

Phylogenomic and mitogenomic data can accelerate inventorying of tropical beetles during the current biodiversity crisis

Michal Motyka[†], Dominik Kusy[†], Matej Bocek, Renata Bilkova, Ladislav Bocak*

Laboratory of Biodiversity and Molecular Evolution, Czech Advanced Technology Research Institute, Centre of Region Hana, Olomouc, Czech Republic

Abstract Conservation efforts must be evidence-based, so rapid and economically feasible methods should be used to quantify diversity and distribution patterns. We have attempted to overcome current impediments to the gathering of biodiversity data by using integrative phylogenomic and three mtDNA fragment analyses. As a model, we sequenced the Metriorrhynchini beetle fauna, sampled from ~700 localities in three continents. The species-rich dataset included ~6500 terminals, ~1850 putative species delimited at 5% uncorrected pairwise threshold, possibly ~1000 of them unknown to science. Neither type of data could alone answer our questions on biodiversity and phylogeny. The phylogenomic backbone enabled the integrative delimitation of robustly defined natural genus-group units that will inform future research. Using constrained mtDNA analysis, we identified the spatial structure of species diversity, very high species-level endemism, and a biodiversity hotspot in New Guinea. We suggest that focused field research and subsequent laboratory and bioinformatic workflow steps would substantially accelerate the inventorying of any hyperdiverse tropical group with several thousand species. The outcome would be a scaffold for the incorporation of further data from environmental sequencing and ecological studies. The database of sequences could set a benchmark for the spatiotemporal evaluation of biodiversity, would support evidence-based conservation planning, and would provide a robust framework for systematic, biogeographic, and evolutionary studies.

*For correspondence:
ladislav.bocak@upol.cz

[†]These authors contributed
equally to this work

Competing interest: The authors
declare that no competing
interests exist.

Funding: See page 16

Received: 02 July 2021

Preprinted: 15 July 2021

Accepted: 18 December 2021

Published: 20 December 2021

Reviewing Editor: David
Donoso, Escuela Politécnica
Nacional, Ecuador

© Copyright Motyka *et al.* This
article is distributed under the
terms of the [Creative Commons
Attribution License](https://creativecommons.org/licenses/by/4.0/), which
permits unrestricted use and
redistribution provided that the
original author and source are
credited.

Editor's evaluation

This manuscript provides clear ideas regarding the usage of next-generation sequencing data, and of more traditional mtDNA markers, to rapidly increase biodiversity inventories. You demonstrate how biodiversity information analyses done in the Metriorrhynchini, a hyperdiverse tropical insect group, can be rapidly expanded via targeted field research and large-scale sequencing. The study sets a benchmark for the spatiotemporal evaluation of tropical biodiversity, supports evidence-based conservation planning, and provides a robust framework for systematic, biogeographic, and evolutionary studies.

Introduction

The number of known insects surpasses that of all other terrestrial groups (*Mora et al., 2011*), and we need much more detailed information to fully understand their diversity. Currently, the available biodiversity data are far from complete, and the majority of insect species remain undescribed (*Novotny et al., 2006; Srivathsan et al., 2019*). In addition, robust phylogenetic hypotheses are lacking for

most lineages, and the genera and tribes are often artificial assemblages which are not relevant to evolutionary and biodiversity research. Therefore, we need to gather new information in order to advance our understanding of evolutionary and genetic relationships, and to build a phylogenetic scaffold for comprehensive taxonomic, biogeographic, and evolutionary studies that would be indispensable for biodiversity management.

Descriptive, morphology-based insect systematics is not keeping pace with the rapid loss and degradation of natural habitats (Theng *et al.*, 2020), and with the ongoing decline in insect abundance as a result of human activities and climate change (van Klink *et al.*, 2020). The largest taxonomic journal, *Zootaxa*, published almost 30,000 studies describing >60,000 new species and these represent over a quarter of all new animal species reported in 2001–2020 (Zhang, 2021). Although these numbers are impressive, they also show how labor-intensive is taxonomic research if, in average, only two new species are reported in a publication. To accelerate the cataloguing of biodiversity, it is vital to gather new material suitable for molecular analyses and to combine available molecular methods with traditional approaches (Riedel *et al.*, 2013; Srivathsan *et al.*, 2019; Yeo *et al.*, 2020; Sharkey *et al.*, 2021). DNA data are indisputably a valuable source for modern biodiversity research, and they can address both shallow and deep relationships (Tautz *et al.*, 2003; Hajibabaei *et al.*, 2007). There are two principal sources of short-fragment data: voucher-based DNA sequences typically produced by systematists (Riedel *et al.*, 2013; Yeo *et al.*, 2020; Sharkey *et al.*, 2021), and DNA sequences produced by an ecosystem-based sequencing that does not associate individual samples with Linnean names (Andújar *et al.*, 2015; Srivathsan *et al.*, 2019). It is the responsibility of systematic biologists to assemble the natural system, that is, we need to reliably delimit genus- and tribe-level taxa, to make their ecological and distribution attributes informative. But the short fragments are often unsuitable for the building of deep phylogenies and large genomic datasets need to be used (McKenna *et al.*, 2019; Baca *et al.*, 2021). These include transcriptomes, whole genome

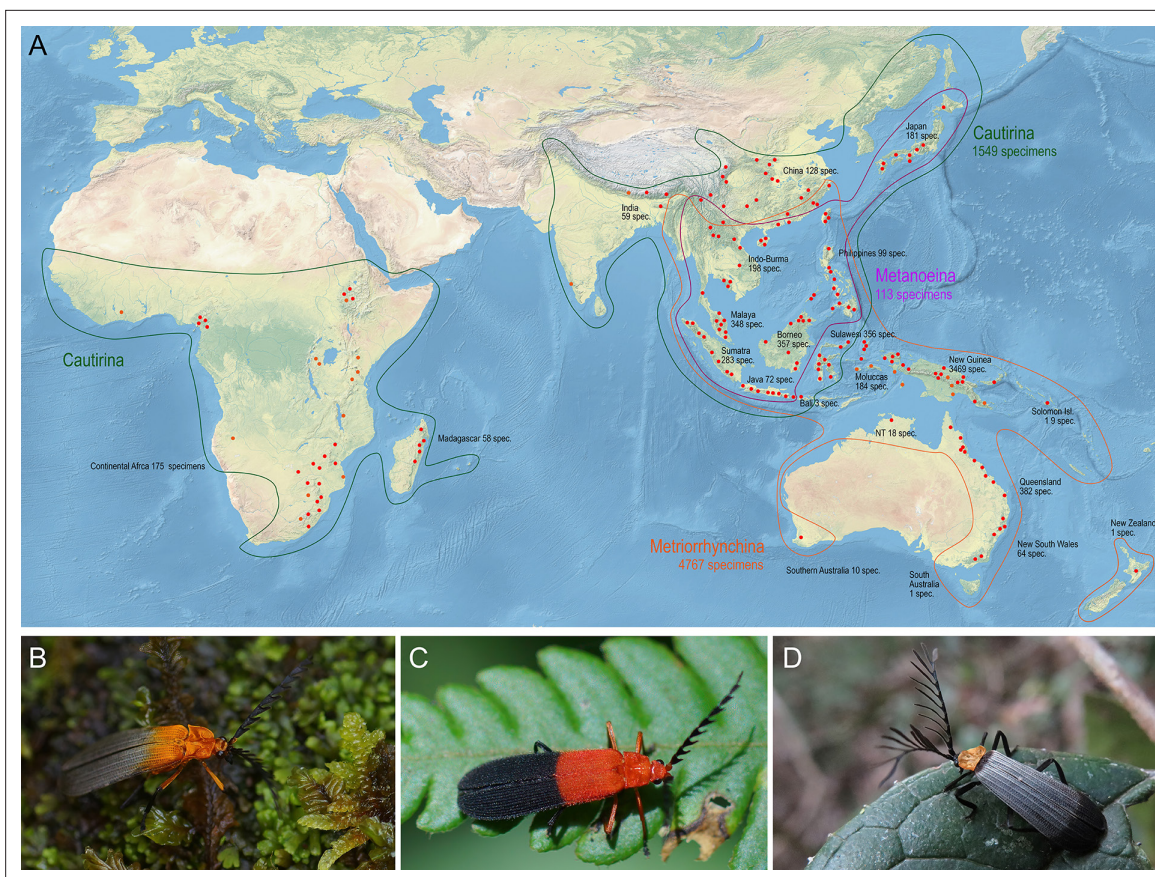


Figure 1. Distribution and appearance of metriorrhynchine net-winged beetles. **(A)** Distribution of Metriorrhynchini with major sampled localities designated by red dots. The numbers of analyzed specimens from individual regions are shown for regions and subtribes. **(B–D)** – General appearance of Metriorrhynchini.

sequences and anchored hybrid capture datasets. Using all information, a robust and stable natural classification will significantly facilitate detailed research into the spatial and temporal distribution of biodiversity (Morrison *et al.*, 2009; Thomson *et al.*, 2018). As an ultimate goal, we should attempt to construct a complete tree of life, or at least its backbone, which is invaluable in aiding the selection of groups for more detailed analyses (Chesters, 2017; McKenna *et al.*, 2019). With a well-defined high-level classification, it is paramount to exploit all accessible data. We assume that voucher-based molecular phylogenies provide much-needed tools to researchers working on site-based biodiversity assessments (Andújar *et al.*, 2015; Srivathsan *et al.*, 2019) and that, in turn, the data produced by environmental and ecosystem-focused sequencing contribute to building the tree-of-life (Arribas *et al.*, 2016; Bocak *et al.*, 2016).

We have used hyperdiverse tropical metriorrhynchine beetles (Coleoptera, Lycidae, Metriorrhynchini) as our model. This net-winged beetle tribe contains >1500 recognised species, mostly found in the Old-World tropics (Figure 1A), and their classification is complicated by the complex taxonomic history (Bocak *et al.*, 2020; see Appendix introductory information). The phenetic plasticity of Metriorrhynchini is relatively high (Figure 1B–D), but many distant species resemble each other due to convergent selection in Mullerian rings (Bocek *et al.*, 2019; Motyka *et al.*, 2020; Motyka *et al.*, 2021). Therefore, unrelated taxa were often assumed to be closely related due to misleading morphological similarities. Although there are over 40 genera in the tribe, three-quarters of the species have been described in five ambiguously defined genera (*Xylobanus*, *Cautires*, *Trichalus*, *Metriorrhynchus*, and *Cladophorus*). Sometimes a single genus contains species from different subtribes (Bocak *et al.*, 2020). In this respect, the Metriorrhynchini is a typical species-rich tropical insect group without well-founded classification and the paucity and inaccuracy of available data (Letsch *et al.*, 2020). As a result, unlike vertebrates, these poorly known insect groups have not been considered for use in large-scale, integrative projects and data meta-analyses (Myers *et al.*, 2000; Holt *et al.*, 2013) and have contributed little to our understanding of global biodiversity patterns.

The principal objective of this study is to demonstrate how biodiversity information for a hyperdiverse tropical group can be rapidly expanded via targeted field research and large-scale sequencing of transcriptomes, genomes, and short mtDNA fragments. Our investigation comprised four distinct steps. First, we assembled material from several hundred localities on three continents (Figure 1, Table 1). Second, as hyperdiverse groups are difficult to tackle and the current classification is unreliable, we attempted to compartmentalise diversity using phylogenomics. We then produced a tree,

Table 1. The numbers of sampled localities per region. Details in **Appendix 1—table 1**.

Area	Localities	Area	Localities
Australian region	298	Sino-Jap. region	79
Australia	118	China	51
New Guinea & Solomons	179	Japan	28
New Zealand	1		
Wallacea	49		
Moluccas	15	Oriental region	206
Sulawesi	34	S.India & Ceylon	3
		E.India & Burma	12
Afrotropical Region	64	E.Indo-Burma	44
West Africa	1	Malay Peninsula	57
Guinean Gulf	11	Sumatra	23
Ethiopia	6	Java & Bali	15
East Africa	10	Philippines	33
South Africa	25		
Madagascar	11	Total	696

using all available data, to estimate species limits, intraspecific genetic variability, and species ranges. Finally, the tree was pruned and used to estimate shallow phylogenetic relationships, total and regional species diversity, and endemicity, and to define generic ranges and continental-scale range shifts. The applied methods of transcriptome and mtDNA analyses are widely used. The genomic datasets dominate among works focusing on deep relationships (transcriptomes and anchored hybrid capture; *Misof et al., 2014; Kusy et al., 2019; McKenna et al., 2019; Baca et al., 2021*). The mitochondrial markers have been used mainly to study the phylogeny of restricted clades (*Toussaint et al., 2014; Sharkey et al., 2021*). Until now, the genomic and mitochondrial data have seldom been combined to get simultaneously the phylogenetic backbone for the mid-rank classification (subtribes, groups of genera, generic limits) and the estimations of species diversity (e.g. *Talavera et al., 2021*). Our information and phylogenetic hypotheses can be a resource for higher level phylogenetics, population genetics, phylogeographic studies, and biodiversity estimation. At the same time, we want to show how limited our taxonomical knowledge is and how this lack is hindering biodiversity research and management (*Thomson et al., 2018*).

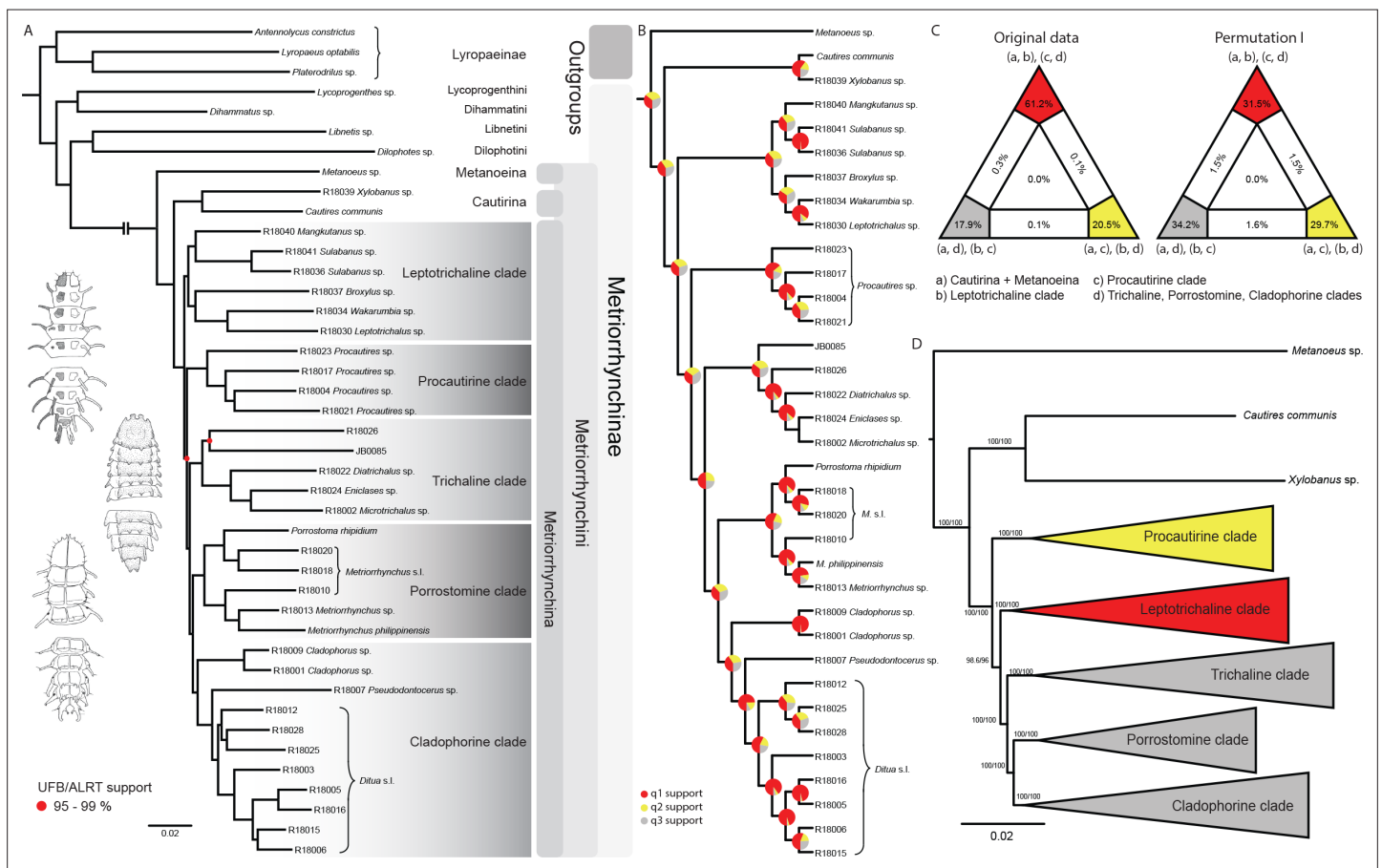


Figure 2. Topologies recovered by phylogenomic analyses. **(A)** Phylogenetic relationships of Metriorrhynchinae based on the ML analyses of the concatenated amino-acid sequence data of supermatrix F-1490-AA-Bacoca-decisive. Unmarked branches are supported by 100/100 UFB/alrt; red circles depict lower phylogenetic branch support. **(B)** Phylogenetic relationships of Metriorrhynchini recovered by the coalescent phylogenetic analysis with ASTRAL when analysing the full set of gene trees (4109 gene trees inferred at the nucleotide level). Pie charts on branches show ASTRAL quartet support (quartet-based frequencies of alternative quadripartition topologies around a given internode). Outgroups taxa are not shown. **(C)** Results of FcLM analyses for selected phylogenetic hypotheses applied at the amino-acid sequence level (supermatrix F). **(D)** Alternative phylogenetic relationships of Metriorrhynchinae based on the ML analyses of the concatenated amino-acid sequence data of supermatrix A-4109-AA. Numbers depict phylogenetic branch support values based on 5000 ultrafast bootstrap replicates.

Results

Sampling of the Metriorrhynchini range

In total, we monitored almost 800 localities, 696 of them with occurrences of the Metriorrhynchini (Tabs. 1, **Appendix 1—table 1**). The distribution of sampling sites was partly biased due to the large extent of the Metriorrhynchini range, limited time and funds, different goals of various expeditions, and logistic problems (inaccessible regions, legal obstacles). The densest sampling is available from the Sundaland and New Guinea, while India and the Afrotropical region are under-sampled (**Figure 1A**).

Assembly of the phylogenomic tree

The phylogenomic dataset contained 35 Metriorrhynchini terminals (**Appendix 1—table 2**), seven outgroups, and ~4200 orthologs ($1.9\text{--}5.7 \times 10^6$ aligned positions; **Supplementary file 1**;

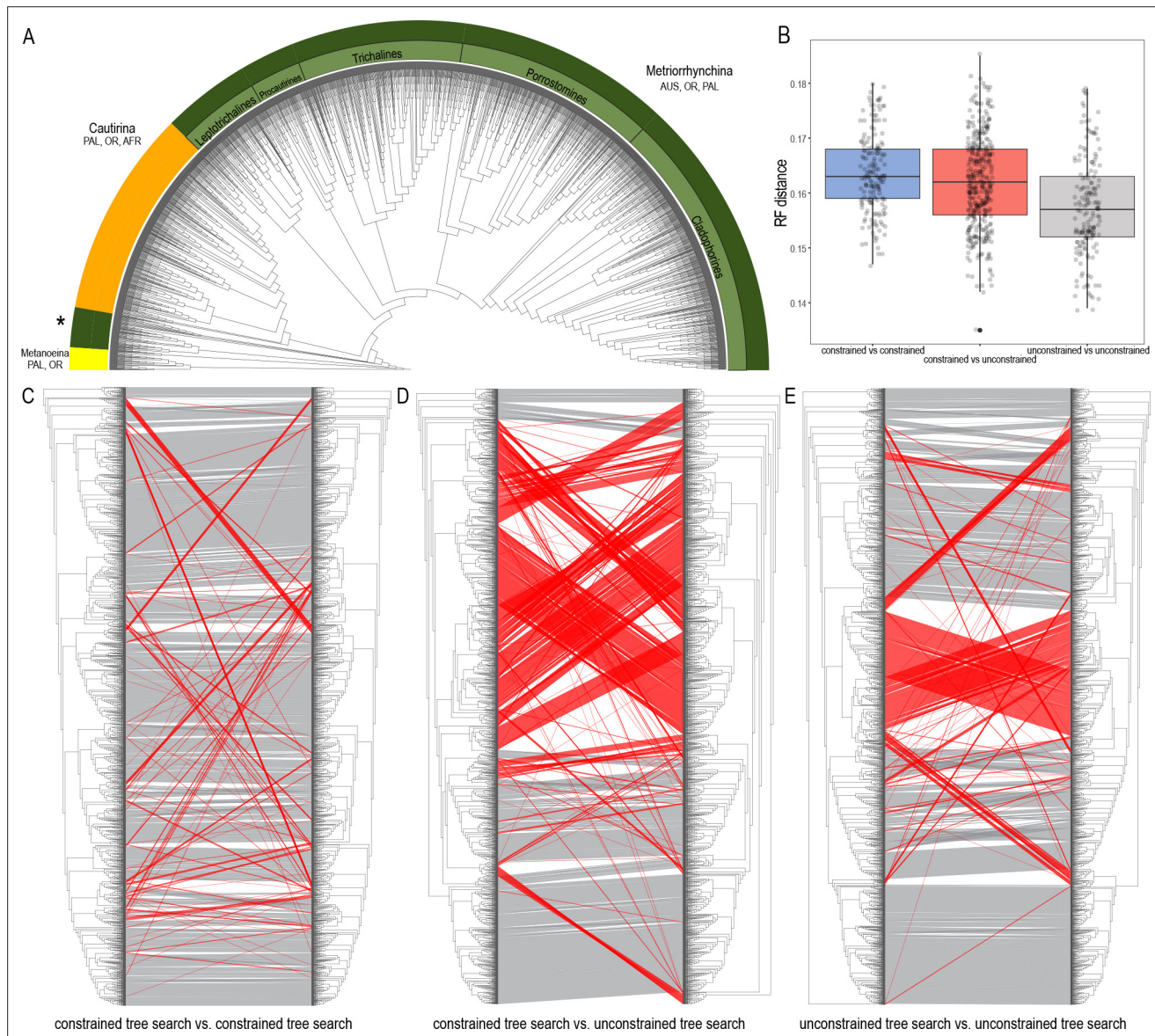


Figure 3. Topologies recovered by mitogenomic analyses. **(A)** Relationships of Metriorrhynchini recovered by the constrained analysis of the pruned dataset at 2% distance. (The full resolution tree is shown in **Source data 2** along with a tree recovered from the analysis of a complete dataset of 6429 terminals in **Source data 1**), asterisk designates a grade of Metriorrhynchina-like taxa found in a position in conflict with their morphology; **(B)** A chart of Robinson-Foulds distances among topologies inferred by repeated runs of the constrained and unconstrained analyses; **(C)** A comparison of the results obtained by two runs of the constrained analysis; **(D)** A comparison of trees inferred with/without the phylogenomic backbone; **(E)** A comparison of results obtained by two runs of the unconstrained analysis. The red lines designate terminals with conflicting positions in compared trees.

Appendix 1—table 3; Appendix 1—table 4). The tree shown in **Figure 2A** was produced using maximum likelihood (ML) analyses, whereas the coalescent method produced the topology shown in **Figure 2B**; additional trees are shown in **Appendix 1—figures 1–8**. For details on the data sets' characteristics see **Appendix 1—figures 9–12**. Phylogenomic analyses resolved three subtribes (Metanoeina [Metriorrhynchina, Cautirina]), and five clades were regularly recovered within the Metriorrhynchina, that is, the hereby defined procautirines, leptotrichalines, trichalines, porrostomines, and cladophorines. Different settings (see Materials and methods) produced slightly different topologies and shifted the positions of the leptotrichalines and procautirines (**Figure 3D**). However, the monophyly of major subclades was not affected. The FcLM analysis favored a deeper position for the leptotrichaline clade (61.2%; **Figure 2C, Appendix 1—figure 13**). The position of the remaining terminals was stable across all analyses. All phylogenomic analyses question the definitions of some species rich genera (**Appendix 1—figures 1–8**) that are either polyphyletic (e.g. *Cladophorus*; 131 described species, most of them recovered in the *Ditua* subclade) or paraphyletic (*Metriorrhynchus* as a grade and *Porrostoma* in the terminal position; 194 species, see **Figure 2A and B, Appendix 1—figures 1–8**).

Constrained mitogenomics

The mtDNA database contained >11,500 mtDNA fragments (5935 *cox1*, 2381 *rrnL*, and 3205 *nad5*) representing 6429 terminals (2930 aligned positions). Using these data, we inferred additional trees using the constrained positions of 35 terminals whose relationships were determined through

Table 2. The numbers of described species and identified mOTUs (molecular operational taxonomic units) at 2% and 5% thresholds per region and the total number of OTUs identified for subtribes. Based on morphological identification, the OTUs of the kassemiine and other deeply rooted clades are included in Metriorrhynchina.

