

# A molecular phylogeny for all 21 families within Chiroptera (bats)

Xiangyu HAO,<sup>1,2</sup> Qin LU<sup>1</sup> and Huabin ZHAO<sup>1</sup>

<sup>1</sup>College of Life Sciences, Wuhan University, Wuhan, Hubei, China and <sup>2</sup>College of Plant Protection, Northwest A&F University, Yangling, Shaanxi, China

#### Abstract

Bats, members of the Chiroptera order, rank as the second most diverse group among mammals. Recent molecular systematic studies on bats have successfully classified 21 families within two suborders: Yinpterochiroptera and Yangochiroptera. Nevertheless, the phylogeny within these 21 families has remained a subject of controversy. In this study, we have employed a balanced approach to establish a robust family-level phylogenetic hypothesis for bats, utilizing a more comprehensive molecular dataset. This dataset includes representative species from all 21 bat families, resulting in a reduced level of missing genetic information. The resulting phylogenetic tree comprises 21 lineages that are strongly supported, each corresponding to one of the bat families. Our findings support to place the Emballonuroidea superfamily as the basal lineage of Yangochiroptera, and that Myzopodidae should be situated as a basal lineage of Emballonuroidea, forming a sister relationship with the clade consisting of Nycteridae and Emballonuridae. Finally, we have conducted dating analyses on this newly resolved phylogenetic tree, providing divergence times for each bat family. Collectively, our study has employed a relatively comprehensive molecular dataset to establish a more robust phylogeny encompassing all 21 bat families. This improved phylogenetic framework will significantly contribute to our understanding of evolutionary processes, ecological roles, disease dynamics, and biodiversity conservation in the realm of bats.

### **INTRODUCTION**

Bats (members of order Chiroptera), consisting of over 1400 species and accounting for approximately one-fifth of the world's living mammal species, are the second-most speciose group of mammals (Wilson & Mittermeier 2020; Lu *et al.* 2021). Understanding the phylogeny of bats is crucial, as it informs evolutionary processes, ecological roles, disease dynamics, and biodiversity conservation efforts (Ramírez-Fráncel *et al.* 2021; Sharnuud & Ameca 2023). Current molecular systematic studies have established the classification of 21 families within two suborders: Yinpterochiroptera and Yangochiroptera (Teeling

*Correspondence*: Qin Lu and Huabin Zhao, College of Life Sciences, Wuhan University, Wuhan, Hubei, China. Email: luqin@whu.edu.cn and huabinzhao@whu.edu.cn et al. 2000; Van Den Bussche & Hoofer 2004; Teeling et al. 2018), including the three newly confirmed families Miniopteridae, Cistugidae and Rhinonycteridae, in addition to the traditional 18 subfamilies (Hoofer & Van Den Bussche 2003; Lack et al. 2010; Foley et al. 2015). However, the family-level relationships within Yangochiroptera are still controversial. For example, one phylogenetic hypothesis supported that the superfamily Emballonuroidea, containing the two families Emballonuridae and Nycteridae, is the basal lineage of Yangochiroptera; it is also the sister group to other yangochiropteran taxa (i.e. superfamilies Noctilionoidea and Vespertilionoidea) (Teeling et al. 2005, 2018) (Fig. 1a). In comparison, other phylogenetic hypotheses argued that the two superfamilies, Emballonuroidea and Noctilionoidea, are sister groups (Meredith et al. 2011; Amador et al. 2018; Álvarez-Carretero et al. 2022) (Fig. 1b-d). Moreover, the

<sup>© 2023</sup> International Society of Zoological Sciences, Institute of Zoology/ Chinese Academy of Sciences and John Wiley & Sons Australia, Ltd.



