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, Palacky University, Slechtitelu 27, 783 71 Olomouc, Czech Republic
#The authors equally contributed to the project
*corresponding author

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 ~6,500 terminals, >2,300 putative species, more than a half of them unknown to science. The phylogenomic backbone enabled the integrative delimitation of robustly defined natural units that will inform future research. Using constrained mtDNA analysis, we identified the spatial structure of α-diversity, very high species-level endemism, a biodiversity hotspot in New Guinea, and high phylogenetic diversity in the Sundaland. We suggest that ~20 person months of 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. 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.

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; Sriwathsan *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). To accelerate the cataloguing of biodiversity, it is vital to use innovative methods (Riedel *et al*., 2013; Sriwathsan *et al.*, 2019; Yeo *et al.*, 2020; Sharkey *et al*., 2021). DNA data are indisputably a valuable source for modern biodiversity research (Tautz *et al*., 2003; Hajibabaei *et al*. 2007). There are two principal sources of 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; Sriwathsan *et al.*, 2019). It is the responsibility of systematic biologists to assemble the natural system, i.e., we need to reliably delimit genus- and tribe-level taxa, to make their ecological and distribution attributes informative. Then, 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; Sriwathsan *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 >1,500 recognised species, mostly found in the Old-World tropics (Fig. 1A), and their classification is

74 complicated by the complex taxonomic history (Bocak *et al.*, 2020). The phenetic plasticity
75 of Metriorrhynchini is relatively high (Fig. 1B–D), but many distant species resemble each
76 other due to convergent selection in Mullerian rings (Bocek *et al.*, 2019; Motyka *et al.*,
77 2020, 2021). Therefore, unrelated taxa were often assumed to be closely related due to
78 misleading morphological similarities. Although there are over 40 genera in the tribe,
79 three-quarters of the species have been described in five ambiguously defined genera
80 (*Xylobanus*, *Cautires*, *Trichalus*, *Metriorrhynchus*, and *Cladophorus*). Sometimes a single
81 genus contains species from different subtribes (Bocak *et al.*, 2020). In this respect, the
82 Metriorrhynchini is a typical species-rich tropical insect group without well-founded
83 classification and the paucity and inaccuracy of available data (Letsch *et al.*, 2020). As a
84 result, unlike vertebrates, these poorly known insect groups have not been considered for
85 use in large-scale, integrative projects and data metanalyses (Myers *et al*., 2000; Holt *et*
86 *al.*, 2013) and have contributed little to our understanding of global biodiversity patterns.
87 The principal objective of this study is to demonstrate how biodiversity information for
88 a hyperdiverse tropical group can be rapidly expanded via targeted field research and
89 large-scale sequencing. Our investigation comprised four distinct steps, aiming for a DNA-
90 based evaluation of diversity and evolution. First, we assembled material from several
91 hundred localities on three continents (Fig. 1, Tab. 1). Second, as hyperdiverse groups
92 are difficult to tackle and the current classification is unreliable, we attempted to
93 compartmentalise diversity using phylogenomics. We then produced a tree, using all
94 available data, to estimate species limits, intraspecific genetic variability, and species
95 ranges. Finally, the tree was pruned and used to estimate shallow phylogenetic
96 relationships, total and regional α-diversity, and endemicity, and to define generic ranges
97 and continental-scale range shifts. Our information and phylogenetic hypotheses can be a
98 resource for higher-level phylogenetics, population genetics, phylogeographic studies,
99 and biodiversity estimation. At the same time, we want to show how limited our
100 taxonomical knowledge is and how this lack is hindering biodiversity research and
101 management (Thompson *et al.*, 2018).
102
103

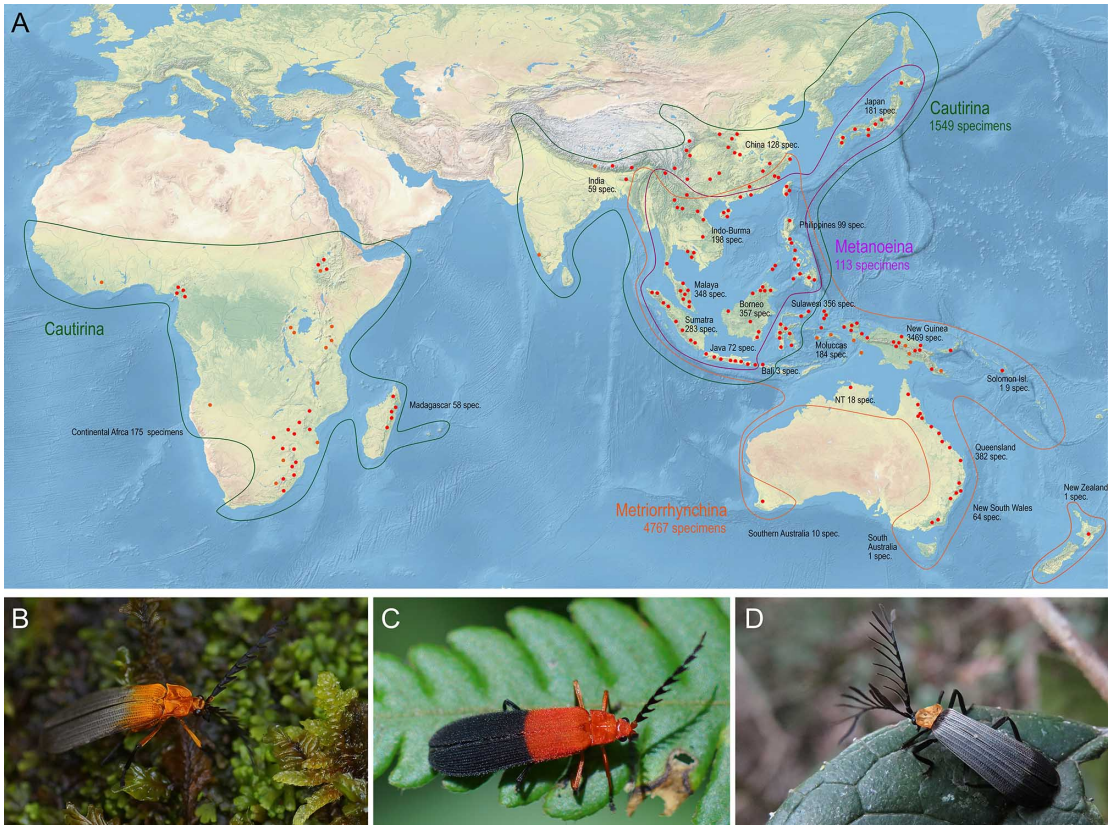


Figure 1. A – Distribution of Metriorrhynchini with major sampled localities designated by red dots. The numbers of analysed specimens from individual regions are shown for regions and subtribes. B–D – General appearance of Metriorrhynchini.

Results

Sampling of the Metriorrhynchini range

In total, we monitored almost 800 localities, 696 of them with occurrences of the Metriorrhynchini (Tabs. 1, S1). 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.

Table 1. The numbers of sampled localities per region.

Area	Localities	Area	Localities
Australian region	347	Palearctic region	79
Australia	118	China	51
New Zealand	1	Japan	28
New Guinea	175	Russian Far East	0
New Britain & Ireland	0		
Solomon Isl.	4		
Moluccas	15	Oriental region	206
Sulawesi	34	Himalayas	0
		S India & Ceylon	3
Afrotropical Region	64	E. India & Burma	12
West Africa	1	E. Indo-Burma	44
Guinean Gulf	9	Malay Peninsula	57
Congo Basin	2	Sumatra	23
Sahel	0	Borneo	19
Ethiopia & Somalia	6	Java & Bali	15
East African Highl.	10	Philippines	33
South Africa	25		
Madagascar	11	Total	696

Assembly of the phylogenomic tree

The phylogenomic dataset contained 35 Metriorrhynchini terminals, seven outgroups, and ~4,200 orthologs (1.9–5.7 × 10^6 aligned positions; Tab. S5). The tree shown in Fig. 2A was produced using maximum likelihood (ML) analyses, whereas the coalescent method produced the topology shown in Fig. 2B; additional trees are shown in Figs. S1–S8. For details on the data sets' characteristics see Figs. S9–S12. Phylogenomic analyses resolved three subtribes (Metanoeina (Metriorrhynchina, Cautirina)), and five clades were regularly recovered within the Metriorrhynchina, i.e., the procautirines, leptotrichalines, trichalines, porrostomines, and cladophorines. Different settings (see Methods) produced slightly different topologies and shifted the positions of the leptotrichalines and procautirines (Fig. 3D). However, the monophyly of major subclades was not affected. The FcLM analysis favoured a deeper position for the leptotrichaline clade (61.2%; Figs. 2C, S13). The position of the remaining terminals was stable across all analyses.

