Comparative Genomics Reveals Insights into the Divergent Evolution of Astigmatic Mites and Household Pest Adaptations

Qing Xiong
School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong
Angel Tsz-Yau Wan
School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong
Xiao-Yu Liu
Shenzhen University
Cathy Sin-Hang Fung
School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong
Xiaojun Xiao
Guangzhou Institute of Respiratory Health
Nat Malainual
Department of Parasitology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand
Jinpao Hou
School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong
Lingyi Wang
School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong
Mingqiang Wang
School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong
Kevin Yang
CUHK https://orcid.org/0000-0002-4200-745X
Yubao Cui
Department of Clinical Laboratory, Wuxi People's Hospital Affiliated to Nanjing Medical University, Wuxi, China
Elaine Leung
Macau Institute for Applied Research in Medicine and Health, State Key Laboratory of Quality Research in Chinese Medicine https://orcid.org/0000-0002-3705-8084
Wenyan Nong
The Chinese University of Hong Kong https://orcid.org/0000-0002-3277-716X
Soo-Kyung Shin
School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong
Shannon Au

Kyoung Yong Jeong
Department of Internal Medicine, Institute of Allergy, Yonsei University College of Medicine, Seoul, Korea https://orcid.org/0000-0001-9887-1426
Fook Tim Chew
National University of Singapore https://orcid.org/0000-0003-1337-5146
Jerome Hui
Chinese University of Hong Kong
Ting Fan Leung
The Chinese University of Hong Kong https://orcid.org/0000-0002-6469-1926
Anchalee Tungtrongchitr
Department of Parasitology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand
Nanshan Zhong
The First Affiliated Hospital of Guangzhou Medical University
Zhigang Liu
Shenzhen Key Laboratory of Allergy and Immunology, School of Medicine, Shenzhen University, China
Stephen Tsui (kwtsui@cuhk.edu.hk)
The Chinese University of Hong Kong

Article

Keywords: Astigmatic mites, Comparative genomics, Horizontal gene transfer, Household pest adaptations, Divergent evolution, Tandem gene duplication

Posted Date: January 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1237176/v1

Version of Record: A version of this preprint was published at Molecular Biology and Evolution on May 10th, 2022. See the published version at https://doi.org/10.1093/molbev/msac097.

Comparative Genomics Reveals Insights into the Divergent Evolution of Astigmatic Mites and Household Pest Adaptations

Authors: Qing Xiong[1,2]†, Angel Tsz-Yau Wan[1,2]†, Xiaoyu Liu[3], Cathy Sin-Hang Fung[1], Xiaojun Xiao[3], Nat Malainual[4], Jinpao Hou[1,5], Lingyi Wang[1], Mingqiang Wang[1,2], Kevin Yi Yang[1,2], Yubao Cui[6], Elaine Lai-Han Leung[7], Wenyan Nong[8], Soo-Kyung Shin[1], Shannon Wing-Ngor Au[8], Kyoung Yong Jeong[9], Fook-Tim Chew[10], Jerome Ho-Lam Hui[8], Ting-Fan Leung[11], Anchalee Tungtrongchitr[4], Nanshan Zhong[12], Zhigang Liu[3], Stephen Kwok-Wing Tsui[1,2,5] *

Affiliations:

1 School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong.
2 Hong Kong Bioinformatics Centre, The Chinese University of Hong Kong, Hong Kong.
3 Shenzhen Key Laboratory of Allergy and Immunology, School of Medicine, Shenzhen University, China.
4 Department of Parasitology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand.
5 Centre for Microbial Genomics and Proteomics, The Chinese University of Hong Kong, Hong Kong.
6 Department of Clinical Laboratory, Wuxi People's Hospital Affiliated to Nanjing Medical University, Wuxi, China.
7 Macau Institute for Applied Research in Medicine and Health, Macau University of Science and Technology, Macau.
8 School of Life Sciences, The Chinese University of Hong Kong, Hong Kong.
9 Institute of Allergy, Department of Internal Medicine, College of Medicine, Yonsei University, Seoul, Korea.
10 Department of Biological Sciences, National University of Singapore, Singapore.
11 Department of Paediatrics, The Chinese University of Hong Kong, Hong Kong.
12 State Key Laboratory of Respiratory Disease, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China.

* Correspondence to Professor Stephen Kwok-Wing Tsui, School of Biomedical Sciences, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong. E-mail: kwtsui@cuhk.edu.hk; Tel: 852-3943 6381; Fax: 852-2603 5123.

† These authors contributed equally to this work.

Abstract

Highly diversified astigmatic mites comprise many medically important human household pests such as house dust mites causing roughly 1–2% of the allergic diseases globally; however, their evolutionary origin, diverse lifestyles including reversible parasitism and quick adaptation to rather new human household environments have not been illustrated at genomic level, which hamper the allergy prevention and our exploration of these household pests. Using six high-quality assembled and annotated genomes, this comparative genomics study not only refuted the monophyly of mites and ticks, but also thoroughly explored the divergence of Acariformes and the divergent evolution of astigmatic mites. In the monophyletic Acariformes, Prostigmata known as notorious plant pests first evolved, then rapidly evolving Astigmata diverged from soil oribatid mites. Within astigmatic mites, a wide range of gene families rapidly expanded via tandem gene duplications, including ionotropic glutamate receptors, triacylglycerol lipases, serine proteases and UDP glucuronosyltransferases (UGTs), which enriched their capacities of adapting to rapidly changing household environments. The gene diversification after tandem duplications provided plenty of genetic resources for their adaptations of sensing environmental signals, digestion, and detoxification. Whilst many gene decay events only occurred in the skin-burrowing parasitic mite *Sarcoptes scabiei*. Throughout the evolution of Acariformes, massive horizontal gene transfer events occurred in gene families such as UGTs and several important fungal cell wall lytic enzymes, which enable the detoxification and associated digestive functions and provide perfect drug targets for pest control. Our comparative study sheds light on the rapid divergent evolution of astigmatic mites from the divergence of Acariformes to their diversification and provides novel insights into the genetic adaptations and even control of human household pests.

Keywords

Astigmatic mites, Comparative genomics, Horizontal gene transfer, Household pest adaptations, Divergent evolution, Tandem gene duplication

Introduction

Astigmatic mites (suborder: Astigmata) are a group of rapidly evolving and successfully radiating small organisms including many medically important human household pests [1-5]. Most astigmatic mites are commensals or parasites on hosts ranging from insects to birds and mammals, in which many free-living species (especially dust mites) are major human allergy related mites and approximately 1–2% of the world population (65–130 million people) suffers from allergic diseases caused by these mites [6-10]. Astigmata was considered evolving from oribatid mites (suborder: Oribatida) which are the numerically dominant microarthropods in soil environments, but the evolutionary dynamics underlying their emergence is still largely unclear. Moreover, their diversification into various lifestyles including reversal parasitism, and how they adapt to recently established human household environments and become household pests remain to be thoroughly investigated at genomic level. In this study, the high-quality genomes of six representative astigmatic mites, comprising two house dust mites *Dermatophagoides* (*D.*) *farinae* and *D. pteronyssinus,* two parasitic mites *Psoroptes* (*P.*) *ovis* and *Sarcoptes* (*S.*) *scabiei*, and two canonical storage mites *Blomia* (*B.*) *tropicalis* and *Tyrophagus* (*T.*) *putrescentiae* (Fig. 1A) were constructed for comparative genomics analysis.

Our phylogenomic analysis refuted the monophyly of mites and ticks (subclass: Acari) [5,11-13] and confirmed astigmatic mites as a rapidly evolving lineage under the monophyletic Acariformes. Massive horizontal gene transfer (HGT) events enabled gene novelties throughout the divergence of Acariformes with the involved genes (HGT genes) being perfect drug targets [14-16]. After diverged from soil mites, astigmatic mites underwent rapid genome evolution and diversification through extensive gene family variations and frequent tandem gene duplications to adapt in different niches of household environments. These comprehensive findings should be useful for the future exploration of these astigmatic mites and the prevention of associated allergic diseases. Therefore, this comparative genomics study not only sheds light on the evolutionary origin and trajectory of allergic astigmatic mites, but also paves the way for understanding the adaptation fundamentals and even designing the control strategies of human household pests.

Results

Genome assembly and annotation

D. farinae and *D. pteronyssinus* are canonical house dust mites that cause human allergic diseases [17-20] (Fig. 1A). *P. ovis* and *S. scabiei* are both parasitic mites of mammals, in which *P. ovis* is an ectoparasitic and non-burrowing mite mainly infesting in domesticated sheep [21,22], while *S. scabiei* could burrow into the skin of hosts such as human and domestic dog [23,24] (Fig. 1A). *B. tropicalis* was previously regarded as a storage mite, but now also considered as a house dust mite mainly in tropical and subtropical areas [25] (Fig. 1A). *T. putrescentiae* is referred to as a mold mite or cheese mite and considered as a storage mite [26] (Fig. 1A). Besides of *D. farinae*, *D. pteronyssinus*, *B. tropicalis* and *T. putrescentiae* known as free-living allergic indoor mites, two parasitic mites *P. ovis* and *S. scabiei* have been also reported as human allergy related [27,28].

Highly contiguous reference-quality genomes of four astigmatic mites (Fig. 1A), *D. farinae*, *D. pteronyssinus*, *B. tropicalis* and *T. putrescentiae* were constructed using four next- and third-sequencing platforms (Table S1). The *P. ovis* genome deposited in NCBI database (BioProject accession: PRJNA521406) was downloaded and reannotated, whereas raw sequencing data for the *S. scabiei* (*var. canis*) genome were downloaded from the database (BioProject accession: PRJNA268368 and PRJNA304361), followed by reassembly and annotation. Completeness of these six genome assemblies ranged from 89.0% to 91.5% and the N50 lengths ranged from 253,843 (*S. scabiei*) bp to 8,981,490 bp (*D. farinae*) (Table 1). The genome sizes of the six astigmatic mites ranged from 56 Mb (*S. scabiei*) to 97 Mb (*T. putrescentiae*).

Consistent annotation pipeline was performed on six genomes of mites to avoid systematic biases. The completeness of the six proteomes ranged from 90.7% to 92.9%, showing the consistent high quality of annotations among the six astigmatic mite genomes (Table 1). The protein-coding gene numbers of the six astigmatic mites ranged from 14,688 (*P. ovis*) to 23,793 (*T. putrescentiae*) (Table 1). Notably, the canonical storage mite *T. putrescentiae* has around 55% more protein-coding genes than the other five mites. Compared with the other five mites, *T. putrescentiae* has a broader living environment which may be related to its larger genome size and more protein-coding genes.

Phylogenomic analysis

It is controversial whether mites and ticks (subclass: Acari) comprising of two major groups Acariformes and Parasitiformes are a monophyletic group [5,11-13]. A phylogenomic analysis was performed with these six genome assemblies and other 22 publicly available genomes (Table 4, Supplementary Text 1). In Fig. 1B, the phylogenetic tree of 26 mites and ticks, the pseudoscorpion *Cordylochernes scorpioides* and the fruit fly *Drosophila melanogaster* was constructed based on 13,133 conserved amino-acid residues in 47 overlapped single and complete BUSCO proteins and suggested that the phylogeny of Acariformes and Parasitiformes were interrupted by the pseudoscorpion which refuted the monophyly of mites and ticks. This phylogenetic tree also supported astigmatic mites evolved from Oribatida. Meanwhile, the mean substitution rate (MSR) of conserved amino-acid residues in the five lineages of mites and ticks were compared and the significantly higher MSR (0.5632, $p < 0.001$, Fig. 1B) of Astigmata confirmed astigmatic mites as a group of rapidly evolving species [5].

General genomic features including genome size were compared in Fig. S1. Except *T. putrescentiae* (97 Mb) and *Euroglyphus* (*E.*) *maynei* (in low completeness), the other astigmatic mites possess similar genome sizes of 60 ± 4 Mb, which are the smallest genome sizes among all sequenced genomes of mites and ticks (Fig. S1). In addition, repeat contents were compared (Fig. S2) and the genome sizes of five psoroptid mites (parvorder: Psoroptidia, *D. farinae*, *D. pteronyssinus*, *P. ovis*, *S. scabiei var. suis* and *var. canis*) were linearly correlated with total repeat contents or interspersed repeat contents (Fig. S3, R=0.99 and 0.98 respectively). The publicly available genome of *S. scabiei var. suis* has a higher continuity but a lower completeness than that of *S. scabiei var. canis* (Fig. S2), so the subsequent study still used the reassembled genome of *S. scabiei var. canis* for comparison.

Based on conserved amino-acid alignment, the ultrametric time tree was constructed and revealed that these astigmatic mites evolved from oribatid mites at around 418 million years ago (MYA) and started divergence at around 278 MYA (Fig. S4).

Gene family evolution

To establish a well resolved evolutionary analysis of gene family, six proteomes generated from our high-quality annotated genomes, eight publicly available proteomes of oribatid, prostigmatic and mesostigmatic mites (Table S5) and the proteome of *D. melanogaster* were assigned into orthogroups (or gene families) according to amino acid sequence similarities.

