Naturalis Repository

Phylogenetic divergences of the true bugs (Insecta: Hemiptera: Heteroptera), with emphasis on the aquatic lineages: the last piece of the aquatic insect jigsaw originated in the Late Permian/Early Triassic

Yan-hui Wang, Ying Cui, Dávid Rédei, Petr Baňař, Qiang Xie, Pavel Štys, Jakob Damgaard, Ping-ping Chen, Wen-bo Yi, Ying Wang, Kai Dang, Chuan-ren Li, Wen-jun Bu

Downloaded from:

https://doi.org/10.1111/cla.12137

Article 25fa Dutch Copyright Act (DCA) - End User Rights

This publication is distributed under the terms of Article 25fa of the Dutch Copyright Act (Auteurswet) with consent from the author. Dutch law entitles the maker of a short scientific work funded either wholly or partially by Dutch public funds to make that work publicly available following a reasonable period after the work was first published, provided that reference is made to the source of the first publication of the work.

This publication is distributed under the Naturalis Biodiversity Center 'Taverne implementation' programme. In this programme, research output of Naturalis researchers and collection managers that complies with the legal requirements of Article 25fa of the Dutch Copyright Act is distributed online and free of barriers in the Naturalis institutional repository. Research output is distributed six months after its first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and copyrights owner(s) of this work. Any use of the publication other than authorized under this license or copyright law is prohibited.

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the department of Collection Information know, stating your reasons. In case of a legitimate complaint, Collection Information will make the material inaccessible. Please contact us through email: collectie.informatie@naturalis.nl. We will contact you as soon as possible.

WILEY-
BLACKWELL

Cladistics 32 (2016) 390–405

Cladistics

10.1111/cla.12137

Phylogenetic divergences of the true bugs (Insecta: Hemiptera: Heteroptera), with emphasis on the aquatic lineages: the last piece of the aquatic insect jigsaw originated in the Late Permian/Early Triassic

Yan-hui Wang[a,†], Ying Cui[b,†], Dávid Rédei[a,†], Petr Baňař[c,†], Qiang Xie[a,*], Pavel Štys[d,*], Jakob Damgaard[e], Ping-ping Chen[f,*], Wen-bo Yi[a], Ying Wang[a], Kai Dang[a], Chuan-ren Li[g] and Wen-jun Bu[a,*]

[a]*Institute of Entomology, College of Life Sciences, Nankai University, 94 Weijin Road, Nankai District Tianjin, 300071, China;* [b]*Tianjin State Key Laboratory of Modern Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin 300193, China;* [c]*Department of Entomology, Moravian Museum, Hviezdoslavova 29, CZ-627 00, Czech Republic;* [d]*Faculty of Science, Department of Zoology, Charles University in Prague, Viničná 7, CZ-128 44, Praha 2, Czech Republic;* [e]*Natural History Museum of Denmark, Universitetsparken 15, 2100, Copenhagen Ø, Denmark;* [f]*Netherlands Centre of Biodiversity Naturalis, 2300, RA Leiden, Netherlands;* [g]*College of Agriculture, Yangtze University, Jingzhou, Hubei 434025, China*

Accepted 17 July 2015

Abstract

Heteroptera are among the most diverse hemimetabolous insects. Seven infraorders have been recognized within this suborder of Hemiptera. Apart from the well-established sister-group relationship between Cimicomorpha and Pentatomomorpha (= Terheteroptera), the two terminal lineages, the relationships among the other five infraorders are still controversial, of which three (Gerromorpha, Nepomorpha and Leptopodomorpha) are intimately connected to aquatic environments. However, the various and often conflicting available phylogeny hypotheses do not offer a clear background for a connection between diversification and palaeoenvironments. In this study, a molecular data set representing 79 taxa and 10 149 homologous sites is used to infer the phylogenetic relationships within Heteroptera. Bayesian inference, maximum-likelihood and maximum parsimony analyses were employed. The results of phylogenetic inferences largely confirm the widely accepted phylogenetic context. Estimation of the divergence time based on the phylogenetic results revealed that Gerromorpha, Nepomorpha and Leptopodomorpha originated successively during the period from the Late Permian to Early Triassic (269–246 Ma). This timescale is consistent with the origin and radiation time of various aquatic holometabolans. Our results indicate that the aquatic and semi-aquatic true bugs evolved under environmental conditions of high air temperature and humidity in an evolutionary scenario similar to that of the aquatic holometabolans.

Background

Aquatic insects account for nearly one-tenth of all insects (Morse, 2009; Zhang, 2011), with the majority belonging to nine orders: Odonata, Ephemeroptera, Plecoptera, Hemiptera, Megaloptera, Coleoptera, Trichoptera, Lepidoptera and Diptera. Among these, Ephemeroptera, Odonata, and Plecoptera apparently originated and diversified earlier than the holometabolan insects (Scherbakov and Popov, 2002; Misof et al., 2014). According to molecular clock-based estimations of divergence times, the aquatic groups of Holometabola most likely arose from the Late Permian to Early

---

*Corresponding authors.

E-mail addresses: qiangxie@nankai.edu.cn; pavelstys@gmail.com; p.chen@minlnv.nl and wenjunbu@nankai.edu.cn

†These authors contributed equally to this work.

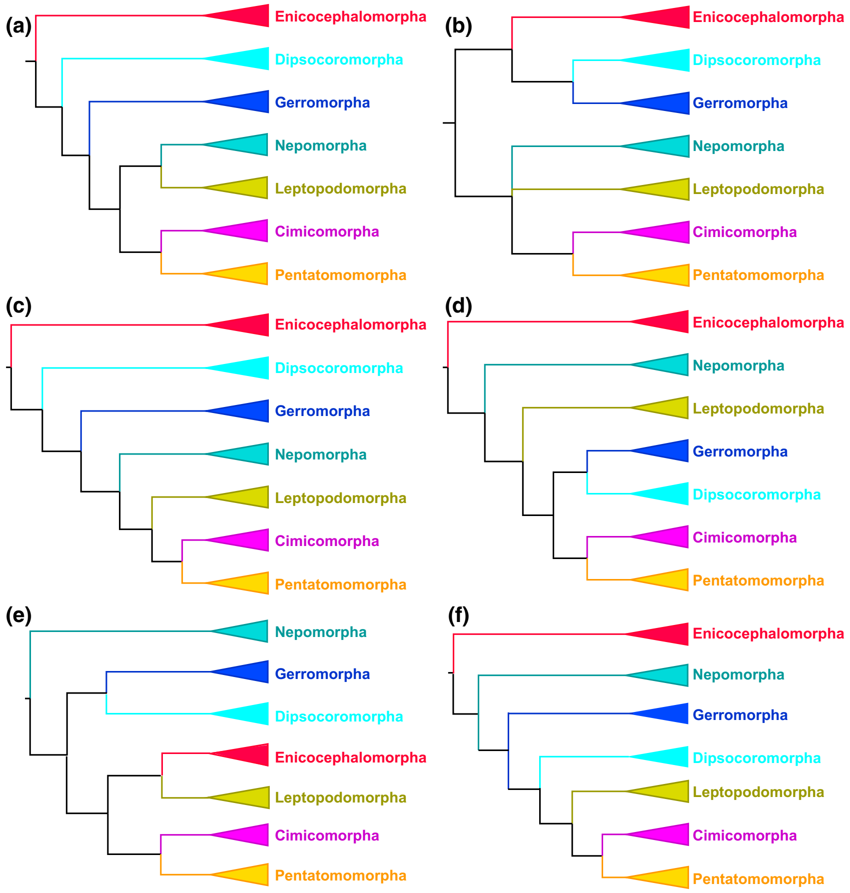

Fig. 1. Alternative hypotheses about the infraordinal relationships of Heteroptera. (a) after Schuh (1979), (b) after Zrzavý (1992), (c) after Wheeler et al. (1993), (d) after Xie et al. (2008), (e) after Li et al. (2012) and (f) after Weirauch and Štys (2014).

Triassic (Hunt et al., 2007; Bertone and Wiegmann, 2009; Winterton et al., 2010; Wiegmann et al., 2011). Compared to those aquatic insects, our understanding of the origin and divergence times of the aquatic lineages within Hemiptera has not been well recognized.

Within Hemiptera, nearly all of the aquatic groups belong to the suborder Heteroptera (true bugs). Heteroptera consists of more than 40 000 described species in approximately 89 families worldwide (Weirauch and Schuh, 2011) and is composed of seven infraorders (Štys and Kerzhner, 1975), of which Gerromorpha, Nepomorpha and Leptopodomorpha, and Dipsocoridae in the infraorder Dipsocoromorpha, are—with only a few exceptions—closely tied to aquatic environments. Most members of the semi-aquatic Gerromorpha are typically pleustonic, meaning that they have the ability to walk on the water surface, such as Mesoveliidae, Hebridae, Paraphrynoveliidae, Macroveliidae, Gerridae, Veliidae, Hydrometridae and Hermatobatidae. Some of them even live on the water surface, such as Hermatobatidae, most Gerridae and many Veliidae. Most members of Nepomorpha, or true water bugs, are fully aquatic in all life stages; these comprise the families Corixidae, Micronectidae, Diaprepocoridae, Nepidae, Belostomatidae, Naucoridae, Aphelocheiridae, Potamocoridae, Notonectidae, Pleidae and Helotrephidae. The nepomorphan families Gelastocoridae and Ochteridae are exceptional, as they mostly occur in riparian habitats. Leptopodomorpha, or shorebugs, including four families (Saldidae, Aepophilidae, Leptopodidae, Omaniidae) are usually associated with habitats along water margins. For convenience, these three infraorders will be referred to as "aquatic infraorders" in the following text, using this term in a broad sense. Among the three aquatic infraorders of Heteroptera, Nepomorpha shows the strongest morphological and biological adaptation to aquatic habitats; those of Gerromorpha and Leptopodomorpha are relatively weaker. A better recognition on the phylogenetic positions of these three lineages is crucial for understanding the origin and diversification of them.

In earlier studies based on morphological characters or combined partial 18S rDNA, these three aquatic infraorders appeared as successively branched lineages (Schuh, 1979; Wheeler et al., 1993) (Fig. 1a, c).

