

Research



Cite this article: Li Y-D, Engel MS, Tihelka E, Cai C. 2023 Phylogenomics of weevils revisited: data curation and modelling compositional heterogeneity. *Biol. Lett.* **19**: 20230307. <https://doi.org/10.1098/rsbl.2023.0307>

Received: 2 July 2023
Accepted: 29 August 2023

Subject Category:
Evolutionary biology

Subject Areas:
evolution

Keywords:
Curculionoidea, Belidae, phylogeny, compositional heterogeneity

Author for correspondence:
Chenyang Cai
e-mail: cycal@nigpas.ac.cn

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.6825533>.

Phylogenomics of weevils revisited: data curation and modelling compositional heterogeneity

Yan-Da Li^{1,2}, Michael S. Engel³, Erik Tihelka² and Chenyang Cai^{1,2}

¹State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences, Nanjing 210008, People's Republic of China

²Bristol Palaeobiology Group, School of Earth Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK

³Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024-5192, USA

id Y-DL, 0000-0002-9439-202X; MSE, 0000-0003-3067-077X; ET, 0000-0002-5048-5355; CC, 0000-0002-9283-8323

Weevils represent one of the most prolific radiations of beetles and the most diverse group of herbivores on land. The phylogeny of weevils (Curculionoidea) has received extensive attention, and a largely satisfactory framework for their interfamilial relationships has been established. However, a recent phylogenomic study of Curculionoidea based on anchored hybrid enrichment (AHE) data yielded an abnormal placement for the family Belidae (strongly supported as sister to Nemomychidae + Anthribidae). Here we reanalyse the genome-scale AHE data for Curculionoidea using various models of molecular evolution and data filtering methods to mitigate anticipated systematic errors and reduce compositional heterogeneity. When analysed with the infinite mixture model CAT-GTR or using appropriately filtered datasets, Belidae are always recovered as sister to the clade (Attelabidae, (Caridae, (Brentidae, Curculionidae))), which is congruent with studies based on morphology and other sources of molecular data. Although the relationships of the 'higher Curculionidae' remain challenging to resolve, we provide a consistent and robust backbone phylogeny of weevils. Our extensive analyses emphasize the significance of data curation and modelling across-site compositional heterogeneity in phylogenomic studies.

1. Introduction

Weevils (superfamily Curculionoidea) represent a hyperdiverse and globally distributed group of phytophagous beetles, with approximately 62 000 described species in 5800 genera [1,2]. Most species of Curculionoidea can be easily recognized by their elongate rostrum, which has been suggested as a key innovation responsible for their success [3–6]. Indeed, the rostrum permitted weevils to access a greater variety of plant tissues than their relatives, particularly when coupled with the later appearance of an elongate scape (mostly in Curculionidae) allowing the antenna to fold posteriorly and for the rostrum to extend deeper into plant tissues while simultaneously preserving some motility from the pedicel to apex. These two characters together allowed weevils to further extend their feeding habits as well as to coopt the rostrum for chewing deep oviposition sites, and with this the higher weevils proliferated in species and dietary breadth. Although there are still some different opinions (e.g. [6–8]), most recent classifications have reached a consensus on the familial division of Curculionoidea (see [1,2]). Clarke *et al.* [9] recognized eight extant curculionoid families, i.e. Cimberididae, Nemomychidae, Anthribidae, Belidae, Attelabidae, Caridae, Brentidae and Curculionidae, plus one extinct family, Mesophyletidae, known exclusively from the mid-Cretaceous Burmese amber. The more plesiomorphic Nemomychidae, Belidae and Caridae are commonly

associated with gymnosperms, whereas the most species-rich Brentidae and Curculionidae are predominantly associated with angiosperms [2]. This niche shift is apparently associated with the radiation of angiosperms, which probably contributed to the success of weevils (e.g. [10–12]).

The phylogeny of Curculionoidea has been investigated extensively based on both morphological and molecular data (e.g. [13–21]). These studies have generally established a congruent framework for the interfamilial relationships of curculionoids. However, in a recent anchored hybrid enrichment (AHE)-based study by Shin *et al.* [21], Belidae were strongly recovered as the sister group of the clade encompassing Nemonychidae + Anthribidae when analysed using amino acid (AA) sequences (their preferred topology). This position of Belidae in Shin *et al.* [21] was unexpected and contradicted previous studies based on either morphological or molecular data sets. In particular, in the most comprehensive phylogenomic studies of beetles based on dozens of nuclear protein-coding genes [22,23] or genome-scale data [12], Belidae were consistently and strongly resolved as sister to the group composed of Attelabidae, Caridae, Brentidae and Curculionidae.

Anchored hybrid enrichment data [24,25], typically represented by hundreds of loci from across the genome, have provided invaluable and novel insights into both deeper and shallower phylogenies of vertebrates [26–29], invertebrates [30–33] and plants [34,35]. AHE-based phylogenomics often yield a congruent tree with the topology inferred from other large-scale data sets (e.g. transcriptomes and ultraconserved elements), when the process of molecular evolution is more adequately modelled [36]. As such, the abnormal phylogenetic position of Belidae recovered in Shin *et al.* [21] could have resulted from model misspecification and/or data curation, since both aspects have been shown as crucial for tree inference in the phylogenomic era [37–40]. In terms of modelling molecular evolution of large-scale data sets, the site-heterogeneous models such as CAT-GTR, provide a consistently better fit to the data and outperform the site-homogeneous substitution models, at the expense of a higher computational cost.

