Genomes of trombidid mites reveal novel predicted allergens and laterally-transferred genes associated with secondary metabolism

Xiaofeng Dong[1,2,3,4], Kittipong Chaisiri[4,5], Dong Xia[4,6], Stuart D. Armstrong[4], Yongxiang Fang[1], Martin J. Donnelly[7], Tatsuhiko Kadowaki[2], John W. McGarry[8], Alistair C. Darby[1†], Benjamin L. Makepeace[4†*]

[1]Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, United Kingdom, [2]Department of Biological Sciences, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China, [3]School of Life Sciences, Jiangsu Normal University, Xuzhou 221116, China, [4]Institute of Infection & Global Health, University of Liverpool, L3 5RF, United Kingdom, [5]Faculty of Tropical Medicine, Mahidol University, Ratchathewi Bangkok 10400, Thailand, [6]The Royal Veterinary College, London NW1 0TU, United Kingdom, [7]Department of Vector Biology, Liverpool School of Tropical Medicine, Liverpool L3 5QA, United Kingdom, [8]Institute of Veterinary Science, University of Liverpool, Liverpool L3 5RP, United Kingdom.

†Contributed equally.

*Correspondence address: Ben Makepeace, Department of Infection Biology, Institute of infection & Global Health, Liverpool Science Park IC2, 146 Brownlow Hill, Liverpool L3 5RF, United Kingdom. Tel: +44 151-7941586; Fax: +44 151 7950236; E-mail: blm1@liv.ac.uk

Abstract

Background

Trombidid mites have a unique lifecycle in which only the larval stage is ectoparasitic. In the superfamily Trombiculoidea ("chiggers"), the larvae feed preferentially on vertebrates, including humans. Species in the genus *Leptotrombidium* are vectors of a potentially fatal bacterial infection, scrub typhus, which affects 1 million people annually. Moreover, chiggers can cause pruritic dermatitis (trombiculiasis) in humans and domesticated animals. In the Trombidioidea (velvet mites), the larvae feed on other arthropods and are potential biological control agents for agricultural pests. Here, we present the first trombidid mites genomes, obtained both for a chigger, *Leptotrombidium deliense*, and for a velvet mite, *Dinothrombium tinctorium*.

Results

Sequencing was performed using Illumina technology. A 180 Mb draft assembly for *D. tinctorium* was generated from two paired-end and one mate-pair library using a single adult specimen. For *L. deliense,* a lower-coverage draft assembly (117 Mb) was obtained using pooled, engorged larvae with a single paired-end library. Remarkably, both genomes exhibited evidence of ancient lateral gene transfer from soil-derived bacteria or fungi. The transferred genes confer functions that are rare in animals, including terpene and carotenoid synthesis. Thirty-seven allergenic protein families were predicted in the *L. deliense* genome, of which nine were unique. Preliminary proteomic analyses identified several of these putative allergens in larvae.

Conclusions

Trombidid mite genomes appear to be more dynamic than those of other acariform mites. A priority for future research is to determine the biological function of terpene synthesis in this taxon and its potential for exploitation in disease control.

41 **Keywords**

42 Chigger, trombiculid, scrub typhus, terpenes, isoprenoids, horizontal transfer, *Leptotrombidium*,

43 *Dinothrombium*, *Tetranychus,* Trombidiformes.

Background

The Acari (mites and ticks) is the most speciose group within the subphylum Chelicerata, with approximately 55,000 described species in both terrestrial and aquatic habitats, and an estimated total diversity of up to 1 million species [1]. This assemblage is paraphyletic and is composed of two major divisions, the Parasitiformes and the Acariformes, which both contain species of medical, veterinary and agricultural importance. For example, the Parasitiformes harbours predatory mites used in the control of agricultural pests (*e.g.*, *Metaseiulus occidentalis*); ectoparasites of honey bees (*Varroa destructor* and *Tropilaelaps mercedesae*) that transmit pathogenic viruses; and most famously, the ticks (Ixodida). The Acariformes includes major ectoparasites and sources of allergens for humans and other animals, such as the scabies mite (*Sarcoptes scabiei*) and the dust mites (*Dermatophagoides* spp. and *Euroglyphus maynei*).

The role of lateral gene transfer (LGT) in the evolution of animals has remained controversial since the first metazoan genomes were sequenced. Despite the explosion of new genomic resources across a wider variety of metazoan phyla in recent years, some claims of large-scale LGT in animal genomes have been shown to be the result of flawed data-analysis methods that failed to exclude sequences from bacterial contaminants [2-5]. Attempts to conduct meta-analyses across diverse metazoan genomes using consistent criteria have also been criticized for reliance on unsound assumptions [6, 7]. However, it is irrefutable that several metazoan taxa are reliant on functions obtained via LGT for essential physiological processes, including digestion of complex and/or toxic materials (especially in the case of herbivores and plant parasites) and the avoidance of host defences [8-11]. One group of acariform mites, the spider mites (superfamily Tetranychoidea: order Trombidiformes; Fig. 1), are major pests of various crops and rely on well-characterised lateral gene transfers of β-cyanoalanine synthase (from bacteria) and carotenoid biosynthesis enzymes (from fungi) to detoxify hydrogen cyanide in their diet [12] and to control diapause [13], respectively. Currently, it is unknown whether

LGT is a key feature of the evolution of the spider mites only, or is more widespread among the Trombidiformes.

In addition to the Tetranychoidea, the Trombidiformes contains two other superfamilies of economic or clinical importance, the Trombidioidea (the velvet mites) and Trombiculoidea (the "chiggers" and related groups) [14]. These two taxa, known collectively as trombidid mites (Fig. 1), have a unique natural history among arthropods in that only the larval stage is ectoparasitic, whereas the nymphs and adults are predators of other arthropods (Fig. 2). However, the Trombidioidea and Trombiculoidea differ in their host preferences. Larvae of the Trombidioidea are exclusively parasites of other arthropods and some species feed on insects of medical, veterinary and agricultural importance, including mosquitoes [15], the New World screwworm fly [16], and aphids [17] (Fig. 2). On dipteran hosts, heavy infestations can reduce flight ability, while certain aphid species can be killed in a few days by as few as two feeding larvae [18, 19].

The more heavily-studied larval stages of the Trombiculoidea (commonly referred to as chiggers or berry bugs) primarily feed on terrestrial vertebrates, including humans [20], although some little-known taxa in this superfamily are ectoparasites of invertebrate hosts in common with the Trombidioidea [21-23] (Fig. 2). Importantly, the only major mite-transmitted disease of humans, scrub typhus or tsutsugamushi disease, is vectored by chiggers in the genus *Leptotrombidium* [24], while other chigger genera have only been implicated epidemiologically as locally-important vectors [25]. Scrub typhus is a severe febrile illness caused by infection with an obligate intracellular bacterium (*Orientia* spp.) in the order *Rickettsiales* and features an epidemiological cycle that includes wild small mammals, which are the primary hosts for many chigger species [26]. This disease has a fatality rate of 6% if not treated promptly with antibiotics [27] and has increased in incidence globally in recent years, with the annual minimum incidence reaching >17/100,000 in South Korea and Thailand in 2012 – 2013 [28]. In the so-called "tsutsugamushi triangle" within the Asia-Pacific region, scrub typhus has a median seroprevalence of 22.2% [28]; but endemic scrub typhus has emerged in several other parts

of the world within the past decade, including South America [29], the Middle East [30], and possibly sub-Saharan Africa [31]. Chiggers have also been implicated in the transmission of hantaviruses [32], *Bartonella* spp. [33] and *Rickettsia* spp. [34]. Furthermore, chiggers have direct impacts worldwide by causing trombiculiasis, which is a highly pruritic dermatitis that can afflict humans, companion animals and domestic ruminants, potentially leading to severe hypersensitivity [35-38].

A remarkable second unique feature of trombidid mites is that the larvae induce the formation of a feeding tube or "stylostome" at the attachment site that is extraneous to the larval mouthparts [39]. These larvae are not blood feeders, but ingest tissue exudates (in the case of vertebrate hosts) or arthropod haemolymph [40]. The life history of trombidid nymphs and adults has been poorly studied. However, in the Trombiculoidea, the eggs of Collembola (springtails) and other arthropods are an important part of the diet [41] (Fig. 2). Arthropod eggs may also serve as food items for adults and nymphs in the Trombidioidea [42], although some species have potential roles in biological control, as they feed on pest arthropods such as spider mites, scale insects, aphids and termites [19, 43-45].

To date, research on trombidid mites has suffered from a dearth of molecular data that could facilitate studies on speciation; population structure; host-vector and vector-pathogen interactions; and life-history evolution in this group. To address this deficit, here we present a comparative analysis of the genomes of the chigger *Leptotrombidium deliense* (the primary scrub typhus vector in South-East Asia [46]) and the giant red velvet mite, *Dinothrombium tinctorium* (the world's largest acarine species [47]). We show that these trombidid mites form a distinct branch of the Trombidiformes that exhibit two classes of LGT for secondary metabolism: the previously identified carotenoid biosynthesis enzymes of fungal origin, and much larger terpene synthase gene families, which probably derive from soil-associated bacteria. We also identify unique clusters of predicted allergens in *L. deliense* that may contribute to the symptoms of trombiculiasis in humans and domestic animals.

Data description

Since the specimens were highly disparate in physical size (adult *D. tinctorium* can reach ~16 mm in length, whereas larval *L. deliense* do not normally exceed 250 µm, even when engorged), a tailored approach to sequencing was necessary in each case. For the velvet mite, DNA from a single adult was used to generate two Illumina TruSeq libraries (insert sizes, 350 bp and 550 bp) and one Nextera mate-pair library (insert size, 3 Kb). The TruSeq libraries were barcoded, indexed and paired-end (PE)-sequenced (2 × 100 bp) on one lane, and the Nextera library was PE-sequenced (2 × 250 bp) on an additional lane, both on the Illumina MiSeq platform. For the chiggers, DNA from a pool of engorged larvae (obtained from Berdmore's ground squirrels in Thailand) was used to produce one NEB Next Ultra DNA library (insert size, 550 bp) and PE-sequenced (2 × 150 bp) on the Illumina MiSeq.

The total number of trimmed reads generated was ~362 million for *D. tinctorium* and ~38 million for *L. deliense.* For the former, PE reads were assembled using Abyss (v. 1.5.2) [48, 49]. For the chigger data, a preliminary assembly to contig level was performed using Velvet (v. 1.2.07) [50]. Reads derived from mammalian host genomic DNA were removed from the preliminary genome assembly using blobtools (v0.9.19) , which generates a GC-coverage plot (proportion of GC bases and node coverage) [51] (Fig. 3). The *L. deliense* genome was then reassembled using SPAdes assembler (v. 3.7.1) [52] with default settings. For gene calling, the MAKER pipeline [53] was used to integrate *ab initio* gene predictions from Augustus (v. 3.2.2) [54], SNAP (v. 2013-11-29) [55] and GeneMark (v. 2.3e) [56] with evidence-based gene models. The “Methods” section provides more details on how downstream genome analyses were performed.

Similarly to the genome sequencing strategy, proteomic analysis of the trombidid mites was customised to the sample types available. A single adult *D. tinctorium* was subjected to protein extraction in SDS buffer and tryptic digestion using the filter-aided sample preparation method [57]. The digested sample was split into eight fractions using a High pH Reversed-Phase Peptide Fractionation Kit (Pierce) prior to nanoLC MS ESI MS/MS analysis on a Q-Exactive mass spectrometer

(Thermo Fisher Scientific). A total of 137,638 spectra were generated across the eight fractions. For *L. deliense*, a soluble protein extract was obtained from a small pool of ethanol-fixed engorged larvae (*n* = 10) collected from several species of wild rodents in Thailand [26]. Following tryptic digestion, downstream analyses proceeded as for *D. tinctorium*, producing 18,059 spectra. The "Methods" section provides details on how MS spectra searches and Pfam enrichment analyses were performed.

Analyses

Genome statistics and phylogenomics

Assembled genome sizes were 180.41 Mb for *D. tinctorium* and 117.33 Mb for *L. deliense* (Additional file 1), whereas *k*-mer analysis placed the genome size estimates much closer together but was of a similar scale to the assemblies (143.52 – 147.09 and 158.31 – 160.95 Mb, respectively; Additional file 2). The repeat content of these genomes presented a significant challenge, with unclassified repeats alone accounting for 19 - 23% of the total size; approximately double the proportion of the *S. scabiei* [58] and *Tetranychus urticae* (two-spotted spider mite [59]) genomes (Additional file 3). As expected from sequencing a pool of *L. deliense* larvae compared with a single adult of *D. tinctorium*, and the corresponding DNA library strategies employed, the chigger genome was considerably less contiguous than that of the velvet mite. Nevertheless, the estimated completeness of the predicted gene set for *L. deliense* based on the Benchmarking Universal Single-Copy Orthologues (BUSCO) criteria [60] compared favourably with that of other arachnid genomes (Additional file 1). Notably, the *D. tinctorium* genome contained the greatest number of protein-coding genes from the Acariformes sequenced to date (Additional file 1).

Maximum likelihood and Bayesian phylogenomic analyses based on >500 one-to-one orthologous genes exhibited complete concordance, placing *L. deliense* and *D. tinctorium* together as sister taxa, and *T. urticae* as their closest relative amongst sequenced species (Fig. 1, Fig. 4). Our divergence time estimates accord closely with those previously published for arachnids [61], with the Parasitiformes and Acariformes separating approximately 416 million years ago (MYA) (Fig. 4). We estimate that the trombidid mites *sensu stricto* (velvet mites and chiggers) diverged from the phytophagous Tetranychoidea 250 MYA, and finally the Trombidioidea and Trombiculoidea last shared a common ancestor approximately 123 MYA (Fig. 4).

Gene family expansions

When gene families were compared among the Acariformes and reference invertebrate genomes, the *D. tinctorium* genome was shown to contain a substantially greater number of unique paralogous groups than the other sequenced acariform mites, while only 68 gene clusters were shared among all Acariformes (Additional file 4). Moreover, the *D. tinctorium* genome displayed a greater number of multi-copy ("N:N:N"), patchy, and arachnid-specific orthologues when analysed alongside the published arachnid genomes (Fig. 4). The gene family expansion in *D. tinctorium* also dwarfed that seen among other members of the Arachnida, including *L. deliense* (Additional file 5).

Relative to other arachnids, *D. tinctorium* exhibited a significant expansion of 38 gene families, including a large family (ORTHOMCL98) of uncharacterised proteins containing 45 members in this species but no representatives in the other arachnid genomes (Additional file 6). Examination of conserved domains in this gene family revealed a major facilitator superfamily domain with some weak but significant similarity by BLASTP (~25% amino-acid identity, >95% coverage) to feline leukemia virus subgroup C receptor-related protein (FLVCR)-1 from various Metazoa. Two other orthologous clusters displayed an identity of up to 30% with FLVCR2 and showed a statistically significant expansion in the velvet mite, containing 27 (ORTHOMCL265) and 20 (ORTHOMCL412) members, compared with only one and three members in *L. deliense*, respectively (Additional file 6). Moreover, a gene family annotated as 4-coumarate:CoA ligases was significantly expanded in *D. tinctorium*, with 27 members; while *L. deliense* also showed a large repertoire (14 members) relative to other arachnids, although this was not statistically significant (Additional file 6). This gene family represents a group of long-chain fatty acyl-CoA synthetases related to firefly luciferase (but without luciferase activity) that are found in multicellular eukaryotes other than vertebrates [62]. These enzymes are associated with peroxisomes and perform catabolism of very long-chain fatty acids by β-oxidation.

Only a single gene, SPARC-related modular calcium-binding protein 1 (SMOC1) from *L. deliense*, was predicted to be under positive selection in either of the trombidid genomes with strong statistical confidence (d_N/d_S = 1.0663). The SMOC genes have been little studied in invertebrates, but in

196 vertebrates they encode for matricellular proteins of the basement membranes and are involved in
197 bone morphogenesis and ocular development [63]. The association with morphological development
198 is conserved in *Drosophila,* in which the SMOC orthologue *pentagone* is expressed in the wing imaginal
199 discs [64].

200 Overrepresented protein families detected by LC-MS/MS

201 From the extremely small sample of engorged *L. deliense*, 522 mite proteins were identified by
202 shotgun LC-MS/MS with at least one unique peptide against a background of 292 proteins of putative
203 host origin (Additional file 7). Of the *L. deliense* proteins, 290 were considered as high-confidence
204 identifications (≥2 unique peptides) and were subjected to Pfam domain enrichment analysis. The
205 most overrepresented protein domains were derived from ATP synthase (PF00006 and PF02874) and
206 both muscle and non-muscle isoforms of myosin or paramyosin (PF01576), although key enzymes of
207 the citric acid cycle (ATP citrate synthase and succinate-CoA ligase, PF00549) were also well
208 represented (Fig. 5). Several other proteins with probable origins from muscle tissue contained
209 calponin homology (CH) domains (PF00307) and/or spectrin repeat domains (PF00435), including
210 muscle protein 20 [65], myophilin, actinin, and spectrin subunits (Fig. 5). Myophilin is an invertebrate-
211 specific protein that has previously been characterised as an immunogenic muscle component from
212 parasitic platyhelminths [66], suggesting that it may contribute to the inflammatory response during
213 trombiculiasis.

214 Proteomic analysis of a single adult *D. tinctorium* led to the robust identification (≥2 unique peptides)
215 of 1,636 proteins, as the large size of the specimen was conducive to peptide fractionation prior to LC-
216 MS/MS (Additional file 8). Laminin B (PF00052) was the most enriched domain (Fig. 5), which was
217 present in several isoforms of perlecan (the basement membrane-specific heparan sulfate
218 proteoglycan core protein). Perlecan is highly expressed during embryonic development in *Drosophila*
219 [67], suggesting that this mite specimen may have contained fertilised eggs. Indeed, this specimen
220 was undoubtedly a female, as it contained vitellogenins (yolk proteins) (Additional file 8). Lyase-1

domains (PF00206) were also highly enriched (Fig. 5) and were found in mitochondrial fumarate hydratase (an enzyme of the citric acid cycle) and in adenylosuccinate lyase of the purine nucleotide cycle. The peptidase M20 (PF01546) and M20 dimer domains (PF07687) were both present in cytosol non-specific dipeptidases, which have a key role in protein digestion in the midgut [68], and in N-fatty-acyl-amino acid synthase-hydrolases, which regulate thermogenesis via uncoupled respiration [69]. Another group of proteins with a putative role in thermoregulation were the α-crystallin-like small heat shock proteins (PF00525); in ticks, these are highly immunogenic proteins expressed in tick salivary glands and exhibit thermoprotective activity [70]. Finally, the inhibitor I29 domain (PF08246) represented digestive cysteine proteinases [71] and cathepsin L-like proteases (Fig. 5), which are highly expressed in feeding stages of *T. urticae* [72] and are a major protein component of spider mite faeces [73].

Lateral gene transfers and mobile elements

To determine the origin of the bright colouration of the trombidid mites, we searched both genomes for the fused carotenoid synthases-cyclases that were reported to have been laterally transferred into the *T. urticae* genome from zycomycete fungi, perhaps via aphids [59]. In common with *T. urtic*ae, two of these carotenoid synthases-cyclases were observed in the *L. deliense* genome, while *D. tinctorium* harboured 12 copies (Additional file 9). However, the single largest gene family expansion observed in the *L. deliense* genome was within an orthologous cluster annotated as “pentalenene synthase”, which contained 41 members (Additional file 6). This cluster (ORTHOMCL47) also contained 21 genes in the *D. tinctorium* genome, but lacked orthologues in the genomes of other arachnids. A second orthologous cluster (ORTHOMCL881) of terpene synthases contained 17 members and was unique to *L. deliense* (Additional file 6).

The capacity to generate terpenoids (also known as isoprenoids) *de novo* in metazoans is extremely unusual. While some millipedes (for instance, the Japanese species *Niponia nodulosa*) are known to produce terpenes such as geosmin and 2-methylisoborneol in defensive secretions, these secondary

metabolites are assumed to be derived from microbial symbionts [74]. The absence of terpene synthases in the fully sequenced diplopod genome from the rusty millipede, *Trigoniulus corallinus*, certainly supports this interpretation [75]. Dust mites also produce a monoterpene, neryl formate, which has been demonstrated to act as an aggregation pheromone [76]. However, BLAST analysis of the *Dermatophagoides farinae* genome assembly [77] using the trombidid terpene synthases failed to identify significant homologues, suggesting that dust mites rely on microbial symbionts for terpene production, or that terpene synthases in mite genomes are evolving too rapidly to be identified by homology searches. To the best of our knowledge, the only animals known to harbour terpene synthase genes in their nuclear genomes are a very restricted number of beetle species [predominantly flea beetles of the subfamily Galerucinae, which produce (6*R*,7*S*)-himachala-9,11-diene as a male aggregation pheromone [78]] and the collembolan *Folsomia candida*, in which the metabolites produced and their function are unknown [79]. Whilst terpene synthases are widespread in plants, fungi and bacteria, the terpene synthases from flea beetles do not resemble those from non-metazoan taxa and appear to have evolved from arthropod *trans*-isoprenyl diphosphate synthases [78]. Moreover, the terpene synthases of *F. candida* are more similar to those from the flea beetles than they are to the non-metazoan enzymes [79].

We generated phylogenies for the trombidid terpene synthases, which clearly showed close affinities to bacterial and fungal homologues and not to those from other arthropods (Fig. 6, Fig. 7). For ORTHOMCL47, the *L. deliense* and *D. tinctorium* enzymes formed distinct groups, and the closest homologues from other taxa included a monoterpene synthase from *Micromonospora* spp. (phylum Actinobacteria), a genus that is known to synthesise 2-methylenebornane [80], as well as related genes from agaricomycete fungi (Fig. 6). In the case of the chigger-specific ORTHOMCL881, the nearest homologues were distributed among Actinobacteria and other bacterial phyla, including the Chloroflexi, Proteobacteria and Bacteroidetes (Fig. 7). These terpene synthases were clearly separated from the flea beetle proteins and consisted predominantly of germacrene or geosmin synthases (Fig. 7). However, for both clusters of trombidid terpene synthases, the amino acid identity with their

nearest bacterial or fungal homologues was low (≤30%). Nevertheless, the majority fulfilled the criteria of Crisp *et al*. [6] as high-confidence ("class A") lateral gene transfers due to the absence of sufficiently closely-related homologues in other Metazoa.

To exclude the possibility that the trombidid terpene synthases were contaminating sequences of bacterial or fungal origin from the environment, or derived from microbial symbionts, we examined the genomic context of each terpene synthase to determine if they were sometimes found adjacent to an incontrovertible metazoan gene. Due to the low contiguity of the *L. deliense* assembly, it was not possible to find other genes on the same contig as a terpene synthase. However, one of the members of ORTHOMCL47 in *D. tinctorium* was located 5.2 kb downstream of a gene with a top BLAST hit to a translational elongation factor-2 mRNA from *Dinothrombium pandorae* [81] (Additional file 10). Furthermore, blob-plot analysis revealed that the GC content and read coverage of the contigs containing terpene synthases lay close to the overall mean for both trombidid genomes (Fig. 8). Importantly, sequences of unambiguous bacterial origin were very rare in both genomes (Additional file 11), with no evidence for a high-titre symbiont or environmental contaminant that may have impacted significantly on the genome assemblies. Of these candidate laterally-transferred genes in the trombidid mites, expression at the protein level could not be detected in the small chigger sample. However, in *D. tinctorium*, a high-confidence identification of a single terpene synthase from ORTHOMCL47 was achieved based on two unique peptides, while a third peptide was shared with an additional terpene synthase (Additional file 12).

In addition to these lateral gene transfers, the trombidid mite genomes exhibited further evidence for dynamism in the form of endogenous retroviruses (ERVs). The *D. tinctorium* genome showed significant expansions of reverse ribonuclease integrases and Pol polyprotein-like genes, whereas in the *L. deliense* genome, a 23-member family of Gag polyprotein-like genes was apparent (Additional file 6). Interestingly, the closest homologues of the Gag-like polyproteins in *L. deliense* were found in rodents, bats, lagomorphs, small carnivores, and colugos (a taxon restricted to South-East Asia [82])

(Fig. 9); all of which are known or likely hosts for chigger mites. Unfortunately, the low contiguity of the *L. deliense* genome and the metazoan context of ERVs (with similar GC content to the host) militated against bioinformatic attempts to exclude the possibility of an origin from host contamination; *i.e.*, from squirrel-derived cells on mite mouthparts or in the gut.

In the *D. tinctorium* genome, most members of the two ERV protein families each clustered in a monophyletic clade, suggesting expansion within the mite genome from a single origin (Additional files 13 and 14). The closest homologues of the Pol-like polyproteins were found mainly in other arthropods (especially cladocerans and ticks) and more distantly in fungi (Additional file 14); whereas the reverse ribonuclease integrases were most similar to those from ants, moths, nematodes and other mites (Additional file 13). Thus, endogenous retroviruses in *D. tinctorium* might originate from larval hosts, prey or soil microorganisms, although the fact that some velvet mites feed on other Acari [18] renders phylogenetic analyses of potential lateral gene transfers especially problematic. Uniquely among chelicerate genomes sequenced to date, the *D. tinctorium* genome also contained hepatitis D ribozyme-like genes (Rfam RF01787; Additional file 15), which in *Anopheles* mosquitoes, have been suggested to be involved in processing of non-LTR retrotransposons [83].

Immune system

Many Acari act as vectors of plant or animal pathogens and their life histories expose them to a multitude of microorganisms in their diets and in the environment. Thus, how they interact with pathogens and commensals via their immune system is likely to be a critical aspect determining their success as a group. The canonical humoral immune response gene networks in *Drosophila* are the Toll signalling pathway (responding to β-1,3-glucans from fungi and lysine-type peptidoglycan from Gram-positive bacteria) and the immune deficiency (IMD) pathway (responding to diaminopimelic acid-type peptidoglycan from Gram-negative bacteria). These pathways are activated when upstream transmembrane receptors [peptidoglycan recognition proteins (PGRPs) and β-glucan recognition proteins] bind to the pathogen-derived molecules [84]. Recently, the expanding number of arthropod

genomes from outside the class Insecta has highlighted key disparities in the immune pathway genes between the Pancrustacea (Hexapoda and Crustacea) *versus* the Chelicerata and Myriapoda. The most striking difference pertains to the IMD signalling pathway, which was thought to be absent in chelicerates [85]. However, genomic analyses and experimental data from *Ixodes scapularis* have revealed an alternative IMD pathway, in which interactions between IMD and Fas-associated protein with a death domain (both absent in the tick) are complemented by a E3 ubiquitin ligase (X-linked inhibitor of apoptosis protein, XIAP) and its ligand, the E2 conjugating enzyme Bendless [86]. This pathway recognises bacterial-derived lipids and restricts the growth of *Anaplasma phagocytophilum* and *Borrelia burgdorferi* in ticks. Although these data indicate that ticks (and perhaps other Parasitiformes) have a parallel IMD pathway distinct from that characterised in insects, we were unable to identify an XIAP homologue in acariform mites, including the trombidid genomes.

There are two non-exclusive scenarios that can be postulated to explain the apparent absence of an IMD pathway in acariform mites. The first is that these taxa might use the Toll pathway to respond to Gram-negative bacteria as well as Gram-positive bacteria and fungi. Indeed, crosstalk and synergistic immune responses to individual pathogens in *Drosophila* indicate that the two pathways are functionally interconnected even in insects [84], and the IMD pathway may have become redundant during the evolution of acariform mites. Second, expansions in other gene families associated with the immune response may provide alternative pathogen recognition and signalling pathways to tackle Gram-negative bacterial infections. This second scenario is supported in the trombidid mite genomes by large repertoires of Dscam genes (Additional file 16), which have previously been described to have undergone expansions in the Chelicerata and Myriapoda compared to the Pancrustacea [85]. In insects, Dscam is involved in phagocytosis of bacteria by haemocytes, and the *D. melanogaster* Dscam-hv gene exhibits a remarkable capacity to generate >150,000 alternatively-spliced isoforms, perhaps conferring some level of specificity to the insect immune response [although this remains highly controversial [87]]. Even relative to other acarine genomes, those of the trombidid mites display

a substantially greater complement of Dscam genes (~40 in *D. tinctorium*; Additional file 16), rivalling the 60 gene family members observed in the *Strigamia maritima* (coastal centipede) genome [85, 88].