Nevertheless, this view has been challenged by some morphological, molecular and palaeoentomological studies (Mahner, 1993; Scherbakov and Popov, 2002; Li et al., 2012). These three papers offered a topology with Nepomorpha located at the basal position within Heteroptera (Fig. 1e), and the corresponding chronogram suggested that Nepomorpha diverged roughly 250 Ma. However, the other two aquatic infraorders appeared neither in sister-group positions nor as successively branched lineages next to Nepomorpha, indicating that the aquatic habits of Gerromorpha, Nepomorpha and Leptopodomorpha most likely originated independently, in separated timespans. Moreover, other hypotheses still exist for the phylogenetic relationships between the infraorders of Heteroptera (Zrzavý, 1992; Xie et al., 2008; Weirauch and Štys, 2014) (Fig. 1b, d, f). In both the works of Xie et al. (2008), focusing on the phylogenetic relationships between the suborders of Hemiptera, and Weirauch and Štys (2014), focusing on the phylogeny of Dipsocoromorpha, the taxon sampling covered all seven infraorders of Heteroptera and thus provided some direct evidence for the phylogeny of Heteroptera. Although Nepomorpha is not the basal lineage in these studies, the phylogenetic positions of the three aquatic infraorders are not consistent and are always entangled with Dipsocoromorpha. Therefore, after 35 years of efforts to infer the higher level phylogeny of Heteroptera, the phylogenetic relationships among the seven infraorders remain controversial except for the sister relationship between Cimicomorpha and Pentatomomorpha (Schuh, 1979; Zrzavý, 1992; Wheeler et al., 1993; Fischer et al., 2000; Xie et al., 2008; Schuh et al., 2009; Cassis and Schuh, 2010; Li et al., 2012; Weirauch and Štys, 2014) (Fig. 1).

From a palaeontological view, the oldest definitive fossil records of Heteroptera almost all belong to Nepomorpha. They can be dated back to the Late Triassic approximately 230 Ma and include Belostomatidae, Naucoridae, Notonectidae and Ochteroidea (Scherbakov and Popov, 2002; Grimaldi and Engel, 2005). By comparison, most fossil records of the other six infraorders appeared much later. Fossil records of Leptopodomorpha appeared at the Triassic/Jurassic boundary, while the remaining five infraorders appeared in the middle or late Jurassic. The absence of fossil Gerromorpha from the Triassic is a mystery because they should be at least of the same age or, according to the currently most widely accepted theory, even older than Nepomorpha and Leptopodomorpha (Schuh, 1979; Wheeler et al., 1993) (Fig. 1a, c), and have a similar chance of fossilization because they live in the same environment. The extant families of aquatic bugs had already existed since the Late Triassic, but the time of the origin of Heteroptera and its major clades is still unknown. Until we find abundant definitive heteropteran fossil records from or before the Early Triassic, the molecular-based divergence time estimation can be used as an alternative method for understanding the early evolutionary history of heteropterans.

As a result, inconsistent topologies in various studies containing phylogenetic inferences of Heteroptera have prevented estimations of the divergence times of the major lineages of the suborder. Although it appears that researchers had employed the same molecular markers—rDNAs—in the aforementioned studies, differences in the length of the molecular markers, alignment methods and whether they are combined with mitochondrial genes can result in different phylogenetic inferences (Wheeler et al., 1993; Xie et al., 2008; Li et al., 2012; Weirauch and Štys, 2014). Although several complete sequences of 18S rDNA for true bugs are currently available from GenBank, only six complete sequences of 28S rDNA exist. Furthermore, no record of mitochondrial genomes for the Dipsocoromorpha is yet available.

In addition to the uncertainties in the phylogenetic studies, several weak points in the analyses of divergence times might also bias the estimates. First, hard bounds on priors probably can result in high uncertainties of the deduced timescale mainly due to the difficulties in arbitrarily determining a proper maximum hard bound, in which the underestimation can be too exclusive while the overestimation can be too uninformative (Ho and Phillips, 2009; Wheat and Wahlberg, 2013). Second, a single calibration point cannot accommodate possible inaccuracy of the record for calibration (Hug and Roger, 2007; Ho and Phillips, 2009), and the deviations of the estimated time for the nodes distant from the calibrated one can be very large (Linder et al., 2005). Third, excessive restrictions on priors are not beneficial either, because they would result in overly informative priors, and then the function of the dataset *per se* in divergence time estimation would be weakened (Welch et al., 2005; Sanders and Lee, 2007; Heled and Drummond, 2012).

Here, we used Bayesian inference (BI), maximum-likelihood (ML) and maximum parsimony (MP) to infer the phylogeny of the seven infraorders of Heteroptera based on the complete sequences of two nrDNAs (18S rDNA and 28S rDNA) and all 13 protein coding genes (PCGs) in mitochondrial genomes. Each infraorder has at least three representative groups in the taxon sampling, with two in the outgroup Coleorrhyncha. Furthermore, an established Bayesian relaxed-clock approach (Drummond et al., 2006) was used to infer a comprehensive timescale for the evolution of Heteroptera, with special attention to the aquatic infraorders. To avoid analytical biases, multiple calibrations and soft bounds on priors were used.

Materials and methods

Taxon sampling

We sampled a total of 79 species, 77 of which are heteropterans and 2 of which are species of Coleorrhyncha serving as outgroups (Table S1). The sampled species included most superfamilies of the seven infraorders, including some very poorly known and rarely collected taxa, such as Aenictopecheidae (Enicocephalomorpha) (a new genus and species from Tibet currently under study by Štys et al.). For rDNAs, 75 complete 18S and 28S sequences were newly sequenced in this study. Furthermore, mitogenomes of three heteropterans are reported here for the first time: *Kokeshia xiei* (Dipsocoromorpha: Schizopteridae) (Accession number: JN989542), *Ptilomera tigrina* (Gerromorpha: Gerridae) (Accession number: KP400583) and *Entomovelia* sp. (Gerromorpha: Veliidae) (Accession number: KP400582).

Molecular experiments

All of the specimens collected for the experiments had been preserved in absolute ethanol. Total genomic DNA was extracted by a CTAB-based method (Reineke et al., 1998) from either muscle tissue of the thorax or the whole body (excluding the abdomen) according to the individual size of the sampled species. The primer sets used for amplification as well as for sequencing are listed in Table S2. The primers Ns1 and Ns8 were from Barker et al. (2003), 18SP3 and 18SP2/18SP4 were from Campbell et al. (1995), while the other primer sets were designed for this study. The thermal cycling program for the polymerase chain reaction of 28S rDNAs and mitochondrial fragments included an initial denaturation at 94 °C for 1 min, denaturation of 30–35 cycles for 20–30 s at 94 °C, annealing for 30 s–1 min at 47–60 °C, elongation for 1–10 min at 72 °C depending on the size of the products, and a final elongation for 10 min at 72 °C. The thermal cycling program for 18S rDNA followed Johnson and Clayton (2000).

Alignment and secondary structure reconstruction

All sequences were initially aligned by the program Muscle embedded within MEGA 6.06 (Tamura et al., 2013). The alignments of nrDNAs were then checked and manually corrected on the basis of the secondary structure models of heteropteran 18S rRNA (Fig. S1) and 28S rRNA (Figs S2 and S3), which refer to the general model of 18S and 28S rRNAs for Hexapoda (Xie et al., 2009; Wang et al., 2013), respectively. The secondary structure reconstructions for 18S and 28S rRNAs were realized by thermodynamic folding using RNA structure 5.3 (Reuter and Mathews, 2010) and comparative analyses, which have been repeatedly described in previous studies (Kjer, 1995, 2004; Gutell et al., 2002; Misof et al., 2006; Xie et al., 2009; Wang et al., 2013). Five length-variable regions exist in 18S rDNA for which the positional homology of nucleotides cannot be determined (Fig. S1, File S1), while 10 similar regions are found in 28S rDNA (Figs S2, S3 and File S1). These length-variable regions were eliminated before the process of phylogenetic reconstruction and the estimation of divergence times. The finally concatenated alignment consists of 10 149 sites, of which 6416 sites are nucleotides from 18S and 28S rDNAs, and 3733 sites are amino acids from the 13 PCGs in mitochondrial genomes (File S2, available at http://dx.doi.org/10.5061/dryad.1b0q8).

Phylogenetic reconstructions

Phylogenetic analyses were conducted utilizing BI, ML and MP. The programs jModeltest 2.1.1 (Posada, 2008) and Treefinder (Jobb, 2011) were used to infer the best substitution model for the nrDNAs and amino acids, respectively. The BI were performed in MrBayes 3.2.5 using the Beagle library on CPUs (Ronquist et al., 2011). In MrBayes, the GTR substitution model for the partition of nrDNAs and the mixed substitution model for the partition of amino acids were used under a gamma distribution (+G) with a proportion of invariable sites (+I) to account for among-site rate variation. Other parameters were set as follows: generations = 5 000 000, samplefreq = 1000, printfreq = 1000, nchains = 4, and the generations with values of the standard deviation greater than 0.01 were discarded. The numbers of burned generations were also checked with the help of Tracer 1.5 (available at http://beast.bio.ed.ac.uk/Tracer). The ML analysis was undertaken using RAxML 8.0.12 in PThreads version (Stamatakis, 2014). For RAxML, the substitution model GTR+G+I for the partition of nrDNAs; amino acid substitution models mtArt+G+I for the partition of COI, COII, ND1 and ND4L; and mtZOA+G+I for the partition of the remaining 9 PCGs turned out to be the most appropriate ones. The best ML tree was calculated from 100 RAxML runs, followed by 2000 bootstrap replicates. Bootstrap proportions were then mapped onto the best ML phylogram. The MP analysis was performed using Tree analysis using New Technology (TNT) (Goloboff et al., 2008). All characters were equally weighted and gaps were treated as missing characters (Ogden and Rosenberg, 2007). Clade robustness was evaluated by using the jackknife resampling (independent character removal) and the bootstrap reweighting (Poisson independent). The traditional searches for trees were performed under TBR (tree bisection reconnection) branch swapping and 100 random replicates. Jackknife and bootstrap resampling were calculated with 2000 replicates.

