Phylogenomics and the evolution of hemipteroid insects

Author accepted version

Online final version: https://doi.org/10.1073/pnas.1815820115

[1]Kevin P. Johnson, [1]Christopher H. Dietrich, [2]Frank Friedrich, [3]Rolf G. Beutel, [3,4]Benjamin Wipfler, [4]Ralph S. Peters, [1,5]Julie M. Allen, [6]Malte Petersen, [6]Alexander Donath, [7]Kimberly K. O. Walden, [8]Alexey Kozlov, [6,9]Lars Podsiadlowski, [6]Christoph Mayer, [6,10,11]Karen Meusemann, [6]Alexandros Vasilikopoulos, [12]Robert M. Waterhouse, [13]Stephen Cameron, [14]Christiane Weirauch, [1]Daniel R. Swanson, [15]Diana M. Percy, [16]Nathaniel Hardy, [17]Irene Terry, [18]Shanlin Liu, [19]Xin Zhou, [6]Bernhard Misof, [7]Hugh M. Robertson, [20]Kazunori Yoshizawa

[1]Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Champaign, Illinois 61820 USA

[2]Institut für Zoologie, Universität Hamburg, Hamburg, Germany

[3]Institut für Zoologie und Evolutionsforschung, Friedrich-Schiller-Universität Jena, 07743 Jena, Germany.

[4]Center of Taxonomy and Evolutionary Research, Arthropoda Department, Zoological Research Museum Alexander Koenig, 53113 Bonn, Germany

[5]Department of Biology, University of Nevada, Reno, Nevada, USA

[6]Center for Molecular Biodiversity Research, Zoological Research Museum Alexander Koenig, 53113 Bonn, Germany

[7]Department of Entomology, University of Illinois, Urbana, Illinois, USA

[8]Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

[9]Institute of Evolutionary Biology and Ecology, University of Bonn, 53121 Bonn, Germany

[10]Evolutionary Biology and Ecology, Institute for Biology I (Zoology), University of Freiburg, 79104 Freiburg, Germany

[11]Australian National Insect Collection, CSIRO National Research Collections Australia, Canberra, Australia

[12]Department of Ecology and Evolution, University of Lausanne and Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

[13]Department of Entomology, Purdue University, West Lafayette, Indiana, USA

[14]Department of Entomology, University of California, Riverside, California, USA

[15]Department of Life Sciences, Natural History Museum, London, United Kingdom

[16]Department of Entomology and Plant Pathology, Auburn University, Auburn, Alabama, USA

[17]Department of Biology, University of Utah, Salt Lake City, Utah, USA

[18]BGI-Shenzhen, Shenzhen, Guangdong Province, 518083, People's Republic of China

[19]Department of Entomology, China Agricultural University, Beijing 100193, People's Republic of China

[20]Systematic Entomology, Hokkaido University, Sapporo, Japan.

Classification: Biological Sciences, Evolution

Short Title: Phylogenomics of hemipteroid insects

Corresponding Author: Kevin P. Johnson, Illinois Natural History Survey, Prairie Research Institute, University of Illinois, 1816 South Oak Street, Champaign, Illinois, 61820 USA, phone: 217-244-9267, email: kpjohnso@illinois.edu

Keywords: phylogeny, systematics, transcriptomes, Hemiptera, Thysanoptera, Psocodea

Abstract

Hemipteroid insects (Paraneoptera), with over 10% of all known insect diversity, are a major component of terrestrial and aquatic ecosystems. Previous phylogenetic analyses have not consistently resolved the relationships among major hemipteroid lineages. We provide maximum likelihood-based phylogenomic analyses of a taxonomically comprehensive dataset comprising sequences of 2,395 single-copy, protein-coding genes for 193 samples of hemipteroid insects and outgroups. These analyses yield a well-supported phylogeny for hemipteroid insects. Monophyly of each of the three hemipteroid orders (Psocodea, Thysanoptera, and Hemiptera) is strongly supported, as are most relationships among suborders and families. Thysanoptera (thrips) is strongly supported as sister to Hemiptera. However, as in a recent large-scale analysis sampling all insect orders, trees from our data matrices support Psocodea (bark lice and parasitic lice) as the sister group to the holometabolous insects (those with complete metamorphosis). In contrast, four-cluster likelihood mapping of these data does not support this result. A molecular dating analysis using 23 fossil calibration points suggests hemipteroid insects began diversifying before the Carboniferous, over 365 million years ago. We also explore implications for understanding the timing of diversification, the evolution of morphological traits, and the evolution of mitochondrial genome organization. These results provide a phylogenetic framework for future studies of the group.

Significance Statement

Hemipteroid insects constitute a major fraction of insect diversity, comprising three orders and over 120,000 described species. We used a comprehensive sample of the diversity of this group involving 193 genome-scale datasets and sequences from 2,395 genes to uncover the evolutionary tree for these insects and provide a timescale for their diversification. Our results indicated that thrips (Thysanoptera) are the closest living relatives of true bugs and allies (Hemiptera) and that these insects started diversifying before the Carboniferous period, over 365 million years ago. The evolutionary tree from this research provides a backbone framework for future studies of this important group of insects.

Introduction

The hemipteroid insect orders, Psocodea (bark lice and parasitic lice), Thysanoptera (thrips), and Hemiptera (true bugs and allies; i.e. hemipterans), with over 120,000 described species, comprise well over 10% of known insect diversity. However, the evolutionary relationships among the major lineages of these insects are not yet resolved. Recent phylogenomic analyses questioned the monophyly of this group (1) demanding a reconsideration of the evolution of hemipteroid and holometabolous insects. We assess these prior results, which placed Psocodea as the sister taxon to Holometabola (insects with complete metamorphosis; e.g. wasps, flies, beetles, butterflies), and uncover relationships within and among hemipteroid insect orders by analyzing a large phylogenomic data set covering all major lineages of hemipteroid insects.