Several other expanded gene families in one or both of the trombidid mite genomes may have roles in the immune response. In common with *I. scapularis*, these genomes lack the transmembrane PGRPs that are activated in the presence of peptidoglycan in insects, but contain several other PGRP genes with putative extracellular or intracellular roles. However, these soluble PGRP genes are present in larger numbers in the trombidid mite genomes than in those of *I. scapularis* and *T. urticae* (Additional file 17). Since soluble PGRP fragments can have a co-receptor function as shown in insects [89], they might work in concert with as-yet-unidentified components of the acarine immune system to recognise pathogens. This is particularly important in the case of chiggers, as evidence for a peptidoglycan-like structure has recently been reported for *Orientia tsutsugamushi* [90]. Moreover, a much larger expansion in a second class of proteins with putative roles in the immune system, the C-type lectin domain (CTLD) proteins, was apparent in *L. deliense* (Additional file 6; Table 1). The CTLD protein family is a large and diverse group, most members of which do not bind carbohydrates and are thus not lectins [91]. If a CTLD protein does have lectin activity, the carbohydrate-recognition domain usually contains the amino acid motif “WND”, together with “EPN” if the specificity is for mannose, and QPD if the specificity is for galactose. However, several exceptions to this pattern do exist [91]. The expanded *L. deliense* CTLD proteins belong to four orthologous groups containing a total of 91 genes, of which one cluster (ORTHOMCL880) lacks any signatures of carbohydrate-binding activity (Table 1). The other three groups mainly contain proteins with EPN motifs, suggesting specificity for mannose, although a small proportion of QPD-motif CTLD proteins were apparent in two of the clusters, which might bind galactose (Table 1). The majority of the *L. deliense* CTLD proteins that were predicted to bind carbohydrates exhibited classical or internal secretion signatures, while only a small proportion (10 – 20%) contained transmembrane domains (Table 1). In common with many members of the CTLD protein family, including those in other arthropods, *N*-glycosylation sites were predicted in 10 – 20% of the *L. deliense* CTLD proteins [92] (Table 1).

A final orthologous cluster with a potential role in innate immunity and which has undergone a significant expansion in *L. deliense* was annotated as "double-stranded (ds)RNA-activated kinase" (Additional file 6) or protein kinase R (PKR), an interferon-inducible enzyme that can block viral protein synthesis following binding to dsRNA substrates [93]. However, canonical PKR is restricted to mammals [94], and the 13 proteins from this cluster in the *L. deliense* genome lack a dsRNA-binding domain (InterPro identifier: IPR014720). Unexpectedly, the closest homologues to the PKR-like proteins in *L. deliense* were not from other arthropods but from a variety of different eukaryotic taxa, including protists and fungi (Additional file 18). However, the sequence similarity between the *L. deliense* genes and homologues in other Metazoa (principally in cnidarians and mammals) was sufficient to prevent fulfilment of the Crisp *et al*. [6] criteria for lateral gene transfer. Notwithstanding the absence of RNA-binding domains, these *L. deliense* proteins resemble PKR-like endoplasmic reticulum kinase (PERK), widespread in eukaryotes, which is a key component of the unfolded protein response during periods of endoplasmic reticulum stress [95].

Photoreceptor and chemosensory systems

Unlike insects, chelicerates lack compound eyes. Mites and ticks may be eyeless, or can possess one or more pairs of simple dorsal ocelli. The Parasitiformes sequenced to date are all eyeless species (*I. scapularis*, *M. occidentalis* and *T. mercedesae*), whereas the trombidid mites and *T. urticae* have two pairs of ocelli on the prodorsum in the adult stage ([47]. However, the genomes of both eyeless and eyed Acari exhibit a variable complement of opsins, which in combination with the chromophore retinal, form light-sensitive proteins termed rhodopsins. The genomes of eyeless ticks and mites, as well as that of *T. urticae*, contain one or more genes of the "all-*trans*-retinal" peropsin class, which in spiders have been shown to encode non-visual photosensitive pigments with combined G-protein coupled receptor and retinal photoisomerase activity [96]. Since even the eyeless species show evidence for reproductive and diapause behaviours that respond to day-length, it has been suggested that peropsins are important for the maintenance of circadian rhythms [97, 98]. Notably, we found no

398 evidence of peropsin genes in the trombidid mite genomes, but did find orthologues of *T. urticae*
399 rhodopsin-1 and -7 in both *L. deliense* and *D. tinctorium* (Fig. 10). In the velvet mite, an additional four
400 rhodopsin-7-like paralogues were apparent, three of which were identical at the amino-acid level (Fig.
401 10).

402 In contrast with insects but in common with crustaceans and myriapods, the Acari appear to have a
403 scant repertoire of chemosensory protein classes, lacking both odorant-binding proteins (OBPs) and
404 odorant receptors. Moreover, the small chemosensory proteins that have expanded considerably in
405 some insect orders (especially Lepidoptera [99]) are completely absent in the mite genomes, although
406 a gene encoding one such protein was identified in the *I. scapularis* genome (Table 2). Thus, mites rely
407 primarily on gustatory and ionotropic receptors for chemosensation. The repertoire of gustatory
408 receptors (GRs) in *L. deliense* (42 members) and *D. tinctorium* (105 members) was in a similar range
409 to most mites and ticks (albeit from the Parasitiformes) and for the Mandibulata (Table 2); hence,
410 there was no evidence for the massive expansion in this gene family recently reported for the *T. urticae*
411 genome, with almost 700 members [100].

412 Ionotropic glutamate receptors (iGluRs) are glutamate-gated ion channels that are divided into two
413 subtypes based on sensitivity to N-methyl-D-aspartic acid (NMDA). The canonical iGLuRs do not have
414 direct roles in chemosensation. Rather, at least in *D. melanogaster*, the NMDA-sensitive channels are
415 expressed in the brain and are involved in associative learning and memory [101]. The non-NMDA
416 channels have fundamental roles in synaptic transmission in the neuromuscular junction within
417 muscle tissue or in the nervous system [102], and certain receptor subunits have been shown to be
418 involved in the regulation of sleep (GluR1 [103]) or vision (Clumsy, CG5621, CG3822, CG11155, CG9935
419 [104, 105]). Strikingly, the *D. tinctorium* genome harboured seven NMDA-type iGluRs and 61 non-
420 NMDA iGluRs, representing substantially greater repertoires than those observed for the *L. deliense,*
421 *T. urticae* and *D. melanogaster* genomes (especially for the non-NMDA iGluRs) (Fig. 11). The
422 chemosensory ionotropic receptors (IRs), which exhibit sequence similarity to iGluRs but do not bind

glutamate [106], also showed interesting differences in gene family size compared with *T. urticae* and *D. melanogaster.* Notably, while *D. melanogaster* has one gene encoding an IR25a protein, *T. urticae* has three such genes and the trombidid mites have five copies each (Fig. 11). The *D. melanogaster* IR25a is a widely-expressed co-receptor that couples with stimulus-specific IRs to facilitate sensitivity to a diverse range of acids and amines. Recently, IR25a in combination with IR21a and IR93a were demonstrated to function as a thermosensory complex expressed by the dorsal organ cool cells of *D. melanogaster* larvae, which mediates avoidance behaviour to cool temperatures (<20°C) [107, 108]. Sequences that cluster with *D. melanogaster* IR21a and IR93a in the "antennal and first leg" class of IRs were identified in the trombidid mite genomes, with one copy in *D. tinctorium* (as for *T. urticae*) and three copies in *L. deliense* (Fig. 11). Although chelicerates lack antennae, the orthologues of IR93a and/or IR25a have been shown to be highly expressed exclusively in the first pair of legs in *T. mercedesae* [98] and *Varroa destructor* [109], suggesting functional parallels between insects and mites.

Predicted allergens

Although the propensity of chiggers to cause pruritic dermatitis is well recognised in humans and other animals [35-38], the identity of the allergens involved has not been established [110]. The *L. deliense* and *D. tinctorium* genomes were predicted to encode 37 and 33 groups of protein allergens, respectively; substantially more than other sequenced mites in the Acariformes with the exception of the dust mites, *D. farinae* [77] and *E. maynei* [111]. Since velvet mites rarely come into contact with humans, only the chigger allergens were subjected to further analysis. The *L. deliense* predicted allergen clusters included nine groups that were unique to this species and six that were shared with *D. tinctorium* only (Fig. 12), while a further 28 putative allergen genes in the *L. deliense* genome did not cluster in orthologous groups (Additional file 19). These chigger-unique groups included five distinct clusters of trypsin-like serine proteases and one cluster each of subtilases, papain-like cysteine proteases, enolases, and cyclophilins, all of which could be classified into recognised allergen families

448 listed in the AllFam database [112] (Fig. 13). The non-clustered allergens belonged to a variety of
449 structural and enzymatic protein groups, but cathepsins, serine proteases and peptidylprolyl
450 isomerases were the most common annotations (Additional file 19).

451 The major allergens in *D. farinae* are the 25-kDa Der f 1, a papain-like cysteine protease; and Der f 2,
452 a 14-kDa uncharacterised protein with a ML (lipid-binding) domain [77]. However, many other minor
453 allergens have been detected by immunoproteomic studies [113, 114] or predicted by homology
454 searches in the *D. farinae* genome [77]. In *L. deliense*, five distinct clusters of papain-like cysteine
455 proteases were identified (AllFam AF030), of which three were shared with *D. farinae,* one was shared
456 only with *D. tinctorium*, and one was unique (Fig. 13). No orthologue of Der f 2 (AF111) was apparent.

457 Recently, an alpha-enolase has been reported as a novel minor allergen in *D. farinae* [114]. However,
458 the two enolases (AF031) with predicted allergenic properties in the *L. deliense* genome formed an
459 orthologous cluster that was absent from other mites sequenced to date, with homologues in parasitic
460 nematodes and distant chelicerate relatives (*e.g.,* horseshoe crabs; Additional file 19). A similar
461 pattern was observed for the cyclophilins (AF038), which have previously been considered a class of
462 allergens restricted to fungal and plant sources [115], although these peptidyl-prolyl *cis-trans*
463 isomerases are universally present across all domains of life. In an immunoproteomic study, a
464 cyclophilin was newly identified as a dust mite allergen, Der f 29 [113], but this was not closely related
465 to the *L. deliense* cyclophilins, which exhibited a greater affinity (~75% identity) to homologues in fungi
466 and fish (Additional file 19). The *L. deliense* subtilases (serine proteases with a peptidase S8/S53
467 domain, AF021) were also absent from other mite genomes but showed 40 – 50% identity to subtilases
468 from fungi and bacteria (Additional file 19). This class of proteases have been identified as major
469 allergens produced by ascomycete fungi such as *Curvularia lunata* [116] and *Trichophyton* spp. [117].
470 Finally, the five clusters of trypsin-like serine proteases (AF024) exhibited closest homologues (40 –
471 50% identity) in a diverse range of organisms, including *T. urticae* (but too distant to cluster in
472 ORTHOMCL31), Diptera and scorpions (ORTHOMCL32), fish and lizards (ORTHOMCL88), bugs and ants

(ORTHOMCL89), and acorn worms and Diptera (ORTHOMCL90) (Additional file 19). Thus, these predicted allergens were distinct from the *D. farinae* molecules classified in AF024 [Der f 3, 6 and 9 [118]], although within ORTHOMCL29, ORTHOMCL30, and ORTHOMCL43, *L. deliense* does possess additional trypsin-like proteases that are orthologous to these *D. farinae* allergens (Figure 13).

A label-free quantitative analysis of protein content in the chiggers indicated that muscle-derived allergens related to the *D. farinae* paramyosin Der f 11 (AF100 [119]) and to *T. urticae* tropomyosin isoforms (AF054) were most abundant (Fig. 13). Although single unique peptides were detected for several *L. deliense*-specific allergen clusters and unclustered allergenic proteins, only one *L. deliense*-specific allergen was present in quantifiable amounts (an enolase in AF031), and this was considerably less abundant than the shared allergens (Additional file 7, Fig. 13). However, as allergenicity is not dictated entirely by allergen quantity and can vary markedly between individuals, validating the identity of the most important allergens in chiggers will require screening of sera from trombiculiasis patients.

Putative salivary proteins

Due to the diminutive size of chiggers and the absence of any artificial feeding mechanism for laboratory colonies that might allow collection of saliva, the chigger sialome has not been characterised to date. However, numerous high-throughput studies of tick saliva have been conducted on several genera and multiple lifecycle stages [120-125], and recently an elegant proteomic analysis of *T. urticae* saliva was published, in which mite salivary secretions were collected in an artificial diet substrate [126]. Using proteomic datasets from this *T. urticae* study and a recent *I. scapularis* sialome analysis conducted over several time-points [120], we identified one-to-one orthologues of the salivary proteins from both sources in the tick, spider mite, and trombidid mite genomes. We reasoned that as *T. urticae* is phylogenetically close to the trombidid mites while *I. scapularis* is very distant, protein families shared by the tick and the trombidid mites but not present in the *T. urticae* genome are likely to represent proteins required for ectoparasitism on animal hosts (as opposed to

498 phytophagy). Indeed, 24 orthologous clusters were shared among the animal ectoparasites but not
499 with *T. urticae*, whereas only five clusters were shared between all mites at the exclusion of the tick
500 (Figure 14). These 24 animal-ectoparasite clusters are candidates as key salivary components of
501 trombidid mites. An additional two clusters were shared exclusively by *I. scapularis* and *L. deliense,*
502 suggesting that they might be important for feeding on vertebrate hosts (Fig. 14).

503 To feed successfully, ticks must suppress local immune responses and prevent the clotting of blood.
504 Although trombidid mites feed on tissue exudates or haemolymph rather than blood, and do not feed
505 for as long as some hard tick species, they face similar challenges as ectoparasites that provoke an
506 inflammatory response in their hosts. Interestingly, in accordance with low levels of haem in the
507 trombidid mite diet, we did not find orthologues of tick salivary proteins involved in haem
508 detoxification (ferritins and hemelipoproteins [122]) in the trombidid mite genomes. Several lipocalins
509 with histamine-binding activity have been identified in tick saliva from multiple different species [120,
510 122, 123], but orthologues of these small proteins were also not present. However, genes encoding
511 two enzymes involved in catabolism of the histamine precursor histidine, urocanate hydratase and
512 formiminotransferase-cyclodeaminase, were detected in both trombidid mite genomes (Table 3). The
513 degradation of histidine feeds into the one carbon pool by folate, and this process is mediated in part
514 by formyltetrahydrofolate dehydrogenase, an enzyme that is also present in multiple copies in the *D.*
515 *tinctorium* genome (Table 3). While the presence of folate biosynthesis enzymes in tick saliva has been
516 reported previously (and not only for *I. scapularis* [122]), the functional significance of their secretion
517 is unclear. One possibility is that ectoparasitic Acari not only utilise bacterial symbionts as folate
518 "factories" [127, 128], but can scavenge it at source from precursors in their B-vitamin-deficient diets.

519 Several other protein clusters with potential roles in immune evasion or the regulation of salivation
520 and the ingestion of host fluids were identified in the trombidid mite genomes. The presence of an
521 expanded acetylcholinesterase gene family in tick genomes has been noted previously and
522 acetylcholinesterases have been detected in the saliva of *Rhipicephalus microplus* [129] and

Amblyomma americanum [122], as well as *I. scapularis* [120]. It has been proposed that salivary acetylcholinesterases could interfere with cholinergic signalling between host immune cells and might facilitate pathogen establishment [129]. However, the trombidid mites have only a single gene copy each that clusters with the *I. scapularis* acetylcholinesterases (Table 3). Similarly, ATPase inhibitors (Table 3) in saliva could impact on local immune responses [122], since extracellular purine metabolites act as "alarmins" [130]. The massive expansion of sulfotransferases in the *I. scapularis* genome and the secretion of some members of this family in saliva is particularly enigmatic, but recently it has been proposed that sulfotransferases could control salivation and feeding cycles in ticks by sulphating the neurotransmitters dopamine and octopamine [131]. Alternatively or in addition, they might be involved in increasing the activity of small cysteine-free thrombin inhibitors in tick saliva by sulfation of tyrosine residues [132]. Notably, only two of these sulfotransferases were present in each trombidid mite genome compared with >70 members of this family in *I. scapularis* (Table 3).

Of the two salivary protein families restricted to *L. deliense* and *I. scapularis* (Table 3), the secreted trypsin inhibitor-like cysteine-rich domain proteins are among are wide diversity of serine protease inhibitors produced by ticks [133]. This specific class of trypsin inhibitor-like proteins includes ixodidin, an antimicrobial peptide expressed in the haemocytes of *R. microplus* [134] and BmSI-7 and BmSI-6, two peptides from the same species of tick that inhibit cuticle-penetrating proteases secreted by entomopathogenic fungi [135]. The BmSI-7 peptide is expressed in multiple tissues, including the salivary glands [135], but its role in saliva is unknown. However, one possibility is that it helps prevent the tick bite site from becoming infected. In contrast with *I. scapularis*, which harbours 27 trypsin inhibitor-like proteins in its genome, only one orthologue was identified in *L. deliense* (Table 3). The second cluster restricted to *I. scapularis* and *L. deliense*, a signal sequence receptor subunit (Table 3), was unexpected as it has a canonical function in trafficking secretory proteins through the ER [136]. This appears to have a moonlighting role in tick saliva, since it generates strong immune responses in rabbits parasitized by *A. americanum* [122].

Feeding ticks secure their mouthparts in the skin of the host for days or weeks using a cement-like substance that forms a cone in the bite wound. Superficially, the stylostome generated at the feeding site of trombidid mites resembles the tick cement cone, although the structure is tubular (in the Trombiculoidea) or highly branched (in the Trombidioidea) [40]. In both of the trombidid mite genomes, we found an orthologue of a glycine-rich protein present in the sialome of *Rhipicephalus pulchellus* [124] (Table 3). The tick glycine-rich proteins are related to spider silk proteins and form the main structural component of tick cement as determined by proteomic studies [137]. To determine if the trombidid mite genomes may contain other cement-associated proteins not detected in the *I. scapularis* salivary proteomics study [120], we searched for orthologues of all tick cement proteins in the National Center for Biotechnology Information (NCBI) database. We found orthologues of an *I. scapularis* glycine-rich cement protein in both *L. deliense* (one copy) and *D. tinctorium* (three copies) that was distinct from the *R. pulchellus* orthologue; moreover, the velvet mite also possessed an orthologue of a second *I. scapularis* cement protein (Additional file 20). In addition, both trombidid mites harboured a gene related to a cement protein transcript identified in the sialotranscriptome of *Amblyomma triste* [125] (Additional file 20). Finally, orthologues of *A. americanum* acidic chitinases involved in conferring stability to the tick cement cone were present in both *D. tinctorium* (four copies) and *L. deliense* (one copy) [138] (Additional file 20).

Discussion

Genome features and trombidid mite evolution

In this study, we exploited the close phylogenetic relationship between the Trombidioidea and the Trombiculoidea in order to obtain a genome from a single adult velvet mite that could be used to corroborate data derived from a suboptimal trombiculid mite sample (*i.e.*, a pool of engorged larvae). This strategy proved successful because in almost all cases, the unusual features of the trombidid mite genomes were shared between the two sequenced taxa. In contrast with other acariform mites, the trombidid mite genomes were substantially larger, contained a greater proportion of repeats, and

exhibited expansions of mobile elements. These features, coupled with heterozygosity and host contamination in the case of *L. deliense*, proved challenging for accurate genome size estimation but did not prevent the annotation of protein-coding genes, which was sufficient (even for *L. deliense*) for an initial protein expression study. Our *k*-mer-based estimate of genome size for *L. deliense* was slightly smaller than those determined for *Leptotrombidium pallidum* and *Leptotrombidium scutellare* using DNA from laboratory-reared adult specimens, which were 191 ± 7 Mb and 262 ± 13 Mb (by qPCR), or 175 Mb and 286 Mb (by *k*-mer analysis), respectively [139]. In future, genome sequencing from individual *L. deliense* adults would help resolve the variation in genome size across the *Leptotrombidium* genus, which is likely to be driven by repeat content.

As previously reported on publication of the first spider genomes [61], the Acari are polyphyletic, with the superorder Parasitiformes (*i.e.*, the ticks together with mesostigmatid and holothyrid mites) evidently more closely aligned to the spiders (order Araneae) than it is to the Acariformes (Figure S1). However, at lower taxonomic scales, the phylogenomic analyses supported the conventional morphology-based taxonomy within the Acariformes, confirming that the trombidid mites are closely related to the phytophagous Tetranychoidea ([14]. Interestingly, within the Trombidiae, the divergence between the velvet mites and the chiggers occurred much later (~123 MYA) than the emergence of the earliest terrestrial vertebrates (395 MYA [140]), coinciding perhaps with the appearance of crown-group mammals. If this scenario is correct, the ectoparasitism of non-mammalian vertebrates by chiggers that occurs today may be a product of secondary adaptation, as suggested by other authors [20]. Although the fossil record is devoid of trombidid mite specimens predating the Eocene [141], it has been speculated from palaeogeographical and comparative morphological evidence that trombiculid mites fed initially on other arthropods, with larval ectoparasitism on vertebrates evolving during the Paleocene, leading to an increase in chigger diversity [20]. Our data challenges this hypothesis, because it implies that something other than host choice in the larval stage drove the split between the Trombidioidea and the Trombiculoidea 60 million

598 years before the latter began feeding on vertebrates. Only the discovery of more ancient trombidid
599 fossils will help to resolve these uncertainties.

600 Potential roles for terpenes in trombidid mite biology

601 The most striking finding in the trombidid mite genomes was the presence of large families of laterally-
602 transferred terpene synthases. As the level of amino acid identity between the trombidid terpene
603 synthases and their closest homologues in microbes is quite low, their end products cannot be inferred
604 with any confidence. However, the question of whether compounds such as 2-methylisoborneol,
605 geosmin or germacrene might confer adaptive advantages to trombidid mites helps to frame
606 hypotheses for experimental testing. Interestingly, all of these compounds are associated with odours
607 and tastes that humans, and some arthropods, may sense as unpleasant or aversive. For instance,
608 2-methylisoborneol has a musty odour that humans associate with ripe cheeses [142]; whereas
609 geosmin confers the smell of moist soil and is the cause of the muddy, "off" taste that the human
610 olfactory system detects in spoiled water, wine and the flesh of certain freshwater fish [143-145].
611 More importantly, geosmin released by *Penicillium* spp. or *Streptomyces* spp. on rotting fruit is
612 strongly aversive to *Drosophila* because these organisms produce secondary metabolites that are
613 directly toxic to the fly or to its primary food source, yeast [146]. Germacrene has also been implicated
614 as an arthropod repellent amongst complex sesquiterpene mixtures found in the essential oils of
615 various plants, which have been shown to be effective against several acarines, including the ticks
616 *Rhipicephalus microplus* and *Ixodes ricinus* [147, 148], and the poultry red mite *Dermanyssus gallinae*
617 [148]. Although it has not been assayed in isolation, germacrene is additionally a significant
618 component of essential oils or crude leaf extracts that exhibit toxic effects against phytophagous mites
619 (*Brevipalpus phoenicis* in the Trombidiformes [149]) and ants (*Solenopsis invicta* [150]).

620 These potential repelling and/or toxic effects of terpenoids would align closely with the apparent
621 aposomatic nature of trombidid mites, most species of which are brightly coloured due to their
622 carotenoid content. These mites have few natural enemies and have been reported to be rapidly

regurgitated if offered to predators in the laboratory [151]. However, cannibalism between adults and even ectoparasitism of the free-living stages by trombidid larvae can occur, underlining the relevance of chemical communication within species and between closely related species [18]. To the best of our knowledge, only one report of a parasitoid affecting trombidid mites has been published (the acrocerid fly *Pterodontia flavipes* attacking *Podothrombium* spp. [152]), but the dearth of research on free-living trombidid stages means that no doubt other parasitoids exploiting these hosts do exist. It is important to note that not all mites are repelled by terpenoid compounds. Many plants use terpenoids as defence compounds to signal to the natural enemies of pest arthropods that the plant is under attack. For example, *Lotus japonicus* infested with *T. urticae* releases several terpenoids, including germacrene, which attract the predatory mite *Phytoseiulus persimilis* [153]. Communication by sex pheromones is also known to occur in *T. urticae*, although molecules other than terpenoids are suspected to mediate this [154]. Moreover, as noted above, neryl formate is an aggregation pheromone in dust mites [76]. In conclusion, while the conferment of a foul taste (and perhaps odour) to the aposomatic trombidids appears to be the most likely evolutionary driver of *de novo* terpenoid synthesis capability, it is also possible that these compounds are used to communicate with potential mates during courtship, or to repel members of the same (or closely-related) species to reduce competition or deter cannibalistic behaviour.

Diapause and seasonality

Many of the putative functions of laterally transferred and/or expanding gene families in the trombidid mites appear to be relevant to the temporal regulation of the lifecycle and the switching of metabolic demands between dormant and active stages. The life history of trombidid mites features alternation between immobile calyptostases (the deutovum, protonymph and tritonymph) and the active instars (larva, deutonymph and adult) [155]. The calyptostases typically persist for 25 – 30 days, while the active stages in temperate species can undergo hibernation over the winter months, including larvae that have not fed until late in the autumn. Diapause of eggs is common in temperate

species and can exceed one year in the chigger, *Hirsutiella zachvatkini*, without loss of viability [20]; while larvae of this species can also overwinter on their rodent hosts [156]. The lifecycle of trombidid mites in tropical and subtropical regions has been little studied, but several generations per year are possible [20]. In the case of *Dinothrombium* spp., although the adults are positively phototactic and diurnal if humidity is high (becoming crepuscular during drier conditions), circadian cycles of activity were maintained if mites were transferred to constant darkness in the laboratory [151]. Remarkably, adult *D. pandorae* of the Californian deserts may only emerge from their burrows during rainstorms to feed and mate for a few hours each year, migrating to the deepest extent of their subterranean refuges during the height of summer [43]. However, overcoming torpor by rapidly adjusting metabolic rate after the cold desert night to the early morning warmth, when termite prey become active, is critical to the lifecycle of *D. pandorae* [43]. It may also be important for chiggers to avoid cool microclimates (and thus maintain peak metabolism) when questing for small mammals, since their hosts are highly motile and a suitable location for attachment must be targeted rapidly before grooming behaviour leads to ingestion of the mite. Hence, the small expansions in the IR repertoire (IR25a, IR21a and IR93a) that we observed in the trombidid mite genomes might reflect more acute sensitivity to cool temperatures than for the phytophagous *T. urticae*.

