Received: 1 November 2024 | Accepted: 7 April 2025

DOI: 10.1111/syen.12685

ORIGINAL ARTICLE

Systematic Entomology

Royal Entomological Society

Whole genome shotgun phylogenomics resolve the diving beetle tree of life

Johannes Bergsten[1,2] | Johan A. A. Nylander[3] | Oscar E. Ospina[4] | Alan R. Lemmon[5] | Kelly B. Miller[6]

[1]Department of Zoology, Swedish Museum of Natural History, Stockholm, Sweden

[2]Department of Zoology, Stockholm University, Stockholm, Sweden

[3]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden

[4]Department of Biostatistics and Bioinformatics, Moffitt Cancer Center, Tampa, Florida, USA

[5]Department of Scientific Computing, Florida State University, Tallahassee, Florida, USA

[6]Department of Biology and Museum of Southwestern Biology, University of New Mexico, Albuquerque, New Mexico, USA

Correspondence
Johannes Bergsten, Department of Zoology, Swedish Museum of Natural History, Box 50007, SE-10405 Stockholm, Sweden.
Email: johannes.bergsten@nrm.se

Kelly B. Miller, Department of Biology and Museum of Southwestern Biology, University of New Mexico, Albuquerque, NM 87131, USA.
Email: kbmiller@unm.edu

Funding information
National Science Foundation

Editor: Gael J. Kergoat

Abstract

Diving beetles (Dytiscidae) are important generalist predators in freshwater ecosystems that have been around since the Jurassic. Previous phylogenetic studies have identified a largely stable set of monophyletic named groups (subfamilies, tribes and subtribes); however, backbone relationships among these have remained elusive. Here we use whole genome sequencing to reconstruct the phylogeny of Dytiscidae. We mine de novo assemblies and combine them with others available from transcriptome studies of Adephaga to compile a dataset of 149 taxa and 5364 orthologous genes. Species tree and concatenated maximum likelihood methods provide largely congruent results, resolving in agreement all but two inter-subfamily nodes. All 11 subfamilies are monophyletic, supporting previous results; possibly also all tribes, but Hydroporini is recovered as paraphyletic with weak support and monophyly of Dytiscini is method dependent. One large clade includes eight of 11 subfamilies (excluding Laccophilinae, Lancetinae and Coptotominae). Matinae is sister to Hydrodytinae + Hydroporinae, in contrast with previous studies that have hypothesized Matinae as sister to the remaining Dytiscidae. Copelatinae belong in a clade with Cybistrinae, Dytiscinae, Agabinae and Colymbetinae. Strongly confirmed sister group relationships of subfamilies include Cybistrinae + Dytiscinae, Agabinae + Colymbetinae, Lancetinae + Coptotominae and Hydrodytinae + Hydroporinae. Remaining problems include resolving with confidence the basal ingroup trichotomy and relationships between tribes in Hydroporinae. Resolution of tribes in Dytiscinae is affected by methodological inconsistencies. Platynectini, new tribe, is described and Hydrotrupini redefined within subfamily Agabinae. This study is a step forward towards completely resolving the backbone phylogeny of Dytiscidae, which we hope will stimulate further work on remaining challenges.

KEYWORDS
classification, Coleoptera, Dytiscidae, phylogenomics, phylogeny, whole genome sequencing

INTRODUCTION

Diving beetles, the Coleoptera family Dytiscidae, is the largest aquatic diversification of beetles in the world with currently over 4600 described species (Nilsson & Hájek, 2024), but well over 5000 expected (Nilsson-Ortman & Nilsson, 2010). Originating in the Jurassic, diving beetles experienced an early burst in body size evolution in the early Cretaceous (Désamoré et al., 2018) and adults range from

<1 mm to almost 5 cm (Miller & Bergsten, 2016). As a group of generalist freshwater predators, only teleost fish, dragonflies and aquatic bugs match their diversity. Across every continent, except Antarctica, and in almost every type of aquatic habitat, anyone dipping an aquatic net in freshwater is likely to encounter diving beetles. Consequently, the importance of diving beetles in freshwater ecosystems cannot be overestimated. Carl von Linné listed *Dytiscus* Linnaeus ('great diving beetles') second after *Blatta* Linnaeus (cockroach) in the Insect column of the 1st edition of *Systema Naturae* (Linnæus, 1735). Darwin's theory of sexual selection (Darwin, 1871) could have, from its beginning, diversified beyond female choice and male–male competition had he only understood the antagonistic rather than cooperative aspects of secondary sexual characters in diving beetles (Bergsten et al., 2001; Miller, 2003; Miller & Bergsten, 2014b, 2023b). As model taxa, diving beetles have figured in various disciplines such as DNA barcoding (Bergsten et al., 2012; Hendrich et al., 2010), Biogeography (Balke et al., 2009; Bukontaite et al., 2015; Morinière et al., 2014, 2016; Toussaint et al., 2013, 2017), sexual conflict (Arnqvist & Rowe, 2005; Bergsten et al., 2001; Bergsten & Miller, 2007; Miller & Bergsten, 2014b, 2023b), macroecological patterns (Abellán & Ribera, 2011; Hjalmarsson et al., 2015; Ribera, 2008; Ribera et al., 2001, 2003; Ribera & Vogler, 2000), diversification dynamics (Désamoré et al., 2018; Toussaint et al., 2014; Villastrigo et al., 2021), mosquito and Malaria control (Choo et al., 2021; Lundkvist et al., 2003; Ohba & Ushio, 2015), thermal tolerance in relation to climate change (Calosi et al., 2010; Calosi, Bilton, & Spicer, 2008; Calosi, Bilton, Spicer, & Atfield, 2008), subterranean and terrestrial adaptation and diversification (Austin et al., 2023; Leys et al., 2003; Toussaint, Hendrich, et al., 2016; Villastrigo et al., 2023), ecosystem functioning and predator–prey interactions (Culler et al., 2014; Yee, 2010), freshwater conservation strategies and indicator species (Bilton et al., 2006; Foster & Bilton, 2014; Foster & Eyre, 1992; Hjalmarsson et al., 2013; Isambert et al., 2011), to name a few. Despite their ubiquitousness and popularity as study organisms, the backbone phylogeny of diving beetles has not yet been resolved. Eleven natural clades at the subfamily rank have been identified and some subfamilies (such as Hydroporinae) have numerous natural clades at the tribe rank which are largely stable across morphological and molecular datasets (Miller & Bergsten, 2014a, 2016, 2023a), but their interrelationships have remained poorly resolved. This pattern of a natural set of well-defined clades, but uncertain inter-clade relationship, is not unusual in other taxa, and genomic data have helped to resolve many similar situations (Misof et al., 2014; Smith et al., 2011). More unusual is the situation, such as at the base of Neoaves, where despite intense phylogenomic study, some researchers now interpret the lack of resolution as a 'hard polytomy' (Suh, 2016).

Modern genomic data offer a completely new quantitative level of DNA sequence data to apply to phylogenetic problems. Among the different approaches to assemble a genomic dataset, methods of reduced-representation are the most popular and have dominated genomic studies within the beetle suborder Adephaga to which Dytiscidae belong (Baca et al., 2021; Gustafson et al., 2020; Vasilikopoulos et al., 2019; Vasilikopoulos et al., 2021). These are cost-effective in terms of sequencing, but more labour-intensive during pre-sequencing. Among reduced-representation methods, transcriptomes require high-quality tissue optimally preserved in special solutions (e.g. RNAlater) and hence a career-long collection of alcohol-preserved taxa is unfit for data acquisition. Anchored hybrid enrichment (AHE) (Lemmon et al., 2012; Lemmon & Lemmon, 2013) and Ultraconserved elements (UCE) (Faircloth et al., 2012) require probe sets and a complex, time-intensive library preparation endeavour, but, of benefit, are more straightforward to process once sequenced since target regions are limited and decided upon a priori. An alternative to reduced-representation methods is simple random fragmentation ('shotgun strategy') and NG sequencing of the whole genome, or low-coverage whole genome sequencing (WGS) (Zhang et al., 2019). The main advantages to WGS are straightforward library preparation and the richness of data that can be mined over and over again, well beyond extracting a standard orthologous gene dataset. For instance, more and more studies show that much of the genomic data discarded due to not passing the single-copy orthologous gene criteria actually carry much useful phylogenetic information (Smith & Hahn, 2021). Negative aspects to WGS include a higher sequencing cost per sample and more laborious post-sequencing bioinformatic processing of the data. As massive sequencing is getting ever cheaper and bioinformatic pipelines for WGS phylogenomics improve, the negatives are more than outweighed by the positives, at least currently for taxa not much larger than one gigabase pair (GB) in genome size.

Morphology still plays a vital role in the phylogenomic era. Thousands of genes can easily reconstruct an incorrect, yet highly supported phylogeny if, for example, model assumptions are violated. Reciprocal illumination is key (Gustafson et al., 2021). Morphology is useful to check genomic results, whereas the latter often provides data where morphology is equivocal. For example, in Dytiscidae, the large-bodied Dytiscinae and Cybistrinae have historically been regarded as closely related based on many larval and adult characters, but datasets of a handful of genes indicated alternative hypotheses that were difficult to reconcile with morphology (Miller & Bergsten, 2014a; Ribera et al., 2008). Lately, genomic studies have again re-established the sister group relationship evident originally by morphology (Gustafson et al., 2020; Vasilikopoulos et al., 2021). Also, medium-sized Agabinae and Colymbetinae, diverse in the nemoral and boreal zones, have long been associated as closely related although Colymbetinae had to go through a taxonomic refining process wherein taxa such as *Lancetes* Sharp, *Coptotomus* Say, *Matus* Aubé and *Agabetes* Crotch were excluded (see Miller, 2001; Miller & Bergsten, 2014a). Whereas not always supported as sister groups with limited datasets (e.g. Miller, 2001; Ribera, Hogan, & Vogler, 2002), and synapomorphies are not obvious, later analyses, including those with genomic data, have reconfirmed Agabinae and Colymbetinae as sister groups (Gustafson et al., 2020; Miller & Bergsten, 2014a; Ribera et al., 2008; Vasilikopoulos et al., 2021). The most recently described subfamily, Hydrodytinae, was proposed as sister to similarly small-bodied, but vastly more diverse Hydroporinae (Miller, 2001), and this hypothesis has withstood the test of genomic data (Gustafson et al., 2020, Vasilikopoulos et al., 2021). Apart from

these three phylogenetic couplets, the placement of circumtropical and megadiverse Copelatinae and Laccophilinae, the New World Coptotominae and the biogeographically disjunct (New World and Australia) Lancetinae and Matinae have remained unsettled as have the relationships between these taxa and the couplets noted above.

The few genomic insights available for Dytiscidae come from studies of Adephaga relationships using reduced-representation methods and a limited taxon sampling for Dytiscidae. For example, UCE data convincingly supported a fourth clade of Coptotominae + Lancetinae (Baca et al., 2021; Gustafson et al., 2020). This relationship was also suggested by a comprehensive analysis of larval characters (Michat et al., 2017, also see Brinck, 1948). More recently, an analysis combining transcriptomes and, with a larger taxon sampling, targeted genes through an exon-capture approach supported the four couplets described above but otherwise disagreed substantially regarding inter-subfamily relationships with the UCE-based study (Vasilikopoulos et al., 2021). Since the exon capture study included much more extensive taxon sampling for Dytiscidae than the UCE-based study, these results may be more trustworthy. But the conflict between the two studies emphasizes that depending on the types of genomic data analysed and taxon sampling, there are still unsolved issues for convincingly resolving the diving beetle tree of life.

The objective of this study is to further improve our understanding of Dytiscidae phylogeny, especially the phylogenetic backbone of the family, by increasing both gene and taxon sampling to the largest dytiscid dataset analysed to date. In particular, we test the position of four couplets of subfamilies (Figure 1) in relation to each other and in relation to Matinae, Copelatinae and Laccophilinae. These are the outstanding remaining uncertainties in Dytiscidae subfamilial relationships (Figure 1). Our sampling is sufficient to also at least partially test relationships among tribes in the largest subfamily, Hydroporinae. Finally, our whole genome sequencing approach and bioinformatic pipeline offer an alternative approach to phylogenomic studies within Adephaga, open for assessment, comparison and scrutiny.

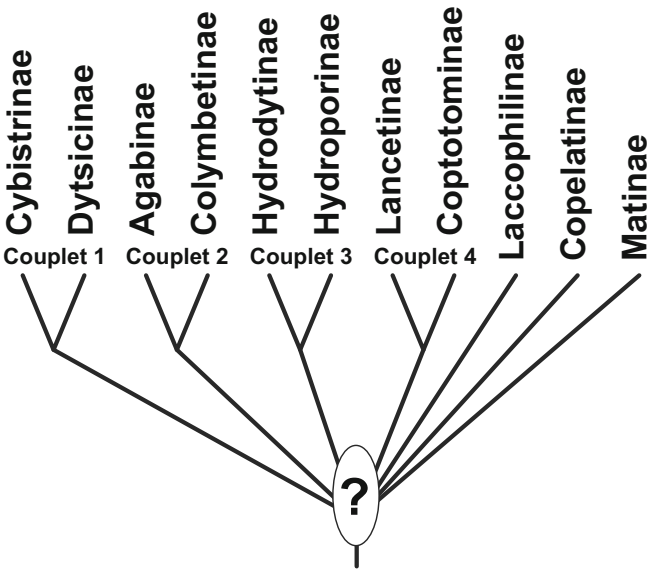

FIGURE 1 Summary consensus cladogram of Dytiscidae based on cumulative phylogenetic literature. Four sister group 'couplets' of subfamily relationships are supported in Dytiscidae. The remaining backbone resolution of the diving beetle tree of life has remained elusive.

METHODS

Taxon sampling

Details about ingroup and outgroup species and specimens are provided in Table S1. In addition to taxa newly sequenced here, we downloaded genomic data from four previous studies and extracted all Adephaga taxa from those datasets (McKenna et al., 2019; Vasilikopoulos et al., 2019, 2020, 2021). All subfamilies, tribes and subtribes of Dytiscidae are represented except the tribe Laccornellini. All subfamilies except Hydrodytinae are represented with multiple representatives. Outgroups include representatives of all Adephagan families except Meruidae, and the phylogeny is rooted between Gyrinidae and the remaining tree based on phylogenetic evidence (Beutel et al., 2020; Vasilikopoulos et al., 2021). Newly sequenced samples were selected with due diligence to applicable requirements.

Extractions

Specimen DNAs were extracted using Qiagen DNEasy or Puregene kits (Valencia, California, USA) using the animal tissue protocols. With large specimens, an incision was made in the side of the thorax and muscle tissue was removed from the body cavity with fine forceps to be extracted. The specimen was then retained for vouchering. Smaller specimens were extracted by removing the abdomen where it joins the metathorax and placing the remaining portions of the specimen (head and thorax) in buffer for extraction. Portions of the specimens remaining after extraction (including the abdomen) were retained for vouchering. All specimens had been collected and stored in ethanol prior to extraction. Vouchers and DNAs are deposited in the Museum of Southwestern Biology (K.B. Miller, curator).

Library preparations, sequencing and pre-assembly

DNA extractions of 14 samples representing all subfamilies except Lancetinae were sent to Science for Life Laboratory, Stockholm and prepared with Chromium Genome kit to generate linked reads with 10× Genomics technology. The 14-sample library was sequenced on 8 lanes of Illumina HiSeqX using a 2 × 151 bp setup and the 'HiSeq X SBS' chemistry. The demultiplexing and FastQ conversion was performed using bcl2fastq v2.19.1.403. Due to a pause in sequencing after cycle 1 the GEM barcode became truncated from 16 to 15 bp which interfered with downstream analysis in its raw form. The

manufacturer suggested prepending an ambiguous 'N' base to each 'R1' which was a workaround for the failing analysis but with a cost of slightly diminished barcode assignment rate. Appending an 'N' to each read1 where a sequencing cycle was dropped was applied using bioawk (https://github.com/lh3/bioawk) in the following manner:

bioawk -cfastx '{print "@"$name" "$comment"\nN"$seq"\n+\n#"$qual}' input_R1.fastq.gz > output_R1.fastq.

To further improve assemblies, 12 of the 14 samples were re-sequenced in a second run with identical run parameters as the previous one but on 6 lanes, and the assembly for these 12 samples is based on merging the data from runs 1 and 2.

Illumina libraries were prepared for an additional 62 samples by the FSU Center for Anchored Phylogenomics following Prum et al. (2015). In short, a Covaris ultrasonicator was used to fragment extracted DNA to a size range of 200–700 bp. Using a Beckman-Coulter Biomek FXp liquid-handling robot, we performed blunt-end repair followed by size selection to 200–400 bp using SPRI select beads (Beckman-Coulter Inc.; 0.9× ratio of bead to sample volume). Adapters containing sample-specific indexes were also ligated (for details, see Prum et al., 2015). After assessing DNA concentration using Qubit, we pooled libraries equally in groups of ~16 and verified library quality using qPCR.

Sequencing was performed at the FSU Translational Laboratory in the College of Medicine. Initial sequencing took place on an Illumina NovaSeq6000 S2 flow cell (shared with 38 other samples), with the PE150bp protocol and dual 8 bp indexing. After assessing sequencing coverage from this initial run, we re-pooled the libraries (to optimize coverage uniformity) and collected additional reads (same protocol) on a portion of an S4 flow cell. The re-pooling/re-sequencing process was repeated twice more using SP flow cells. The total sequencing effort across the 62 samples was approximately 6.6 billion read pairs.

The fastq files resulting from the four Illumina NovaSeq runs were examined using FastQC v0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to obtain a first look at the read quality. Next, we trimmed sequencing adapters from the reads using Trim Galore! v0.6.4 (https://github.com/FelixKrueger/TrimGalore/tree/master) and kept only sequences with length >100 bp. After trimming, sequences were checked for quality again using FastQC.

De novo assembly and QC

For samples prepared using 10× Genomics, draft de novo assemblies were generated using supernova v.2.1.1 with non-default parameters '--nopreflight' and '--accept-extreme-coverage' (Weisenfeld et al., 2017). Merged reads from both runs of *Coptotomus interrogatus* (Fabricius) were downsampled to 250 M reads in order not to exceed the 2 TB memory allocation for the Supernova assembler. A quantitative assessment of the assemblies was done with Quast v.4.5.4 (Gurevich et al., 2013), and completeness was assessed with BUSCO (Simão et al., 2015) using the endopterygota_odb10 dataset.

For remaining samples, processed reads from the four sequenced lanes were concatenated and used as input for Abyss v2.2 in paired-end mode. After testing several k-mer sizes, we decided to use a k-mer size of 48 based on the quality of resulting alignments and appropriate length of resulting contigs (N50). We also set Abyss to run using a Bloom filter size of 100G with three hash functions (-H argument) and a k-mer count threshold of 3 (−kc).

Gene selection and extraction

We used three previous transcriptome-based studies focusing on Coleoptera (McKenna et al., 2019), Dytiscoidea (Vasilikopoulos et al., 2019) or Neuropterida (Vasilikopoulos et al., 2020) to assemble an orthologous gene dataset. We chose this approach in order to leverage published data and maximize our taxon sampling by including relevant terminals from these studies, complementary to our newly sequenced taxa. All three used OrthoDB (v. 7 or v. 9 (Zdobnov et al., 2020)) to extract a reference set of single-copy genes orthologous at the level of Endopterygota (Holometabola). Each study used a different set of reference taxa with *Tribolium castaneum* (Herbst) as the only common denominator. Vasilikopoulos et al. (2020) used four taxa (*T. castaneum*, *Drosophila melanogaster* Linnaeus, *Bombyx mori* Linnaeus and *Acromyrmex echinatior* Forel) as did McKenna et al. (2019) (*T. castaneum*, *D. melanogaster*, *Nasonia vitripennis* (Walker) and *Danaus plexippus* Linnaeus). Vasilikopoulos et al. (2019) used six taxa (*T. castaneum*, *Anopheles gambiae* Giles, *Harpegnathos saltator* Jerdon, *N. vitripennis*, *B. mori* and *D. plexippus*). Vasilikopoulos et al. (2019), Vasilikopoulos et al. (2020) required the presence of a single copy in all reference taxa whereas Mckenna et al. (2019) included genes present in three out of four taxa, conditional on that *T. castaneum* was included. The difference in selected reference taxa, and criteria for inclusion, as well as some additional filtering of the original extraction, resulted in 4818 (McKenna et al., 2019), 3983 (Vasilikopoulos et al., 2020) and 3083 (Vasilikopoulos et al., 2019) final sets of orthologous genes, respectively. Since they are orthologous at the level of Endopterygota, the three gene datasets overlap to a large extent, but each dataset also has unique genes not present in the other two.

Each orthologous gene is identified by an OrthoDB code and we used OrthoDB V.10 and a translation table to match the codes between OrthoDB V.7 (McKenna et al., 2019) (Vasilikopoulos et al., 2020) and V.9 (Vasilikopoulos et al., 2019) (translation table file in data availability package). The matching and merging of the datasets resulted in 6413 preliminary reference genes. The exon-capture study of Adephaga (Vasilikopoulos et al., 2021) is a subset of Vasilikopoulos et al. (2019) since it targeted 651 of the 3085 genes in the latter study and the 651 genes from the Adephaga terminals were downloaded, matched and included as well.

The amino acid alignments from the published data were subsequently used as baits when extracting corresponding regions from the new genome assemblies (scaffolds files). The gene extraction was made using the ALiBaSeq workflow v1.2 (Knyshov et al., 2021). The workflow performs sequence extraction based on a local alignment search. We used tblastn v2.12.0+ (Camacho et al., 2009) with an E-value set to 1e-10, and ALiBaSeq ran with alibaseqPy3.py -x a -f M

-b blast_results -t assemblies -e 1e-10 --is --amalgamate-hits --ac tdna-tdna.

The extracted sequences from the ALiBaSeq-step were placed in gene-specific files, merged with a subset of the bait- or reference sequences. More specifically, we removed duplicate terminal species between datasets based on the completeness of gene presence. We also removed a number of superfluous outgroup terminals, especially among Caraboidea, to keep only one representative per subfamily or genus (Haliplidae). The list of taxa, including information on the data source, protocol and raw data accession codes for new samples, is provided in Table S1.

The combined nucleotide data set was translated to amino acids and aligned using the program suite MACSE v10.02 (Ranwez et al., 2011). This process included multiple sequence alignment with MAFFT v7.271 (Katoh et al., 2002), at both nucleotide and amino acid levels, and both pre- and post-alignment filtering steps using HMMCleaner v1.8.VR2 (Di Franco et al., 2019), where longer indel regions, shorter isolated codons and frameshifts are identified and masked, as well as trimming alignments at the ends. Some randomly selected multiple sequence alignments of loci with data from both OrthoDB v. 7 and v. 9 coded datasets, and from our new assemblies, were examined by eye. The MACSE workflow omm_macse was run with the default settings using the Singularity image file omm_macse_v10.02.sif.

Phylogenetic analyses

The gene files were aligned with MAFFT v7.453 (option --auto). The multiple sequence alignment was then filtered using BMGE v1.12 (Criscuolo & Gribaldo, 2010) with default settings (removing sites with an entropy score below 0.5 (−h 0.5) and a gap proportion below 0.2 (−g 0.2), and only if these form a block of at least 5 sites with these properties (−b 5)). Maximum likelihood phylogenies were then estimated using RAxML-NG v1.1.0 (Kozlov et al., 2019) with a fixed substitution model (LG + G8 + F) (Le & Gascuel, 2008; Yang, 1994). These trees were used together with the multiple sequence alignment as input to TreeShrink v1.3.9 (Mai & Mirarab, 2018) (with default settings), which can filter sequences based on whether a terminal appears as an outlier in a tree as determined by its branch length. The resulting, filtered alignment was then re-aligned with MAFFT and subjected to a new tree inference with RAxML-NG, this time with automatic selection of the substitution model using ModelTest-NG v.0.2.0 (Darriba et al., 2019). The final set of gene trees was used as input to ASTRAL-III v5.6.3 (Zhang et al., 2018), a quartet-based summary method that calculates the optimal species tree from input gene trees under the multispecies coalescent model (MSC, Rannala & Yang, 2003, Rannala et al., 2020). Efficient use of computer resources was facilitated by the use of GNU parallel (Tange, 2011) and ParGenes (Morel et al., 2018). The complete align-to-species tree workflow was implemented in ATPW v0.8.0, available at https://github.com/nylander/Align-and-trees-parallel-workflow.

In addition to the species tree inference, where gene trees are inferred individually and then combined, we also performed a concatenated maximum likelihood (CML) analysis based on combined gene alignments. 5364 alignments containing amino acid sequences were concatenated and the resulting file (149 sequences of length 825,452 positions) was analysed using IQ-TREE v2.1.2 (Nguyen et al., 2015). Specifically, we used the options to perform standard model selection (Kalyaanamoorthy et al., 2017), limiting the set of models in IQ-TREE to WAG, LG, JTT, mtART, mtInv and automatic partitioning (using --rclusterf 10), followed by tree inference and ultrafast bootstrapping (Hoang et al., 2017).

To address potential bias caused by heterogeneity in the amino acid substitution process between sites, we combined a protein matrix with a profile mixture model (Le et al., 2008). More specifically, we applied a fast approximation known as the posterior mean site frequency (PMSF) method (Wang et al., 2017). In practice, PMSF is applied as a two-step procedure using the software IQ-TREE (Minh, Hahn, & Lanfear, 2020). In the first step, a mixture model is fitted to a preliminary tree. We used the result of our first CML analysis as input. In this first step, we used the empirical amino acid exchange rate matrix (Minh et al., 2021) estimated for insects (Misof et al., 2014), which was implemented in recent versions (>v2.1.3) of IQ-TREE. Preliminary fitting of substitution matrices to individual gene data sets shows that this insect-specific matrix (named Q.insect in IQ-TREE) had a best fit in the vast majority of data partitions (analyses not shown). The Q.insect matrix was combined with the options F (empirical AA frequencies from the data) and G (Yang, 1994) to model rate heterogeneity across sites. For the choice of profile mixture model, we had to resort to the C10 profile mixture model (Le et al., 2008) due to limitations in available hardware (RAM requirement for the C10 mixture model was 0.65 TB, compared to, e.g., C20 which required 1.25 TB). In the second step of the PMSF method, the estimated site frequencies for the model are applied in a more thorough tree search (again using Q.insect, F and G), and we also added a round of non-parametric bootstrapping (Hoang et al., 2017) to estimate node support. The commands for the steps above are:

1. CML: infer tree using automatic model- and partitions-selection

   iqtree2 -s data.faa --seqtype AA -p data.partitions -T AUTO -m TESTMERGE -mset WAG, LG, JTT, mtART, mtInv -rclusterf 10 -B 1000 --prefix mtest.