Knowledge of the phylogeny of these insects is important for several reasons. First, major transitions between the mandibulate (chewing) mouthpart insect groundplan and piercing-sucking mouthparts occurred in this group. In particular, thrips and hemipterans, and some ectoparasite lice in Psocodea, have highly modified mouthparts adapted for feeding on fluids and, hence, differ markedly from their mandibulate ancestors. Through a series of remarkable modifications, hemipteroids acquired a "piercing-sucking" mode of feeding in both immature and adult stages that enabled them to feed not only on plant vascular fluids, but also on blood and other liquid diets. Resolution of the evolutionary tree of hemipteroid insects is needed to provide a framework for understanding morphological transitions that occurred in this group, as well as to provide a timeframe over which these changes occurred.

In addition, several lineages of hemipteroid insects (particularly thrips and Psocodea) underwent major reorganizations of their mitochondrial genomes, including the emergence of

minicircles (2). Understanding how these changes in mitochondrial genome organization occurred requires knowledge of evolutionary relationships to document in which lineages these changes first arose. Finally, hemipteroids are among the most abundant insects (3) and are therefore key components of terrestrial and aquatic food webs (4). Thus, a robust backbone phylogenetic framework is needed to place ecological studies in their evolutionary context and for use in comparative genomic and macroevolutionary analyses.

Despite their importance, relatively few studies have addressed the relationships among the major groups of hemipteroid insects (Paraneoptera, *sensu stricto* [excluding Zoraptera], also termed Acercaria). While a recent large transcriptome-based phylogenomic analysis of insects (1) provided a well resolved and strongly supported phylogenetic framework for the insect orders in general, it did not sample intensively within individual orders and recovered some unexpected relationships. Among the most puzzling was the non-monophyly of the hemipteroid insects, with Psocodea as the sister taxon of holometabolous insects rather than as sister to thrips plus hemipterans (Condylognatha). Although this result was congruent with one earlier analysis based on three nuclear protein-coding genes (5), it had not been proposed in other molecular phylogenetic or morphological studies. Previous morphological studies indicated monophyly of hemipteroid insects with Psocodea sister to thrips plus hemipterans (6–9); or sometimes a group comprising thrips plus Psocodea (10, 11).

Another unexpected relationship recovered by Misof et al. (1) was the placement of moss bugs (Coleorrhyncha) as sister to a group comprising leafhoppers, cicadas, and relatives (Auchenorrhyncha) instead of sister to true bugs (Heteroptera). A recent morphological study also found some support for moss bugs sister to Auchenorrhyncha (12). In contrast, prior analyses based on morphology (e.g. 9) and DNA sequence data (e.g. 13) consistently placed

moss bugs as sister to true bugs. An analysis of a reduced gene set from transcriptome data (14) also recovered moss bugs as sister to true bugs, while the full gene set placed moss bugs as sister to Auchenorrhyncha. Analysis of mitochondrial genomes (15) produced an even more unconventional result, with moss bugs placed as the sister taxon of planthoppers (Fulgoroidea), making Auchenorrhyncha paraphyletic. Thus, it is important to investigate the placement of moss bugs in more detail with both expanded taxon and gene sampling.

We evaluated these possible conflicts among analyses by analyzing a more comprehensive dataset comprising an increased number of clusters of orthologous sequence groups (2,395 protein-coding, single-copy genes) as well as an increased taxon sample within hemipteroid insects: 160 samples vs. 22 sampled by Misof et al. (1). We included representatives of all major hemipteroid lineages (sub- and infraorders). Outgroups comprised 33 species of holometabolous and non-holometabolous insect orders. This data set enabled us to test the hypothesis of non-monophyly of hemipteroid insects and also provides a more detailed backbone framework for the hemipteroid phylogeny. We evaluate the implications of this phylogeny for understanding the evolution of feeding strategy, morphology, and mitochondrial genome organization of this major group of insects.

Results

Phylogeny of Hemipteroid Insect Orders

Separate amino-acid sequence alignments of the 2,395 single-copy genes across 193 terminal taxa (SI Appendix, Tables S1-S4) yielded a concatenated supermatrix of 859,518 aligned amino-acid positions, which was used in subsequent phylogenetic analyses. A concatenated nucleotide sequence supermatrix of only first and second codon positions resulted

in ~1.72 million aligned nucleotide sequence sites. Tree reconstructions based on the nucleotide sequence data supported a phylogenetic tree (Fig. 1, SI Appendix: Figs. S1 - S2) with 172/190 (~90%) of all nodes supported in 100% of bootstrap replicates. The tree based on amino-acid sequence data (SI Appendix: Fig. S3) was highly concordant with that based on nucleotide data. Analysis of an optimized amino-acid dataset (see Materials and Methods) produced a tree (SI Appendix: Fig. S4) that was identical to that based on all amino-acids with respect to relationships among orders, suborders, infraorders, and superfamilies, but had some minor rearrangements within these groups.

Considering relationships within and among orders in more detail, the thrips (Thysanoptera) were recovered with 100% bootstrap support as the sister taxon of Hemiptera (i.e., monophyletic Condylognatha), although only 68% of quartets supported this result in Four-cluster Likelihood Mapping (FcLM; SI Appendix: Tables S5 - S6). As in the study of Misof et al. (1), Psocodea were placed as the sister taxon of Holometabola in 100% of bootstrap replicates, rendering hemipteroid insects paraphyletic. However, only 25% of quartets supported Psocodea as sister to Holometabola, compared to 67% of the quartets supporting hemipteroid insect monophyly. Results from the FcLM imply that the placement of Psocodea as sister to Holometabola is unstable and may be due to confounding phylogenetic signal (e.g., from heterogeneous composition of amino-acid sequences, non-stationarity of substitution processes, or non-random distribution of missing data) and is also dependent on the taxon sample. However, permutation tests of these results suggested the impact of these potential confounding signals on the topology was minor (SI Appendix: Table S6). To evaluate whether the parasitic lice in particular (Phthiraptera), which have elevated substitution rates compared to other hemipteroids (16), were a possible source of conflicting signal, we compared quartets with and

without these ectoparasitic insects as the representative of Psocodea. However, the support from FcLM for monophyly of hemipteroid insects was highly similar whether parasitic lice were included (66%) or not (67%).

