300 Million years of coral treaders (Insecta: Heteroptera: Hermatobatidae) back to the ocean in the phylogenetic context of Arthropoda

Yan-hui Wang, Yun-xia Luan, Jiu-yang Luo, Yu Men, Michael S. Engel, Jakob Damgaard, Abderrahman Khila, Ping-ping Chen, FelipeFerraz Figueiredo Moreira, José A. Rafael and Qiang Xie

DOI:
https://doi.org/10.1098/rspb.2023.0855

PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Wang Y *et al.* 2023 300 Million years of coral treaders (Insecta: Heteroptera: Hermatobatidae) back to the ocean in the phylogenetic context of Arthropoda. *Proc. R. Soc. B* **290**: 20230855. https://doi.org/10.1098/rspb.2023.0855

Received: 13 December 2022
Accepted: 6 June 2023

Subject Category:
Evolution

Subject Areas:
evolution, genomics, microbiology

Keywords:
origin of marine insects, genome, phylogenomics, metagenome, symbiotic bacteria, arthropod phylogeny

Author for correspondence:
Qiang Xie
e-mail: xieq8@mail.sysu.edu.cn

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.6700019.

THE ROYAL SOCIETY PUBLISHING

300 Million years of coral treaders (Insecta: Heteroptera: Hermatobatidae) back to the ocean in the phylogenetic context of Arthropoda

Yan-hui Wang[1], Yun-xia Luan[2], Jiu-yang Luo[1], Yu Men[1], Michael S. Engel[3], Jakob Damgaard[4], Abderrahman Khila[5], Ping-ping Chen[6], Felipe Ferraz Figueiredo Moreira[7], José A. Rafael[8] and Qiang Xie[1]

[1]State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, 135 Xingangxi Road, Guangzhou 510275, Guangdong, China
[2]Guangdong Provincial Key Laboratory of Insect Developmental Biology and Applied Technology, Institute of Insect Science and Technology, School of Life Sciences, South China Normal University, 55 West of Zhongshan Avenue, Guangzhou 510631, China
[3]Division of Entomology, Natural History Museum, and Department of Ecology and Evolutionary Biology, University of Kansa, 1501 Crestline Drive – Suite 140, Lawrence, KS 66045, USA
[4]Natural History Museum of Denmark, Zoological Museum, Universitetsparken 15, 2100 Copenhagen Ø, Denmark
[5]Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, Université Claude Bernard Lyon1, CNRS UMR 5242, Ecole Normale Supérieure de Lyon, 46, allée d'Italie, 69364 Lyon Cedex 07, France
[6]Netherlands Centre of Biodiversity Naturalis, 2300 RA, Leiden, The Netherlands
[7]Laboratório de Biodiversidade Entomológica, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil
[8]Instituto Nacional de Pesquisas da Amazônia, INPA, Caixa Postal 478, 69011-970 Manaus, Amazonas, Brazil

YL, 0000-0003-3573-7144; AK, 0000-0003-0908-483X; QX, 0000-0001-6376-8808

Among hundreds of insect families, Hermatobatidae (commonly known as coral treaders) is one of the most unique. They are small, wingless predaceous bugs in the suborder Heteroptera. Adults are almost black in colour, measuring about 5 mm in body length and 3 mm in width. Thirteen species are known from tropical coral reefs or rocky shores, but their origin and evolutionary adaptation to their unusual marine habitat were unexplored. We report here the genome and metagenome of *Hermatobates lingyangjiaoensis*, hitherto known only from its type locality in the South China Sea. We further reconstructed the evolutionary history and origin of these marine bugs in the broader context of Arthropoda. The dated phylogeny indicates that Hexapoda diverged from their marine sister groups approximately 498 Ma and that Hermatobatidae originated 192 Ma, indicating that they returned to an oceanic life some 300 Myr after their ancestors became terrestrial. Their origin is consistent with the recovery of tropical reef ecosystems after the end-Triassic mass extinction, which might have provided new and open niches for them to occupy and thrive. Our analyses also revealed that both the genome changes and the symbiotic bacteria might have contributed to adaptations necessary for life in the sea.

1. Background

Among the over one million described species of Hexapoda, the marine insects comprise no more than 0.05% [1]. Hermatobatidae (Insecta: Heteroptera: Gerromorpha) is one of an exceptionally few insect families that include only marine species. They live in the intertidal zone of rocky and coral reefs with cavities [2]. During low tides, they tread very rapidly on the sea surface. As a result, Carpenter [3] coined the name for the sole included genus *Hermatobates* and therefore the family after the coral reefs and the god Hermes, who is the fastest

in all the Greek pantheon and has the ability to freely move between the mortal and divine worlds. Observations of the first author show that some individuals of *H. lingyangjiaoensis* can be found under dead shells together with moulted cuticles. These intriguing insects are carnivorous, usually preying on small marine arthropods [4]. The first author has observed *H. lingyangjiaoensis* prey on *Halovelia* (Veliidae) and *Halobates* (Gerridae), even dead crickets under laboratory feeding conditions.

Hermatobatidae contains a single genus *Hermatobates*, with 13 described species [5,6]. They are limited to the tropical regions of the world with only a few species occurring beyond the boundaries of this zone. Some species, like *H. weddi* and *H. djiboutensis* are widely distributed; some species, like *H. singaporensis* and *H. hawaiienses*, are regional; and some species, like *H. lingyangjiaoensis*, *H. armatus*, and *H. palmyra* are likely endemic. Hermatobatids have a dark, compact body, and are wingless, thereby resembling representatives of other gerromorphan families with marine members, such as Halobatinae and Trepobatinae (both Gerridae) and Haloveliinae (Veliidae). However, hermatobatids have some unique characteristics, including a reduced number of nymphal instars (four compared to normally five) [7], and modified double tripod gait (compared to synchronous rowing of Gerridae and Veliidae) [8]. During high tide, they retreat to the crevices of coral reefs, and remain secluded for several days when coinciding with typhoons or other extreme weather (personal observation of the first and the third authors in 2019). Conversely, during low tide they swim on the sea surface in a tripod rowing mode, and adults even set foot on the open sea. Unlike other marine bugs, such as *Halobates* and *Halovelia*, the behaviour of coral treaders is largely controlled by the tidal cycles [2,4,9]. On the one hand, this biology is likely to favour optimizing resources both in the intertidal zone and open sea [4], while on the other, it means that hermatobatids face dual challenges from both low oxygen concentrations when they were flooded by high tide [2,7] and strong ultraviolet radiation when they swim on the open sea without shelter [10,11]. Noting that the zone of suitable habitats for the coral treaders is restricted to a narrow (approx. 20 cm) vertical band [7]. As global warming and sea levels rise, their habitats will gradually disappear. To date, little is known about the underlying molecular mechanism of their adaptive evolution to the extreme of marine habitats.

Extant arthropods comprise three widely known lineages—chelicerates, myriapods, pancrustaceans, which comprises hexapods and "crustaceans" [12]. Among Chelicerata, Pycnogonida (sea spiders) and Xiphosura (horseshoe crabs) consist of exclusively marine species. In Myriapoda and Hexapoda, all of the extant groups are strictly terrestrial. The "crustaceans", by contrast, include a vast diversity of mostly marine lineages, with Remipedia as the putative sister group to Hexapoda [12,13]. For Chelicerata and/or "Crustacea", the origin and evolutionary history of marine groups have been extensively explored in context of broader arthropod phylogeny [12–14]. By contrast, the origin and macroevolutionary factors driving marine insect diversification have remained unexplored, a reality that is at least partly due to the rarity and scarcity of strictly marine insects. Although strictly marine insects are rare considering the huge diversity of insect species, such a habitat is relatively common for semi-aquatic bugs (Gerromorpha). They have a considerable number of taxa associated with marine environments ranging from estuaries, mangroves, river deltas, lagoons and other coastal habitats to the tropical oceans, where these remarkable species, as the only pelagic insects, live their entire lives hundreds of kilometres from the nearest coast [1,15]. The marine lineages of Gerromorpha mainly belong to the superfamily Gerroidea (including Hermatobatidae, Gerridae and Veliidae), among which Hermatobatidae is one of the few exclusively marine family. Andersen [15] investigated the multiple transitions from freshwater to seawater in Gerromorpha, and hypothesized that Hermatobatidae was the first lineage of Gerromorpha to colonize the marine environment.

The phylogenetic affinities of Hermatobatidae are obscure among semi-aquatic lineages of true bugs (Heteroptera). Morphological evidence supports a close relationship between the families Hermatobatidae, Gerridae and Veliidae (as superfamily Gerroidea) [15,16]. Likewise, phylogenetic analyses based a combined analysis of morphology and two short DNA fragments (16S + 28S rRNAs) also supported a sister relationship between Hermatobatidae and the clade (Gerridae + Veliidae) [17]. However, when increasing the sampling of critical gerromorphan taxa and adding more DNA fragments (COI + II), Damgaard [18] recovered Hermatobatidae as sister to Hydrometridae instead. In order to better refine the phylogenetic context of hermatobatid marine evolution, a more integrated multiomic approach is necessary.