A gene gain/loss analysis was performed in Fig. 2A. We started our analysis from the monophyletic Acariformes, in which Prostigmata first evolved with four rapid expanding gene families including titin and sodium-dependent glucose transporter (Table S6), while only one uncharacterized protein family expanded in Oribatida and Astigmata (Table S7). Then oribatid mites further expanded 101 gene families, which covered diverse detoxification and digestion gene families (Table S8). As for the astigmatic mites, 10 gene families rapidly expanded and enriched gene ontology (GO) terms related to ABC transporter (ABC), zinc finger and ionotropic glutamate receptor (iGluR) (Table S9). Of glucosylceramidase, the contraction in astigmatic mites and even gene decay in *S. scabiei* could be the result of food source changes (Table S10 and Supplementary Text 2).

Within astigmatic mites, that higher numbers of expanding and contracting gene families were found in *T. putrescentiae* and *S. scabiei* respectively (Fig. 2B). GO enrichment analysis of critical orthogroups involved in the evolution of the four psoroptid mites (*D. farinae*, *D. pteronyssinus*, *P. ovis*, *S. scabiei*) and the two canonical storage mites (*B. tropicalis* and *T. putrescentiae*) suggested divergent evolution of serine protease, ionotropic glutamate receptor and various detoxification gene families including ABC transporter and cytochrome P450 (CYP) (Table S11–21, Supplementary Text 2). The gene loss of a phosphoenolpyruvate synthase (Table S15) in the psoroptid mites may be related to their ancestral parasitism resulted metabolism changes [1]. A wide range of rapid expanding gene families were identified in *T. putrescentiae*, including several glycoside hydrolases, serine protease, heat shock protein 70 (HSP70), and detoxification gene families including UDP glucuronosyltransferase (UGT), CYP and ABC (Table 14).

Species-specific orthogroups were identified in Venn diagram (Fig. 2B) and analyzed in gene ontology (GO) enrichment (Table S22–28). Notably, *T. putrescentiae* have much bigger number of species-specific orthogroups including many proteases, glycoside hydrolase, CYP, ionotropic glutamate receptor, suggesting that this storage mite underwent more sophisticated

adaptation during evolution. Besides, GO enrichment revealed species-specific serine proteases in *B. tropicalis*. Based on the analysis of gene family evolution, many gene families were further compared, especially the frequently enriched ionotropic glutamate receptor, digestive enzymes, and detoxification gene families.

Ionotropic glutamate receptors

The ionotropic glutamate receptor (iGluR) family is well known in the sensing to a wide range of environmental changes or signals such as temperature, moisture and taste in both vertebrates and invertebrates [29,30]. All iGluRs of six mites were collected for phylogenetic analysis (Fig. 3A), in which *P. ovis* and *S. scabiei* displayed significantly fewer iGluR genes than other four mites. The iGluRs in small clusters a and b (Fig. 3A) were identified as IR25a and IR93a for the phylogenetic analysis (Fig. 3B) and their highly conserved regions showed homology to those of IR25a and IR93a of *D. melanogaster* (Fig. S6), respectively. Interestingly, IR25a and IR93a were adjacent in all six astigmatic mites (Fig. 3B) but are located on two different chromosomes in the genome of *D. melanogaster* (Chromosome 2L and 3R, respectively).

Frequent tandem gene duplications were observed in the iGluRs (Fig. 3A), in which of a total of 285 iGluRs, 116 genes (40.7%) were tandemly arrayed, and 37 genes (13.0%) were proximally arrayed (Fig. 3A). Most tandem duplications occurred at the cluster Y, while gene clusters were evenly distributed among the six mites in the cluster X (Fig. 3A), indicating that cluster Y is a more recently diversified cluster of iGluRs than cluster X and played an important role in the diversification of astigmatic mites.

In the cluster 1, more iGluR genes were identified in the two canonical storage mites than in the two house dust mites, while only one iGluR gene in *P. ovis* and none in *S. scabiei* were identified. In the cluster 2, massive tandem duplications were observed, but only one iGluR gene was identified in *T. putrescentiae*. Since *T. putrescentiae* is the only storage mite and far from human or mammals unlike other five mites, we propose the cluster 2 iGluRs play functions in sensing signals related to human or mammals. In the subcluster 2-1, gene synteny alignment identified tandemly arrayed iGluRs in *D. farinae*, *D. pteronyssinus*, *S. scabiei* and *B. tropicalis*, but only one iGluR each in *P. ovis* and *T. putrescentiae* (Fig. 3C). Notably, nine divergent and tandemly arrayed iGluR genes were identified in *B. tropicalis*, while only one was identified in *T. putrescentiae* at the same gene locus (Fig. 3C).

Among all identified iGluRs in this study, IR25a and IR93a (Fig. 3B) are particularly important, because the essential function of these two genes in temperature and humidity sensing of honeybee parasitic mite, *T. mercedesae*, has been demonstrated by rescuing temperature and humidity preference defects in fruit fly *D. melanogaster* IR25a and IR93a mutants [31]. The gene family variations of HSP70s and aquaporins were explored in Fig. S7 and Table S29 (Supplementary Text 3).

Major digestive enzymes

Major digestive enzymes of the six astigmatic mites are summarized in Table S30. Alpha amylase (EC 3.2.1.1) acts in the first step in the hydrolysis of starch to dextrin and then alpha glucosidase (EC 3.2.1.20) catalyzes the hydrolysis of dextrin to glucose. No alpha amylase gene was identified in *S. scabiei* because of gene synteny (Fig. S8), while the other four astigmatic mites except *T. putrescentiae* have duplicated alpha amylases (Fig. S8A). As a canonical storage mite*, T. putrescentiae* possessed more alpha glucosidases (21 genes) than the other five astigmatic mites, and gene expansion and diversification observed in the phylogenetic analysis (Fig. S8C) are consistent with its preference for high-carbohydrate diet, as a canonical storage mite [32]. It is unexpected that alpha amylase was duplicated in *D. farinae*, *D. pteronyssinus*, *P. ovis*, and *B. tropicalis* but not the storage mite *T. putrescentiae* (Fig. S8). According to our metagenomic analysis, *Neurospora crassa* (a starch utilizer with highly stable amylase) [33] was identified as the most abundant microorganism in *T. putrescentiae* (5.26% in abundance) and but not found in the other five astigmatic mites, implying that the alpha amylase of *Neurospora crassa* is an alternative to the alpha amylase gene duplication in *T. putrescentiae*.

More genes of triacylglycerol lipase were identified in *T. putrescentiae* (58 genes), which was consistent with the fact that *T. putrescentiae* is referred to as a cheese mite and prefers food with high fat content [34]. Phylogenetic tree of all triacylglycerol lipases of the six astigmatic mites confirmed the gene expansion in *T. putrescentiae* (Fig. 4A). Within all 195 triacylglycerol lipase genes, 67 genes (34.4%) were tandemly arrayed and 11 genes (5.6%) were proximally arrayed (Fig. 4A). In the marked cluster PTL1, massive tandem gene duplications were identified in the four psoroptid mites, including 7 tandemly arrayed genes in *S. scabiei*. Of the total 360 serine protease genes, 81 genes (22.5%) were tandemly arrayed and 33 genes (9.2%) were proximally arrayed (Fig. 4B). In the phylogenetic tree (Fig. 4B),

two clusters of *B. tropicalis*-specific (CT1 and T2) and one cluster of *S. scabiei*-specific (T1) serine proteases were identified. None of the serine proteases in the clusters T1 and T2 possessed a complete and active catalytic triad, while 24 out of 28 serine proteases in the cluster CT1 possessed complete and active catalytic triads [35] (Fig. 4C).

Detoxification gene families

In all the five major gene families responsible for detoxification, *T. putrescentiae* has the largest number of genes, while *S. scabiei* has the smallest (Fig. S9). The huge increase in the number of detoxification genes in the storage mite indicates its higher probability to come across xenobiotics in their living environments. Concurrent with the results of GO enrichment in *T. putrescentiae*, 85 CYP genes were identified, which was more than those in the other five mites. In the CYP family, the clans CYP3 and CYP4 are mainly responsible for detoxification of xenobiotics [36]. Significant expansion in the CYP3 and CYP4 clans was observed in *T. putrescentiae*, while expansion only in the CYP3 clan was observed in *B. tropicalis* (Fig. S10A). *S. scabiei* has the fewest genes in both CYP3 and CYP4, with gene decay (Fig. S10B). In the ATP-binding cassette (ABC) transporter gene family, *D. farinae* has more genes when compared with *D. pteronyssinus* and *P. ovis*. The gene number variation mainly occurred in the G and C families, both of which are considered to be related to drug resistance [37]. As visible in the highlighted cluster 1 in Fig. S10C, distinct and independent gene expansions of the ABCG family genes occurred among the six astigmatic mites. Similarly, an ABCB gene decayed in *S. scabiei* (Fig. S10D).

The variation in gene numbers was particularly obvious in the UDP glucuronosyltransferase (UGT) gene family. There were only 5 UGT genes identified in *S. scabiei*, but 73 UGT genes in *T. putrescentiae*. To explore how the gene number varied among the six astigmatic mites, phylogenetic analysis was performed on all UGT genes of the six astigmatic mites (Fig. 4D). Of the total 172 UGTs, 73 genes (42.2%) were tandemly arrayed and 9 genes (5.2%) were proximally arrayed. Among them, 28 UGTs were single-exon genes (16.3% of the total 172 UGTs, highlighted with red labels). When we divided all the UGTs of the six astigmatic mites into three large clusters (UGT1–3), remarkable expansions in the two canonical storage mites *B. tropicalis* and *T. putrescentiae* were observed in UGT2 and UGT3, while significant gene expansions in the two house dust mites *D. farinae* and *D. pteronyssinus* were identified mainly in UGT3. However, the functional difference among these UGT clusters is still unknown.

Horizontal gene transfer

Horizontal gene transfer (HGT), also known as lateral gene transfer, is the movement of genetic information across mating barriers, such as DNA sharing between bacterial and animal genome [38,39]. Massive HGT events were identified in the genomes of six mites and 16 HGT genes were classified into 6 functional categories (Table 2, Supplementary Text 4). All these HGT genes have been excluded possibility of contamination by necessary metagenomic and phylogenetic analysis. Of these 16 HGT genes, the best match hits of 9 genes were from species mainly dwelling in soil environments, including Actinobacteria, Firmicutes and slime mold, supporting astigmatic mites evolved from within soil oribatid mites. The conservation of these HGT genes in other species was listed in Table S31. Considering the important adaptive functions of HGT genes like UGTs and chitinases, we propose that these HGT events facilitated the rapid divergent evolution of astigmatic mites.

Detoxification and UGTs

Phylogenetic analysis (Fig. 5A) revealed that all three clusters of UGTs in the six astigmatic mites showed high similarity to the UGTs of bacterial species, two oribatid mites and three prostigmatic mites, but low similarity to the closest UGTs in other arthropods. In the phylogenetic tree (Fig. 5A), UGT1s were even closer to those of prostigmatic mites than the other two UGT clusters. Therefore, we posit that all the UGTs in astigmatic mites were gained via HGT from some bacteria and UGT1s are the most ancient cluster. Combining with the tandemly arrayed UGTs from *B. tropicalis* and *T. putrescentiae* in different clusters (Fig. 4D), we speculate that the other two clusters of UGTs evolved from UGT1s through expansion and diversification in six astigmatic mites. Since UGTs are important conjugative enzymes in detoxification, we conclude that the HGT of UGTs in astigmatic mites play a vital role in their adaptation of tolerance to toxins. Notably, there is a UGT of the springtail *Folsomia* clustered with those of bacteria (Fig. 5A).

Lysis of fungal cell wall

HGTs involving a group of fungal cell wall lytic enzymes were also identified in astigmatic mites, including two chitinases (Fig. S11), three beta-1,3 glucanases (Fig. S12), and one chitosanase (Fig. S13). It was reported in cotton that both chitinases and beta-1,3 glucanases participated in the resistance to fungi [40], and chitosanase breaks down chitosan, one of the major components of fungal cell wall [41].

Two clusters of chitinases in astigmatic mites were suggested to be gained via HGT (Fig. S11A). Chitinases in cluster 1 are close to chitinases in Actinobacteria with those of oribatid mites (Fig. 5B), while chitinases in cluster 2 have protein homologs to those in other mites (oribatid mites and prostigmatic mites) and springtails (Table S31). Conserved gene syntenies and splice sites (locations of exon-intron boundaries) (Fig. S11A) supported that both the chitinases were gained via two single HGT events in the ancestral mite, respectively. Gene synteny (Fig. S11B) also revealed that a gene decay event (complete deletion) of chitinase in cluster 1 occurred in the genome of *S. scabiei*.

All beta-1,3 glucanases of mites could be divided into three clusters (Fig. S12A), of which cluster 3 was expanded in *D. farinae*, *D. pteronyssinus*, *B. tropicalis* and *T. putrescentiae*. Notably, all beta-1,3 glucanases in cluster 3 are single-exon genes, except 7 genes of *T. putrescentiae*. In gene synteny alignment, we identified that an HGT-gained peptidoglycan endopeptidase was adjacent to the beta-1,3 glucanase in cluster 1 (Fig. S12B). In gene synteny alignment, we confirmed a gene decay event of cluster 3-1 beta-1,3 glucanase in *S. scabiei* (Fig. S12C), and triple tandemly arrayed beta-1,3 glucanases (cluster 3-2) specific in *D. farinae* and *D. pteronyssinus* (Fig. S12C). The triple tandemly arrayed beta-1,3 glucanases were considered as nonexistent in *D. farinae* in a previous report [42].

