer exchanged into 50 mM MES pH 5.5 for cation exchange separation. Glucanases bound the column (HiTrap SP Xl) and eluted between 60 and 90 mM sodium chloride. The resultant β-1-3 glucanase active fraction contained 20 µg protein with relative specific activity of 670 mU/mg. Typical values given.

Table 4. LFQ LC-MS/MS of Proteins Excreted by *D. pteronyssinus* airmid cultured With and Without Yeast.

Protein	Y+	**Y-**	**Peptides**	**Sequence coverage [%]**	**MS/MS count**	**Log_2 Difference**	**Percentage Change**	***P*-value**
Glu1	32.494	32.096	17	67.6	316	0.398	+31.7%	0.044
Glu2	30.941	30.672	9	58.7	133	0.268	+20.4%	0.033
Glu3	25.825	25.349	5	25.8	30	0.467	+39.0%	0.012

Proteins from SM and MB from *D. pteronyssinus* airmid fed with yeast (Y+) or without yeast (Y-) were extracted, normalised and analysed by LC-MS/MS. Glu1, Glu2 and Glu3 were present in higher abundance ($P \leq 0.05$) in both in SM and MB of *D. pteronyssinus* airmid fed with yeast. LFQ intensity: Log (2), mean of four replicates. *P*-value: Students T-test. Y+: Growth media containing bakers' yeast. Y-: Growth media without bakers' yeast. SM: Spent Culture Medium. MB: Mite Body.

Supplementary Data

Appendix 2. Electronic Files (Excel)

S1 Data.

Proteins (*n* = 753) from SM from *D. pteronyssinus* airmid fed with yeast (Y+) or without yeast (Y-) were extracted, normalised and analysed by LC-MS/MS. Proteins with significant differences in abundance (Students T-test, $P \leq 0.05$) between SM Y+ and SM Y- are marked with + in column "C: Student's T-test Significant SM-_SM+".

S2 Data.

Subset of S1 Data containing proteins uniquely present in SM Y+ or SM Y-. Unique: present in three of four replicates in one condition and absent from three of four replicates in the other condition.

5.0 Discussion

5.1 Summary

When this study commenced, *D. pteronyssinus* genomic and protein sequence data was limited to a discrete number of genes and a select number of proteins. Thus, the work presented in this thesis represents the first large-scale peer reviewed study of the genome and proteome of *D. pteronyssinus* using the airmid strain. Specifically, genome sequencing, assembly and gene identification led to the high-quality publication of the first *D. pteronyssinus* genome sequence (Chapter 2). This genome assembly facilitated extensive characterisation of the predicted proteome, allergenome and proteome of laboratory reared *D. pteronyssinus* airmid with reference to the proteome of wildtype *D. pteronyssinus* (Chapter 3). Genomic data enabled the identification of 4,002 gene products, confirming their presence and identifying their sites of localisation. Combined proteomic and genomic data led to the discovery of three novel Glycosyl hydrolase family 16 proteins (Chapter 4). Recombinant expression and extensive bioinformatic, biochemical and proteomic analysis revealed these to be involved in digestion of yeasts. The data presented in this thesis, provides new insights into the molecular physiology and pathogenicity of *D. pteronyssinus.* The principal research achievements from the work presented in this thesis are summarised in Figure 5.1

This discussion will consider (i) Implications of expanding the genome of *D. pteronyssinus* and generation of a high quality predicted proteome, (ii) Novel insights from orphan genes, (iii) new insights into the biochemical pathways that allow *D. pteronyssinus* to persist in the home, (iv) implications of the intra-mite proteome for allergy (v) implications of the excretome for allergy (vi) use the wildtype proteome as a basis for interpreting laboratory proteomic studies (vii) enzymes capable of digesting fungal components (viii) implications of this research in the development of biocontrol strategies, and diagnostic products and therapeutic extracts. This thesis will conclude with a consideration of how this work may be expanded in the future.

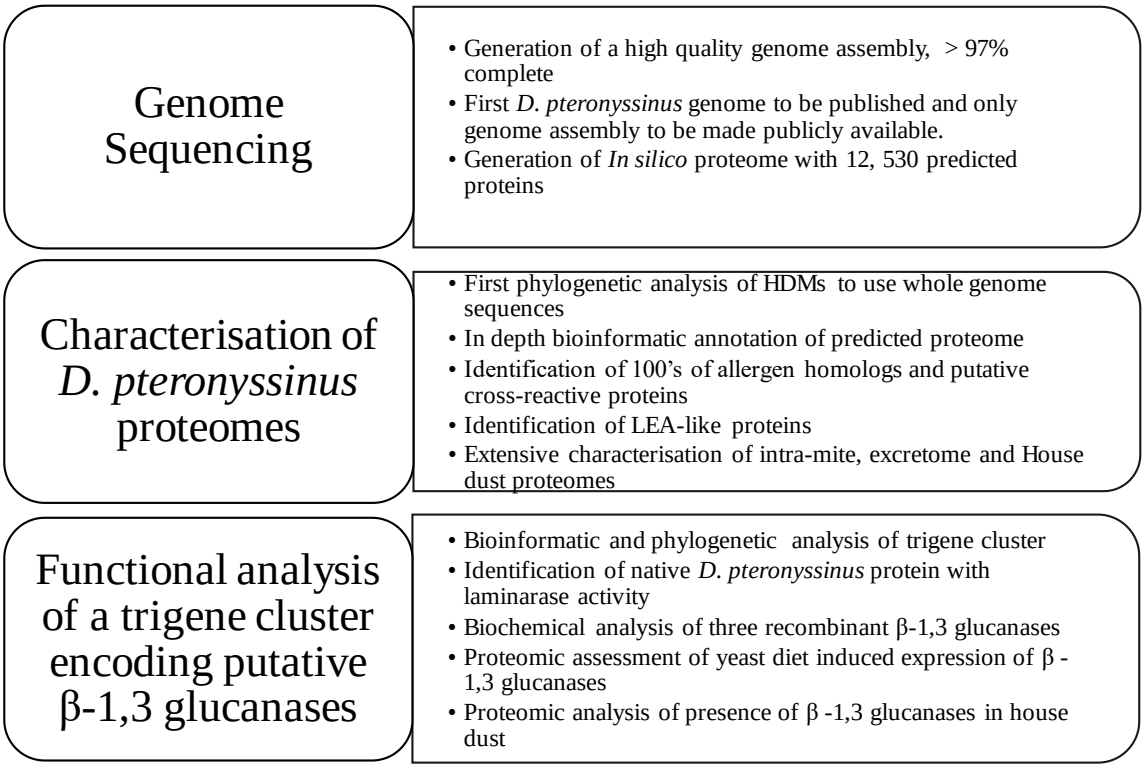


Figure 5.1. Summary of Primary Research Outputs. Main research outputs from each research area.

5.2 Implications of Expanding the Genome of *D. pteronyssinus* and Generation of a High Quality Predicted Proteome

Without the generation of a high-quality complete genome assembly, it would have been impossible to conduct proteomic experiments to their full potential or generate large-scale novel information about this species.

The genome sequencing of *D. pteronyssinus* airmid in the early phase of this project enabled the first comprehensive proteogenomic study of *D. pteronyssinus* to be conducted. This sophisticated computational strategy provides an interface between genomics and proteomics, integrating nucleotide sequencing and MS data from the same sample (Ruggles *et al.*, 2017). Resultant information can be used to assess genome completeness, accuracy of gene models and compare genome assemblies to determine (i) the relative completeness of the query genomes and (ii) the most appropriate model for further proteomics studies.

We employed a proteogenomics strategy to exploit data derived from the *D. pteronyssinus* airmid genome sequencing project, coupling it with extensive proteomic data compiled during the course of this study to (i) assess *D. pteronyssinus* airmid genome completeness and (ii) validate gene models.

The *D. pteronyssinus* airmid assembly was estimated to be 96.86% complete, containing 419 of the 429 CEGMA eukaryotic core genes. A*b initio* gene prediction discovered 12,530 gene models, containing 48,371 exons in total (Waldron *et al.*, 2017). Proteogenomic interrogation of a six-frame translation of the genome facilitated the mapping of 615,150 LC-MS/MS spectra (28,001 non-redundant) to the predicted proteome and 402,998 (21,505 non-redundant) spectra to the assembly. The higher number of spectra mapped to the predicted proteome is to be expected, as peptides spanning intron-exon junctions are mapped only to the predicted proteome, and cannot be identified from the genome assembly alone.

What is most informative about using this approach is to identify peptides that were exclusively matched to the genome, as this gives an insight into the degree to which gene

calling was successful. Our analysis showed that 96.2 % of peptides that were identified in searches against the genome assembly were also located within predicted protein coding genes, indicating a high degree of gene call accuracy. This result further supports the CEGMA estimation of the D. *pteronyssinus* airmid genome assembly completeness, and both estimates put the Waldron *et al.* (2017) assembly and predicted proteome at approximately 96 % complete.

Among the remaining 822 non-redundant spectra, that matched exclusively to the genome, 65 were located adjacent to protein coding genes, indicating these gene models may need extending. The remaining peptides were matched to loci outside of open reading frames. As these represented only a tiny proportion of the total number of protein coding genes, re-evaluation of the genome assembly isn't warranted at this point.