Divergence time estimation

The program BEAST 2.1.3 was applied to simultaneously calculate divergence times and reconstruct heteropteran phylogenetic relationships (Bouckaert et al., 2014). The program BEAUti 2.1.3 (Bouckaert et al., 2014) was used to generate the file used in the BEAST analysis. The substitution model used for nrDNAs in BEAST was the same as that in the ML analysis, while the mtREV substitution model was used for amino acids due to the limited models available in BEAST. The birth–death speciation process of priors (Stadler et al., 2013) and the relaxed clock allowing the substitution rate to vary according to an uncorrelated lognormal distribution (Drummond et al., 2006) were implemented. Initial runs with BEAST showed that Notonectidae + Pleoidea, Ochteroidea, Nepoidea, Naucoroidea and Lygaeoidea did not remain monophyletic. As the monophylies of Notonectidae + Pleoidea, Ochteroidea and Nepoidea (Mahner, 1993; Hebsgaard et al., 2004; Li et al., 2014), the monophyly of Naucoroidea (Mahner, 1993; Schuh and Slater, 1995; Hua et al., 2009; Li et al., 2014), and the monophyly of Lygaeoidea (Schuh and Slater, 1995; Henry, 1997) have been widely accepted, we constrained the corresponding clades in BEAST analyses. All other priors, except calibration points described below, were left to the defaults in BEAST.

Fifteen fossil records from various lineages were used in calibration (Table 1). An attempt was made to use the oldest convincing fossil record known for each group (Tillyard, 1926; Popov, 1971; Lin, 1985; Popov et al., 1994; Fraser et al., 1996; Azar et al., 1999; Perrichot et al., 2005; Yao et al., 2006, 2007a,b; Damgaard, 2008; Azar and Nel, 2010; Hou et al., 2012). According to the criteria suggested by Parham et al., 2012, we provided information on fossil taxon descriptions, localities, age, mode of preservation and references for each fossil record (Table 1). All calibration points were set as normal distributions (Table 1). The mean was set to the middle value of the stage/age in which the corresponding fossil was located, while the bounds for the 95% confidence interval (CI) were set as the starting time of one stage/age earlier and the ending time of one stage/age later. Each fossil record was used to calibrate the most adjacent ancestral node with definitiveness. Among the calibrated nodes, each one was corresponding to one fossil record, except for the root node, at which two fossil records were used. For the clade Heteroptera + Coleorrhyncha, *Actinoscytina belmontensi* is the oldest definitive fossil record discovered up to now, and it can be dated back to 254.0–252.3 Ma. As it is among the stem groups of Coleorrhyncha, it was used to calibrate the root node in a consistent way of calibration as the other nodes. The mean of the 95% CI of the root node was set as the middle value of the timespan 254.0–252.3 Ma. However, the setting of the standard deviation value for the root node is a little bit different from the other calibrated nodes for two reasons. For one thing, the time intervals corresponding to the stage/age of the fossil record *Actinoscytina belmontensi* and its adjacent two are very narrow, which may result in excessive restrictions on the root node. For the other, several definitively the oldest fossil records of Heteroptera in various extant families from the Late Triassic, approximately 230 Ma, have been described. Combining the information from the earliest fossil records of Coleorrhyncha and Heteroptera, the prior distribution of the origin and early diversification of Heteroptera is proposed to be 253 ± 23 Ma. Therefore, one of the earliest fossil records of Heteroptera was selected to serve as the minimum bound of the 95% CI of the root node.

The analysis was run for 1 000 000 000 generations and sampled every 1000 generations. Effective sample sizes (ESS), marginal densities of priors and posteriors, and all other parameters were checked in the Tracer v.1.5 program (available at http://beast.bio.ed.ac.uk/Tracer). TreeAnnotator (Bouckaert et al., 2014) was used to summarize the set of post-burn-in trees and their parameters, to produce a maximum clade credibility (MCC) chronogram showing mean divergence time estimates with 95% high posterior density (HPD) intervals. All of the ESSs were above the recommended threshold of 200, indicating that the parameter space had been sufficiently sampled.

Two additional analyses were also conducted. The first was a “null analysis” which ran under the priors without sequence data only to assess whether the prior distribution was sufficiently wide—not too informative to accommodate the impact of sequence data. In the null analysis, the only difference was that the nucleotide and amino acid sequences were replaced with a series of missing states: NNNN (Bouckaert et al., 2014). The results of null analysis were compared with those obtained with the sequence data to assess whether our sequence data had also contributed to the posterior estimation. The second analysis was used to assess the impact of missing data on divergence time estimation. Only those species that had both the complete nrDNA and all amino acid sequences of the 13 PCGs were selected to constitute a new matrix. This new matrix, consisting of 32 objects, was used in a separate BEAST analysis with the same settings as described above.

Table 1
Fossil taxa with ages of divergence (Million years ago, Ma) and prior probability distributions of the calibration points used in the divergence time estimation

Calibration nodes	Taxonomic group	Fossil records	Age (Ma)	Mode of preservation	Localities	Prior	References
Root	Progonocimicidae	*Actinoscytina belmontensi*	254.0–252.3	Impression fossil	Belmont	Normal mean = 253, SD = 11	Tillyard (1926)
	Belostomatidae	Belostomatidae indet.	227–208.5	Impression fossil	North Carolina		Fraser et al. (1996)
Cimiciformes	Cimicoidea: Vetanthocoridae	*Pumilanthocoris gracilis*	166.1–157.3	Compression fossil	Daohugou	Normal mean = 161.7, SD = 3.3	Hou et al. (2012)
Coreoidea	Rhopalidae	*Originicorizus pyriformis*	166.1–157.3	Compression fossil	Daohugou	Normal mean = 161.7, SD = 3.3	Yao et al. (2006)
Corixoidea	Corixidae	*Liassocorixa dorsetica*	199.3–190.8	Impression fossil	Flatstones, Stonebarrow, Charmouth	Normal mean = 195.05, SD = 3.2	Popov et al. (1994)
Dipsocoromorpha	Schizopteridae	*Libanohypselosoma popovi*	125.0–113.0	Amber	Mdeyrij-Hammana, Casa Baabda	Normal mean = 119, SD = 5.3	Azar and Nel (2010)
Enicocephalomorpha	Enicocephalidae	*Enicocephalinus acragrimaldii*	125.0–113.0	Amber	Mdeyrij-Hammana, Casa Baabda	Normal mean = 119, SD = 5.3	Azar et al. (1999)
Gerroidea	Gerridae	*Cretogerris albianus*	113–100.5	Amber	Archingeay-Les-Nouillers	Normal mean = 107, SD = 6.7	Perrichot et al. (2005); Damgaard (2008)
Leptopodomorpha	Archegocimicidae	*Britannicola senilis*	201.3–199.3	Impression fossil	Apperley	Normal mean = 200.3, SD = 4.2	Popov et al. (1994)
Lygaeoidea	Lygaeoidea	*Leipolygaeus similis*	170.3–166.1	Impression fossil	Pengzhuang village, Hanshan County	Normal mean = 168.2, SD = 2	Lin (1985)
Miroidea	Miridae	*Mirivena robusta*	166.1–157.3	Compression fossil	Daohugou	Normal mean = 161.7, SD = 3.3	Yao et al. (2007b)
Nepoidea	Belostomatidae	*Tarsabedus menkei*	199.3–190.8	Impression fossil	0.6 km east Charmouth	Normal mean = 195.05, SD = 3.2	Popov et al. (1994)
Notonectidae+ Pleoidae	Notonectidae	*Liadonecta tomiensis*	182.7–170.3	Compression fossil	Chernyy Etap II	Normal mean = 176.5, SD = 4.5	Popov (1971)
Ochteroidea	Ochteridae	*Propreocoris maculatus*	199.3–190.8	Impression fossil	Woodstones, Black Ven, Charmouth	Normal mean = 195.05, SD = 3.2	Popov et al. (1994)
Pentatomoidea	Cydnidae	*Cilicydnus robustispinus*	145.0–139.8	Compression fossil	Huangbanjigou, Chaomidian Village	Normal mean = 142.4, SD = 5	Yao et al. (2007a)

Confidence test of the selected node in the ML result

We assessed the confidence of the sister relationship of Enicocephalomorpha and Dipsocoromorpha obtained in the ML analysis with the program CONSEL (Shimodaira and Hasegawa, 2001). We first constrained the monophyly of Euheteroptera (including the other six infraorders without Enicocephalomorpha) and inferred the best ML tree based on this topology-constrained tree using RAxML 8.0.12 (Stamatakis, 2014). Subsequently, both the constrained and nonconstrained best ML trees were used for the site-wise log-likelihood analysis. The only major difference between the two best ML trees was the position of Dipsocoromorpha. *P*-values were calculated by the main approximately unbiased (AU) test and various supplementary tests, which reflected the confidence of the selected topology.

Results

Phylogenetic analyses

Phylogenetic analyses using BI, ML and MP yielded largely congruent trees with high posterior probabilities and bootstrap values. The resulting trees are presented in Fig. 2 and Figs S4–S6, and contain detailed information about node support values. All seven infraorders were found to be monophyletic with high node support values, except the lower ones for Cimicomorpha and Nepomorpha in the MP analysis. In addition, Panheteroptera, Leptopodomorpha + Terheteroptera (*sensu* Weirauch and Štys, 2014, equal to Cimicomorpha + Pentatomomorpha), and Terheteroptera were all shown to be monophyletic groups with high support values. In addition, Gerromorpha was strongly supported as a sister group to Panheteroptera (= Nepomorpha + (Leptopodomorpha + (Cimicomorpha + Pentatomomorpha))) in the MP analysis and with moderate bootstrap values in the ML analysis.

