

Research



Cite this article: Li H, Leavengood Jr JM, Chapman EG, Burkhardt D, Song F, Jiang P, Liu J, Zhou X, Cai W. 2017 Mitochondrial phylogenomics of Hemiptera reveals adaptive innovations driving the diversification of true bugs. *Proc. R. Soc. B* **284**: 20171223. <http://dx.doi.org/10.1098/rspb.2017.1223>

Received: 31 May 2017

Accepted: 25 July 2017

Subject Category:

Evolution

Subject Areas:

evolution, taxonomy and systematics

Keywords:

Hemiptera, phylogeny, evolutionary history, ancestral character state reconstruction, mitochondrial genome

Authors for correspondence:

Xuguo Zhou

e-mail: xuguozhou@uky.edu

Wanzhi Cai

e-mail: caiwz@cau.edu.cn

Electronic supplementary material is available online at <https://dx.doi.org/10.6084/m9.figshare.c.3849793>.

Mitochondrial phylogenomics of Hemiptera reveals adaptive innovations driving the diversification of true bugs

Hu Li^{1,2}, John M. Leavengood Jr², Eric G. Chapman², Daniel Burkhardt³, Fan Song¹, Pei Jiang¹, Jinpeng Liu⁴, Xuguo Zhou² and Wanzhi Cai¹

¹Key Laboratory of Pest Monitoring and Green Management, Ministry of Agriculture, Department of Entomology, China Agricultural University, Beijing 100193, People's Republic of China

²Department of Entomology, University of Kentucky, Lexington, KY 40546, USA

³Naturhistorisches Museum, Augustinergasse 2, 4001 Basel, Switzerland

⁴Markey Cancer Center, University of Kentucky, Lexington, KY 40536, USA

XZ, 0000-0002-2385-8224

Hemiptera, the largest non-holometabolous order of insects, represents approximately 7% of metazoan diversity. With extraordinary life histories and highly specialized morphological adaptations, hemipterans have exploited diverse habitats and food sources through approximately 300 Myr of evolution. To elucidate the phylogeny and evolutionary history of Hemiptera, we carried out the most comprehensive mitogenomics analysis on the richest taxon sampling to date covering all the suborders and infra-orders, including 34 newly sequenced and 94 published mitogenomes. With optimized branch length and sequence heterogeneity, Bayesian analyses using a site-heterogeneous mixture model resolved the higher-level hemipteran phylogeny as (Sternorrhyncha, (Auchenorrhyncha, (Coleorrhyncha, Heteroptera))). Ancestral character state reconstruction and divergence time estimation suggest that the success of true bugs (Heteroptera) is probably due to angiosperm coevolution, but key adaptive innovations (e.g. prognathous mouthpart, predatory behaviour, and haemelytron) facilitated multiple independent shifts among diverse feeding habits and multiple independent colonizations of aquatic habitats.

1. Introduction

Ernst Mayr defined evolutionary novelty as 'any newly acquired structure or property that permits the performance of a new function, which, in turn, will open a new adaptive zone' [1]. Driven by adaptive modifications and the colonization of new ecospace, evolutionary radiations of animals and plants have long been recognized as driving today's biodiversity. Tracking the evolutionary origins of morphological novelty has fascinated biologists for over a century [2]. Even though stochastic factors lead to the development of new lineages, only a fraction of these have successfully diversified over time. Some of the major Metazoan radiations, such as true flies [3] and beetles [4], have been well documented; however, other mega-diverse invertebrate clades have not received the attention they deserve.

With an estimated 97 000–103 590 known species [5,6], Hemiptera represents approximately 7% of metazoan diversity. The biodiversity of Hemiptera includes, but is not limited to, plant lice, cicadas, planthoppers, moss bugs, and true bugs. Heteroptera (true bugs) has evolved diverse life histories and specialized morphological adaptations enabling them to colonize both terrestrial and aquatic habitats, and to exploit various food sources ranging from plants, fungi, small arthropods, and vertebrate blood [7]. Although its monophyly is well supported, in particular, by the synapomorphic segmented, piercing-sucking mouthparts with elaborate food and salivary pumps that permit fluid-feeding specializations

[6] (see the electronic supplementary material, figure S1a–d), the higher-level relationships within Hemiptera have been debated for over two and a half centuries [8–10]. Traditionally, Hemiptera has been categorized into ‘Homoptera’ and Heteroptera, sometimes with ordinal status, based on the presence or absence of a gula [8]. More recently, Hemiptera has been subdivided into four major suborders, Sternorrhyncha (Psylloidea, Aleyrodoidea, Aphidoidea, and Coccoidea) (e.g. electronic supplementary material, figure S1a,b), Auchenorrhyncha (Cicadomorpha and Fulgoromorpha) (e.g. electronic supplementary material, figure S1c), Coleorrhyncha (with the only extant family Peloridiidae), and Heteroptera (seven infraorders) (e.g. electronic supplementary material, figure S1d,e). The sister group relationship between Sternorrhyncha and the remainder of Hemiptera has received strong support from both morphological and molecular evidence [9–12]. However, the monophyly of Auchenorrhyncha has been questioned [10,13,14], and the phylogenetic position of Coleorrhyncha is ambiguous [10,12,14,15]. In addition, relationships among the basal infraorders of Heteroptera are poorly understood [6].

Phylogenetic analysis of Hemiptera based solely on morphology has been challenging. The sedentary lifestyles coupled with phloem-feeding behaviours in some Auchenorrhyncha and especially Sternorrhyncha (behaving as plant parasites) have spurred morphological reductions and losses, neotenus females, extreme sexual dimorphism, and convergently derived morphological characters that would otherwise be useful in phylogenetic analyses [16,17]. The confusion of convergent character states with synapomorphies has contributed to the taxonomic reshufflings of superfamily composition within ‘Homoptera’ [16]. Owing to a large number of morphological features unique to Hemiptera (e.g. the labium forming a sheath for the remaining mouthparts), some of the important characters cannot be readily homologized with structures in the more inclusive groups, resulting in ambiguous or even erroneous ancestral state reconstructions.

Historically, some hemipterists assumed that the ancestor of Hemiptera was phytophagous [18], whereas the ancestor of Heteroptera was considered to be predaceous [19]. The presumed diet of ‘Homoptera’ was intuitive, because the vast majority are plant feeders. The predaceous ancestor of Heteroptera was inferred by the predatory behaviour exhibited by the putative ‘basal’ infraorders, Enicocephalomorpha, Dipsocoromorpha, and Gerromorpha [19]. It is understood that after the Permian-Triassic (P-T) extinction events, many previously exploited niches once again became available for resource partitioning [20]. Heteroptera constitutes approximately 40% of Hemiptera and represents the vast majority of behavioural diversity in terms of diet and habitat. The other three suborders are entirely terrestrial and predominantly phytophagous [7]. Hypotheses of selective forces underlying the diversification of higher-level hemipteran lineages have not yet been substantiated outside of morphology and fossil-based extrapolation [21,22].

With the advent of the Genomics Era, recent analyses have increasingly embraced the molecular resources to advance our understanding of the phylogeny of Hemiptera [10–13,15]. Nevertheless, major issues such as the phylogenetic status (monophyly versus paraphyly) of Auchenorrhyncha and the position of Coleorrhyncha are still unsettled [10,12]. With a recent influx of genomic information, including mitochondrial genomes (mitogenomes), new phylogenetic hypotheses are emerging. Although representing only a subset of the genomic

information (approx. 16 000 nucleotides), mitogenomic data have made substantial contributions to resolve intraordinal relationships in insects [3,23,24].