With relevance to both regulation of metabolism and circadian cycles, the homologues of FLVCR proteins have been little studied in arthropods, but in vertebrates they export haem from the cytoplasm to the extracellular milieu (for FLVCR1) and from mitochondria into the cytoplasm (for FLVCR2) [157]. A FLVCR gene homologue in *Drosophila melanogaster*, *CG1358*, is involved in maintenance of circadian rhythms in the absence of light together with other genes with roles in iron metabolism [158]. Thus, it is intriguing that FLVCR homologues were significantly expanded in the *D. tinctorium* genome but not that of *L. deliense,* as adult chiggers in tropical environments exhibit more regular activity above ground than do *Dinothrombium* spp. [20]. To the best of our knowledge, the function of the 4-coumarate:CoA ligase family has not been explored in the Arachnida, but in adult females of the kissing bug *Rhodnius prolixus,* RNAi of long-chain acyl-CoA synthetase 2 led to a 90%

674 decrease in fatty acid β-oxidation and substantial reductions in oviposition and egg hatching, as well
675 as marked abnormalities in the remaining eggs and hatched nymphs [159]. The large gene expansions
676 in acyl-CoA synthetases observed here in *D. tinctorium* suggest that β-oxidation is a particularly
677 important facet of its metabolism. Furthermore, the proteomic analysis of *D. tinctorium* revealed
678 overrepresentation of putative digestive enzymes with peptidase M20 and inhibitor I29 domains,
679 which is consistent with an imperative for adult velvet mites to obtain food reserves rapidly while
680 foraging briefly above ground [43]. An elevated metabolic rate for *L. deliense* larvae was also
681 suggested by the preponderance of mitochondrial enzymes responsible for aerobic energy production
682 in protein extracts. Further studies on trombidid mite metabolism are clearly warranted, as most
683 metabolic studies in the Trombidiformes have focused on winter diapause in *T. urticae* [160], creating
684 a knowledge gap around the physiology of tropical species.

685 In spider mites, carotenoids are essential for the control of diapause and sexual reproduction.
686 However, reproductive behaviour between the Tetranychoidea and the trombidid mites is radically
687 different, since male *T. urticae* become developmentally arrested in close proximity to dormant
688 female deutonymphs on leaf surfaces. This “guarding” behaviour is stimulated by the intensity of the
689 yellow colouration of the dormant females (derived from their carotenoid pigments), and allows the
690 male to mate with the adult female immediately after ecdysis [161]. In contrast, in order to be
691 inseminated, adult females of trombidid mites must collect a spermatophore deposited on the ground
692 by the male, the location of which is signposted by signalling threads. While colouration is not known
693 to be factor during courtship, the males of some species deposit spermatophores in specially
694 constructed “gardens” and perform encircling dances with the female [18]. However, it has recently
695 been discovered using genetic manipulation of a laterally-transferred phytoene desaturase that
696 carotenoids have a second, distinct function in the regulation of diapause in *T. urticae* [13]. A lack of
697 phytoene desaturase activity not only results in albinism, but prevents overwintering strains from
698 entering diapause, probably due to disrupted light (and thus photoperiod) perception caused by
699 vitamin A deficiency.

To the best of our knowledge, anatomical and experimental studies on trombidid mite vision have not been performed, but spider mites once again provide a closely-related template. The eyes of *T. urticae* have been partially characterised, and show biconvex lenses in the anterior pair and simplex convex lenses in the posterior pair [162]. It has been proposed that the anterior eyes respond to UV and green light, whereas the posterior pair are sensitive to UV only. However, laser ablation experiments have demonstrated that either pair can receive sufficient information to control the photoperiodic termination of diapause, while removal of both pairs prevents diapausal exit [163]. In *T. urticae*, the expression pattern of rhodopins in the eyes has not been determined, but in the jumping spider *Hasarius adansoni,* green-sensitive rhodopsin-1 expressed in the lateral ocelli is important for the monochromatic detection of movement [164]. Until recently, the function of rhodopsin-7 was enigmatic, but experiments in *D. melanogaster* have shown that it operates in the circadian pacemaker neurons of the central brain and is responsible for their highly sensitive response to violet light [165]. Taken together, these experimental findings from other arthropods suggest that trombidid mites might depend on rhodopsin-7 homologues rather than peropsins for control of circadian rhythms (although we failed to identify orthologues of the *Drosophila* clock gene in Acari). The relative roles of FLVCR gene homologues, phytoene desaturases and the rhodopsin pigments in the control of diapause and other life history traits in trombidid mites is evidently a key priority for future experimental studies.

Immune response and vector biology

As trombidid mites are edaphic organisms with a parasitic larval stage, they are exposed to soil microorganisms, the exterior flora of their hosts, and pathogens contained in ingested body fluids. In the case of certain trombiculid mites (especially *Leptotrombidium* spp.), their role as biological vectors of *O. tsutsugamushi* highlights a specific infectious challenge that has resulted from feeding on small mammals. In other arthropods, CTLD proteins with lectin activity act as transmembrane or secreted pattern recognition receptors that bind to carbohydrates on the surface of pathogens. They can

function as opsonins in the haemolymph that agglutinate unicellular pathogens and facilitate their phagocytosis by haemocytes [166], or may be expressed on the surface of tissues that form a barrier to infectious assaults, such as in the gut or on the gills of crustaceans [92, 167]. These findings are compatible with roles as secreted opsonins for most of the CTLD proteins identified in the current study, with a smaller number perhaps operating as immune surveillance receptors on the surface of cells or extracellular matrices.

It is unclear why the *L. deliense* genome harbours a variety of PERK-like proteins. However, the alpha subunit of eukaryotic translation-initiation factor-2 is the main substrate of PERK and is inactivated by phosphorylation, leading to inhibition of global protein synthesis. As the unfolded protein response bridges the cellular response to viral infection and metabolic homeostasis [168], the manipulation of PERK by dengue virus [169] suggests that an expanded repertoire of these kinases might be beneficial in the evolutionary arms race with viral pathogens. Indeed, we identified a striking diversity of ERV-related sequences in the *L. deliense* genome, apparently reflecting a substantial degree of exposure to viral nucleic acids. The relatively large number of these Gag-like polyproteins, the absence of similar quantities of other retroviral proteins in the *L. deliense* assembly, and the closer relationship of the Gag-like polyproteins to ERV elements in non-sciurid mammals, all rendered host DNA contamination in the mite gut or on mouthparts as an unlikely source for these sequences; although this possibility cannot be excluded entirely. Despite this caveat, lateral transfers in the distant past originating from mammalian body fluids during the brief ectoparasitic stage is a working hypothesis that can be tested when a more contiguous chigger genome becomes available. In contrast, the horizontal transfer of the long-interspersed element BovB is postulated to have involved an opposite transmission route; that is, between vertebrates by ticks [170].

To the best of our knowledge, no experimental studies on the immune response of trombidid mites have been performed to date. However, a recent study in *T. urticae* involving experimental challenge with bacteria demonstrated that the spider mites were highly susceptible to systemic infection [171].

This contrasted with a more robust response to bacterial challenge in another acariform mite, *Sancassania berlesei*, which unlike *T. urticae* has a saprophytic lifestyle. The authors of this study concluded that the ecology of spider mites, in which all lifecycle stages feed on plant phloem (a relatively aseptic food source), has led to a high degree of susceptibility to pathogen exposure. This failure to overcome infectious insults was associated with an apparent absence of many antimicrobial protein effectors in the spider mite genome. In support of the hypothesis that spider mites have adapted to an environment characterised by very low levels of pathogen challenge, we found that compared with the *L. deliense* genome, the *T. urticae* genome displays a relative paucity of PGRP (Additional file 17), CTLD (Additional file 6) and Dscam genes (Additional file 16). Thus, the common ancestor of the Trombidiformes may have harboured a diverse immune gene repertoire that was selectively lost in the branch leading to the Tetranychoidea, and/or the Trombiculoidea has undergone more recent immune gene family expansions. The intermediate immune gene repertoire of the Trombidioidea between that of the spider mites and the chiggers (*D. tinctorium* has considerably fewer CTLD proteins and PERKs than *L. deliense*) suggests that both immune-related gene losses and gains have occurred during the evolution of the Trombidiformes in response to their radically different natural histories. Indeed, in terms of the degree of exposure to pathogen diversity and abundance, the euedaphic velvet mites are likely to encounter greater infectious challenges than spider mites, and the feeding behaviour of chiggers on vertebrates is likely to exacerbate this exposure further compared to their relatives that are ectoparasitic on other invertebrates only.

Conclusions

This first analysis of trombidid mite genomes has revealed their dynamic nature relative to those of other acariform mites, including expansions in laterally-transferred gene families and mobile elements. These genomes provide a foundation for fundamental experimental studies on mite immune responses, host-seeking behaviour and feeding, and environmental impacts on lifecycle progression. The function of the laterally-transferred terpene synthases will become a major research

theme for chigger biology, as only experimental exposure of chiggers and their potential natural enemies to mite terpene extracts will be able to determine if these unique aspects of secondary metabolism have evolved to attract conspecifics; or conversely, to repel predators, parasitoids and/or competitors. From an applied perspective, the identification of predicted allergens in the *L. deliense* genome sets the scene for immunoproteomic studies of trombiculiasis in both humans and domestic animals, with the potential for immunotherapeutic approaches to be developed as for dust mite allergy [172]. Finally, the successful development of recombinant vaccines against ticks [173] and the promising progress of recombinant vaccine development for both sheep scab [174] and poultry red mite [175] indicate that a similar approach could be explored for chiggers, which has the potential to interrupt, or at least reduce, the transmission of *O. tsutsugamushi* to humans in scrub typhus-endemic areas. Considering the high strain variability of the scrub typhus agent [176], a chigger vaccine utilising mite salivary or gut antigens could provide a much-needed breakthrough against this intractable disease.

Materials and methods

Sample collection and DNA extraction

Adult specimens of giant red velvet mites were collected within the grounds of the UK Medical Research Council Field Station at Wali Kunda, The Gambia (13°34'N, 14°55'W). Mites were sampled from flowerbeds following heavy rains in June 2010 and stored in 95% ethanol at -80°C. They were identified as *Dinothrombium tinctorium* by Joanna Mąkol (Wrocław University of Environmental and Life Sciences, Poland). Approximately 5 µg of DNA was extracted from a single individual using a Genomic-tip Kit (Qiagen) according to the manufacturer's instructions. Integrity of the DNA was confirmed by agarose gel electrophoresis, which showed a single band of ~20 kb.

For *L. deliense*, engorged larvae were collected from two Berdmore's ground squirrels (*Menetes berdmorei*) captured in Udonthani Province, Thailand, in September 2015. Trapping and euthanasia of

small mammals followed the CERoPath (Community Ecology of Rodents and their Pathogens in a changing environment) project protocols [177]. Chiggers were located inside the ears and the inguinal area of the squirrels and stored in absolute ethanol at -20°C. A subsample of the mites was selected and mounted in clearing medium, Berlese fluid (TCS Bioscience, UK), prior to species identification under a compound microscope. Fifty unmounted larvae were pooled and ~30 ng of genomic DNA was extracted using a DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions. The DNA was partially degraded, but a dominant band of ~5 kb was apparent by agarose gel electrophoresis.

Library preparation and sequencing

These steps were performed at the Centre for Genomic Research at the University of Liverpool. The *D. tinctorium* DNA was used to generate two Illumina TruSeq libraries and one Nextera mate-pair library. For the former, bead-based size selection using 100 ng and 200 ng of DNA as input into the TruSeq DNA LT Sample Prep Kit with 350 and 550 bp inserts, respectively, was applied. Following eight cycles of amplification, libraries were purified using Agencourt AMPure XP beads (Beckman Coulter). Each library was quantified using a Qubit fluorimeter (Life Technologies) and the size distribution was assessed using a 2100 Bioanalyzer (Agilent). The final libraries were pooled in equimolar amounts using the Qubit and Bioanalyzer data. The quantity and quality of each pool was assessed on the Bioanalyzer and subsequently by qPCR using the Illumina Library Quantification Kit (KAPA Biosystems) on a LightCycler 480 instrument II (Roche Molecular Diagnostics) according to the manufacturer's instructions. The pool of libraries was sequenced on one lane of the HiSeq 2000 with 2 × 100 bp PE sequencing and v3 chemistry.

The *D. tinctorium* mate-pair library was constructed using the Nextera Mate Pair Kit (Illumina) with 3 kb inserts. The DNA (3 µg) was tagmented as described in the manufacturer's protocol and cleaned using a Genomic DNA Clean & Concentrator column (Zymo Research). The sample was then subjected to strand displacement and cleaned with AMPure XP beads. A 0.6% Certified Megabase Agarose gel

(Bio-Rad) was used to separate the fragments, and those in the range of 2 – 5 kb were extracted and recovered using a Zymoclean Large Fragment DNA Recovery Kit (Zymo Research). The recovered DNA was quantified and transferred into a circularisation reaction at 16°C overnight. After purification with AMPure XP beads, DNA was sonicated into ~500 bp fragments using a focused ultrasonicator (Covaris) and recovered with AMPure XP beads as before. Samples were bound to Dynabeads M-280 Streptavidin (Thermo Fisher Scientific) and all subsequent reactions (end repair, A-tailing, and adapter ligation) were bead-based. Samples were amplified with 10 cycles of PCR, recovered by AMPure XP beads at a 1:1 ratio, and quantified using the Qubit dsDNA HS Assay Kit. The library was then subjected to quality control on the Bioanalyzer and LightCycler as for the TruSeq libraries above. The library was sequenced on one run of the MiSeq with 2 × 250 bp PE sequencing.

For *L. deliense*, the DNA sample was sheared to 550 bp using a Bioruptor Pico sonication device (Diagenode) and purified using an AxyPrep FragmentSelect-I Kit (Axygen). The sample was then quantified using a Qubit dsDNA HS Assay Kit on the Qubit fluorimeter, and the size distribution was ascertained on the Bioanalyzer using a High Sensitivity DNA chip (Agilent). The entire sample was used as input material for the NEB Next Ultra DNA Library Preparation Kit. Following nine PCR cycles, the library was purified using the AxyPrep kit and quantified as before by Qubit. Library size was determined on the Bioanalyzer (Agilent). The quality and quantity of the pool was assessed as described above for *D. tinctorium*. The sequencing was conducted on one lane of an Illumina MiSeq with 2 × 150 bp PE sequencing and v2 chemistry.

Assembly and annotation

For both genomes, base-calling and de-multiplexing of indexed reads was performed by bcl2fastq v. 1.8.4 (Illumina) to produce sequence data in fastq format. The raw fastq files were trimmed to remove Illumina adapter sequences using Cutadapt version 1.2.1 [178]. The option "-O 3" was set, so the 3' end of any reads which matched the adapter sequence over at least 3 bp was trimmed off. The reads were further trimmed to remove low quality bases, using Sickle version 1.200 [179] with a minimum

window quality score of 20. After trimming, reads shorter than 10 bp were removed. If both reads from a pair passed this filter, each was included in the R1 (forward reads) or R2 (reverse reads) file. If only one of a read pair passed this filter, it was included in the R0 (unpaired reads) file.

The Kraken taxonomic sequence classification system (v. 1.0) [180] was used to assign taxonomic annotations to the reads and specifically to estimate the level of bacterial contamination in the raw data. For genome size estimations, *k*-mers were counted by the Jellyfish program (v. 2.0) [181] and the resultant histograms were uploaded to GenomeScope (v. 1.0) [182] to visualise the *k*-mer distributions.

For *D. tinctorium*, the PE reads were assembled using Abyss (v. 1.5.2) [48, 49], Allpaths-LG (v. r51279) [183], SOAPdenovo2 (v. 2.04-r240) [184] and Discovar (v. r51454) [185]. When running Abyss, *k*-mer sizes were from 35 bp to 80 bp with an interval of 5, and the output "*k*-mer size = 80 bp" was selected as this produced the optimal assembly. Allpaths-LG and Discovar specify *k*-mer size automatically, whereas "*k*-mer size = 63" was selected for SOAPdenovo2, as suggested by the developer. Discovar requires read lengths of 250 bp, so was applied only to the data generated from the mate-pair library. Assessment of the completeness of the genome assemblies was based on the percentage alignment obtained against the reads from the TruSeq 350-bp insert library using bowtie2 (v. 2.0.10) [186] and the predicted core gene content determined by CEGMA (Core Eukaryotic Genes Mapping Approach) (v. 2.5) [187]. The final, optimum assembly was created by Abyss (97.4% of scaffolds >500 bp were mapped; 99.2% of Key Orthologs for eukaryotic Genomes were present) and included all scaffolds of ≥1,000 bp, which constituted ~80% of the total length of the raw assembly.

For *L. deliense*, a preliminary genome assembly at contig level was performed using Velvet (v. 1.2.07) [50], with parameters of 'best *k*-mer 99' and '-ins_length 500'. Reads derived from mammalian host genomic DNA were removed from the preliminary genome assembly using blobtools (v0.9.19), which generates a GC-coverage plot (proportion of GC bases and node coverage) [51] (Fig. 3). The filtered *L.*

873 *deliense* reads were then reassembled using SPAdes assembler (v3.7.1) [52] with default settings and
874 the length cut-off for scaffolds was set at 500 bp.

875 To find, classify and mask repeated sequences in the mite genome assemblies, a *de novo* repeat library
876 was first built using RepeatModeler (v. 1.0.8) [188] with the '-database' function, followed by
877 application of RepeatMasker (v. 4.0.6) [189] using default settings for *de novo* repeated sequences
878 prediction. Then, a homology-based prediction of repeated sequences in the genome was achieved
879 using RepeatMasker with default settings to search against the RepBase repeat library (issued on
880 August 07, 2015). For non-interspersed repeated sequences, RepeatMasker was run with the '-noint'
881 option, which is specific for simple repeats, microsatellites, and low-complexity repeats.

882 Thee *ab initio* gene prediction programs, including Augustus (v. 3.2.2) [54], SNAP (v. 2013-11-29) [55]
883 and GeneMark (v. 2.3e) [56] were used for *de novo* gene predictions in each genome assembly.
884 Augustus and SNAP were trained based on the gene structures generated by CEGMA (v. 2.5) [187],
885 whereas GeneMark [56] was self-trained with the '--BP OFF' option. All three *ab initio* gene prediction
886 programs were run with default settings. We also generated an integrated gene set for each genome
887 assembly using the MAKER v. 2.31.8 [53] pipeline. The MAKER pipeline runs Augustus, SNAP and
888 GeneMark to produce *de novo* gene predictions, and integrates them with evidence-based
889 predictions. These were generated by aligning the invertebrate RefSeq protein sequences
890 (downloaded on March 31, 2016 from NCBI) to the masked mite genomes by BLASTX. The MAKER
891 pipeline was run with '-RM_off' option to turn all repeat masking options off, and all parameters in
892 control files were left in their default settings. Genes identified by *de novo* prediction, which did not
893 overlap with any genes in the integrated gene sets, were also added to the final gene set for each
894 genome assembly if they could be annotated by InterProScan (v. 4.8) [190] with the InterPro
895 superfamily database (v. 43.1) using '-appl superfamily -nocrc' options.

896 The Blast2GO pipeline (v. 2.5) [191] was used to annotate proteins by Gene Ontology (GO) terms. In
897 the first step, all protein sequences were searched against the nr database with BLASTP. The E-value

cutoff was set at 1×10^{-6} and the best 20 hits were used for annotation. Based on the BLAST results, the Blast2GO pipeline then predicted the functions of the sequences and assigned GO terms to the BLAST-based annotations. Metabolic pathways were constructed using KAAS (KEGG Automatic Annotation Server) [192] with the recommended eukaryote sets, all other available insects, and *I. scapularis*. The pathways in which each gene product might be involved were derived from the best KO hit with the BBH (bi-directional best hit) method.

Phylogenetics

Protein data sets of the following arthropod genomes were used as references: *D. melanogaster* (fruit fly; GOS release: 6.11) [193], *A. mellifera* (honey bee; GOS release: 3.2) [194], *T. mercedesae* (bee mite; GOS release: v. 1.0) [98], *T. urticae* (spider mite; GOS release: 20150904) [59], *Stegodyphus mimosarum* (velvet spider; GOS release: 1.0) [61], *I. scapularis* (blacklegged tick; GOS release: 1.4) [195], *M. occidentalis* (predatory mite; GOS release: 1.0) [97]. *Caenorhabditis elegans* (nematode; GOS release: WS239) [196] was used as the outgroup. For gene-family phylogenetics, we first aligned orthologous protein sequences with Mafft (v7.309) [197] or Kalign (v2.0) [198]. We manually trimmed the aligned sequences for large gene sets. The best substitution models of amino-acid substitution were determined for the alignments by ProtTest (v3.4) with parameters set to "-all-matrices, -all-distributions, -AIC" [199]. Then, phylogenetic trees were constructed using maximum likelihood methods (Phyml, v3.1) [200]. In addition, a neighbour-joining method was used for building the distance-based trees using MEGA (7.021) [201].

For species-level phylogenetics, the rapid evolution of acariform mites may challenge phylogenetic analyses due to long-branch attraction [202]. Thus, we used a very strict E-value (1×10^{-50}) when performing a reciprocal BLASTP to exclude the most variant orthologous genes across all genomes tested. The reciprocal BLAST search resulted in identification of a total of 926 highly conserved one-to-one orthologues in all eight genomes. Each of these orthologous groups was aligned using Mafft with the "-auto" option. These alignments were trimmed by Gblocks (v. 0.91b) [203] and concatenated

into unique protein super-alignments. ProtTest determined the best-fit substitution model of LG with invariant sites (0.131) and gamma distributed rates (0.913) using parameters as above prior to conducting the phylogenetic analysis with Phyml and Bayesian methods (MrBayes, v. 3.2.6) [204]. Based on the topology defined by this phylogenetic analysis, we estimated the divergence time of each species using the Bayesian MCMC method in the PAML package (v. 4.9a) [205] with the correction of several fossil records (time expressed in MYA): tick-spider: 311–503 [61] (oldest spider from coal, UK), *T. urticae*-tick-spider: 395–503 [61] (oldest Acari), *A. mellifera-D. melanogaster*: 238-307 [206] and nematode-arthropods: 521–581 [206].

Analysis of gene family expansions

Orthologous gene clusters of *D. tinctorium*, *L. deliense* and the other reference genomes described above were defined based on OrthoMCL (v. 1.4) [207]. We used CAFE (v. 3.1) [208] to infer the gene family expansion and contraction in *D. tinctorium* and *L. deliense* against other Acariformes (*T. urticae* and *S. scabiei*). The ultrametric species tree used in the CAFE analyses was created as described for gene-family phylogenetics above. We also calculated ω (d_N/d_S) ratios for 454 one-to-one orthologues defined by OrthoMCL using codeml in the PAML package [205] with the free-ratio model. Branches with ω >1 are considered to be under positive selection. The null model used for the branch test was the one-ratio model (nssites = 0; model = 0) where ω was the same for all branches. Kappa and omega values were automatically estimated from the data, with the clock entirely free to change among branches. The *P* value was determined twice using the log-likelihood difference between the two models, compared to a χ^2 distribution with the difference in number of parameters between the one-ratio and free-ratio models. To estimate significance with the *P* value, a likelihood-ratio test was used to compare lnl values for each model and test if they were significantly different. The differences in log-likelihood values between two models were compared to a χ^2 distribution with degrees of freedom equal to the difference in the number of parameters for the two models. Measurement of d_S was

947 assessed for substitution saturation, and only d_S values <3.0 were maintained in the analysis for
948 positive selection. Genes with high ω (>10) were also discarded.

949 Analysis of candidate lateral gene transfers

950 We used a modification of the Crisp method [6] for examination of LGTs in the two mite genomes.
951 Each mite protein dataset was aligned with BLASTP against two databases derived from the NCBI nr
952 database, one consisting of metazoan proteins (excluding proteins from species in the same phylum
953 as the studied species - Arthropoda) and the other of non-metazoan proteins. The LGT index, *h*, was
954 calculated by subtracting the bit-score of the best metazoan match from that of the best non-
955 metazoan match. The genes can be classified into class C if they gained an *h* index ≥ 30 and a best non-
956 metazoan bit-score of ≥100. For each class C gene, its average *h* value (h_{orth}) and that of its paralogous
957 genes in each OrthoMCL cluster defined above was determined. If *h* was ≥30, the best non-metazoan
958 bit-score was ≥100, and the h_{orth} value was ≥30, the gene was considered to be a class B gene. Class A
959 genes were defined as a subset of class B genes with *h* ≥30, a best non-metazoan bit-score of ≥100, an
960 h_{orth} value of ≥30, and a best metazoan bit-score of <100.

961 Analysis of immune-related gene families

962 A search for mite immune-related genes was initially preformed with a BLASTP search (E-value, $<1 \times$
963 10^{-5}) against each mite protein set using immune-related genes defined by Palmer & Jiggins [85]. The
964 identified potentially immune-related genes were then manually checked using BLASTP online at NCBI.
965 For analysis of CTLD proteins, FASTA sequences of proteins in ORTHOMCL136, ORTHOMCL510,
966 ORTHOMCL582, and ORTHOMCL880 were analysed by SecretomeP 2.0 [209], TMHMM 2.0 [210] and
967 NetNGlyc 1.0 [211] using default settings to identify respective sequence features. Protein sequences
968 were also submitted to InterPro [212] for domain structure analysis. The CTLDs extracted from
969 individual protein sequences were then manually searched for the amino-acid motifs “EPN”, “WND”
970 and “QPD”.

Analysis of chemosensory and photoreceptor gene families

A search for *D. tinctorium* and *L. deliense* OBPs was initially preformed using TBLASTN (E-value, $<1 \times 10^{-3}$) against their genome assemblies using *D. melanogaster, Drosophila mojavensis, Anopheles gambiae, Bombyx mori, Tribolium castaneum, A. mellifera, Pediculus humanus humanus* and *Acyrthosiphon pisum* OBPs (identified by Vieira and Rozas [213]) as queries. No OBPs were found in the *D. tinctorium* and *L. deliense* genome assemblies. Because OBPs are very divergent in terms of the amino-acid sequences within the family, and the sequence identities between the family members from the different species can be as low as 8% [213], a TBLASTN search has limited power to identify these genes. A search for OBPs was therefore performed again with BLASTP (E-value, $<1 \times 10^{-3}$) to search the automated protein predictions from the mite genome assemblies. A search for small chemosensory proteins is *D. tinctorium* and *L. deliense* was preformed using the same methods as for the OBPs. The query sequences were also based on the study of Vieira and Rozas [213], using chemosensory protein sequences from *D. melanogaster, D. mojavensis, A. gambiae, B. mori, T. castaneum, A. mellifera, P. humanus humanus, A. pisum, I. scapularis*, and *Daphnia pulex*. For odorant receptors, both TBLASTN and BLASTP searches were performed using *D. melanogaster* and *A. mellifera* sequences (identified by Nozawa and Nei [214] and Robertson and Wanner [215], respectively) as queries.

The *D. tinctorium* and *L. deliense* GR gene families were manually annotated according to TBLASTN and BLASTP searches (both with an E-value cutoff of $<1 \times 10^{-3}$) against their genome assemblies and predicted protein coding genes, respectively, using all *D. melanogaster* [216], *A. mellifera* [215], *I. scapularis* [195], *T. urticae* [100], *T. mercedesae* [98] and *M. occidentalis* [97] GRs as queries. An iterative search was also conducted with termite GRs as queries until no new genes were identified in each major subfamily or lineage. For phylogenetic analysis, *D. tinctorium* and *L. deliense* GRs were aligned with *D. melanogaster* GRs by Kalign [198]with default settings. Poorly-aligned and variable N-terminal and C-terminal regions, as well as several internal regions of highly variable sequences, were

996 excluded from the phylogenetic analysis. Other regions of potentially uncertain alignment between
997 these highly divergent proteins were retained, as removing these regions could potentially
998 compromise subfamily relationships. Based on the trimmed alignment, a PhyML tree was constructed
999 using the substitution model of LG determined by ProtTest [199]. Here, the SH-like local support
1000 method was used to assess the significance of phylogenetic clustering.

1001 The *D. tinctorium and L. deliense* iGluRs and IRs were manually annotated according to a TBLASTN (E-
1002 value cutoff, $<1 \times 10^{-3}$) search against the *D. tinctorium and L. deliense* genome assemblies using all
1003 iGluRs and IRs identified by Croset *et al*. [217] across vertebrates and invertebrates, as well as those
1004 identified in the recent *T. mercedesae* genome project [98]. Iterative searches were also conducted
1005 with termite iGluRs and IRs as queries until no new genes were identified in each major subfamily or
1006 lineage. In the phylogenetic analysis, all manually annotated *D. tinctorium and L. deliense* IRs and
1007 iGluRs were aligned with the *D. melanogaster* IRs and iGluRs by Mafft [197] using default settings.
1008 Phylogenetic analysis proceeded as for the GR genes described above, using the more conserved
1009 iGluRs to root the tree.

