

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

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

6 <sup>1</sup>Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, United Kingdom,

7 <sup>2</sup>Department of Biological Sciences, Xi'an Jiaotong-Liverpool University, Suzhou 215123, China,

8 <sup>3</sup>School of Life Sciences, Jiangsu Normal University, Xuzhou 221116, China, <sup>4</sup>Institute of Infection &

9 Global Health, University of Liverpool, L3 5RF, United Kingdom, <sup>5</sup>Faculty of Tropical Medicine,

10 Mahidol University, Ratchathewi Bangkok 10400, Thailand, <sup>6</sup>The Royal Veterinary College, London

11 NW1 0TU, United Kingdom, <sup>7</sup>Department of Vector Biology, Liverpool School of Tropical Medicine,

12 Liverpool L3 5QA, United Kingdom, <sup>8</sup>Institute of Veterinary Science, University of Liverpool, Liverpool

13 L3 5RP, United Kingdom.

14 †Contributed equally.

15 \*Correspondence address: Ben Makepeace, Department of Infection Biology, Institute of infection &

16 Global Health, Liverpool Science Park IC2, 146 Brownlow Hill, Liverpool L3 5RF, United Kingdom. Tel:

17 +44 151-7941586; Fax: +44 151 7950236; E-mail: [blm1@liv.ac.uk](mailto:blm1@liv.ac.uk)

## 18 **Abstract**

### 19 *Background*

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

### 28 *Results*

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

### 37 *Conclusions*

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

41 **Keywords**

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

43 *Dinothrombium*, *Tetranychus*, Trombidiformes.

## 44 **Background**

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

55 The role of lateral gene transfer (LGT) in the evolution of animals has remained controversial since the  
56 first metazoan genomes were sequenced. Despite the explosion of new genomic resources across a  
57 wider variety of metazoan phyla in recent years, some claims of large-scale LGT in animal genomes  
58 have been shown to be the result of flawed data-analysis methods that failed to exclude sequences  
59 from bacterial contaminants [2-5]. Attempts to conduct meta-analyses across diverse metazoan  
60 genomes using consistent criteria have also been criticized for reliance on unsound assumptions [6,  
61 7]. However, it is irrefutable that several metazoan taxa are reliant on functions obtained via LGT for  
62 essential physiological processes, including digestion of complex and/or toxic materials (especially in  
63 the case of herbivores and plant parasites) and the avoidance of host defences [8-11]. One group of  
64 acariform mites, the spider mites (superfamily Tetranychoidae: order Trombidiformes; Fig. 1), are  
65 major pests of various crops and rely on well-characterised lateral gene transfers of  $\beta$ -cyanoalanine  
66 synthase (from bacteria) and carotenoid biosynthesis enzymes (from fungi) to detoxify hydrogen  
67 cyanide in their diet [12] and to control diapause [13], respectively. Currently, it is unknown whether

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

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

80 The more heavily-studied larval stages of the Trombiculoidea (commonly referred to as chiggers or  
81 berry bugs) primarily feed on terrestrial vertebrates, including humans [20], although some little-  
82 known taxa in this superfamily are ectoparasites of invertebrate hosts in common with the  
83 Trombidioidea [21-23] (Fig. 2). Importantly, the only major mite-transmitted disease of humans, scrub  
84 typhus or tsutsugamushi disease, is vectored by chiggers in the genus *Leptotrombidium* [24], while  
85 other chigger genera have only been implicated epidemiologically as locally-important vectors [25].  
86 Scrub typhus is a severe febrile illness caused by infection with an obligate intracellular bacterium  
87 (*Orientia* spp.) in the order *Rickettsiales* and features an epidemiological cycle that includes wild small  
88 mammals, which are the primary hosts for many chigger species [26]. This disease has a fatality rate  
89 of 6% if not treated promptly with antibiotics [27] and has increased in incidence globally in recent  
90 years, with the annual minimum incidence reaching >17/100,000 in South Korea and Thailand in 2012  
91 – 2013 [28]. In the so-called “tsutsugamushi triangle” within the Asia-Pacific region, scrub typhus has  
92 a median seroprevalence of 22.2% [28]; but endemic scrub typhus has emerged in several other parts

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

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

106 To date, research on trombidid mites has suffered from a dearth of molecular data that could facilitate  
107 studies on speciation; population structure; host-vector and vector-pathogen interactions; and life-  
108 history evolution in this group. To address this deficit, here we present a comparative analysis of the  
109 genomes of the chigger *Leptotrombidium deliense* (the primary scrub typhus vector in South-East Asia  
110 [46]) and the giant red velvet mite, *Dinotrombium tinctorium* (the world’s largest acarine species  
111 [47]). We show that these trombidid mites form a distinct branch of the Trombidiformes that exhibit  
112 two classes of LGT for secondary metabolism: the previously identified carotenoid biosynthesis  
113 enzymes of fungal origin, and much larger terpene synthase gene families, which probably derive from  
114 soil-associated bacteria. We also identify unique clusters of predicted allergens in *L. deliense* that may  
115 contribute to the symptoms of trombiculiasis in humans and domestic animals.

## 116 **Data description**

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

126 The total number of trimmed reads generated was ~362 million for *D. tinctorium* and ~38 million for  
127 *L. deliense*. For the former, PE reads were assembled using Abyss (v. 1.5.2) [48, 49]. For the chigger  
128 data, a preliminary assembly to contig level was performed using Velvet (v. 1.2.07) [50]. Reads derived  
129 from mammalian host genomic DNA were removed from the preliminary genome assembly using  
130 blobtools (v0.9.19) , which generates a GC-coverage plot (proportion of GC bases and node coverage)  
131 [51] (Fig. 3). The *L. deliense* genome was then reassembled using SPAdes assembler (v. 3.7.1) [52] with  
132 default settings. For gene calling, the MAKER pipeline [53] was used to integrate *ab initio* gene  
133 predictions from Augustus (v. 3.2.2) [54], SNAP (v. 2013-11-29) [55] and GeneMark (v. 2.3e) [56] with  
134 evidence-based gene models. The “Methods” section provides more details on how downstream  
135 genome analyses were performed.

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

141 (Thermo Fisher Scientific). A total of 137,638 spectra were generated across the eight fractions. For *L.*  
142 *deliense*, a soluble protein extract was obtained from a small pool of ethanol-fixed engorged larvae (*n*  
143 = 10) collected from several species of wild rodents in Thailand [26]. Following tryptic digestion,  
144 downstream analyses proceeded as for *D. tinctorium*, producing 18,059 spectra. The “Methods”  
145 section provides details on how MS spectra searches and Pfam enrichment analyses were performed.  
146



## 147 **Analyses**

### 148 **Genome statistics and phylogenomics**

149 Assembled genome sizes were 180.41 Mb for *D. tinctorium* and 117.33 Mb for *L. deliense* (Additional  
150 file 1), whereas *k*-mer analysis placed the genome size estimates much closer together but was of a  
151 similar scale to the assemblies (143.52 – 147.09 and 158.31 – 160.95 Mb, respectively; Additional file  
152 2). The repeat content of these genomes presented a significant challenge, with unclassified repeats  
153 alone accounting for 19 - 23% of the total size; approximately double the proportion of the *S. scabiei*  
154 [58] and *Tetranychus urticae* (two-spotted spider mite [59]) genomes (Additional file 3). As expected  
155 from sequencing a pool of *L. deliense* larvae compared with a single adult of *D. tinctorium*, and the  
156 corresponding DNA library strategies employed, the chigger genome was considerably less contiguous  
157 than that of the velvet mite. Nevertheless, the estimated completeness of the predicted gene set for  
158 *L. deliense* based on the Benchmarking Universal Single-Copy Orthologues (BUSCO) criteria [60]  
159 compared favourably with that of other arachnid genomes (Additional file 1). Notably, the *D.*  
160 *tinctorium* genome contained the greatest number of protein-coding genes from the Acariformes  
161 sequenced to date (Additional file 1).

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

### 170 **Gene family expansions**

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

178 Relative to other arachnids, *D. tinctorium* exhibited a significant expansion of 38 gene families,  
179 including a large family (ORTHOMCL98) of uncharacterised proteins containing 45 members in this  
180 species but no representatives in the other arachnid genomes (Additional file 6). Examination of  
181 conserved domains in this gene family revealed a major facilitator superfamily domain with some  
182 weak but significant similarity by BLASTP (~25% amino-acid identity, >95% coverage) to feline  
183 leukemia virus subgroup C receptor-related protein (FLVCR)-1 from various Metazoa. Two other  
184 orthologous clusters displayed an identity of up to 30% with FLVCR2 and showed a statistically  
185 significant expansion in the velvet mite, containing 27 (ORTHOMCL265) and 20 (ORTHOMCL412) members,  
186 compared with only one and three members in *L. deliense*, respectively (Additional file 6). Moreover,  
187 a gene family annotated as 4-coumarate:CoA ligases was significantly expanded in *D. tinctorium*, with  
188 27 members; while *L. deliense* also showed a large repertoire (14 members) relative to other  
189 arachnids, although this was not statistically significant (Additional file 6). This gene family represents  
190 a group of long-chain fatty acyl-CoA synthetases related to firefly luciferase (but without luciferase  
191 activity) that are found in multicellular eukaryotes other than vertebrates [62]. These enzymes are  
192 associated with peroxisomes and perform catabolism of very long-chain fatty acids by  $\beta$ -oxidation.

193 Only a single gene, SPARC-related modular calcium-binding protein 1 (SMOC1) from *L. deliense*, was  
194 predicted to be under positive selection in either of the trombidid genomes with strong statistical  
195 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 ( $\geq 2$  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 ( $\geq 2$  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

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

## 232 Lateral gene transfers and mobile elements

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

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

246 metabolites are assumed to be derived from microbial symbionts [74]. The absence of terpene  
247 synthases in the fully sequenced diplopod genome from the rusty millipede, *Trigoniulus corallinus*,  
248 certainly supports this interpretation [75]. Dust mites also produce a monoterpene, neryl formate,  
249 which has been demonstrated to act as an aggregation pheromone [76]. However, BLAST analysis of  
250 the *Dermatophagoides farinae* genome assembly [77] using the trombidid terpene synthases failed to  
251 identify significant homologues, suggesting that dust mites rely on microbial symbionts for terpene  
252 production, or that terpene synthases in mite genomes are evolving too rapidly to be identified by  
253 homology searches. To the best of our knowledge, the only animals known to harbour terpene  
254 synthase genes in their nuclear genomes are a very restricted number of beetle species  
255 [predominantly flea beetles of the subfamily Galerucinae, which produce (6R,7S)-himachala-9,11-  
256 diene as a male aggregation pheromone [78]] and the collembolan *Folsomia candida*, in which the  
257 metabolites produced and their function are unknown [79]. Whilst terpene synthases are widespread  
258 in plants, fungi and bacteria, the terpene synthases from flea beetles do not resemble those from non-  
259 metazoan taxa and appear to have evolved from arthropod *trans*-isoprenyl diphosphate synthases  
260 [78]. Moreover, the terpene synthases of *F. candida* are more similar to those from the flea beetles  
261 than they are to the non-metazoan enzymes [79].

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

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

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

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

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

301 In the *D. tinctorium* genome, most members of the two ERV protein families each clustered in a  
302 monophyletic clade, suggesting expansion within the mite genome from a single origin (Additional  
303 files 13 and 14). The closest homologues of the Pol-like polyproteins were found mainly in other  
304 arthropods (especially cladocerans and ticks) and more distantly in fungi (Additional file 14); whereas  
305 the reverse ribonuclease integrases were most similar to those from ants, moths, nematodes and  
306 other mites (Additional file 13). Thus, endogenous retroviruses in *D. tinctorium* might originate from  
307 larval hosts, prey or soil microorganisms, although the fact that some velvet mites feed on other Acari  
308 [18] renders phylogenetic analyses of potential lateral gene transfers especially problematic. Uniquely  
309 among chelicerate genomes sequenced to date, the *D. tinctorium* genome also contained hepatitis D  
310 ribozyme-like genes (Rfam RF01787; Additional file 15), which in *Anopheles* mosquitoes, have been  
311 suggested to be involved in processing of non-LTR retrotransposons [83].

## 312 Immune system

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

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

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



347 a substantially greater complement of Dscam genes (~40 in *D. tinctorium*; Additional file 16), rivalling  
348 the 60 gene family members observed in the *Strigamia maritima* (coastal centipede) genome [85, 88].  
349 Several other expanded gene families in one or both of the trombidid mite genomes may have roles  
350 in the immune response. In common with *I. scapularis*, these genomes lack the transmembrane PGRPs  
351 that are activated in the presence of peptidoglycan in insects, but contain several other PGRP genes  
352 with putative extracellular or intracellular roles. However, these soluble PGRP genes are present in  
353 larger numbers in the trombidid mite genomes than in those of *I. scapularis* and *T. urticae* (Additional  
354 file 17). Since soluble PGRP fragments can have a co-receptor function as shown in insects [89], they  
355 might work in concert with as-yet-unidentified components of the acarine immune system to  
356 recognise pathogens. This is particularly important in the case of chiggers, as evidence for a  
357 peptidoglycan-like structure has recently been reported for *Orientia tsutsugamushi* [90]. Moreover, a  
358 much larger expansion in a second class of proteins with putative roles in the immune system, the C-  
359 type lectin domain (CTLN) proteins, was apparent in *L. deliense* (Additional file 6; Table 1). The CTLN  
360 protein family is a large and diverse group, most members of which do not bind carbohydrates and  
361 are thus not lectins [91]. If a CTLN protein does have lectin activity, the carbohydrate-recognition  
362 domain usually contains the amino acid motif “WND”, together with “EPN” if the specificity is for  
363 mannose, and QPD if the specificity is for galactose. However, several exceptions to this pattern do  
364 exist [91]. The expanded *L. deliense* CTLN proteins belong to four orthologous groups containing a  
365 total of 91 genes, of which one cluster (ORTHOMCL880) lacks any signatures of carbohydrate-binding  
366 activity (Table 1). The other three groups mainly contain proteins with EPN motifs, suggesting  
367 specificity for mannose, although a small proportion of QPD-motif CTLN proteins were apparent in  
368 two of the clusters, which might bind galactose (Table 1). The majority of the *L. deliense* CTLN proteins  
369 that were predicted to bind carbohydrates exhibited classical or internal secretion signatures, while  
370 only a small proportion (10 – 20%) contained transmembrane domains (Table 1). In common with  
371 many members of the CTLN protein family, including those in other arthropods, *N*-glycosylation sites  
372 were predicted in 10 – 20% of the *L. deliense* CTLN proteins [92] (Table 1).

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

## 386 Photoreceptor and chemosensory systems

387 Unlike insects, chelicerates lack compound eyes. Mites and ticks may be eyeless, or can possess one  
388 or more pairs of simple dorsal ocelli. The Parasitiformes sequenced to date are all eyeless species (*I.*  
389 *scapularis*, *M. occidentalis* and *T. mercedesae*), whereas the trombidid mites and *T. urticae* have two  
390 pairs of ocelli on the prodorsum in the adult stage ([47]. However, the genomes of both eyeless and  
391 eyed Acari exhibit a variable complement of opsins, which in combination with the chromophore  
392 retinal, form light-sensitive proteins termed rhodopsins. The genomes of eyeless ticks and mites, as  
393 well as that of *T. urticae*, contain one or more genes of the “all-*trans*-retinal” peropsin class, which in  
394 spiders have been shown to encode non-visual photosensitive pigments with combined G-protein  
395 coupled receptor and retinal photoisomerase activity [96]. Since even the eyeless species show  
396 evidence for reproductive and diapause behaviours that respond to day-length, it has been suggested  
397 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

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

#### 436 Predicted allergens

437 Although the propensity of chiggers to cause pruritic dermatitis is well recognised in humans and other  
438 animals [35-38], the identity of the allergens involved has not been established [110]. The *L. deliense*  
439 and *D. tinctorium* genomes were predicted to encode 37 and 33 groups of protein allergens,  
440 respectively; substantially more than other sequenced mites in the Acariformes with the exception of  
441 the dust mites, *D. farinae* [77] and *E. maynei* [111]. Since velvet mites rarely come into contact with  
442 humans, only the chigger allergens were subjected to further analysis. The *L. deliense* predicted  
443 allergen clusters included nine groups that were unique to this species and six that were shared with  
444 *D. tinctorium* only (Fig. 12), while a further 28 putative allergen genes in the *L. deliense* genome did  
445 not cluster in orthologous groups (Additional file 19). These chigger-unique groups included five  
446 distinct clusters of trypsin-like serine proteases and one cluster each of subtilases, papain-like cysteine  
447 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

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

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

#### 486 Putative salivary proteins

487 Due to the diminutive size of chiggers and the absence of any artificial feeding mechanism for  
488 laboratory colonies that might allow collection of saliva, the chigger sialome has not been  
489 characterised to date. However, numerous high-throughput studies of tick saliva have been conducted  
490 on several genera and multiple lifecycle stages [120-125], and recently an elegant proteomic analysis  
491 of *T. urticae* saliva was published, in which mite salivary secretions were collected in an artificial diet  
492 substrate [126]. Using proteomic datasets from this *T. urticae* study and a recent *I. scapularis* sialome  
493 analysis conducted over several time-points [120], we identified one-to-one orthologues of the  
494 salivary proteins from both sources in the tick, spider mite, and trombidid mite genomes. We reasoned  
495 that as *T. urticae* is phylogenetically close to the trombidid mites while *I. scapularis* is very distant,  
496 protein families shared by the tick and the trombidid mites but not present in the *T. urticae* genome  
497 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

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

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



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

## 565 Discussion

### 566 Genome features and trombidid mite evolution

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

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

582 As previously reported on publication of the first spider genomes [61], the Acari are polyphyletic, with  
583 the superorder Parasitiformes (*i.e.*, the ticks together with mesostigmatid and holothyrid mites)  
584 evidently more closely aligned to the spiders (order Araneae) than it is to the Acariformes (Figure S1).  
585 However, at lower taxonomic scales, the phylogenomic analyses supported the conventional  
586 morphology-based taxonomy within the Acariformes, confirming that the trombidid mites are closely  
587 related to the phytophagous Tetranychoida ([14]. Interestingly, within the Trombidiaea, the  
588 divergence between the velvet mites and the chiggers occurred much later (~123 MYA) than the  
589 emergence of the earliest terrestrial vertebrates (395 MYA [140]), coinciding perhaps with the  
590 appearance of crown-group mammals. If this scenario is correct, the ectoparasitism of non-  
591 mammalian vertebrates by chiggers that occurs today may be a product of secondary adaptation, as  
592 suggested by other authors [20]. Although the fossil record is devoid of trombidid mite specimens  
593 predating the Eocene [141], it has been speculated from palaeogeographical and comparative  
594 morphological evidence that trombiculid mites fed initially on other arthropods, with larval  
595 ectoparasitism on vertebrates evolving during the Paleocene, leading to an increase in chigger  
596 diversity [20]. Our data challenges this hypothesis, because it implies that something other than host  
597 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

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

#### 640 Diapause and seasonality

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

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

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

674 decrease in fatty acid  $\beta$ -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  $\beta$ -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 Tetranychoida 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.