Region	Metriorrhynchina described/ analyzed at 2%/5%		Cautirina described/ analyzed at 2%/5%	Metanoeina described/ analyzed at 2%/5%	Metriorrhynchini described/ analyzed at 2%/5%	Ratio Analyzed/ described
Australian region	639/1608/1239			639/1608/1239	2.52–1.93	
Australia	196/167/133			196/167/131	0.85–0.67	
New Guinea	423/1434/1105			423/1434/1105	3.39–2.61	
Solomon Isl.	21/9/9			21/9/9	0.43	
Wallacea	162/174/162	14/10/9		176/184/171	1.05–0.97	
Philippines	51/18/18	45/12/12	8/3/3	104/33/33	0.32	
Continental Asia	43/52/42	331/330/257	30/34/31	404/416/330	1.03–0.82	
Sundaland	36/44/39	201/184/146	24/19/17	261/247/202	0.95–0.77	
Indo-Burma	6/7/7	62/52/42	3/4/4	74/63/53	0.85–0.72	
China, Japan	1/1/1	53/75/58	1/11/11	55/87/70	1.58–1.27	
India		35/19/18	2/0/0	37/19/18	0.51–0.49	
Afrotropical region	231/104/94			231/104/94	0.46–0.41	
Sub-Saharan Africa	178/74/65			178/74/65	0.42–0.37	
Madagascar	53/30/29			53/30/29	0.57	
Total number of OTUs	895/1852/1445	641/456/369	38/37/34	1574/2345/1848	1.50–1.17	

phylogenomic analyses, and the free positions of the other ~6400 terminals (**Source data 1**). The units based on uncorrected pairwise distances represent molecular operational taxonomic units (mOTUs), considered to be putative species, or 'species' for short. Depending on the applied 2% and 5% thresholds, we identified 34–37 mOTUs in the Metanoecina clade and 369–456 mOTUs in Cautirina. The major Metriorrhynchina clade (1376–1763 mOTUs) included procautirines, leptotrichalines, trichalines, porrostomines, and cladophorines. In addition, we identified several deeply rooted lineages, the kassemiines, and another five small clades (69–89 mOTUs in total; **Source data 2**), each of which comprised a limited number of species. As phylogenomic data for these terminals are still lacking, their positions were determined based only on mtDNA data and they are included in Metriorrhynchina, based on morphological traits (**Table 2**). The number of mOTUs does not include ~50 mOTUs for which *cox1* was unavailable.

Pruned mitogenomic tree with and without constraints

The dataset was subsequently pruned to a single terminal per mOTU based on 2% and 5% distance (see below) and was analyzed both with and without topological constraints (**Figure 3A**; **Source data 2 and 3**) show the pruned trees at 2% levels to capture the intraspecific genetic variability within the clusters of closely related mOTUs). Repeated runs with different starting seeds identified terminals with unstable positions (**Figure 3A–C**). The major clades were generally stable, whereas small, deeply rooted clades were prone to 'wandering' around the tree, as were distinct singletons. The trees that resulted from each of the seed-specific 19 ML runs differed slightly; tree similarity was thus evaluated using the Robinson-Foulds index, with values ranging from 0.180 (most similar) to 0.147 (most distant; **Appendix 1—table 4**).

Tree congruence

The degree of incongruence between selected topologies is shown in **Figure 3C–D**. The unconstrained analysis of mitochondrial data yielded a topology with a high number of terminals that were recovered in positions incongruent with their morphology (**Figure 3D and E**, **Source data 3**). The same dataset, when analyzed using the constrained position of 35 terminals (based on their relative relationships inferred by prior phylogenomic analyses), produced a topology with a much lower proportion of terminals in dubious positions (**Figure 3C**, **Source data 2**). The composition of the constituent clades is based on the topology recovered by the constrained mtDNA analyses and the position of genera was validated by morphological comparisons of vouchers with the type species of earlier described genera (all redescribed by **Bocak, 2002**). The named genera assigned into individual subclades are shown individual clades are characterized in Appendix results.

Species diversity

To investigate the total and regional species diversity of the Metriorrhynchini, we analyzed a dataset comprising 5935 of the 6429 terminals for which the *cox1* mtDNA fragment was available (**Figure 3A**; **Supplementary file 1**). For the Metriorrhynchini, we identified 1848 and 2345 mOTUs using the 5% and 2% thresholds, respectively (**Appendix 1—figure 14**). We disregarded the presence of ~50 mOTUs (494 terminals) for which *cox1* was missing. The number of mOTUs based on the *cox1* analysis varied by thresholds and the number of delimited OTUs increased relatively slowly with decreasing threshold values from 1% to 10% (**Appendix 1—figure 14**).

Using an earlier published literature review (**Bocak et al., 2020**), we updated species lists for the Cautirina (641 spp. described species), Metanoecina (38 spp.), and Metriorrhynchina by adding taxa described in 2020 and 2021. By analysing DNA data, we identified 34–37 putative spp. of Metanoecina, 369–456 spp. of Cautirina, and 1445–1852 spp. of Metriorrhynchina, depending on the applied mtDNA uncorrected pairwise 5% and 2% mtDNA distance thresholds. The numbers of species per subregion, along with the estimated ratios between formally described and estimated species diversity, are shown in **Table 2** for the 2% and 5% threshold (further information in **Appendix 1—figure 14**). Using both thresholds, 2% and 5%, the numbers of putative species surpass the numbers of species reported in the literature.

We observed very high species turnover even if 5% threshold was applied for delimitation. Only four mOTUs have been recorded in two landmasses separated by a deep-sea (> 200 m). The faunas of Sulawesi and the islands across Wallace's and Weber's lines share two mOTUs, one mOTU was

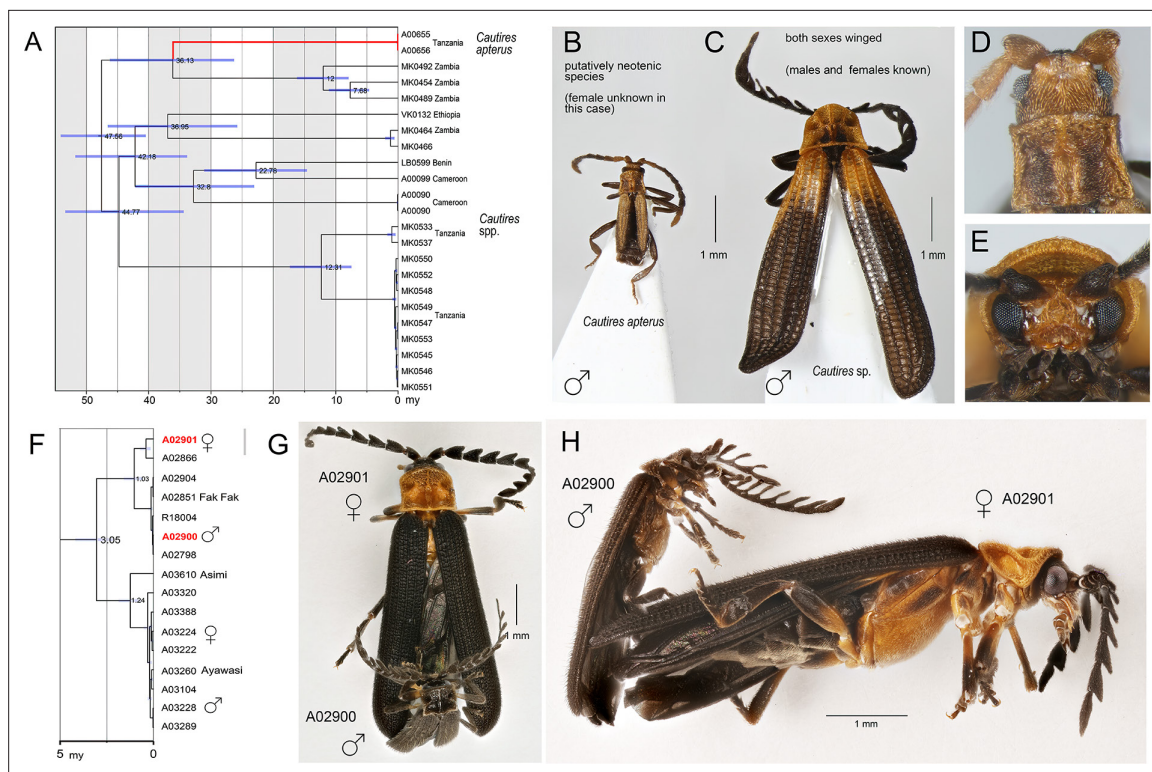


Figure 4. Identification of sexual dimorphism by large-scale biodiversity inventory. (A) Relationships of lineages with modified ontogeny, the dated tree; (B, D) General appearance and head of *Cautires apterus*, a putative neotenic species; (C, E) ditto of the close relative with both sexes winged. Mimetic sexual dimorphism identified during diversity survey. (F) The dated tree, red colored terminal labels designate the individuals shown in G and H; (G) Dorsal view of individuals in copula; (H) Ditto, lateral view. Except of collecting individuals in copula, DNA-based assessment of relationships is the only option as the species are sexually dimorphic and no morphological traits indicate their conspecificity.

simultaneously identified in Laos and Luzon and one species in New Guinea and the Solomon islands. Similarly, only sixteen mOTUs were distributed across two landmasses separated by an inundated shelf (sea depth <100 m). Nine mOTUs were distributed in two or more islands of Southeast Asia and seven species were found in both New Guinea and Australia. The centres of species diversity of the Metriorrhynchini are New Guinea (1,105 putative spp. at 5% threshold) and the seasonally to perennially humid areas of the Sundaland (202 spp. at 5% threshold). The results suggest substantial modifications to the generic limits and ranges for numerous taxa that had been previously delimited (**Appendix 1—figure 15**).

Having the extensive the mtDNA topologies, we looked for examples of the evolution of neoteny and mimetic polymorphism. The detailed inspection of trees identifies the closest available relative of a putative neotenic, *Cautires apterus* (Cautirina). This species is morphologically very distinct (**Figure 4B and D**). The dated subtree indicates the recent origin of morphological divergence (**Figure 4A**). The mtDNA analyses recover some species with pronounced sexual dimorphism, such as an unidentified genus and species of the procautirine clade (**Figure 4F–H**). The origin of the polymorphism is putatively very recent (**Figure 4F**).

Discussion

In the context of the present loss of biodiversity (**Sodhi et al., 2004; Hallmann et al., 2017; Theng et al., 2020**), large-scale genomic resources are urgently needed for biodiversity assessment and conservation (**Hajibabaei et al., 2007; Krehenwinkel et al., 2019**). Molecular data cannot replace morphology-based taxonomy (**Figure 3C–D; Thomson et al., 2018**), but the analyses of our dataset complement and facilitate traditional biodiversity research in several directions. Our first step is to compartmentalize hyperdiverse Metriorrhynchini into manageable natural units (**Figure 2**). The densely sampled phylogeny identifies tribal and generic limits. It provides a useful foundation for

detailed taxonomic research through the identification of weak areas in earlier classifications and points out the clades with undescribed diversity and non-monophyletic genera (e.g. several hundred species of *Ditua* and paraphyletic *Metriorrhynchus*; **Figure 3, Appendix 1—figure 15**). Furthermore, the analyses of species-rich datasets identify the areas with high species diversity as one of the critical conservation value parameters (**Table 2; Baselga, 2010; Srivathsan et al., 2019**). Traditional taxonomic research costs time and money, and the number of newly described species is relatively low if we consider the enormous diversity of tropical insects (**Novotny et al., 2006; Sangster and Luksenburg, 2015**). Therefore, we use DNA-based units as a provisional descriptor of species diversity (**Hebert et al., 2003; Monaghan et al., 2009**), and subsequently as a source for integrative taxonomy (**Source data 1–3; Srivathsan et al., 2019**). The presented large-scale monitoring project provides information on relationships (**Figures 2 and 3**), genetic divergence (**Source data 1–3**), turnover (**Table 2**), the extent of generic and species ranges (**Appendix 1—figure 15, Source data 1–3**), and on evolutionary phenomena that are usually studied using a few model organisms (**Figure 4**). Using phylogenomics and voucher-based sequencing, we show that taxonomic literature has provided insufficient and sometimes erroneous information, even after the formal consolidation of scattered descriptions (**Bocak et al., 2020**). We show that a taxon-focused continental scale project can effectively assemble comprehensive data for diversity of tropical insects.

Continent-wide taxon-specific monitoring of biodiversity: feasibility and impediments

Tissue and DNA archives have become critical in the assessment of biodiversity status (**Hajibabaei et al., 2007; Blom, 2021**). Although museomics is a potentially valuable source of data (**Gauthier et al., 2020**), in our case, museum collections are insufficient for filling data gaps due to the scarcity of material. For example, the *Metriorrhynchini* collection deposited in the Natural History Museum in London contains <3000 specimens, whereas there are ~6500 terminals in our dataset. At the beginning of our study, we faced critical absence of primary data. Therefore, we conducted intensive field research to obtain samples for a realistic assessment of the extant *Metriorrhynchini* diversity. We processed samples from our expeditions (most of which were focused on a range of topics over two decades between 2001 and 2019) and samples obtained through extensive collaboration with other researchers, both local and visiting, and with local naturalists whose contribution has increased with the growing number of citizen science projects (**Jaskula et al., 2021; MacPhail and Colla, 2020**). In such a way, we assembled a *Metriorrhynchini* tissue collection from almost 700 localities in three continents (**Table 1, Figure 1**). For several reasons our sampling is partly biased. We noted the serious loss of natural habitat in many regions. Previously described species were often collected in vicinity of seaports, but the lowland ecosystems are rapidly disappearing due to human exploitation. Therefore, type localities of many described species could not be sampled during recent expeditions and species known from museum collections are missing in our DNA dataset (**Jiruskova et al., 2019**). The habitat loss in South East Asia also affects other animal groups, and lowland primary forests are seriously endangered in the whole region (**Sodhi et al., 2004; Theng et al., 2020**). Further sampling bias is a consequence of the unsafe conditions and logistic problems in large areas of West Africa, Sahel, and the Congo Basin (**Figure 1A**), where net-winged beetles have not been systematically studied since the 1930 s. Additional data gaps are caused by strict biodiversity research restrictions (**Prathapan, 2018**). Regardless of these limitations, we believe that the assembled dataset is a foundation for a robust classification framework and a soundly based assessment of biodiversity. Our results show the importance of field research for biodiversity studies and systematics (**Basset and Lamarre, 2019**).

Phylogenetic relationships: a scaffold for targeted research

Unresolved taxonomy is a common reason for the exclusion of specific groups from biodiversity research projects and this omission has an effect on conservation policies (**Gutiérrez and Helgen, 2013**). The current phylogenomic and mitogenomic phylogenetic hypotheses (**Figures 2 and 3; Source data 1–3**) supersede the morphology-based topologies (**Bocak, 2002**). The phylogenomic analysis incorporates a large amount of information, and we favor this method over morphological traits and short DNA sequences, both of which contain uncertainties (**McKenna et al., 2019**). Phylogenomics has resolved subtribe relationships and their internal structures. The analyzed 35 transcripts and low-coverage genomes were sufficient to identify five major *Metriorrhynchina* clades with

several hundred putative species each and also to identify the limits of genera, which can be tested using traditional taxonomic methods (**Figures 2 and 3A**; **Source data 1–3**).

The sampling strategy is critical for building a phylogenomic backbone. Our goal was to cover as many deep lineages as possible and simultaneously to limit the number of sequenced RNA samples to avoid high costs. Therefore, we sequenced RNAlater preserved tissues and conspecific vouchers prior to assigning tissue samples for transcriptomic analyses. In this way, two rounds of sequencing provided us with critical information based on evenly distributed anchor taxa. In the next step, we re-analyzed the short-fragment dataset (**Supplementary file 1**) using constrained positions for taxa whose relationships had already been recovered through phylogenomics (**Figures 2 and 3**). A stabilized phylogenomic backbone is critical for inferring a robust topology because the large-scale analyses of short mtDNA fragments are sensitive, even to the application of starting seeds, and they often produce topologies incongruent with morphological traits (**Figure 3B and E**; **Sklenarova et al., 2014**). Only several small lineages have remained unanchored by genomic data, owing to a lack of properly fixed samples (**Source data 2 and 3**). For example, four small clades are much more deeply rooted than their morphology suggests (**Figure 3A**, **Source data 2**) and additional data are needed to place them in a phylogenetic context. Despite some contentious relationships that need further investigation, 35 genomic samples, that is under 2% of species, sufficiently supported relationships among most terminal clades that approximately represent genera, groups of genera and subtribes. **Talavera et al., 2021** have shown that only six nuclear markers for 5%–10% of terminals can similarly stabilize the phylogeny of a species rich model group. We assume, that the combination of genomic and short DNA data can be valuable for building of the species-level trees of life.

We identified a substantial conflict between phylogenomic analyses, morphology-based classifications (**Bocak, 2002** and earlier studies cited therein), and the analyses of a few short DNA fragments (**Sklenarova et al., 2014**, Appendix introductory information). Our analyses confirm the monophyly of the recently described *Cautirina* and *Metanoeina* and redefined *Metriorrhynchina* except several unanchored lineages (**Sklenarova et al., 2013**; **Sklenarova et al., 2014**), but, for the first time, we can robustly recover subtribal relationships (**Figure 2A and B**). We reject most internal splits suggested by morphological and mtDNA analyses (**Bocak, 2002**; **Sklenarova et al., 2013**; **Sklenarova et al., 2014** and earlier studies cited therein). The present delimitation of five monophyla within *Metriorrhynchina* resolves the backbone of the subtribe that was contentious due to high levels of homoplasy in Sanger and morphology-based datasets. Similarly, some generic concepts are questioned as they have been mostly based on highly homoplastic traits (**Sklenarova et al., 2014**; **Kusy et al., 2019**). The taxonomic studies must consider the morphology along with molecular hypotheses. As morphology is not described in this study, we do not discuss the limits of individual genera and report only short information on newly defined subclades (Appendix results).

Our approach yielded a constrained phylogeny with 6429 terminals and almost 2000 mOTUs using 5% mtDNA distance threshold, and this provides the basis for the approximation of species diversity for constituent subclades and geographic regions (**Figure 3**; **Source data 2 and 3**). Concerning the extent of diversity, phylogenomic and mitochondrial data must be simultaneously analysed to provide a strong foundation for subsequent investigations (**Figure 3C**, **Talavera et al., 2021**). Phylogenomics cannot deal with thousands of species, and mitogenomic data are insufficient for the construction of robust deep relationships. The final steps are morphological validation of the proposed generic groups and genera (see Appendix results) and, in the future, formal descriptions of biodiversity using the Linnean classification. In such a way, the results of phylogenomic and mitogenomic inventory should be incorporated in the formal classification (**Godfray and Knapp, 2004**).

Species diversity: literature data and reality

Here, we deal with a species rich tropical beetle tribe (> 1500 described species), and therefore we use the uncorrected pairwise distance thresholds for our diversity estimation (**Table 2**, **Appendix 1—figure 14**). The application of any threshold is a compromise between estimation accuracy, speed, and sequencing costs, taking into account the feasibility of inventorying a hyperdiverse group within a limited time frame (**Hebert et al., 2003**; **Dupuis et al., 2012**; **Eberle et al., 2020**). Several taxonomical works on *Metriorrhynchini* have simultaneously analysed mtDNA fragments, nuclear genes, morphology, and ecology (e.g. **Bocak and Yagi, 2010**; **Kalousova and Bocak, 2017**; **Bocek et al., 2019**; **Jiruskova et al., 2019**), but their results cannot robustly defend an application of a distance

threshold for the whole tribe. To avoid a possibility of diversity overestimation, we base further discussion on the 5% distance. We assume that such a threshold might be sufficiently cautious. Future taxonomic revisions are surely needed for the validation of here presented data on species diversity.

When analysing the *cox1* mtDNA fragment, we identified 1,848 mOTUs at 5% threshold and the numbers of delimited mOTUs indicate that, the substantial part of species diversity remains undescribed (**Table 2, Appendix 1—figure 14**). Additionally, the slope representing the relationship between the number of mOTUs and distance thresholds was gradual (**Appendix 1—figure 14**) due to a high number of genetically distant, indisputably distinct lineages in the dataset (**Appendix 1—figure 14**).

Our approach provides information about the diversity of the internal lineages. Metriorrhynchina is by far the most diverse group, within which the cladophorines comprise the largest clade (490 mOTUs, e.g. *Ditua* historically has included 2 spp., now ~250 spp.). The porrostomine clade is the next diverse group (373 mOTUs) and contains the speciose *Porrostoma* (126 mOTUs) and a paraphyletic series of lineages whose species have conventionally been placed in *Metriorrhynchus*. The differences between previously published data and our results are substantial and they question any reanalyses of literature data without prior verification (**Bocak, 2002; Bocak et al., 2020** references therein; **Source data 2 and 3**).

The numbers of mOTUs must be interpreted in the context of the sampling activity in each region. We did not use standardized protocols due to the long-term character of our research, the incorporation of some samples provided by other researchers, and the necessity to apply appropriate collecting methods to maximize the number of recorded species in various ecosystems and under different weather conditions. We identified only 94 mOTUs from the Afrotropical region, mainly due to the limited number of collecting trips by authors (64 localities) and the inaccessibility of some areas. Despite intensive field research (33 localities), we collected from the Philippines less than one third of the species described (33 mOTUs). Our collection activities in the Philippines were hindered by substantial loss of natural habitats, and this is soon expected to be the case in other regions (**Sodhi et al., 2004**). The number of species known from the Sundaland (114 localities) was approximately equal to the number of sequenced mOTUs despite disproportionately intensive collecting effort by the authors. Even after numerous expeditions to the Sundaland, many regions remain unsampled. As metriorrhynchine species ranges are small (**Jiruskova et al., 2019; Motyka et al., 2020**), the number of species will probably increase in the future. The proportion of new species was regionally ~70% if DNA data and morphology were considered in detailed taxonomic studies (e.g. **Jiruskova et al., 2019**). While these regions house numerous unknown species, we found New Guinea to be exceptionally diverse, with almost three times the number of species reported in the literature (1,105 mOTUs at 5% threshold; 175 localities, **Table 1** and **Table 2**). Despite the relatively large number of sampled localities, many areas of New Guinea remain unexplored and many places were only superficially sampled by colleagues and never visited by the authors (**Figure 1A**). Additional species were added to the dataset with each batch of sequenced samples from New Guinea and the area possibly houses much higher diversity than documented by the present study.

We observed a high turnover between regions, and few species had ranges which included landmasses separated by shallow seas (seven spp. Queensland / New Guinea, 9 spp. in Southeast Asia; **Source data 1**). Poorly dispersing lycids generally have very small ranges, except for the few genera that visit flowers and fly in open areas (**Motyka et al., 2021**). A similar small-scale turnover has recently been reported along altitudinal gradients (**Bocek et al., 2019; Motyka et al., 2020; Motyka et al., 2021**). A high turnover indicates a large proportion of hidden diversity, especially in tropical mountains (**Merckx et al., 2015; Mastretta-Yanes et al., 2018**). Mountain fauna is especially vulnerable to climate change and its inventorying is urgently needed.

The Metriorrhynchini has recently received considerable attention in taxonomic studies, and 302 species have been described by several authors over the past three decades, making a total of 1574 formally described species (**Table 1, Kazantsev, 2010; Bocak et al., 2020**; Appendix introductory information). Although the recent 24% increase in described diversity appears substantial, the distance-based analysis indicates the presence of almost 2000 mOTUs (**Appendix 1—figure 14**). An additional ~50 putative species (494 terminals) were identified, but this identification was only based on divergent morphology because of the absence of *cox1*. We assume that our sampling represents only a subset of all known species (< 50%). It means that the dataset contains ~1000 undescribed

species. At the current rate, formal morphological descriptions of an additional 1000 species would take decades or hundreds of years. This is a very long time in the context of the ongoing deforestation and fragmentation of natural habitats, and currently undocumented diversity might be lost long before it can be catalogued (Brooks et al., 2002; Sodhi et al., 2004; Ceballos et al., 2015; Theng et al., 2020). The rapid DNA-based inventory is an effective shortcut for obtaining basic information on the true diversity of tropical beetles and for setting a benchmark for future biodiversity re-evaluations.

The results reveal major biodiversity hotspots in New Guinea and the Sundaland. Tropical rainforests currently cover most of New Guinea, a tectonically young island that has not been considered a biodiversity hotspot for vertebrates (Myers et al., 2000; Hall, 2011; Toussaint et al., 2014). In the case of net-winged beetles, we show that the New Guinean fauna is phylogenetically diverse, spatially heterogeneous, and extremely rich as regards both the number of species and the endemic genera (Table 1). Additionally, the large clades of New Guinean species indicate that the diversification of major lineages preceded the uplift of the islands, and possibly started on the northern margin of the Australian craton and adjacent islands. Southeast Asia is a centre of phylogenetic diversity at the tribal level; its fauna contains all principal lineages and the highest diversity of Caurina but is smaller than those of New Guinea. The Afrotropical and Palearctic regions represent only recently populated low-diversity outposts.