2. First PMSF step: estimate the conditional amino acid frequency profile

   iqtree2 -s data.faa --seqtype AA -T AUTO -m Q.insect+C10+F +G -n 0 --tree-freq mtest.contree --prefix PMSF.

3. Second PMSF step: apply the sitefreq file to a ML search with ultrafast bootstrapping

   iqtree2 -s data.faa --seqtype AA -T AUTO -m Q.insect+C10+F +G -fs PMSF.sitefreq --ufboot 1000.

Gene and site concordance factor (gCF, sCF)

It has been shown that gene and site concordance factors (gCF, sCF) are complementary measures of clade support to the bootstrap; gCF and sCF measure the underlying variance at gene tree or site level in support of a focal branch, whereas bootstrap values measure the sampling variance (Minh, Schmidt, et al., 2020). Bootstrap values increase with the addition of loci while the gCF and sCF are unaffected apart from larger estimation error when datasets are small (Minh, Hahn, & Lanfear, 2020). gCF should be interpreted as the proportion of input gene trees in a dataset that contain a focal node out of those that could (*decisive* gene trees for that node sensu Minh, Schmidt, et al., 2020). Similarly, the site concordance factor is the average proportion of sites that contain a focal node out of those that could, but sCF is quartet-based and uses parsimony criteria (Minh, Hahn, & Lanfear, 2020). We calculated gCF values in IQ-TREE (option --gcf) and site concordance factors using IQ-TREE option --scf 100. Site concordance factor was only monitored and reported for a specific clade where gene concordance factor showed signs of bias (see Section 4).

Likelihood mapping

Except for Hydroporinae, which will need a denser taxon sampling in the future, we identified four conflicting nodes between analysis types in the backbone phylogeny of Dytiscidae. All four conflicts revolve around the resolution of trichotomous nodes. To evaluate the relative support for each of the three possible resolutions, or more specifically between the ASTRAL and CML resolution, we used four-cluster likelihood mapping (FcLM) (Strimmer & von Haeseler, 1997) as implemented in IQ-TREE (Minh, Schmidt, et al., 2020). For each problem, we defined four clusters, assumed monophyletic, as follows: (1) 'Early Dytiscidae trichotomy problem': (i) Laccophilinae, (ii) Lancetinae + Coptotominae, (iii) Remaining Dytiscidae, (iv) Outgroups; (2) 'Copelatinae problem': (i) Copelatinae, (ii) Agabinae + Colymbetinae, (iii) Cybistrinae + Dytiscinae, (iv) remaining Dytiscidae + outgroups; (3) '*Hyderodes* Hope problem': (i) *Hyderodes*, (ii) *Dytiscus*, (iii) remaining Dytiscinae, (iv) remaining Dytiscidae + outgroups; (4) '*Notaticus* Zimmermann problem': (i) *Notaticus*, (ii) *Hydaticus* Leach, (iii) Aciliini + Eretini, (iv) remaining Dytiscidae + outgroups. We based our analyses on the concatenated amino acid data applying a single substitution model ('Q.insect+I+G4' (Minh, Hahn, & Lanfear, 2020, Misof et al., 2014)). Likelihood mapping in IQ-TREE was performed with '-lmap 10,000', sampling 10,000 quartets.

Assessing the impact of missing data

As our genomic dataset includes loci and terminals from multiple origins and sequencing methods (transcriptomes, genomes, exon-capture), creating a non-random pattern of missing data cells, we performed a series of analyses to evaluate the effect of missing data. These analyses follow three strategies: (i) gene occupancy filtering and exclusion, (ii) gene block exclusion based on legacy dataset overlap and idiosyncrasy and (iii) taxon block exclusion based on inferior locus occupancy from a single legacy dataset. All analyses were performed as species tree inferences with ASTRAL using the same settings as above. For the first two strategies, gene trees as inferred in the main analysis were reused, whereas the third strategy required re-inferring the gene trees using fewer taxa. In addition, the gene block and taxon block exclusions were also analysed with concatenated maximum likelihood analyses using the Q.insect+I+G4 model in IQ-TREE (Minh, Schmidt, et al., 2020; Misof et al., 2014). We evaluated the results by monitoring the trend of clade support using both local posterior probability and gene concordance factor for a set of reference clades (subfamilies and the four subfamily couplets) and a set of target clades of interest with weak or conflicting results in the main analyses. We also performed four-cluster likelihood mapping analyses for the four conflicting nodes and monitored changes in the relative support of alternative resolutions. For computational reasons, the concatenated maximum likelihood analyses were only monitored by comparing clade topologies (i.e. no resampling schemes were applied).

Gene occupancy filtering and exclusion

We created five data subsets with decreasing numbers of loci and missing data but increasing occupancy by assessing per gene taxon loci occupancy and extracting the top 100, 90, 75, 50, 25 and 10% of genes when ranked in order of occupancy (datasets referred to as D100, D90 etc). This resulted in datasets with 5364 (full, D100), 4826 (D90), 4022 (D75), 2681 (D50), 1340 (D25) and 536 (D10) loci included. We performed a similar filtering and exclusion based on occupancy on an amino acid site level instead of loci level, but the results were very similar, and we only report the filtering on loci level.

Gene block exclusion

Our full dataset consists of taxa from six different sources: four previous publications using either transcriptomics (McKenna et al., 2019; Vasilikopoulos et al., 2019; Vasilikopoulos et al., 2020) or targeted exon capture (Vasilikopoulos et al., 2021) and two different approaches for our newly sequenced taxa (10× genomics and standard whole genome shotgun sequencing). These methods vary in performance and targets, which unites different taxon blocks in the dataset. The exon-capture study by Vasilikopoulos et al. (2021) stands out in terms of missing data, as it targeted a mere 11% of the genes in our full dataset. In addition, our selection of loci was done by matching and merging the transcriptomic datasets of (McKenna et al., 2019; Vasilikopoulos et al., 2019; Vasilikopoulos et al., 2020). This also creates a pattern of non-random taxon-by-locus missing gene blocks, which may influence phylogenetic reconstruction (Xi et al., 2015). To assess the impact of these non-random patterns of missing data, we created three data

subsets based on overlap between gene sources. First, we removed all genes unique to a single source dataset, which yielded a subset of 3202 loci. Second, we extracted the genes in common for all three transcriptomic studies, which yielded a dataset of 1529 loci. Finally, we extracted the genes in common between all four source studies, including the exon-capture study, which yielded a very small dataset of 369 loci. We refer to these datasets as All (5364 loci), nounique (3202 loci), common3 (1529 loci) and common4 (369 loci). Average locus occupancy increases across these datasets as follows: 35% (All), 40% (nounique) 44% (common3) and 55% (common4) (see Table S2).

Taxon block exclusion

Terminals included from the exon-capture study by Vasilikopoulos et al. (2021) have a significantly higher proportion of missing data (average occupancy 8%) compared to remaining terminals from the transcriptomic studies and our newly genome-sequenced taxa (average occupancy 51%). To assess if this can have an undesirable effect on phylogenetic reconstruction and clade support, we excluded all terminals from that study and reconstructed a reduced phylogeny for the three datasets All, nounique and common3. Average locus occupancies from these datasets were 51% (All), 55% (nounique) and 59% (common3). Resulting trees are not fully comparable to phylogenies from analyses with all taxa included, especially within Hydroporinae where Hydrovatini, Pachydrini and Vatellini are not represented. We can evaluate however if any of the backbone resolutions between remaining subfamilies and tribes are altered although it may prove difficult to assign any alteration to the effect of fewer taxa versus higher occupancy within Hydroporinae.

Filtering genes based on phylogenetic informativeness

Some studies have shown that species tree inference may be negatively influenced by the inclusion of poorly resolved gene trees (Hosner et al., 2015). To assess the impact of gene filtering based on phylogenetic informativeness, we mirrored the gene occupancy filtering and exclusion strategy and used the measure *fraction of supported quartets* recovered from the FcLM analysis and extracted the top 100%, 90%, 75%, 50%, 25% and 10% of genes when ranked in order of this measure of phylogenetic informativeness. The resulting number of loci in each subset is very similar to the occupancy ranking (+/− 20 loci due to ties). Datasets are similarly referred to as D100, D90, etc. although the gene loci composition of each is different. These analyses were only performed using ASTRAL and reusing the same gene trees as inferred in the main analysis.

Reference clade-based gene filtering

One proposed method where morphology can aid phylogenomics is through reference clade-based gene filtering, or 'node-control strategy' (Chen et al., 2015). The monophyly of Dytiscidae is unquestioned based on ample morphological evidence as well as phylogenomic results (Baca et al., 2021; Gustafson et al., 2020; Michat et al., 2017; Miller, 2001; Miller & Bergsten, 2014a, 2016; Ruhnau, 1986; Vasilikopoulos et al., 2019; Vasilikopoulos et al., 2021). To resolve a basal trichotomy in Dytiscidae we filtered genes based on presence of the Dytiscidae node in gene trees under the premise that the genes carrying information of the nearby 'correct' branch may also provide better than average information for a neighbouring branch. In contrast genes that do not recover the Dytiscidae node are regarded as likely subject to random or systematic error (Chen et al., 2015). Specifically, we excluded gene trees from the final set of gene trees that did not recover a monophyletic Dytiscidae from the input to ASTRAL to test if the basal trichotomy is resolved with stronger support. Whereas a higher clade support value is not by itself related to accuracy, the logic behind reference clade-based subsampling is attractive (Chen et al., 2015).

Long-branch extraction

In order to assess the potential sensitivity to long-branch artefacts (Bergsten, 2005), we repeated the phylogenetic inference based on the concatenated amino acid data after exclusion of certain taxa (Siddall & Whiting, 1999). Specifically, the species *Hyderodes shuckardi* Hope and *Notaticus fasciatus* Zimmermann were excluded (one at the time, while including the other). We also tested whether the two *Dytiscus* terminals with nearly an order of magnitude difference in occupancy (452 loci for *D. marginalis* vs. 3253 loci for *D. dauricus*) could influence the positions of *Hyderodes* and *Notaticus* by excluding either or both. These repeated analyses were performed with FastTreeMP v2.1.11 (Price et al., 2010), while applying the LG model (Le & Gascuel, 2008) combined with the CAT method (20 rate categories) for accommodating rate variation over sites (Lartillot & Philippe, 2004). The resulting tree topologies were then manually compared and evaluated for changes in crucial regions.

Rogue taxon exclusion

In our gene filtering and missing data analyses, we discovered that the single *Pachydrus* terminal representing Pachydrini could be characterized as a rogue taxon as it jumped around in species trees from some of the smaller datasets (D25, D10, common 3, common4) between very distant positions within Hydroporinae. Gene occupancy data reveals this taxon has the highest amount of missing data among all ingroup terminals, second in the dataset only to the outgroup terminal *Andogyrus* sp. (Table S2). We suspected this might pull down clade support to some of the intertribal backbone nodes within Hydroporinae. We therefore excluded *Pachydrus* from a last set of ASTRAL and IQ-Tree analyses as above using all, nounique and common3 datasets and monitored its effect on clade support and resolution within Hydroporinae.

RESULTS

The 14 samples prepared with Chromium Genome kit (10× Genomics) provided an average of 524 M reads (262–881) and an assembly length of 0.75 Gbp (0.29–1.56) (Quast statistics). The Supernova assemblies had on average an N50 Scaffold size of 17 Mbp (12–36), an effective coverage of 33× (10–55) and retrieved 83% (48–97) of BUSCO reference genes (complete and fragmented). The 62 shotgun-sequenced samples resulted in an average of 110 M reads (35–379), an estimated coverage of 19× (13–76) and draft assemblies had an N50 value of 1.3Kbp (0.7–6.4). Assembly completeness as assessed with BUSCO reference genes was as expected lower compared with the 10× Genomics sequenced taxa and averaged 56% (24–96). Per taxon BUSCO statistics are available in Table S2.

The processing pipeline with quality filtering steps resulted in a final dataset with 149 taxa (106 ingroup and 43 outgroup taxa) and 825,452 aa positions from 5364 gene loci. 76 mainly ingroup taxa were newly sequenced and 73 taxa were reused including the majority of outgroups. Gene locus presence per taxon averaged 1898 (202–3983) and taxon completeness per gene locus 53 (4–147), equalling 35%–36% completeness on both gene locus and taxon level. On amino acid level, missing data per taxon averaged 0.63 (0.24–0.98). Per taxon statistics of locus occupancy can be found in Table S2.

Species tree

We recovered a well-resolved species tree with outgroup families each well-supported as monophyletic and resolved as expected (Figure 2). Rooted using Gyrinidae, Caraboidea is resolved as sister group to Haliploidea + Dytiscoidea. In Caraboidea, Trachypachidae is sister to Cicindelidae + Carabidae. In Haliploidea, *Peltodytes* Régimbart is sister to *Brychius* Thomson + *Haliplus* Latreille + *Algophilus* Zimmermann. In Dytiscoidea, Noteridae is sister to Dytiscidae and the smaller families, Paelobiidae, Aspidytidae and Amphizoidae. The much-investigated resolution between these four families (Alarie, Short, et al., 2011; Balke et al., 2005, 2008; Gustafson et al., 2020, 2021; Hawlitschek et al., 2012; Ribera, Beutel, et al., 2002; Toussaint, Beutel, et al., 2016; Vasilikopoulos et al., 2019; Vasilikopoulos et al., 2021) is resolved as (Dytiscidae (Paelobiidae (Aspidytidae + Amphizoidae))), although support for monophyly of the three smaller families is less than maximal (0.95; Figure 2). In Noteridae, Notomicrinae is sister to Noterinae and the previously challenged monophyly of Aspidytidae is here firmly supported (Figure 2).

In the ingroup, over 80% of the nodes are maximally supported (Figure 2). This includes support for each subfamily, each of the four subfamily couplets (Figure 1), and all inter-subfamily backbone branches except one. Dytiscidae is monophyletic, as expected, but a second well-supported backbone node (Figure 2, Clade A) includes eight of the 11 subfamilies, excluding Laccophilinae and Couplet 4 (Lancetinae + Coptotominae) (Figure 2). Clade A, Laccophilinae and Couplet 4 form a basal trichotomy. The resolution is a very short and weakly supported (0.54) branch that places Laccophilinae as sister to Couplet 4 + Clade A. Within clade A, Matinae is sister group to Couplet 3 (Hydrodytinae + Hydroporinae) (Figure 2, Clade B), and Copelatinae is sister group to Couplet 1 (Dytiscinae + Cybistrinae) + Couplet 2 (Agabinae + Colymbetinae) (Figure 2, Clade C). Also, the monophyletic sister group relationship between Couplets 1 and 2 (Figure 2) is maximally supported. Within Hydroporinae, a large clade (Figure 2, Clade D) includes five large tribes: Bidessini, Vatellini, Hydroporini, Hygrotini and Hyphydrini, and excludes Methlini, Hydrovatini, Pachydrini and Laccornini. The resolution between these four 'basal' tribes and with Clade D is uncertain since support is <1. All Hydroporinae except the deviant Methlini, however, form a clade with 0.80 in support. Within Clade D, Hydroporini + Hygrotini + Hyphydrini (Figure 2, Clade E) is maximally supported, whereas the position of Bidessini and Vatellini is unresolved. All subfamilies, tribes and subtribes sampled with multiple individuals are maximally supported as monophyletic except Hydroporini. In Hydroporini, the Australian subtribe Sternopriscina is in an unresolved polytomy together with Hygrotini, Hyphydrini and a clade with the remaining subtribes of Hydroporini (Hydroporina + Deronectina + Siettitiina). This last group of three subtribes of Hydroporini is maximally supported as monophyletic (Figure 2).

Maximum likelihood tree

The concatenated maximum likelihood (CML) analysis resulted in a similarly well-resolved tree with maximal support for the majority of backbone relationships (Figure S1). The configuration is largely in agreement with the ASTRAL tree, but a few resolutions differ. Outgroup families are resolved as in the ASTRAL analysis except Aspidytidae and Amphizoidae are not together monophyletic. Instead, the resolution of the problematic three smaller families and Dytiscidae is resolved as (Paelobiidae (Aspidytidae (Amphizoidae + Dytiscidae))). Again, one grouping (Aspidytidae + Dytiscidae + Amphizoidae) has less than maximal support (0.70), highlighting this notoriously difficult phylogenetic problem.

In the ingroup, the monophyly of all subfamilies, of the four couplets (Figure 1) and all inter-subfamily backbone branches are maximally supported. This includes the large Clade A, excluding Laccophilinae and Couplet 4, and Matinae as sister group to Hydrodytinae + Hydroporinae in agreement with the ASTRAL analysis. However, the basal resolution, unresolved in the ASTRAL analysis, is here maximally supported as Couplet 4 + (Laccophilinae + Clade A). The second point of disagreement is more noteworthy as it entails conflicting positions of Copelatinae in relation to Couplets 1 and 2, despite being maximally supported in each analysis. The CML analysis supports Copelatinae as sister to Couplet 1 (Cybistrinae + Dytiscinae) whereas the ASTRAL analysis placed Copelatinae as sister to Couplet 1 + Couplet 2 (Agabinae + Colymbetinae).

Also within two subfamilies the relationships among tribes differ from the ASTRAL analysis. In Dytiscinae, Dytiscini is paraphyletic with *Hyderodes* recovered as sister to a clade containing Aciliini, Hydaticini, Aubehydrini and Eretini. Within the latter clade Aubehydrini

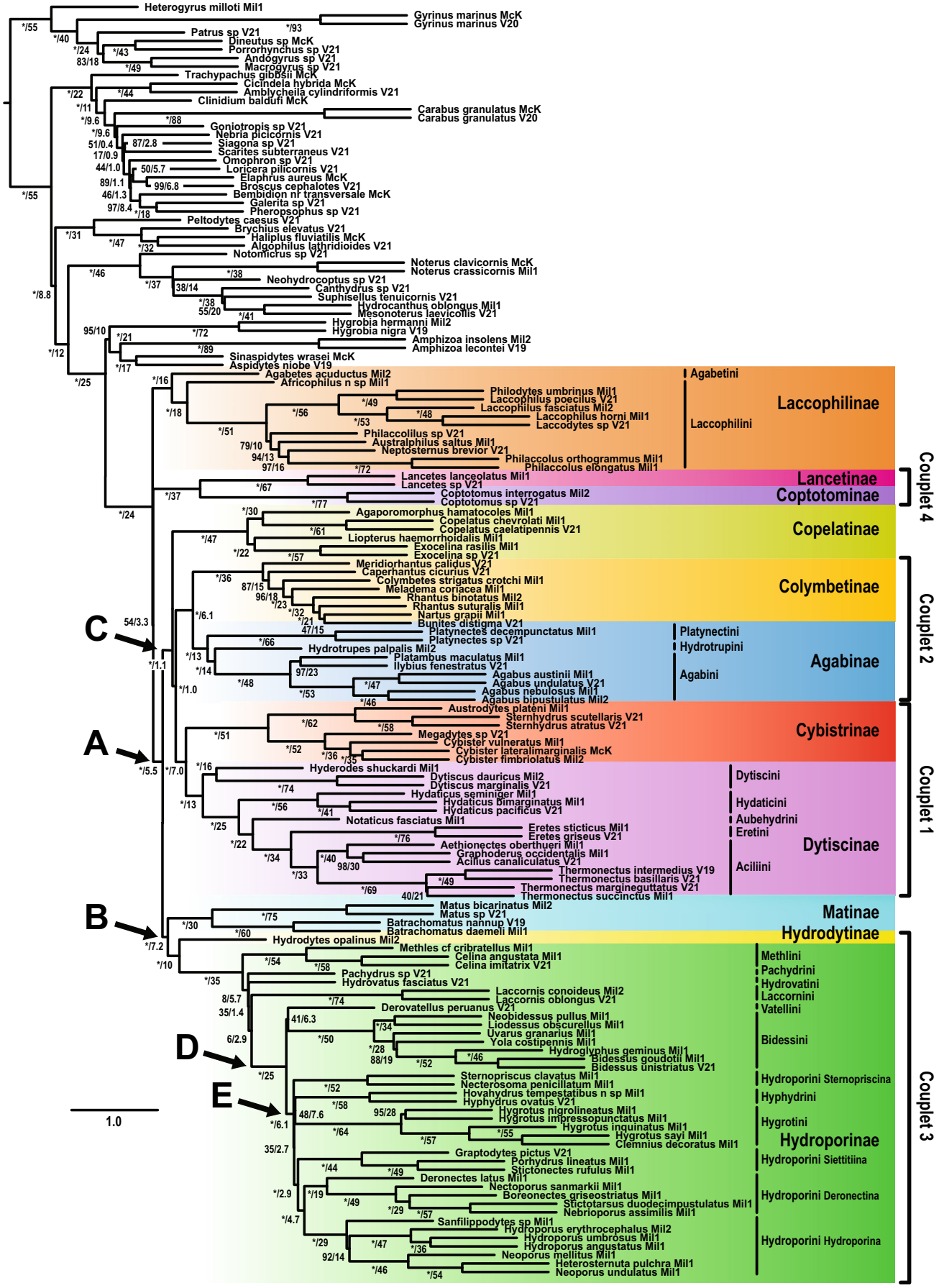

FIGURE 2 Legend on next page.

(*Notaticus*) and Hydaticini have swapped positions. Dytiscini was maximally supported as monophyletic in the ASTRAL analysis as was the alternative position of Aubehydrini (*Notaticus*) and Hydaticini (Figure 2). *Notaticus* and *Hyderodes* are the longest single terminal branches within Dytiscinae, and, therefore, we performed long-branch extraction (Siddall & Whiting, 1999) to evaluate if the presence or absence of one affects the position of the other due to long-branch attraction. This is not the case, however, and their relative positions persist even when excluding each other (Figures S2 and S3). Including either or neither of the two *Dytiscus* terminals likewise had no effect (not shown).

The second subfamily with some contrasting tribal resolutions compared with the ASTRAL analysis is Hydroporinae (Figure 2; Figure S1). The CML analysis similarly recovered Clades D and E and also resulted in maximum support for Vatellini + Bidessini. The clade with all Hydroporinae excluding Methlini has increased support (ufboot 96). The CML analysis also recovered Hydrovatini + Pachydrini with near maximum support (99), a relationship not recovered by the ASTRAL analysis. A monophyletic Hydroporinae without both Methlini and Laccornini (96) was also different from the ASTRAL analysis. Finally, within Clade E there are strongly supported resolutions that make Hydroporini paraphyletic, with Hyphydrini grouped with the Hydroporini subtribe Siettitiina and Hygrotini grouped with the Hydroporini subtribe Sternopriscina. The ASTRAL analysis included three subtribes of Hydroporini in a clade with strong support (Deronectina + Hydroporina + Siettitiina).

Gene concordance factor

Comparisons of the absolute values of gene concordance factors (gCF) between clades and between alternative resolutions reveal distinctions in gene tree support between many of the clades indistinguishable by local posterior probability or ultrafast bootstrap since they have 'maxed out' (Figure 2 and Figure S1). For instance, gCF for subfamilies ranges from 12.6 (Agabinae) to 76.62 (Coptotominae) and for the subfamily couplets from 6.08 (Agabinae + Colymbetinae) to 37.2 (Lancetinae + Coptotominae). Clades A, B, C and E all have gCF below 10 (clade C lowest at 1.1) while clade D, Dytiscini and Notaticus + Aciliini + Eretini have comparatively high gCF at 24.9, 15.9 and 21.9 respectively. The lowest gCF of all inter-subfamily backbone nodes is found in clade C excluding Copelatinae at 0.91. gCF ranges from 0 to 100 and measures the percentage of gene trees *decisive* for the node that also contains the node. Less than 1% of the input gene trees that could have contained this node do in this case. This is in fact significantly lower than the gCF (2.69) for the alternative resolution (Copelatinae + Cybistrinae + Dytiscinae) found in the CML tree despite maximal local posterior probability support for the ASTRAL resolution. gCF for Dytiscini (15.9) is somewhat higher than the paraphyletic resolution (14.0) found in the CML analysis, while the ASTRAL resolution of *Notaticus* with Aciliini + Eretini has more than twice as high gCF (21.9) compared to the CML resolution (9.9). gCF is slightly higher for the CML resolution (4.27) over the ASTRAL resolution (3.31) for the basal ingroup trichotomy. Other well-supported clades in agreement between the ASTRAL and CML analyses have gCF values that strongly support the existing over alternative resolutions except for clade C. While clade C is maximally supported in both the ASTRAL and CML analyses, gCF is very low (1.1 and 1.3, respectively) and alternative resolutions have slightly higher gCF (1.3 and 1.5, respectively).

Likelihood mapping

All four trichotomy problems where we assessed relative support for each alternative resolution using FcLM (Figure 3) resulted in support for the ASTRAL resolution (Figure 2) over the CML resolution (Figure S1). FcLM yielded the strongest support (51%) for Laccophilinae as sister to the remaining Dytiscidae in the early Dytiscidae trichotomy problem (Figure 3a). Of the alternative resolutions, Lancetinae + Coptotominae as sister to remaining Dytiscidae (CML, Figure S1) scored 26.4%, and a sister group relationship between Laccophilinae and Lancetinae + Coptotominae received 19.7% in support (Figure 3a). Copelatinae is resolved outside of a clade consisting of Agabinae + Colymbetinae + Cybistrinae + Dytiscinae with 47.3% in support (Figure 3). The two potential resolutions (Copelatinae sister to Agabinae + Colymbetinae or sister to Cybistrinae + Dytiscinae) had nearly equal and lower support but slightly favoured the CML resolution (with Cybistrinae + Dytiscinae) (Figure 3b). The strongest differentiation of alternative resolutions was found for the *Hyderodes* problem where a monophyletic *Dytiscus* + *Hyderodes* received 76.7% in support, the CML resolution (*Hyderodes* + Dytiscinae except *Dytiscus*) 21.9% and negligible support (0.5%) for the third alternative (*Hyderodes* + other Dytiscidae and outgroups) (Figure 3c). Finally, FcLM favoured the sister group relationship of *Notaticus* and Aciliini + Eretini (49.6%) (ASTRAL, Figure 2), over alternative resolutions in the *Notaticus* problem, again with the CML resolution (Figure S1) second at 27.2% and third resolution with 21% (Figure 3d).