Morphological character mapping over three possible alternative topologies (SI Appendix: Fig. S5) revealed no apomorphies supporting Psocodea + Holometabola. In contrast, there are 14 potential apomorphies for the monophyly of Paraneoptera. These results indicate that there is more agreement between morphology and the FcLM results, as compared to the supermatrix analyses with all taxa. For Coleorrhyncha (moss bugs), three characters are apomorphies for a sister relationship to Auchenorrhyncha (leafhoppers and relatives) but two other characters appear to support a sister relationship to Heteroptera (true bugs).

In general, the phylogenetic results from transcriptomes are congruent with the generally accepted classification schemes within these insect orders. Bark lice and parasitic lice (Psocodea) together are monophyletic. As has been suggested based on both morphological (17) and molecular (16, 18) analyses, the parasitic lice are embedded within free-living bark lice, being the sister taxon of book lice (Liposcelididae), which makes the bark lice ("Psocoptera") paraphyletic. In contrast to results based on 18S rDNA sequences (18), parasitic lice (Phthiraptera) were supported as a monophyletic group in our analyses, which included representatives of all four suborders of parasitic lice.

The thrips (Thysanoptera) were found to be monophyletic. The thrips family Phlaeothripidae was recovered as the sister taxon to the remaining thrips (Aeolothripidae + Thripidae), congruent with previous molecular analyses and the current classification of Thysanoptera into the suborders Tubulifera (i.e. Phlaeothripidae) and Terebrantia (all other thrips) (19).

The order Hemiptera was also monophyletic. Within Hemiptera, Sternorrhyncha (whiteflies, psyllids, scales, and aphids) was recovered as the sister taxon of the remaining hemipterans. Recent classification schemes (20) and prior molecular studies (13, 21) have placed the enigmatic moss bugs as the sister taxon of true bugs. However, our results recover moss bugs as the sister taxon of Auchenorrhyncha (leafhoppers, planthoppers, and relatives), which was also found by Misof et al. (1). In FcLM analyses, 96% of quartets placed moss bugs with Auchenorrhyncha, suggesting little underlying conflict in the data for this result (Table S6).

Within Sternorrhyncha, whiteflies (Aleyrodoidea) were sister to the remainder of the suborder, and psyllids (Psylloidea) were sister to a clade composed of aphids (Aphidoidea) + scale insects (Coccoidea), also supported by 91% of quartets in FcLM analyses. Previous phylogenetic analyses of Sternorrhyncha have tended to focus within particular superfamilies or families (e.g. 22–24) rather than addressing relationships among major lineages (superfamilies).

The earliest molecular phylogenetic analyses of Hemiptera (e.g. 25, 26) failed to recover Auchenorrhyncha as a monophyletic group, as has a more recent analysis of mitochondrial genomes (15). However, our analyses provided strong support for monophyly of this group, corroborating results of other studies based on multiple loci (13, 14). Within Auchenorrhyncha, our results strongly support the taxonomic status of the two recognized infraorders Fulgoromorpha (i.e. Fulgoroidea, planthoppers) and Cicadomorpha (leafhoppers/treehoppers, spittlebugs, and cicadas) as monophyletic, as found previously (13). However, relationships among the three superfamilies of Cicadomorpha were inconsistently resolved. Cicadas (Cicadoidea) plus spittlebugs (Cercopoidea) were sister to leafhoppers/treehoppers (Membracoidea) in the analysis of nucleotide sequences (Fig. 1, FcLM 52% of quartets), but cicadas were sister to spittlebugs plus leafhoppers/treehoppers in the analysis of amino-acid

sequence data (SI Appendix: Fig. S1), which was also found in 48% of quartets of nucleotide data in FcLM analyses.

Relationships among the earlier diverging lineages of true bugs (Heteroptera) have not been resolved consistently across previous analyses (14, 27–29), in which the deepest divergences received low statistical branch support and recovered different relationships among infraorders. In our analysis, which included representatives of all seven currently recognized infraorders, the four infraorders for which more than one species was included were found to be monophyletic. Like two recent studies based on combined molecular and morphological data (29) and transcriptome data (14), we found 100% bootstrap support for 1) a clade comprising litter bugs (Dipsocoromorpha), unique-headed bugs (Enicocephalomorpha), and semi-aquatic bugs (Gerromorpha) (also found in 100% of quartets in FcLM analyses) and 2) shore bugs (Leptopodomorpha) as the sister to Cimicomorpha + Pentatomomorpha (also found in 100% of quartets in FcLM analyses).

Divergence Time Analysis

The estimate of the root age for our tree, the split between Palaeoptera (dragonflies, damselflies, and mayflies) and Neoptera (all other insects) at 437 million years ago (mya) (95% CI 401-486) was only slightly older than that estimated for this node by Misof et al. (1), at 406 mya. Divergence dates for more interior nodes tended to be older than those estimated by Misof et al. (1) and more similar to those of Tong et al. (30), possibly due either to much denser sampling of minimum age fossil calibration points throughout this part of the insect tree or to different methodology (e.g., MCMCtree versus BEAST; or different prior distributions of expected ages for Bayesian analyses). Analyses of divergence times postulated a common

ancestor of thrips and hemipterans as early as the Devonian (~407 mya, 95% CI 373-451). Radiation within Hemiptera is also inferred to have begun in this period (~386 mya, 95% CI 354-427), with radiations within Sternorrhyncha, Auchenorrhyncha, and Heteroptera having commenced by the late Carboniferous (all before 300 mya). Radiation within modern Psocodea dates to the Carboniferous (328 mya, 95% CI 292-376), with divergence of this lineage from other insects as early as 404 mya (95% CI 367-451).