Both the genome and metagenome can reflect the adaptation of multicellular organisms to certain environments [19]. Among insects, a few species have been studied from both perspectives [20–23]. We expect to uncover the evolutionary adaptation of coral treaders to harsh marine habitat from aspects of genomic changes and metagenomic functions. Here, we report the high-quality genome of *Hermatobates lingyangjiaoensis*, which is the first such data for an intertidal insect. We also provided transcriptomic data for seven species. We then assembled a large-scale phylogenomic dataset, including 3052 orthologous genes from 156 representative species to investigate the phylogeny of Hermatobatidae in the broader context of Arthropoda. Comparative genomic analyses with other insects indicate significantly contracted gene families functioning in anatomical development and morphogenesis. We also assessed the biodiversity of symbiotic microorganisms and assembled two near-complete genomes of bacteria from genomic and metagenomic data, i.e. *Wolbachia* sp. (wHlin) and *Serratia* sp. (sHlin). The symbiotic bacteria are vital in host carbohydrate, lipid and vitamin metabolism.

2. Material and methods

(a) Sample collection and genome sequencing

The adults of *Hermatobates lingyangjiaoensis* were collected from the Xisha Islands in the South China Sea (16.4473° N, 111.6061° E, Hainan, China). A total of 20 female individuals were prepared for genome sequencing. Genomic DNA extraction, library preparation, and sequencing were carried out at Biomarker Technologies (Beijing, China). The paired-end libraries for short-read sequencing with an insert size of 350 bp were constructed according to the manufacturer's protocols and sequenced (2 × 150) using the Illumina NovaSeq 6000 platform. The single molecule real-time SMRTbell libraries for PacBio sequencing were constructed with an insert size of 20 kb according to the manufacturer's protocols and sequenced on the PacBio Sequel II platform.

(b) Genome assembly and gene annotation

De novo genome assembly with continuous long reads (CLR) was performed using Canu v. 1.5 [24] and wtdbg2 [25]. Then the quality of the assembly was measured using Quast v. 5.0.2 [26]. To assess the completeness of assemblies, we applied Benchmarking Universal Single-Copy Orthologous (BUSCO) analyses using BUSCO v. 5.2.2 with OrthoDB v. 10 datasets [27]. In addition, we also mapped Illumina short reads to the final genome assembly with Minimap v. 2.23 [28]. Three approaches were used to predict protein-coding genes (PCGs): (1) *ab initio* gene prediction, (2) homology-based prediction, and (3) RNA-sequencing annotation. Detailed gene prediction and functional annotation methods see electronic supplementary materials and methods file S1. Moreover, a draft genome of *Sinentomon* sp. (Hexapoda: Protura) was assembled and the PCGs were predicted with the same method (see electronic supplementary material, file S1).

(c) Sample preparation, assembly and binning of symbiotic bacteria

In order to avoid contamination from environmental microorganisms, we dissected the abdomen and removed the whole intestinal tract for each individual. The intestinal tracts of 60 adults were mixed into a single sample. Genomic DNA extraction and library preparation followed the standard protocol provided by ONT (Oxford Nanopore Technologies) and then was sequenced on the PromethION48 platform (see electronic supplementary material, file S1). The raw sequencing data were firstly preprocessed to remove adapters, low-quality reads and sequences shorter than 2 kb. Then the suspected host and human genomes were filtered out using Minimap and SAMtools v. 1.14 [29]. The former software was used to successively map the processed clean data to host and human genomes and the latter was used to extract the reads that could not be matched. Moreover, we extracted the Illumina short reads of *H. lingyangjiaoensis* which could not be aligned to host genome assemblies using the same method as mentioned above. Compositions of these reads were assessed using Kraken v. 2.1.2 [30].

Both short- and long-reads was assembled using the hybrid assembler OPERA-MS v. 0.9.0 [31]. PCGs of the final assembled contigs were predicted using Prodigal v. 2.6.3 [32] 'option: -p meta'. Genome binning was carried with MetaBat2, MaxBin2 and Concoct, which were imbedded within MetaWRAP v. 1.3.2 [33]. For detailed methods, see electronic supplementary materials and methods file S1.

(d) Taxon sampling, transcriptome sequencing and assembly

Our taxon sampling included 156 arthropod species (electronic supplementary material, table S1), which covered all of the three main groups of Arthropoda, especially all extent orders of Hexapoda. Among the 156 species, 35 species have genomes (including the newly sequenced genome of *H. lingyangjiaoensis*), which were also used to conduct gene-family expansion and contraction analyses, while the remaining species have transcriptomes. We newly sequenced transcriptomes for the following seven species: *Zorotypus shannoni* (Zorotypidae), *Lucihormetica fenestrate* (Blaberidae), *Coleotroctellus burckhardti* (Troctopsocidae), *Gerris* sp. (Gerridae), *Metrocoris* sp. (Gerridae), *Amemboa* sp. (Gerridae), *Perittopus* sp. (Veliidae) (electronic supplementary material, table S2). For each species, total RNA was extracted from adults with abdomen and wings removed. The quality control, cDNA library preparation, transcriptome sequencing and raw data preprocessing followed Wang *et al*. [34,35]. The remaining high-quality reads were assembled to long contigs using Trinity v. 2.0.6 [36] and then further clustered to non-redundant transcript sequences using the TGI Clustering tool v. 2.1 [37].

(e) Orthology prediction

As the number of taxa in a dataset increases, the number of single-copy genes tends to decrease because of gene duplication or mis-annotation of genes in any given lineage; thus, low-copy nuclear genes (LCN) were also selected as phylogenetic markers in this study. LCN genes were identified using OrthoFinder v. 2.5.2 [38] (option: -M msa). The homologues were obtained from the following 12 reference species: *Daphnia pulex*, *Catajapyx aquilonaris*, *Folsomia candida*, *Sinentomon* sp., *Acyrthosiphon pisum*, *Thrips palmi*, *Hermatobates lingyangjiaoensis*, *Pediculus humanus*, *Apis mellifera*, *Blattella germanica*, *Drosophila melanogaster*, *Zootermopsis nevadensis*, which were categorized into three groups successively. LCN homologous groups (HGs) needed to meet the following requirements: copy number less than or equal to two; at least one species for each group present in each HG (see flowchart in electronic supplementary material, figure S1). There were 3952 high-quality putative target homologous genes (including 1339 single-copy and 2613 two-copy genes) that were used as references to search for homologues in the 156 arthropod genomes and transcriptomes using Orthograph v. 0.6.3 [39] with the best-reciprocal hit (BRH) criterion. Moreover, we applied a non-strict reciprocal search and allowed frameshift correction and open reading frame (ORF) extension of each transcript (option: extend-orf = 1). All other Orthograph parameters were left at default values. Finally, the accompanying script summarize_orthograph_results.pl was used to summarize the sequences from multiple Orthograph output directories for all taxa. During this process, the sequences of the reference taxa were deleted (option: -d), and we defined the PCGs hereafter as orthologous groups (OGs).

Then, we first filtered the preliminary OGs and selected only those genes with more than half representatives. Then, sequences for each OG were aligned with MAFFT v. 7.222 using the L-INS-i algorithm [40] at the translational level and generated the corresponding nucleotide multiple-sequence alignments using a local version of the software TranslatorX v. 1.1 [41]. We removed the ambiguously aligned sites from both protein and nucleotide alignments using GBlocks v. 0.91b [42] (options: −b3 = 8, −b4 = 5, −b5 = h) in TranslatorX and further trimmed low-quality aligned sequences using trimAL v. 1.4 [43]. Next, we inspected each tree for rogue taxa with unrealistically long branches using RAxML v. 8.2.10 [44] (options: -f a, -# 200). Finally, we retained 3052 PCGs in the following analyses, among which 1013 PCGs were single-copy genes in the 12 species mentioned above. Detailed information for orthologous genes analysed for each species in the final concatenated matrix were provided (electronic supplementary material, figure S2).

(f) Phylogenetic analyses

We reconstructed the Arthropoda phylogeny using both protein and nucleotide datasets using concatenation and coalescent methods. As the third codon positions exhibited a high level of variation among lineage compositional heterogeneity (electronic supplementary material, figure S3), only the first two codon positions were used. We used ModelFinder, which is embedded within IQ-TREE v. 1.6.10 [45,46] to identify the best partitioning schemes and determine the best amino acid substitution model for each partition with the corrected Akaike information criterion (AICc) (see electronic supplementary material, file S1). The nucleotide partitioning scheme had the same partitions as the amino acid partitioning scheme. ML analyses were conducted using RAxML v. 8.2.10 with the Pthreads version [44]. The best ML and bootstrap trees were inferred through a rapid bootstrap algorithm (option: -f a) with 200 replicates. For coalescent-based

analyses, we determined the best-fitting model for each gene using IQ-TREE as well. Gene trees were estimated using RAxML with 200 rapid bootstrap replicates and then used by ASTRAL v. 5.7.1 [47] to infer species trees with posterior probabilities. This approach is thought to be more robust to incomplete lineage sorting (ILS) or deep coalescence than concatenation and works quickly on genome-scale datasets.