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

## 718 Immune response and vector biology

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

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

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

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



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

## 769 **Conclusions**

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

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

## 788 **Materials and methods**

### 789 **Sample collection and DNA extraction**

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

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

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

### 807 Library preparation and sequencing

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

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

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

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

### 843 Assembly and annotation

844 For both genomes, base-calling and de-multiplexing of indexed reads was performed by bcl2fastq v.  
845 1.8.4 (Illumina) to produce sequence data in fastq format. The raw fastq files were trimmed to remove  
846 Illumina adapter sequences using Cutadapt version 1.2.1 [178]. The option “-O 3” was set, so the 3'  
847 end of any reads which matched the adapter sequence over at least 3 bp was trimmed off. The reads  
848 were further trimmed to remove low quality bases, using Sickle version 1.200 [179] with a minimum

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

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

857 For *D. tinctorium*, the PE reads were assembled using Abyss (v. 1.5.2) [48, 49], Allpaths-LG (v. r51279)  
858 [183], SOAPdenovo2 (v. 2.04-r240) [184] and Discover (v. r51454) [185]. When running Abyss, *k*-mer  
859 sizes were from 35 bp to 80 bp with an interval of 5, and the output “*k*-mer size = 80 bp” was selected  
860 as this produced the optimal assembly. Allpaths-LG and Discover specify *k*-mer size automatically,  
861 whereas “*k*-mer size = 63” was selected for SOAPdenovo2, as suggested by the developer. Discover  
862 requires read lengths of 250 bp, so was applied only to the data generated from the mate-pair library.  
863 Assessment of the completeness of the genome assemblies was based on the percentage alignment  
864 obtained against the reads from the TruSeq 350-bp insert library using bowtie2 (v. 2.0.10) [186] and  
865 the predicted core gene content determined by CEGMA (Core Eukaryotic Genes Mapping Approach)  
866 (v. 2.5) [187]. The final, optimum assembly was created by Abyss (97.4% of scaffolds >500 bp were  
867 mapped; 99.2% of Key Orthologs for eukaryotic Genomes were present) and included all scaffolds of  
868  $\geq 1,000$  bp, which constituted ~80% of the total length of the raw assembly.

869 For *L. deliense*, a preliminary genome assembly at contig level was performed using Velvet (v. 1.2.07)  
870 [50], with parameters of ‘best *k*-mer 99’ and ‘-ins\_length 500’. Reads derived from mammalian host  
871 genomic DNA were removed from the preliminary genome assembly using blobtools (v0.9.19), which  
872 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 ‘-noit’  
881 option, which is specific for simple repeats, microsatellites, and low-complexity repeats.

882 Three *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

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

## 904 Phylogenetics

905 Protein data sets of the following arthropod genomes were used as references: *D. melanogaster* (fruit  
906 fly; GOS release: 6.11) [193], *A. mellifera* (honey bee; GOS release: 3.2) [194], *T. mercedesae* (bee mite;  
907 GOS release: v. 1.0) [98], *T. urticae* (spider mite; GOS release: 20150904) [59], *Stegodyphus*  
908 *mimosarum* (velvet spider; GOS release: 1.0) [61], *I. scapularis* (blacklegged tick; GOS release: 1.4)  
909 [195], *M. occidentalis* (predatory mite; GOS release: 1.0) [97]. *Caenorhabditis elegans* (nematode; GOS  
910 release: WS239) [196] was used as the outgroup. For gene-family phylogenetics, we first aligned  
911 orthologous protein sequences with Mafft (v7.309) [197] or Kalign (v2.0) [198]. We manually trimmed  
912 the aligned sequences for large gene sets. The best substitution models of amino-acid substitution  
913 were determined for the alignments by ProtTest (v3.4) with parameters set to “-all-matrices, -all-  
914 distributions, -AIC” [199]. Then, phylogenetic trees were constructed using maximum likelihood  
915 methods (Phyml, v3.1) [200]. In addition, a neighbour-joining method was used for building the  
916 distance-based trees using MEGA (7.021) [201].

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

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

### 931 Analysis of gene family expansions

932 Orthologous gene clusters of *D. tinctorium*, *L. deliense* and the other reference genomes described  
933 above were defined based on OrthoMCL (v. 1.4) [207]. We used CAFE (v. 3.1) [208] to infer the gene  
934 family expansion and contraction in *D. tinctorium* and *L. deliense* against other Acariformes (*T. urticae*  
935 and *S. scabiei*). The ultrametric species tree used in the CAFE analyses was created as described for  
936 gene-family phylogenetics above. We also calculated  $\omega$  ( $d_N/d_S$ ) ratios for 454 one-to-one orthologues  
937 defined by OrthoMCL using codeml in the PAML package [205] with the free-ratio model. Branches  
938 with  $\omega > 1$  are considered to be under positive selection. The null model used for the branch test was  
939 the one-ratio model (nssites = 0; model = 0) where  $\omega$  was the same for all branches. Kappa and omega  
940 values were automatically estimated from the data, with the clock entirely free to change among  
941 branches. The  $P$  value was determined twice using the log-likelihood difference between the two  
942 models, compared to a  $\chi^2$  distribution with the difference in number of parameters between the one-  
943 ratio and free-ratio models. To estimate significance with the  $P$  value, a likelihood-ratio test was used  
944 to compare  $\ln l$  values for each model and test if they were significantly different. The differences in  
945 log-likelihood values between two models were compared to a  $\chi^2$  distribution with degrees of freedom  
946 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  $\omega$  ( $>10$ ) 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  $\geq 30$  and a best non-  
956 metazoan bit-score of  $\geq 100$ . For each class C gene, its average  $h$  value ( $h_{\text{orth}}$ ) and that of its paralogous  
957 genes in each OrthoMCL cluster defined above was determined. If  $h$  was  $\geq 30$ , the best non-metazoan  
958 bit-score was  $\geq 100$ , and the  $h_{\text{orth}}$  value was  $\geq 30$ , the gene was considered to be a class B gene. Class A  
959 genes were defined as a subset of class B genes with  $h \geq 30$ , a best non-metazoan bit-score of  $\geq 100$ , an  
960  $h_{\text{orth}}$  value of  $\geq 30$ , 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 performed 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”.

## 971 Analysis of chemosensory and photoreceptor gene families

972 A search for *D. tinctorium* and *L. deliense* OBPs was initially preformed using TBLASTN (E-value,  $<1 \times$   
973  $10^{-3}$ ) against their genome assemblies using *D. melanogaster*, *Drosophila mojavensis*, *Anopheles*  
974 *gambiae*, *Bombyx mori*, *Tribolium castaneum*, *A. mellifera*, *Pediculus humanus humanus* and  
975 *Acyrtosiphon pisum* OBPs (identified by Vieira and Rozas [213]) as queries. No OBPs were found in  
976 the *D. tinctorium* and *L. deliense* genome assemblies. Because OBPs are very divergent in terms of the  
977 amino-acid sequences within the family, and the sequence identities between the family members  
978 from the different species can be as low as 8% [213], a TBLASTN search has limited power to identify  
979 these genes. A search for OBPs was therefore performed again with BLASTP (E-value,  $<1 \times 10^{-3}$ ) to  
980 search the automated protein predictions from the mite genome assemblies. A search for small  
981 chemosensory proteins in *D. tinctorium* and *L. deliense* was preformed using the same methods as for  
982 the OBPs. The query sequences were also based on the study of Vieira and Rozas [213], using  
983 chemosensory protein sequences from *D. melanogaster*, *D. mojavensis*, *A. gambiae*, *B. mori*, *T.*  
984 *castaneum*, *A. mellifera*, *P. humanus humanus*, *A. pisum*, *I. scapularis*, and *Daphnia pulex*. For odorant  
985 receptors, both TBLASTN and BLASTP searches were performed using *D. melanogaster* and *A. mellifera*  
986 sequences (identified by Nozawa and Nei [214] and Robertson and Wanner [215], respectively) as  
987 queries.

988 The *D. tinctorium* and *L. deliense* GR gene families were manually annotated according to TBLASTN  
989 and BLASTP searches (both with an E-value cutoff of  $<1 \times 10^{-3}$ ) against their genome assemblies and  
990 predicted protein coding genes, respectively, using all *D. melanogaster* [216], *A. mellifera* [215], *I.*  
991 *scapularis* [195], *T. urticae* [100], *T. mercedesae* [98] and *M. occidentalis* [97] GRs as queries. An  
992 iterative search was also conducted with termite GRs as queries until no new genes were identified in  
993 each major subfamily or lineage. For phylogenetic analysis, *D. tinctorium* and *L. deliense* GRs were  
994 aligned with *D. melanogaster* GRs by Kalign [198] with default settings. Poorly-aligned and variable N-  
995 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

1021 sequences were submitted to Pfam (EBI, v.31.0 [220]) to be assigned to protein families. Mapped Pfam  
1022 domain identifications were then searched against the AllFam database [112] to retrieve  
1023 corresponding AllFam identifiers for allergen families. Venn diagrams were constructed using  
1024 InteractiVenn [221].

#### 1025 Salivary and cement protein analysis

1026 A total of 159 non-redundant tick cement proteins were retrieved from the NCBI database and then  
1027 clustered with all *D. tinctorium* and *L. deliense* amino-acid sequences using OrthoMCL in order to  
1028 identify the tick cement orthologues in the new genomes. Salivary proteins of *I. scapularis* and *T.*  
1029 *urticae* have been previously identified using proteomic methods by Kim *et al.* [120] and Jonckheere  
1030 *et al.* [126], respectively. Proteins that might be present in the saliva of *D. tinctorium* and *L. deliense*  
1031 were identified by clustering all predicted protein-coding sequences in the new genomes with these  
1032 *T. urticae* and *I. scapularis* salivary proteins using OrthoMCL [207]. Venn diagrams were constructed  
1033 using InteractiVenn [221].

#### 1034 Sample preparation for proteomics

1035 Engorged larvae of *L. deliense* (10 specimens) were pooled from infested rodent specimens (*Bandicota*  
1036 *indica*, *Bandicota savilei* and *Rattus tanezumi*) captured across several provinces in Thailand during  
1037 the field studies of the CERoPath project [177]. The chigger samples were fixed in 70% ethanol and  
1038 identified as *L. deliense* using autofluorescence microscopy [222]. For *D. tinctorium*, proteomic  
1039 analysis was performed on a single ethanol-fixed individual from the same collection used for genome  
1040 sequencing.

1041 The chiggers were washed with chilled 50 mM ammonium bicarbonate. Soluble protein extracts were  
1042 prepared by homogenisation in 0.1% w/v, Rapigest (Waters) in 50 mM ammonium bicarbonate using a  
1043 polypropylene mini-pestle. This was followed by three cycles of sonication on ice (Vibra-cell 130PB  
1044 sonicator at 20 Hz with microprobe; 10 sec of sonication alternating with 30 sec of incubation on ice).

1045 Samples were centrifuged at  $13,000 \times g$  for 10 min at  $4^{\circ}\text{C}$ . The supernatant was removed and retained.  
1046 The *D. tinctorium* individual was homogenised using a mini-pestle in lysis buffer (4% SDS, Tris-  
1047 hydrochloride, pH 7.6) and sonicated and centrifuged as above. Supernatants from both mite  
1048 preparations were stored at  $-80^{\circ}\text{C}$ .

1049 Protein concentrations of the samples were determined using a Bradford protein assay (Thermo Fisher  
1050 Scientific). The *L. deliense* protein extract was reduced with 3 mM dithiothreitol (Sigma) at  $60^{\circ}\text{C}$  for 10  
1051 min, cooled, then alkylated with 9 mM iodoacetamide (Sigma) at room temperature for 30 min in the  
1052 dark; all steps were performed with intermittent vortex-mixing. Proteomic-grade trypsin (Sigma) was  
1053 added at a protein:trypsin ratio of 50:1 and incubated at  $37^{\circ}\text{C}$  overnight. Rapigest was removed by  
1054 adding trifluoroacetic acid (TFA) to a final concentration of 0.5% (v/v). Peptide samples were  
1055 centrifuged at  $13,000 \times g$  for 30 min to remove precipitated Rapigest. The *D. tinctorium* protein extract  
1056 was reduced, alkylated and digested with trypsin using the filter-aided sample preparation approach  
1057 [57]. Peptides from *D. tinctorium* were split into eight fractions using the Pierce High pH Reversed-  
1058 Phase Peptide Fractionation Kit according to the manufacturer's instructions. Each digest and fraction  
1059 was concentrated and desalted using C18 Stage tips (Thermo Fisher Scientific), then dried using a  
1060 centrifugal vacuum concentrator (Eppendorf) and re-suspended in a 0.1% (v/v) TFA, 3% (v/v)  
1061 acetonitrile solution.

## 1062 Mass spectrometry

1063 Peptides were analysed by on-line nanoflow LC using the Ultimate 3000 nano system (Dionex/Thermo  
1064 Fisher Scientific). Samples were loaded onto a trap column (Acclaim PepMap 100, 2 cm  $\times$  75  $\mu\text{m}$  inner  
1065 diameter, C18, 3  $\mu\text{m}$ , 100  $\text{\AA}$ ) at 9  $\mu\text{l}$  /min with an aqueous solution containing 0.1% (v/v) TFA and 2%  
1066 (v/v) acetonitrile. After 3 min, the trap column was set in-line to an analytical column (Easy-Spray  
1067 PepMap<sup>®</sup> RSLC 50 cm  $\times$  75  $\mu\text{m}$  inner diameter, C18, 2  $\mu\text{m}$ , 100  $\text{\AA}$ ) fused to a silica nano-electrospray  
1068 emitter (Dionex). The column was operated at a constant temperature of  $35^{\circ}\text{C}$  and the LC system was  
1069 coupled to a Q-Exactive mass spectrometer (Thermo Fisher Scientific). Chromatography was

1070 performed with a buffer system consisting of 0.1 % formic acid (buffer A) and 80 % acetonitrile in 0.1  
1071 % formic acid (buffer B). The peptides were separated by a linear gradient of 3.8 – 50 % buffer B over  
1072 90 minutes (*D. tinctorium* and *L. deliense* whole digests) or 30 min (*D. tinctorium* fractions) at a flow  
1073 rate of 300 nl/min. The Q-Exactive was operated in data-dependent mode with survey scans acquired  
1074 at a resolution of 70,000 at m/z 200. Scan range was 300 to 2000m/z. Up to the top 10 most abundant  
1075 isotope patterns with charge states +2 to +5 from the survey scan were selected with an isolation  
1076 window of 2.0 Th and fragmented by higher energy collisional dissociation with normalized collision  
1077 energies of 30. The maximum ion injection times for the survey scan and the MS/MS scans were 250  
1078 and 50 ms, respectively, and the ion target value was set to  $1 \times 10^6$  for survey scans and  $1 \times 10^4$  for  
1079 the MS/MS scans. The MS/MS events were acquired at a resolution of 17,500. Repetitive sequencing  
1080 of peptides was minimized through dynamic exclusion of the sequenced peptides for 20 sec.

#### 1081 Protein identification, quantification and enrichment analysis

1082 Thermo RAW files were imported into Progenesis LC–MS (version 4.1, Nonlinear Dynamics). Peaks  
1083 were picked by the software using default settings and filtered to include only peaks with a charge  
1084 state between +2 and +7. Spectral data were converted into .mgf files with Progenesis LC–MS and  
1085 exported for peptide identification using the Mascot (version 2.3.02, Matrix Science) search engine as  
1086 described above. Tandem MS data were searched against a database including translated ORFs from  
1087 either the *D. tinctorium* genome (DinoT\_V2\_aug2016, 19,258 sequences; 8,386,445 residues) and a  
1088 contaminant database (cRAP, GPMDB, 2012) (119 sequences; 40,423 residues), or the *L. deliense*  
1089 genome (L\_deliense\_V2\_Aug16, 15,096 sequences; 5,183,596 residues), the *Rattus norvegicus*  
1090 genome (UniProt, Apr16 7,948 sequences; 4,022,300 residues) and a contaminant database (cRAP,  
1091 GPMDB, 2012) (119 sequences; 40,423 residues). The search parameters were as follows: the  
1092 precursor mass tolerance was set to 10 ppm and the fragment mass tolerance was set as 0.05 Da. Two  
1093 missed tryptic cleavages were permitted. Carbamidomethylation (cysteine) was set as a fixed  
1094 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  $\geq 2$  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; SMOG: 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.

1144 **Tables**1145 **Table 1:** Characteristics of C-type lectin domain proteins in the trombidid mite genomes.

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

1146 <sup>a</sup>Determined by SecretomeP 2.0 [209].1147 <sup>b</sup>Determined by TMHMM 2.0 [210].1148 <sup>c</sup>Determined by NetNGlyc 1.0 [211].

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

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

1151 <sup>a</sup>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.

Orthologous cluster	Number of genes in cluster			Representative gene ID	Representative annotation <sup>a</sup>
	<i>I. scapularis</i>	<i>L. deliense</i>	<i>D. tinctorium</i>		
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 [ <i>Rhipicephalus pulchellus</i> ]
ORTHOMCL3018	1	1	3	EEC14134	Calponin
ORTHOMCL5249	2	1	1	JAB84323	Stretch regulated skeletal muscle protein [ <i>Ixodes ricinus</i> ]
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 [ <i>Ixodes ricinus</i> ]
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 [ <i>Ixodes ricinus</i> ]
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

1154 <sup>a</sup>Annotations are from *I. scapularis* unless otherwise stated. Protein domain information was obtained from the Conserved Domain Database [228]. Tick salivary protein data  
1155 were obtained from Kim *et al.* [120].

## 1156 **Figure legends**

1157 **Figure 1:** Phylogenetic tree based on the amino acid sequences of 527 one-to-one orthologous genes  
1158 in 11 species of Ecdysozoa using Bayesian methods. The taxonomy of the trombidid mites follows the  
1159 scheme of Lindquist *et al.* [14].

1160 **Figure 2:** Simplified lifecycle of trombidid mites. In the Trombidoidea (velvet mites), the larvae are  
1161 parasitic on other arthropods; whereas in the Trombiculoidea (“chiggers”), the larvae feed on a variety  
1162 of vertebrates or (more rarely) other invertebrates. The deutonymph and adult are free-living, edaphic  
1163 stages that predate soft-bodied arthropods (*e.g.*, termites, springtails and other mites) or consume  
1164 their eggs. Trombidid eggs are laid in the environment and produce questing larvae that congregate  
1165 and seek a host. For clarity, only the deutonymph and active instars are shown; the protonymph  
1166 (between the larva and deutonymph stages) and the tritonymph (between the deutonymph and adult  
1167 stages) are calyptostatic.

1168 **Figure 3:** Blob-plot of contigs assembled from sequence data derived from engorged *Leptotrombidium*  
1169 *deliense* larvae. Blue = Eukaryota; red = Bacteria; white = other; grey = no hit.

1170 **Figure 4:** Estimated divergence times using a relaxed molecular clock with fossil calibration time and  
1171 classification of protein-coding genes between 11 species of Ecdysozoa. *Caenorhabditis elegans* was  
1172 used as the outgroup and the bootstrap value was set as 10,000,000. The 1:1:1 orthologs comprise  
1173 the common orthologs with the same copy numbers in different species, and the N:N:N orthologs  
1174 comprise the common orthologs with different copy numbers in these species. Patchy orthologs are  
1175 shared between more than one, but not all species (excluding those belonging to the previous  
1176 categories). Unclustered genes are those that cannot be clustered into gene families.