In the present study, we reanalyse the data by Shin *et al.* [21] using various models of molecular evolution and data filtering methods, aiming to build a backbone phylogeny of weevils and clarify the phylogenetic position of Belidae.

2. Material and methods

(a) Data preparation

Our analyses were conducted on four datasets prepared based on the concatenated AA sequences by Shin *et al.* [21]. We did not use the nucleotide (NT) data set because of the presence of compositional biases and apparently saturated third-codon positions [21]. More importantly, it has long been shown that AA-based phylogenetic results are more robust for resolving deep divergences (e.g. [23,41]). Dataset 1 was the same as the original matrix by Shin *et al.* [21]. Datasets 2–4 were trimmed with different methods to remove ambiguously aligned regions and reduce compositional heterogeneity. Dataset 2 was trimmed with BMGE 1.1 [42] using BLOSUM95 similarity matrix and the default entropy score cut-off ($-h$ 0.5). An inspection revealed some misaligned regions unrecovered in the BMGE-trimmed dataset 2. Figure S1 (see electronic supplementary material) provides an example of these misaligned regions, where two obviously discrete sequence modes can be seen: the predominant one starting with FEEVQL, and the less common one starting with XEGYPV. The XEGYPV-type sequences are

present in unrelated lineages. According to NCBI BLAST searches, these two types of sequences are different parts of the same RNA helicase, which are clearly non-orthologous. Likewise, most other misaligned regions are also represented by different parts of the same protein. Therefore, for dataset 3, after trimming by BMGE, these incorrectly aligned regions were further manually removed. Dataset 4 was trimmed with Gblocks 0.91 [43] under a relaxed setting (minimum number of sequences for a flank position: minimum; maximum number of contiguous non-conserved positions: maximum; minimum length of a block: minimum; allowed gap positions: with half), as the default settings have been suggested to be too stringent [42]. The distant outgroup *Aethina* (Nitiduloidea: Nitidulidae) was removed in datasets 2 and 3.

(b) Testing compositional heterogeneity

To assess the compositional heterogeneity across taxa, the normalized relative composition frequency variability (nRCFV) was calculated using RCFV Reader v.1 [44]. To assess the compositional heterogeneity across sites, we tested the model fitness (based on the Bayesian information criterion, BIC) of various site-homogeneous models alongside the mixture models LG4X [45] and LG + C20 + F [46] using IQ-TREE 1.6.2 or 2.1.3 [47,48]. The best fitting site-homogeneous model was determined with ModelFinder implemented in IQ-TREE 2.1.3 [49], where several new clade-specific AA replacement matrices were available [50].

(c) Phylogenetic analyses

The maximum likelihood (ML) phylogenetic trees were inferred with IQ-TREE 1.6.2 or 2.1.3 under both the best fitting site-homogeneous model and the mixture models LG4X, LG + C20 + F, and LG + C40 + F (for dataset 2 only), as well as the posterior mean site frequency (PMSF) approximation of LG + C20 + F, LG + C40 + F and LG + C60 + F [51]. The PMSF profiles were computed based on the guide tree estimated under LG4X (datasets 1, 2, 4) or LG + C20 + F (dataset 3). For each ML analysis, ultrafast bootstrap support values were calculated from 1000 replicates [52]. The gene concordance factor (gCF) and site concordance factor (sCF) were calculated for the LG4X tree based on dataset 1 [53].

The Bayesian phylogenies were inferred under the infinite mixture model CAT-GTR + G4 [54] in PhyloBayes MPI 1.7 [55]. Two independent chains were computed for each Bayesian analysis, with their convergence evaluated using the bpcomp program. Ideally the analyses were stopped when the maximum difference across bipartitions (maxdiff) was at least lower than 0.3 (but see also Results). The trees were visualized using the online tool iTOL 6.8 [56].

The coalescent-based phylogenies were also estimated. Since the sequences of individual genes were not directly supplied by Shin *et al.* [21], we extracted the aligned protein-coding genes (in AA form) from the supermatrix based on the partition file provided, using R 4.1.0 [57] and R packages Biostrings 2.62.0 [58] and seqinr 4.2-30 [59]. The gene trees were estimated under the LG4X model using IQ-TREE 2.1.3. The species tree was then inferred using ASTRAL 5.7.8 [60], first with all 521 gene trees as the input. As some of the gene fragments contain clearly misaligned regions (as discussed above), an additional species tree was estimated based on 364 gene trees (157 trees based on misaligned fragments were excluded). The species trees were visualized using R package ggtree 3.2.1 [61], with the quartet supports mapped using R package ggmame 0.3.3 [62].