In all of the phylogenetic analyses, Enicocephalomorpha together with Dipsocoromorpha formed a monophyletic group with high support values. This clade was strongly supported as the basal heteropteran lineage. For the ML analysis, the *P*-value of the AU test of the clade (Enicocephalomorpha + Dipsocoromorpha) was 0.995, and the *P*-value for the clade Euheteroptera was 0.005, which is less than the significance level of 0.05. The other *P*-values obtained by Kishino–Hasegawa (KH), Shimodaira–Hasegawa (SH), and the weighted Kishino–Hasegawa (WKH) and weighted Shimodaira–Hasegawa (WSH) tests for the clade Euheteroptera were 0.006. As the *P*-values are lower than the significance level of 0.05, the alternative topology can be rejected. That means that the sister relationship between Enicocephalomorpha and Dipsocoromorpha obtained using the ML analysis is robust. Our estimated topology is therefore largely congruent with that reported by Wheeler et al. (1993), differing only in the position of the Dipsocoromorpha (Fig. 2).

Within Nepomorpha, the superfamilies Nepoidea (= Belostomatidae + Nepidae), Ochteroidea (= Gelastocoridae + Ochteridae) and Pleoidea (= Pleidae + Helotrephidae) are monophyletic groups with high node support values in all analyses, which agrees with the opinions of Popov (1971), Rieger (1976), Mahner (1993), Hebsgaard et al. (2004) and Li et al. (2014). Besides, the monophyly of Notonectidae + Pleoidea was strongly supported in the BI and ML analyses. For Corixoidea (according to Chen et al., 2005) and Naucoroidea (according to Mahner, 1993; Schuh and Slater, 1995), including two more families such as Diaprepocoridae and Potamocoridae in the future will further facilitate recognition of the monophyly of these two superfamilies. As for Cimicomorpha, our results showed that this infraorder is split into three main lineages: Miriformes (consisting of Tingidae, Miridae and Thaumastocoridae), Cimiciformes (consisting of Nabidae, Anthocoridae and Cimicidae), and Reduvioidea (Reduviidae). Such a context is largely congruent with Schuh and Štys (1991) and Schuh et al. (2009). Aradidae was consistently recovered as the sister group to the rest of Pentatomomorpha. This result supports the widely accepted recognition of Pentatomomorpha (cf. Schuh and Slater, 1995) and conflicts with the opinion of Sweet (1996, 2006) who excluded Aradidae from Pentatomomorpha and recognized a sister-group relationship between Leptopodomorpha + Pentatomomorpha (without Aradidae). At the same time it also offers strong support for the monophyly of Trichophora in accordance with the broadly held views of previous authors (Štys and Kerzhner, 1975; Schaefer, 1993; Hua et al., 2008; Tian et al., 2011; Yao et al., 2012).

Divergence times estimation

The MCC tree is displayed in Fig. 3 with a median node height and a 95% HPD interval for each main clade. We estimated that the divergence time between Gerromorpha and Panheteroptera was 269 Ma, with a 95% HPD of 254–284 Ma. The next divergence between Nepomorpha and Leptopodomorpha + Terheteroptera is estimated to have occurred 258 Ma (95% HPD 244–272 Ma) in the Late Permian and Early Triassic, followed by the divergence between Leptopodomorpha and Terheteroptera 246 Ma (95% HPD 231–260 Ma) in the Early Triassic. The split between Cimicomorpha and Pentatomomorpha occurred 231 Ma (95% HPD 217–245 Ma) in the Carnian stage of the Triassic. Within Nepomorpha,

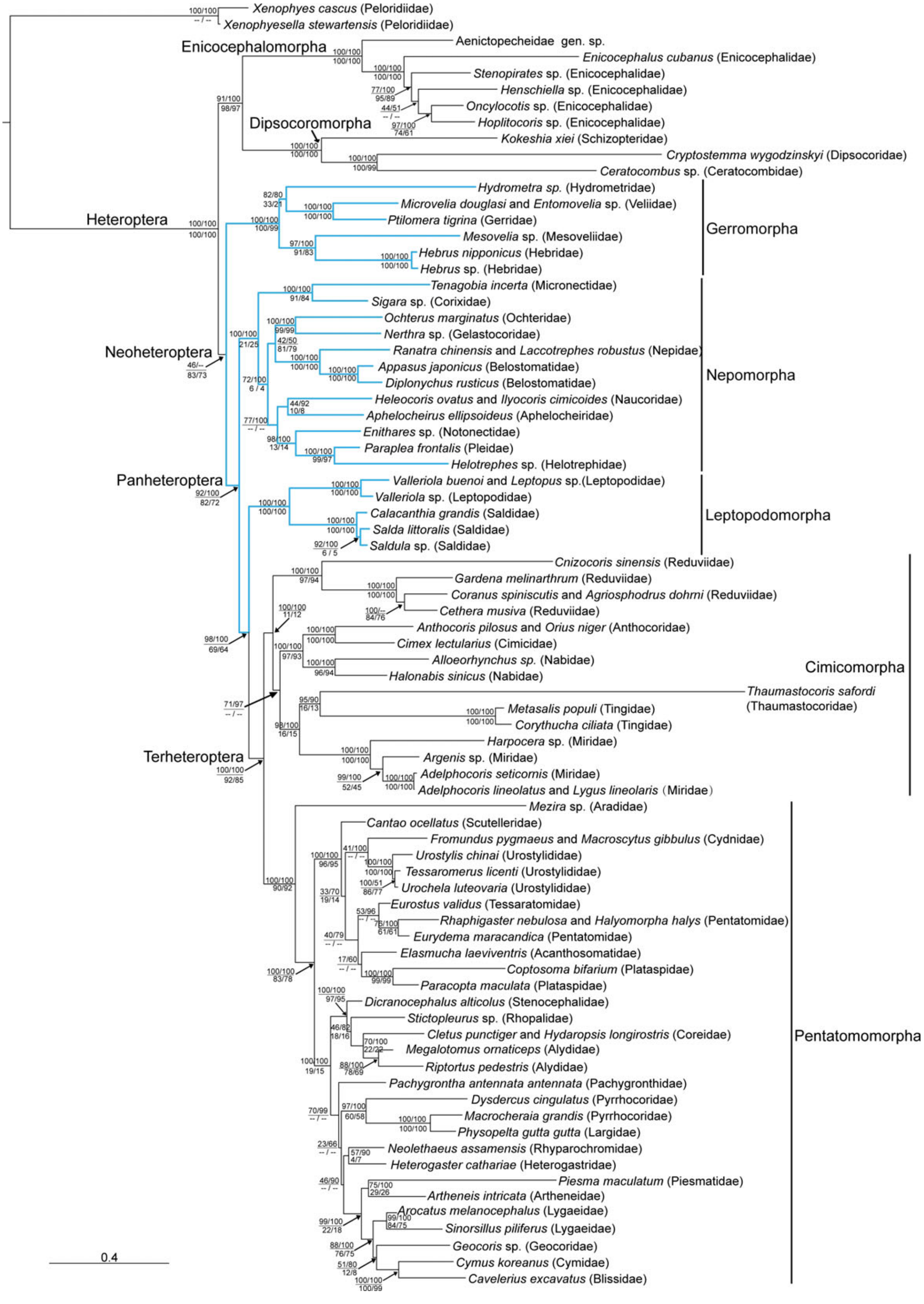

Fig. 2. Phylogenetic tree based on complete taxon sampling. Numbers above each node indicate maximum-likelihood (ML) bootstrap values and Bayesian posterior probabilities values. Numbers below each node indicate maximum parsimony (MP) jackknife values and bootstrap values. A dash is shown if the topology is not shown in the maximum parsimony analysis. The lengths of the branches follow the phylogram of the ML tree. Internal branches not supported by MP analysis were drawn as dashed lines.