We assessed the support or conflict for alternative hypotheses by four-cluster likelihood mapping (FcLM) analysis using IQ-TREE. We evaluated five phylogenetic hypotheses concerning the backbone of arthropods and the groups for the FcLM analyses are given (electronic supplementary material, table S3). The likelihood of each quartet is presented as a triangle, the corners of which represent the three possible alternative topologies. We applied the LG + Γ model in the FcLM test with amino acid sequences (four gamma rates), while employing the GTR + Γ in nucleotide sequence sites with 2000 quartets.

(g) Divergence time estimation

To estimate the evolutionary timescale of Arthropoda, we calibrated a relaxed molecular clock using 23 fossil-based age constraints throughout the tree (electronic supplementary material, table S4). Explicit descriptions for each fossil record are provided in the electronic supplementary material, file S2. For the age justification of each fossil record, we mainly referred to the Paleobiology Database (https://paleobiodb.org/) and the work of Wolfe *et al*. [48]. To increase the site coverage and alleviate the impact of missing data on divergence time estimation, only the genes present in 95% of the species and that have a length of greater than or equal to 100 amino acids were engaged in dating analyses. Finally, the reduced datasets retained 462 orthologous genes, among which 204 OGs (50 398 amino acid sites) are single copy genes in the 12 reference species. To exclude possible noise from paralogous genes, only the 204 OGs were used to date the deep phylogeny of Arthropoda.

We performed a Bayesian phylogenomic dating analysis using the program MCMCTree in the PAML package v. 4.9j [49]. An approximate likelihood calculation method was used to calculate the branch lengths and reduce the computational burden [50]. We checked for convergence by running the analysis in duplicate and checked for sufficient sampling using Tracer v. 1.6 (http://beast.bio.ed.ac.uk/Tracer).

(h) Gene-family expansion and contraction identification

Amino acid sequences for the genomes of the 34 arthropod species mentioned in the phylogeny part, were retrieved from the NCBI, Ensemble, i5k and USDA databases (electronic supplementary material, table S5). After filtering redundant alternative splicing events, the protein dataset containing non-redundant transcripts was used to find the homologous pairs of sequences using OrthoFinder with blast as the sequence aligner. The processed genome of *H. lingyangjiaoensis* was compared with the other 34 genomes. CAFE v. 4.2 [51] was used to study gene gains and losses in gene families, in which a random birth and death model with the global parameter λ was employed. Species trees and dated time trees were generated according to the methods mentioned above. For the gene families with significant expansions and contractions in *H. lingyangjiaoensis*, we conducted GO and KEGG functional annotation via local eggNOG-mapper v. 2.1.7 [52] and domain annotation with InterProScan v. 5.54-87.0 [53].

To understand gene-family expansion or contraction in *H. lingyangjiaoensis* compared with that in other Arthropoda, the mean gene-family size was calculated for all gene families (excluding single-copy genes and species-specific families). The number of genes per species for each family was transformed into a matrix of *z*-scores to centre and normalize the data. The first 100 families with the largest gene-family size were selected and visualized using TBtools v. 1.098761 [54].

3. Results

(a) Genome assembly and annotation

Both of the Illumina short reads and PacBio Sequel II long reads have a high-depth coverage (greater than 100×) (electronic supplementary material, file S1). The initial draft genome assemblies were 914.67 Mb and 1973 contigs, among which 20 contigs were suspected contaminants. After filtering out contaminants, the final genome assemblies were 912.63 Mb and 1953 contigs (with a contig N50 of 2.56 Mb, L50 of 81, GC content: 31.5%) (electronic supplementary material, table S6). Our assembly had a medium genome size among species of true bugs (0.68–1.55 Gb). Assembly completeness was estimated to be 96.5% ($n = 977$) against an Arthropoda dataset ($n = 1013$). Similarly, the results showed a high-quality assembly against Hemiptera and Insecta databases (electronic supplementary material, table S7). Comparison with other heteropteran genome assemblies, this genome also showed a high completeness (electronic supplementary material, table S8). In addition, 95.44% of the Illumina short reads can be mapped back to the final genome assembly. With a genome-guided strategy, we obtained 26 467 transcripts with a completeness of greater than or equal to 79% (electronic supplementary material, file S1).

Using a combination of *ab initio*, homology-, and transcriptome-based approaches, 14 242 protein-coding genes were predicted, with a mean of 6.37 exons and 5.37 introns per gene (table 1; electronic supplementary material, table S9). Exons and introns had a mean length of 258.49 bp and 2419.51 bp, respectively. For the predicted genes, 12 363 PCGs (86.80% of total) were supported by RNA-seq and homology-based gene prediction. BUSCO analysis identified 966 (95.4%) complete, 18 (1.8%) fragmented and 29 (2.8%) missing BUSCOs comparing with the Arthropoda database. Compared to the Hemiptera and Insecta databases, the results also showed high-quality gene structures (electronic supplementary material, table S7). InterProScan identified protein domains for 11 657 (81.98%) genes. A total of 11 639 genes were successfully annotated by at least one database among KOG, KEGG, GO, Swiss-Prot and TrEMBL.

(b) Phylogenomic analyses and divergence time analyses

Phylogenomic analyses of nucleotide and amino acid sequences with both coalescent and concatenation methods produced largely congruent results in the backbone of Arthropoda (electronic supplementary material, figures S4–S11). For certain conflicting hypotheses generated from different inference methods and previous studies, the quartet supports for three possible topologies are shown in electronic supplementary figures S12 and S13.

The monophyly of each main group of Arthropoda, including Chelicerata, Myriapoda and Pancrustacea was recovered and well supported. Within Pancrustacea, Remipedia, Branchiopoda and Cephalocarida were placed as

Table 1. Statistics for genome annotation of *H. lingyangjiaoensis*.

	type	number/length
gene annotations	genes	14 242 (208.49 Mb)
	gene mean length	14 638.89 bp
	exons	90 651 (23.43 Mb)
	exon mean length	258.49 bp
	average exon number per gene	6.37
	introns	76 409 (184.87 Mb)
	intron mean length	2419.51 bp
	average intron number per gene	5.37

successive sister groups to Hexapoda, congruent with other phylogenomic evidence [12,13]. For early diverging Hexapoda, Collembola and Protura constituted a monophyletic group or two independent lineages, and was sister to the clade Diplura + Insecta (i.e. Euinsecta) in all ASTRAL analyses. Meanwhile, almost all major clades of insects were found to be monophyletic, consistent with previous phylogenomic evidence based on 1000 + genes [55,56]. Our phylogenomic analyses recovered the monophyly of Paraneoptera based on the coalescent approach using amino acid matrices, which is consistent with the topology of Thomas *et al*. [57]. A sister relationship between Thysanoptera and Hemiptera and the internal relationships of Hemiptera were recovered, in congruence with the results of Johnson *et al*. [58]. While Coleorrhyncha (moss bugs) was revealed as the sister taxon to Auchenorrhyncha (leafhoppers, cicadas and relatives), instead of sister to Heteroptera as in Wang *et al*. [34]. Within Heteroptera–Gerromorpha, Hermatobatidae was found to be sister to the remaining Gerromorpha in both concatenation and coalescent analyses, but with moderate posterior probabilities (61–83%) in the ASTRAL analyses. The four-cluster likelihood mapping (FcLM) analysis supports a sister relationship between Hermatobatidae and the clade "Veliidae" + Gerridae with 97.8% quartets using amino acids of 3052 genes and with 98.7% quartets using the first two codons of 3052 genes (electronic supplementary material, figures S12 and S13).

Molecular dating of arthropod lineages using a stringent set of 204 genes and with age calibrations based on 23 fossils inferred the crown age of Arthropoda at 592 million years ago (Ma) (figure 1), with detailed information for each node supplemented (see electronic supplementary material, figures S14 and S15). Divergence times for most nodes were consistent with previous studies (electronic supplementary material, table S10). Radiation of Chelicerata and Pancrustacea started in the Early Cambrian (about 531 Ma and 540 Ma, respectively), while that of Myriapoda occurred in the late Cambrian (about 508 Ma). Hexapods diverged from among their "crustacean" relatives in the Late Cambrian (498 Ma, 95% credibility interval: 484–511 Ma). Divergence times for these deep nodes of Arthropoda were consistent with the results of Giribet & Edgecombe [12]. Diversification of crown-group Hexapoda occurred in the Early Ordovician (483 Ma, 95% credibility interval: 468–497 Ma), which coincided with the result of Schwentner *et al*. [13] and Misof *et al*. [55]. We dated the diversification of Pterygota to the Late Silurian (421 Ma) and that of Neoptera to the Early Devonian (406 Ma), which were a little earlier than the results of Misof *et al*. [55] and Wang *et al*. [59]. Moreover, our estimated divergence times for major clades within Hemiptera were earlier than that of the results of Wang *et al*. [34]. This discrepancy was likely to be biased by the different topology of Hemiptera. The split between Hermatobatidae and "Veliidae" + Gerridae was estimated to have occurred at the Early Jurassic (192 Ma, 95% credibility interval: 165–217 Ma).