Figure 1 Four different phylogenetic hypotheses proposed by previous studies. (a) Teeling *et al.* (2018) Annual Review of Animal Biosciences; (b) Meredith *et al.* (2011) *Science*; (c) Amador *et al.* (2018) Journal of Mammalian Evolution; (d) Álvarez-Carretero *et al.* (2022) *Nature*.

taxonomic status of the family Myzopodidae remains ambiguous. Specifically, the work of Teeling *et al.* (2018) and a recent species-level study from Álvarez-Carretero *et al.* (2022) supported that Myzopodidae should be positioned at the base of the superfamily Noctilionoidea (Fig. 1a,d), whereas Amador *et al.* (2018) placed Myzopodidae into Emballonuroidea, which is the sister group to Noctilionoidea (Fig. 1c). In addition, Meredith *et al.* (2011) argued that Myzopodidae should be positioned in the basal clade of the superfamily Vespertilionoidea (Fig. 1b).

The lack of genomic data has forced phylogeneticists to make trade-offs between higher coverage of taxonomic sampling and more complete data matrices. From the four phylogenetic hypotheses depicted (Fig. 1), Amador *et al.* (2018) and Álvarez-Carretero *et al.* (2022) sampled 799 and 890 species, respectively, and reconstructed the species-level evolutionary relationships of the relatively complete bat taxa, but with over 60% (using 9 genes) and 95% (the available molecular characters are still small although they claimed to use 182 loci) missing data matrices, which are calculated by the ratio of the total number of missing genes of all species in the concatenated super-gene matrix to the total number of genes in this matrix. On the contrary, Teeling *et al.* (2005) and Meredith *et al.* (2011) included relatively incomplete taxonomic sampling to obtain relatively complete molecular datasets. Previous simulations demonstrated that the estimates of tree topology and branch lengths may be biased by nonrandomly distributed missing data

in maximum likelihood (ML) and Bayesian inference (BI) approaches (Agnarsson & May-Collado 2008; Lemmon *et al.* 2009; Simmons 2012; Xi *et al.* 2016). A common view that seems to apply to all phylogenetic methods is that high levels of missing data are problematic for phylogenies based on small datasets (i.e. overall number of characters is small) (Philippe *et al.* 2004; Wiens & Morrill 2011). In addition, in Wiens's argument, incomplete taxa may be misleading in phylogenetic inference due to long branches, especially for model-based methods (e.g. Bayesian analysis, likelihood, and neighbor-joining), and adding taxa could lead to dramatic increases in accuracy (Wiens 2006).

Currently, the complete molecular data matrix without missing genes for reconstructing the family-level phylogeny of bats is unavailable. In this study, we therefore have used a compromise strategy to provide a robust family-level phylogenetic hypothesis of bats based on a concatenated molecular data matrix, which contains representative species of all 21 bat families and a low level of missing genes.

#### MATERIALS AND METHODS

Our molecular dataset is retrieved from the supermatrix in the study of Amador et al. (2018), including four mitochondrial genes (Cyt-b, ND1, 12S rRNA, and 16S rRNA) and five nuclear genes (BRCA1, DMP1, RAG1, RAG2, and vWF). The taxon sampling strategy is that the species must contain six or more of these nine genes, with a few exceptions when some families only contain one or few species for selection. For example, species of the following families have only five genes available in the super-matrix: Cistugo seabrae (Cistugidae), Craseonycteris thonglongvai (Craseonycteridae). Rhvnchonycteris naso (Emballonuridae), Taphozous nudiventris (Emballonuridae), Hipposideros diadema (Hipposideridae), Nyctinomops femorosaccus (Molossidae), Chilonatalus micropus (Natalidae), and Mormoops blainvillii and M. megalophylla (Mormoopidae). Notably, the only species Paratriaenops furculus in the family Rhinonycteridae just has three available genes (Table S1, Supporting Information). Our samples finally included 151 bat species covering 21 families. Furthermore, additional data were obtained to this initial data matrix described above to cover the missing genes as much as possible. We retrieved all currently available whole genomes, complete mitochondrial genomes, and the Sequence Read Archive (SRA) data of all selected bat species in the National Center for Biotechnology Information (NCBI) database. We eventually successfully extracted and/or assembled 95 genes

(including partial genes) from 20 whole genomes, 19 mitochondrial genomes, and 18 SRA data, covering a total of 50 species (Table S1, Supporting Information). Newly obtained sequences in this study are provided in the Dataset S1, Supporting Information. The accession numbers of all genes are provided in Table S1, Supporting Information. Accession numbers for SRA data that failed to assemble into genes of interest (i.e. the four mitochondrial and the five nuclear genes) were not provided. Human (*Homo sapiens*), Chinese pangolin (*Manis pentadactyla*), and West Indian manatee (*Trichechus manatus*) are selected as outgroups.

