*Insect Systematics and Diversity*, 8(4), 3; 2024, ixae015
https://doi.org/10.1093/isd/ixae015
Research

ENTOMOLOGICAL SOCIETY OF AMERICA
SHARING INSECT SCIENCE GLOBALLY

OXFORD

Molecular phylogenetics, phylogenomics, and phylogeography

# Skimming the skaters: genome skimming improves phylogenetic resolution of Halobatinae (Hemiptera: Gerridae)

**Jia Jin Marc Chang[1,*], Michael J. Raupach[2], Lanna Cheng[3,4], Jakob Damgaard[5], Watcharapong Hongjamrassilp[6], Yin Cheong Aden Ip[1], Matthew Hui-Chieh Ng[1], Wan Wen Rochelle Chan[1], Ismael Kunning[7], Bryna Jia Ying Liang[1], Davide Maggioni[8,9], Ralph R. Mana[7], Himanshu Mishra[10], Maxine A.D. Mowe[1], Benjamin J. Wainwright[1,11], Jonathan L. Whitney[12], Kennedy Wolfe[13], Darren C.J. Yeo[1,4,14], Danwei Huang[1,4,14,15]**

[1]Department of Biological Sciences, National University of Singapore, 16 Science Drive 4, Singapore 117558, Singapore, [2]Section Hemiptera, Bavarian State Collection of Zoology, Münchhausenstraße 21, München 81247, Germany, [3]Scripps Institution of Oceanography, University of California San Diego, La Jolla, San Diego, CA 92039, USA, [4]Lee Kong Chian Natural History Museum, National University of Singapore, 2 Conservatory Drive, Singapore 117377, Singapore, [5]Natural History Museum of Denmark, Zoological Museum, Copenhagen 2100 Ø, Denmark, [6]Center of Excellence for Marine Biotechnology, Department of Marine Science, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand, [7]Biological Sciences, School of Natural and Physical Sciences, University of Papua New Guinea, Waigani Campus, Port Moresby, Papua New Guinea, [8]Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milano 20126, Italy, [9]Marine Research and Higher Education (MaRHE) Center, University of Milano-Bicocca, Faafu Magoodhoo Island 12030, Maldives, [10]Environmental Science and Engineering (EnSE) Program, Biological and Environmental Science and Engineering (BESE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia, [11]Yale-NUS College, National University of Singapore, 16 College Avenue West, Singapore 138527, Singapore, [12]Pacific Islands Fisheries Science Center, National Oceanic & Atmospheric Administration, Honolulu, Hawaii, USA, [13]School of the Environment, The University of Queensland, St. Lucia, Queensland 4072, Australia, [14]Tropical Marine Science Institute, National University of Singapore, 18 Kent Ridge Road Singapore 119227, Singapore, [15]Centre for Nature-based Climate Solutions, National University of Singapore, 6 Science Drive 2, Singapore 117546, Singapore *Corresponding author, mail: jjmc@u.nus.edu

Subject Editor: Kazunori Yoshizawa

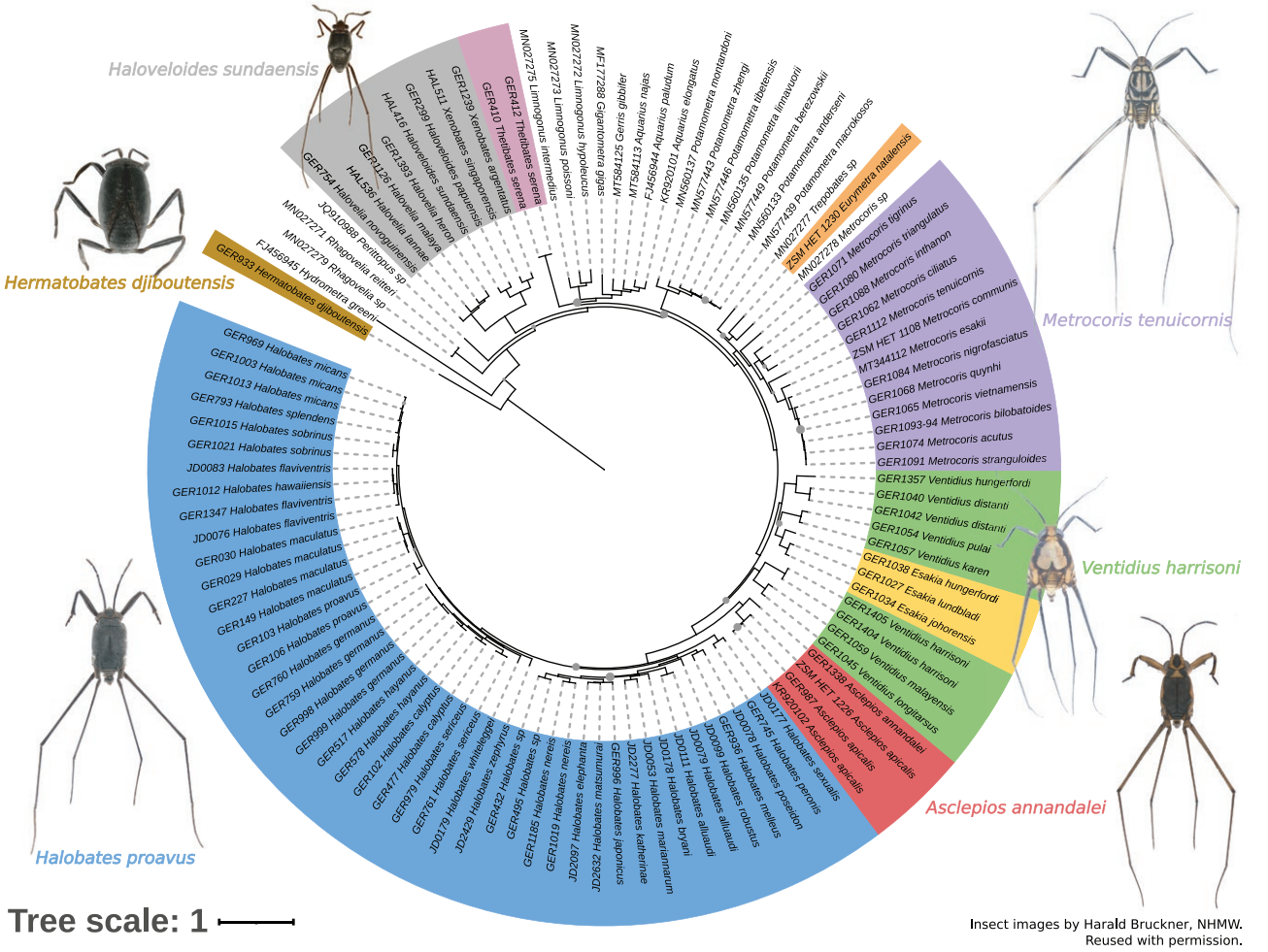
Received on 19 December 2023; revised on 19 April 2024; accepted on 17 June 2024

Gerromorpha Popov, 1971 is a fascinating and diverse insect lineage that evolved about 200 Mya to spend their entire life cycle on the air–water interface and have since colonized all types of aquatic habitats. The subfamily Halobatinae Bianchi, 1896 is particularly interesting because some species have adapted to life on the open ocean—a habitat where insects are very rarely found. Several attempts have been made to reconstruct the phylogenetic hypotheses of this subfamily, but the use of a few partial gene sequences recovered only a handful of well-supported relationships, thus limiting evolutionary inferences. Fortunately, the emergence of high-throughput sequencing technologies has enabled the recovery of more genetic markers for phylogenetic inference. We applied genome skimming to obtain mitochondrial and nuclear genes from low-coverage whole-genome sequencing of 85 specimens for reconstructing a well-supported phylogeny, with particular emphasis on Halobatinae. Our study confirmed that Metrocorini Matsuda, 1960, is paraphyletic, whereas *Esakia* Lundblad, 1933, and *Ventidius* Distant, 1910, are more closely related to Halobatini Bianchi, 1896, than *Metrocoris* Mayr, 1865, and *Eurymetra* Esaki, 1926. We also found that *Ventidius* is paraphyletic and in need of a taxonomic revision. Ancestral state reconstruction suggests that Halobatinae evolved progressively from

limnic to coastal habitats, eventually attaining a marine lifestyle, especially in the genus *Halobates* Eschscholtz, 1822, where the oceanic lifestyle evolved thrice. Our results demonstrate that genome skimming is a powerful and straightforward approach to recover genetic loci for robust phylogenetic analysis in non-model insects.

Key words: aquatic insect, Indo-Pacific, mitogenome, phylogenomics, whole-genome sequencing

## Graphical Abstract




## Introduction

It is estimated that there are ~5.5 million insect species in the world (Gaston 1991, Stork et al. 2015, Stork 2018). Of these, an estimated 50,000–76,000 species are considered aquatic, and even less (~25,000 species) are found in saline habitats (Cheng 1976, Lévêque et al. 2005, Balian et al. 2008, Cheng and Mishra 2022). The infraorder Gerromorpha Popov, 1971 (Insecta: Hemiptera: Heteroptera), comprising more than 2,000 extant species, evolved about 200 Mya to live on the air–water interface (Andersen 1982, Damgaard 2008a, Armisén and Khila 2022, Armisén et al. 2022). Gerromorpha currently comprises 4 superfamilies and 8 families: Mesoveloidea (Mesoveliidae), Hebroidea (Hebridae), Hydrometroidea (Hydrometridae, Paraphrynoveliidae, Macroveliidae), and Gerroidea (Hermatobatidae, Veliidae, Gerridae) (Andersen 1982; Polhemus et al. (2008), but in a study combining DNA sequence data and morphological characters, Damgaard (2008b) questioned the established relationships of superfamilies and found that Veliidae, comprising approximately 50% of all gerromorphan bugs, is paraphyletic with regards to Gerridae. Since then, more studies have questioned phylogenetic relationships among families in Gerromorpha (Damgaard 2013). Most recently, Armisén et al. (2022) found Hydrometridae to be polyphyletic based on transcriptome data.

Owing to their unique pleustonic lifestyle (Cheng 1975), Gerromorpha has attracted considerable research attention on evolutionary adaptations and mechanisms to life on the water surface (Andersen 1982, Hu and Bush 2010, Crumière et al. 2016, Mahadik et al. 2020, Cheng and Mishra 2022). This group is also unique because its members have been able to colonize all types of aquatic habitats—limnic, brackish, marine, and even oceanic—in every continent except Antarctica (Andersen 1982). It has been estimated that members of the Gerromorpha have independently evolved from limnic to marine environments at least 14 times (Andersen 1999).

Examples of diverse marine genera include *Hermatobates* Carpenter, 1892 (Gerrormorpha: Hermatobatidae), *Halovelia* Bergroth, 1893 (Gerromorpha: Veliidae), *Thetibates* Polhemus & Polhemus, 1996 (Gerromorpha: Gerridae), and *Halobates* Eschscholtz, 1822 (Gerromorpha: Gerridae). It is notable that *Hermatobates* represents the first (and oldest) lineage of the Gerromorpha to invade the marine environment (Andersen 1999, Wang et al. 2023), while 5 species of *Halobates* from Halobatinae Bianchi, 1896, have adapted to live a truly oceanic lifestyle, thousands of kilometers away from shore (Andersen and Cheng 2004).