3. Results

(a) Compositional heterogeneity in datasets

The normalized RCFV value measures the compositional heterogeneity across taxa. The original dataset displayed

Table 1. Comparison among the differently filtered datasets. BIC, Bayesian information criterion; nRCFV, normalized relative composition frequency variability.

		dataset 1	dataset 2	dataset 3	dataset 4
number of amino acid sites		225 254	25 468	23 329	39 572
data occupancy		22.1%	86.4%	86.6%	78.4%
nRCFV		0.003760	0.000979	0.000935	0.001243
BIC score	LG + R6	3 661 622.351	774 637.579	625 241.560	1 335 347.454
	Q.insect + R6	3 645 847.536	769 033.791	620 442.561	1 327 483.076
	LG4X	3 674 201.090	777 927.453	629 723.247	1 339 488.312
	LG + C20 + F	3 593 307.319	750 270.935	610 507.371	1 294 237.132
	LG + C40 + F	/	747 484.080	/	/

a relatively high nRCFV value, indicating a pronounced compositional heterogeneity (table 1). In the trimmed datasets, the across-taxa compositional heterogeneity was reduced to some extent and consequently, the data occupancy was increased significantly from 22.1% (dataset 1, the original AA dataset) to 86.4% (dataset 2), 86.6% (dataset 3) and 78.4% (dataset 4) (table 1). As shown by the BIC values (table 1), for each of the datasets, the best fitting site-homogeneous model outperformed the LG4X model. However, overall the mixture model LG + C20 + F (and also LG + C40 + F when computed) always had better fitness than the best site-homogeneous model selected by ModelFinder implemented in IQ-TREE. This implies a high level of compositional heterogeneity across sites, and shows the necessity of analysing with the site-heterogeneous model.

(b) Phylogeny of Curculionoidea

The Bayesian analyses under the CAT-GTR model did not fully converge for datasets 1 and 4 (maxdiff = 1). However, the failure to converge was due only to discrepancies between chains about the nodes within the family Curculionidae, and all the nodes related to the interfamilial relationships of Curculionoidea had reached a posterior probability support of 1.00 in these two analyses (electronic supplementary material, figures S8 and S30). According to [63], these results should be acceptable as the discrepancies do not directly affect our nodes of interest.

The relationships among most families were stable across our analyses with different datasets and models, except for the positions of Cimberididae and Belidae (figures 1 and 2). A backbone of ((Nemonychidae, Anthribidae), (Attelabidae, (Caridae, (Brentidae, Curculionidae)))) was consistently recovered, which was accordant with the original results by Shin *et al.* [21]. In most analyses, Cimberididae were resolved as sister to the remaining Curculionoidea (figure 2). In a few analyses, Cimberididae appeared as sister to the chrysomeloid outgroups Orsodacnidae + Chrysomelidae, sister to the remaining Curculionoidea except Nemonychidae + Anthribidae (figure 2), or sister to Nemonychidae + Anthribidae (electronic supplementary material, figure S33), which we suppose was an artefact. The internal relationships of Curculionidae were somewhat variable among analyses and not well supported, suggesting that the present data are insufficient to resolve these issues.

(c) Position of Belidae

Belidae were recovered as sister to Nemonychidae + Anthribidae in most of the results from datasets 1 and 2, except when analysed with the infinite mixture model CAT-GTR (figure 2). However, under the CAT-GTR model, Belidae turned out to be sister to the 'ACBC clade' of (Attelabidae, (Caridae, (Brentidae, Curculionidae))) (figure 2). In all the results based on datasets 3 and 4, even including those based on site-homogeneous models, Belidae were also recovered as sister to the ACBC clade (figure 2). In addition, at least in the results under the CAT-GTR model, the placement of Belidae was always strongly supported (BPP = 0.99–1.00). Our coalescent analyses supported Belidae as sister to the ACBC clade as well, although the quartet support was not significantly higher than the alternative hypotheses (electronic supplementary material, figures S33, S34) [64].

4. Discussion

In the partitioned maximum-likelihood analysis of AA data by Shin *et al.* [21], Belidae were grouped together with Nemonychidae and Anthribidae. This topology conflicted with the NT-based results by themselves and also previous studies [16,20] (see also [65]), where Belidae were often grouped with the remaining Curculionoidea except Nemonychidae and Anthribidae. Shin *et al.* [21] suggested that the position of Belidae in the NT-based results might be an artefact resulting from codon-usage bias. However, the overall low concordance factors of the ML tree suggest that the phylogenetic signal is quite weak (electronic supplementary material, figures S31, S32), and particularly, the sCF supporting Belidae as sister to Nemonychidae + Anthribidae is even lower than the discordance factor for an alternative quartet. In this case, the removal of phylogenetic noise and the use of appropriate evolutionary models could be of vital importance. Our re-analysis of the AA dataset by Shin *et al.* [21] shows that Belidae are grouped together with the clade (Attelabidae, (Caridae, (Brentidae, Curculionidae))), as long as the molecular data are properly filtered and (or) the infinite mixture model is adopted. This topology is in accordance with recent large-scale phylogenomic studies [12,22,23]. It is noteworthy that the coalescent tree based on AA data revealed the same interfamilial relationship (electronic supplementary material, figure S34; fig. S14 in [21]) as our preferred tree (figure 1). Actually, as shown by previous studies, methods



using coalescent or site-heterogeneous models can often reconstruct more plausible phylogenies, as they better reflect the real evolutionary process (e.g. [66–68]).