We extended our analysis to examine other recently published *D. pteronyssinus* genomes to determine if *D. pteronyssinus* airmid was the best model for our proteomic characterisation studies. The Randall *et al.* (2018) and Liu *et al.* (2018) assemblies reported higher numbers of protein coding genes than the Waldron *et al.* (2017) assembly. However, interrogation of all three gene sets with LC-MS/MS spectra revealed the highest number of proteins were identified using the Waldron *et al.* (2017) proteome, resulting in 4,581 proteins identified by 2 or more peptide spectrum matches. This equated to 36.5% of the predicted proteome being positively matched to LC-MS/MS spectra. The Randall *et al.* (2018) gene set contained 19,368 predicted proteins, our analysis positively identified 4,416 of these, a comparable number to the Waldron *et al.* (2017) assembly, however this only accounted for 22.8% of their gene set. A similar trend was seen for the Liu *et al.* (2018) gene set, which contained 15,846 predicted proteins, of which only 3,408 were identified, accounting for 21.5% of their gene set. Moreover, despite having fewer predicted proteins than the other two available *D. pteronyssinus* genome assemblies, the *D. pteronyssinus* airmid gene set contained approximately 20% more amino acids with longer protein coding genes.

These results highlight the variation that can result from differences in gene calling methodologies. Unbiased proteogenomic analysis of the available *D. pteronyssinus*

genome assemblies revealed the Waldron *et al.* (2017) assembly and predicted proteome to be the most comprehensive, incorporating more a.a, longer protein coding genes and resulting in more protein identifications than the other assemblies. Thus, providing compelling evidence of its strength for future proteomic studies on *D pteronyssinus* airmid. Moreover, as the only publicly available *D. pteronyssinus* genome assembly, this data will no doubt prove to be an invaluable resource to other HDM researchers. A similar analysis should be employed by other research groups to determine which *D. pteronyssinus* genome assembly offers the best gene models for their particular strain of mite or proteomic methodology. As all three genome assemblies contained genes not found in the other strains, the genome offering the highest number of proteomic identifications should be experimentally evaluated and used for subsequent studies.

A high-coverage genome assembly merely provides the foundation for proteomic studies, the highest benefit comes from using an extensively annotated predicted proteome. Compilation of such a proteomic database was critical to achieving our goal of generating a multidimensional view of *D. pteronyssinus*. The *D. pteronyssinus* airmid predicted proteome was subjected to a multi-step protein annotation workflow using BLAST2GO Version 5.0 software package, to perform (i) BLAST homology annotation, (ii) Assign Gene Ontology (GO) terms, (iii) identify; families, domains, sites and repeats in predicted proteins, (iv) Identify predicted secretion peptides and (v) map GO terms to enzyme codes. This combined annotation strategy resulted in 96% of predicted proteins being assigned one or more annotation. Once complete, the proteomic database, facilitated the exploitation of proteomic data to identify important protein families, enzymes, allergens and predicted allergens expressed by this medically important HDM.

5.3 Novel Insights from Orphan Genes

Taxonomically restricted genes or orphan genes can offer huge insight into lineage-specific adaptations. These genes lack homologs in other closely related species and typically account for 10% – 30% of genes (Wissler *et al.*, 2013). We identified orphan genes by further examining a subset of predicted proteins, those that had revealed no blast homology to proteins in NCBInr (n = 3,906) and searching them against full gene

sets of 11 arachnids (S1 Table). Many had homologues in other arachnids, however a subset of 1,848 protein coding genes were unique to *D. pteronyssinus*. Almost 80% of these *D. pteronyssinus* orphan genes were specific to the airmid strain, as they were not found in the other *D. pteronyssinus* assemblies (Randall *et al.*, 2018, Liu *et al.*, 2018).

These data provide significant insight into the genetic mechanisms underpinning *D. pteronyssinus* populations, showing that 88.3% of the *D. pteronyssinus* airmid genome represents core protein coding genes, found in all three *D. pteronyssinus* assemblies, with the remaining 11.7% being airmid strain-specific. Some uncharacterised *D. pteronyssinus* airmid-specific proteins may represent adaptations, while others are a likely consequence of genetic drift occurring in isolated populations (Tutar, 2012). Compilation of proteomic data from all proteins identified in this study (representative proteome analysis) validated more than 10 % of these airmid strain orphan genes as true protein coding genes, showing they are expressed under normal laboratory conditions. The functions of these strain specific proteins are intriguing, their role in strain specific adaptions may be discovered through further proteomic investigation.

5.4 New Insights into the Biochemical Pathways that Allow *D. pteronyssinus* to Persist in the Home

Hygrothermal conditions dictate duration of life cycle, fecundity, survival, feeding and allergen production in *D. pteronyssinus* and other HDMs (Arlian *et al.*, 1998, Colloff, 1992). The ability of mites to withstand periods of environmental extremes is key to their continued persistence in the home (Colloff *et al.*, 1992, Calderón *et al.*, 2015). Thus, identifying protein coding genes that confer hygrothermal stress tolerance may provide much needed insight into biochemical pathways that enable such stress resistance. Disruption of these pathways could curtail *D. pteronyssinus* populations.

In this study we used a similar bioinformatic approach to Gusev *et al.* (2014) to perform the first study of its' kind in HDMs, to identify a subset of *D. pteronyssinus* airmid predicted proteins which may have a role in desiccation and cold tolerance. These proteins, 18 in total, were indirectly assessed through homology with known late embryogenesis abundant proteins (LEAPs) demonstrated to allow other organisms to

survive hygrothermal extremes (Hunault & Jaspard, 2010). Expression of seven of these LEA-like proteins was validated by proteomics, all appeared to be expressed at low levels except for two, DERPT_G12026 and DERPT_G404. These highly abundant LEA-like proteins are expressed under optimal non-desiccating laboratory growth conditions and were identified in both mite body (MB) and spent media (SM).

Anhydrobiotic organisms can tolerate extreme water loss exceeding 97%, there is evidence for similar mechanisms in *D. pteronyssinus* as the protonymph stage is highly resistant to desiccation, surviving many months at 0% RH (Arlian, 1992). Future studies may focus on the protonymph life stage as a model for desiccation resistance in *D. pteronyssinus*. Proteomic investigation of desiccated and non-desiccated protonymphs may yield important information on whether these LEA-like proteins aid in stabilisation of proteins, membrane structure and osmoprotection during hygrothermal stress (Sinclair *et al.*, 2013). Moreover, expansion of the LEA work presented in this study may eventually allow for development of much needed secondary prevention strategies for mite-induced diseases (Sanchez-Borges *et al.*, 2017).

5.5 Implications of the Mite Body Proteome for Allergy

A key goal of this study was to better understand the internal proteome of *D. pteronyssinus* and how this may relate to allergen exposure and allergy. Thus, a characterisation study was undertaken to catalogue proteins associated with *D. pteronyssinus* airmid under normal laboratory growth conditions. A fractionation approach was taken to improve identification of low abundant proteins. Coupling LC-MS/MS with pre-fractionation facilitated the identification of 248 proteins from the mite body (MB) that were not detected through direct proteomic analysis. Following removal of proteins also identified from spent culture media extracts, 1,076 proteins were found to be exclusively present in the body of *D. pteronyssinus* airmid.

All proteins are synthesised in the MB, and they are then either retained in the MB or excreted. However, MB localised proteins may still be found in the SM but are typically

low abundance in these extracts compared to the MB. Direct quantification of the 1000's of proteins identified during this study is not possible, therefore a comparative proteomic strategy was employed to establish protein localisation to the MB or SM. Where a protein was identified in both MB and SM extracts, localisation was inferred by higher LFQ intensity and ms count.

The route by which a sensitised individual is exposed to HDM allergens has significant implications for the types of allergens they react to. We frequently think of exposure in terms of inhalation, however skin contact exposure is important for certain allergic patients, particularly those with atopic dermatitis (AD). Bedding and clothing provide a means for HDMs to come into direct contact with the skin and cause transdermal sensitisation (Teplitsky *et al.*, 2008, Clarke *et al.*, 2015a, Tovey *et al.*, 1995). Distinct patterns of allergen reactivity is seen in patients with AD symptoms compared to those with respiratory symptoms (Banerjee *et al.*, 2015). AD patients show greater reactivity to Der p 10, 11 and 14, than patients with respiratory symptoms only (Banerjee *et al.*, 2015). Our localisation analysis identified these allergens to be some of the most abundant MB proteins in *D. pteronyssinus* airmid. These allergens are non-enzymatic, therefore the propensity to induce IgE responses may be linked to their high abundance. If so, other newly identified and highly abundant proteins in *D. pteronyssinus* MB may also cause sensitisation and therefore may represent interesting targets for future studies on AD patients.

In addition, the MB was abundant in proteins showing distinct homology to known allergens, our analysis shows that more than 11% of all the proteins identified from the MB of *D. pteronyssinus* airmid could elicit immune responses either directly, or as a result of homology based cross-reactivity. The degree to which these putative allergens induce allergic reactions in HDM sensitised individuals remains a topic for future studies.