Assessing the impact of missing data

Filtering genes based on loci occupancy revealed that almost all our ASTRAL results are stable and supported across all datasets from All (D100) to only using the top 10% (D10) of loci in terms of taxon occupancy. Support for all our reference clades (subfamilies and four

FIGURE 2 Species tree recovered from ASTRAL analysis based on 5364 gene trees as input. Values at nodes represent local posterior probability support (* if 1.0)/gene concordance factor (gCF), and branch lengths (scale bar) are given in coalescent units (Sayyari & Mirarab, 2016). Capital letters (A–E) indicate clades referenced in the text.

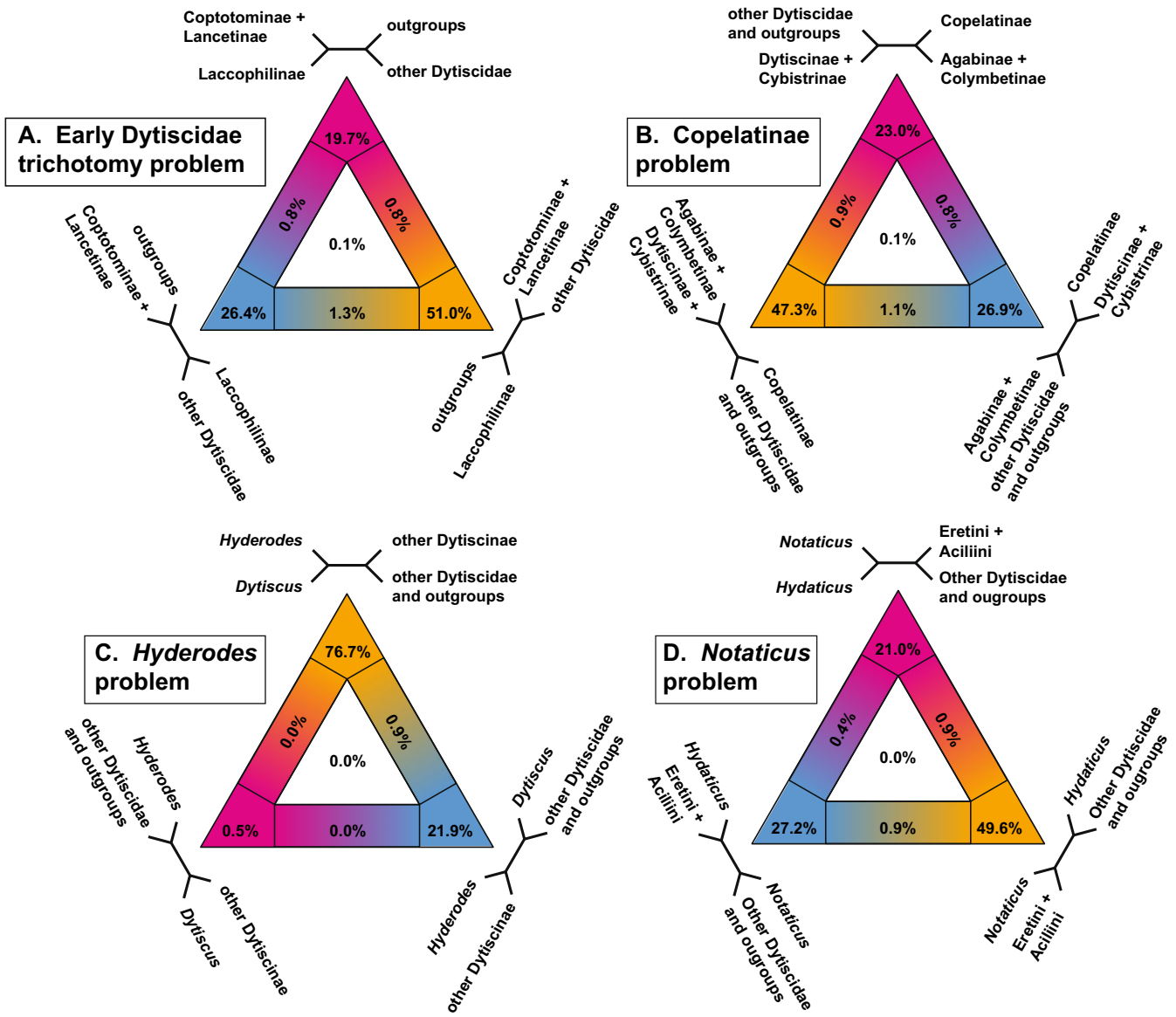

FIGURE 3 Four-cluster likelihood mapping support (Strimmer & von Haeseler, 1997) for alternative resolutions of four nodes where species tree (ASTRAL, Figure 2) and concatenated maximum likelihood analyses (CML, Figure S1) conflict. Subpanels are four phylogenetic 'problems' (Figure 5) related to: (a) basal ingroup trichotomy, (b) position of Copelatinae, (c) position of *Hyderodes*, (d) position of *Notaticus*.

subfamily couplets) remained at 1.0 in local posterior probability. Gene concordance factors (gCF) remained largely stable across D100, D90 and D75, peaked most commonly at D25 (few times at D50 or D10) and were often lowest at D10 (Figure 4a,b). gCF peaked at D25 for all four subfamily couplet clades (Figure 4b). Clades A, B, C, D and E were recovered with maximal (or in two cases at D25 or D10 near maximal) support in all analyses as well except in the most reduced dataset D10, which did not recover clades B and C. This was caused by Matinae moving to a position within clade C as sister to Colymbetinae + Agabinae. gCF peaked for all five clades at D25 as in the reference clades but clade C deviated with a distinctly lower gCF at D90, D75 and D50 compared with at D100 (Figure 4c). Support for a monophyletic Dytiscini and for a clade with *Notaticus* + Aciliini + Eretini was likewise recovered with maximal support across all datasets and with a gCF that peaked at D25 (Figure 4d). The weakly supported clade of all Dytiscidae except Laccophilinae was recovered in all analyses but with support remaining very low (0.27–0.56) and with gCF peaking at D50 instead of at D25 (Figure 4d). Clade C except Copelatinae gradually decreased in support from 1, 1, 0.99, 0.97, 0.82 and NA (not recovered) while gCF, in contrast to all other clades, peaked at the full dataset D100 and then stayed lower for all reduced datasets (Figure 4d). At D10, Copelatinae was recovered as sister to Cybistrinae + Dytiscinae as in the full CML analysis, albeit with low support (0.64). Relative support, as assessed by FcLM, remained largely constant in each of the four conflicting nodes when genes were filtered based on locus occupancy (Figures S4–S7). The ASTRAL resolution remained the most supported across D100–D10 and at largely the same level as originally across all four nodes.

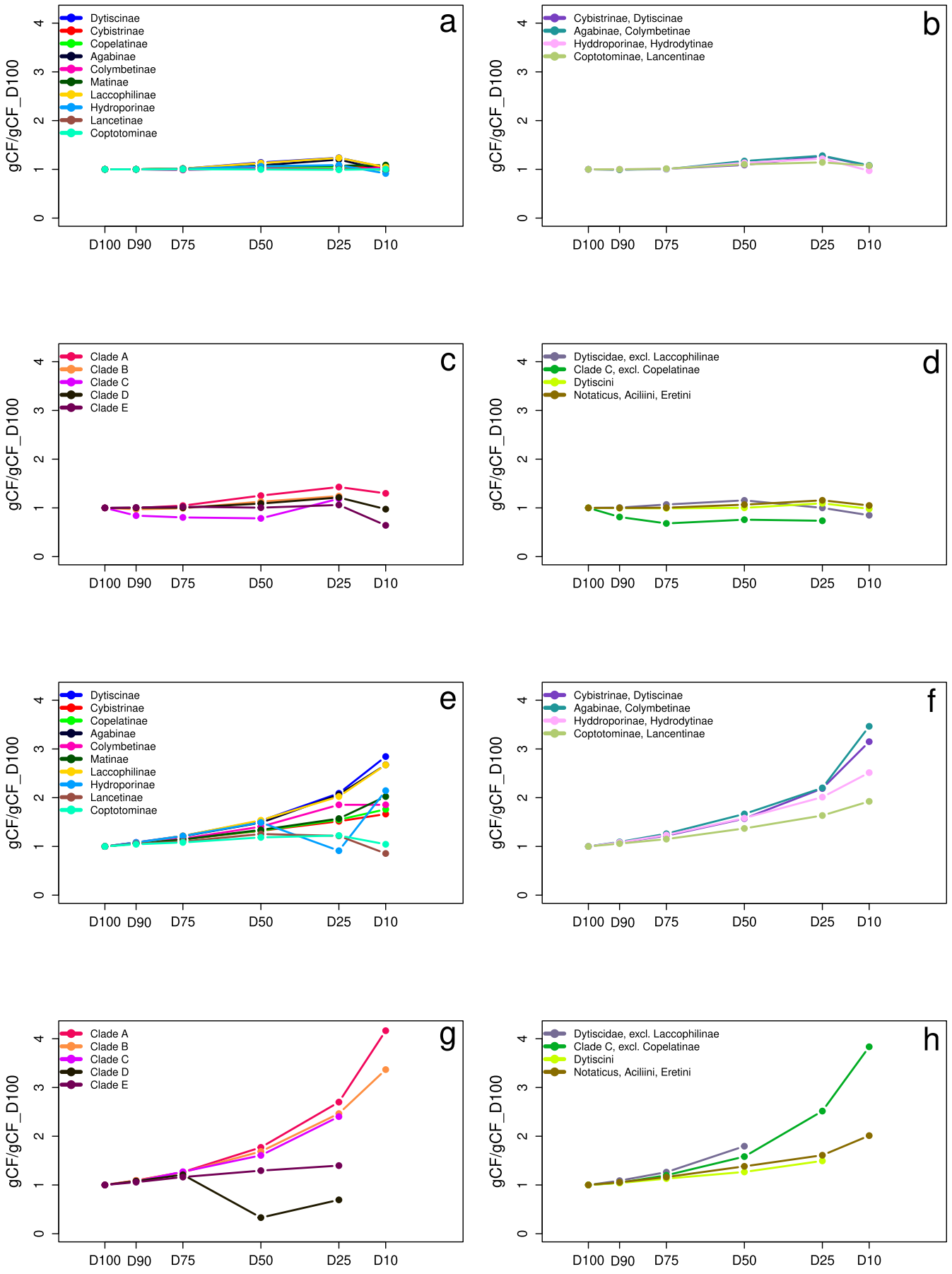

FIGURE 4 Legend on next page.

TABLE 1 Support from four-cluster likelihood mapping analyses of alternative resolutions for the four problem nodes using reduced datasets (all, nounique, common3, common4).

	Basal trich. Problem			Copelatinae problem			*Hyderodes* problem			*Notaticus* problem		
	ASTRAL	CML	THIRD	ASTRAL	CML	THIRD	ASTRAL	CML	THIRD	ASTRAL	CML	THIRD
149 taxa												
All	**52**	27.5	20.5	**48.4**	27.9	23.7	**77.1**	22.4	0.5	**50.4**	27.9	21.7
Nounique	**51.8**	28	20.2	**51**	26.7	22.3	**77.7**	21.6	0.8	**48.7**	30	21.3
Common3	**54.6**	26.6	18.7	**54.2**	28.2	17.6	**75.8**	23.6	0.7	**57.6**	25.7	16.7
Common4	**53**	27.2	19.9	**58.7**	28.4	12.8	**84.7**	15.2	0.1	**59.9**	22.3	17.8
95 taxa												
All	**62.5**	23.3	14.2	**40.4**	35.2	24.4	47.6	**49.8**	2.6	35.1	**51.2**	13.7
Nounique	**53.8**	30.9	15.2	**40.5**	32.1	27.4	**53.4**	43.9	2.8	31.2	**54**	14.8
Common3	**50**	32.9	17.1	**44.4**	27.8	27.8	47.1	**49.3**	3.6	39.9	**49.2**	10.9

Note: 149-taxa dataset has all terminals, 95-taxa dataset has excluded the 54 taxa from the exon capture study of Vasilikopoulos et al. (2021). Percentages correspond to the equilateral triangle divided into 3, not 7, basins of attraction and sum to 100 ((Strimmer and von Haeseler 1997). The ASTRAL resolution is found in Figure 2, the CML resolution in Figure S1 and the THIRD resolution was never recovered in any analysis. Highest support for each dataset and problem in bold.

Increasing dataset occupancy by using only common genes between legacy datasets in different ways (all (Figure 2), nounique (Figure S8), common3 (Figure S9), common4 (Figure S10)) revealed that in the most reduced dataset with only 369 loci (average per taxon loci occupancy 212) support for a few reference clades was reduced to less than maximal in the ASTRAL analysis (Figure S10): Agabinae (0.86), Dytiscinae + Cybistrinae (0.68) and Hydroporinae + Hydrodytinae (0.86). For all other reference clades and for also these three with the larger datasets (all, nounique, common3) support remained maximal (Figure 2 and Figures S8–S10). For several of the other nodes a pattern could be discerned that dataset all and nounique behaved similarly while common3 sometimes behaved like common4. For instance, clades A, B, C, D and E remained maximally supported by all and nounique, while clades B and C were not recovered at all by datasets common3 and common4 and support for clade D was reduced to 0.85 in both. Clades B and C collapsed due to a shift of Matinae to a position sister to Agabinae + Colymbetinae. Support for Clade A (0.27), Dytiscini (0.99) and Notaticus + Aciliini + Eretini (0.67) decreased from maximal only with the smallest dataset (Figure S10). Concatenated maximum likelihood analyses (all (Figure S1), nounique (Figure S11), common3 (Figure S12) and common4 (Figure S13)) remained in conflict with the ASTRAL analyses across all four datasets regarding the basal ingroup trichotomy resolution, position of Copelatinae in clade C and paraphyly of Dytiscini. However, the congruent clade *Notaticus* + Eretini + Aciliini appeared with the common3 and common4 datasets and with common3 also the congruent clade Deronectina + Hydroporina + Siettitiina was recovered (Figures S12 and S13). FcLM assessment remained in support of the ASTRAL resolution for all four conflicting nodes (Table 1). Support for the ASTRAL resolution also increased as the datasets increased in completeness and was highest with the common4 dataset for three of the four nodes: support for a monophyletic Dytiscini reached 84.7%, *Notaticus* + Aciliini + Eretini 59.9%, ASTRAL position of Copelatinae 58.7% and a monophyletic Dytiscidae except Laccophilinae 54.6%, the last with the common3 dataset (Table 1).

Exclusion of all 54 terminals with low average loci occupancy (8%) associated with the target capture study of Vasilikopoulos et al. (2021) increased average per taxon locus occupancy of the full dataset from 35% to 51%. With the full loci dataset, all our well-supported ASTRAL results remained stable except for the loss of a monophyletic Deronectina + Hydroporina + Siettitiina, despite this significant reduction of terminals (Figure S14). The basal trichotomy was resolved differently with Lancetinae + Coptotominae as sister to the remaining ingroup, like in the CML analysis but likewise unsupported (0.39). Support for clade C dropped marginally from 1.0 to 0.99, while support for a clade C except Copelatinae remained at 1.0 as did support for Dytiscini and *Notaticus* + Aciliini + Eretini. A maximally supported Hydroporinae except Methlini was recovered, but it should be remembered that several tribes within Hydroporinae were not represented in this taxon-reduced dataset, namely early diverging lineages Hydrovatini and Pachydrini as well as the more derived Vatellini. Relationships within clade E were unresolved except for Deronectina + Hydroporina (1.0). We also ran analyses with the reduced datasets nounique (Figure S15) and common3 (Figure S16)

FIGURE 4 Gene concordance factor (gCF) trends for a set of reference clades (a,b,e,f) and a set of target clades of interest (c,d,g,h) when loci are filtered and excluded based on occupancy (a–d) or phylogenetic informativeness (e–h). Datasets D100, D90, D75, D50, D25 and D10 are equivalent to retaining all (D100) or top 90%, 75%, 50%, 25% or 10% of loci when ranked in order of occupancy (a–d) or phylogenetic informativeness (e–h).

where the same 54 taxa were excluded. Dataset nounique gave a very similar result but where support for the basal resolution increased slightly (0.68) while the support for clade C decreased slightly (0.96) (Figure S15). With the common3 dataset, support for clade C dropped further to 0.69 and Clade C except Copelatinae decreased to 0.95. Also, support for Deronectina + Hydroporina dropped from 1.0 to 0.94 and the basal trichotomy resolution to 0.4 (Figure S16). Notably, monophyly of Dytiscini collapsed with dataset common3, but with 0 in support of Dytiscinae excluding *Dytiscus*, while the position of *Notaticus* remained stable (Figure S16). Concatenated maximum likelihood analyses with the same 54 terminals excluded (Figures S17–S19) remained in conflict with the ASTRAL analysis across all three datasets (all, nounique, common3) regarding the basal ingroup trichotomy resolution, position of Copelatinae in clade C, and paraphyly of Dytiscini, but with the congruent clades *Notaticus* + Eretini + Aciliini and Deronectina + Hydroporina + Siettitiina appearing with the common3 dataset (Figure S19). Here, the phylogenetic differences between data subsets follow the same pattern independent of the inclusion or exclusion of this large taxon block with low occupancy. FcLM assessment of the four problem nodes remained in support of the ASTRAL resolution for the basal ingroup trichotomy problem and for the Copelatinae problem (Table 1). However, the conflicting resolution of Dytiscini came to a tie between the ASTRAL and CML resolutions, and for the *Notaticus* problem, support changed to favour the CML resolution (Table 1).

Gene filtering based on phylogenetic informativeness

Filtering genes based on phylogenetic informativeness had a significantly larger effect on gCF compared with filtering based on locus occupancy (Figure 4e–h). Nonetheless, all reference clades were recovered as maximally supported across all datasets (D100-D10) with ASTRAL. In contrast to when genes were filtered based on locus occupancy, gCF increased gradually and peaked at D10 in almost all reference clades (Figure 4e,f). The only exceptions were Lancetinae and Coptotominae which peaked at D50 and D25, respectively, and Hydroporinae experienced a notable dip to the lowest gCF score at D25 but still peaked at D10 (Figure 4e). All four subfamily couplets showed a unanimous gradual increasing pattern of gCF and peaked at D10 (Figure 4f). Clades A-E were all recovered as maximally supported across D100-D25 except clade D which dropped in support to 0.96 for D50 and D25. With the most reduced dataset D10 only clades A and B were recovered. Clade C was not recovered since Copelatinae moved to a poorly supported (0.34) position as sister to clade B while clades D and E collapsed because Pachydrini surprisingly became nested within clade E with strong support (clade E incl. Pachydrini: 0.97, clade D including Pachydrini: 1.0). At D25 Pachydrini was instead recovered as sister to the remaining Hydroporinae and it was the rogue behaviour of this single terminal that caused the deviant gCF pattern of both Hydroporinae (Figure 4e) and clade D (Figure 4g). Several clades were maximally (or near maximally, Dytiscini 0.99 at D25) supported across D100-D25 but were not recovered at all with the most reduced dataset D10, including Dytiscini clades C and E. For all these gCF still gradually increased from D100 to D25 like in the reference clades (Figure 4g,h). Clade Vatellini + Bidessini deviated from all other monitored clades in that also local posterior probability gradually increased along with gCF and peaked instead of dropped at D10 (local posterior probability increased from 0.41 at D100 to 0.88 at D10). When filtered by phylogenetic informativeness, FcLM analyses changed to marginally support the CML resolution over the ASTRAL resolution for both the basal ingroup trichotomy at D25 (42.8%) (Figure S4) and for the Copelatinae position at D10 (36.7%) (Figure S5). For the *Hyderodes* and *Notaticus* problems, FcLM support remained in favour of the ASTRAL resolution, increasingly so at D25 (60.4%) and D10 (56.1%) for the *Notaticus* problem (Figure S7), but slightly reduced for the Dytiscini problem at D10 (63.8%) (Figure S6).

Reference clade-based gene filtering

To investigate the basal trichotomy in Dytiscidae among Couplet 4 + Clade A + Laccophilinae, we used the Dytiscidae ingroup branch as a reference clade and filtered away all gene trees that did not recover this node from the input to ASTRAL. This resulted in retaining 1687 gene trees. The basal trichotomy in the resulting species tree was resolved as Couplet 4 + (Laccophilinae + Clade A) (Figure S20), but support was even lower (0.39) compared to the unfiltered analysis (0.54) for the conflicting Laccophilinae + (Clade A + Couplet 4). The poorly supported resolution is in agreement with the CML tree where this configuration received maximal support. Interestingly, the node-filtered ASTRAL tree also recovered Copelatinae in the same position as in the CML tree as sister to Cybistrinae + Dytiscinae with strong support (0.98), and a non-monophyletic Dytiscini with the clade *Hyderodes* + Dytiscinae except *Dytiscus*, as in the CML tree, though with weak support (0.42). *Notaticus* + Aciliini + Eretini remained maximally supported (Figure S20).

Rogue taxon exclusion

The single Pachydrini terminal (*Pachydrus* sp.) was identified as a rogue taxon based on both irrational behaviour in the phylogenetic informativeness gene filtering exercise and on having the lowest loci occupancy (5%) of all ingroup taxa (Table S2), and was therefore excluded in a series of analyses. Exclusion of *Pachydrus* resulted in increased support for the clade of Hydroporinae excluding Methlini from 0.8 to 0.98 in the ASTRAL analysis of the full and nounique dataset and even to 1.0 in the common3 dataset (tree from full dataset shown in Figure S21). No other noteworthy differences were detected with the exclusion of *Pachydrus*. Concatenated maximum likelihood analyses behaved similarly to respective dataset (all, nounique and common3) with *Pachydrus* excluded (not shown).

DISCUSSION

We assembled and analysed the largest genomic dataset to date for Dytiscidae in order to resolve the backbone relationships between subfamilies and tribes. A large body of phylogenetic literature using adult, and/or larval, morphological characters, smaller numbers of Sanger-sequenced genes and combined analyses with morphology and DNA, has largely identified a set of well-supported monophyletic clades today recognised at the subfamily, tribe or subtribe levels by mostly named taxa (Miller & Bergsten, 2016). However, the relationships between these named groups have been notoriously hard to resolve, and little consensus exists apart from four couplets of subfamilies (Figure 1) that are supported in recent genomic analyses focused on Adephaga beetles or the superfamily Dytiscoidea (Baca et al., 2021; Gustafson et al., 2020; Vasilikopoulos et al., 2021). With the large, whole genome analysis presented here, we take a major step forward and confidently resolve all but two backbone branches of inter-subfamily relationships.

We used two fundamentally different approaches of phylogenetic inference, both of which have their merits and shortcomings. Concatenated maximum likelihood (CML) analyses ignore gene tree vs. species tree conflicts due to incomplete lineage sorting, which causes inconsistent behaviour in certain parts of parameter space (Kubatko & Degnan, 2007; Roch & Steel, 2015). The ability to handle incomplete lineage sorting with the multispecies coalescent model was a paradigm shift for phylogenetics (Edwards, 2009; Edwards et al., 2016; Mirarab et al., 2021). However, in the realm of finite datasets, each method can be superior to the other depending on factors such as the degree of incomplete lineage sorting and gene tree estimation error, which are data dependent (Mirarab et al., 2014; Molloy & Warnow, 2017; Roch & Warnow, 2015). Our approach was to compare the outcome of the two methods, be confident in relationships recovered by both with high support and evaluate conflicts with specific tests and external evidence such as morphology.

Our FcLM test of relative support for alternative resolutions of the four analysed problem areas supported the ASTRAL over the CML resolution in all cases (Figure 3), which requires reflection. Phylogenomic datasets, including ours, commonly include a moderate to high proportion of missing data (Portik et al., 2023; Xi et al., 2015) and it is natural to consider if missing data may have influenced any inference. Missing data may affect some methods negatively, especially when missing data is non-random (Xi et al., 2015). However, there are a number of studies that show little benefit from filtering and excluding gene loci based on missing data (Hosner et al., 2015; Molloy & Warnow, 2017; Streicher et al., 2016) or that such an approach may even worsen accuracy (Jiang et al., 2014; Streicher et al., 2016). There is also overwhelming evidence that even taxa with many missing characters can be convincingly positioned in a phylogeny if sufficient informative characters are available (Wiens & Morrill, 2011). If a consensus is emerging on this long-standing topic, it could arguably be that missing data does not need to be intrinsically problematic (e.g. Portik et al., 2023), but also that effects are method dependent. We used the species tree method as implemented in ASTRAL (Mirarab et al., 2014) not only for its ability to model incomplete lineage sorting, but also for its greater robustness to missing data. Several simulation studies have shown that ASTRAL is more robust to missing data and, in contrast to several other methods, remains statistically consistent under models of missing data (Nute et al., 2018; Rhodes et al., 2020; Xi et al., 2015). Empirical reports that have found ASTRAL to produce more spurious results are more often related to short or uninformative loci rather than missing data per se (Hosner et al., 2015). Moreover, the statistical consistency of ASTRAL under models of missing data means that species tree accuracy increases with the *addition of incomplete loci* and vice versa; removing loci based on missing data decreases accuracy (Jiang et al., 2014; Molloy & Warnow, 2017; Nute et al., 2018; Vachaspati & Warnow, 2015; Xi et al., 2015).