1010 Reference opsin genes were collected based on the work of Nagata *et al*. [96]. Opsin-like sequences
1011 in *M. occidentalis, I. scapularis* and *T. urticae* were obtained from the NCBI database. These opsin
1012 genes were classified by phylogenetic analysis using the neighbour-joining method with 1,000
1013 bootstraps. The multiple alignment of the amino acid sequences was carried out using Mafft [197]
1014 with the '--auto' option. The gaps deletion of the alignment was set to 75% in MEGA7 [201].

1015 Prediction of allergenic gene families

1016 Allergenic protein-coding genes in the genomes of acariform mites (*D. tinctorium*, *L. deliense, T.*
1017 *urticae, D. farinae, S. scabiei,* and *E. maynei*) were predicted using a standalone version of Allerdictor
1018 (v. 1.0) [218]. Because the predicted proteome of *D. farinae* is not publicly available, we used protein
1019 sequences identified from a new Trinity (v. 2.4) [219] assembly for the prediction of allergenic genes.
1020 The allergenic gene clusters were constructed using OrthoMCL (v. 1.4) [207] and individual protein

sequences were submitted to Pfam (EBI, v.31.0 [220]) to be assigned to protein families. Mapped Pfam domain identifications were then searched against the AllFam database [112] to retrieve corresponding AllFam identifiers for allergen families. Venn diagrams were constructed using InteractiVenn [221].

Salivary and cement protein analysis

A total of 159 non-redundant tick cement proteins were retrieved from the NCBI database and then clustered with all *D. tinctorium* and *L. deliense* amino-acid sequences using OrthoMCL in order to identify the tick cement orthologues in the new genomes. Salivary proteins of *I. scapularis* and *T. urticae* have been previously identified using proteomic methods by Kim *et al*. [120] and Jonckheere *et al*. [126], respectively. Proteins that might be present in the saliva of *D. tinctorium* and *L. deliense* were identified by clustering all predicted protein-coding sequences in the new genomes with these *T. urticae* and *I. scapularis* salivary proteins using OrthoMCL [207]. Venn diagrams were constructed using InteractiVenn [221].

Sample preparation for proteomics

Engorged larvae of *L. deliense* (10 specimens) were pooled from infested rodent specimens (*Bandicota indica, Bandicota savilei* and *Rattus tanezumi*) captured across several provinces in Thailand during the field studies of the CERoPath project [177]. The chigger samples were fixed in 70% ethanol and identified as *L. deliense* using autofluorescence microscopy [222]. For *D. tinctorium*, proteomic analysis was performed on a single ethanol-fixed individual from the same collection used for genome sequencing.

The chiggers were washed with chilled 50 mM ammonium bicarbonate. Soluble protein extracts were prepared by homgenisation in 0.1% w/v, Rapigest (Waters) in 50 mM ammonium bicarbonate using a polypropylene mini-pestle. This was followed by three cycles of sonication on ice (Vibra-cell 130PB sonicator at 20 Hz with microprobe; 10 sec of sonication alternating with 30 sec of incubation on ice).

Samples were centrifuged at 13,000 × g for 10 min at 4°C. The supernatant was removed and retained. The *D. tinctorium* individual was homogenised using a mini-pestle in lysis buffer (4% SDS, Tris-hydrochloride, pH 7.6) and sonicated and centrifuged as above. Supernatants from both mite preparations were stored at -80°C.

Protein concentrations of the samples were determined using a Bradford protein assay (Thermo Fisher Scientific). The *L. deliense* protein extract was reduced with 3 mM dithiothreitol (Sigma) at 60°C for 10 min, cooled, then alkylated with 9 mM iodoacetamide (Sigma) at room temperature for 30 min in the dark; all steps were performed with intermittent vortex-mixing. Proteomic-grade trypsin (Sigma) was added at a protein:trypsin ratio of 50:1 and incubated at 37°C overnight. Rapigest was removed by adding trifluoroacetic acid (TFA) to a final concentration of 0.5% (v/v). Peptide samples were centrifuged at 13,000 × g for 30 min to remove precipitated Rapigest. The *D. tinctorium* protein extract was reduced, alkylated and digested with trypsin using the filter-aided sample preparation approach [57]. Peptides from *D. tinctorium* were split into eight fractions using the Pierce High pH Reversed-Phase Peptide Fractionation Kit according to the manufacturer's instructions. Each digest and fraction was concentrated and desalted using C18 Stage tips (Thermo Fisher Scientific), then dried using a centrifugal vacuum concentrator (Eppendorf) and re-suspended in a 0.1% (v/v) TFA, 3% (v/v) acetonitrile solution.

Mass spectrometry

Peptides were analysed by on-line nanoflow LC using the Ultimate 3000 nano system (Dionex/Thermo Fisher Scientific). Samples were loaded onto a trap column (Acclaim PepMap 100, 2 cm × 75 µm inner diameter, C18, 3 µm, 100 Å) at 9µl /min with an aqueous solution containing 0.1 %(v/v) TFA and 2% (v/v) acetonitrile. After 3 min, the trap column was set in-line to an analytical column (Easy-Spray PepMap® RSLC 50 cm × 75 µm inner diameter, C18, 2 µm, 100 Å) fused to a silica nano-electrospray emitter (Dionex). The column was operated at a constant temperature of 35°C and the LC system was coupled to a Q-Exactive mass spectrometer (Thermo Fisher Scientific). Chromatography was

performed with a buffer system consisting of 0.1 % formic acid (buffer A) and 80 % acetonitrile in 0.1 % formic acid (buffer B). The peptides were separated by a linear gradient of 3.8 – 50 % buffer B over 90 minutes (*D. tinctorium* and *L. deliense* whole digests) or 30 min (*D. tinctorium* fractions) at a flow rate of 300 nl/min. The Q-Exactive was operated in data-dependent mode with survey scans acquired at a resolution of 70,000 at m/z 200. Scan range was 300 to 2000m/z. Up to the top 10 most abundant isotope patterns with charge states +2 to +5 from the survey scan were selected with an isolation window of 2.0 Th and fragmented by higher energy collisional dissociation with normalized collision energies of 30. The maximum ion injection times for the survey scan and the MS/MS scans were 250 and 50 ms, respectively, and the ion target value was set to 1×10^6 for survey scans and 1×10^4 for the MS/MS scans. The MS/MS events were acquired at a resolution of 17,500. Repetitive sequencing of peptides was minimized through dynamic exclusion of the sequenced peptides for 20 sec.

Protein identification, quantification and enrichment analysis

Thermo RAW files were imported into Progenesis LC–MS (version 4.1, Nonlinear Dynamics). Peaks were picked by the software using default settings and filtered to include only peaks with a charge state between +2 and +7. Spectral data were converted into .mgf files with Progenesis LC–MS and exported for peptide identification using the Mascot (version 2.3.02, Matrix Science) search engine as described above. Tandem MS data were searched against a database including translated ORFs from either the *D. tinctorium* genome (DinoT_V2_aug2016, 19,258 sequences; 8,386,445 residues) and a contaminant database (cRAP, GPMDB, 2012) (119 sequences; 40,423 residues), or the *L. deliense* genome (L_deliense_V2_Aug16, 15,096 sequences; 5,183,596 residues), the *Rattus norvegicus* genome (UniProt, Apr16 7,948 sequences; 4,022,300 residues) and a contaminant database (cRAP, GPMDB, 2012) (119 sequences; 40,423 residues). The search parameters were as follows: the precursor mass tolerance was set to 10 ppm and the fragment mass tolerance was set as 0.05 Da. Two missed tryptic cleavages were permitted. Carbamidomethylation (cysteine) was set as a fixed modification and oxidation (methionine) set as a variable modification. Mascot search results were

1095 further validated using the machine learning algorithm Percolator embedded within Mascot. The
1096 Mascot decoy database function was utilised and the false discovery rate was <1%, while individual
1097 percolator ion scores >13 indicated identity or extensive homology (*P* <0.05). Mascot search results
1098 were imported into Progenesis LC–MS as XML files. Fractions were combined using the Progenesis
1099 "combine analysed fractions" workflow. Relative protein abundance was calculated by the Hi-3 default
1100 method in Progenesis. Mass spectrometric data were deposited to the ProteomeXchange Consortium
1101 (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository [223] with the
1102 dataset identifier PXD008346.

1103 Enrichment of protein domains was assessed using Pfam (EBI, v.27.0 [220]) as previously described
1104 [224] using the gathering threshold as a cut-off. Briefly, a hypergeometric test for enrichment of Pfam
1105 domains in the observed proteome (for identifications supported by ≥2 unique peptides only) relative
1106 to the complete search database was performed using R (phyper). The Benjamini & Hochberg step-up
1107 false-discovery rate-controlling procedure was applied to the calculated *P* values [225], and
1108 enrichment was considered statistically significant where *P* <0.05.

1109 Availability of data and materials

1110 The data sets supporting the results of this article are available in the GigaDB repository associated
1111 with this publication.

1112 Abbreviations

1113 CTLD: C-type lectin domain; ERV: endogenous retrovirus; FLVCR: feline leukemia virus subgroup C
1114 receptor-related protein; GR: gustatory receptor; IMD: immune deficiency; IR: ionotropic receptor;
1115 iGluR: ionotropic glutamate receptors; LGT: lateral gene transfer; MYA: million years ago; NCBI:
1116 National Center for Biotechnology Information; OBP: odorant-binding protein; PE: paired-end; PERK:
1117 PKR-like endoplasmic reticulum kinase; PGRP: peptidoglycan recognition protein; SMOC: SPARC-

1118 related modular calcium-binding protein; TFA: trifluoroacetic acid; XIAP: X-linked inhibitor of
1119 apoptosis protein.

1120 Animal ethics

1121 Wild rodents trapped during the CERoPath project [177] and used as a source of chigger material were
1122 euthanized by inhaled anaesthetic overdose according to guidelines published by the American
1123 Veterinary Medical Association Council on Research [226] and the Canadian Council on Animal Care
1124 [227].

1125 Competing interests

1126 The authors declare that they have no competing interests.

1127 Funding

1128 This project was funded by Bayer PLC (Animal Health division) and the University of Liverpool. KC was
1129 the recipient of a Mahidol-Liverpool Chamlong Harinasuta Scholarship. The funding bodies had no role
1130 in the design of the study; the collection, analysis, and interpretation of data; or in the writing of the
1131 manuscript and the decision to publish.

1132 Authors' contributions

1133 Conceptualisation: A.C.D., B.L.M. Formal analysis: X.D., D.X., S.D.A., Y.F., A.C.D. Funding acquisition:
1134 J.W.M., A.C.D., B.L.M. Investigation: K.C., S.D.A., M.J.D. Project administration: A.C.D., B.L.M.
1135 Supervision: A.C.D., J.W.M., T.K., B.L.M. Validation: X.D., A.C.D., D.X., B.L.M. Visualization: X.D., D.X.,
1136 S.D.A., A.C.D., B.L.M. Writing (original draft): B.L.M., X.D. Writing (review and editing): B.L.M., X.D.,
1137 K.C., T.K., M.J.D., D.X., S.D.A., A.C.D. All authors have read and approved the final manuscript.

1138 Acknowledgements

1139 The authors are grateful to Serge Morand (Kasetsart University, Thailand) for managing the CERoPath
1140 project field collections that provided chigger material and to Joanna Mąkol (Wrocław University of
1141 Environmental and Life Sciences, Poland) for identification of the *D. tinctorium* specimens. We also
1142 thank the core research staff at the Centre for Genomic Research (University of Liverpool) for DNA
1143 library preparations, sequencing and initial data quality control.

Tables

Table 1: Characteristics of C-type lectin domain proteins in the trombidid mite genomes.

			No. of genes with characteristic			Motifs		
Cluster	**Species**	**No. genes in cluster**	**Secretion (%)[a]**	**Transmembrane domains (%)[b]**	***N*-glycosylation (%)[c]**	**EPN (%)**	**WND (%)**	**QPD (%)**
ORTHOMCL136	*D. tinctorium*	3	2 (66.7)	3 (100)	2 (66.7)	3 (100)	0	0
ORTHOMCL136	*L. deliense*	33	22 (66.7)	6 (18.2)	5 (15.2)	6 (18.2)	1 (3)	2 (6.1)
ORTHOMCL510	*L. deliense*	21	16 (76.2)	0	2 (9.5)	15 (71.4)	12 (57.1)	1 (4.8)
ORTHOMCL582	*L. deliense*	20	13 (65.0)	3 (15.0)	2 (10.0)	14 (70.0)	13 (65.0)	0
ORTHOMCL880	*L. deliense*	17	11 (64.7)	2 (11.8)	3 (17.6)	0	0	0

[a]Determined by SecretomeP 2.0 [209].

[b]Determined by TMHMM 2.0 [210].

[c]Determined by NetNGlyc 1.0 [211].

1149 **Table 2:** Comparison of chemosensory receptor repertoires between trombidid mites and 11 other
1150 arthropods.

	Chemosensory receptor[a]				
Species	**GR**	**OR**	**IR**	**OBP**	**CSP**
D. tinctorium	105	0	7	0	0
L. deliense	42	0	8	0	0
T. urticae	689	0	4	0	0
T. mercedesae	5	0	8	0	0
M. occidentalis	64	0	65	0	0
I. scapularis	60	0	22	0	1
S. maritima	77	0	60	0	2
D. pulex	53	0	85	0	3
D. melanogaster	73	62	66	51	4
A. mellifera	10	163	10	21	6
B. mori	56	48	18	44	18
A. pisum	53	48	11	15	13
P. humanus humanus	8	10	12	5	7

1151 [a]GR, gustatory receptor; OR, olfactory receptor; IR, ionotropic receptor; OBP, odorant-binding protein; CSP,
1152 chemosensory protein. Data for other genomes were obtained from [88, 97, 98, 100, 213].

1153 **Table 3**: Orthologues of *Ixodes scapularis* salivary proteins in trombidid mite genomes.

	Number of genes in cluster				
Orthologous cluster	***I. scapularis***	***L. deliense***	***D. tinctorium***	**Representative gene ID**	**Representative annotation[a]**
ORTHOMCL5799	1	2	1	EEC06447	CNDP dipeptidase
ORTHOMCL8488	1	1	1	EEC03480	Uncharacterized protein (peptidase M17, leucine aminopeptidase/peptidase B domain])
ORTHOMCL9	73	2	2	EEC08659	Sulfotransferase
ORTHOMCL2940	2	1	2	ISCW018873	Short-chain alcohol dehydrogenase
ORTHOMCL8622	1	1	1	JAA64643	Glycine-rich cell wall structural protein, partial [*Rhipicephalus pulchellus*]
ORTHOMCL3018	1	1	3	EEC14134	Calponin
ORTHOMCL5249	2	1	1	JAB84323	Stretch regulated skeletal muscle protein [*Ixodes ricinus*]
ORTHOMCL7670	1	1	1	EEC14106	Multifunctional chaperone
ORTHOMCL8293	1	1	1	EEC09272	Uncharacterised protein (aspartate dehydrogenase domain)
ORTHOMCL262	5	5	5	ISCW002881	UDP-sugar hydrolase
ORTHOMCL2418	1	1	4	EEC13206	Cytochrome C
ORTHOMCL5815	2	1	1	ISCW011750	Succinyl-CoA ligase beta subunit
ORTHOMCL3911	1	1	3	JAB73948	Formyltetrahydrofolate dehydrogenase, partial [*Ixodes ricinus*]
ORTHOMCL3015	2	1	2	ISCW001951	ATPase inhibitor
ORTHOMCL8	66	12	7	EEC10817	Acyl-CoA synthetase
ORTHOMCL260	13	1	1	ISCW001079	Acetylcholinesterase
ORTHOMCL6164	2	1	1	ISCW022662	NS1-binding protein
ORTHOMCL4904	2	1	1	ISCW006538	60S acidic ribosomal protein LP1
ORTHOMCL907	6	1	1	EEC05404	Translation initiation inhibitor UK114/IBM1
ORTHOMCL4900	2	1	1	JAB75945	Ribosomal protein LP2 [*Ixodes ricinus*]
ORTHOMCL3820	3	1	1	ISCW002028	IMP-GMP specific 5'-nucleotidase
ORTHOMCL4006	1	1	3	EEC13628	Urocanate hydratase
ORTHOMCL3946	1	1	3	EEC20451	Uncharacterised protein (pseudouridine-5'-phosphate glycosidase domain)

ORTHOMCL5111	2	1	1	ISCW010907	Formiminotransferase-cyclodeaminase
ORTHOMCL75	27	1	0	EEC02489	Cysteine-rich secreted protein (trypsin inhibitor-like cysteine rich domain)
ORTHOMCL8850	2	1	0	ISCW012704	Signal sequence receptor beta

[a]Annotations are from *I. scapularis* unless otherwise stated. Protein domain information was obtained from the Conserved Domain Database [228]. Tick salivary protein data were obtained from Kim *et al.* [120].

Figure legends

Figure 1: Phylogenetic tree based on the amino acid sequences of 527 one-to-one orthologous genes in 11 species of Ecdysozoa using Bayesian methods. The taxonomy of the trombidid mites follows the scheme of Lindquist *et al*. [14].

Figure 2: Simplified lifecycle of trombidid mites. In the Trombidioidea (velvet mites), the larvae are parasitic on other arthropods; whereas in the Trombiculoidea ("chiggers"), the larvae feed on a variety of vertebrates or (more rarely) other invertebrates. The deutonymph and adult are free-living, edaphic stages that predate soft-bodied arthropods (*e.g.,* termites, springtails and other mites) or consume their eggs. Trombidid eggs are laid in the environment and produce questing larvae that congregate and seek a host. For clarity, only the deutovum and active instars are shown; the protonymph (between the larva and deutonymph stages) and the tritonymph (between the deutonymph and adult stages) are calyptostatic.

Figure 3: Blob-plot of contigs assembled from sequence data derived from engorged *Leptotrombidium deliense* larvae. Blue = Eukaryota; red = Bacteria; white = other; grey = no hit.

Figure 4: Estimated divergence times using a relaxed molecular clock with fossil calibration time and classification of protein-coding genes between 11 species of Ecdysozoa. *Caenorhabditis elegans* was used as the outgroup and the bootstrap value was set as 10,000,000. The 1:1:1 orthologs comprise the common orthologs with the same copy numbers in different species, and the N:N:N orthologs comprise the common orthologs with different copy numbers in these species. Patchy orthologs are shared between more than one, but not all species (excluding those belonging to the previous categories). Unclustered genes are those that cannot be clustered into gene families.

Figure 5: Overrepresented Pfam domains in proteomic datasets generated from a single adult *Dinothrombium tinctorium* and a pool of engorged *Leptotrombidium deliense* larvae. The colour scale represents fold-enrichment.

Figure 6: Phylogeny of terpene synthases from cluster ORTHOMCL47 of trombidid mites alongside related genes from bacteria and fungi. The tree was constructed using a maximum likelihood method; poorly-supported nodes are highlighted in red font.

Figure 7: Phylogeny of terpene synthases from cluster ORTHOMCL881 of *Leptotrombidium deliense* alongside related genes from bacteria. Metazoan terpene synthases from galerucid beetles are shown as an outgroup. The tree was constructed using a maximum likelihood method; poorly-supported nodes are highlighted in red font.

Figure 8: Blob-plot of read coverage and GC content for terpene synthase genes in *Dinothrombium tinctorium* (Dt) and *Leptotrombidium deliense* (Ld). Terpene synthases (in red, "Bacteria") are shown in relation to all other mite genes (in blue).

Figure 9: Phylogeny of Gag-like polyproteins from *Leptotrombidium deliense* in relation to homologous sequences from small mammals. The tree was constructed using a maximum likelihood method; poorly-supported nodes are highlighted in red font.

Figure 10: Phylogeny and classification of metazoan opsins. The tree was constructed using a neighbour-joining method. Poorly-supported nodes are highlighted in red font.

Figure 11: Phylogeny of *Dinothrombium tinctorium*, *Leptotrombidium deliense*, *Tetranychus urticae* and *Drosophila melanogaster* ionotropic receptors and ionotropic glutamate receptors. The tree was constructed using a maximum-likelihood method.

Figure 12: Venn diagram of orthologous clusters of predicted allergens from six species of acariform mites.

Figure 13: Orthologous clusters of predicted allergens in *Leptotrombidium deliense* as classified by the AllFam database. "Peptides" refers to the number of unique peptides from each allergen detected by mass spectrometric analysis of a pool of *L. deliense* larvae. The heat-map indicates the number of orthologues for each predicted allergen in the genomes of six acariform mites. Ld, *Leptotrombidium*

1204 *deliense*; Dt, *Dinothrombium tinctorium*; Df, *Dermatophagoides farinae*; Tu, *Tetranychus urticae*; Ss,

1205 *Sarcoptes scabiei*; Em, *Euroglyphus maynei*.

1206 **Figure 14:** Venn diagram of orthologous clusters of putative salivary proteins in four species of Acari.

1207 One-to-one orthologues of salivary proteins from *Ixodes scapularis* (Kim et al.) and *Tetranychus urticae*

1208 (Jonckheere) were identified in the genomes of *Leptotrombidium deliense* and *Dinothrombium*

1209 *tinctorium.* Thumbnail images indicate representative host species.

Additional files

Additional file 1: Genome assembly and gene set statistics compared with 14 other arachnids (table in .xlsx format).

Additional file 2: *K*-mer distributions for *Leptotrombidium deliense* (*A*) and *Dinothrombium tinctorium* (*B*) plotted by GenomeScope (figure in .pdf format).

Additional file 3: Identification of repetitive sequences in the *Dinothrombium tinctorium* and *Leptotrombidium deliense* assemblies compared with other acariform mites (table in .pdf format).

Additional file 4: The number of gene families shared among acariform mites (*Dinothrombium tinctorium, Leptotrombidium deliense, Tetranychus urticae* and *Sarcoptes scabiei*); alongside other references including *Drosophila melanogaster, Apis mellifera, Tropilaelaps mercedesae, Metaseiulus occidentalis, Ixodes scapularis, Stegodyphus mimosarum* and *Caenorhabditis elegans* by the OrthoMCL classification algorithm (figure in .pdf format).

Additional file 5: Gene family contraction and expansion in 11 species of Ecdysozoa (figure in .pdf format).

Additional file 6: Changes of gene family size in trombidid mites in comparison with three other acariform mites (table in .xlsx format).

Additional file 7: High-confidence protein identifications and abundance scores for *Leptotrombidium deliense* engorged larvae (spreadsheet in .xlsx format).

Additional file 8: High-confidence protein identifications and abundance scores for a single adult *D. tinctorium* individual (spreadsheet in .xlsx format).

Additional file 9: Phylogeny of carotenoid synthases-cyclases from trombidid mites, spider mites, aphids and fungi (figure in .pdf format).

1232 **Additional file 10**: Genomic scaffold of *Dinothrombium tinctorium* containing a putative lateral gene
1233 transfer adjacent to an incontrovertible mite gene (figure in .pdf format).

1234 **Additional file 11**: Microbial reads identified in the trombidid genomic data by the Kraken taxonomic
1235 sequence classification system (table in .pdf format).

1236 **Additional file 12**: Peptides detected by mass spectrometry from two terpene synthases in an adult
1237 specimen of *Dinothrombium tinctorium* (figure in .pdf format).

1238 **Additional file 13**: Phylogeny of reverse ribonuclease integrases in *Dinothrombium tinctorium* and
1239 their closest homologues in other taxa (figure in .pdf format).

1240 **Additional file 14**: Phylogeny of Pol-like polyproteins in trombidid mites and their closest homologues
1241 in other taxa (figure in .pdf format).

1242 **Additional file 15**: RNA families identified in the Rfam database in 10 arthropod genomes (table in
1243 .xlsx format).

1244 **Additional file 16**: Phylogeny of Dscam protein-coding sequences in *Dinothrombium tinctorium,*
1245 *Leptotrombidium deliense, Tetranychus urticae* and *Ixodes scapularis* (figure in .pdf format).

1246 **Additional file 17**: Phylogeny of peptidoglycan recognition protein sequences in *Dinothrombium*
1247 *tinctorium, Leptotrombidium deliense, Tetranychus urticae* and *Ixodes scapularis* alongside
1248 homologous sequences from insects.

1249 **Additional file 18**: Phylogeny of PKR-like endoplasmic reticulum kinase sequences in *Leptotrombidium*
1250 *deliense* and their closest homologues in other taxa (figure in .pdf format).

1251 **Additional file 19**: Predicted allergens in the *Leptotrombidium deliense* genome with top BLAST hit,
1252 Pfam domains and AllFam classifications (spreadsheet in .xlsx format).

1253 **Additional file 20**: Orthologous clusters of tick cement proteins in the genomes of *Dinothrombium*
1254 *tinctorium* and *Leptotrombidium deliense* (table in .pdf format).

References

1. Krantz GW. Introduction. In: Krantz GW, Walter, D.E., editor. A Manual of Acarology. Lubbock, Texas: Texas Tech University Press; 2009. p. 1-2.
2. Danchin EGJ. Lateral gene transfer in eukaryotes: tip of the iceberg or of the ice cube? Bmc Biol. 2016;14.
3. Ku C and Martin WF. A natural barrier to lateral gene transfer from prokaryotes to eukaryotes revealed from genomes: the 70 % rule. Bmc Biol. 2016;14 1:89. doi:10.1186/s12915-016-0315-9.
4. Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C, et al. No evidence for extensive horizontal gene transfer in the genome of the tardigrade Hypsibius dujardini. Proc Natl Acad Sci U S A. 2016;113 18:5053-8. doi:10.1073/pnas.1600338113.
5. Boothby TC, Tenlen JR, Smith FW, Wang JR, Patanella KA, Nishimura EO, et al. Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. Proc Natl Acad Sci U S A. 2015;112 52:15976-81. doi:10.1073/pnas.1510461112.
6. Crisp A, Boschetti C, Perry M, Tunnacliffe A and Micklem G. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. Genome Biol. 2015;16:50. doi:10.1186/s13059-015-0607-3.
7. Salzberg SL. Horizontal gene transfer is not a hallmark of the human genome. Genome Biol. 2017;18 1:85. doi:10.1186/s13059-017-1214-2.
8. Acuna R, Padilla BE, Florez-Ramos CP, Rubio JD, Herrera JC, Benavides P, et al. Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. Proc Natl Acad Sci U S A. 2012;109 11:4197-202. doi:10.1073/pnas.1121190109.
9. Haegeman A, Jones JT and Danchin EG. Horizontal gene transfer in nematodes: a catalyst for plant parasitism? Mol Plant Microbe Interact. 2011;24 8:879-87. doi:10.1094/MPMI-03-11-0055.
10. Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, et al. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell. 2013;153 7:1567-78. doi:10.1016/j.cell.2013.05.040.
11. Wybouw N, Pauchet Y, Heckel DG and Van Leeuwen T. Horizontal Gene Transfer Contributes to the Evolution of Arthropod Herbivory. Genome Biol Evol. 2016;8 6:1785-801. doi:10.1093/gbe/evw119.
12. Wybouw N, Dermauw W, Tirry L, Stevens C, Grbic M, Feyereisen R, et al. A gene horizontally transferred from bacteria protects arthropods from host plant cyanide poisoning. Elife. 2014;3:e02365. doi:10.7554/eLife.02365.
13. Bryon A, Kurlovs AH, Dermauw W, Greenhalgh R, Riga M, Grbic M, et al. Disruption of a horizontally transferred phytoene desaturase abolishes carotenoid accumulation and diapause in Tetranychus urticae. Proc Natl Acad Sci U S A. 2017;114 29:E5871-E80. doi:10.1073/pnas.1706865114.
14. Lindquist EE, Krantz, G.W., Walter, D.E. Classification. In: Krantz GW, Walter, D.E., editor. A Manual of Acarology. Lubbock, Texas: Texas Tech University Press; 2009. p. 97-103.
15. Snell AE and Heath ACG. Parasitism of mosquitoes (Diptera : Culicidae) by larvae of Arrenuridae and Microtrombidiidae (Acari : Parasitengona) in the Wellington region, New Zealand. New Zeal J Zool. 2006;33 1:9-15.
16. Mcgarry JW, Gusbi AM, Baker A, Hall MJR and Elmegademi K. Phoretic and Parasitic Mites Infesting the New-World Screwworm Fly, Cochliomyia-Hominivorax, Following Sterile Insect Releases in Libya. Med Vet Entomol. 1992;6 3:255-60.
17. Goldarazena A and Zhang ZQ. Seasonal abundance of Allothrombium monochaetum and Allothrombium pulvinum in Navarra-Nafarroa (northern Spain), with notes on larval host preference and rate of parasitism. Exp Appl Acarol. 1999;23 12:987-93.