Constrained mitogenomics

The mtDNA database contained >11,500 mtDNA fragments (5,935 *cox1*, 2,381 *rrnL*, and 3,205 *nad5*) representing 6,429 terminals (2,930 aligned positions). Using these data, we inferred additional trees using the constrained positions of 35 terminals whose relationships were determined through phylogenomic analyses, and the free positions of the other ~6,400 terminals (Fig. S14). The units based on uncorrected pairwise distances represent molecular operational taxonomic units (mOTUs), considered to be putative species, or 'species' for short. We identified 37 mOTUs in the Metanoeina clade and 456

mOTUs in Cautirina. The major Metriorrhynchina clade (1,763 mOTUs) included procautirines, leptotrichalines, trichalines, porrostomines, and cladophorines. In addition, we identified several deeply rooted lineages, the kassemiines, and another five small clades (89 mOTUs in total; Fig. S15), each of which comprised a limited number of species. As phylogenomic data for these terminals are still lacking, their terminal positions were determined based only on mtDNA data. The number of mOTUs does not include ~50 mOTUs for which *cox1* was unavailable.

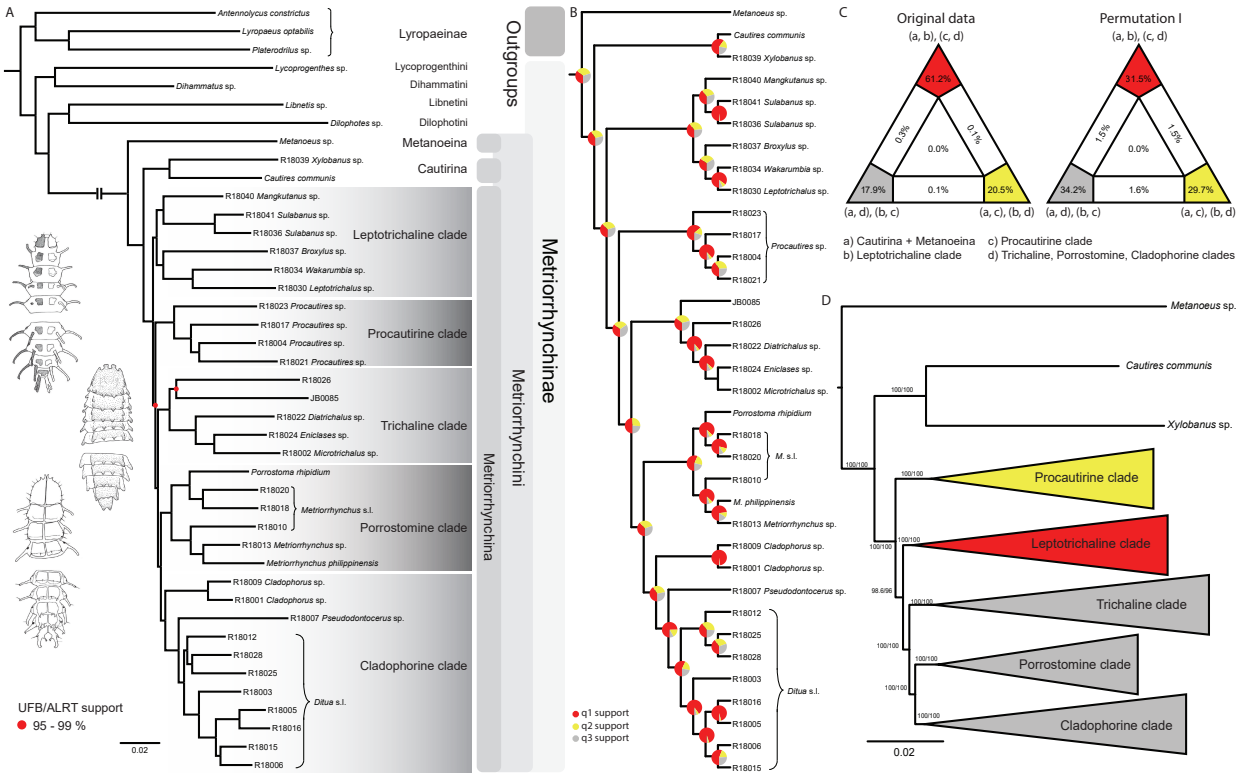


Figure 2. A – Phylogenetic relationships of Metriorrhynchinae based on the ML analyses of the concatenated amino-acid sequence data of supermatrix F-1490-AA-Bacoca-decisive. Unmarked branches are supported by 100/100 UFB/alrt; red circles depict lower phylogenetic branch support. B – Phylogenetic relationships 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.

Pruned mitogenomic tree with and without constraints

The dataset was subsequently pruned to a single terminal per mOTU (see below) and was analysed both with and without topological constraints (Figs. 3A, S15, S16). Repeated runs with different starting seeds identified terminals with unstable positions (Figs. 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; Tab. S7).

Tree congruence

The degree of incongruence between selected topologies is shown in Fig. 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 (Fig. 3D, E, S16). The same dataset, when analysed 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 (Fig. 3C, S15). The composition of the constituent clades is shown in Tab. S8; individual clades are characterised in the Supplementary Text.

Alpha-diversity

To investigate the total and regional α-diversity of the Metriorrhynchini, we analysed a dataset comprising 5,935 of the 6,429 terminals for which the *cox1* mtDNA fragment was available (Fig. 3A; Tab. S3). We identified 2,345 mOTUs using a 2% distance threshold. 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 threshold. For the Metriorrhynchini, we identified 1,848 and 2,356 mOTUs using thresholds of 5% and 2%, respectively (Fig. S17).