We explored the effect of removing loci based on missing data in different ways, but interpretations are hampered by the fact that an increase in overall dataset occupancy is achieved by a reduction of loci numbers. The single factor 'missing data' cannot be extracted and its effect for phylogenetic reconstruction studied in isolation through filtering exercises. There is overwhelming support for increased accuracy and confidence with the addition of loci (Mirarab et al., 2014; Molloy & Warnow, 2017), as we have seen above even with the addition of incomplete loci for ASTRAL (Nute et al., 2018; Xi et al., 2015). Counterintuitively, while overall dataset occupancy improves, the problem caused by missing data may in fact increase through filtering exercises. Specifically, the same level of missing data may become more of an issue the smaller the dataset (Xi et al., 2015). Topological changes from our filtering exercises could therefore be interpreted as stemming from (i) the higher overall occupancy in the dataset; (ii) the greater missing data effect; or (iii) the reduction in overall data evidence, specifically the number of loci. If interpreted as stemming from the first alternative, one would assume improved accuracy with filtering, but with the latter two, reduced accuracy. Our approach was to define a set of reference clades (subfamilies and four subfamily couplets) assumed to be true, and when any of these were not recovered or support decreased, changes were interpreted as worse reconstructions.

Our main finding is therefore that endeavours to increase occupancy by excluding large sets of the most poorly represented loci (datamatrices common3, common4, D25 and D10) is a poor strategy that worsen phylogenetic accuracy. When reduced to a dataset with less than 400 loci (common4, 369 loci), we find support for well-established clades reduced such as Agabinae, Dytiscinae + Cybistrinae and Hydroporinae + Hydrodytinae. With the most reduced datasets we also find the collapse of some clades (B, C, Deronectina + Hydroporina) that become unanimously and maximally supported across all other larger datasets and inference methods. Regarding the four conflicts analysed with FcLM, we found that filtering loci based on occupancy had no effect at all on the preference for the ASTRAL solution over the CML reconstruction (Figures S4–S7) and neither did removing non-random blocks of loci with lower occupancy (Table 1). This shows that missing data per se is not biasing the FcLM support of the ASTRAL resolution of the four nodes. In

contrast, removing a third of all terminal taxa (representing a taxon block with low loci occupancy) did affect the FcLM preference for two of the four problems (Table 1), as did the most severe loci filtering based on phylogenetic informativeness (Figures S4 and S5). These dataset alterations affect other important properties for phylogenetic reconstruction than just overall occupancy and loci numbers.

Our exercise with filtering and reducing our full dataset based on occupancy gave a couple of important insights, however. Large amounts of missing data in the single terminal *Pachydrus* did cause a reduction in support in the ASTRAL analysis to 0.8 for the node defining Hydroporinae excl. Methlini, even in the full dataset. This was discovered from erratic behaviour of the terminal in the filtering exercises, causing atypical gCF graphs for Hydroporinae and clade D (Figure 4). Once identified as a rogue taxon due to missing data and excluded, alone or along with all taxa from the exon target capture study (Vasilikopoulos et al., 2021), support for Hydroporinae excl. Methlini was maximal (1.0: *Pachydrus* excluded common3, Vasili-21 taxa excluded All, nounique, common3) or near maximal (0.98: *Pachydrus* excluded All, *Pachydrus* excluded nounique). Excluding all terminals from the exon capture study which form a taxon block of extra low occupancy improved overall dataset occupancy from 35% to 51% but did not change any of our main conclusions. The only well-supported clade that was lost was Deronectina + Hydroporina + Siettitiina, but this was more likely a result of losing the *Graptodytes* terminal, a deep branch in the Siettitiina clade complementary to the remaining representative terminals *Porhydrus* and *Stictonectes* (Villastrigo et al. 2021).

A second interesting effect was discovered by the exercise of reducing the dataset in ways so that overlapping genes from legacy datasets made up a larger share. The common3 dataset with still a hefty 1529 genes behaved for some nodes (collapse of clades B, C and lower support for clade D) similar to the common4 dataset with a mere 369 loci whereas our expectation would be that it should behave more like the nounique (3202 loci) and all (5364 loci) datasets. A likely explanation was found in the per taxon locus occupancy statistics which show that the 54 taxa from the exon capture study have exactly 260 loci on average in both the common3 and common4 datasets but exactly 411 loci on average in the all and nounique datasets (the fact that locus occupancy for these terminals do not change between pairs of datasets is because loci of Vasilikopoulos et al. (2021) is a subset of loci in the dataset by Vasilikopoulos et al. (2019) and have no unique loci to be removed). Here the interaction of a non-random taxon block by gene block absence might have more adversely affected the outcome even if the dataset was larger (1529 loci) than a dataset D25 (1340 loci) filtered 'randomly' by loci occupancy which did recover clades B, C and D with maximal support. At least such an explanation is in line with simulation studies showing that non-random patterns of missing data is more problematic than randomly distributed missing data (Xi et al., 2015).

The conflicting but in both cases maximally supported resolution of clade C by ASTRAL and CML is an intriguing problem. Filtering based on loci occupancy (D100-D10) revealed an opposite gCF trend of Clade C excluding Copelatinae in relation to the behaviour for all reference clades and all other well-supported clades. There is no biological reason gCF should systematically decrease when datasets become more complete (more informative) except towards low loci numbers where estimation errors increase (Minh, Hahn, & Lanfear, 2020). Also, the gene concordance vs. discordance factors for the full dataset oddly show that the ASTRAL resolution is not the best supported. FcLM do however support the ASTRAL resolution for the full dataset and remain stable after filtering on loci occupancy. So how can the odd behaviour of gCF be explained? Lanfear and Hahn (2024) provide an interpretation for an identical empirical example in birds where a very short maximally supported node of Columbaves has a very low gCF but higher support (discordance) for an alternative resolution. Citing a demonstration by a collegue, an unequal rate of non-monophyly of two included clades will lead to unequal numbers of nondecisive trees which will lead to a biased concordance vector for one of the resolutions of that node (Lanfear & Hahn, 2024). When the number of undecisive trees is very high (95.4% in the case here, 97.6% in the case of Columbaves), this bias can inflate the discordance factor of a minor alternative topology, an artefact rather than any biological process (Lanfear & Hahn, 2024). We also calculated the site concordance factor since this measure, being quartet-based, is immune to nonmonophyly biases (Lanfear & Hahn, 2024) and here the clade did not differ in trend from reference clades, had a similar support (36.08) as several other clades in the tree and with site discordance factors lower (31.52 and 32.4). This provides a satisfactory explanation to the gCF paradox on the resolution of clade C as well as similar but smaller trends for recovery of clade C itself and the resolution of the basal ingroup trichotomy.

As we are unable to explain the two conflicting inter-subfamily resolutions between ASTRAL and CML analyses to problems of missing data, we suspect that method assumptions are the culprit. Based on the FcLM results (Figure 3) we here assume that the ASTRAL resolution is correct, which then points towards violated assumptions by the CML analyses. The firsthand suspect is then naturally the ignorance of incomplete lineage sorting (ILS) by the CML strategy. Jiang et al. (2020) for instance, showed that the concatenation assumption of topologically congruent gene trees is rejected by a much larger proportion of loci in empirical phylogenomic datasets across many animal groups, including insects, compared with the proportion rejecting the MSC model. The basal ingroup trichotomy and the resolution of clade C both involve a very short internode. With three lineages developing within such a short timeframe, in both cases there is little time for a signature of the order of the split to be registered in the genome, supported by the observation that the vast majority of gene trees are undecisive for the two nodes. Accounting for ILS for the few gene trees that are decisive is likely imperative, as the short timeframe likewise paves the way for a high level of ILS, a situation where species tree methods are generally more accurate than concatenation methods (Edwards, 2009; Mirarab et al., 2014; Mirarab & Warnow, 2015). In certain parts of parameter space (the anomaly zone) where ILS is high, most genes coalesce in the longer ancestral internode, and the most common gene tree will not be the species tree (Degnan & Rosenberg, 2006). Yet concatenation of the gene

alignments and not accounting for ILS will lead to the reconstruction of the most common gene tree with increasing support as more genes are sampled (Kubatko & Degnan, 2007).

The two conflicting tribal nodes within Dytiscinae are somewhat different and more surprising as they do not involve very short internodes. With longer internodes, ILS should be lower and congruence between concatenated and species tree methods is expected. We suspected the long terminal branch of *Notaticus* might be pulled towards the base of Dytiscinae or towards the *Hyderodes* terminal, but taxon exclusion experiments showed their positions remained stable (Figures S2 and S3). This test does not exclude artefacts due to the long terminal branches, but neither supports it. An alternative explanation, still based on assuming the validity of the ASTRAL resolution following the FcLM results and also on morphological evidence (see below), is that the CML analysis does in fact suffer from non-random missing data patterns in combination with high rate heterogeneity across sites, loci or branches. Simulations have shown that CML analyses may become inconsistent under some of these conditions, even in the absence (Simmons, 2012; Xia, 2014) or under low levels of ILS (Xi et al., 2015). The CML analysis with the common3 dataset did recover the ASTRAL position of *Notaticus*, which changed with the addition of blocks of genes with a non-random pattern of missing data across terminals. But *Notaticus* has a high loci occupancy (64%) in the full dataset, as does *Hyderodes* (50%), one of the two *Dytiscus* terminals (61%) and two of the three *Hydaticus* terminals (40%–42%). Why this part of the tree would be affected is therefore not very clear other than if conditions apply, an inconsistent method by definition moves towards an incorrect tree with the addition of similar data. The long terminal branch of *Notaticus* does indicate large rate heterogeneity across branches in this part of the tree, which may often be as compromising for accurate reconstruction as rate heterogeneity across sites or loci if not correctly modelled. Admittedly, species tree methods like ASTRAL, involving an ML estimation step of gene trees, are not immune against long-branch attraction with empirical finite-length loci (Roch et al., 2018), yet species tree methods appear more robust to both missing data (Nute et al., 2018; Xi et al., 2015) and to long branches with elevated substitution rates (Liu et al., 2015). As the FcLM analyses remain in strong favour of the ASTRAL resolution in both cases and under all loci filtering and exclusion datasets (Figures S4–S7, Table 1 except with exclusion of a large number of terminals), and gCF support also favours the ASTRAL over the CML resolutions, the two problems in Dytiscinae seem more a problem of explaining the methodological inconsistency than resolving the nodes with confidence.

Phylogenetic conclusions

Since Miller's (2001) comprehensive Dytiscidae phylogeny, those groups that currently have names at the family group level have, with few exceptions, been consistently supported by other analyses and are strongly supported here (Figure 2, e.g. Baca et al., 2021; Désamoré et al., 2018; Gustafson et al., 2020; Michat et al., 2017; Miller & Bergsten, 2014a, 2023a; Ribera et al., 2008; Ribera, Hogan, & Vogler, 2002; Vasilikopoulos et al., 2021). Notable exceptions include Hydrotrupini (see Toussaint et al., 2017 and below for resolution), Hydroporini, which was not recovered as monophyletic here (Figure 2 and Figure S1), and the methodological inconsistencies affecting Dytiscini in our and previous analyses (Figure 2 and Figure S1; Miller, 2000, Miller & Bergsten, 2014a; see Section 4 below). For discussion of synapomorphies and taxonomic composition of the named subfamilies, tribes and genera, previous phylogenetic discussions by Miller (2001), Miller and Bergsten (2014a, 2016, 2023a) and Michat et al. (2017) are available. Analyses and phylogenetic insights within many of the named Dytiscidae groups have also been done recently, including within Agabinae (Alarie & Michat, 2020, 2022; Okada et al., 2019; Toussaint et al., 2017), Colymbetinae (Balke, Hajek, & Hendrich, 2017; Barman et al., 2014; Michat et al., 2023; Morinière et al., 2014), Cybistrinae (Michat et al., 2015, 2019; Miller et al., 2024; Miller, Bergsten, & Whiting, 2007), Copelatinae (Bilton et al., 2015; Toussaint, Balke, et al., 2016), Laccophilinae (Alarie, Watanabe, & Michat, 2023; Benetti et al., 2019; Michat & Toledo, 2015; Toledo & Michat, 2015), Aciliini (Alarie, Michat, Bergsten, & Hájek, 2023; Alarie, Michat, Shaverdo, & Hájek, 2023; Bukontaite et al., 2015), Sternopriscina (Alarie et al., 2018, 2019, 2021; Hendrich et al., 2014; Toussaint et al., 2014; Toussaint, Hendrich, et al., 2016; Villastrigo et al., 2023), Hygrotini (Villastrigo et al., 2017; Villastrigo et al., 2018), Hyphydrini (Alarie et al., 2017, 2022), Hydroporini (Fery & Bouzid, 2016; Queney et al., 2020; Villastrigo et al., 2021), Deronectina (Fery & Ribera, 2018), Siettitiina (Kanda et al., 2016; Ribera & Reboleira, 2019) and Bidessini (Balke, Bergsten, et al., 2017; Hendrich et al., 2020; Miller, 2016; Miller & Short, 2015; Miller & Wheeler, 2015; Watts et al., 2023).

Below we discuss well-supported phylogenetic relationships among those named subfamilies and tribes based on this analysis in light of morphological evidence and other studies

1. Cybistrinae + Dytiscinae

Cybistrinae has long been closely associated with the Dytiscinae as a tribe, Cybistrini, within the subfamily (Alarie, Michat, & Miller, 2011; Burmeister, 1976; Michat et al., 2017; Miller, 2000, 2001, 2003; Ruhnau & Brancucci, 1984; Sharp, 1882). However, in mainly molecular phylogenetic analyses, Cybistrinae were found resolved elsewhere from other Dytiscinae (Miller & Bergsten, 2014a, 2023a; Ribera et al., 2008; Ribera, Hogan, & Vogler, 2002; Vasilikopoulos et al., 2019). Miller and Bergsten (2014a) elevated the group to subfamily rank. More recently, however, both UCE and exon-capture datasets have recovered Cybistrinae and Dytiscinae as sister taxa (Gustafson et al., 2020; Vasilikopoulos et al., 2021), although this has been at least partly analysis dependent (Baca et al., 2021).

Here, we consistently recovered Cybistrinae and Dytiscinae as sister taxa with maximal support using both methods (Figures 2 and 5, Figure S1). In addition, morphological support is strong for Cybistrinae + Dytiscinae including, in adults: (1) the anterior

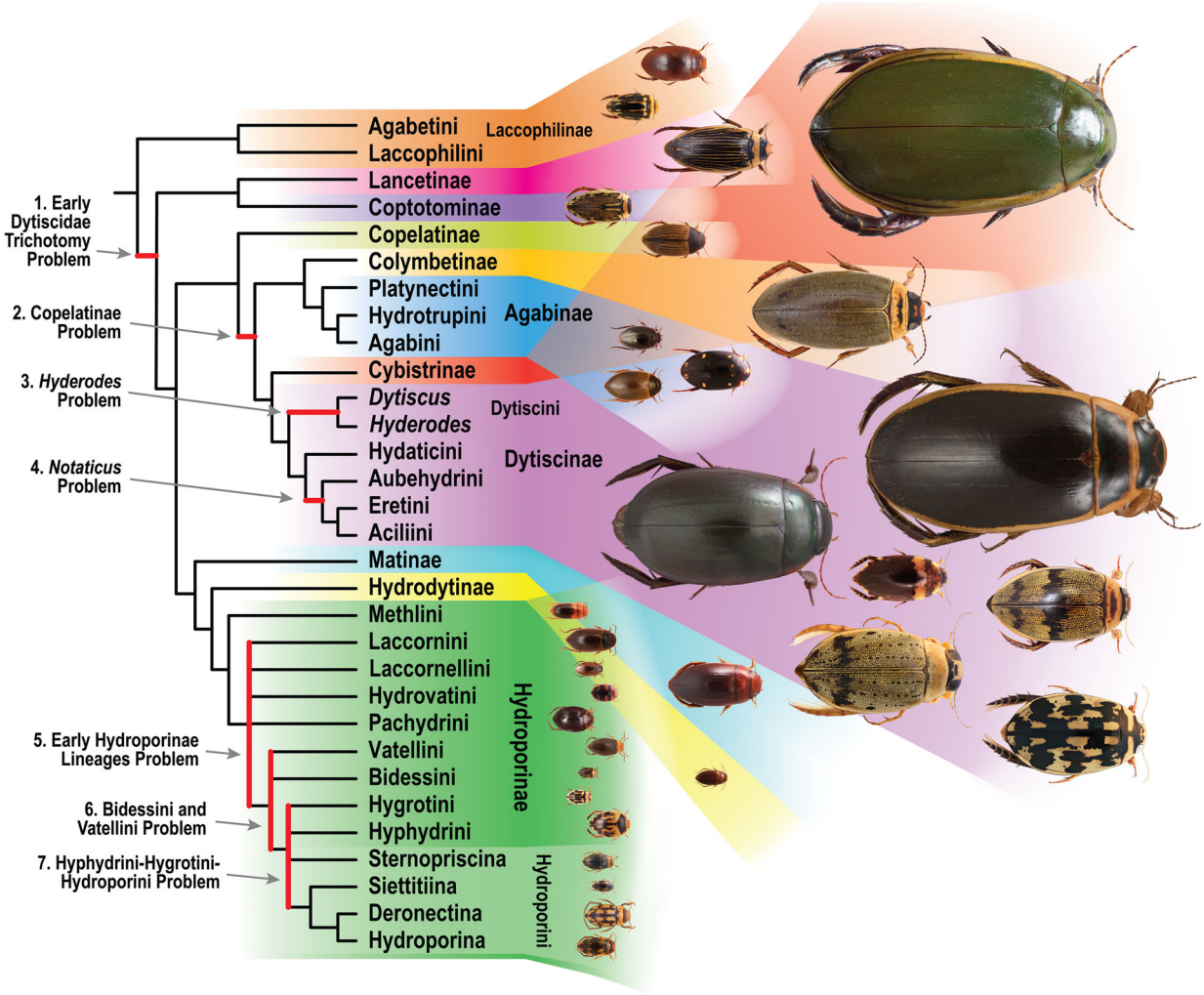

FIGURE 5 Summary cladogram of results from current study. Red internodes represent nodes where ASTRAL (Figure 2) and CML (Figure S1) analyses were in conflict but FcLM unambiguously supported the here accepted (ASTRAL) resolution (Figure 3; Table 1). Red polytomies in Hydroporinae represent remaining uncertain resolutions that require further taxon sampling to resolve. Note that specimens of Laccornellini were not included in this study, but the tribe is placed in the polytomy of early Hydroporinae lineages based on previous studies (Miller & Bergsten, 2014a, 2023a). Beetle photos reproduced from Miller & Bergsten (2016).

margins of the eyes rounded, not emarginate; (2) the median lobe of the male aedeagus bilaterally symmetrical with an elongate ventral sclerite; (3) females with one genital opening for both reception of sperm and oviposition; and (4) fusion of the female gonocoxae, among others (Miller, 2000, 2001; Miller, Bergsten, & Whiting, 2007). Larvae have: (1) abdominal segments VII–VIII with lateral fringes of long, natatory setae, (2) the antennomeres and maxillary and labial palpomeres variously subsegmented, (3) the anterior clypeal margin trilobed, (4) the premaxillary lobes well developed and anteriorly projected, (5) additional setae present on antennomeres I–II, and several more features (Alarie, Michat, & Miller, 2011, Michat et al., 2017). Nevertheless, cybistrines are morphologically different in significant ways from Dytiscinae (Miller et al., 2024; Miller & Bergsten, 2016; Miller, Bergsten, & Whiting, 2007), so there is utility in continuing to recognize them at the subfamily rank which we do here.

2. Agabinae + Colymbetinae

These two groups were historically placed together as separate tribes in the paraphyletic subfamily Colymbetinae until Miller (2001) elevated each to subfamily rank along with a major restructuring of Colymbetinae, mainly elevation of numerous tribes to subfamily. In that morphological analysis, the two groups were not resolved together as monophyletic. This conclusion was partially supported by molecular analyses by Ribera, Hogan, and Vogler (2002) and Ribera et al. (2008) but lately phylogenomic studies based on UCE and transcriptomes have consistently recovered Agabinae and Colymbetinae as sister groups (Baca et al., 2021; Gustafson et al., 2020; Vasilikopoulos et al., 2021), as we do with maximal support in both ASTRAL and CML analyses (Figures 2 and 5, Figure S1).

Nilsson (1997) and Miller (2001) recognized two main groups in Agabinae, the *Agabus*-group of genera and the *Platynectes*-group of

genera. In their sense, the *Agabus*-group includes mainly Holarctic taxa, whereas the *Platynectes*-group includes taxa from South and Central America, southeast Asia and Australia. However, Ribera, Hogan, and Vogler (2002) and Ribera et al. (2008) found Agabinae to not be monophyletic, with the *Platynectes*-group not related to the *Agabus*-group. Later analysis of Agabinae more closely by Ribera et al. (2004) also resulted in two clear groups.

A problematic taxon has been the North American and east Asian genus *Hydrotrupes* Sharp. Miller (2001) found *Hydrotrupes* sister to the *Platynectes*-group. Roughley (2000) erected the subfamily Hydrotrupinae for the genus (until then formally classified as a genus in the tribe Agabini) based on larval features (Beutel, 1994) that suggested the genus is sister to all Dytiscidae except Copelatinae. This was not supported in the analyses by Miller (2001), Alarie (1998) or Ribera et al. (2008), which found *Hydrotrupes* related to all of Agabinae or to the *Platynectes*-group of genera. Miller and Bergsten (2014a, 2023a) found Agabinae, as historically defined (i.e. the *Agabus*-group + the *Platynectes* group + *Hydrotrupes*) to be monophyletic with moderate support and recognized two tribes, Agabini (the *Agabus*-group) and Hydrotrupini (the *Platynectes*-group + *Hydrotrupes*). Three groups were also recognized by Toussaint et al. (2017), Agabini (the *Agabus*-group), 'Platynectini' (the *Platynectes*-group) and Hydrotrupini (*Hydrotrupes*) based on results of their analysis suggesting that Hydrotrupini is sister to Agabini, not to the *Platynectes*-group (contra Miller & Bergsten, 2014a, 2023a). Michat et al. (2017) found a monophyletic Agabinae based on larval morphology but found Agabini to be polyphyletic. Clearly, relationships among the three distinctive components of this subfamily have not been easily clarified.

Agabinae sensu lato is here maximally supported as monophyletic, with three groups resolved as *Platynectes*-group + (Hydrotrupini + Agabini) (Figures 2 and 5, Figure S1), as found by Toussaint et al. (2017). The main synapomorphy for Agabinae is the presence of a linear series of stout setae at the anteroventral apical angle of the metafemur. Miller (2001) regarded Agabinae to be among the most poorly supported subfamilies of Dytiscidae.

We note here that Toussaint et al. (2017) justifiably proposed the tribe Platynectini for the *Platynectes*-group of genera (*Andonectes* Guéorguiev and *Platynectes* Régimbart, not including *Hydrotrupes*), a group first proposed by Nilsson (1997). However, their proposed tribe name lacks a description and is not available based on Article 13 of the Code of Zoological Nomenclature (ICZN, 1999) (see Miller and Bergsten (2023a) for further discussion). We therefore make space available here for Emmanuel Toussaint and Michael Balke to provide a description to make the name Platynectini available from this publication:

> Platynectini Toussaint & Balke, new tribe
>
> Type genus: *Platynectes* Régimbart, 1879.
>
> Description and diagnosis: Members of this tribe of subfamily Agabinae are characterized by a pair of elliptical, sublateral clypeal foveae and female metatibia and metatarsus with a ventral fringe of natatory setae present. Hydrotrupini is restricted to a narrower definition in relation to Miller and Bergsten (2014a, 2023a) to only include the type genus *Hydrotrupes* Sharp, 1882. As such, both sexes lack a ventral fringe of natatory setae on metatibia and metatarsus in Hydrotrupini and are otherwise diagnosed by several adaptations to a hygropetric habitat including a somewhat Hydrophilid-like body shape (Larson et al. 2000; Miller & Perkins, 2012). Agabini is characterized by members having linear clypeal foveae near anterolateral angles or along the entire anterior margin and having female metatibia and metatarsus without a ventral fringe of natatory setae (except in *Ilybius discidens* Sharp, 1882 where this fringe is present in females but otherwise have all characters of the *subaeneus* group of *Ilybius* Erichson) (Nilsson, 1997; Nilsson, 2000).

With the advent of cladistic analysis, the historical composition of Colymbetinae (i.e. including Copelatinae, Coptotominae, Lancetinae, Hydrodytinae, Colymbetinae sensu stricto, Agabinae and portions of Laccophilinae (i.e. *Agabetes* Crotch)) became questionable (e.g. Burmeister, 1976). Miller (2001) circumscribed Colymbetinae much more narrowly than it had been to include only taxa in the tribe Colymbetini. Colymbetinae have adults with: (1) the eyes anteriorly emarginate, (2) the male median lobe bilaterally asymmetrical, (3) the lateral lobes bilaterally symmetrical, (4) the female gonocoxae usually dorsoventrally flattened and apically rounded, (5) the prosternum and prosternal process together in the same plane in lateral aspect, (6) the apices of the elytra evenly rounded, (7) the metatarsal claws unequal in length, and (except in *Rhantus tristanicola* (Brinck) and *Rhantus selkirki* Jäch, Balke & Michat) (8) abdominal pleurite II with transverse rugae (not visible with elytra closed) (Miller, 2001; Miller & Bergsten, 2014a, 2016, 2023a). Here, Colymbetinae sensu Miller (2001) was found to be monophyletic (Figures 2 and 5, Figure S1). Morinière et al. (2014) showed that the odd and isolated island taxa *R. tristanicola* and *R. selkirki*, previously recognised in a tribe, Anisomeriini, were deeply nested in the Colymbetinae genus *Rhantus*. The tribe Carabdytini, erected by Pederzani (1995), has also been shown to be unjustified and based on a highly derived species adapted to lotic habitats, *Carabdytes upin* Balke, Hendrich and Wewalka (Balke, 2001; Balke et al., 2009; Balke, Hajek, & Hendrich, 2017; Morinière et al., 2016).

There has not been much historical controversy over the monophyly of Colymbetinae sensu stricto (see above), but its relationship with Agabinae has been uncertain. Miller (2001) found them to not be together monophyletic, as did Ribera, Hogan, and Vogler (2002) and Michat et al. (2017), though Ribera et al. (2008) and Miller and Bergsten (2014a, 2023a) did find them to be resolved together, as did recent phylogenomic studies, albeit with limited taxon sampling (Baca et al., 2021, Gustafson et al., 2020, Vasilikopoulos et al., 2021). In our analysis, Colymbetinae and Agabinae are each monophyletic and together form a well-supported clade (couplet 2, Figures 2 and 5, Figure S1). Other than some generalized similarities and plesiomorphies, Agabinae and Colymbetinae do not share known distinctive morphological synapomorphies either in larvae or adults, despite their long history of close association and evident monophyletic relationship with each other based on genomic data. They were historically associated because of plesiomorphies (Miller, 2001). Now that this

relationship is better established, perhaps morphological evidence will be forthcoming.