1177 **Figure 5:** Overrepresented Pfam domains in proteomic datasets generated from a single adult  
1178 *Dinothrombium tinctorium* and a pool of engorged *Leptotrombidium deliense* larvae. The colour scale  
1179 represents fold-enrichment.

1180 **Figure 6:** Phylogeny of terpene synthases from cluster ORTHOMCL47 of trombidid mites alongside  
1181 related genes from bacteria and fungi. The tree was constructed using a maximum likelihood method;  
1182 poorly-supported nodes are highlighted in red font.

1183 **Figure 7:** Phylogeny of terpene synthases from cluster ORTHOMCL881 of *Leptotrombidium deliense*  
1184 alongside related genes from bacteria. Metazoan terpene synthases from galerucid beetles are shown  
1185 as an outgroup. The tree was constructed using a maximum likelihood method; poorly-supported  
1186 nodes are highlighted in red font.

1187 **Figure 8:** Blob-plot of read coverage and GC content for terpene synthase genes in *Dinothrombium*  
1188 *tinctorium* (Dt) and *Leptotrombidium deliense* (Ld). Terpene synthases (in red, “Bacteria”) are shown  
1189 in relation to all other mite genes (in blue).

1190 **Figure 9:** Phylogeny of Gag-like polyproteins from *Leptotrombidium deliense* in relation to homologous  
1191 sequences from small mammals. The tree was constructed using a maximum likelihood method;  
1192 poorly-supported nodes are highlighted in red font.

1193 **Figure 10:** Phylogeny and classification of metazoan opsins. The tree was constructed using a  
1194 neighbour-joining method. Poorly-supported nodes are highlighted in red font.

1195 **Figure 11:** Phylogeny of *Dinothrombium tinctorium*, *Leptotrombidium deliense*, *Tetranychus urticae*  
1196 and *Drosophila melanogaster* ionotropic receptors and ionotropic glutamate receptors. The tree was  
1197 constructed using a maximum-likelihood method.

1198 **Figure 12:** Venn diagram of orthologous clusters of predicted allergens from six species of acariform  
1199 mites.

1200 **Figure 13:** Orthologous clusters of predicted allergens in *Leptotrombidium deliense* as classified by the  
1201 AllFam database. “Peptides” refers to the number of unique peptides from each allergen detected by  
1202 mass spectrometric analysis of a pool of *L. deliense* larvae. The heat-map indicates the number of  
1203 orthologues for each predicted allergen in the genomes of six acariform mites. Ld, *Leptotrombidium*

1204 *deliense*; Dt, *Dinotrombium 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 *Dinotrombium*  
1209 *tinctorium*. Thumbnail images indicate representative host species.



1210 **Additional files**

1211 **Additional file 1:** Genome assembly and gene set statistics compared with 14 other arachnids (table  
1212 in .xlsx format).

1213 **Additional file 2:** K-mer distributions for *Leptotrombidium deliense* (A) and *Dinotrombium tinctorium*  
1214 (B) plotted by GenomeScope (figure in .pdf format).

1215 **Additional file 3:** Identification of repetitive sequences in the *Dinotrombium tinctorium* and  
1216 *Leptotrombidium deliense* assemblies compared with other acariform mites (table in .pdf format).

1217 **Additional file 4:** The number of gene families shared among acariform mites (*Dinotrombium*  
1218 *tinctorium*, *Leptotrombidium deliense*, *Tetranychus urticae* and *Sarcoptes scabiei*); alongside other  
1219 references including *Drosophila melanogaster*, *Apis mellifera*, *Tropilaelaps mercedesae*, *Metaseiulus*  
1220 *occidentalis*, *Ixodes scapularis*, *Stegodyphus mimosarum* and *Caenorhabditis elegans* by the OrthoMCL  
1221 classification algorithm (figure in .pdf format).

1222 **Additional file 5:** Gene family contraction and expansion in 11 species of Ecdysozoa (figure in .pdf  
1223 format).

1224 **Additional file 6:** Changes of gene family size in trombidid mites in comparison with three other  
1225 acariform mites (table in .xlsx format).

1226 **Additional file 7:** High-confidence protein identifications and abundance scores for *Leptotrombidium*  
1227 *deliense* engorged larvae (spreadsheet in .xlsx format).

1228 **Additional file 8:** High-confidence protein identifications and abundance scores for a single adult *D.*  
1229 *tinctorium* individual (spreadsheet in .xlsx format).

1230 **Additional file 9:** Phylogeny of carotenoid synthases-cyclases from trombidid mites, spider mites,  
1231 aphids and fungi (figure in .pdf format).

- 1232 **Additional file 10:** Genomic scaffold of *Dinotrombium 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 *Dinotrombium tinctorium* (figure in .pdf format).
- 1238 **Additional file 13:** Phylogeny of reverse ribonuclease integrases in *Dinotrombium 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 *Dinotrombium tinctorium*,  
1245 *Leptotrombidium deliense*, *Tetranychus urticae* and *Ixodes scapularis* (figure in .pdf format).
- 1246 **Additional file 17:** Phylogeny of peptidoglycan recognition protein sequences in *Dinotrombium*  
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 *Dinotrombium*  
1254 *tinctorium* and *Leptotrombidium deliense* (table in .pdf format).

## References

- 1255  
1256  
1257 1. Krantz GW. Introduction. In: Krantz GW, Walter, D.E., editor. A Manual of Acarology.  
1258 Lubbock, Texas: Texas Tech University Press; 2009. p. 1-2.  
1259 2. Danchin EGJ. Lateral gene transfer in eukaryotes: tip of the iceberg or of the ice cube? *Bmc*  
1260 *Biol.* 2016;14.  
1261 3. Ku C and Martin WF. A natural barrier to lateral gene transfer from prokaryotes to  
1262 eukaryotes revealed from genomes: the 70 % rule. *Bmc Biol.* 2016;14 1:89.  
1263 doi:10.1186/s12915-016-0315-9.  
1264 4. Koutsovoulos G, Kumar S, Laetsch DR, Stevens L, Daub J, Conlon C, et al. No evidence for  
1265 extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*.  
1266 *Proc Natl Acad Sci U S A.* 2016;113 18:5053-8. doi:10.1073/pnas.1600338113.  
1267 5. Boothby TC, Tenlen JR, Smith FW, Wang JR, Patanella KA, Nishimura EO, et al. Evidence for  
1268 extensive horizontal gene transfer from the draft genome of a tardigrade. *Proc Natl Acad*  
1269 *Sci U S A.* 2015;112 52:15976-81. doi:10.1073/pnas.1510461112.  
1270 6. Crisp A, Boschetti C, Perry M, Tunnacliffe A and Micklem G. Expression of multiple  
1271 horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes.  
1272 *Genome Biol.* 2015;16:50. doi:10.1186/s13059-015-0607-3.  
1273 7. Salzberg SL. Horizontal gene transfer is not a hallmark of the human genome. *Genome Biol.*  
1274 2017;18 1:85. doi:10.1186/s13059-017-1214-2.  
1275 8. Acuna R, Padilla BE, Florez-Ramos CP, Rubio JD, Herrera JC, Benavides P, et al. Adaptive  
1276 horizontal transfer of a bacterial gene to an invasive insect pest of coffee. *Proc Natl Acad*  
1277 *Sci U S A.* 2012;109 11:4197-202. doi:10.1073/pnas.1121190109.  
1278 9. Haegeman A, Jones JT and Danchin EG. Horizontal gene transfer in nematodes: a catalyst  
1279 for plant parasitism? *Mol Plant Microbe Interact.* 2011;24 8:879-87. doi:10.1094/MPMI-03-  
1280 11-0055.  
1281 10. Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, et al. Horizontal gene transfer from  
1282 diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell.*  
1283 2013;153 7:1567-78. doi:10.1016/j.cell.2013.05.040.  
1284 11. Wybouw N, Pauchet Y, Heckel DG and Van Leeuwen T. Horizontal Gene Transfer  
1285 Contributes to the Evolution of Arthropod Herbivory. *Genome Biol Evol.* 2016;8 6:1785-  
1286 801. doi:10.1093/gbe/evw119.  
1287 12. Wybouw N, Dermauw W, Tirry L, Stevens C, Grbic M, Feyereisen R, et al. A gene horizontally  
1288 transferred from bacteria protects arthropods from host plant cyanide poisoning. *Elife.*  
1289 2014;3:e02365. doi:10.7554/eLife.02365.  
1290 13. Bryon A, Kurlovs AH, Dermauw W, Greenhalgh R, Riga M, Grbic M, et al. Disruption of a  
1291 horizontally transferred phytoene desaturase abolishes carotenoid accumulation and  
1292 diapause in *Tetranychus urticae*. *Proc Natl Acad Sci U S A.* 2017;114 29:E5871-E80.  
1293 doi:10.1073/pnas.1706865114.  
1294 14. Lindquist EE, Krantz, G.W., Walter, D.E. Classification. In: Krantz GW, Walter, D.E., editor. A  
1295 Manual of Acarology. Lubbock, Texas: Texas Tech University Press; 2009. p. 97-103.  
1296 15. Snell AE and Heath ACG. Parasitism of mosquitoes (Diptera : Culicidae) by larvae of  
1297 Arrenuridae and Microtrombidiidae (Acari : Parasitengona) in the Wellington region, New  
1298 Zealand. *New Zeal J Zool.* 2006;33 1:9-15.  
1299 16. McGarry JW, Gusbi AM, Baker A, Hall MJR and Elmegademi K. Phoretic and Parasitic Mites  
1300 Infesting the New-World Screwworm Fly, *Cochliomyia-Hominivorax*, Following Sterile  
1301 Insect Releases in Libya. *Med Vet Entomol.* 1992;6 3:255-60.  
1302 17. Goldarazena A and Zhang ZQ. Seasonal abundance of *Allothrombium monochaetum* and  
1303 *Allothrombium pulvinum* in Navarra-Nafarroa (northern Spain), with notes on larval host  
1304 preference and rate of parasitism. *Exp Appl Acarol.* 1999;23 12:987-93.

- 1305 18. Zhang ZQ. Biology and ecology of trombiculid mites (Acari : Trombidioidea). *Experimental &*  
1306 *Applied Acarology*. 1998;22 3:139-55.
- 1307 19. Zhang ZQ and Xin JL. Biology of *Allothrombium-Pulvinum* (Acariformes, Trombididae), a  
1308 Potential Biological-Control Agent of Aphids in China. *Experimental & Applied Acarology*.  
1309 1989;6 2:101-8.
- 1310 20. Shatrov AB, Kudryashova, N.I. Taxonomy, life cycles and the origin of parasitism in  
1311 trombiculid mites. In: Morand S, Krasnov, B.R, Poulin, R, editor. *Micromammals and*  
1312 *Macroparasites: From Evolutionary Ecology to Management*. Tokyo: Springer Japan; 2006.  
1313 p. 119-40.
- 1314 21. Nadchatram M. Correlation of habitat, environment and color of chiggers, and their  
1315 potential significance in the epidemiology of scrub typhus in Malaya (Prostigmata:  
1316 Trombiculidae). *J Med Entomol*. 1970;7 2:131-44.
- 1317 22. Lakshana P. A new species of trombiculid mite infesting scorpions in Thailand (Acarina,  
1318 Trombiculidae). *J Med Entomol*. 1966;3 3:258-60.
- 1319 23. Audy JR. Trombiculid mites infesting birds, reptiles, and arthropods in Malaya, with a  
1320 taxonomic revision, and descriptions of a new genus, two new subgenera, and six new  
1321 species. *Bulletin of the Raffles Museum*. 1956;28:27-80.
- 1322 24. Xu G, Walker DH, Jupiter D, Melby PC and Arcari CM. A review of the global epidemiology  
1323 of scrub typhus. *Plos Neglect Trop D*. 2017;11 11.
- 1324 25. Tilak R, Kunwar R, Wankhade UB and Tilak VW. Emergence of *Schoengastiella ligula* as the  
1325 vector of scrub typhus outbreak in Darjeeling: has *Leptotrombidium deliense* been  
1326 replaced? *Indian J Public Health*. 2011;55 2:92-9. doi:10.4103/0019-557X.85239.
- 1327 26. Chaisiri K, Stekolnikov AA, Makepeace BL and Morand S. A Revised Checklist of Chigger  
1328 Mites (Acari: Trombiculidae) From Thailand, with the Description of Three New Species. *J*  
1329 *Med Entomol*. 2016;53 2:321-42. doi:10.1093/jme/tjv244.
- 1330 27. Taylor AJ, Paris DH and Newton PN. A Systematic Review of Mortality from Untreated Scrub  
1331 Typhus (*Orientia tsutsugamushi*). *PLoS Negl Trop Dis*. 2015;9 8:e0003971.  
1332 doi:10.1371/journal.pntd.0003971.
- 1333 28. Bonell A, Lubell Y, Newton PN, Crump JA and Paris DH. Estimating the burden of scrub  
1334 typhus: A systematic review. *PLoS Negl Trop Dis*. 2017;11 9:e0005838.  
1335 doi:10.1371/journal.pntd.0005838.
- 1336 29. Weitzel T, Dittrich S, Lopez J, Phuklia W, Martinez-Valdebenito C, Velasquez K, et al.  
1337 Endemic Scrub Typhus in South America. *N Engl J Med*. 2016;375 10:954-61.  
1338 doi:10.1056/NEJMoa1603657.
- 1339 30. Izzard L, Fuller A, Blacksell SD, Paris DH, Richards AL, Aukkanit N, et al. Isolation of a novel  
1340 *Orientia* species (*O. chuto* sp. nov.) from a patient infected in Dubai. *J Clin Microbiol*.  
1341 2010;48 12:4404-9. doi:10.1128/JCM.01526-10.
- 1342 31. Maina AN, Farris CM, Odhiambo A, Jiang J, Laktabai J, Armstrong J, et al. Q Fever, Scrub  
1343 Typhus, and Rickettsial Diseases in Children, Kenya, 2011-2012. *Emerg Infect Dis*. 2016;22  
1344 5:883-6. doi:10.3201/eid2205.150953.
- 1345 32. Yu XJ and Tesh RB. The role of mites in the transmission and maintenance of Hantaan virus  
1346 (Hantavirus: Bunyaviridae). *J Infect Dis*. 2014;210 11:1693-9. doi:10.1093/infdis/jiu336.
- 1347 33. Kabeya H, Colborn JM, Bai Y, Lerthusnee K, Richardson JH, Maruyama S, et al. Detection  
1348 of *Bartonella tamiae* DNA in Ectoparasites from Rodents in Thailand and Their Sequence  
1349 Similarity with Bacterial Cultures from Thai Patients. *Vector-Borne Zoonot*. 2010;10 5:429-  
1350 34.
- 1351 34. Huang Y, Zhao L, Zhang Z, Liu M, Xue Z, Ma D, et al. Detection of a Novel Rickettsia From  
1352 *Leptotrombidium scutellare* Mites (Acari: Trombiculidae) From Shandong of China. *J Med*  
1353 *Entomol*. 2017;54 3:544-9. doi:10.1093/jme/tjw234.

- 1354 35. Smith GA, Sharma V, Knapp JF and Shields BJ. The summer penile syndrome: seasonal acute  
1355 hypersensitivity reaction caused by chigger bites on the penis. *Pediatr Emerg Care.* 1998;14  
1356 2:116-8.
- 1357 36. Little SE, Carmichael KP and Rakich PM. Trombidiosis-induced dermatitis in white-tailed  
1358 deer (*Odocoileus virginianus*). *Vet Pathol.* 1997;34 4:350-2.  
1359 doi:10.1177/030098589703400412.
- 1360 37. Leone F, Di Bella A, Vercelli A and Cornegliani L. Feline trombiculosis: a retrospective study  
1361 in 72 cats. *Vet Dermatol.* 2013;24 5:535-e126. doi:10.1111/vde.12053.
- 1362 38. Faccini JL, Santos AC, Santos SB, Jacinavicius FC, Bassini-Silva R and Barros-Battesti DM.  
1363 Trombiculiasis in domestic goats and humans in the state of Maranhao, Brazil. *Rev Bras*  
1364 *Parasitol Vet.* 2017;26 1:104-9. doi:10.1590/S1984-29612016088.
- 1365 39. Shatrov AB. Stylostome formation in trombiculid mites (Acariformes: Trombiculidae). *Exp*  
1366 *Appl Acarol.* 2009;49 4:261-80.
- 1367 40. Shatrov AB and Felska M. Comparative stylostome ultrastructure of *Hirsutiella zachvatkini*  
1368 (*Trombiculidae*) and *Trombidium holosericeum* (*Trombidiidae*) larvae. *Exp Appl Acarol.*  
1369 2017;72 4:339-65.
- 1370 41. Minter DM. Some field and laboratory observations on the British harvest mite, *Trombicula*  
1371 *autumnalis* Shaw. *Parasitology.* 1957;47 1-2:185-93.
- 1372 42. Singer G. *Neotrombidium leonardi* (Acarina: Trombidoidea), Pt. 2 : Biology of *N. beeri* and  
1373 other species of the genus. *Acarologia.* 1971;13:143-51.
- 1374 43. Tevis L and Newell IM. Studies on Biology and Seasonal Cycle of Giant Red Velvet Mite,  
1375 *Dinotrombium-Pandorae* (Acari, Trombidiidae). *Ecology.* 1962;43 3:497-&.
- 1376 44. Zhang ZQ. Notes on the Occurrence and Distribution of the Biocontrol Agent,  
1377 *Allothrombium-Pulvinum* Ewing (Acari, Trombidiidae), in a Peach Orchard in China. *J Appl*  
1378 *Entomol.* 1992;113 1:13-7.
- 1379 45. Wiggins GJ, Grant JF and Welbourn WC. *Allothrombium mitchelli* (Acari : Trombidiidae) in  
1380 the Great Smoky Mountains National Park: Incidence, seasonality, and predation on beech  
1381 scale (Homoptera : Eriococcidae). *Ann Entomol Soc Am.* 2001;94 6:896-901.
- 1382 46. Traub R and Wisseman CL, Jr. Ecological considerations in scrub typhus. 2. Vector species.  
1383 *Bull World Health Organ.* 1968;39 2:219-30.
- 1384 47. Walter DE, Lindquist, E.E., Smith, I.M., Cook, D.R., Krantz, G.W. Order Trombidiformes. In:  
1385 Krantz GW, Walter, D.E., editor. *A Manual of Acarology.* Lubbock, Texas: Texas Tech  
1386 University Press; 2009. p. 233-419.
- 1387 48. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM and Birol I. ABySS: A parallel  
1388 assembler for short read sequence data. *Genome Res.* 2009;19 6:1117-23.
- 1389 49. Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, et al. De novo assembly and  
1390 analysis of RNA-seq data. *Nat Methods.* 2010;7 11:909-U62.
- 1391 50. Zerbino DR and Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn  
1392 graphs. *Genome Res.* 2008;18 5:821-9.
- 1393 51. Kumar S, Jones M, Koutsovoulos G, Clarke M and Blaxter M. Blobology: exploring raw  
1394 genome data for contaminants, symbionts and parasites using taxon-annotated GC-  
1395 coverage plots. *Front Genet.* 2013;4:237. doi:10.3389/fgene.2013.00237.
- 1396 52. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new  
1397 genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.*  
1398 2012;19 5:455-77. doi:10.1089/cmb.2012.0021.
- 1399 53. Cantarel B, Korf I, Robb S, Parra G, Ross E, Moore B, et al. MAKER: An easy-to-use  
1400 annotation pipeline designed for emerging model organism genomes. *Genome Research.*  
1401 2008;18 1:188-96. doi:10.1101/gr.6743907.
- 1402 54. Stanke M and Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes  
1403 that allows user-defined constraints. *Nucleic Acids Res.* 2005;33 Web Server issue:W465-7.  
1404 doi:10.1093/nar/gki458.