PhyloBayes analyses, all nodes of their AA and NT trees were maximally supported (BPP = 1). Such results are suspect, because a maxdiff of 1 means that at least one clade is inferred with a posterior probability of 1 in one chain and 0 in the other chain and therefore all BPPs of 1 in the consensus trees of the two chains are impossible by definition. It therefore seems possible that their maximally supported trees were derived from only one chain of each PhyloBayes analysis, but this is an inappropriate practice. Contrary to Shin *et al.* [21], in our CAT-GTR analysis of the same original AA dataset, Belidae were strongly supported as the sister group of the clade (Attelabidae, (Caridae, (Brentidae, Curculionidae)).

The genomic data used by Shin *et al.* [21] were obtained through anchored hybrid enrichment. While efficient for collecting a large amount of data, such target sequence capture

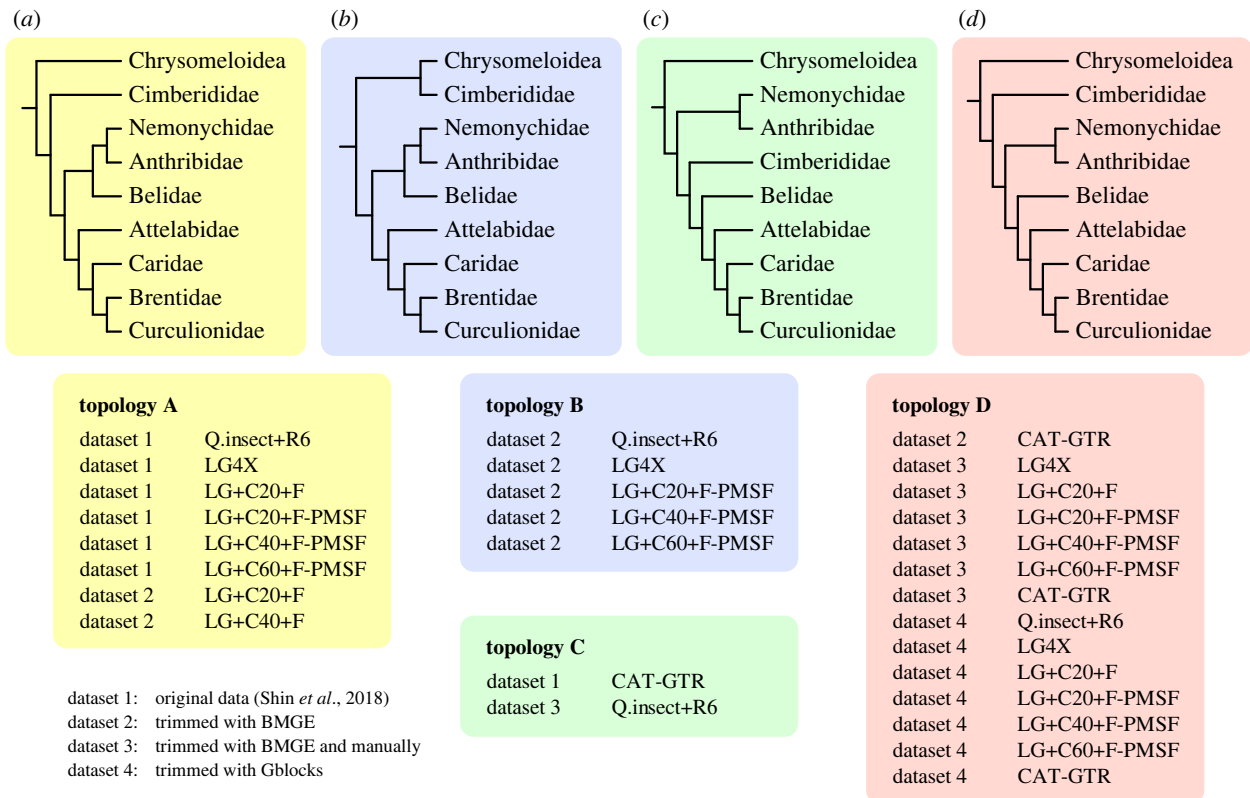


Figure 2. Comparison among tree topologies generated with different models and datasets. Topology D is our preferred topology for weevil phylogeny.

approaches could introduce considerable missing data [24], which might affect the accuracy of alignment and phylogenetic reconstruction [69]. As shown in table 1, the original concatenated datasets were very gappy and the data occupancy was low (22.1%). This is likely because they retained the flanking regions of a distant outgroup (*Aethina*), included some incorrectly aligned loci, and skipped the data filtering step before tree reconstruction. The impact of alignment trimming on downstream phylogenetic analysis is often not very clear [70], as it is challenging to achieve balance between removing potentially misleading alignment regions and removing phylogenetically informative regions [71,72]. Nevertheless, in our analyses based on the trimmed datasets 3 and 4, Belidae are always recovered in a stable position regardless of model choice and were congruent with analyses using the heterogeneous CAT-GTR model, indicating that proper trimming or data curation is crucial at least for the present dataset.