Despite extensive efforts, previous mitogenomic analyses in Hemiptera did not cover all the suborders and infraorders, and had limited resolution due to the substitution saturation and the compositional heterogeneity of mitogenomes [11,13,15]. Here, we sequenced 34 mitogenomes to complement the existing mitogenomic data derived from 94 hemipteran species. Using a holistic sampling approach, we included the mitogenomes from all four suborders and all seven heteropteran infraorders, covering all four superfamilies of Auchenorrhyncha, three of the four superfamilies of Sternorrhyncha (excluding Coccoidea), the only superfamily of Coleorrhyncha, and 19 of the 23 superfamilies of Heteroptera. Using a fossil-calibrated divergence dating analysis, we also carried out the first order-wide diversification study in Hemiptera to track the timing of major cladogenetic events. Equipped with the most comprehensive mitochondrial phylogenomic analysis in Hemiptera and informed by the ancestral state reconstruction of morphological characters, habitat preference, and feeding behaviours, we address the following questions: (i) what is the timing of key morphological adaptations that led to the diversification of habitat utilization and feeding behaviour in Heteroptera? (ii) Was the ancestor of Heteroptera predatory or phytophagous? (iii) What extinction and/or rapid radiation events coincide with the diversification of the major lineages in Hemiptera?

2. Material and methods

(a) Taxon sampling

Previous studies assessed mitochondrial phylogenetic signal limits in Paraneoptera and detected long-branch attraction artefacts among Phthiraptera, Thysanoptera, and Sternorrhyncha [11,25]. Thus, Phthiraptera and Thysanoptera were not included in the taxon sampling of outgroups. We included six outgroup species to represent other paraneopteran lineages as well as the putatively more ancient lineages Blattodea and Mantodea (electronic supplementary material, table S1). As ingroups for phylogenetic analysis, 34 hemipteran species were sequenced in this study, and the sequences of 94 hemipterans were obtained from the National Center for Biotechnology Information (NCBI) database. All 128 mitogenomes represent each of the major hemipteran suborders (with coverage of extant taxa) (electronic supplementary material, table S1).

(b) Complete mitogenome sequence generation

Specimens of 34 hemipterans were collected in 95–100% ethanol and stored at -20°C in the Entomological Museum of China Agricultural University (Beijing, China). Genomic DNA was extracted from the thoracic muscle tissue using the DNeasy blood and tissue kit (Qiagen) following the animal tissue protocol. Whole mitogenomes were generated by amplification, sequencing, and assembly of overlapping PCR fragments, employing general insect mitochondrial primers (electronic supplementary material, table S2). Species-specific primers were designed based on the sequenced fragments to bridge gaps when general primers failed to produce a usable product. Details of the amplification conditions and sequencing strategies were described in our previous study [26].

(c) Assembly, annotation, and alignment

Sequences from each genome were assembled into contigs using SEQUENCER v5.1 (Gene Codes, Ann Arbor, MI, USA).

Protein-coding genes (PCGs) and rRNA genes were identified using BLAST searches of GenBank and alignment with homologous sequences. The tRNAs were identified with tRNAscan-SE v1.21 [27]. Sequences of each PCG (excluding stop codons) were aligned individually based on codon-based multiple alignments using the MAFFT algorithm implemented in the TranslatorX online platform [28]. Ambiguously aligned sites were removed from the protein alignment before back-translating to nucleotides using GBLOCKS in TranslatorX with default settings. Sequences of each RNA gene were individually aligned using the MAFFT v7.0 online server with G-INS-i strategy [29] and ambiguously aligned sites were omitted using GBLOCKS v0.91b [30] with default settings. All alignments were then checked and corrected manually in MEGA v6.0 [31] for quality.

(d) Phylogenetic analyses

Recent phylogenomic studies have shown the ability of site-heterogeneous models (e.g. CAT-based models) to reduce artefacts resulting from mutational saturation and unequal patterns of substitution, which are major problems when analysing genomic data and ancient events [24,25,32–36]. The heterogeneity of sequence divergence within the dataset (e.g. each codon position of PCG and sequences of RNA genes) was analysed using AliGROOVE [37] with the default sliding window size. Indels in the nucleotide dataset were treated as ambiguity and a BLOSUM62 matrix was used as the default amino acid substitution matrix. To account for the strong sequence heterogeneity of the third codon position of the PCGs found in the results of AliGROOVE analysis (electronic supplementary material, figure S2), three datasets were concatenated for phylogenetic analysis: (i) the AA matrix, including amino acid sequences of the 13 PCGs (total of 3123 amino acids), (ii) the protein-coding plus RNA gene (PCGRNA) matrix, including all three codon positions of the 13 PCGs, two rRNA genes, and 17 tRNA genes (total of 11 652 bp), (iii) the PCG12RNA matrix, including the first and second codon positions of the 13 PCGs, two rRNA genes, and 17 tRNA genes (total of 8528 bp). Five tRNAs (Ala, Ile, Met, Gln, and Ser) were not found in many nearly complete mitogenomes and therefore were excluded from our analyses.

Bayesian cross-validation was performed to test the fit of two site-heterogeneous mixture models (CAT and CAT + GTR) and site-homogeneous model (GTR) to our mitogenomic data using PhyloBayes 3.3f [38]. The cross-validation was performed according to the PhyloBayes manual in 10 replicates each with 1100 cycles and the first 100 cycles being discarded as burn-in. The CAT + GTR model was found to be the best fitting model for all datasets (electronic supplementary material, table S3). We then inferred phylogenies from three datasets using PhyloBayes MPI 1.4f [39], with the CAT + GTR model and a discrete gamma distribution with 4 rate categories. In each individual analysis, two independent chains starting from a random tree were run and a consensus tree was calculated by pooling sampled trees from two independent runs, with all analyses satisfactorily converged (maxdiff less than 0.3). The number of discarding trees (burn-in) was calculated case by case to minimize the maxdiff statistics. All analyses were carried out on the CIPRES Science Gateway (<https://www.phylo.org>) and at the High Performance Computing Cluster at the University of Kentucky Analytics and Technologies (UKAT).

(e) Ancestral character state reconstruction

Ancestral states for feeding and living habits and morphological characters were reconstructed in MESQUITE v2.75 (<http://mesquite-project.org>) with Maximum Likelihood (ML) methods. We based ancestral state reconstruction on the tree from PhyloBayes analysis of the PCGRNA dataset with Heteropterodea (Heteroptera + Coleorrhyncha) constrained to be monophyletic. For the ML

optimizations, the ‘Markov k-state 1 parameter model’ (MK1 model in which ‘forward’ and ‘backward’ transition rates are equal) was used. Sources of data for feeding habit, living habitat, mouthpart placement, and the presence of hemelytra are listed in the electronic supplementary material, table S4. To make decisions regarding the significance of ancestral character state reconstructions, we followed the recommendation that ancestral character state estimates with a log-likelihood of two or more units lower than the best state estimate be rejected [40]. For ease of interpretation, likelihoods of ancestral states are reported as proportional likelihoods (PL; scaled to add up to 1, thus expressed as a per cent of total likelihood).