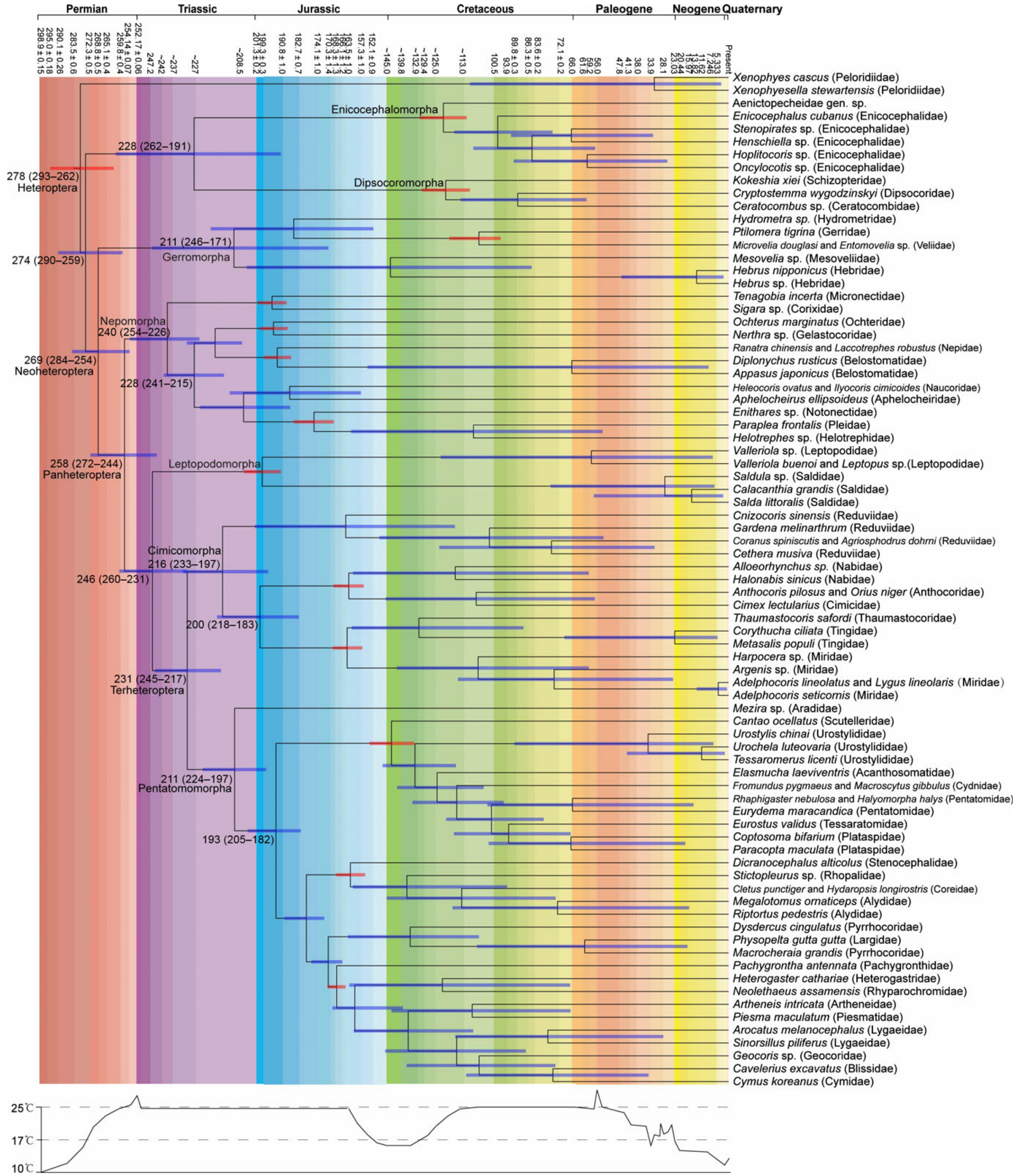

Fig. 3. Estimated divergence times of major Heteroptera clades based on complete taxon sampling. Blue bars indicate 95% mean confidence intervals. Red bars denote the calibration points. The chronological coordinates in the chronogram are referenced against the International Chronostratigraphic Chart (Gradstein et al., 2012). The evolution of temperature variation over time is shown below the chronogram and is based on http://www.scotese.com/climate.htm.

Table 2
Ages of the major clades of Heteroptera (HPD, higher posterior density)

	Mean divergence time and 95% HPD (Ma)	
Clades	79 taxa	32 taxa
Neoheteroptera	269 (254–284)	267 (252–282)
Panheteroptera	258 (244–272)	254 (240–269)
Leptopodomorpha + Terheteroptera	246 (231–260)	243 (229–257)
Terheteroptera	231 (217–245)	228 (214–243)
Cimicomorpha	216 (197–233)	214 (197–231)
Pentatomomorpha	211 (197–224)	210 (195–224)
Gerromorpha	211 (171–246)	195 (158–229)
Nepomorpha	240 (226–254)	236 (223–249)
Split between Enicocephalomorpha and Dipsocoromorpha	228 (191–262)	229 (189–264)
Origin of Miroidea	200 (183–218)	200 (183–217)

the first diversification is estimated to have occurred 240 Ma (95% HPD 226–254 Ma); the superfamilies then further diversified in the Late Triassic. Almost all of the extant families had already appeared approximately by the early Jurassic.

An assessment of prior bias on posterior estimates

In order to avoid a distortion of the time estimates by a systematic misplacement of fossil calibration points, all of the calibration points can be placed within a reliable systematic context. These calibration points were distributed across all seven infraorders of Heteroptera, and all of these calibration points were set as normal distributions. After comparing the prior and posterior distributions of the 14 calibration points, we found that the posterior distributions of the age estimations were approximately the same as the prior distributions for four of the nodes (Cimiciformes, Corixoidea, Leptopodomorpha and Pentatomoidea). However, they were older for five of the nodes (root, Dipsocoromopha, Enicocephalomorpha, Notonectidae + Pleoidea and Miroidea) and younger for the remaining five nodes (Coreoidea, Gerroidea, Lygaeoidea, Ochteroidea and Nepoidea) (File S3). The results indicated that our data did contribute to the posterior estimates, that is, the 14 soft priors did not result in overparameterization and thus dominate the posterior estimates (Heled and Drummond, 2012; Wheat and Wahlberg, 2013). Alternatively said, the divergence time had been inferred based on the dataset in combination with prior settings.

Effect of missing data on divergence time estimates

After removing the species having only nrDNA sequences from the original dataset, the divergence times of the main clades were re-estimated (Fig. S7). The effect of data reduction on the divergence time estimations was minimal. Mean estimates and their 95% HPD for the main clades of interest were tabulated for easy access (Table 2). The 95% HPD for almost all of the clades correspondingly overlapped. Mean divergence times for most nodes fluctuated by a few million years compared with the results obtained from the complete dataset. The one exception was the Gerromorpha node, which appeared to be 16 Myr younger than the result from the original dataset, but the 95% HPD of this node in the original dataset is skewed within the 95% HPD of the minor dataset.

Discussion

Phylogenetic relationships

In this study, we made considerable efforts in taxon sampling for the outgroup Coleorrhyncha and the ingroup of all seven infraorders of Heteroptera, especially for Enicocephalomorpha in which the Aenictopecheidae was incorporated for the first time in a phylogenetic study of Heteroptera. In addition, three families of Dipsocoromorpha (i.e. Dipsocoridae, Schizopteridae and Ceratocombidae) were included. The taxon sampling included 77 species from 54 families of the seven infraorders. Seventy-five newly sequenced and complete 18S rDNA and 28S rDNA sequences were provided for the first time in this study. In addition, data for the mitochondrial PCGs also come from Coleorrhyncha and all seven infraorders of the true bugs, among which the mitochondrial genome of Dipsocoromorpha is reported here for the first time. These newly obtained sequences make it possible to reconstruct the phylogenetic relationships between the different infraorders of Heteroptera based on complete sequences of both nrDNAs and mitochondrial PCGs.

There is little agreement on the phylogenetic position of Nepomorpha within Heteroptera in previous studies. Five competing hypotheses have been proposed (Fig. 1), some of them suggesting a sister-group relationship with either Leptopodomorpha (Schuh, 1979) or Leptopodomorpha + Terheteroptera (Wheeler et al., 1993). Others recover the infraorder as the basal one of Heteroptera (Mahner, 1993; Scherbakov and Popov, 2002; Li et al., 2012) or the second branch next to Enicocephalomorpha (Xie et al., 2008; Weirauch and Štys, 2014). Mahner (1993) placed Nepomorpha as the basal clade based on a set of morphological characters; this hypothesis later received support from the fossil record and morphological characters (Scherbakov and Popov, 2002). A recent work based on molecular data also arrived at the same conclusion (Li et al., 2012). Among the molecular markers used in that study, nrDNAs played a key role in determining topology; the mitochondrial 16S rDNA and *COI* were also likely saturated and provided little information for addressing the infraordinal relationships. The weak contributions of mitochondrial 16S rDNA and *COI* were also observed in a case study of Cimicomorpha (Tian et al., 2008). Compared to the molecular phylogenetic studies of Heteroptera that only recruited nrDNAs (Wheeler et al., 1993; Xie et al., 2008), the basal position of Nepomorpha needs to be cautiously treated. In this study, complete nucleotide sequences of the 18S and 28S rDNAs and amino acid sequences of all 13 PCGs in mitochondrial genomes have been combined. The monophyly of Panheteroptera was consistently recovered with 100% Bayesian posterior probability, 92% likelihood bootstrap value, and 82% parsimony jackknife and 72% parsimony bootstrap values. This is congruent with the hypothesis of Schuh (1979), Wheeler et al. (1993) and Schuh et al. (2009). The traditionally well-established clade Terheteroptera was reinforced in our analyses. In addition, Leptopodomorpha was placed as a sister group of Terheteroptera, which had also been shown by several previous studies (Wheeler et al., 1993; Cassis and Schuh, 2010).

A taxon corresponding to the modern Nepomorpha was first proposed (as Hydrocorisae) by Latreille (1802) and it has been universally accepted as a monophyletic clade by subsequent authors (Popov, 1971; Rieger, 1976; Mahner, 1993; Hebsgaard et al., 2004; Li et al., 2012, 2014). The only exception is Hua et al. (2009) who elevated Pleoidea to infraorder rank (as "Plemorpha," correctly Pleomorpha) due to its sister-group relationship to the remaining groups of Nepomorpha, Leptopodomorpha, Cimicomorpha and Pentatomomorpha inferred from the result of phylogenetic reconstruction based on mitochondrial genomes. However, the position of Pleoidea in the result of Hua et al. (2009) is controversial and it may have resulted from an improper taxon sampling of outgroups and incomplete taxon sampling of ingroups (Li et al., 2014). In this study, the monophyly of Nepomorpha was recovered by the phylogenetic reconstructions based on BI, ML and MP. Furthermore, the sister relationship between Notonectidae and Pleoidea were strongly supported in both the BI and ML analyses, which agreed with the widely accepted opinion (Popov, 1971; Mahner, 1993; Hebsgaard et al., 2004) and suggested the single origin of the back-swimming behaviour in Notonectidae, Pleidae and Helotrephidae.

The position of Gerromorpha as a sister group to Panheteroptera was revealed by MP analysis with the values of 83% in jackknife and 73% in bootstrap. This clade was also supported in the ML phylogram but with just a moderate bootstrap value. While in the phylogram of BI, the position of Gerromorpha was unresolved as one of the trifurcated branches together with (Enicocephalomorpha + Dipsocoromorpha) and Panheteroptera. The most likely reason for the unresolved status of Gerromorpha in the BI analysis is the limited application of the substitution models of amino acids. The substitution model mtArt was suggested to be the best one for the amino acid sequences of COI, COII, ND1 and ND4L; and mtZOA was suggested to be the best one for those corresponding to the remaining nine PCGs. However, only the mixed substitution model can be applied to the amino acids as a whole. Therefore, the interrelationships between the five infraorders of Neoheteroptera were recovered as (Gerromorpha + (Nepomorpha + (Leptopodomorpha + (Cimicomorpha + Pentatomomorpha)))) in a consensus, in accord with Wheeler et al. (1993).

Enicocephalomorpha and Dipsocoromorpha were recovered as the basal clade of Heteroptera. The sister relationship between them was not only supported by high support values but also by the result of CONSEL (Shimodaira and Hasegawa, 2001). This result contradicts the opinion that Enicocephalomorpha and Dipsocoromorpha represent two stand-alone basal groups of Heteroptera. To distinguish the robustness of these two competing hypotheses, phylogenomic analyses with complete taxon sampling are expected in the future.