A detailed account of the taxonomic history of Halobatinae can be found in Roman-Palacios et al. (2020) and is thus not repeated here. The subfamily is composed of 2 tribes, Halobatini Bianchi, 1896 and Metrocorini Matsuda, 1960, with Halobatini being marine and Metrocorini being limnic (Matsuda 1960). The current classification based on molecular sequence data of the 2 tribes suggests there are 9 genera (Table 1), though further taxonomic revision is needed (Damgaard 2008b, Román-Palacios et al. 2020). Studies on Halobatinae to date have traditionally focused on classical taxonomy (Cheng 1965, 1966, Chen and Nieser 1993a, 1993b, Chen and Zettel 1998, Andersen and Cheng 2004, Tran et al. 2023), whereas genomic sequencing work has been restricted to *Halobates* (Andersen et al. 2000, Damgaard et al. 2000, Leo et al. 2012, Wang et al. 2021a, Chang et al. 2022), with some publications on phylogenetic reconstructions of Gerromorpha that included other members of Halobatinae (Damgaard et al. 2005, Damgaard 2008b).

The most comprehensive phylogenetic reconstruction of Halobatinae was performed by Roman-Palacios et al. (2020) based on 4 partial genes—mitochondrial COI, COII, 16S rRNA, and nuclear 28S rRNA. Their 4-gene phylogeny confirmed that: (1) the limnic lifestyle of Metrocorini was ancestral within Halobatinae, with Metrocorini likely to be paraphyletic; and (2) the marine lifestyle evolved from the common ancestor of *Asclepios* + *Halobates*. While the authors noted other potential hypotheses, such as the possibility of 3 independent invasions of *Halobates* into the open ocean, their branch support values were too low to provide firm conclusions. A workaround would be to increase the number of loci used, which is known to be correlated with phylogenetic accuracy (Rokas and Carroll 2005).

**Table 1.** Present classification of Halobatinae Bianchi, 1896 (Gerromorpha: Gerridae). Number of known species within each genus (species) and number of species sequenced in this study (sequenced)

| Subfamily Halobatinae Bianchi, 1896 | Species | Sequenced |
|---|---|---|
| Tribe Halobatini Bianchi, 1896 | | |
| Genus *Asclepios* Distant, 1915 | 3 | 2 |
| Genus *Halobates* Eschscholtz, 1822 | 50 | 27 |
| Tribe Metrocorini Matsuda, 1960 | | |
| Genus *Esakia* Lundblad, 1933 | 10 | 3 |
| Genus *Eurymetra* Esaki, 1926 | 7 | 1 |
| Genus *Eurymetropsiella* Poisson, 1950 | 3 | 0 |
| Genus *Eurymetropsielloides* Poisson, 1956 | 1 | 0 |
| Genus *Eurymetropsis* Poisson, 1948 | 2 | 0 |
| Genus *Metrocoris* Mayr, 1865 | 80 | 12 |
| Genus *Ventidius* Distant, 1910 | | |
| Subgenus *Ventidioides* Hungerford & Matsuda, 1960 | 9 | 2 |
| Subgenus *Ventidiopsis* Miyamoto, 1967 | 2 | 0 |
| Subgenus *Ventidius* Distant, 1910 | 12 | 5 |

Advancements in high-throughput sequencing technology have revolutionized the ease with which genomic data can be generated for phylogenomic inferences and even population genomics (Johnson 2019, Quek and Huang 2022). Consequently, a myriad of different approaches have arisen in recent years (Dodsworth 2015, Wachi et al. 2018), e.g., target enrichment via bait capture (Blaimer et al. 2016, Call et al. 2021, Pauli et al. 2021), genome subsampling methods like RAD-seq (Storer et al. 2017), transcriptome sequencing and assembly (Wang et al. 2021b, Armisén et al. 2022), and genome skimming (Linard et al. 2015, Trevisan et al. 2019, Zheng et al. 2020). Genome skimming involves shallow sequencing of genomic DNA to retrieve high-copy fractions of the genome (Straub et al. 2012) and requires the least effort and reagent costs (Dodsworth 2015). Mitochondrial DNA makes up 0.5–5% of genomic DNA extracts (Arribas et al. 2016, Crampton-Platt et al. 2016), so it is a suitable target marker for retrieval from low-coverage, genome skims. Other candidate loci include chloroplasts for plants (Bakker et al. 2015, Nevill et al. 2020), as well as the ribosomal RNA tandem cluster (Grandjean et al. 2017).

The primary goal of our study was to recover high-copy number genes for phylogenetic analysis with genome skimming. Specifically, we tested if the inclusion of more mitochondrial and nuclear genes would provide better branch support and resolve phylogenetic relationships among aquatic insects. We generated and analyzed sequence data for several genera of marine insects, but focused on Halobatinae. We also included more freshwater species of *Metrocoris* Mayr, 1865, *Esakia* Lundblad, 1933, and *Ventidius* Distant, 1910 to better resolve relationships between genera. The sequence data generated improved the availability of genetic resources for aquatic insects, particularly marine bugs, which were poorly studied and underrepresented in sequence databases (Hotaling et al. 2020).

## Materials and Methods

### Taxon Sampling

Eighty-five specimens were obtained for this study, which included freshly collected and museum loaned specimens (Supplementary File S1), and data from 15 low-coverage whole genome libraries from Chang et al. (2022). Fresh specimens were collected from Singapore (Permits NP/RP21-056 and NP/RP22-102), Australia (Permit G22/47448.1), and Papua New Guinea (Permit AA927408) by sweeping the water surface with hand nets and preserved in either molecular-grade ethanol or salt-saturated dimethyl sulfoxide before DNA extractions.

Multiple individuals of the same species were sequenced where possible, to capture baselines for geographic and intraspecific variability or to test species hypothesis (see Supplementary File S2 for more information). Despite the limited sampling effort, results could still be useful for detecting potential cryptic speciation. For oceanic *Halobates* species, particularly *Halobates germanus* White, 1883, *Halobates micans* Eschscholtz, 1822, and *Halobates sericeus* Eschscholtz, 1822, we sequenced samples collected across their known geographic ranges to account for potential genetic differentiation or cryptic speciation. Past studies have noted certain genetic differences among *H. micans* across Atlantic, Indian, and Pacific Oceans and *H. sericeus* populations from Northern and Southern Pacific Ocean (Andersen et al. 2000, Leo et al. 2012). We also tested deeper sequencing to discern if subpopulations formed distinct clades

by sequencing 2 specimens each of *H. germanus* from the Indian and Pacific Oceans, one specimen each of *H. micans* from the Atlantic, Indian, and Pacific Oceans, and 1 specimen each of *H. sericeus* from the Northern and Southern Pacific Ocean.

Our taxon coverage of Halobatinae included 52 species across 7 genera (including subgenus), 6 of which were featured in Roman-Palacios et al. (2020) (Table 1). We were unable to obtain specimens from 3 Afrotropical genera, *Eurymetropsiella* Poisson, 1950, *Eurymetropsielloides* Poisson, 1956, and *Eurymetropsis,* Poisson, 1948. Fourteen *Halobates* samples analyzed here were re-sequenced from Roman-Palacios et al. (2020) but *Halobates darwini* Herring, 1961, *Halobates fijiensis* Herring, 1958, *Halobates mjobergi* Hale, 1925, and *Halobates rivularis* (Andersen and Weir 1994) were not re-sequenced as the DNA extracts were completely expended. The same was true for *Ventidius* (*Ventidiopsis*) *yangae* Chen & Zettel, 1999, from Roman-Palacios et al. (2020), but we successfully acquired other specimens of *Ventidius* (*Ventidioides*) Hungerford & Matsuda, 1960.

## DNA Extractions and Library Preparation

We followed the DNA extraction protocol by Chang et al. (2022). Briefly, genomic DNA was extracted using DNeasy Blood and Tissue Kit (Qiagen), following the manufacturer protocol for insects, with final elution of genomic DNA (gDNA) in molecular-grade water.

Low-coverage whole-genome libraries were prepared in 2 ways, depending on the DNA quantity. For specimens with ≥ 100 ng of gDNA, gDNA was sheared using the BioRuptor Pico (Diagenode) for either 7 or 13 cycles, of 30/30 s ON/OFF to target an insert size of ~300 and ~200 bp, respectively (manufacturer's recommendations). The former insert size was targeted for freshly caught specimens, while the latter was used for loaned museum specimens as the extracted DNA tended to be highly degraded. For specimens with < 100 ng of gDNA, the libraries were prepared with NEBNext Ultra II FS DNA Library Prep Kit (New England Biolabs) instead, in which gDNA was fragmented enzymatically. Samples were incubated for 13 min to obtain an insert size of ~200 bp.

All other libraries were prepared using NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) following the manufacturer protocols, but with TruSeq CD Dual Indexes (Illumina). Adapter-ligated libraries were PCR-enriched according to manufacturer recommendations. DNA clean-up and size-selection was done using either AMPure XP beads (Beckman Coulter), SPRIselect (Beckman Coulter), using manufacturer-recommended ratios, or with E-gel SizeSelect II 2% agarose gels (ThermoFisher Scientific). The libraries were sequenced either on Illumina HiSeq X or NovaSeq 6000 (150 bp paired-end). We targeted a minimum of 3 Gbp of sequence reads per sample, based on 2–3 × sequencing coverage recommended by Tan et al. (2021) for genome skimming, and the known ~1 Gbp *Gerris buenoi* Kirkaldy, 1911 genome size (Armisén et al. 2018).

## Sequence Quality Check and Assembly

Raw sequence reads were processed using default settings in fastp v0.23.2 (Chen et al. 2018). Assemblies were performed using SPAdes v3.13.0 (Bankevich et al. 2012), with default settings and error correction enabled (--careful mode). We identified and filtered potential contaminants (archaea, bacteria, virus, plasmid, human, UniVec_Core, protozoa, plant, and fungi) in assembled SPAdes contigs using kraken2 v.2.1.3 (Wood et al. 2019) against the kraken2 PlusPFP database (updated 5 June 2023; downloaded from https://benlangmead.github.io/aws-indexes/k2).

## Mitochondrial Genes

Mitogenomes were extracted from SPAdes assembly graphs with the get_organelle_from_assembly.py script in GetOrganelle v1.7.7.0 (Jin et al. 2020). We supplied closely related gerromorphan mitogenomes as references—*Asclepios apicalis* (Esaki, 1924) (KR920102) from Liu et al. (2021) was used for *Halobates* and *Asclepios* samples; *Metrocoris esakii* Chen & Nieser, 1993 (MT344112) from Ye et al. (2020) for *Esakia*, *Metrocoris*, and *Ventidius* samples; *Trepobates* sp. (MN027277) and *Rhagovelia reitteri* Reuter, 1882 (MN027271) from Esemu et al. (2019) for *Thetibates* and the Haloveliinae species, respectively; and finally, *Hydrometra greeni* Kirkaldy, 1898 (FJ456945) from Hua et al. (2009) for *Hermatobates djiboutensis* Coutiere & Martin, 1901. Preliminary tests revealed only minor increase in assembly coverage when congeneric sequences were supplied as reference (i.e., closely related species are effective as reference sequences).