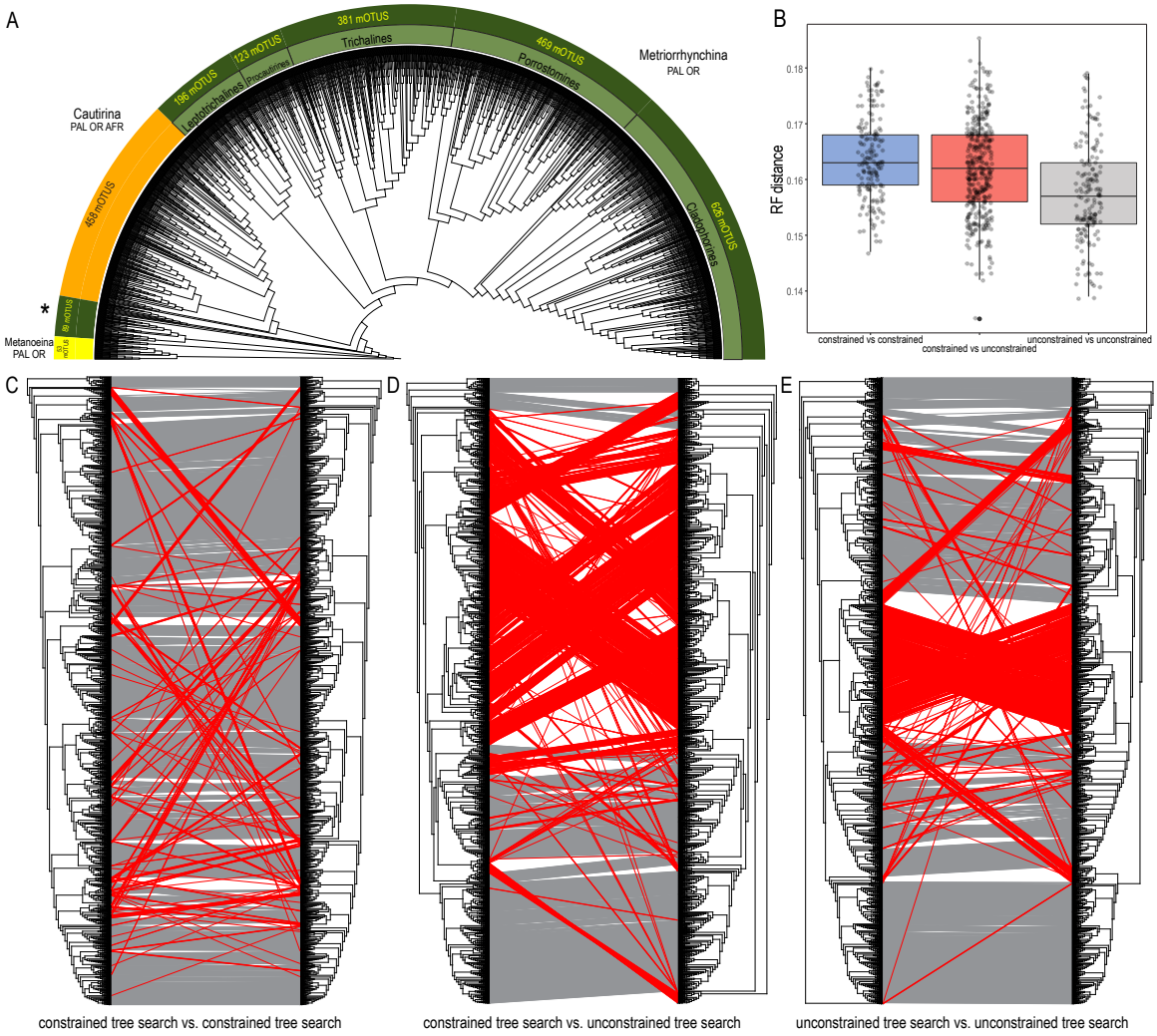


Figure 3. A – Relationships of 2,345 Metriorrhynchini species recovered by the constrained analysis of the pruned dataset (The full resolution tree is shown in Fig. S15 along with a tree recovered from the analysis of a complete dataset of 6,429 terminals in Fig. S14), asterisk designates a grade of 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.

Following an extensive literature review, we updated species lists for the Cautirina (641 spp.), Metanoeina (38 spp.), and Metriorrhynchina (895 spp.; Bocak *et al.*, 2020). By analysing DNA data, we identified 456 spp. of Cautirina, 37 spp. of Metanoeina, and 1,852 spp. of Metriorrhynchina. The numbers of species per subregion, along with the estimated ratios between formally described and estimated α-diversity, are shown in Tab.

1. Even using a threshold of 5%, the number of putative species surpasses the number of species reported in the literature.

We observed very high species turnover, and no species has been recorded in two landmasses separated by a deep-sea (>200 m). Similarly, the faunas of Sulawesi and of the surrounding large islands do not overlap. Only thirteen species were distributed across two landmasses separated by an inundated shelf (sea depth <100 m). Eleven species were distributed in two or more islands of the Great Sundas and two species were found in both New Guinea and Australia. The centres of α-diversity of the Metriorrhynchini are New Guinea (1,434 spp.) and the seasonally to perennially humid areas of the Sundaland (261 spp.). The results suggest substantial modifications to the generic limits and ranges for numerous taxa that had been previously delimited (Fig. S18).

Table 2. The number of described species and identified mOTUs (molecular operational taxonomic units at 2% difference) per region. * – some species are shared by two regions.

Region	Metriorrhynchina	Cautirina	Metanoeina	Metriorrhynchini	Seq./Descr.
Australian region	**639/1608**			**639/1608**	**2.52**
Australia	196/167*			196/167	0.85
New Guinea	423/1434*			423/1434	3.39
Solomon	21*/9			21/9	0.43
Wallacea	**162/174**	**14/10**		**176/184**	**1.05**
Philippines	**51/18**	**45/12**	**8/3**	**104/33**	**0.32**
Continental Asia	**43/52**	**331/330**	**30/34**	**404/416**	**1.03**
Sundaland	36/44	201/184	24/19	261/247	0.95
Indo-Burma	6/7	62/52	3/4	74/63	0.85
China, Japan	1/1	53/75	1/11	55/87	1.58
India		35/19	2/0	37/19	0.51
Afrotropical region		**231/104**		**231/104**	**0.46**
Africa		178/74		178/74	0.42
Madagascar		53/30		53/30	0.57
Total	**895/1852**	**641/456**	**38/37**	**1574/2345**	**1.50**

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 (Fig. 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 compartmentalise hyperdiverse Metriorrhynchini into manageable natural units (Fig. 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 (Figs. 3, S18). Furthermore, the analyses of species-rich datasets identify the areas with high α-diversity as one of the critical conservation value parameters (Tab. 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 & Luksenburg, 2014). Therefore, we use DNA-based units as a provisionary descriptor of α-diversity (Hebert *et al*., 2003; Monaghan *et al*., 2009), and subsequently as a source for integrative taxonomy (Figs. S14–S16; Srivathsan *et al*., 2019). The presented large-scale monitoring project provides information on relationships (Figs. 2, 3), genetic divergence (Figs. S14–S16), turnover (Tab. 2), the extent of generic and species ranges (Fig. S14–S16, S18), and on evolutionary phenomena that are usually studied using a few model organisms (Fig. 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 <3,000 specimens, whereas there are ~6,500 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, & Colla, 2020). In such a way we assembled a Metriorrhynchini tissue collection from almost 700 localities in three continents (Tab. 1, Fig. 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 (Fig. 1A), where net-winged beetles have not been systematically studied since the 1930s. Additional data gaps are caused by strict biodiversity research restrictions (Prathapan *et al*., 2018; Laird *et al*., 2020). 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 & 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 (Gutierrez & Helgen, 2013). The current phylogenomic and mitogenomic phylogenetic hypotheses (Figs. 2, 3; S14–S16) supersede the morphology-based topologies (Bocak, 2002). The phylogenomic analysis incorporates a large amount of information, and we favour 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 analysed 35 transcriptomes and low-coverage genomes were sufficient to identify five major Metriorrhynchina clades (with 100–600 putative species each) and also to identify the limits of genera, which can be tested using traditional taxonomic methods (Figs. 2, 3A, S14–S16).

The sampling strategy is critical for building a phylogenomic backbone, and our goal was to cover as many deep lineages as possible. 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-analysed the short-fragment dataset (Tab. S3) using constrained positions for taxa whose relationships had already been recovered through phylogenomics (Figs. 2, 3). A stabilised phylogenomic backbone is critical for inferring a robust topology because the analyses of short mtDNA fragments are sensitive, even to the application of starting seeds, and they often produce topologies incongruent with morphological traits (Figs. 3B, E). Only several small lineages have remained unanchored by genomic data, owing to a lack of properly fixed samples (Figs. S15, S16). For example, four small clades are much more deeply rooted than their morphology suggests (Figs. 3A, S15) and additional data are needed to place them in a phylogenetic context.

Our approach yielded a phylogeny with 6,429 terminals and 2,345 mOTUs, and this provides the basis for the approximation of α-diversity (Figs. 3, S15, S16). Concerning the extent of diversity, phylogenomic and mitochondrial data must be simultaneously analysed to provide a strong foundation for subsequent investigations (Fig. 3C). Phylogenomics cannot deal with thousands of species, and mitogenomic data are insufficient for the construction of robust relationships. The final steps are morphological validation (see Supplementary Text) 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 Linnean classification (Godfray & Knapp, 2004).