In order to test the robustness of the infraordinal relationships within Heteroptera, we carried out additional analyses based on a smaller data matrix consisting only of the taxa having both nucleotide sequences of nrDNAs and amino acid sequences of the mitochondrial PCGs. The topologies of the BI, ML and MP analyses are the same in the basal parts of the results as those based on the original dataset except for the position of Gerromorpha in the BI and ML analyses (Figs S8 and S9). In the BI analysis, the sister relationships between Gerromorpha and the clade formed by Enicocephalomorpha and Dipsocoromorpha received an 86% posterior probability, which likely suffered from the improper application of amino acids substitution models as well. Meanwhile, in the ML analysis, Gerromorpha formed a group with Enicocephalomorpha and Dipsocoromorpha with a bootstrap value of 46%. However, according to the CONSEL results, the *P*-value of the AU test of the clade (Gerromorpha + (Enicocephalomorpha + Dipsocoromorpha)) was 0.560. Moreover, the *P*-value for the clade Neoheteroptera was 0.440, which is much higher than the significance level of 0.05. Thus, the possibility of Gerromorpha forming a monophyletic group with Panheteroptera is not rejected in the ML analysis. In addition, Cimicomorpha was not recovered as a monophyletic group in the results of MP analysis with reduced taxon sampling. The changes in the positions of Gerromorpha in the ML analysis and Reduviidae in the MP analysis may have been a consequence of fewer sampled taxa. Thus, in general, the missing mitochondrial genomes of some species have only a little impact on the phylogenetic reconstruction of deep nodes within Heteroptera. The topological robustness provides a sound foundation for the temporal estimation of the divergence of Heteroptera, especially the deep nodes of aquatic bugs.

Divergence dates and the evolution of aquatic bugs

Efforts have been made in two areas to avoid the negative impacts of single calibration points (Ho and Phillips, 2009) and hard bound or excessive restrictions on priors (Welch et al., 2005; Sanders and Lee, 2007; Ho and Phillips, 2009; Heled and Drummond, 2012). One is the multiple calibration based on 15 fossil records, and the other is the soft bounds on priors. These improvements help to obtain a chronogram with relatively small-scale 95% CIs. At the same time, such multiple calibrations did not result in the possible side effect of overparameterization (File S2). Moreover, the missing data had no significant effects on the estimation of divergence times.

Our estimates of the timespans for the deeper divergences of Gerromorpha, Nepomorpha and Leptopodomorpha within Heteroptera occurred within a short time span, from 269 to 246 Ma. The extant superfamilies of Nepomorpha diversified from a common ancestor in the Middle and Late Triassic, and further diversification within each superfamily occurred in the early Jurassic. These results closely correspond to those drawn from a thorough study of the fossil record by Popov (1971). They are also consistent with the evolutionary backgrounds of the aquatic holometabolans, such as the aquatic lineages of extant lower Diptera (Bertone and Wiegmann, 2009; Wiegmann et al., 2011), which also originated in the Late Permian (*c.* 267 Ma) and radiated during the Middle and Late Triassic. According to the fossil records of both the adults and larvae, megalopterans have been in existence since the Late Permian (Sinitshenkova, 2003). In addition, even though the divergence time of the Hydradephaga beetles (237 ± 2.6 Ma) (Hunt et al., 2007; Bertone and Wiegmann, 2009) occurred somewhat later than for other aquatic insects, it is also in accordance with the palaeoenviromental changes across this time span.

The simultaneously convergent evolution in various lineages of aquatic heteropterans and holometabolans indicate the emergence of more aquatic niches during that timespan. Although the rising air temperature may not necessarily have resulted in environments rich in water for some regions, it is true that highly diversified aquatic habitats appeared more frequently in warmer regions from a global view. Therefore, it is still of interest to notice that the average air temperature increased from 21 °C to approximately 27 °C within a period of 16 Myr, as observed from the trend in the Late Permian (Fig. 3) (http://www.scotese.com/climate.htm). The Late Permian and Early Triassic witnessed one of the hottest climates in the Phanerozoic (Sun et al., 2012), which thus probably caused the emergence of further aquatic environments, at least in some peripheral regions of the Pangaea. The newly emerging aquatic habitats were rapidly colonized by heteropterans of three aquatic infraorders through adaptive radiation, very probably during the Late Permian and Early Triassic.

In a recent study that also explored deep divergence times within Heteroptera, the estimations based on BEAST were calibrated by five fossil records, two of which had uniform distributions and two with normal distributions; the root was given a fixed age (Li et al., 2012). The selected fossil records were not the earliest ones with certainty, and the uniform distributions and the fixed time for the root were hard bounds. The results of that study suggested that Nepomorpha originated approximately 250 Ma in the earliest Triassic. Among the several problems with this study method, the most serious one is the fixed time for the root. According to the phylograms of Li et al. (2012), the infraorder Nepomorpha is the basal clade within Heteroptera, and the earliest divergence of Heteroptera and the origin of Nepomorpha they discussed is in fact the value they set *a priori*. Their phylogenetic hypothesis is also quite different from the generally accepted one (Schuh, 1979; Wheeler et al., 1993) and ours, except for the sister relationship between Cimicomorpha and Pentatomomorpha. The main aquatic lineages within Heteroptera were interrupted, which resulted in the suggestion of much younger origins of Gerromorpha and Leptopodomorpha (about 217 Ma or much younger), and more complicated evolutionary pathways on the origins and diversifications of the aquatic lineages in Heteroptera.

Within the phylogenetic context of Heteroptera, Terheteroptera represent the most derived infraorders next to the positions of the aquatic lineages. Together, they account for more than 85% of the species diversity of Heteroptera (Henry, 2009) and most of them are phytophagous. For the split between Cimicomorpha and Pentatomomorpha, our estimated timespan is in the Carnian stage (231 Ma, 95% HPD 217–245), which is older than the corresponding result of 185.3 ± 34 Ma in a previous study by Li et al. (2012). The estimated time of the radiation within either Cimicomorpha or Pentatomomorpha is also earlier than the corresponding result in the same paper. Most of the current superfamilies arose during the Jurassic (Fig. 3). The radiation of the phytophagous holometabolans leading to most of the current superfamilies also occurred in the Jurassic (Hunt et al., 2007; Bertone and Wiegmann, 2009; Winterton et al., 2010; Wiegmann et al., 2011). This timescale is consistent with the hypothesized late Jurassic origin of angiosperms based on molecular dating (Smith et al., 2010). The diversification of phytophagous insects would have benefitted substantially from the nutritious resources, as well as from the various heterogeneous niches provided by the early radiation of angiosperm. Therefore, as is the case for the aquatic groups, the timescale of the diversification within Terheteroptera is again concordant with that of the phytophagous holometabolan groups. Together with the result for the aquatic groups, such congruence strengthens the overall credibility of the timescale estimated in this study.

Aquatic Heteroptera originated and established across a short timescale from the Late Permian to the Early Triassic (269–246 Ma). This means that the aquatic habits of true bugs have evolved under a continuous and steady evolutionary background characterized by high air temperature and a water-rich environment. As the estimated timespan is in good accordance with that of the aquatic holometabolans, the origin and diversification of the aquatic bugs within Heteroptera may serve as the last jigsaw piece for the aquatic insects, which originated from the Late Permian and Early Triassic.

Acknowledgements

We thank Cuiqing Gao, Zhen Ye, Rui Wang, Xiping Zhang and Yiran Mu for their identification of species. We also thank Gexia Qiao and Sanan Wu for supporting specimens. This project was supported by the National Science Foundation Projects of China (grant numbers: 31222051, J1210005, 31372240) and the Czech Science Foundation (PŠ: grant number P505/11/1459). The authors declare that they have no conflicting interests.

References

Azar, D., Nel, A., 2010. The earliest fossil schizopterid bug (Insecta: Heteroptera) in the Cretaceous, Lower amber of Lebanon. Ann. Soc. Entomol. Fr. 46, 193–197.

Azar, D., Fleck, G., Nel, A., Solignac, M., 1999. A new enicocephalid bug, *Enicocephalinus acragrimaldii* gen. nov., sp. nov., from the Lower Cretaceous amber of Lebanon (Insecta, Heteroptera, Enicocephalidae). Estud. Mus. Cienc. Nat. Álava 14, 217–230.

Barker, S.C., Whiting, M., Johnson, K.P., Murrell, A., 2003. Phylogeny of the lice (Insecta, Phthiraptera) inferred from small subunit rRNA. Zool. Scr. 32, 407–414.

Bertone, M.A., Wiegmann, B.M., 2009. True flies (Diptera). In: Hedges, S.B., Kumar, S. (Eds.), The Timetree of Life. Oxford University Press, Oxford, pp. 270–277.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537. doi:10.1371/journal.pcbi.1003537.

Campbell, B.C., Steffen-Campbell, J.D., Sorensen, J.T., Gill, R.J., 1995. Paraphyly of Homoptera and Auchenorrhyncha inferred from 18S rDNA nucleotide sequences. Syst. Entomol. 20, 175–194.

Cassis, G., Schuh, R.T., 2010. Systematic methods, fossils, and relationships within Heteroptera (Insecta). Cladistics 26, 262–280.

Chen, P., Nieser, N., Zettel, H., 2005. The Aquatic and Semi-Aquatic Bugs (Heteroptera: Nepomorpha & Gerromorpha) of Malesia. Fauna Malesiana Handbook Vol. 5. Brill, Leiden.

Damgaard, J., 2008. Evolution of the semi-aquatic bugs (Hemiptera: Heteroptera: Gerromorpha) with a re-interpretation of the fossil record. Acta Entomol. Musei Nat. Prag. 4, 251–268.

Drummond, A., Ho, S.Y., Phillips, M., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88.

Fischer, C., Mahner, M., Wachmann, E., 2000. The rhabdom structure in the ommatidia of the Heteroptera (Insecta), and its phylogenetic significance. Zoomorphology 120, 1–13.

Fraser, N.C., Grimaldi, D.A., Olsen, P.E., Axsmith, B., 1996. A Triassic Lagerstatte from eastern North America. Nature 380, 615–619.