The systematic placement of Belidae based on our reanalyses is congruent with morphology-based studies [13–15] (see also [17,73]). Potential synapomorphies for the clade including Belidae, Attelabidae, Caridae, Brentidae and Curculionidae include: hypopharyngeal bracon without sclerome, frontoclypeal suture distinct, mandibular molar absent and maxillary articulatory lobes absent in larvae; and clypeolabral suture indistinct, mandibular pharyngeal process at least as long as mandible, mandibular molar absent, mandible with teeth at incisor area, maxillary palp short and spermathecal duct and gland contiguous or subcontiguous on spermathecal body in adults [13–15]. Actually, it is not uncommon for a counterintuitive (i.e. contradictory to morphology) phylogeny based on simple models to be later proved as an artefact by analyses with the better-fitting site-heterogeneous models (e.g. [37,74,75]).

Although the interfamilial relationships of weevils (Curculionidae) are now confidently resolved, the relationships within the ‘higher Curculionidae’ [21], especially the CCCMS clade, remain unsettled due to ancient rapid radiation likely during the Late Cretaceous as indicated by exceptionally short internodes. A more focused phylogenomic study of Curculionidae based on genome-scale data (e.g. UCEs, nuclear genomes) would be welcome to unravel the mysterious rapid radiation of derived weevils.

5. Concluding remarks

Weevils represent a classical case study on adaptive radiation associated with the diversification of flowering plants [10,11]. A valid phylogeny would be crucial to reconstructing the sequence of character acquisitions that enabled weevils to attain their present-day mega-diversity, and the timescale of their radiation. Here we provide a well-resolved backbone phylogeny of weevils, clarifying the interfamilial relationships of extant weevils. As recovered in other recent phylogenomic studies [12,23], Belidae are confirmed as the sister group of (Attelabidae, (Caridae, (Brentidae, Curculionidae))) based on the re-analysis of the AHE AA data by Shin *et al.* [21]. This topology is well accordant with previous studies based on morphology and other sources of molecular data. Our results further demonstrate the importance of data curation and modelling compositional heterogeneity in genomic-scale phylogenetic analyses of insects.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. The data are provided in the electronic supplementary material [76].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. Y.-D.L.: data curation, formal analysis, investigation, visualization, writing—original draft, writing—review and editing; M.S.E.: investigation, writing—review and editing; E.T.: investigation, writing—review and editing; C.C.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, supervision, writing—original draft, writing—review and editing.

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

Conflict of interest declaration. We declare we have no competing interests.

Funding. Financial support was provided by the Strategic Priority Research Program of the Chinese Academy of Sciences (grant no. XDB26000000) and the National Natural Science Foundation of China (grant nos 42222201, 42288201). Y.-D.L. is supported by a scholarship granted by the China Scholarship Council (grant no. 202108320010).

Acknowledgements. Zhen-Hua Liu kindly provided the habitus photo of *Rhinotia* sp. used in figure 1. Two reviewers provided helpful comments to improve the manuscript.