Other important non-allergenic pathogenic proteins were identified in the MB of *D. pteronyssinus,* including phospholipase A2. Recent research has identified HDM phospholipase A2 to be a key driver in the pathogenesis of AD, shown to generate antigenic neo-lipids, that activate T-cells, causing inflammation (Jarrett *et al.*, 2016). The

predicted proteome of *D. pteronyssinus* airmid contained 19 predicted phospholipid A2 enzymes, however only one was identified (DERPT_G9584), it was observed in the MB dataset at low LFQ intensity. Phospholipid A2 enzymes contain a number of hydrophobic amino acid residues, allowing them to interact with the lipid substrate, this may result in the proteins being unfavourable for the aqueous protein extraction methods employed in this study (Dennis *et al.*, 2011). Optimising proteomic extraction methodologies for the identification of phospholipase A2 proteins from the MB of *D. pteronyssinus* may identify the specific proteins responsible for this pathology.

The data generated from characterising the MB of *D. pteronyssinus* sheds new light into the relationship between allergen localisation in the mite and the route of allergen exposure, presenting clear evidence that intra-mite proteins also represent potential targets for the immune system via transdermal exposure. Further proteomic investigation of *D. pteronyssinus* MB utilising alternative protein extraction methods and alternative culture media will no doubt expand upon the number of proteins identified in this study.

5.6 Implications of the Excretome for Allergy

The faecal particles of HDMs are the primary vector for allergens in patients with respiratory symptoms. Due to their small size, they are inhaled deeply into the lungs, coming into direct contact with the lung epithelium. In the context of sensitisation, proteins present in the faeces are of particular importance, therefore predicting what types of proteins may be excreted is equally as important (Tovey *et al.*, 1981, Platts-Mills *et al.*, 1986, Jacquet, 2013).

It has been estimated that a third of all proteins expressed by an organism are secreted (Orfanoudaki *et al.*, 2017). In our analysis, eukaryotic secretion signals were present in 10.3% of predicted proteins, and 12% of validated proteins in the representative proteome. However, they were significantly over represented in the SM (P = 1E-26) where more than 22 % of all proteins identified contained eukaryotic secretion signals. This data shows that the presence of a secretion signal is a strong indicator that a protein may be present in the faeces, and therefore may be used as a predictor for route of exposure when assessing new allergens.

Given what we know about HDM digestive physiology, it is therefore no surprise that over 52% of proteins identified in the SM proteome have predicted enzyme activity. However, specific types of enzymes were over-represented in this dataset, most notedly ones assigned the term Acting on peptide bonds (P = 6.73E-31). The large presence of peptidases in SM was surprising. Despite making up only 3% of the predicted proteome, over 10% of all proteins in the SM had predicted peptidase activity. It is well-established that enzymes have a propensity to induce allergic responses, peptidases are the best established example of this (Huby *et al.*, 2000). Potent peptidases, such as Der p 1, disrupt epithelia barrier function and alter numerous immune system processes (Reithofer & Jahn-Schmid, 2017, Huby *et al.*, 2000). Given this new insight into the enzymatic content of SM, it is not difficult to understand how inhaling such an enzyme-rich substance can elicit the enormous levels of HDM sensitisation reported. Therefore, in the context of patient exposure, *D. pteronyssinus* proteins with predicted enzyme activity should be considered allergen candidates worthy of further study, particularly if they contain predicted secretion peptides, as they are more likely to be excreted into house dust.

5.7 The Wildtype Proteome as a Basis for Interpreting Laboratory Proteomic Studies

Establishing a laboratory model of *D. pteronyssinus*, that shows similar protein expression and excretion profiles, to that of wild-type *D. pteronyssinus* is essential to progress knowledge of this and other species of HDMs. Researchers frequently purchase commercially available *D. pteronyssinus* protein extracts to conduct studies into the mechanisms underlying HDM allergy. However, extracts from different companies are not standardised, varying substantially in the ratios of some allergens and lacking others entirely. How these extracts compare to the types of proteins or abundance of allergens found in house dust is unknown (Thomas, 2018). Our comparative proteomic analysis of house dust revealed a surprising degree of similarity between the excretome of laboratory reared *D. pteronyssinus* airmid mites to *D. pteronyssinus* proteins found in house dust. All but 13 proteins identified in house dust were also identified in the SM dataset. However, further development of the laboratory model is needed, as demonstrated for allergens Der p 5 and Der p 21. Initial proteomic assessment failed to detect Der p 5 and

Der p 21 in the SM, suggesting they were not excreted under laboratory conditions. However, proteomic analysis of house dust revealed Der p 5 to be among the top 10 most abundant proteins and Der p 21 the 63rd most abundant. This serves to demonstrate that mechanisms other than excretion can lead to allergen accumulation in house dust. It is likely that in the wildtype environment, as the mite bodies begin to degrade they release MB localised proteins. This was not observed in the laboratory model, as accumulation of dead mites is avoided by regular sub-culturing. This work demonstrates the utility of high sensitivity protein MS as a novel way to identify HDM products in the wildtype environment, contrasting them with those found in laboratory reared mites to contextualise laboratory studies.

5.8 Digestion of Fungi

The nature of HDMs-fungal interactions has been the source of much debate since the early days of HDM research (van Bronswijk, 1973, Hay *et al.*, 1993, Hay *et al.*, 1992, Naegele *et al.*, 2013). However, recent research has unambiguously demonstrated that *D. pteronyssinus* engages in mycophagy, utilising *Aspergillus spp*, *Penicillium spp*. and yeast for nutrition (Molva *et al.*, 2019). In the course of this study, we observed *D. pteronyssinus* protein extracts to exhibit laminarinase activity and devised a series of experiments to investigate if this β-1,3 glucanase activity was linked to consuming fungi.

Following positive identification of a β-1,3 glucanase active protein from *D. pteronyssinus* protein extracts, bioinformatic analysis was conducted to examine the genomes of other Acari in search of similar proteins in close relatives. This analysis revealed β-1,3 glucanases are recent adaptions in acariformes. Moreover, genes encoding β-1,3 glucanases were expanded in *D. pteronyssinus* compared to *E. maynei* and *D. farinae*. This additional Glu2 protein may allow adaptation of *D. pteronyssinus* towards mycophagy, exploiting fungal components in house dust more efficiently than other HDMs in this complex ecosystem. This novel finding may in part explain the co-existence of multiple HDM species in house dust, and why the presence of one species does not influence the numbers found of the other (Zock *et al.*, 2006).

Most research groups use dried yeast to supplement diets and improve mite population growth (Andersen, 1991, Hubert *et al.*, 2016, Eraso *et al.*, 1997, Arlian & Morgan, 2015, Hart *et al.*, 2007). Therefore, we used *S. cerevisiae* in our study as a standardised fungal medium on which to feed *D. pteronyssinus* and unambiguously demonstrate that fungal feedstuffs quantitatively increase glucanase activity and abundance in dust mite body and spent media. Further, recombinant expression of these Glu proteins confirmed the hydrolytic activity of Glu1 and Glu2 towards β-1,3 glucans in laminarin, pachyman and barley glucan. These Glu proteins were expressed and excreted basally (without yeast diet) by *D. pteronyssinus* in response to microbes naturally present in both their laboratory and wildtype environment (Molva *et al.*, 2018). A number of other enzymes capable of digesting fungal components were also identified, uniquely present in the SM of yeast-fed mites, including three alpha-mannosidase, and a fourth alpha-mannosidase was significantly increased in abundance. It is probable that Glu1, Glu2 and Glu3 act in conjunction with these enzymes to digest fungal components, and further research is needed to examine this hypothesis.

5.9 Implications of this Research in the Development of Biocontrol Strategies, Diagnostic Products and Therapeutic Extracts.

The process of chitin synthesis and remodelling is an integral part of the growth and development of all arthropods. Chitin is also a major component in the peritrophic membrane of HDMs, therefore plays an essential role in both cuticle formation and digestive physiology of HDMs. Dysregulation of the chitin synthase pathway would have a catastrophic effect on the ability of *D. pteronyssinus* to grow and feed in wild-type habitats. This chitin synthase pathway is highly conserved, utilising chitinase, β-N-acetylhexosaminidase and chitin synthase (Merzendorfer & Zimoch, 2003). Proteomic profiling of *D. pteronyssinus* airmid facilitated identification of eight predicted chitinases, five β-N-acetylhexosaminidases and two chitin synthases, putatively involved in chitin remodelling. As chitin is absent from vertebrates, dysregulation of these enzymes could provide much needed biocontrol strategies to alleviate HDM colonization of homes (Merzendorfer & Zimoch, 2003, Gøtzsche & Johansen, 2008). Future proteomic studies examining the regulation of these enzymes by *D. pteronyssinus* and in

response to exposure to various test compounds, may provide a high-throughput testing method for screening new acaricides.

International experts have outlined a number of critical areas of HDM allergy research and therapeutics development that need to be addressed in the global efforts to combat house dust mite induced diseases, those most relevant to this study are outlined below along with how the research conducted in this study may form the foundation upon which these can be addressed.

1. Development of objective methods to assess allergen exposure and environmental control outcomes (Sanchez-Borges *et al.*, 2017).