Seven chitosanases of the six astigmatic mites (two genes in *D. pteronyssinus*, while only one each in the other five species) were identified as HGT genes. To further confirm the HGT, the 7 chitosanases of astigmatic mites and 11 chitosanases retrieved from the UniProt database were phylogenetically analyzed (Fig. S13A). Chitosanases of astigmatic mites shared a very high similarity to that of *Bacillus subtilis* (O07921), which is even higher than that between the chitosanases of *Bacillus subtilis* (O07921) and *Bacillus circulans* (P33673) (Fig. S13A). This unexpectedly high similarity strongly supported that chitosanases in astigmatic mites were horizontally transferred from *Bacillus subtilis*, which is commonly found in soil. Gene synteny alignment and conserved splice sites confirmed that the HGT event occurred in the ancestral astigmatic mites (Fig. S13).

Lysis of bacterial cell wall

Peptidoglycan is the main component of bacterial cell wall [43]. Peptidoglycan endopeptidase and endolytic peptidoglycan transglycosylase are related to the lysis of peptidoglycan [44,45]. Conserved gene synteny (Fig. S12B) revealed that the peptidoglycan endopeptidase and

adjacent beta-1,3 glucanase in cluster 1 were incorporated into the genome of the ancestor of astigmatic mites via possibly a single HGT event. We also found that the peptidoglycan endopeptidase was tandemly duplicated in both *P. ovis* and *T. putrescentiae*, while decayed (complete deletion) in *S. scabiei*. This peptidoglycan endopeptidase was reported in both *D. farinae* and *D. pteronyssinus* [46-48]. These HGT-gained lytic enzymes for fungal and bacterial cell wall are considered to play vital roles in the digestive functions of astigmatic mites, particularly house dust mites, *D. farinae* and *D. pteronyssinus* feeding on human skin flakes colonized by microbes including fungi and bacteria [10,49,50].

Other HGTs

Two bacterial resistance-related genes, stress response protein [51] and D-alanyl-D-alanine dipeptidase [52] were also identified as HGT genes in astigmatic mites. All the stress response protein genes of the six astigmatic mites are single-exon gene and highly conserved with those from the genus *Bacillus*. The N-terminal sequences of the stress response proteins from the six astigmatic mites were well conserved with the *Bacillus* reference gene (Fig. S14A). As seen in Fig. S14B, the highly conserved gene syntenies of the stress response protein among the six mites supported that this gene was conserved among astigmatic mites. D-alanyl-D-alanine dipeptidase is known as the vancomycin resistance protein [52]. It is interesting that only the two canonical storage mites, *B. tropicalis* and *T. putrescentiae* have D-alanyl-D-alanine dipeptidases, but the function of D-alanyl-D-alanine dipeptidases in these two mites is unclear and may be related to symbiosis with some bacteria. Terpene synthase was previously reported as an HGT gene in spider mite and trombidiid mites [53,54] and likewise, we identified it as an HGT gene in the six astigmatic mites (Table 2). Additionally, two methyltransferases and two nucleotide enzymes were also identified as HGT genes (Table 2).

Discussion

Mites and ticks (subclass: Acari) are comprised of a wide range of pests of humans, other animals, and plants [55,56]. Their monophyly was argued with more genetic resources available [5,11-13] and refuted by our phylogenomic analysis (Fig. 1B). Our study proposes Acariformes as an independent monophyletic group from Parasitiformes and further explored its divergence (Fig. 6). Solifugae was suggested as the sister group of Acariformes [5], but the absence of genomic resource impedes our phylogenetic validation. Although the monophyly of Acari was reclaimed in a recent study [13], we have more evidence as below to refute it, besides of our phylogenetic tree in Fig. 1B. In the ultrametric tree (Fig. S4), the estimated divergence time of Acariformes and Parasitiformes was as short as only 1.93 MYA after the root age with *D. melanogaster*, which could not well support the monophyly like the situation in Deuterostomia [57]. Moreover, the three Parasitiformes mites even shared more overlapped orthogroup numbers with *D. melanogaster* than those with the Acariformes (Fig. S5), which suggest the huge gene family difference between these two groups. Likewise, the extremely large number of gene families contracted in Parasitiformes, but very few common gene family variations occurred at the root node of the two groups (Fig. 2A).

The evolutionary history of astigmatic mite species was illustrated from the divergence of the monophyletic Acariformes to the diversification, as shown in Fig. 6. Two HGT genes of UGT and chitinase 2 are conserved with springtail species known as abundant soil dwelling microarthropods (Table S31), but taxonomy and our phylogenomic analysis dost not support that springtail is the sister group of Acariformes; thus, we suggest that the primitive Acariformes species shared the common living environments of soil with springtail species, so that they have the conserved HGT genes. In the monophyletic Acariformes, Prostigmata (including notorious plant pests) first evolved, then rapidly evolving astigmatic mites diverged from Oribatida (soil mites) (Fig. 6). Intriguingly, two gene families of titin and sodium-dependent glucose transporter rapidly expanded in Prostigmata (Table S6), which may contribute to their better locomotion for climbing on plants and absorption of glucose respectively. As to other two lineages, Oribatida has more rapid expanding gene families (Table S8) including a range of detoxification gene families, which reflected they came across huge ecological challenges in soil environments especially when the biological process GO term of response to DDT was significantly enriched (Table S8). The poor assembly continuities of the two annotated Oribatida genomes (<7,500 bp in scaffold N50, Fig. S1) impede our further gene family comparison.

Massive HGT events occurred throughout the divergence history of Acariformes (Fig. 6, Table 2 and S26). The UGT, a bacterial chitinase (cluster 2 in Fig. S11A) and a terpene synthase were gained through HGT in the ancestral Acariformes mite. After the divergence of Prostigmata into plant pests, the common ancestor of Oribatida and Astigmata acquired another bacterial chitinase (cluster 1 in Fig. S11A), a chitosanase (Fig. S14), beta-1,3 glucanases (Fig. S12A) and other four HGT genes (Fig. 6). Chitinase, chitosanase and beta-1,3 glucanases play key roles in the digestion of fungal cell walls. A malonyl-ACP O-methyltransferase and an RNA 2'-phosphotransferase are HGT genes only exist in astigmatic mites, and a D-alanyl-D-alanine dipeptidase is the specific HGT gene in two canonical storage mites *B. tropicalis* and *T. putrescentiae* (Fig. 6 and Table S31). Except the peptidoglycan endopeptidase reported in both *D. farinae* and *D. pteronyssinus* [46-48], none of other HGT events was reported in astigmatic mites previously. UGTs, chitinase and other genes were reported as HGT genes in spider mite [58-60]. Therefore, we identified many novel HGT events in astigmatic mites for the first time. These HGT events provide perfect drug targets for control of these mites [14-16] and were possibly mediated by some endosymbionts such as *Wolbachia* [39,61,62].

Astigmatic mites are a group of rapidly evolving species for their significantly higher mean substitution rates (Fig. 1B) than those of other linages. Astigmatic mites evolved from Oribatida with rapid expansions of iGluRs and ABCs, entered nests of birds and mammals and possibly became commensals of the host animals firstly. We propose that later astigmatic mites experienced two rounds of divergence. In the first round (around 278 MYA, Fig. S4), the ancestor of free-living storage mites including Glycyphagoidea (e.g., *B. tropicalis*) and Acaroidea (e.g., *T. putrescentiae*) branched out and split from the ancestor of psoroptid mites (parvorder: Psoroptidia) which are mostly parasitic mites. This divergence may be associated with the development of their phoretic behavior [63], in which the ancestor of storage mites was carried and settled down in the storage food of some animals, while the ancestor of psoroptid mites developed reversible ectoparasitic lifestyle. As for the second round (around 212 MYA, Fig. S4), psoroptid mites diversified, in which through gene decay, the skin burrowing mite *S. scabiei* developed its more obligate parasitic lifestyle than the sheep scab mite *P. ovis*, an ectoparasitic and non-burrowing mite mainly infesting on fleece of sheep; while afterwards via gene duplications, the ancestor of house dust mites (e.g., *D. farinae* and *D. pteronyssinus*) diverged within Psoroptidia and possibly from an ectoparasitic mite closely

linked to *P. ovis*. Since psoroptid mites are closely related to the bodies of birds and mammals [64], the second round of divergence (around 212 MYA, Fig. S4) was considered as a result of the emergence of feather and hair, especially when the origin of feathers was estimated as ~165–250 MYA [65] and mammals appeared at least 178 MYA [66]. Because non-parasitic mites in the family Pyroglyphidae (including house dust mites) mainly live in the nests of birds and mammals [67], we propose that massive gene duplications enabled their reversal (at around 145 MYA, Fig. S4) from parasitic on animal bodies to free-living in surrounding environments [1]. Since the emergence and diversification of these astigmatic mites occurred much earlier before the establishment of human household environment (Fig. S4), these early evolutionary events should be the genetic basis of their later adaptation to human household environment, but not the response to household environments.

In the diversification of astigmatic mites, a wide range of genetic variations in the genomes of six mites, including many gene family expansions via tandem gene duplications, enabled their rapid genome evolution by acquiring new genes and quick adaptation to rather newly established human household environments, such as the expansions of five detoxification gene families could enable their tolerance to the relatively high levels of toxins and contaminants in the environments. The gene family expansion of iGluRs frequently occurred in different mite lineages, which contributed to their more specified sensing to signals in the living environments. Compared with four psoroptid mites, two canonical storage mites underwent more gene family expansions, especially in detoxification gene families (Fig. 6). Besides, a wide of digestion enzymes especially triacylglycerol lipase and serine protease expanded in astigmatic mites (Fig. 4A and B). Extensive tandem gene duplication of triacylglycerol lipases in the six astigmatic mites facilitates the survival of these mites on a relatively high fat content in the human household environments (Fig. 4A). Regarding the species-specific inactive serine proteases identified in *S. scabiei* and *B. tropicalis* (Fig. 4B and C), the role of those proteases in *S. scabiei* has been reported to be associated with immune evasion or inhibition to the host immune system [35,68,69]. It is possible that these proteases assist the canonical storage mite but now tropical house dust mite, *B. tropicalis* to simultaneously adapt to the human household environments [25,70]. In the same line of thought, a new cluster of active serine proteases in *B. tropicalis* (Fig. 4B) could be responsible for cleavage of tight junctions between epithelial cells in human skin, thereby contributing to the rising rates of allergy in patients in the past decades [70,71]. Meanwhile, many gene decay events occurred in both digestion and detoxification gene families of the skin burrowing mite

S. scabiei (Fig. 6), because of its parasitic fate. Unlike cross-species HGT and gene duplication events of which the occurrence time could be estimated by divergence of species (Fig. 6), species-specific gene duplications and decays may occur at very recent but unknown time points, and possibly generated as results of the adaptation to human household environments. There should be more diverse gene family variations participating in mite evolution, but this study could only cover these significantly enriched gene families.

Conclusions

In this study, high-quality genomes of six allergic astigmatic mites comprising four free-living mites and two parasitic mites contribute to a comparative genomics model for exploration of the evolution and diversification of astigmatic mites and understanding the adaptations of human household pests. In the monophyletic Acariformes independent from Parasitiformes, prostigmatic mites known as plant pests first evolved from the primitive mites dwelling in soil environments; then the rapidly evolving astigmatic mites diverged from soil oribatid mites and later diversified into storage mites, parasitic mites, and house dust mites. The divergence of Acariformes illustrates a concise diagram of pest group evolution, in which plant pests (Prostigmata) and then human household pests (Astigmata) evolved from their relatives in soil environments. From emergence and to diversification of astigmatic mites, many HGT events introduced functionally important genes into their genomes and enriched their adaptation capacities especially in detoxification and digestion and provided perfect drug targets for pest control, and a wide range of gene family expansions via tandem duplications especially in free-living mites facilitated their rapid divergent evolution and quick adaptation to the rather new and rapidly changing human household environments, while many genes decayed in the parasitic mite *S. scabiei* as results of parasitism.

This comparative genomics study comprehensively illustrated the evolutionary dynamics of these allergic astigmatic mites and revealed the genetic evidence for how these mites diverge from soil mites, rapidly evolve, and adapt to human household environments, which vigorously expanded our knowledge to these medically important species and would ultimately facilitate the prevention of associated allergic diseases. Massive novel genomic insights into the adaptation of human household pests were provided by this study and would be important for further exploration and even designing effective pest control strategies.

Methods

Mite culture and purity check

D. farinae, *D. pteronyssinus* and *T. putrescentiae* were cultured in Shenzhen University (Shenzhen, China). The culture methods of *D. farinae* and *D. pteronyssinus* could be found in publications [46,72], and *T. putrescentiae* was cultured in the same way. Mite culture of *B. tropicalis* was performed in the Siriraj Dust Mite Center for Services and Research, Siriraj Hospital, Bangkok, Thailand. *B. tropicalis* mites were fed in the mixture of rat chow and wheat germ and cultured under 25 ± 3 °C and 75 ± 5% relative humidity. After the harvest of pure mite bodies of four species from the mite culture, monospecies of mites was confirmed during cultivation and at the final stage by observation of the morphology under a light microscope and verified with the pictorial keys [73]. Further purity check was confirmed by singleplex PCR with species-specific primers after genomic DNA extraction.