3. Hydroporinae + Hydrodytinae

A sister group relationship between Hydrodytinae and Hydroporinae was first proposed by Miller (2001) after finding that a subset of species previously placed in *Agaporomorphus* Zimmermann (Copelatinae) was its own separate group with characteristics similar to hydroporines. The species were placed in their own genus, *Hydrodytes* Miller, and a new subfamily sister to Hydroporinae (Miller, 2001, 2002). Hydrodytinae have plesiomorphic features such as an externally visible scutellum (with the elytra closed) and the prosternal process not declivous, different from many Hydroporinae. However, hydroporines and hydrodytines also share several morphological synapomorphies including an elongate apodeme at the anterior end of the female gonocoxa and several features of the metafurca (Miller, 2001, 2002). This relationship has been supported in subsequent analyses involving combined data (Miller & Bergsten, 2014a, 2023a) and by reduced representation genomic datasets (Baca et al., 2021, Gustafson et al., 2020, Vasilikopoulos et al., 2021). The relationship is further maximally supported here (Couplet 3, Figures 2 and 5, Figure S1).

4. Coptotominae + Lancetinae

There has been little previous support for this relationship from adult morphology, though these two distinctive subfamilies have specimens superficially quite similar with elongate, streamlined bodies. Brinck (1948), presciently, proposed this relationship based on mainly general and plesiomorphic features placing *Coptotomus* Say and *Lancetes* Sharp together in the same tribe, Coptotomini. A comprehensive analysis of combined larval data also recovered this relationship with strong support based in part on a reduced number of *lamellae clypeales*, specialized spatulate setae at the anterior margin of the frontoclypeus (Bertrand, 1972) (convergent with Laccophilinae, Michat et al., 2017). Previous molecular or combined phylogenetic analyses with a handful of genes have not recovered this relationship (Désamoré et al., 2018; Miller & Bergsten, 2014a). Based on our analysis, the two groups are together monophyletic with strong support (Couplet 4, Figures 2 and 5, Figure S1) and this has been a consistent pattern in recent phylogenomic studies of Adephaga with included *Coptotomus* and *Lancetes* terminals (Baca et al., 2021, Gustafson et al., 2020, Vasilikopoulos et al., 2021). Again, this should hopefully spur reciprocal illuminative reinvestigations of adult and larval morphology in search of synapomorphies. It is surprising that this couplet was not more confidently established earlier, given that it is the strongest supported subfamily couplet by far based on gCF (37.2 compared with 6.08, 6.97 and 10.37 for the other three).

5. Clade A: (all subfamilies except Laccophilinae, Lancetinae, Coptotominae)

Previous phylogenetic analyses have regarded various other taxa as sister to the remaining Dytiscidae (see Problem 1 below). Clade A was not recovered by Gustafson et al. (2020) based on UCE loci. An expanded dataset of the same Adephaga UCE probe set (Baca et al., 2021), as well as the exon target capture study by Vasilikopoulos et al. (2021) did recover Clade A in concatenated maximum likelihood analyses but not in ASTRAL analyses, and hence, the support remained inconclusive. Here we find maximum support for Clade A for the first time based on both CML and ASTRAL analyses (Figures 2 and 5, Figure S1). There are currently no known morphological features supporting this configuration, though future investigation may find them.

6. Clade B: Matinae + (Hydrodytinae + Hydroporinae)

Matinae were historically placed in Colymbetinae sensu lato until Miller (2001) found them resolved as sister to the rest of Dytiscidae (though not without some reservations) and elevated the group to subfamily rank. That resolution was supported by Miller and Bergsten (2014a, 2023a), though not by Michat et al. (2017). However, here they are maximally supported as sister to Hydrodytinae + Hydroporinae by both ASTRAL and CML analyses (Figures 2 and 5, Figure S1). This clade was not found in UCE-based studies (Baca et al., 2021; Gustafson et al., 2020) nor by ASTRAL-based analyses by Vasilikopoulos et al. (2021) all of which largely associated Matinae with Agabinae or with Agabinae + Colymbetinae. Vasilikopoulos et al. (2021) did, however, find Clade B supported in their preferred CML analysis, which we here add strong species tree support to as well. There are no obvious morphological synapomorphies corresponding to this cladistic grouping, but there is also no other more clear relationship of matines to any other group of Dytiscidae. The members of Matinae are from eastern North America (*Matus* Aubé) and Australia (*Batrachomatus* Clark), a unique biogeographic relationship that warrants further investigation to understand their evolutionary history, particularly if they are, as well-supported here, the sister to such a large group as Hydrodytinae + Hydroporinae.

7. Clade C: Copelatinae + (Cybistrinae + Dytiscinae) + (Agabinae + Colymbetinae)

Colymbetinae sensu lato and Dytiscinae sensu lato have had a long history of close association (e.g. Sharp, 1882) mainly based on symplesiomorphies (Miller, 2001). Miller (2001) broke up the subfamily Colymbetinae into separate subfamilies, including Agabinae, but that analysis resulted in Dytiscinae (including Cybistrinae), Lancetinae, Colymbetinae and Agabine in a clade, albeit with little morphological support. A clade including these subfamilies was not recovered by Michat et al. (2017) based on larvae nor a combined analysis by Miller and Bergsten (2014a, 2023a). This clade (without Lancetinae but including Copelatinae) was recovered in this analysis with maximal support in both ASTRAL and CML analyses (Figures 2 and 5, Figure S1). None of the previous phylogenomic studies on Adephaga

recovered this clade, though ASTRAL analyses often recovered a similar clade but with Matinae included in one study (Vasilikopoulos et al., 2021) as did concatenated maximum likelihood analysis in another (Gustafson et al., 2020). There is no clear morphological synapomorphy from either adult or larvae supporting the relationship, though features may be discovered in the future. Within the group, analyses contrast regarding the relationship between copelatines and the other two couplets (see Section 4 below).

8. Hydroporinae

Hydroporinae is the largest dytiscid subfamily in numbers of species and diversity of morphology in the Dytiscidae and includes numerous tribes which are largely phylogenetically well-supported (Michat et al., 2017; Miller, 2001; Miller & Bergsten, 2014a). The concept of the subfamily has changed relatively little in taxonomic composition or morphological description since its establishment. It has several morphological synapomorphies including: (1) declivity of the prosternal process (the medial portion of the prosternum and the apical portion of the prosternal process in distinctly different planes), (2) pseudotetramerous pro- and mesotarsi in both sexes (a few groups with pro- and mesotarsomeres IV secondarily more clearly visible), (3) the scutellum not visible with the elytra closed (convergent with Laccophilini and Aubehydrini and potentially secondarily visible in *Celina* Aubé and few other species) and (4) the male median lobe bilaterally symmetrical (convergent with Dytiscinae and secondarily asymmetrical in several hydroporine taxa) and, in larvae, (5) presence of a nasale and obliquely oriented mandibles (de Marzo & Nilsson, 1988; Michat et al., 2017).

Here, Hydroporinae is maximally supported as monophyletic (Figures 2 and 5, Figure S1), consistent with many previous studies. Although Hydroporinae is well-supported and most tribes within Hydroporinae are also well-support (Michat et al., 2017, Miller, 2001, Miller & Bergsten, 2014a), relationships among the tribes are not as clear or well-supported (see Problems below). Two large and important clades within Hydroporinae are recovered with confidence however and discussed next.

9. Clade D: 'Higher Hydroporinae': (Hydroporini, Hyphydrini, Hygrotini, Bidessini and Vatellini), excluding the 'Plesiotypic Hydroporinae', (Methlini, Laccornini, Pachydrini, Hydrovatini and Laccornellini)

As found in early cladistic analyses (e.g. Wolfe, 1985, 1988), Hydroporinae includes two larger clusters of tribes. The first is a paraphyletic assemblage including Laccornini, Methlini, Laccornellini, Pachydrini and Hydrovatini (see Problem 5 below). These are early-diverging lineages to a large diverse clade including the other tribes of Hydroporinae. The other tribes, the 'Higher Hydroporinae' include Hydroporini, Hyphydrini, Hygrotini, Bidessini and Vatellini. The 'Higher Hydroporinae' together form an extremely diverse and well-supported clade (Figures 2 and 5, Figure S1), which together make up over 40% of species diversity in Dytiscidae (Nilsson & Hájek, 2024). Morphologically, they are characterized especially by loss of the female laterotergite (Miller, 2001), though these are also absent in Methlini, Laccornellini and the highly modified female genitalia of Hydrovatini (Miller et al., 2006). Members of the group also have the metacoxal lobes reduced in various ways or absent (Miller, 2001), whereas most of the 'Plesiotypic Hydroporinae' have these lobes larger and rounded, similar to non-Hydroporinae Dytiscidae. However, Pachydrini and Hydrovatini also have the lobes entirely or somewhat reduced.

10. Clade E: Hydroporini + Hygrotini + Hyphydrini

These three speciose tribes are superficially somewhat similar to each other, but there are, as yet, no clear known morphological synapomorphies uniting them. They have been consistently resolved near each other in other analyses based mainly on molecular data or homoplasious morphological data (Michat et al., 2017; Miller, 2001; Miller & Bergsten, 2014a, 2023a) and are well-supported as a clade here (Figures 2 and 5, Figure S1). Relationships among them are problematic; however, see Problem 7 below.

Problems remaining in the phylogenetic backbone of Dytiscidae

Based on our analyses, there are seven remaining problems in the phylogenetic backbone of Dytiscidae, partly discussed above from a general and theoretical method perspective. Two are at the level of subfamilies and five at the level of tribes. Problems are here defined as having conflicting resolutions between the ASTRAL and CML analyses, or having a support of less than 0.95/95 in either analysis, using the full dataset. We can further divide the problems into two categories: methodological inconsistencies and unresolved nodes due to insufficient data sampling (Figure 5).

Problem 1 is a basal trichotomy within Dytiscidae: Laccophilinae + (Coptotominae + Lancetinae) + other Dytiscidae. Problem 2 is the position of Copelatinae in relation to (Agabinae + Colymbetinae) and (Cybistrinae + Dytiscinae). Problem 3 is whether Dytiscini including *Hyderodes* is mono- or paraphyletic. Problem 4 is the position of *Notaticus* in relation to Hydaticini and Eretini + Aciliini. We consider these four problems to be methodological inconsistencies at this point with data in favour of the ASTRAL resolution (Figures 3 and 5, Table 1). Although not impossible, it is not clear that additional data sampling (genes or taxa) will solve the conflict. The basal trichotomy is the most uncertain of these as local posterior probability was low in the ASTRAL analysis but given the extremely short internode this may be unavoidable (Figure 2). The position of Copelatinae is the second most uncertain as gCF of the resolution is very low indeed despite maximum support in the ASTRAL analysis. FcLM support is consistently in favour of the ASTRAL resolution in both cases unless the

dataset is severely reduced in taxa or genes (Figures S4 and S5, Table 1).

Problem 5 is the lack of resolution among early-branching Hydroporinae lineages, including Methlini, Laccornini, Pachydrini, Hydrovatini and Laccornellini (this last tribe was not included in our analyses but is evidently in this part of the phylogeny based on previous studies (Miller & Bergsten, 2014a, 2023a)). Problem 6 is the position of Bidessini and Vatellini, which are resolved together with the Hydroporini + Hygrotini + Hyphydrini clade and not with the early Hydroporinae lineages, but relationships among these three clades are uncertain. Problem 7 is the Hydroporini + Hygrotini + Hyphydrini problem, including whether Hydroporini, with four subtribes (Miller & Bergsten, 2014a, 2023a), is monophyletic or not. We consider these three problems to be unresolved nodes due to insufficient data sampling at this point. Several tribes related to problems 5–6 were represented by single terminals with poor locus occupancy (Vatellini, Pachydrini, Hydrovatini) or were not represented at all (Laccornellini). Problem 7 relates to a very diverse clade, which even at the genus level was inadequately sampled to address the phylogenetic relationship. This inadequacy was illustrated by Siettitiina, which failed to group with Hydroporina + Deronectina when one of the three representative terminals was removed (Figure S14). The fact that Sternopriscina did not group with the remaining Hydroporini could very well be due to the fact that only two out of 11 described genera were represented. Hyphydrini was likewise only represented by two out of 14 recognized genera. Sampling additional strategic taxa, and for problems 5–6 also more loci for existing terminals, will be crucial to resolve these nodes.

Five of the seven problems are trichotomy problems, each limited to three possible resolutions. If we assume the monophyly of Hydroporini, as in Villastrigo et al. (2021), Problem 6 also becomes a trichotomy problem (though we discourage such an assumption based on current analyses) leaving only the early Hydroporinae lineages as a more complicated issue. Although phylogeneticists generally strive to interpret evolution as a strictly bifurcating process, there is certainly the real possibility of a fourth 'resolution' to a trichotomy problem and that is what is called a 'hard polytomy' (Hoelzer & Meinick, 1994; Whitfield & Lockhart, 2007). It has been argued, for instance, that the earliest diversification of Neoaves immediately following the K-Pg mass extinction is an empirical example of a hard polytomy containing eight near-simultaneous and 'unresolvable' speciation events (Suh, 2016). While the problems within Hydroporinae first and foremost require testing with additional taxon sampling, and the problems within Dytiscinae are certainly solvable as branch lengths are not short, the basal trichotomy in Dytiscidae and the Copelatinae problem are potential candidates of hard polytomies, especially the former. A hard polytomy is a zero-length branch and it is clear from our analyses that any resolution of the basal trichotomy contains an exceedingly short internode (Figure 2, also see Vasilikopoulos et al., 2021). A zero-length branch predicts equal proportions of gene trees in favour of each of the three possible resolutions by chance (Slowinski, 2001). The basal trichotomy in Dytiscidae fulfills this prediction rather comfortably as measured with the gene concordance factor (3.31, 3.19 and 4.27, respectively, for the three resolutions). The ASTRAL resolution of the Copelatinae problem, as discussed above, is also a very short internode with a gene concordance factor of 0.91 (on a scale from 0 to 100) meaning that >99% of the gene trees in the dataset are either uninformative about, or in conflict with, the resolution (uninformative here defined as either undecisive or decisive but one or more of the four target clades are paraphyletic, gDFp sensu Minh, Schmidt, et al., 2020). Gene concordance factors may be a rather crude measure in these cases however, since most gene trees are regarded as uninformative. An alternative and more powerful approach making more efficient use of the data is the quartet-based likelihood mapping method (Strimmer & von Haeseler, 1997). A zero-length branch, or star phylogeny, is expected to give a graphical representation of quartet probability vectors concentrated in the centre of the equilateral triangle and extending with equal proportions into the three corners (Strimmer & von Haeseler, 1997). Instead, what we see is a pattern of a dataset informative on both the basal ingroup trichotomy node and the Copelatinae problem node (centre receives but 0.1% in each case), with one resolution clearly favoured over the other two (Figure 3). This speaks against a hard polytomy interpretation in both cases (but see Nieselt-Struwe & von Haeseler, 2001 regarding likelihood mapping's over-liberal tendency to suggest a resolved tree even when the underlying model is a star-tree).

We provide further comments below for each of the seven problems

Problem 1. Early Dytiscidae Trichotomy Problem (Trichotomy members: Laccophilinae, Lancetinae + Coptotominae, all other Dytiscidae)

The basal trichotomy in dytiscids consists of Laccophilinae + (Lancetinae + Coptotominae) + all other Dytiscidae (Figure 4), and this seems to be a difficult problem to resolve. This is a relatively new arrangement in the history of dytiscid phylogenetics. In particular, Matinae is not resolved among these earliest lineages (Figures 2 and 5) as found in earlier morphological and combined analyses (Désamoré et al., 2018; Miller, 2001; Miller & Bergsten, 2014a, 2023a). Michat et al. (2017) recovered Laccophilinae as sister to the remaining Dytiscidae based on larval characters but with Lancetinae + Coptotominae further from the root. Others have inferred Hydroporinae (Michat & Torres, 2009; Nilsson, 1988; Ribera, Hogan, & Vogler, 2002), Hydrodytinae + Hydroporinae (Gustafson et al., 2020) or Hydrodytinae + Laccophilinae (Balke et al., 2004) as sister to the remaining Dytiscidae. Several studies have resulted in a larger polytomy at the root node (Ribera et al., 2008). But the latest phylogenomic analyses have, like us, recovered these three lineages at the earliest ingroup branch with various poorly supported resolutions or configurations that are analysis-dependent (Baca et al., 2021; Vasilikopoulos et al., 2021). Our two analysis methods likewise result in conflicting resolutions (Figure 2, Figure S1). The node-filtered ASTRAL tree (Figure S4) agrees with the concatenated ML tree (Figure S1) in placing Lancetinae + Coptotominae as sister to the remaining dytiscids as found by Vasilikopoulos et al. (2021), but

the FcLM analysis supports the main ASTRAL resolution which places Laccophilinae as sister to the remaining Dytiscidae (Figures 2 and 3). In contrast to weak support in the ASTRAL tree (Figure 2) and in Vasilikopoulos et al. (2021), the CML tree is maximally supported (Figure S1). We consider this resolution still not settled with confidence but with a preference in our data for the ASTRAL resolution placing Laccophilinae as sister to all other dytiscids (Figures 3 and 5, Figure S4, Table 1). There are several larval characters uniting Laccophilinae with Lancetinae + Coptotominae (Michat & Alarie, 2013; Michat & Torres, 2009; Nilsson, 1988), and that resolution of the basal trichotomy was found in the phylogenomic study by Baca et al. (2021) under some settings. It is difficult at this time to determine whether these features are plesiomorphies, apomorphic, or convergent relative to these groups, however. One example is the 'four-peg' pattern of the lamellae clypeales in first instar larvae as discussed by Michat et al. (2017). Ruhnau and Brancucci (1984) suggested this pattern was derived three times in Dytiscidae (Laccophilinae, Lancetinae and Coptotominae), whereas Nilsson (1988) and Michat and Alarie (2013) suggested the 'four-peg' pattern was synapomorphic for the three groups. However, the 'four-peg' pattern appears to also be present in numerous other families of Adephaga (Alarie & Bilton, 2005; Alarie, Short, et al., 2011; Michat et al., 2017) suggesting the feature may be plesiomorphic in Dytiscidae.

Problem 2. Copelatinae problem
(*Trichotomy members*: *Copelatinae*, *Agabinae + Colymbetinae*, *Dytiscinae + Cybistrinae*)

Copelatinae has been notoriously difficult to place among other subfamilies prior to the molecular era. The group was long part of Colymbetinae as a tribe (e.g. Sharp, 1882) until Miller (2001) elevated the group to its own subfamily. Unique among dytiscids (except for *Hydrotrupes*), Copelatine larvae lack the characteristic mandibular channel of other Dytiscidae and ingest solid food, which has compelled several authors to place Copelatinae as the earliest sister lineage to the remaining dytiscids (de Marzo & Nilsson, 1986; Ruhnau & Brancucci, 1984). However, most analyses since, including molecular, morphological and combined, have not supported this, but they have also failed to reach a consensus about copelatine placement in the phylogeny (Baca et al., 2021; Balke et al., 2004; Désamoré et al., 2018; Gustafson et al., 2020; Michat et al., 2017; Michat & Torres, 2009; Miller & Bergsten, 2014a, 2023a; Nilsson, 1988; Ribera et al., 2008; Ribera, Hogan, & Vogler, 2002; Vasilikopoulos et al., 2021).

Here we found that the subfamily belongs in a clade with the two couplets Agabinae + Colymbetinae and Dytiscinae + Cybistrinae (Figure 2 and Figure S1). Our two types of analyses each placed Copelatinae here but with well-supported conflicting resolutions (Figures 2 and 5, Figure S1). FcLM analyses favoured the ASTRAL resolution where Copelatinae is sister to remaining four subfamilies across all analyses (except with the most reduced dataset D10 from phylogenetic informativeness filtering) albeit with non-significant support also for the other two resolutions (Figure 3 and Figure S5, Table 1). This position is contrary to other recent phylogenomic studies which place Copelatinae as sister to Hydroporinae + Hydrodytinae (Baca et al., 2021) or sister to Matinae + Hydrodytinae + Hydroporinae (Vasilikopoulos et al., 2021). However, our conclusion agrees with studies of larval characters supporting the affinity of Copelatinae with especially Agabinae (Michat & Torres, 2009; Nilsson, 1988). Studying primary setae and pores on the legs of first instar larvae of Dytiscidae, Nilsson refuted the early origin of Copelatinae and stated that '…with respect to my results *Copelatus* Erichson is almost identical with most Agabini…' (Nilsson, 1988: 2292). Scrutinizing the various supplementary analyses in Vasilikopoulos et al. (2021) and Baca et al. (2021) it is actually clear that all their ASTRAL analyses also supported affinity of Copelatinae with the larger clade Agabinae + Colymbetinae and Dytiscinae + Cybistrinae, but then also with Matinae included (Baca et al., 2021, Vasilikopoulos et al., 2021).

Problem 3. Hyderodes Problem
(*Trichotomy members*: *Dytiscus*, *Hyderodes*, *other Dytiscinae*)

Whether the Holarctic *Dytiscus* and Australian *Hyderodes* are together monophyletic or whether *Hyderodes* is in a clade with other Dytiscinae has been surprisingly hard to resolve. *Hyderodes* was originally placed together with *Dytiscus* by Sharp (1882). Phylogenetic analyses based on adult morphology did not support such a clade (Miller, 2000, 2001), prompting Miller (2000) to erect the tribe Hyderodini for *Hyderodes*. However, analyses based on larval characters supported the monophyly of *Dytiscus* and *Hyderodes* (Alarie, Michat, & Miller, 2011; Michat et al., 2017) as did a combined analysis of morphological data and nine genes (Miller & Bergsten, 2014a, 2023a), with strong support in all three studies. This prompted Miller and Bergsten (2014a) to synonymize Hyderodini with Dytiscini. The phylogenomic analysis by Vasilikopoulos et al. (2021) included both taxa and failed to recover a monophyletic Dytiscini in the preferred CML analysis. Support was weak, however, for the placement of *Hyderodes* as sister to Hydaticini + Aciliini + Eretini, and a monophyletic Dytiscini was recovered in their supplementary ASTRAL analyses. Morphological features uniting the two groups historically were mainly symplesiomorphies within Dytiscinae, and few clear synapomorphies have been discovered uniting these two genera in adults or larvae (Alarie, Michat, & Miller, 2011; Michat et al., 2017; Miller, 2000, 2001; Miller & Bergsten, 2014a, 2023a).

In our analyses, we find the ASTRAL (Figure 2) and CML (Figure S1) methods disagree. ASTRAL recovers a monophyletic *Dytiscus* + *Hyderodes* with strong support (Figure 2), whereas CML places *Hyderodes* in a clade with Aubehydrini + Hydaticini + Eretini + Aciliini, also with strong support (Figure S1). The FcLM test, however, delivered the strongest preference for one solution over the other two among all four problems analysed. A monophyletic Dytiscini received 77% in support compared with 22% for the other two combined (Figure 3) and the support remained strong across all gene filtering exercises (Figure S6).

The conflicting placement of *Hyderodes* in the ASTRAL (Figure 2) vs. CML (Figure S1) analyses is similar to the alternating position of *Notaticus* (see below). Since they occur proximally in the tree and represent some of the longest terminal branches in the phylogeny, we performed a long-branch extraction test (Bergsten, 2005; Siddall & Whiting, 1999) to evaluate if their relative placements affect each other. Their relative positions in the ML tree were unaffected by the presence or absence of the other (Figures S2 and S3). When the 54 taxa from the exon capture study of Vasilikopoulos et al. (2021) were excluded (incl. one of the two *Dytiscus* terminals), FcLM support evened out between the ASTRAL and CML analyses (Table 1). O'Connor et al. (2010) criticized the effectiveness of long-branch extraction to detect long-branch attraction and we still suspect that LBA is at play in this region of the tree affecting both CML and FcLM with the reduced taxa datasets (Table 1).

Problem 4. Notaticus Problem
(*Trichotomy members*: Notaticus, *Hydaticini*, *Aciliini* + *Eretini*)

The Neotropical genus *Notaticus*, the only genus in the tribe Aubehydrini (formerly its own subfamily Aubehydrinae) has an odd combination of characters including the scutellum concealed with the elytra closed, pentamerous pro- and mesotarsi, and unlobed metatarsomeres. A cladistic analysis by Miller (2000) dispositively placed *Notaticus* well within Dytiscinae, as did analyses of *Notaticus* larvae (Alarie, Michat, & Miller, 2011; Michat et al., 2017; Michat & Alarie, 2009; Miller, Alarie, & Whiting, 2007).

Placement of *Notaticus* well within Dytiscinae is confirmed by our analyses (Figures 2 and 5, Figure S1). However, *Notaticus* is part of a difficult trichotomy together with Hydaticini and Eretini + Aciliini. Our CML analysis recovered Aubehydrini as sister to Hydaticini + (Eretini + Aciliini) (Figure S1) in agreement with Miller (2000), Alarie, Michat, and Miller (2011) and Michat et al. (2017). The ASTRAL tree, however, recovered Aubehydrini as sister group to Eretini + Aciliini (Figure 2) in agreement with Miller et al. (2009), Michat and Alarie (2009) and Miller and Bergsten (2014a, 2023a). The latter resolution unifies three tribes that share a larval body shape adapted to a more active, nektonic lifestyle (Miller, Alarie, & Whiting, 2007). It is also supported by the FcLM analyses (Figure 3 and Figure S7) and by more convincing discrete larval characters than the alternative resolution (see Michat & Alarie, 2009). With the exclusion of a large set of taxa, FcLM support switched to the CML resolution (Table 1) but as noted above, we suspect LBA is at play in this region of the tree affecting both *Notaticus* and Dytiscini, even if when subjected to a long-branch attraction test their positions were not affected (Figures S2 and S3).