2. Mandatory documentation of exposure to the relevant allergens and "it is recommended to document the exposure level for the individual patient especially for the evaluation periods to evaluate the variation of indoor allergens" (EMA, 2008).

Protocols and techniques for the large-scale unbiased proteomic characterisation of *D. pteronyssinus* products were developed, identifying over 4,000 proteins in laboratory reared mites and 150 proteins from house dust. Extensive bioinformatic analysis identified hundreds of *D. pteronyssinus* proteins that may play a role in allergy and identifying the expression of many of these, providing the most comprehensive allergenome analysis of this medically important mite to date. Expanding these MS studies to incorporate quantitative analysis of allergenic and putative allergenic components in house dust to develop multiplex assays would provide much needed methods to address the deficits in allergen exposure knowledge highlighted above. Moreover, this would enable generation of diagnostic and therapeutic HDM extracts with allergen content and ratios that reflect natural allergen exposure.

3. Development of secondary prevention strategies for mite-induced diseases (Sanchez-Borges *et al.*, 2017).

Our molecular characterisation of *D. pteronyssinus* airmid identified potential protein targets for dysregulation of crucial physiological activities including digestion, cuticle

remodelling and hygrothermal stress tolerance. Proteomic methodologies developed in the course of this study could be implemented to assess disruption of these and other essential pathways, providing a means to measure the efficacy of environmental control measures.

5.10 Concluding Remarks

To conclude, this thesis presents the first large-scale molecular investigation of the medically important mite, *D. pteronyssinus*, in terms of genome annotation, identification of putative allergens, intra-mite proteome and excretome, proteins involved in cuticle remodelling and hygrothermal stress tolerance. Moreover, the thesis describes, in detail, novel experimental strategies for further study of this allergen vector. Overall, this work has yielded significant insight into the molecular physiology of HDMs.

Bibliography

Aalberse, R.C. (2000) Structural biology of allergens. *Journal of Allergy and Clinical Immunology* **106**: 228-238.

An, S., Chen, L.L., Long, C.B., Liu, X.Y., Xu, X.M., Lu, X.R., Rong, M.Q., Liu, Z.G., and Lai, R. (2013) Dermatophagoides farinae Allergens Diversity Identification by Proteomics. *Molecular & Cellular Proteomics* **12**: 1818-1828.

Andersen, A. (1991) Nutritional value of yeast for Dermatophagoides pteronyssinus (Acari: Epidermoptidae) and the antigenic and allergenic composition of extracts during extended culturing. *J Med Entomol* **28**: 487-491.

Arlian, L.G. (1975) Dehydration and survival of the European house dust mite, Dermatophagoides pteronyssinus. *Journal of Medical Entomology* **12**: 437-442.

Arlian, L.G. (1992) WATER-BALANCE AND HUMIDITY REQUIREMENTS OF HOUSE DUST MITES. *Experimental & Applied Acarology* **16**: 15-35.

Arlian, L.G., Bernstein, D., Bernstein, I.L., Friedman, S., Grant, A., Lieberman, P., Lopez, M., Metzger, J., Platts-Mills, T., Schatz, M., and et al. (1992) Prevalence of dust mites in the homes of people with asthma living in eight different geographic areas of the United States. *J Allergy Clin Immunol* **90**: 292-300.

Arlian, L.G., and Morgan, M.S. (2015) Reproductive biology of Euroglyphus maynei with comparisons to Dermatophagoides farinae and D. pteronyssinus. *Exp Appl Acarol* **66**: 1-9.

Arlian, L.G., Neal, J.S., and Bacon, S.W. (1998) Survival, fecundity, and development of Dermatophagoides farinae (Acari : Pyroglyphidae) at fluctuating relative humidity. *Journal of Medical Entomology* **35**: 962-966.

Arlian, L.G., and Platts-Mills, T.A.E. (2001) The biology of dust mites and the remediation of mite allergens in allergic disease. *Journal of Allergy and Clinical Immunology* **107**.

Arlian, L.G., Rapp, C.M., and Ahmed, S.G. (1990) Development of Dermatophagoides pteronyssinus (Acari: Pyroglyphidae). *Journal of Medical Entomology* **27**: 1035-1040.

Asher, M.I., Twiss, J., and Ellwood, E., (2012) 44 - The Epidemiology of Asthma. In: Kendig & Chernick's Disorders of the Respiratory Tract in Children (Eighth Edition). R.W. Wilmott, T.F. Boat, A. Bush, V. Chernick, R.R. Deterding & F. Ratjen (eds). Philadelphia: W.B. Saunders, pp. 647-676.

Avula-Poola, S., Morgan, M.S., and Arlian, L.G. (2012) Diet Influences Growth Rates and Allergen and Endotoxin Contents of Cultured <b><i>Dermatophagoides

farinae and Dermatophagoides pteronyssinus House Dust Mites. *International Archives of Allergy and Immunology* **159**: 226-234.

Banerjee, S., Resch, Y., Chen, K.W., Swoboda, I., Focke-Tejkl, M., Blatt, K., Novak, N., Wickman, M., van Hage, M., Ferrara, R., Mari, A., Purohit, A., Pauli, G., Sibanda, E.N., Ndlovu, P., Thomas, W.R., Krzyzanek, V., Tacke, S., Malkus, U., Valent, P., Valenta, R., and Vrtala, S. (2015) Der p 11 Is a Major Allergen for House Dust Mite-Allergic Patients Suffering from Atopic Dermatitis. *Journal of Investigative Dermatology* **135**: 102-109.

Bantz, S.K., Zhu, Z., and Zheng, T. (2014) The Atopic March: Progression from Atopic Dermatitis to Allergic Rhinitis and Asthma. *Journal of clinical & cellular immunology* **5**: 202.

Barberan, A., Dunn, R.R., Reich, B.J., Pacifici, K., Laber, E.B., Menninger, H.L., Morton, J.M., Henley, J.B., Leff, J.W., Miller, S.L., and Fierer, N. (2015) The ecology of microscopic life in household dust. *Proc Biol Sci* **282**.

Barton, P.J., and Buckingham, M.E. (1985) The myosin alkali light chain proteins and their genes. *The Biochemical journal* **231**: 249-261.

BLF, (2016) The battle for breath: the impact of lung disease in the UK. In. British Lung Foundation, pp.

Bolognesi, R., Terra, W.R., and Ferreira, C. (2008) Peritrophic membrane role in enhancing digestive efficiency: Theoretical and experimental models. *Journal of Insect Physiology* **54**: 1413-1422.

Boner, A., Pescollderungg, L., and Silverman, M., (2002) *The role of house dust mite elimination in the management of childhood asthma: An unresolved issue*, p. 23-31.

Boquete, M., Carballas, C., Carballada, F., Iraola, V., Carnes, J., and Fernandez-Caldas, E. (2006) In vivo and in vitro allergenicity of the domestic mite Chortoglyphus arcuatus. *Ann Allergy Asthma Immunol* **97**: 203-208.

Bordas-Le Floch, V., Le Mignon, M., Bussieres, L., Jain, K., Martelet, A., Baron-Bodo, V., Nony, E., Mascarell, L., and Moingeon, P. (2017) A combined transcriptome and proteome analysis extends the allergome of house dust mite Dermatophagoides species. *PLoS One* **12**: e0185830.

Brown, S.J. (2016) Atopic eczema. *Clinical medicine (London, England)* **16**: 66-69.

Browne, J.A., Dolan, K.M., Tyson, T., Goyal, K., Tunnacliffe, A., and Burnell, A.M. (2004) Dehydration-Specific Induction of Hydrophilic Protein Genes in the Anhydrobiotic Nematode Aphelenchus avenae. *Eukaryotic Cell* **3**: 966-975.

Calderón, M.A., Linneberg, A., Kleine-Tebbe, J., De Blay, F., Hernandez Fernandez de Rojas, D., Virchow, J.C., and Demoly, P. (2015) Respiratory allergy caused by house

dust mites: What do we really know? *Journal of Allergy and Clinical Immunology* **136**: 38-48.

Casset, A., Mari, A., Purohit, A., Resch, Y., Weghofer, M., Ferrara, R., Thomas, W.R., Alessandri, C., Chen, K.W., de Blay, F., Valenta, R., and Vrtala, S. (2012) Varying allergen composition and content affects the in vivo allergenic activity of commercial Dermatophagoides pteronyssinus extracts. *Int Arch Allergy Immunol* **159**: 253-262.

Chan, T.F., Ji, K.M., Yim, A.K.Y., Liu, X.Y., Zhou, J.W., Li, R.Q., Yang, K.Y., Li, J., Li, M., Law, P.T.W., Wu, Y.L., Cai, Z.L., Qin, H., Bao, Y., Leung, R.K.K., Ng, P.K.S., Zou, J., Zhong, X.J., Ran, P.X., Zhong, N.S., Liu, Z.G., and Tsui, S.K.W. (2015) The draft genome, transcriptome, and microbiome of Dermatophagoides farinae reveal a broad spectrum of dust mite allergens. *Journal of Allergy and Clinical Immunology* **135**: 539-548.

Chandramouli, K., and Qian, P.-Y. (2009) Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Human genomics and proteomics : HGP* **2009**: 239204.