Impact of biodiversity inventorying on biogeographical and evolutionary research

Detailed data on Metriorrhynchini diversity indicate low dispersal propensity and this makes Metriorrhynchini a promising model for biogeographic studies (Ikeda et al., 2012). Our densely sampled phylogeny did not find any long-distance dispersal events, in contrast to many studies of flying beetles (Balke et al., 2009; Jordal, 2015). Most recovered overseas dispersal events are limited to distances

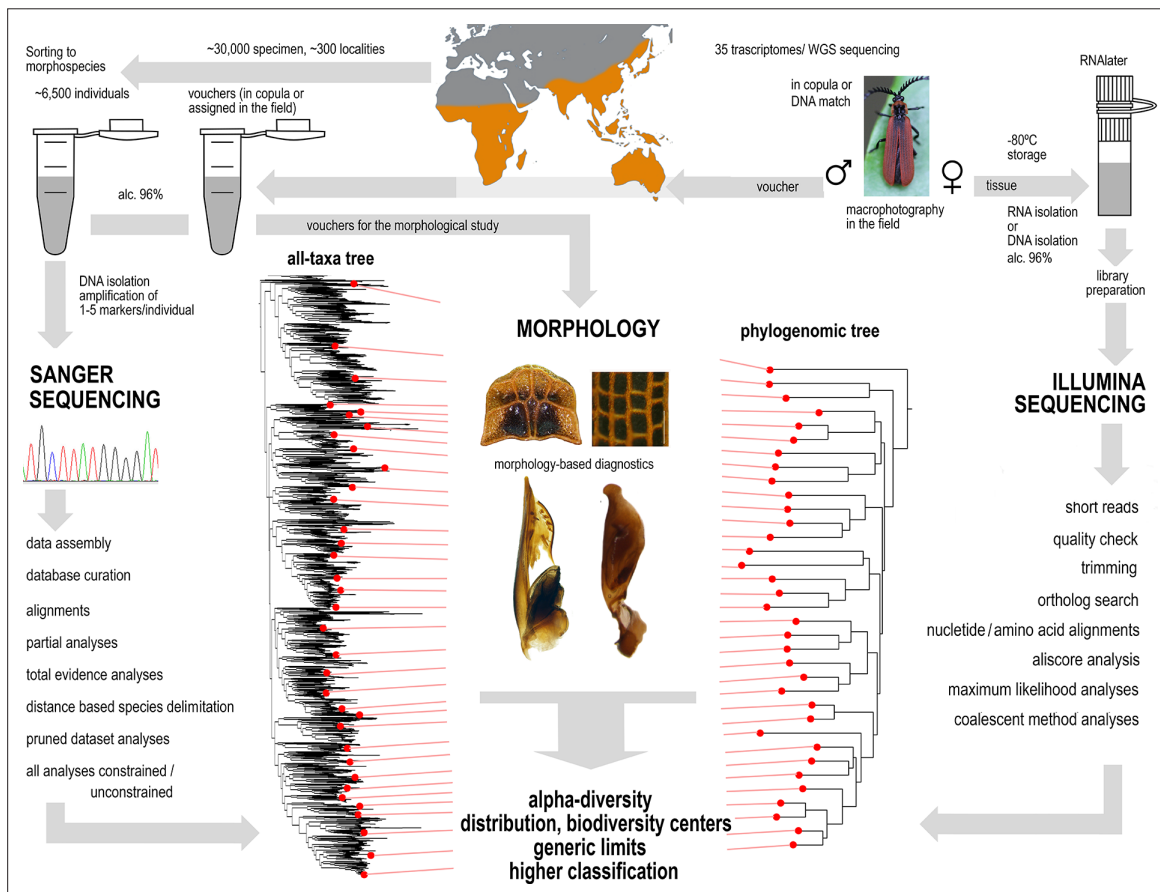


Figure 5. A sequence of applied methods from sampling to hypotheses.

of less than 100 km and are commonly accompanied by speciation (**Source data 2 and 3**). The high proportion of erroneous placement of many taxa (**Appendix 1—figure 15; Bocak et al., 2020**) renders the distribution data cited in previous literature unsuitable for phylogeographic investigations, and revision of the classification is important in order to understand the true distribution of individual taxa. The original and revised ranges of selected genera are compared in **Appendix 1—figure 15** as examples.

Intensive biodiversity research has the potential to fill knowledge gaps concerning evolutionary phenomena that are mainly studied using a small number of model species, and the research can identify the unique attributes of other potential models. We document the contribution of a large-scale biodiversity inventory to evolutionary studies with two examples.

Net-winged beetles include several lineages in which females have lost the ability to completely metamorphose (**Bocak et al., 2008; McMahon and Hayward, 2016**). If a putative neotenic species is discovered, a comprehensive reference database of the group may identify its closest relatives. We used our data to place the East African *Cautires apterus* in a phylogenetic context, and the results indicated that it may be the youngest neotenic taxon of all net-winged beetles (36.1 my, **Figure 4**).

Our extensive DNA database of metriorrhynchine diversity may also play an important role in the study of the evolution of mimicry. Our inventory identified an extreme and previously unknown aposematic dimorphism in New Guinean metriorrhynchines (**Figure 4; Figure 5**). The placement of sexually dimorphic species in the phylogeny suggests that the shift to dimorphism was very recent (3.0 mya at the earliest) and began when both sexes were small-bodied. Mimetic sexual polymorphism is well understood in butterflies with non-mimetic males and mimetic females (**Kunte, 2008**), but the advergence of males and females to different aposematic models has only recently been reported in two subfamilies of net-winged beetles (**Motyka et al., 2018; Motyka et al., 2020; Motyka et al., 2021**). Divergent evolution in Müllerian systems appears to be more common in multi-pattern aposematic rings than was previously believed when morphology was the sole source of information.

Conclusion

Priority areas for global conservation have usually been identified based on richness, species endemism and vulnerability of vertebrates (**Myers et al., 2000; Holt et al., 2013**). We assume that different patterns of biodiversity distribution can be revealed if other animal groups are studied. Reliable information on additional groups can focus our conservation efforts on valuable regions (**Morrison et al., 2009; Thomson et al., 2018**). Our model, beetles, is the most speciose group of animals but is much less known than vertebrates. Therefore, new data must be generated, and our research workflow must use innovations to economically produce the large-scale phylogenetic hypothesis for a high number of species. We conducted a worldwide sampling in ~700 localities, analysed transcriptomes, genomes, and mitochondrial markers, and validated our results with morphology. We show that the constrained position of less than 2% terminals increases the stability of tree topology and the congruence of molecular hypotheses with morphological traits (Appendix results). By the simultaneous consideration of genomic and mitochondrial phylogenetic signal, we achieved substantial progress with respect to the development of a Metriorrhynchini tree of life (**Chesters, 2017; Linard et al., 2016**). The voucher-based DNA entries established a framework for classifying samples from other studies, such as environmental sequencing (**Linard et al., 2016; Andújar et al., 2015; Arribas et al., 2016**) and for subsequent morphology-based studies. Despite limited time and funding, we identified almost 2000 mOTUs which indicate that there are at least twice more species than the number reported in the literature. This means that, at a conservative estimate, ~ 1000 species in the dataset were previously unknown to science. Furthermore, we identified New Guinea as a biodiversity hotspot, which is in clear contrast with studies identifying the biodiversity patterns of vertebrates. Our large-scale inventory shows that the literature records of tropical beetles cannot be used for biodiversity conservation and meta-analyses without critical revision. We suggest that if focused field research is conducted even by a small research group and subsequent workflow steps are applied to any hyperdiverse tropical group, the results can set a benchmark for future evaluation of spatiotemporal changes in biodiversity.

Materials and methods

Field research

The analyzed individuals had been accumulated by numerous expeditions to various regions of the *Metriorrhynchini* range (**Figure 1A**, **Appendix 1—table 1**). The distribution of sampling sites was partly biased, and no samples are available from West Africa, Congo Basin, Sahel, Sri Lanka, and the Lesser Sundas. About 10% of samples were provided by other researchers.

Tissues for transcriptomic analyses were fixed in the field. As field identification is generally unreliable, we preferred to collect pairs *in copula*, then the female was fixed using RNAlater, and the male kept separately in 96% ethanol for Sanger sequencing and the voucher collection. Alternatively, the morphologically similar individual from the same place was fixed in ethanol and the identity of an individual assigned for transcriptomic analysis was confirmed by sequencing *cox1* mtDNA using tissue from the specimen preserved in RNAlater and putatively conspecific voucher (**Figure 2**). About 100 tissue samples were fixed and thirty-five of them were used for sequencing (**Appendix 1—table 2**). Earlier published transcriptomes were added (**McKenna et al., 2019; Kusy et al., 2019**). Due to the inaccessibility of properly fixed tissue, the two critical samples were shotgun sequenced using isolated DNA.

Almost 7000 samples from 696 localities (**Table 1**) were included in the sequencing program to obtain short mtDNA fragments. In total, 6429 yielding at least a single fragment were included in the analysis (**Supplementary file 1**). The analyzed data set contained some previously published sequences (e.g. **Sklenarova et al., 2013; Bocek and Bocak, 2019**). Voucher specimens are deposited in the collection of the Laboratory of Biodiversity and Molecular Evolution, CATRIN-CRH, Olomouc.

Genomic and transcriptomic sequencing, data analysis

Libraries for thirty transcriptomes were prepared by Novogene Co., Ltd. (Beijing, China) and sequenced on the HiSeq X-ten platform (Illumina Inc, San Diego, CA). The removal of low-quality reads and TruSeq adaptor sequences were performed using fastp v.0.20.0 (**Chen et al., 2018**) with the following parameters: `-q 5 u 50 l 50 n 15`. All paired-end transcriptomic reads were assembled using SOAPdenovo-Trans-31mer (**Xie et al., 2014**).

Additionally, the total DNA (~33 Gb each) of *Metanoeus* sp. (Voucher code G19002) and an unidentified sample *Metriorrhynchina* species (Voucher JB0085) was shotgun-sequenced on the same platform. Reads were filtered with fastp using the same settings as above and quality was visualized with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). The draft genomes were assembled using SPAdes v.3.13.1 (**Bankevich et al., 2012**), with k-mer sizes of 21, 33, 55, 77, and 99. Obtained contigs were used to train Augustus (**Stanke and Waack, 2003**) for species-specific gene models with BUSCO (**Waterhouse et al., 2018**). Predicted species-specific gene models were then used for ab initio gene predictions in Augustus and predicted protein-coding sequences were used for subsequent analyses. Outgroup taxa were reported in previous studies (**Kusy et al., 2018; Kusy et al., 2019; McKenna et al., 2019**).

The ortholog set was collated by searching the OrthoDB 9.1 (**Zdobnov et al., 2017**) for single copy orthologs in six beetle genomes (**Appendix 1—table 3**). We used Orthograph v.0.6.3 (**Petersen et al., 2017**) with default settings to search in our assemblies for the presence of specified single copy orthologs. From the recovered 4193 orthologs, terminal stop codons were removed, and internal stop codons at the translational and nucleotide levels were masked. The amino acid sequences were aligned using MAFFT v.7.471 with the L-INS-i algorithm (**Katoh and Standley, 2013**). The alignments from each ortholog group were then checked for the presence of outliers. To identify random or ambiguous similarities within amino acid alignments, we used Aliscore v.2.076 with the maximum number of pairwise comparisons $-r 10^{27}$, option `-e`. and we masked them using Alicut v.2.3 (**Kück et al., 2010**). Alinuc.pl was then used to apply the Aliscore results to match amino acids to the nucleotide data. MARE v.0.1.2-rc was used to calculate the information content of each gene partition (**Misof et al., 2013**). Partitions with zero information content were removed at both levels. Finally, the remaining 4109 alignments were retained for subsequent multispecies coalescent analyses, and different concatenated datasets were generated for both amino acid and nucleotide levels using FasConCat-G v.1.4 (**Kück and Longo, 2014; Appendix 1—table 4** and Appendix methods). The degree of missing data and overall completeness scores (Ca) across all datasets were inspected using AliStat v.1.7 (**Wong et al., 2020**).

Compositional heterogeneity tests

To explore the effect of among species compositional heterogeneity and its possible bias to tree reconstruction, we inspected the data with BaCoCa v.1.105 (Kück and Struck, 2014) to identify the gene partitions that strongly deviate from compositional homogeneity using relative composition frequency variation value (RCFV). Following Vasilikopoulos et al., 2019, we considered compositional heterogeneity among species in a given partition to be high when RCFV ≥ 0.1 . The heterogeneous partitions were excluded from the data to generate a more compositionally homogeneous dataset. We used Maximum Symmetry Test (Naser-Khdour et al., 2019) to identify the partitions that strongly deviate from compositional homogeneity at the nucleotide level (p -value cut off < 0.05), and partitions below the threshold were excluded. The software SymTest v.2.0.49 (Ott, 2019) was used to calculate the overall deviation from stationarity, reversibility, and homogeneity (SRH) (Ababneh et al., 2006).

Phylogenomic maximum likelihood analyses

For all datasets, phylogenetic reconstruction was performed using the maximum likelihood (ML) criterion with IQ-TREE 2.1.2 (Minh et al., 2020). First, we analyzed all datasets using the original gene partition boundary. The model selection for each gene was performed with ModelFinder (Chernomor et al., 2016; Kalyaanamoorthy et al., 2017) implemented in IQ-TREE2 (-MFP option). GTR model was considered for nucleotide supermatrices. For the amino acid supermatrices, the substitution models LG, DCMUT, JTT, JTTDCMUT, DAYHOFF, WAG, and free rate models LG4X and LG4M were tested. All possible combinations of modeling rate heterogeneity among sites were allowed (options: -mrate E,I,G,I + G,R -gmedian -merit BIC). We used the edge-linked partitioned model for tree reconstructions (-spp option) allowing each gene to have its own rate. The optimized partition schemes and best-fitting models were inferred for some datasets using -m MFP+ MERGE and the considering same substitution models as above. The fast-relaxed clustering algorithm was used to speed up computation during partition-scheme optimization (Lanfear et al., 2017). Ultrafast bootstrap (Hoang et al., 2018) and SH-like approximate likelihood ratio test (SH-aLRT) were calculated in IQ-TREE2 (options -bb 5000 and -alrt 5000) to assess nodal supports for focal relationships.

Coalescent analyses and analyses of the confounding and alternative signal

To account for variation among gene trees owing to incomplete lineage sorting and to account for potential gene tree heterogeneity and discordance (Edwards, 2009), the data were also analyzed using the coalescent-based species-tree method. For every single gene partition, we calculated an ML gene tree in IQ-TREE2, with 5,000 ultrafast bootstrap replicates (-bb option) and using the same substitution models as predicted by ModelFinder in the above described partitioned analyses. For subsequent coalescent species tree estimation, the Accurate Species Tree Algorithm (ASTRAL-III v.5.7.3; Zhang et al., 2018) was used. To account for very poorly resolved branches on gene trees, branches with ultrafast bootstrap ≤ 10 were collapsed using newick utilities v.1.6 (Junier and Zdobnov, 2010) in every ASTRAL analysis. Local posterior probabilities (Sayyari and Mirarab, 2016) and quartet frequencies of the internal branches in every species tree were calculated using the parameter ' $t = 2$ '. Pie charts representing quartet scores for the given topology and two alternatives were plotted to the resulting species trees in R using plot_Astral_trees (Bellot, 2021).

Additionally, we studied the effect of potentially confounding signals, like non-random distribution of data coverage and violations of SRH conditions, on our phylogenetic reconstructions with the Four-cluster likelihood mapping (FCLM) approach (Strimmer and Haeseler, 1997) implemented in IQ-TREE2. Based on the results of our tree reconstructions, we tested the hypotheses about the alternative placement of leptotrichaline and procautirine clades.

Mitochondrial DNA sequencing and data analysis

Total DNA was extracted from the metathorax with a Wizard SV96 kit (Promega Corp., Madison, WI). The yield was measured using a NanoDrop-1000 Spectrophotometer (Thermo Fisher Scientific Inc, Waltham, MA). The PCR settings and cycle sequencing conditions were the same as those used by Bocak et al., 2008. Three fragments of mitochondrial genome were sequenced: *cox1*+ tRNA *Leu*+ *cox2* (~1100 bp), *rrnL*+ tRNA *Leu*+ *nad1* (~800 bp), and ~1210 bp of *nad5* and adjacent tRNA-*Phe*, tRNA-*Glu*, and tRNA-*Ser* mtDNA (the mtDNA fragments are further mentioned as *rrnL*, *cox1*, and

nad5). The PCR products were purified using PCRu96 Plates (Merck Millipore Inc, Burlington, MA) and sequenced by an ABI 3130 (Applied Biosystems, Waltham, MA) sequencer using the BigDye Terminator Cycle Sequencing Kit 1.1 (Applied Biosystems, Waltham, MA). Sequences were edited using Sequencher v.4.9 software (Gene Codes Corp., Ann Arbor, MI). Altogether 6476 individuals were analyzed including some previously published (Sklenarova et al., 2013; Bocek and Bocak, 2019).

The *cox1* gene fragment was used to OTUs delimitation (Blaxter et al., 2005) using CD-hit-est (Fu et al., 2012) and different thresholds (from similarity 0.99–0.90 by 0.05 steps). Therefore, we assembled two datasets: (A) the dataset containing all sequenced individuals and (B) all OTUs delineated by 0.98 similarity of the *cox1* gene. The *rrnL* and *tRNAs* were aligned using MAFFT 7.2 with Q-INS-I algorithm (Kato and Standley, 2013), protein-coding genes were eye-checked for stop codons and aligned using Trans-Align (Bininda-Emonds, 2005). All fragments were concatenated using FasConCat (Kück and Longo, 2014) and analyzed under maximum-likelihood criterion in IQ-TREE v.2.1.2 (Minh et al., 2020; Appendix 1—table 5). To assess the branch supports values, we used SH-aLRT test with 1000 iterations. ModelFinder tool implemented in IQ-TREE was used to identify the best fit models using the Bayesian Information Criterion (Chernomor et al., 2016). The results of the TSA/WGS analyses were used to constrain basal topology among major clades of Metriorhynchini in both analyses of datasets A and B. Further, we ran unconstrained analyses of the above-mentioned datasets with identical settings except -g option to compare results. We replicated constrained tree search nineteen-times and compared resulting trees using Robinson–Foulds distances in R package phangorn (Schliep, 2011; Appendix 1—table 2). Randomly chosen trees were then compared using cophylo script (phytools; Revell, 2012) with argument rotate = TRUE.

Acknowledgements

Several colleagues provided samples from their respective home countries or their own field research. We are obliged to G Monteith and A Slipinski for their support during our research in Australia. The field research was enabled by the permits from the Government of Papua New Guinea, the Queensland Ministry of Environment, Malaysian Ministry of Natural Resources, and UP Los Baños. The trips to Papua New Guinea were supported by the Binatang Research Centre, Nagada and we are obliged to H Maraia and J Kua for their field assistance. A substantial part of the field research before the start of the Czech Science Foundation project was funded by the senior author and the Palacky University is acknowledged for granting the leave. L Harmackova advised with some analyses in R. We cordially acknowledge I Frebort and P Banas for their encouragement and support of this project after the liquidation of the laboratory at the Palacky University. Funding: This research was funded by The Czech Science Foundation, grant number 18–14942 S.

Additional information

Funding

Funder	Grant reference number	Author
Grantová Agentura České Republiky	18-14942S	Ladislav Bocak

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Michal Motyka, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Visualization, Writing – original draft, Writing – review and editing; Dominik Kusy, Conceptualization, Data curation, Formal analysis, Methodology, Resources, Visualization, Writing – original draft, Writing – review and editing; Matej Bocek, Renata Bilkova, Data curation, Investigation, Resources, Writing – review and editing; Ladislav Bocak, Conceptualization, Data curation, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing

Author ORCIDsMatej Bocek  <http://orcid.org/0000-0002-3398-6078>Ladislav Bocak  <http://orcid.org/0000-0001-6382-8006>**Decision letter and Author response**Decision letter <https://doi.org/10.7554/eLife.71895.sa1>Author response <https://doi.org/10.7554/eLife.71895.sa2>

Additional files**Supplementary files**

- Transparent reporting form
- Supplementary file 1. Maximum likelihood tree recovered by the analysis of the full dataset (mitochondrial fragments); numbers above branches represent SH-ahlrt support values.
- Source data 1. Maximum likelihood tree recovered by the analysis of the full dataset (mitochondrial fragments). Depicted numbers above branches represent SH-ahlrt support values.
- Source data 2. Maximum likelihood tree recovered by the analysis of the reduced dataset (98% similarity OTUs). Depicted numbers above branches represent SH-aLRT support values. The taxa with the constrained position in the tree are marked with the red star.
- Source data 3. Maximum likelihood tree recovered by the analysis of the reduced dataset (mitochondrial fragments, 98% similarity OTUs) with unconstrained backbone. Depicted numbers above branches represent SH-ahlrt support values.

Data availabilityAll datasets are deposited in the Mendeley Data repository <https://doi.org/10.17632/ntgg6k4fjx.1>.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Motyka M, Kusy D, Bocek M, Bilkova R, Bocak L	2021	Data for: Phylogenomic and mitogenomic data can accelerate inventorying of tropical beetles during the current biodiversity crisis	https://doi.org/10.17632/ntgg6k4fjx.1	Mendeley Data, 10.17632/ntgg6k4fjx.1

References

- Ababneh F**, Jermin LS, Ma C, Robinson J. 2006. Matched-pairs tests of homogeneity with applications to homologous nucleotide sequences. *Bioinformatics* **22**:1225–1231. DOI: <https://doi.org/10.1093/bioinformatics/btl064>
- Andújar C**, Arribas P, Ruzicka F, Crampton-Platt A, Timmermans MJTN, Vogler AP. 2015. Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics. *Molecular Ecology* **24**:3603–3617. DOI: <https://doi.org/10.1111/mec.13195>
- Arribas P**, Andújar C, Hopkins K, Shepherd M, Vogler AP, Yu D. 2016. Metabarcoding and mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil. *Methods in Ecology and Evolution* **7**:1071–1081. DOI: <https://doi.org/10.1111/2041-210X.12557>
- Baca SM**, Gustafson GT, Alexander AM, Gough HM, Toussaint EFA. 2021. Integrative phylogenomics reveals a Permian origin of Adephaga beetles. *Systematic Entomology* **46**:968–990. DOI: <https://doi.org/10.1111/syen.12506>
- Balke M**, Ribera I, Hendrich L, Miller MA, Sagata K, Posman A, Vogler AP, Meier R. 2009. New Guinea highland origin of a widespread arthropod supertramp. *Proceedings. Biological Sciences* **276**:2359–2367. DOI: <https://doi.org/10.1098/rspb.2009.0015>, PMID: 19364747
- Bankevich A**, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**:455–477. DOI: <https://doi.org/10.1089/cmb.2012.0021>, PMID: 22506599
- Baselga A**. 2010. Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography* **19**:134–143. DOI: <https://doi.org/10.1111/j.1466-8238.2009.00490.x>
- Basset Y**, Lamarre GPA. 2019. Toward a world that values insects. *Science* **364**:1230–1231. DOI: <https://doi.org/10.1126/science.aaw7071>, PMID: 31249044