- 1405 55. Korf I. Gene finding in novel genomes. *BMC Bioinformatics*. 2004;5:59. doi:10.1186/1471-  
1406 2105-5-59.
- 1407 56. Lukashin AV and Borodovsky M. GeneMark.hmm: new solutions for gene finding. *Nucleic*  
1408 *Acids Res*. 1998;26 4:1107-15.
- 1409 57. Wisniewski JR, Zougman A, Nagaraj N and Mann M. Universal sample preparation method  
1410 for proteome analysis. *Nat Methods*. 2009;6 5:359-62. doi:10.1038/nmeth.1322.
- 1411 58. Rider SD, Jr., Morgan MS and Arlian LG. Draft genome of the scabies mite. *Parasit Vectors*.  
1412 2015;8:585. doi:10.1186/s13071-015-1198-2.
- 1413 59. Grbic M, Van Leeuwen T, Clark RM, Rombauts S, Rouze P, Grbic V, et al. The genome of  
1414 *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature*. 2011;479 7374:487-92.  
1415 doi:10.1038/nature10640.
- 1416 60. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV and Zdobnov EM. BUSCO: assessing  
1417 genome assembly and annotation completeness with single-copy orthologs.  
1418 *Bioinformatics*. 2015;31 19:3210-2. doi:10.1093/bioinformatics/btv351.
- 1419 61. Sanggaard KW, Bechsgaard JS, Fang X, Duan J, Dyrlynd TF, Gupta V, et al. Spider genomes  
1420 provide insight into composition and evolution of venom and silk. *Nat Commun*.  
1421 2014;5:3765. doi:10.1038/ncomms4765.
- 1422 62. Oba Y, Sato M, Ojika M and Inouye S. Enzymatic and genetic characterization of firefly  
1423 luciferase and *Drosophila* CG6178 as a fatty acyl-CoA synthetase. *Biosci Biotechnol*  
1424 *Biochem*. 2005;69 4:819-28. doi:10.1271/bbb.69.819.
- 1425 63. Okada I, Hamanoue H, Terada K, Tohma T, Megarbane A, Chouery E, et al. SMOC1 is  
1426 essential for ocular and limb development in humans and mice. *Am J Hum Genet*. 2011;88  
1427 1:30-41. doi:10.1016/j.ajhg.2010.11.012.
- 1428 64. Thomas JT, Eric Dollins D, Andrykovich KR, Chu T, Stultz BG, Hursh DA, et al. SMOC can act  
1429 as both an antagonist and an expander of BMP signaling. *Elife*. 2017;6  
1430 doi:10.7554/eLife.17935.
- 1431 65. Ayme-Southgate A, Lasko P, French C and Pardue ML. Characterization of the gene for  
1432 mp20: a *Drosophila* muscle protein that is not found in asynchronous oscillatory flight  
1433 muscle. *J Cell Biol*. 1989;108 2:521-31.
- 1434 66. Huang Y, Li W, Huang L, Hu Y, Chen W, Wang X, et al. Identification and characterization of  
1435 myophilin-like protein: a life stage and tissue-specific antigen of *Clonorchis sinensis*.  
1436 *Parasitol Res*. 2012;111 3:1143-50. doi:10.1007/s00436-012-2946-2.
- 1437 67. Friedrich MV, Schneider M, Timpl R and Baumgartner S. Perlecan domain V of *Drosophila*  
1438 *melanogaster*. Sequence, recombinant analysis and tissue expression. *Eur J Biochem*.  
1439 2000;267 11:3149-59.
- 1440 68. Klinkowstrom AM, Terra WR and Ferreira C. Midgut Dipeptidases from *Rhynchosciara-*  
1441 *Americana* (Diptera) Larvae - Properties of Soluble and Membrane-Bound Forms. *Insect*  
1442 *Biochem Molec*. 1995;25 3:303-10.
- 1443 69. Long JZ, Svensson KJ, Bateman LA, Lin H, Kamenecka T, Lokurkar IA, et al. The Secreted  
1444 Enzyme PM20D1 Regulates Lipidated Amino Acid Uncouplers of Mitochondria. *Cell*.  
1445 2016;166 2:424-35. doi:10.1016/j.cell.2016.05.071.
- 1446 70. Shahein YE, El-Rahim MT, Hussein NA, Hamed RR, El-Hakim AE and Barakat MM. Molecular  
1447 cloning of a small heat shock protein (sHSP11) from the cattle tick *Rhipicephalus* (*Boophilus*)  
1448 *annulatus* salivary gland. *Int J Biol Macromol*. 2010;47 5:614-22.  
1449 doi:10.1016/j.ijbiomac.2010.08.006.
- 1450 71. Matsumoto I, Watanabe H, Abe K, Arai S and Emori Y. A putative digestive cysteine  
1451 proteinase from *Drosophila melanogaster* is predominantly expressed in the embryonic  
1452 and larval midgut. *Eur J Biochem*. 1995;227 1-2:582-7.
- 1453 72. Santamaria ME, Hernandez-Crespo P, Ortego F, Grbic V, Grbic M, Diaz I, et al. Cysteine  
1454 peptidases and their inhibitors in *Tetranychus urticae*: a comparative genomic approach.  
1455 *BMC Genomics*. 2012;13:307. doi:10.1186/1471-2164-13-307.

- 1456 73. Santamaria ME, Gonzalez-Cabrera J, Martinez M, Grbic V, Castanera P, Diaz L, et al.  
1457 Digestive proteases in bodies and faeces of the two-spotted spider mite, *Tetranychus*  
1458 *urticae*. *J Insect Physiol.* 2015;78:69-77. doi:10.1016/j.jinsphys.2015.05.002.
- 1459 74. Kuwahara Y, Ichiki Y, Morita M, Tanabe T and Asano Y. Chemical Polymorphism in Defense  
1460 Secretions during Ontogenetic Development of the Millipede *Niponia nodulosa*. *J Chem*  
1461 *Ecol.* 2015;41 1:15-21.
- 1462 75. Kenny NJ, Shen X, Chan TT, Wong NW, Chan TF, Chu KH, et al. Genome of the Rusty  
1463 Millipede, *Trigoniulus corallinus*, Illuminates Diplopod, Myriapod, and Arthropod Evolution.  
1464 *Genome Biol Evol.* 2015;7 5:1280-95. doi:10.1093/gbe/evv070.
- 1465 76. Skelton AC, Cameron MM, Pickett JA and Birkett MA. Identification of neryl formate as the  
1466 airborne aggregation pheromone for the American house dust mite and the European  
1467 house dust mite (Acari: Epidermoptidae). *J Med Entomol.* 2010;47 5:798-804.
- 1468 77. Chan TF, Ji KM, Yim AK, Liu XY, Zhou JW, Li RQ, et al. The draft genome, transcriptome, and  
1469 microbiome of *Dermatophagoides farinae* reveal a broad spectrum of dust mite allergens.  
1470 *J Allergy Clin Immunol.* 2015;135 2:539-48. doi:10.1016/j.jaci.2014.09.031.
- 1471 78. Beran F, Rahfeld P, Luck K, Nagel R, Vogel H, Wielsch N, et al. Novel family of terpene  
1472 synthases evolved from trans-isoprenyl diphosphate synthases in a flea beetle. *Proc Natl*  
1473 *Acad Sci U S A.* 2016;113 11:2922-7. doi:10.1073/pnas.1523468113.
- 1474 79. Faddeeva-Vakhrusheva A, Kraaijeveld K, Derks MFL, Anvar SY, Agamennone V, Suring W, et  
1475 al. Coping with living in the soil: the genome of the parthenogenetic springtail *Folsomia*  
1476 *candida*. *BMC Genomics.* 2017;18 1:493. doi:10.1186/s12864-017-3852-x.
- 1477 80. Komatsu M, Tsuda M, Omura S, Oikawa H and Ikeda H. Identification and functional analysis  
1478 of genes controlling biosynthesis of 2-methylisoborneol. *Proc Natl Acad Sci U S A.* 2008;105  
1479 21:7422-7. doi:10.1073/pnas.0802312105.
- 1480 81. Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, et al. Arthropod relationships  
1481 revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature.* 2010;463  
1482 7284:1079-83. doi:10.1038/nature08742.
- 1483 82. Mason VC, Li G, Minx P, Schmitz J, Churakov G, Doronina L, et al. Genomic analysis reveals  
1484 hidden biodiversity within colugos, the sister group to primates. *Sci Adv.* 2016;2  
1485 8:e1600633. doi:10.1126/sciadv.1600633.
- 1486 83. Webb CH and Luptak A. HDV-like self-cleaving ribozymes. *RNA Biol.* 2011;8 5:719-27.  
1487 doi:10.4161/rna.8.5.16226.
- 1488 84. Valanne S, Wang JH and Ramet M. The *Drosophila* Toll signaling pathway. *J Immunol.*  
1489 2011;186 2:649-56. doi:10.4049/jimmunol.1002302.
- 1490 85. Palmer WJ and Jiggins FM. Comparative Genomics Reveals the Origins and Diversity of  
1491 Arthropod Immune Systems. *Mol Biol Evol.* 2015;32 8:2111-29.  
1492 doi:10.1093/molbev/msv093.
- 1493 86. Shaw DK, Wang X, Brown LJ, Chavez AS, Reif KE, Smith AA, et al. Infection-derived lipids  
1494 elicit an immune deficiency circuit in arthropods. *Nat Commun.* 2017;8:14401.  
1495 doi:10.1038/ncomms14401.
- 1496 87. Brites D and Du Pasquier L. Somatic and Germline Diversification of a Putative  
1497 Immunoreceptor within One Phylum: Dscam in Arthropods. *Results Probl Cell Differ.*  
1498 2015;57:131-58. doi:10.1007/978-3-319-20819-0\_6.
- 1499 88. Chipman AD, Ferrier DE, Brena C, Qu J, Hughes DS, Schroder R, et al. The first myriapod  
1500 genome sequence reveals conservative arthropod gene content and genome organisation  
1501 in the centipede *Strigamia maritima*. *PLoS Biol.* 2014;12 11:e1002005.  
1502 doi:10.1371/journal.pbio.1002005.
- 1503 89. Kaneko T, Yano T, Aggarwal K, Lim JH, Ueda K, Oshima Y, et al. PGRP-LC and PGRP-LE have  
1504 essential yet distinct functions in the *drosophila* immune response to monomeric DAP-type  
1505 peptidoglycan. *Nat Immunol.* 2006;7 7:715-23. doi:10.1038/ni1356.

- 1506 90. Atwal S, Giengkam S, Chaemchuen S, Dorling J, Kosaisawe N, VanNieuwenhze M, et al.  
1507 Evidence for a peptidoglycan-like structure in *Orientia tsutsugamushi*. *Mol Microbiol.*  
1508 2017;105 3:440-52. doi:10.1111/mmi.13709.
- 1509 91. Zelensky AN and Gready JE. The C-type lectin-like domain superfamily. *FEBS J.* 2005;272  
1510 24:6179-217. doi:10.1111/j.1742-4658.2005.05031.x.
- 1511 92. Alenton RR, Koiwai K, Miyaguchi K, Kondo H and Hirono I. Pathogen recognition of a novel  
1512 C-type lectin from *Marsupenaeus japonicus* reveals the divergent sugar-binding specificity  
1513 of QAP motif. *Sci Rep.* 2017;7:45818. doi:10.1038/srep45818.
- 1514 93. Gale M, Jr. and Katze MG. Molecular mechanisms of interferon resistance mediated by  
1515 viral-directed inhibition of PKR, the interferon-induced protein kinase. *Pharmacol Ther.*  
1516 1998;78 1:29-46.
- 1517 94. Pomar N, Berlanga JJ, Campuzano S, Hernandez G, Elias M and de Haro C. Functional  
1518 characterization of *Drosophila melanogaster* PERK eukaryotic initiation factor 2alpha  
1519 (eIF2alpha) kinase. *Eur J Biochem.* 2003;270 2:293-306.
- 1520 95. Chakraborty R, Baek JH, Bae EY, Kim WY, Lee SY and Kim MG. Comparison and contrast of  
1521 plant, yeast, and mammalian ER stress and UPR. *Appl Biol Chem.* 2016;59 3:337-47.
- 1522 96. Nagata T, Koyanagi M, Tsukamoto H and Terakita A. Identification and characterization of  
1523 a protostome homologue of peropsin from a jumping spider. *J Comp Physiol A.* 2010;196  
1524 1:51-9.
- 1525 97. Hoy MA, Waterhouse RM, Wu K, Estep AS, Ioannidis P, Palmer WJ, et al. Genome  
1526 Sequencing of the Phytoseiid Predatory Mite *Metaseiulus occidentalis* Reveals Completely  
1527 Atomized Hox Genes and Superdynamic Intron Evolution. *Genome Biology and Evolution.*  
1528 2016;8 6:1762-75.
- 1529 98. Dong XF, Armstrong SD, Xia D, Makepeace BL, Darby AC and Kadowaki T. Draft genome of  
1530 the honey bee ectoparasitic mite, *Tropilaelaps mercedesae*, is shaped by the parasitic life  
1531 history. *Gigascience.* 2017;6 3.
- 1532 99. Ozaki K, Utoguchi A, Yamada A and Yoshikawa H. Identification and genomic structure of  
1533 chemosensory proteins (CSP) and odorant binding proteins (OBP) genes expressed in  
1534 foreleg tarsi of the swallowtail butterfly *Papilio xuthus*. *Insect Biochem Molec.* 2008;38  
1535 11:969-76.
- 1536 100. Ngoc PCT, Greenhalgh R, Dermauw W, Rombauts S, Bajda S, Zhurov V, et al. Complex  
1537 Evolutionary Dynamics of Massively Expanded Chemosensory Receptor Families in an  
1538 Extreme Generalist Chelicerate Herbivore. *Genome Biology and Evolution.* 2016;8 11:3323-  
1539 39.
- 1540 101. Xia SZ, Miyashita T, Fu TF, Lin WY, Wu CL, Pyzocha L, et al. NMDA receptors mediate  
1541 olfactory learning and memory in *Drosophila*. *Curr Biol.* 2005;15 7:603-15.
- 1542 102. Frank CA. Homeostatic plasticity at the *Drosophila* neuromuscular junction.  
1543 *Neuropharmacology.* 2014;78:63-74.
- 1544 103. Robinson JE, Paluch J, Dickman DK and Joiner WJ. ADAR-mediated RNA editing suppresses  
1545 sleep by acting as a brake on glutamatergic synaptic plasticity. *Nature Communications.*  
1546 2016;7.
- 1547 104. Karupudurai T, Lin TY, Ting CY, Pursley R, Melnattur KV, Diao FQ, et al. A Hard-Wired  
1548 Glutamatergic Circuit Pools and Relays UV Signals to Mediate Spectral Preference in  
1549 *Drosophila*. *Neuron.* 2014;81 3:603-15.
- 1550 105. Hu W, Wang TT, Wang X and Han JH. I-h Channels Control Feedback Regulation from  
1551 Amacrine Cells to Photoreceptors. *Plos Biology.* 2015;13 4.
- 1552 106. Rytz R, Croset V and Benton R. Ionotropic Receptors (IRs): Chemosensory ionotropic  
1553 glutamate receptors in *Drosophila* and beyond. *Insect Biochem Molec.* 2013;43 9:888-97.
- 1554 107. Ni L, Klein M, Svec KV, Budelli G, Chang EC, Ferrer AJ, et al. The Ionotropic Receptors IR21a  
1555 and IR25a mediate cool sensing in *Drosophila*. *Elife.* 2016;5.



- 1556 108. Knecht ZA, Silbering AF, Ni LN, Klein M, Budelli G, Bell R, et al. Distinct combinations of  
1557 variant ionotropic glutamate receptors mediate thermosensation and hygrosensation in  
1558 *Drosophila*. *Elife*. 2016;5.
- 1559 109. Eliash N, Singh NK, Thangarajan S, Sela N, Leshkowitz D, Kamer Y, et al. Chemosensing of  
1560 honeybee parasite, *Varroa destructor*: Transcriptomic analysis. *Sci Rep-Uk*. 2017;7.
- 1561 110. Wright SM, Wikel SK and Wrenn WJ. Host Immune Responsiveness to the Chigger,  
1562 *Eutrombicula-Cinnabaris*. *Ann Trop Med Parasit*. 1988;82 3:283-93.
- 1563 111. Rider SD, Jr., Morgan MS and Arlian LG. Allergen homologs in the *Euroglyphus maynei* draft  
1564 genome. *Plos One*. 2017;12 8:e0183535. doi:10.1371/journal.pone.0183535.
- 1565 112. Radauer C, Bublin M, Wagner S, Mari A and Breiteneder H. Allergens are distributed into  
1566 few protein families and possess a restricted number of biochemical functions. *J Allergy*  
1567 *Clin Immunol*. 2008;121 4:847-52 e7. doi:10.1016/j.jaci.2008.01.025.
- 1568 113. An S, Chen L, Long C, Liu X, Xu X, Lu X, et al. *Dermatophagoides farinae* allergens diversity  
1569 identification by proteomics. *Mol Cell Proteomics*. 2013;12 7:1818-28.  
1570 doi:10.1074/mcp.M112.027136.
- 1571 114. Choopong J, Reamtong O, Sookrung N, Seesuyay W, Indrawattana N, Sakolvaree Y, et al.  
1572 Proteome, Allergenome, and Novel Allergens of House Dust Mite, *Dermatophagoides*  
1573 *farinae*. *J Proteome Res*. 2016;15 2:422-30. doi:10.1021/acs.jproteome.5b00663.
- 1574 115. Fluckiger S, Fijten H, Whitley P, Blaser K and Cramer R. Cyclophilins, a new family of cross-  
1575 reactive allergens. *Eur J Immunol*. 2002;32 1:10-7. doi:10.1002/1521-  
1576 4141(200201)32:1<10::AID-IMMU1062>3.0.CO;2-I.
- 1577 116. Tripathi P, Nair S, Singh BP and Arora N. Molecular and immunological characterization of  
1578 subtilisin like serine protease, a major allergen of *Curvularia lunata*. *Immunobiology*.  
1579 2011;216 3:402-8. doi:10.1016/j.imbio.2010.06.009.
- 1580 117. Woodfolk JA, Wheatley LM, Piyasena RV, Benjamin DC and Platts-Mills TA. Trichophyton  
1581 antigens associated with IgE antibodies and delayed type hypersensitivity. Sequence  
1582 homology to two families of serine proteinases. *J Biol Chem*. 1998;273 45:29489-96.
- 1583 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  
1584 allergens from *Dermatophagoides farinae*. *Mol Med Rep*. 2015;11 5:3559-64.  
1585 doi:10.3892/mmr.2015.3166.
- 1586 119. Tsai LC, Chao PL, Hung MW, Sun YC, Kuo IC, Chua KY, et al. Protein sequence analysis and  
1587 mapping of IgE and IgG epitopes of an allergenic 98-kDa *Dermatophagoides farinae*  
1588 paramyosin, Der f 11. *Allergy*. 2000;55 2:141-7.
- 1589 120. Kim TK, Tirloni L, Pinto AF, Moresco J, Yates JR, 3rd, da Silva Vaz I, Jr., et al. *Ixodes scapularis*  
1590 Tick Saliva Proteins Sequentially Secreted Every 24 h during Blood Feeding. *PLoS Negl Trop*  
1591 *Dis*. 2016;10 1:e0004323. doi:10.1371/journal.pntd.0004323.
- 1592 121. Tirloni L, Kim TK, Pinto AFM, Yates JR, 3rd, da Silva Vaz I, Jr. and Mulenga A. Tick-Host Range  
1593 Adaptation: Changes in Protein Profiles in Unfed Adult *Ixodes scapularis* and *Amblyomma*  
1594 *americanum* Saliva Stimulated to Feed on Different Hosts. *Front Cell Infect Microbiol*.  
1595 2017;7:517. doi:10.3389/fcimb.2017.00517.
- 1596 122. Radulovic ZM, Kim TK, Porter LM, Sze SH, Lewis L and Mulenga A. A 24-48 h fed *Amblyomma*  
1597 *americanum* tick saliva immuno-proteome. *BMC Genomics*. 2014;15:518.  
1598 doi:10.1186/1471-2164-15-518.
- 1599 123. Esteves E, Maruyama SR, Kawahara R, Fujita A, Martins LA, Righi AA, et al. Analysis of the  
1600 Salivary Gland Transcriptome of Unfed and Partially Fed *Amblyomma sculptum* Ticks and  
1601 Descriptive Proteome of the Saliva. *Front Cell Infect Microbiol*. 2017;7:476.  
1602 doi:10.3389/fcimb.2017.00476.
- 1603 124. Tan AW, Francischetti IM, Slovak M, Kini RM and Ribeiro JM. Sexual differences in the  
1604 sialomes of the zebra tick, *Rhipicephalus pulchellus*. *J Proteomics*. 2015;117:120-44.  
1605 doi:10.1016/j.jprot.2014.12.014.