(f) Divergence time estimation

Recent molecular dating analyses have questioned the adequacy of the uncorrelated models of molecular clock relaxation parameters for estimating divergence times with large phylogenomic datasets [41–43]. Based on Bayes factor comparisons, Lepage *et al.* [41] showed that the autocorrelated models provide a significantly better fit than the uncorrelated gamma model for phylogenomic data. Our divergence time estimates were calculated for the two nucleotide and amino acid datasets using PHYLOBAYES 3.3f [38], the best fitting relaxed clock models, and the optimal tree used in the analysis of ancestral character state reconstruction. We used Bayes factor (calculated using thermodynamic integration) in PhyloBayes to compare three widely used relaxed models, the autocorrelated Lognormal and Cox–Ingersoll–Ross (CIR) process and uncorrelated gamma multipliers (UGAM) [41]. In PhyloBayes, Bayes factor analysis was conducted by running 10 000 points, sampling every 10 points after a burn-in of 1000. The uncorrelated UGAM model fell into the same category as the models implemented in BEAST, and this model is shown to fit the data more poorly than two autocorrelated models (CIR and Lognormal). As the Bayes factors for the CIR and Lognormal models were similar (electronic supplementary material, table S5), ‘-auto’ analyses (see PhyloBayes manual) were used to compare these two models. For all molecular clock analyses, a birth–death prior on divergence time and the root age of Hemiptera was constrained to prior 306 to 311 Ma, corresponding to the early Hemiptera fossils from the Moscovian age (e.g. *Aviorrhyncha magnifica* and *Protoprobole straeleni*) [44]. Additionally, 12 fossil calibrations were used with soft bounds, and the details of these fossil calibrations are provided in electronic supplementary material, table S6. We allocated 10% of the probability mass to lie outside each calibration interval. All calculations were performed by running 20 000 generations and sampled every 10 generations (after a burn-in of 2 000 generations).

3. Results and discussion

(a) Phylogeny of Hemiptera

The results of our phylogenetic study based on two nucleotide datasets (PCGRNA and PCG12RNA) produced nearly identical topology with high nodal support values (figure 1; see the electronic supplementary material, figures S3 and S4). The monophyly of Hemiptera was strongly supported, with Sternorrhyncha forming the sister group to all the remaining hemipterans (PP = 1.0 and 0.85). Five long-recognized groups were recovered within Hemiptera: Sternorrhyncha, Cicadomorpha, Fulgoromorpha, Coleorrhyncha, and Heteroptera. However, Auchenorrhyncha was recovered as paraphyletic, with Cicadomorpha forming the sister group to (Fulgoromorpha + Coleorrhyncha). Within Heteroptera, all infraorders were recovered as monophyletic with high support values, except for Cimicomorpha, which was paraphyletic in all

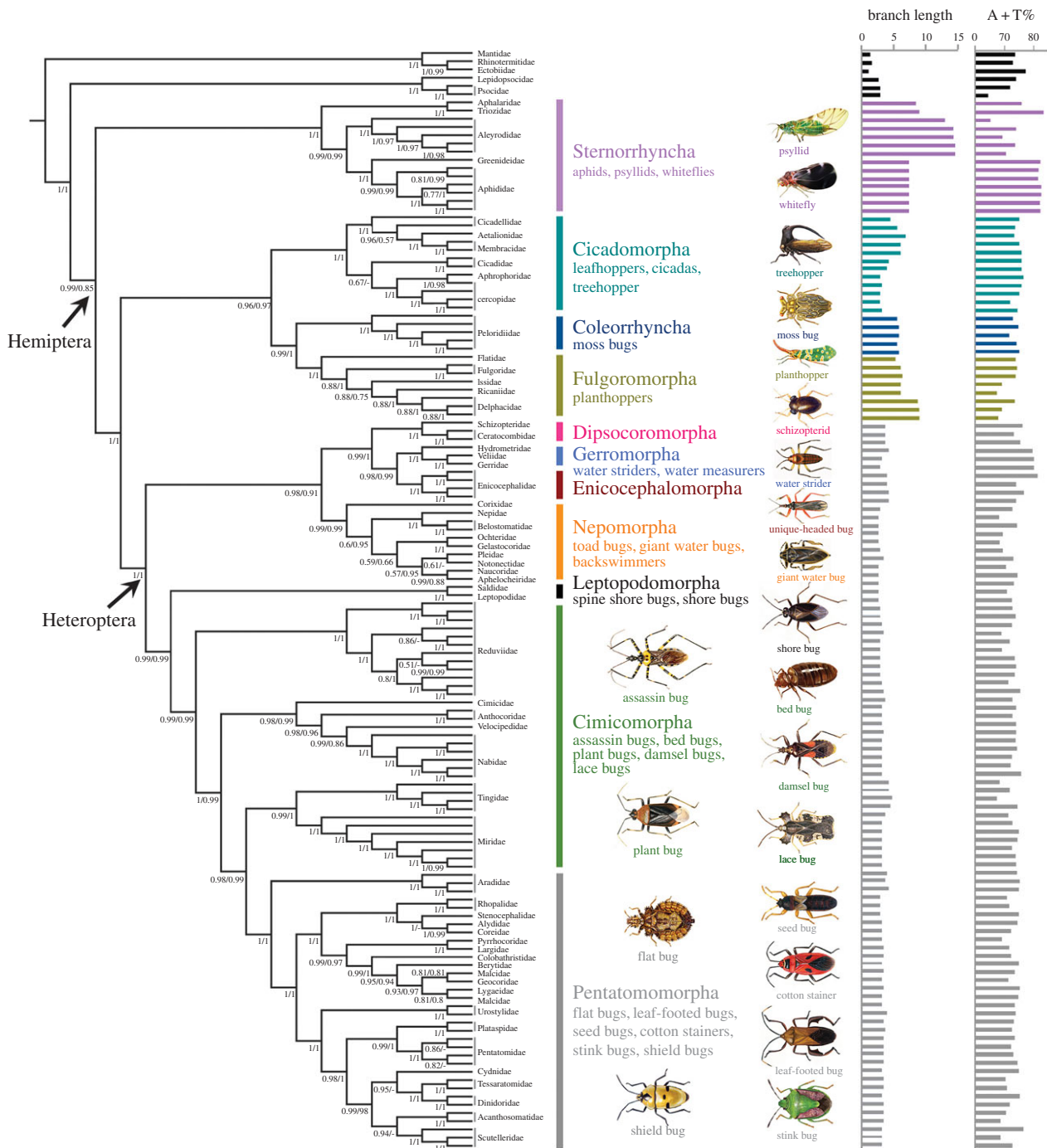


Figure 1. Phylogeny of Hemiptera as inferred from PhyloBayes analyses of the PCGRNA and PCG12RNA datasets under the CAT + GTR mixture model. We show a schematic cladogram depicting the family-level relationships of Hemiptera. Values at nodes are Bayesian posterior probability (PP) using the PCGRNA (left) and PCG12RNA (right) datasets. Dashes indicate PPs less than 0.5. The histogram on the right indicates the branch length of terminal taxa from the Bayesian tree of the PCGRNA dataset and A + T content of terminal taxa from the PCGRNA dataset.

analyses. The Cimicomorpha was the closest extant relative of Pentatomomorpha. Leptopodomorpha was recovered as the sister to Cimicomorpha and Pentatomomorpha. The remaining infraorders formed a clade: (Nepomorpha, (Dipsocoromorpha, (Gerromorpha, Enicocephalomorpha))).