Problem 5. Early Hydroporinae Lineages Problem

Within Hydroporinae, there are several named groups that have been long recognized as characterized by plesiomorphies beginning with Wolfe (1985, 1988) who called them the 'plesiotypic hydroporines', including *Laccornis* Gozis (at the time including species now in *Laccornellus* Roughley and Wolfe), Methlini (*Celina* and *Methles*), *Canthyporus* Zimmermann and Hydrovatini (*Hydrovatus* and *Queda*). Based on adult morphology he found *Laccornis* resolved as sister to all other Hydroporinae with the clade Methlini + Hydrovatini next, and *Canthyporus* (later in Laccornellini with the genus *Laccornellus*) next (Wolfe, 1985, 1988). Based on larval data, *Laccornis* was found to be characterized by plesiomorphies by Nilsson (1988). Laccornini and Laccornellini were later separated, but both were recognized as early-branching by Wolfe and Roughley (1990). *Laccornis* and Methlini were found by Biström et al. (1997) in a similar early-branching configuration based on adult morphology (*Hydrovatus* not included), with Pachydrini placed in a basal position as well. Additional, similar conclusions based on adult and larval morphology were found by Miller (2001), Michat and Torres (2008), Alarie and Michat (2007a, 2007b), Michat (2006), Alarie and Harper (1990), Alarie (1991), Miller (2003) and Miller and Bergsten (2014a, 2023a). Some morphological characters supporting the association of these tribes within the subfamily are the plesiomorphic presence of a female laterotergite in Laccornini and Pachydrini and, in most of the tribes, well-developed metacoxal lobes, though these are absent in Pachydrini and reduced in Hydrovatini (Wolfe, 1985, 1988).

Methlini has been regarded as part of the 'Plesiotypic Hydroporinae' lineages for some time (Miller, 2001; Wolfe, 1985, 1988). Notably, it has even been regarded as a subfamily within Dytiscidae based in large part on the plesiomorphic presence in *Celina* of a visible scutellum with the elytra closed (Falkenström, 1938; Guignot, 1936; Guignot, 1959; Pederzani, 1995; van den Branden, 1885). However, *Celina* and *Methles* are clearly closely related (Figure 2 and Figure S1), and in *Methles* the scutellum is concealed with the elytra closed. Michat et al. (2017) also found Methlini sister to all other Hydroporinae based on a large larval morphology dataset. In contrast, Miller and Bergsten (2014a, 2023a) and Désamoré et al. (2018) recovered Methlini nested higher up within Hydroporinae, even polyphyletic in the absence of morphological data (Désamoré et al., 2018).

Based on several phylogenetic analyses, Laccornini has been found to be the sister to all other Hydroporinae (Miller, 2001; Miller & Bergsten, 2014a, 2023a; Wolfe, 1985, 1988) with Laccornellini as sister to other Hydroporinae except Laccornini, and Methlini closely, but somewhat ambiguously, related to these groups (e.g. Miller & Bergsten, 2014a, 2023a). Based on similarly acuminate posterior apices of the body, Wolfe (1985, 1988) regarded Methlini and Hydrovatini as related. However, *Queda* (Hydrovatini) lacks acuminate posterior apices, placing that relationship in doubt (Miller, 2001, and see above).

Pachydrini and Hydrovatini are quite distinctive. Pachydrini (the genera *Pachydrus* Sharp and *Heterhydrus* Fairmaire) were originally in or near Hyphydrini (e.g. Sharp, 1882) until Biström et al. (1997) questioned that relationship, though Miller (2001) placed these groups back together albeit with some reservations. A more comprehensive analysis of dytiscid larvae found it to also be close to Hyphydrini (Michat et al., 2017). Hydrovatini were regarded as part of the Plesiotypic Hydroporinae by Wolfe (1985, 1988) who based this in part on

the similarly apically acuminate body form in *Celina* and *Hydrovatus* Motschulsky. However, as noted above, this is a problematic character. Miller (2001) found Hydrovatini sister to Hygrotini, and Miller and Bergsten (2014a, 2023a) found Hydrovatini sister to Pachydrini, and this clade sister to Hygrotini. Together with *Methles*, which was separated from *Celina* in the pure molecular analysis by Désamoré et al. (2018), Hydrovatini and Hygrotini were recovered together.

Here we find strong support for the several tribes of Plesiotypic Hydroporinae in an unresolved or paraphyletic grade separate from a maximally supported clade of tribes in the 'Higher Hydroporinae' (see above, Figures 2 and 5, Figure S1). Support for a clade of Hydroporinae to the exclusion of Methlini (Figure 2 and Figure S1) is close to maximal in the CML analysis (Figure S1) and with ASTRAL once the rogue *Pachydrus* terminal is excluded (Figure S21). Recognition of the group as a subfamily of Dytiscidae or a tribe within Hydroporinae is a subjective decision. However, support for the relationship is potentially still somewhat questionable until the other plesiotypic lineages are more properly sampled, and removing the group from Hydroporinae would make the morphological definition of the rest of the subfamily difficult, so we retain the group as a tribe within the subfamily. Similarly, although Pachydrini and Hydrovatini are found either as a clade or grade among other 'Plesiotypic Hydroporine', their exclusion from the 'Higher Hydroporinae' is maximally supported by both types of analyses (Figures 2 and 5, Figure S1). Placement of Pachydrini among the 'Plesiotypic Hydroporinae' is supported by the presence of distinct gonocoxae in females (Miller, 2001). Supporting Wolfe (1985, 1988), Hydrovatini is among the 'Plesiotypic Hydroporinae' based on this analysis (Figures 2 and 5, Figure S1). A close relationship between Pachydrini and Hydrovatini was also recovered in the CML analysis by Vasilikopoulos et al. (2021). Relationships among the tribes of the 'Plesiotypic Hydroporinae' remain one of the most perplexing problems remaining within Dytiscidae, and a focused genomic sampling effort in this part of the tree (especially improved taxon and loci representation of Pachydrini, Hydrovatini and Laccornellini) should be rewarding.

Problem 6. Bidessini and Vatellini Problem

No compelling conclusions about the placement of Bidessini and Vatellini have been proposed in previous analyses. Each of these tribes is quite distinct and really different from other groups of derived Hydroporinae. Vatellini has a large number of highly unusual apomorphies for Dytiscidae (Miller, 2005). Bidessini includes about 16% of the species of Dytiscidae (Miller & Bergsten, 2016) and has several very distinctive synapomorphies within the family (Miller et al., 2006). Bidessini was historically associated with Hyphydrini and Pachydrini based on the metacoxae connate with the basal abdominal sternites (Sharp, 1882), but that has not been generally supported with more data (Miller, 2001, Miller & Bergsten, 2014a, 2023a). Vatellini has been recovered in various positions including as sister to Hydroporini (Miller, 2001), as sister to a large clade of various Hydroporinae (Miller & Bergsten, 2014a, 2023a), as sister to Hyphydrini (Désamoré et al., 2018) or to Pachydrini + Hyphydrini (Michat et al., 2017).

In our analyses, we find Bidessini + Vatellini maximally supported as sister to Clade E within Clade D, either as a clade or possibly consecutively branching (Figures 2 and 5, Figure S1, see above). We also consistently recover Bidessini + Vatellini (Figure 2, Figure S1), though support is very weak in the ASTRAL analysis (Figure 2). A sister group relationship between these two highly distinctive and unusual groups within Dytiscidae is interesting but needs further confirmation. Interestingly, Vatellini and Bidessini were recovered in close approximation in one previous analysis but then also included Hyphydrini in the clade (Désamoré et al., 2018). Vatellini was represented by a single terminal with poor loci occupancy (6%) and its phylogenetic position in relation to both clade E and Bidessini needs testing with improved sampling of taxa and genes.

Problem 7. Hyphydrini-Hygrotini-Hydroporini Problem (Including problems among subtribes in Hydroporini)

The three groups Hydroporini (with four distinctive subtribes), Hygrotini and Hyphydrini have many superficial similarities. They lack the features of the 'Problem 5 taxa' and the very unique features of Vatellini and Bidessini. But relationships among the groups are ambiguous. Within Hydroporini, there are four distinctive but difficult groups based on inconspicuous morphology that Miller and Bergsten (2014a) recognized as formal subtribes. These groups only inconsistently resolve together as monophyletic (Miller & Bergsten, 2014a) in relation to each other, Hygrotini and Hyphydrini. They were not resolved as monophyletic by Miller (2001), Michat et al. (2017) or Désamoré et al. (2018). Also, notably, Pachydrini was historically placed with Hyphydrini (e.g. Sharp, 1882) but excluded by Biström et al. (1997) with some support for this conclusion presented by Miller (2001). Pachydrini are here resolved among the Plesiotypic Hydroporinae (see above).

Hyphydrini + Hygrotini + Hydroporini is here maximally supported as monophyletic (Clade E, Figures 2 and 5, Figure S1, see above). Hygrotini and Hyphydrini are also maximally supported as monophyletic (Figure 2, Figure S1). Hydroporini is, however, more problematic with relationships among the four subtribes and Hyphydrini and Hygrotini not clear (Figures 2 and 5, Figure S1, Miller & Bergsten, 2014a, 2023a). An impressively sampled molecular phylogeny of Hydroporini was presented by Villastrigo et al. (2021). Monophyly of Hydroporini was not tested, but using a clock model to root the tree they recovered Siettitiina as sister to remaining three subtribes followed by Deronectina as sister to Sternopriscina + Hydroporini, all maximally supported in the bayesian analysis. Their supplementary maximum likelihood analysis however, rooted on Siettitiina, recovered instead Deronectina + Hydroporina as sisters (Villastrigo et al., 2021) which agrees with our ASTRAL and CML results (Figure 2, Figure S1). Our ASTRAL analyses also support a clade with Siettitiina + (Deronectina + Hydroporina) (Figure 2), a clade that was lost when low-occupancy terminals were removed along with one of the three

representatives of Siettitiina (Figure S14). To resolve clade E and properly test the naturalness of Hydroporini an improved taxon sampling for especially Hyphydrini, Sternopriscina and Siettitiina is necessary.

Family group classification of Dytiscidae

Our results support the following higher-level classification of the family Dytiscidae (Figure 5).

Agabinae Thomson
Agabini Thomson
Hydrotrupini Roughley
Platynectini Tousssaint & Balke, **new tribe**
Colymbetinae Erichson
Copelatinae Branden
Coptotominae Branden
Cybistrinae Sharp
Dytiscinae Leach
Aciliini Thomson
Aubehydrini Guignot
Dytiscini[1] Leach
Eretini Crotch
Hyaticini Sharp
Hydrodytinae Miller
Hydroporinae Aubé
Bidessini Sharp
Hydroporini[2] Aubé
Deronectina Galewski
Hydroporina Aubé
Siettitiina Smrž
Sternopriscina Branden
Hydrovatini Sharp
Hygrotini Portevin
Hyphydrini Gistel
Laccornellini Miller and Bergsten
Laccornini Wolfe and Roughley
Methlini Branden
Pachydrini Young
Vatellini Sharp
Laccophilinae Gistel
Agabetini Branden
Laccophilini Gistel
Lancetinae Branden
Matinae Branden

[1]Dytiscini (*Dytiscus* + *Hyderodes*) was not recovered as monophyletic by CML analyses, but ASTRAL, GCF and FcLM analyses all support a monophyletic Dytiscini (see Section 4).

[2]Hydroporini (*Sternopriscina* + *Deronectina* + *Siettitiina* + *Hydroporina*) was recovered as paraphyletic (CML) or unresolved (ASTRAL), however neither the taxon sampling nor the support of paraphyly is sufficient from this analysis to change the current status (see Section 4).

CONCLUSIONS

The current possibility of assembling large genomic datasets of thousands of loci has immensely improved the ability to resolve deeper backbone nodes in the tree of life. However, phylogenetic accuracy does not lean on the number of genes alone. Sufficient taxon sampling and appropriate method and model choice for data analyses remain critical also in the age of phylogenomics (Bernot et al., 2023; Steenwyk et al., 2023; Young & Gillung, 2020). The genomic-level taxon sampling for Dytiscidae has been building up from several phylogenomic studies mainly with a focus on Adephaga (Gustafson et al., 2020; McKenna et al., 2019; Vasilikopoulos et al., 2019; Vasilikopoulos et al., 2021). Given the diversity of Dytiscidae and previous studies' focus on more inclusive clades, these studies were limited in their conclusions about the family. With a deeper sampling, we were able to convincingly and cross-methodologically resolve all but two inter-subfamily nodes in Dytiscidae, with our data clearly favouring one resolution also for the remaining two nodes (Figure 5). These configurations are generally more congruent with transcriptomic and exon-capture studies (Vasilikopoulos et al., 2021) than with previous UCE-based studies (Baca et al., 2021; Gustafson et al., 2020) which may be related to respective studies's taxon sampling within Dytiscidae. Hypotheses based on morphological evidence (e.g. Michat et al., 2017; Miller, 2000, 2001) are largely supported, particularly in resolving named clades at the family group rank, but we also provide resolutions of nodes where morphological evidence has failed to reach a consensus about relationships among these clades. Previous phylogenomic studies have also largely preferred concatenated maximum likelihood analyses over conflicting results from species tree methods (Baca et al., 2021, Gustafson et al., 2020, Vasilikopoulos et al., 2021), but we were able to reconcile for the first time the methodological approaches in several cases of inter-subfamily relationships.

A less scrutinized aspect of data gathering choice relates to the future value of datasets. In contrast to the recent phylogenomic studies of Adephagan evolution that used reduced representation methods, we engaged with whole genome sequencing, weighing the added future value against sequencing costs. Baca et al. (2021) capitalized on the possibility of in silico extraction of homologous UCE loci from UCE-enriched, transcriptomic and genomic raw reads. Unsurprisingly, genomic reads recovered the greatest proportion of loci, illustrating the superior re-harvesting potential of whole genome sequencing data for purposes not even considered when first generated (Baca et al., 2021). We hope our effort to resolve the backbone phylogeny of Dytiscidae will not only spur forthcoming studies tackling remaining challenges, but also that data generated will be reused for many novel purposes in the future.

AUTHOR CONTRIBUTIONS

Johannes Bergsten: Conceptualization; investigation; formal analysis; validation; resources; writing – original draft; methodology. **Johan A. A. Nylander:** Data curation; formal analysis; writing – original draft; validation; methodology. **Oscar E. Ospina:** Data curation;

writing – review and editing; methodology. **Alan R. Lemmon:** Resources; writing – review and editing; methodology. **Kelly B. Miller:** Conceptualization; investigation; funding acquisition; writing – original draft; resources; data curation; project administration; visualization; methodology.

ACKNOWLEDGEMENTS

The authors acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. Special thanks for bioinformatic support to Remi-André Olsen at the SciLifeLab and Martin Sköld at the Swedish Museum of Natural History. Miller thanks G. Courtney, D. Post, A.E.Z. Short, R. Sites and G.W. Wolfe for help with acquiring specimens or finding collecting localities. Portions of this project were supported by grants to KB Miller from the US National Science Foundation (#DEB–1353426, #DEB–0845984, #DEB–0816904, #DEB–0515924, #DEB–0329115 and #DEB–0073088).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Raw data for new taxa in the form of untrimmed reads in FastQ format is made available at the European Nucleotide Archive (ENA) (accession numbers in Table S1). The concatenated dataset (aligned and trimmed) in fasta format, a partition file defining the separate genes and a gene code translation table are made available at SciLifeLab data repository https://www.scilifelab.se/data/repository/ powered by Figshare. DOI number for collection of files: https://doi.org/10.17044/scilifelab.28868468.v1.

ORCID

Johannes Bergsten https://orcid.org/0000-0002-6153-4431

REFERENCES

Abellán, P. & Ribera, I. (2011) Geographic location and phylogeny are the main determinants of the size of the geographical range in aquatic beetles. *BMC Evolutionary Biology Supplement*, 11(1–15), 344. Available from: https://doi.org/10.1186/1471-2148-11-344

Alarie, Y. (1991) Primary setae and pores on the cephalic capsule and head appendages of larval Hydroporinae (Coleoptera: Dytiscidae: Hydroporinae). *Canadian Journal of Zoology*, 69, 2255–2265. Available from: https://doi.org/10.1139/z91-316

Alarie, Y. (1998) Phylogenetic relationships of Nearctic Colymbetinae (Coleoptera: Adephaga: Dytiscidae) based on chaetotaxic and porotaxic analysis of head capsule and appendages of larvae. *Canadian Entomologist*, 130, 803–824. Available from: https://doi.org/10.4039/Ent130803-6

Alarie, Y. & Bilton, D.T. (2005) Larval morphology of Aspidytidae (Coleoptera: Adephaga) and its phylogenetic implications. *Annals of the Entomological Society of America*, 98, 417–430.

Alarie, Y. & Harper, P.P. (1990) Primary setae and pores on the last abdominal segment and the urogomphi of larval Hydroporinae (Coleoptera: Adephaga: Dytiscidae), with notes on other dytiscid larvae. *Canadian Journal of Zoology*, 68, 368–374. Available from: https://doi.org/10.1139/z90-052

Alarie, Y., Michat, C.M., Bergsten, J. & Hájek, J. (2023) Morphology of the larvae of *Rhantaticus congestus* (Klug, 1833) and phylogenetic comparison with other known Aciliini (Coleoptera: Dytiscidae: Dytiscinae). *Zootaxa*, 5380, 247–264. Available from: https://doi.org/10.11646/zootaxa.5380.3.3

Alarie, Y., Michat, C.M., Hendrich, L. & Watts, C.H. (2018) Larval description and phylogenetic placement of the Australian endemic genus *Barretthydrus* lea, 1927 (Coleoptera: Dytiscidae: Hydroporinae: Hydroporini: Sternopriscina). *The Coleopterists Bulletin*, 72, 639–661. Available from: https://doi.org/10.1649/0010-065X-72.4.639

Alarie, Y., Michat, C.M., Shaverdo, H. & Hájek, J. (2023) Morphology of the larvae of *Sandracottus femoralis* Heller, 1934, and *S. Mixtus* (Blanchard, 1843) and phylogenetic comparison with other known Aciliini (Coleoptera: Dytiscidae, Dytiscinae). *Zootaxa*, 5263, 301–334. Available from: https://doi.org/10.11646/zootaxa.5263.3.1

Alarie, Y. & Michat, M.C. (2007a) Phylogenetic analysis of Hydroporinae (Coleoptera: Dytiscidae) based on larval morphology, with description of first instar of *Laccornellus lugubris*. *Annals of the Entomological Society of America*, 100, 655–665. Available from: https://doi.org/10.1603/0013-8746(2007)100[655:PAOHCD]2.0.CO;2

Alarie, Y. & Michat, M.C. (2007b) Primary setae and pores on the maxilla of larvae of the subfamily Hydroporinae (Coleoptera: Adephaga: Dytiscidae): ground plan pattern reconsidered. *The Coleopterists Bulletin*, 61, 310–317. Available from: https://doi.org/10.1649/0010-065X(2007)61[310:PSAPOT]2.0.CO;2

Alarie, Y. & Michat, M.C. (2020) Larval morphology of Agabinae (Coleoptera: Dytiscidae): descriptions of three species of the subgenus *Agabus* s. str. Leach, 1817 with phylogenetic considerations. *Aquatic Invasions*, 41, 197–226. Available from: https://doi.org/10.1080/01650424.2020.1741646

Alarie, Y. & Michat, M.C. (2022) Larval morphology of *Ilybiosoma* crotch, 1873 (Coleoptera: Dytiscidae: Agabinae): description of five Nearctic species with phylogenetic considerations. *Aquatic Invasions*, 43, 99–129. Available from: https://doi.org/10.1080/01650424.2021.1975759

Alarie, Y., Michat, M.C. & Challet, G.L. (2017) Larval description and phylogenetic placement of the south African endemic genera *Coelhydrus* Sharp and *Darwinhydrus* Sharp (Coleoptera: Dytiscidae: Hydroporinae: Hyphydrini). *The Coleopterists Bulletin*, 71, 389–401. Available from: https://doi.org/10.1649/0010-065X-71-2.389

Alarie, Y., Michat, M.C. & Miller, K.B. (2011) Notation of primary setae and pores on larvae of Dytiscinae (Coleoptera: Dytiscidae), with phylogenetic considerations. *Zootaxa*, 3087, 1–55. Available from: https://doi.org/10.11646/ZOOTAXA.3087.1.1

Alarie, Y., Michat, M.C., Ranarilalatiana, T. & Bergsten, J. (2022) Larval morphology of the Madagascan endemic diving beetle genus Hovahydrus Biström, 1982 (Coleoptera: Dytiscidae) and phylogenetic comparison with other known Hyphydrini. *Zootaxa*, 5219(3), 227–246. Available from: https://doi.org/10.11646/zootaxa.5219.3.2

Alarie, Y., Michat, M.C. & Watts, C.H. (2021) Description of the mature larva of four species of the Australasian endemic genus *Sternopriscus* Sharp, 1882 (Coleoptera: Dytiscidae) with phylogenetic considerations. *Aquatic Insects*, 42, 197–215. Available from: https://doi.org/10.1080/01650424.2021.1919716

Alarie, Y., Michat, M.C. & Watts, C.H.S. (2019) Larval morphology of the Australian radiation Sternopriscina (Coleoptera: Dytiscidae, Hydroporinae): description and phylogenetic placement of the genus *Chostonectes* Sharp, 1882. *Aquatic Insects*, 40, 328–354. Available from: https://doi.org/10.1080/01650424.2019.1610570

Alarie, Y., Short, A.E.Z., García, M. & Joly, L. (2011) Larval morphology of Meruidae (Coleoptera: Adephaga) and its phylogenetic implications. *Annals of the Entomological Society of America*, 104, 25–36.

Alarie, Y., Watanabe, K. & Michat, C.M. (2023) The very rare Japanese endemic diving beetle *Japanolaccophilus niponensis* (Kamiya, 1939), (Coleoptera: Dytiscidae, Laccophilinae): larval morphology and phylogenetic comparison with other known Laccophilini. *Zootaxa*, 5285, 116–132. Available from: https://doi.org/10.11646/zootaxa.5285.1.4

Arnqvist, G. & Rowe, L. (2005) *Sexual conflict*. Princeton: Princeton University Press.

Austin, A., Guzik, M., Jones, K., Humphreys, W., Watts, C. & Cooper, S. (2023) The unique Australian subterranean Dytiscidae: diversity, biology, and evolution. In: Yee, D.A. (Ed.) *Ecology, systematics, and the natural history of predaceous diving beetles (Coleoptera: Dytiscidae)*. Cham: Springer International Publishing, pp. 401–425. Available from: https://doi.org/10.1007/978-3-031-01245-7_9

Baca, S.M., Gustafson, G.T., Alexander, A.M., Gough, H.M. & Toussaint, E.F.A. (2021) Integrative phylogenomics reveals a Permian origin of Adephaga beetles. *Systematic Entomology*, 46, 968–990. Available from: https://doi.org/10.1111/syen.12506

Balke, M. (2001) Biogeography and classification of new Guinean Colymbetini (Coleoptera: Dytiscidae: Colymbetinae). *Invertebrate Taxonomy*, 2001, 259–275. Available from: https://doi.org/10.1071/IT98008

Balke, M., Bergsten, J., Wang, L.-J. & Hendrich, L. (2017) A new genus and two new species of southeast Asian Bidessini as well as new synonyms for Oceanian species (Coleoptera, Dytiscidae). *Zookeys*, 647, 137–151. Available from: https://doi.org/10.3897/zookeys.647.11231

Balke, M., Hajek, J. & Hendrich, L. (2017) Generic reclassification of species formerly included in *Rhantus* Dejean (Coleoptera, Dytiscidae, Colymbetinae). *Zootaxa*, 4258, 91–100. Available from: https://doi.org/10.11646/zootaxa.4258.1.7

Balke, M., Ribera, I., Beutel, R., Viloria, A., Garcia, M. & Vogler, A.P. (2008) Systematic placement of the recently discovered beetle family Meruidae (Coleoptera: Dytiscoidea) based on molecular data. *Zoologica Scripta*, 37, 647–650. Available from: https://doi.org/10.1111/j.1463-6409.2008.00345.x

Balke, M., Ribera, I. & Beutel, R.G. (2005) The systematic position of Aspidytidae, the diversification of Dytiscoidea (Coleoptera, Adephaga) and the phylogenetic signal of third codon positions. *Journal of Zoological Systematics and Evolutionary Research*, 43, 223–242.

Balke, M., Ribera, I., Hendrich, L., Miller, M.A., Sagata, K., Posman, A. et al. (2009) New Guinea highland origin of a widespread arthropod supertramp. *Proceedings of the Royal Society Biological Sciences Series B*, 276, 2359–2367. Available from: https://doi.org/10.1098/rspb.2009.0015

Balke, M., Ribera, I. & Vogler, A.P. (2004) MtDNA phylogeny and biogeography of Copelatinae, a highly diverse group of tropical diving beetles (Dytiscidae). *Molecular Phylogenetics and Evolution*, 32, 866–880. Available from: https://doi.org/10.1016/j.ympev.2004.03.014

Barman, E.H., Michat, M.C., Alarie, Y. & Wolfe, G.W. (2014) A description of the first instar of *Hoperius planatus* fall, 1927 (Coleoptera: Dytiscidae: Colymbetinae: Colymbetini), with phylogenetic implications. *Coleopterists Bulletin*, 68, 321–330. Available from: https://doi.org/10.1649/0010-065X-68.2.321

Benetti, C.J., Short, A.E.Z. & Michat, M.C. (2019) *Hamadiana chapadensis*, a new genus and species of diving beetle from Brazil (Coleoptera, Dytiscidae, Laccophilinae, Laccophilini). *Zootaxa*, 4615(1), zootaxa.4615.1.10. Available from: https://doi.org/10.11646/zootaxa.4615.1.10

Bergsten, J. (2005) A review of long-branch attraction. *Cladistics*, 21, 163–193. Available from: https://doi.org/10.1111/j.1096-0031.2005.00059.x

Bergsten, J., Bilton, D.T., Fujisawa, T., Elliott, M., Monaghan, M.T., Balke, M. et al. (2012) The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology*, 61, 851–869. Available from: https://doi.org/10.1093/sysbio/sys037

Bergsten, J. & Miller, K.B. (2007) Phylogeny of diving beetles reveals a coevolutionary arms race between the sexes. *PLoS One*, 2, e522. Available from: https://doi.org/10.1371/journal.pone.0000522

Bergsten, J., Töyrä, A. & Nilsson, A.N. (2001) Intraspecific variation and intersexual correlation in secondary sexual characters of three diving beetles (Coleoptera: Dytiscidae). *Biological Journal of the Linnean Society*, 73, 221–232. Available from: https://doi.org/10.1111/j.1095-8312.2001.tb01359.x

Bernot, J.P., Owen, C.L., Wolfe, J.M., Meland, K., Olesen, J. & Crandall, K.A. (2023) Major revisions in Pancrustacean phylogeny and evidence of sensitivity to taxon sampling. *Molecular Biology and Evolution*, 40, msad175. Available from: https://doi.org/10.1093/molbev/msad175

Bertrand, H.P.I. (1972) *Larves et nymphes des Coléoptères aquatiques du globe avec tableaux de détermination des genres*, Vol. 78. Paris: E. Piallart, pp. 89–97.