Alpha-diversity: literature data and reality

Here, we deal with a tropical beetle tribe with >1,500 described species, and our results indicate that regionally up to 3.4 times that number of species remain undescribed (Tab. 2). When analysing *cox1* mtDNA, we identified 2,345 mOTUs based on an arbitrary distance threshold of 2 % uncorrected pairwise distance (Hebert *et al.*, 2003). The application of the threshold is a compromise between estimation accuracy, speed, and sequencing costs, taking into account the feasibility of inventorying a hyperdiverse group within a limited timeframe (Dupuis *et al.*, 2012; Eberle *et al.*, 2020). We had previously explored a subset of Metriorrhynchini to estimate the congruence between mtDNA and morphology, and between nextRAD- and mtDNA-based species limits. We found similar numbers of species regardless of the approach that was applied (e.g., Bocek *et al.*, 2019, Jiruskova et al. 2019). Additionally, the slope representing the relationship between the number of mOTUs and distance thresholds was gradual (Fig. S17) due to a high number of genetically distant, indisputably distinct lineages (Figs. S15, S16). Therefore, we

assume that our estimates of α-diversity are realistic, although future taxonomic revisions are needed for validation.

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 (626 mOTUs, *Ditua* historically has included 2 spp. now ~300 spp.). The porrostomine clade is the next diverse group (346 spp.) and contains the speciose *Porrostoma* (156 spp.) 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 (Figs. S15, S16).

The numbers of mOTUs must be interpreted in the context of the sampling activity in each region. We identified only 104 mOTUs from the Afrotropical region, mainly due to the limited number of collecting trips (5 person months; 64 localities) and the inaccessibility of some areas. Despite intensive field research (4 person months, 33 localities), we collected from the Philippines less than one third of the species described. 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 (16 person months, 114 localities) was approximately equal to the number of sequenced mOTUs. Many regions remain unsampled and species ranges are small (Jiruskova *et al*., 2019), so this number will 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 3.4 times the number of species reported in the literature (1,434 mOTUs; 7 person months; 175 localities). Despite the relatively large number of sampled localities, many areas of New Guinea remain unexplored (Fig. 1A), and additional species were added to the dataset with each batch of sequenced samples from other localities.

We observed a high turnover between regions, and few species had ranges which included landmasses separated by shallow seas (2 spp. Queensland / New Guinea, 11 spp. Sundaland islands; Fig. S14). Poorly dispersing lycids generally have very small ranges, except for the few genera that visit flowers and fly in open areas (Kusy *et al.*, 2021). A similar small-scale turnover has recently been reported along altitudinal gradients (Bocek *et al.*, 2019; Motyka *et al.*, 2020, 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 1,574 formally described species (Tab. 1, Bocak et al. 2020; Supplementary Text and References). Although the recent 24% increase in described diversity appears substantial, the distance-based analysis indicates the presence of 2,345 mOTUs (Fig. S17). 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 1,000–1,500 undescribed species. At the current rate, formal morphological descriptions of an additional 1,000+ species would take 100 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 (Tab. 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 norther 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 Cautirina 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. 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 *et al.*, 2015). Most recovered overseas dispersal events are limited to distances of less than 100 km and are commonly accompanied by speciation (Fig. S15, S16). The high proportion of erroneous placement of many taxa (Fig. S18; 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 Fig. S18 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 & 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, Fig. 4).

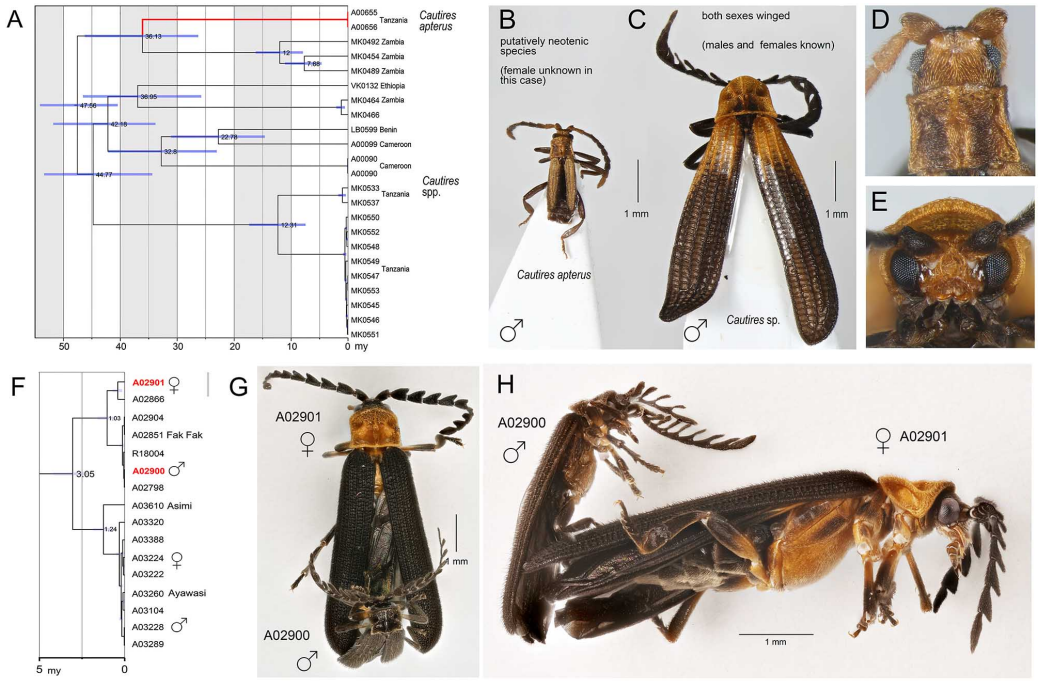


Figure 4. Identification of sexual dimorphism by large scale biodiversity inventory. A – Relationships of lineages with modified ontogeny, the dated tree; B, D – general appearance and head of Cautires apterus, a putative neotenic species; C, E – ditto of the close relative with both sexes winged. Mimetic sexual dimorphism identified during diversity survey. F – the dated tree, red coloured 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 conspecifity.

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 (Fig. 4). 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, 2020, 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 be innovative. We conducted a worldwide sampling in ~700 localities, analysed transcriptomes, genomes, and mitochondrial markers (>2,300 species), and validated our results with morphology. We achieved substantial progress with respect to the development of a Metriorrhynchini tree of life (Chesters, 2017; Linard *et al.*, 2018). The voucher-based DNA entries established a framework for classifying samples from other studies, such as environmental sequencing (Linard *et al*., 2016; Andujar *et al*., 2015; Arribas *et al*., 2016). Despite limited time and funding, we identified >2,300 mOTUs which indicate that there are at least twice more species than the number previously reported in the literature. This means that, at a conservative estimate, 1,000–1,500 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 accelerated inventory shows that the literature records of tropical beetles cannot be used for biodiversity conservation and metanalyses without critical revision. We suggest that if ~20 person months of focused field research 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.

Material and methods

Field research

The analysed individuals had been accumulated by numerous expeditions to various regions of the Metriorrhynchini range (Fig. 1A, Tab. S1). 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 (Fig. 2). About 100 tissue samples were fixed and thirty-five of them were used for sequencing (Tab. S2). 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.