The sister relationship between Sternorrhyncha and all the remaining hemipterans and the monophyletic Auchenorrhyncha have been well resolved in recent studies based on two mitochondrial DNAs and five nuclear loci [10], 1478 nuclear genes [12], and morphological characters [14]. Mitochondrial phylogenomic analyses, however, suggested that the species with accelerated substitution rates always fall together in one group, e.g. the grouping of Sternorrhyncha

with Fulgoromorpha in Song *et al.* [13], Fulgoromorpha with Coleorrhyncha in Cui *et al.* [15], and Sternorrhyncha with Fulgoromorpha and Coleorrhyncha in this study (electronic supplementary material, figure S5). These unexpected groupings were probably caused by the high degree of compositional heterogeneity and, in particular, a significantly accelerated rate in Sternorrhyncha, Fulgoromorpha, and Coleorrhyncha (figure 1; electronic supplementary material, figure S6). The inclusion of rRNA genes in the nucleotide dataset improved the phylogenetic inferences under the site-heterogeneous model, and correctly recovered the majority of deep branches within Hemiptera phylogeny. The sister relationship of Fulgoromorpha and Coleorrhyncha, which exhibited

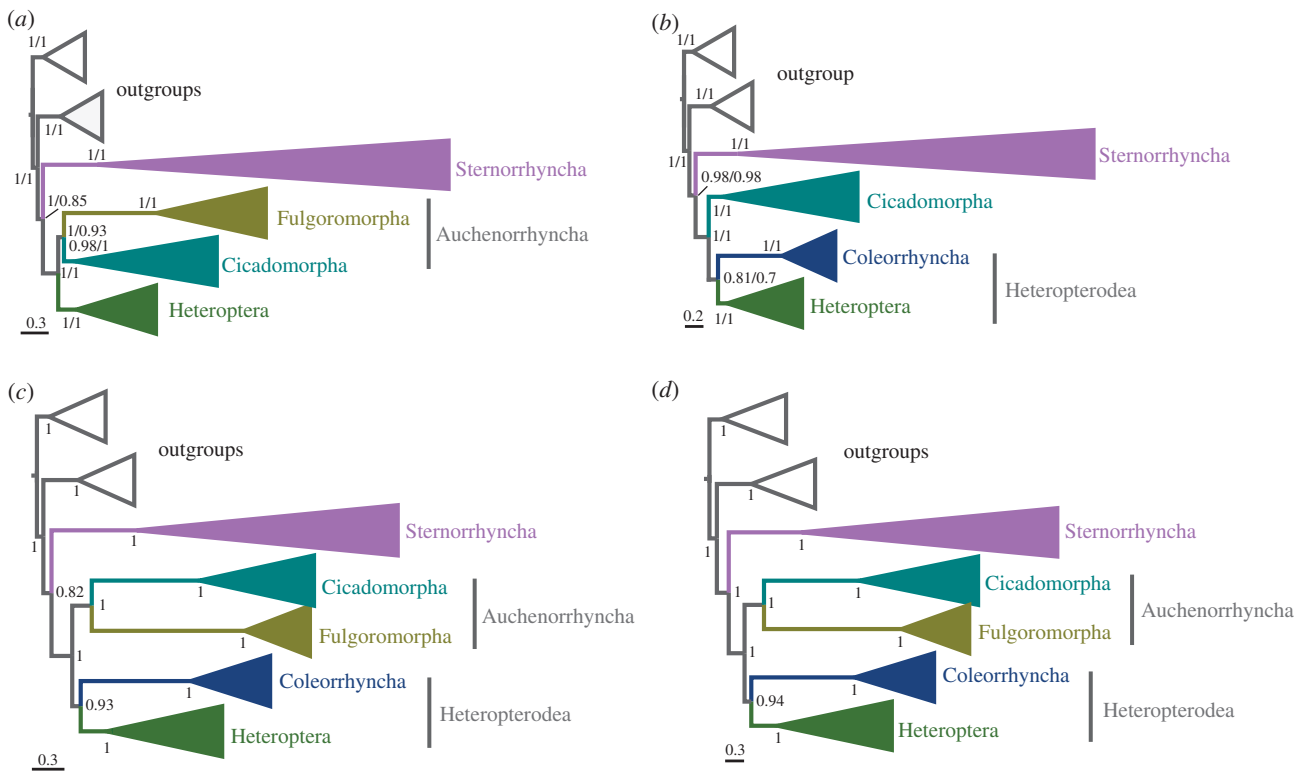


Figure 2. Phylogenetic trees obtained from PhyloBayes analyses of datasets with improved taxon sampling under the CAT + GTR mixture model. (a) Datasets with the removal of Coleorrhyncha. (b) Datasets with the removal of Fulgoromorpha. In (a,b), values at nodes are Bayesian PPs from the modified PCGRNA (left) and PCG12RNA (right) datasets. (c) PCG12RNA and (d) PCGRNA datasets with the removal of 17 species according to the branch length and the heterogeneity in sequence divergence (see the electronic supplementary material, figure S11). We show a schematic version of the Bayesian trees with some suborders and infraorders collapsed for clarity.

long branches compared with species from Cicadomorpha and Heteroptera, was highly supported. We used the ‘long-branch extraction’ method [45] to sequentially remove Coleorrhyncha and then Fulgoromorpha from the Bayesian analyses of PCGRNA and PCG12RNA datasets using a CAT + GTR model. When coleorrhynchans were excluded, the monophyly of Auchenorrhyncha was recovered in all analyses, with Fulgoromorpha forming the sister group to Cicadomorpha (figure 2a; see the electronic supplementary material, figures S7 and S8). When fulgoromorphs were excluded, Coleorrhyncha grouped with Heteroptera (figure 2b; see the electronic supplementary material, figures S9 and S10). The four resulting trees showed the identical relationships of heteropteran infraorders as those obtained from the original analyses. These results suggest that the grouping of Fulgoromorpha and Coleorrhyncha is probably an artefact. With the removal of five species with the longest branches in Sternorrhyncha, seven species with the shortest branches in Cicadomorpha, three species with the longest branches in Fulgoromorpha (electronic supplementary material, figure S11a), and the two moss bug species in Coleorrhyncha with substantial heterogeneity in their sequence divergence (electronic supplementary material, figure S11b), we generated a 117-taxa dataset for the subsequent phylogenetic analysis (see the electronic supplementary material, figure S11). Using PhyloBayes with a CAT + GTR model, the monophyly of Auchenorrhyncha (PP = 1.0) and the sister relationship of Heteroptera and Coleorrhyncha (PP = 0.93 and 0.94) were both recovered by the datasets PCG12RNA (figure 2c; electronic supplementary material, figure S12) and PCGRNA (figure 2d; electronic supplementary material, figure S13).

With the most comprehensive hemipteran mitogenome sampling to date, the site-heterogeneous mixture model produces an almost fully resolved tree except for the paraphyletic Cicadomorpha. Our results demonstrate that mitogenomes have considerable resolving power in a phylogenetic study because of the ease of sequencing, the feasibility of large taxon sampling, and the use of comprehensive evolutionary models [24,25,32,36].