References

- Oberprieler RG, Anderson RS, Marvaldi AE. 2014 Curculionoidea Latreille, 1802: Introduction, Phylogeny. In *Handbook of zoology, Arthropoda: Insecta, Coleoptera, beetles, Vol 3: morphology and systematics (Phytophaga)* (eds NP Kristensen, RG Beutel), pp. 285–300. Berlin, Germany: Walter de Gruyter. (doi:10.1515/9783110274462.285)
- Oberprieler RG, Marvaldi AE, Anderson RS. 2007 Weevils, weevils, weevils everywhere. *Zootaxa* **1668**, 491–520. (doi:10.11646/zootaxa.1668.1.24)
- Anderson RS. 1995 An evolutionary perspective on diversity in Curculionoidea. *Mem. Entomol. Soc. Wash.* **14**, 103–118.
- Davis SR. 2017 The weevil rostrum (Coleoptera: Curculionoidea): internal structure and evolutionary trends. *Bull. Am. Mus. Nat. Hist.* **2017**, 1–76. (doi:10.1206/0003-0090-416.1.1)
- Davis SR. 2018 Developmental genetics in a complex adaptive structure, the weevil rostrum. *bioRxiv*, 287185. (doi:10.1101/287185)
- Davis SR. 2014 *Morphology, phylogeny, and evolutionary development in the weevils (Insecta: Coleoptera: Curculionoidea)*. Lawrence, KS: University of Kansas.
- Legalov A. 2015 Fossil Mesozoic and Cenozoic weevils (Coleoptera, Obrienoidea, Curculionoidea). *Paleontol. J.* **49**, 1442–1513. (doi:10.1134/S0031030115130067)
- Legalov A. 2018 Annotated key to weevils of the world. Part 1. Families Nemonychidae, Anthribidae, Belidae, Ithyceridae, Rhynchitidae, Brachyceridae and Brentidae. *Ukr. J. Ecol.* **8**, 780–831. (https://www.ujecology.com/abstract/annotated-key-to-weevils-of-the-world-part-1-families-nemonychidae-anthribidae-belidae-ithyceridae-rhynchitidae-brachyce-975.html)
- Clarke DJ, Limaye A, McKenna DD, Oberprieler RG. 2019 The weevil fauna preserved in Burmese amber—snapshot of a unique, extinct lineage (Coleoptera: Curculionoidea). *Diversity* **11**, 1. (doi:10.3390/d11010001)
- Farrell BD. 1998 'Inordinate fondness' explained: why are there so many beetles? *Science* **281**, 555–559. (doi:10.1126/science.281.5376.555)
- McKenna DD, Sequeira AS, Marvaldi AE, Farrell BD. 2009 Temporal lags and overlap in the diversification of weevils and flowering plants. *Proc. Natl Acad. Sci. USA* **106**, 7083–7088. (doi:10.1073/pnas.0810618106)
- McKenna DD *et al.* 2019 The evolution and genomic basis of beetle diversity. *Proc. Natl Acad. Sci. USA* **116**, 24 729–24 737. (doi:10.1073/pnas.1909655116)
- Kuschel G. 1995 A phylogenetic classification of Curculionoidea to families and subfamilies. *Mem. Entomol. Soc. Wash.* **14**, 5–33.
- Marvaldi AE, Morrone JJ. 2000 Phylogenetic systematics of weevils (Coleoptera: Curculionoidea): a reappraisal based on larval and adult morphology. *Insect Syst. Evol.* **31**, 43–58. (doi:10.1163/187631200X00309)
- Marvaldi AE, Sequeira AS, O'Brien CW, Farrell BD. 2002 Molecular and morphological phylogenetics of weevils (Coleoptera, Curculionoidea): do niche shifts accompany diversification? *Syst. Biol.* **51**, 761–785. (doi:10.1080/10635150290102465)
- Marvaldi AE, Duckett CN, Kjer KM, Gillespie JJ. 2009 Structural alignment of 18S and 28S rDNA sequences provides insights into phylogeny of Phytophaga (Coleoptera: Curculionoidea and Chrysomeloidea). *Zool. Scr.* **38**, 63–77. (doi:10.1111/j.1463-6409.2008.00360.x)
- Morimoto K, Kojima H. 2003 Morphologic characters of the weevil head and phylogenetic implications (Coleoptera, Curculionoidea). *Esakia* **43**, 133–169. (doi:10.5109/2677)
- Haran J, Timmermans MJTN, Vogler AP. 2013 Mitogenome sequences stabilize the phylogenetics of weevils (Curculionoidea) and establish the monophyly of larval ectophagy. *Mol. Phylogenet. Evol.* **67**, 156–166. (doi:10.1016/j.ympev.2012.12.022)
- Gillett CPDT, Crampton-Platt A, Timmermans MJTN, Jordal BH, Emerson BC, Vogler AP. 2014 Bulk de novo mitogenome assembly from pooled total DNA elucidates the phylogeny of weevils (Coleoptera: Curculionoidea). *Mol. Biol. Evol.* **31**, 2223–2237. (doi:10.1093/molbev/msu154)
- Gunter NL, Oberprieler RG, Cameron SL. 2016 Molecular phylogenetics of Australian weevils (Coleoptera: Curculionoidea): exploring relationships in a hyperdiverse lineage through comparison of independent analyses. *Austral Entomol.* **55**, 217–233. (doi:10.1111/aen.12173)
- Shin S *et al.* 2018 Phylogenomic data yield new and robust insights into the phylogeny and evolution of weevils. *Mol. Biol. Evol.* **35**, 823–836. (doi:10.1093/molbev/msx324)
- Zhang S-Q, Che L-H, Li Y, Liang D, Pang H, Ślipiński A, Zhang P. 2018 Evolutionary history of Coleoptera revealed by extensive sampling of genes and species. *Nat. Commun.* **9**, 205. (doi:10.1038/s41467-017-02644-4)
- Cai C *et al.* 2022 Integrated phylogenomics and fossil data illuminate the evolution of beetles. *R. Soc. Open Sci.* **9**, 211771. (doi:10.1098/rsos.211771)
- Lemmon AR, Emme SA, Lemmon EM. 2012 Anchored hybrid enrichment for massively high-throughput phylogenomics. *Syst. Biol.* **61**, 727–744. (doi:10.1093/sysbio/sys049)
- Lemmon EM, Lemmon AR. 2013 High-throughput genomic data in systematics and phylogenetics. *Ann. Rev. Ecol. Syst.* **44**, 99–121. (doi:10.1146/annurev-ecolsys-110512-135822)
- Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, Lemmon AR. 2015 A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **526**, 569–573. (doi:10.1038/nature15697)
- Hamilton CA, Lemmon AR, Lemmon EM, Bond JE. 2016 Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. *BMC Evol. Biol.* **16**, 212. (doi:10.1186/s12862-016-0769-y)
- Burbrink FT *et al.* 2020 Interrogating genomic-scale data for Squamata (lizards, snakes, and amphisbaenians) shows no support for key traditional morphological relationships. *Syst. Biol.* **69**, 502–520. (doi:10.1093/sysbio/syz062)
- Hime PM *et al.* 2021 Phylogenomics reveals ancient gene tree discordance in the amphibian tree of life. *Syst. Biol.* **70**, 49–66. (doi:10.1093/sysbio/syaa034)
- Dietrich CH, Allen JM, Lemmon AR, Lemmon EM, Takiya DM, Evangelista O, Walden KKO, Grady PGS, Johnson KP. 2017 Anchored hybrid enrichment-based phylogenomics of leafhoppers and treehoppers (Hemiptera: Cicadomorpha: Membracidae). *Insect Syst. Div.* **1**, 57–72. (doi:10.1093/isd/ixx003)
- Haddad S, Shin S, Lemmon AR, Lemmon EM, Svacha P, Farrell B, Ślipiński A, Windsor D, McKenna DD. 2018 Anchored hybrid enrichment provides new insights into the phylogeny and evolution of longhorned beetles (Cerambycidae). *Syst. Entomol.* **43**, 68–89. (doi:10.1111/syen.12257)