Beutel, R.G. (1994) On the systematic position of *Hydrotrupes palpalis* Sharp (Coleoptera: Dytiscidae). *Aquatic Invasions*, 16, 157–164. Available from: https://doi.org/10.1080/01650429409361550

Beutel, R.G., Ribera, I., Fikáček, M., Vasilikopoulos, A., Misof, B. & Balke, M. (2020) The morphological evolution of the Adephaga (Coleoptera). *Systematic Entomology*, 45, 378–395. Available from: https://doi.org/10.1111/syen.12403

Bilton, D.T., McAbendroth, L., Bedford, A. & Ramsay, P.M. (2006) How wide to cast the net? Cross-taxon congruence of species richness, community similarity and indicator taxa in ponds. *Freshwater Biology*, 51, 578–590. Available from: https://doi.org/10.1111/j.1365-2427.2006.01505.x

Bilton, D.T., Toussaint, E.F.A., Turner, C.R. & Balke, M. (2015) *Capelatus prykei* gen. Et sp.n. (Coleoptera: Dytiscidae: Copelatinae) – a phylogenetically isolated diving beetle from the Western cape of South Africa. *Systematic Entomology*, 40, 520–531. Available from: https://doi.org/10.1111/syen.12128

Biström, O., Nilsson, A.N. & Wewalka, G. (1997) A systematic review of the tribes Hyphydrini Sharp and Pachydrini n. trib. (Coleoptera, Dytiscidae). *Entomologica Fennica*, 8, 57–82. Available from: https://doi.org/10.33338/ef.83921

Brinck, P. (1948) Coleoptera of Tristan da Cunha. *Results of the Norwegian Scientific Expedition to Tristan da Cunha 1937–1938*., 17, 1–22.

Bukontaite, R., Ranarilalatiana, T., Randriamihaja, J.H. & Bergsten, J. (2015) In or out of Madagascar? Colonization patterns for large bodied diving beetles (Coleoptera: Dytiscidae). *PLoS One*, 10, e0120777. Available from: https://doi.org/10.1371/journal

Burmeister, E.-G. (1976) Der ovipositor der Hydradephaga (Coleoptera) und seine phylogenetische Bedeutung unter besonderer Berucksichtigung der Dytiscidae. *Zoomorphologie*, 85, 165–257. Available from: https://doi.org/10.1007/BF00993515

Calosi, P., Bilton, D.T. & Spicer, J.I. (2008) Thermal tolerance, acclimatory capacity and vulnerability to global climate change. *Biology Letters*, 4, 99–102. Available from: https://doi.org/10.1098/rsbl.2007.0408

Calosi, P., Bilton, D.T., Spicer, J.I. & Atfield, A. (2008) Thermal tolerance and geographical range size in the *Agabus brunneus* group of European diving beetles (Coleoptera: Dytiscidae). *Journal of Biogeography*, 35, 295–305. Available from: https://doi.org/10.1111/j.1365-2699.2007.01787.x

Calosi, P., Bilton, D.T., Spicer, J.I., Votier, S.C. & Atfield, A. (2010) What determines a species' geographical range? Thermal biology and latitudinal range size relationships in European diving beetles (Coleoptera: Dytiscidae). *Journal of Animal Ecology*, 79, 194–204. Available from: https://doi.org/10.1111/j.1365-2656.2009.01611.x

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, 10, 421. Available from: https://doi.org/10.1186/1471-2105-10-421

Chen, M.-Y., Liang, D. & Zhang, P. (2015) Selecting question-specific genes to reduce incongruence in phylogenomics: a case study of jawed vertebrate backbone phylogeny. *Systematic Biology*, 64, 1104–1120. Available from: https://doi.org/10.1093/sysbio/syv059

Choo, M.Z.J., Low, B.W. & Yeo, D.C.J. (2021) Predaceous diving beetles: a potential alternative mosquito biocontrol agent to dragonflies. *Journal of Vector Ecology*, 46, 226–229 224. Available from: https://doi.org/10.52707/1081-1710-46.2.226

Criscuolo, A. & Gribaldo, S. (2010) BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology*, 10, 210. Available from: https://doi.org/10.1186/1471-2148-10-210

Culler, L.E., Ohba, S.-y. & Crumrine, P. (2014) Predator-prey interactions of dytiscids. In: Yee, D.A. (Ed.) *Ecology, systematics, and natural history of predaceous diving beetles (Coleoptera: Dytiscidae)*. New York: Springer, pp. 363–386.

Darriba, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B. & Flouri, T. (2019) ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular Biology and Evolution*, 37, 291–294. Available from: https://doi.org/10.1093/molbev/msz189

Darwin, C. (1871) *The descent of man, and selection in relation to sex*. London: John Murray.

de Marzo, L. & Nilsson, A. (1988) Structure of the "preoral comb" in larvae of European Agabini (Coleoptera: Dytiscidae). *Entomologica (Bari)*, 23, 191–205. Available from: https://doi.org/10.15162/0425-1016/621

de Marzo, L. & Nilsson, A.N. (1986) Morphological variation and fine structure of some head structures in larvae of Dytiscidae (Coleoptera). *Entomologica Basiliensia*, 11, 29–42. Available from: https://doi.org/10.5169/seals-980543

Degnan, J.H. & Rosenberg, N.A. (2006) Discordance of species trees with their Most likely gene trees. *PLoS Genetics*, 2(5), e68. Available from: https://doi.org/10.1371/journal.pgen.0020068

Désamoré, A., Laenen, B., Miller, K.B. & Bergsten, J. (2018) Early burst in body size evolution is uncoupled from species diversification in diving beetles (Dytiscidae). *Molecular Ecology*, 27, 979–993. Available from: https://doi.org/10.1111/mec.14492

Di Franco, A., Poujol, R., Baurain, D. & Philippe, H. (2019) Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evolutionary Biology*, 19, 21. Available from: https://doi.org/10.1186/s12862-019-1350-2

Edwards, S.V. (2009) Is a new and general theory of molecular systematics emerging? *Evolution*, 63, 1–19. Available from: https://doi.org/10.1111/j.1558-5646.2008.00549.x

Edwards, S.V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C. et al. (2016) Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution*, 94, 447–462. Available from: https://doi.org/10.1016/j.ympev.2015.10.027

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T. & Glenn, T.C. (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, 61, 717–726. Available from: https://doi.org/10.1093/sysbio/sys004

Falkenström, G. (1938) Die Arthropodoenfauna von Madeira nach den Ergebnissen der Reise von Prof. Dr. O. Lundblad Juil-August 1935. IX. Coleoptera: Dytsicidae. *Arkiv för Zoologi*, 30A, 1–19.

Fery, H. & Bouzid, S. (2016) Notes on *Graptodytes* SEIDLITZ, 1887, reinstatement of *G. laeticulus* (SHARP, 1882) as valid species and description of *Tassilodytes* nov.gen. From Algeria (Coleoptera, Dytiscidae, Hydroporinae, Siettitiina). *Linzer Biologische Beiträge*, 1, 451–481. Available from: https://doi.org/10.5281/zenodo.13146895

Fery, H. & Ribera, I. (2018) Phylogeny and taxonomic revision of Deronectina Galewski, 1994 (Coleoptera: Dytiscidae: Hydroporinae: Hydroporini). *Zootaxa*, 4474, 1–104. Available from: https://doi.org/10.11646/zootaxa.4474.1.1

Foster, G.N. & Bilton, D. (2014) The conservation of predaceous diving beetles: knowns, unknowns and anecdotes. In: Yee, D.A. (Ed.) *Ecology, systematics, and natural history of predaceous diving beetles (Coleoptera: Dytiscidae)*. New York: Springer, pp. 437–462.

Foster, G.N. & Eyre, M.D. (1992) Classification and ranking of water beetle communities. *UK Nature Conservation*, 1, 1–110.

Guignot, F. (1936) Fascicule 31. Coleoptera. 10. Haliplidae et Dytiscidae (1re partie). In: Jeannel, R. (Ed.) *Mission scientifique de l'Omo. Tome IV. Zoologie*. Paris: Mémoires du Muséum National d'Histoire Naturelle Paris, Nouvelle Série 8, pp. 1–75.

Guignot, F. (1959) Revision des Hydrocanthares d'Afrique (Coleoptera Dytiscoidea). Première partie. *Annales du Musée Royal du Congo Belge Tervuren (Belgique) (Ser. 8) Sciences Zoologiques*, 70, 1–316.

Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29, 1072–1075. Available from: https://doi.org/10.1093/bioinformatics/btt086

Gustafson, G.T., Baca, S.M., Alexander, A.M. & Short, A.E.Z. (2020) Phylogenomic analysis of the beetle suborder Adephaga with comparison of tailored and generalized ultraconserved element probe performance. *Systematic Entomology*, 45, 552–570. Available from: https://doi.org/10.1111/syen.12413

Gustafson, G.T., Miller, K.B., Michat, M.C., Alarie, Y., Baca, S.M., Balke, M. et al. (2021) The enduring value of reciprocal illumination in the era of insect phylogenomics: a response to Cai et al. (2020). *Systematic Entomology*, 46, 473–486. Available from: https://doi.org/10.1111/syen.12471

Hawlitschek, O., Hendrich, L. & Balke, M. (2012) Molecular phylogeny of the squeak beetles, a family with disjunct Palearctic-Australian range. *Molecular Phylogenetics and Evolution*, 62, 550–554.

Hendrich, L., Pons, J., Ribera, I. & Balke, M. (2010) Mitochondrial Cox1 sequence data reliably uncover patterns of insect diversity but suffer from high lineage-idiosyncratic error rates. *PLoS One*, 5(12), e14448. Available from: https://doi.org/10.1371/journal.pone.0014448

Hendrich, L., Toussaint, E.F.A. & Balke, M. (2014) A new genus of Hydroporini from south-western Australia (Coleoptera, Dytiscidae). *Spixiana*, 37, 103–109.

Hendrich, L., Watts, C.H.S. & Balke, M. (2020) The "minute diving beetles" of southern Australia – taxonomic revision of *Gibbidessus* Watts, 1978, with description of six new species (Coleoptera, Dytiscidae, Bidessini). *Zookeys*, 975, 11–49. Available from: https://doi.org/10.3897/zookeys.975.55456

Hjalmarsson, A.E., Bergsten, J. & Monaghan, M.T. (2015) Dispersal is linked to habitat use in 59 species of water beetles (Coleoptera: Adephaga) on Madagascar. *Ecography*, 38, 732–739. Available from: https://doi.org/10.1111/ecog.01138

Hjalmarsson, A.E., Bukontaite, R., Ranarilalatiana, T., Randriamihaja, J.H. & Bergsten, J. (2013) Taxonomic revision of Madagascan *Rhantus* (Coleoptera, Dytiscidae, Colymbetinae) with an emphasis on Manjakatompo as a conservation priority. *Zookeys*, 350, 21–45. Available from: https://doi.org/10.3897/zookeys.350.6127

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. & Vinh, L.S. (2017) UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35, 518–522. Available from: https://doi.org/10.1093/molbev/msx281

Hoelzer, G.A. & Meinick, D.J. (1994) Patterns of speciation and limits to phylogenetic resolution. *Trends in Ecology & Evolution*, 9, 104–107. Available from: https://doi.org/10.1016/0169-5347(94)90207-0

Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L. & Kimball, R.T. (2015) Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). *Molecular Biology and Evolution*, 33, 1110–1125. Available from: https://doi.org/10.1093/molbev/msv347

ICZN. (1999) *International code of zoological nomenclature*, Fourth edition. London, UK: The International Trust for Zoological Nomenclature.

Isambert, B., Bergsten, J., Monaghan, M.T., Andriamizehy, H., Ranarilalatiana, T., Ratsimbazafy, M. et al. (2011) Endemism and evolutionary history in conflict over Madagascar's freshwater conservation priorities. *Biological Conservation*, 144, 1902–1909. Available from: https://doi.org/10.1016/j.biocon.2011.04.016

Jiang, W., Chen, S.-Y., Wang, H., Li, D.-Z. & Wiens, J.J. (2014) Should genes with missing data be excluded from phylogenetic analyses? *Molecular Phylogenetics and Evolution*, 80, 308–318. Available from: https://doi.org/10.1016/j.ympev.2014.08.006

Jiang, X., Edwards, S.V. & Liu, L. (2020) The multispecies coalescent model outperforms concatenation across diverse Phylogenomic data sets. *Systematic Biology*, 69, 795–812. Available from: https://doi.org/10.1093/sysbio/syaa008

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. & Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14, 587–589. Available from: https://doi.org/10.1038/nmeth.4285

Kanda, K., Gomez, R.A., Van Driesche, R., Miller, K.B. & Maddison, D.R. (2016) Phylogenetic placement of the Pacific northwest subterranean endemic diving beetle *Stygoporus oregonensis* Larson & LaBonte (Dytiscidae, Hydroporinae). *Zookeys*, 632, 75–91. Available from: https://doi.org/10.3897/zookeys.632.9866

Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30, 3059–3066. Available from: https://doi.org/10.1093/nar/gkf436

Knyshov, A., Gordon, E.R.L. & Weirauch, C. (2021) New alignment-based sequence extraction software (ALiBaSeq) and its utility for deep level phylogenetics. *PeerJ*, 9, e11019. Available from: https://doi.org/10.7717/peerj.11019

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. (2019) RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35, 4453–4455. Available from: https://doi.org/10.1093/bioinformatics/btz305

Kubatko, L.S. & Degnan, J.H. (2007) Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, 56, 17–24. Available from: https://doi.org/10.1080/10635150601146041

Lanfear, R. & Hahn, M.W. (2024) The meaning and measure of concordance factors in Phylogenomics. *Molecular Biology and Evolution*, 41(11), msae214. Available from: https://doi.org/10.1093/molbev/msae214

Lartillot, N. & Philippe, H. (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution*, 21, 1095–1109. Available from: https://doi.org/10.1093/molbev/msh112

Larson, D.J., Alarie Y. & Roughley, R.E. (2000) *Predaceous Diving Beetles (Coleoptera: Dytiscidae) of the Nearctic Region with emphasis on the fauna of Canada and Alaska*. Ottawa, Canada: NRC Research Press.

Le, S.Q. & Gascuel, O. (2008) An improved general amino acid replacement matrix. *Molecular Biology and Evolution*, 25, 1307–1320. Available from: https://doi.org/10.1093/molbev/msn067

Le, S.Q., Gascuel, O. & Lartillot, N. (2008) Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics*, 24, 2317–2323. Available from: https://doi.org/10.1093/bioinformatics/btn445

Lemmon, A.R., Emme, S.A. & Lemmon, E.M. (2012) Anchored hybrid enrichment for massively highthroughput phylogenomics. *Systematic Biology*, 61, 727–744. Available from: https://doi.org/10.1093/sysbio/sys049

Lemmon, E.M. & Lemmon, A.R. (2013) High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, 44, 99–121. Available from: https://doi.org/10.1146/annurev-ecolsys-110512-135822

Leys, R., Watts, C.H.S., Cooper, S.J.B. & Humphreys, W.F. (2003) Evolution of subterranean diving beetles (Coleoptera: Dytiscidae: Hydroporini, Bidessini) in the arid zone of Australia. *Evolution*, 57, 2819–2834. Available from: https://doi.org/10.1554/03-165

Linnæus, C. (1735) *Systema naturæ, sive regna tria naturæ systematice proposita per classes, ordines, genera, & species*. Lugduni Batavorum: Haak.

Liu, L., Xi, Z. & Davis, C.C. (2015) Coalescent methods are robust to the simultaneous effects of long branches and incomplete lineage sorting. *Molecular Biology and Evolution*, 32, 791–805. Available from: https://doi.org/10.1093/molbev/msu331

Lundkvist, E., Landin, J., Jackson, M. & Svensson, C. (2003) Diving beetles (Dytiscidae) as predators of mosquito larvae (Culicidae) in field experiments and in laboratory tests of prey preference. *Bulletin of Entomological Research*, 93, 219–226. Available from: https://doi.org/10.1079/BER2003237

Mai, U. & Mirarab, S. (2018) TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics*, 19, 272. Available from: https://doi.org/10.1186/s12864-018-4620-2

McKenna, D.D., Shin, S., Ahrens, D., Balke, M., Beza-Beza, C., Clarke, D.J. et al. (2019) The evolution and genomic basis of beetle diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 24729–24737. Available from: https://doi.org/10.1073/pnas.1909655116

Michat, M., Alarie, W. & Balke, M. (2023) *Bunites distigma* (Brullé, 1837): discovery of the second- and third-instar larvae and phylogenetic relationships within the Colymbetinae (Coleoptera: Dytiscidae). *Zootaxa*, 5353, 551–566. Available from: https://doi.org/10.11646/zootaxa.5353.6.3

Michat, M. & Toledo, M. (2015) Phylogenetic relationships and larval morphology of the recently described diving beetle genus *Laccomimus* (Coleoptera: Dytiscidae: Laccophilinae). *European Journal of Endocrinology*, 112, 831–843. Available from: https://doi.org/10.14411/eje.2015.085

Michat, M.C. (2006) The phylogenetic position of *Hydrovatus* Motschulsky: evidence from larval morphology of *H. caraibus* Sharp (Coleoptera: Dytiscidae: Hydroporinae). *Insect Systematics & Evolution*, 37, 419–432. Available from: https://doi.org/10.1163/187631206788831416

Michat, M.C. & Alarie, Y. (2009) Phylogenetic relationships of *Notaticus* (Coleoptera: Dytiscidae) based on larval morphology. *Annals of the Entomological Society of America*, 102, 797–808. Available from: https://doi.org/10.1603/008.102.0506

Michat, M.C. & Alarie, Y. (2013) Phylogenetic relationships, larval morphology, and chaetotaxy of the subfamily Coptotominae (Coleoptera: Dytiscidae). *Canadian Entomologist*, 145, 247–264. Available from: https://doi.org/10.4039/tce.2012.100

Michat, M.C., Alarie, Y. & Miller, K.B. (2017) Higher-level phylogeny of diving beetles (Coleoptera: Dytiscidae) based on larval characters. *Systematic Entomology*, 42, 734–767. Available from: https://doi.org/10.1111/syen.12243

Michat, M.C., Alarie, Y. & Watts, C.H. (2015) Phylogenetic placement of the genus *Sternhydrus* (Coleoptera: Dytiscidae: Cybistrini) based on larval morphology, with description of *S. atratus*. *Annals of the Entomological Society of America*, 108, 881–892. Available from: https://doi.org/10.1093/aesa/sav067

Michat, M.C., Alarie, Y. & Watts, C.H.S. (2019) Dealing with a hairy beast-larval morphology and chaetotaxy of the Australian endemic diving

beetle genus *Spencerhydrus* (Coleoptera, Dytiscidae, Cybistrini). *Zookeys*, 884, 53–67. Available from: https://doi.org/10.3897/zookeys.884.38391

Michat, M.C. & Torres, P.L.M. (2008) On the systematic position of the diving-beetle genus *Pachydrus* (Coleoptera: Dytiscidae: Hydroporinae): evidence from larval chaetotaxy and morphology. *European Journal of Entomology*, 105, 737–750. Available from: https://doi.org/10.14411/eje.2008.100

Michat, M.C. & Torres, P.L.M. (2009) A preliminary study on the phylogenetic relationships of *Copelatus* Erichson (Coleoptera: Dytiscidae: Copelatinae) based on larval chaetotaxy and morphology. *Hydrobiologia*, 632, 309–327. Available from: https://doi.org/10.1007/s10750-009-9853-2

Miller, K.B. (2000) Cladistic analysis of the tribes of Dytiscinae and the phylogenetic position of the genus *Notaticus* Zimmermann (Coleoptera; Dytiscidae). *Insect Systematics & Evolution*, 31, 165–177. Available from: https://doi.org/10.1163/187631200X00363

Miller, K.B. (2001) On the phylogeny of the Dytiscidae (Coleoptera) with emphasis on the morphology of the female reproductive tract. *Insect Systematics & Evolution*, 32, 45–92. Available from: https://doi.org/10.1163/187631201X00029

Miller, K.B. (2002) Revision of the subfamily Hydrodytinae Miller with description of a new genus (Coleoptera: Dytiscidae). *Insect Systematics & Evolution*, 33, 1–8. Available from: https://doi.org/10.1163/187631202X00019

Miller, K.B. (2003) The phylogeny of diving beetles (Coleoptera: Dytiscidae) and the evolution of sexual conflict. *Biological Journal of the Linnaean Society*, 79, 359–388. Available from: https://doi.org/10.1046/j.1095-8312.2003.00195.x

Miller, K.B. (2005) Revision of the New World and southeast Asian Vatellini (Coleoptera: Dytiscidae: Hydroporinae) and phylogenetic analysis of the tribe. *Zoological Journal of the Linnean Society*, 144, 415–510. Available from: https://doi.org/10.1111/j.1096-3642.2005.00180.x

Miller, K.B. (2016) *Novadessus viracocha*, a new genus and species of Bidessini Sharp from Peru (Coleoptera, Adephaga, Dytiscidae, Hydroporinae). *Zookeys*, 623, 125–130. Available from: https://doi.org/10.3897/zookeys.623.10018

Miller, K.B., Alarie, Y. & Whiting, M.F. (2007) Description of the larva of *Notaticus fasciatus* Zimmermann (Coleoptera: Dytiscidae) associated with adults using DNA sequence data. *Annals of the Entomological Society of America*, 100, 787–797. Available from: https://doi.org/10.1603/0013-8746(2007)100[787:DOTLON]2.0.CO;2

Miller, K.B. & Bergsten, J. (2014a) The phylogeny and classification of diving beetles (Coleoptera: Dytiscidae). In: Yee, D.A. (Ed.) *Ecology, systematics, and natural history of predaceous diving beetles (Coleoptera: Dytiscidae)*. New York: Springer, pp. 49–172.

Miller, K.B. & Bergsten, J. (2014b) Predaceous diving beetle sexual systems. In: Yee, D.A. (Ed.) *Ecology, systematics, and natural history of predaceous diving beetles (Coleoptera: Dytiscidae)*. New York: Springer, pp. 199–234.

Miller, K.B. & Bergsten, J. (2016) *Diving beetles of the world*. Baltimore, Maryland: Johns-Hopkins University Press.

Miller, K.B. & Bergsten, J. (2023a) The phylogeny and classification of diving beetles (Coleoptera: Dytiscidae). In: Yee, D.A. (Ed.) *Ecology, systematics, and natural history of predaceous diving beetles (Coleoptera: Dytiscidae)*, 2nd edition. New York: Springer, pp. 55–185.

Miller, K.B. & Bergsten, J. (2023b) Predaceous diving beetle sexual systems. In: Yee, D.A. (Ed.) *Ecology, systematics, and natural history of predaceous diving beetles (Coleoptera: Dytiscidae)*, 2nd edition. New York: Springer, pp. 187–224.

Miller, K.B., Bergsten, J. & Whiting, M.F. (2007) Phylogeny and classification of diving beetles in the tribe Cybistrini (Coleoptera, Dytiscidae, Dytiscinae). *Zoologica Scripta*, 36, 41–59. Available from: https://doi.org/10.1111/j.1463-6409.2006.00254.x

Miller, K.B., Bergsten, J. & Whiting, M.F. (2009) Phylogeny and classification of the tribe Hydaticini (Coleoptera: Dytiscidae): partition choice for Bayesian analysis with multiple nuclear and mitochondrial protein-coding genes. *Zoologica Scripta*, 38, 591–615. Available from: https://doi.org/10.1111/j.1463-6409.2009.00393.x

Miller, K.B., Michat, M.C. & Ferreira, N., Jr. (2024) Reclassification of Cybistrinae Sharp, 1880 in the Neotropical region (Coleoptera, Adephaga, Dytiscidae), with description of new taxa. *ZooKeys*, 1188, 125–168. Available from: https://doi.org/10.3897/zookeys.1188.110081

Miller, K.B. & Perkins, P.D. (2012) The diving beetle species *Hydrotrupes palpalis* Sharp, 1882 (Coleoptera: Dytiscidae). *The Coleopterist Bulletin*, 66(4), 371–377.