- Bellot S.** 2021. plot_Astral_trees_v2.R. 2. GitHub. https://github.com/sidonieB/scripts/blob/master/plot_Astral_trees_v2.R
- Bininda-Emonds ORP.** 2005. transAlign: using amino acids to facilitate the multiple alignment of protein-coding DNA sequences. *BMC Bioinformatics* **6**:156. DOI: <https://doi.org/10.1186/1471-2105-6-156>, PMID: 15969769
- Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E.** 2005. Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **360**:1935–1943. DOI: <https://doi.org/10.1098/rstb.2005.1725>, PMID: 16214751
- Blom MPK.** 2021. Opportunities and challenges for high-quality biodiversity tissue archives in the age of long-read sequencing. *Molecular Ecology* **30**:5935–5948. DOI: <https://doi.org/10.1111/mec.15909>
- Bocak L.** 2002. Revision and phylogenetic analysis of Metriorrhynchinae. *European Journal of Entomology* **99**:315–351. DOI: <https://doi.org/10.14411/eje.2002.043>
- Bocak L, Matsuda K.** 2003. Review of the immature stages of the family Lycidae (Insecta: Coleoptera). *Journal of Natural History* **37**:1463–1507. DOI: <https://doi.org/10.1080/00222930210125362>
- Bocak L, Bocakova M, Hunt T, Vogler AP.** 2008. Multiple ancient origins of neoteny in Lycidae(Coleoptera): consequences for ecology and macroevolution. *Proceedings. Biological Sciences* **275**:2015–2023. DOI: <https://doi.org/10.1098/rspb.2008.0476>, PMID: 18477542
- Bocak L, Yagi T.** 2010. Evolution of mimicry patterns in Metriorrhynchus (Coleoptera: Lycidae): the history of dispersal and speciation in southeast Asia. *Evolution; International Journal of Organic Evolution* **64**:39–52. DOI: <https://doi.org/10.1111/j.1558-5646.2009.00812.x>, PMID: 19674098
- Bocak L, Kundrata R, Fernández CA, Vogler AP.** 2016. The discovery of Iberobaeniidae (Coleoptera: Elateroidea): a new family of beetles from Spain, with immatures detected by environmental DNA sequencing. *Proceedings. Biological Sciences* **283**:20152350. DOI: <https://doi.org/10.1098/rspb.2015.2350>, PMID: 27147093
- Bocak L, Motyka M, Kusy D, Bilkova R.** 2020. Biodiversity Inventory and Distribution of Metriorrhynchina Net-Winged Beetles (Coleoptera: Lycidae), with the Identification of Generic Ranges. *Insects* **11**:E710. DOI: <https://doi.org/10.3390/insects11100710>, PMID: 33081413
- Bocek M, Bocak L.** 2017. The molecular phylogeny and classification of trichaline net-winged beetles (Lycidae: Metriorrhynchini). *PeerJ* **5**:e3963. DOI: <https://doi.org/10.7717/peerj.3963>
- Bocek M, Bocak L.** 2019. The origins and dispersal history of the trichaline net-winged beetles in Southeast Asia, Wallacea, New Guinea and Australia. *Zoological Journal of the Linnean Society* **185**:1079–1094. DOI: <https://doi.org/10.1093/zoolinnean/zly090>
- Bocek M, Kusy D, Motyka M, Bocak L.** 2019. Persistence of multiple patterns and intraspecific polymorphism in multi-species Müllerian communities of net-winged beetles. *Frontiers in Zoology* **16**:38. DOI: <https://doi.org/10.1186/s12983-019-0335-8>, PMID: 31636689
- Breinholt JW, Kawahara AY.** 2013. Phylotranscriptomics: saturated third codon positions radically influence the estimation of trees based on next-gen data. *Genome Biology and Evolution* **5**:2082–2092. DOI: <https://doi.org/10.1093/gbe/evt157>, PMID: 24148944
- Brooks TM, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Rylands AB, Konstant WR, Flick P, Pilgrim J, Oldfield S, Magin G, Hilton-Taylor C.** 2002. Habitat Loss and Extinction in the Hotspots of Biodiversity. *Conservation Biology* **16**:909–923. DOI: <https://doi.org/10.1046/j.1523-1739.2002.00530.x>
- Calder A.** 1998. Coleoptera: Elateroidea. Wells A (Ed). *Zoological Catalogue of Australia*. CSIRO Publishing. p. 9–248.
- Ceballos G, Ehrlich PR, Barnosky AD, Pringle RM, Palmer TM.** 2015. Accelerated modern human – induced species losses: entering the sixth mass extinction. *Science Advances* **1**:e1400253. DOI: <https://doi.org/10.1126/sciadv.1400253>
- Chen S, Zhou Y, Chen Y, Gu J.** 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**:i884–i890. DOI: <https://doi.org/10.1093/bioinformatics/bty560>
- Chernomor O, von Haeseler A, Minh BQ.** 2016. Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. *Systematic Biology* **65**:997–1008. DOI: <https://doi.org/10.1093/sysbio/syw037>, PMID: 27121966
- Chesters D.** 2017. Construction of a Species-Level Tree of Life for the Insects and Utility in Taxonomic Profiling. *Systematic Biology* **66**:426–439. DOI: <https://doi.org/10.1093/sysbio/syw099>, PMID: 27798407
- Drummond AJ, Suchard MA, Xie D, Rambaut A.** 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**:1969–1973. DOI: <https://doi.org/10.1093/molbev/mss075>, PMID: 22367748
- Dupuis JR, Roe AD, Sperling FAH.** 2012. Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Molecular Ecology* **21**:4422–4436. DOI: <https://doi.org/10.1111/j.1365-294X.2012.05642.x>, PMID: 22891635
- Eberle J, Ahrens D, Mayer C, Niehuis O, Misof B.** 2020. A Plea for Standardized Nuclear Markers in Metazoan DNA Taxonomy. *Trends in Ecology & Evolution* **35**:336–345. DOI: <https://doi.org/10.1016/j.tree.2019.12.003>
- Edwards SV.** 2009. Is a new and general theory of molecular systematics emerging? *Evolution; International Journal of Organic Evolution* **63**:1–19. DOI: <https://doi.org/10.1111/j.1558-5646.2008.00549.x>
- Eisner T, Schroeder FC, Snyder N, Grant JB, Aneshansley DJ, Utterback D, Meinwald J, Eisner M.** 2008. Defensive chemistry of lycid beetles and of mimetic cerambycid beetles that feed on them. *Chemoecology* **18**:109–119. DOI: <https://doi.org/10.1007/s00049-007-0398-4>
- Fernández R, Edgecombe GD, Giribet G.** 2016. Exploring Phylogenetic Relationships within Myriapoda and the Effects of Matrix Composition and Occupancy on Phylogenomic Reconstruction. *Systematic Biology* **65**:871–889. DOI: <https://doi.org/10.1093/sysbio/syw041>, PMID: 27162151

- Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**:3150–3152. DOI: <https://doi.org/10.1093/bioinformatics/bts565>
- Gauthier J, Pajkovic M, Neuenschwander S, Kaila L, Schmid S, Orlando L, Alvarez N. 2020. Museomics identifies genetic erosion in two butterfly species across the 20th century in Finland. *Molecular Ecology Resources* **20**:1191–1205. DOI: <https://doi.org/10.1111/1755-0998.13167>, PMID: 32304133
- Godfray HCJ, Knapp S. 2004. Introduction. Taxonomy for the twenty-first century. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **359**:559–569. DOI: <https://doi.org/10.1098/rstb.2003.1457>, PMID: 15253344
- Guérin-Méneville FE. 1838. Crustacés, arachnides et insectes. In Voyage autour du monde sur la Corvette de Sa Majesté, La Coquille, pendant les années 1822, 1823, 1824 et 1825. Bertrand Arthus (Ed). Zoologie. Vol 2. Paris, France: Duperrey, L.I.
- Gutiérrez EE, Helgen KM. 2013. Outdated taxonomy blocks conservation. *Nature* **495**:314. DOI: <https://doi.org/10.1038/495314e>
- Hajibabaei M, Singer GAC, Hebert PDN, Hickey DA. 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics* **23**:167–172. DOI: <https://doi.org/10.1016/j.tig.2007.02.001>, PMID: 17316886
- Hall R. 2011. Australia–SE Asia collision: plate tectonics and crustal flow. *Geological Society, London, Special Publications* **355**:75–109. DOI: <https://doi.org/10.1144/SP355.5>
- Hallmann CA, Sorg M, Jongejans E, Siepel H, Hofland N, Schwan H, Stenmans W, Müller A, Sumser H, Hören T, Goulson D, de Kroon H, Lamb EG. 2017. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLOS ONE* **12**:e0185809. DOI: <https://doi.org/10.1371/journal.pone.0185809>
- Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B* **270**:313–321. DOI: <https://doi.org/10.1098/rspb.2002.2218>
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* **35**:518–522. DOI: <https://doi.org/10.1093/molbev/msx281>, PMID: 29077904
- Holt BG, Lessard JP, Borregaard MK, Fritz SA, Rahbek C. 2013. An Update of Wallace’s Zoogeographic Regions of the World. *Science* **339**:74–78. DOI: <https://doi.org/10.1126/science.1228282>
- Ikeda H, Nishikawa M, Sota T. 2012. Loss of flight promotes beetle diversification. *Nature Communications* **3**:648. DOI: <https://doi.org/10.1038/ncomms1659>, PMID: 22337126
- Jaskula R, Kolanowska M, Michalski M, Schwerk A. 2021. From Phenology and Habitat Preferences to Climate Change: Importance of Citizen Science in Studying Insect Ecology in the Continental Scale with American Red Flat Bark Beetle, *Cucujus clavipes*, as a Model Species. *Insects* **12**:369. DOI: <https://doi.org/10.3390/insects12040369>
- Jiruskova A, Motyka M, Bocak M, Bocak L. 2019. The Malacca Strait separates distinct faunas of poorly-flying Cauties net-winged beetles. *PeerJ* **7**:e6511. DOI: <https://doi.org/10.7717/peerj.6511>
- Jordal BH. 2015. Molecular phylogeny and biogeography of the weevil subfamily Platypodinae reveals evolutionarily conserved range patterns. *Molecular Phylogenetics and Evolution* **92**:294–307. DOI: <https://doi.org/10.1016/j.ympev.2015.05.028>
- Junier T, Zdobnov EM. 2010. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* **26**:1669–1670. DOI: <https://doi.org/10.1093/bioinformatics/btq243>
- Kalousova R, Bocak L. 2017. Species delimitation of colour polymorphic Cladophorus (Coleoptera: Lycidae) from New Guinea. *Zootaxa* **4320**:505. DOI: <https://doi.org/10.11646/zootaxa.4320.3.6>
- Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler A, Jermini LS. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**:587–589. DOI: <https://doi.org/10.1038/NMETH.4285>
- Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**:772–780. DOI: <https://doi.org/10.1093/molbev/mst010>
- Kazantsev SV. 2010. New taxa of Lycidae from Samoa, Fiji and Tonga (Coleoptera: Lycidae). *Russian Journal of Entomology* **18**:195–199.
- Keeling CI, Yuen MMS, Liao NY, Docking TR, Chan SK, Taylor GA, Palmquist DL, Jackman SD, Nguyen A, Li M, Henderson H, Janes JK, Zhao Y, Pandoh P, Moore R, Sperling FAH, Huber DPW, Birol I, Jones SJM, Bohlmann J. 2013. Draft genome of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major forest pest. *Genome Biology* **14**:R27. DOI: <https://doi.org/10.1186/gb-2013-14-3-r27>, PMID: 23537049
- Kleine R. 1933. Lycidae. Hincks WD (Ed). *Coleopterorum Catalogus Auspiciis et Auxilio W. Junk Editus a Schenkling*. Berlin: W. Junk. p. 123–145.
- Krehenwinkel H, Pomerantz A, Henderson JB, Kennedy SR, Lim JY, Swamy V, Shoobridge JD, Graham N, Patel NH, Gillespie RG, Prost S. 2019. Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. *GigaScience* **8**:giz006. DOI: <https://doi.org/10.1093/gigascience/giz006>, PMID: 30824940
- Kubeczek V, Bray TC, Bocak L. 2015. Molecular phylogeny of Metanoecina net-winged beetles identifies Ochinoeus, a new genus from China and Laos (Coleoptera: Lycidae). *Zootaxa* **3955**:113–122. DOI: <https://doi.org/10.11646/zootaxa.3955.1.6>, PMID: 25947840
- Kück P, Meusemann K, Dambach J, Thormann B, von Reumont BM, Wägele JW, Misof B. 2010. Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Frontiers in Zoology* **7**:10. DOI: <https://doi.org/10.1186/1742-9994-7-10>, PMID: 20356385

- Kück P**, Longo GC. 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Frontiers in Zoology* **11**:81. DOI: <https://doi.org/10.1186/s12983-014-0081-x>, PMID: 25426157
- Kück P**, Struck TH. 2014. BaCoCa – A heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. *Molecular Phylogenetics and Evolution* **70**:94–98. DOI: <https://doi.org/10.1016/j.ympev.2013.09.011>
- Kunte K**. 2008. Mimetic butterflies support Wallace’s model of sexual dimorphism. Proceedings of the Royal Society B – Biological Sciences. 1617–1624. DOI: <https://doi.org/10.1098/rspb.2008.0171>
- Kusy D**, Motyka M, Bocek M, Vogler AP, Bocak L. 2018. Genome sequences identify three families of Coleoptera as morphologically derived click beetles (Elateridae). *Scientific Reports* **8**:17084. DOI: <https://doi.org/10.1038/s41598-018-35328-0>, PMID: 30459416
- Kusy D**, Motyka M, Bocek M, Masek M, Bocak L. 2019. Phylogenomic analysis resolves the relationships among net-winged beetles (Coleoptera: Lycidae) and reveals the parallel evolution of morphological traits. *Systematic Entomology* **44**:911–925. DOI: <https://doi.org/10.1111/syen.12363>
- Lanfear R**, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Molecular Biology and Evolution* **34**:772–773. DOI: <https://doi.org/10.1093/molbev/msw260>, PMID: 28013191
- Le SQ**, Gascuel O. 2008. An improved general amino acid replacement matrix. *Molecular Biology and Evolution* **25**:1307–1320. DOI: <https://doi.org/10.1093/molbev/msn067>, PMID: 18367465
- Letsch H**, Balke M, Toussaint EFA, Narakusumo RP, Fiedler K, Riedel A. 2020. Transgressing Wallace’s Line brings hyperdiverse weevils down to earth. *Ecography* **43**:1329–1340. DOI: <https://doi.org/10.1111/ecog.05128>
- Linard B**, Arribas P, Crampton-Platt A, Vogler AP. 2016. Lessons from genome skimming of arthropod-preserving ethanol. *Molecular Ecology Resources* **16**:1365–1377. DOI: <https://doi.org/10.1111/1755-0998.12539>
- MacPhail VJ**, Colla SR. 2020. Power of the people: A review of citizen science programs for conservation. *Biological Conservation* **249**:108739. DOI: <https://doi.org/10.1016/j.biocon.2020.108739>
- Masek M**, Motyka M, Kusy D, Bocek M, Li Y, Bocak L. 2018. Molecular Phylogeny, Diversity and Zoogeography of Net-Winged Beetles (Coleoptera: Lycidae). *Insects* **9**:E154. DOI: <https://doi.org/10.3390/insects9040154>, PMID: 30388727
- Mastretta-Yanes A**, Xue AT, Moreno-Letelier A, Jorgensen TH, Alvarez N, Piñero D, Emerson BC. 2018. Long-term in situ persistence of biodiversity in tropical sky islands revealed by landscape genomics. *Molecular Ecology* **27**:432–448. DOI: <https://doi.org/10.1111/mec.14461>, PMID: 29226496
- McKenna DD**, Scully ED, Pauchet Y, Hoover K, Kirsch R, Richards S. 2017. Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle–plant interface. *Genome Biology* **17**:227. DOI: <https://doi.org/10.1186/s13059-016-1088-8>
- McKenna DD**, Shin S, Ahrens D, Balke M, Beza-Beza C, Clarke DJ, Donath A, Escalona HE, Friedrich F, Letsch H, Liu S, Maddison D, Mayer C, Misof B, Murin PJ, Niehuis O, Peters RS, Podsiadlowski L, Pohl H, Scully ED, et al. 2019. The evolution and genomic basis of beetle diversity. *PNAS* **116**:24729–24737. DOI: <https://doi.org/10.1073/pnas.1909655116>, PMID: 31740605
- McMahon DP**, Hayward A. 2016. Why grow up? A perspective on insect strategies to avoid metamorphosis. *Ecological Entomology* **41**:505–515. DOI: <https://doi.org/10.1111/een.12313>
- Merckx VSFT**, Hendriks KP, Beentjes KK, Mennes CB, Becking LE, Peijnenburg KTCA, Afendy A, Arumugam N, de Boer H, Biun A, Buang MM, Chen P-P, Chung AYC, Dow R, Feijen FAA, Feijen H, Feijen-van Soest C, Geml J, Geurts R, Gravendeel B, et al. 2015. Evolution of endemism on a young tropical mountain. *Nature* **524**:347–350. DOI: <https://doi.org/10.1038/nature14949>, PMID: 26266979
- Minh BQ**, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R, Teeling E. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**:1530–1534. DOI: <https://doi.org/10.1093/molbev/msaa015>
- Misof B**, Meyer B, von Reumont BM, Kück P, Misof K, Meusemann K. 2013. Selecting informative subsets of sparse supermatrices increases the chance to find correct trees. *BMC Bioinformatics* **14**:348. DOI: <https://doi.org/10.1186/1471-2105-14-348>, PMID: 24299043
- Misof B**, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, Niehuis O, Petersen M, Izquierdo-Carrasco F, Wappler T, Rust J, Aberer AJ, Aspöck U, Aspöck H, Bartel D, Blanke A, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**:763–767. DOI: <https://doi.org/10.1126/science.1257570>, PMID: 25378627
- Monaghan MT**, Wild R, Elliot M, Fujisawa T, Balke M, Inward DJG, Lees DC, Ranaivosolo R, Eggleton P, Barraclough TG, Vogler AP. 2009. Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Systematic Biology* **58**:298–311. DOI: <https://doi.org/10.1093/sysbio/syp027>, PMID: 20525585
- Mora C**, Tittensor DP, Adl S, Simpson AGB, Worm B. 2011. How many species are there on Earth and in the ocean? *PLoS Biology* **9**:e1001127. DOI: <https://doi.org/10.1371/journal.pbio.1001127>, PMID: 21886479
- Morrison WR**, Lohr JL, Duchon P, Wilches R, Trujillo D, Mair M, Renner SS. 2009. The impact of taxonomic change on conservation: Does it kill, can it save, or is it just irrelevant? *Biological Conservation* **142**:3201–3206. DOI: <https://doi.org/10.1016/j.biocon.2009.07.019>
- Motyka M**, Kampova L, Bocak L. 2018. Phylogeny and evolution of Müllerian mimicry in aposematic Dilophotes: evidence for advergence and size-constraints in evolution of mimetic sexual dimorphism. *Scientific Reports* **8**:3744. DOI: <https://doi.org/10.1038/s41598-018-22155-6>

- Motyka M**, Bocek M, Kusy D, Bocak L. 2020. Interactions in multi-pattern Mullerian communities support origins of new patterns, false structures, imperfect resemblance and mimetic sexual dimorphism. *Scientific Reports* **10**:11193. DOI: <https://doi.org/10.1038/s41598-020-68027-w>
- Motyka M**, Kusy D, Masek M, Bocek M, Li Y, Bilkova R, Bocak L. 2021. Conspicuousness, phylogenetic structure, and origins of Müllerian mimicry in 4000 lycid beetles from all zoogeographic regions. *Scientific Reports* **11**:5961. DOI: <https://doi.org/10.1038/s41598-021-85567-x>
- Myers N**, Mittermeier RA, Mittermeier CG, Kent J. 2000. Biodiversity hotspots for conservation priorities. *Nature* **403**:853–858. DOI: <https://doi.org/10.1038/35002501>
- Naser-Khdour S**, Minh BQ, Zhang W, Stone EA, Lanfear R. 2019. The Prevalence and Impact of Model Violations in Phylogenetic Analysis. *Genome Biology and Evolution* **11**:3341–3352. DOI: <https://doi.org/10.1093/gbe/evz193>, PMID: 31536115
- Novotny V**, Drozd P, Miller SE, Kulfan M, Janda M, Basset Y, Weiblen GD. 2006. Why are there so many species of herbivorous insects in tropical rainforests? *Science* **313**:1115–1118. DOI: <https://doi.org/10.1126/science.1129237>, PMID: 16840659
- Ott M**. 2019. SymTest. v.2.0.49. GitHub. <https://github.com/ottmi/symtest>
- Peters RS**, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, Kozlov A, Podsiadlowski L, Petersen M, Lanfear R, Diez PA, Heraty J, Kjer KM, Klopstein S, Meier R, Polidori C, Schmitt T, Liu S, Zhou X, Wappler T, et al. 2017. Evolutionary History of the Hymenoptera. *Current Biology* **27**:1013–1018. DOI: <https://doi.org/10.1016/j.cub.2017.01.027>, PMID: 28343967
- Petersen M**, Meusemann K, Donath A, Dowling D, Liu S, Peters RS, Podsiadlowski L, Vasilikopoulos A, Zhou X, Misof B, Niehuis O. 2017. Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. *BMC Bioinformatics* **18**:111. DOI: <https://doi.org/10.1186/s12859-017-1529-8>, PMID: 28209129
- Poelchau M**, Childers C, Moore G, Tsavatapalli V, Evans J, Lee CY, Lin H, Lin JW, Hackett K. 2015. The i5k Workspace@NAL--enabling genomic data access, visualization and curation of arthropod genomes. *Nucleic Acids Research* **43**:D714–D719. DOI: <https://doi.org/10.1093/nar/gku983>, PMID: 25332403
- Prathapan KD**. 2018. When the cure kills-CBD limits biodiversity research National laws fearing biopiracy squelch taxonomy studies. *Science* **360**:1405–1406. DOI: <https://doi.org/10.1126/science.aat9844>
- Revell LJ**. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**:217–223. DOI: <https://doi.org/10.1111/j.2041-210X.2011.00169.x>
- Richards S**, Gibbs RA, Weinstock GM, Brown SJ, Denell R, Beeman RW, Gibbs R, Beeman RW, Brown SJ, Bucher G, Friedrich M, Grimmelikhuijzen CJP, Klingler M, Lorenzen M, Richards S, Roth S, Schröder R, Tautz D, Zdobnov EM, Muzny D, et al. 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**:949–955. DOI: <https://doi.org/10.1038/nature06784>, PMID: 18362917
- Riedel A**, Sagata K, Suhardjono YR, Tänzler R, Balke M. 2013. Integrative taxonomy on the fast track - towards more sustainability in biodiversity research. *Frontiers in Zoology* **10**:15. DOI: <https://doi.org/10.1186/1742-9994-10-15>, PMID: 23537182
- Sangster G**, Luksenburg JA. 2015. Declining rates of species described per taxonomist: slowdown of progress or a side-effect of improved quality in taxonomy? *Systematic Biology* **64**:144–151. DOI: <https://doi.org/10.1093/sysbio/syu069>, PMID: 25190593
- Sayyari E**, Mirarab S. 2016. Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. *Molecular Biology and Evolution* **33**:1654–1668. DOI: <https://doi.org/10.1093/molbev/msw079>, PMID: 27189547
- Schliep KP**. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**:592–593. DOI: <https://doi.org/10.1093/bioinformatics/btq706>, PMID: 21169378
- Sharkey MJ**, Janzen DH, Hallwachs W, Chapman EG, Smith MA, Dapkey T, Brown A, Ratnasingham S, Naik S, Manjunath R, Perez K, Milton M, Hebert P, Shaw SR, Kittel RN, Solis MA, Metz MA, Goldstein PZ, Brown JW, Quicke DLJ, et al. 2021. Minimalist revision and description of 403 new species in 11 subfamilies of Costa Rican braconid parasitoid wasps, including host records for 219 species. *ZooKeys* **1013**:1–665. DOI: <https://doi.org/10.3897/zookeys.1013.55600>, PMID: 34512087
- Shelton JM**, Coleman MC, Herndon N, Lu N, Lam ET, Anantharaman T, Sheth P, Brown SJ. 2015. Tools and pipelines for BioNano data: molecule assembly pipeline and FASTA super scaffolding tool. *BMC Genomics* **16**:734. DOI: <https://doi.org/10.1186/s12864-015-1911-8>
- Sklenarova K**, Chesters D, Bocak L. 2013. Phylogeography of Poorly Dispersing Net-Winged Beetles: A Role of Drifting India in the Origin of Afrotropical and Oriental Fauna. *PLOS ONE* **8**:e67957. DOI: <https://doi.org/10.1371/journal.pone.0067957>, PMID: 23840793
- Sklenarova K**, Kubecek V, Bocak L. 2014. Subtribal classification of Metriorrhynchini (Insecta: Coleoptera: Lycidae): an integrative approach using molecular phylogeny and morphology of adults and larvae. *Arthropod Systematics & Phylogeny* **72**:37–54.
- Sodhi NS**, Koh LP, Brook BW, Ng PKL. 2004. Southeast Asian biodiversity: an impending disaster. *Trends in Ecology & Evolution* **19**:654–660. DOI: <https://doi.org/10.1016/j.tree.2004.09.006>
- Srivathsan A**, Hartop E, Puniamoorthy J, Lee WT, Kutty SN, Kurina O, Meier R. 2019. Rapid, large-scale species discovery in hyperdiverse taxa using 1D MinION sequencing. *BMC Biology* **17**:96. DOI: <https://doi.org/10.1186/s12915-019-0706-9>, PMID: 31783752
- Stanke M**, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**:ii215–ii225. DOI: <https://doi.org/10.1093/bioinformatics/btg1080>