(c) Gene family evolution

Gene families were identified among 35 arthropod species (electronic supplementary material, table S5). A total of 93% (560 182) genes were assigned into 16 287 gene families (except species-specific orthogroups). Moreover, 1812 gene families were shared by all species and 29 were single-copy orthogroups. For *H. lingyangjiaoensis*, 12 340 (87.5%) genes were clustered into 8044 gene families, among which 140 gene families (including 383 genes) were species-specific, and the statistics for the remaining arthropod species are provided in electronic supplementary material, table S11. Gene family evolution (gain and loss or expansion and contraction) were analysed using CAFE. Expansions and contractions of gene families for all 35 species are shown in figure 2 and electronic supplementary material, figure S16. Additionally, expansions and contractions for the first 100 families of each species with the largest gene-family size are visualized in electronic supplementary material, figure S17.

The estimated gene average birth rate (lambda) was 0.00134, accounting for duplications/gene/million years. A total of 239 gene families experienced significant expansion or contraction events across the tree of arthropods, with a family-wide p-value < 0.05. By comparing the *H. lingyangjiaoensis* genome with that of the remaining Arthropoda, 30 rapidly evolving gene families were identified, including four that significantly expanded and 26 that significantly contracted (table 2; electronic supplementary material, table S12). The significantly expanded gene families were trypsin, ABC transporter and FecCD transport family. Their function primarily involved in biological process relating to symbiotic interactions, nitrogen compound processes, organic substance processing and primary metabolic processes. Simultaneously, the gene families, ankyrin repeats, integrase zinc-binding domain and zinc finger were significantly contracted, and these families functioned in anatomical structure development and morphogenesis, cell cycle, moulting cycle, circadian rhythm and flight. *Hermatobates lingyangjiaoensis* specifically shared 712 orthogroups with the other five true bugs, among which 140 orthogroups were shared exclusively between *H. lingyangjiaoensis* and *G. buenoi* (Gerromorpha: Gerridae).

(d) Analysis of symbiotic bacteria

After quality control and deleting host and human contaminants, we obtained 367.86 Mb long reads and 5.61 Gb clean short reads (electronic supplementary material, table S13). These reads mainly comprised *Wolbachia* spp., *Serratia* spp., *Spiroplasma* spp. and *Bacillus* sp. (electronic supplementary material, figures S18 and S19). Via hybrid assembling and

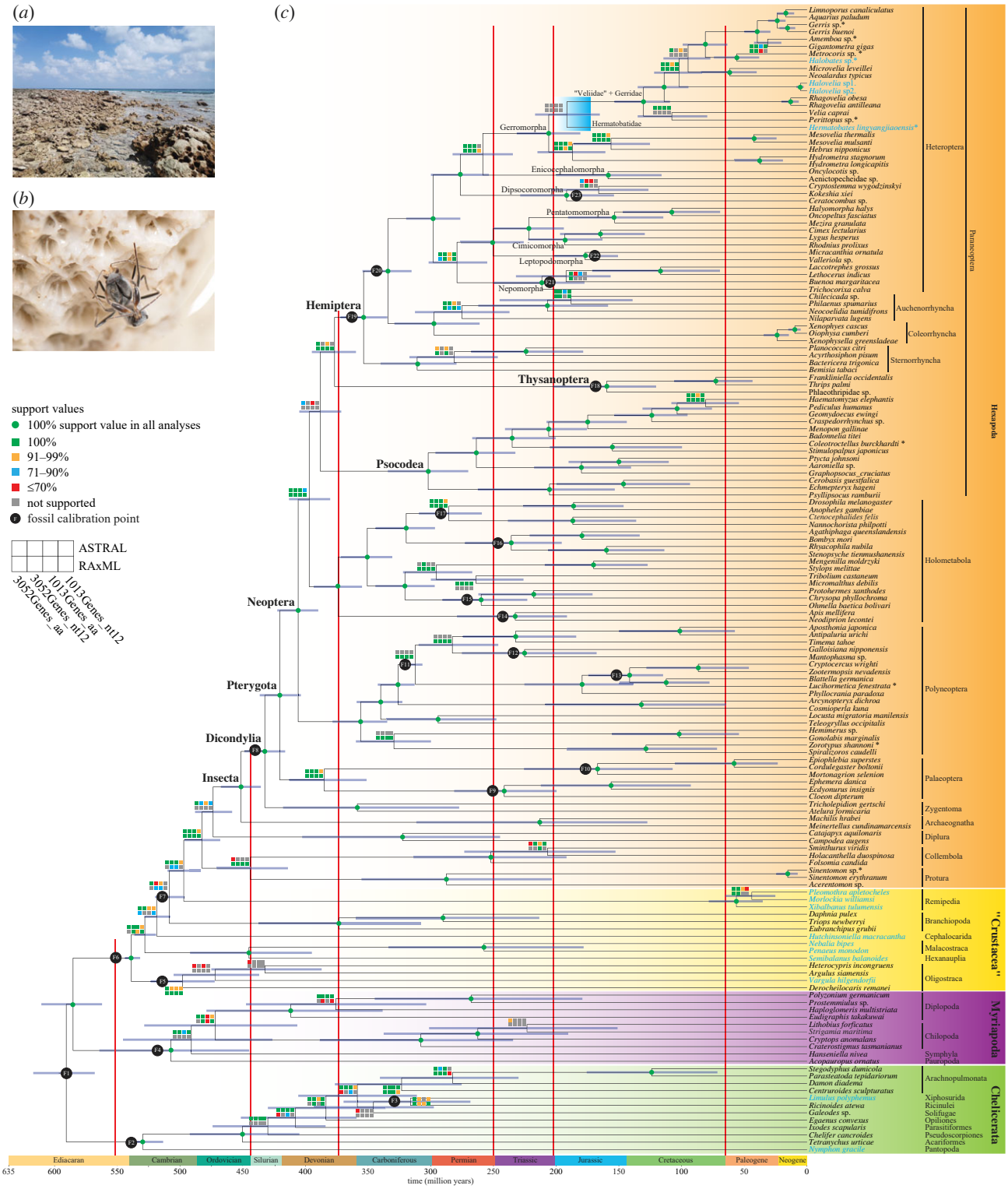

Figure 1. Dated phylogenomic relationships of Arthropoda. The time tree was constructed with 50 398 aligned amino acid sites. (*a*) Natural habitat of *Hermatobates lingyangjiaoensis*; (*b*) *in situ* photo of *H. lingyangjiaoensis*; (*c*) summary phylogeny and timescale of 156 Arthropoda species. Internal tree nodes are labelled with coloured dots summarizing the branch support from four maximum likelihood and four ASTRAL trees. Colour-coded square matrices show supports from different analyses for the branches that were not supported with 100% support values. Higher taxa are indicated as taxon labels on the right of the tree. Blue bars at each node represent 95% credibility intervals of the estimated date. Red lines indicate the six mass extinctions and asterisks indicate the newly sequenced transcriptomes and genomes. Species in cyan are marine. The gradient of cyan in Jurassic indicates the recovery time span of tropical reef ecosystems.

five rounds of polishing, we obtained a total of 1007 contigs (with a N50 of 24 690 kb) (electronic supplementary material, table S14). The main composition was the same as that of raw reads (electronic supplementary material, figure S20). Functional gene annotation showed that these symbiotic bacteria may play important roles in the carbohydrate, lipid and vitamin metabolism of the host (figure 3; electronic supplementary material, table S15). As shown in figure 3, the functional genes of symbiotic bacteria supplement or collaborate with host genes in many steps to accomplish a hybrid metabolic pathway.