When the assembly graph was too complicated to resolve, we used the get_organelle_from_reads.py script instead. Briefly, bowtie2 v2.3.5.1 (Langmead and Salzberg 2012) was used to map the filtered reads onto the same references for recruiting and extending mitogenome-associated reads. All target-associated reads were then assembled with SPAdes, followed by blastn in NCBI-BLAST+ v2.13.0 (Camacho et al. 2009) to the local, default, GetOrganelle database to isolate mitochondrial contigs, before circularizing into a complete mitogenome (Jin et al. 2020). There are no differences in mitogenome length and accuracy when retrieved from "get_organelle_from_assembly.py" or "get_organelle_from_reads.py" (Supplementary File S3).

Assembled mitogenomes were annotated for RNAs and protein-coding genes (PCGs) using the MITOS2 web server (Bernt et al. 2013, Donath et al. 2019), using the Invertebrate Mitochondrial Code (Code 5) and the RefSeq89 Metazoa reference for the annotation. Several annotations were manually rectified after MAFFT v7.511 (Katoh and Standley 2013) alignment of PCGs in frame with other GenBank mitogenome annotations, as described by Quek et al. (2021). Confirmatory Sanger sequencing was performed to confirm that any internal gaps within the alignment were not due to mis-assembly (Supplementary File S4). We also performed blastn searches of the genome-skimmed COI barcode against the *nt* database (downloaded 15 Feb 2023) to verify species identities.

## Nuclear Ribosomal Genes

For skimming of high-copy number ribosomal RNA reads, we used barrnap v0.9 (https://github.com/tseemann/barrnap) to isolate 18S, 5.8S, and 28S rRNA genes from the kraken2-filtered contigs. These were matched against the *nt* nucleotide database and any rRNA sequences that did not hit Gerromorpha were discarded. The rRNA sequences were visualized and edited (if needed) on Geneious Prime v2023.1.1 (http://www.geneious.com/).

## Phylogenetic Analyses

Our phylogenetic matrix comprised 13 mitochondrial PCGs and 2 rRNAs (12S and 16S), along with 3 nuclear rRNA genes (18S, 5.8S, and 28S) of our samples together with 26 Gerromorpha mitogenomes from GenBank (Hua et al. 2009, Li et al. 2017, Sun et al. 2018, Esemu et al. 2019, Ye et al. 2020, Zheng et al. 2020, Cui et al. 2021, Liu et al. 2021) (Supplementary File S5). Full-length sequences were aligned by gene using the L-INS-i algorithm in MAFFT, with nuclear rRNA genes were treated as missing data for the 26 GenBank mitogenome sequences. The rRNA alignments were further curated with Gblocks v0.91b (Castresana 2000, Talavera

and Castresana 2007), allowing for gap positions in the final blocks. Finally, sequences were concatenated in Geneious Prime to form the phylogenetic matrix that comprised 110 samples and 18,436 bp. We partitioned the matrices by gene and selected the best evolutionary models with ModelTest-NG v0.1.7 (Darriba et al. 2020).

We conducted maximum likelihood (ML) and Bayesian inference (BI) reconstructions, providing evolutionary models for each gene partition based on the Akaike information criterion (AIC). *Stenopirates* sp. Walker, 1873 (KP406518) was selected as the outgroup for both ML and BI analyses, based on past phylogenetic work (Li et al. 2017, Armisén et al. 2022). ML trees were inferred using RAxML-NG v1.2.0 (Kozlov et al. 2019), with 100 starting trees (50 random and 50 parsimony-based). Node supports were quantified using 1,000 nonparametric bootstrap pseudoreplicates (Felsenstein 1985) and assessed to have converged at the default 3% cutoff before mapping onto the best tree (Pattengale et al. 2010). BI analyses were performed with MrBayes v3.2.7a (Ronquist et al. 2012), where we implemented 4 Markov chains Monte Carlo (MCMC) of 20,000,000 generations across 3 runs, sampling one tree per 100 generations. MCMC convergence was assessed using Tracer v1.7 (Rambaut et al. 2018), after discarding the first 20,001 trees as burn-in. Trees were visualized using the Interactive Tree of Life (iTOL) web tool (Letunic and Bork 2021).

### Ancestral State Reconstruction

Reconstruction of habitat and salinity preference was performed using R v4.3.3 in RStudio v2023.06.2, with the packages ape v5.7-1 (Paradis and Schliep 2018), phytools v.2.1-1 (Revell 2012), and geiger v2.0.11 (Pennell et al. 2014). We pruned the input tree, leaving only single species representatives of Halobatinae. We assigned each taxon one of 3 discrete categories: "Limnic," "Coastal," and "Oceanic" based on their known habitat and salinity preference—"Limnic" for freshwater habitats; "Coastal" for coastal habitats like reef flats or mangroves; and "Oceanic" for strictly open ocean species. We then ran 3 different transition models of trait evolution using the ace function in ape—equal-rates model (ER), symmetrical model (SYM), and all-rates-different model (ARD). Model fit was assessed by comparing the AIC scores from fitDiscrete function in geiger (999 iterations). We also adopted an MCMC approach using the make.simmap function in phytools for ancestral state reconstruction using 300 simulations, with Q set to MCMC. The output was compared for consistency and results were mapped onto the phylogenetic reconstruction using phytools.

## Results

### Genome Skimming for High-Copy Number Genes

A total of 2,420,224,795 raw reads (~363 Gbp) were generated from 85 low-coverage genome libraries. Reads varied between 3,971,488 and 92,737,764 reads per sample (Supplementary File S1). We retrieved 76 complete mitogenomes that measured ~15,000 bp in length, with 3–44,623 × coverage obtained per mitogenome (Fig. 1A and Supplementary File S1). Mitogenome coverage was weakly correlated with sequenced reads and specimen age (Supplementary Fig. S1). The 76 mitogenomes were successfully annotated for 37 known genes (13 PCGs, 2 rRNAs, and 22 tRNAs) with MITOS2. All mitogenomes adhered to the ancestral insect mitogenome gene order (designated "Type A") (Hua et al. 2008). Initiation codons for the 13 PCGs also adhered to canonical codons under the Invertebrate Mitochondrial Code (code 5). Start codons for genes like atp6 and cox3 were invariable across our dataset, with only 1 initiation codon observed, while other genes like nad3 had 4 possible variations in its start codon (Fig. 1B).

The 8 species for which we were unable to obtain the complete mitogenome were: *Esakia johorensis* Cheng, 1966, *Esakia lundbladi* Cheng, 1966, *Eurymetra natalensis* Distant, 1903, *Halobates elephanta* Andersen & Foster, 1992, *Halobates japonicus* Esaki, 1924, *Metrocoris bilobatoides* Chen & Nieser, 1993, *Ventidius* (*Ventidioides*) *karen* Lansbury, 1990, and *Ventidius* (*Ventidioides*) *pulai* Cheng, 1965. These samples had low coverage of mitochondrial contigs (3.7–65.4×), likely due to a low proportion of mitochondrial reads in the prepared libraries. In any case, we were still able to retrieve mitochondrial genes with varying degrees of success, with species like *E. johorensis* and *E. natalensis* only missing a partial segment of a single mitochondrial gene, to more fragmented assemblies like *H. elephanta*, where we only obtained 8 partial mitochondrial gene segments (Table 2). We merged data from GER1094 (nad2 and cox1) and GER1093 (all remaining genes) to obtain the complete set of mitochondrial genes for *M. bilobatoides*.

Our genome skimming effort for ribosomal RNA genes was successful; we recovered all 3 nuclear rRNA genes in all samples, with varying degrees of completeness (in terms of gene length). There were 5 samples for which we were unable to obtain the full-length 5.8S rRNA gene (155 bp). The 18S gene (full length ~1.8 kbp) was recovered across all samples at ≥ 80% completeness and was only highly fragmented for samples GER029 and GER030. Success rate for the 28S rRNA genes was the most varied; we recovered the full-length 28S rRNA gene (~4 kbp) for 33 samples, and there were 12 samples where we recovered < 50% of the complete 28S rRNA gene. We retained 95%, 94%, and 69% of the respective 18S, 5.8S rRNA, and 28S gene alignments post Gblocks-curation for phylogenetic analyses.

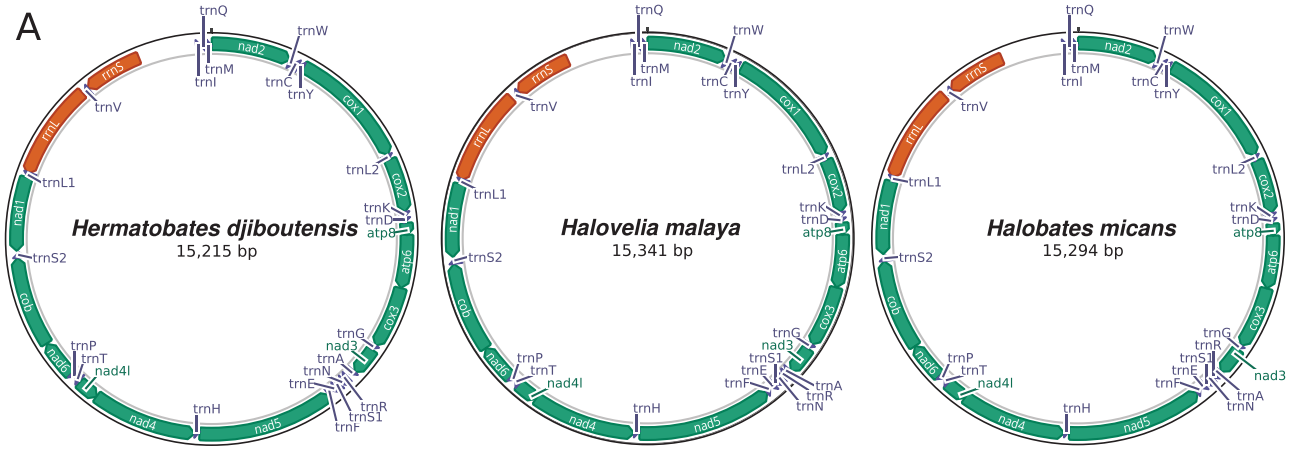
### Phylogenetic Analyses

Analyses with ML and BI on 110 gerromorphan insects using 15 mitochondrial genes and 3 nuclear rRNA genes generated identical phylogenies, with most branches maximally supported (Fig. 2). Analyses conducted using only single species representatives also returned identical tree topology (Supplementary Fig. S2). *Hermatobates djiboutensis* is the first branching marine lineage on the tree (Fig. 2A).

Our results indicate that "Veliidae" is paraphyletic, with Haloveliinae more closely related to Gerridae (Branch support: 94/1.00) than to *Rhagovelia* or *Perittopus*. There is, however, maximum support for the monophyly of Haloveliinae (Fig. 2B).