- Strimmer K, Haeseler A.** 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *PNAS* **94**:6815–6819. DOI: <https://doi.org/10.1073/pnas.94.13.6815>
- Suyama M, Torrents D, Bork P.** 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* **34**:W609–W612. DOI: <https://doi.org/10.1093/nar/gkl315>
- Talavera G, Lukhtanov V, Pierce NE, Vila R.** 2021. DNA barcodes combined with multi-locus data of representative taxa can generate reliable higher-level phylogenies. *Systematic Biology*:syab038. DOI: <https://doi.org/10.1093/sysbio/syab038>, PMID: 34022059
- Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP.** 2003. A plea for DNA taxonomy. *Trends in Ecology & Evolution* **18**:70–74. DOI: [https://doi.org/10.1016/S0169-5347\(02\)00041-1](https://doi.org/10.1016/S0169-5347(02)00041-1)
- Theng M, Jusoh WFA, Jain A, Huertas B, Tan DJX, Tan HZ, Kristensen NP, Meier R, Chisholm RA.** 2020. A comprehensive assessment of diversity loss in a well-documented tropical insect fauna: Almost half of Singapore's butterfly species extirpated in 160 years. *Biological Conservation* **242**:108401. DOI: <https://doi.org/10.1016/j.biocon.2019.108401>
- Thomson SA, Pyle RL, Ah Yong ST, Alonso-Zarazaga M, Ammirati J, Araya JF, Ascher JS, Audisio TL, Azevedo-Santos VM, Bailly N, Baker WJ, Balke M, Barclay MVL, Barrett RL, Benine RC, Bickstaff JRM, Bouchard P, Bour R, Bourgoin T, Boyko CB, et al.** 2018. Taxonomy based on science is necessary for global conservation. *PLOS Biology* **16**:e2005075. DOI: <https://doi.org/10.1371/journal.pbio.2005075>
- Toussaint EFA, Hall R, Monaghan MT, Sagata K, Ibalim S, Shaverdo HV, Vogler AP, Pons J, Balke M.** 2014. The towering orogeny of New Guinea as a trigger for arthropod megadiversity. *Nature Communications* **5**:4001. DOI: <https://doi.org/10.1038/ncomms5001>
- van Klink R, Bowler DE, Gongalsky KB, Swengel AB, Gentile A, Chase JM.** 2020. Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science* **368**:417–420. DOI: <https://doi.org/10.1126/science.aax9931>, PMID: 32327596
- Vasilikopoulos A, Balke M, Beutel RG, Donath A, Podsiadlowski L, Pflug JM, Waterhouse RM, Meusemann K, Peters RS, Escalona HE, Mayer C, Liu S, Hendrich L, Alarie Y, Bilton DT, Jia F, Zhou X, Maddison DR, Niehuis O, Misof B.** 2019. Phylogenomics of the superfamily Dytiscoidea (Coleoptera: Adephaga) with an evaluation of phylogenetic conflict and systematic error. *Molecular Phylogenetics and Evolution* **135**:270–285. DOI: <https://doi.org/10.1016/j.ympev.2019.02.022>
- Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM.** 2018. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Molecular Biology and Evolution* **35**:543–548. DOI: <https://doi.org/10.1093/molbev/msx319>, PMID: 29220515
- Wong TKF, Kalyaanamoorthy S, Meusemann K, Yeates DK, Misof B, Jermini LS.** 2020. A minimum reporting standard for multiple sequence alignments. *NAR Genom Bioinform* **2**. DOI: <https://doi.org/10.1093/nargab/lqaa024>
- Wong T.** 2021. AliStat. 1.7. GitHub. <https://github.com/thomaskf/AliStat>
- Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, Zhou X, Lam TW, Li Y, Xu X, Wong GKS, Wang J.** 2014. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* **30**:1660–1666. DOI: <https://doi.org/10.1093/bioinformatics/btu077>, PMID: 24532719
- Yeo D, Srivathsan A, Meier R.** 2020. Longer is Not Always Better: Optimizing Barcode Length for Large-Scale Species Discovery and Identification. *Systematic Biology* **69**:999–1015. DOI: <https://doi.org/10.1093/sysbio/syaa014>, PMID: 32065638
- Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, Ioannidis P, Seppey M, Loetscher A, Kriventseva EV.** 2017. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Research* **45**:D744–D749. DOI: <https://doi.org/10.1093/nar/gkw1119>, PMID: 27899580
- 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. DOI: <https://doi.org/10.1186/s12859-018-2129-y>, PMID: 29745866
- Zhang ZQ.** 2021. Contributions of Zootaxa to biodiversity discovery: an overview of the first twenty years. *Zootaxa* **4979**:6–16. DOI: <https://doi.org/10.11646/zootaxa.4979.1.3>

Appendix 1

Appendix introductory information

Overview of Metriorrhynchini systematics and natural history

Lycidae (net-winged beetles) is a diverse, mostly tropical family of Elateroidea and Metriorrhynchini is the largest lycid tribe with almost 1600 species (**Masek et al., 2018**). The history of their classification started with Kleine's proposals of tribes and/or subfamilies Metriorrhynchini/inae, Trichalini/inae, Cladophorini/inae, and Dilolycini/inae (see an overview by **Kleine, 1933**). The supergeneric classification of net-winged beetles was later revised. Conderini were merged with Metriorrhynchini and Trichalini in the widely defined Metriorrhynchinae and Cladophorini and Dilolycini were synonymized to Metriorrhynchini (**Bocak, 2002** and references therein). **Calder, 1998** considered Metriorrhynchinae Kleine, 1926 (based on *Metriorrhynchus* Guérin-Méneville, 1838) to be a homonym of the crocodylian subfamily Metriorrhynchinae Meyer, 1832 and used Dilolycinae Kleine, 1926 as a replacement name. Trichalinae were placed as a tribe Trichalini in Metriorrhynchinae or a subtribe Trichalina in Metriorrhynchini (**Bocak, 2002**). Additionally, Hemiconderina **Bocak et al., 2008** were erected for *Hemiconderis* and related genera. Detailed information on the metriorrhynchine classification history was published in the latest revision of the subtribal classification (**Sklenarova et al., 2014**) and the genera were assigned to subtribes by **Kubecek et al., 2015** and **Bocak et al., 2020**.

The alpha-taxonomy is mostly based on the original, often uninformative descriptions (catalogued by **Kleine, 1933; Bocak et al., 2020**). Recently, **Calder, 1998** compiled the catalogue of Australian net-winged beetles and proposed *Metriorrhynchus* Guérin-Méneville, 1838 as a junior homonym of *Metriorrhynchus* Meyer, 1832. Subsequently, *Porrostoma* Castelnau, 1838 was used as a replacement name for *Metriorrhynchus* Guérin-Méneville, 1838 and new combinations and new replacement names were proposed (**Calder, 1998**). **Bocak, 2002** reviewed the status of lycid taxa described by **Guérin-Méneville, 1838** and proposed to replace *Metriorrhynchus* Guérin-Méneville, 1838 by *Metriorrhynchus* Gemminger et Harold, 1869. **Bocak, 2002** additionally considered *Porrostoma* Castelnau, 1838 as a valid name for a separate genus in Metriorrhynchini. These studies resulted in repeated transfers of several hundred of species between *Metriorrhynchus* and *Porrostoma* (**Calder, 1998; Bocak et al., 2020**).

There are five Metriorrhynchini genera which house the majority of described species: *Xylobanus* (231 spp.), *Cautires* (437 spp.), *Trichalus* (120 spp.) *Metriorrhynchus* (194 spp.), and *Cladophorus* (131 spp.; **Kleine, 1933; Calder, 1998; Bocak, 2002; Bocak et al., 2020**), although their limits have been ambiguous. *Trichalus* was recovered as a much smaller clade than *Microtrichalus* Pic, 1921 (**Bocek and Bocak, 2017**). Australian *Metriorrhynchus* transferred to *Porrostoma* Castelnau, 1838 by **Calder, 1998** were formally returned to *Metriorrhynchus* by *Metriorrhynchus* by **Bocak et al., 2020**. *Xylobanus* was recovered as a clade that contains most Oriental Cautirina with four primary costae, but Afrotropical *Xylobanus* were recovered as a distantly related terminal clade within Afrotropical *Cautires* (**Sklenarova et al., 2013; Sklenarova et al., 2014**). Some Oriental *Xylobanus* belong to Metanoeina and New Guinean species to Metriorrhynchina (**Kubecek et al., 2015; Bocak et al., 2020**). An additional typologically defined genus, *Procautires* Kleine, 1925, was reported from Asia, Africa and Australasia (**Kleine, 1933**), but the type species belongs to Metriorrhynchina in contrast to the Afrotropical and Asian species which are close to *Cautires* (Cautirina).

Geographic distribution

The tribe Metriorrhynchini dominates the Australian net-winged beetle fauna with ~85% (196 spp.) of the named taxa. They are similarly common in New Guinea (423 spp.) and the Wallacea (176 spp.). Although highly diversified (261 spp. and 231 spp., respectively), they represent only about 30% of described net-winged beetle species in the Oriental and Afrotropical regions where similar numbers of species belong to Platerodini, Calochromini and Lycini, or several small tribes (**Masek et al., 2018**).

The tribe Metriorrhynchini has a Gondwanan distribution, and the range reaches to the Palearctic East Asia (Russian Far East, Cautirina and Metanoeina). Additionally, the tribe is well represented in the Sino-Japanese realm sensu **Holt et al., 2013** (China, Japan, Cautirina and Metanoeina in the whole area, Metriorrhynchina in southernmost China only) and in the Oriental realm (most of India, Sri Lanka, the lower elevations of the Himalayas, Indo-Burma, Malaya, the

Great and Lesser Sundas, the Philippines). Further, the tribe is widespread in the Oceanian realm, especially in its western part (the Moluccas, New Guinea, the Solomon Islands and a few islands further east) and in the Australian realm (Australia, Tasmania, and one introduced species in New Zealand). The Afrotropical realm houses relatively diversified fauna in forest and savannah habitats, but all species belong to Cautirina and represent only two clades which correspond with putative independent colonization events from drifting India to Madagascar and continental Africa (Sklenarova et al., 2013).

The subtribes differ in their distribution. Metriorrhynchina is mostly Oceanian and Australian, relatively rich Metriorrhynchina fauna is known from the Sulawesi, much less species occur in the Philippines and only a few species of *Diatrichalus*, *Leptotrichalus*, *Trichalus*, *Microtrichalus*, and *Metriorrhynchus* in the eastern part of the Oriental realm (Bocak and Yagi, 2010; Bocak and Bocak, 2019; Bocek et al., 2019) and southern Yunnan in the Sino-Japanese realm. The Cautirina occurs in the Afrotropical, Oriental, Palearctic, and Sino-Japanese realms, but they do not cross Weber's line that was proposed as a border between Oriental and Oceanian realm (Holt et al., 2013). The *Metanoeina* is Oriental, with a limited number of species in China and Japan. Only *Metanoeus* occurs in the Philippines.

The Metriorrhynchini is the species-rich tribe and the almost 200 years of taxonomic research accumulated data that need serious revision (Bocak et al., 2020). In the taxonomy of this group, we face the burden of uninformative descriptions (e.g. the studies by the French entomologist M. Pic in the first half of the 20th century). The careless classification of newly described species, inaccessibility of type-holding collections or their poor organization, and legal regulations limit taxonomic research (Prathapan, 2018; Bocak et al., 2020).

Biology

The adults are volatile, but they are poor dispersers due to their soft-bodiedness and many species remain in shaded situations under the canopy, usually sitting on leaves. Only some species (*Porrostoma*, *Metriorrhynchus*, *Leptotrichalus*, and *Trichalus*) visit flowers. The flower frequenting species are more common in semi-dry areas where nectar represents a source of water for adults instead of water on leaves or in rotten wood. All Metriorrhynchini are unpalatable for predators and most are aposematically coloured (Eisner et al., 2008; Motyka et al., 2021). The color patterns are usually geographically restricted and the sympatrically occurring species form Müllerian mimicry rings (Bocak and Yagi, 2010; Motyka et al., 2018; Motyka et al., 2020; Motyka et al., 2021; Bocek et al., 2019). The larvae of *Metanoeus*, *Cautires*, *Leptotrichalus*, *Metriorrhynchus*, and *Porrostoma* were described by Bocak and Matsuda, 2003. They live in rotten wood including trunks and logs on soil surface or in rotten roots in soil in arid regions.

Appendix results

Metriorrhynchina subclades based on phylogenomic analyses and their morphological characteristics and distribution.

The procautirine clade

The phylogenomic analysis contained four taxa that clustered as the first or second deepest serial branch in Metriorrhynchina. One of the internal clades contains species that are morphologically highly similar to the type-species of *Procautires*. They have reduced secondary elytral costae in the middle part of the elytra. Other procautirine species have fully developed secondary costae unlike *Procautires*. The only morphological trait supporting the monophyly of the procautirine clade is a setose patch in the internal sac of the male genitalia.

The leptotrichaline clade

The phylogenomic backbone identified close relationships of three endemic genera from the Sulawesi (*Mangkutanus*, *Wakarumbia*, and *Broxylus*), and two genera, *Sulabanus*, *Leptotrichalus*, known from Sulawesi and the Philippines (Bocak et al., 2020). The mtDNA dataset identified as closely related the Australian and New Guinean *Synchonnus* and New Guinean *Falsolucidota*. Two lineages of the clade putatively colonized the Oriental region: *Leptotrichalus* (the Philippines, Sundas, Indo-Burma and southernmost China) and *Sulabanus* (only the Philippines, 2 spp. from Mindanao and Sibuyan identified herein, either undescribed species or currently placed in Cautirina: *Xylobanus*). We have not found any reliable external diagnostic morphological character

which would support the relationships of these genera. Only *Sulabanus* and *Mangkutanus* have a complete set of lateral pronotal carinae. Other genera have simplified or absent carinae except those that form the median lanceolate areola (**Bocak, 2002**). The group contains genera with the highly diverse arrangement of elytral costae.

The trichaline clade

The genomic dataset contained representatives of three genera: *Diatrichalus*, *Microtrichalus*, and *Eniclases*, all of them sharing the characteristic morphological traits of the trichaline clade, that is a single median areola in the pronotum and the shortened first primary elytral costa. We identified as close relatives two lineages of small-bodied Metriorrhynchina with conspicuous median pronotal areola and either absent or simplified lateral carinae, but with the full-length elytral costa 1. The structure of elytral costae is diverse in the clade: most trichalines sensu stricto have well-developed secondary costae, that is, nine costae at humeri, only a few *Diatrichalus* have only four primary costae in the humeral part. All genera of the here delimited trichaline clade, except *Diatrichalus*, share similar male genitalia (**Bocak, 2002**). The deepest branches of the trichaline clade contain mostly New Guinean species and several lineages of the trichalines colonized the Sulawesi, Philippines, Greater Sundas, Malaya, Indo-Burma, and southernmost China.

The porrostomine clade

Six transcriptomes were analyzed to recover the backbone of the clade. *Porrostoma* and *Metriorrhynchus* are distantly related and easy to distinguish with the genital morphology (**Bocak, 2002**). Numerous small bodied, morphologically diverse Metriorrhynchina species were identified as their close relatives. Most *Metriorrhynchus* and their closest relatives come from New Guinea, but several species of a terminal clade crossed Lydderker's, Weber's, and Wallace's lines and colonized Sulawesi, the Philippines, Greater Sundas, Malaya, and Indo-Burma (**Bocak and Yagi, 2010**).

The cladophorine clade

Altogether eleven taxa were included in the genomic analysis, and we found *Cladophorus* as a sister to *Pseudodontocerus* (= *Carathrix*) and *Ditua*. Most genera have seven pronotal areoles, flabellate to pectinate male antennae, and nine well developed costae in the elytra. Mitochondrial dataset recovered this clade as the most speciose group. The characteristic large-bodied species resembling the type-species of *Cladophorus* were found as a terminal subclade among several lineages of small-bodied Metriorrhynchina which do not resemble them in general appearance. Further splits contain the *Pseudodontocerus* clade and a large clade containing representatives of earlier described genera *Ditua* and *Cautiromimus*. The species of the *Cladophorus* clade are most diverse in New Guinea, but some species of the *Ditua* clade were recorded from the Philippines and Sulawesi, a limited number of species is native to continental Australia.

Impact of molecular phylogenetics on the assessment of morphological evolution

With the robust, data-rich phylogeny, we can test earlier hypotheses on morphological evolution. Relationships among subtribes have been based on the similarity of larvae and ambiguously supported by Sanger data (**Sklenarova et al., 2013; Sklenarova et al., 2014**). Our results imply that bipartite larval terga can be inferred either as a synapomorphy of Metriorrhynchini with reversal to entire tergites in Cautirina or we must alternatively consider their independent origin in Metanoeina and Metriorrhynchina. The mapping of the character on the phylogenomic tree needs two instead of a single step as earlier supposed (**Sklenarova et al., 2014**). When most larvae are unknown and their morphology variable, the phylogenetic hypotheses based on incomplete morphological data and uncertain polarity urgently need validation from independent sources of information.

Due to quite high morphological disparity and ambiguous interpretation of some morphological characters, there were proposed subfamilies, tribes, and/or subtribes for various subsets of genera now placed in the Metriorrhynchina (Metriorrhynchinae/ini, Cladophorinae/ini, Trichalinae/ini, Dilolycinae/ini, Hemiconderina, and possibly Melanerotini, now in Lycidae *incertae sedis*; **Bocak, 2002; Kazantsev, 2010; Kusy et al., 2019**). Phylogenomics reject the possibility to assign a high rank to Trichalini and Hemiconderina as numerous coordinated taxa would have to be erected to make all groups reciprocally monophyletic and keep the earlier proposed names valid.

The phylogenomic analyses reinforce the previously identified monophyly of the trichaline clade (=earlier Trichalini part; **Bocek and Bocak, 2017**), but simultaneously suggest their delayed origin as a terminal clade and the presence of some 'non-trichaline' lineages as serial sister-groups to the trichalines in the earlier sense. The cladophorine clade contains numerous taxa which had been placed until recently into *Cautires* (Cautirina), a distant lineage in respect to cladophorines. Herein, we further delimit the porrostomine clade which contains *Metriorrhynchus* and *Porrostoma*, but along with them several deeply rooted lineages which have been incorrectly placed to various genera.

For the first time, we define the leptotrichaline clade that contains the genera earlier placed into the trichalines (*Leptotrichalus*), Hemiconderina (*Falsolucidota*), Lycinae: Calopterini (*Broxylus*), and Cautirina (some *Xylobanus*). The procautirine clade is also a newly defined group. *Procautires* had been for long considered as related to *Cautires*, but it is one of earliest Metriorrhynchina.

Until recently, the classification had been based almost exclusively on the arrangement of pronotal carinae, elytral costae and the shape of antennae. These traits are easily observable and sometimes define a clade, but in many cases their presence and/or absence is misleading as they are highly plastic even within restricted clusters of species (**Sklenarova et al., 2014; Bocek and Bocak, 2017; Kusy et al., 2019**). Before the availability of molecular data, the homoplasy of external morphological characters could not be tested. As a result, the earlier studies coped with uncertain relationships of newly described species at the levels from a tribe down to a genus (**Kleine, 1933; Calder, 1998; Bocak et al., 2020**).

The list of Metriorrhynchini genera with complete synonymy:

Subtribe Metanoeina **Sklenarova et al., 2014**

Matsudanoeus **Sklenarova et al., 2014**

Metanoeus Waterhouse, 1878

Ochinoeus Kubecek et al., 2015

Xylometanoeus Sklenarova et al. 2014

Subtribe Cautirina Sklenarova et al. 2014

Subtribe Cautirina Sklenarova et al. 2015

Caenioxylobanus Waterhouse, 1878

Cautires Waterhouse, 1878

Prometanoeus Kleine, 1925

= *Tapromenoeus* Bocak et Bocakova, 1989

Paracautires Kazantsev, 2012

Spartoires Kazantsev, 2012

Tricautires Kazantsev, 2006

Xylobanus Waterhouse, 1878

Subtribe Metriorrhynchina Kleine, 1926

Procautirine clade

Procautires Kleine, 1925b

Xylothrix Kazantsev, 2015

Leptotrichaline clade

Falsolucidota Pic, 1921

= *Hemiconderis* Kleine, 1926

Broxylus Waterhouse, 1879

= *Samanga* Pic, 1921

Mangkutanus Kubecek et al., 2011

Sulabanus Dvorak & Bocak, 2007

Leptotrichalus Kleine, 1925

Synchonnus Waterhouse, 1879

= *Achras* Waterhouse, 1879

= *Enylus* Waterhouse, 1879

= *Strophicus* Waterhouse, 1879

Wakarumbia Bocak, 1999

Cladophorine clade

Cautiromimus Pic, 1926

Cladophorinus Kleine, 1926

Cladophorus Guérin-Méneville, 1830

= *Odontocerus* Guérin-Méneville, 1838

= *Spacekia* Strand, 1936
 subgenus *Cladophorus* s. str.
 subgenus *Falsocautires* Pic, 1926
Ditua Waterhouse, 1879
Marena Kazantsev, 2007
Pseudodontocerus Pic, 1921
 = *Carathrix* Kleine, 1926
 Trichaline clade
Diatrichalus Kleine, 1926
 = *Mimoxylabanus* Pic, 1921
Eniclases Waterhouse, 1879
 = *Trichalolus* Pic, 1923
Flabellotrichalus Pic, 1921
 = *Stereotrichalus* Kleine, 1926
 = *Villosotrichalus* Pic, 1921
 subgen. *Flabellotrichalus* s. str.
 subgen. *Maibrius* Bocak et Bocak, 2017
Lobatang Bocak, 1998
 subgenus *Lobatang* s. str.
 subgenus *Spinotrichalus* Kazantsev, 2010
Microtrichalus Pic, 1921
 = *Falsoenylus* Pic, 1926
Schizotrichalus Kleine, 1926
Trichalus Waterhouse, 1877
 = *Xantheros* Fairmaire, 1877
 Porrostomine clade
Metriorrhynchus Gemminger et Harold, 1869
 = *Metriorrhynchus* Guérin-Méneville, 1838
 = *Dilolycus* Kleine, 1926
 = *Flabelloporrostoma* Pic, 1923
Porrostoma Laporte, 1838
Oriomum Bocak, 1999
Metriorrhynchoides Kleine, 1926
Stadenus Waterhouse, 1879
Metriorrhynchina incertae sedis
Malacolycus Kleine, 1943
Xylobanomimus Kleine, 1926
Xylobanomorphus Kleine, 1935
Kassemia Bocak, 1998

Appendix methods

Genomic and transcriptomic sequencing, data analysis, individual datasets
 Libraries for thirty transcriptomes were prepared by Novogene Co., Ltd. (Beijing, China) and sequenced on the HiSeq X-ten platform (Illumina Inc, San Diego, USA). The removal of low-quality reads and TruSeq adaptor sequences were performed using fastp v.0.20.0 (Chen et al., 2018) with the following parameters: -q 5 u 50 l 50 n 15. All paired-end transcriptomic reads were assembled using SOAPdenovo-Trans-31mer (Xie et al., 2014).

Additionally, the total DNA (~33 Gb each) of *Metanoëus* sp. and an unidentified sample JB0085 was shotgun-sequenced on the same platform by Novogene Co., Ltd. (Beijing, China) for 150 bp paired-end reads. Read were filtered with fastp using same setting as above and quality was visualized with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>).

The draft genomes were assembled using SPAdes v.3.13.1 (Bankevich et al., 2012), with all parameters set to default values and k-mer sizes of 21, 33, 55, 77 and 99. Obtained contig sequences were used to train Augustus (Stanke and Waack, 2003) for species specific gene models with BUSCO v.3 (Waterhouse et al., 2018), -long option, Endopterygota set of conserved genes (n = 2442) and -sp tribolium2012 as the closest relative. Predicted species specific gene models were then used for *ab initio* gene predictions in Augustus and predicted protein coding sequences were used for subsequent analyses. Seven outgroup taxa, *Porrostoma* and

Mangkutanus transcriptomes were reported in previous studies (Kusy et al., 2018, Kusy et al., 2019; McKenna et al., 2019). The filtering and assembling methods for these samples have been described by Kusy et al., 2019.

The ortholog set was collated by searching the OrthoDB v.9.1 database (Zdobnov et al., 2017) for single copy orthologs in six beetle genomes *Onthophagus taurus* (Poelchau et al., 2015), *Tribolium castaneum* (Shelton et al., 2015; Richards et al., 2008) *Dendroctonus ponderosae* (Keeling et al., 2013), *Anoplophora glabripennis* (McKenna et al., 2017), *Leptinotarsa decemlineata* (Poelchau et al., 2015) and *Agrilus planipennis* (Poelchau et al., 2015; Appendix 1—table 3). OrthoDB v.9.1 predicted 4225 single copy orthologs for beetle species and Coleoptera (Polyphaga) reference node. We used Orthograph v.0.6.3 (Petersen et al., 2017) with default settings to search in our assemblies for the presence of specified single copy orthologs. From the recovered 4193 orthologs, terminal stop codons were removed, and internal stop codons at the translational and nucleotide levels were masked using the Perl script summarize_orthograph_results.pl (Petersen et al., 2017). The amino acid sequences were aligned using MAFFT v.7.471 with the L-INS-i algorithm (Kato and Standley, 2013). The alignments from each ortholog group were then checked for the presence of outliers using the script checker_complete.1.3.1.2.pl, according to previously published methods (Misof et al., 2014; Peters et al., 2017). Identified outlier sequences were removed from amino acid and nucleotide alignments. After this step, the sequences of reference taxa and all gap-only sites were pruned. Corresponding multiple sequence alignments of nucleotides were generated using Pal2Nal (Suyama et al., 2006). To identify random or ambiguous similarities within amino acid alignments, we used Aliscore v. 2.076 with the maximum number of pairwise comparisons $-r$ 1027, a special scoring approach for gap-filled amino acid sites option $-e$, and other parameters set to default values. Any random or ambiguous similarities were masked using Alicut 2.3 (Kück et al., 2010). Alinuc.pl was then used to apply the Aliscore results to match amino acids to the nucleotide data (Peters et al., 2017). Additionally, in each gene alignment, the short randomly aligned fragments were replaced with gaps and sequences with $\geq 80\%$ missing data (calculated as percentage of '-' and 'X' in the amino acid alignments and '-' and 'N' in the nucleotide alignments) were removed using Python scripts (Zhang, 2021). MARE v. 0.1.2-rc was used to calculate the information content of each gene partition in terms of amino acid coding (Misof et al., 2013). Partitions with zero information content (IC0) were removed at both amino acid and nucleotide level.