Discussion

Analysis of 2,395 protein-coding, single-copy genes derived from transcriptomes of hemipteroid insects and outgroups provided strong support for a backbone tree of hemipteroid insects largely congruent with previous analyses and classification schemes. In particular, we recovered with strong support monophyly of the three orders of hemipteroid insects: Psocodea, Thysanoptera, and Hemiptera. We also recovered monophyly of most currently recognized suborders, infraorders, and superfamilies within these groups as well as resolving relationships among these major groups. Although the unconventional result of a sister relationship between Psocodea and Holometabola of Misof et al. (1) appeared to be robust to our substantially increased taxon sampling based on maximum likelihood bootstrapping, it was not supported by Four-cluster Likelihood Mapping analyses. FcLM, which can detect potentially confounding signal, suggests extensive underlying conflict for this result, with the majority of quartets placing Psocodea with thrips and hemipterans, which would imply monophyly of Paraneoptera in rooted trees. However, permutations appear to rule out several possible types of confounding signal (e.g. among-lineage heterogeneity or non-random distribution of missing data) in our dataset. Recent work has suggested that bootstrap support from very large data sets may provide an

overestimate of confidence for phylogenetic results (31–33). Thus, the position of Psocodea in the insect tree is still an open question. Monophyly of hemipteroid insects is supported by several morphological autapomorphies (34); therefore, non-monophyly of the group would imply homoplasy in these traits. In addition, there is no known morphological apomorphy supporting Psocodea + Holometabola (SI Appendix: Fig. S5). In contrast, the other less conventional relationship, a clade comprising Coleorrhyncha and Auchenorrhyncha uncovered by Misof et al. (1), was recovered by our trees with increased taxon sampling and is supported by 96% of quartets in the FcLM analyses and three morphological apomorphies, suggesting that this result is robust.

Divergence time estimates using a dense sampling of 23 fossil calibration points suggest that the radiation of the hemipteroid insect orders is relatively ancient, beginning before the early Carboniferous, considerably older than initial expectations based on available fossils. However, the insect fossil record of this period is extremely fragmentary, and relatively old fossils of modern lineages that are used as calibration points imply that branches uniting these lineages must be older still, given that fossil ages represent minimum ages.

Implications for Evolution of Feeding Strategy

Our phylogenetic results generally agree with evidence from the fossil record that the earliest hemipteroids fed on detritus, pollen, fungi, or spores (as in most modern barklice and thrips). Plant-fluid feeding probably coincided with the origin of Hemiptera and was independently derived in thrips. Today, Hemiptera is the fifth largest insect order, surpassed only by the four major holometabolous orders (Hymenoptera, Coleoptera, Lepidoptera, and Diptera). It remains one of the most abundant and diverse groups of plant-feeding insects. Within

Hemiptera, the origin of true bugs apparently coincided with a shift from herbivory to predation, with subsequent shifts back to herbivory (29, 35) in the more derived lineages (Pentatomomorpha and Cimicomorpha). The two other large suborders of Hemiptera (Auchenorrhyncha and Sternorrhyncha) feed almost exclusively on vascular plant fluids.

Our results also suggest that the earliest hemipterans fed preferentially on phloem. Phloem feeding remains predominant in extant plant-feeding hemipterans, including nearly all Sternorrhyncha and most Auchenorrhyncha (36), while modern moss bugs feed on phloem-like tissues in mosses (37). A shift to xylem feeding appears to have coincided with the origin of Cicadomorpha (at least the crown group of this lineage), in which all cicadas and spittlebugs retain this preference. This is also supported by the fossil record in which the earliest leafhoppers had inflated faces (38), indicating a preference for xylem feeding, despite the predominance of phloem feeding among modern leafhoppers and treehoppers (Membracoidea). A shift to phloem feeding apparently occurred early in the evolution of Membracoidea but at least one reversal to xylem feeding (in Cicadellinae–sharpshooters) has been inferred previously (39), consistent with our results.

Implications for Morphological Evolution

Based on the conflicting statistical support between the supermatrix analysis and Four-cluster Likelihood Mapping, the position of lice (Psocodea) appears to be unstable. Morphological evidence, in contrast, supports the monophyly of hemipteroid insects (Paraneoptera). Our parsimony mapping of 142 morphological characters (SI Appendix: Fig. S5) found no apomorphies supporting Psocodea + Holometabola but 14 apomorphies supporting hemipteroid insect monophyly. Some of these are reductions or losses, including the reduced

number of tarsomeres (three in modern hemipteroids), reduced number of Malpighian tubules (four), and presence of only one abdominal ganglionic complex. Nevertheless, these characters, together with characters of the forewing base, still appear to support the sister group relationship between Psocodea and thrips plus hemipterans (11, 34, 40). Thus, the phylogenetic position of Psocodea requires further study of morphological and molecular data.

In contrast to the equivocal support for Paraneoptera, Condylognatha is strongly supported not only in the phylogenomic analyses, but also with six morphological apomorphies. The origin of this group apparently coincided with a distinct shift in mouthpart morphology and feeding habits toward piercing and sucking. These changes include anterior shifting of tentorial pits, elongated and slender mandibles, stylet-like laciniae, and a narrowed labium (SI Appendix: Fig. S5). Subsequent evolutionary transformations led to the very distinct and unique piercing-sucking mouthparts of hemipterans that facilitate ingestion of liquid from plant or animal tissues.

The sister-group relationship that we found between moss bugs (Coleorrhyncha) and Auchenorrhyncha has not, to our knowledge, been proposed previously in any explicit phylogenetic analysis other than in recent phylogenomic analyses of transcriptomes (1, 14). Traditionally, moss bugs were treated as one of three suborders of "Homoptera" (along with Sternorrhyncha and Auchenorrhyncha), largely based on the structure of the head. The mouthparts of moss bugs arise posteroventrally (41), as in leafhoppers and relatives, rather than anteriorly as in true bugs (42). Nevertheless, morphological evidence from fossil and living moss bugs, primarily from wing structure and musculature, suggested a closer relationship to true bugs (9, 41, 43). However, a recent comparative morphological study (12) revealed that moss bugs share a unique derived feature of the wing base with Auchenorrhyncha; a membranous proximal median plate. The same study also showed that some previously suggested morphological

synapomorphies of moss bugs and true bugs (SI Appendix: Fig. S5C) are either ambiguous or have been misinterpreted (12). Prior molecular evidence supporting moss bugs plus true bugs was also somewhat equivocal (13: ML bootstrap 83% and MP bootstrap 63%). Our results support those of other transcriptome studies (1, 14) in placing Coleorrhyncha sister to Auchenorrhyncha.