The monophyly of Gerridae is moderately supported in our analysis (Branch support: 79/1.00), but maximally supported for the subfamilies Gerrinae, Ptilomerinae, and Halobatinae. There is also maximum support for the sister relationship between Ptilomerinae and Halobatinae. Within Halobatinae, relationships are congruent with Roman-Palacios et al. (2020), albeit with stronger support values. There are 2 major clades within Halobatinae (Fig. 2C); the first comprising *Eurymetra* and *Metrocoris* (Fig. 2D) and a second containing *Asclepios*, *Esakia*, *Halobates*, and *Ventidius* (Fig. 2E). The traditionally defined Metrocorini is thus paraphyletic (cf. Table 1). *Metrocoris* is monophyletic, but does not conform to traditional species groups delineated by morphology (see Supplementary Table S1). *Esakia* is embedded within *Ventidius*, and sister to *Ventidius* (*Ventidioides*) samples GER1054 and GER1057. *Ventidius sensu stricto* is paraphyletic; the *Ventidius distanti* (formerly *V. modulatus*) group (Fig. 2F) is sister to the *Ventidius* (*Ventidioides*) + *Esakia* clade, whereas the *Ventidius aquarius* group (Fig. 2G) is sister to

A




B

| *nad2* | *cox1* | *cox2* | *atp8* | *atp6* | *cox3* | *nad3* | *nad5* | *nad4* | *nad4l* | *nad6* | *cob* | *nad1* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ATA<br>ATT | TTG<br>ATG<br>GTG | ATT<br>ATA<br>ATC | ATA<br>ATT<br>ATC | ATG | ATG | ATT<br>ATA<br>ATC<br>ATG | TTG<br>ATG<br>ATT | ATG | ATT | ATG<br>ATA<br>ATT | ATG | ATA<br>ATT<br>ATG |

**Fig. 1.** A) Representative circular mitogenome maps of aquatic Gerromorpha: left, *Hermatobates djiboutensis*; middle, *Halovelia malaya*; right, *Halobates micans*, annotated for protein-coding, ribosomal RNA, and transfer RNA genes. Mitogenome maps were created with Geneious Prime v2023.1.1 (http://www.geneious.com/). B) Gene order of protein-coding genes is also expressed linearly, with observed initiation codons listed. The gene order [coined "Type A" by Hua et al. (2008)] adheres to the ancestral insect mitogenome gene order.

the *Asclepios* + *Halobates* clade instead (Branch support: 71/1.00). *Asclepios* is sister to *Halobates* with maximum support, and each genus is monophyletic.

For *Halobates*, we recovered similar cladistic relationships as Roman-Palacios et al. (2020). Similarities included the recovery of the following group relationships: (1) *Halobates katherinae* Herring, 1958, *Halobates mariannarum* Esaki, 1924, and *Halobates bryani* Herring, 1961; (2) *Halobates hayanus* and *Halobates sericeus*; and (3) *Halobates proavus* and *Halobates micans*. Species relationships within each of the abovementioned clades are similar to Roman-Palacios et al. (2020). Most clades also conform to traditional species groups delineated with genital morphology (see Supplementary Table S1). There are, however, 3 exceptions found in our species groups: (1) *Halobates regalis* is paraphyletic; *Halobates whiteleggei* Skuse, 1891 is sister to *H. hayanus* and *H. sericeus*, while the other *H. regalis* group members, *Halobates peronis* Herring, 1961 and *Halobates sexualis* Distant, 1903, form their own clade as the first branching lineage in *Halobates*. Previously, *H. peronis* and *H. sexualis* are sister to what was previously known as the *Halobates* (*Hillelia*) subgenus and are embedded within *Halobates* (Román-Palacios et al. 2020). (2) *Halobates japonicus* nests within *Halobates matsumurai*; and (3) *H. hayanus* nests within *H. sericeus* (maximum support) (Fig. 2). The 3 oceanic *Halobates* species—*H. germanus*, *H. micans*, and *H. sericeus* sampled from the various ocean basins, all form monophyletic clades with maximum support. Interestingly, *Halobates flaviventris* Eschscholtz, 1822 from 3 different locations do not form a monophyletic clade (Fig. 2), with GER1347 and JD0083 more closely related to *Halobates hawaiiensis* Usinger, 1938.

### Ancestral State Reconstruction of Habitat and Salinity Preference

All 3 models generated similar ancestral state reconstructions in Halobatinae (Fig. 3 and Supplementary Fig. S3), but we selected the SYM model, which scored the lowest AIC value (SYM: 42.015; ARD: 44.985; ER: 47.050). Halobatinae evolved from limnic to coastal, and eventually to an oceanic lifestyle, with Metrocorini being limnic, and the transition to coastal habitats occurred in the ancestor of *Asclepios* + *Halobates*, while transition to an oceanic lifestyle occurred within *Halobates* (Fig. 3). Interestingly, our analysis also suggests that the oceanic lifestyle of present-day *Halobates* evolved 3 times: twice independently for *H. sericeus* and *H. germanus*, and once more at the common ancestor for *H. micans*, *H. sobrinus*, and *H. splendens*.

## Discussion

In this study, we reconstructed the phylogeny of 110 aquatic insects based on a 18,436 bp matrix comprising 15 mitochondrial genes and 3 nuclear rRNA genes. Overall, phylogenetic relationships inferred were well-supported across most branches (Fig. 2) and corroborated results from Damgaard (2008b). We present here taxonomic and evolutionary implications of our results and assess the utility of genome skimming for phylogenetics.

### Taxonomy of Halobatini and Metrocorini

The monophyly of Matsuda's (1960) classification has been questioned (Damgaard 2008b), but no formal decisions were made due to low phylogenetic resolution achieved with few molecular markers. Even in Roman-Palacios et al. (2020), the topologies of *Eurymetra* and *Metrocoris* differed between ML and BI analyses. Broader sequencing efforts and inclusion of more genes for phylogenetic analyses in this study, however, enabled recovery of 2 major clades within Halobatinae; *Eurymetra* and *Metrocoris* forming one, and *Asclepios*, *Esakia*, *Halobates*, and *Ventidius* forming the other. Given the maximum support values from our analysis, we propose reorganizing Metrocorini to comprise *Eurymetra* and *Metrocoris*,

**Table 2.** Mitochondrial genes that were present (P), incomplete (I), or missing (M) in species where complete mitogenome could not be assembled. *Metrocoris bilobatoides* was omitted because all mitochondrial genes could be found after merging data from 2 samples

| Species / Gene | nad2 | cox1 | cox2 | atp8 | atp6 | cox3 | nad3 | nad5 | nad4 | nad4l | nad6 | cob | nad1 | rrnL | rrnS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Esakia jobrensis* | I | P | P | P | P | P | P | P | P | P | P | P | P | P | P |
| *Esakia lundbladi* | M | I | P | P | P | P | I | I | P | P | P | P | P | P | P |
| *Eurymetra natalensis* | P | P | P | P | P | P | P | P | P | P | P | I | P | P | P |
| *Halobates elephanta* | M | I | I | M | I | I | M | I | I | M | M | I | I | I | M |
| *Halobates japonicus* | M | P | P | M | I | P | I | I | M | M | I | P | P | P | P |
| *Ventidius karen* | P | I | P | P | I | P | P | P | P | P | P | P | P | P | P |
| *Ventidius pulai* | P | P | P | P | P | P | P | P | P | P | P | P | P | P | P |

and for *Esakia* and *Ventidius* to be reclassified under Halobatini. Indeed, *Ventidius* species have been observed at intertidal zones (Yang et al. 1999), indicating a degree of salinity tolerance much like *Asclepios* and *Halobates*.

Our analyses further confirmed the paraphyly of 2 groups in *Ventidius* sensu stricto; *Ventidius aquarius* is sister to *Asclepios* + *Halobates*, whereas *Ventidius distanti* is sister to *Esakia* + *Ventidioides*. If monophyly of *Ventidius* is to be maintained, then *Ventidius* sensu stricto must be restricted to encompass only the *V. aquarius* group (type species: *Ventidius aquarius* Distant, 1910), and a new genus named for the *V. distanti* group. Chen and Zettel (1998) have previously suggested that the species groups could be different genera as each group has clearly defined apomorphic characters, e.g., long and slender antennal segments with stout middle and hind femora in *V. aquarius* group, whereas the *V. distanti* group are characterized by shorter and less slender antennal segments and slender middle and hind femora. There is thus corroborative morphological evidence to support our phylogenetic results. It is also likely that *Ventidioides* could be raised to genus level (type species: *Ventidius kuiterti* Hungerford & Matsuda, 1960), given its sister relationship to *Esakia* (Fig. 2). We are unable to confirm the relationships between *Esakia*, *Ventidioides*, and *Ventidiopsis*, although *Ventidiopsis* and *Esakia* are sister groups according to Roman-Palacios et al. (2020). Further sampling from the *Ventidioides kuiterti* and *Ventidiopsis* groups will be needed to determine their phylogenetic placements and nomenclatural revision of *Ventidius*.

## Phylogeny of *Halobates* and Evolution into the Open Ocean

The genus *Halobates* is separated into coastal and oceanic groups based on morphology and habitat preference (Herring 1961), with the coastal species further delineated into several species groups by genitalia morphology and single-gene sequences (Andersen 1991, Damgaard et al. 2000). Relationships between *Halobates* species groups largely agree with our phylogenetic results and mirrored Roman-Palacios et al. (2020). However, the *Halobates regalis* group is paraphyletic here, with *H. peronis* and *H. sexualis* distinct from *H. whiteleggei* and *H. zephyrus*. The former 2 form the first branching lineage of *Halobates* in our dataset (Fig. 2, maximum support); previously, the *H. regalis* group is more closely related to the *H. sericeus*, *H. proavus*, and *H. micans* species groups (Román-Palacios et al. 2020). This new position is noteworthy as they are closely related to the recently synonymized subgenus *Halobates* (*Hilliella*) China, 1957 (Román-Palacios et al. 2020), which that the subgenus occupies a similar position by association. This finding reignites discussions of *Hilliella* as the primitive form of *Halobates* (Andersen and Weir 1994). We have been unable to re-sequence *Halobates* (*Hilliella*) *mjobergi* Hale, 1925 from Roman-Palacios et al. (2020) to test this hypothesis, and recommend future work to prioritize sequencing of species in this group.

Critically, we have obtained well-supported relationships for the clade of *Halobates* that contain the 5 open ocean species, with *H. micans* group sister to *H. proavus* group, and both clades in turn sister to a clade containing *H. hayanus* and *H. sericeus* groups. We propose a merger of *H. hayanus* and *H. sericeus* groups since they form a well-supported monophyletic clade. Moreover, males of this group all possess black spines on the lateral sides of the proctiger, which is likely an apomorphy for this new *H. sericeus* clade (Andersen 1991, Andersen and Cheng 2004).