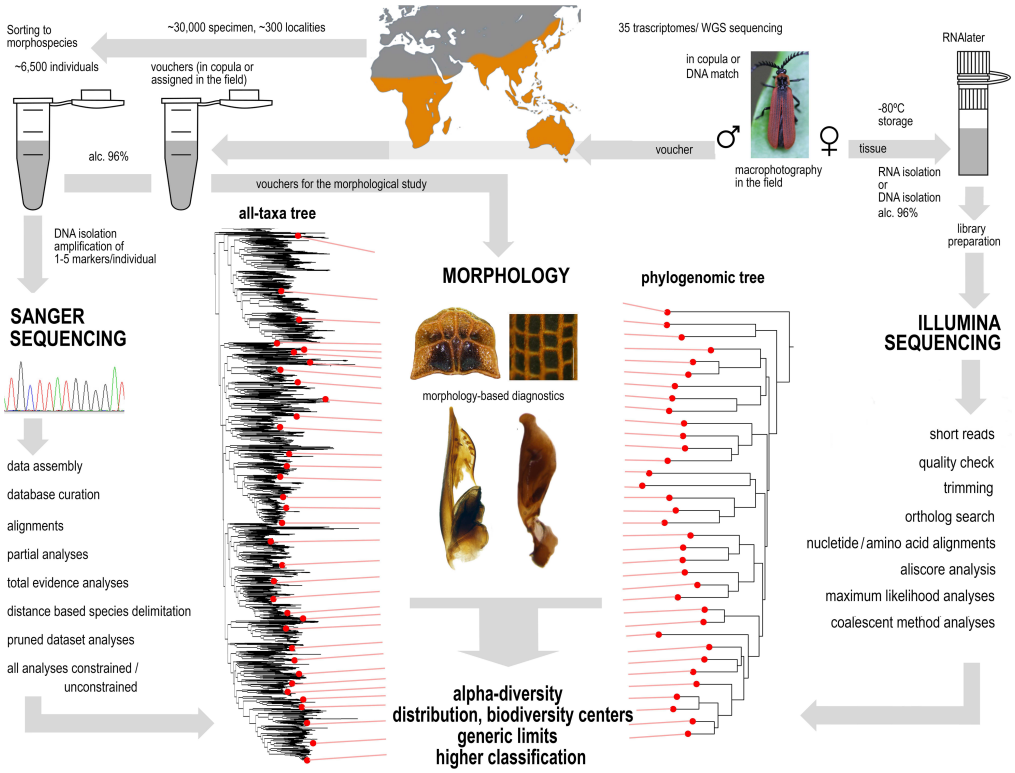
Almost 7,000 samples from 696 localities (Tab. 1) were included in the sequencing program to obtain short mtDNA fragments. In total, 6,429 yielding at least a single fragment were included in the analysis (Tab. S3). The analysed data set contained some previously published sequences (e.g., Sklenarova *et al*., 2013; Bocek & 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. 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

533 *et al.*, 2012), with k-mer sizes of 21, 33, 55, 77, and 99. Obtained contigs were used to
534 train Augustus (Stanke & Waack, 2003) for species-specific gene models with BUSCO.
535 Predicted species-specific gene models were then used for *ab initio* gene predictions in
536 Augustus and predicted protein-coding sequences were used for subsequent analyses.
537 Outgroup taxa were reported in previous studies (Kusy *et al.*, 2019; McKenna *et al.*,
538 2019).
539



540
541 Figure 5. A sequence of applied methods from sampling to hypotheses.
542
543 The ortholog set was collated by searching the OrthoDB 9.1 (Zdobnov *et al.*, 2016) for
544 single copy orthologs in six beetle genomes (Tab. S4). We used Orthograph v.0.6.3
545 (Petersen *et al.*, 2017) with default settings to search in our assemblies for the presence
546 of specified single copy orthologs. From the recovered 4,193 orthologs, terminal stop
547 codons were removed, and internal stop codons at the translational and nucleotide levels
548 were masked. The amino acid sequences were aligned using MAFFT v.7.471 with the L-
549 INS-i algorithm (Katoh & Standley, 2013). The alignments from each ortholog group were
550 then checked for the presence of outliers. To identify random or ambiguous similarities
551 within amino acid alignments, we used Aliscore v.2.076 with the maximum number of
552 pairwise comparisons –r 10^{27}, option -e. and we masked them using Alicut v.2.3 (Kück *et*
553 *al.*, 2010). Alinuc.pl was then used to apply the Aliscore results to match amino acids to
554 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 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 & Longo, 2014) (Tab. S5 and supplementary methods). The degree of missing data and overall completeness scores (Ca) across all datasets were inspected using AliStat v.1.7 (https://github.com/thomaskf/AliStat).

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 & 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 (https://github.com/ottmi/symtest) 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 analysed 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 modelling 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.*, 2017) and SH-like

592 approximate likelihood ratio test (SH-aLRT) were calculated in IQ-TREE2 (options -bb
593 5000 and -alrt 5000) to assess nodal supports for focal relationships.
594
595 *Coalescent analyses and analyses of the confounding and alternative signal*
596 To account for variation among gene trees owing to incomplete lineage sorting and to
597 account for potential gene tree heterogeneity and discordance (Edwards, 2009), the data
598 were also analysed using the coalescent-based species-tree method. For every single
599 gene partition, we calculated an ML gene tree in IQ-TREE2, with 5000 ultrafast bootstrap
600 replicates (-bb option) and using the same substitution models as predicted by
601 ModelFinder in the above described partitioned analyses. For subsequent coalescent
602 species tree estimation, the Accurate Species Tree Algorithm (ASTRAL-III v.5.7.3; Zhang
603 *et al.*, 2018) was used. To account for very poorly resolved branches on gene trees,
604 branches with ultrafast bootstrap ≤ 10 were collapsed using newick utilities v.1.6 (Junier &
605 Zdobnov, 2010) in every ASTRAL analysis. Local posterior probabilities (Erfan & Mirarab,
606 2016) and quartet frequencies of the internal branches in every species tree were
607 calculated using the parameter '-t=2'. Pie charts representing quartet scores for the given
608 topology and two alternatives were plotted to the resulting species trees in R using
609 https://github.com/sidonieB/scripts/blob/master/plot_Astral_trees_v2.R.
610 Additionally, we studied the effect of potentially confounding signals, like non-random
611 distribution of data coverage and violations of SRH conditions, on our phylogenetic
612 reconstructions with the Four-cluster likelihood mapping (FcLM) approach (Strimmer &
613 von Haeseler, 1997) implemented in IQ-TREE2. Based on the results of our tree
614 reconstructions we tested the hypotheses about the alternative placement of
615 leptotrichaline and procautirine clades.
616
617 *Mitochondrial DNA sequencing and data analysis*
618 Total DNA was extracted from the metathorax with a Wizard SV96 kit (Promega Corp.,
619 Madison, WI). The yield was measured using a NanoDrop-1000 Spectrophotometer
620 (Thermo Fisher Scientific Inc., Waltham, MA). The PCR settings and cycle sequencing
621 conditions were the same as those used by Bocak *et al.* (2008). Three fragments of
622 mitochondrial genome were sequenced: *cox1* + tRNA-*Leu* + *cox2* (~1100 bp), *rrnL* +
623 tRNA-*Leu* + *nad1* (~800 bp), and ~1210 bp of *nad5* and adjacent tRNA-*Phe*, tRNA-*Glu*,
624 and tRNA-*Ser* mtDNA (the mtDNA fragments are further mentioned as *rrnL*, *cox1*, and
625 *nad5*). The PCR products were purified using PCRu96™ Plates (Merck Millipore Inc.,
626 Burlington, MA) and sequenced by an ABI 3130 (Applied Biosystems, Waltham, MA)
627 sequencer using the BigDye® Terminator Cycle Sequencing Kit 1.1 (Applied Biosystems,
628 Waltham, MA). Sequences were edited using Sequencher v.4.9 software (Gene Codes

Corp., Ann Arbor, MI). Altogether 6,476 individuals were analysed including some previously published (Sklenarova *et al*., 2013; Bocek & 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 to 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 (Katoh & 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 & Longo, 2014) and analysed under maximum-likelihood criterium in IQ-TREE v.2.1.2 (Minh *et al.*, 2020; Tab. S6). To assess the branch supports values, we used SH-aLRT test with 1,000 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 Metriorrhynchini 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; Tab. S7). Randomly chosen trees were then compared using cophylo script (phytools; Revell, 2012) with argument rotate = TRUE.