(b) Ancestral state reconstructions

Results of ancestral state reconstructions suggest that the common ancestors of Hemiptera, Sternorrhyncha, Auchenorrhyncha, and Coleorrhyncha are all phytophagous and terrestrial with significant PL in all cases, whereas the common ancestor of Heteroptera is predaceous and terrestrial (figure 3; electronic supplementary material, figures S14 and S15). Within Heteroptera, there was a transition from predation to phytophagy in the common ancestor of Miridae + Tingidae (Cimicomorpha, in part) and Pentatomomorpha, and a reversal from phytophagy to predation (Geocoridae). If we include predatory Pentatomidae and Miridae, at least two additional independent reversals would be expected. Omnivory arose twice independently, once from a predaceous ancestor within Nepomorpha (Corixidae) and once from a phytophagous ancestor within Miridae. There were two independent transitions (Reduviidae and Cimicidae) from predation to haematophagy (blood feeding) in Heteroptera. Fungivory in adults arose once from a phytophagous ancestor in Aradidae (and in nymphs of some Auchenorrhyncha). All aquatic, water surface-dwelling, and

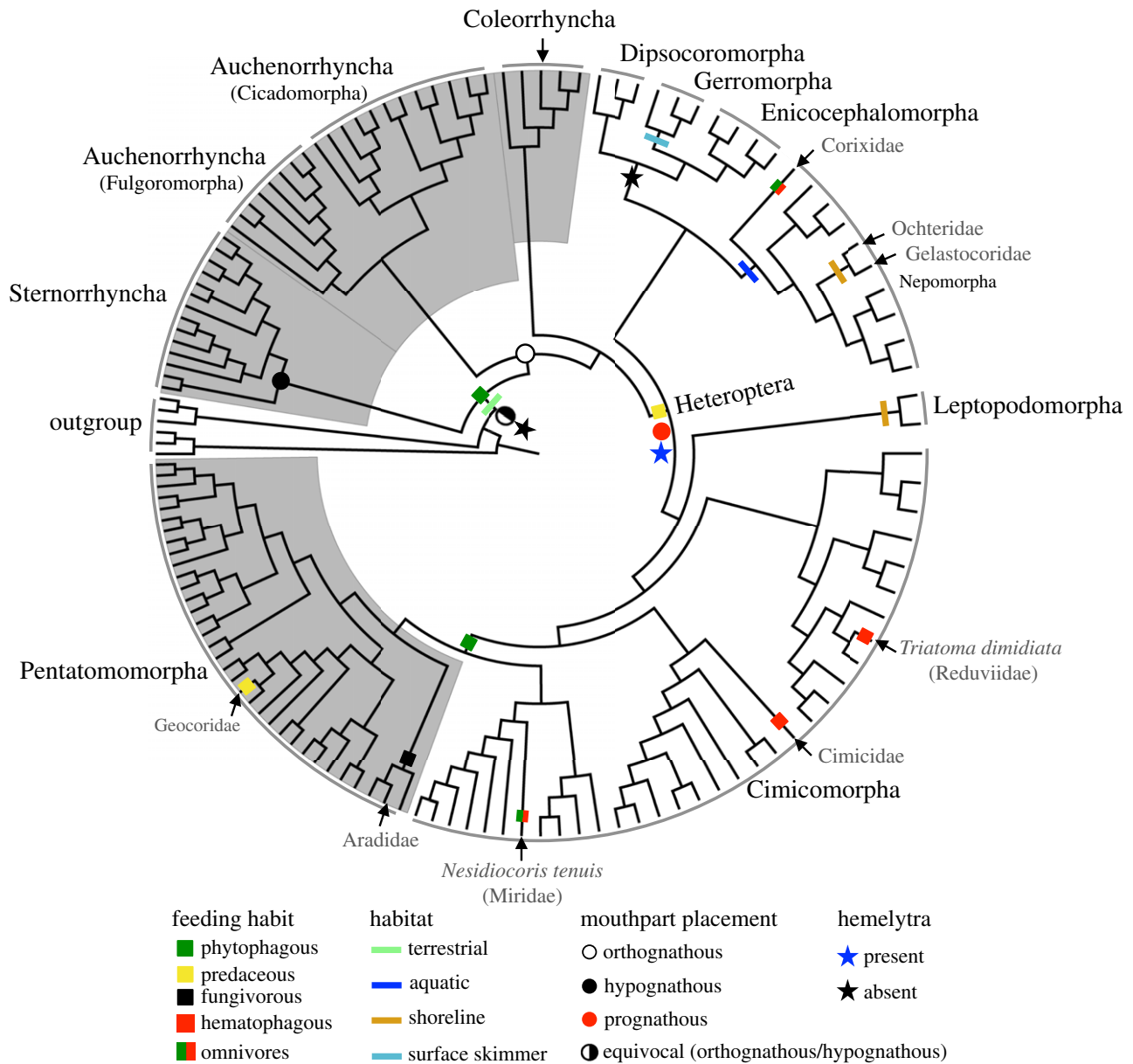


Figure 3. Summary of character state transitions for four characters of hemipteran insects. We based ancestral state reconstruction on the tree from PhyloBayes analysis of the PCGRNA dataset with Heteropterodea (Heteroptera + Coleorrhyncha) constrained to be monophyletic. All character state transitions are judged to be significant by ML methods except where otherwise noted (equivocal or unknown).

litter-dwelling infraorders (incidentally all predators except the omnivorous Corixidae) were recovered as a monophyletic group.

There were two independent transitions to shoreline habitat (one from a terrestrial ancestor and one from an aquatic ancestor within Nepomorpha), one transition to surface skimmers from a terrestrial ancestor (Gerromorpha), and one transition from terrestrial to aquatic habitat (Nepomorpha). Optimization of mouthpart origin (see the electronic supplementary material, figure S1a–d) indicates that the ancestor of Sternorrhyncha had hypognathous mouthparts, the ancestors of Auchenorrhyncha and Coleorrhyncha had orthognathous mouthparts, and the ancestor of Heteroptera had prognathous mouthparts that arose from an ancestor with orthognathous mouthparts (figure 3; electronic supplementary material, figure S16). The presence of hemelytra (electronic supplementary material, figure S1e) arose once in the common ancestor of all Heteroptera (figure 3; electronic supplementary material, figure S17). This character state was lost in the

common ancestor to Dipsocoromorpha, Enicocephalomorpha, and Gerromorpha.

(c) Adaptive innovations driving the diversification of true bugs

Divergence data estimates were not significantly different between datasets using an autocorrelated CIR model (electronic supplementary material, table S7). Hemiptera shares a common ancestor with the remaining Paraneoptera about 328 Ma (confidence interval (CI), 340–318 Ma; figure 4). Subsequently, Hemiptera diversified into Sternorrhyncha and the remaining Hemiptera approximately 309 Ma (CI 311–306 Ma), at the end of the radiation of spermatophytes (seed plants) 385–299 Ma [46]. Our analyses suggest that a Permian diversification of hemipteran suborders was immediately followed by a Triassic diversification of heteropteran infraorders (figure 4). From a Carboniferous origin, early terrestrial lineages of Hemiptera radiated soon after the