Finally, the remaining 4,109 alignments were retained for subsequent multispecies coalescent analyses, and different concatenated datasets were generated for both amino acid and nucleotide levels using FasConCat-G v.1.4 (Kück and Longo, 2014). With all 4109 genes, we generated datasets designated as A-4199-AA and B-4199-NT (name-number of partitions-level). To reduce the effect of saturation on phylogenetic reconstruction (Breinholt and Kawahara, 2013) we created the dataset C-4199-NT12 and E-4109-NT2, with third-codon or third and first-codon positions excluded. The degree of missing data and overall completeness scores (Ca) across all datasets were inspected using AliStat v.1.7 (Wong, 2021) and heatmaps of pairwise completeness scores for all analysed datasets were generated (see Appendix 1—table 3). Further Information content and saturation values (the overall degree of data coverage with respect to gene presence or absence) were calculated with MARE v.0.1.2-rc (MAtrix REduction) (Misof et al., 2013) for each dataset.

Compositional heterogeneity tests

To explore the effect of among species compositional heterogeneity and its possible bias to tree reconstruction, we inspected the data with BaCoCa v.1.105 (Kück and Struck, 2014) to identify the gene partitions that strongly deviate from compositional homogeneity using relative composition frequency variation value (RCFV). Following Fernández et al., 2016 and Vasilikopoulos et al., 2019, we considered compositional heterogeneity among species in a partition to be high when the overall RCFV value was ≥ 0.1 . The compositionally heterogeneous partitions were excluded from the dataset A-4109-AA to generate more compositionally homogeneous dataset D-3370-AA. To create more homogenous and decisive dataset, we reduced the dataset D-3370-AA to contain only partitions with all taxa and created dataset F-1490-AA. Additionally, we used MARE to increase information content of dataset D-3370-AA and default output supermatrix is designed as dataset J-2129-AA_MARE. We used tree matched-pairs

tests of homogeneity the MPTS (matched- pairs test of symmetry), MPTMS (matched-pairs test of marginal symmetry), and MPTIS (matched- pairs test of internal symmetry) (**Naser-Khdour et al., 2019**) implemented in IQ-TREE2 to identify the partitions that strongly deviate from compositional homogeneity at nucleotide level (P-value cutoff <0.05) and partitions below the threshold were excluded. And following datasets generated: G-NT-1767_MaxSymTest, H-NT-1645_MaxSymTestmarginal and I-NT-3905_MaxSymTestInternal.

The software SymTest v.2.0.49 (<https://github.com/ottmi/symtest>) was used to calculate the overall deviation from stationarity, reversibility, and homogeneity (SRH) (**Ababneh et al., 2006**) between the amino acid and nucleotide sequences. Heatmaps were generated for all datasets to visualize the pairwise deviations from SRH conditions.

Phylogenetic maximum likelihood analyses

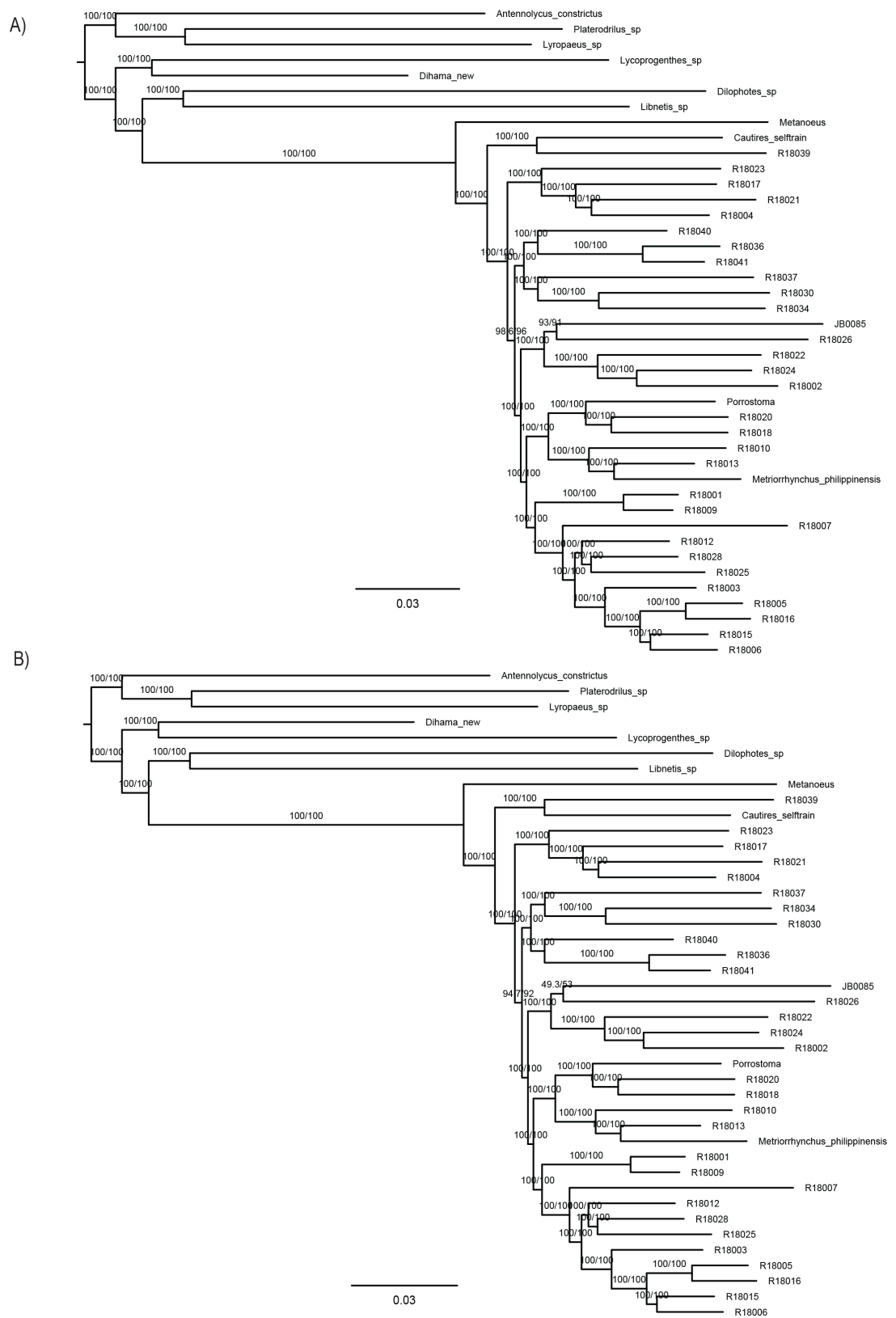
For all datasets phylogenetic reconstruction was performed using maximum likelihood (ML) criterion with IQ-TREE 2.1.2 (**Minh et al., 2020**). First, we analyzed all datasets using original gene partition boundary. The model selection for each gene was performed with ModelFinder (**Chernomor et al., 2016; Kalyaanamoorthy et al., 2017**) implemented in IQ-TREE2 (-MFP option). GTR model were considered for nucleotide supermatrices. For the amino acid supermatrices, the substitution models LG, DCMUT, JTT, JTTDCMUT, DAYHOFF, WAG and free rate models LG4X and LG4M were tested. All possible combinations of modeling rate heterogeneity among sites were allowed (options: -mrate E,I,G,I + G,R -gmedian -merit BIC). We used the edge-linked partitioned model for tree reconstructions (-spp option) allowing each gene to have its own rate. The optimized partition schemes and best-fitting models were inferred for dataset A-4109-AA using -m MFP+ MERGE and considering same substitution models as above. The fast-relaxed clustering algorithm was used to speed up computation during partition-scheme optimization (**Lanfear et al., 2017**). The top 10% of partitions schemes --rclusterf 10 and maximum 12 327 partitions pairs (three times the number of partitions) --rcluster-max 12 327 were considered. Dataset A-4109-AA was also analyzed without any prior partitioning and under best fitting JTT + F + R9 substitution model. Ultrafast bootstrap (**Hoang et al., 2018**) and SH-like approximate likelihood ratio test (SH-aLRT) were calculated in IQ-TREE2 (options -bb 5000 and -alrt 5000) to assess nodal supports for focal relationships (**Appendix 1—table 3**).

Analyses of the confounding and alternative signal with four-cluster likelihood mapping (FcLM) and sequence data permutations

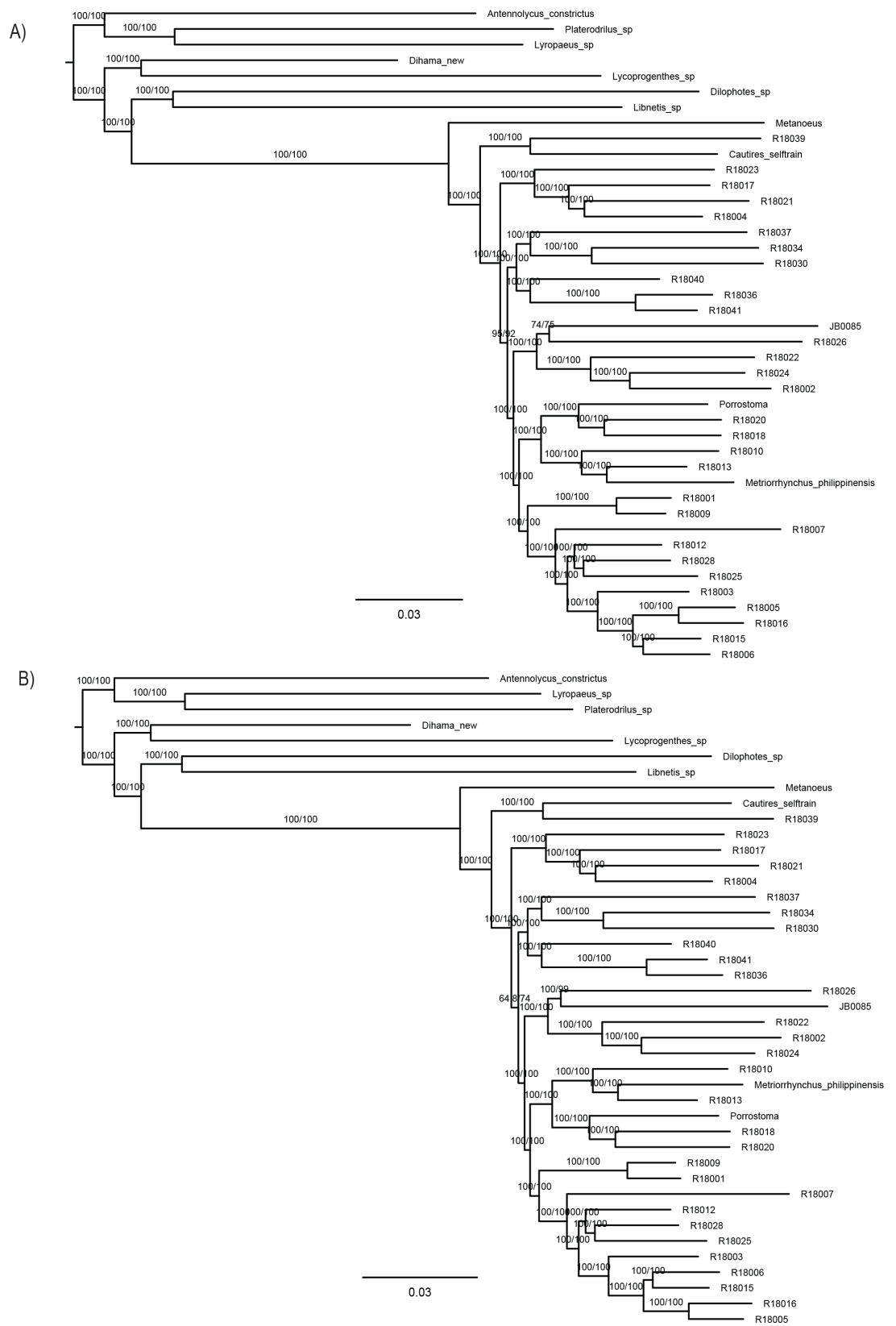
Additionally, we studied the effect of potentially confounding signal, like non-random distribution of data coverage and violations of SRH conditions, on our phylogenetic reconstructions with the Four- cluster likelihood mapping (FcLM) approach (**Strimmer and Haeseler, 1997**) as described by **Misof et al., 2014** and implemented in IQ-TREE2. Based on the results of our ML tree reconstructions we tested the hypotheses about alternative placement of the leptotrichaline and procautirine clades. Permutation of data were executed as described in **Misof et al., 2014** Our approach only differs in the use of the LG substitution matrix (**Le and Gascuel, 2008**) for permuting the supermatrices instead of WAG substitution matrix. For the analysis of the permuted supermatrices we used the option -q for the partition model in IQ-TREE2 and we used the LG model (for amino-acid alignments) and GTR model (for the nucleotide alignments). With original data, we performed the FcLM analyses by applying previously estimated substitution models (edgeloinked models, -spp) with IQ-TREE2.

For estimation of *Cautires apterus* and sexually dimorphic *Metriorrhynchus* sp. splits were used subsets of the samples which represent the most common lineages. Dating of forementioned diversification was analyzed in BEAST v.1.8.1 (**Drummond et al., 2012**) applying HKY + I + G substitution model, uncorrelated relaxed clock model and birth–death process tree prior and 0.0115 rate for *cox1* gene as applied earlier by **Bocak and Yagi, 2010**. The stationary phase was checked in Tracer v.1.5 (<http://beast.bio.ed.ac.uk/Tracer>). The consensus tree was computed with TreeAnnotator v.1.8.2 (<http://beast.bio.ed.ac.uk/treeannotator>) after eliminating a part of trees as a burn-in after evaluating ESS in tracer v.1.5. The final trees were visualised in Figtree v.1.4.4.

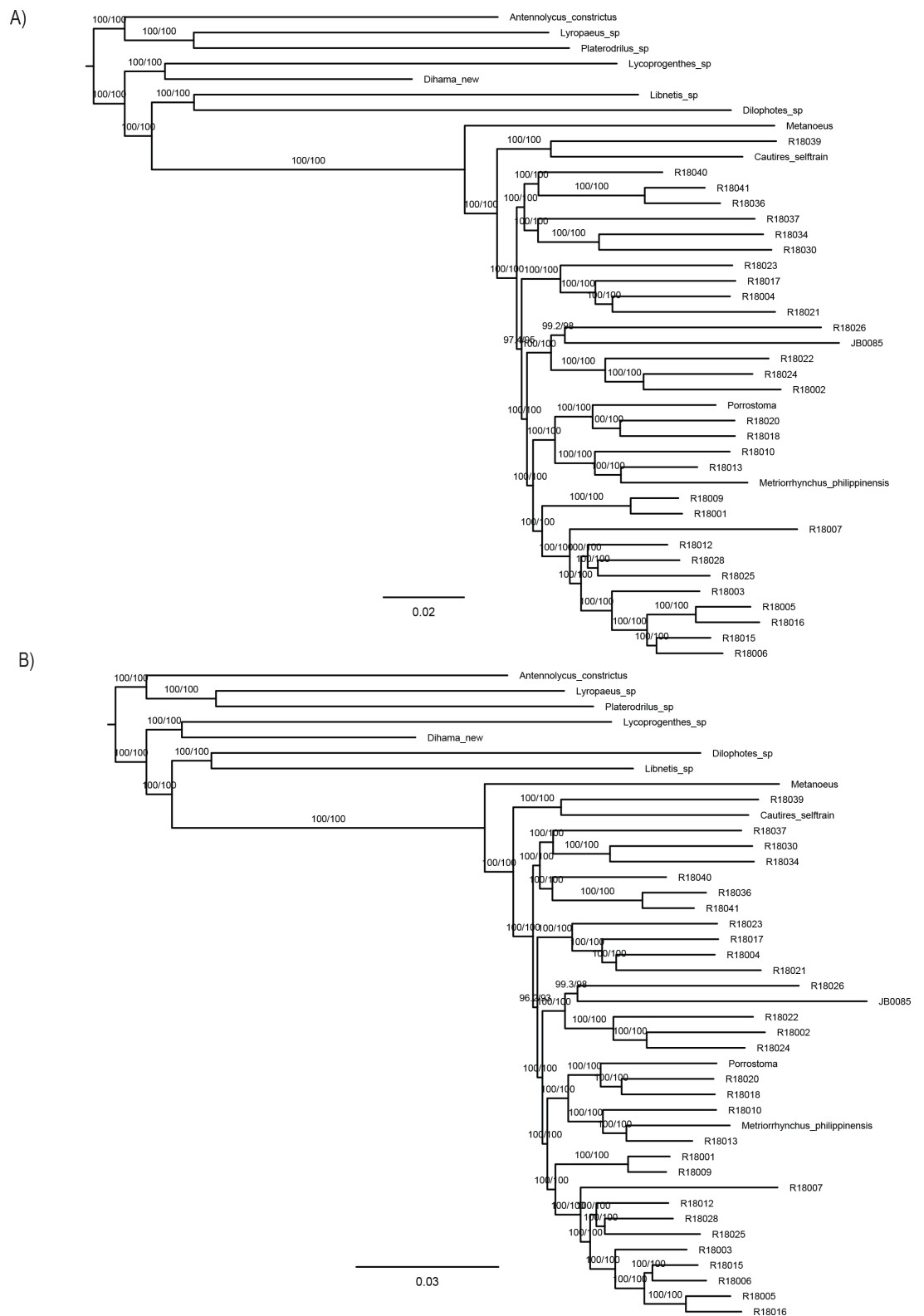
Data deposition: The additional supporting data and all the analysed supermatrices are deposited in the Mendeley Data repository DOI:<https://doi.org/10.17632/ntgg6k4fjx.1>.



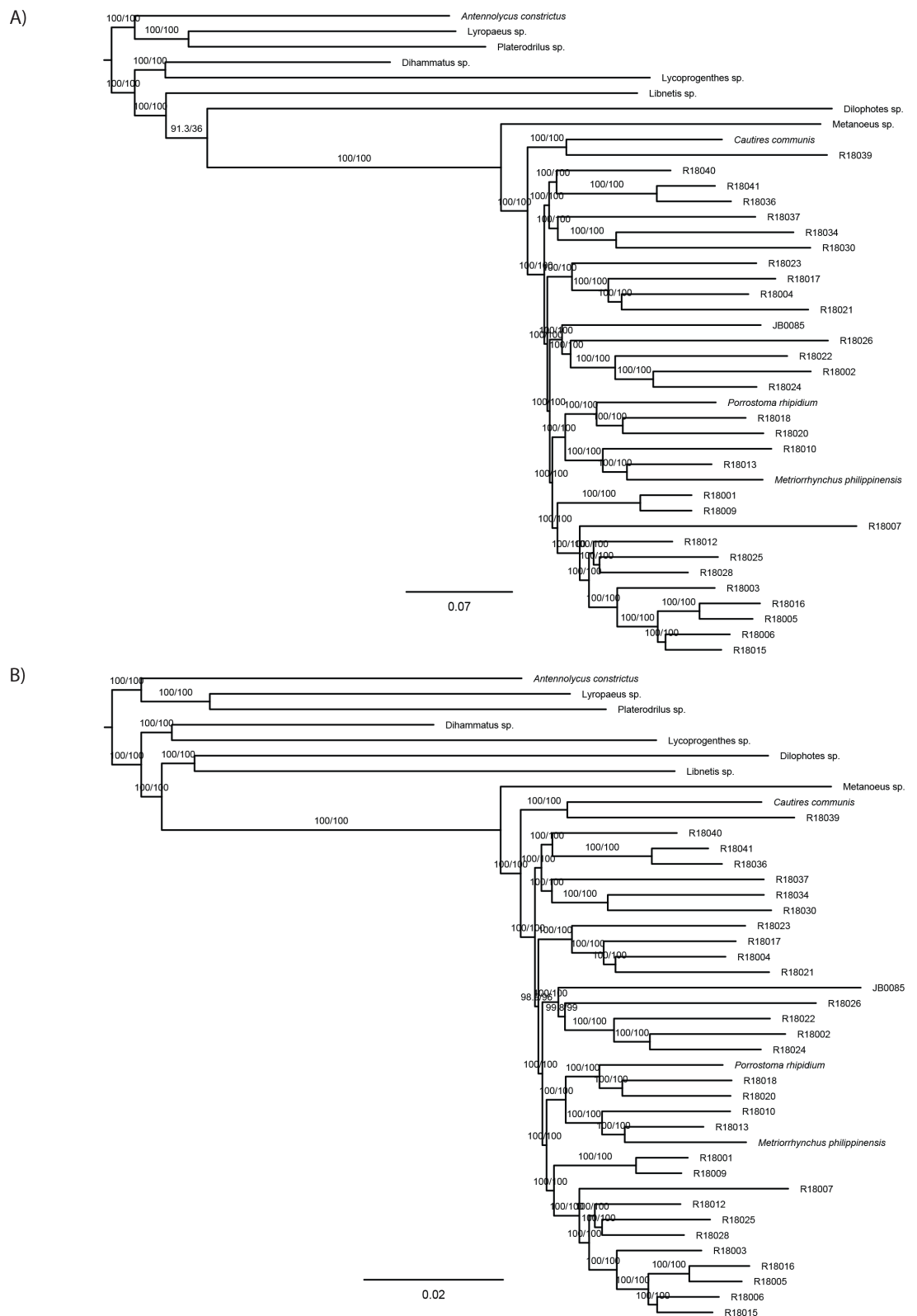
Appendix 1—figure 1. Maximum likelihood trees from IQ-TREE amino acid analysis of dataset A –4109-AA. (A) With partitioning by gene and (B) without partitioning. The depicted branch support values represent SH-aLRT and ultrafast bootstrap.



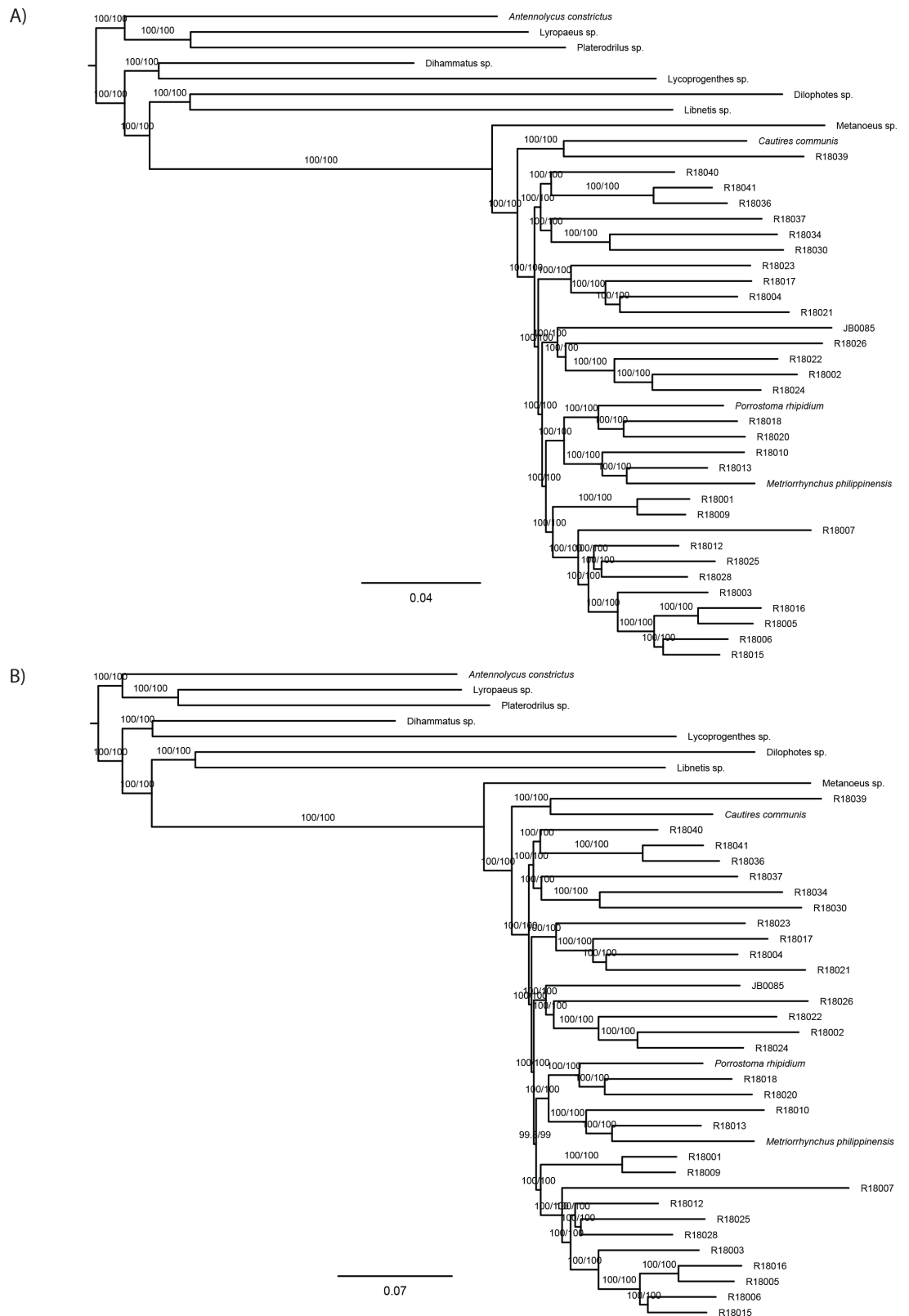
Appendix 1—figure 2. Maximum likelihood trees from IQ-TREE amino acid analysis. **(A)** Analysis of the dataset A-4109-AA with optimization of the partitioning scheme and **(B)** analysis of the dataset D-3370-AA_Bacoca with partitioning by gene. The depicted branch support values represent SH-aLRT and ultrafast bootstrap.



