Phylogenomics and the evolution of hemipteroid insects

Kevin P. Johnson¹,1, Christopher H. Dietrich⁵, Frank Friedrich⁴, Rolf G. Beutel¹, Benjamin Wipfler⁵, Ralph S. Peters⁶, Juliane M. Allen⁵, Malte Petersen⁶, Alexander Donath⁷, Kimberly K. O. Walden⁸, Alexey M. Kozlov⁹, Lars Podsiadlowski¹⁰, Christoph Mayer¹¹, Karen Meusemann¹²,¹³, Alexey M. Kozlov, Lars Podsiadlowski, Robert M. Waterhouse, Stephen L. Cameron, Christiane Weirauch, Daniel R. Swanson, Diana M. Percy, Nate B. Hardy, Irene Terry, Shalin Liu, Xin Zhou, Bernhard Misof, Hugh M. Robertson, and Kazunori Yoshizawa

¹Illinois Natural History Survey, Prairie Research Institute, University of Illinois at Urbana–Champaign, Champaign, IL 61820; ²Institut für Zoologie, Universität Hamburg, 20146 Hamburg, Germany; ³Institut für Zoologie und Evolutionsforschung, Friedrich-Schiller-Universität Jena, 07743 Jena, Germany; ⁴Center of Taxonomy and Evolutionary Research, Arthropoda Department, Zoological Research Museum Alexander Koenig, 53113 Bonn, Germany; ⁵Department of Biology, University of Nevada, Reno, NV 89557; ⁶Center for Molecular Biodiversity Research, Zoological Research Museum Alexander Koenig, 53113 Bonn, Germany; ⁷Department of Entomology, University of Illinois at Urbana–Champaign, Urbana, IL 61801; ⁸Scientific Computing Group, Heidelberg Institute for Theoretical Studies, 69118 Heidelberg, Germany; ⁹Institute of Evolutionary Biology and Ecology, University of Bonn, 53121 Bonn, Germany; ¹⁰Evolutionary Biology and Ecology, Institute for Biology I (Zoolgy), University of Freiburg, 79104 Freiburg, Germany; ¹¹Australian National Insect Collection, Commonwealth Scientific and Industrial Research Organisation National Research Collections Australia, Acton, ACT 2601 Canberra, Australia; ¹²Department of Ecology and Evolution, University of Lausanne and Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland; ¹³Department of Entomology, Purdue University, West Lafayette, IN 47907; ¹⁴Department of Entomology, University of California, Riverside, CA 92521; ¹⁵Department of Life Sciences, Natural History Museum, London, SW7 5BD United Kingdom; ¹⁶Department of Botany, University of British Columbia, Vancouver V6T 1Z4, Canada; ¹⁷Department of Entomology and Plant Pathology, Auburn University, Auburn, AL 36849; ¹⁸School of Biological Sciences, University of Utah, Salt Lake City, UT 84112; ¹⁹BGI-Shenzhen, Shenzhen, 518083 Guangdong Province, People’s Republic of China; ²⁰Department of Entomology, China Agriculture University, 100193 Beijing, People’s Republic of China; and ²¹Systematic Entomology, Hokkaido University, Sapporo, 060-8589 Japan

Edited by David M. Hillis, The University of Texas at Austin, Austin, TX, and approved October 25, 2018 (received for review September 13, 2018)

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 phylogenetic framework for future studies of the group. Hemipteroid insects constitute a major lineage of Hemiptera (true bugs and allies; i.e., hemipterans), with over 120,000 described species, comprising well over 10% of known insect diversity. However, the evolutionary tree from this research provides a backbone framework for future studies of this important group of insects.

Significance

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.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission. Published under the PNAS license.

Data deposition: The data reported in this paper have been deposited in NCBI (accession nos. SRA SRR1821891–SRR1821980, SRR2051465–SRR2051515, and SRR921611–SRR921666). Gene sets, alignments, trees, quartet likelihood mapping results, morphological data matrices, and dating analyses results were deposited in Dryad repository, 10.5061/dryad.4fclgb8.¹

¹To whom correspondence should be addressed. Email: kpjohnso@illinois.edu.

This article contains supporting information online at www.pnas.org/lookup/supp/doi/10.1073/pnas.1815820115/-/DCSupplemental.

Published online November 26, 2018.
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 nonmonophyly of the hemipteroid insects, with Psocodea as the sister taxon of holometabolous insects rather than as sister to thrips plus hemipterans (Cordylognatha). 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., ref. 9) and DNA sequence data (e.g., ref. 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 orthologous protein-coding genes (16). Concatenation of these 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 nonholometabolous insect orders. This dataset enabled us to test the hypothesis of nonmonophyly 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 and SI Appendix, Figs. S1 and 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 (SI Appendix, Supplemental 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 Cordylognatha), although only 68% of quartets supported this result in four-cluster likelihood mapping (FcLM) (SI Appendix, Tables S5 and S6). As in the study of Misof et al. (1), Psocodea was 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 with 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 substitution processes, nonstationarity of substitution processes, or nonrandom 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 (Pithiraptera), which have elevated substitution rates compared with 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 were 14 potential apomorphies for the monophyly of Paraneoptera. These results indicate that there is more agreement between morphology and the FcLM results, compared with 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 tree-living bark lice, being the sister taxon of book lice (Liposcelididae), which makes the bark lice ("Psocodea") paraphyletic. In contrast to results based on 18S rDNA sequences (18), parasitic lice (Pithiraptera) 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 (Aelothripidae + 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 recovered 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 no underlying conflict in the data for this result (SI Appendix, 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., refs. 22–24).
rather than addressing relationships among major lineages (superfamilies).