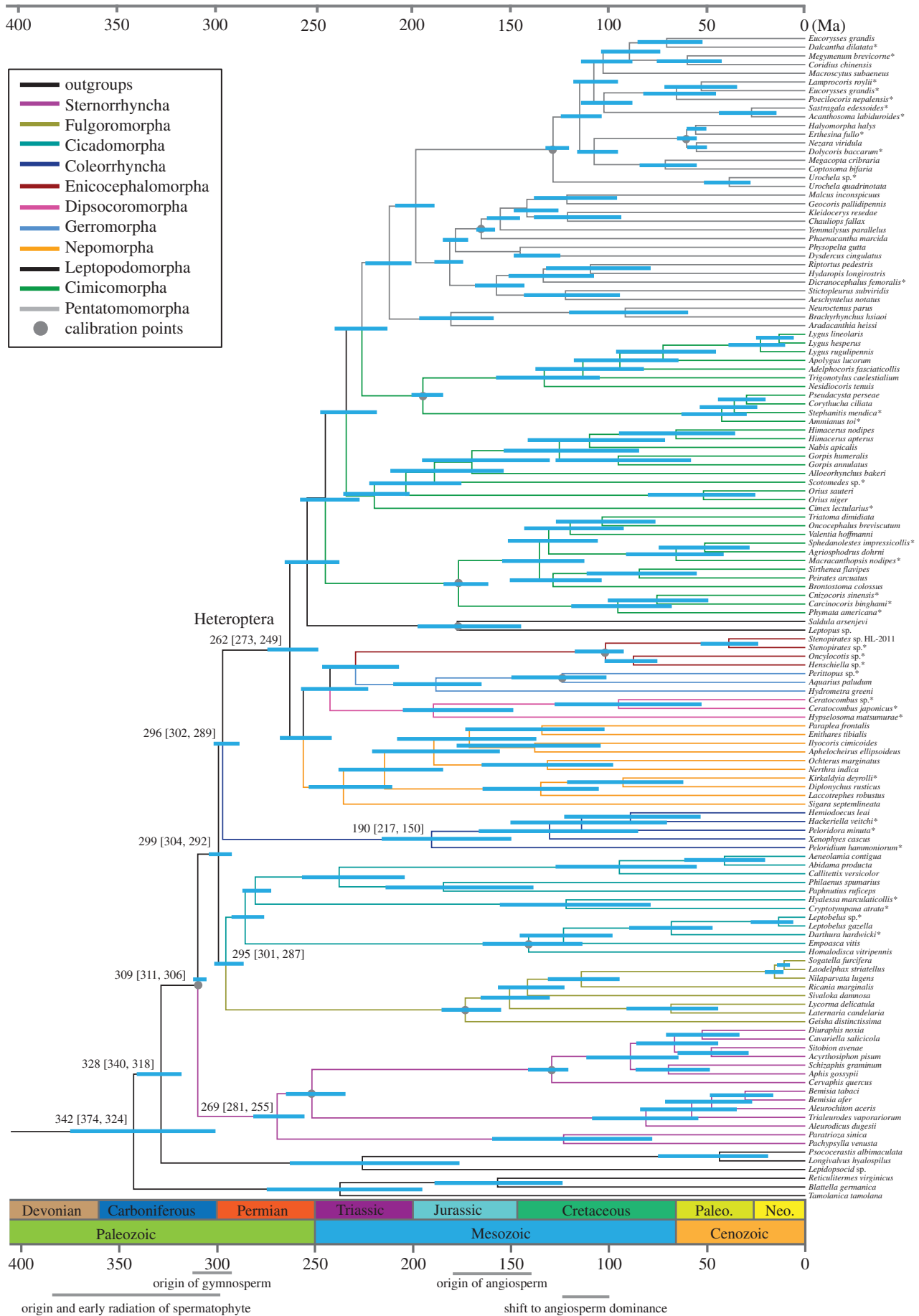


Figure 4. Chronogram showing hemipteran phylogeny and divergence time estimates. Consensus tree presenting divergence dates produced by the PhyloBayes analysis of the PCGRNA dataset (with Heteropteroidea, Heteroptera + Coleorrhyncha, constrained to be monophyletic) using 13 fossil calibration points, the CIR autocorrelated process, the site-heterogeneous mixture CAT + GTR substitution model, and soft bound 10%. Blue bars indicate 95% mean confidence intervals of each node. A geological timescale is shown at the bottom. New mitogenomes are highlighted using an asterisk close to the species name. Divergence date estimates based on the PCG12RNA and AA under the CIR model are summarized in the electronic supplementary material, table S7.

hypothesized origin of gymnosperms [47], and formed Sternorrhyncha, Auchenorrhyncha, Coleorrhyncha, and Heteroptera in the Permian.

With the exception of some mycophagous nymphs, Sternorrhyncha, Auchenorrhyncha, and Coleorrhyncha are entirely phytophagous, feeding on fluids of phloem, xylem, or cambium, with some inducing galls (some Psylloidea, Aphidoidea, and Coccoidea) [21]. Extinct hemipteran taxa that formed the ancestral stock of today's major lineages were consistently linked to gymnosperms. Shcherbakov [48] inferred 'such short-legged Archescytinidae (primitive Hemiptera) either lived in confined spaces of gymnosperm reproductive organs or clung tightly to the plant surface'. Small, usually dorsoventrally depressed hoppers and their flattened cryptic nymphs (a body form possibly adapted to living between cone scales) [49] probably fed on phloem of thick gymnosperm stems [21]. The first xylem-feeding Hemiptera existed in the gymnosperm-dominated Permian and Triassic forests, while the large and clumsily built early Permian boreoscytids possibly fed on large gymnosperm ovules [21]. Furthermore, fossils representing the early members of Sternorrhyncha and Auchenorrhyncha were recovered from the same Kungurian beds (275–270 Ma), coincident with gymnosperm dominance [21]. However, most family-level diversification events in extant Sternorrhyncha seem to coincide with the angiosperm radiation, as indicated in our analysis (Psylloidea may be exceptional in that all eight extant families are not older than the Eocene) [50]. Ortiz-Rivas *et al.* [51] likewise linked angiosperm and aphid tribe diversification, producing angiosperm-feeding taxa. As all extant superfamilies of Sternorrhyncha (scales, aphids, whiteflies) feed on angiosperms and gymnosperms, yet evolved from gymnosperm feeders [17,51–53], it is difficult to deduce the finer mechanisms governing their evolution. This notion is especially complicated considering that well after the era of gymnosperm replacement with angiosperms (beginning 150 Ma), there was an increase in gymnosperm diversification rates persisting over the last 100 Ma [54].

The evolution and diversification of seed plants give rise to vast ecological niches [46]. The evolution of the seed has not only promoted the evolutionary success of plants for nearly 400 Ma but also probably initiated and facilitated the subsequent success of Hemiptera. Among the four hemipteran suborders, Heteroptera displays the greatest diversity in their habitat and behaviour, as well as species diversity. The origin of Heteroptera (approx. 262 Ma; CI, 273–249 Ma) coincided with the evolution of the apically produced labium (electronic supplementary material, figure S1*d*; i.e. a gula permitting a prognathous rostrum position), predatory behaviour, and the novel protective forewing. The true bug infraorders diversified in the Late Permian and Triassic (262–226 Ma). We propose that the diversification of potential prey species following the P-T extinction (252 Ma) [55] may have paved the way for the diversification of the arthropod-feeding heteropteran lineages. Two key adaptations that facilitated the rapid family-level radiation of Heteroptera coincide with the shift from gymnosperm- to angiosperm-dominance. The evolution of prognathous mouthparts and the novel hemelytron probably facilitated multiple independent evolutions of predatory behaviour from a phytophagous ancestor and, consequently, multiple transitions to aquatic and semi-aquatic habitats. The prognathous mouthparts clearly facilitated the development of a more versatile suite of feeding behaviours including

predation, blood feeding, and mycophagy, none of which occurs in the other three predominantly phytophagous and entirely terrestrial suborders. This behavioural diversity may explain the higher rates of diversification (of extant lineages) in Heteroptera, composing more number of families and species than the other three hemipteran suborders.