- 1606 125. Garcia GR, Gardinassi LG, Ribeiro JM, Anatriello E, Ferreira BR, Moreira HN, et al. The  
1607 sialotranscriptome of *Amblyomma triste*, *Amblyomma parvum* and *Amblyomma*  
1608 *cajennense* ticks, uncovered by 454-based RNA-seq. *Parasit Vectors*. 2014;7:430.  
1609 doi:10.1186/1756-3305-7-430.
- 1610 126. Jonckheere W, Dermauw W, Zhurov V, Wybouw N, Van den Bulcke J, Villarroel CA, et al.  
1611 The Salivary Protein Repertoire of the Polyphagous Spider Mite *Tetranychus urticae*: A  
1612 Quest for Effectors. *Mol Cell Proteomics*. 2016;15 12:3594-613.  
1613 doi:10.1074/mcp.M116.058081.
- 1614 127. Hunter DJ, Torkelson JL, Bodnar J, Mortazavi B, Laurent T, Deason J, et al. The Rickettsia  
1615 Endosymbiont of *Ixodes pacificus* Contains All the Genes of De Novo Folate Biosynthesis.  
1616 *Plos One*. 2015;10 12:e0144552. doi:10.1371/journal.pone.0144552.
- 1617 128. Rio RV, Attardo GM and Weiss BL. Grandeur Alliances: Symbiont Metabolic Integration and  
1618 Obligate Arthropod Hematophagy. *Trends Parasitol*. 2016;32 9:739-49.  
1619 doi:10.1016/j.pt.2016.05.002.
- 1620 129. Temeyer KB and Tuckow AP. Tick Salivary Cholinesterase: A Probable Immunomodulator of  
1621 Host-parasite Interactions. *J Med Entomol*. 2016;53 3:500-4. doi:10.1093/jme/tjv252.
- 1622 130. Junger WG. Immune cell regulation by autocrine purinergic signalling. *Nat Rev Immunol*.  
1623 2011;11 3:201-12. doi:10.1038/nri2938.
- 1624 131. Pichu S, Yalcin EB, Ribeiro JM, King RS and Mather TN. Molecular characterization of novel  
1625 sulfotransferases from the tick, *Ixodes scapularis*. *BMC Biochem*. 2011;12:32.  
1626 doi:10.1186/1471-2091-12-32.
- 1627 132. Thompson RE, Liu X, Ripoll-Rozada J, Alonso-Garcia N, Parker BL, Pereira PJB, et al. Tyrosine  
1628 sulfation modulates activity of tick-derived thrombin inhibitors. *Nat Chem*. 2017;9 9:909-  
1629 17. doi:10.1038/nchem.2744.
- 1630 133. Blisnick AA, Foulon T and Bonnet SI. Serine Protease Inhibitors in Ticks: An Overview of  
1631 Their Role in Tick Biology and Tick-Borne Pathogen Transmission. *Front Cell Infect*  
1632 *Microbiol*. 2017;7:199. doi:10.3389/fcimb.2017.00199.
- 1633 134. Fogaca AC, Almeida IC, Eberlin MN, Tanaka AS, Bulet P and Daffre S. Ixodidin, a novel  
1634 antimicrobial peptide from the hemocytes of the cattle tick *Boophilus microplus* with  
1635 inhibitory activity against serine proteinases. *Peptides*. 2006;27 4:667-74.  
1636 doi:10.1016/j.peptides.2005.07.013.
- 1637 135. Sasaki SD, de Lima CA, Lovato DV, Juliano MA, Torquato RJ and Tanaka AS. BmSI-7, a novel  
1638 subtilisin inhibitor from *Boophilus microplus*, with activity toward Pr1 proteases from the  
1639 fungus *Metarhizium anisopliae*. *Exp Parasitol*. 2008;118 2:214-20.  
1640 doi:10.1016/j.exppara.2007.08.003.
- 1641 136. Rapoport TA. Protein transport across the endoplasmic reticulum membrane: facts,  
1642 models, mysteries. *Faseb J*. 1991;5 13:2792-8.
- 1643 137. Bullard R, Allen P, Chao CC, Douglas J, Das P, Morgan SE, et al. Structural characterization  
1644 of tick cement cones collected from in vivo and artificial membrane blood-fed Lone Star  
1645 ticks (*Amblyomma americanum*). *Ticks Tick Borne Dis*. 2016;7 5:880-92.  
1646 doi:10.1016/j.ttbdis.2016.04.006.
- 1647 138. Kim TK, Curran J and Mulenga A. Dual silencing of long and short *Amblyomma americanum*  
1648 acidic chitinase forms weakens the tick cement cone stability. *J Exp Biol*. 2014;217 Pt  
1649 19:3493-503. doi:10.1242/jeb.107979.
- 1650 139. Kim JH, Roh JY, Kwon DH, Kim YH, Yoon KA, Yoo S, et al. Estimation of the genome sizes of  
1651 the chigger mites *Leptotrombidium pallidum* and *Leptotrombidium scutellare* based on  
1652 quantitative PCR and k-mer analysis. *Parasit Vectors*. 2014;7:279. doi:10.1186/1756-3305-  
1653 7-279.
- 1654 140. Niedzwiedzki G, Szrek P, Narkiewicz K, Narkiewicz M and Ahlberg PE. Tetrapod trackways  
1655 from the early Middle Devonian period of Poland. *Nature*. 2010;463 7277:43-8.  
1656 doi:10.1038/nature08623.

- 1657 141. Konikiewicz M and Makol J. A fossil Paratrombiinae mite (Actinotrichida: Trombidoidea)  
1658 from the Rovno amber, Ukraine. Zootaxa. 2014;3847 4:583-9.  
1659 doi:10.11646/zootaxa.3847.4.8.
- 1660 142. Karahadian C, Josephson DB and Lindsay RC. Volatile Compounds from *Penicillium* Sp  
1661 Contributing Musty Earthy Notes to Brie and Camembert Cheese Flavors. J Agr Food Chem.  
1662 1985;33 3:339-43.
- 1663 143. Gerber NN. Volatile substances from actinomycetes: their role in the odor pollution of  
1664 water. CRC Crit Rev Microbiol. 1979;7 3:191-214.
- 1665 144. Gerber NN and Lechevalier HA. Geosmin, an earthy-smelling substance isolated from  
1666 actinomycetes. Appl Microbiol. 1965;13 6:935-8.
- 1667 145. Tabachek JAL and Yurkowski M. Isolation and Identification of Blue-Green-Algae Producing  
1668 Muddy Odor Metabolites, Geosmin, and 2-Methylisoborneol, in Saline Lakes in Manitoba.  
1669 J Fish Res Board Can. 1976;33 1:25-35.
- 1670 146. Stensmyr MC, Dweck HK, Farhan A, Ibba I, Strutz A, Mukunda L, et al. A conserved dedicated  
1671 olfactory circuit for detecting harmful microbes in *Drosophila*. Cell. 2012;151 6:1345-57.  
1672 doi:10.1016/j.cell.2012.09.046.
- 1673 147. Ashitani T, Garbouli SS, Schubert F, Vongsombath C, Liblikas I, Palsson K, et al. Activity  
1674 studies of sesquiterpene oxides and sulfides from the plant *Hyptis suaveolens* (Lamiaceae)  
1675 and its repellency on *Ixodes ricinus* (Acari: Ixodidae). Exp Appl Acarol. 2015;67 4:595-606.  
1676 doi:10.1007/s10493-015-9965-5.
- 1677 148. Birkett MA, Abassi SA, Krober T, Chamberlain K, Hooper AM, Guerin PM, et al.  
1678 Antiectoparasitic activity of the gum resin, gum haggard, from the East African plant,  
1679 *Commiphora holtziana*. Phytochemistry. 2008;69 8:1710-5.  
1680 doi:10.1016/j.phytochem.2008.02.017.
- 1681 149. Baldin ELL, Souza ES, Silva JPGF, Pavarini DP, Lopes NP, Lopes JLC, et al. TOWARDS NEW  
1682 BOTANICAL PESTICIDES: THE TOXIC EFFECT OF *Eremanthus goyazensis* (Asteraceae) LEAVES  
1683 ESSENTIAL OIL AGAINST *Brevipalpus phoenicis* (Acari: Tenuipalpidae). Quim Nova. 2012;35  
1684 11:2254-7.
- 1685 150. Huang RL, Li ZH, Wang SY, Fu JT, Cheng DM and Zhang ZX. Insecticidal effect of volatile  
1686 compounds from plant materials of *Murraya exotica* against Red Imported Fire Ant  
1687 Workers. Sociobiology. 2016;63 2:783-91.
- 1688 151. Cloudsley-Thompson JL. Some aspects of the physiology and behaviour of *Dinothrombium*  
1689 (Acari). Entomologia Experimentalis et Applicata. 1962;5:69-73.
- 1690 152. Sferra NJ. 1st Record of *Pterodontia-Flavipes* (Diptera, Acroceridae) Larvae in the Mites  
1691 *Podothrombium* (Acari, Trombidiidae) and *Abrolophus* (Acari, Erythraeidae). Entomol  
1692 News. 1986;97 3:121-3.
- 1693 153. Ozawa R, Shimoda T, Kawaguchi M, Arimura G, Horiuchi J, Nishioka T, et al. *Lotus japonicus*  
1694 infested with herbivorous mites emits volatile compounds that attract predatory mites. J  
1695 Plant Res. 2000;113 1112:427-33.
- 1696 154. Royalty RN, Phelan PL and Hall FR. Arrestment of Male 2-Spotted Spider-Mite Caused by  
1697 Female Sex-Pheromone. J Chem Ecol. 1992;18 2:137-53.
- 1698 155. Wohltmann A. The evolution of life histories in Parasitengona (Acari: Prostigmata).  
1699 Acarologia. 2000;41 1-2:145-204.
- 1700 156. Moniuszko H and Makol J. Host-parasite association in trombiculid mites (Actinotrichida:  
1701 Trombiculidae) of temperate zone - the case of *Hirsutiella zachvatkini* (Schluger, 1948); are  
1702 we dealing with prolonged contact with the host? Parasit Vectors. 2016;9:61.  
1703 doi:10.1186/s13071-016-1339-2.
- 1704 157. Khan AA and Quigley JG. Heme and FLVCR-related transporter families SLC48 and SLC49.  
1705 Mol Aspects Med. 2013;34 2-3:669-82. doi:10.1016/j.mam.2012.07.013.
- 1706 158. Mandilaras K and Missirlis F. Genes for iron metabolism influence circadian rhythms in  
1707 *Drosophila melanogaster*. Metallomics. 2012;4 9:928-36. doi:10.1039/c2mt20065a.

- 1708 159. Alves-Bezerra M, Klett EL, De Paula IF, Ramos IB, Coleman RA and Gondim KC. Long-chain  
1709 acyl-CoA synthetase 2 knockdown leads to decreased fatty acid oxidation in fat body and  
1710 reduced reproductive capacity in the insect *Rhodnius prolixus*. *Biochim Biophys Acta*.  
1711 2016;1861 7:650-62. doi:10.1016/j.bbali.2016.04.007.
- 1712 160. Bryon A, Kurlovs AH, Van Leeuwen T and Clark RM. A molecular-genetic understanding of  
1713 diapause in spider mites: current knowledge and future directions. *Physiol Entomol*.  
1714 2017;42 3:211-24.
- 1715 161. Royalty RN, Phelan PL and Hall FR. Comparative Effects of Form, Color, and Pheromone of  
1716 2-Spotted Spider-Mite Quiescent Deutonymphs on Male Guarding Behavior. *Physiol*  
1717 *Entomol*. 1993;18 3:303-16.
- 1718 162. Goto SG. Physiological and molecular mechanisms underlying photoperiodism in the spider  
1719 mite: comparisons with insects. *J Comp Physiol B*. 2016;186 8:969-84. doi:10.1007/s00360-  
1720 016-1018-9.
- 1721 163. Hori Y, Numata H, Shiga S and Goto SG. Both the anterior and posterior eyes function as  
1722 photoreceptors for photoperiodic termination of diapause in the two-spotted spider mite.  
1723 *J Comp Physiol A Neuroethol Sens Neural Behav Physiol*. 2014;200 2:161-7.  
1724 doi:10.1007/s00359-013-0872-0.
- 1725 164. Terakita A and Nagata T. Functional properties of opsins and their contribution to light-  
1726 sensing physiology. *Zoolog Sci*. 2014;31 10:653-9. doi:10.2108/zs140094.
- 1727 165. Ni JD, Baik LS, Holmes TC and Montell C. A rhodopsin in the brain functions in circadian  
1728 photoentrainment in *Drosophila*. *Nature*. 2017;545 7654:340-4. doi:10.1038/nature22325.
- 1729 166. Tanji T, Ohashi-Kobayashi A and Natori S. Participation of a galactose-specific C-type lectin  
1730 in *Drosophila* immunity. *Biochem J*. 2006;396 1:127-38. doi:10.1042/BJ20051921.
- 1731 167. Costa FH, Valenca NS, Silva AR, Bezerra GA, Cavada BS and Radis-Baptista G. Cloning and  
1732 molecular modeling of *Litopenaeus vannamei* (Penaeidae) C-type lectin homologs with  
1733 mutated mannose binding domain-2. *Genet Mol Res*. 2011;10 2:650-64.  
1734 doi:10.4238/vol10-2gmr999.
- 1735 168. Nakamura T, Furuhashi M, Li P, Cao H, Tuncman G, Sonenberg N, et al. Double-stranded  
1736 RNA-dependent protein kinase links pathogen sensing with stress and metabolic  
1737 homeostasis. *Cell*. 2010;140 3:338-48. doi:10.1016/j.cell.2010.01.001.
- 1738 169. Pena J and Harris E. Dengue virus modulates the unfolded protein response in a time-  
1739 dependent manner. *J Biol Chem*. 2011;286 16:14226-36. doi:10.1074/jbc.M111.222703.
- 1740 170. Walsh AM, Kortschak RD, Gardner MG, Bertozzi T and Adelson DL. Widespread horizontal  
1741 transfer of retrotransposons. *Proc Natl Acad Sci U S A*. 2013;110 3:1012-6.  
1742 doi:10.1073/pnas.1205856110.
- 1743 171. Santos-Matos G, Wybouw N, Martins NE, Zele F, Riga M, Leitao AB, et al. *Tetranychus*  
1744 *urticae* mites do not mount an induced immune response against bacteria. *Proc Biol Sci*.  
1745 2017;284 1856 doi:10.1098/rspb.2017.0401.
- 1746 172. Yang L and Zhu R. Immunotherapy of house dust mite allergy. *Hum Vaccin Immunother*.  
1747 2017;13 10:2390-6. doi:10.1080/21645515.2017.1364823.
- 1748 173. de la Fuente J and Contreras M. Tick vaccines: current status and future directions. *Expert*  
1749 *Rev Vaccines*. 2015;14 10:1367-76. doi:10.1586/14760584.2015.1076339.
- 1750 174. Burgess ST, Nunn F, Nath M, Frew D, Wells B, Marr EJ, et al. A recombinant subunit vaccine  
1751 for the control of ovine psoroptic mange (sheep scab). *Vet Res*. 2016;47:26.  
1752 doi:10.1186/s13567-016-0315-3.
- 1753 175. Bartley K, Wright HW, Huntley JF, Manson ED, Inglis NF, McLean K, et al. Identification and  
1754 evaluation of vaccine candidate antigens from the poultry red mite (*Dermanyssus gallinae*).  
1755 *Int J Parasitol*. 2015;45 13:819-30. doi:10.1016/j.ijpara.2015.07.004.
- 1756 176. Wongprompitak P, Duong V, Anukool W, Sreyrath L, Mai TT, Gavotte L, et al. *Orientia*  
1757 *tsutsugamushi*, agent of scrub typhus, displays a single metapopulation with maintenance