Implications for Evolution of Mitochondrial Genome Organization

Several groups of hemipteroid insects have been shown to have highly rearranged mitochondrial genomes (2). The sister relationship between thrips and hemipterans indicates that the heightened rates of mitochondrial (mt) genome rearrangements observed in the lice (44) and thrips (45) are the result of convergence between these two clades. Even if Psocodea is sister to thrips plus hemipterans, and not to holometabolous insects, recent analyses indicating that the ancestor of all Psocodea had a generally standard insect mitochondrial gene order still result in an interpretation involving convergence (46). This phylogenetic evidence is also consistent with the absence of any shared, derived gene arrangements between Psocodea and thrips, as both have independently diverged from the inferred ancestral insect mt genome arrangement (2, 45).

An interpretation involving convergence is also consistent with the varying degrees of rearrangement observed within each order. Within Psocodea, mt genomes vary wildly across different taxonomic scales, from a single derived arrangement found in all Psocomorpha (46), to wide variation within a single genus (*Liposcelis*, 47), and between closely related species of parasitic lice. In contrast, for the thrips, mitochondrial genome arrangements are relatively consistent at the family level (with only tRNA rearrangements observed), albeit still highly rearranged relative to the ancestral insect mt genome (48). Very few rearrangements of any type

are observed in the Hemiptera, with the vast majority of families possessing the inferred ancestral arrangement (2).

In summary, although the exact phylogenetic position of Psocodea remains to be resolved convincingly, our results based on transcriptomes for hemipteroid insects provide a strong new phylogenetic framework for future studies of genomic, morphological, ecological, and behavioral characteristics of this important group of insects.

Materials and Methods

Our general approach closely followed methods described previously by Misof et al. (1) and Peters et al. (49) for phylogenomic analyses of insect transcriptomes (SI Appendix, Dryad accession pending acceptance). Transcriptomes of 140 samples of Paraneoptera were newly sequenced with 100 bp paired-end reads for this study using Illumina HiSeq2000 or HiSeq2500 machines to achieve at least 2.5 Gbp per taxon. The final taxon sample of 193 includes representatives of 97 hemipteroid families with several larger families represented by multiple subfamily representatives.

All paired-end reads were assembled with SOAPdenovo-Trans (version 1.01; 50) and the assembled transcripts were filtered for possible contaminants (SI Appendix: Table S2) as described in Peters et al. (49). The raw reads and filtered assemblies were submitted to the NCBI SRA and TSA archives (SI Appendix: Table S1). We searched the assemblies for transcripts of 2,395 protein-coding genes that the OrthoDB v7 database (51) suggested to be single-copy across the genomes of six species (SI Appendix: Table S3) using the software Orthograph (version beta4, 52; results of orthology search see Table S4). Orthologous transcripts were aligned with MAFFT (version 7.123; 53) at the translational (amino acid) level. Corresponding

nucleotide MSAs were generated with a modified version of the software Pal2Nal (54) (version 14).

Alignment sections that could not be discriminated from randomly aligned regions at the amino-acid level of each gene were identified with Aliscore version 1.2 (55, 56). To maximize the fit of our substitution models, we identified for each gene the protein domains (clans, families) and unannotated regions using the Pfam database (1, 57; Supplemental Materials and Methods). The phylogenetic information content of each data block was assessed with MARE (version 0.1.2-rc) (58), and all uninformative data blocks (IC=0) were removed. We subsequently used PartitionFinder (developer version 2.0.0-pre14, 59) to simultaneously infer the best partitioning scheme and amino acid or nucleotide (removing third positions because of heterogeneity, SI Appendix: Fig. S6) substitution models, using the rclusterf algorithm.

Phylogenetic trees were inferred using a Maximum Likelihood approach with ExaML vers. 3.0.17 (60) for both the nucleotide and amino-acid data sets. We performed 50 non-parametric bootstrap replicates mapping the support on the best ML tree after checking for bootstrap convergence with the default bootstopping criteria (61). An optimized dataset, which requires the presence of at least one species from a given taxonomic group (SI Appendix: Table S5) in each data block of the supermatrix (62), was used for testing the possible impact of missing data at the partition level. Four-cluster likelihood mapping (63) was used for assessing the phylogenetic signal for alternative phylogenetic relationships (SI Appendix: Tables S5 - S6). Permutation tests in these analyses assessed the impact of heterogeneous amino-acid sequence composition among lineages, non-stationarity of substitution processes, and non-random distribution of missing data on the inferred phylogenetic tree (1).

To understand the morphological transformations underlying the evolution of the hemipteroid groups and to identify potential shared derived characters (synapomorphies), we used the morphological data matrix of Friedemann et al. (9) with 118 characters of the entire body (with modifications from 14) and additionally 25 characters associated with the wing base (8). By tracing characters over the tree using maximum parsimony using Winclada (64), we evaluated three possible phylogenetic alternatives: 1) paraphyletic Paraneoptera and Coleorrhyncha sister to Auchenorrhyncha (result from ML analysis of transcriptomes), 2) monophyletic Paraneoptera (as suggested by FcLM analyses), and 3) paraphyletic Paraneoptera, but with Coleorrhyncha sister to Heteroptera (as suggested in previous literature).

To estimate divergence dates, we used the topology resulting from ML analysis of first and second position nucleotides as the input tree and assigned 23 ingroup fossil calibration points (65) throughout the tree (SI Appendix: Table S7). These calibrations were used as minimum ages in soft bound uniform priors with a root age of 406 mya (1) as a soft bound maximum. These priors were used in a Bayesian MCMCTree (66) molecular dating analysis of a first and second position nucleotide data set for which sites were present in at least 95% of taxa.

Acknowledgments

Data reported in this paper is deposited in NCBI (SI Appendix: see Table S1) and Dryad (accession number upon acceptance). We thank E. Anton, M. Bowser, C. Bramer, T. Catanach, D. H. Clayton, J. R. Cooley, G. Gibbs, A. Hansen, E. Hdez, A. Katz, K. Kjer, J. Light, A. Melber, B. Morris, D. Papura, H. Pohl, R. Rakitov, C. Ray, S. Schneider, K. Schütte, W. Smith, K-Q. Song, T. Sota, N. Szucsich, G. Taylor, S. Taylor, S. Tiwari, and X. Tong for assistance with obtaining specimens. We thank G. Meng and BGI staff for their efforts in data curation and

O. Niehuis for assistance preparing the ortholog gene set. RMW was supported by Swiss National Science Foundation grant PP00P3_1706642. KM was supported by David Yeates, the Schlinger Endowment, CSIRO NRC Australia, J. Korb, and University of Freiburg. This work was also supported by NSF DEB-1239788 to KPJ, CHD, and HMR.

