

Appendix S1. Supplementary analyses to test for the impact of missing data.

S1.1. Supplementary Material and Methods.

To explore the potential impact of missing data, three datasets with different percentages of missing data were analysed (see Appendix S1.2): (i) a species-level dataset that includes all genes (*i.e.* the species-level dataset presented in the main text of the manuscript), (ii) a species-level dataset without *cox3*, and (iii) a species-level dataset without *cox3*, *18S* and *28S*. Percentages of missing data in each dataset were calculated using AMAS (Borowiec, 2016). All analyses were performed under maximum likelihood (ML) using IQ-TREE v2.1.3 (Minh *et al.*, 2020), with the partition strategy and settings presented in the main text of the manuscript. The Bayesian Information Criterion (BIC) implemented in IQ-TREE was used to choose the best-fit substitution models and partition schemes. Clade support was estimated using 1,000 replicates for both SH-like approximate likelihood ratio tests (SH-aLRT; Guindon *et al.*, 2010) and ultrafast bootstraps (uBV; Minh *et al.*, 2013); in addition, Transfer Bootstrap Expectation (TBE) branch support metrics were calculated with BOOSTER (Lemoine *et al.*, 2018), using 100 bootstrap trees generated with IQ-TREE as input. For each metric, we followed the recommendations of authors and relied on the following thresholds: SH-aLRT values $\geq 80\%$ (see Guindon *et al.*, 2010), uBV values $\geq 95\%$ (see Minh *et al.*, 2013) and TBE values $\geq 70\%$ (see Lemoine *et al.*, 2018).

S1.2. Percentages of missing data for the three species-level datasets; percentages were calculated for all species ('global'), for Palparidiini and Palparini species ('ingroups') or for all species with the exception of Palparidiini and Palparini species ('outgroups').

Dataset	% missing data		
	global	ingroups	outgroups
1	57 %	62 %	53 %
2	57 %	63 %	52 %
3	46 %	47 %	45 %

S1.3. Supplementary Results and Discussion

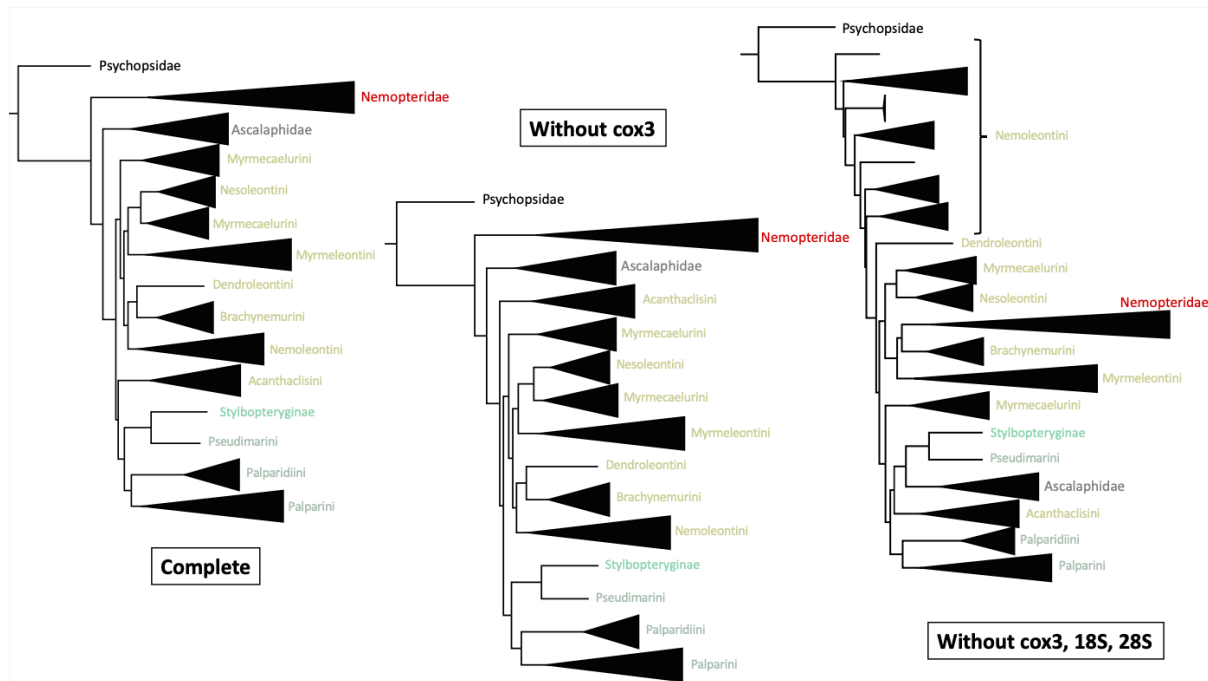
The results of our analyses indicate that - regardless of the metric used - branch support values are always higher for the focal clade of this study (see the ‘ingroups’ columns in Table S1.4). The analyses based on the dataset with the lowest level of missing data (46% instead of 57%) also yield branch supports that are lower, with the exception of uBV values when either considering all species or ingroup taxa only (see Appendix S.1.4).