- 1758 of ancestral haplotypes throughout continental South East Asia. *Infect Genet Evol.*  
1759 2015;31:1-8. doi:10.1016/j.meegid.2015.01.005.
- 1760 177. Auffray JC, Blasdel, K.R., Bordes, F., Chabé, M., Chaisiri, K., Charbonnel, N., Claude, J., Dei-  
1761 Cas, E., Desquesnes, M., Dobigny, G., Douangboupha, B., Galan, M., Haukialmi, V.,  
1762 Henttonen, H., Hugot, J.P., Jiyipong, T., Latinne, A., Michaux, J., Milocco, C., Morand, S.,  
1763 Pagès, M., Phoophitpong, D., Pumhom, P., Ribas Salvador, A., Soonchan, S.,  
1764 Suputtamongkol, Y., Waengsothorn, S., Waywa, D., Xuereb, A. *Protocols for field and*  
1765 *laboratory rodent studies.* Bangkok: Kasetsart University Press; 2011.
- 1766 178. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.  
1767 *EMBnetjournal.* 2011;17 1:10-2.
- 1768 179. Joshi NA, Fass, J.N.: Sickle: A sliding-window, adaptive, quality-based trimming tool for  
1769 FastQ files. <https://github.com/najoshi/sickle> (2011). Accessed 24/01/18.
- 1770 180. Wood DE and Salzberg SL. Kraken: ultrafast metagenomic sequence classification using  
1771 exact alignments. *Genome Biol.* 2014;15 3:R46. doi:10.1186/gb-2014-15-3-r46.
- 1772 181. Marcais G and Kingsford C. A fast, lock-free approach for efficient parallel counting of  
1773 occurrences of k-mers. *Bioinformatics.* 2011;27 6:764-70.  
1774 doi:10.1093/bioinformatics/btr011.
- 1775 182. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al.  
1776 *GenomeScope: fast reference-free genome profiling from short reads.* *Bioinformatics.*  
1777 2017;33 14:2202-4. doi:10.1093/bioinformatics/btx153.
- 1778 183. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft  
1779 assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad*  
1780 *Sci U S A.* 2011;108 4:1513-8. doi:10.1073/pnas.1017351108.
- 1781 184. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved  
1782 memory-efficient short-read de novo assembler. *Gigascience.* 2012;1 1:18.  
1783 doi:10.1186/2047-217X-1-18.
- 1784 185. Weisenfeld NI, Yin S, Sharpe T, Lau B, Hegarty R, Holmes L, et al. Comprehensive variation  
1785 discovery in single human genomes. *Nat Genet.* 2014;46 12:1350-5. doi:10.1038/ng.3121.
- 1786 186. Langmead B and Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.*  
1787 2012;9 4:357-9. doi:10.1038/nmeth.1923.
- 1788 187. Parra G, Bradnam K and Korf I. CEGMA: a pipeline to accurately annotate core genes in  
1789 eukaryotic genomes. *Bioinformatics.* 2007;23 9:1061.
- 1790 188. Smit A, Hubley, R.: RepeatModeler: De-novo repeat discovery tool.  
1791 <https://github.com/rmhuley/RepeatModeler> (2017). Accessed 24/01/18.
- 1792 189. Tarailo-Graovac M and Chen N. Using RepeatMasker to identify repetitive elements in  
1793 genomic sequences. *Curr Protoc Bioinformatics.* 2009;Chapter 4:Unit 4 10.  
1794 doi:10.1002/0471250953.bi0410s25.
- 1795 190. Zdobnov EM and Apweiler R. InterProScan--an integration platform for the signature-  
1796 recognition methods in InterPro. *Bioinformatics.* 2001;17 9:847-8.
- 1797 191. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M and Robles M. Blast2GO: a universal  
1798 tool for annotation, visualization and analysis in functional genomics research.  
1799 *Bioinformatics.* 2005;21 18:3674-6. doi:10.1093/bioinformatics/bti610.
- 1800 192. Kanehisa M and Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids*  
1801 *Research.* 2000;27 1:29-34(6).
- 1802 193. Adams M, Celniker S, Holt R, Evans C, Gocayne J, Amanatides P, et al. The genome sequence  
1803 of *Drosophila melanogaster.* *Science.* 2000;287 5461:2185-95.  
1804 doi:10.1126/science.287.5461.2185.
- 1805 194. Weinstock G, Robinson G, Gibbs R, Worley K, Evans J, Maleszka R, et al. Insights into social  
1806 insects from the genome of the honeybee *Apis mellifera.* *Nature.* 2006;443 7114:931-49.  
1807 doi:10.1038/nature05260.

- 1808 195. Gulia-Nuss M, Nuss AB, Meyer JM, Sonenshine DE, Roe RM, Waterhouse RM, et al. Genomic  
1809 insights into the Ixodes scapularis tick vector of Lyme disease. *Nat Commun.* 2016;7:10507.  
1810 doi:10.1038/ncomms10507.
- 1811 196. Stein L, Sternberg P, Durbin R, Thierry-Mieg J and Spieth J. WormBase: network access to  
1812 the genome and biology of *Caenorhabditis elegans*. *Nucleic Acids Research.* 2001;29 1:82-  
1813 6. doi:10.1093/nar/29.1.82.
- 1814 197. Katoh K and Standley DM. MAFFT multiple sequence alignment software version 7:  
1815 improvements in performance and usability. *Mol Biol Evol.* 2013;30 4:772-80.  
1816 doi:10.1093/molbev/rmt010.
- 1817 198. Lassmann T, Frings O and Sonnhammer EL. Kalign2: high-performance multiple alignment  
1818 of protein and nucleotide sequences allowing external features. *Nucleic Acids Res.* 2009;37  
1819 3:858-65. doi:10.1093/nar/gkn1006.
- 1820 199. Darriba D, Taboada GL, Doallo R and Posada D. ProtTest 3: fast selection of best-fit models  
1821 of protein evolution. *Bioinformatics.* 2011;27 8:1164-5.  
1822 doi:10.1093/bioinformatics/btr088.
- 1823 200. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W and Gascuel O. New algorithms  
1824 and methods to estimate maximum-likelihood phylogenies: assessing the performance of  
1825 PhyML 3.0. *Syst Biol.* 2010;59 3:307-21. doi:10.1093/sysbio/syq010.
- 1826 201. Kumar S, Stecher G and Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis  
1827 Version 7.0 for Bigger Datasets. *Molecular Biology & Evolution.* 2016;33 7:1870.
- 1828 202. Dabert M, Witalinski W, Kazmierski A, Olszanowski Z and Dabert J. Molecular phylogeny of  
1829 acariform mites (Acari, Arachnida): strong conflict between phylogenetic signal and long-  
1830 branch attraction artifacts. *Mol Phylogenet Evol.* 2010;56 1:222-41.  
1831 doi:10.1016/j.ympev.2009.12.020.
- 1832 203. Castresana J. Selection of conserved blocks from multiple alignments for their use in  
1833 phylogenetic analysis. *Mol Biol Evol.* 2000;17 4:540-52.
- 1834 204. Ronquist F and Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed  
1835 models. *Bioinformatics.* 2003;19 12:1572-4.
- 1836 205. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24  
1837 8:1586-91. doi:10.1093/molbev/msm088.
- 1838 206. Benton M, Donoghue, PCJ, Asher, RJ. Calibrating and constraining molecular clocks. In:  
1839 Hedges SB, Kumar, S., editor. *The timetree of life.* Oxford: Oxford University Press; 2009. p.  
1840 35 - 86.
- 1841 207. Li L, Stoeckert CJ and Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic  
1842 genomes. *Genome Res.* 2003;13 9:2178-89. doi:10.1101/gr.1224503.
- 1843 208. De Bie T, Cristianini N, Demuth JP and Hahn MW. CAFE: a computational tool for the study  
1844 of gene family evolution. *Bioinformatics.* 2006;22 10:1269-71.  
1845 doi:10.1093/bioinformatics/btl097.
- 1846 209. Bendtsen JD, Jensen LJ, Blom N, Von Heijne G and Brunak S. Feature-based prediction of  
1847 non-classical and leaderless protein secretion. *Protein Eng Des Sel.* 2004;17 4:349-56.  
1848 doi:10.1093/protein/gzh037.
- 1849 210. Krogh A, Larsson B, von Heijne G and Sonnhammer EL. Predicting transmembrane protein  
1850 topology with a hidden Markov model: application to complete genomes. *J Mol Biol.*  
1851 2001;305 3:567-80. doi:10.1006/jmbi.2000.4315.
- 1852 211. Gupta R, Jung, E., Brunak, S.: NetNGlyc 1.0 server.  
1853 <http://www.cbs.dtu.dk/services/NetNGlyc/> (2004). Accessed 25/01/18.
- 1854 212. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017-  
1855 beyond protein family and domain annotations. *Nucleic Acids Res.* 2017;45 D1:D190-D9.  
1856 doi:10.1093/nar/gkw1107.
- 1857 213. Sánchez-Gracia A, Vieira FG, Almeida FC and Rozas J. *Comparative Genomics of the Major*  
1858 *Chemosensory Gene Families in Arthropods.* John Wiley & Sons, Ltd; 2011.

- 1859 214. Nozawa M and Nei M. Evolutionary dynamics of olfactory receptor genes in *Drosophila*  
1860 species. *Proceedings of the National Academy of Sciences of the United States of America*.  
1861 2007;104 17:7122.
- 1862 215. Robertson HM and Wanner KW. The chemoreceptor superfamily in the honey bee, *Apis*  
1863 *mellifera*: Expansion of the odorant, but not gustatory, receptor family. *Genome Research*.  
1864 2006;16 11:1395.
- 1865 216. Robertson HM, Warr CG and Carlson JR. Molecular evolution of the insect chemoreceptor  
1866 gene superfamily in *Drosophila melanogaster*. *Proceedings of the National Academy of*  
1867 *Sciences of the United States of America*. 2003;100 Suppl 2 100 Suppl 2:14537.
- 1868 217. Croset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, et al. Ancient protostome  
1869 origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste  
1870 and olfaction. *Plos Genetics*. 2010;6 8:e1001064.
- 1871 218. Dang HX and Lawrence CB. Allerdicator: fast allergen prediction using text classification  
1872 techniques. *Bioinformatics*. 2014;30 8:1120-8. doi:10.1093/bioinformatics/btu004.
- 1873 219. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length  
1874 transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*.  
1875 2011;29 7:644-52. doi:10.1038/nbt.1883.
- 1876 220. Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein  
1877 families database. *Nucleic Acids Res*. 2014;42 Database issue:D222-30.  
1878 doi:10.1093/nar/gkt1223.
- 1879 221. Heberle H, Meirelles GV, da Silva FR, Telles GP and Minghim R. InteractiVenn: a web-based  
1880 tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics*. 2015;16:169.  
1881 doi:10.1186/s12859-015-0611-3.
- 1882 222. Park SW, Ha NY, Ryu B, Bang JH, Song H, Kim Y, et al. Urbanization of scrub typhus disease  
1883 in South Korea. *PLoS Negl Trop Dis*. 2015;9 5:e0003814. doi:10.1371/journal.pntd.0003814.
- 1884 223. Ternent T, Csordas A, Qi D, Gomez-Baena G, Beynon RJ, Jones AR, et al. How to submit MS  
1885 proteomics data to ProteomeXchange via the PRIDE database. *Proteomics*. 2014;14  
1886 20:2233-41. doi:10.1002/pmic.201400120.
- 1887 224. Armstrong SD, Babayan SA, Lhermitte-Vallarino N, Gray N, Xia D, Martin C, et al.  
1888 Comparative analysis of the secretome from a model filarial nematode (*Litomosoides*  
1889 *sigmodontis*) reveals maximal diversity in gravid female parasites. *Mol Cell Proteomics*.  
1890 2014;13 10:2527-44. doi:10.1074/mcp.M114.038539.
- 1891 225. Benjamini Y and Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful  
1892 Approach to Multiple Testing. *J Roy Stat Soc B Met*. 1995;57 1:289-300.
- 1893 226. Golab GC, Patterson-Kane, E. AVMA Guidelines for the Euthanasia of Animals (2013  
1894 Edition). 2013.0.1 ed. Schaumburg, Illinois: American Veterinary Medical Association; 2013.
- 1895 227. Charbonneau R, Niel, L., Olfert, E., von Keyserlingk, M., Griffin, G. CCAC guidelines on:  
1896 euthanasia of animals used in science. Ottawa: Canadian Council on Animal Care; 2010.
- 1897 228. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: functional  
1898 classification of proteins via subfamily domain architectures. *Nucleic Acids Res*. 2017;45  
1899 D1:D200-D3. doi:10.1093/nar/gkw1129.

1900

Figure 1

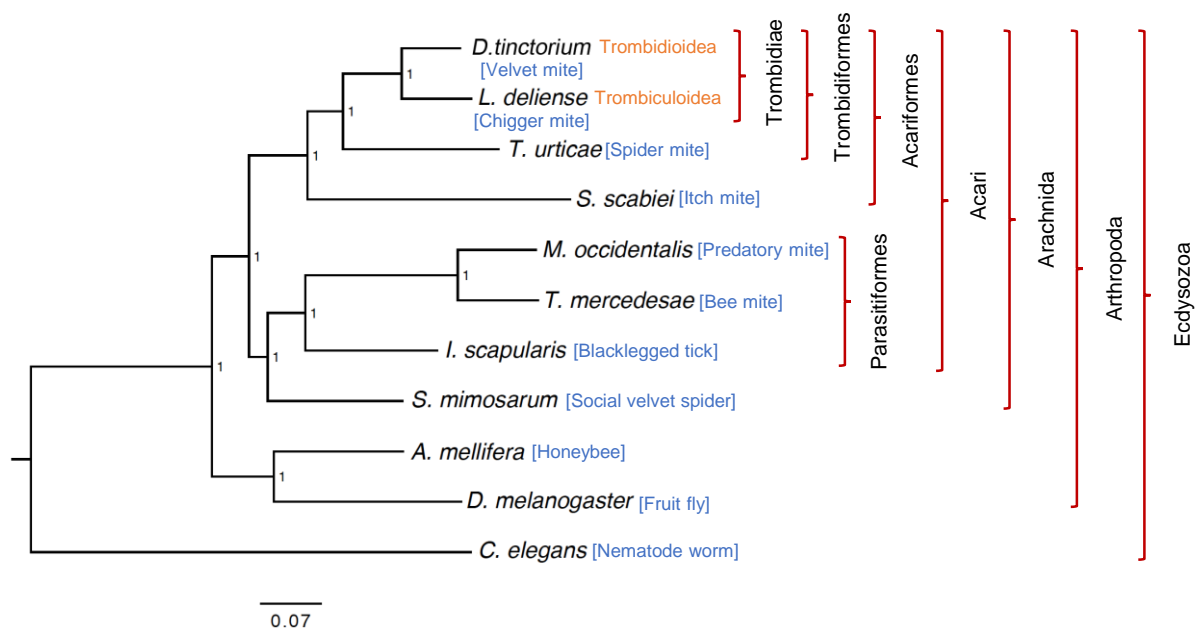
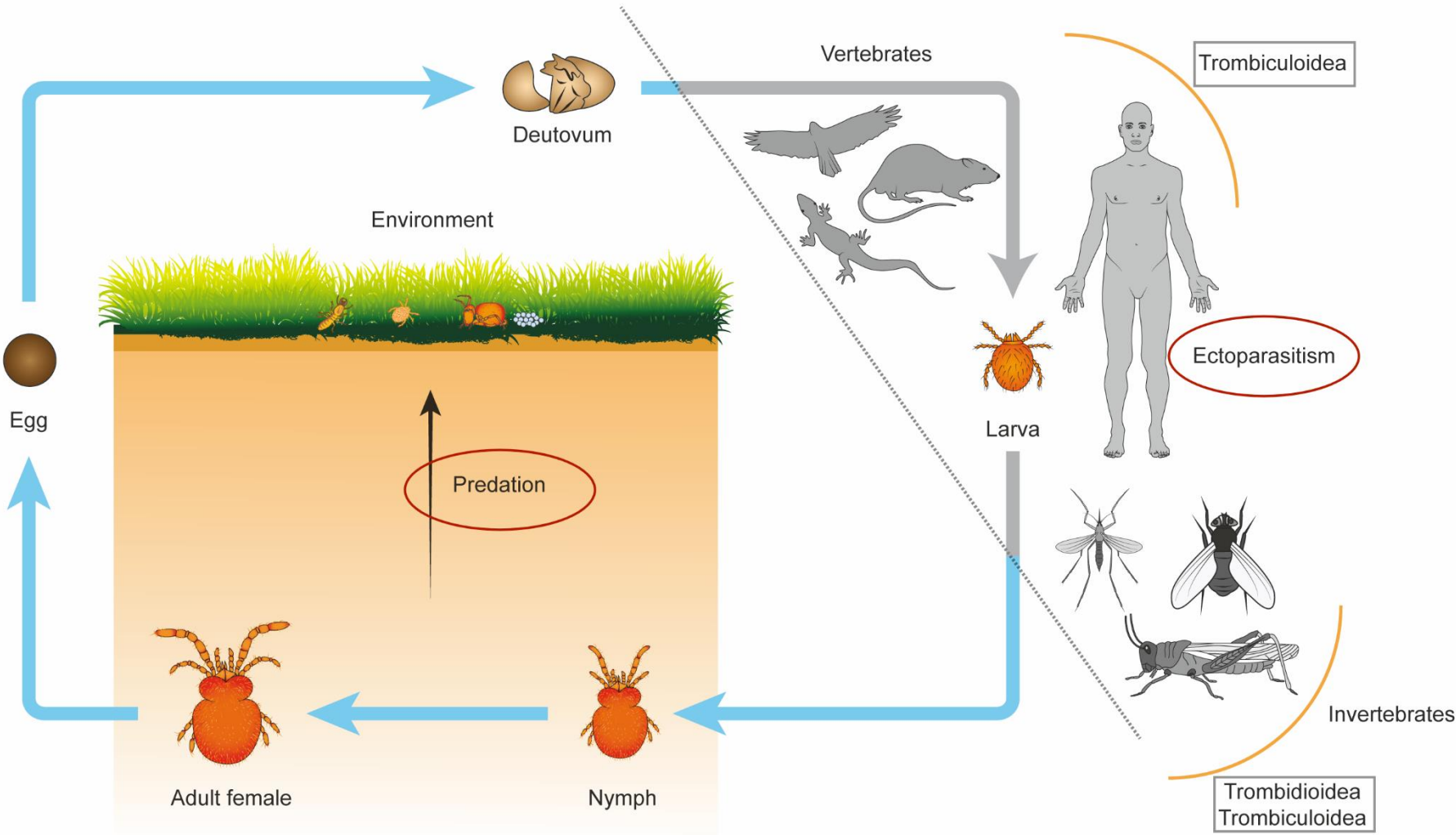
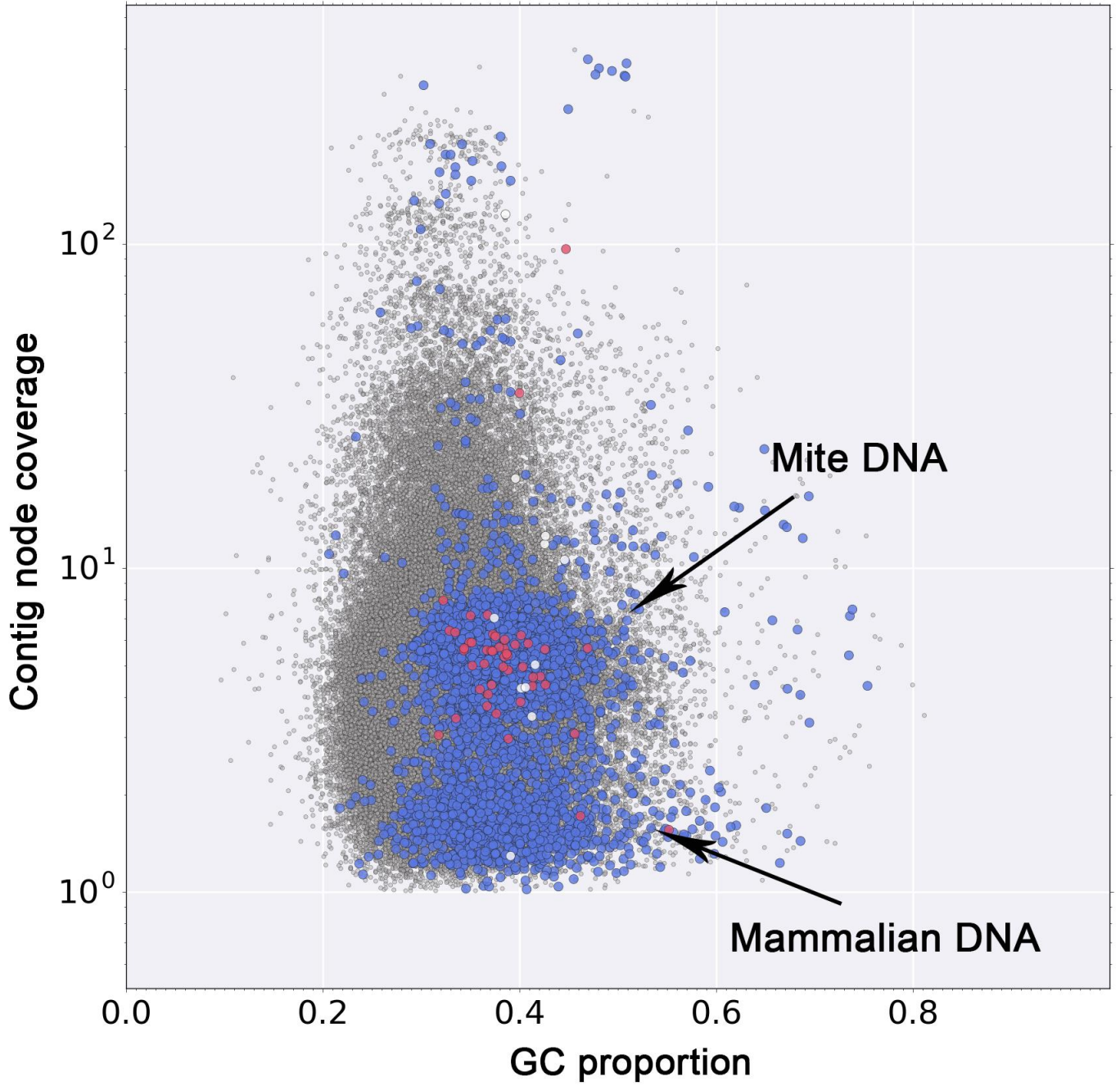
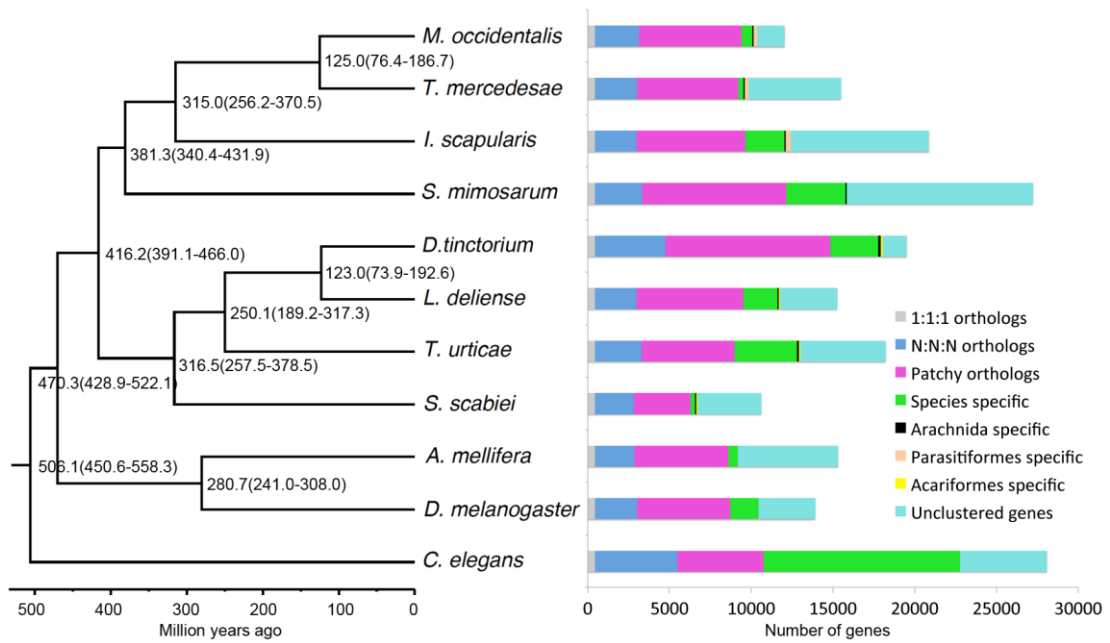