Miller, K.B. & Short, A.E.Z. (2015) *Belladessus* Miller and Short (Coleoptera: Dytiscidae: Hydroporinae: Bidessini), new genus for two new species from northern South America: parthenogenetic diving beetles? *The Coleopterists Bulletin*, 69(3), 498–503. Available from: https://doi.org/10.1649/0010-065X-69.3.498

Miller, K.B. & Wheeler, Q.D. (2015) *Zimpherus nancae* n. gen. And n. sp., from Venezuela (Coleoptera: Dytiscidae: Hydroporinae: Bidessini). *The Coleopterists Bulletin*, 69, 507–511. Available from: https://doi.org/10.1649/0010-065X-69.3.507

Miller, K.B., Wolfe, G.W. & Biström, O. (2006) Phylogeny of the Hydroporinae and classification of the genus *Peschetius* Guignot, 1942 (Coleoptera: Dytiscidae). *Insect Systematics & Evolution*, 37, 257–279. Available from: https://doi.org/10.1163/187631206788838617

Minh, B.Q., Dang, C.C., Vinh, L.S. & Lanfear, R. (2021) QMaker: fast and accurate method to estimate empirical models of protein evolution. *Systematic Biology*, 70, 1046–1060. Available from: https://doi.org/10.1093/sysbio/syab010

Minh, B.Q., Hahn, M.W. & Lanfear, R. (2020) New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution*, 37, 2727–2733. Available from: https://doi.org/10.1093/molbev/msaa106

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A. et al. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37, 1530–1534. Available from: https://doi.org/10.1093/molbev/msaa015

Mirarab, S., Nakhleh, L. & Warnow, T. (2021) Multispecies coalescent: theory and applications in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, 52, 247–268. Available from: https://doi.org/10.1146/annurev-ecolsys-012121-095340

Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S. & Warnow, T. (2014) ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, 30, i541–i548. Available from: https://doi.org/10.1093/bioinformatics/btu462

Mirarab, S. & Warnow, T. (2015) ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics*, 31, i44–i52. Available from: https://doi.org/10.1093/bioinformatics/btv234

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C. et al. (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science*, 346, 763–767. Available from: https://doi.org/10.1126/science.1257570

Molloy, E.K. & Warnow, T. (2017) To include or mot to include: the impact of gene filtering on species tree estimation methods. *Systematic Biology*, 67, 285–303. Available from: https://doi.org/10.1093/sysbio/syx077

Morel, B., Kozlov, A.M. & Stamatakis, A. (2018) ParGenes: a tool for massively parallel model selection and phylogenetic tree inference on thousands of genes. *Bioinformatics*, 35, 1771–1773. Available from: https://doi.org/10.1093/bioinformatics/bty839

Morinière, J., Michat, M.C., Jäch, M.A., Bergsten, J., Hendrich, L. & Balke, M. (2014) Anisomeriini diving beetles—an Atlantic–Pacific

Island disjunction on Tristan da Cunha and Robinson Crusoe Island, Juan Fernández? *Cladistics*, 31, 166–176. Available from: https://doi.org/10.1111/cla.12074

Morinière, J., Van Dam, M.H., Hawlitschek, O., Bergsten, J., Michat, M.C., Hendrich, L. et al. (2016) Phylogenetic niche conservatism explains an inverse latitudinal diversity gradient in freshwater arthropods. *Scientific Reports*, 6, 26340. Available from: https://doi.org/10.1038/srep26340

Nguyen, L.-T., Schmidt, H.A., Haeseler, A.v. & Minh, B.Q. (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution*, 32, 268–274. Available from: https://doi.org/10.1093/molbev/msu300

Nieselt-Struwe, K. & von Haeseler, A. (2001) Quartet-mapping, a generalization of the likelihood-mapping procedure. *Molecular Biology and Evolution*, 18, 1204–1219. Available from: https://doi.org/10.1093/oxfordjournals.molbev.a003907

Nilsson, A.N. (1988) A review of primary setae and pores on legs of larval Dytiscidae (Coleoptera). *Canadian Journal of Zoology*, 66, 2283–2294. Available from: https://doi.org/10.1139/z88-339

Nilsson, A.N. (1997) A redefinition and revision of the *Agabus optatus*-group (Coleoptera, Dytiscidae); an example of Pacific intercontinental disjunction. *Entomologica Basiliensia*, 19, 621–651.

Nilsson, A.N. (2000) A new view on the generic classification of the Agabus-group of genera of the Agabini, aimed at solving the problem with a paraphyletic Agabus (Coleoptera: Dytiscidae). *Koleopterologische Rundschau*, 70, 17–36.

Nilsson, A.N. & Hájek, J. (2024) A world catalogue of the family Dytiscidae, or the diving beetles (Coleoptera, Adephaga), 1.I.2024:1-323.

Nilsson-Ortman, V. & Nilsson, A.N. (2010) Using taxonomic revision data to estimate the global species richness and characteristics of undescribed species of diving beetles (Coleoptera: Dytiscidae). *Biodiversity Informatics*, 7, 1–16. Available from: https://doi.org/10.17161/bi.v7i1.3631

Nute, M., Chou, J., Molloy, E.K. & Warnow, T. (2018) The performance of coalescent-based species tree estimation methods under models of missing data. *BMC Genomics*, 19, 286. Available from: https://doi.org/10.1186/s12864-018-4619-8

O'Connor, T., Sundberg, K., Carroll, H., Clement, M. & Snell, Q. (2010) Analysis of long branch extraction and long branch shortening. *BMC Genomics*, 11, S14. Available from: https://doi.org/10.1186/1471-2164-11-S2-S14

Ohba, S.-y. & Ushio, M. (2015) Effect of water depth on predation frequency by diving beetles on mosquito larvae prey. *Entomological Science*, 18, 519–522. Available from: https://doi.org/10.1111/ens.12143

Okada, R., Alarie, Y. & Michat, M.C. (2019) Description of the larvae of four Japanese *Platambus* Thomson, 1859 (Coleoptera: Dytiscidae: Agabinae) with phylogenetic considerations. *Zootaxa*, 4646, 401–433. Available from: https://doi.org/10.11646/zootaxa.4646.3.1

Pederzani, F. (1995) Keys to the identification of the genera and subgenera of adult Dytiscidae (*sensu lato*) of the world (Coleoptera Dytiscidae). *Atti Della Accademia Roveretana Degli Agiati*, 4, 5–83.

Portik, D.M., Streicher, J.W., Blackburn, D.C., Moen, D.S., Hutter, C.R. & Wiens, J.J. (2023) Redefining possible: combining phylogenomic and supersparse data in frogs. *Molecular Biology and Evolution*, 40, msad109. Available from: https://doi.org/10.1093/molbev/msad109

Price, M.N., Dehal, P.S. & Arkin, A.P. (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One*, 5, e9490. Available from: https://doi.org/10.1371/journal.pone.0009490

Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M. et al. (2015) A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, 526, 569–573. Available from: https://doi.org/10.1038/nature15697

Queney, P., Lemaire, J.-M. & Ferrand, M. (2020) A new genus and species of subterranean diving beetle from Laos (Coleoptera, Dytiscidae, Hydroporinae, Hydroporini). *Annales de la Societe Entomologique de France (N.S.)*, 125, 407–416. Available from: https://doi.org/10.32475/bsef_2144

Rannala, B., Leache, A., Edwards, S. & Yang, Z. (2020) The multispecies coalescent model and species tree inference. In: Scornavacca, C., Delsuc, F. & Galtier, N. (Eds.) *Phylogenetics in the genomic era*. Montpellier, France: Self Published, pp. 3.3:1–3.3:21.

Rannala, B. & Yang, Z. (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, 164, 1645–1656. Available from: https://doi.org/10.1093/genetics/164.4.1645

Ranwez, V., Harispe, S., Delsuc, F. & Douzery, E.J.P. (2011) MACSE: multiple alignment of coding sequences accounting for frameshifts and stop codons. *PLoS One*, 6, e22594. Available from: https://doi.org/10.1371/journal.pone.0022594

Régimbart, M. (1879) Étude sur la classification des Dytiscidae. *Annales de la Société Entomologique de France (5)*, 8(1878), 447–466.

Rhodes, J.A., Nute, M.G. & Warnow, T. (2020) NJst and ASTRID are not statistically consistent under a random model of missing data. arXiv: 2001.07844 [q-bio.PE]. https://doi.org/10.48550/arXiv.2001.07844

Ribera, I. (2008) Chapter 15: habitat constraints and the generation of diversity in freshwater macroinvertebrates. In: Lancaster, J. & Briers, R.A. (Eds.) *Aquatic insects: challenges to populations*. London: CAB International UK, pp. 289–311. Available from: https://doi.org/10.1079/9781845933968.0000

Ribera, I., Barraclough, T.G. & Vogler, A.P. (2001) The effect of habitat type on speciation rates and range movements in aquatic beetles: inferences from species-level phylogenies. *Molecular Ecology*, 10, 721–735. Available from: https://doi.org/10.1046/j.1365-294x.2001.01218.x

Ribera, I., Beutel, R.G., Balke, M. & Vogler, A.P. (2002) Discovery of Aspidytidae, a new family of aquatic Coleoptera. *Proceedings of the Royal Society Biological Sciences Series B*, 269, 2351–2356. Available from: https://doi.org/10.1098/rspb.2002.2157

Ribera, I., Foster, G.N. & Vogler, A.P. (2003) Does habitat use explain large scale diversity patterns in European water beetles? *Ecography*, 26, 145–152. Available from: https://doi.org/10.1034/j.1600-0587.2003.03271.x

Ribera, I., Hogan, J.E. & Vogler, A.P. (2002) Phylogeny of hydradephagan water beetles inferred from 18S rRNA sequences. *Molecular Phylogenetics and Evolution*, 23, 43–62. Available from: https://doi.org/10.1006/mpev.2001.1080

Ribera, I., Nilsson, A.N. & Vogler, A.P. (2004) Phylogeny and historical biogeography of Agabinae diving beeetles (Coleoptera) inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, 30, 545–562. Available from: https://doi.org/10.1016/S1055-7903(03)00224-0

Ribera, I. & Reboleira, A.S.P.S. (2019) The first stygobiont species of Coleoptera from Portugal, with a molecular phylogeny of the Siettitia group of genera (Dytiscidae, Hydroporinae, Hydroporini, Siettitiina). *ZooKeys*, 813, 21–38. Available from: https://doi.org/10.3897/zookeys.813.29765

Ribera, I. & Vogler, A.P. (2000) Habitat type as a determinant of species range sizes: the example of lotic-lentic differences in aquatic Coleoptera. *Biological Journal of the Linnean Society*, 71, 33–52. Available from: https://doi.org/10.1111/j.1095-8312.2000.tb01240.x

Ribera, I., Vogler, A.P. & Balke, M. (2008) Phylogeny and diversification of diving beetles (Coleoptera: Dytiscidae). *Cladistics*, 24, 563–590. Available from: https://doi.org/10.1111/j.1096-0031.2007.00192.x

Roch, S., Nute, M. & Warnow, T. (2018) Long-branch attraction in species tree estimation: inconsistency of partitioned likelihood and topology-based summary methods. *Systematic Biology*, 68, 281–297. Available from: https://doi.org/10.1093/sysbio/syy061

Roch, S. & Steel, M. (2015) Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically

inconsistent. *Theoretical Population Biology*, 100, 56–62. Available from: https://doi.org/10.1016/j.tpb.2014.12.005

Roch, S. & Warnow, T. (2015) On the robustness to gene tree estimation error (or lack thereof) of coalescent-based species tree methods. *Systematic Biology*, 64, 663–676. Available from: https://doi.org/10.1093/sysbio/syv016

Roughley, R.E. (2000) Subfamily Hydrotrupinae Roughley, new subfamily. In: Larson, D.J., Alarie, Y. & Roughley, R.E. (Eds.) *Predaceous diving beetles (Coleoptera: Dytiscidae) of the Nearctic region, with emphasis on the fauna of Canada and Alaska*. Ottawa, Ontario, Canada: National Research Council of Canada Research Press, pp. 52–54.

Ruhnau, S. (1986) Phylogenetic relations within the Hydradephaga (Coleoptera) using larval and pupal characters. *Entomologica Basiliensia*, 11, 231–271.

Ruhnau, S. & Brancucci, M. (1984) Studies on the genus *Lancetes*. 2. Analysis of its phylogenetic position using preimaginal characters (Coleoptera, Dytiscidae). *Entomologica Basiliensia*, 9, 80–107.

Sayyari, E. & Mirarab, S. (2016) Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution*, 33, 1654–1668. Available from: https://doi.org/10.1093/molbev/msw079

Sharp, D. (1882) On aquatic carnivorous Coleoptera or Dytiscidae. *Scientific Transactions of the Royal Dublin Society*, 2, 179–1003.

Siddall, M.E. & Whiting, M.F. (1999) Long branch abstractions. *Cladistics*, 15, 9–24. Available from: https://doi.org/10.1111/j.1096-0031.1999.tb00391.x

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. & Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31, 3210–3212. Available from: https://doi.org/10.1093/bioinformatics/btv351

Simmons, M.P. (2012) Misleading results of likelihood-based phylogenetic analyses in the presence of missing data. *Cladistics*, 28, 208–222. Available from: https://doi.org/10.1111/j.1096-0031.2011.00375.x

Slowinski, J.B. (2001) Molecular polytomies. *Molecular Phylogenetics and Evolution*, 19, 114–120. Available from: https://doi.org/10.1006/mpev.2000.089

Smith, M.L. & Hahn, M.W. (2021) New approaches for inferring phylogenies in the presence of paralogs. *Trends in Genetics*, 37, 174–187. Available from: https://doi.org/10.1016/j.tig.2020.08.012

Smith, S.A., Wilson, N.G., Goetz, F.E., Feehery, C., Andrade, S.C.S., Rouse, G.W. et al. (2011) Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature*, 480, 364–367. Available from: https://doi.org/10.1038/nature10526

Steenwyk, J.L., Li, Y., Zhou, X., Shen, X.-X. & Rokas, A. (2023) Incongruence in the phylogenomics era. *Nature Reviews Genetics*, 24, 834–850. Available from: https://doi.org/10.1038/s41576-023-00620-x

Streicher, J.W., Schulte, J.A., II & Wiens, J.J. (2016) How should genes and taxa be sampled for phylogenomic analyses with missing data? An empirical study in iguanian lizards. *Systematic Biology*, 65, 128–145. Available from: https://doi.org/10.1093/sysbio/syv058

Strimmer, K. & von Haeseler, A. (1997) Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 6815–6819. Available from: https://doi.org/10.1073/pnas.94.13.6815

Suh, A. (2016) The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. *Zoologica Scripta*, 45, 50–62. Available from: https://doi.org/10.1111/zsc.12213

Tange, O. (2011) GNU parallel – the command-line power tool. *The USENIX Magazine*, 36, 42–47.

Toledo, M. & Michat, M.C. (2015) Description of *Laccomimus* gen. n. and eleven new species from the Neotropical region (Coleoptera, Dytiscidae, Laccophilinae). *Zootaxa*, 3990, 301–354. Available from: https://doi.org/10.11646/zootaxa.3990.3.1

Toussaint, E., Hendrich, L., Escalona, H., Porch, N. & Balke, M. (2016) Evolutionary history of a secondary terrestrial Australian diving beetle (Coleoptera, Dytiscidae) reveals a lineage of high morphological and ecological plasticity. *Systematic Entomology*, 41, 650–657. Available from: https://doi.org/10.1111/syen.12182

Toussaint, E.F.A., Balke, M., Garcia, M. & Short, A.E.Z. (2016) Molecular systematics of the Neotropical diving beetle genus *Rugosus* García, 2001 (Coleoptera: Dytiscidae: Copelatinae). *The Coleopterists Bulletin*, 70, 53–58. Available from: https://doi.org/10.1649/072.070.0105

Toussaint, E.F.A., Beutel, R.G., Moriniere, J., Jia, F.L., Xu, S.Q., Michat, M.C. et al. (2016) Molecular phylogeny of the highly disjunct cliff water beetles from South Africa and China (Coleoptera: Aspidytidae). *Zoological Journal of the Linnean Society*, 176(3), 537–546. Available from: https://doi.org/10.1111/zoj.12332

Toussaint, E.F.A., Condamine, F.L., Hawlitschek, O., Watts, C.H., Porch, N., Hendrich, L. et al. (2014) Unveiling the diversification dynamics of Australasian predaceous diving beetles in the Cenozoic. *Systematic Biology*, 64, 3–24. Available from: https://doi.org/10.1093/sysbio/syu067

Toussaint, E.F.A., Hendrich, L., Hájek, J., Michat, M.C., Panjaitan, R., Short, A.E.Z. et al. (2017) Evolution of Pacific rim diving beetles sheds light on Amphi-Pacific biogeography. *Ecography*, 40, 500–510. Available from: https://doi.org/10.1111/ecog.02195

Toussaint, E.F.A., Sagata, K., Surbakti, S., Hendrich, L. & Balke, M. (2013) Australasian sky islands act as a diversity pump facilitating peripheral speciation and complex reversal from narrow endemic to widespread ecological supertramp. *Ecology and Evolution*, 3, 1031–1049. Available from: https://doi.org/10.1002/ece3.517

Vachaspati, P. & Warnow, T. (2015) ASTRID: accurate species TRees from internode distances. *BMC Genomics*, 16(Suppl 10), S3. Available from: https://doi.org/10.1186/1471-2164-16-S10-S3

van den Branden, C. (1885) Catalogue des coléoptères carnassiers aquatiques (Haliplidae, Amphizoïdae, Pelobiidae et Dytiscidae). *Annales de la Société Entomologique de Belgique*, 29, 5–118.

Vasilikopoulos, A., Balke, M., Beutel, R.G., Donath, A., Podsiadlowski, L., Pflug, J.M. et al. (2019) Phylogenomics of the superfamily Dytiscoidea (Coleoptera: Adephaga) with an evaluation of phylogenetic conflict and systematic error. *Molecular Phylogenetics and Evolution*, 135, 270–285. Available from: https://doi.org/10.1016/j.ympev.2019.02.022

Vasilikopoulos, A., Balke, M., Kukowka, S., Pflug, J.M., Martin, S. & Meusemann, K. (2021) Phylogenomic analyses clarify the pattern of evolution of Adephaga (Coleoptera) and highlight phylogenetic artefacts due to model misspecification and excessive data trimming. *Systematic Entomology*, 46, 991–1018. Available from: https://doi.org/10.1111/syen.12508

Vasilikopoulos, A., Meusemann, K., Beutel, R., Niehuis, O., Lieberz, D., Aspöck, H. et al. (2020) Phylogenomic insights into the evolution of Neuropterida (Insecta: Holometabola).

Villastrigo, A., Abellán, P. & Ribera, I. (2021) Habitat preference and diversification rates in a speciose lineage of diving beetles. *Molecular Phylogenetics and Evolution*, 159, 107087. Available from: https://doi.org/10.1016/j.ympev.2021.107087

Villastrigo, A., Deharveng, L. & Balke, M. (2023) New Caledonia's enigmatic terrestrial diving beetle Typhlodessus monteithi is a derived species of Paroster. *Zoologica Scripta*, 52, 176–183. Available from: https://doi.org/10.1111/zsc.12581

Villastrigo, A., Fery, H., Manuel, M., Millan, A. & Ribera, I. (2018) Evolution of salinity tolerance in the diving beetle tribe Hygrotini (Coleoptera, Dytiscidae). *Zoologica Scripta*, 47, 63–71. Available from: https://doi.org/10.1111/zsc.12255

Villastrigo, A., Ribera, I., Manuel, M., Millán, A. & Fery, H. (2017) A new classification of the tribe Hygrotini Portevin, 1929 (Coleoptera: Dytiscidae: Hydroporinae). *Zootaxa*, 4317, 499–529. Available from: https://doi.org/10.11646/zootaxa.4317.3.4

Wang, H.-C., Minh, B.Q., Susko, E. & Roger, A.J. (2017) Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Systematic Biology*, 67, 216–235. Available from: https://doi.org/10.1093/sysbio/syx068

Watts, C.H.S., Villastrigo, A., Langille, B.L., Stringer, D.N., Bradford, T.M., Humphreys, W.F. et al. (2023) Phylogenetic placement and description of *Ngaliadessus humphreysi* gen. Et sp. nov., Watts & Villastrigo (Coleoptera: Dytiscidae), a subterranean diving beetle from the Ngalia Basin in central Australia. *Austral Entomology*, 62, 300–309. Available from: https://doi.org/10.1111/aen.12649

Weisenfeld, N.I., Kumar, V., Shah, P., Church, D.M. & Jaffe, D.B. (2017) Direct determination of diploid genome sequences. *Genome Research*, 27, 757–767. Available from: https://doi.org/10.1101/gr.214874.116

Whitfield, J.B. & Lockhart, P.J. (2007) Deciphering ancient rapid radiations. *Trends in Ecology & Evolution*, 22, 258–265. Available from: https://doi.org/10.1016/j.tree.2007.01.012

Wiens, J.J. & Morrill, M.C. (2011) Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. *Systematic Biology*, 60, 719–731. Available from: https://doi.org/10.1093/sysbio/syr025

Wolfe, G.W. (1985) A phylogenetic analysis of plesiotypic hydroporine lineages with an emphasis on *Laccornis* des Gozis (Coleoptera: Dytiscidae). *Proceedings of the Academy of Natural Sciences of Philadelphia*, 137, 132–155.

Wolfe, G.W. (1988) A phylogenetic investigation of *Hydrovatus*, Methlini and other plesiotypic hydroporines (Coleoptera: Dytiscidae). *Psyche*, 95, 327–344. Available from: https://doi.org/10.1155/1988/17890

Wolfe, G.W. & Roughley, R.E. (1990) A taxonomic, phylogenetic, and zoogeographic analysis of *Laccornis* Gozis (Coleoptera: Dytiscidae) with the description of Laccornini, a new tribe of Hydroporinae. *Quaestiones Entomologicae*, 26, 273–354.

Xi, Z., Liu, L. & Davis, C.C. (2015) The impact of missing data on species tree estimation. *Molecular Biology and Evolution*, 33, 838–860. Available from: https://doi.org/10.1093/molbev/msv266

Xia, X. (2014) Phylogenetic bias in the likelihood method caused by missing data coupled with among-site rate variation: an analytical approach. In: Basu, M., Pan, Y. & Wang, J. (Eds.) *Bioinformatics research and applications. ISBRA 2014*. Lecture Notes in Computer Science(), Vol. 8492. Cham: Springer. Available from: https://doi.org/10.1007/978-3-319-08171-7_2

Yang, Z. (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution*, 39, 306–314. Available from: https://doi.org/10.1007/BF00160154

Yee, D.A. (2010) Behavior and aquatic plants as factors affecting predation by three species of larval predaceous diving beetles (Coleoptera: Dytiscidae). *Hydrobiologia*, 637, 33–43.

Young, A.D. & Gillung, J.P. (2020) Phylogenomics — principles, opportunities and pitfalls of big-data phylogenetics. *Systematic Entomology*, 45, 225–247. Available from: https://doi.org/10.1111/syen.12406

Zdobnov, E.M., Kuznetsov, D., Tegenfeldt, F., Manni, M., Berkeley, M. & Kriventseva, E.V. (2020) OrthoDB in 2020: evolutionary and functional annotations of orthologs. *Nucleic Acids Research*, 49, D389–D393. Available from: https://doi.org/10.1093/nar/gkaa1009

Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. (2018) ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19, 153. Available from: https://doi.org/10.1186/s12859-018-2129-y

Zhang, F., Ding, Y., Zhu, C.-D., Zhou, X., Orr, M.C., Scheu, S. et al. (2019) Phylogenomics from low-coverage whole-genome sequencing. *Methods in Ecology and Evolution*, 10, 507–517. Available from: https://doi.org/10.1111/2041-210X.13145

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Figure S1. Concatenated Maximum likelihood tree based on all loci (5364). Values at nodes are ultrafast bootstrap values (UFB)(* if 100)/gene concordance factor (gCF).

Figure S2. Concatenated Maximum likelihood tree without *Hyderodes shuckardi* (long-branch extraction). Node values are local SH-like support values from FastTreeMP.

Figure S3. Concatenated Maximum likelihood tree without *Notaticus fasciatus* (long-branch extraction). Node values are local SH-like support values from FastTreeMP.

Figure S4. Four-cluster likelihood mapping of the basal ingroup trichotomy when genes are ranked by either occupancy (a) or phylogenetic informativeness (b) and increasing portions (0, 10, 25, 50, 75 90%) excluded.

Figure S5. Four-cluster likelihood mapping of the Copelatinae position when genes are ranked by either occupancy (a) or phylogenetic informativeness (b) and increasing portions (0, 10, 25, 50, 75 90%) excluded.

Figure S6. Four-cluster likelihood mapping of the *Hyderodes* problem when genes are ranked by either occupancy (a) or phylogenetic informativeness (b) and increasing portions (0, 10, 25, 50, 75 90%) excluded.

Figure S7. Four-cluster likelihood mapping of the *Notaticus* problem when genes are ranked by either occupancy (a) or phylogenetic informativeness (b) and increasing portions (0, 10, 25, 50, 75 90%) excluded.

Figure S8. ASTRAL tree based on dataset nounique (3202 loci).

Figure S9. ASTRAL tree based on dataset common3 (1529 loci).

Figure S10. ASTRAL tree based on dataset common4 (369 loci).

Figure S11. Concatenated maximum likelihood tree based on dataset nounique (3202 loci).

Figure S12. Concatenated maximum likelihood tree based on dataset common3 (1529 loci).

Figure S13. Concatenated maximum likelihood tree based on dataset common4 (369 loci).

Figure S14. ASTRAL tree based on all loci (5364) where 54 taxa from Vasilikopoulos et al. (2021) are excluded.

Figure S15. ASTRAL tree based on dataset nounique (3202 loci) where 54 taxa from Vasilikopoulos et al. (2021) are excluded.

Figure S16. ASTRAL tree based on dataset common3 (1529 loci) where 54 taxa from Vasilikopoulos et al. (2021) are excluded.

Figure S17. Concatenated maximum likelihood tree based on all loci (5364) where 54 taxa from Vasilikopoulos et al. (2021) are excluded.

Figure S18. Concatenated maximum likelihood tree based on dataset nounique (3202 loci) where 54 taxa from Vasilikopoulos et al. (2021) are excluded.

Figure S19. Concatenated maximum likelihood tree based on dataset common3 (1529 loci) where 54 taxa from Vasilikopoulos et al. (2021) are excluded.

Figure S20. ASTRAL tree based on reference clade-based gene filtering.

Figure S21. ASTRAL tree based on all 5364 loci but *Pachydrus* sp. is excluded.

Table S1. Metadata on included specimens.

Table S2. Statistics of per taxon locus occupancy in full dataset and three reduced datasets as well as per taxon BUSCO statistics for newly sequenced taxa. C = Complete genes, S = Single Copy Complete Genes, D=Duplicated Complete Genes, F = Fragmented Genes, subclass 1: only a portion of the gene is present in the assembly, and the rest of the gene cannot be aligned. I=Fragmented Genes, subclass 2: a section of the gene aligns to one position in the assembly, while the remaining part aligns to another position, M = Missing Genes, N = number of BUSCO reference genes.

How to cite this article: Bergsten, J., Nylander, J.A.A., Ospina, O.E., Lemmon, A.R. & Miller, K.B. (2025) Whole genome shotgun phylogenomics resolve the diving beetle tree of life. *Systematic Entomology*, 1–35. Available from: https://doi.org/10.1111/syen.12685