Supplementary Materials: The supplementary files are available online here.

Data Depositories. The mitogenomic dataset and all the analysed supermatrices are deposited in the Mendeley Data repository DOI: 10.17632/ntgg6k4fjx.1.

Author Contributions: Conceptualization, L.B., DK, MM; formal analyses, M.M. (Sanger data), D.K. (phylogenomic data); data production and curation, L.B., M. B. and R.B.; writing, original draft preparation, L.B., M.M., D.K.; writing—review and editing, L.B., M.M., D.K., M.B., R.B.; visualization, L.B., M.M., D.K.; funding acquisition, L.B., M.M., D.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The Czech Science Foundation, grant number 18-14942S.

Acknowledgements

Several colleagues provided samples from their respective home countries or their own field research. The field research was enabled by the permits from the Government of Papua New Guinea and 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 research before the start of the project was funded by the senior author and Palacky University is acknowledged for granting the leave. L. Harmackova advised with some analyses in R.

Conflicts of Interest: The authors declare no conflict of interest.

References

Ababneh, F., Jermiin, L. S., Ma, Ch. & Robinson, J. (2006). Matched-pairs tests of homogeneity with applications to homologous nucleotide sequences. *Bioinformatics*, *22*, 1225–1231. https://doi.org/10.1093/bioinformatics/btl064

Andújar, C., Arribas, P., Ruzicka, F., Crampton-Platt, A., Timmermans, M. J. T. N. & Vogler, A. P. (2015). Phylogenetic community ecology of soil biodiversity using mitochondrial metagenomics. *Molecular Ecology*, *24*, 3603–3617. https://doi.org/10.1111/mec.13195

Arribas, P., Andújar, C., Hopkins, K., Shepherd, M. & Vogler, A. P. (2016). Metabarcoding and mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil. *Methods in Ecology & Evolution, 7*, 1071–1081. https://doi.org/10.1111/2041-210X.12557

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

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., … Pevzner P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*, 455–477. https://doi.org/10.1089/cmb.2012.0021

Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, *19*, 134–143. https://doi.org/10.1111/j.1466-8238.2009.00490.x

Basset, Y. & Lamarre, G. P. A. (2019). Toward a world that values insects. *Science*, 364 (6447), 1230–1231. https://doi.org/10.1126/science.aaw7071

Bininda-Emonds, O. R. (2005). transAlign: using amino acids to facilitate the multiple alignment of protein-coding DNA sequences. *BMC Bioinformatics*, *6*, 156. https://doi.org/10.1186/1471-2105-6-156

Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., … Abebe, E. (2005). Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society B – Biological Sciences*, *360*, 1935–1943. https://doi.org/10.1098/rstb.2005.1725

Blom, M. P. K. (2021). Opportunities and challenges for high-quality biodiversity tissue archives in the age of long-read sequencing. *Molecular Ecology*. https://doi.org/10.1111/mec.15909.

Bocak, L. (2002). Revision and phylogenetic analysis of Metriorrhynchinae. *European Journal of Entomology*, *99*, 315–351. https://doi.org/10.14411/eje.2002.043

Bocak, L., Bocakova, M., Hunt, T. & Vogler, A. P. (2008). Multiple ancient origins of neoteny in Lycidae (Coleoptera): consequences for ecology and macroevolution. *Proceedings of the Royal Society – Biological Sciences*, *275*, 2015–2023. https://doi.org/10.1098/rspb.2008.0476

Bocak, L., Kundrata, R., Andújar, C. & Vogler A. P. (2016). The discovery of Iberobaeniidae (Coleoptera: Elateroidea): a new family of beetles from Spain, with immatures detected by environmental DNA sequencing. *Proceedings of the Royal Society of London B – Biological Sciences*, *283*, 20152350. https://doi.org/10.1098/rspb.2015.2350.

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*, 710. https://doi.org/10.3390/insects11100710

Bocek, M. & Bocak, L. (2019). The origins and dispersal history of the trichaline net-winged beetles in South East Asia, Wallacea, New Guinea and Australia. *Zoological journal of the Linnean Society*, *185*, 1079–1094. 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. https://doi.org/10.1186/s12983-019-0335-8

Brooks, T. M., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., Rylands, A. B., … Hilton-Taylor, C. (2002). Habitat loss and extinction in the hotspots of

738 biodiversity. *Conservation Biology*, *16*, 909–923. https://doi.org/10.1046/j.1523-
739 1739.2002.00530.x
740 Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M. & Palmer, T. M.
741 (2015). Accelerated modern human – induced species losses: entering the sixth
742 mass extinction. *Science Advances*, *1*, e1400253.
743 Chen, S., Zhou, Y., Chen, Y. & Gu, J. (2018). Fastp: an ultra-fast all-in-one FASTQ
744 preprocessor. *Bioinformatics*, *34*, i884–i890.
745 https://doi.org/10.1093/bioinformatics/bty560
746 Chernomor, O., von Haeseler, A. & Minh, B. Q. (2016). Terrace aware data structure for
747 phylogenomic inference from supermatrices. *Systematic Biology*, *65*, 997–1008.
748 https://doi.org/10.1093/sysbio/syw037
749 Chesters, D. (2017). Construction of a Species-Level Tree of Life for the Insects and
750 Utility in Taxonomic Profiling. *Systematic Biology*, *66*, 426–439
751 https://10.1093/sysbio/syw099
752 Dupuis, J. R., Roe, A. D. & Sperling, F. A. H. (2012). Multi-locus species delimitation in
753 closely related animals and fungi: one marker is not enough. *Molecular Ecology*,
754 *21*, 4422–4436. https://doi.org/10.1111/j.1365-294X.2012.05642.x
755 Eberle, J., Ahrens, A., Mayer, C., Niehuis, O. & Misof, B. (2020). Plea for Standardized
756 Nuclear Markers in Metazoan DNA Taxonomy. *Trends in Ecology & Evolution*, *35*,
757 336–345. https://doi.org/10.1016/j.tree.2019.12.003
758 Edwards, S. V. (2009). Is a new and general theory of molecular systematics emerging?
759 *Evolution*, *63*, 1–19. https://doi.org/10.1111/j.1558-5646.2008.00549.x
760 Erfan, S. & Mirarab, S. (2016). Fast coalescent-based computation of local branch
761 support from quartet frequencies. *Molecular Biology & Evolution*, *33*, 1654–1668.
762 https://doi.org/10.1093/molbev/msw079
763 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. (2012) CD-HIT: accelerated for clustering the next
764 generation sequencing data. *Bioinformatics*, *28*, 3150–3152.
765 https://doi.org10.1093/bioinformatics/bts565.
766 Gauthier, J., Pajkovic, M., Neuenschwander, S., Kaila, L., Schmid, S., Orlando, L. &
767 Alvarez, N. (2020). Museomics identifies genetic erosion in two butterfly species
768 across the 20th century in Finland. *Molecular Ecology Resources*, *20*, 1191–1205.
769 https://doi.org/10.1111/1755-0998.13167
770 Godfray H. C. J. & Knapp S. (2004). Taxonomy for the twenty- first century. Introduction.
771 *Philosophical Transactions of the Royal Society London B*, *359*, 559–569.
772 https://doi.org/10.1098/rstb.2003.1457
773 Gutierrez E. E. & Helgen K. M. (2013). Outdated taxonomy blocks conservation. *Nature*,
774 *495*, 314. https://doi.org/10.1038/495314e