Goloboff, P.A., Farris, J.S., Nixon, K.C., 2008. TNT, a free program for phylogenetic analysis. Cladistics 24, 774–786.

Gradstein, F.M., Ogg, J.G., Schmitz, M., Ogg, G., 2012. The Geologic Time Scale 2012. Elsevier, Amsterdam, The Netherlands.

Grimaldi, D., Engel, M.S., 2005. Evolution of the Insects. Cambridge University Press, New York.

Gutell, R.R., Lee, J.C., Cannone, J.J., 2002. The accuracy of ribosomal RNA comparative structure models. Curr. Opin. Struct. Biol. 12, 301–310.

Hebsgaard, M.B., Andersen, N.M., Damgaard, J., 2004. Phylogeny of the true water bugs (Nepomorpha: Hemiptera-Heteroptera) based on 16S and 28S rDNA and morphology. Syst. Entomol. 29, 488–508.

Heled, J., Drummond, A.J., 2012. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. Syst. Biol. 61, 138–149.

Henry, T.J., 1997. Phylogenetic analysis of family groups within the infraorder Pentatomomorpha (Hemiptera: Heteroptera), with emphasis on the Lygaeoidea. Ann. Entomol. Soc. Am. 90, 275–301.

Henry, T.J., 2009. Biodiversity of the Heteroptera. In: Foottit, R.G., Adler, P.H. (Eds.), Insect Biodiversity: Science and Society. Wiley-Blackwell, Oxford, pp. 223–263.

Ho, S.Y., Phillips, M.J., 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Syst. Biol. 58, 367–380.

Hou, W.J., Yao, Y.Z., Zhang, W.T., Ren, D., 2012. The earliest fossil flower bugs (Heteroptera: Cimicomorpha: Cimicoidea: Vetanthocoridae) from the Middle Jurassic of Inner Mongolia, China. Eur. J. Entomol. 109, 281–288.

Hua, J., Li, M., Dong, P., Cui, Y., Xie, Q., Bu, W., 2008. Comparative and phylogenomic studies on the mitochondrial genomes of Pentatomomorpha (Insecta: Hemiptera: Heteroptera). BMC Genom. 9, 610.

Hua, J., Li, M., Dong, P., Cui, Y., Xie, Q., Bu, W., 2009. Phylogenetic analysis of the true water bugs (Insecta: Hemiptera: Heteroptera: Nepomorpha): evidence from mitochondrial genomes. BMC Evol. Biol. 9, 134.

Hug, L.A., Roger, A.J., 2007. The impact of fossils and taxon sampling on ancient molecular dating analyses. Mol. Biol. Evol. 24, 1889–1897.

Hunt, T., Bergsten, J., Levkanicova, Z., Papadopoulou, A., St. John, O., Wild, R., Hammond, P.M., Ahrens, D., Balke, M., Caterino, M.S., Gómez-Zurita, J., Ribera, I., Barraclough, T.G., Bocakova, M., Bocak, L., Vogler, A.P., 2007. A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. Science 318, 1913–1916.

Jobb, G., 2011. TREEFINDER version of March 2011. Munich, Germany. http://www.treefinder.de/ (accessed on 18 March 2011).

Johnson, K.P., Clayton, D.H., 2000. Nuclear and mitochondrial genes containsimilar phylogenetic signal for pigeons and doves (Aves: Columbiformes). Mol. Phylogenet. Evol. 14, 141–151.

Kjer, K.M., 1995. Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs. Mol. Phylogenet. Evol. 4, 314–330.

Kjer, K.M., 2004. Aligned 18S and insect phylogeny. Syst. Biol. 53, 506–514.

Latreille, P.A., 1802. Histoire naturelle, générale et particulière des crustacés et des insectes, 3, i–xii, 1–467. Dufan, Paris.

Li, M., Tian, Y., Zhao, Y., Bu, W., 2012. Higher level phylogeny and the first divergence time estimation of Heteroptera (Insecta: Hemiptera) based on multiple genes. PLoS ONE 7, e32152.

Li, T., Hua, J., Wright, A.M., Cui, Y., Xie, Q., Bu, W., Hillis, D.M., 2014. Long-branch attraction and the phylogeny of true water bugs (Hemiptera: Nepomorpha) as estimated from mitochondrial genomes. BMC Evol. Biol. 14, 99.

Lin, Q.B., 1985. Insect fossils from the Hanshan Formation at Hanshan County, Anhui Province. Acta Palaeontol. Sin. 24, 300–315.

Linder, H.P., Hardy, C.R., Rutschmann, F., 2005. Taxon sampling effects in molecular clock dating: an example from the African Restionaceae. Mol. Phylogenet. Evol. 35, 569–582.

Mahner, M., 1993. Systema cryptoceratorum phylogeneticum (Insecta, Heteroptera). Zoologica (Stuttgart) 48, 1–302.

Misof, B., Niehuis, O., Bischoff, I., Rickert, A., Erpenbeck, D., Staniczek, A., 2006. A hexapod nuclear SSU rRNA secondary-structure model and catalog of taxon-specific structural variation. J. Exp. Zool. B 306B, 70–88.

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A.J., Aspöck, U., Aspöck, H., Bartel, D., Blanke, A., Berger, S., Böhm, A., Buckley, T.R., Calcott, B., Chen, J., Friedrich, F., Fukui, M., Fujita, M., Greve, C., Grobe, P., Gu, S., Huang, Y., Jermiin, L.S., Kawahara, A.Y., Krogmann, L., Kubiak, M., Lanfear, R., Letsch, H., Li, Y., Li, Z., Li, J., Lu, H., Machida, R., Mashimo, Y., Kapli, P., McKenna, D.D., Meng, G., Nakagaki, Y., Navarrete-Heredia, J.L., Ott, M., Ou, Y., Pass, G., Podsiadlowski, L., Pohl, H., von Reumont, B.M., Schütte, K., Sekiya, K., Shimizu, S., Slipinski, A., Stamatakis, A., Song, W., Su, X., Szucsich, N.U., Tan, M., Tan, X., Tang, M., Tang, J., Timelthaler, G., Tomizuka, S., Trautwein, M., Tong, X., Uchifune, T., Walzl, M.G., Wiegmann, B.M., Wilbrandt, J., Wipfler, B., Wong, T.K., Wu, Q., Wu, G., Xie, Y., Yang, S., Yang, Q., Yeates, D.K., Yoshizawa, K., Zhang, Q., Zhang, R., Zhang, W., Zhang, Y., Zhao, J., Zhou, C., Zhou, L., Ziesmann, T., Zou, S., Li, Y., Xu, X., Zhang, Y., Yang, H., Wang, J., Wang, J., Kjer, K.M., Zhou, X., 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346, 763–767.

Morse, J.C., 2009. Biodiversity of aquatic insects. In: Foottit, R.G., Adler, P.H. (Eds.), Insect Biodiversity: Science and Society. Blackwell, Oxford, pp. 165–184.

Ogden, T.H., Rosenberg, M.S., 2007. How should gaps be treated in parsimony? A comparison of approaches using simulation Mol. Phylogenet. Evol. 42, 817–826.

Parham, J.F., Donoghue, P., Bell, C.J., Calway, T.D., Head, J.J., Holroyd, P.A., Inoue, J.G., Irmis, R.B., Joyce, W.G., Ksepka, D.T., Patané, J.S., Smith, N.D., Tarver, J.E., van Tuinen, M., Yang, Z., Angielczyk, K.D., Greenwood, J.M., Hipsley, C.A., Jacobs, L., Makovicky, P.J., Müller, J., Smith, K.T., Theodor, J.M., Warnock, R.C., Benton, M.J., 2012. Best practices for justifying fossil calibrations. Syst. Biol. 61, 346–359.

Perrichot, V., Nel, A., Neraudeau, D., 2005. Gerromorphan bugs in Early Cretaceous French amber (Insecta: Heteroptera): first representatives of Gerridae and their phylogenetic and palaeontological implication. Cretaceous Res. 26, 793–800.

Popov, Y.A., 1971. [Historical development of the hemipteran infraorder Nepomorpha (Heteroptera).] Trudy Paleontol. Inst. Akad. Nauk SSSR 129: 1–230 [in Russian].

Popov, Y.A., Dolling, W.R., Whalley, P.E.S., 1994. British Upper Triassic and Lower Jurassic Heteroptera and Coleorrhyncha (Insecta: Hemiptera). Genus 5, 307–347.

Posada, D., 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 5, 1253–1256.

Reineke, A., Karlovsky, P., Zebitz, C.P.W., 1998. Preparation and purification of DNA from insects for AFLP analysis. Insect Mol. Biol. 7, 95–99.

Reuter, J., Mathews, D., 2010. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics 11, 129.

Rieger, Ch., 1976. Skelett und Muskulatur des Kopfes und Prothorax von Ochterus marginatus Latreille. Beitrag zur Klärung der phylogenetischen Verwandtschaftsbeziehungen der Ochteridae (Insecta, Heteroptera). Zoomorphologie 83, 109–191.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2011. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.

Sanders, K.L., Lee, M.S., 2007. Evaluating molecular clock calibrations using Bayesian analyses with soft and hard bounds. Biol. Lett. 3, 275–279.

Schaefer, C.W., 1993. The Pentatomomorpha (Hemiptera: Heteroptera): an annotated outline of its systematic history. Eur. J. Entomol. 90, 105–122.

Scherbakov, D.E., Popov, Y.A., 2002. Superorder Cimicidea Laicharting, 1781. Order Hemiptera Linné, 1758. The bugs, cicadas, plantlice, scale insects, etc. In: Rasnitsyn, A.P., Quicke, D.L.J. (Eds.), History of Insects. Kluwer Academic, Dordrecht, pp. 143–157.

Schuh, R.T., 1979. [Review of] Evolutionary Trends in Heteroptera. Part II. Mouthpart-structures and Feeding Strategies, by R. H. Cobben. Syst. Zool. 28, 653–656.