The trees resulting from the analyses of the two species-level datasets that did not include all genes present some noticeable levels of incongruence with the topology resulting from the analysis of all available gene fragments. For example, the analysis of the dataset excluding *cox3*, *18s* and *28S* does not recover the monophyly of Myrmeleontidae due to the placement of Nemopteridae within them (see Appendix S1.5). Another incongruence relates to the placement of Acanthaclisini which are sister of Palparinae and Stylbopteryginae in the analysis of the complete dataset; by contrast when analysing the dataset excluding *cox3* Acanthaclisini are recovered sister of all remaining Myrmeleontidae whereas when analysing the dataset excluding *cox3*, *18s* and *28S* Acanthaclisini are found sister to Ascalaphidae, Pseudimarini and Stylbopteryginae (see Appendix S1.5). When looking at our focal clade, excluding *cox3* results in having the Nemoleontini clustering in a derived position within the Palparini (see Appendix S.1.6). These three examples of unequivocally misplaced lineages exemplify the interest of not excluding genes with high levels of missing data (see Table S2). Specifically, it appears that *18S* and *28S* are really useful at stabilizing deep relationships for our dataset. These additional analyses ultimately indicate that - for our dataset - removing gene fragments to lower the amount of missing data is detrimental to phylogenetic accuracy and robustness.

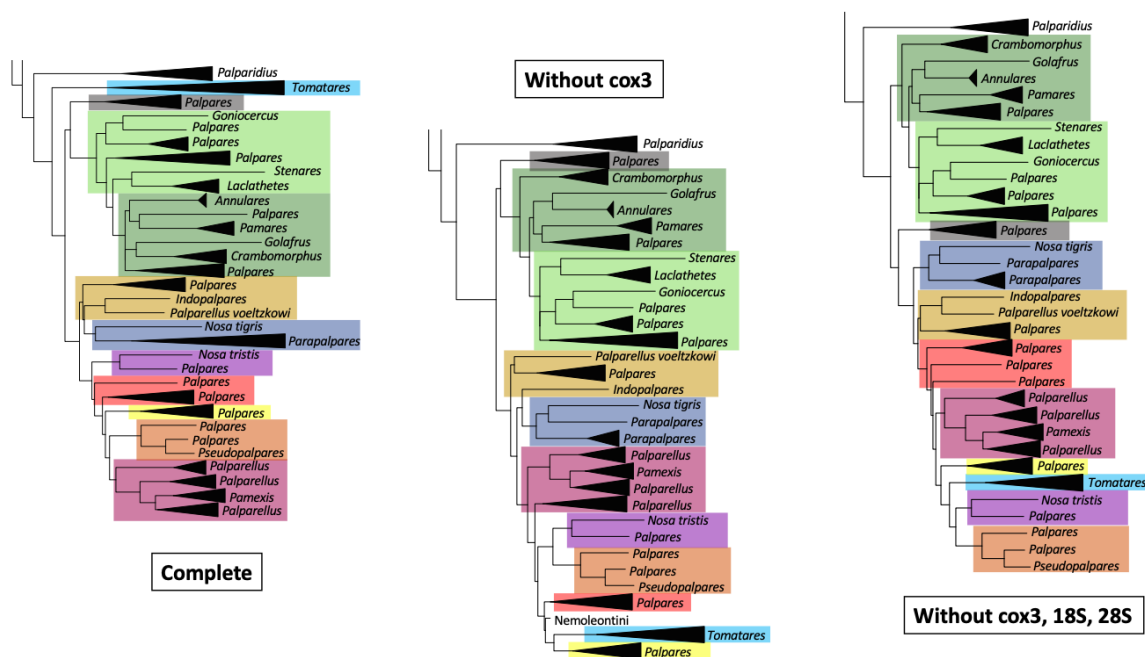
S1.4. Table summarizing the support values resulting from the analyses of the three species-level datasets.

Dataset	% SH-aLRT			% uBV			% TBE		
	global	ingroups	outgroups	global	ingroups	outgroups	global	ingroups	outgroups
1	63 %	73 %	56 %	46 %	47 %	45 %	51 %	53 %	49 %
2	62 %	69 %	56 %	49 %	51 %	47 %	47 %	53 %	43 %
3	57 %	66 %	51 %	50 %	69 %	36 %	43 %	58 %	32 %

S1.5. Simplified topologies focusing on outgroups for the three species-level datasets. For the dataset without *cox3*, *18S* and *28S* some Nemoleontini are recovered within Acanthaclisini, Myrmeleontini and Nesoleontini but are not represented here.



S1.6. Simplified topologies of ingroups for the three species-level datasets. For more clarity, some clades have been arbitrarily highlighted with distinct colors.



Additional data availability statement

The data supporting the findings of those supplementary analyses are openly available in Figshare at <https://doi.org/10.6084/m9.figshare.21904746>.

Additional references

Borowiec, M.L. (2016) AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* **4**, e1660.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, **59**(3), 307–321.

Lemoine, F., Domelevo Entfellner, J.-B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T. & Gascuel, O. (2018) Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature*, **556**, 452–456.

Minh, B.Q., Nguyen, M.A.T. & von Haeseler, A. (2013) Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, **30**(5), 1188–1195.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A. & Lanfear, R. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, **37**(5), 1530–1534.