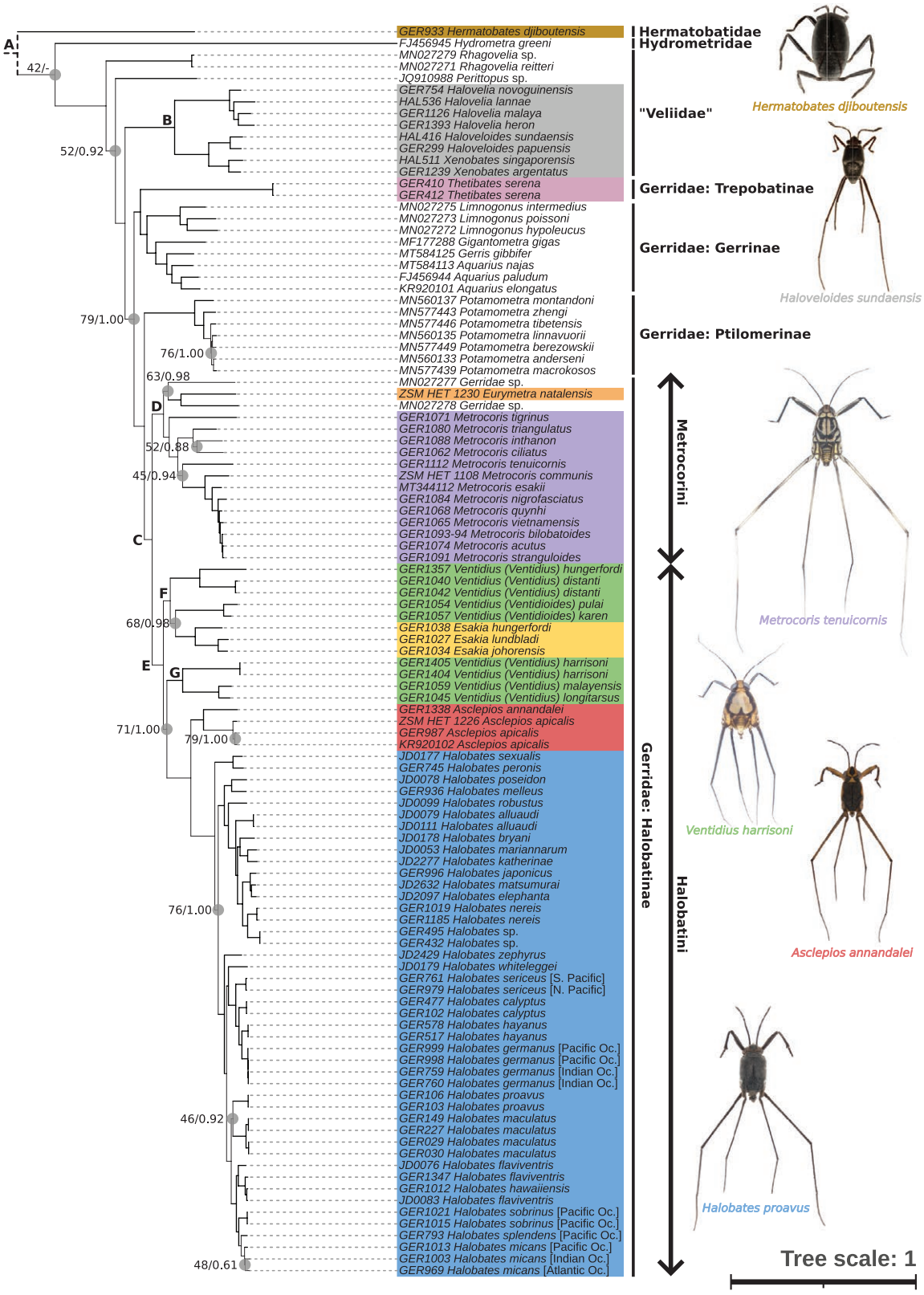



**Fig. 2.** Phylogenetic reconstruction of 110 Gerromorpha samples based on a concatenated matrix (18,436 bp) comprising 15 mitochondrial genes (13 PCGs and 2 rRNA) and 3 nuclear rRNA genes. Samples from this study are shaded, while JD-coded specimens were resequenced from Román-Palacios et al. (2020). Enicocephalomorpha outgroups removed for figure clarity. A) *Hermatobates djiboutensis* is the first branching marine insect lineage in this study; B) Clade Haloveliinae; C) Clade Halobatinae; D) Clade Metrocorini; E) Clade Halobatini; F) *Ventidius* is paraphyletic; G) *Ventidius aquarius* group. Proposed re-classification of Metrocorini and Halobatini are bounded by arrows. Names for MN027277 (*Trepobates* sp.) and MN027278 (*Metrocoris* sp.) were amended to Gerridae sp. to avoid confusion due to potential misidentification. Only branches with < 85 maximum likelihood bootstrap support and < 0.90 posterior probability, respectively, are depicted with circles. *Asclepios annandalei, Metrocoris tenuicornis, Halobates proavus,* and *Ventidius harrisoni* were imaged by Harald Bruckner, NHMW, and reused with permission; all other images are the author's own.

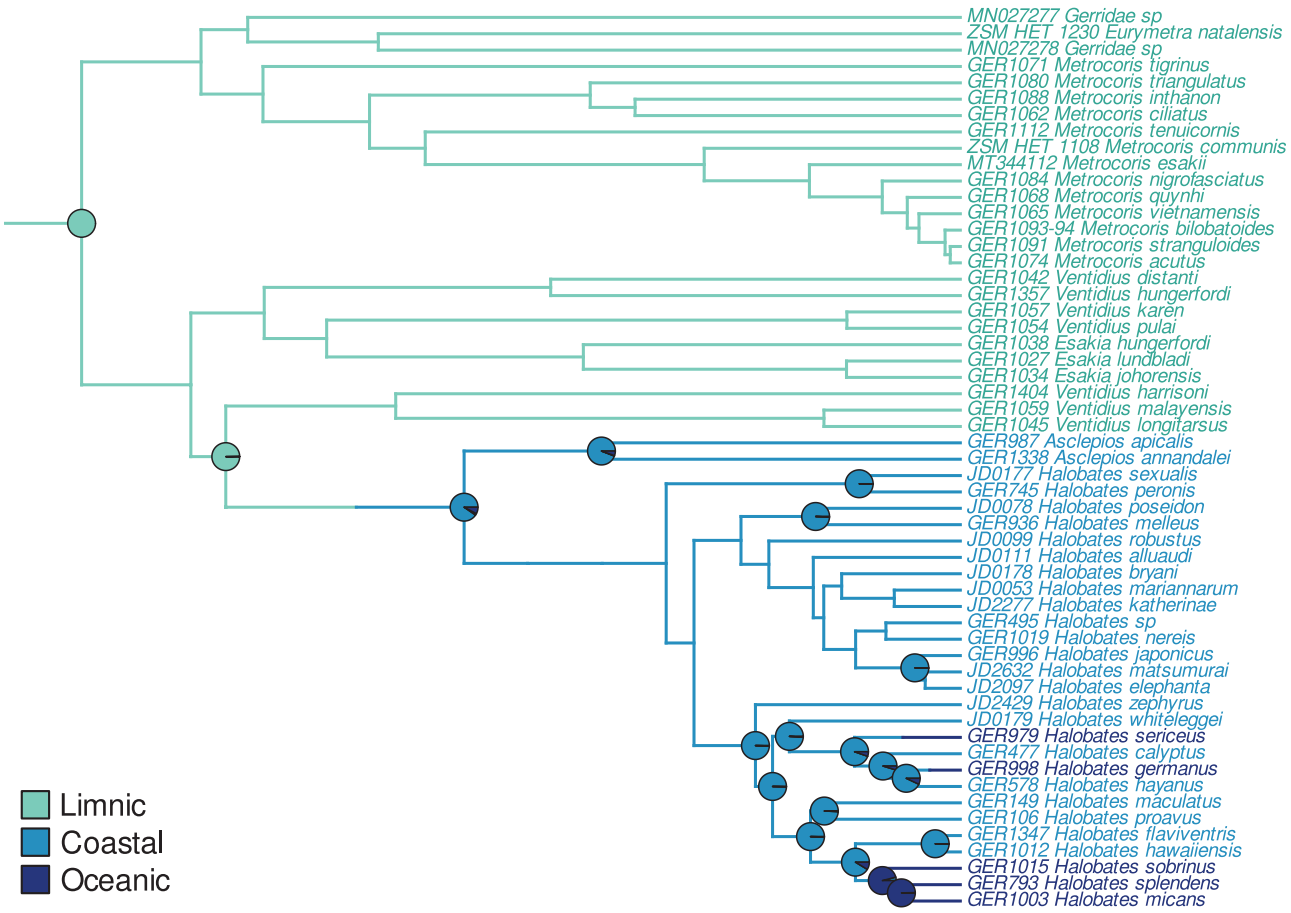



**Fig. 3.** Ancestral state reconstruction of salinity and habitat preference in Halobatinae, reconstructed using the make.simmap function in phytools, in R, using 300 Markov chains Monte Carlo simulations of the SYM model, for the 3 discrete categorical states: "Limnic," "Coastal," and "Oceanic." Pies represent the probability of ancestral character state, and those that reflect only 1 state are removed for figure clarity.

For *H. flaviventris*, computed pairwise *p*-distances (Supplementary File S6) show that GER1347 and JD0083 differ from JD0076 by ~6%, but are only ~3% different from *H. hawaiiensis* (GER1012). Given the known ~3–4% maximum intraspecific variability in *Halobates*, JD0076 likely represents a different lineage. Unfortunately, we are unable to verify the morphologies of JD0076 and JD0083 (vouchers cannot be tracked down) to eliminate potential misidentification. GER1347 and GER1012, however, match their respective species descriptions, with the latter possessing dense, stiff, black bristles on the hind acetabula—a key identifying character to separate *H. hawaiiensis* from *H. flaviventris* (Andersen and Cheng 2004). This could imply that either the trait is not useful for defining species or that introgression or incomplete lineage sorting resulted in similar mitogenomes. More in-depth sampling of these 2 species is needed to untangle this observation.

Our phylogeny also lends strong support to the hypothesis that the oceanic lifestyle evolved at least 3 times within *Halobates*. One habitat transition occurred at the common ancestor of *H. micans* + *H. splendens* + *H. sobrinus*, with 2 other independent invasions by *H. sericeus* and *H. germanus* (Fig. 3). It was previously thought that all 5 species form a single monophyletic clade, given their smaller size, similar color patterning, and oceanic lifestyle (Herring 1961). However, a close examination of their genital morphology together with the mitochondrial COI locus revealed this to be unlikely, with *H. micans*, *H. splendens*, and *H. sobrinus* more closely related to *H. flaviventris* (coastal), whereas *H. sericeus* and *H. germanus* are more closely related to *H. hayanus* group (coastal), implying independent invasions to the oceans (Andersen 1991, Damgaard et al. 2000). What all 3 independent coastal-to-oceanic transitions share is that they evolved from coastal marine lineages with transoceanic rather than endemic or nearshore distributions (Ikawa et al. 2012). The fact that nearshore relatives such as *H. calyptus*, *H. flaviventris*, *H. maculatus*, and *H. proavus* are more commonly found at fore reefs (Polhemus and Polhemus 1991, 2006), where sea surface conditions are undoubtedly harsher (relative to bays and mangroves), does lend support to the more gradual evolution of the oceanic lifestyle (Damgaard et al. 2000, Andersen and Cheng 2004). It is also highly plausible that the Indo-Pacific is the geographic origin of the oceanic lifestyle in *Halobates*, given that all 5 oceanic species are found in the Pacific, and the Pacific population of *H. micans* is sister to Indian and Atlantic Ocean *H. micans* lineages (Fig. 2; maximum support). Furthermore, of more than 40 coastal *Halobates* species described, only 9 are known from the Indian Ocean, while none had been found from the Atlantic Ocean (Ikawa et al. 2012). These evolutionary patterns lend credence to the hypothesis that *H. micans* evolved somewhere in the Indo-Pacific and dispersed later to the other oceans. A more detailed study with more specimens from the 3 oceans is needed to interrogate this theory.