18. Zhang ZQ. Biology and ecology of trombidiid mites (Acari : Trombidioidea). Experimental & Applied Acarology. 1998;22 3:139-55.
19. Zhang ZQ and Xin JL. Biology of Allothrombium-Pulvinum (Acariformes, Trombidiidae), a Potential Biological-Control Agent of Aphids in China. Experimental & Applied Acarology. 1989;6 2:101-8.
20. Shatrov AB, Kudryashova, N.I. Taxonomy, life cycles and the origin of parasitism in trombiculid mites. In: Morand S, Krasnov, B.R, Poulin, R, editor. Micromammals and Macroparasites: From Evolutionary Ecology to Management. Tokyo: Springer Japan; 2006. p. 119-40.
21. Nadchatram M. Correlation of habitat, environment and color of chiggers, and their potential significance in the epidemiology of scrub typhus in Malaya (Prostigmata: Trombiculidae). J Med Entomol. 1970;7 2:131-44.
22. Lakshana P. A new species of trombiculid mite infesting scorpions in Thailand (Acarina, Trombiculidae). J Med Entomol. 1966;3 3:258-60.
23. Audy JR. Trombiculid mites infesting birds, reptiles, and arthropods in Malaya, with a taxonomic revision, and descriptions of a new genus, two new subgenera, and six new species. Bulletin of the Raffles Museum. 1956;28:27-80.
24. Xu G, Walker DH, Jupiter D, Melby PC and Arcari CM. A review of the global epidemiology of scrub typhus. Plos Neglect Trop D. 2017;11 11.
25. Tilak R, Kunwar R, Wankhade UB and Tilak VW. Emergence of Schoengastiella ligula as the vector of scrub typhus outbreak in Darjeeling: has Leptotrombidium deliense been replaced? Indian J Public Health. 2011;55 2:92-9. doi:10.4103/0019-557X.85239.
26. Chaisiri K, Stekolnikov AA, Makepeace BL and Morand S. A Revised Checklist of Chigger Mites (Acari: Trombiculidae) From Thailand, with the Description of Three New Species. J Med Entomol. 2016;53 2:321-42. doi:10.1093/jme/tjv244.
27. Taylor AJ, Paris DH and Newton PN. A Systematic Review of Mortality from Untreated Scrub Typhus (Orientia tsutsugamushi). PLoS Negl Trop Dis. 2015;9 8:e0003971. doi:10.1371/journal.pntd.0003971.
28. Bonell A, Lubell Y, Newton PN, Crump JA and Paris DH. Estimating the burden of scrub typhus: A systematic review. PLoS Negl Trop Dis. 2017;11 9:e0005838. doi:10.1371/journal.pntd.0005838.
29. Weitzel T, Dittrich S, Lopez J, Phuklia W, Martinez-Valdebenito C, Velasquez K, et al. Endemic Scrub Typhus in South America. N Engl J Med. 2016;375 10:954-61. doi:10.1056/NEJMoa1603657.
30. Izzard L, Fuller A, Blacksell SD, Paris DH, Richards AL, Aukkanit N, et al. Isolation of a novel Orientia species (O. chuto sp. nov.) from a patient infected in Dubai. J Clin Microbiol. 2010;48 12:4404-9. doi:10.1128/JCM.01526-10.
31. Maina AN, Farris CM, Odhiambo A, Jiang J, Laktabai J, Armstrong J, et al. Q Fever, Scrub Typhus, and Rickettsial Diseases in Children, Kenya, 2011-2012. Emerg Infect Dis. 2016;22 5:883-6. doi:10.3201/eid2205.150953.
32. Yu XJ and Tesh RB. The role of mites in the transmission and maintenance of Hantaan virus (Hantavirus: Bunyaviridae). J Infect Dis. 2014;210 11:1693-9. doi:10.1093/infdis/jiu336.
33. Kabeya H, Colborn JM, Bai Y, Lerdthusnee K, Richardson JH, Maruyama S, et al. Detection of Bartonella tamiae DNA in Ectoparasites from Rodents in Thailand and Their Sequence Similarity with Bacterial Cultures from Thai Patients. Vector-Borne Zoonot. 2010;10 5:429-34.
34. Huang Y, Zhao L, Zhang Z, Liu M, Xue Z, Ma D, et al. Detection of a Novel Rickettsia From Leptotrombidium scutellare Mites (Acari: Trombiculidae) From Shandong of China. J Med Entomol. 2017;54 3:544-9. doi:10.1093/jme/tjw234.

35. Smith GA, Sharma V, Knapp JF and Shields BJ. The summer penile syndrome: seasonal acute hypersensitivity reaction caused by chigger bites on the penis. Pediatr Emerg Care. 1998;14 2:116-8.
36. Little SE, Carmichael KP and Rakich PM. Trombidiosis-induced dermatitis in white-tailed deer (Odocoileus virginianus). Vet Pathol. 1997;34 4:350-2. doi:10.1177/030098589703400412.
37. Leone F, Di Bella A, Vercelli A and Cornegliani L. Feline trombiculosis: a retrospective study in 72 cats. Vet Dermatol. 2013;24 5:535-e126. doi:10.1111/vde.12053.
38. Faccini JL, Santos AC, Santos SB, Jacinavicius FC, Bassini-Silva R and Barros-Battesti DM. Trombiculiasis in domestic goats and humans in the state of Maranhao, Brazil. Rev Bras Parasitol Vet. 2017;26 1:104-9. doi:10.1590/S1984-29612016088.
39. Shatrov AB. Stylostome formation in trombiculid mites (Acariformes: Trombiculidae). Exp Appl Acarol. 2009;49 4:261-80.
40. Shatrov AB and Felska M. Comparative stylostome ultrastructure of Hirsutiella zachvatkini (Trombiculidae) and Trombidium holosericeum (Trombidiidae) larvae. Exp Appl Acarol. 2017;72 4:339-65.
41. Minter DM. Some field and laboratory observations on the British harvest mite, Trombicula autumnalis Shaw. Parasitology. 1957;47 1-2:185-93.
42. Singer G. Neotrombidium leonardi (Acarina: Trombidioidea), Pt. 2 : Biology of N. beeri and other species of the genus. Acarologia. 1971;13:143-51.
43. Tevis L and Newell IM. Studies on Biology and Seasonal Cycle of Giant Red Velvet Mite, Dinothrombium-Pandorae (Acari, Trombidiidae). Ecology. 1962;43 3:497-&.
44. Zhang ZQ. Notes on the Occurrence and Distribution of the Biocontrol Agent, Allothrombium-Pulvinum Ewing (Acari, Trombidiidae), in a Peach Orchard in China. J Appl Entomol. 1992;113 1:13-7.
45. Wiggins GJ, Grant JF and Welbourn WC. Allothrombium mitchelli (Acari : Trombidiidae) in the Great Smoky Mountains National Park: Incidence, seasonality, and predation on beech scale (Homoptera : Eriococcidae). Ann Entomol Soc Am. 2001;94 6:896-901.
46. Traub R and Wisseman CL, Jr. Ecological considerations in scrub typhus. 2. Vector species. Bull World Health Organ. 1968;39 2:219-30.
47. Walter DE, Lindquist, E.E., Smith, I.M., Cook, D.R., Krantz, G.W. Order Trombidiformes. In: Krantz GW, Walter, D.E., editor. A Manual of Acarology. Lubbock, Texas: Texas Tech University Press; 2009. p. 233-419.
48. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM and Birol I. ABySS: A parallel assembler for short read sequence data. Genome Res. 2009;19 6:1117-23.
49. Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, et al. De novo assembly and analysis of RNA-seq data. Nat Methods. 2010;7 11:909-U62.
50. Zerbino DR and Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18 5:821-9.
51. Kumar S, Jones M, Koutsovoulos G, Clarke M and Blaxter M. Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. Front Genet. 2013;4:237. doi:10.3389/fgene.2013.00237.
52. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19 5:455-77. doi:10.1089/cmb.2012.0021.
53. Cantarel B, Korf I, Robb S, Parra G, Ross E, Moore B, et al. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Research. 2008;18 1:188-96. doi:10.1101/gr.6743907.
54. Stanke M and Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 2005;33 Web Server issue:W465-7. doi:10.1093/nar/gki458.

55. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004;5:59. doi:10.1186/1471-2105-5-59.
56. Lukashin AV and Borodovsky M. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. 1998;26 4:1107-15.
57. Wisniewski JR, Zougman A, Nagaraj N and Mann M. Universal sample preparation method for proteome analysis. Nat Methods. 2009;6 5:359-62. doi:10.1038/nmeth.1322.
58. Rider SD, Jr., Morgan MS and Arlian LG. Draft genome of the scabies mite. Parasit Vectors. 2015;8:585. doi:10.1186/s13071-015-1198-2.
59. Grbic M, Van Leeuwen T, Clark RM, Rombauts S, Rouze P, Grbic V, et al. The genome of Tetranychus urticae reveals herbivorous pest adaptations. Nature. 2011;479 7374:487-92. doi:10.1038/nature10640.
60. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV and Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31 19:3210-2. doi:10.1093/bioinformatics/btv351.
61. Sanggaard KW, Bechsgaard JS, Fang X, Duan J, Dyrlund TF, Gupta V, et al. Spider genomes provide insight into composition and evolution of venom and silk. Nat Commun. 2014;5:3765. doi:10.1038/ncomms4765.
62. Oba Y, Sato M, Ojika M and Inouye S. Enzymatic and genetic characterization of firefly luciferase and Drosophila CG6178 as a fatty acyl-CoA synthetase. Biosci Biotechnol Biochem. 2005;69 4:819-28. doi:10.1271/bbb.69.819.
63. Okada I, Hamanoue H, Terada K, Tohma T, Megarbane A, Chouery E, et al. SMOC1 is essential for ocular and limb development in humans and mice. Am J Hum Genet. 2011;88 1:30-41. doi:10.1016/j.ajhg.2010.11.012.
64. Thomas JT, Eric Dollins D, Andrykovich KR, Chu T, Stultz BG, Hursh DA, et al. SMOC can act as both an antagonist and an expander of BMP signaling. Elife. 2017;6 doi:10.7554/eLife.17935.
65. Ayme-Southgate A, Lasko P, French C and Pardue ML. Characterization of the gene for mp20: a Drosophila muscle protein that is not found in asynchronous oscillatory flight muscle. J Cell Biol. 1989;108 2:521-31.
66. Huang Y, Li W, Huang L, Hu Y, Chen W, Wang X, et al. Identification and characterization of myophilin-like protein: a life stage and tissue-specific antigen of Clonorchis sinensis. Parasitol Res. 2012;111 3:1143-50. doi:10.1007/s00436-012-2946-2.
67. Friedrich MV, Schneider M, Timpl R and Baumgartner S. Perlecan domain V of Drosophila melanogaster. Sequence, recombinant analysis and tissue expression. Eur J Biochem. 2000;267 11:3149-59.
68. Klinkowstrom AM, Terra WR and Ferreira C. Midgut Dipeptidases from Rhynchosciara-Americana (Diptera) Larvae - Properties of Soluble and Membrane-Bound Forms. Insect Biochem Molec. 1995;25 3:303-10.
69. Long JZ, Svensson KJ, Bateman LA, Lin H, Kamenecka T, Lokurkar IA, et al. The Secreted Enzyme PM20D1 Regulates Lipidated Amino Acid Uncouplers of Mitochondria. Cell. 2016;166 2:424-35. doi:10.1016/j.cell.2016.05.071.
70. Shahein YE, El-Rahim MT, Hussein NA, Hamed RR, El-Hakim AE and Barakat MM. Molecular cloning of a small heat shock protein (sHSPII) from the cattle tick Rhipicephalus (Boophilus) annulatus salivary gland. Int J Biol Macromol. 2010;47 5:614-22. doi:10.1016/j.ijbiomac.2010.08.006.
71. Matsumoto I, Watanabe H, Abe K, Arai S and Emori Y. A putative digestive cysteine proteinase from Drosophila melanogaster is predominantly expressed in the embryonic and larval midgut. Eur J Biochem. 1995;227 1-2:582-7.
72. Santamaria ME, Hernandez-Crespo P, Ortego F, Grbic V, Grbic M, Diaz I, et al. Cysteine peptidases and their inhibitors in Tetranychus urticae: a comparative genomic approach. BMC Genomics. 2012;13:307. doi:10.1186/1471-2164-13-307.

73. Santamaria ME, Gonzalez-Cabrera J, Martinez M, Grbic V, Castanera P, Diaz L, et al. Digestive proteases in bodies and faeces of the two-spotted spider mite, Tetranychus urticae. J Insect Physiol. 2015;78:69-77. doi:10.1016/j.jinsphys.2015.05.002.

74. Kuwahara Y, Ichiki Y, Morita M, Tanabe T and Asano Y. Chemical Polymorphism in Defense Secretions during Ontogenetic Development of the Millipede Niponia nodulosa. J Chem Ecol. 2015;41 1:15-21.

75. Kenny NJ, Shen X, Chan TT, Wong NW, Chan TF, Chu KH, et al. Genome of the Rusty Millipede, Trigoniulus corallinus, Illuminates Diplopod, Myriapod, and Arthropod Evolution. Genome Biol Evol. 2015;7 5:1280-95. doi:10.1093/gbe/evv070.

76. Skelton AC, Cameron MM, Pickett JA and Birkett MA. Identification of neryl formate as the airborne aggregation pheromone for the American house dust mite and the European house dust mite (Acari: Epidermoptidae). J Med Entomol. 2010;47 5:798-804.

77. Chan TF, Ji KM, Yim AK, Liu XY, Zhou JW, Li RQ, et al. The draft genome, transcriptome, and microbiome of Dermatophagoides farinae reveal a broad spectrum of dust mite allergens. J Allergy Clin Immunol. 2015;135 2:539-48. doi:10.1016/j.jaci.2014.09.031.

78. Beran F, Rahfeld P, Luck K, Nagel R, Vogel H, Wielsch N, et al. Novel family of terpene synthases evolved from trans-isoprenyl diphosphate synthases in a flea beetle. Proc Natl Acad Sci U S A. 2016;113 11:2922-7. doi:10.1073/pnas.1523468113.

79. Faddeeva-Vakhrusheva A, Kraaijeveld K, Derks MFL, Anvar SY, Agamennone V, Suring W, et al. Coping with living in the soil: the genome of the parthenogenetic springtail Folsomia candida. BMC Genomics. 2017;18 1:493. doi:10.1186/s12864-017-3852-x.

80. Komatsu M, Tsuda M, Omura S, Oikawa H and Ikeda H. Identification and functional analysis of genes controlling biosynthesis of 2-methylisoborneol. Proc Natl Acad Sci U S A. 2008;105 21:7422-7. doi:10.1073/pnas.0802312105.

81. Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, et al. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. Nature. 2010;463 7284:1079-83. doi:10.1038/nature08742.

82. Mason VC, Li G, Minx P, Schmitz J, Churakov G, Doronina L, et al. Genomic analysis reveals hidden biodiversity within colugos, the sister group to primates. Sci Adv. 2016;2 8:e1600633. doi:10.1126/sciadv.1600633.

83. Webb CH and Luptak A. HDV-like self-cleaving ribozymes. RNA Biol. 2011;8 5:719-27. doi:10.4161/rna.8.5.16226.

84. Valanne S, Wang JH and Ramet M. The Drosophila Toll signaling pathway. J Immunol. 2011;186 2:649-56. doi:10.4049/jimmunol.1002302.

85. Palmer WJ and Jiggins FM. Comparative Genomics Reveals the Origins and Diversity of Arthropod Immune Systems. Mol Biol Evol. 2015;32 8:2111-29. doi:10.1093/molbev/msv093.

86. Shaw DK, Wang X, Brown LJ, Chavez AS, Reif KE, Smith AA, et al. Infection-derived lipids elicit an immune deficiency circuit in arthropods. Nat Commun. 2017;8:14401. doi:10.1038/ncomms14401.

87. Brites D and Du Pasquier L. Somatic and Germline Diversification of a Putative Immunoreceptor within One Phylum: Dscam in Arthropods. Results Probl Cell Differ. 2015;57:131-58. doi:10.1007/978-3-319-20819-0_6.

88. Chipman AD, Ferrier DE, Brena C, Qu J, Hughes DS, Schroder R, et al. The first myriapod genome sequence reveals conservative arthropod gene content and genome organisation in the centipede Strigamia maritima. PLoS Biol. 2014;12 11:e1002005. doi:10.1371/journal.pbio.1002005.

89. Kaneko T, Yano T, Aggarwal K, Lim JH, Ueda K, Oshima Y, et al. PGRP-LC and PGRP-LE have essential yet distinct functions in the drosophila immune response to monomeric DAP-type peptidoglycan. Nat Immunol. 2006;7 7:715-23. doi:10.1038/ni1356.

90. Atwal S, Giengkam S, Chaemchuen S, Dorling J, Kosaisawe N, VanNieuwenhze M, et al. Evidence for a peptidoglycan-like structure in Orientia tsutsugamushi. Mol Microbiol. 2017;105 3:440-52. doi:10.1111/mmi.13709.
91. Zelensky AN and Gready JE. The C-type lectin-like domain superfamily. FEBS J. 2005;272 24:6179-217. doi:10.1111/j.1742-4658.2005.05031.x.
92. Alenton RR, Koiwai K, Miyaguchi K, Kondo H and Hirono I. Pathogen recognition of a novel C-type lectin from Marsupenaeus japonicus reveals the divergent sugar-binding specificity of QAP motif. Sci Rep. 2017;7:45818. doi:10.1038/srep45818.
93. Gale M, Jr. and Katze MG. Molecular mechanisms of interferon resistance mediated by viral-directed inhibition of PKR, the interferon-induced protein kinase. Pharmacol Ther. 1998;78 1:29-46.
94. Pomar N, Berlanga JJ, Campuzano S, Hernandez G, Elias M and de Haro C. Functional characterization of Drosophila melanogaster PERK eukaryotic initiation factor 2alpha (eIF2alpha) kinase. Eur J Biochem. 2003;270 2:293-306.
95. Chakraborty R, Baek JH, Bae EY, Kim WY, Lee SY and Kim MG. Comparison and contrast of plant, yeast, and mammalian ER stress and UPR. Appl Biol Chem. 2016;59 3:337-47.
96. Nagata T, Koyanagi M, Tsukamoto H and Terakita A. Identification and characterization of a protostome homologue of peropsin from a jumping spider. J Comp Physiol A. 2010;196 1:51-9.
97. Hoy MA, Waterhouse RM, Wu K, Estep AS, Ioannidis P, Palmer WJ, et al. Genome Sequencing of the Phytoseiid Predatory Mite Metaseiulus occidentalis Reveals Completely Atomized Hox Genes and Superdynamic Intron Evolution. Genome Biology and Evolution. 2016;8 6:1762-75.
98. Dong XF, Armstrong SD, Xia D, Makepeace BL, Darby AC and Kadowaki T. Draft genome of the honey bee ectoparasitic mite, Tropilaelaps mercedesae, is shaped by the parasitic life history. Gigascience. 2017;6 3.
99. Ozaki K, Utoguchi A, Yamada A and Yoshikawa H. Identification and genomic structure of chemosensory proteins (CSP) and odorant binding proteins (OBP) genes expressed in foreleg tarsi of the swallowtail butterfly Papilio xuthus. Insect Biochem Molec. 2008;38 11:969-76.
100. Ngoc PCT, Greenhalgh R, Dermauw W, Rombauts S, Bajda S, Zhurov V, et al. Complex Evolutionary Dynamics of Massively Expanded Chemosensory Receptor Families in an Extreme Generalist Chelicerate Herbivore. Genome Biology and Evolution. 2016;8 11:3323-39.
101. Xia SZ, Miyashita T, Fu TF, Lin WY, Wu CL, Pyzocha L, et al. NMDA receptors mediate olfactory learning and memory in Drosophila. Curr Biol. 2005;15 7:603-15.
102. Frank CA. Homeostatic plasticity at the Drosophila neuromuscular junction. Neuropharmacology. 2014;78:63-74.
103. Robinson JE, Paluch J, Dickman DK and Joiner WJ. ADAR-mediated RNA editing suppresses sleep by acting as a brake on glutamatergic synaptic plasticity. Nature Communications. 2016;7.
104. Karuppudurai T, Lin TY, Ting CY, Pursley R, Melnattur KV, Diao FQ, et al. A Hard-Wired Glutamatergic Circuit Pools and Relays UV Signals to Mediate Spectral Preference in Drosophila. Neuron. 2014;81 3:603-15.
105. Hu W, Wang TT, Wang X and Han JH. I-h Channels Control Feedback Regulation from Amacrine Cells to Photoreceptors. Plos Biology. 2015;13 4.
106. Rytz R, Croset V and Benton R. Ionotropic Receptors (IRs): Chemosensory ionotropic glutamate receptors in Drosophila and beyond. Insect Biochem Molec. 2013;43 9:888-97.
107. Ni L, Klein M, Svec KV, Budelli G, Chang EC, Ferrer AJ, et al. The Ionotropic Receptors IR21a and IR25a mediate cool sensing in Drosophila. Elife. 2016;5.

108. Knecht ZA, Silbering AF, Ni LN, Klein M, Budelli G, Bell R, et al. Distinct combinations of variant ionotropic glutamate receptors mediate thermosensation and hygrosensation in Drosophila. Elife. 2016;5.

109. Eliash N, Singh NK, Thangarajan S, Sela N, Leshkowitz D, Kamer Y, et al. Chemosensing of honeybee parasite, Varroa destructor: Transcriptomic analysis. Sci Rep-Uk. 2017;7.

110. Wright SM, Wikel SK and Wrenn WJ. Host Immune Responsiveness to the Chigger, Eutrombicula-Cinnabaris. Ann Trop Med Parasit. 1988;82 3:283-93.

111. Rider SD, Jr., Morgan MS and Arlian LG. Allergen homologs in the Euroglyphus maynei draft genome. Plos One. 2017;12 8:e0183535. doi:10.1371/journal.pone.0183535.

112. Radauer C, Bublin M, Wagner S, Mari A and Breiteneder H. Allergens are distributed into few protein families and possess a restricted number of biochemical functions. J Allergy Clin Immunol. 2008;121 4:847-52 e7. doi:10.1016/j.jaci.2008.01.025.

113. An S, Chen L, Long C, Liu X, Xu X, Lu X, et al. Dermatophagoides farinae allergens diversity identification by proteomics. Mol Cell Proteomics. 2013;12 7:1818-28. doi:10.1074/mcp.M112.027136.

114. Choopong J, Reamtong O, Sookrung N, Seesuay W, Indrawattana N, Sakolvaree Y, et al. Proteome, Allergenome, and Novel Allergens of House Dust Mite, Dermatophagoides farinae. J Proteome Res. 2016;15 2:422-30. doi:10.1021/acs.jproteome.5b00663.

115. Fluckiger S, Fijten H, Whitley P, Blaser K and Crameri R. Cyclophilins, a new family of cross-reactive allergens. Eur J Immunol. 2002;32 1:10-7. doi:10.1002/1521-4141(200201)32:1<10::AID-IMMU10>3.0.CO;2-I.

116. Tripathi P, Nair S, Singh BP and Arora N. Molecular and immunological characterization of subtilisin like serine protease, a major allergen of Curvularia lunata. Immunobiology. 2011;216 3:402-8. doi:10.1016/j.imbio.2010.06.009.

117. Woodfolk JA, Wheatley LM, Piyasena RV, Benjamin DC and Platts-Mills TA. Trichophyton antigens associated with IgE antibodies and delayed type hypersensitivity. Sequence homology to two families of serine proteinases. J Biol Chem. 1998;273 45:29489-96.

118. Teng F, Yu L, Bian Y, Sun J, Wu J, Ling C, et al. In silico structural analysis of group 3, 6 and 9 allergens from Dermatophagoides farinae. Mol Med Rep. 2015;11 5:3559-64. doi:10.3892/mmr.2015.3166.

119. Tsai LC, Chao PL, Hung MW, Sun YC, Kuo IC, Chua KY, et al. Protein sequence analysis and mapping of IgE and IgG epitopes of an allergenic 98-kDa Dermatophagoides farinae paramyosin, Der f 11. Allergy. 2000;55 2:141-7.

120. Kim TK, Tirloni L, Pinto AF, Moresco J, Yates JR, 3rd, da Silva Vaz I, Jr., et al. Ixodes scapularis Tick Saliva Proteins Sequentially Secreted Every 24 h during Blood Feeding. PLoS Negl Trop Dis. 2016;10 1:e0004323. doi:10.1371/journal.pntd.0004323.

121. Tirloni L, Kim TK, Pinto AFM, Yates JR, 3rd, da Silva Vaz I, Jr. and Mulenga A. Tick-Host Range Adaptation: Changes in Protein Profiles in Unfed Adult Ixodes scapularis and Amblyomma americanum Saliva Stimulated to Feed on Different Hosts. Front Cell Infect Microbiol. 2017;7:517. doi:10.3389/fcimb.2017.00517.

122. Radulovic ZM, Kim TK, Porter LM, Sze SH, Lewis L and Mulenga A. A 24-48 h fed Amblyomma americanum tick saliva immuno-proteome. BMC Genomics. 2014;15:518. doi:10.1186/1471-2164-15-518.

123. Esteves E, Maruyama SR, Kawahara R, Fujita A, Martins LA, Righi AA, et al. Analysis of the Salivary Gland Transcriptome of Unfed and Partially Fed Amblyomma sculptum Ticks and Descriptive Proteome of the Saliva. Front Cell Infect Microbiol. 2017;7:476. doi:10.3389/fcimb.2017.00476.

124. Tan AW, Francischetti IM, Slovak M, Kini RM and Ribeiro JM. Sexual differences in the sialomes of the zebra tick, Rhipicephalus pulchellus. J Proteomics. 2015;117:120-44. doi:10.1016/j.jprot.2014.12.014.

125. Garcia GR, Gardinassi LG, Ribeiro JM, Anatriello E, Ferreira BR, Moreira HN, et al. The sialotranscriptome of Amblyomma triste, Amblyomma parvum and Amblyomma cajennense ticks, uncovered by 454-based RNA-seq. Parasit Vectors. 2014;7:430. doi:10.1186/1756-3305-7-430.

126. Jonckheere W, Dermauw W, Zhurov V, Wybouw N, Van den Bulcke J, Villarroel CA, et al. The Salivary Protein Repertoire of the Polyphagous Spider Mite Tetranychus urticae: A Quest for Effectors. Mol Cell Proteomics. 2016;15 12:3594-613. doi:10.1074/mcp.M116.058081.

127. Hunter DJ, Torkelson JL, Bodnar J, Mortazavi B, Laurent T, Deason J, et al. The Rickettsia Endosymbiont of Ixodes pacificus Contains All the Genes of De Novo Folate Biosynthesis. Plos One. 2015;10 12:e0144552. doi:10.1371/journal.pone.0144552.