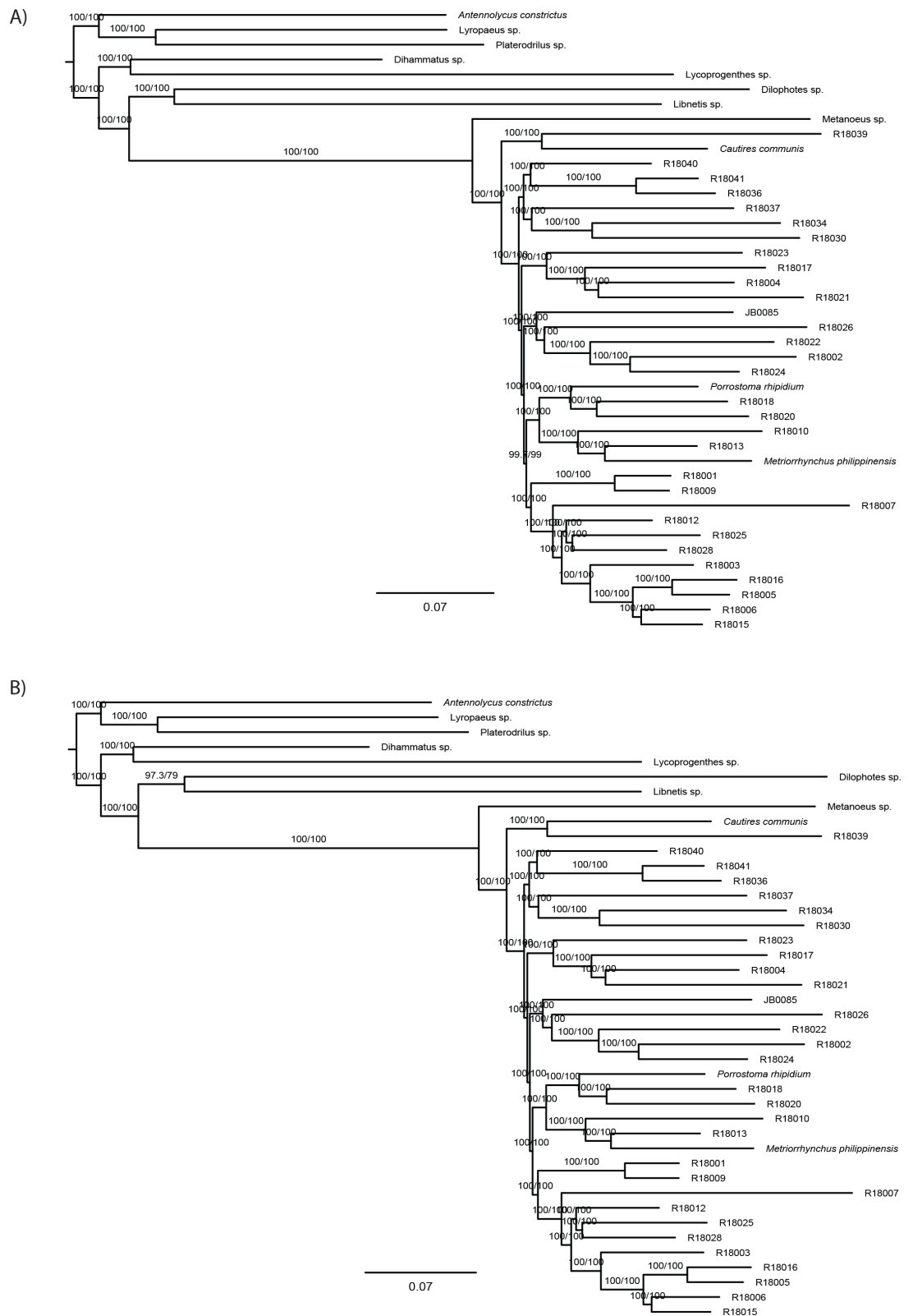
Appendix 1—figure 3. Maximum likelihood trees from IQ-TREE amino acid analysis. (A) analysis of the dataset A-F-1490-AA_Bacoca_decisive and (B) analysis of the dataset J-2129-AA_MARE. Both datasets were partitioned by gene. The depicted branch support values represent SH-aLRT and ultrafast bootstrap.



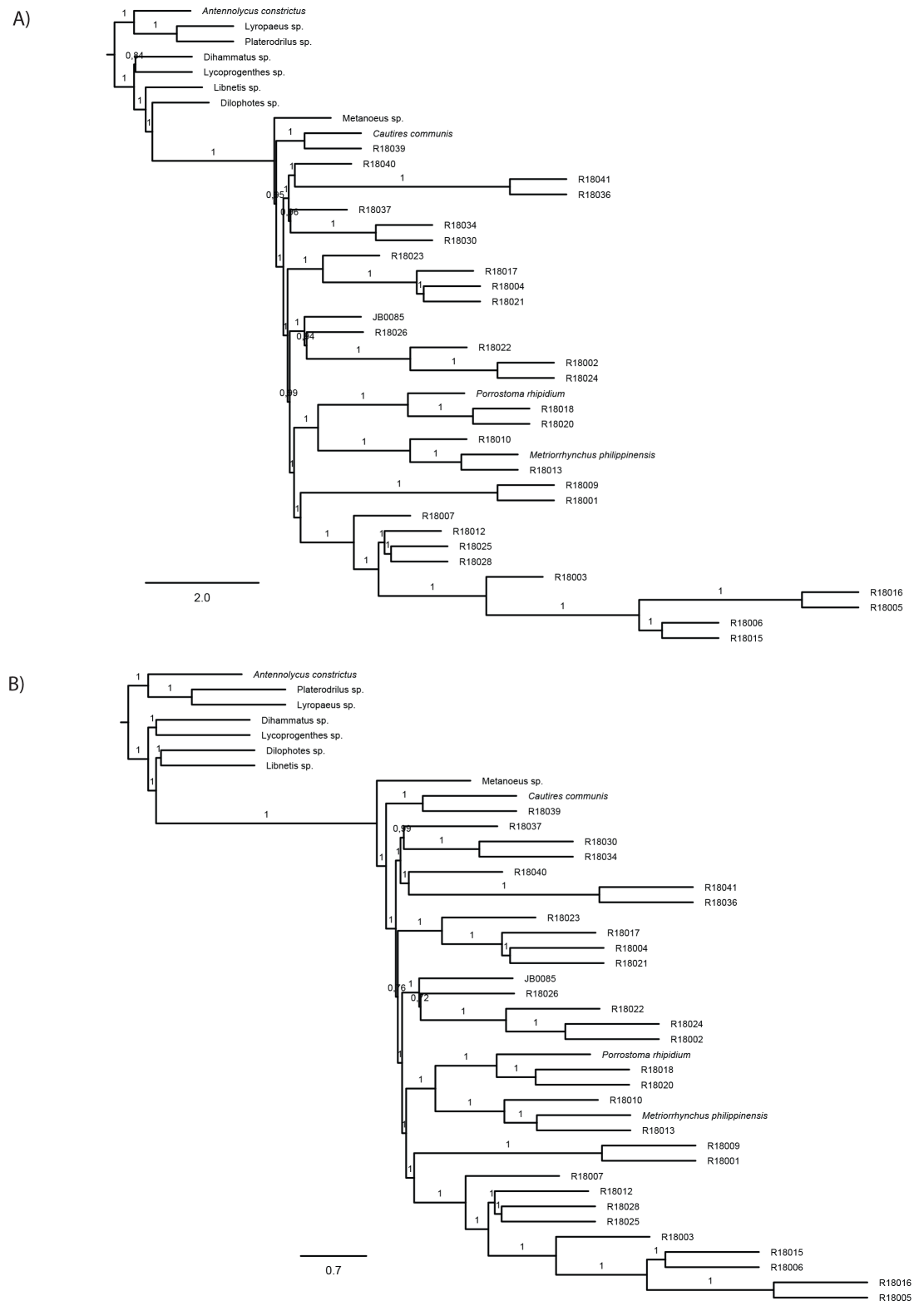
Appendix 1—figure 4. Maximum likelihood trees from IQ-TREE nucleotide analysis. **(A)** Analysis of the dataset B-4109-NT and **(B)** analysis of the dataset E-4109-NT2 using only second codon positions. Both datasets were partitioned by gene. The depicted branch support values represent SH-aLRT and ultrafast bootstrap.



Appendix 1—figure 5. Maximum likelihood trees from IQ-TREE nucleotide analysis. (A) analysis of the dataset C-4109-NT12 using codon positions 1 + 2 and (B) analysis of the dataset G-NT-1767_MaxSymTest. Both datasets were partitioned by gene. The depicted branch support values represent SH-aLRT and ultrafast bootstrap.



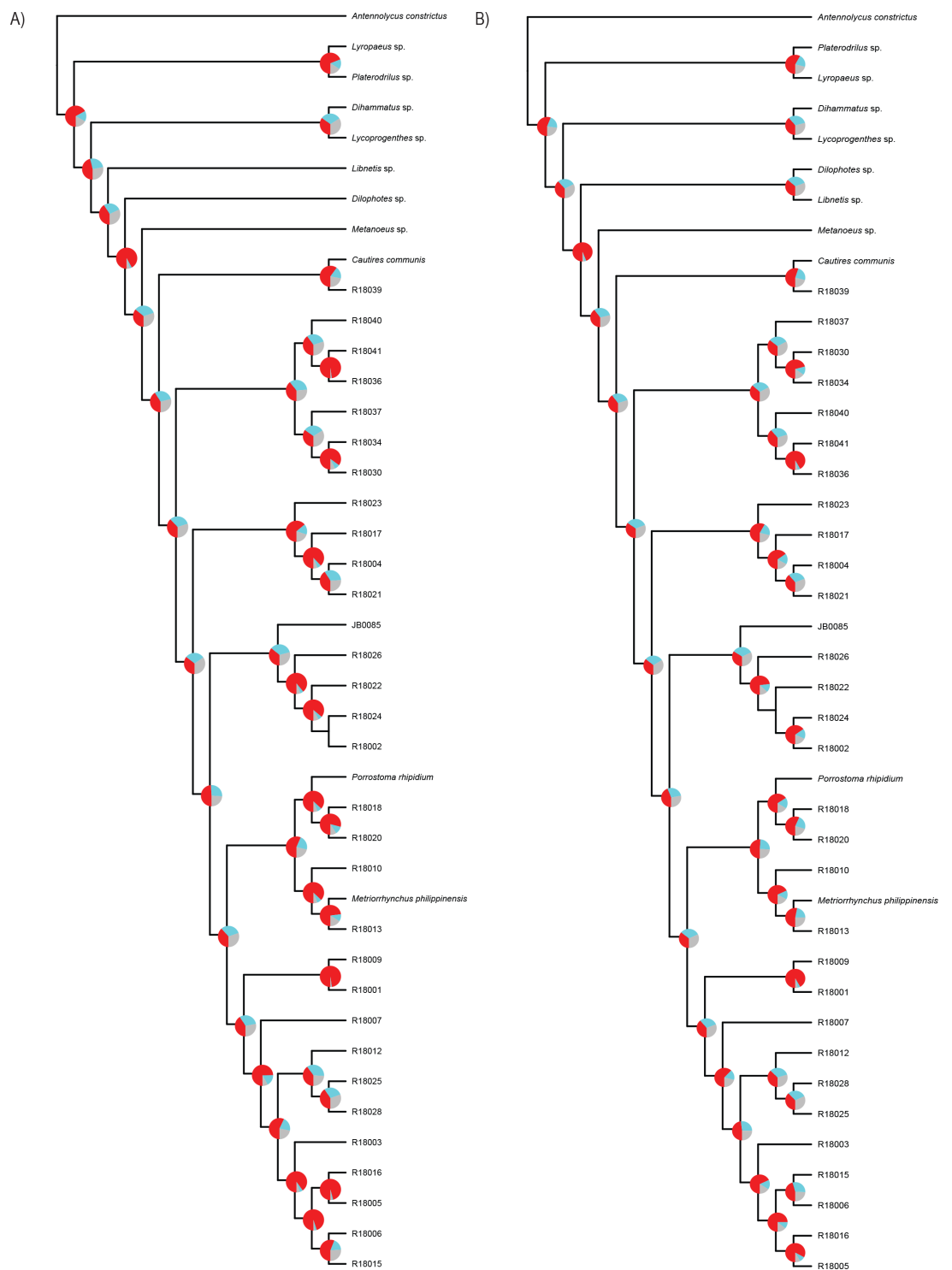
Appendix 1—figure 6. Maximum likelihood trees from IQ-TREE nucleotide analysis. **(A)** analysis of the dataset H-NT-1645_MaxSymTestmarginal and **(B)** analysis of the dataset I-NT-3905_MaxSymTestInternal. Both datasets were partitioned by gene. The depicted branch support values represent SH-aLRT and ultrafast bootstrap.



Appendix 1—figure 7. Topologies recovered by Astral analyses. **(A)** ASTRAL species trees with branch lengths in coalescent units as resulted from the analyses of individual IQ-TREE maximum likelihood gene trees of nucleotide dataset B-4109-NT and **(B)** ASTRAL species trees with branch lengths in coalescent units as resulted from the analyses of individual IQ-TREE maximum likelihood gene trees of amino acid dataset A-4199-AA. Numbers on nodes show local posterior probabilities
Appendix 1—figure 7 continued on next page

Appendix 1—figure 7 continued

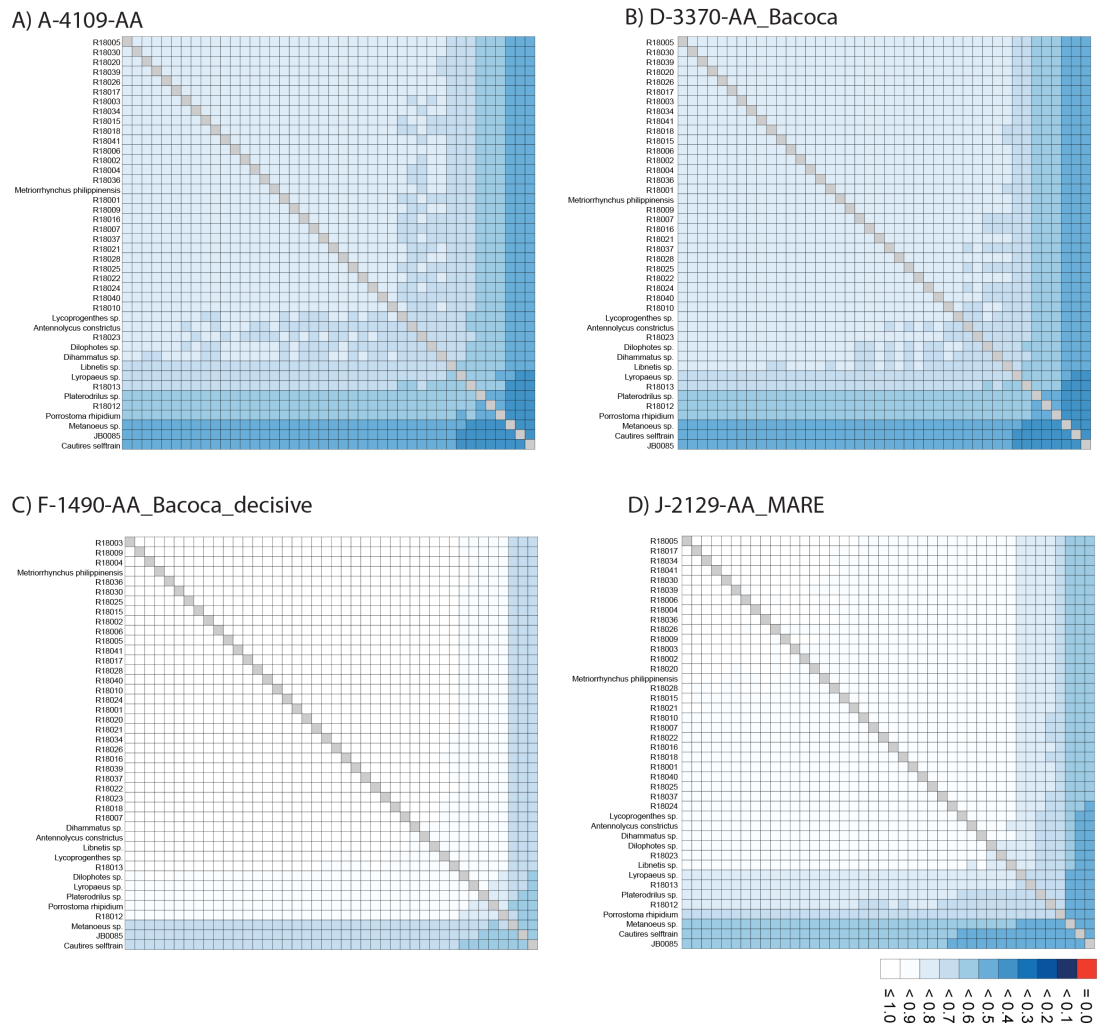
(pp1). Quartet support for the alternative topologies (q1, q2, and q3), total number of induced quartet trees in the gene trees that support the alternative topologies (f1, f2, f3) and local posterior probabilities (pp1, pp2, pp3) are available at Dryad repository.



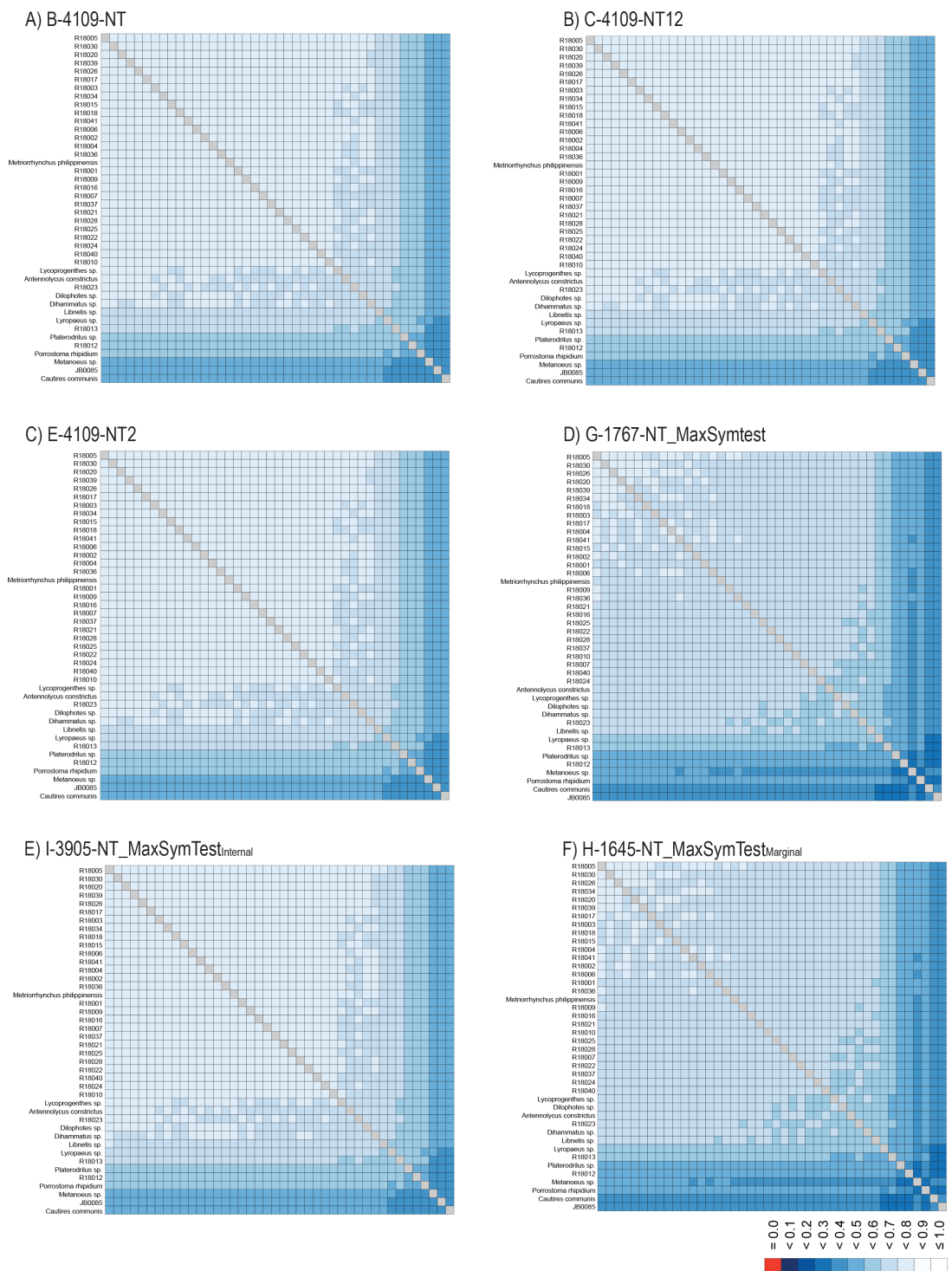
Appendix 1—figure 8. Phylogenomic topologies. (A) Phylogenetic relationships of Metriorrhynchini, resulted from the summary coalescent phylogenetic analysis with ASTRAL, when analyzing the full Appendix 1—figure 8 continued on next page

Appendix 1—figure 8 continued

set of gene trees (4109 gene trees) inferred at the nucleotide level. (B) doitto at amino acids level. Pie charts on branches show ASTRAL quartet supports q1,q2,q3 (quartet-based frequencies of alternative quadripartition topologies around a given internode). Tree topologies correspond with **Appendix 1—figure 7**.

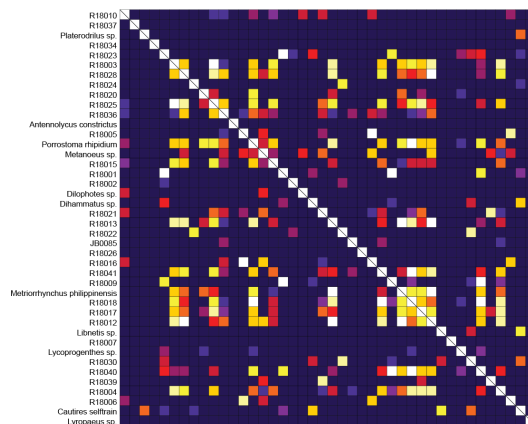


Appendix 1—figure 9. Alistat heatmaps. (A–D) AliStat rectangular heatmaps showing pairwise alignment completeness scores for all species included in the analyzed amino acid supermatrices. The abbreviations of datasets correspond to **Appendix 1—table 4**. Values closer to one indicate higher completeness scores for the pairwise sequence comparisons.

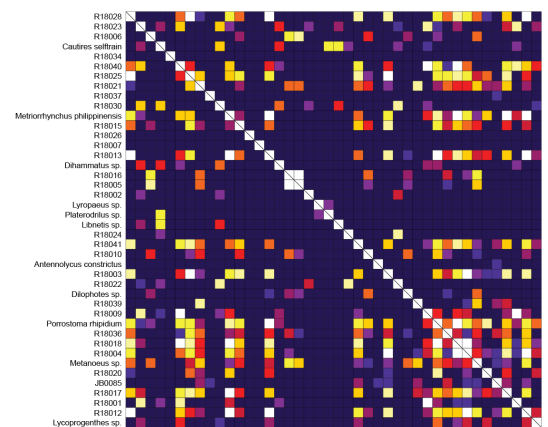


Appendix 1—figure 10. AliStat heatmaps. (A–F) AliStat rectangular heatmaps showing pairwise alignment completeness scores for all species included in the analyzed nucleotide supermatrices. The abbreviations of datasets correspond to **Appendix 1—table 4**. Values closer to one indicate higher completeness scores for the pairwise sequence comparisons.