References

1. Misof B, et al. (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346:763-767.

2. Cameron, SL (2014) Insect mitochondrial genomics: Implications for evolution and phylogeny. *Ann Rev Entomol* 59:95-117.

3. Adis J, Lubin YD, Montgomery GG (1984) Arthropods from the canopy of inundated and terra firme forests near Manaus, Brazil, with critical considerations on the pyrethrum-fogging technique. *Studies Neotrop Fauna Environ* 19:223-236.

4. Schaefer CW, Panizzi AR (2000) *Heteroptera of Economic Importance*. CRC Press, Boca Raton, New York.

5. Ishiwata K, Sasaki G, Ogawa J, Miyata T, Su Z-H (2011) Phylogenetic relationships among insect orders based on three nuclear protein-coding gene sequences. *Molecular Phylog Evol* 58:169-180.

6. Beutel RG, Gorb SN (2001) Ultrastructure of attachment specializations of hexapods (Arthropoda): evolutionary patterns inferred from a revised ordinal phylogeny. *J Zool Syst Evol Res* 39:177-207.

7. Wheeler WC, Whiting M, Wheeler QD, Carpenter JM (2001) The phylogeny of extant hexapod orders. *Cladistics* 17: 113-169.

8. Yoshizawa K, Saigusa T (2001) Phylogenetic analysis of paraneopteran orders (Insecta: Neoptera) based on forewing base structure, with comments on the monophyly of Auchenorrhyncha (Hemiptera). *Syst Ent* 26:1-13.

9. Friedemann K, Spangenberg R, Yoshizawa K, Beutel RG (2014) Evolution of attachment structure in the highly diverse Acercaria (Hexapoda). *Cladistics* 30:170–201.

10. Whiting MF, Carpenter JC, Wheeler QD, Wheeler WC (1997) The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Syst Biol* 46:1-68.

11. Kristensen NP (1991) Phylogeny of extant hexapods pp. 125-140. In: CSIRO *The Insects of Australia.* Melbourne University Press.

12. Yoshizawa K, Ogawa N, Dietrich CH (2017) Wing base structure supports Coleorrhyncha + Auchenorrhyncha (Insecta: Hemiptera). *J Zool Syst Evol Res* 55:199-207.

13. Cryan JR, Urban JM (2012) Higher-level phylogeny of the insect order Hemiptera: is Auchenorrhyncha really paraphyletic? *Syst Ent* 37:7-21.

14. Wang Y-H, et al. (2017) When did the ancestor of true bugs become stinky? Disentangling the phylogenomics of Hemiptera-Heteroptera. *Cladistics* In press.

15. Li H, et al. (2017) Mitochondrial phylogenomics of Hemiptera reveals adaptive innovations driving the diversification of true bugs. *Proc Roy Soc Lond* B 284:20171223.

16. Yoshizawa K., Johnson KP (2013) Changes in base composition bias of nuclear and mitochondrial genes in lice (Insecta: Psocodea). *Genetica* 141:491-499.

17. Lyal CHC (1985) Phylogeny and classification of the Psocodea, with particular reference to the lice (Psocodea: Phthiraptera). *Syst Ent* 10:145-165.

18. Johnson K, Yoshizawa K, Smith VS (2004) Multiple origins of parasitism in lice. *Proc Roy Soc Lond* B 271:1771-1776.

19. Buckman RS, Mound LA, Whiting MF (2013) Phylogeny of thrips (Insecta: Thysanoptera) based on five molecular loci. *Syst Ent* 38:123-133.

20. Bourgoin T, Campbell BC (2002) Inferring a phylogeny for Hemiptera: falling into the 'autapomorphic trap'. *Denisia* 4:67–82.

21. Ouvrard D, Campbell BC, Bourgoin T, Chan KL (2000) 18S rRNA secondary structure and phylogenetic position of Peloridiidae (Insecta, Hemiptera). *Mol Phylog Evol* 16:403-417.

22. Von Dohlen CD, Moran NA (2000) Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation. *Biol J Linn Soc* 71:689-717.

23. Gullan PJ, Cook LG (2007) Phylogeny and higher classification of the scale insects (Hemiptera: Sternorrhyncha: Coccoidea). *Zootaxa* 1668:413-425.

24. Percy DM et al. (2018) Resolving the psyllid tree of life: phylogenomic analyses of the superfamily Psylloidea (Hemiptera). *Syst Ent* 43:762-776.

25. Campbell BC, Steffen-Campbell JD, Sorensen HT, Gill RJ (1995) Paraphyly of Homoptera and Auchenorrhyncha inferred from 18S rDNA nucleotide sequences. *Syst Ent* 20:175-194.

26. von Dohlen CD, Moran NA (1995) Molecular phylogeny of the Homoptera: A paraphyletic taxon. *J Mol Evol* 41:211-223.

27. Li H, et al. (2012) The complete mitochondrial genome and novel gene arrangement of the unique-headed bug *Stenopirates* sp. (Hemiptera: Enicocephalidae). *PLoS One* 7:e29419.

28. Weirauch C, Štys P (2014) Litter bugs exposed: phylogenetic relationships of Dipsocoromorpha (Hemiptera: Heteroptera) based on molecular data. *Insect Syst Evol* 45:351-370.

29. Weirauch C, Schuh RT, Cassis G, Wheeler WC (2018) Revisiting habitat and lifestyle transitions in Heteroptera (Insecta: Hemiptera): insights from combined morphological and molecular phylogeny. *Cladistics* In press.

30. Tong KJ, Duchêne S, Ho SYW, Lo N (2015) Comment on "Phylogenomics resolves the timing and pattern of insect evolution." *Science* 349:487.

31. Salichos L, Rokas A (2013) Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* 497:327–331.

32. Salichos L, Rokas A, Stamatakis A (2016) Computing the internode certainty and related measures from partial gene trees. *Mol Biol Evol* 33:1606–1617.