128. Rio RV, Attardo GM and Weiss BL. Grandeur Alliances: Symbiont Metabolic Integration and Obligate Arthropod Hematophagy. Trends Parasitol. 2016;32 9:739-49. doi:10.1016/j.pt.2016.05.002.

129. Temeyer KB and Tuckow AP. Tick Salivary Cholinesterase: A Probable Immunomodulator of Host-parasite Interactions. J Med Entomol. 2016;53 3:500-4. doi:10.1093/jme/tjv252.

130. Junger WG. Immune cell regulation by autocrine purinergic signalling. Nat Rev Immunol. 2011;11 3:201-12. doi:10.1038/nri2938.

131. Pichu S, Yalcin EB, Ribeiro JM, King RS and Mather TN. Molecular characterization of novel sulfotransferases from the tick, Ixodes scapularis. BMC Biochem. 2011;12:32. doi:10.1186/1471-2091-12-32.

132. Thompson RE, Liu X, Ripoll-Rozada J, Alonso-Garcia N, Parker BL, Pereira PJB, et al. Tyrosine sulfation modulates activity of tick-derived thrombin inhibitors. Nat Chem. 2017;9 9:909-17. doi:10.1038/nchem.2744.

133. Blisnick AA, Foulon T and Bonnet SI. Serine Protease Inhibitors in Ticks: An Overview of Their Role in Tick Biology and Tick-Borne Pathogen Transmission. Front Cell Infect Microbiol. 2017;7:199. doi:10.3389/fcimb.2017.00199.

134. Fogaca AC, Almeida IC, Eberlin MN, Tanaka AS, Bulet P and Daffre S. Ixodidin, a novel antimicrobial peptide from the hemocytes of the cattle tick Boophilus microplus with inhibitory activity against serine proteinases. Peptides. 2006;27 4:667-74. doi:10.1016/j.peptides.2005.07.013.

135. Sasaki SD, de Lima CA, Lovato DV, Juliano MA, Torquato RJ and Tanaka AS. BmSI-7, a novel subtilisin inhibitor from Boophilus microplus, with activity toward Pr1 proteases from the fungus Metarhizium anisopliae. Exp Parasitol. 2008;118 2:214-20. doi:10.1016/j.exppara.2007.08.003.

136. Rapoport TA. Protein transport across the endoplasmic reticulum membrane: facts, models, mysteries. Faseb J. 1991;5 13:2792-8.

137. Bullard R, Allen P, Chao CC, Douglas J, Das P, Morgan SE, et al. Structural characterization of tick cement cones collected from in vivo and artificial membrane blood-fed Lone Star ticks (Amblyomma americanum). Ticks Tick Borne Dis. 2016;7 5:880-92. doi:10.1016/j.ttbdis.2016.04.006.

138. Kim TK, Curran J and Mulenga A. Dual silencing of long and short Amblyomma americanum acidic chitinase forms weakens the tick cement cone stability. J Exp Biol. 2014;217 Pt 19:3493-503. doi:10.1242/jeb.107979.

139. Kim JH, Roh JY, Kwon DH, Kim YH, Yoon KA, Yoo S, et al. Estimation of the genome sizes of the chigger mites Leptotrombidium pallidum and Leptotrombidium scutellare based on quantitative PCR and k-mer analysis. Parasit Vectors. 2014;7:279. doi:10.1186/1756-3305-7-279.

140. Niedzwiedzki G, Szrek P, Narkiewicz K, Narkiewicz M and Ahlberg PE. Tetrapod trackways from the early Middle Devonian period of Poland. Nature. 2010;463 7277:43-8. doi:10.1038/nature08623.

141. Konikiewicz M and Makol J. A fossil Paratrombiinae mite (Actinotrichida: Trombidioidea) from the Rovno amber, Ukraine. Zootaxa. 2014;3847 4:583-9. doi:10.11646/zootaxa.3847.4.8.
142. Karahadian C, Josephson DB and Lindsay RC. Volatile Compounds from Penicillium Sp Contributing Musty Earthy Notes to Brie and Camembert Cheese Flavors. J Agr Food Chem. 1985;33 3:339-43.
143. Gerber NN. Volatile substances from actinomycetes: their role in the odor pollution of water. CRC Crit Rev Microbiol. 1979;7 3:191-214.
144. Gerber NN and Lechevalier HA. Geosmin, an earthly-smelling substance isolated from actinomycetes. Appl Microbiol. 1965;13 6:935-8.
145. Tabachek JAL and Yurkowski M. Isolation and Identification of Blue-Green-Algae Producing Muddy Odor Metabolites, Geosmin, and 2-Methylisoborneol, in Saline Lakes in Manitoba. J Fish Res Board Can. 1976;33 1:25-35.
146. Stensmyr MC, Dweck HK, Farhan A, Ibba I, Strutz A, Mukunda L, et al. A conserved dedicated olfactory circuit for detecting harmful microbes in Drosophila. Cell. 2012;151 6:1345-57. doi:10.1016/j.cell.2012.09.046.
147. Ashitani T, Garboui SS, Schubert F, Vongsombath C, Liblikas I, Palsson K, et al. Activity studies of sesquiterpene oxides and sulfides from the plant Hyptis suaveolens (Lamiaceae) and its repellency on Ixodes ricinus (Acari: Ixodidae). Exp Appl Acarol. 2015;67 4:595-606. doi:10.1007/s10493-015-9965-5.
148. Birkett MA, Abassi SA, Krober T, Chamberlain K, Hooper AM, Guerin PM, et al. Antiectoparasitic activity of the gum resin, gum haggar, from the East African plant, Commiphora holtziana. Phytochemistry. 2008;69 8:1710-5. doi:10.1016/j.phytochem.2008.02.017.
149. Baldin ELL, Souza ES, Silva JPGF, Pavarini DP, Lopes NP, Lopes JLC, et al. TOWARDS NEW BOTANICAL PESTICIDES: THE TOXIC EFFECT OF Eremanthus goyazensis (Asteraceae) LEAVES ESSENTIAL OIL AGAINST Brevipalpus phoenicis (Acari: Tenuipalpidae). Quim Nova. 2012;35 11:2254-7.
150. Huang RL, Li ZH, Wang SY, Fu JT, Cheng DM and Zhang ZX. Insecticidal effect of volatile compounds from plant materials of Murraya exotica against Red Imported Fire Ant Workers. Sociobiology. 2016;63 2:783-91.
151. Cloudsley-Thompson JL. Some aspects of the physiology and behaviour of Dinothrombium (Acari). Entomologia Experimentalis et Applicata. 1962;5:69-73.
152. Sferra NJ. 1st Record of Pterodontia-Flavipes (Diptera, Acroceridae) Larvae in the Mites Podothrombium (Acari, Trombidiidae) and Abrolophus (Acari, Erythraeidae). Entomol News. 1986;97 3:121-3.
153. Ozawa R, Shimoda T, Kawaguchi M, Arimura G, Horiuchi J, Nishioka T, et al. Lotus japonicus infested with herbivorous mites emits volatile compounds that attract predatory mites. J Plant Res. 2000;113 1112:427-33.
154. Royalty RN, Phelan PL and Hall FR. Arrestment of Male 2-Spotted Spider-Mite Caused by Female Sex-Pheromone. J Chem Ecol. 1992;18 2:137-53.
155. Wohltmann A. The evolution of life histories in Parasitengona (Acari: Prostigmata). Acarologia. 2000;41 1-2:145-204.
156. Moniuszko H and Makol J. Host-parasite association in trombiculid mites (Actinotrichida: Trombiculidae) of temperate zone - the case of Hirsutiella zachvatkini (Schluger, 1948); are we dealing with prolonged contact with the host? Parasit Vectors. 2016;9:61. doi:10.1186/s13071-016-1339-2.
157. Khan AA and Quigley JG. Heme and FLVCR-related transporter families SLC48 and SLC49. Mol Aspects Med. 2013;34 2-3:669-82. doi:10.1016/j.mam.2012.07.013.
158. Mandilaras K and Missirlis F. Genes for iron metabolism influence circadian rhythms in Drosophila melanogaster. Metallomics. 2012;4 9:928-36. doi:10.1039/c2mt20065a.

159. Alves-Bezerra M, Klett EL, De Paula IF, Ramos IB, Coleman RA and Gondim KC. Long-chain acyl-CoA synthetase 2 knockdown leads to decreased fatty acid oxidation in fat body and reduced reproductive capacity in the insect Rhodnius prolixus. Biochim Biophys Acta. 2016;1861 7:650-62. doi:10.1016/j.bbalip.2016.04.007.

160. Bryon A, Kurlovs AH, Van Leeuwen T and Clark RM. A molecular-genetic understanding of diapause in spider mites: current knowledge and future directions. Physiol Entomol. 2017;42 3:211-24.

161. Royalty RN, Phelan PL and Hall FR. Comparative Effects of Form, Color, and Pheromone of 2-Spotted Spider-Mite Quiescent Deutonymphs on Male Guarding Behavior. Physiol Entomol. 1993;18 3:303-16.

162. Goto SG. Physiological and molecular mechanisms underlying photoperiodism in the spider mite: comparisons with insects. J Comp Physiol B. 2016;186 8:969-84. doi:10.1007/s00360-016-1018-9.

163. Hori Y, Numata H, Shiga S and Goto SG. Both the anterior and posterior eyes function as photoreceptors for photoperiodic termination of diapause in the two-spotted spider mite. J Comp Physiol A Neuroethol Sens Neural Behav Physiol. 2014;200 2:161-7. doi:10.1007/s00359-013-0872-0.

164. Terakita A and Nagata T. Functional properties of opsins and their contribution to light-sensing physiology. Zoolog Sci. 2014;31 10:653-9. doi:10.2108/zs140094.

165. Ni JD, Baik LS, Holmes TC and Montell C. A rhodopsin in the brain functions in circadian photoentrainment in Drosophila. Nature. 2017;545 7654:340-4. doi:10.1038/nature22325.

166. Tanji T, Ohashi-Kobayashi A and Natori S. Participation of a galactose-specific C-type lectin in Drosophila immunity. Biochem J. 2006;396 1:127-38. doi:10.1042/BJ20051921.

167. Costa FH, Valenca NS, Silva AR, Bezerra GA, Cavada BS and Radis-Baptista G. Cloning and molecular modeling of Litopenaeus vannamei (Penaeidae) C-type lectin homologs with mutated mannose binding domain-2. Genet Mol Res. 2011;10 2:650-64. doi:10.4238/vol10-2gmr999.

168. Nakamura T, Furuhashi M, Li P, Cao H, Tuncman G, Sonenberg N, et al. Double-stranded RNA-dependent protein kinase links pathogen sensing with stress and metabolic homeostasis. Cell. 2010;140 3:338-48. doi:10.1016/j.cell.2010.01.001.

169. Pena J and Harris E. Dengue virus modulates the unfolded protein response in a time-dependent manner. J Biol Chem. 2011;286 16:14226-36. doi:10.1074/jbc.M111.222703.

170. Walsh AM, Kortschak RD, Gardner MG, Bertozzi T and Adelson DL. Widespread horizontal transfer of retrotransposons. Proc Natl Acad Sci U S A. 2013;110 3:1012-6. doi:10.1073/pnas.1205856110.

171. Santos-Matos G, Wybouw N, Martins NE, Zele F, Riga M, Leitao AB, et al. Tetranychus urticae mites do not mount an induced immune response against bacteria. Proc Biol Sci. 2017;284 1856 doi:10.1098/rspb.2017.0401.

172. Yang L and Zhu R. Immunotherapy of house dust mite allergy. Hum Vaccin Immunother. 2017;13 10:2390-6. doi:10.1080/21645515.2017.1364823.

173. de la Fuente J and Contreras M. Tick vaccines: current status and future directions. Expert Rev Vaccines. 2015;14 10:1367-76. doi:10.1586/14760584.2015.1076339.

174. Burgess ST, Nunn F, Nath M, Frew D, Wells B, Marr EJ, et al. A recombinant subunit vaccine for the control of ovine psoroptic mange (sheep scab). Vet Res. 2016;47:26. doi:10.1186/s13567-016-0315-3.

175. Bartley K, Wright HW, Huntley JF, Manson ED, Inglis NF, McLean K, et al. Identification and evaluation of vaccine candidate antigens from the poultry red mite (Dermanyssus gallinae). Int J Parasitol. 2015;45 13:819-30. doi:10.1016/j.ijpara.2015.07.004.

176. Wongprompitak P, Duong V, Anukool W, Sreyrath L, Mai TT, Gavotte L, et al. Orientia tsutsugamushi, agent of scrub typhus, displays a single metapopulation with maintenance

of ancestral haplotypes throughout continental South East Asia. Infect Genet Evol. 2015;31:1-8. doi:10.1016/j.meegid.2015.01.005.

177. Auffray JC, Blasdell, K.R., Bordes, F., Chabé, M., Chaisiri, K., Charbonnel, N., Claude, J., Dei-Cas, E., Desquesnes, M., Dobigny, G., Douangboupha, B., Galan, M., Haukisalmi, V., Henttonen, H., Hugot, J.P., Jiyipong, T., Latinne, A., Michaux, J., Milocco, C., Morand, S., Pagès, M., Phoophitpong, D., Pumhom, P., Ribas Salvador, A., Soonchan, S., Suputtamongkol, Y., Waengsothorn, S., Waywa, D., Xuereb, A. Protocols for field and laboratory rodent studies. Bangkok: Kasetsart University Press; 2011.

178. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011;17 1:10-2.

179. Joshi NA, Fass, J.N.: Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. https://github.com/najoshi/sickle (2011). Accessed 24/01/18.

180. Wood DE and Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15 3:R46. doi:10.1186/gb-2014-15-3-r46.

181. Marcais G and Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27 6:764-70. doi:10.1093/bioinformatics/btr011.

182. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 2017;33 14:2202-4. doi:10.1093/bioinformatics/btx153.

183. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A. 2011;108 4:1513-8. doi:10.1073/pnas.1017351108.

184. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012;1 1:18. doi:10.1186/2047-217X-1-18.

185. Weisenfeld NI, Yin S, Sharpe T, Lau B, Hegarty R, Holmes L, et al. Comprehensive variation discovery in single human genomes. Nat Genet. 2014;46 12:1350-5. doi:10.1038/ng.3121.

186. Langmead B and Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9 4:357-9. doi:10.1038/nmeth.1923.

187. Parra G, Bradnam K and Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 2007;23 9:1061.

188. Smit A, Hubley, R.: RepeatModeler: De-novo repeat discovery tool. https://github.com/rmhubley/RepeatModeler (2017). Accessed 24/01/18.

189. Tarailo-Graovac M and Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. 2009;Chapter 4:Unit 4 10. doi:10.1002/0471250953.bi0410s25.

190. Zdobnov EM and Apweiler R. InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics. 2001;17 9:847-8.

191. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M and Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21 18:3674-6. doi:10.1093/bioinformatics/bti610.

192. Kanehisa M and Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research. 2000;27 1:29-34(6).

193. Adams M, Celniker S, Holt R, Evans C, Gocayne J, Amanatides P, et al. The genome sequence of Drosophila melanogaster. Science. 2000;287 5461:2185-95. doi:10.1126/science.287.5461.2185.

194. Weinstock G, Robinson G, Gibbs R, Worley K, Evans J, Maleszka R, et al. Insights into social insects from the genome of the honeybee Apis mellifera. Nature. 2006;443 7114:931-49. doi:10.1038/nature05260.

195. Gulia-Nuss M, Nuss AB, Meyer JM, Sonenshine DE, Roe RM, Waterhouse RM, et al. Genomic insights into the Ixodes scapularis tick vector of Lyme disease. Nat Commun. 2016;7:10507. doi:10.1038/ncomms10507.

196. Stein L, Sternberg P, Durbin R, Thierry-Mieg J and Spieth J. WormBase: network access to the genome and biology of Caenorhabditis elegans. Nucleic Acids Research. 2001;29 1:82-6. doi:10.1093/nar/29.1.82.

197. Katoh K and Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30 4:772-80. doi:10.1093/molbev/mst010.

198. Lassmann T, Frings O and Sonnhammer EL. Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features. Nucleic Acids Res. 2009;37 3:858-65. doi:10.1093/nar/gkn1006.

199. Darriba D, Taboada GL, Doallo R and Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics. 2011;27 8:1164-5. doi:10.1093/bioinformatics/btr088.

200. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W and Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59 3:307-21. doi:10.1093/sysbio/syq010.

201. Kumar S, Stecher G and Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology & Evolution. 2016;33 7:1870.

202. Dabert M, Witalinski W, Kazmierski A, Olszanowski Z and Dabert J. Molecular phylogeny of acariform mites (Acari, Arachnida): strong conflict between phylogenetic signal and long-branch attraction artifacts. Mol Phylogenet Evol. 2010;56 1:222-41. doi:10.1016/j.ympev.2009.12.020.

203. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000;17 4:540-52.

204. Ronquist F and Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2003;19 12:1572-4.

205. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24 8:1586-91. doi:10.1093/molbev/msm088.

206. Benton M, Donoghue, PCJ, Asher, RJ. Calibrating and constraining molecular clocks. In: Hedges SB, Kumar, S., editor. The timetree of life. Oxford: Oxford University Press; 2009. p. 35 - 86.

207. Li L, Stoeckert CJ and Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13 9:2178-89. doi:10.1101/gr.1224503.

208. De Bie T, Cristianini N, Demuth JP and Hahn MW. CAFE: a computational tool for the study of gene family evolution. Bioinformatics. 2006;22 10:1269-71. doi:10.1093/bioinformatics/btl097.

209. Bendtsen JD, Jensen LJ, Blom N, Von Heijne G and Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng Des Sel. 2004;17 4:349-56. doi:10.1093/protein/gzh037.

210. Krogh A, Larsson B, von Heijne G and Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305 3:567-80. doi:10.1006/jmbi.2000.4315.

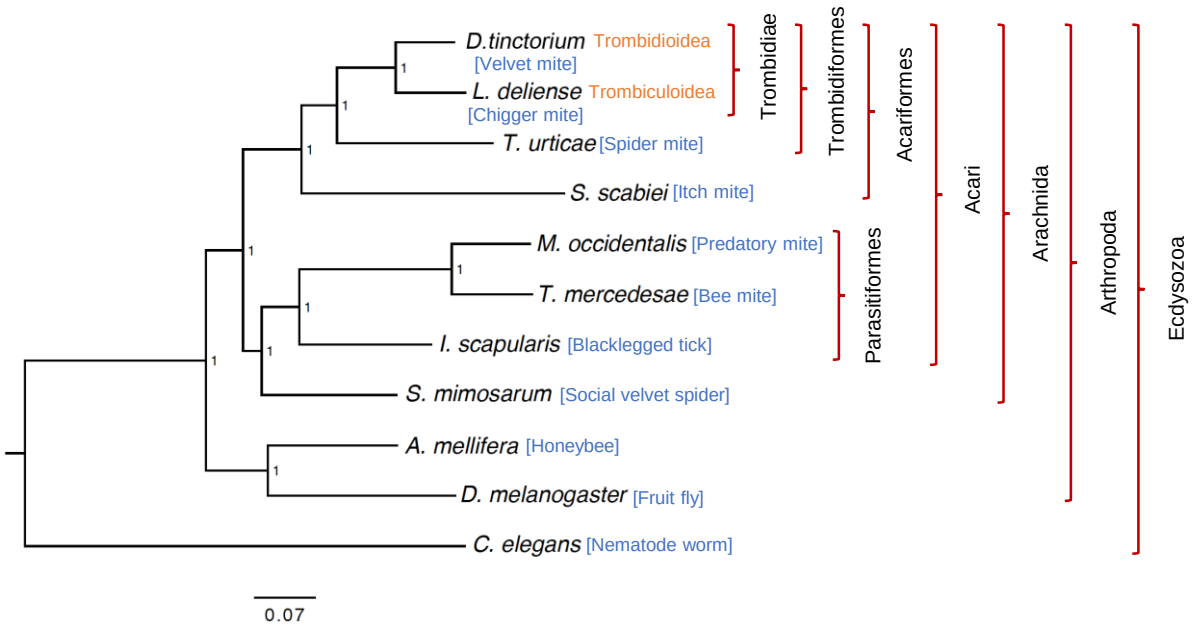
211. Gupta R, Jung, E., Brunak, S.: NetNGlyc 1.0 server. http://www.cbs.dtu.dk/services/NetNGlyc/ (2004). Accessed 25/01/18.

212. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res. 2017;45 D1:D190-D9. doi:10.1093/nar/gkw1107.

213. Sánchez-Gracia A, Vieira FG, Almeida FC and Rozas J. Comparative Genomics of the Major Chemosensory Gene Families in Arthropods. John Wiley & Sons, Ltd; 2011.

214. Nozawa M and Nei M. Evolutionary dynamics of olfactory receptor genes in Drosophila species. Proceedings of the National Academy of Sciences of the United States of America. 2007;104 17:7122.
215. Robertson HM and Wanner KW. The chemoreceptor superfamily in the honey bee, Apis mellifera: Expansion of the odorant, but not gustatory, receptor family. Genome Research. 2006;16 11:1395.
216. Robertson HM, Warr CG and Carlson JR. Molecular evolution of the insect chemoreceptor gene superfamily in Drosophila melanogaster. Proceedings of the National Academy of Sciences of the United States of America. 2003;100 Suppl 2 100 Suppl 2:14537.
217. Croset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, et al. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. Plos Genetics. 2010;6 8:e1001064.
218. Dang HX and Lawrence CB. Allerdictor: fast allergen prediction using text classification techniques. Bioinformatics. 2014;30 8:1120-8. doi:10.1093/bioinformatics/btu004.
219. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29 7:644-52. doi:10.1038/nbt.1883.
220. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42 Database issue:D222-30. doi:10.1093/nar/gkt1223.
221. Heberle H, Meirelles GV, da Silva FR, Telles GP and Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. BMC Bioinformatics. 2015;16:169. doi:10.1186/s12859-015-0611-3.
222. Park SW, Ha NY, Ryu B, Bang JH, Song H, Kim Y, et al. Urbanization of scrub typhus disease in South Korea. PLoS Negl Trop Dis. 2015;9 5:e0003814. doi:10.1371/journal.pntd.0003814.
223. Ternent T, Csordas A, Qi D, Gomez-Baena G, Beynon RJ, Jones AR, et al. How to submit MS proteomics data to ProteomeXchange via the PRIDE database. Proteomics. 2014;14 20:2233-41. doi:10.1002/pmic.201400120.
224. Armstrong SD, Babayan SA, Lhermitte-Vallarino N, Gray N, Xia D, Martin C, et al. Comparative analysis of the secretome from a model filarial nematode (Litomosoides sigmodontis) reveals maximal diversity in gravid female parasites. Mol Cell Proteomics. 2014;13 10:2527-44. doi:10.1074/mcp.M114.038539.
225. Benjamini Y and Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met. 1995;57 1:289-300.
226. Golab GC, Patterson-Kane, E. AVMA Guidelines for the Euthanasia of Animals (2013 Edition). 2013.0.1 ed. Schaumburg, Illinois: American Veterinary Medical Association; 2013.
227. Charbonneau R, Niel, L., Olfert, E., von Keyserlingk, M., Griffin, G. CCAC guidelines on: euthanasia of animals used in science. Ottawa: Canadian Council on Animal Care; 2010.
228. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 2017;45 D1:D200-D3. doi:10.1093/nar/gkw1129.

Figure 1


D.tinctorium Trombidioidea
[Velvet mite]
L. deliense Trombiculoidea
[Chigger mite]
T. urticae [Spider mite]
S. scabiei [Itch mite]
M. occidentalis [Predatory mite]
T. mercedesae [Bee mite]
I. scapularis [Blacklegged tick]
S. mimosarum [Social velvet spider]
A. mellifera [Honeybee]
D. melanogaster [Fruit fly]
C. elegans [Nematode worm]
1
0.07
Trombidiae
Trombidiformes
Acariformes
Parasitiformes
Acari
Arachnida
Arthropoda
Ecdysozoa

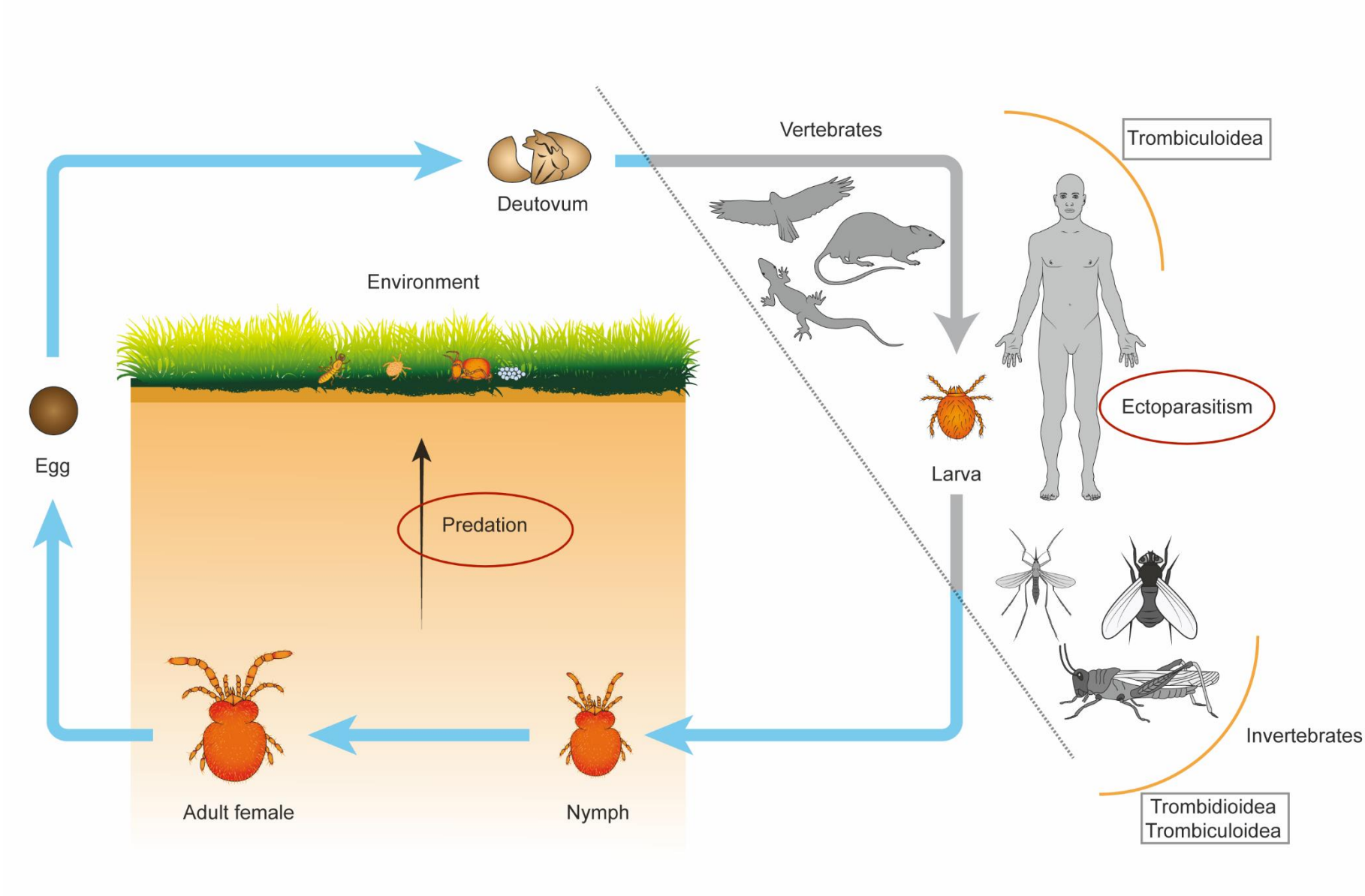

Deutovum
Vertebrates
Trombiculoidea
Environment
Egg
Predation
Larva
Ectoparasitism
Adult female
Nymph
Invertebrates
Trombidioidea
Trombiculoidea