Chanez, P., and Bourdin, A., (2008) Chapter 3 - Pathophysiology of Asthma. In: Clinical Asthma. M. Castro & M. Kraft (eds). Philadelphia: Mosby, pp. 23-34.

Chang, C.-F., Wu, F.F.-S., Chen, C.-Y., Crane, J., and Siebers, R. (2011) Effect of freezing, hot tumble drying and washing with eucalyptus oil on house dust mites in soft toys. *Pediatric Allergy and Immunology* **22**: 638-641.

Chapman, M., (2008) Allergen Nomenclature. In: Allergens and Allergen Immunotherapy. R. Lockey & D. Ledford (eds). New York: Informa Healthcare, pp. 47-58.

Clarke, D., Burke, D., Gormally, M., and Byrne, M. (2015a) Dynamics of house dust mite transfer in modern clothing fabrics. *Ann Allergy Asthma Immunol* **114**: 335-340.

Clarke, D., Dix, E., Liddy, S., Gormally, M., and Byrne, M. (2016) The influence of household pets on the composition and quantity of allergenic mite fauna within Irish homes: a preliminary investigation. *Medical and Veterinary Entomology* **30**: 21-30.

Clarke, D., Gormally, M., Sheahan, J., and Byrne, M. (2015b) Child car seats - a habitat for house dust mites and reservoir for harmful allergens. *Annals of Agricultural and Environmental Medicine* **22**: 17-22.

Colloff, M.J. (1987) Mite fauna of dust from passenger trains in Glasgow. *Epidemiology and infection* **98**: 127-130.

Colloff, M.J. (1992) AGE STRUCTURE AND DYNAMICS OF HOUSE DUST MITE POPULATIONS. *Experimental & Applied Acarology* **16**: 49-74.

Colloff, M.J., (2009) *Dust Mites*. CSIRO Publishing, Springer, Van Godewijckstraat 303311 GX Dordrecht The Netherlands.

Colloff, M.J., Ayres, J., Carswell, F., Howarth, P.H., Merrett, T.G., Mitchell, E.B., Walshaw, M.J., Warner, J.O., Warner, J.A., and Woodcock, A.A. (1992) The control of allergens of dust mites and domestic pets: a position paper. *Clin Exp Allergy* **22 Suppl 2**: 1-28.

Crowther, D., Wilkinson, T., Biddulph, P., Oreszczyn, T., Pretlove, S., and Ridley, I. (2006) A simple model for predicting the effect of hygrothermal conditions on populations of house dust mite Dermatophagoides pteronyssinus (Acari : Pyroglyphidae). *Experimental and Applied Acarology* **39**: 127-148.

Cui, Y., Gao, C., Zhou, Y., Zhou, P., Peng, M., Lin, Y., and Peng, J. (2010) Phylogenetic analysis of house dust mites. *Central European Journal of Medicine* **5**: 69-74.

Cunningham (1998) Direct measurements of temperature and humidity in dust mite microhabitats. *Clinical & Experimental Allergy* **28**: 1104-1112.

Custovic, A., Taggart, S.C., Francis, H.C., Chapman, M.D., and Woodcock, A. (1996) Exposure to house dust mite allergens and the clinical activity of asthma. *J Allergy Clin Immunol* **98**: 64-72.

Dannemiller, K.C., Weschler, C.J., and Peccia, J. (2017) Fungal and bacterial growth in floor dust at elevated relative humidity levels. *Indoor Air* **27**: 354-363.

De Boer, R., Kuller, K., and Kahl, O. (1998) Water balance of Dermatophagoides pteronyssinus (Acari : Pyroglyphidae) maintained by brief daily spells of elevated air humidity. *Journal of Medical Entomology* **35**: 905-910.

deBoer, R., and Kuller, K. (1997) Mattresses as a winter refuge for house-dust mite populations. *Allergy* **52**: 299-305.

Dennis, E.A., Cao, J., Hsu, Y.-H., Magrioti, V., and Kokotos, G. (2011) Phospholipase A2 enzymes: physical structure, biological function, disease implication, chemical inhibition, and therapeutic intervention. *Chemical reviews* **111**: 6130-6185.

Douwes, J., Doekes, G., Montijn, R., Heederik, D., and Brunekreef, B. (1997) An immunoassay for the measurement of (1-->3)-beta-D-glucans in the indoor environment. *Mediators of inflammation* **6**: 257-262.

Du, D., Zhang, Q., Cheng, T., Pan, H., Yang, W., and Sun, L. (2013) Genome-wide identification and analysis of late embryogenesis abundant (LEA) genes in Prunus mume. *Molecular Biology Reports* **40**: 1937-1946.

ElRamlawy, K.G., Fujimura, T., Baba, K., Kim, J.W., Kawamoto, C., Isobe, T., Abe, T., Hodge-Hanson, K., Downs, D.M., Refaat, I.H., Beshr Al-Azhary, D., Aki, T., Asaoku, Y., Hayashi, T., Katsutani, T., Tsuboi, S., Ono, K., and Kawamoto, S. (2016) Der f 34, a Novel Major House Dust Mite Allergen Belonging to a Highly Conserved Rid/YjgF/YER057c/UK114 Family of Imine Deaminases. *J Biol Chem* **291**: 21607-21615.

EMA, (2008) Guideline on the Clinical Development of Products for Specific Immunotherapy for the Treatment of Allergic Diseases. In: CHMP/EWP/18504/2006. E.M. Agency (ed). European Medicines Agency, pp.

Eraso, E., Guisantes, J.A., Martinez, J., Saenz-de-Santamaria, M., Martinez, A., Palacios, R., and Cisterna, R. (1997) Kinetics of allergen expression in cultures of house dust mites, Dermatophagoides pteronyssinus and D. farinae (Acari: Pyroglyphidae). *J Med Entomol* **34**: 684-689.

Erban, T., Erbanova, M., Nesvorna, M., and Hubert, J. (2009) The importance of starch and sucrose digestion in nutritive biology of synanthropic acaridid mites: alpha-amylases and alpha-glucosidases are suitable targets for inhibitor-based strategies of mite control. *Arch Insect Biochem Physiol* **71**: 139-158.

Erban, T., Harant, K., and Hubert, J. (2017) 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**: 11-19.

Erban, T., and Hubert, J. (2008) Digestive function of lysozyme in synanthropic acaridid mites enables utilization of bacteria as a food source. *Exp Appl Acarol* **44**: 199-212.

Erban, T., and Hubert, J., (2012) Digestive Physiology of Synanthrophic Mites (Acari: Acaridida). In.: SOAJ Entomological Studies, pp. 1-32.

Evans, G.W., and McCoy, J.M. (1998) WHEN BUILDINGS DON'T WORK: THE ROLE OF ARCHITECTURE IN HUMAN HEALTH. *Journal of Environmental Psychology* **18**: 85-94.

FAO/WHO, (2001) Joint FAO/WHO Expert Consultation on Allergenicity of Foods Derived from Biotechnology. In. W.H.O. Food and Agriculture Organization of the United Nations (ed). pp.

Ferguson, P., and Broide, D.H. (1995) Environmental and bronchoalveolar lavage Dermatophagoides pteronyssinus antigen levels in atopic asthmatics. *Am J Respir Crit Care Med* **151**: 71-74.

Fesel, P.H., and Zuccaro, A. (2016) β-glucan: Crucial component of the fungal cell wall and elusive MAMP in plants. *Fungal Genetics and Biology* **90**: 53-60.

Fonslow, B.R., Stein, B.D., Webb, K.J., Xu, T., Choi, J., Park, S.K., and Yates, J.R. (2013) Digestion and depletion of abundant proteins improves proteomic coverage. *Nature methods* **10**: 54-56.

Ghazy, N.A., and Suzuki, T. (2014) Desiccation tolerance in diapausing spider mites Tetranychus urticae and T. kanzawai (Acari: Tetranychidae). *Experimental and Applied Acarology* **63**: 49-55.

Grantham, N.S., Reich, B.J., Pacifici, K., Laber, E.B., Menninger, H.L., Henley, J.B., Barberán, A., Leff, J.W., Fierer, N., and Dunn, R.R. (2015) Fungi identify the geographic origin of dust samples. *PloS one* **10**: e0122605-e0122605.

Gregory, L.G., and Lloyd, C.M. (2011) Orchestrating house dust mite-associated allergy in the lung. *Trends Immunol* **32**: 402-411.

Gøtzsche, P., and Johansen, H., (2008) House dust mite control measures for asthma. In: Cochrane Database of Systematic Reviews. The Cochrane Library: Cochrane Collaboration, pp.

Hales, B.J., Martin, A.C., Pearce, L.J., Laing, I.A., Hayden, C.M., Goldblatt, J., Le Souëf, P.N., and Thomas, W.R. (2006) IgE and IgG anti–house dust mite specificities in allergic disease. *Journal of Allergy and Clinical Immunology* **118**: 361-367.

Hand, S.C., Menze, M.A., Toner, M., Boswell, L., and Moore, D. (2011) LEA Proteins During Water Stress: Not Just for Plants Anymore. *Annual Review of Physiology* **73**: 115-134.