Figure 2







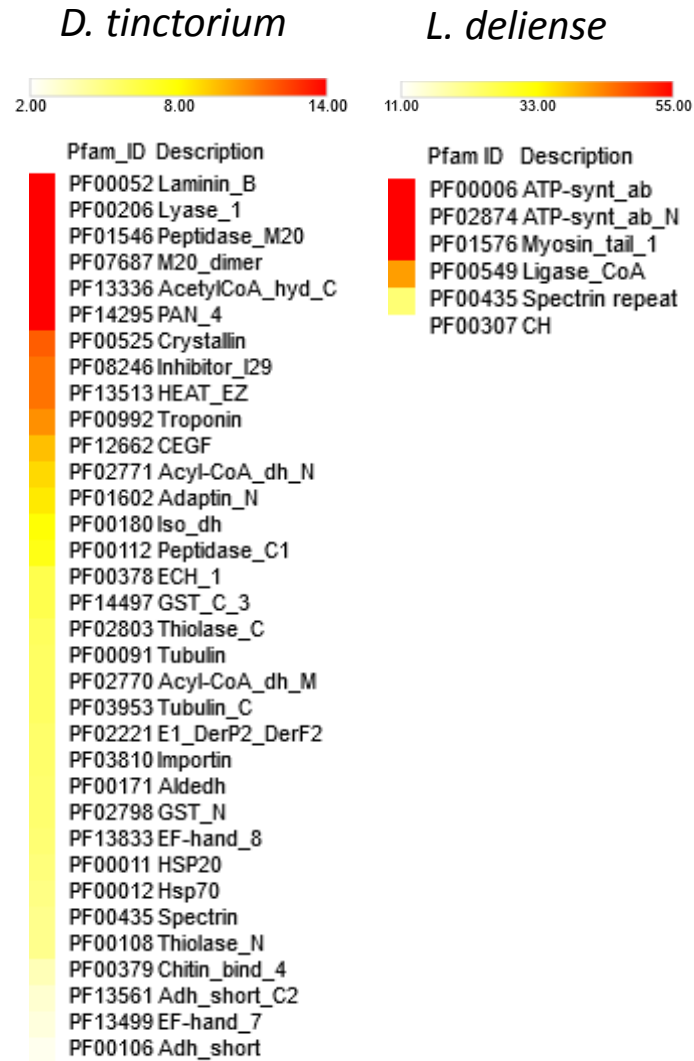


Figure 6

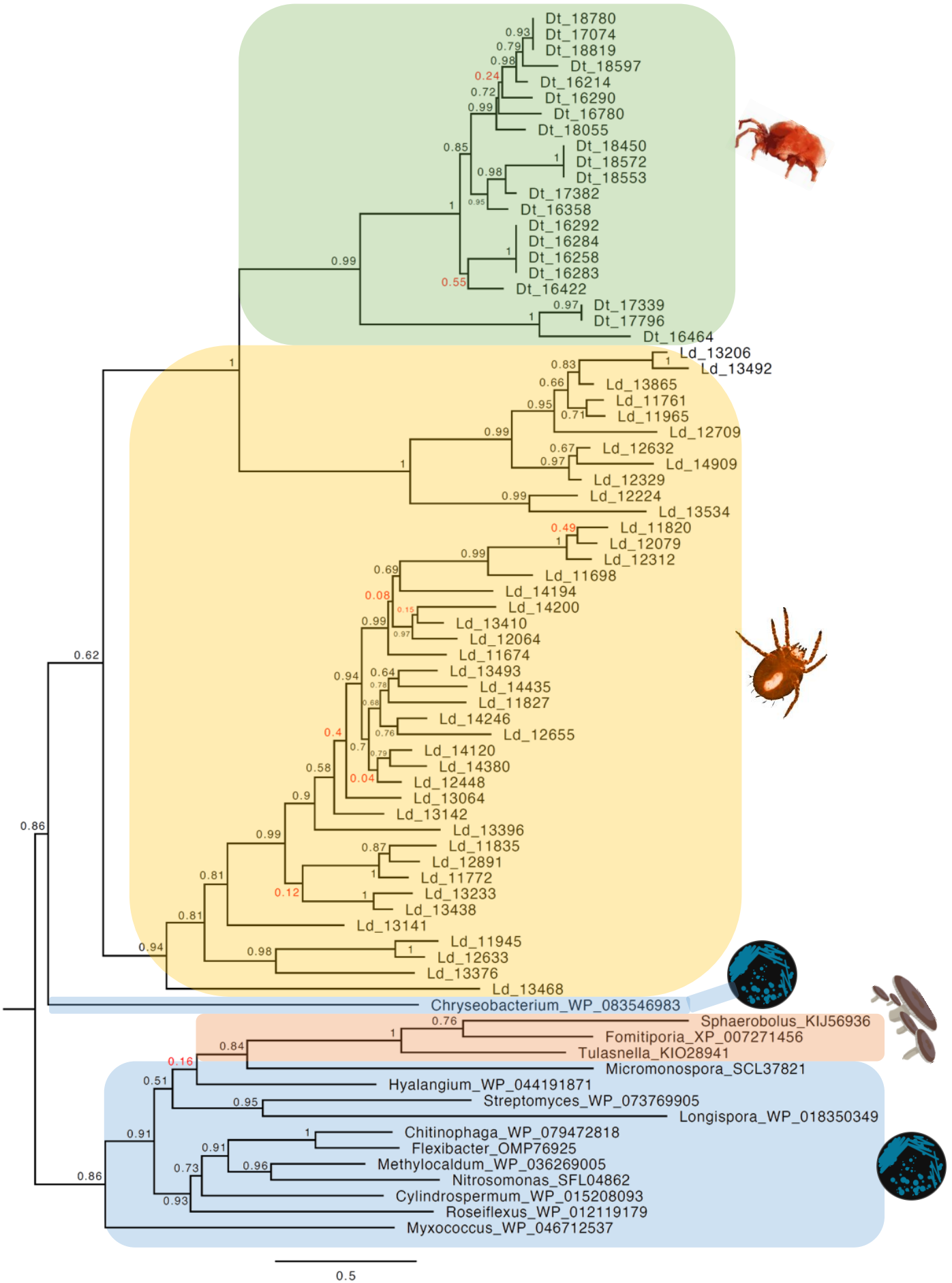
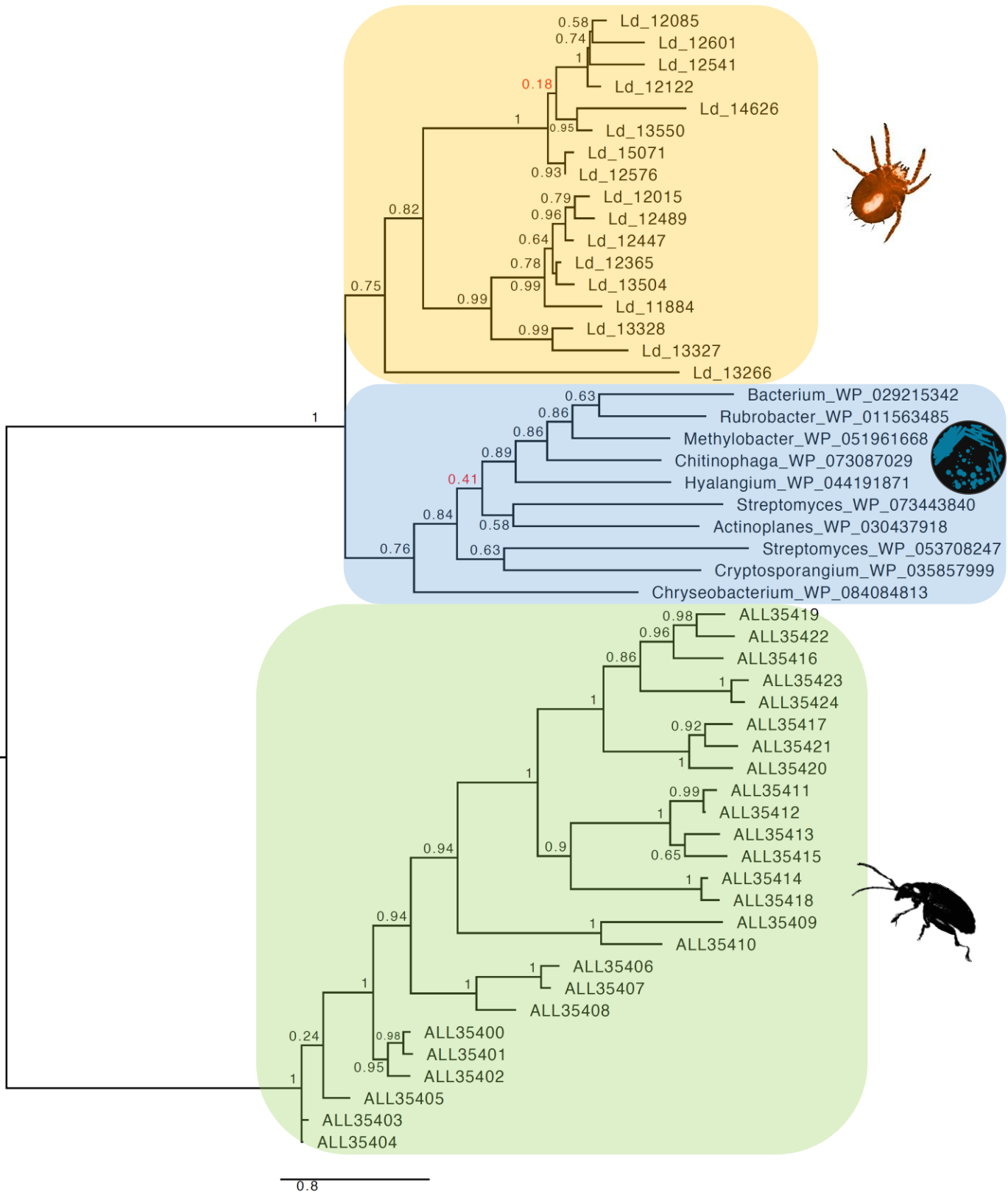
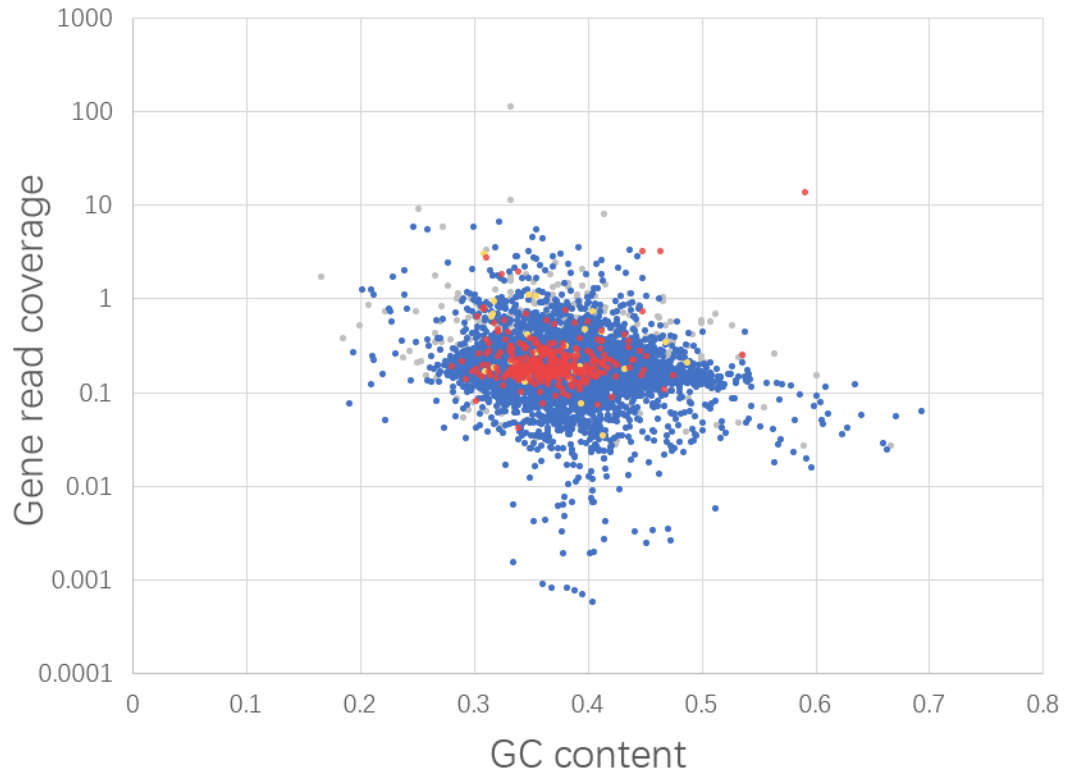