Figure 3

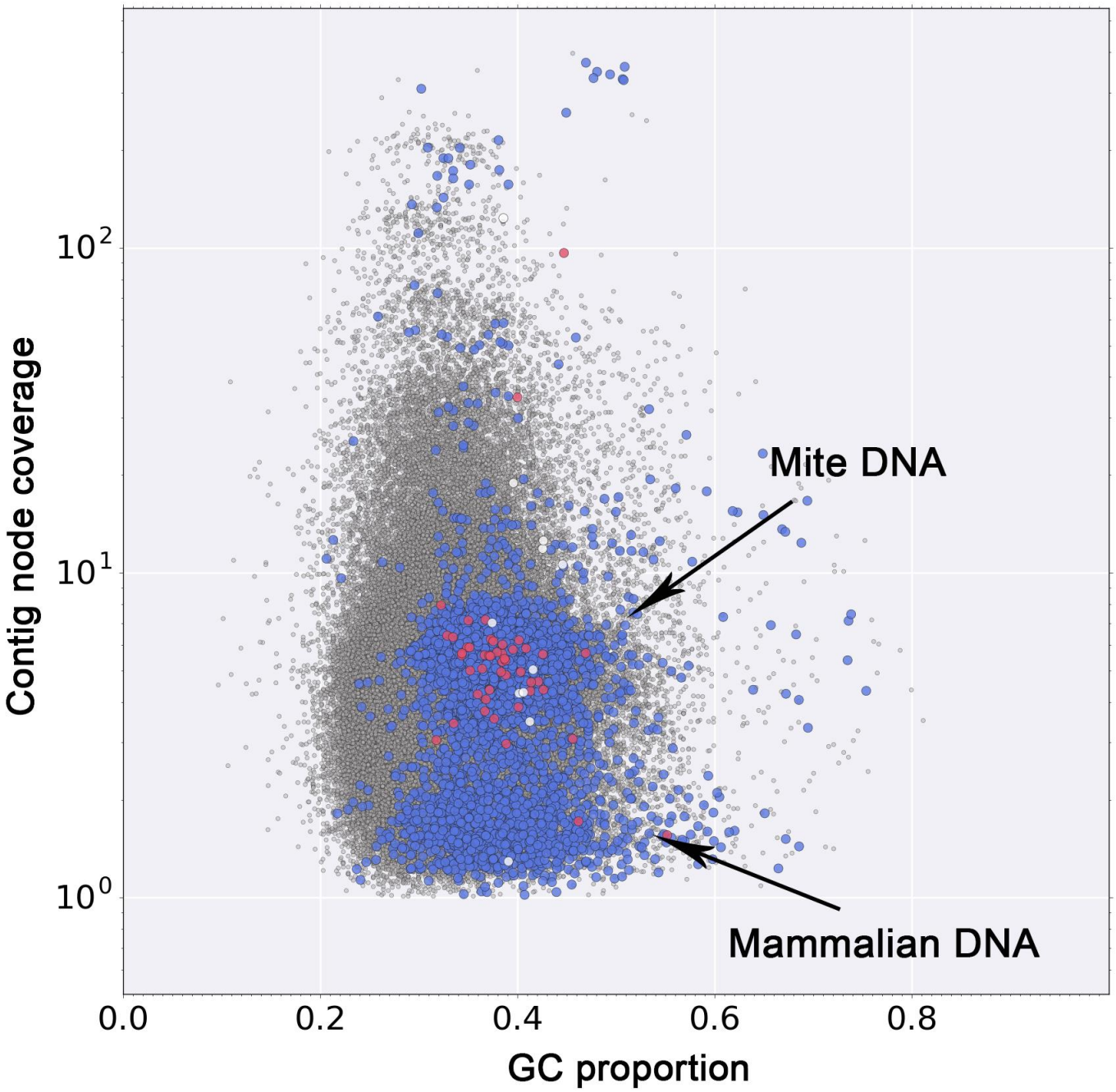
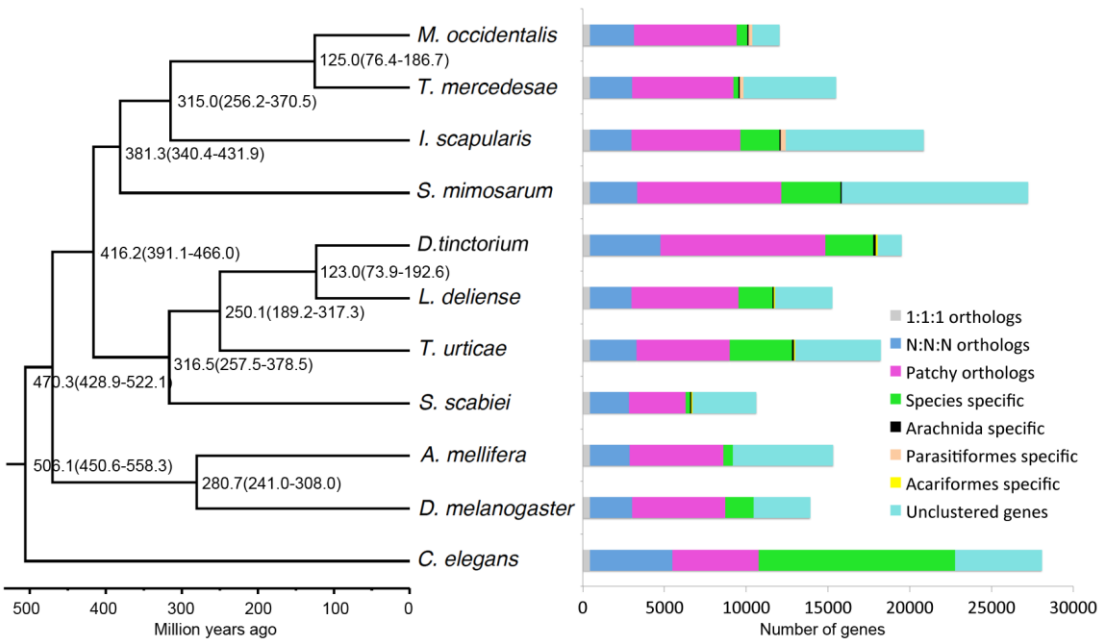

Contig node coverage
10^2
10^1
10^0
Mite DNA
Mammalian DNA
0.0
0.2
0.4
0.6
0.8
GC proportion

Figure 4


M. occidentalis
T. mercedesae
I. scapularis
S. mimosarum
D.tinctorium
L. deliense
T. urticae
S. scabiei
A. mellifera
D. melanogaster
C. elegans
125.0(76.4-186.7)
315.0(256.2-370.5)
381.3(340.4-431.9)
416.2(391.1-466.0)
123.0(73.9-192.6)
250.1(189.2-317.3)
316.5(257.5-378.5)
470.3(428.9-522.1)
506.1(450.6-558.3)
280.7(241.0-308.0)
500
400
300
200
100
0
Million years ago
1:1:1 orthologs
N:N:N orthologs
Patchy orthologs
Species specific
Arachnida specific
Parasitiformes specific
Acariformes specific
Unclustered genes
0
5000
10000
15000
20000
25000
30000
Number of genes

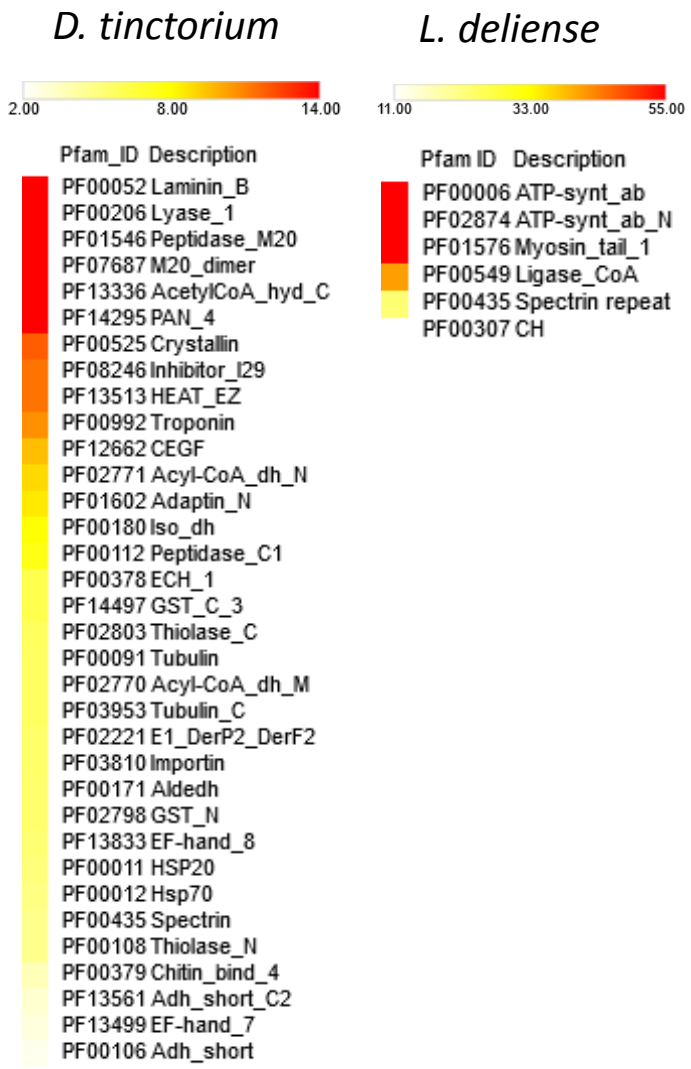

D. tinctorium
2.00
8.00
14.00
Pfam_ID Description
PF00052 Laminin_B
PF00206 Lyase_1
PF01546 Peptidase_M20
PF07687 M20_dimer
PF13336 AcetylCoA_hyd_C
PF14295 PAN_4
PF00525 Crystallin
PF08246 Inhibitor_I29
PF13513 HEAT_EZ
PF00992 Troponin
PF12662 CEGF
PF02771 Acyl-CoA_dh_N
PF01602 Adaptin_N
PF00180 Iso_dh
PF00112 Peptidase_C1
PF00378 ECH_1
PF14497 GST_C_3
PF02803 Thiolase_C
PF00091 Tubulin
PF02770 Acyl-CoA_dh_M
PF03953 Tubulin_C
PF02221 E1_DerP2_DerF2
PF03810 Importin
PF00171 Aldedh
PF02798 GST_N
PF13833 EF-hand_8
PF00011 HSP20
PF00012 Hsp70
PF00435 Spectrin
PF00108 Thiolase_N
PF00379 Chitin_bind_4
PF13561 Adh_short_C2
PF13499 EF-hand_7
PF00106 Adh_short
L. deliense
11.00
33.00
55.00
Pfam ID Description
PF00006 ATP-synt_ab
PF02874 ATP-synt_ab_N
PF01576 Myosin_tail_1
PF00549 Ligase_CoA
PF00435 Spectrin repeat
PF00307 CH

Figure 6

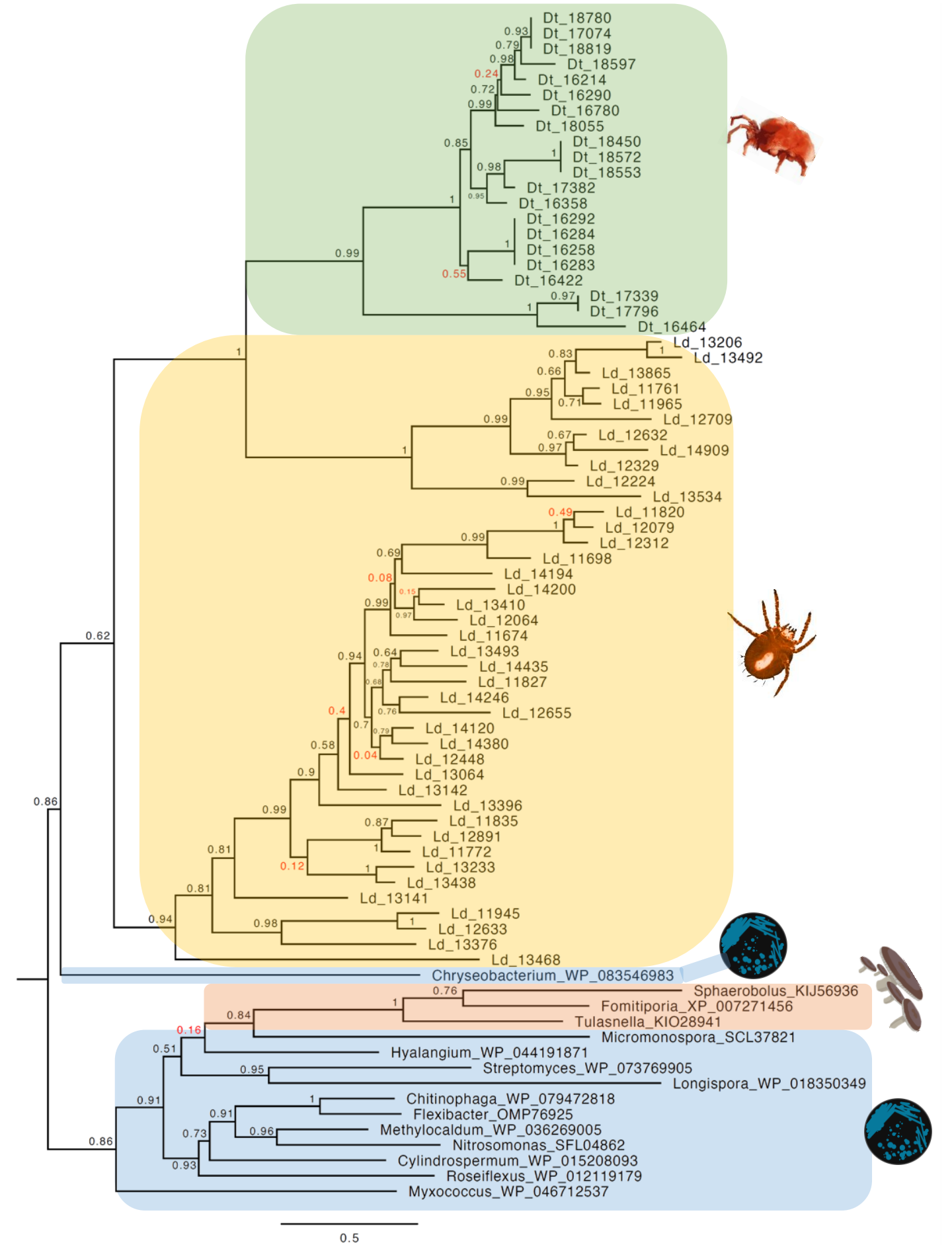

Dt_18780
Dt_17074
Dt_18819
Dt_18597
Dt_16214
Dt_16290
Dt_16780
Dt_18055
Dt_18450
Dt_18572
Dt_18553
Dt_17382
Dt_16358
Dt_16292
Dt_16284
Dt_16258
Dt_16283
Dt_16422
Dt_17339
Dt_17796
Dt_16464
Ld_13206
Ld_13492
Ld_13865
Ld_11761
Ld_11965
Ld_12709
Ld_12632
Ld_14909
Ld_12329
Ld_12224
Ld_13534
Ld_11820
Ld_12079
Ld_12312
Ld_11698
Ld_14194
Ld_14200
Ld_13410
Ld_12064
Ld_11674
Ld_13493
Ld_14435
Ld_11827
Ld_14246
Ld_12655
Ld_14120
Ld_14380
Ld_12448
Ld_13064
Ld_13142
Ld_13396
Ld_11835
Ld_12891
Ld_11772
Ld_13233
Ld_13438
Ld_13141
Ld_11945
Ld_12633
Ld_13376
Ld_13468
Chryseobacterium_WP_083546983
Sphaerobolus_KIJ56936
Fomitiporia_XP_007271456
Tulasnella_KIO28941
Micromonospora_SCL37821
Hyalangium_WP_044191871
Streptomyces_WP_073769905
Longispora_WP_018350349
Chitinophaga_WP_079472818
Flexibacter_OMP76925
Methylocaldum_WP_036269005
Nitrosomonas_SFL04862
Cylindrospermum_WP_015208093
Roseiflexus_WP_012119179
Myxococcus_WP_046712537
0.5

Figure 7

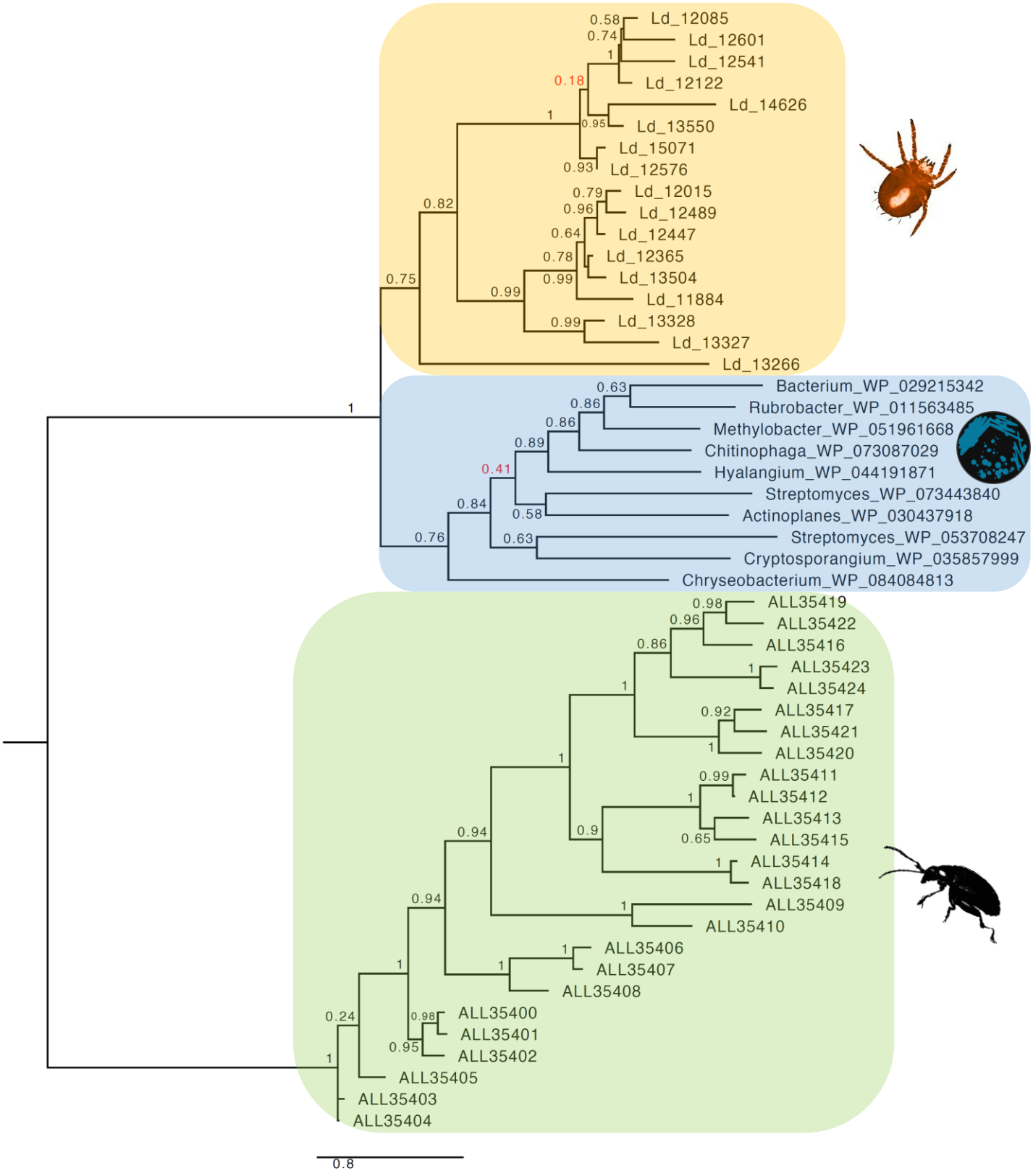
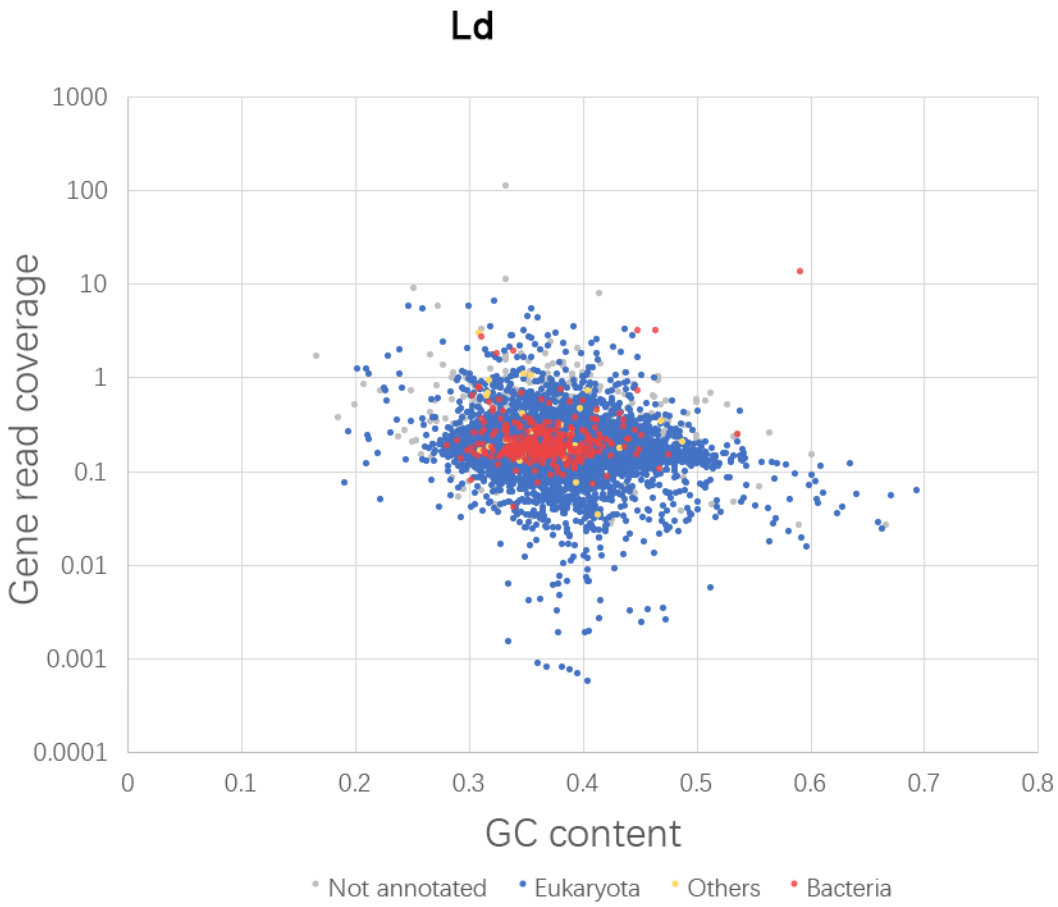

Ld_12085
Ld_12601
Ld_12541
Ld_12122
Ld_14626
Ld_13550
Ld_15071
Ld_12576
Ld_12015
Ld_12489
Ld_12447
Ld_12365
Ld_13504
Ld_11884
Ld_13328
Ld_13327
Ld_13266
Bacterium_WP_029215342
Rubrobacter_WP_011563485
Methylobacter_WP_051961668
Chitinophaga_WP_073087029
Hyalangium_WP_044191871
Streptomyces_WP_073443840
Actinoplanes_WP_030437918
Streptomyces_WP_053708247
Cryptosporangium_WP_035857999
Chryseobacterium_WP_084084813
ALL35419
ALL35422
ALL35416
ALL35423
ALL35424
ALL35417
ALL35421
ALL35420
ALL35411
ALL35412
ALL35413
ALL35415
ALL35414
ALL35418
ALL35409
ALL35410
ALL35406
ALL35407
ALL35408
ALL35400
ALL35401
ALL35402
ALL35405
ALL35403
ALL35404
0.8

Figure 8


Ld
1000
100
10
1
0.1
0.01
0.001
0.0001
Gene read coverage
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
GC content
Not annotated
Eukaryota
Others
Bacteria

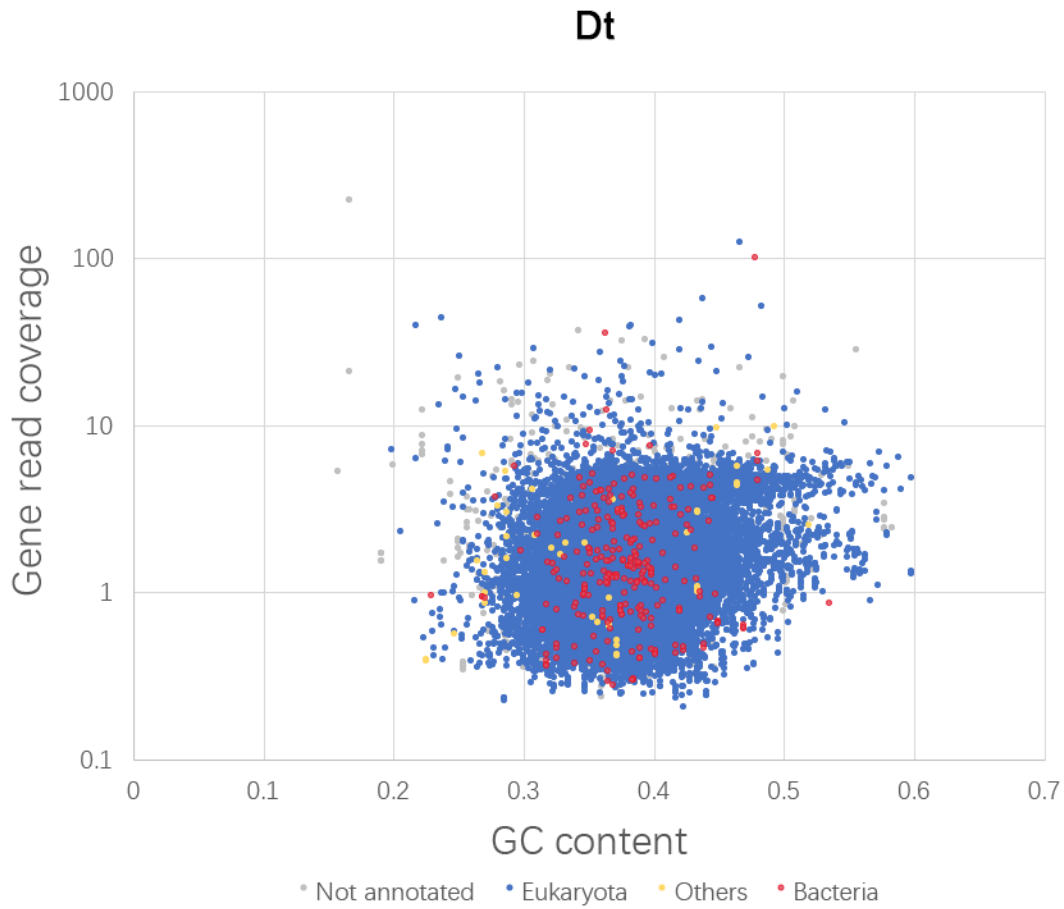

Dt
1000
100
10
1
0.1
Gene read coverage
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
GC content
Not annotated
Eukaryota
Others
Bacteria