Hart, B.J., Crowther, D., Wilkinson, T., Biddulph, P., Ucci, M., Pretlove, S., Ridley, I., and Oreszczyn, T. (2007) Reproduction and development of laboratory and wild house dust mites (Acari : Pyroglyphidae) and their relationship to the natural dust ecosystem. *Journal of Medical Entomology* **44**: 568-574.

Hart, B.J., and Whitehead, L. (1990) Ecology of house dust mites in Oxfordshire. *Clin Exp Allergy* **20**: 203-209.

Hauser, M., Roulias, A., Ferreira, F., and Egger, M. (2010) Panallergens and their impact on the allergic patient. *Allergy, Asthma, and Clinical Immunology : Official Journal of the Canadian Society of Allergy and Clinical Immunology* **6**: 1-1.

Hay, D.B., Hart, B.J., and Douglas, A.E. (1993) EFFECTS OF THE FUNGUS ASPERGILLUS-PENICILLIOIDES ON THE HOUSE-DUST MITE DERMATOPHAGOIDES-PTERONYSSINUS - AN EXPERIMENTAL REEVALUATION. *Medical and Veterinary Entomology* **7**: 271-274.

Hay, D.B., Hart, B.J., Pearce, R.B., Kozakiewicz, Z., and Douglas, A.E. (1992) How relevant are house dust mite-fungal interactions in laboratory culture to the natural dust system? *Exp Appl Acarol* **16**: 37-47.

Hogrebe, A., von Stechow, L., Bekker-Jensen, D.B., Weinert, B.T., Kelstrup, C.D., and Olsen, J.V. (2018) Benchmarking common quantification strategies for large-scale phosphoproteomics. *Nature Communications* **9**: 1045.

Holgate, S.T. (2013) Mechanisms of asthma and implications for its prevention and treatment: a personal journey. *Allergy, asthma & immunology research* **5**: 343-347.

Hubert, J., Kopecky, J., Sagova-Mareckova, M., Nesvorna, M., Zurek, L., and Erban, T. (2016) Assessment of Bacterial Communities in Thirteen Species of Laboratory-Cultured Domestic Mites (Acari: Acaridida). *J Econ Entomol* **109**: 1887-1896.

Huby, R.D.J., Dearman, R.J., and Kimber, I. (2000) Why Are Some Proteins Allergens? *Toxicological Sciences* **55**: 235-246.

Hunault, G., and Jaspard, E. (2010) LEAPdb: a database for the late embryogenesis abundant proteins. *BMC Genomics* **11**: 221-221.

Hyndman, S.J., Vickers, L.M., Htut, T., Maunder, J.W., Peock, A., and Higenbottam, T.W. (2000) A randomized trial of dehumidification in the control of house dust mite. *Clin Exp Allergy* **30**: 1172-1180.

Ichikawa, S., Takai, T., Yashiki, T., Takahashi, S., Okumura, K., Ogawa, H., Kohda, D., and Hatanaka, H. (2009) Lipopolysaccharide binding of the mite allergen Der f 2. *Genes Cells* **14**: 1055-1065.

Jacquet, A. (2013) Innate immune responses in house dust mite allergy. *ISRN Allergy* **2013**: 735031.

Jarrett, R., Salio, M., Lloyd-Lavery, A., Subramaniam, S., Bourgeois, E., Archer, C., Cheung, K.L., Hardman, C., Chandler, D., Salimi, M., Gutowska-Owsiak, D., de la Serna, J.B., Fallon, P.G., Jolin, H., McKenzie, A., Dziembowski, A., Podobas, E.I., Bal, W., Johnson, D., Moody, D.B., Cerundolo, V., and Ogg, G. (2016) Filaggrin inhibits generation of CD1a neolipid antigens by house dust mite-derived phospholipase. *Science translational medicine* **8**: 325ra318-325ra318.

Jiang, C., Fan, X., Li, M., Xing, P., Liu, X., Wu, Y., Zhang, M., Yang, P., and Liu, Z. (2015) Characterization of Der f 29, a new allergen from dermatophagoides farinae. *American Journal of Translational Research* **7**: 1303-1313.

Kaarakainen, P., Rintala, H., Vepsalainen, A., Hyvarinen, A., Nevalainen, A., and Meklin, T. (2009) Microbial content of house dust samples determined with qPCR. *Science of the Total Environment* **407**: 4673-4680.

Karvonen, A.M., Hyvärinen, A., Rintala, H., Korppi, M., Täubel, M., Doekes, G., Gehring, U., Renz, H., Pfefferle, P.I., Genuneit, J., Keski-Nisula, L., Remes, S., Lampi, J., von Mutius, E., and Pekkanen, J. (2014) Quantity and diversity of environmental microbial exposure and development of asthma: a birth cohort study. *Allergy* **69**: 1092-1101.

Kidon, M.I., Chiang, W.C., Liew, W.K., Ong, T.C., Tiong, Y.S., Wong, K.N., Angus, A.C., Ong, S.T., Gao, Y.F., Reginald, K., Bi, X.Z., Shang, H.S., and Chew, F.T. (2011) Mite component-specific IgE repertoire and phenotypes of allergic disease in childhood: the tropical perspective. *Pediatr Allergy Immunol* **22**: 202-210.

Kikawada, T., Nakahara, Y., Kanamori, Y., Iwata, K.I., Watanabe, M., McGee, B., Tunnacliffe, A., and Okuda, T. (2006) Dehydration-induced expression of LEA

proteins in an anhydrobiotic chironomid. *Biochemical and Biophysical Research Communications* **348**: 56-61.

Klimov, P.B., Bochkov, A.V., and O'Connor, B.M. (2015) Phylogenetic position of the house dust mite subfamily Guatemalichinae (Acariformes: Pyroglyphidae) based on integrated molecular and morphological analyses and different measures of support. *Cladistics* **32**: 261-275.

Klimov, P.B., and O'Connor, B. (2013) Is Permanent Parasitism Reversible?-Critical Evidence from Early Evolution of House Dust Mites. *Systematic Biology* **62**: 411-423.

Klont, F., Bras, L., Wolters, J.C., Ongay, S., Bischoff, R., Halmos, G.B., and Horvatovich, P. (2018) Assessment of Sample Preparation Bias in Mass Spectrometry-Based Proteomics. *Anal Chem* **90**: 5405-5413.

Kong, H.H. (2011) Skin microbiome: genomics-based insights into the diversity and role of skin microbes. *Trends in molecular medicine* **17**: 320-328.

Koren, L.G.H., and Eckhardt, M.D., (1995) Suitability of Furnishings in use for Mite Population Development. In. Eindhoven University: Eindhoven: Technische Universiteit Eindhoven, pp.

Korpi, A., Pasanen, A.L., Pasanen, P., and Kalliokoski, P. (1997) Microbial growth and metabolism in house dust. *International Biodeterioration & Biodegradation* **40**: 19-27.

Liu, X.Y., Yang, K.Y., Wang, M.Q., Kwok, J.S., Zeng, X., Yang, Z., Xiao, X.J., Lau, C.P., Li, Y., Huang, Z.M., Ba, J.G., Yim, A.K., Ouyang, C.Y., Ngai, S.M., Chan, T.F., Leung, E.L., Liu, L., Liu, Z.G., and Tsui, S.K. (2018) High-quality assembly of Dermatophagoides pteronyssinus genome and transcriptome reveals a wide range of novel allergens. *J Allergy Clin Immunol* **141**: 2268-2271.e2268.

Martinez, J., Eraso, E., Palacios, R., and Guisantes, J.A. (1999) Enzymatic analyses of house dust mite extracts from Dermatophagoides pteronyssinus and Dermatophagoides farinae (Acari: Pyroglyphidae) during different phases of culture growth. *J Med Entomol* **36**: 370-375.

McKenna, O.E., Asam, C., Araujo, G.R., Roulias, A., Goulart, L.R., and Ferreira, F. (2016) How relevant is panallergen sensitization in the development of allergies? *Pediatric Allergy and Immunology* **27**: 560-568.

Merzendorfer, H., and Zimoch, L. (2003) Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *Journal of Experimental Biology* **206**: 4393.

Miller, J.D., and Young, J.C. (1997) The use of ergosterol to measure exposure to fungal propagules in indoor air. *Am Ind Hyg Assoc J* **58**: 39-43.

Mills, S., Wickens, K., Crane, J., and Fitzharris, P., (2002) House Dust Mite Allergen Levels in Individual Bedding Components in New Zealand. In.: New Zealand Medical Journal, pp. 151-153.

Mitchell, W.F., Wharton, G.W., Larson, D.G., and Modic, R. (1969) House dust, mites, and insects. *Ann Allergy* **27**: 93-99.

Molva, V., Bostlova, M., Nesvorna, M., and Hubert, J. (2018) Do the microorganisms from laboratory culture spent growth medium affect house dust mite fitness and microbiome composition? *Insect Sci*.

Molva, V., Nesvorna, M., and Hubert, J. (2019) Feeding Interactions Between Microorganisms and the House Dust Mites Dermatophagoides pteronyssinus and Dermatophagoides farinae (Astigmata: Pyroglyphidae). *Journal of Medical Entomology*.