Figure 7

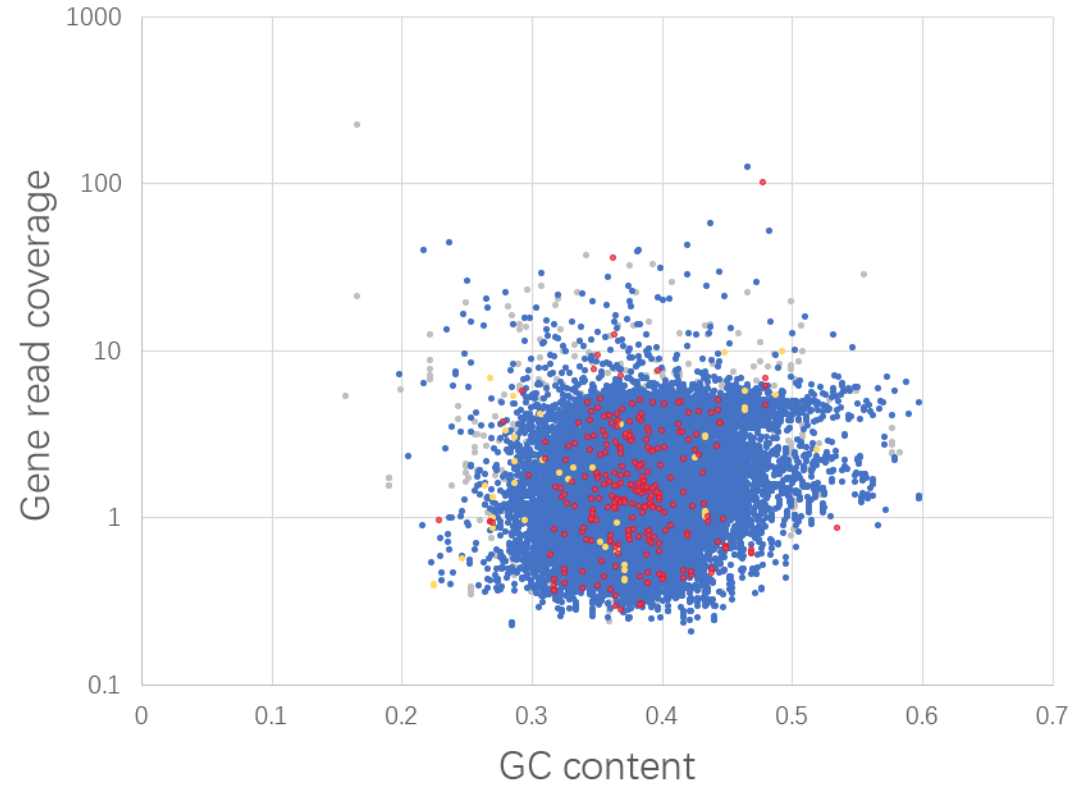


Ld



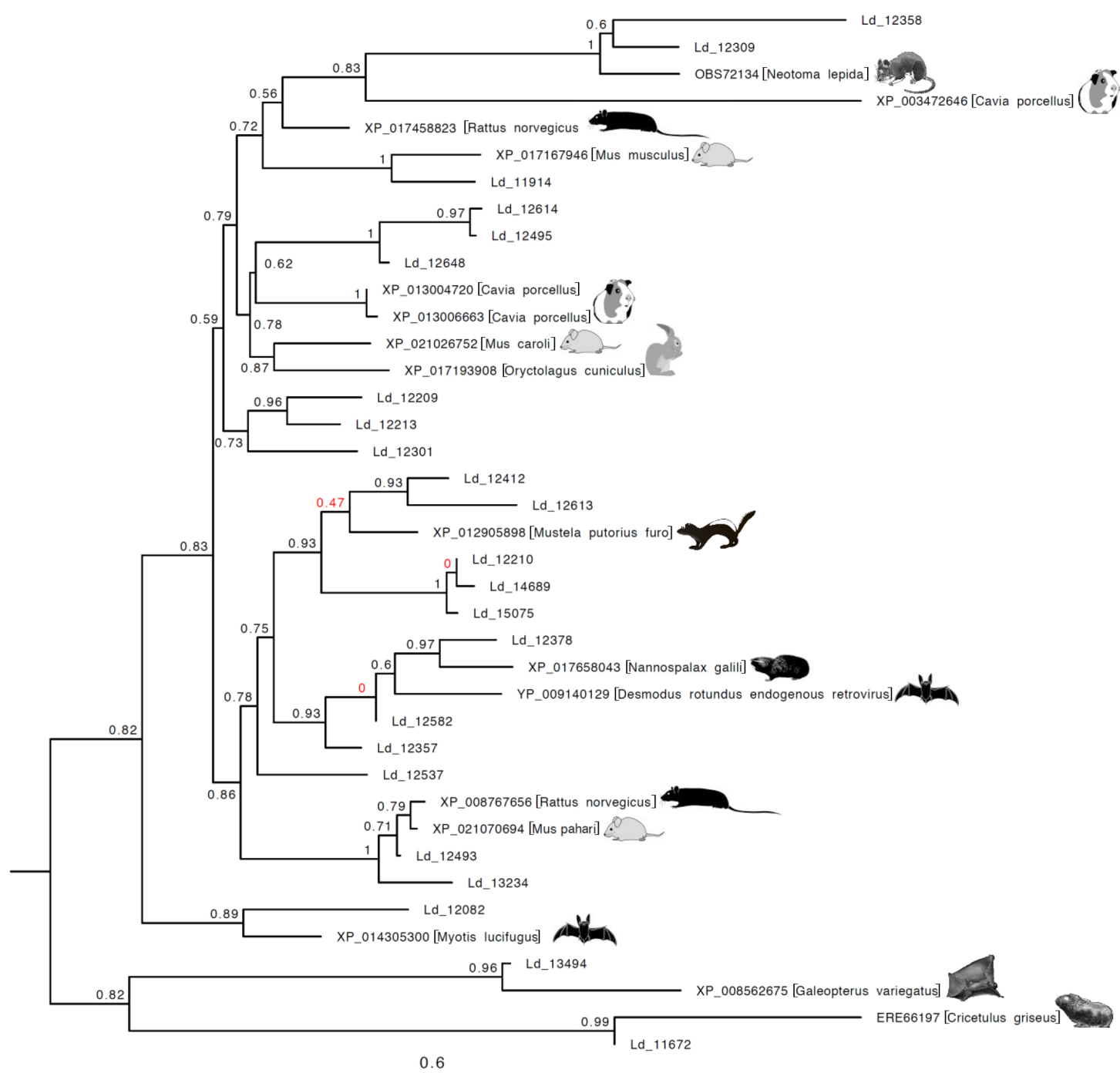
• Not annotated • Eukaryota • Others • Bacteria

Dt

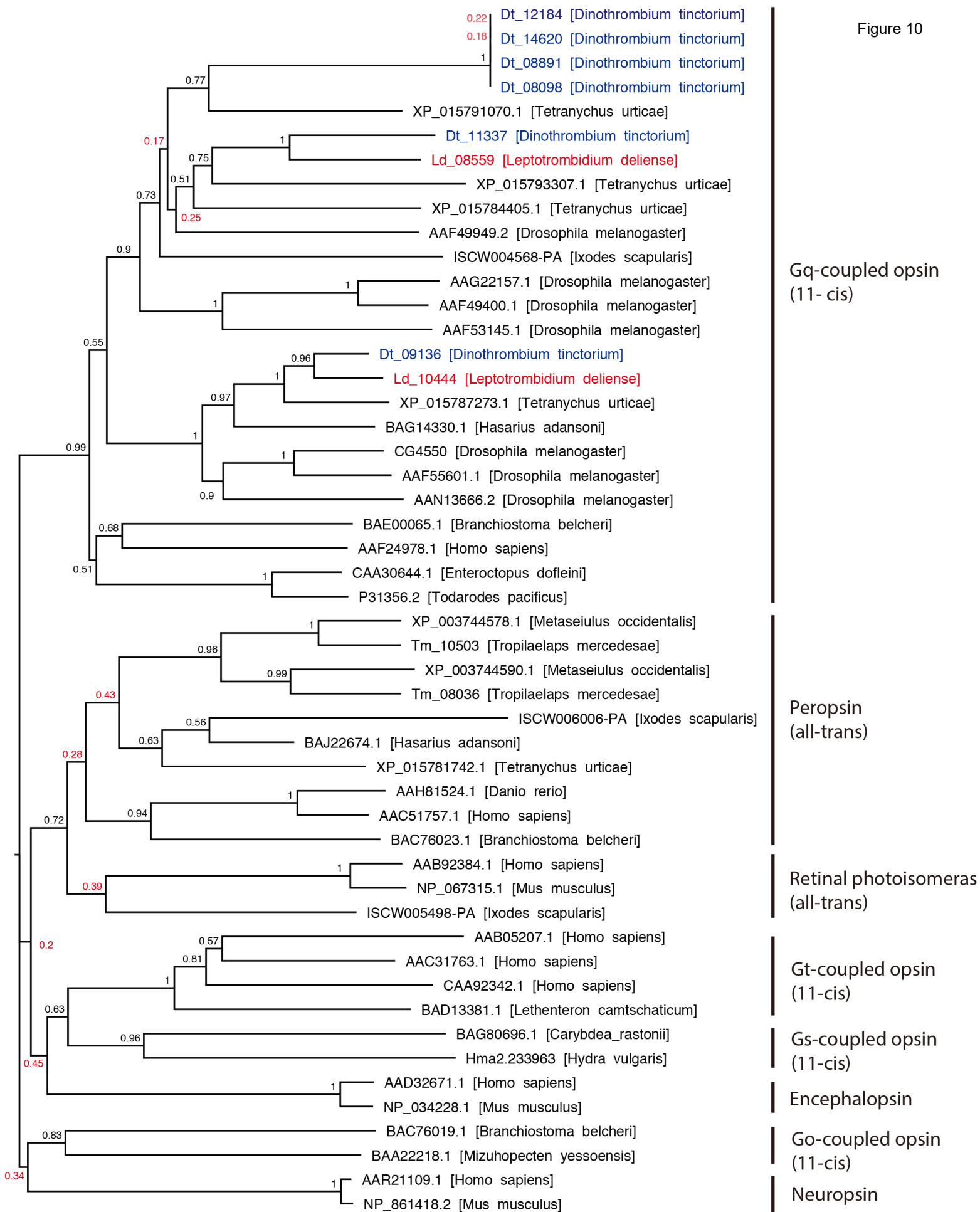


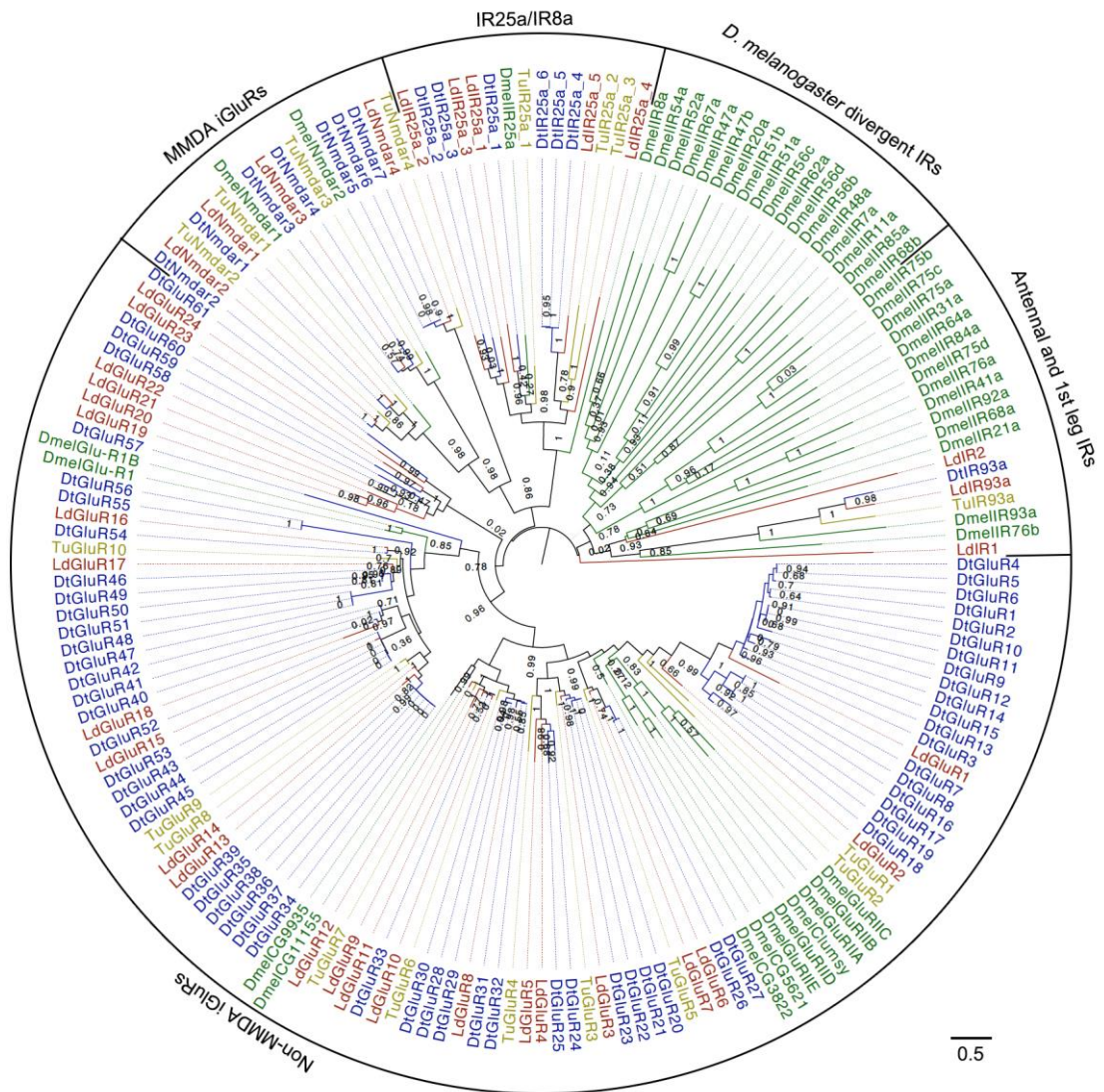
• Not annotated • Eukaryota • Others • Bacteria

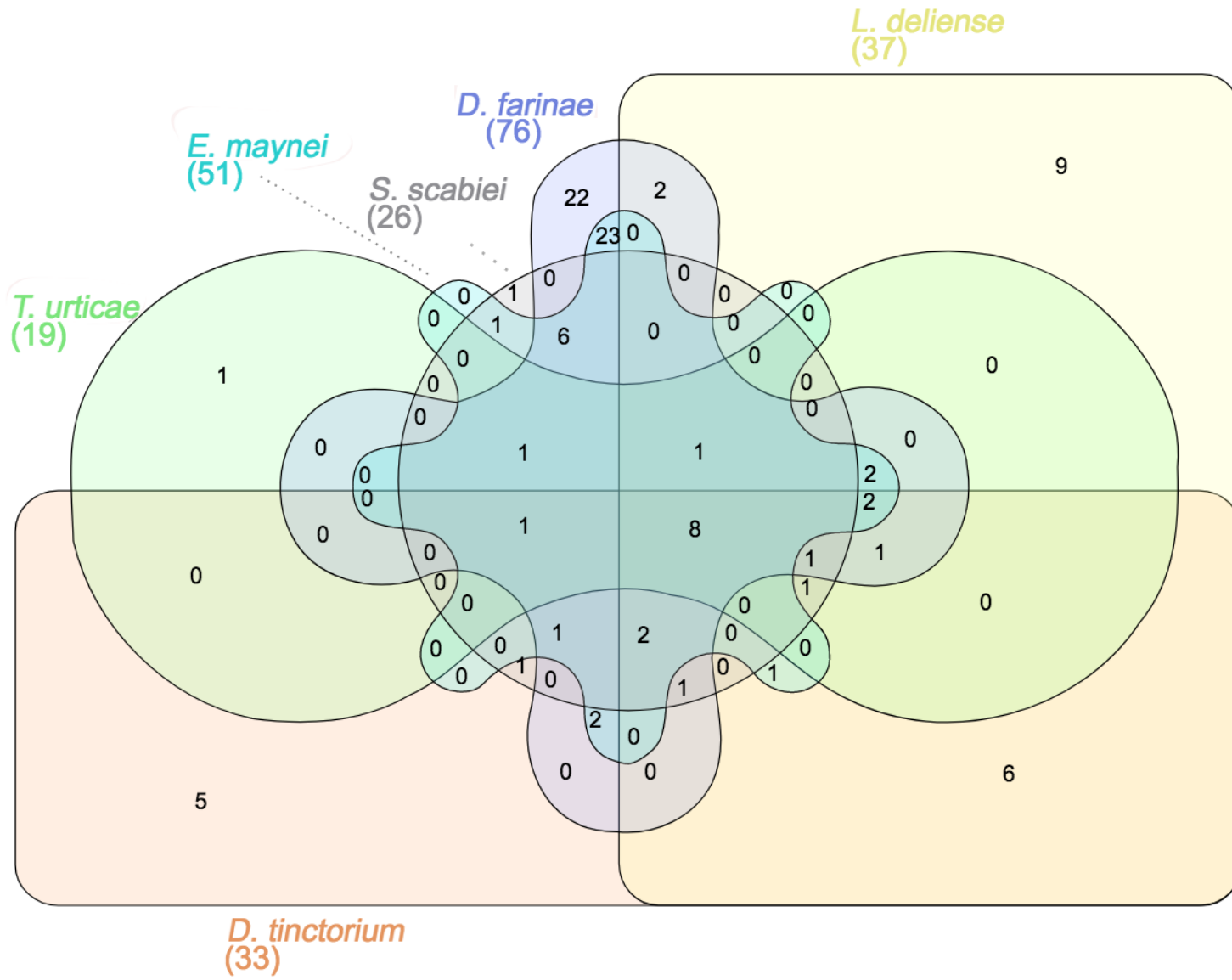
Figure 9











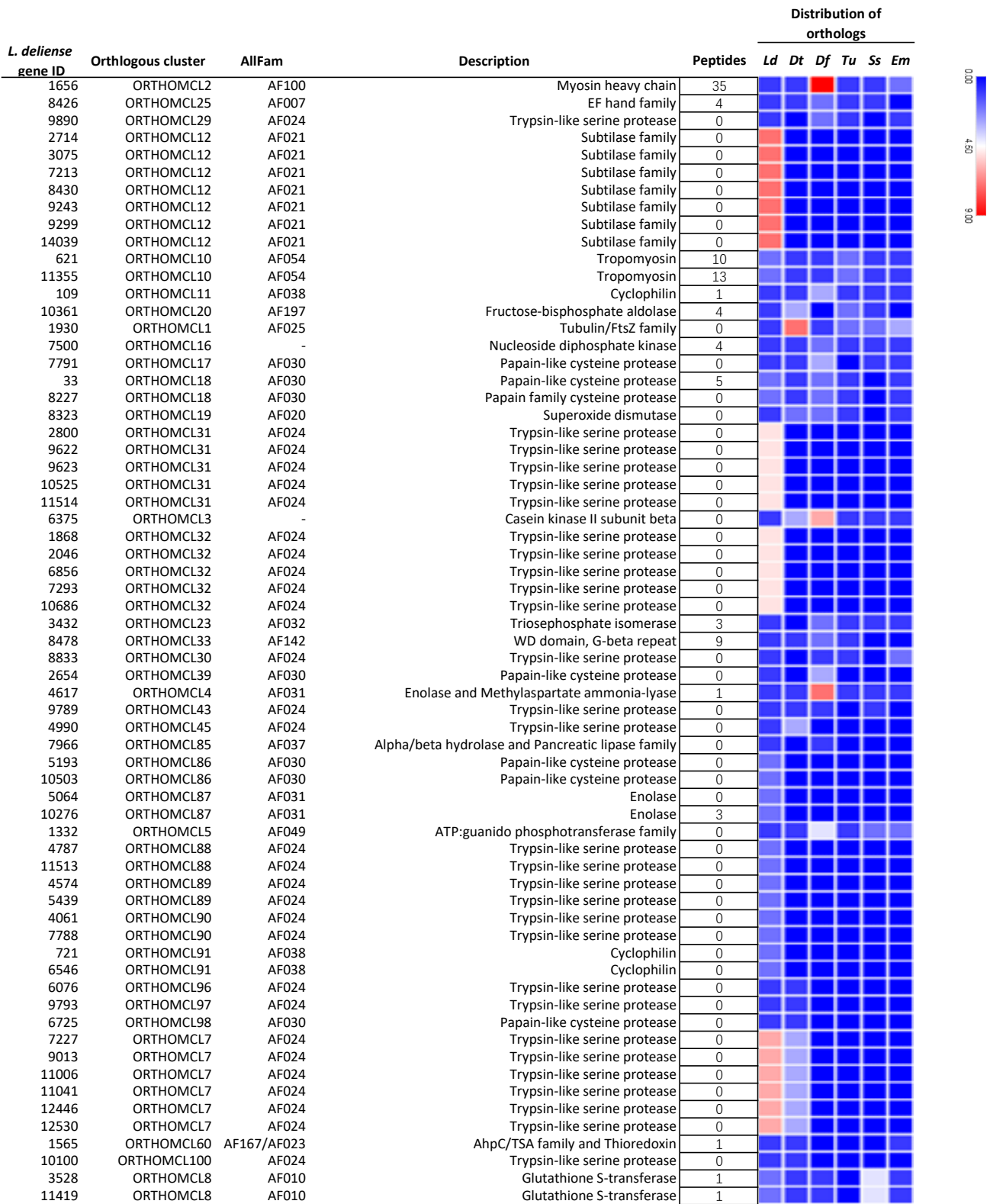


Figure 14