775 Hall, R. (2011). Australia-SE Asia collision: plate tectonics and crustal flow. In: Hall, R.,
776 Cottam, M. A. & Wilson, M. E. J., editors. The SE Asian Gateway: History and
777 Tectonics of the Australia-Asia Collision. *Geological Society Special Publications*,
778 *355*, 75–109.
779 Hallmann, C. A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., ... Goulson,
780 D. (2017). More than 75 percent decline over 27 years in total flying insect
781 biomass in protected areas. *PLoS One, 12*, e0185809.
782 Hajibabaei, M., Singer, G. A. C., Hebert, P. D. N. & Hickey, D. A. (2007). DNA barcoding:
783 how it complements taxonomy, molecular phylogenetics and population genetics.
784 *Trends in Genetics*, *2*3, 167-172. https://doi.org/10.1016/j.tig.2007.02.001
785 Hebert, P. D. N. Cywinska, A., Ball, S. L. & deWaard J. R. (2003). Biological
786 identifications through DNA barcodes. *Proceedings of the Royal Society B –*
787 *Biological Sciences*, *270*, 313–321. https://doi.org/10.1098/rspb.2002.2218
788 Hoang, D. T., Chernomor, O., Haeseler A. von, Minh B. Q. & Vinh L. S. (2017). UFBoot2:
789 Improving the ultrafast bootstrap approximation. *Molecular Biology & Evolution*,
790 *35*, 518–522. https://doi.org/10.1093/molbev/msx281
791 Holt, B. G., Lessard, J.-P., Borregaard, M. K., Fritz, S. A., Araújo, M. B., … Rahbek, C.
792 (2013). An Update of Wallace's Zoogeographic Regions of the World. *Science*,
793 *339*, 74–78. https://doi.org/10.1126/science.1228282
794 Jaskula, R., Kolanowska, M., Michalski, M. & Schwerk, A. (2021). From Phenology and
795 Habitat Preferences to Climate Change: Importance of Citizen Science in Studying
796 Insect Ecology in the Continental Scale with American Red Flat Bark Beetle,
797 Cucujus clavipes, as a Model Species. *Insects*, *12*, 369.
798 https://doi.org/10.3390/insects12040369
799 Jiruskova, A., Motyka, M., Bocek, M. & Bocak, L. (2019). The Malacca Strait separates
800 distinct faunas of poorly-flying *Cautires* net-winged beetles. *PeerJ*, *7*, e6511.
801 https://doi.org/10.7717/peerj.6511
802 Jordal, B. H. (2015). Molecular phylogeny and biogeography of the weevil subfamily
803 Platypodinae reveals evolutionarily conserved range patterns. *Molecular*
804 *Phylogenetics & Evolution*, *92*, 294–307.
805 https://doi.org/10.1016/j.ympev.2015.05.028
806 Junier, T. & Zdobnov, E. M. (2010). The Newick utilities: high-throughput phylogenetic
807 tree processing in the UNIX shell. *Bioinformatics*, *26*, 1669–1670.
808 https://doi.org/10.1093/bioinformatics/btq243
809 Katoh, K. & Standley, D. M. (2013). MAFFT multiple sequence alignment software version
810 7: improvements in performance and usability. *Molecular Biology & Evolution*, *30*,
811 772–780. https://doi.org/10.1093/molbev/mst010

812 Klink, R. van, Bowler, D. E., Gongalsky, K. B., Swenge, A. B., Gentile, A., Chase, J. M.
813 (2020). Meta-analysis reveals declines in terrestrial but increases in freshwater
814 insect abundances. *Science*, *368*, 417–420.
815 https://doi.org/10.1126/science.aax9931
816 Krehenwinkel, H., Pomerantz, A., Henderson, J. B.; Kennedy, S. R., Lim, J. Y., Swamy,
817 V., Shoobridge, J. D., Graham, N., Patel, N. H., Gillespie, R. G. & Prost, S. (2019).
818 Nanopore sequencing of long ribosomal DNA amplicons enables portable and
819 simple biodiversity assessments with high phylogenetic resolution across broad
820 taxonomic scale. *Gigascience*, *8*, giz006.
821 https://doi.org/10.1093/gigascience/giz006
822 Kück, P. & Longo, G. C. (2014). FASconCAT-G: extensive functions for multiple
823 sequence alignment preparations concerning phylogenetic studies. *Frontiers in*
824 *Zoology*, *11*, 81. https://doi.org/10.1186/s12983-014-0081-x
825 Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B.M., … Misof, B.
826 (2010). Parametric and non-parametric masking of randomness in sequence
827 alignments can be improved and leads to better resolved trees. *Frontiers in*
828 *Zoology*, *7*, 10. https://doi.org/10.1186/1742-9994-7-10
829 Kück, P. & Struck, T. H. (2014). BaCoCa – a heuristic software tool for the parallel
830 assessment of sequence biases in hundreds of gene and taxon partitions.
831 *Molecular Phylogenetics & Evolution*, *70*, 94–98.
832 https://doi.org/10.1016/j.ympev.2013.09.011
833 Kunte, K. (2008). Mimetic butterflies support Wallace's model of sexual dimorphism.
834 *Proceedings of the Royal Society B – Biological Sciences*, *275*, 1617–1624.
835 https://doi.org/10.1098/rspb.2008.0171
836 Kusy, D., Motyka, M., Bocek, M., Masek, M. & Bocak L. (2019). Phylogenomic analysis
837 resolves the relationships among net-winged beetles (Coleoptera: Lycidae) and
838 reveals the parallel evolution of morphological traits. *Systematic Entomology*, *44*,
839 911–925. https://doi.org/10.1111/syen.12363
840 Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. (2017).
841 PartitionFinder 2: new methods for selecting partitioned models of evolution for
842 molecular and morphological phylogenetic analyses. *Molecular Biology &*
843 *Evolution*, *34*, 772–773. https://doi.org/10.1093/molbev/msw260
844 Letsch, H., Balke, M., Toussaint, E. F. A., Narakusumo, R. P., Fiedler, K. & Riedel, A.
845 (2020). Transgressing Wallace's Line brings hyperdiverse weevils down to earth.
846 *Ecography*, *43*, 1329–1340. https://doi.org/10.1111/ecog.05128

847 Linard, B., Arribas, P., Andújar, C., Crampton-Platt, A., Vogler, A. P. (2016). Lessons from
848 genome skimming of arthropod-preserving ethanol. *Molecular Ecology Resources*,
849 *16*, 1365–1377. https://doi.org/10.1111/1755-0998.12539
850 MacPhail, V. J. & Colla, S. R. (2020) Power of the people: A review of citizen science
851 programs for conservation. *Biological Conservation*, *249*, 108739.
852 https://doi.org/10.1016/j.biocon.2020.108739
853 Mastretta-Yanes, A. Xue, AT., Moreno-Letelier, A., Jorgensen, T. H., Alvarez, N., Pinero,
854 D. & Emerson, B. C. (2018). Long-term in situ persistence of biodiversity in tropical
855 sky islands revealed by landscape genomics. *Molecular Ecology, 27*, 432–448.
856 https://doi.org/10.1111/mec.14461
857 McMahon, D. P. & Hayward, A. (2016). Why grow up? A perspective on insect strategies
858 to avoid metamorphosis. *Ecological Entomology*, *41*, 505–515.
859 https://doi.org/10.1111/een.12313
860 McKenna, D. D., Shin, S., Ahrens, D., Balke, M., Beza-Beza, C., … Beutel, R. G. (2019).
861 The evolution and genomic basis of beetle diversity. *Proceedings of the National*
862 *Academy of Sciences of the United States of America*, *116*, 24729–24737.
863 https://doi.org/10.1073/pnas.1909655116
864 Merckx, V. S. F. T *et al*. (2015). Evolution of endemism on a young tropical mountain.
865 *Science*, *524*, 347–350. https://doi.org/10.1038/nature1494
866 Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., … Lanfear,
867 R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic
868 inference in the genomic era. *Molecular Biology & Evolution*, *37*, 1530–1534.
869 https://doi.org/10.1093/molbev/msaa015
870 Misof, B., Meyer, B., von Reumont, B. M., Kück, P., Misof, K. & Meusemann, K. (2013).
871 Selecting informative subsets of sparse supermatrices increases the chance to
872 find correct trees. *BMC Bioinformatics*, *14*, 348. https://doi.org/10.1186/1471-
873 2105-14-348
874 Monaghan, M. T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D. J., Lees, D. C.,
875 Ranaivosolo, R., Eggleton, P., Barraclough, T. G. & Vogler, A. P. (2009).
876 Accelerated Species Inventory on Madagascar Using Coalescent-Based Models
877 of Species Delineation. *Systematic Biology*, *58*, 298–311.
878 https://doi.org/10.1093/sysbio/syp027
879 Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B. & Worm, B. (2011). How many
880 species are there on Earth and in the ocean? *PLoS Biology*, *9*, e1001127.
881 https://doi.org/10.1371/journal.pbio.1001127
882 Morrison, W. R., Lohra, J. L., Duchen, P., Wilches R., Trujillo D., … Renner S. S. (2009).
883 The impact of taxonomic change on conservation: Does it kill, can it save, or is it