Murray, A.B., and Ferguson, A.C. (1983) Dust-free bedrooms in the treatment of asthmatic children with house dust or house dust mite allergy: a controlled trial. *Pediatrics* **71**: 418-422.

Naegele, A., Reboux, G., Scherer, E., Roussel, S., and Millon, L. (2013) Fungal food choices of Dermatophagoides farinae affect indoor fungi selection and dispersal. *International Journal of Environmental Health Research* **23**: 91-95.

Nagakura, T., Yasueda, H., Obata, T., Kanmuri, M., Masaki, T., Ihara, N., and Maekawa, K. (1996) Major Dermatophagoides mite allergen, Der 1, in soft toys. *Clinical and Experimental Allergy* **26**: 585-589.

Nankervis, H., Pynn, E.V., Boyle, R.J., Rushton, L., Williams, H.C., Hewson, D.M., and Platts-Mills, T. (2015) House dust mite reduction and avoidance measures for treating eczema. *Cochrane Database Syst Rev* **1**: Cd008426.

NHQRS, (2017) National Healthcare

Quality Reporting System. In. D.o. Health (ed). Department of Health, pp.

Nielsen, M.G., Gadagkar, S.R., and Gutzwiller, L. (2010) Tubulin evolution in insects: gene duplication and subfunctionalization provide specialized isoforms in a functionally constrained gene family. *BMC Evolutionary Biology* **10**: 113.

Niven, R.M., Fletcher, A.M., Pickering, A.C., Custovic, A., Sivour, J.B., Preece, A.R., Oldham, L.A., and Francis, H.C. (1999) Attempting to control mite allergens with mechanical ventilation and dehumidification in British houses. *Journal of Allergy and Clinical Immunology* **103**: 756-762.

O'Connor, B.M., (1979) EVOLUTIONARY ORIGINS OF ASTIGMATID MITES INHABITING STORED PRODUCTS. In: Recent Advances in Acarology. J.G. Rodriguez (ed). Academic Press, pp. 273-278.

O'Neil, S.E., Heinrich, T.K., Hales, B.J., Hazell, L.A., Holt, D.C., Fischer, K., and Thomas, W.R. (2006) The chitinase allergens Der p 15 and Der p 18 from Dermatophagoides pteronyssinus. *Clin Exp Allergy* **36**: 831-839.

Orfanoudaki, G., Markaki, M., Chatzi, K., Tsamardinos, I., and Economou, A. (2017) MatureP: prediction of secreted proteins with exclusive information from their mature regions. *Scientific Reports* **7**: 3263.

Pauchet, Y., Freitak, D., Heidel-Fischer, H.M., Heckel, D.G., and Vogel, H. (2009) Immunity or digestion: glucanase activity in a glucan-binding protein family from Lepidoptera. *J Biol Chem* **284**: 2214-2224.

Platts-Mills, T.A., Heymann, P.W., Longbottom, J.L., and Wilkins, S.R. (1986) Airborne allergens associated with asthma: particle sizes carrying dust mite and rat allergens measured with a cascade impactor. *J Allergy Clin Immunol* **77**: 850-857.

Pomés, A., Davies, J.M., Gadermaier, G., Hilger, C., Holzhauser, T., Lidholm, J., Lopata, A.L., Mueller, G.A., Nandy, A., Radauer, C., Chan, S.K., Jappe, U., Kleine-Tebbe, J., Thomas, W.R., Chapman, M.D., van Hage, M., van Ree, R., Vieths, S., Raulf, M., and Goodman, R.E. (2018) WHO/IUIS Allergen Nomenclature: Providing a common language. *Molecular Immunology* **100**: 3-13.

Prasad, T.S., Mohanty, A.K., Kumar, M., Sreenivasamurthy, S.K., Dey, G., Nirujogi, R.S., Pinto, S.M., Madugundu, A.K., Patil, A.H., Advani, J., Manda, S.S., Gupta, M.K., Dwivedi, S.B., Kelkar, D.S., Hall, B., Jiang, X., Peery, A., Rajagopalan, P., Yelamanchi, S.D., Solanki, H.S., Raja, R., Sathe, G.J., Chavan, S., Verma, R., Patel, K.M., Jain, A.P., Syed, N., Datta, K.K., Khan, A.A., Dammalli, M., Jayaram, S., Radhakrishnan, A., Mitchell, C.J., Na, C.H., Kumar, N., Sinnis, P., Sharakhov, I.V., Wang, C., Gowda, H., Tu, Z., Kumar, A., and Pandey, A. (2017) Integrating transcriptomic and proteomic data for accurate assembly and annotation of genomes. *Genome Res* **27**: 133-144.

Radauer, C., (2017) AllFam Database of allergen families. In.: Medical University of Vienna , Austria, pp.

Radauer, C., Bublin, M., Wagner, S., Mari, A., and Breiteneder, H. (2008) Allergens are distributed into few protein families and possess a restricted number of biochemical functions. *Journal of Allergy and Clinical Immunology* **121**: 847-852.e847.

Rains, N., Siebers, R., Crane, J., and Fitzharris, P. (1999) House dust mite allergen (Der p 1) accumulation on new synthetic and feather pillows. *Clinical and Experimental Allergy* **29**: 182-185.

Randall, T.A., Mullikin, J.C., and Mueller, G.A. (2018) The Draft Genome Assembly of Dermatophagoides pteronyssinus Supports Identification of Novel Allergen Isoforms in Dermatophagoides Species. *Int Arch Allergy Immunol* **175**: 136-146.

Reithofer, M., and Jahn-Schmid, B. (2017) Allergens with Protease Activity from House Dust Mites. *International journal of molecular sciences* **18**: 1368.

Resch, Y., Blatt, K., Malkus, U., Fercher, C., Swoboda, I., Focke-Tejkl, M., Chen, K.-W., Seiberler, S., Mittermann, I., Lupinek, C., Rodriguez-Dominguez, A., Zieglmayer, P., Zieglmayer, R., Keller, W., Krzyzanek, V., Valent, P., Valenta, R., and Vrtala, S. (2016) Molecular, Structural and Immunological Characterization of Der p 18, a Chitinase-Like House Dust Mite Allergen. *PloS one* **11**: e0160641-e0160641.

Rider, S.D., Jr., Morgan, M.S., and Arlian, L.G. (2017) Allergen homologs in the Euroglyphus maynei draft genome. *PLOS ONE* **12**: e0183535.

Rintala, H., Pitkaranta, M., and Taubel, M. (2012) Microbial communities associated with house dust. *Adv Appl Microbiol* **78**: 75-120.

Rolinck-Werninghaus, C., Keil, T., Kopp, M., Zielen, S., Schauer, U., von Berg, A., Wahn, U., and Hamelmann, E. (2008) Specific IgE serum concentration is associated with symptom severity in children with seasonal allergic rhinitis. *Allergy* **63**: 1339-1344.

Rowland-Seymour, A., (2010) Rhinitis. In: Decision Making in Medicine (Third Edition). S.B. Mushlin & H.L. Greene (eds). Philadelphia: Mosby, pp. 32-35.

Ruggles, K.V., Krug, K., Wang, X., Clauser, K.R., Wang, J., Payne, S.H., Fenyö, D., Zhang, B., and Mani, D.R. (2017) Methods, Tools and Current Perspectives in Proteogenomics. *Molecular & cellular proteomics : MCP* **16**: 959-981.

Sanchez-Borges, M., Fernandez-Caldas, E., Thomas, W.R., Chapman, M.D., Lee, B.W., Caraballo, L., Acevedo, N., Chew, F.T., Ansotegui, I.J., Behrooz, L., Phipatanakul, W., Gerth van Wijk, R., Pascal, D., Rosario, N., Ebisawa, M., Geller, M., Quirce, S., Vrtala, S., Valenta, R., Ollert, M., Canonica, G.W., Calderon, M.A., Barnes, C.S., Custovic, A., Benjaponpitak, S., and Capriles-Hulett, A. (2017) International consensus (ICON) on: clinical consequences of mite hypersensitivity, a global problem. *World Allergy Organ J* **10**: 14.

Satitsuksanoa, P., Kennedy, M., Gilis, D., Le Mignon, M., Suratannon, N., Soh, W.T., Wongpiyabovorn, J., Chatchatee, P., Vangveravong, M., Rerkpattanapipat, T., Sangasapaviliya, A., Piboonpocanun, S., Nony, E., Ruxrungtham, K., and Jacquet, A. (2016) The minor house dust mite allergen Der p 13 is a fatty acid-binding protein and an activator of a TLR2-mediated innate immune response. *Allergy* **71**: 1425-1434.

Schei, M.A., Hessen, J.O., and Lund, E. (2002) House-dust mites and mattresses. *Allergy* **57**: 538-542.

Schiavone, M., Vax, A., Martin-Yken, H., François, J.M., Dague, E., and Formosa, C. (2014) A combined chemical and enzymatic method to determine quantitatively the polysaccharide components in the cell wall of yeasts. *FEMS Yeast Research* **14**: 933-947.

Sidenius, K.E., Hallas, T.E., Poulsen, L.K., and Mosbech, H. (2001) Allergen cross-reactivity between house-dust mites and other invertebrates. *Allergy* **56**: 723-733.