32. Piekarski PK, Carpenter JM, Lemmon AR, Lemmon EM, Sharanowski BJ. 2018 Phylogenomic evidence overturns current conceptions of social evolution in wasps (Vespidae). *Mol. Biol. Evol.* **35**, 2097–2109. (doi:10.1093/molbev/msy124)
33. Buenaventura E, Szpila K, Cassel BK, Wiegmann BM, Pape T. 2020 Anchored hybrid enrichment challenges the traditional classification of flesh flies (Diptera: Sarcophagidae). *Syst. Entomol.* **45**, 281–301. (doi:10.1111/syen.12395)
34. Frago-Martínez I, Salazar GA, Martínez-Gordillo M, Magallón S, Sánchez-Reyes L, Lemmon EM, Lemmon AR, Szatormil F, Mendoza CG. 2017 A pilot study applying the plant anchored hybrid enrichment method to New World sages (*Salvia* subgenus *Calospatha*; Lamiaceae). *Mol. Phylogenet. Evol.* **117**, 124–134. (doi:10.1016/j.ympev.2017.02.006)
35. Liu Y *et al.* 2019 Resolution of the ordinal phylogeny of mosses using targeted exons from organellar and nuclear genomes. *Nat. Commun.* **10**, 1485. (doi:10.1038/s41467-019-09454-w)
36. Cai C-Y, Tihelka E, Liu X-Y, Engel MS. 2023 Improved modelling of compositional heterogeneity reconciles phylogenomic conflicts among lacewings. *Palaeoentomology* **6**, 49–57. (doi:10.11646/palaeoentomology.6.1.8)
37. Cai C, Tihelka E, Pisani D, Donoghue PCJ. 2020 Data curation and modeling of compositional heterogeneity in insect phylogenomics: a case study of the phylogeny of Dytiscoidea (Coleoptera: Adephaga). *Mol. Phylogenet. Evol.* **147**, 106782. (doi:10.1016/j.ympev.2020.106782)
38. Tihelka E, Cai C, Giacomelli M, Pisani D, Donoghue PCJ. 2020 Integrated phylogenomic and fossil evidence of stick and leaf insects (Phasmatodea) reveal a Permian–Triassic co-origination with insectivores. *R. Soc. Open Sci.* **7**, 201689. (doi:10.1098/rsos.201689)
39. Tihelka E, Giacomelli M, Huang D-Y, Pisani D, Donoghue PCJ, Cai C-Y. 2020 Fleas are parasitic scorpionflies. *Palaeoentomology* **3**, 641–653. (doi:10.11646/palaeoentomology.3.6.16)
40. Tihelka E, Cai C, Giacomelli M, Lozano-Fernandez J, Rota-Stabelli O, Huang D, Engel MS, Donoghue PCJ, Pisani D. 2021 The evolution of insect biodiversity. *Curr. Biol.* **31**, R1299–R1311. (doi:10.1016/j.cub.2021.08.057)
41. Inagaki Y, Simpson AGB, Dacks JB, Roger AJ. 2004 Phylogenetic artifacts can be caused by leucine, serine, and arginine codon usage heterogeneity: dinoflagellate plastid origins as a case study. *Syst. Biol.* **53**, 582–593. (doi:10.1080/10635150490468756)
42. Criscuolo A, Gribaldo S. 2010 BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210. (doi:10.1186/1471-2148-10-210)
43. Castresana J. 2000 Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552. (doi:10.1093/oxfordjournals.molbev.a026334)
44. Fleming JF, Struck TH. 2023 nRCFV: a new, dataset-size-independent metric to quantify compositional heterogeneity in nucleotide and amino acid datasets. *BMC Bioinform.* **24**, 145. (doi:10.1186/s12859-023-05270-8)
45. Le SQ, Dang CC, Gascuel O. 2012 Modeling protein evolution with several amino acid replacement matrices depending on site rates. *Mol. Biol. Evol.* **29**, 2921–2936. (doi:10.1093/molbev/mss112)
46. Le SQ, Gascuel O, Lartillot N. 2008 Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* **24**, 2317–2323. (doi:10.1093/bioinformatics/btn445)
47. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. 2015 IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. (doi:10.1093/molbev/msu300)
48. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. 2020 IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534. (doi:10.1093/molbev/msaa015)
49. Kalyanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermini LS. 2017 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589. (doi:10.1038/nmeth.4285)
50. Minh BQ, Dang CC, Vinh LS, Lanfear R. 2021 QMaker: fast and accurate method to estimate empirical models of protein evolution. *Syst. Biol.* **70**, 1046–1060. (doi:10.1093/sysbio/syab010)
51. Wang H-C, Minh BQ, Susko E, Roger AJ. 2018 Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Syst. Biol.* **67**, 216–235. (doi:10.1093/sysbio/syx068)
52. Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. 2018 UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522. (doi:10.1093/molbev/msx281)
53. Minh BQ, Hahn MW, Lanfear R. 2020 New methods to calculate concordance factors for phylogenomic datasets. *Mol. Biol. Evol.* **37**, 2727–2733. (doi:10.1093/molbev/msaa106)
54. Lartillot N, Philippe H. 2004 A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095–1109. (doi:10.1093/molbev/msh112)
55. Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013 PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* **62**, 611–615. (doi:10.1093/sysbio/syt022)
56. Letunic I, Bork P. 2021 Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296. (doi:10.1093/nar/gkab301)
57. R Core Team. 2021 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
58. Pages H, Aboyoun P, Gentleman R, DebRoy S. 2021 Biostrings: string objects representing biological sequences, and matching algorithms. *R package version 2.62.0*.
59. Charif D, Lobry JR. 2007 SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In *Structural approaches to sequence evolution: molecules, networks, populations* (eds U Bastolla, M Porto, HE Roman, M Vendruscolo), pp. 207–232. Berlin, Germany: Springer. (doi:10.1007/978-3-540-35306-5_10)
60. Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018 ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.* **19**, 15–30. (doi:10.1186/s12859-018-2021-9)
61. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017 ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36. (doi:10.1111/2041-210X.12628)
62. Yu G. 2020 Using ggtree to visualize data on tree-like structures. *Curr. Protoc. Bioinform.* **69**, e96. (doi:10.1002/cpbi.96)
63. Lartillot N. 2020 PhyloBayes: Bayesian phylogenetics using site-heterogeneous models. In *Phylogenetics in the genomic era* (eds C Scornavacca, F Delsuc, N Galtier), pp. 1.5:1–1.5:16. See <https://hal.inria.fr/PGE>
64. Sayyari E, Mirarab S. 2018 Testing for polytomies in phylogenetic species trees using quartet frequencies. *Genes* **9**, 132. (doi:10.3390/genes9030132)
65. Bocak L, Barton C, Crampton-Platt A, Chesters D, Ahrens D, Vogler AP. 2014 Building the Coleoptera tree-of-life for >8000 species: composition of public DNA data and fit with Linnaean classification. *Syst. Entomol.* **39**, 97–110. (doi:10.1111/syen.12037)
66. Schwentner M, Combosch DJ, Nelson JP, Giribet G. 2017 A phylogenomic solution to the origin of insects by resolving crustacean–hexapod relationships. *Curr. Biol.* **27**, 1818–1824. (doi:10.1016/j.cub.2017.05.040)
67. Williams TA, Cox CJ, Foster PG, Szöllösi GJ, Embley TM. 2020 Phylogenomics provides robust support for a two-domains tree of life. *Nat. Ecol. Evol.* **4**, 138–147. (doi:10.1038/s41559-019-1040-x)
68. Yu D, Ding Y, Tihelka E, Cai C, Hu F, Liu M, Zhang F. 2022 Phylogenomics of elongate-bodied springtails reveals independent transitions from aboveground to belowground habitats in deep time. *Syst. Biol.* **71**, 1023–1031. (doi:10.1093/sysbio/syaa024)
69. Lemmon AR, Brown JM, Stanger-Hall K, Lemmon EM. 2009 The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Syst. Biol.* **58**, 130–145. (doi:10.1093/sysbio/syp017)
70. Kapli P, Yang Z, Telford MJ. 2020 Phylogenetic tree building in the genomic age. *Nat. Rev. Genet.* **21**, 428–444. (doi:10.1038/s41576-020-0233-0)
71. Talavera G, Castresana J. 2007 Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577. (doi:10.1080/10635150701472164)