## Genome Skimming for Phylogenetic Reconstruction

We applied genome skimming to extract high-copy number loci (i.e., mitogenomes and nuclear ribosomal RNA) for phylogenetic

reconstruction of the Halobatinae and successfully retrieved all 3 nuclear ribosomal RNA genes from all samples (to varying degrees of completion) and recovered 76 complete mitogenomes. Phylogenetic reconstructions with a concatenated matrix of 18 genes resulted in a generally well-resolved tree, with maximum support values for most branches (Fig. 2). Our study joins a growing body of work demonstrating the value of genome skimming for retrieving and analyzing more higher-resolution phylogenetic markers (Linard et al. 2015, Chen et al. 2019, Raupach et al. 2022, Duan et al. 2023).

Importantly, we found genome skimming to be generally forgiving in its input DNA requirements. We were able to work with highly degraded DNA and/or very small input amounts (~10 ng), thus allowing us to use museum-preserved collections, which may not necessarily be optimally preserved but still recovering phylogenetically-informative loci at sufficient depth. For instance, 22 out of 24 (91.6%) of our freshwater Halobatinae samples were obtained from museum collections where the DNA was generally degraded, while 14 samples re-sequenced from Roman-Palacios et al. (2020) (coded with JD prefix) had very little input DNA for library construction (0.8–200 ng, average ~40 ng). Yet, we were still able to retrieve complete mitogenomes for 20/24 and 13/14 samples, respectively, and all 3 nuclear genes for all samples. Given the near-global distribution of Halobatinae, it would have been logistically challenging to collect fresh specimens for sequencing. We also found that sample age had a very weak positive correlation with mitogenome coverage (Supplementary Fig. S1B), further demonstrating the powerful potential of genome skimming for getting "more from less" as aptly described by Tan et al. (2021). Furthermore, genome skimming could be carried out affordably vis-à-vis other methods like target enrichment or RAD-seq, which require additional treatments of genomic DNA; or even transcriptomes, which typically require freshly collected specimens and high RNA integrity.

Nevertheless, genome skimming is a random, almost non-targeted approach to extract phylogenetically informative loci (Supplementary Fig. S1A). Despite incomplete mitogenomes for samples GER996 and JD2097, we were able to retrieve full-length ribosomal RNA genes, while the opposite was true for samples GER029 and GER030 (i.e., complete mitogenomes but highly fragmented rRNAs). The design of more efficient capture or amplification methods for target genes could thus help lower sequencing costs in the future. We note that there are hybrid-capture bait sets designed to work on Hemiptera (Faircloth 2017), but the locus capture success rate was lower than expected when tested on actual samples (Kieran et al. 2019), so refinements to the methodology are needed. Given the vast improvements in the accuracy of third-generation, long-read sequencing technology, we could potentially see the revival in popularity of long-range PCRs for targeted amplification of long amplicons. Past studies have attempted this for long ribosomal RNA in arthropods (Krehenwinkel et al. 2019). The field of long read sequencing of mitogenomes is still relatively nascent for insects, though there are publications on fish (Ramón-Laca et al. 2023), other vertebrates (Karin et al. 2023), and even environmental DNA (Deiner et al. 2017). The generation of more genomic resources like those generated in this study will go a long way into building more comprehensive databases that can not only support more confident species identifications but also aid in other aspects such as the design of suitable primer binding sites.

## Conclusion

In this study, we have constructed a well-resolved phylogeny for marine insects in Gerromorpha with emphasis on Halobatinae. Importantly, Metrocorini is paraphyletic and *Esakia* and *Ventidius* are transferred to Halobatini. *Ventidius* is also paraphyletic, but we suggest greater sampling before any nomenclatural decisions are made. Our findings corroborate previous studies that the ancestor of *Asclepios* + *Halobates* was likely limnic or mangrove-dwelling, and further confirm that *Halobates* invaded the oceans on 3 independent occasions. Our results highlight the utility of genome skimming for recovering more molecular markers to generate well-supported phylogenies, which can be further applied to help advance our understanding of the evolution of aquatic insects.

Reconstructing relationships within the Halobatinae and Gerromorpha remains a work in progress. For Halobatinae, we recommend increased sampling effort for African and Madagascan genera *Eurymetropsiella*, *Eurymetropsielloides*, and *Eurymetropsis*, to better understand how these genera fit within Halobatinae. Likewise, we recommend better sample representation of Australian *Halobates* species. Australia is the only known locality of freshwater *Halobates*—*Halobates acherontis* J. Polhemus, 1982, *Halobates robinsoni* Andersen & Weir, 2003, and *Halobates rivularis* Andersen & Weir, 1994 (Polhemus and Cheng 1982, Andersen and Weir 1994, 2003). It was thought that *H. rivularis* represents the freshwater lineage that reverted from a mangrove-dwelling ancestor (Andersen and Weir 1994, Damgaard et al. 2000), but recent molecular evidence suggested that *H. rivularis* is nested within *Halobates* (Román-Palacios et al. 2020). Knowing where these freshwater *Halobates* species place phylogenetically in relation to the sequenced taxa would help address the evolutionary origins of the freshwater habitat and whether these freshwater reversions represent single, or multiple evolutionary events. Such extensive research will provide general insight into the evolutionary biogeography of *Halobates* which, to date, remains uncertain.

## Specimen Collection Statement

The authors attest that all legal and regulatory requirements, including export and import collection permits, have been followed for the collection of specimens from source populations at any international, national, regional, or other geographic level for all relevant field specimens collected as part of this study.

## Acknowledgments

We are grateful to the following staff and organizations for assisting with specimen loans, collection, or deposition: Wei Song Hwang, Paul Tien, and Maosheng Foo (Lee Kong Chian Natural History Museum, LKCNHM); Olena Samonina (King Abdullah University of Science and Technology); and Moira Decima and Linsey Sala (Pelagic Invertebrate Collection, Scripps Institute of Oceanography). We would like to thank Harald Bruckner and Herbert Zettel (Naturhistorisches Museum Wien) and Jonathan Ho (LKCNHM) for permission to reuse images for this publication. We also acknowledge the National Supercomputing Centre (NSCC), Singapore and National University of Singapore's High Performance Computing (HPC) for granting access to their computing resources for analysis.

## Funding

This research was funded by a Tier 1 Academic Research Fund from the Singapore Ministry of Education (A-0008517-00-00), the Marine Science R&D Programme of the National Research Foundation (MSRDP-P38 and P47), and the Ocean Park Conservation Foundation, Hong Kong (OT01.1819). The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Author Contributions

Jia Jin Marc Chang (Conceptualization [lead], Data curation [lead], Formal analysis [lead], Methodology [lead], Project administration [lead], Validation

[lead], Visualization [lead], Writing—original draft [lead]), Michael Raupach (Conceptualization [equal], Formal analysis [lead], Investigation [lead], Methodology [equal], Writing—review & editing [lead]), Lanna Cheng (Conceptualization [lead], Formal analysis [lead], Funding acquisition [supporting], Methodology [equal], Supervision [lead], Writing—review & editing [lead]), Jakob Damgaard (Formal analysis [supporting], Investigation [equal], Resources [lead], Writing—review & editing [equal]), Watcharapong Hongjamrassilp (Formal analysis [equal], Resources [lead], Writing—review & editing [equal]), Yin Cheong Aden Ip (Investigation [equal], Visualization [equal]), Matthew Hui-Chieh Ng (Investigation [equal], Visualization [equal]), Rochelle Chan (Resources [equal], Writing—review & editing [supporting]), Ismael Kunning (Resources [supporting]), Bryna Jia Ying Liang (Resources [equal], Writing—review & editing [supporting]), Davide Maggioni (Resources [equal], Writing—review & editing [equal]), Ralph Mana (Resources [equal]), Himanshu Mishra (Resources [supporting]), Maxine A.D. Mowe (Resources [supporting]), Benjamin Wainwright (Funding acquisition [lead], Resources [equal], Writing—review & editing [supporting]), Jonathan L. Whitney (Resources [supporting]), Kennedy Wolfe (Resources [supporting], Writing—review & editing [equal]), Darren Yeo (Resources [supporting], Writing—review & editing [equal]), and Danwei Huang (Formal analysis [equal], Funding acquisition [lead], Project administration [lead], Resources [lead], Supervision [lead], Writing—review & editing [lead])

## Supplementary Material

Supplementary data are available at *Insect Systematics and Diversity* online.

## Data Availability

Specimen vouchers have been deposited at either the Lee Kong Chian Natural History Museum or the Pelagic Invertebrate Collection, Scripps Institute of Oceanography. Sequence data are also publicly available on NCBI GenBank. Supplementary File S1 contains all the relevant accession numbers for vouchers and sequences. Raw sequence reads are also available on the NCBI Sequence Read Archive under BioProjects PRJNA752803 and PRJNA1026719. The datasets, codes, and output files are also available at Zenodo (https://zenodo.org/doi/10.5281/zenodo.10992549).

## References


Andersen NM. The semiaquatic bugs (Hemiptera, Gerromorpha): phylogeny, adaptations, biogeography, and classification. Denmark: Scandinavian Science Press; 1982. p. 455.

Andersen NM. Marine insects: genital morphology, phylogeny and evolution of sea skaters, genus *Halobates* (Hemiptera: Gerridae). Zool J Linn Soc. 1991:103(1):21–60. https://doi.org/10.1111/j.1096-3642.1991.tb00896.x

Andersen NM. The evolution of marine insects: phylogenetic, ecological and geographical aspects of species diversity in marine water striders. Ecography. 1999:22(1):98–111. https://doi.org/10.1111/j.1600-0587.1999.tb00458.x

Andersen NM, Cheng L. The marine insect *Halobates* (Heteroptera: Gerridae): biology, adaptations, distribution, and phylogeny. In: Gibson RN, Atkinson RJA, Gordon JDA, editors. Oceanography and marine biology: an annual review. Vol. 42. London: CRC Press; 2004. p. 119–180.