884 just irrelevant? *Biological Conservation*, *142*, 3201–3206.
885 https://doi.org/10.1016/j.biocon.2009.07.019
886 Motyka, M., Kampova, L. & Bocak, L. (2018). Phylogeny and evolution of Müllerian
887 mimicry in aposematic *Dilophotes*: evidence for advergence and size-constraints
888 in evolution of mimetic sexual dimorphism. *Scientific Reports*, *8*, 3744.
889 https://doi.org/10.1038/s41598-018-22155-6
890 Motyka, M., Bocek, M., Kusy, D., Bocak, L. (2020). Interactions in multi-pattern Mullerian
891 communities support origins of new patterns, false structures, imperfect
892 resemblance and mimetic sexual dimorphism *Scientific Reports*, *10*, 11193.
893 https://doi.org/10.1038/s41598-020-68027-w
894 Motyka, M., Kusy, D., Masek, M., Bocek, M., Li, Y., Bilkova, R., … Bocak, L. (2021).
895 Conspicuousness, phylogenetic structure, and origins of Müllerian mimicry in 4000
896 lycid beetles from all zoogeographic regions. *Scientific Reports*, *11*, 5961.
897 https://doi.org/10.1038/s41598-021-85567-x
898 Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca G. A. B. & Kent J. (2000).
899 Biodiversity hotspots for conservation priorities. *Nature*, *403*, 853–858.
900 https://doi.org/10.1038/35002501
901 Naser-Khdour, S., Minh, B. Q., Zhang, W., Stone, S. A. & Lanfear, R. (2019). The
902 prevalence and impact of model violations in phylogenetic analysis. *Genome*
903 *Biology & Evolution*, *11*, 3341–3352. https://doi.org/10.1093/gbe/evz193
904 Novotny, V., Drozd, P., Miller, S. E., Kulfan, M., Janda, M., Basset, Y., Weiblen, G. D.
905 (2006). Why are there so many species of herbivorous insects in tropical
906 rainforests? *Science, 313*, 1115–1118. https://doi.org/10.1126/science.1129237
907 Petersen, M., Meusemann, K., Donath, A., Dowling, D., Liu, S., … Niehuis, O. (2017).
908 Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters
909 of orthologous genes. *BMC Bioinformatics*, *18*, 111.
910 https://doi.org/10.1186/s12859-017-1529-8
911 Prathapan, K. D. *et al*. (2018). When the cure kills-CBD limits biodiversity research
912 National laws fearing biopiracy squelch taxonomy studies. *Science*, *360*, 1405–
913 1406. https://doi.org/10.1126/science.aat9844
914 Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and
915 other things). *Methods in Ecology & Evolution*, *3*, 217–223.
916 https://doi.org/10.1111/j.2041-210X.2011.00169.x
917 Riedel, A., Sagata, K., Suhardjono, Y. R., Tänzler, R. & Balke, M. (2013). Integrative
918 taxonomy on the fast track-towards more sustainability in biodiversity research.
919 *Frontiers in Zoology*, *10*, 15. https://doi. org/10.1186/1742-9994-10-15

Sangster, G., Luksenburg, J. A. (2014). Declining rates of species described per taxonomist: slowdown of progress or a side-effect of improved quality in taxonomy? *Systematic Biology*, *64*, 144–151. https://doi.org/10.1093/sysbio/syu069

Schliep, K. P. (2011). phangorn: phylogenetic analysis in R. *Bioinformatics*, *27*, 592–593. https://doi.org/10.1093/bioinformatics/btq706

Sharkey, M. J., Janzen, D. H., Hallwachs, W., … Burns J. M. (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. https://doi.org/10.3897/zookeys.1013.55600

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*(6), e67957. https://doi.org/10.1371/journal.pone.0067957

Sodhi, N. S., Koh, L. P., Brook, B. W. & Ng P. K. L. (2004). Southeast Asian biodiversity: an impending disaster. *Trend in Ecology & Evolution*, *19*, 654–660. https://doi.org/10.1016/j.tree.2004.09.006

Srivathsan, A., Hartop, E., Puniamoorthy, J., Lee, W. T., Kutty, S. N., … Meier, R. (2019). Rapid, large-scale species discovery in hyperdiverse taxa using 1D MinION sequencing. *BMC Biology*, *17*, 96. https://doi.org/10.1186/s12915-019-0706-9

Stanke, M. & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*, *19*, 215–225. https://doi.org/10.1093/bioinformatics/btg1080

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

Tautz, D., Arctander, P., Minelli, A., Thomas, R. H., Vogler, A. P. (2003). A plea for DNA taxonomy. *Trends in Ecology & Evolution*, *18*, 70–74. https://doi.org/10.1016/S0169-5347(02)00041-1

Theng, M., Jusoh, W. F. A., Jain, A., Huertas, B., Tan, D. J. X., Tan, H. Z., Kristensen, N. P., Meier, R., Chisholm, R. A. (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. https://doi.org/10.1016/j.biocon.2019.108401

Thomson, S. A., Pyle, R. L., Ahyong, S. T., Alonso-Zarazaga, M., Ammirati, J., … Zhou, H.-Z. (2018). Taxonomy based on science is necessary for global conservation. *PLoS Biology*, *16*, 1–12. https://doi.org/10.1371/journal.pbio.2005075

957 Toussaint, E. F. A., Hall, R., Monaghan, M. T., Sagata, K., Ibalim, S., … Balke, M. (2014).
958 The towering orogeny of New Guinea as a trigger for arthropod megadiversity.
959 *Nature Communications*, *5*, 4001. https://doi.org/10.1038/ncomms5001
960 Vasilikopoulos, A., Balke, M., Beutel, R.G., Donath A., Podsiadlowski, … Misof, B. (2019).
961 Phylogenomics of the superfamily Dytiscoidea (Coleoptera: Adephaga) with an
962 evaluation of phylogenetic conflict and systematic error. *Molecular Phylogenetics*
963 *& Evolution*, *135*, 270–285. 10.1016/j.ympev.2019.02.022
964 Xie, Y., Wu, G., Tang, J., Luo, R.,Patterson, J., ... Wang, J. (2014). Trans: de novo
965 transcriptome assembly with short RNA-Seq reads. *Bioinformatics*, *30*, 1660–
966 1666. https://doi.org/10.1093/bioinformatics/btu077
967 Yeo, D., Srivathsan, A. & Meier, R. (2020). Longer is not always better: optimizing
968 barcode length for large-scale species discovery and identification. *Systematic*
969 *Biology*, *69*, 999–1015. https://doi.org/10.1093/sysbio/syaa014
970 Zdobnov, E. M., Tegenfeldt, F., Kuznetsov, D., Waterhouse, R. M., Simão F. A., ...
971 Kriventseva, E. V. (2016). OrthoDBv9.1: cataloging evolutionary and functional
972 annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs.
973 *Nucleic Acids Research*, *45*, D744–D749. https://doi.org/10.1093/nar/gkw1119
974 Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. (2018). ASTRAL-III: Polynomial time
975 species tree reconstruction from partially resolved gene trees. *BMC*
976 *Bioinformatics*, *19*(S6), 153. https://doi.org/10.1186/s12859-018-2129-y