One high-quality metagenome-assembled genome (MAG) (size: 4.96 Mb) and one medium-quality MAG (size: 1.03 Mb) were achieved through binning analysis (electronic

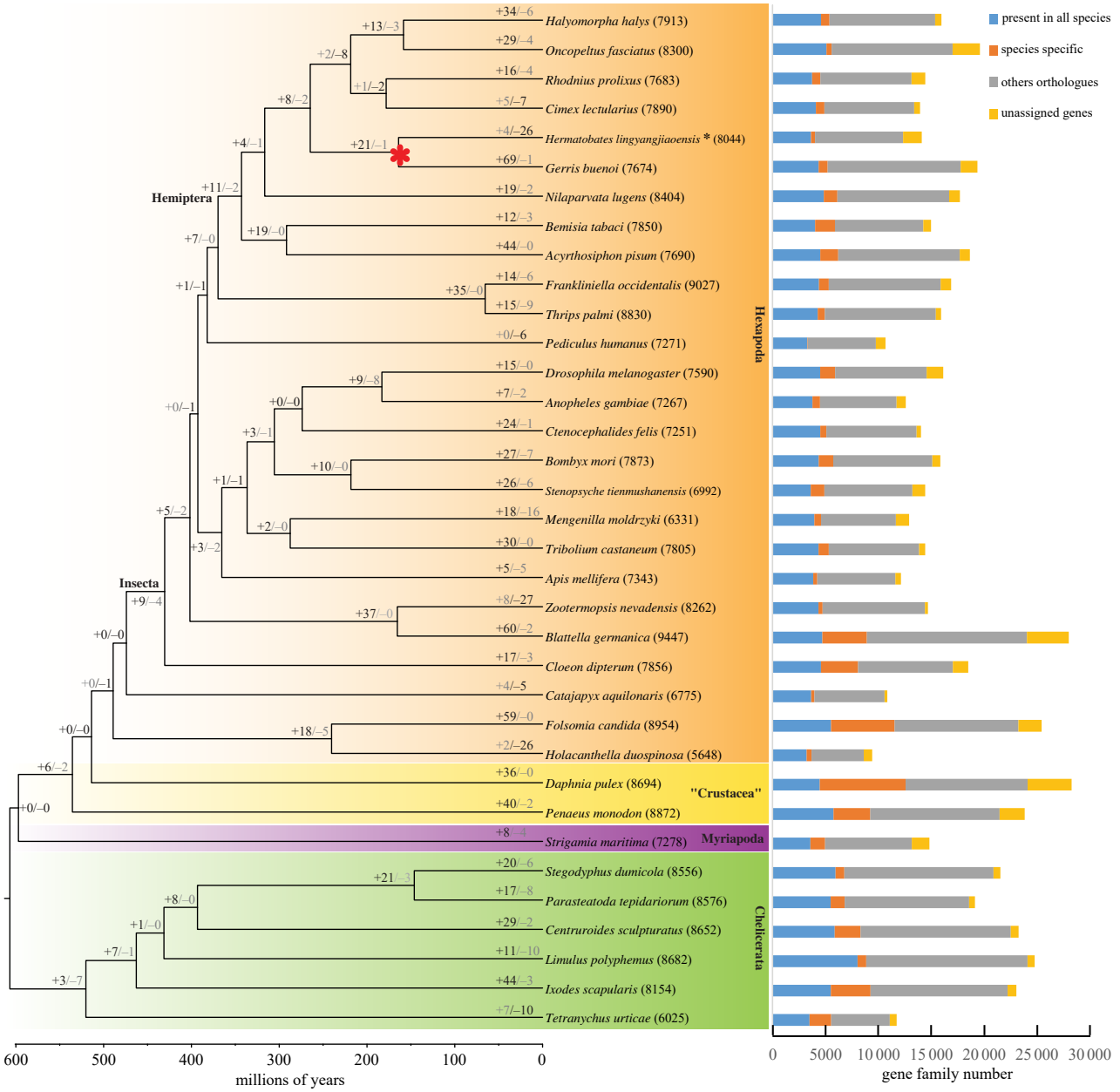

Figure 2. Gene family expansion and contraction among genomes of *H. lingyangjiaoensis* and other 34 arthropod species. Expansion and contraction are indicated with symbol + and −, and the numbers indicate significantly expansion and contraction for each node and crown group. Numbers of gene families for each species are shown following specific epithet. The asterisk indicates the newly sequenced genome of *H. lingyangjiaoensis*. The time tree was generated with MCMCtree and the gene family expansion and contraction analyses were conducted with CAFE. Branch lengths are measured in millions of years.

supplementary material, table S16). Phylogenetic analysis based on genome taxonomy database showed that the high-quality MAG is closely related to the genus *Serratia* (Proteobacteria: Gammaproteobacteria: Enterobacteriaceae). This genome had 95.15% completeness and 0.97% contamination. While the medium-quality MAG was closely related to the genus *Wolbachia* (Proteobacteria: Alphaproteobacteria: Rickettsiales: Anaplasmataceae), and with completeness at 89.99% and zero contamination. This is, for the first time, a near-complete genome of *Serratia* assembled from a marine insect, and we tentatively named this bacterial strain as *Serratia* sp. (hereafter: sHlin). The overall G + C content of sHlin is 59.7% (electronic supplementary material, figure S21 and table S16).

Additionally, from the assembled draft genome we identified 20 bacterial-origin contigs from genera *Serratia*, *Wolbachia*, *Corynebacterium* and *Formosa* (electronic supplementary material, table S17). Among the bacterial-origin contigs, the longest contig (Contig00166, 1.17 Mb) is a near-complete *Wolbachia* genome (hereafter: wHlin), which has 97.62% completeness and only 0.21% contamination. Comparing with the medium-quality MAG, the average nucleotide identity (ANI) based on MUMmer is 99.81, which is higher than the 95% threshold and means that they belong to the same species. The overall G + C content of wHlin is 36.2% (electronic supplementary material, figure S22 and table S16). Phylogenomic analyses based on amino acids of 176 single-copy genes showed a close relationship between the *Wolbachia* strains of *H. lingyangjiaoensis* and *Cimex lectularius* (Heteroptera, Cimicidae), belonging to the F supergroup (electronic supplementary material, figure S23). KEGG pathway analyses show wHlin genome possesses partial pathways for biotin (vitamin B7) and thiamine (vitamin B1) found in the genome of *Wolbachia* strain of *C. lectularius*.

Table 2. Significantly expanded and contracted gene families of *H. lingyangjiaoensis*.

	gene no.	gene copy no. in *H. lingyangjiaoensis*	gene copy no. in *Gerris buenoi*
significantly	OG0000132	16	12
expanded	OG0000454	5	0
gene	OG0001600	24	1
families	OG0001733	4	1
significantly	OG0000015	5	28
contracted	OG0000016	8	29
gene	OG0000021	6	18
families	OG0000032	12	35
	OG0000044	2	26
	OG0000046	24	95
	OG0000057	2	52
	OG0000064	0	54
	OG0000068	5	32
	OG0000076	7	95
	OG0000079	0	51
	OG0000086	2	46
	OG0000087	0	107
	OG0000091	9	71
	OG0000098	0	21
	OG0000107	0	15
	OG0000108	1	30
	OG0000118	0	21
	OG0000122	3	47
	OG0000250	1	17
	OG0000319	1	79
	OG0000338	1	61
	OG0000358	0	40
	OG0000484	0	81
	OG0000884	0	31
	OG0002013	0	50

4. Discussion

Based on the results of this and prior studies (e.g. [55]), stem-group insects originated in the Early Ordovician and thereafter successfully occupied all conceivable terrestrial niches. There is no doubt that the marine environment was the last habitat invaded by insects. Unlike their "crustacean" relatives who had so successfully established themselves in the oceans, insects had to contend with various physical, physiological and biological factors to return to the sea [60]. Although strictly marine insects are rare, considering the huge diversity of insect species, such a habitat is not uncommon for semi-aquatic bugs, among which Hermatobatidae is one of the few exclusively marine family. Therefore, reconstructing the phylogenomics of Hermatobatidae in the broad context of Arthropoda and dating their origin may help resolve evolutionary patterns and the palaeoecological forces allowing for such a habitat shift and specialization. Simultaneously, deciphering the genome and metagenome of a coral treader may provide an opportunity to understand their adaptation at a molecular genetic level. It should be stressed that obtaining a high-quality genome is in the central status to improve phylogenetic reconstruction, to have comparative genomic analyses and to explore the potential complementary functions between insects and symbiotic bacteria.

The backbone phylogeny of extant arthropods recovered here is largely congruent with other phylogenomic studies [12,13,55]. This robust result provides a solid foundation for investigating the evolution of marine lineages along the phylogeny and corresponding genomic changes. A main finding in the present study is the recovery of Paraneoptera being monophyletic by ASTRAL analyses using the amino acid sequences of 1013 and 3052 genes. Although the posterior probabilities are not high, there is corroborating evidence from morphological synapomorphies and prior molecular studies [59]. The improved phylogenetic inference of insects may encourage higher integration between the knowledge of evolution and development of insects [61–63]. As for the special attention to the phylogenetic position of Hermatobatidae within Gerromorpha, FcLM analysis supports the sister relationship between Hermatobatidae and the clade "Veliidae" + Gerridae with overwhelming quartets, suggesting underlying strong signals in the data for this result. Furthermore, there are several morphological synapomorphies in support of such a relationship, e.g. labial segments with intercalary sclerites between segments three and four, the female gynatrial sac with glandular cells, and each ovarium with four ovarioles [15]. However, before any taxonomic conclusions are drawn, it is important to note that our taxon sampling is modest and that families such as Macroveliidae and Paraphrynoveliidae, as well as representatives of many subfamilies are missing. This is especially problematic in the Hydrometridae, which is considered to be the closest extant relatives of Hermatobatidae [18].