Sample preparation of genomic DNA and RNA

Genomic DNAs were extracted using the Qiagen Blood & Cell Culture DNA Maxi Kit (Qiagen, Germany). Firstly, pure mite bodies were washed twice with Phosphate-buffered saline (PBS, pH 7.4) and homogenized into powder form by mortar and pestle with liquid nitrogen for keeping low temperature. The homogenized mite samples were then incubated at 50 °C for proteolysis following the manufacturer's protocol. Genomic DNA was bound to the column, washed, and eluted by elution buffer, and precipitated in 70% ethanol cold in 4 °C. Finally, air-dry pellets of genomic DNA were dissolved overnight in UltraPure™ DNase/RNase-free distilled water (Thermo Fisher Scientific, USA) overnight at room temperature. The integrity of genomic DNA was determined by electrophoresis in 0.5% agarose gel and analyzed by Agilent 2100 Bioanalyzer (Agilent Technologies, USA), and the quantity was detected by Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, USA) and Qubit Fluorometer (Thermo Fisher Scientific, USA).

For RNA extraction, the homogenized mite powder was transferred into tubes with TRIzol reagent (Invitrogen, USA), followed by phenol-chloroform extraction. The colorless upper aqueous phase was transferred to the column from the PureLink RNA Mini Kit (Thermo Fisher Scientific, USA) following the manufacturer's protocol. After extraction, the concentration of RNA samples was quantified by Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, USA) and Qubit Fluorometer (Thermo Fisher Scientific, USA). Then, with the RNA samples, SMARTer™ PCR cDNA Synthesis Kit (Takara Bio, Japan)

was used to synthesize double-stranded cDNA. Both integrities of RNA and double-stranded cDNA were qualified by Agilent 2100 Bioanalyzer (Agilent Technologies, USA).

Genome and transcriptome sequencing

To assemble the genomes of *D. farinae*, *D. pteronyssinus*, *T. putrescentiae* and *B. tropicalis*, a variety of sequencing platforms were employed, and all data of genomic DNA sequencing are listed in Table S1. The genome sequencing data of *S. scabiei* (*var. canis*) were downloaded from NCBI database (BioProject accession: PRJNA268368). Besides, genome sequencing data of *D. farinae* and *D. pteronyssinus* (BioProject accession: PRJNA174061 and PRJNA388362, respectively) were also downloaded.

For transcriptome sequencing of *B. tropicalis* and *T. putrescentiae*, both RNA sequencing and cDNA sequencing were performed on Illumina HiSeq 2500 by Groken Bioscience (Hong Kong) and paired-end 150-bp reads were generated. All data of transcriptome are listed in Table S2, while transcriptome sequencing data of *D. farinae*, *D. pteronyssinus*, *P. ovis* and *S. scabiei* were downloaded from NCBI database (BioProject accession: PRJNA174061, PRJNA388362, PRJNA521406 and PRJNA304361, respectively). Quality control of both sequencing reads of genomic DNA and transcriptome was performed by FastQC v0.11.8.

Genome and transcriptome assembly

For *D. farinae*, the initial genome assembly was constructed with PacBio long reads using Flye v2.6 [74]. Then, scaffolding was performed by SSPACE Basic v2.0 [75] with paired-end Illumina short reads [76] and SSPACE LongRead v1.1 [77] with PacBio long reads [78]. Sequence polishing was finished by Pilon v1.22 [79] with all Illumina short reads. Further scaffolding was performed with SSPACE-LongRead v.1.1 [77] using Oxford Nanopore Technologies (ONT) sequencing reads [80] and gaps were filled with the raw ONT reads using LR_Gapcloser v1.0 [81].

For *D. pteronyssinus*, ONT sequencing reads [80] were used to build an initial assembly using Canu v1.8 [82]. The preliminary draft assembly were polished for three rounds of self-correction using Medaka v1.2.0 [83] with ONT sequencing reads. To further improve the accuracy of the assembly, reads from Illumina sequencing were aligned with BWA [84] and additional three rounds of consensus correction were performed by Pilon v1.22 [79]. The resulting contigs underwent a scaffolding procedure by SSPACE-LongRead v.1.1 [77] with

default parameters using ONT reads . The gaps within the scaffolds were further filled with the raw ONT reads using LR_Gapcloser v1.0 [81].

For *S. scabiei*, all Illumina short reads were assembled with SPAdes v3.13.1 [85] into the first draft genome. Then, scaffolding was performed by SSPACE Basic v2.0 [75] with paired-end Illumina short reads, followed by gap filling using GapFiller v1.10 [86] and GapFinisher v1.1 [87]. Finally, sequence polishing was completed by Pilon v1.22 [79] with all reads and the final genome assembly was generated.

For *B. tropicalis* and *T. putrescentiae*, first draft genome assemblies were constructed with PacBio long reads using Canu v1.8 [82], Flye v2.6 [74] and Miniasm v0.2.0 [88]-Racon v1.0.0 [89] pipeline. Then, scaffolding was performed by SSPACE Basic v2.0 [75] with paired-end Illumina short reads and SSPACE LongRead v1.1 [77] with PacBio long reads, followed by gap filling using GapFiller v1.10 [86], GapFinisher v1.1 [87], GMcloser v1.6.2 [90] and FGAP v1.8.1 [91]. Finally, sequence polishing was finished by Pilon v1.22 [79] with all NGS short reads, generating the final genome assemblies.

All genome assemblies were assessed by QUAST v5.0.2 [92] in continuity and the completeness of genome assembly was assessed by BUSCO v3.1.0 [93] with database arthropoda_odb9. All these statistics were summarized in Fig. S1, Table 1 and S3. All genome assemblies were assessed by QUAST v5.0.2 [92] in continuity and the completeness of genome assembly was assessed by BUSCO v3.1.0 [93] with database arthropoda_odb9. All these statistics were summarized in Fig. S1.

Two transcriptome assemblies were constructed, namely *de novo* assembly and reference-mapping based assembly. For the *de novo* assembly, Trinity v2.8.4 [94] was used to assemble transcriptome sequences with the RNA and cDNA sequencing reads. For the reference-mapping based assembly, Hisat2 v2.0.4 [95] was used for mapping transcriptome reads to the genome assembly. Then, Samtools v1.9 [96], StringTie v1.3.0 [95], and Gffread [97] converted reads mapping results into transcriptome sequences.

Genome annotation

With genome and transcriptome assembled, genome annotation was performed. Firstly, repeat masking was performed by *de novo* prediction with RepeatModeler v2.0.1 [98] and

masking with RepeatMasker v4.0.8 [99] (RepBase edition 20181026). In the *de novo* prediction with RepeatModeler v2.0.1 [98], RECON v1.05 [100] and RepeatScout v1.0.6 [101] were used for the *de novo* prediction of repeat family in the genome.

Then, genome annotation was performed by Maker pipeline v2.31 [102]. In the Maker pipeline, transcriptome assemblies and homologous proteins were used as evidence for alignment by Exonerate v2.4.0 [103], whilst gene prediction was completed by SNAP (lib v2017-03-01) [104], GeneMark v4.38 [105] and Augustus v3.3.1 [106]. The quality of genome annotation was assessed by BUSCO v3.1.0 [93] with database arthropoda_odb9. The statistics of annotations of the six astigmatic mites were summarized in Table 1 and S3.

Collection of genome assemblies

Genome assemblies of other 20 species of mites and ticks, and the pseudoscorpion *Cordylochernes scorpioides* were directly downloaded from NCBI assembly database (Table S4). The genome assembly of *Drosophila* (*D.*) *melanogaster* (NCBI accession: GCF_000001215.4) was used as an outgroup. Completeness of genome assemblies was assessed by BUSCO v3.1.0 [93] with database arthropoda_odb9 (Fig. S1). Other assembly statistics was assessed by QUAST v5.0.2 [92] (Fig. S1). Repeat annotation was performed by *de novo* prediction with RepeatModeler v2.0.1 and masking with RepeatMasker v4.0.8 [99] (RepBase edition 20181026). In the *de novo* prediction step in RepeatModeler v2.0.1 [98], RECON v1.05 [100], and RepeatScout v1.0.6 [101] were used for the prediction of repeat families in the genome assemblies.

Phylogenetic analysis of genome assemblies

Based on the genome completeness assessment above, 47 common single and complete BUSCO proteins of 28 genomes were collected and aligned by the MAFFT [107] and then edited in Gblocks [108] with the options '-b4=5 -b5=h' to generate a sequence alignment of 13,133 conserved amino-acid residues. Then the sequence alignment was used to construct the phylogenetic tree (Fig. S1) in maximum likelihood algorithm and 100 bootstrap replicates by RAxML v8.2.12 with the options '-m PROTCATWAG -f a -# 100'. The mean substitution rates (MSRs) among five phylogenetic lineages (Fig. 1B) were compared with the outgroup *D. melanogaster* by relative-rate tests using RRTree v1.1.11 [109].

Then the ultrametric time tree (Fig. S4) was constructed by BEAST v2.6.6 [110] using Yule Model and calibrated by the divergence times of *Drosophila melanogaster-Tetranychus urticae* (mean time: 605 MYA) and *Tetranychus urticae-Platynothrus peltifer* (mean time: 526 MYA) provided by TIMETREE [111], and final edited by the online tool Interactive Tree of Life (iTOL) [112].

Collection of proteomes

The genomes of *D. farinae*, *D. pteronyssinus*, *P. ovis*, *S. scabiei* (*var. canis*), *B. tropicalis* and *T. putrescentiae* were all annotated by our group. Combined with the other 8 proteomes of mites downloaded from NCBI GenBank or UniProt database, a total of 14 proteomes of mites were collected, and the quality was assessed by BUSCO v3.1.0 [93] with database arthropoda_odb9 (Table S5). For *D. tinctorium*, a one-protein-per-gene proteome (UniProt ID: UP000285301) was downloaded from UniProt database. The high duplication rates of *T. urticae* (18.8%), *V. destructor* (52.2%), *V. jacobsoni* (48.3%) were caused by too many redundant isoforms included in the proteomes, so their protein-per-gene proteomes were extracted for analysis. Besides, the high duplication rate in *D. tinctorium* (9.8%) was probably caused by the high heterozygosity of the sample, similar as that of the genome.

Phylogenomic analysis and gene ontology enrichment analysis

Phylogenomic orthology analysis was performed among 14 proteomes of mites and *D. melanogaster* (UniProt ID: UP000000803) as an outgroup by OrthoFinder v2.5.4 [113,114] and CAFÉ v4.2 [115]. With the proteome of *D. melanogaster* as outgroup, all 14 proteomes of mites were assigned into orthogroups according to the protein similarities on sequence level (Fig. S5). These orthogroups were also considered as gene families.

Gene ontology (GO) enrichment analysis was performed by the online tool DAVID (david.ncifcrf.gov). Rapidly evolving gene families and species-specific orthogroups were collected, of which the genes were mapped to the proteome of *D. melanogaster* (UniProt ID: UP000000803) to acquire UniProt accession identifiers for GO enrichment in BLASTP v2.9.0 [116] with E-value cutoff as 1E-6, to understand the evolutionary history of gene families.

Collection and comparison of gene families

All proteins of the six astigmatic mites were searched by BLASTP v2.9.0 [116] with reference proteins in Swiss-Prot database at E-value cutoff of 1E-6. After manual curation based on transcriptome data especially intron-exon split sites and filtering out proteins with shorter than 50% average length of the gene family, all proteins of the six astigmatic mites identified in the target gene families were aligned and drawn into a phylogenetic tree. Sequence alignment was performed by CLUSTAL W [117] and MUSCLE [118] in MEGA v10.2.2 [119], and all phylogenetic trees were constructed by MEGA v10.2.2 [119] with maximum likelihood (ML) algorithm in the JTT (Jones-Taylor-Thornton) model, 80% site coverage and 100 bootstrap replicates, and then edited by online tool Interactive Tree of Life (iTOL) [112]. The similarity matrix (BLOSUM62) of genes was analyzed and generated by the online tool SIAS (Sequence Identity And Similarity, http://imed.med.ucm.es/Tools/sias.html) using default parameters.

If two genes were located adjacently on genome and no other gene was located between them, these two genes were both considered as tandemly arrayed genes generated via tandem duplication [120]. If two genes were proximally arrayed (separated by no more than 10 genes) but no transposable element was identified as adjacent to these proximally arrayed genes, we consider these genes as generated via anciently tandem duplication events [120]. Frequent tandemly arrayed genes were identified in gene families, like ionotropic glutamate receptors of astigmatic mites, and connected with curve lines in phylogenetic trees. If one gene is both tandemly arrayed with a gene and proximally arrayed with another gene, this gene was classified as tandemly arrayed gene.

In addition, to explore the potential symbiont microorganisms in the six astigmatic mites, Illumina short reads of genomic DNA were analyzed using Biobakery workflows [121], in which MetaPhlAn (Metagenomic Phylogenetic Analysis) v2.0 [122] was employed for the taxonomical profiling of microbial communities.

Identification of HGT events

A modified method based on Alastair Crisp et al., 2015 [123] was performed to identify the HGT events in the six astigmatic mites. All proteins from bacteria and other eukaryotes (excluding all metazoa) in UniRef50 database (UniProt) were collected and built as reference databases, while all proteins from other metazoa (excluding all arthropods) were collected and built as a comparison database.