33. Shen X-X, Hittinger CT, Rokas A (2017) Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nature Ecol Evol* 1:0126.

34. Yoshizawa K, Lienhard C (2016) Bridging the gap between chewing and sucking in the hemipteroid insects: insights from Cretaceous amber. *Zootaxa* 4979:229-245.

35. Cobben RH (1979) On the original feeding habits of the Hemiptera (Insecta): a reply to Merrill Sweet. *Ann Ent Soc Am* 72:711–715.

36. Backus EA (1988) Sensory systems and behaviors which mediate hemipteran plant-feeding: a taxonomic overview. *J Ins Physiol* 34:151-165.

37. Cronk QCB, Forest F (2017) The evolution of Angiosperm trees: From Palaeobotany to genomics. Pp. 1-17 In: *Comparative and Evolutionary Genomics of Angiosperm Trees* (A. Groover and Q. Cronk, eds.). Springer, New York, NY.

38. Shcherbakov D (1996) Origin and evolution of the Auchenorrhyncha as shown by the fossil record. In: Schaeffer CW (ed.) *Studies on Hemipteran phylogeny*. Entomol Soc Am, Lanham, MD, USA.

39. Dietrich CH, et al. (2017) Anchored enrichment-based phylogenomics of leafhoppers and treehoppers (Hemiptera: Cicadomorpha: Membracoidea). *Insect Syst Diver* 1: 57-72.

40. Beutel RG, Friedrich F, Ge S-Q Yang X-K (2014) *Insect Morphology and Phylogeny: A Textbook for Students of Entomology*. Walter de Gruyter.

41. Spangenberg R, et al. (2013) The cephalic morphology of the Gondwanan key taxon *Hackeriella* (Coleorrhyncha, Hemiptera). *Arthr Str Dev* 42:315-337.

42. Spangenberg R, Friedemann K, Weirauch C, Beutel RG (2013) The head morphology of the potentially basal heteropteran lineages Enicocephalomorpha and Dipsocoromorpha (Insecta: Hemiptera: Heteroptera). *Arthropod Syst Phy* 71:103-136.

43. Shcherbakov D, Popov YA (2002) Superorder Cimicidea Laicharting, 1781, Order Hemiptera Linné, 1758. The bugs, cicadas, plantlice, scale insects, etc. Pp. 143-157. In, Rasnitsyn A P, Quicke DLJ (eds.) *History of Insects*. Kluwer, Dordrecht.

44. Cameron SL, Yoshizawa K, Mizukoshi A, Whiting MF, Johnson KP (2011) Mitochondrial genome deletions and mini-circles are common in lice (Insecta: Phthiraptera). *BMC Genomics* 12:394.

45. Dickey AM, et al. (2015) A novel mitochondrial genome architecture in thrips (Insecta: Thysanoptera): Extreme size asymmetry among chromosomes and possible recent control region duplication. *BMC Genomics* 16:439.

46. Yoshizawa K, et al. (2018) Mitochondrial phylogenomics and rearrangements in barklice (Insecta: Psocodea). *Mol Phylog Evol* 119:118-127.

47. Shi Y, et al. (2016) The mitochondrial genome of booklouse, *Liposcelis sculptilis* (Psocoptera: Liposcelididae) and the evolutionary timescale of *Liposcelis*. *Sci Rep* 6:30660.

48. Yan D, et al. (2014) The mitochondrial genome of *Frankliniella intosa*: Insights into the evolution of mitochondrial genomes at lower taxonomic levels in Thysanoptera. *Genomics* 104:306-312.

49. Peters RS, et al. (2017) Evolutionary history of the Hymenoptera. *Current Biol* 27:1013-1018.

50. Xie Y, et al. (2014). SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinf* 30:1660–1666.

51. Kriventseva EV, Rahman N, Espinosa O, Zdobnov EM (2008) OrthoDB: the hierarchical catalog of eukaryotic orthologs. *Nuc Acids Res* 36:D271-5.

52. Petersen M, et al. (2017) Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. *BMC Bioinf* 18:111.

53. Katoh K, Standley DM (2016) A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics* 32:1933-1942.

54. Suyama M, Torrents D, Bork P (2006) PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nuc Acids Res* 34:W609–W612.

55. Misof B, Misof K (2009) A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. *Syst Biol* 58:21–34.

56. Kück P, et al. (2010) Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Front Zool* 7:10.

57. Finn RD, et al. (2014) Pfam: the protein families database. *Nuc Acids Res*. 42:D222–D230.

58. Misof B, et al. (2013) Selecting informative subsets of sparse supermatrices increases the chance to find correct trees. *BMC Bioinf* 14:348.

59. Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A (2014) Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evol Biol* 14:82.

60. Kozlov AM, Aberer AJ, Stamatakis A (2015) ExaML version 3: a tool for phylogenomic analyses on supercomputers. *Bioinf* 31:2577–2579.

61. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A (2010) How many bootstrap replicates are necessary? *J Comp Biol* 17:337-354.

62. Dell'Ampio E, et al. (2014) Decisive data sets in phylogenomics: Lessons from studies on the phylogenetic relationships of primarily wingless insects. *Mol Biol Evol* 31:239–249.