72. Tan G, Muffato M, Ledergerber C, Herrero J, Goldman N, Gil M, Dessimoz C. 2015 Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Syst. Biol.* **64**, 778–791. (doi:10.1093/sysbio/syv033)
73. Zherikhin V, Gratshev V. 1995 A comparative study of the hind wing venation of the superfamily Curculionoidea, with phylogenetic implications. In *Biology, phylogeny and classification of Coleoptera: papers celebrating the 80th birthday of Roy A. Crowson, Vol. 2* (eds J Pakaluk, A Ślipiński), pp. 633–777. Warszawa: Muzeum i Instytut Zoologii Polska Akademia Nauk.
74. Philippe H *et al.* 2009 Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* **19**, 706–712. (doi:10.1016/j.cub.2009.02.052)
75. Campbell LI *et al.* 2011 MicroRNAs and phylogenomics resolve the relationships of Tardigrada and suggest that velvet worms are the sister group of Arthropoda. *Proc. Natl Acad. Sci. USA* **108**, 15 920–15 924. (doi:10.1073/pnas.1105499108)
76. Li Y-D, Engel MS, Tihelka E, Cai C. 2023 Phylogenomics of weevils revisited: data curation and modeling compositional heterogeneity. Figshare. (doi:10.6084/m9.figshare.c.6825533)