The HGT index was calculated by dividing the bit score of the best bacteria or other eukaryotes (excluding all metazoa) match by that of the best other metazoa (excluding all arthropods) match. If one gene has a HGT index ≥ 2, or has match hits in bacteria or other eukaryotes (excluding all metazoa) databases but not in other metazoa (excluding all arthropods) database with BLASTP v2.9.0 [116] (E-value cutoff set as 1E-6), this gene was considered as a candidate HGT gene. All candidate HGT genes were further validated via BLASTP [116] to non-redundant protein (NR) database (online version). If the top non-Acariformes matches of one candidate HGT gene to NR database are from bacteria or other eukaryotes (excluding all metazoa), but not other arthropods or other metazoa (excluding all arthropods), this candidate HGT gene was considered as acquired via HGT. In this study, massive horizontal gene transferring events in astigmatic mites were identified and summarized in Table 2.

Data availability

All data are accessible under NCBI BioProject numbers (PRJNA174061 for *D. farinae*, PRJNA388362 for *D. pteronyssinus*, PRJNA702011 for *B. tropicalis*, PRJNA706095 for *T. putrescentiae*).

References

1 Klimov, P. B. & Oconnor, B. Is Permanent Parasitism Reversible?—Critical Evidence from Early Evolution of House Dust Mites. *Systematic Biology* **62**, 411-423, doi:10.1093/sysbio/syt008 (2013).

2 Domes, K., Althammer, M., Norton, R. A., Scheu, S. & Maraun, M. The phylogenetic relationship between Astigmata and Oribatida (Acari) as indicated by molecular markers. *Experimental & applied acarology* **42**, 159-171, doi:10.1007/s10493-007-9088-8 (2007).

3 Norton, R. A. Morphological evidence for the evolutionary origin of Astigmata (Acari: Acariformes). *Experimental & applied acarology* **22**, 559-594 (1998).

4 Sakata, T. & Norton, R. A. Opisthonotal gland chemistry of early-derivative oribatid mites (Acari) and its relevance to systematic relationships of Astigmata. *International Journal of Acarology* **27**, 281-292 (2001).

5 Dabert, M., Witalinski, W., Kazmierski, A., Olszanowski, Z. & Dabert, J. Molecular phylogeny of acariform mites (Acari, Arachnida): Strong conflict between phylogenetic signal and long-branch attraction artifacts. *Molecular Phylogenetics and Evolution* **56**, 222-241, doi:https://doi.org/10.1016/j.ympev.2009.12.020 (2010).

6 Calderón, M. A. *et al.* Respiratory allergy caused by house dust mites: What do we really know? *Journal of Allergy and Clinical Immunology* **136**, 38-48, doi:https://doi.org/10.1016/j.jaci.2014.10.012 (2015).

7 Miller, J. D. The Role of Dust Mites in Allergy. *Clinical reviews in allergy & immunology* **57**, 312-329, doi:10.1007/s12016-018-8693-0 (2019).

8 Thomas, W. R., Hales, B. J. & Smith, W. A. House dust mite allergens in asthma and allergy. *Trends in molecular medicine* **16**, 321-328, doi:10.1016/j.molmed.2010.04.008 (2010).

9 Arlian, L. G. & Platts-Mills, T. A. The biology of dust mites and the remediation of mite allergens in allergic disease. *The Journal of allergy and clinical immunology* **107**, S406-413, doi:10.1067/mai.2001.113670 (2001).

10 Colloff, M. J. *Dust Mites*. (Springer, 2009).

11 Pepato, A. R., da Rocha, C. E. F. & Dunlop, J. A. Phylogenetic position of the acariform mites: sensitivity to homology assessment under total evidence. *BMC Evolutionary Biology* **10**, 235, doi:10.1186/1471-2148-10-235 (2010).

12 Pepato, A. R. & Klimov, P. B. Origin and higher-level diversification of acariform mites – evidence from nuclear ribosomal genes, extensive taxon sampling, and secondary structure alignment. *BMC Evolutionary Biology* **15**, 178, doi:10.1186/s12862-015-0458-2 (2015).

13 Lozano-Fernandez, J. *et al.* Increasing species sampling in chelicerate genomic-scale datasets provides support for monophyly of Acari and Arachnida. *Nature Communications* **10**, 2295, doi:10.1038/s41467-019-10244-7 (2019).

14 Manna, S. & Harman, A. Horizontal gene transfer of a Chlamydial tRNA-guanine transglycosylase gene to eukaryotic microbes. *Molecular Phylogenetics and Evolution* **94**, 392-396, doi:https://doi.org/10.1016/j.ympev.2015.09.022 (2016).

15 Peixoto, L. & Roos, D. S. Genomic scale analysis of lateral gene transfer in Apicomplexan parasites: insights into early eukaryotic evolution, host-pathogen interaction and drug target development. *BMC Bioinformatics* **8**, S5, doi:10.1186/1471-2105-8-S8-S5 (2007).

16 Eme, L., Gentekaki, E., Curtis, B., Archibald, J. M. & Roger, A. J. Lateral Gene Transfer in the Adaptation of the Anaerobic Parasite Blastocystis to the Gut. *Current biology : CB* **27**, 807-820, doi:10.1016/j.cub.2017.02.003 (2017).

17 Calderon, M. A. *et al.* House Dust Mite Respiratory Allergy: An Overview of Current

Therapeutic Strategies. *The journal of allergy and clinical immunology. In practice* **3**, 843-855, doi:10.1016/j.jaip.2015.06.019 (2015).
18 Voorhorst, R., Spieksma, F. T. M., Varekamp, H., Leupen, M. & Lyklema, A. The house-dust mite (Dermatophagoides pteronyssinus) and the allergens it produces. Identity with the house-dust allergen. *Journal of Allergy* **39**, 325-339 (1967).
19 Peat, J. K. *et al.* House dust mite allergens. A major risk factor for childhood asthma in Australia. *American journal of respiratory and critical care medicine* **153**, 141-146 (1996).
20 Thomas, W. R., Smith, W.-A., Hales, B. J., Mills, K. L. & O'Brien, R. M. Characterization and immunobiology of house dust mite allergens. *International archives of allergy and immunology* **129**, 1-18 (2002).
21 Sinclair, A. & Kirkwood, A. Feeding behaviour of Psoroptes ovis. (1983).
22 Hamilton, K., Nisbet, A., Lehane, M., Taylor, M. & Billingsley, P. F. A physiological and biochemical model for digestion in the ectoparasitic mite, Psoroptes ovis (Acari: Psoroptidae). *International journal for parasitology* **33**, 773-785 (2003).
23 Arlian, L. G. & Morgan, M. S. A review of Sarcoptes scabiei: past, present and future. *Parasites & Vectors* **10**, 297, doi:10.1186/s13071-017-2234-1 (2017).
24 Arlian, L. G. Biology, host relations, and epidemiology of Sarcoptes scabiei. *Annual review of entomology* **34**, 139-159 (1989).
25 Guilleminault, L. & Viala-Gastan, C. [Blomia tropicalis: A house dust mite in the tropics]. *Revue des maladies respiratoires* **34**, 791-801, doi:10.1016/j.rmr.2016.10.877 (2017).
26 Green, W. F. & Woolcock, A. J. Tyrophagus putrescentiae: an allergenically important mite. *Clinical allergy* **8**, 135-144, doi:10.1111/j.1365-2222.1978.tb00458.x (1978).
27 Burgess, S. T. G. *et al.* A genomic analysis and transcriptomic atlas of gene expression in Psoroptes ovis reveals feeding- and stage-specific patterns of allergen expression. *BMC Genomics* **20**, 756, doi:10.1186/s12864-019-6082-6 (2019).
28 Shen, N. *et al.* Comparative analysis of the allergenic characteristics and serodiagnostic potential of recombinant chitinase-like protein-5 and -12 from Sarcoptes scabiei. *Parasites & Vectors* **14**, 148, doi:10.1186/s13071-021-04654-0 (2021).
29 Rytz, R., Croset, V. & Benton, R. Ionotropic receptors (IRs): chemosensory ionotropic glutamate receptors in Drosophila and beyond. *Insect biochemistry and molecular biology* **43**, 888-897, doi:10.1016/j.ibmb.2013.02.007 (2013).
30 van Giesen, L. & Garrity, P. A. More than meets the IR: the expanding roles of variant Ionotropic Glutamate Receptors in sensing odor, taste, temperature and moisture. *F1000Research* **6**, 1753, doi:10.12688/f1000research.12013.1 (2017).
31 Lei, J., Liu, Q. & Kadowaki, T. Honey Bee Parasitic Mite Contains the Sensilla-Rich Sensory Organ on the Foreleg Tarsus Expressing Ionotropic Receptors With Conserved Functions. *Frontiers in Physiology* **10**, doi:10.3389/fphys.2019.00556 (2019).
32 Erban, T., Erbanova, M., Nesvorna, M. & Hubert, J. 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. *Archives of insect biochemistry and physiology* **71**, 139-158, doi:10.1002/arch.20312 (2009).
33 Tavano, O. *et al.* Stabilization of an Amylase from Neurospora crassa by Immobilization on Highly Activated Supports. *Food Biotechnology* **22**, 262-275, doi:10.1080/08905430802262616 (2008).
34 Erban, T., Rybanska, D. & Hubert, J. Population Growth of the Generalist Mite Tyrophagus putrescentiae (Acari: Acaridida) Following Adaptation to High- or Low-Fat and High- or Low-Protein Diets and the Effect of Dietary Switch. *Environmental*

entomology **44**, 1599-1604, doi:10.1093/ee/nvv129 (2015).
35 Holt, D. C. *et al.* Mechanisms for a novel immune evasion strategy in the scabies mite sarcoptes scabiei: a multigene family of inactivated serine proteases. *J Invest Dermatol* **121**, 1419-1424, doi:10.1046/j.1523-1747.2003.12621.x (2003).
36 Zhang, W. *et al.* Identification and Characterization of Three New Cytochrome P450 Genes and the Use of RNA Interference to Evaluate Their Roles in Antioxidant Defense in Apis cerana cerana Fabricius. *Front Physiol* **9**, 1608, doi:10.3389/fphys.2018.01608 (2018).
37 Liu, X. in *Drug Transporters in Drug Disposition, Effects and Toxicity* (eds Xiaodong Liu & Guoyu Pan) 13-100 (Springer Singapore, 2019).
38 Soucy, S. M., Huang, J. & Gogarten, J. P. Horizontal gene transfer: building the web of life. *Nature Reviews Genetics* **16**, 472-482, doi:10.1038/nrg3962 (2015).
39 Husnik, F. & McCutcheon, J. P. Functional horizontal gene transfer from bacteria to eukaryotes. *Nature Reviews Microbiology* **16**, 67-79, doi:10.1038/nrmicro.2017.137 (2018).
40 Xu, X. *et al.* Genome-wide characterization of the β-1,3-glucanase gene family in Gossypium by comparative analysis. *Scientific Reports* **6**, 29044, doi:10.1038/srep29044 (2016).
41 Fukamizo, T. & Brzezinski, R. Chitosanase from Streptomyces sp. strain N174: a comparative review of its structure and function. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **75**, 687-696, doi:10.1139/o97-079 (1997).
42 Waldron, R. *et al.* Characterisation of three novel β-1,3 glucanases from the medically important house dust mite Dermatophagoides pteronyssinus (airmid). *Insect biochemistry and molecular biology* **115**, 103242, doi:10.1016/j.ibmb.2019.103242 (2019).
43 Meroueh, S. O. *et al.* Three-dimensional structure of the bacterial cell wall peptidoglycan. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 4404, doi:10.1073/pnas.0510182103 (2006).
44 Shin, J.-H. *et al.* Structural basis of peptidoglycan endopeptidase regulation. *Proceedings of the National Academy of Sciences* **117**, 11692-11702 (2020).
45 Lee, M. *et al.* Exolytic and endolytic turnover of peptidoglycan by lytic transglycosylase Slt of <em>Pseudomonas aeruginosa</em&gt. *Proceedings of the National Academy of Sciences* **115**, 4393, doi:10.1073/pnas.1801298115 (2018).
46 Chan, T. F. *et al.* The draft genome, transcriptome, and microbiome of Dermatophagoides farinae reveal a broad spectrum of dust mite allergens. *The Journal of allergy and clinical immunology* **135**, 539-548, doi:10.1016/j.jaci.2014.09.031 (2015).
47 Erban, T., Di Presa, C. A., Kopecky, J., Poltronieri, P. & Hubert, J. PCR detection of the 14.5 antibacterial NlpC/P60-like Dermatophagoides pteronyssinus protein in Dermatophagoides farinae (Acari: Pyroglyphidae). *Journal of medical entomology* **50**, 931-933, doi:10.1603/me13027 (2013).
48 Mathaba, L. T. *et al.* Isolation and characterisation of a 13.8-kDa bacteriolytic enzyme from house dust mite extracts: homology with prokaryotic proteins suggests that the enzyme could be bacterially derived. *FEMS Immunology & Medical Microbiology* **33**, 77-88, doi:10.1111/j.1574-695X.2002.tb00576.x (2002).
49 Portnoy, J. *et al.* Environmental assessment and exposure control of dust mites: a practice parameter. *Ann Allergy Asthma Immunol* **111**, 465-507, doi:10.1016/j.anai.2013.09.018 (2013).
50 Colloff, M. J. *et al.* The control of allergens of dust mites and domestic pets: a position paper. *Clinical and experimental allergy : journal of the British Society for Allergy and*

Clinical Immunology **22 Suppl 2**, 1-28, doi:10.1111/j.1365-2222.1992.tb01763.x (1992).

51 Hecker, M. & Völker, U. General stress response of Bacillus subtilis and other bacteria. *Advances in microbial physiology* **44**, 35-91, doi:10.1016/s0065-2911(01)44011-2 (2001).