Andersen NM, Cheng L, Damgaard J, Sperling FAH. Mitochondrial DNA sequence variation and phylogeography of oceanic insects (Hemiptera: Gerridae: *Halobates* spp.). Mar Biol. 2000:136(3):421–430. https://doi.org/10.1007/s002270050701

Andersen NM, Weir TA. *Austrobates rivularis* gen. et sp. nov., a freshwater relative of *Halobates* Eschscholtz (Hemiptera: Gerridae), with a new perspective on the evolution of sea skaters. Invertebr Syst. 1994:8(1):1–15. https://doi.org/10.1071/it9940001

Andersen NM, Weir TA. A new species of sea skaters, *Halobates* Eschscholtz, from Robinson River, Western Australia (Hemiptera-Heteroptera: Gerridae). Aquat Insects. 2003:25(1):9–18. https://doi.org/10.1076/aqin.25.1.9.14020

Armisén D, Khila A. Genomics of the semi-aquatic bugs (Heteroptera; Gerromorpha): recent advances toward establishing a model lineage for the study of phenotypic evolution. Curr Opin Insect Sci. 2022:50:100870. https://doi.org/10.1016/j.cois.2021.12.010

Armisén D, Rajakumar R, Friedrich M, Benoit JB, Robertson HM, Panfilio KA, Ahn S-J, Poelchau MF, Chao H, Dinh H, et al. The genome of the water strider *Gerris buenoi* reveals expansions of gene repertoires associated with adaptations to life on the water. BMC Genomics. 2018:19(1):832. https://doi.org/10.1186/s12864-018-5163-2

Armisén D, Viala S, Cordeiro I da RS, Crumière AJJ, Hendaoui E, Le Bouquin A, Duchemin W, Santos E, Toubiana W, Vargas-Lowman A, et al. Transcriptome-based phylogeny of the semi-aquatic bugs (Hemiptera: Heteroptera: Gerromorpha) reveals patterns of lineage expansion in a series of new adaptive zones. Mol Biol Evol. 2022:39(11):msac229. https://doi.org/10.1093/molbev/msac229

Arribas P, Andújar C, Hopkins K, Shepherd M, Vogler AP. Metabarcoding and mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil. Methods Ecol Evol. 2016:7(9):1071–1081. https://doi.org/10.1111/2041-210x.12557

Bakker FT, Lei D, Yu J, Mohammadin S, Wei Z, van de Kerke S, Gravendeel B, Nieuwenhuis M, Staats M, Alquezar-Planas DE, et al. Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly pipeline. Biol J Linn Soc. 2015:117(1):33–43. https://doi.org/10.1111/bij.12642

Balian EV, Segers H, Lévèque C, Martens K. The Freshwater Animal Diversity Assessment: an overview of the results. Hydrobiologia. 2008:595(1):627–637. https://doi.org/10.1007/s10750-007-9246-3

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012:19(5):455–477. https://doi.org/10.1089/cmb.2012.0021

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF. MITOS: improved *de novo* metazoan mitochondrial genome annotation. Mol Phylogenet Evol. 2013:69(2):313–319. https://doi.org/10.1016/j.ympev.2012.08.023

Blaimer BB, Lloyd MW, Guillory WX, Brady SG. Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. PLoS One. 2016:11(8):e0161531. https://doi.org/10.1371/journal.pone.0161531

Call E, Mayer C, Twort V, Dietz L, Wahlberg N, Espeland M. Museomics: phylogenomics of the moth family Epicopeiidae (Lepidoptera) using target enrichment. Insect Syst Divers. 2021:5(2):6. https://doi.org/10.1093/isd/ixaa021

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinf. 2009:10:421. https://doi.org/10.1186/1471-2105-10-421

Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000:17(4):540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334

Chang JJM, Ip YCA, Cheng L, Kunning I, Mana RR, Wainwright BJ, Huang D. High-throughput sequencing for life-history sorting and for bridging reference sequences in marine Gerromorpha (Insecta: Heteroptera). Insect Syst Divers. 2022:6(1):1. https://doi.org/10.1093/isd/ixab024

Chen P-P, Nieser N. A taxonomic revision of the oriental water strider genus *Metrocoris* Mayr (Hemiptera, Gerridae). Part I. Steenstrupia. 1993a:19(1):1–43.

Chen P-P, Nieser N. A taxonomic revision of the oriental water strider genus *Metrocoris* Mayr (Hemiptera, Gerridae). Part II. Steenstrupia. 1993b:19(1):45–92.

Chen P-P, Zettel H. A taxonomic revision of the oriental water strider genus *Ventidius* Distant (Hemiptera, Gerromorpha, Gerridae). Tijdschr Entomol. 1998:141(1–2):137–208. https://doi.org/10.1163/22119434-99900010

Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018:34(17):i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chen X, Dong Z, Liu G, He J, Zhao R, Wang W, Peng Y, Li X. Phylogenetic analysis provides insights into the evolution of Asian fireflies and adult bioluminescence. Mol Phylogenet Evol. 2019:140:106600. https://doi.org/10.1016/j.ympev.2019.106600

Cheng L. The genus *Ventidius* Distant (Heteroptera: Gerridae) in Malaya, with a description of four new species. Proc R Entomol Soc Lond Ser B Taxon. 1965:34(11–12):153–163. https://doi.org/10.1111/j.1365-3113.1965.tb01693.x

Cheng L. Three new species of *Esakia* Lundblad (Heteroptera: Gerridae) from Malaya. Proc R Entomol Soc Lond Ser B Taxon. 1966:35(1–2):16–22.
Cheng L. Marine pleuston—animals at the sea-air interface. In: Barnes H, editor. Oceanography and marine biology: an annual review. Vol. 13. London: Allen & Unwin; 1975. p. 181–212.
Cheng L. Insects in marine environments. In: Cheng L, editor. Marine insects. Amsterdam (the Netherlands): North-Holland Publishing Company; 1976. p. 1–4.
Cheng L, Mishra H. Why did only one genus of insects, *Halobates*, take to the high seas? PLoS Biol. 2022:20(4):e3001570. https://doi.org/10.1371/journal.pbio.3001570
Crampton-Platt A, Yu DW, Zhou X, Vogler AP. Mitochondrial metagenomics: letting the genes out of the bottle. GigaScience. 2016:5(1):15. https://doi.org/10.1186/s13742-016-0120-y
Crumière AJJ, Santos ME, Sémon M, Armisén D, Moreira FFF, Khila A. Diversity in morphology and locomotory behavior is associated with niche expansion in the semi-aquatic bugs. Curr Biol. 2016:26(24):3336–3342. https://doi.org/10.1016/j.cub.2016.09.061
Cui Y, Yang H, Tian X. The complete mitochondrial genome sequence of *Aquarius elongatus* (Hemiptera: Gerridae). Mitochondrial DNA B Resour. 2021:6(3):1016–1017. https://doi.org/10.1080/23802359.2021.1895001
Damgaard J. Evolution of the semi-aquatic bugs (Hemiptera: Heteroptera: Gerromorpha) with a re-interpretation of the fossil record. Acta Entomol Mus Nat Pragae. 2008a:48(2):251–268.
Damgaard J. Phylogeny of the semiaquatic bugs (Hemiptera-Heteroptera, Gerromorpha). Insect Syst Evol. 2008b:39(4):431–460. https://doi.org/10.1163/187631208788784264
Damgaard J. What do we know about the phylogeny of the semi-aquatic bugs (Hemiptera: Heteroptera: Gerromorpha)? Entomol Am. 2013:118(1):81–98. https://doi.org/10.1664/12-ra-030.1
Damgaard J, Andersen NM, Cheng L, Sperling FAH. Phylogeny of sea skaters, *Halobates* Eschscholtz (Hemiptera, Gerridae), based on mtDNA sequence and morphology. Zool J Linn Soc. 2000:130(4):511–526. https://doi.org/10.1111/j.1096-3642.2000.tb02199.x
Damgaard J, Andersen NM, Meier R. Combining molecular and morphological analyses of water strider phylogeny (Hemiptera-Heteroptera, Gerromorpha): effects of alignment and taxon sampling. Syst Entomol. 2005:30(2):289–309. https://doi.org/10.1111/j.1365-3113.2004.00275.x
Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol Biol Evol. 2020:37(1):291–294. https://doi.org/10.1093/molbev/msz189
Deiner K, Renshaw MA, Li Y, Olds BP, Lodge DM, Pfrender ME. Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA. Methods Ecol Evol. 2017:8(12):1888–1898. https://doi.org/10.1111/2041-210x.12836
Dodsworth S. Genome skimming for next-generation biodiversity analysis. Trends Plant Sci. 2015:20(9):525–527. https://doi.org/10.1016/j.tplants.2015.06.012
Donath A, Jühling F, Al-Arab M, Bernhart SH, Reinhardt F, Stadler PF, Middendorf M, Bernt M. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucleic Acids Res. 2019:47(20):10543–10552. https://doi.org/10.1093/nar/gkz833
Duan Y, Fu S, Ye Z, Bu W. Phylogeny of Urostylididae (Heteroptera: Pentatomoidea) reveals rapid radiation and challenges traditional classification. Zool Scr. 2023:52(3):264–278. https://doi.org/10.1111/zsc.12582
Esemu SN, Dong X, Kfusi AJ, Hartley CS, Ndip RN, Ndip LM, Darby AC, Post RJ, Makepeace BL. Aquatic Hemiptera in Southwest Cameroon: biodiversity of potential reservoirs of *Mycobacterium ulcerans* and multiple *Wolbachia* sequence types revealed by metagenomics. Diversity. 2019:11(12):225. https://doi.org/10.3390/d11120225
Faircloth BC. Identifying conserved genomic elements and designing universal bait sets to enrich them. Methods Ecol Evol. 2017:8(9):1103–1112. https://doi.org/10.1111/2041-210x.12754
Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 1985:39(4):783–791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x
Gaston KJ. The magnitude of global insect species richness. Conserv Biol. 1991:5(3):283–296. https://doi.org/10.1111/j.1523-1739.1991.tb00140.x
Grandjean F, Tan MH, Gan HM, Lee YP, Kawai T, Distefano RJ, Blaha M, Roles AJ, Austin CM. Rapid recovery of nuclear and mitochondrial genes by genome skimming from Northern Hemisphere freshwater crayfish. Zool Scr. 2017:46(6):718–728. https://doi.org/10.1111/zsc.12247
Herring L. The genus *Halobates* (Hemiptera: Gerridae). Pac Insects. 1961:3(2-3):225–305.
Hotaling S, Kelley JL, Frandsen PB. Aquatic insects are dramatically underrepresented in genomic research. Insects. 2020:11(9):601. https://doi.org/10.3390/insects11090601
Hu DL, Bush JWM. The hydrodynamics of water-walking arthropods. J Fluid Mech. 2010:644:5–33. https://doi.org/10.1017/s0022112009992205
Hua J, Li M, Dong P, Cui Y, Xie Q, Bu W. Comparative and phylogenomic studies on the mitochondrial genomes of Pentatomomorpha (Insecta: Hemiptera: Heteroptera). BMC Genomics. 2008:9:610. https://doi.org/10.1186/1471-2164-9-610
Hua J, Li M, Dong P, Cui Y, Xie Q, Bu W. Phylogenetic analysis of the true water bugs (Insecta: Hemiptera: Heteroptera: Nepomorpha): evidence from mitochondrial genomes. BMC Evol Biol. 2009:9:134. https://doi.org/10.1186/1471-2148-9-134
Ikawa T, Okabe H, Cheng L. Skaters of the seas—comparative ecology of nearshore and pelagic *Halobates* species (Hemiptera: Gerridae), with special reference to Japanese species. Mar Biol Res. 2012:8(10):915–936. https://doi.org/10.1080/17451000.2012.705848
Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, Li D-Z. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020:21(1):241. https://doi.org/10.1186/s13059-020-02154-5
Johnson KP. Putting the genome in insect phylogenomics. Curr Opin Insect Sci. 2019:36:111–117. https://doi.org/10.1016/j.cois.2019.08.002
Karin BR, Arellano S, Wang L, Walzer K, Pomerantz A, Vasquez JM, Chatla K, Sudmant PH, Bach BH, Smith LL, et al. Highly-multiplexed and efficient long-amplicon PacBio and Nanopore sequencing of hundreds of full mitochondrial genomes. BMC Genomics. 2023:24:229. https://doi.org/10.1186/s12864-023-09277-6
Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013:30(4):772–780. https://doi.org/10.1093/molbev/mst010
Kieran TJ, Gordon ERL, Forthman M, Hoey-Chamberlain R, Kimball RT, Faircloth BC, Weirauch C, Glenn TC. Insight from an ultraconserved element bait set designed for hemipteran phylogenetics integrated with genomic resources. Mol Phylogenet Evol. 2019:130:297–303. https://doi.org/10.1016/j.ympev.2018.10.026
Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019:35(21):4453–4455. https://doi.org/10.1093/bioinformatics/btz305
Krehenwinkel H, Pomerantz A, Henderson JB, Kennedy SR, Lim JY, Swamy V, Shoobridge JD, Graham N, Patel NH, Gillespie RG, et al. Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. GigaScience. 2019:8(5):giz006. https://doi.org/10.1093/gigascience/giz006
Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012:9(4):357–359. https://doi.org/10.1038/nmeth.1923
Leo SST, Cheng L, Sperling FAH. Genetically separate populations of the ocean-skater *Halobates sericeus* (Heteroptera: Gerridae) have been maintained since the late Pleistocene. Biol J Linn Soc. 2012:105(4):797–805. https://doi.org/10.1111/j.1095-8312.2011.01836.x
Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021:49(W1):W293–W296. https://doi.org/10.1093/nar/gkab301
Lévêque C, Balian EV, Martens K. An assessment of animal species diversity in continental waters. Hydrobiologia. 2005:542(1):39–67. https://doi.org/10.1007/s10750-004-5522-7
Li H, Leavengood JM Jr, Chapman EG, Burkhardt D, Song F, Jiang P, Liu J, Zhou X, Cai W. Mitochondrial phylogenomics of Hemiptera reveals

adaptive innovations driving the diversification of true bugs. Proc Biol Sci. 2017:284(1862):20171223. https://doi.org/10.1098/rspb.2017.1223