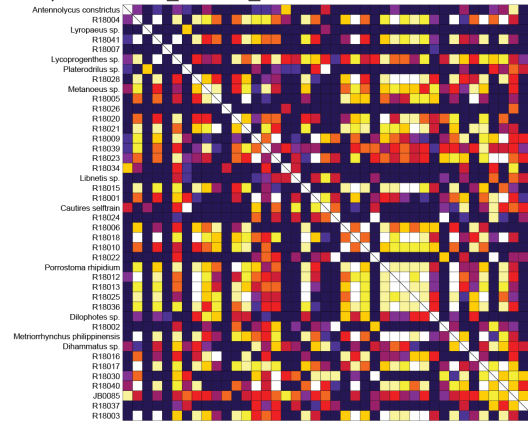
A) A-4109-AA



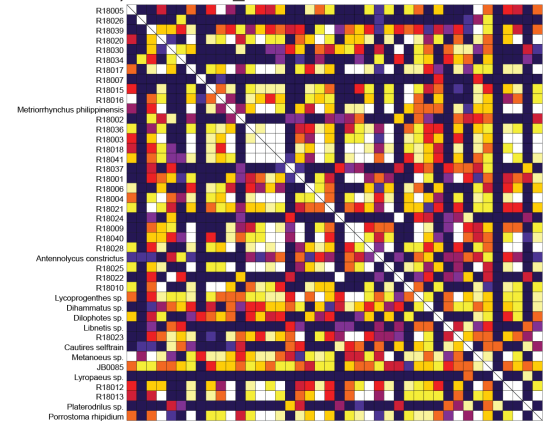
B) D-3370-AA_Bacoca



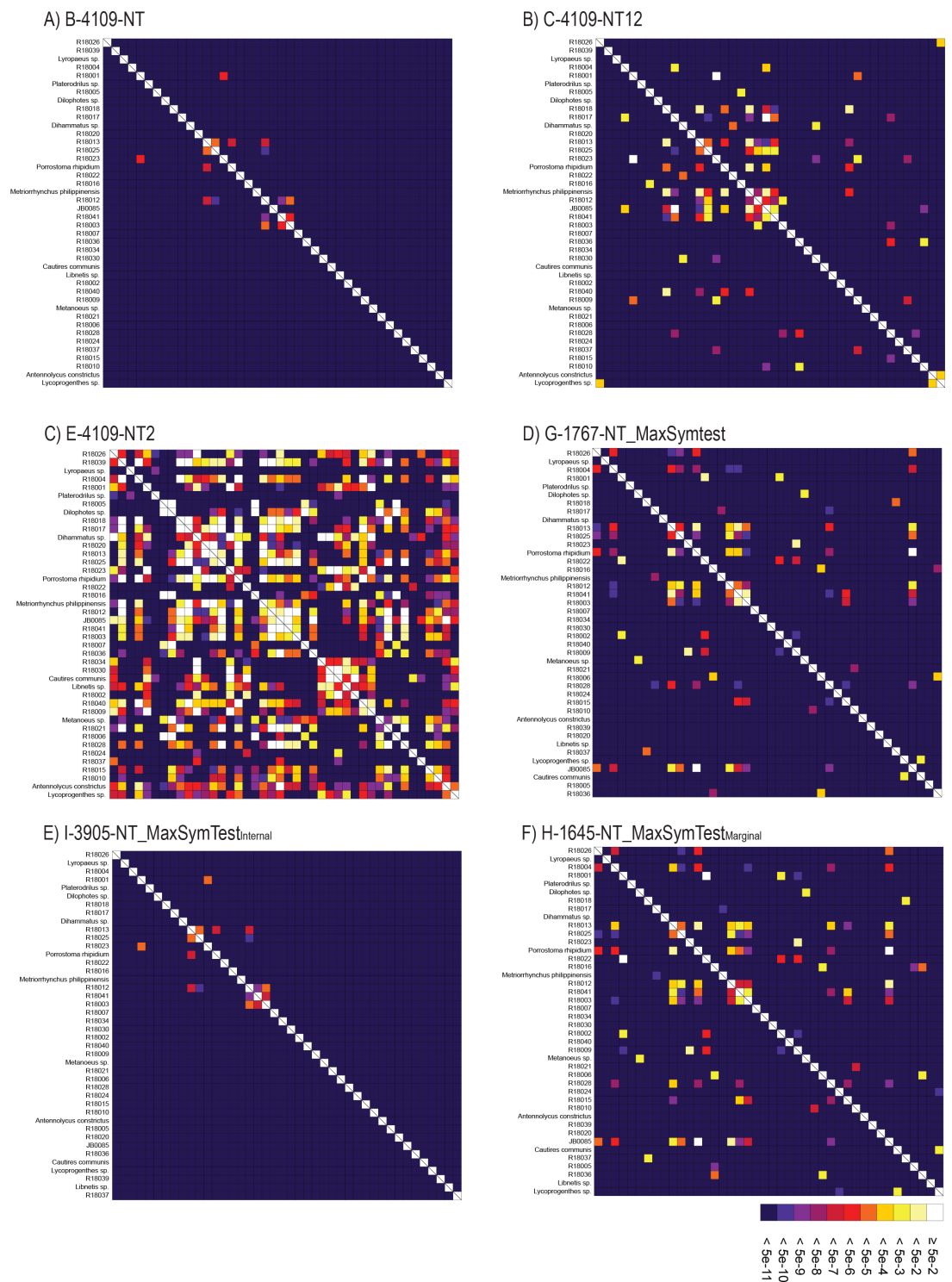
C) F-1490-AA_Bacoca_decisive



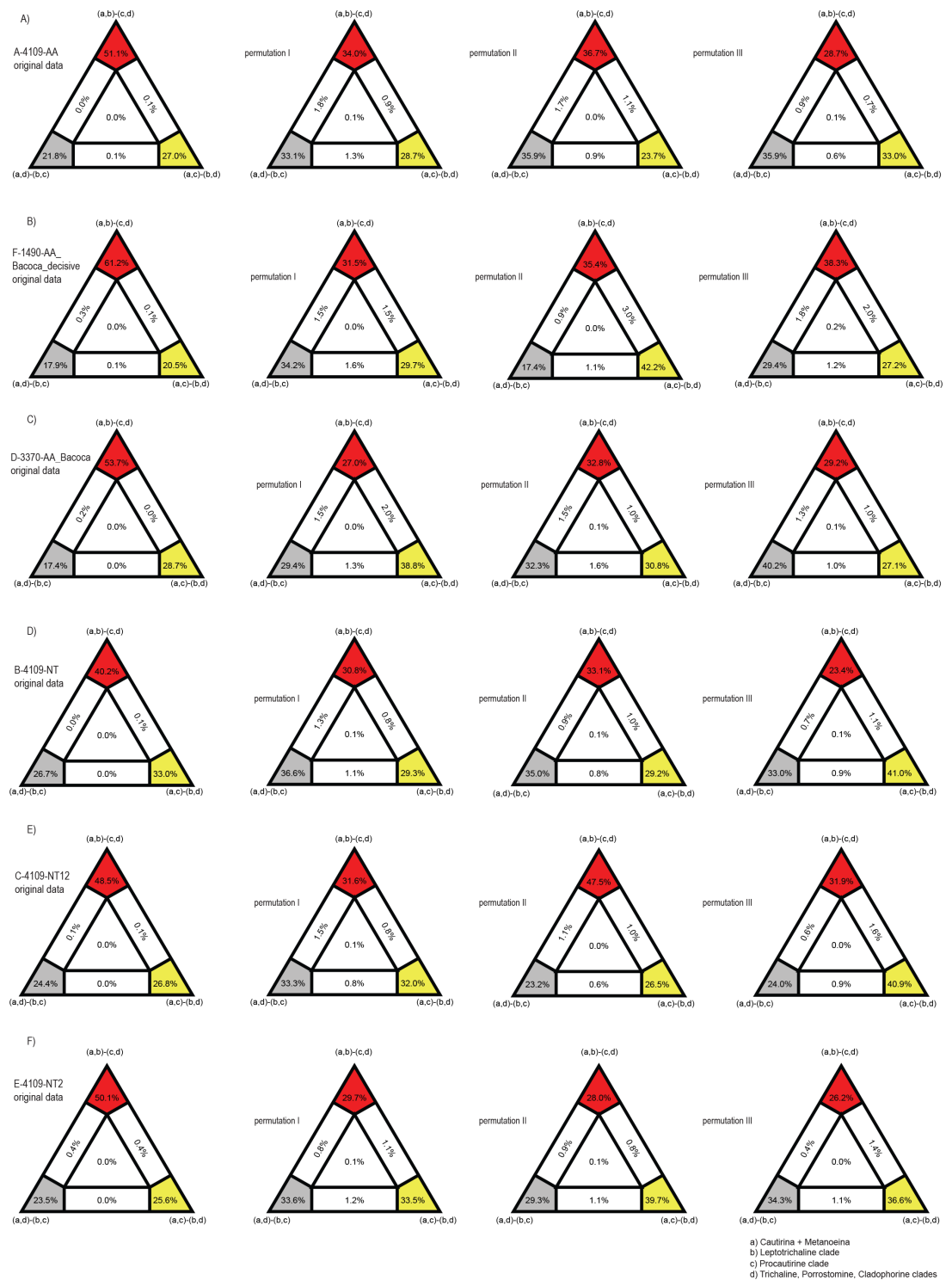
D) J-2129-AA_MARE



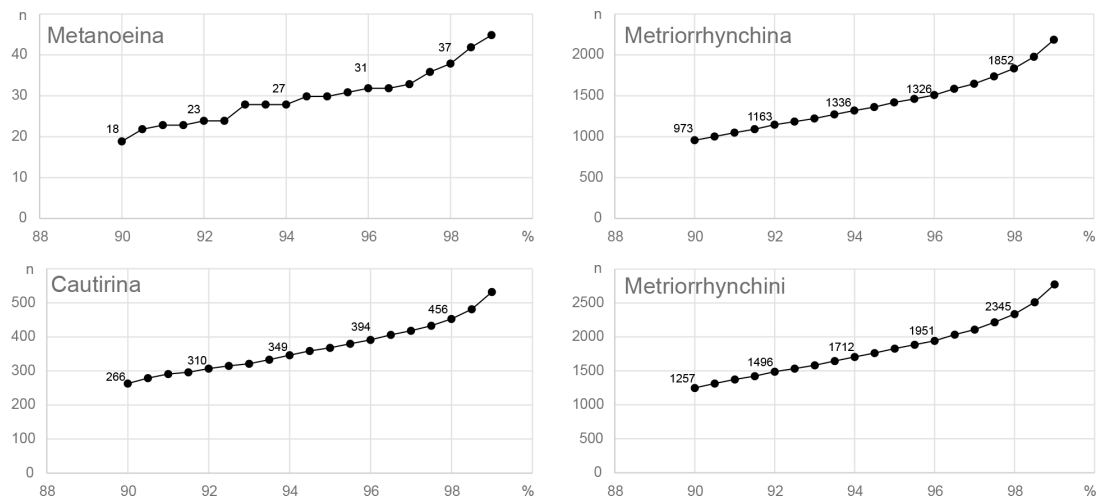
Appendix 1—figure 11. SymTest analyses. (A–D) Rectangular heatmap calculated with SymTest showing p-values for the pairwise Bowker’s tests in the analyzed amino acid supermatrices. Darker boxes indicate lower p-values and thus larger deviation from evolution under SRH conditions.



Appendix 1—figure 12. SymTest analyses. (A–F) Rectangular heat map calculated with SymTest showing p-values for the pairwise Bowker’s tests in the analyzed nucleotide supermatrices. Darker boxes indicate lower p-values and thus larger deviation from the evolution under SRH conditions.

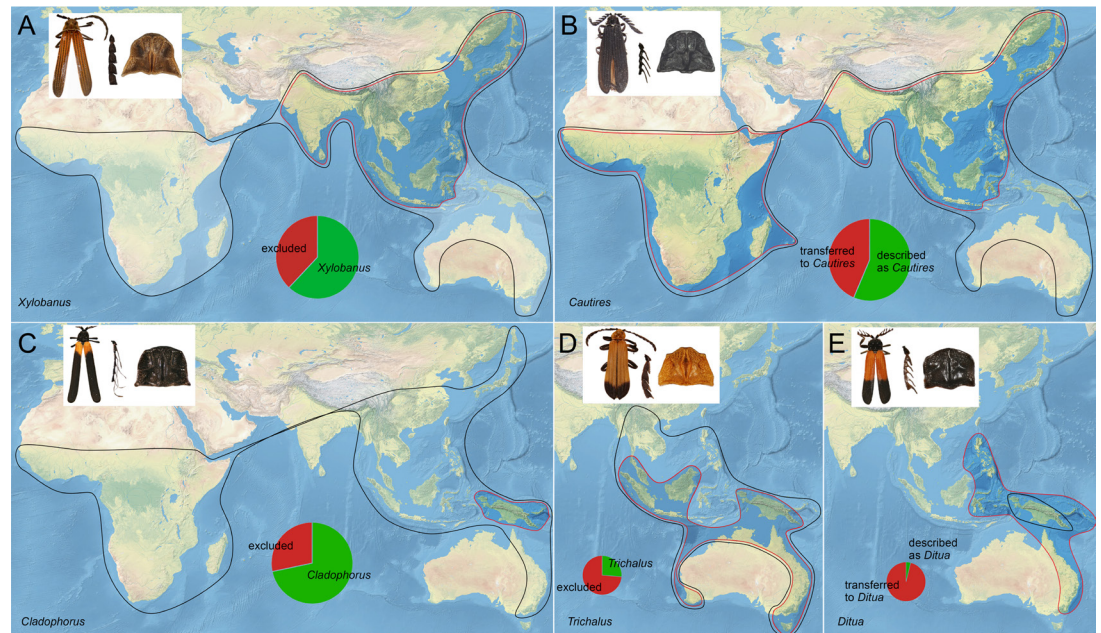


Appendix 1—figure 13. Results of FcLM analyses testing alternative phylogenetic hypotheses about placement of procautirine and leptotrichaline clades applied for various supermatrices. The first column shows the results of FcLM when the original data were analyzed. The second column shows the results of FcLM after the phylogenetic signal had been eliminated from data. The third column show the results of FcLM after elimination of the phylogenetic signal and inhomogeneous amino-acid or nucleotide composition. The fourth column show the results of FcLM after the elimination of phylogenetic signal, inhomogeneous amino-acid or nucleotide composition and with randomized data coverage within all meta-partitions (see Supplementary methods for details).

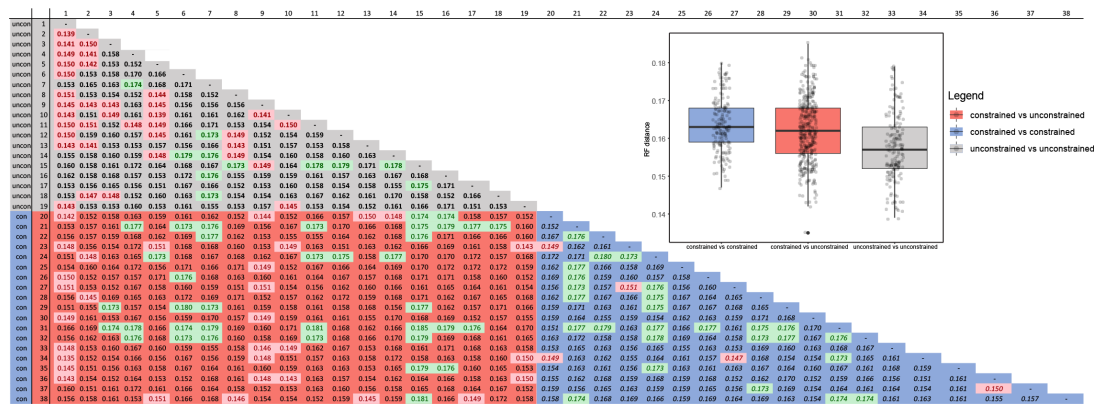


Appendix 1—figure 14. Tribal and subtribal mOTUs delimitation using CD-hit-est.

Axis X represent the number of operational taxonomic units, whereas axis Y shows the delimitation threshold.



Appendix 1—figure 15. Distribution of selected Metriorrhynchini genera (originally assumed and revised).



Appendix 1—figure 16. Comparison of the Robinson-Foulds (RF) distances among tree searches with constrained topology and tree searches with unconstrained topology. Green values represent top 10% RF- distances, and red values shows bottom 10% RF- distances. The blue part of the graph represents constrained trees values, the grey part of the graph represents unconstrained trees values, the red part of the graph represents constrained/unconstrained trees values.

Appendix 1—table 1. Detailed information on regional sampling.

	Metriorrhynchina	Cautirina	Metanoenia	Metriorrhynchini
Region	# of specimens	# of specimens	# of specimens	# of specimens
Australian region	4489			4489
Australian continent	3964			3964
Australia	475			475
New South Wales	64			64
Northern Territory	18			18
Queensland	382			382
South Australia	1			1
Western Australia	10			10
New Zealand	1			1
New Guinea	3461			3461
Solomon Islands	19			19
Aru Islands	8			8
Wallacea	525	15		540
Maluku-Buru	15			15
Halmahera	169			169
Sulawesi	341	15		
Oriental region	278	1,311	113	1,692
Philippines	48	37	14	99
Mindanao	11	3	9	23
Negros	2			
Palawan	29	34	5	68
Sibuyan	4			4
Luzon	2			2

Appendix 1—table 1 Continued on next page

Appendix 1—table 1 Continued

	Metriorrhynchina	Cautirina	Metanoelina	Metriorrhynchini
Continental Asia	230	1264	99	1593
Malaya	34	270	8	348
Java	21	45	6	72
Bali	2	1		3
Sumatra	56	201	26	283
Borneo	82	253	22	357
Cambodia	7	13		20
Indo-Burma	23	145	10	178
China incl. Taiwan	5	100	23	128
India		59		59
Japan		177	4	181
Afrotropical region		233		233
Africa		175		175
Madagascar		58		58
Total	4767	1549	113	6429

Appendix 1—table 2. List of material for phylogenomic analyses.

Ingroup

Voucher	Species	Clade	Geographic origin	Data source	Data type	Tissue type	
R18010	Metriorrhynchus s. l.	Metriorrhynchina	Porrostomine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18013	Metriorrhynchus sp.	Metriorrhynchina	Porrostomine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R20001	<i>Metriorrhynchus philippinensis</i>	Metriorrhynchina	Porrostomine	Philippines	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18018	Metriorrhynchus s. l.	Metriorrhynchina	Porrostomine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18020	Metriorrhynchus s. l.	Metriorrhynchina	Porrostomine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
1_kite	<i>Porrostoma rhipidium</i>	Metriorrhynchina	Porrostomine	Australia	McKenna et al., 2019	ILLUMINA RNA-seq PE-reads	Whole animal
R18001	Cladophorus sp.	Metriorrhynchina	Cladophorine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18009	Cladophorus sp.	Metriorrhynchina	Cladophorine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18007	Pseudodontocerus sp.	Metriorrhynchina	Cladophorine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18012	Ditua s. l.	Metriorrhynchina	Cladophorine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18025	Ditua s. l.	Metriorrhynchina	Cladophorine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18028	Ditua s. l.	Metriorrhynchina	Cladophorine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18003	Ditua s. l.	Metriorrhynchina	Cladophorine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18016	Ditua s. l.	Metriorrhynchina	Cladophorine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18005	Ditua s. l.	Metriorrhynchina	Cladophorine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18015	Ditua s. l.	Metriorrhynchina	Cladophorine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18006	Ditua s. l.	Metriorrhynchina	Cladophorine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18002	Microtrichalus sp.	Metriorrhynchina	Trichaline	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18024	Eniclases sp.	Metriorrhynchina	Trichaline	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal

Appendix 1—table 2 Continued on next page

Appendix 1—table 2 Continued

Ingroup							
Voucher	Species		Clade	Geographic origin	Data source	Data type	Tissue type
R18022	Diatrichalus sp.	Metriorrhynchina	Trichaline	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18026	Trichaline sp.	Metriorrhynchina	Trichaline	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
JB0085	Trichaline sp.	Metriorrhynchina	Trichaline	New Guinea	This study	ILLUMINA WGS-seq PE-reads	Thoracic muscles
R18004	Procautires sp.	Metriorrhynchina	Procautirine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18017	Procautires sp.	Metriorrhynchina	Procautirine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18021	Procautires sp.	Metriorrhynchina	Procautirine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18023	Procautires sp.	Metriorrhynchina	Procautirine	New Guinea	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18037	<i>Broxylus</i> sp.	Metriorrhynchina	Leptotrichaline	Sulawesi	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18030	Leptotrichalus sp.	Metriorrhynchina	Leptotrichaline	Sulawesi	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18034	Wakarumbia sp.	Metriorrhynchina	Leptotrichaline	Sulawesi	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18036	<i>Sulabanus</i> sp.	Metriorrhynchina	Leptotrichaline	Sulawesi	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18041	<i>Sulabanus</i> sp.	Metriorrhynchina	Leptotrichaline	Sulawesi	This study	ILLUMINA RNA-seq PE-reads	Whole animal
R18040	<i>Mangkutanus</i> sp.	Metriorrhynchina	Leptotrichaline	Sulawesi	<i>Kusy et al., 2019</i>	ILLUMINA RNA-seq PE-reads	Whole animal
R18039	Xylobanus sp.	Cautirina		Sulawesi	This study	ILLUMINA RNA-seq PE-reads	Whole animal
AJ0013	<i>Cautires communis</i>	Cautirina		Malaysia	<i>Kusy et al., 2019</i>	ILLUMINA WGS-seq PE-reads	Thoracic muscles
G19002	Metanoes sp.	Metanoeina		Malaysia	This study	ILLUMINA WGS-seq PE-reads	Thoracic muscles
Outgroup							
	Dilophotes sp.			Malaysia	<i>Kusy et al., 2019</i>	ILLUMINA RNA-seq PE-reads	Whole animal
	Dihammatus sp.			Malaysia	<i>Kusy et al., 2019</i>	ILLUMINA RNA-seq PE-reads	Whole animal
	<i>Lycoprogenes</i> sp.			Malaysia	<i>Kusy et al., 2019</i>	ILLUMINA RNA-seq PE-reads	Whole animal
	<i>Libnetis</i> sp.			Malaysia	<i>Kusy et al., 2019</i>	ILLUMINA RNA-seq PE-reads	Whole animal
	Platerodrilus sp.			Malaysia	<i>Kusy et al., 2019</i>	ILLUMINA RNA-seq PE-reads	Whole animal
	<i>Lyropaeus optabilis</i>			Malaysia	<i>Kusy et al., 2019</i>	ILLUMINA RNA-seq PE-reads	Whole animal
	<i>Antennolycus constrictus</i>			Malaysia	<i>Kusy et al., 2019</i>	ILLUMINA RNA-seq PE-reads	Whole animal

Appendix 1—table 3. Overview of official gene sets of six reference species used for ortholog assessment.

Number of genes correspond with OrthoDB 9.1.

Species	Accession	# of contigs	Source	Download date	Reference
Onthophagus taurus	PRJNA167478	17,483	i5K	05.03.2017	1
Tribolium castaneum	PRJNA12540	16,631	iBeetle	05.03.2017	2, 3
Dendroctonus ponderosae	PRJNA360270	13,088	ENS Metazoa	05.03.2017	4
Anoplophora glabripennis	PRJNA167479	22,035	i5K	05.03.2017	5
Leptinotarsa decemlineata	PRJNA171749	24,671	i5K	05.03.2017	1
Agrilus planipennis	PRJNA230921	15,497	i5K	05.03.2017	1

Appendix 1—table 4. Detailed information and statistics of each generated dataset.

Dataset name	Number of taxa	Number of partitions	Number of alignment sites	Completeness score (Ca) AliStat	Percentage of pairwise P-values < 0.05 for the Bowker's test	(SV) MARE Matrix saturation	(IC) MARE Information content	Partition scheme optimization
Amino acids								
A-4109-AA	42	4,109	1892691	0.772127	98.61%	0.888	0.506	Yes
D-3370-AA_Bacoca	42	3,370	1672362	0.780363	97.91%	0.880	0.511	
F-1490-AA_Bacoca_decisive	42	1,490	673,102	0.925082	94.08%	1	0.594	
J-2129-AA_MARE	42	2,129	959,741	0.876331	90.82%	0.964	0.648	
Nucleoides								
B-4109-NT	42	4,109	5678073	0.772133	100%	NA	NA	
C-4109-NT12	42	4,109	3785382	0.772133	99.77%	NA	NA	
E-4109-NT2	42	4,109	1892691	'0.772134	92.45%	NA	NA	
G-NT-1767_MaxSymTest	42	1,767	2413164	'0.713912	99.88%	NA	NA	
H-NT-1645_MaxSymTestmarginal	42	1,645	2233485	'0.712208	99.77%	NA	NA	
I-NT-3905_MaxSymTestInternal	42	3,905	5377449	'0.771645	100%	NA	NA	

Appendix 1—table 5. Characteristics of concatenated super-matrices (mitochondrial fragments) and used models of the DNA evolution (descriptions of columns abbreviation listed bellow tables).

Abbreviations. Seq – Number of sequences, Site – Number of bases, Infor – Number of parsimony-informative sites, Invar – Number of invariant sites, Model – Best-fit model according to BIC using ModelFinder.

Subset	Seqs	Sites	Infor	Invar	Model
Complete dataset					
<i>nad5</i>	3,205	987	809	88	GTR + F + I + G4
<i>cox</i>	5,935	1,068	812	134	GTR + F + I + G4
16 s	2,381	875	577	201	GTR + F + I + G4
98% OTUs delimitation dataset					
<i>nad5</i>	1,252	969	776	115	GTR + F + I + G4
tRNA <i>Leu</i>	2,245	62	34	17	GTR + F + G4
<i>cox1</i>	2,395	783	536	141	GTR + F + I + G4
<i>cox2</i>	2,192	216	194	11	GTR + F + I + G4
16 s	1,115	834	512	237	GTR + F + I + G4