52 Lessard, I. A. & Walsh, C. T. VanX, a bacterial D-alanyl-D-alanine dipeptidase: resistance, immunity, or survival function? *Proceedings of the National Academy of Sciences of the United States of America* **96**, 11028-11032, doi:10.1073/pnas.96.20.11028 (1999).

53 Grbić, M. *et al.* The genome of Tetranychus urticae reveals herbivorous pest adaptations. *Nature* **479**, 487, doi:10.1038/nature10640 https://www.nature.com/articles/nature10640#supplementary-information (2011).

54 Dong, X. *et al.* Genomes of trombidid mites reveal novel predicted allergens and laterally transferred genes associated with secondary metabolism. *Gigascience* **7**, doi:10.1093/gigascience/giy127 (2018).

55 Xiong, Q., Wan, A. T. Y. & Tsui, S. K. A Mini-review of the Genomes and Allergens of Mites and Ticks. *Curr Protein Pept Sci* **21**, 114-123, doi:10.2174/1389203720666190719150432 (2020).

56 Varma, M. R. G. in *Medical Insects and Arachnids* (eds Richard P. Lane & Roger W. Crosskey) 597-658 (Springer Netherlands, 1993).

57 Kapli, P. *et al.* Lack of support for Deuterostomia prompts reinterpretation of the first Bilateria. *Science Advances* **7**, eabe2741, doi:doi:10.1126/sciadv.abe2741 (2021).

58 Ahn, S. J., Dermauw, W., Wybouw, N., Heckel, D. G. & Van Leeuwen, T. Bacterial origin of a diverse family of UDP-glycosyltransferase genes in the Tetranychus urticae genome. *Insect biochemistry and molecular biology* **50**, 43-57, doi:10.1016/j.ibmb.2014.04.003 (2014).

59 Snoeck, S. *et al.* Substrate specificity and promiscuity of horizontally transferred UDP-glycosyltransferases in the generalist herbivore Tetranychus urticae. *Insect biochemistry and molecular biology* **109**, 116-127, doi:10.1016/j.ibmb.2019.04.010 (2019).

60 Wybouw, N., Van Leeuwen, T. & Dermauw, W. A massive incorporation of microbial genes into the genome of Tetranychus urticae, a polyphagous arthropod herbivore. *Insect molecular biology* **27**, 333-351, doi:10.1111/imb.12374 (2018).

61 Klasson, L., Kambris, Z., Cook, P. E., Walker, T. & Sinkins, S. P. Horizontal gene transfer between Wolbachia and the mosquito Aedes aegypti. *BMC genomics* **10**, 1-9 (2009).

62 de Miguel, T., Zhu, O. & Villa, T. G. in *Horizontal Gene Transfer* 227-234 (Springer, 2019).

63 Houck, M. A. & Oconnor, B. M. Ecological and Evolutionary Significance of Phoresy in the Astigmata. *Annual Review of Entomology* **36**, 611-636, doi:10.1146/annurev.en.36.010191.003143 (1991).

64 Proctor, H. C. Feather Mites (Acari: Astigmata): Ecology, Behavior, and Evolution. *Annual Review of Entomology* **48**, 185-209, doi:10.1146/annurev.ento.48.091801.112725 (2003).

65 Benton, M. J., Dhouailly, D., Jiang, B. & McNamara, M. The Early Origin of Feathers. *Trends in Ecology & Evolution* **34**, 856-869, doi:https://doi.org/10.1016/j.tree.2019.04.018 (2019).

66 Pickrell, J. How the earliest mammals thrived alongside dinosaurs. *Nature* **574**, 468-473 (2019).

67 Arlian, L. G. in *Encyclopedia of Insects (Second Edition)* (eds Vincent H. Resh & Ring

T. Cardé) 152-156 (Academic Press, 2009).
68 Reynolds, S. L. *et al.* Scabies mite inactive serine proteases are potent inhibitors of the human complement lectin pathway. *PLoS neglected tropical diseases* **8**, e2872, doi:10.1371/journal.pntd.0002872 (2014).
69 Bergström, F. C. *et al.* Scabies mite inactivated serine protease paralogs inhibit the human complement system. *Journal of immunology (Baltimore, Md. : 1950)* **182**, 7809-7817, doi:10.4049/jimmunol.0804205 (2009).
70 Santos da Silva, E. *et al.* Allergens of Blomia tropicalis: An Overview of Recombinant Molecules. *International archives of allergy and immunology* **172**, 203-214, doi:10.1159/000464325 (2017).
71 Wan, H. *et al.* The transmembrane protein occludin of epithelial tight junctions is a functional target for serine peptidases from faecal pellets of Dermatophagoides pteronyssinus. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* **31**, 279-294, doi:10.1046/j.1365-2222.2001.00970.x (2001).
72 Liu, X. Y. *et al.* High-quality assembly of Dermatophagoides pteronyssinus genome and transcriptome reveals a wide range of novel allergens. *Journal of Allergy and Clinical Immunology* **141**, 2268-2271 (2018).
73 Colloff, M. J. & Spieksma, F. T. Pictorial keys for the identification of domestic mites. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* **22**, 823-830, doi:10.1111/j.1365-2222.1992.tb02826.x (1992).
74 Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology* **37**, 540-546, doi:10.1038/s41587-019-0072-8 (2019).
75 Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578-579, doi:10.1093/bioinformatics/btq683 (2010).
76 Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53-59, doi:10.1038/nature07517 (2008).
77 Boetzer, M. & Pirovano, W. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* **15**, 211, doi:10.1186/1471-2105-15-211 (2014).
78 Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133-138 (2009).
79 Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one* **9**, e112963 (2014).
80 Deamer, D., Akeson, M. & Branton, D. Three decades of nanopore sequencing. *Nature Biotechnology* **34**, 518-524, doi:10.1038/nbt.3423 (2016).
81 Xu, G.-C. *et al.* LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. *GigaScience* **8**, doi:10.1093/gigascience/giy157 (2018).
82 Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research* **27**, 722-736 (2017).
83 Medaka. medaka: Sequence correction provided by ONT Research. https://github.com/nanoporetech/medaka. (Apr. 2020).
84 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
85 Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology* **19**, 455-477 (2012).
86 Nadalin, F., Vezzi, F. & Policriti, A. GapFiller: a de novo assembly approach to fill the gap within paired reads. *BMC bioinformatics* **13**, S8 (2012).

87 Kammonen, J. I. *et al.* gapFinisher: A reliable gap filling pipeline for SSPACE-LongRead scaffolder output. *PloS one* **14**, e0216885 (2019).
88 Li, H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**, 2103-2110, doi:10.1093/bioinformatics/btw152 (2016).
89 Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Racon-Rapid consensus module for raw de novo genome assembly of long uncorrected reads. *Bioinformatics* **18**, 452-464 (2002).
90 Kosugi, S., Hirakawa, H. & Tabata, S. GMcloser: closing gaps in assemblies accurately with a likelihood-based selection of contig or long-read alignments. *Bioinformatics* **31**, 3733-3741, doi:10.1093/bioinformatics/btv465 (2015).
91 Piro, V. C. *et al.* FGAP: an automated gap closing tool. *BMC research notes* **7**, 371 (2014).
92 Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072-1075 (2013).
93 Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).
94 Grabherr, M. G. *et al.* Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature biotechnology* **29**, 644 (2011).
95 Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature protocols* **11**, 1650 (2016).
96 Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
97 Pertea, G. & Pertea, M. GFF Utilities: GffRead and GffCompare. *F1000Research* **9** (2020).
98 Smit, A. F. & Hubley, R. RepeatModeler Open-1.0. *Available fom http://www. repeatmasker. org* (2008).
99 Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current protocols in bioinformatics* **25**, 4.10. 11-14.10. 14 (2009).
100 Bao, Z. & Eddy, S. R. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome research* **12**, 1269-1276 (2002).
101 Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351-i358, doi:10.1093/bioinformatics/bti1018 (2005).
102 Cantarel, B. L. *et al.* MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome research* **18**, 188-196 (2008).
103 Slater, G. S. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31, doi:10.1186/1471-2105-6-31 (2005).
104 Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59, doi:10.1186/1471-2105-5-59 (2004).
105 Borodovsky, M. & Lomsadze, A. Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. *Curr Protoc Bioinformatics* **Chapter 4**, Unit 4.6.1-10, doi:10.1002/0471250953.bi0406s35 (2011).
106 Hoff, K. J. & Stanke, M. Predicting Genes in Single Genomes with AUGUSTUS. *Curr Protoc Bioinformatics* **65**, e57, doi:10.1002/cpbi.57 (2019).
107 Katoh, K., Misawa, K., Kuma, K. i. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059-3066, doi:10.1093/nar/gkf436 (2002).
108 Castresana, J. Selection of Conserved Blocks from Multiple Alignments for Their Use

in Phylogenetic Analysis. *Molecular Biology and Evolution* **17**, 540-552, doi:10.1093/oxfordjournals.molbev.a026334 (2000).
109 Robinson-Rechavi, M. & Huchon, D. RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics* **16**, 296-297, doi:10.1093/bioinformatics/16.3.296 (2000).
110 Bouckaert, R. *et al.* BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLOS Computational Biology* **10**, e1003537, doi:10.1371/journal.pcbi.1003537 (2014).
111 Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution* **34**, 1812-1819, doi:10.1093/molbev/msx116 (2017).
112 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127-128, doi:10.1093/bioinformatics/btl529 (2006).
113 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome biology* **20**, 238, doi:10.1186/s13059-019-1832-y (2019).
114 Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome biology* **16**, 157, doi:10.1186/s13059-015-0721-2 (2015).
115 De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269-1271, doi:10.1093/bioinformatics/btl097 (2006).
116 McGinnis, S. & Madden, T. L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research* **32**, W20-W25, doi:10.1093/nar/gkh435 (2004).
117 Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* **22**, 4673-4680, doi:10.1093/nar/22.22.4673 (1994).
118 Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113, doi:10.1186/1471-2105-5-113 (2004).
119 Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* **35**, 1547-1549, doi:10.1093/molbev/msy096 (2018).
120 Qiao, X. *et al.* Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome biology* **20**, 38, doi:10.1186/s13059-019-1650-2 (2019).
121 McIver, L. J. *et al.* bioBakery: a meta'omic analysis environment. *Bioinformatics* **34**, 1235-1237, doi:10.1093/bioinformatics/btx754 (2017).
122 Truong, D. T. *et al.* MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature methods* **12**, 902-903, doi:10.1038/nmeth.3589 (2015).
123 Crisp, A., Boschetti, C., Perry, M., Tunnacliffe, A. & Micklem, G. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome biology* **16**, 50, doi:10.1186/s13059-015-0607-3 (2015).

Acknowledgments

We would like to thank the staff members in the Shenzhen Key Laboratory of Allergy and Immunology of the Shenzhen University and Siriraj Dust Mite Center for Services and Research of the Mahidol University for the mite authentication and culturing work.

Author contributions

Qing Xiong: Investigation, Methodology, Conceptualization, Writing–Original Draft. Angel Tsz-Yau Wan: Investigation, Methodology, Conceptualization, Writing–Review & Editing. Xiaoyu Liu: Investigation, Data curation, Methodology. Cathy Sin-Hang Fung: Investigation, Data curation. Xiaojun Xiao: Investigation, Methodology. Nat Malainual: Investigation. Jinpao Hou: Investigation. Lingyi Wang: Investigation. Mingqiang Wang: Investigation, Methodology. Kevin Yi Yang: Investigation, Methodology. Yubao Cui: Investigation, Resources. Elaine Lai-Han Leung: Resources. Wenyan Nong: Investigation, Formal analysis. Soo-Kyung Shin: Investigation. Shannon Wing-Ngor Au: Funding acquisition, Resources. Kyoung Yong Jeong: Investigation, Writing–Review & Editing. Fook-Tim Chew: Resources. Jerome Ho-Lam Hui: Conceptualization, Writing–Review & Editing. Ting-Fan Leung: Funding acquisition, Resources. Anchalee Tungtrongchitr: Resources, Supervision, Writing–Review & Editing. Nanshan Zhong: Resources, Supervision. Zhigang Liu: Funding acquisition, Resources, Project administration. Stephen Kwok-Wing Tsui: Funding acquisition, Conceptualization, Project administration, Writing–Original Draft, Writing–Review & Editing.

Competing interests

We have no competing interest to disclose.

Materials & Correspondence

Correspondence and requests for materials should be addressed to Professor Stephen Kwok-Wing Tsui, School of Biomedical Sciences, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong. E-mail: kwtsui@cuhk.edu.hk; Tel: 852-3943 6381; Fax: 852-2603 5123.

Table 1. Overview of the statistics of genome assemblies and annotations of six astigmatic mites

	D. farinae	*D. pteronyssinus*	*P. ovis*	*S. scabiei*	*B. tropicalis*	*T. putrescentiae*
Assembly features						
Assembled genome size (bp)	60,394,945	59,034,585	63,214,126	55,647,270	63,746,680	97,387,311
BUSCO completeness	91.3%	91.0%	90.6%	89.0%	91.3%	91.5%
BUSCO duplication	1.1%	3.0%	0.8%	0.8%	1.3%	3.3%
Assembly type	scaffold	contig	contig	scaffold	scaffold	scaffold
Number of scaffolds or contigs	10	131	93	574	116	176
Maximum length	16,911,843	3,357,599	5,538,194	1,295,475	11,105,103	6,764,852
N50 length (bp)	8,981,490	963,061	2,279,290	253,843	3,687,816	2,913,131
N90 length (bp)	4,381,017	219,916	605,092	58,941	311,152	467,811
Total gap length (bp)	274,781	N/A*	N/A*	19,550	318,601	747,038
Annotation features						
Number of genes	15,457	14,999	14,688	14,763	16,869	23,793
BUSCO completeness	92.0%	91.9%	90.9%	91.1%	92.9%	90.7%
BUSCO duplication	2.7%	4.5%	1.4%	0.9%	2.4%	4.6%
Repeat content	15.79%	14.24%	18.87%	9.03%	9.54%	13.03%

* N/A, not applicable.