Sinclair, B.J., Ferguson, L.V., Salehipour-shirazi, G., and MacMillan, H.A. (2013) Cross-tolerance and Cross-talk in the Cold: Relating Low Temperatures to Desiccation and Immune Stress in Insects. *Integrative and Comparative Biology* **53**: 545-556.

Solomon, A., Salomon, R., Paperna, I., and Glazer, I. (2000) Desiccation stress of entomopathogenic nematodes induces the accumulation of a novel heat-stable protein. *Parasitology* **121**: 409-416.

Sonobe, H., Obinata, T., Minokawa, T., Haruta, T., Kawamura, Y., Wakatsuki, S., and Sato, N. (2016) Characterization of paramyosin and thin filaments in the smooth muscle of acorn worm, a member of hemichordates. *The Journal of Biochemistry* **160**: 369-379.

Sousa, A.C., Almeida, J.R., Pereira, C.C., Ramiro Pastorinho, M., Pereira, A.M., Nogueira, A.J., Taborda-Barata, L., Teixeira, J.P., Correia, A.C., and Alves, A. (2014) Characterization of fungal communities in house dust samples collected from central Portugal-a preliminary survey. *J Toxicol Environ Health A* **77**: 972-982.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006) AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic acids research* **34**: W435-W439.

Steppuhn, H., Langen, U., Scheidt-Nave, C., and Keil, T. (2013) Major comorbid conditions in asthma and association with asthma-related hospitalizations and emergency department admissions in adults: results from the German national health telephone interview survey (GEDA) 2010. *BMC Pulmonary Medicine* **13**: 46.

Taubel, M., Rintala, H., Pitkaranta, M., Paulin, L., Laitinen, S., Pekkanen, J., Hyvarinen, A., and Nevalainen, A. (2009) The occupant as a source of house dust bacteria. *Journal of Allergy and Clinical Immunology* **124**: 834-840.

Teplitsky, V., Mumcuoglu, K.Y., Babai, I., Dalal, I., Cohen, R., and Tanay, A. (2008) House dust mites on skin, clothes, and bedding of atopic dermatitis patients. *International Journal of Dermatology* **47**: 790-795.

Thomas, W.R. (2010) Geography of house dust mite allergens. *Asian Pac J Allergy Immunol* **28**: 211-224.

Thomas, W.R. (2015) Hierarchy and molecular properties of house dust mite allergens. *Allergology International* **64**: 304-311.

Thomas, W.R., (2018) Blueprint for the House Dust Mite. In: Int Arch Allergy Immunol. Switzerland, pp. 121-125.

Thomas, W.R., Hales, B.J., and Smith, W.-A. (2005) Structural biology of allergens. *Current Allergy and Asthma Reports* **5**: 388-393.

Thomas, W.R., Hales, B.J., and Smith, W.-A. (2010) House dust mite allergens in asthma and allergy. *Trends in Molecular Medicine* **16**: 321-328.

Tovey, E.R., Chapman, M.D., and Platts-Mills, T.A. (1981) Mite faeces are a major source of house dust allergens. *Nature* **289**: 592-593.

Tovey, E.R., Mahmic, A., and McDonald, L.G. (1995) Clothing--an important source of mite allergen exposure. *J Allergy Clin Immunol* **96**: 999-1001.

Tsiatsiani, L., and Heck, A.J. (2015) Proteomics beyond trypsin. *Febs j* **282**: 2612-2626.

Tutar, Y. (2012) Pseudogenes. *Comparative and functional genomics* **2012**: 424526-424526.

van Bronswijk, J.E.M.H. (1973) Dermatophagoides Pteronyssinus (Trouessart, 1897) in Mattress and Floor Dust in a Temperate Climate (Acari: Pyroglyphidae). *Journal of Medical Entomology* **10**: 63-70.

van den Bemt, L., de Vries, M.P., van Knapen, L., Jansenw, M., Goossens, M., Muris, J.W.M., and van Schayck, C.P. (2006) Influence of mattress characteristics on house dust mite allergen concentration. *Clinical and Experimental Allergy* **36**: 233-237.

VanBronswijk, J.E.M.H., and Sinha, R.N. (1973) Role of Fungi in the Survival of Dermatophagoides (Acarina: Pyroglyphidae) in House-Dust Environment1. *Environmental Entomology* **2**: 142-145.

Vidal, C., Chomón, B., Pérez-Carral, C., and González-Quintela, A. (1997) Sensitization to Lepidoglyphus destructor, Tyrophagus putrescentiae, and Acarus siro in patients allergic to house dust mites (Dermatophagoides spp.). *Journal of Allergy and Clinical Immunology* **100**: 716-718.

Vidal-Quist, J.C., Ortego, F., Rombauts, S., CastaÑEra, P., and HernÁNdez-Crespo, P. (2017) Dietary shifts have consequences for the repertoire of allergens produced by the European house dust mite. *Medical and Veterinary Entomology* **31**: 272-280.

Waldron, R., McGowan, J., Gordon, N., McCarthy, C., Mitchell, E.B., Doyle, S., and Fitzpatrick, D.A. (2017) Draft Genome Sequence of Dermatophagoides pteronyssinus, the European House Dust Mite. *Genome Announcements* **5**: e00789-00717.

Wan, H., Winton, H.L., Soeller, C., Tovey, E.R., Gruenert, D.C., Thompson, P.J., Stewart, G.A., Taylor, G.W., Garrod, D.R., Cannell, M.B., and Robinson, C. (1999) Der p 1 facilitates transepithelial allergen delivery by disruption of tight junctions. *J Clin Invest* **104**: 123-133.

Wang, J.-Y. (2013) The innate immune response in house dust mite-induced allergic inflammation. *Allergy, asthma & immunology research* **5**: 68-74.

Weghofer, M., Dall'Antonia, Y., Grote, M., Stocklinger, A., Kneidinger, M., Balic, N., Krauth, M.T., Fernandez-Caldas, E., Thomas, W.R., van Hage, M., Vieths, S., Spitzauer, S., Horak, F., Svergun, D.I., Konarev, P.V., Valent, P., Thalhamer, J., Keller,

W., Valenta, R., and Vrtala, S. (2008) Characterization of Der p 21, a new important allergen derived from the gut of house dust mites. *Allergy* **63**: 758-767.

Weghofer, M., Grote, M., Resch, Y., Casset, A., Kneidinger, M., Kopec, J., Thomas, W.R., Fernandez-Caldas, E., Kabesch, M., Ferrara, R., Mari, A., Purohit, A., Pauli, G., Horak, F., Keller, W., Valent, P., Valenta, R., and Vrtala, S. (2013) 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**: 3059-3067.

Whiten, S.R., Eggleston, H., and Adelman, Z.N. (2018) Ironing out the Details: Exploring the Role of Iron and Heme in Blood-Sucking Arthropods. *Frontiers in physiology* **8**: 1134-1134.

WHO/IUIS. In., pp.

Wickens, K., Martin, I., Pearce, N., Fitzharris, P., Kent, R., Holbrook, N., Siebers, R., Smith, S., Trethowen, H., Lewis, S., Town, I., and Crane, J. (1997) House dust mite allergen levels in public places in New Zealand. *J Allergy Clin Immunol* **99**: 587-593.

Wissler, L., Gadau, J., Simola, D.F., Helmkampf, M., and Bornberg-Bauer, E. (2013) Mechanisms and dynamics of orphan gene emergence in insect genomes. *Genome biology and evolution* **5**: 439-455.

Wong, S.F., Chong, A.L., Mak, J.W., Tan, J., Ling, S.J., and Ho, T.M. (2011) Molecular identification of house dust mites and storage mites. *Experimental and Applied Acarology* **55**: 123-133.

Worland, M.R., and Block, W. (2003) Desiccation stress at sub-zero temperatures in polar terrestrial arthropods. *Journal of Insect Physiology* **49**: 193-203.

Wraith, D.G., Cunnington, A.M., and Seymour, W.M. (1979) The role and allergenic importance of storage mites in house dust and other environments. *Clinical & Experimental Allergy* **9**: 545-561.

Wu, G.D., and Lewis, J.D. (2013) Analysis of the human gut microbiome and association with disease. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association* **11**: 774-777.

www.allergenome.org, allergenome.org. In., pp.

www.Allergenonline.org, Allergenonline. In., pp. Allergen Database.

Yusuf, N., Nasti, T.H., Huang, C.-M., Huber, B.S., Jaleel, T., Lin, H.-Y., Xu, H., and Elmets, C.A. (2009) Heat shock proteins HSP27 and HSP70 are present in the skin and are important mediators of allergic contact hypersensitivity. *Journal of immunology (Baltimore, Md. : 1950)* **182**: 675-683.

Zock, J.P., Heinrich, J., Jarvis, D., Verlato, G., Norback, D., Plana, E., Sunyer, J., Chinn, S., Olivieri, M., Soon, A., Villani, S., Ponzio, M., Dahlman-Hoglund, A., Svanes, C., and Luczynska, C. (2006) Distribution and determinants of house dust mite allergens in Europe: the European Community Respiratory Health Survey II. *J Allergy Clin Immunol* **118**: 682-690.

Appendix 3.