In the history of life, there are six well-known great mass extinctions, one at the end of the Ediacaran during the Proterozoic and five in the Phanerozoic. Almost every mass extinction event was followed by the origin of new clades, expansions of morphologic disparity and a rapid diversification of new taxa [64–66]. After the end-Ediacaran, the three main lineages of arthropods began to diversify. According to our results, after the end-Ordovician event, the Dicondylia, particularly the Pterygota, began their initial radiation. Later, after the mass extinction in the Late Devonian, Polyneoptera, Holometabola, and Hemiptera began to radiate, followed by many order-level groups which expanded after the end-Permian event (approx. 250 Ma), including the Gerromorpha. Our results suggest that the end-Triassic mass extinction was likely correlated with the origin and diversification of coral treaders. This extinction event resulted in significant biodiversity loss and ecological crises in both terrestrial and marine ecosystems, and an especially devastating impact on tropical reef ecosystems that did not fully recover until the Middle Jurassic [67]. The recovery of reef ecosystems may have provided many new and open niches into which coral treaders could occupy and thrive.

Our dated phylogeny shows that the Hermatobatidae represent the earliest-diverged marine lineage (192 Ma) among Gerromorpha–Gerroidea. The other marine lineages of semi-

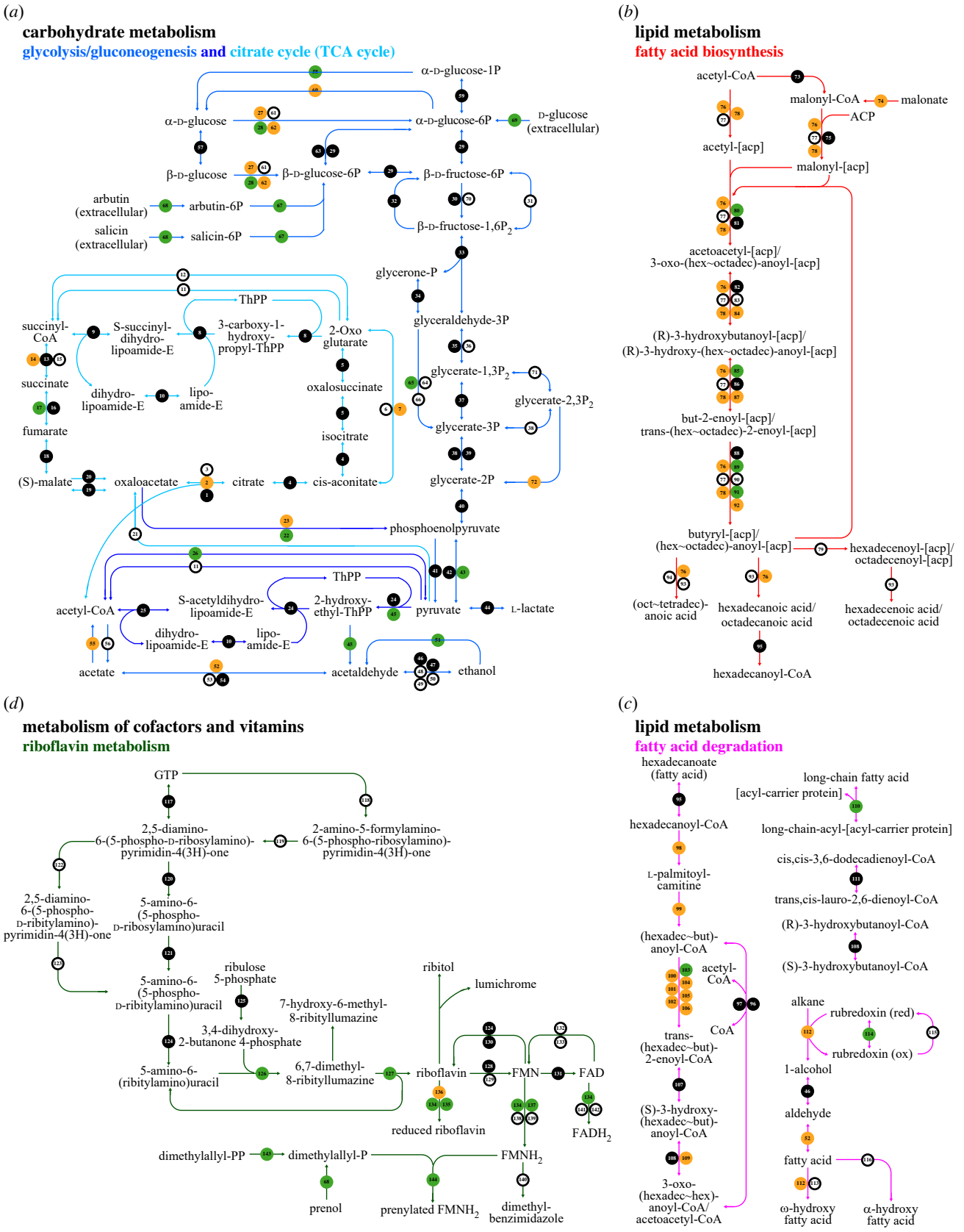

Figure 3. Functional genes of host insect and its symbiotic bacteria in different KEGG pathways. The green circle indicates functional genes of symbiotic bacteria. The orange circle indicates functional genes of insects. The black circle indicates functional genes of both host and symbiotic bacteria, and the hollow circle indicates that both host and symbiotic bacteria were lacking. Detailed gene functions associated with the aforementioned pathways were recorded in electronic supplementary material, table S15.

aquatic bugs are principally found nested within more derived subclades of certain subfamilies of 'Veliidae' + Gerridae and represent transitions from humid terrestrial or freshwater habitats to saline waters, a pattern apparently repeated independently several times throughout the phylogeny of Gerromorpha [15]. These more recent invasions of marine habitats have taken place to a large extent in the levels of species, genera or tribes with predominantly limnetic relatives. These, quite unlike the case for Hermatobatidae, represent transitions from limnetic habitats to estuaries and mangrove

swamps then to intertidal or open ocean environments. If any extinction event is correlated with these shifts, the end-Cretaceous mass extinction or even smaller events during the Cenozoic such as the Eocene thermal maximum or Eocene–Oligocene transition are seemingly most likely.

The genome of *H. lingyangjiaoensis* is of a medium size for true bugs (0.68–1.55 Gb) [68,69] and a similar size to those of other semi-aquatic bugs [70,71]. Comparative genomic analyses with other arthropod species show that the gene families functioning in the moulting cycle and related processes are significantly contracted, which certainly relates to their reduced life cycles. Their life history includes the egg, four nymphal stages, and finally the adult. The loss of a nymphal stage likely increases fitness in an otherwise harsh marine environment. In addition, gene families associated with anatomical structure development and morphogenesis and flight were also contracted, likely associated with their highly simplified and compact abdominal structure and obligatory flightlessness. The flightlessness of all marine Gerromorpha is generally associated with the higher stability of marine environments when compared with freshwater habitats, especially temporary ones, where the maintenance of wings for dispersal (at least in part of the populations) is favoured [15]. Specifically, comparing to the genome of the freshwater species *Gerris buenoi*, with the typical five nymphal instars and wing polymorphism, the corresponding gene families OG0000057, OG0000068 and OG0000250 were significantly contracted, with only two, five, and one genes in *H. lingyangjiaoensis* versus 52, 32 and 17 genes in *G. buenoi*, respectively (table 2; electronic supplementary material, table S12). These gene families play important roles in anatomical structure development and morphogenesis, circadian rhythm and the moulting cycle process and deserve more attention in development studies.

Symbiotic microbes play important and diverse physiological functions for their hosts, including supplementation of nutrition, protection from parasites and pathogens, influence on insect mating and reproduction, and even a contribution to the hosts' adaptation to novel ecological niches [72–74]. The metagenome offers the most direct evidence of the genetic capacity of the microbial partners. Our functional gene annotation for the metagenome of *H. lingyangjiaoensis* shows that their symbiotic bacteria supplement or collaborate with the host's genes in many steps of carbohydrate, lipid and vitamin metabolism pathways. These symbiotic bacteria likely provide significant fitness benefits for the host's survival. Within the gut of coral treaders, *Serratia* sp. is one of the more dominant strains. *Serratia* has a wide distribution, and can be found in the air, in soil, as well as living symbiotically with plants and insects [75]. For example, it has been reported that *Serratia* can confer host resistance to invading parasitoids [76]. Another dominant strain found with *H. lingyangjiaoensis* was the genus *Wolbachia*, which can supply essential nutrition for host reproduction [77] and has the genes to synthesize several B vitamins [21]. The wHlin genome additionally possesses partial pathways for biotin (vitamin B7) and thiamine (vitamin B1) as the case in *Cimex lectularius* [78]. But the precise function of these bacteria in *H. lingyangjiaoensis* remains unknown and deserves further investigation.