Table 2. HGT events identified in the six astigmatic mites

Functional category	Description	Representative	Length/aa	Best hit in NR*			Copy number					
				E-value	Identity	Taxonomy	*Df*#	*Dp*#	*Po*#	*Ss*#	*Bt*#	*Tp*#
Detoxification	UGT	DF_011775.01	430	1E-93	37.7%	Bacteria, Actinobacteria	22	25	17	5	31	73
Lysis of fungal cell wall	Chitinase 1	DF_010318.01	459	1E-107	52.1%	Bacteria, Actinobacteria	1	1	1	0	1	1
	Chitinase 2	DF_000907.01	403	5E-115	59.2%	Bacteria, Actinobacteria	1	1	1	1	1	1
	Beta-1,3 glucanase 1	DF_002571.01	343	1E-20	29.9%	Eukaryota, Slime mold	1	1	1	1	1	1
	Beta-1,3 glucanase 2	DF_014758.01	464	2E-25	31.1%	Bacteria, FCB group	1	1	1	1	1	2
	Beta-1,3 glucanase 3	DF_001158.01	280	8E-66	43.1%	Eukaryota, Slime mold	4	4	1	0	4	8
	Chitosanase	DF_009713.01	259	7E-73	43.1%	Bacteria, Firmicutes	1	2	1	1	1	1
Lysis of bacterial cell wall	Peptidoglycan endopeptidase	DF_002569.01	150	3E-33	49.3%	Eukaryota, Fungi	1	1	2	0	1	3
	Endolytic peptidoglycan transglycosylase	DF_000506.01	111	1E-25	58.2%	Bacteria, Proteobacteria	1	1	0	0	2	3
Bacterial resistance	Stress response protein	DF_013478.01	303	2E-29	49.2	Bacteria, Firmicutes	1	1	1	1	1	1
	D-alanyl-D-alanine dipeptidase	BT_004312.01	352	1E-60	44.3%	Bacteria, Proteobacteria	0	0	0	0	1	2
Terpene synthase	Terpene synthase	DF_008370.02	415	7E-16	22.5%	Bacteria, Actinobacteria	1	1	1	1	1	1
Methyltransferase	Trans-aconitate 2-methyltransferase	DF_011983.01	310	4E-15	39.7%	Bacteria, Actinobacteria	1	1	1	1	1	1
	Malonyl-ACP O-methyltransferase	DF_001833.01	298	2E-5	36.0%	Bacteria, Proteobacteria	1	1	1	1	0	1
Nucleotide metabolism	RNA 2'-phosphotransferase	DF_001392.01	190	6E-16	33.7%	Archaea, TACK group	1	1	1	1	1	1
	Deoxyribonuclease RhsA	TP_001696.01	387	1E-15	38.9%	Bacteria, Proteobacteria	0	0	0	0	0	2

* Best hit in NR database (excluding arthropods) with BLASTP.

Df, *D. farinae*; *Dp*, *D. pteronyssinus*; *Po*, *P. ovis*; *Ss*, *S. scabiei*; *Bt*, *B. tropicalis*; *Tp*, *T. putrescentiae*.

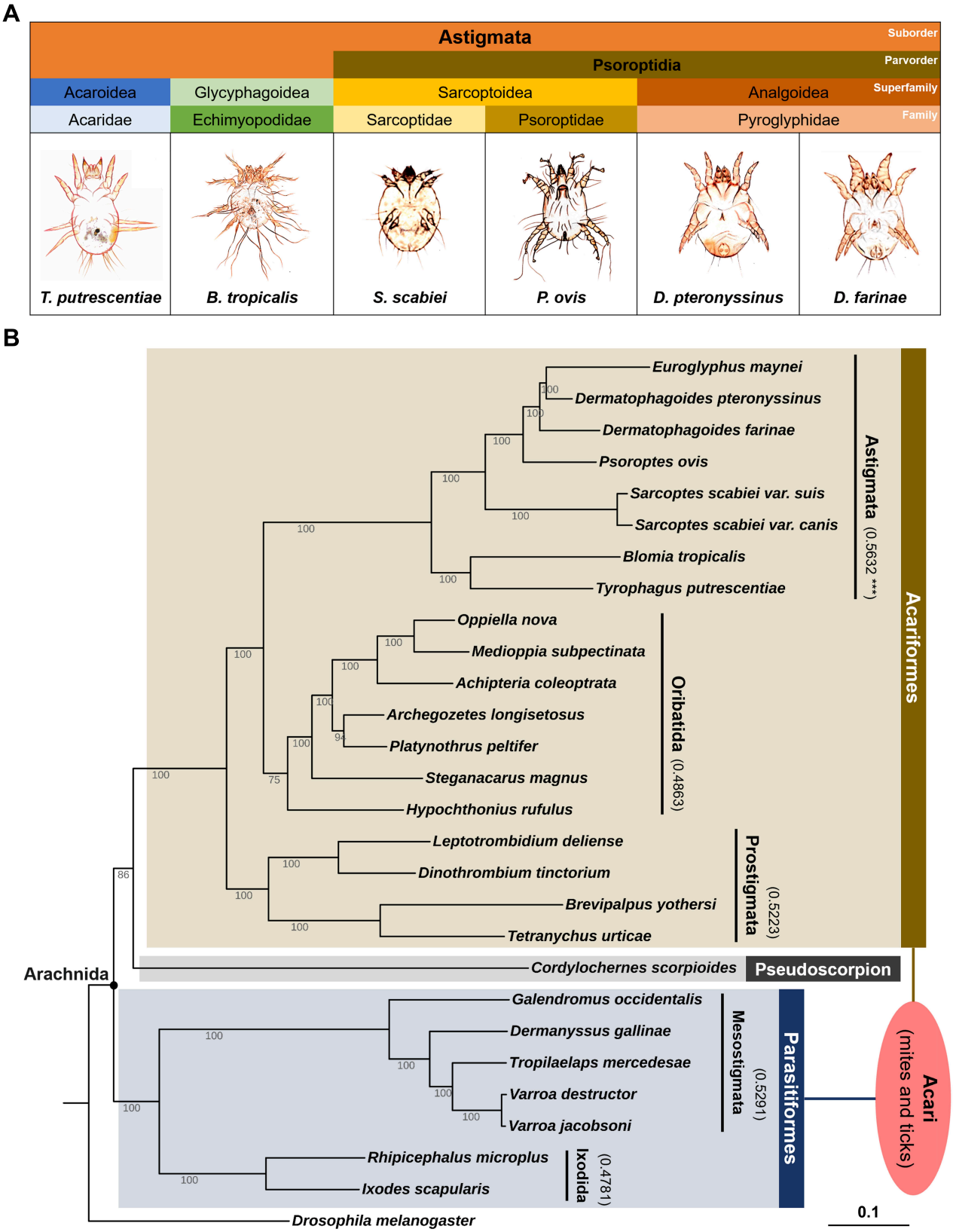

Fig. 1. Morphology of six astigmatic mites and phylogenetic tree of Acari (mites and ticks)

(A) Taxonomy and diverse morphology of six astigmatic mites (manually painted). Six astigmatic mites in diverse morphologies had high-quality genomes assembled and/or annotated in this study. Taxonomy assignment used NCBI taxonomy version. The *S. scabiei* is *var. canis,* if not specified hereafter. **(B)** The phylogenetic tree of mites and ticks (subclass: Acari). Two major groups Acariformes and Parasitiformes in the subclass Acari were interrupted by the pseudoscorpion *Cordylochernes scorpioides*. The phylogenetic tree was constructed based on the alignment of 13,133 conserved amino-acid residues in 47 overlapped single and complete BUSCO protein sequences of 28 genomes by RAxML in maximum likelihood algorithm and 100 bootstrap replicates. The mean substitution rate (MSR) of the five lineages of mites and ticks were marked. *** p<0.001.

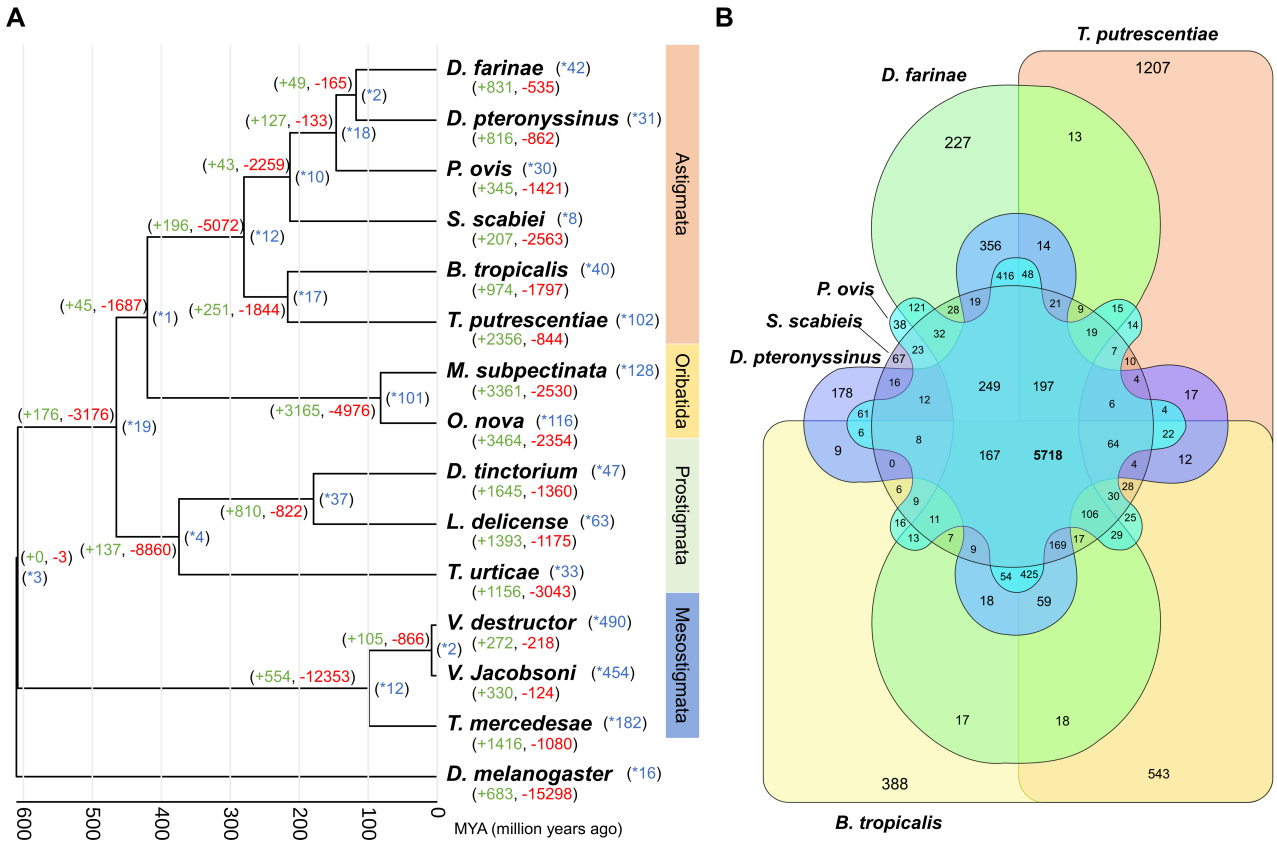

Fig. 2. Phylogenomic analysis revealed a wide range of genetic changes

(A) Analysis on gene gain or loss in the evolution of mites. The numbers in brackets on nodes indicate the number of orthogroups under expansion (+ and green) or contraction (- and red), and rapidly evolving orthogroups (* and blue). The ultrametric time tree was adapted from Fig. S4 and the proteome of *D. melanogaster* (UniProt ID: UP000000803) was used as an outgroup. **(B)** Venn diagram of orthogroups of the six astigmatic mites. Proteomes were assigned into orthogroups using OrthoFinder2 and the overlapped orthogroups of the six astigmatic mites were then presented using a Venn diagram.