Angiosperm coevolution is often the default explanation for major radiations. The family diversifications of Sternorrhyncha (without Psylloidea) and Pentatomomorpha coincide, in large part, with the consequent decline of gymnosperms (i.e. shift to angiosperm dominance; 125–100 Ma) [56]. However, at the family-level, there is little clear association to be made between the radiations of hemipterans and angiosperms, probably because the latter is much older than the former as in the case of Psylloidea. Although angiosperms may have driven familial or intrafamilial diversity in many groups, the diversity of habitat and feeding behaviour observed in Heteroptera can almost entirely be linked to diversification events coincident with an era of angiosperm suppression before 150 Ma. Of course, hypothesizing explanations for such ancient events remains challenging. Future studies focusing on thorough sampling of each suborder/infraorder must be conducted to elucidate finer intrafamilial radiation stories (perhaps for interfamilial relationships as well).

Similarly, Hunt *et al.* [4] failed to directly link the 'superradiation' of beetles (which account for 25% of all metazoans) at 285 Ma with angiosperm coevolution. Like beetles, true bugs exhibit immense versatility with diverse habitat colonization, varied feeding habits, and modified forewings that confer protection [57] and even facilitate plastron (air bubble) retention for aquatic respiration. The modified, protective forewings of beetles and true bugs may account for the rapid lineage diversification and probably facilitated the versatile feeding and habitat colonization, including multiple independent shifts to predation and aquatic habitats [4] that gave rise to the biodiversity before us today. The radiation of angiosperms may have simply facilitated an already bustling process.

Data accessibility. The annotated mitogenome sequences of 34 hemipteran insects have been deposited in GenBank (accession numbers shown in electronic supplementary material, table S1). The data supporting this article have been uploaded as the electronic supplementary material and the datasets used in the phylogenetic analyses are available in the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.13m7r> [58].

Authors' contributions. H.L., X.Z., and W.C. designed this study. H.L., F.S., and P.J. collected insect materials and performed the molecular experiments. H.L., E.C., F.S., and J.P.L. analysed the data. All authors discussed results and implications and read and approved the final manuscript. H.L., J.L., E.C., D.B., X.Z., and W.C. wrote the manuscript.

Competing interests. We declare no competing interests.

Funding. This work was supported by grants from the National Natural Science Foundation of China (nos. 31372229, 31401991 and 31420103902), the National Basic Research Program of China (no. 2013CB127600), the Chinese Universities Scientific Fund (nos. 2017QC100, 2017QC066 and 2017ZB002), and a start-up fund from the University of Kentucky. The information reported in this paper (no. 17-08-055) is part of a project of the Kentucky Agricultural Experiment Station and is published with the approval of the Director. These agencies had no role in study design, data collection/analysis, manuscript preparation, or the decision to publish.

Acknowledgements. We thank Tadashi Ishikawa (University of Tokyo, Japan) for providing samples of Dipsochoromorpha. We also thank Drs Thomas J. Henry (United States National Museum of Natural History, USA), Christiane Weirauch (University of California, USA), Guanyang Zhang (Arizona State University, USA), David Weisrock and John J. Obrycki (University of Kentucky, USA), Nicole Gunter

(Cleveland Museum of Natural History, USA), Colin S. Brent (Arid-Land Agricultural Research Center, USA), Mick Webb, William Rodney Dolling and Diana Percy (Natural History Museum, UK),

and Mr Leonidas-Romanos Davranoglou (Imperial College, UK) for their constructive comments on the earlier versions of the manuscript.

References

- Mayr E. 1963 *Animal species and evolution*, p. 797. Cambridge, MA: Harvard University Press.
- Müller GB, Newman SA. 2005 The innovation triad: an EvoDevo agenda. *J. Exp. Zool. Part B* **304**, 487–503. (doi:10.1002/jez.b.21081)
- Wiegmann BM *et al.* 2011 Episodic radiations in the fly tree of life. *Proc. Natl Acad. Sci. USA* **108**, 5690–5695. (doi:10.1073/pnas.1012675108)
- Hunt T *et al.* 2007 A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* **318**, 1913–1916. (doi:10.1126/science.1146954)
- Zhang Z. 2011 Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness. *Zootaxa* **3148**, 1–237.
- Weirauch C, Schuh RT. 2011 Systematics and evolution of Heteroptera: 25 years of progress. *Annu. Rev. Entomol.* **56**, 487–510. (doi:10.1146/annurev-ento-120709-144833)
- Schuh RT, Slater JA. 1995 *True bugs of the world (Hemiptera: Heteroptera): classification and natural history*, p. 336. Ithaca, NY: Cornell University Press.
- Myers JG, China WE. 1929 The systematic position of the Peloridiidae as elucidated by a further study of the external anatomy of Hemiodoecus leai China (Hemiptera, Peloridiidae). *Ann. Mag. Nat. Hist.* **3**, 282–294. (doi:10.1080/00222932908672971)
- Bourgoin T, Campbell BC. 2002 Inferring a phylogeny for Hemiptera: falling into the 'autapomorphic trap'. *Denisia* **4**, 67–82.
- Cryan JR, Urban JM. 2012 Higher-level phylogeny of the insect order Hemiptera: is *Auchenorrhyncha* really paraphyletic? *Syst. Entomol.* **37**, 7–21. (doi:10.1111/j.1365-3113.2011.00611.x)
- Talavera G, Vila R. 2011 What is the phylogenetic signal limit from mitogenomes? The reconciliation between mitochondrial and nuclear data in the Insecta class phylogeny. *BMC Evol. Biol.* **11**, 315. (doi:10.1186/1471-2148-11-315)
- Misof B *et al.* 2014 Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767. (doi:10.1126/science.1257570)
- Song N, Liang AP, Bu CP. 2012 A molecular phylogeny of Hemiptera inferred from mitochondrial genome sequences. *PLoS ONE* **7**, e48778. (doi:10.1371/journal.pone.0048778)
- Friedemann K, Spangenberg R, Yoshizawa K, Beutel RG. 2014 Evolution of attachment structures in the highly diverse Acercaria (*Hexapoda*). *Cladistics* **30**, 170–201. (doi:10.1111/cla.12030)
- Cui Y, Xie Q, Hua J, Dang K, Zhou J, Liu X, Wang G, Yu X, Bu, W. 2013 Phylogenomics of Hemiptera (Insecta: *Paraneoptera*) based on mitochondrial genomes. *Syst. Entomol.* **38**, 233–245. (doi:10.1111/j.1365-3113.2012.00660.x)
- von Dohlen CD, Moran NA. 1995 Molecular phylogeny of the Homoptera: a paraphyletic taxon. *J. Mol. Evol.* **41**, 211–223. (doi:10.1007/BF00170675)
- Gullan PJ, Kosztarab M. 1997 Adaptations in scale insects. *Annu. Rev. Entomol.* **42**, 23–50. (doi:10.1146/annurev.ento.42.1.23)
- Sweet MH. 1979 On the original feeding habits of the Hemiptera (Insecta). *Ann. Entomol. Soc. Am.* **72**, 575–579. (doi:10.1093/aesa/72.5.575)
- Schuh RT, Cobben RH. 1979 Evolutionary trends in Heteroptera. Part II. Mouthpart-structure and feeding strategies. *Syst. Zool.* **28**, 653–656. (doi:10.2307/2412577)
- Tanner LH, Lucas SG, Chapman MG. 2004 Assessing the record and causes of Late Triassic extinctions. *Earth-Sci. Rev.* **65**, 103–139. (doi:10.1016/S0012-8252(03)00082-5)
- Rasnitsyn AP, Quicke DLJ. 2002 *History of insects*, p. 517. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Grimaldi D, Engel MS. 2005 *Evolution of the insects*, p. 755. New York, NY: Cambridge University Press.
- Cameron SL. 2014 Insect mitochondrial genomics: implications for evolution and phylogeny. *Annu. Rev. Entomol.* **59**, 95–117. (doi:10.1146/annurev-ento-011613-162007)
- Gillett CPDT, Crampton-Platt A, Timmermans MJTN, Jordal BH, Emerson BC, Vogler AP. 2014 Bulk *de novo* mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (*Coleoptera: Curculionoidea*). *Mol. Biol. Evol.* **31**, 2223–2237. (doi:10.1093/molbev/msu154)
- Li H, Shao R, Song N, Song F, Jiang P, Li Z, Cai W. 2015 Higher-level phylogeny of paraneopteran insects inferred from mitochondrial genome sequences. *Sci. Rep.* **5**, 8527. (doi:10.1038/srep08527)
- Li H, Liu H, Shi A, Štys P, Zhou X, Cai W. 2012 The complete mitochondrial genome and novel gene arrangement of the unique-headed bug *Stenopirates* sp. (*Hemiptera: Enicocephalidae*). *PLoS ONE* **7**, e29419. (doi:10.1371/journal.pone.0029419)
- Lowe TM, Eddy SR. 1997 tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964. (doi:10.1093/nar/25.5.0955)
- Abascal F, Zardoya R, Telford MJ. 2010 TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **38**, W7–W13. (doi:10.1093/nar/gkq291)
- Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. (doi:10.1093/molbev/mst010)
- Talavera G, Castresana J. 2007 Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577. (doi:10.1080/10635150701472164)
- Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. 2013 MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729. (doi:10.1093/molbev/mst197)
- Song F, Li H, Jiang P, Zhou X, Liu J, Sun C, Vogler AP, Cai WZ. 2016 Capturing the phylogeny of Holometabola with mitochondrial genome data and Bayesian site-heterogeneous mixture models. *Genome Biol. Evol.* **8**, 1411–1426. (doi:10.1093/gbe/evw086)
- Lartillot N, Philippe H. 2004 A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095–1109. (doi:10.1093/molbev/msh112)
- Delsuc F, Brinkmann H, Philippe H. 2005 Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* **6**, 361–375. (doi:10.1038/nrg1603)
- Morgan CC, Foster PG, Webb AE, Pisani D, McInerney JO, O'Connell MJ. 2013 Heterogeneous models place the root of the placental mammal phylogeny. *Mol. Biol. Evol.* **30**, 2145–2156. (doi:10.1093/molbev/mst117)
- Timmermans MJTN *et al.* 2015 Family-level sampling of mitochondrial genomes in Coleoptera: compositional heterogeneity and phylogenetics. *Genome Biol. Evol.* **8**, 161–175. (doi:10.1093/gbe/evv241)
- Küçk P, Meid SA, Groß C, Wägele JW, Misof B. 2014 AliGROOVE – visualization of heterogeneous sequence divergence within multiple sequence alignments and detection of inflated branch support. *BMC Bioinform.* **15**, 294. (doi:10.1186/1471-2105-15-294)
- Lartillot N, Lepage T, Blanquart S. 2009 PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288. (doi:10.1093/bioinformatics/btp368)
- Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013 PhyloBayes-MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* **62**, 611–615. (doi:10.1093/sysbio/syt022)
- Pagel M. 1999 The maximum likelihood approach to reconstructing ancestral character states on phylogenies. *Syst. Biol.* **48**, 612–622. (doi:10.1080/106351599260184)
- Lepage T, Bryant D, Philippe H, Lartillot N. 2007 A general comparison of relaxed molecular clock