Figure 9

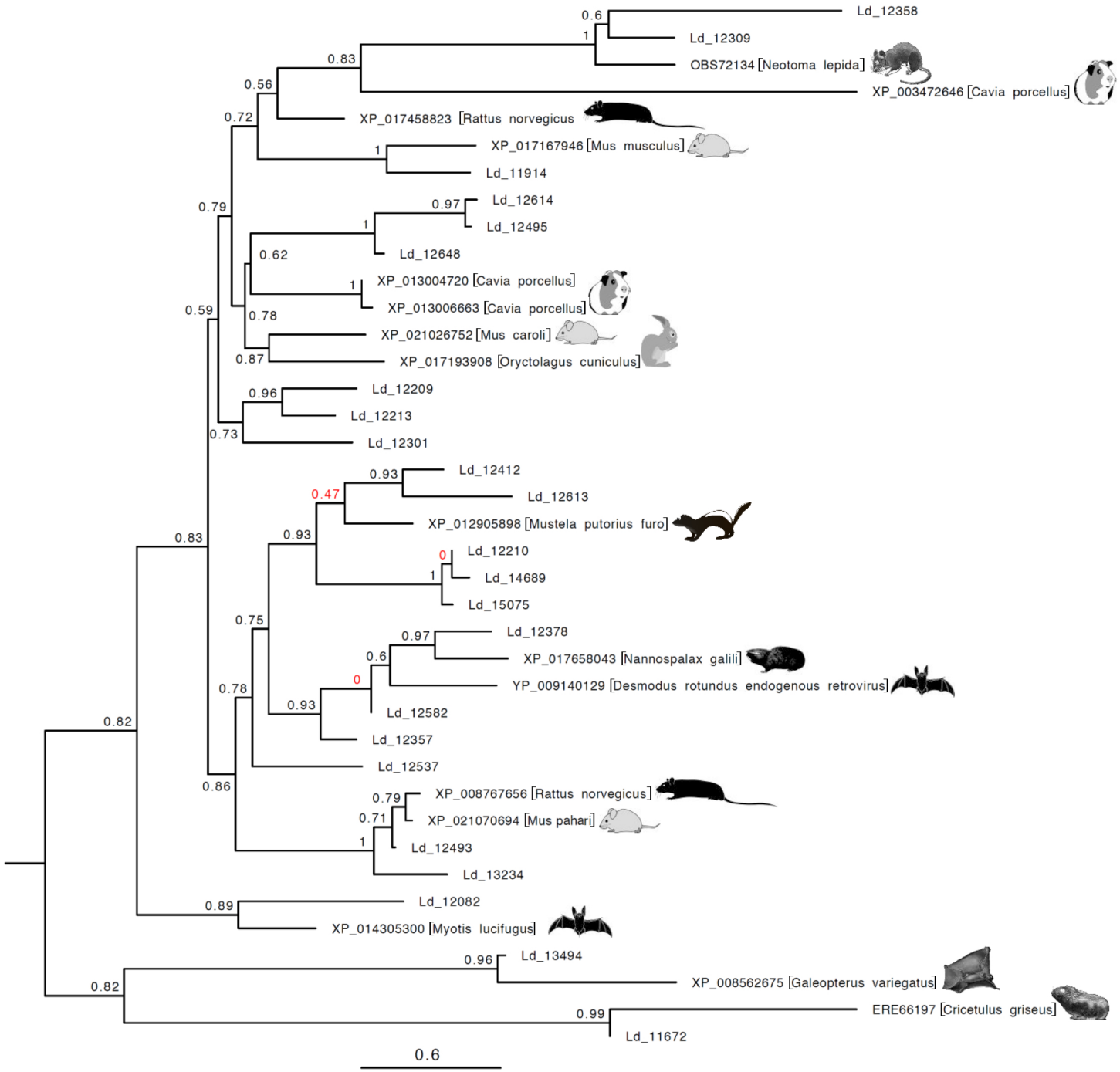

Ld_12358
Ld_12309
OBS72134 [Neotoma lepida]
XP_003472646 [Cavia porcellus]
XP_017458823 [Rattus norvegicus
XP_017167946 [Mus musculus]
Ld_11914
Ld_12614
Ld_12495
Ld_12648
XP_013004720 [Cavia porcellus]
XP_013006663 [Cavia porcellus]
XP_021026752 [Mus caroli]
XP_017193908 [Oryctolagus cuniculus]
Ld_12209
Ld_12213
Ld_12301
Ld_12412
Ld_12613
XP_012905898 [Mustela putorius furo]
Ld_12210
Ld_14689
Ld_15075
Ld_12378
XP_017658043 [Nannospalax galili]
YP_009140129 [Desmodus rotundus endogenous retrovirus]
Ld_12582
Ld_12357
Ld_12537
XP_008767656 [Rattus norvegicus]
XP_021070694 [Mus pahari]
Ld_12493
Ld_13234
Ld_12082
XP_014305300 [Myotis lucifugus]
Ld_13494
XP_008562675 [Galeopterus variegatus]
ERE66197 [Cricetulus griseus]
Ld_11672
0.6

Figure 10

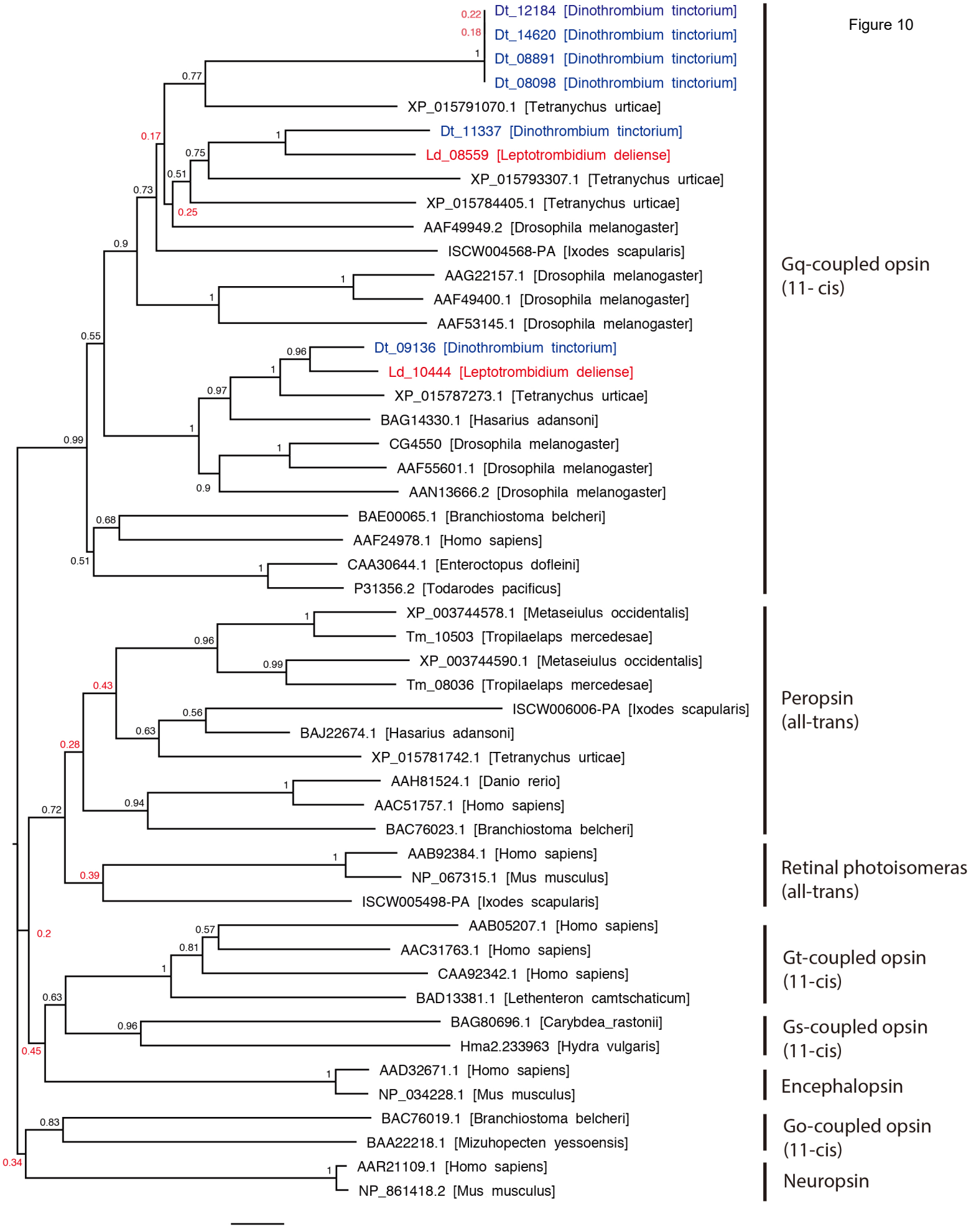

Dt_12184 [Dinothrombium tinctorium]
Dt_14620 [Dinothrombium tinctorium]
Dt_08891 [Dinothrombium tinctorium]
Dt_08098 [Dinothrombium tinctorium]
XP_015791070.1 [Tetranychus urticae]
Dt_11337 [Dinothrombium tinctorium]
Ld_08559 [Leptotrombidium deliense]
XP_015793307.1 [Tetranychus urticae]
XP_015784405.1 [Tetranychus urticae]
AAF49949.2 [Drosophila melanogaster]
ISCW004568-PA [Ixodes scapularis]
AAG22157.1 [Drosophila melanogaster]
AAF49400.1 [Drosophila melanogaster]
AAF53145.1 [Drosophila melanogaster]
Dt_09136 [Dinothrombium tinctorium]
Ld_10444 [Leptotrombidium deliense]
XP_015787273.1 [Tetranychus urticae]
BAG14330.1 [Hasarius adansoni]
CG4550 [Drosophila melanogaster]
AAF55601.1 [Drosophila melanogaster]
AAN13666.2 [Drosophila melanogaster]
BAE00065.1 [Branchiostoma belcheri]
AAF24978.1 [Homo sapiens]
CAA30644.1 [Enteroctopus dofleini]
P31356.2 [Todarodes pacificus]
XP_003744578.1 [Metaseiulus occidentalis]
Tm_10503 [Tropilaelaps mercedesae]
XP_003744590.1 [Metaseiulus occidentalis]
Tm_08036 [Tropilaelaps mercedesae]
ISCW006006-PA [Ixodes scapularis]
BAJ22674.1 [Hasarius adansoni]
XP_015781742.1 [Tetranychus urticae]
AAH81524.1 [Danio rerio]
AAC51757.1 [Homo sapiens]
BAC76023.1 [Branchiostoma belcheri]
AAB92384.1 [Homo sapiens]
NP_067315.1 [Mus musculus]
ISCW005498-PA [Ixodes scapularis]
AAB05207.1 [Homo sapiens]
AAC31763.1 [Homo sapiens]
CAA92342.1 [Homo sapiens]
BAD13381.1 [Lethenteron camtschaticum]
BAG80696.1 [Carybdea_rastonii]
Hma2.233963 [Hydra vulgaris]
AAD32671.1 [Homo sapiens]
NP_034228.1 [Mus musculus]
BAC76019.1 [Branchiostoma belcheri]
BAA22218.1 [Mizuhopecten yessoensis]
AAR21109.1 [Homo sapiens]
NP_861418.2 [Mus musculus]
Gq-coupled opsin (11- cis)
Peropsin (all-trans)
Retinal photoisomeras (all-trans)
Gt-coupled opsin (11-cis)
Gs-coupled opsin (11-cis)
Encephalopsin
Go-coupled opsin (11-cis)
Neuropsin
0.1

Figure 11

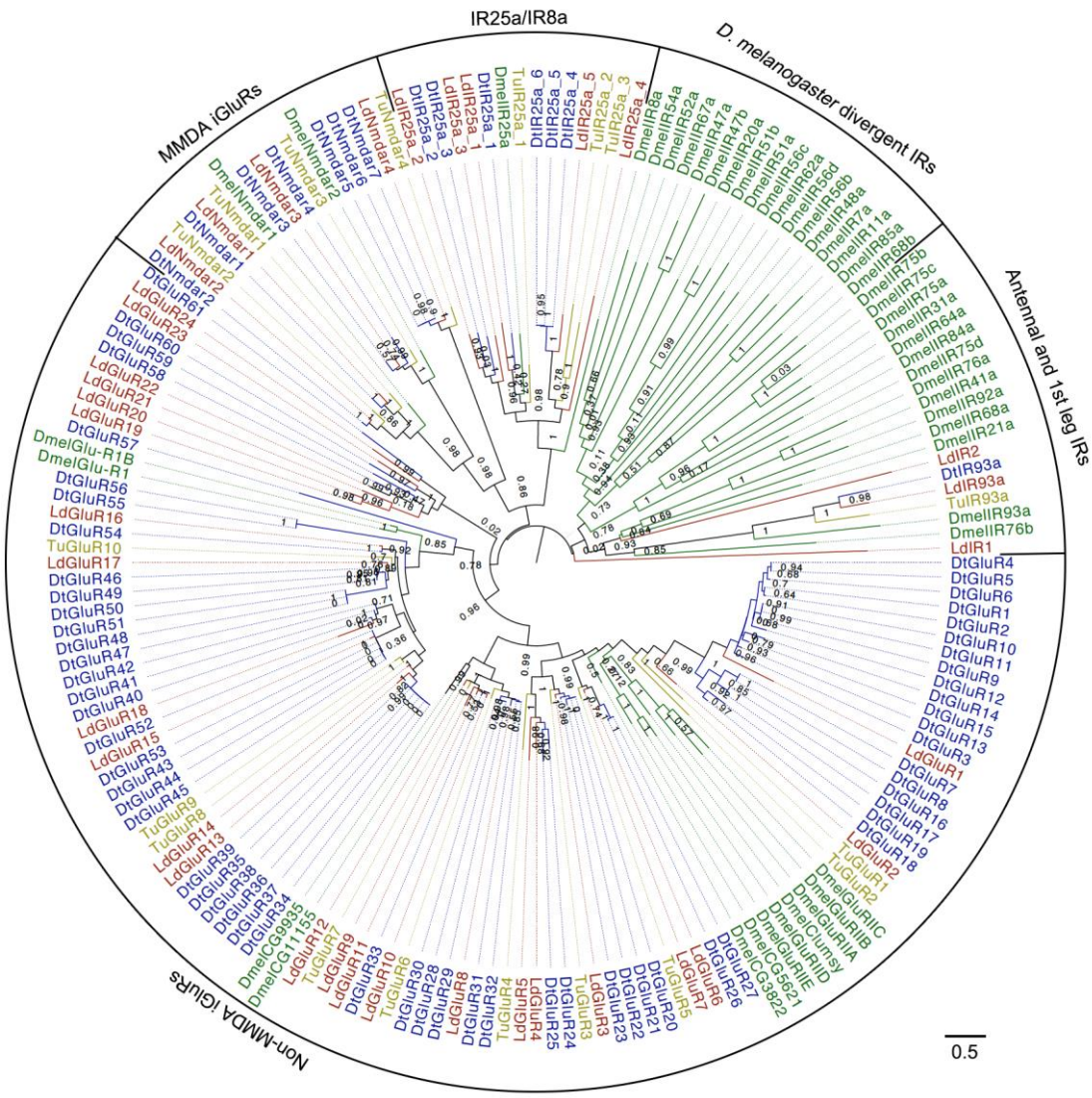
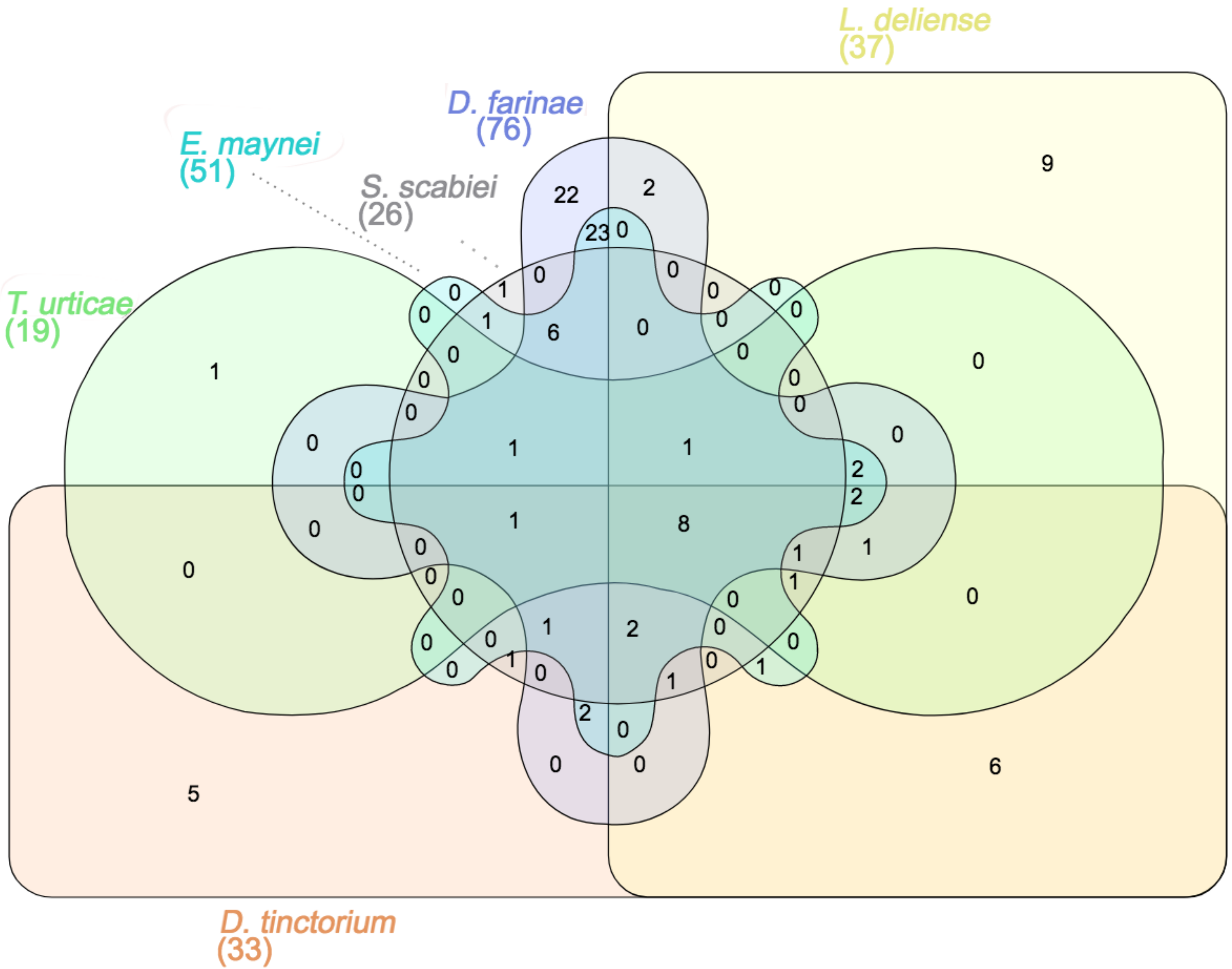

IR25a/IR8a
D. melanogaster divergent IRs
Antennal and 1st leg IRs
Non-NMDA iGluRs
MMDA iGluRs
0.5

Figure 12


L. deliense
(37)
D. farinae
(76)
E. maynei
(51)
S. scabiei
(26)
T. urticae
(19)
D. tinctorium
(33)

L. deliense gene ID	Orthlogous cluster	AllFam	Description	Peptides	Distribution of orthologs *Ld*	*Dt*	*Df*	*Tu*	*Ss*	*Em*
1656	ORTHOMCL2	AF100	Myosin heavy chain	35						
8426	ORTHOMCL25	AF007	EF hand family	4						
9890	ORTHOMCL29	AF024	Trypsin-like serine protease	0						
2714	ORTHOMCL12	AF021	Subtilase family	0						
3075	ORTHOMCL12	AF021	Subtilase family	0						
7213	ORTHOMCL12	AF021	Subtilase family	0						
8430	ORTHOMCL12	AF021	Subtilase family	0						
9243	ORTHOMCL12	AF021	Subtilase family	0						
9299	ORTHOMCL12	AF021	Subtilase family	0						
14039	ORTHOMCL12	AF021	Subtilase family	0						
621	ORTHOMCL10	AF054	Tropomyosin	10						
11355	ORTHOMCL10	AF054	Tropomyosin	13						
109	ORTHOMCL11	AF038	Cyclophilin	1						
10361	ORTHOMCL20	AF197	Fructose-bisphosphate aldolase	4						
1930	ORTHOMCL1	AF025	Tubulin/FtsZ family	0						
7500	ORTHOMCL16	-	Nucleoside diphosphate kinase	4						
7791	ORTHOMCL17	AF030	Papain-like cysteine protease	0						
33	ORTHOMCL18	AF030	Papain-like cysteine protease	5						
8227	ORTHOMCL18	AF030	Papain family cysteine protease	0						
8323	ORTHOMCL19	AF020	Superoxide dismutase	0						
2800	ORTHOMCL31	AF024	Trypsin-like serine protease	0						
9622	ORTHOMCL31	AF024	Trypsin-like serine protease	0						
9623	ORTHOMCL31	AF024	Trypsin-like serine protease	0						
10525	ORTHOMCL31	AF024	Trypsin-like serine protease	0						
11514	ORTHOMCL31	AF024	Trypsin-like serine protease	0						
6375	ORTHOMCL3	-	Casein kinase II subunit beta	0						
1868	ORTHOMCL32	AF024	Trypsin-like serine protease	0						
2046	ORTHOMCL32	AF024	Trypsin-like serine protease	0						
6856	ORTHOMCL32	AF024	Trypsin-like serine protease	0						
7293	ORTHOMCL32	AF024	Trypsin-like serine protease	0						
10686	ORTHOMCL32	AF024	Trypsin-like serine protease	0						
3432	ORTHOMCL23	AF032	Triosephosphate isomerase	3						
8478	ORTHOMCL33	AF142	WD domain, G-beta repeat	9						
8833	ORTHOMCL30	AF024	Trypsin-like serine protease	0						
2654	ORTHOMCL39	AF030	Papain-like cysteine protease	0						
4617	ORTHOMCL4	AF031	Enolase and Methylaspartate ammonia-lyase	1						
9789	ORTHOMCL43	AF024	Trypsin-like serine protease	0						
4990	ORTHOMCL45	AF024	Trypsin-like serine protease	0						
7966	ORTHOMCL85	AF037	Alpha/beta hydrolase and Pancreatic lipase family	0						
5193	ORTHOMCL86	AF030	Papain-like cysteine protease	0						
10503	ORTHOMCL86	AF030	Papain-like cysteine protease	0						
5064	ORTHOMCL87	AF031	Enolase	0						
10276	ORTHOMCL87	AF031	Enolase	3						
1332	ORTHOMCL5	AF049	ATP:guanido phosphotransferase family	0						
4787	ORTHOMCL88	AF024	Trypsin-like serine protease	0						
11513	ORTHOMCL88	AF024	Trypsin-like serine protease	0						
4574	ORTHOMCL89	AF024	Trypsin-like serine protease	0						
5439	ORTHOMCL89	AF024	Trypsin-like serine protease	0						
4061	ORTHOMCL90	AF024	Trypsin-like serine protease	0						
7788	ORTHOMCL90	AF024	Trypsin-like serine protease	0						
721	ORTHOMCL91	AF038	Cyclophilin	0						
6546	ORTHOMCL91	AF038	Cyclophilin	0						
6076	ORTHOMCL96	AF024	Trypsin-like serine protease	0						
9793	ORTHOMCL97	AF024	Trypsin-like serine protease	0						
6725	ORTHOMCL98	AF030	Papain-like cysteine protease	0						
7227	ORTHOMCL7	AF024	Trypsin-like serine protease	0						
9013	ORTHOMCL7	AF024	Trypsin-like serine protease	0						
11006	ORTHOMCL7	AF024	Trypsin-like serine protease	0						
11041	ORTHOMCL7	AF024	Trypsin-like serine protease	0						
12446	ORTHOMCL7	AF024	Trypsin-like serine protease	0						
12530	ORTHOMCL7	AF024	Trypsin-like serine protease	0						
1565	ORTHOMCL60	AF167/AF023	AhpC/TSA family and Thioredoxin	1						
10100	ORTHOMCL100	AF024	Trypsin-like serine protease	0						
3528	ORTHOMCL8	AF010	Glutathione S-transferase	1						
11419	ORTHOMCL8	AF010	Glutathione S-transferase	1						

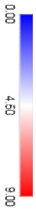
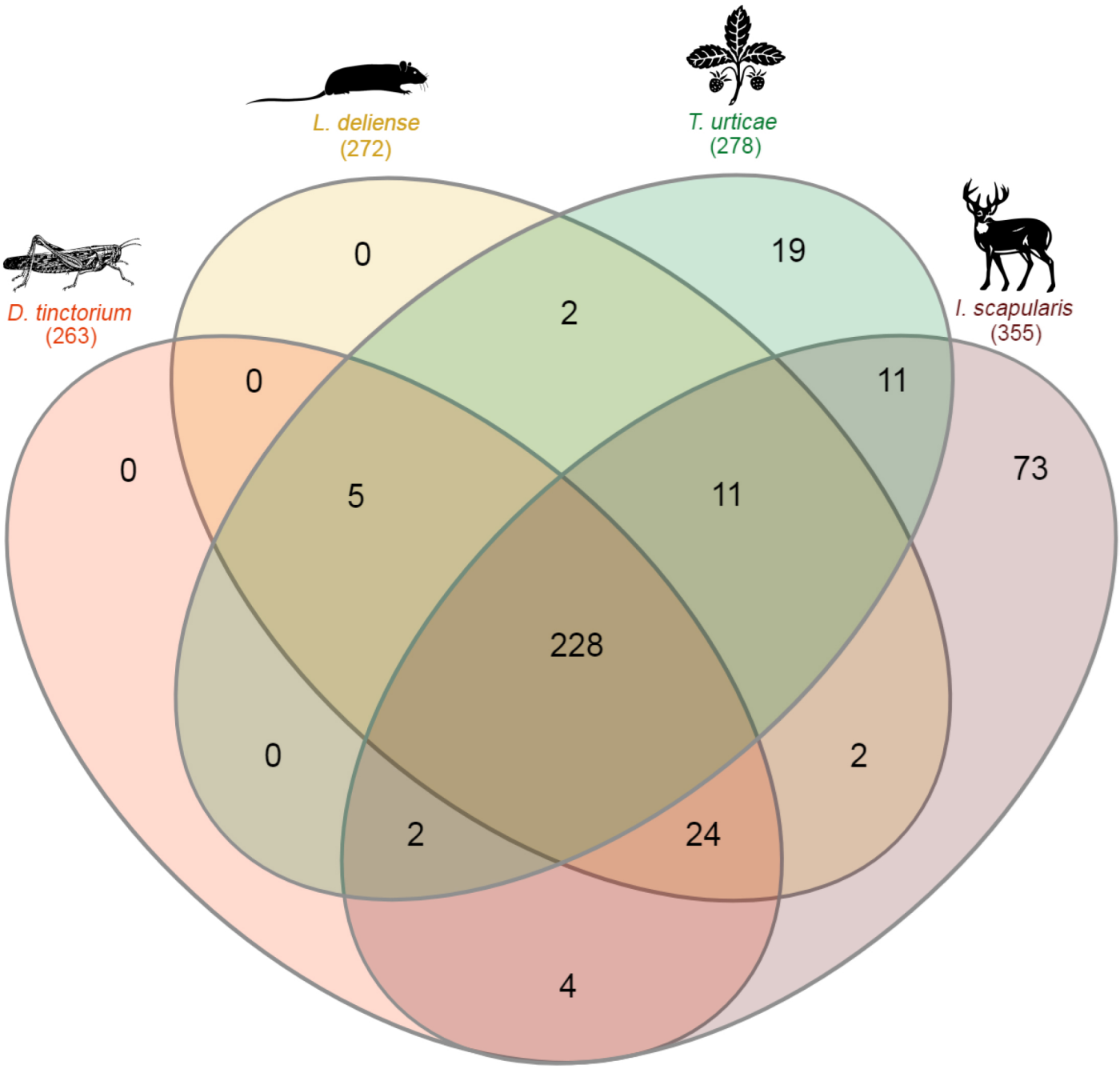


Figure 14


L. deliense
(272)
T. urticae
(278)
D. tinctorium
(263)
I. scapularis
(355)
0
19
2
0
11
0
5
11
73
228
0
2
2
24
4