Schuh, R.T., Slater, J.A., 1995. True bugs of the world (Hemiptera: Heteroptera) classification and natural history. Cornell University Press, Ithaca, NY.

Schuh, R.T., Štys, P., 1991. Phylogenetic analysis of cimicomorphan family relationships (Heteroptera). J. New York Entomol. Soc. 99, 298–350.

Schuh, R.T., Weirauch, C., Wheeler, W.C., 2009. Phylogenetic relationships within the Cimicomorpha (Hemiptera: Heteroptera): a total-evidence analysis. Syst. Entomol. 34, 15–48.

Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247.

Sinitshenkova, N.D., 2003. Main ecological events in aquatic insects history. Acta Zool. Crac. 46, 381–392.

Smith, S.A., Beaulieu, J.M., Donoghue, M.J., 2010. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. Proc. Natl Acad. Sci. USA 107, 5897–5902.

Stadler, T., Kühnert, D., Bonhoeffer, S., Drummond, A.J., 2013. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proc. Natl Acad. Sci. USA 110, 228–233.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Štys, P., Kerzhner, I., 1975. The rank and nomenclature of higher taxa in recent Heteroptera. Acta Entomol. Bohemoslov. 72, 65–79.

Sun, Y., Joachimski, M.M., Wignall, P.B., Yan, Ch., Chen, Y., Jiang, H., Wang, L., Lai, X., 2012. Lethally hot temperatures during the early Triassic greenhouse. Science 338, 366–370.

Sweet, M.H., 1996. Comparative external morphology of the pregenital abdomen of the Hemiptera. In: Schaefer, C.W. (Ed.), Studies on Hemipteran Phylogeny. Thomas Say Publications in Entomology: Proceedings. Entomological Society of America, Lanham, MD, pp. 119–158.

Sweet, M.H., 2006. Justification for the Aradimorpha as an infraorder of the suborder Heteroptera (Hemiptera, Prosorrhyncha) with special reference to the pregenital abdominal structure. Denisia 19, 225–248.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.

Tian, Y., Zhu, W., Li, M., Xie, Q., Bu, W., 2008. Influence of data conflict and molecular phylogeny of major clades in Cimicomorphan true bugs (Insecta: Hemiptera: Heteroptera). Mol. Phylogenet. Evol. 47, 581–597.

Tian, X., Xie, Q., Li, M., Gao, C., Cui, Y., Xi, L., Bu, W., 2011. Phylogeny of pentatomomorphan bugs (Hemiptera-Heteroptera: Pentatomomorpha) based on six Hox gene fragments. Zootaxa 2888, 57–68.

Tillyard, R.J., 1926. Upper Permian insects of New South Wales Part 1. Introduction and the order Hemiptera. Proc. Linn. Soc. N.S.W. 51, 1–30.

Wang, Y., Engel, M.S., Rafael, J.A., Dang, K., Wu, H., Wang, Y., Xie, Q., Bu, W., 2013. A unique box in 28S rRNA is shared by the enigmatic insect order Zoraptera and Dictyoptera. PLoS ONE 8, e53679.

Weirauch, C., Schuh, R.T., 2011. Systematics and evolution of Heteroptera: 25 years of progress. Ann. Rev. Entomol. 56, 487–510.

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

Welch, J.J., Fontanillas, E., Bromham, L., 2005. Molecular dates for the "Cambrian explosion": the influence of prior assumptions. Syst. Biol. 54, 672–678.

Wheat, C.W., Wahlberg, N., 2013. Phylogenomic insights into the Cambrian explosion, the colonization of land and the evolution of flight in arthropoda. Syst. Biol. 62, 93–109.

Wheeler, W.C., Schuh, R.T., Bang, R., 1993. Cladistic relationships among higher groups of Heteroptera: congruence between morphological and molecular data sets. Entomol. Scand. 24, 121–137.

Wiegmann, B.M., Trautwein, M.D., Winkler, I.S., Barr, N.B., Kim, J.W., Lambkin, C., Bertone, M.A., Cassel, B.K., Bayless, K.M., Heimberg, A.M., Wheeler, B.M., Peterson, K.J., Pape, T., Sinclair, B.J., Skevington, J.H., Blagoderov, V., Caravas, J., Kutty, S.N., Schmidt-Ott, U., Kampmeier, G.E., Thompson, F.C., Grimaldi, D.A., Beckenbach, A.T., Courtney, G.W., Friedrich, M., Meier, R., Yeates, D.K., 2011. Episodic radiations in the fly tree of life. Proc. Natl Acad. Sci. USA 108, 5690–5695.

Winterton, S.L., Hardy, N.B., Wiegmann, B.M., 2010. On wings of lace: phylogeny and Bayesian divergence time estimates of Neuropterida (Insecta) based on morphological and molecular data. Syst. Entomol. 35, 349–378.

Xie, Q., Tian, Y., Zheng, L., Bu, W., 2008. 18S rRNA hyper-elongation and the phylogeny of Euhemiptera (Insecta: Hemiptera). Mol. Phylogenet. Evol. 47, 463–471.

Xie, Q., Tian, X.X., Qin, Y., Bu, W.J., 2009. Phylogenetic comparison of local length plasticity of the small subunit of nuclear rDNAs among all Hexapoda orders and the impact of hyper-length-variation on alignment. Mol. Phylogenet. Evol. 50, 310–316.

Yao, Y.Z., Cai, W.Z., Ren, D., 2006. Systematics, in New fossil rhopalids (Heteroptera: Coreoidea) from the Middle Jurassic of Inner Mongolia, China. Zootaxa 1384, 41–58.

Yao, Y.Z., Cai, W.Z., Ren, D., 2007a. The first fossil Cydnidae (Hemiptera: Pentatomoidea) from the Late Mesozoic of China. Zootaxa 1388, 59–68.

Yao, Y.Z., Cai, W.Z., Ren, D., 2007b. The oldest known fossil plant bug (Hemiptera: Miridae), from Middle Jurassic of Inner Mongolia, China. Zootaxa 1442, 37–41.

Yao, Y., Ren, D., Rider, D.A., Cai, W., 2012. Phylogeny of the infraorder Pentatomomorpha based on fossil and extant morphology, with description of a new fossil family from China. PLoS ONE 7, e37289.

Zhang, Z., 2011. Phylum Arthropoda von Siebold, 1848. In: Zhang, Z.-Q. (Ed.), Animal Biodiversity: An Outline of Higher-Level Classification and Survey of Taxonomic Richness. Zootaxa 3148, 99–103.

Zrzavý, J., 1992. Evolution of antennae and historical ecology of the hemipteran insects Paraneoptera. Acta Entomol. Bohemoslov. 89, 77–86.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Secondary structure model of the 18S rRNA of Aenictopecheidae gen.sp. (Accession number: KJ461282). The red and blue bases denote the positions of length-variable regions. The blue bases are removed from the data matrix used for phylogenetic constructions. Base pairing is indicated as follows: standard canonical pairs by lines (C-G, G-C, A-U, U-A); wobble G•U pairs by dots (G•U); A○G and A○C pairs by open circles (A○G, A○C); other noncanonical pairs by filled circles (e.g. U•U).

Fig. S2. Secondary structure model of the 28S rRNA 5′-half of Aenictopecheidae gen.sp. (Accession number: KJ461277). The red and blue bases denote the positions of length-variable regions. The blue bases are removed from the data matrix used for phylogenetic constructions. Base pairing is indicated as follows: standard canonical pairs by lines (C-G, G-C, A-U, U-A); wobble G•U pairs by dots (G•U); A○G and A○C pairs by open circles (A○G, A○C); other noncanonical pairs by filled circles (e.g., U•U).

Fig. S3. Secondary structure model of the 28S rRNA 3′-half of Aenictopecheidae gen.sp. (Accession number: KJ461277). The red and blue bases denote the positions of length-variable regions. The blue bases are removed from the data matrix used for phylogenetic constructions. Base pairing is indicated as follows: standard canonical pairs by lines (C-G, G-C, A-U, U-A); wobble G•U pairs by dots (G•U); A○G and A○C pairs by open circles (A○G, A○C); other noncanonical pairs by filled circles (e.g. U•U).

Fig. S4. Phylogram of BI analysis based on complete taxon sampling. The numbers associated with the nodes are posterior probability values.

Fig. S5. Cladogram of MP analysis using jackknife resampling method based on complete taxon sampling. Jackknife support values are indicated at each node.

Fig. S6. Cladogram of MP analysis using Poisson bootstrap reweighting method based on complete taxon sampling. Bootstrap support values are indicated at each node.

Fig. S7. Estimated divergence times of major Heteroptera clades based on 32 taxa. Blue bars indicate 95% mean confidence intervals.

Fig. S8. Phylograms of BI and ML analyses based on 32 taxa. (A) BI topology, (B) ML topology. Posterior probabilities and bootstrap values are shown at each node, respectively.

Fig. S9. Cladograms of MP analysis using different resampling method based on complete taxon sampling. (A) MP topology based on Jackknife resampling method, (B) MP topology based on Poisson reweighting method. Jackknife support values and bootstrap support values are indicated at each node.

Table S1. Taxa sampling of rDNAs and mitochondria genomes.

Table S2. Primer sets used for amplification and sequencing. Naming of the primer sets for rDNAs has no set rules. The specific primers for the mitochondrial genome are named with the first three letters of the genus name. “Pti” was used for amplifying the sequences of *Ptilomera tigrina*, with “Ent” used for *Entomovelia* sp. and “Kok” for *Kokeshia xiei*.

File S1. rDNA matrix with ambiguous alignment regions. The first 2324 nucleotides are 18S rDNA, and the remaining nucleotides are 28S rDNA. Among which the regions 278–305, 857–1019, 1757–1761, 1769–1820, 2196–2234, 1754–2899, 3014–3082, 3277–3310, 4182–4216, 4528–4603, 4838–5243, 5782–5853, 5919–6010, 6423–6448, and 6620–6824 are ambiguous alignment regions.

File S2. Data matrix used for phylogenetic reconstruction and divergence time estimations.

File S3. Prior and posterior marginal distributions. The blue probability distributions indicate the posterior density distributions, while the grey ones denote the prior density distributions.