- models. *Mol. Biol. Evol.* **24**, 2669–2680. (doi:10.1093/molbev/msm193)
42. Chiari Y, Cahais V, Galtier N, Delsuc F. 2012 Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (*Archosauria*). *BMC Biol.* **10**, 65. (doi:10.1186/1741-7007-10-65)
 43. Rota-Stabelli O, Daley AC, Pisani D. 2013 Molecular timetrees reveal a Cambrian colonization of land and a new scenario for ecdysozoan evolution. *Curr. Biol.* **23**, 392–398. (doi:10.1016/j.cub.2013.01.026)
 44. Nel A *et al.* 2013 The earliest known holometabolous insects. *Nature* **503**, 257–261. (doi:10.1038/nature12629)
 45. Siddall ME, Whiting MF. 1999 Long-branch abstractions. *Cladistics* **15**, 9–24. (doi:10.1111/j.1096-0031.1999.tb00391.x)
 46. Sims HJ. 2012 The evolutionary diversification of seed size: using the past to understand the present. *Evolution* **66**, 1636–1649. (doi:10.1111/j.1558-5646.2011.01527.x)
 47. Smith SA, Beaulieu JM, Donoghue MJ. 2010 An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proc. Natl Acad. Sci. USA* **107**, 5897–5902. (doi:10.1073/pnas.1001225107)
 48. Shcherbakov DE. 2002 The 270 million year history of *Auchenorrhyncha* (Homoptera). *Denisia* **4**, 29–36.
 49. Leeper GW. 1962 *The evolution of living organisms*, p. 469. Melbourne, Australia: Melbourne University Press.
 50. Ouvrard D, Burckhardt D, Azar D, Grimaldi D. 2010 Non-jumping plant-lice in Cretaceous amber (*Hemiptera: Sternorrhyncha: Psylloidea*). *Syst. Entomol.* **35**, 172–180. (doi:10.1111/j.1365-3113.2009.00499.x)
 51. Ortiz-Rivas B, Moya A, Martinez-Torres D. 2004 Molecular systematics of aphids (Homoptera: *Aphididae*): new insights from the long-wavelength opsin gene. *Mol. Phylogenet. Evol.* **30**, 24–37. (doi:10.1016/S1055-7903(03)00155-6)
 52. Drohojowska J, Szewo J. 2015 Early cretaceous aleyrodidae (*Hemiptera: Sternorrhyncha*) from the Lebanese amber. *Cretaceous Res.* **52**, 368–389. (doi:10.1016/j.cretres.2014.03.015)
 53. Shcherbakov DE. 2000 Permian faunas of Homoptera (*Hemiptera*) in relation to phytogeography and the Permian-Triassic Crisis. *Paleontol. J.* **34**, S251–S267.
 54. Burleigh JG, Barbazuk WB, Davis JM, Morse AM, Soltis PS. 2012 Exploring diversification and genome size evolution in extant gymnosperms through phylogenetic synthesis. *J. Bot.* **2012**, 1–6. (doi:10.1155/2012/292857)
 55. Chen ZQ, Benton MJ. 2012 The timing and pattern of biotic recovery following the end-Permian mass extinction. *Nat. Geosci.* **5**, 375–383. (doi:10.1038/ngeo1475)
 56. Berendse F, Scheffer M. 2009 The angiosperm radiation revisited, an ecological explanation for Darwin's 'abominable mystery'. *Ecol. Lett.* **12**, 865–872. (doi:10.1111/j.1461-0248.2009.01342.x)
 57. Crowson RA. 1981 *The biology of the Coleoptera*, p. 802. London, UK: Academic Press.
 58. Li H, Leavengood Jr JM, Chapman EG, Burckhardt D, Song F, Jiang P, Liu J, Zhou X, Cai W. 2017 Data from: Mitochondrial phylogenomics of Hemiptera reveals adaptive innovations driving the diversification of true bugs. Dryad Digital Repository. (<http://dx.doi.org/10.5061/dryad.13m7r>)