63. Strimmer K, von Haeseler A (1997) Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *PNAS* 94:6815–6819.

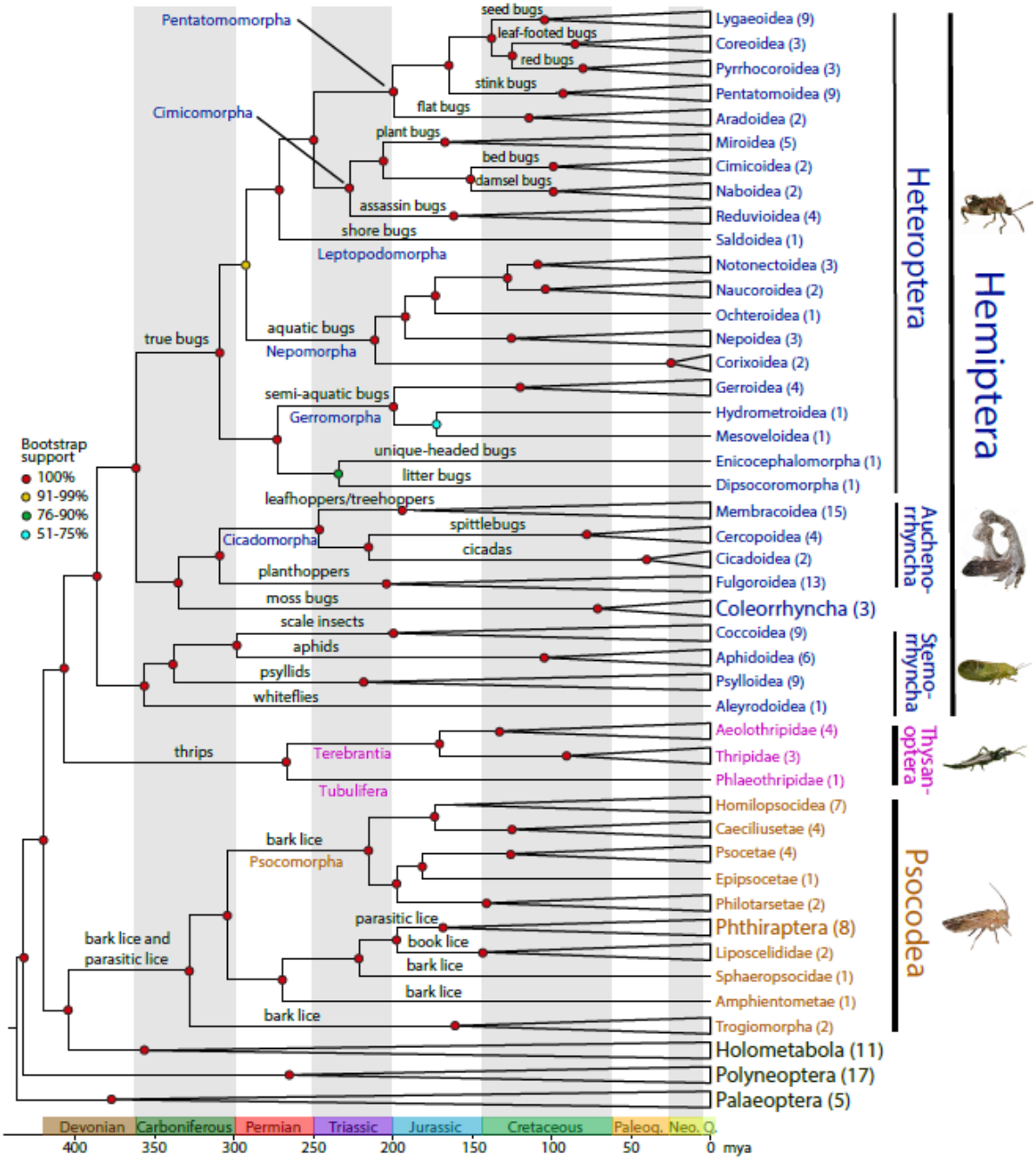
64. Nixon K (2002) Winclada ver 1.00. 08. Ithaca, NY.

65. Parham JF, et al. (2012) Best practices for justifying fossil calibrations. *Syst Biol* 61:346-359.

66. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586-1591.

Figure Legends

Figure 1. Dated phylogeny of hemipteroid insects (Hemiptera, Thysanoptera, Psocodea) based on maximum likelihood analysis of a supermatrix of first and second codon position nucleotides corresponding to 859,518 aligned amino-acid positions from transcriptome or genome sequences of 193 samples. Colored circles indicate bootstrap support. Timescale in millions of years (bottom) estimated from MCMCTree Bayesian divergence time analyses using 23 fossil calibration points and a reduced dataset. Number of species sampled from each group indicated in parentheses. Higher taxa indicated as taxon labels and below branches; most convenient generalized common names above branches. Images represent five major groups: Heteroptera, Auchenorrhyncha, Sternorrhyncha, Thysanoptera, and Psocodea.


Pentatomomorpha
Cimicomorpha
true bugs
Nepomorpha
Gerromorpha
Cicadomorpha
Terebrantia
Tubulifera
Psocomorpha
bark lice and parasitic lice
seed bugs
leaf-footed bugs
red bugs
stink bugs
flat bugs
plant bugs
bed bugs
damsel bugs
assassin bugs
shore bugs
Leptopodomorpha
aquatic bugs
semi-aquatic bugs
unique-headed bugs
litter bugs
leafhoppers/treehoppers
spittlebugs
cicadas
planthoppers
moss bugs
scale insects
aphids
psyllids
whiteflies
thrips
bark lice
parasitic lice
book lice
bark lice
bark lice
bark lice
Bootstrap support
100%
91-99%
76-90%
51-75%
Lygaeoidea (9)
Coreoidea (3)
Pyrrhocoroidea (3)
Pentatomoidea (9)
Aradoidea (2)
Miroidea (5)
Cimicoidea (2)
Naboidea (2)
Reduvioidea (4)
Saldoidea (1)
Notonectoidea (3)
Naucoroidea (2)
Ochteroidea (1)
Nepoidea (3)
Corixoidea (2)
Gerroidea (4)
Hydrometroidea (1)
Mesoveloidea (1)
Enicocephalomorpha (1)
Dipsocoromorpha (1)
Membracoidea (15)
Cercopoidea (4)
Cicadoidea (2)
Fulgoroidea (13)
Coleorrhyncha (3)
Coccoidea (9)
Aphidoidea (6)
Psylloidea (9)
Aleyrodoidea (1)
Aeolothripidae (4)
Thripidae (3)
Phlaeothripidae (1)
Homilopsocidea (7)
Caeciliusetae (4)
Psocetae (4)
Epipsocetae (1)
Philotarsetae (2)
Phthiraptera (8)
Liposcelididae (2)
Sphaeropsocidae (1)
Amphientometae (1)
Trogiomorpha (2)
Holometabola (11)
Polyneoptera (17)
Palaeoptera (5)
Heteroptera
Auchenorrhyncha
Sternorrhyncha
Thysanoptera
Psocodea
Hemiptera
Devonian
Carboniferous
Permian
Triassic
Jurassic
Cretaceous
Paleog.
Neo. Q.
400
350
300
250
200
150
100
50
0 mya