The earliest molecular phylogenetic analyses of Hemiptera (e.g., refs. 25 and 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 infraroorders Fulgoromorpha (i.e., Fulgoroidea, planthoppers) and Cicadomorpha (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 infraroorders. In our analysis, which included representatives of all seven currently recognized infraroorders, the four infraroorders 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 (i) a clade comprising litter bugs (Diplocoromorpha), unique-headed bugs (Enicocephalorhyncha), and semiaquatic bugs (Gerromorpha) (also found in 100% of quartets in FcLM analyses) and (ii) shore bugs (Lepidopteromorpha) 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 Paleoptera (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 354–432), 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 (326 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, infraroorders, 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 counter to our previous results, our 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 datasets 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, nonmonophyly 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 morphology are used as calibration points imply the 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 bark lice 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 predominantly on phloem fluid, and 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 Pso- codea requires further study of morphological and molecular data.

In contrast to the equivocal support for Paraneoptera, Con- dylognatha is strongly supported not only in the phylogenomic analyses, but also with six morphological apomorphies. The or- igin 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 evolution- ary 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 posterovertrally (41), as in leafhoppers and relatives, rather than anteriorly as in true bugs (42). Nevertheless, morpho- logical 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 (42, 55) are not supported. (For results see 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 ([ref. 13: maximum likelihood (ML) bootstrap 83% and maximum parsimony (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 re- lationship 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 conver- gence between these two clades. Even if Psocodea is sister to thrips plus hemipterans, and not to holometabolous insects, re- cent 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, de- rived 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 taxo- nomic scales, from a single derived arrangement found in all Psocomorpha (46), to wide variation within a single genus (Lip- oseilus, ref. 47), and between closely related species of parasitic lice. In contrast, for the thrips, mitochondrial genome arrange- ments 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 contrast to the initial phylogenetic position of Psocodea remains to be resolved convincingly, our results based on transcriptomes for hemipteroid insects provide a strong phylo- genetic framework for future studies of genomic, morphological, eco- logical, 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 repository, 10.5061/dryad.4m8g8). Trans- scriptomes 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 species.

All paired-end reads were assembled with SOAPOdenovo-Trans (version 1.01; ref. 50) and the assembled transcripts were filtered for possible conti- namants (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 ar- chives (SI Appendix, Table S3) for further analysis. We searched the assembled transcriptomes 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 OrthoView (version beta4, ref. 52) for orthology search see SI Appendix, Table S4). Orthologous transcripts were aligned with MAFFT (version 7.123; ref. 53) at the translational (amino acid) level. Cor- respondence of nuclearide nucleotide alignments generated with a modified version of the software Pal2Nal (54) (version 14).

Alignment segments that could not be discriminated from randomly aligned regions at the amino acid level of each gene were identified with Alligator version 1.2 (55, 56). To maximize the fit of our substitution models, we identified for each gene the protein domain (clades, families) and unannotated regions using the Pfam database (refs. 1 and 57 and SI Appendix, Supplemental Materials and Methods). The phylogenetic information content of each data block was assessed with MARE (version 0.1.2-rc5) (58), and all uninformative data blocks (K = 0) were removed. We subsequently used PartitionFinder (developer version 2.0.0-p14, ref. 59) to simultaneously infer the best partitioning scheme and amino acid or nucleotide (removing third positions because of heterogeneity, SI Appendix, Fig. 50) substitution models, using the clustalw algorithm.

Phylogenetic trees were inferred using a maximum likelihood approach with ExaML version 3.17.00 (60) for both the nucleotide and amino acid datasets. We performed 50 nonparametric 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 phyloge- netic relationships (SI Appendix, Tables S5 and S6). Permutation tests in these analyses assessed the impact of heterogeneous amino acid sequence composition among lineages, nonstationarity of substitution processes, and non- random contribution of missing data to the inferred tree (64).

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 ref. 12) and additionally 25 characters associated with the wing base (8). By tracing char- acters over the tree using maximum parsimony with Winclada (64), we evaluated three possible phylogenetic alternatives: (i) paraphyletic Para- neoptera and Coleorrhyncha sister to Auchenorrhyncha (result from ML analysis of transcriptomes); (ii) monophyletic Paraneoptera (as suggested by FeLM analyses); and (iii) 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 MCCMCTree (66) molecular dating analysis of a first and second position nucleotide dataset for which sites were present in at least 95% of taxa.

for assistance preparing the ortholog gene set. R.M.W. was supported by Swiss National Science Foundation Grant P0003P3. 1706642. K.M. was supported by David Yeates, the Schlinger Endowment, CSIRO National Research Council Australia. J. Korb, and the University of Freiburg. This work was also supported by National Science Foundation DEB-1239788 (to K.P.J.H., and H.M.R.).