5. Conclusion

In conclusion, our dated phylogeny indicates that after hexapods left the oceans to diversify on land, coral treaders returned to the seas approximately 300 Myr later. Their adaptive evolution to marine habitats mainly reflects two aspects. First, a significant contraction in some gene families resulting in the reduced life cycle of coral treaders as well as their compact abdominal structure, putatively to increase fitness in their harsh marine environment. Second, the functional genes of their symbiotic bacteria supplement or collaborate with host genes in many steps of nutrient metabolism and energy production. Further understanding of adaptations of these interesting bugs is clearly needed. Next, we also hope to verify whether the candidate gene families are correlated with life cycle via comparative transcriptomics and establishing laboratory colony of wild population.

Data accessibility. All of the raw sequence reads used in this study have been deposited into the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) and Transcriptome Shotgun Assemblies (TSA) under the BioProject accession number PRJNA826332. Genome assemblies have been deposited in BioProject PRJNA897844 and PRJNA953184. Protein and nucleotide dataset, 3052 genes, metagenomes and genome annotations are available in the Dryad Digital Repository [79].

The data are provided in electronic supplementary material [80].

Authors' contributions. Y.W.: data curation, formal analysis, resources, writing—original draft, writing—review and editing; Y.L.: resources, writing—review and editing; J.L.: resources, writing—review and editing; Y.M.: formal analysis, writing—review and editing; M.S.E.: writing—review and editing; J.D.: writing—review and editing; A.K.: writing—review and editing; P.C.: resources, writing—review and editing; F.F.F.M.: resources, writing—review and editing; J.A.R.: resources, writing—review and editing; Q.X.: conceptualization, data curation, funding acquisition, resources, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. The authors have no competing interests to declare.

Funding. This work was supported by the National Natural Science Foundation of China (grant no. 31222051) and the Natural Science Foundation of Guangdong Province - Outstanding Youth Team Project (grant no. 2023B1515040002).

Acknowledgements. We thank the organizers (Sun Yat-sen University) who organized the scientific expeditions of XiSha Islands (Hainan, China) in 2018 and 2019, which have resulted in the finding of this interesting water bugs. We appreciate Dr Lanna Cheng (University of California San Diego) and Mr Marc Jia Jin Chang (National University of Singapore) for valuable suggestions to improve the quality of our manuscript. We are greatly in debt to Dr Li Liu (Sun Yat-sen University) for providing an extra expedition opportunity to the XiSha Islands. We also thank Prof. Jin Xu (Sun Yat-sen University) and Prof. Chung-I Wu (Sun Yat-sen University) for the helpful advices.

References

1. Cheng L, Mishra H. 2022 Why did only one genus of insects, *Halobates*, take to the high seas? *PLoS Biol.* **20**, e3001570. (doi:10.1371/journal.pbio.3001570)
2. Cheng L. 1977 The elusive sea bug *Hermatobates*. *Pan-Pacific Entomol.* **53**, 87–97.

3. Carpenter GH. 1892 Rhynchota from Murray Island and Mabuiag. Reports on the Zoological collections made in Torres Straits by Professor Haddon AC, 1888–1889. *Sci. Proc. R. Dublin Soc.* **7**, 137–146.
4. Andersen NM, Weir TA. 2000 The coral treaders, *Hermatobates* Carpenter (Hemiptera: Hermatobatidae), of Australia and New Caledonia with notes on biology and ecology. *Invertebr. Taxon.* **14**, 327–345. (doi:10.1071/IT99012)
5. Polhemus JT, Polhemus DA. 2012 A review of the genus *Hermatobates* (Heteroptera: Hermatobatidae), with descriptions of two new species. *Entomol. Am.* **118**, 202–241. (doi:10.1664/12-RA-018.1)
6. Luo J, Chen P, Wang Y, Xie Q. 2019 First record of Hermatobatidae from China, with description of *Hermatobates lingyangjiaoensis* sp. n. (Hemiptera: Heteroptera). *Zootaxa* **4679**, 527–538. (doi:10.11646/zootaxa.4679.3.6)
7. Foster WA. 1989 Zonation, behaviour and morphology of the intertidal coral-treader *Hermatobates* (Hemiptera: Hermatobatidae) in the south-west Pacific. *Zool. J. Linn. Soc.* **96**, 87–105. (doi:10.1111/j.1096-3642.1989.tb01822.x)
8. Perez-Goodwyn P, Maezono Y, Takamatsu H, Fujisaki K. 2009 Semiaquatic Heteroptera locomotion: coral treaders (*Hermatobates weddi*, Hermatobatidae), sea skaters (*Halovelia septentrionalis*, Veliidae), and water striders (*Metrocoris histrio*, Gerridae). Usual and unusual gaits. *Hydrobiologia* **630**, 219–229. (doi:10.1007/s10750-009-9794-9)
9. Esaki T. 1947 Notes on *Hermatobates haddonii* Carpenter (Hemiptera, Gerridae). *Mushi* **18**, 49–51.
10. Cheng L, Douek M, Goring DAI. 1978 UV absorbtion by gerrid cuticles. *Limnol. Oceanogr.* **23**, 554–556.
11. Andersen NM, Cheng L. 2004 The marine insect *Halobates* (Heteroptera: Gerridae): biology, adaptations, distribution and phylogeny. *Oceanogr. Mar. Biol.* **42**, 119–180. (doi:10.1201/9780203507810)
12. Giribet G, Edgecombe GD. 2019 The phylogeny and evolutionary history of arthropods. *Curr. Biol.* **29**, R592–R602. (doi:10.1016/j.cub.2019.04.057)
13. Schwentner M, Combosch DJ, Nelson JP, Giribet G. 2017 A phylogenomic solution to the origin of insects by resolving Crustacean–Hexapod relationships. *Curr. Biol.* **27**, 1818–1824.e5. (doi:10.1016/j.cub.2017.05.040)
14. Lozano-Fernandez J *et al.* 2016 A molecular palaeobiological exploration of arthropod terrestrialization. *Phil. Trans. R. Soc. B* **371**, 20150133. (doi:10.1098/rstb.2015.0133)
15. Andersen NM. 1999 The evolution of marine insects: phylogenetic, ecological and geographical aspects of species diversity in marine water striders. *Ecography* **22**, 98–111. (doi:10.1111/j.1600-0587.1999.tb00458.x)
16. Andersen NM, Weir TA. 2004 Mesoveliidae, Hebridae, and Hydrometridae of Australia (Hemiptera: Heteroptera: Gerromorpha), with a reanalysis of the phylogeny of semiaquatic bugs. *Invertebr. Syst.* **18**, 467–522. (doi:10.1071/IS03032)
17. Damgaard J, Andersen NM, Meier R. 2005 Combining molecular and morphological analyses of water strider phylogeny (Hemiptera–Heteroptera, Gerromorpha): effects of alignment and taxon sampling. *Syst. Entomol.* **30**, 289–309. (doi:10.1111/j.1365-3113.2004.00275.x)
18. Damgaard J. 2008 Phylogeny of the semiaquatic bugs (Hemiptera–Heteroptera, Gerromorpha). *Insect Syst. Evol.* **39**, 431–460. (doi:10.1163/187631208788784264)
19. Douglas AE. 2018 Omics and the metabolic function of insect–microbial symbioses. *Curr. Opin. Insect Sci.* **29**, 1–6. (doi:10.1016/j.cois.2018.05.012)
20. McCutcheon JP, McDonald BR, Moran NA. 2009 Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet.* **5**, e1000565. (doi:10.1371/journal.pgen.1000565)
21. Hosokawa T, Koga R, Kikuchi Y, Meng XY, Fukatsu T. 2010 *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proc. Natl Acad. Sci. USA* **107**, 769–774. (doi:10.1073/pnas.0911476107)
22. Shi W *et al.* 2013 Comparative genomic analysis of the microbiome of herbivorous insects reveals eco-environmental adaptations: biotechnology applications. *PLoS Genet.* **9**, e1003131. (doi:10.1371/journal.pgen.1003131)
23. Berasategui A, Salem H, Paetz C, Santoro M, Gershenzon J, Kaltenpoth M, Schmidt A. 2017 Gut microbiota of the pine weevil degrades conifer diterpenes and increases insect fitness. *Mol. Ecol.* **26**, 4099–4110. (doi:10.1111/mec.14186)
24. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017 Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* **27**, 722–736. (doi:10.1101/gr.215087.116)
25. Ruan J, Li H. 2020 Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155–158. (doi:10.1038/s41592-019-0669-3)
26. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013 QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075. (doi:10.1093/bioinformatics/btt086)
27. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021 BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654. (doi:10.1093/molbev/msab199)
28. Li H. 2018 Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100. (doi:10.1093/bioinformatics/bty191)
29. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009 The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079. (doi:10.1093/bioinformatics/btp352)
30. Wood DE, Lu J, Langmead B. 2019 Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257. (doi:10.1186/s13059-019-1891-0)
31. Bertrand D *et al.* 2019 Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nat. Biotechnol.* **37**, 937–944. (doi:10.1038/s41587-019-0191-2)
32. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJT. 2010 Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf.* **11**, 119. (doi:10.1186/1471-2105-11-119)
33. Uritskiy GV, DiRuggiero J, Taylor J. 2018 MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158. (doi:10.1186/s40168-018-0541-1)
34. Wang Y *et al.* 2019 When did the ancestor of true bugs become stinky? Disentangling the phylogenomics of Hemiptera–Heteroptera. *Cladistics* **35**, 42–66. (doi:10.1111/cla.12232)
35. Wang Y, Moreira FFF, Rédei D, Chen P, Kuechler SM, Luo J, Men Y, Wu H, Xie Q. 2021 Diversification of true water bugs revealed by transcriptome-based phylogenomics. *Syst. Entomol.* **46**, 339–356. (doi:10.1111/syen.12465)
36. Grabherr MG *et al.* 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652. (doi:10.1038/nbt.1883)
37. Pertea G *et al.* 2003 TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* **19**, 651–652. (doi:10.1093/bioinformatics/btg034)
38. Emms DM, Kelly S. 2019 OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238. (doi:10.1186/s13059-019-1832-y)
39. Petersen M *et al.* 2017 Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. *BMC Bioinf.* **18**, 111. (doi:10.1186/s12859-017-1529-8)
40. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. (doi:10.1093/molbev/mst010)
41. Abascal F, Zardoya R, Telford MJ. 2010 TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **38**, W7–W13. (doi:10.1093/nar/gkq291)
42. Talavera G, Castresana J. 2007 Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577. (doi:10.1080/10635150701472164)
43. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009 trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973. (doi:10.1093/bioinformatics/btp348)
44. Stamatakis A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. (doi:10.1093/bioinformatics/btu033)

45. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015 IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. (doi:10.1093/molbev/msu300)
46. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589. (doi:10.1038/nmeth.4285)
47. Sayyari E, Mirarab S. 2016 Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* **33**, 1654–1668. (doi:10.1093/molbev/msw079)
48. Wolfe JM, Daley AC, Legg DA, Edgecombe GD. 2016 Fossil calibrations for the arthropod Tree of Life. *Earth Sci. Rev.* **160**, 43–110. (doi:10.1016/j.earscirev.2016.06.008)
49. Yang Z. 2007 PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591. (doi:10.1093/molbev/msm088)
50. dos Reis M, Yang Z. 2011 Approximate likelihood calculation for Bayesian estimation of divergence times. *Mol. Biol. Evol.* **28**, 2161–2172. (doi:10.1093/molbev/msr045)
51. Han MV, Thomas GW, Lugo-Martinez J, Hahn MW. 2013 Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* **30**, 1987–1997. (doi:10.1093/molbev/mst100)
52. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021 eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829. (doi:10.1093/molbev/msab293)
53. Finn RD *et al.* 2017 InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **45**(D1), D190–D199. (doi:10.1093/nar/gkw1107)
54. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. 2020 TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **13**, 1194–1202. (doi:10.1016/j.molp.2020.06.009)
55. Misof B *et al.* 2014 Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767. (doi:10.1126/science.1257570)
56. Wipfler B *et al.* 2019 Evolutionary history of Polyneoptera and its implications for our understanding of early winged insects. *Proc. Natl Acad. Sci. USA* **116**, 3024–3029. (doi:10.1073/pnas.1817794116)
57. Thomas GWC *et al.* 2020 Gene content evolution in the arthropods. *Genome Biol.* **21**, 15. (doi:10.1186/s13059-019-1925-7)
58. Johnson KP *et al.* 2018 Phylogenomics and the evolution of hemipteroid insects. *Proc. Natl Acad. Sci. USA* **115**, 12 775–12 780. (doi:10.1073/pnas.1815820115)
59. Wang Y, Engel MS, Rafael JA, Wu H, Rédei D, Xie Q, Wang G, Liu X, Bu W. 2016 Fossil record of stem groups employed in evaluating the chronogram of insects (Arthropoda: Hexapoda). *Sci. Rep.* **6**, 38939. (doi:10.1038/srep38939)
60. Cheng L. 1976 *Marine insects*. Amsterdam, The Netherlands: North-Holland Publishing Company.
61. Truman JW. 2019 The evolution of insect metamorphosis. *Curr. Biol.* **29**, R1252–R1268. (doi:10.1016/j.cub.2019.10.009)
62. Elias-Neto M, Alvarez N, Ventos-Alfonso A, Belles X. 2022 Flight or protection: the genes *Ultrabithorax* and *apterous* in the determination of membranous and sclerotized wings in insects. *Proc. R. Soc. B* **289**, 20220967. (doi:10.1098/rspb.2022.0967)
63. Okude G, Moriyama M, Kawahara-Miki R, Yajima S, Fukatsu T, Futahashi R. 2022 Molecular mechanisms underlying metamorphosis in the most-ancestral winged insect. *Proc. Natl Acad. Sci. USA* **119**, e2114773119. (doi:10.1073/pnas.2114773119)
64. Wignall PB. 2015 *The worst of times*. Princeton, NJ: Princeton University Press.
65. Darroch SAF, Smith EF, Laflamme M, Erwin DH. 2018 Ediacaran extinction and Cambrian explosion. *Trends Ecol. Evol.* **33**, 653–663. (doi:10.1016/j.tree.2018.06.003)
66. Dal Corso J, Song H, Callegaro S, Chu D, Sun Y, Hilton J, Grasby SE, Joachimski MM, Wignall PB. 2022 Environmental crises at the Permian–Triassic mass extinction. *Nat. Rev. Earth Environ.* **3**, 197–214. (doi:10.1038/s43017-021-00259-4)
67. Dunhill AM, Foster WJ, Sciberras J, Twitchett RJ. 2017 Impact of the Late Triassic mass extinction on functional diversity and composition of marine ecosystems. *Palaeontology* **61**, 133–148. (doi:10.1111/pala.12332)
68. Huang HJ *et al.* 2021 Chromosome-level genome assembly of the bean bug *Riptortus pedestris*. *Mol. Ecol. Resour.* **21**, 2423–2436. (doi:10.1111/1755-0998.13434)
69. Jiang T *et al.* 2021 Chromosomal-level genome assembly of a true bug, *Aspongopus chinensis* Dallas, 1851 (Hemiptera: Dinidoridae). *Genome Biol. Evol.* **13**, evab232. (doi:10.1093/gbe/evab232)
70. Armisén D *et al.* 2018 The genome of the water strider *Gerris buenoi* reveals expansions of gene repertoires associated with adaptations to life on the water. *BMC Genom.* **19**, 832. (doi:10.1186/s12864-018-5163-2)
71. Toubiana W, Armisén D, Dechaud C, Arbore R, Khila A. 2021 Impact of male trait exaggeration on sex-biased gene expression and genome architecture in a water strider. *BMC Biol.* **19**, 89. (doi:10.1186/s12915-021-01021-4)
72. Engel P, Moran NA. 2013 The gut microbiota of insects—diversity in structure and function. *FEMS Microbiol. Rev.* **37**, 699–735. (doi:10.1111/1574-6976.12025)
73. Sudakaran S, Kost C, Kaltenpoth M. 2017 Symbiont acquisition and replacement as a source of ecological innovation. *Trends Microbiol.* **25**, 375–390. (doi:10.1016/j.tim.2017.02.014)
74. Ankrah NYD, Douglas AE. 2018 Nutrient factories: metabolic function of beneficial microorganisms associated with insects. *Environ. Microbiol.* **20**, 2002–2011. (doi:10.1111/1462-2920.14097)
75. Grimont PA, Grimont F. 1978 The genus *Serratia*. *Ann. Rev. Microbiol.* **32**, 221–248. (doi:10.1146/annurev.mi.32.100178.001253)
76. Bai L, Wang L, Vega-Rodríguez J, Wang G, Wang S. 2019 A gut symbiotic bacterium *Serratia marcescens* renders mosquito resistance to *Plasmodium* infection through activation of mosquito immune responses. *Front. Microbiol.* **10**, 1580. (doi:10.3389/fmicb.2019.01580)
77. Ju JF *et al.* 2020 *Wolbachia* supplement biotin and riboflavin to enhance reproduction in planthoppers. *ISME J.* **14**, 676–687. (doi:10.1038/s41396-019-0559-9)
78. Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M, Fukatsu T. 2014 Evolutionary origin of insect–*Wolbachia* nutritional mutualism. *Proc. Natl Acad. Sci. USA* **111**, 10 257–10 262. (doi:10.1073/pnas.1409284111)
79. Wang Y *et al.* 2023 Data from: 300 Million years of coral treaders (Insecta: Heteroptera: Hermatobatidae) back to the ocean in the phylogenetic context of Arthropoda. Dryad Digital Repository. (doi:10.5061/dryad.h18931zqn)
80. Wang Y *et al*. 2023 300 Million years of coral treaders (Insecta: Heteroptera: Hermatobatidae) back to the ocean in the phylogenetic context of Arthropoda. Figshare. (doi:10.6084/m9.figshare.c.6700019)