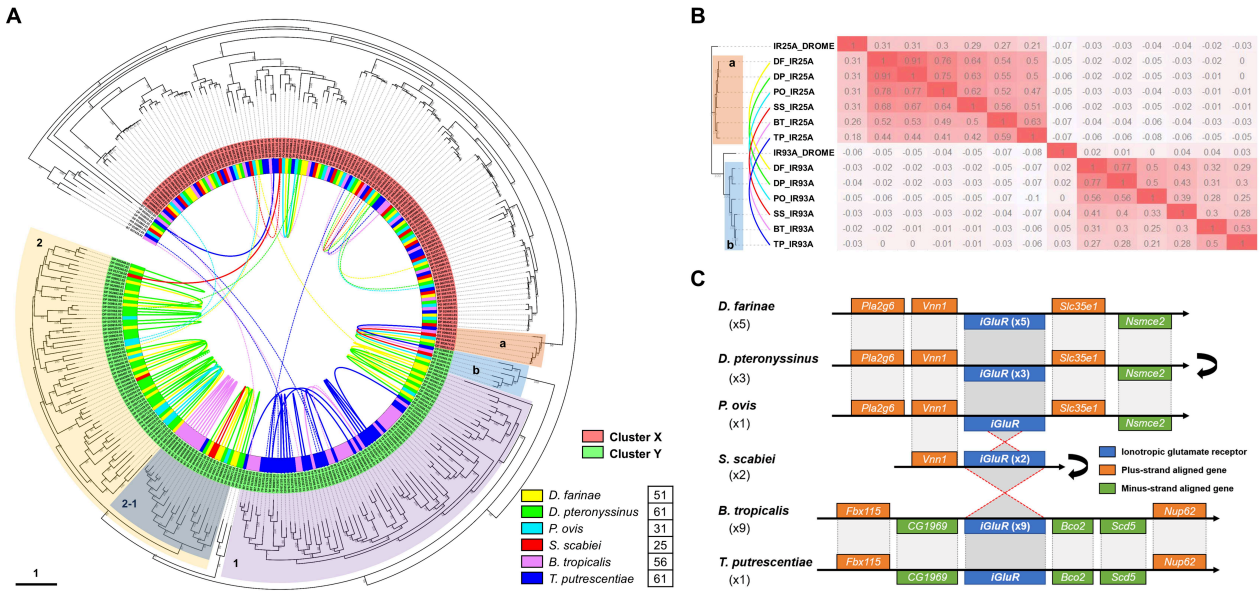

Fig. 3. Frequent tandem gene duplications of ionotropic glutamate receptors (iGluRs)

(A) Phylogenetic analysis of all iGluRs in the six astigmatic mites. All tandemly arrayed genes are connected using curved solid lines, while all proximally arrayed genes (separated by no more than 10 genes) are connected using dotted lines. **(B)** Phylogenetic analysis and similarity matrix of two adjacent iGluRs, IR25a and IR93a of the six astigmatic mites. IR25a and IR93a of *D. melanogaster* were used as references. Adjacent IR25a and IR93a were linked by solid curves in all the six mites. The similarity matrix (BLOSUM62) was generated using the online tool SIAS (Sequence Identity And Similarity) in default parameters. **(C)** Gene synteny alignment of iGluRs in subcluster 2-1. Tandem gene duplications were highlighted, such as five tandemly arrayed iGluRs in *D. farinae* are marked as iGluR (x5).

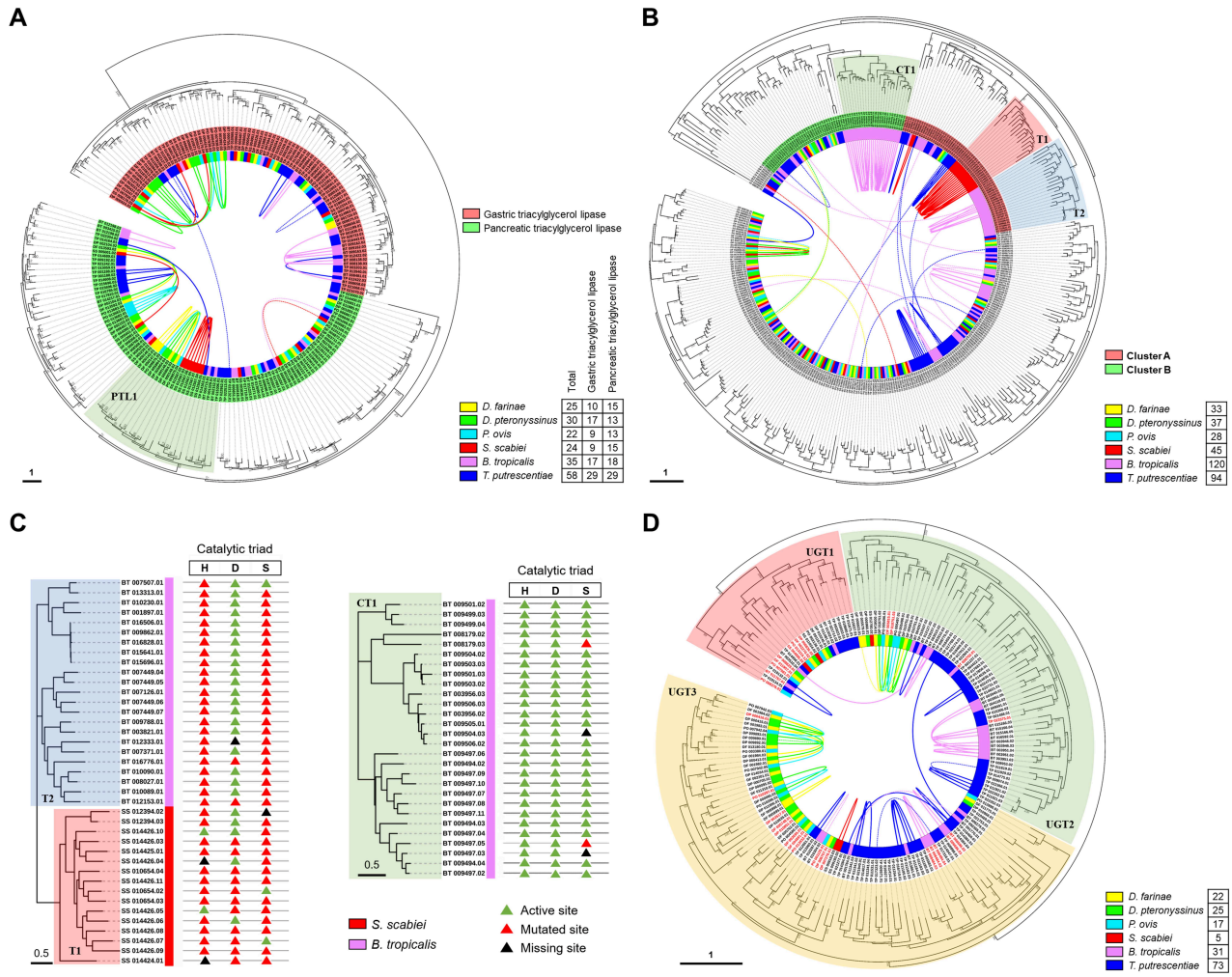

Fig. 4. Remarkable tandem gene duplications in digestion and detoxification gene families

(A) Phylogenetic analysis and tandem gene duplications of triacylglycerol lipases in the six mites. Tandemly and proximally arrayed genes of glutamate receptors in six mites were connected using the curved solid lines and dotted lines, respectively. **(B)** Phylogenetic analysis and tandem gene duplications of serine proteases. All serine proteases of the six astigmatic mites were identified and drawn into a phylogenetic tree. Three species-specific clusters (T1, T2 and CT1) of serine proteases are highlighted. **(C)** Alignment of catalytic triads of serine proteases in three species-specific clusters. Three species-specific clusters of serine proteases in Fig. 4B were collected and aligned for their catalytic triads (H, histidine; D, aspartic acid; S, serine). Active, mutated, and missing sites were marked in colored triangles. **(D)** Phylogenetic analysis and tandem gene duplications of UGTs. UGTs of the six astigmatic mites were drawn into a phylogenetic tree and divided into three large clusters.

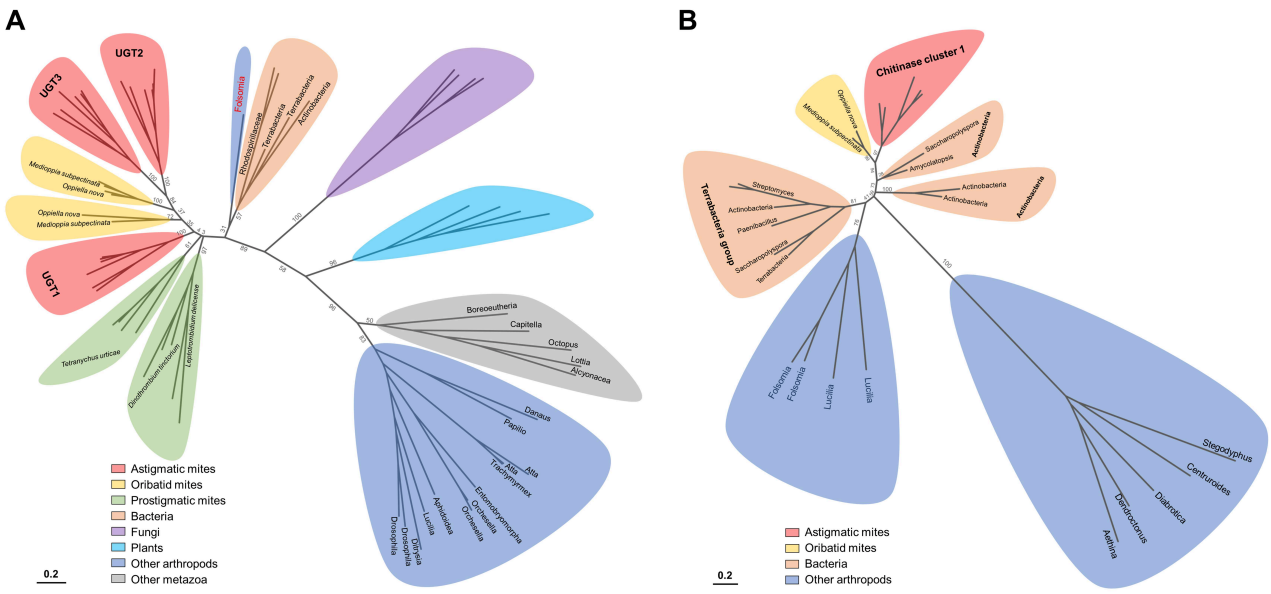

Fig. 5. Phylogenetic analysis of HGT genes, UGTs, and chitinases

(A) Phylogenetic analysis of UGTs from the six astigmatic mites. The closest UGTs in the UniRef50 database from other taxonomic categories were collected for comparison, including those from bacteria, fungi, plants, other metazoa (excluding arthropods), other arthropods (excluding mites and ticks), oribatid mites and prostigmatic mites. **(B)** Phylogenetic analysis of chitinases in cluster 1 (Fig. S11A) from the six astigmatic mites. The closest chitinases in the UniRef50 database from other taxonomic categories were collected for comparison. The phylogenetic trees were constructed by MEGA v10.2.2 with maximum likelihood (ML), 80% site coverage and 100 bootstrap replicates, and finally edited by the online tool iTOL.

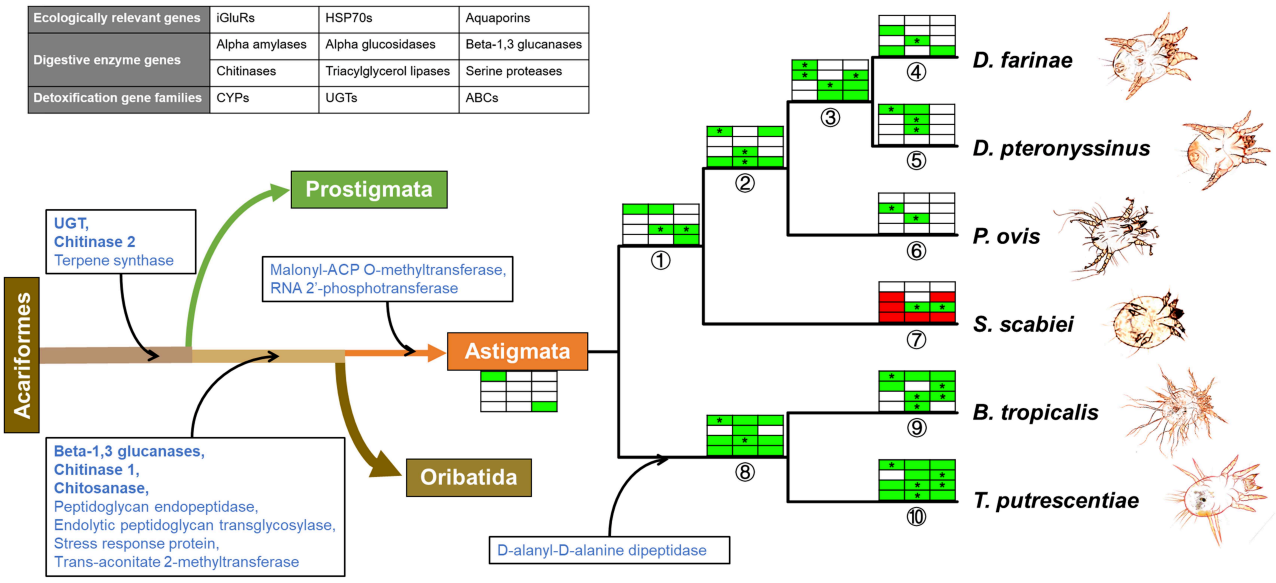

Fig. 6. Graphical illustration of the evolutionary history of astigmatic mites

In the evolutionary history, Prostigmata (e.g., two-spotted spider mite, *Tetranychus urticae*) first branched out, and then Astigmata evolved within Oribatida (known as soil mites). The blue colored genes in black boxes are the identified HGT genes. In the phylogenetic tree, colored tables show the variation events of 12 gene families in three categories (corresponding to the table in the upper left corner), in which the red box means gene loss, green box means gene gain, and star symbol * means tandem gene duplication.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementarymaterialNatureEcologyandEvolution07012022.pdf