Linard B, Crampton-Platt A, Gillett CPDT, Timmermans MJTN, Vogler AP. Metagenome skimming of insect specimen pools: potential for comparative genomics. Genome Biol Evol. 2015:7(6):1474–1489. https://doi.org/10.1093/gbe/evv086

Liu Y, Yang H, Tian X, Cui Y. The complete mitochondrial genome sequence of *Asclepios apicalis* (Gerromorpha: Gerridae). Mitochondrial DNA B Resour. 2021:6(4):1471–1472. https://doi.org/10.1080/23802359.2021.1909436

Mahadik GA, Hernandez-Sanchez JF, Arunachalam S, Gallo Jr. A, Cheng L, Farinha AS, Thoroddsen ST, Mishra H, Duarte CM. Superhydrophobicity and size reduction enabled *Halobates* (Insecta: Heteroptera, Gerridae) to colonize the open ocean. Sci Rep. 2020:10(1):7785. https://doi.org/10.1038/s41598-020-64563-7

Matsuda R. Morphology, evolution and a classification of the Gerridae (Hemiptera-Heteroptera). Univ Kans Sci Bull. 1960:41(2):25–632. https://doi.org/10.5962/bhl.part.15602

Nevill PG, Zhong X, Tonti-Filippini J, Byrne M, Hislop M, Thiele K, van Leeuwen S, Boykin LM, Small I. Large scale genome skimming from herbarium material for accurate plant identification and phylogenomics. Plant Methods. 2020:16:1. https://doi.org/10.1186/s13007-019-0534-5

Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2018:35(3):526–528. https://doi.org/10.1093/bioinformatics/bty633

Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How many bootstrap replicates are necessary? J Comput Biol. 2010:17(3):337–354. https://doi.org/10.1089/cmb.2009.0179

Pauli T, Meusemann K, Kukowka S, Sann M, Donath A, Mayer C, Oeyen JP, Ballesteros Y, Berg A, van den Berghe E, et al. Analysis of RNA-Seq, DNA target enrichment, and sanger nucleotide sequence data resolves deep splits in the phylogeny of cuckoo wasps (Hymenoptera: Chrysididae). Insect Syst Divers. 2021:5(1):1. https://doi.org/10.1093/isd/ixaa018

Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, FitzJohn RG, Alfaro ME, Harmon LJ. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics. 2014:30(15):2216–2218. https://doi.org/10.1093/bioinformatics/btu181

Polhemus DA, Polhemus JT. Distributional data and new synonymy for species of *Halobates* Eschscholtz (Heteroptera: Gerridae) occurring on Aldabra and nearby atolls, Western Indian Ocean. J N Y Entomol Soc. 1991:99(2):217–223.

Polhemus JT, Cheng L. Notes on marine water-striders with descriptions of new species—Part I. Gerridae (Hemiptera). Pac Insects. 1982:24(3-4):219–227.

Polhemus JT, Polhemus DA. The marine Heteroptera of far eastern New Guinea and adjacent archipelagoes (Insecta, Gerromorpha). Denisia. 2006:19:927–982.

Polhemus JT, Polhemus DA. Global diversity of true bugs (Heteroptera; Insecta) in freshwater. In: Balian EV, Lévêque C, Segers H, et al., editors. Freshwater animal diversity assessment. Dordrecht (the Netherlands): Springer Netherlands; 2008. p. 379–391.

Quek ZBR, Chang JJM, Ip YCA, Chan YKS, Huang D. Mitogenomes reveal alternative initiation codons and lineage-specific gene order conservation in echinoderms. Mol Biol Evol. 2021:38(3):981–985. https://doi.org/10.1093/molbev/msaa262

Quek ZBR, Huang D. Application of phylogenomic tools to unravel anthozoan evolution. Coral Reefs. 2022:41(3):475–495. https://doi.org/10.1007/s00338-021-02072-3

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics Using Tracer 1.7. Syst Biol. 2018:67(5):901–904. https://doi.org/10.1093/sysbio/syy032

Ramón-Laca A, Gallego R, Nichols KM. Affordable *de novo* generation of fish mitogenomes using amplification-free enrichment of mitochondrial DNA and deep sequencing of long fragments. Mol Ecol Resour. 2023:23(4):818–832. https://doi.org/10.1111/1755-0998.13758

Raupach MJ, Deister F, Villastrigo A, Balke M. The complete mitochondrial genomes of *Notiophilus quadripunctatus* Dejean, 1826 and *Omophron limbatum* (Fabricius, 1777): new insights into the mitogenome phylogeny of the Carabidae (Insecta, Coleoptera). Insect Syst Evol. 2022:53(3):242–263. https://doi.org/10.1163/1876312x-bja10027

Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol. 2012:3(2):217–223. https://doi.org/10.1111/j.2041-210x.2011.00169.x

Rokas A, Carroll SB. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. Mol Biol Evol. 2005:22(5):1337–1344. https://doi.org/10.1093/molbev/msi121

Román-Palacios C, Damgaard J, Cheng L, Moreira FFF, Ikawa T, Weir TA, Zettel H. Molecular phylogeny of sea-skaters (*Halobates* Eschscholtz, 1822), relationship between Halobatini and Metrocorini, and a catalogue of the subfamily Halobatinae (Hemiptera: Heteroptera: Gerridae). Insect Syst Evol. 2020:51(2):139–240. https://doi.org/10.1163/1876312x-00002197

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012:61(3):539–542. https://doi.org/10.1093/sysbio/sys029

Storer C, Payton A, McDaniel S, Jordal B, Hulcr J. Cryptic genetic variation in an inbreeding and cosmopolitan pest, *Xylosandrus crassiusculus*, revealed using ddRADseq. Ecol Evol. 2017:7(24):10974–10986. https://doi.org/10.1002/ece3.3625

Stork NE. How many species of insects and other terrestrial arthropods are there on Earth? Annu Rev Entomol. 2018:63:31–45. https://doi.org/10.1146/annurev-ento-020117-043348

Stork NE, McBroom J, Gely C, Hamilton AJ. New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. Proc Natl Acad Sci U S A. 2015:112(24):7519–7523. https://doi.org/10.1073/pnas.1502408112

Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. Am J Bot. 2012:99(2):349–364. https://doi.org/10.3732/ajb.1100335

Sun X, Wang Y, Chen PP, Wang H, Lu L, Ye Z, Wu Y, Li T, Bu W, Xie Q. Biased heteroplasmy within the mitogenomic sequences of *Gigantometra gigas* revealed by sanger and high-throughput methods. Zool Syst. 2018:43(4):356–386. https://doi.org/10.11865/zs.201833

Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 2007:56(4):564–577. https://doi.org/10.1080/10635150701472164

Tan MH, Gan HM, Bracken-Grissom H, Chan T-Y, Grandjean F, Austin CM. More from less: genome skimming for nuclear markers for animal phylogenomics, a case study using decapod crustaceans. J Crustac Biol. 2021:41(2):ruab009. https://doi.org/10.1093/jcbiol/ruab009

Tran AD, Le TPL, Nguyen TAN. The taxonomy of the genus *Asclepios* Distant (Heteroptera, Gerromorpha, Gerridae), with description of a new species from Vietnam. J Asia-Pac Entomol. 2023:26(4):102144. https://doi.org/10.1016/j.aspen.2023.102144

Trevisan B, Alcantara DMC, Machado DJ, Marques FPL, Lahr DJG. Genome skimming is a low-cost and robust strategy to assemble complete mitochondrial genomes from ethanol preserved specimens in biodiversity studies. PeerJ. 2019:7:e7543. https://doi.org/10.7717/peerj.7543

Wachi N, Matsubayashi KW, Maeto K. Application of next-generation sequencing to the study of non-model insects. Entomol Sci. 2018:21(1):3–11. https://doi.org/10.1111/ens.12281

Wang WY, Chang JJM, Norris R, Huang D, Cheng L. Distinct population histories among three unique species of oceanic skaters *Halobates* Eschscholtz, 1822 (Hemiptera: Heteroptera: Gerridae) in the Eastern Pacific Ocean. Mar Biol. 2021a:168:147. https://doi.org/10.1007/s00227-021-03944-6

Wang Y, Moreira FFF, Rédei D, Chen PP, Kuechler SM, Luo J, Men Y, Wu H, Xie Q. Diversification of true water bugs revealed by transcriptome-based phylogenomics. Syst Entomol. 2021b:46(2):339–356. https://doi.org/10.1111/syen.12465

Wang Y-H, Luan Y-X, Luo J-Y, Men Y, Engel MS, Damgaard J, Khila A, Chen P-P, Figueiredo Moreira FF, Rafael JA, et al. 300 Million years of coral treaders (Insecta: Heteroptera: Hermatobatidae) back to the ocean in the phylogenetic context of Arthropoda. Proc Biol Sci. 2023:290(2001):20230855. https://doi.org/10.1098/rspb.2023.0855

Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019:20(1):257. https://doi.org/10.1186/s13059-019-1891-0

Yang CM, Wong TCM, Lua HK, Koh LM. A checklist of aquatic and semi-aquatic bugs (Insecta: Hemiptera: Heteroptera) from Pulau Tioman, Peninsular Malaysia. Raffles Bull Zool Suppl. 1999:6:277–288.

Ye Z, Chen D, Yuan J, Zheng C, Yang X, Wang W, Zhang Y, Wang S, Jiang K, Bu W. Are population isolations and declines a threat to island endemic water striders? A lesson from demographic and niche modelling of *Metrocoris esakii* (Hemiptera: Gerridae). Mol Ecol. 2020:29(23):4573–4587. https://doi.org/10.1111/mec.15669

Zheng C, Ye Z, Zhu X, Zhang H, Dong X, Chen PP, Bu W. Integrative taxonomy uncovers hidden species diversity in the rheophilic genus *Potamometra* (Hemiptera: Gerridae). Zool Scr. 2020:49(2):174–186. https://doi.org/10.1111/zsc.12401