We used two strategies to restore the missing data: (1) We re-retrieved the complete mitochondrial genomes (mitogenomes) and the available assembled whole-genomes in the NCBI database. Missing mitochondrial genes were extracted from complete mitogenomes using PhyloSuite (Zhang et al. 2020; Xiang et al. 2023), and nuclear genes were acquired in the whole genomes using the tblastn program (Altschul et al. 1990). The protein sequences from Amador et al. (2018) and intact proteincoding genes (PCGs) in NCBI were used as queries with the tblastn program. Open reading frames (ORFs) were corrected by GeneWise (Birney et al. 2004). (2) We searched all available SRA data in NCBI to assemble the missing genes. GetOrganelle (Jin et al. 2020) was used to assemble raw reads into complete mitogenomes. Subsequently, the target genes with intact ORFs were identified by multiple sequence alignments. For nuclear genes, raw reads were mapped to reference sequences using Geneious R9 (Biomatters, Auckland, New Zealand).

Reconstruction of phylogeny was based on four different sub-datasets: (1) PR matrix, containing all codon positions of the seven PCGs (including Cyt-b, ND1, BRCA1, DMP1, RAG1, RAG2, vWF), and the two rRNAs (12S rRNA and 16S rRNA); (2) P12R matrix, containing the first and second codon positions of the seven PCGs, and the two rRNAs; (3) P3R matrix, containing the third codon positions of the seven PCGs, and the two rRNAs; (4) AA matrix, including all amino acid sequences of the seven PCGs. The phylogenetic trees were reconstructed as follows: Multiple sequence alignments were conducted using MAFFT (Katoh et al. 2002; Katoh & Standley 2013) with L-INS-i (accurate) strategy under the codon alignment mode for PCGs and the nucleotide mode for RNAs. Poorly aligned regions were removed using Gblocks (Castresana 2000). Individual genes were then concatenated into molecular data matrices. Based on the four sub-datasets, phylogenetic trees were reconstructed using the ML and BI methods. The optimal partitioning

<sup>© 2023</sup> International Society of Zoological Sciences, Institute of Zoology/ Chinese Academy of Sciences and John Wiley & Sons Australia, Ltd.

schemes of genes, codon positions, and nucleotide substitution models of ML and BI analyses were selected using PartitionFinder2 (Lanfear et al. 2017) with the greedy algorithm (Table S2, Supporting Information). ML analvses were inferred by IO-TREE (Nguyen et al. 2015) with the Ultrafast bootstrap (UFB) algorithm (Hoang et al. 2018), and the bootstrap support value of each node was estimated with 10 000 UFB replicates. BI method was conducted using MrBayes (Ronquist et al. 2012). Two sets of independent runs were conducted, each consisting of 20 million generations. Simultaneously, four independent Markov Chain Monte Carlo (MCMC) runs were performed, with sampling occurring every 5000 generations. A consensus tree was obtained after the initial 25% of the trees were discarded as burn-in in each MCMC run. The final consensus tree was considered to have reached convergence when the average standard deviation of split frequencies became smaller than 0.01, and the confidence value of each node was shown as the Bayesian posterior probability (BPP). To test the impact of sequence heterogeneity on phylogeny, we additionally used AliGROOVE (Kück et al. 2014) to visualize the heterogeneity levels of combined subsets of mitochondrial and nuclear genes (Fig. S1, Supporting Information), with ambiguous parameters for DNA indels and blocks of amino acid substitution matrix 62 (BLOSUM62), respectively. PhyloBayes-MPI (Lartillot et al. 2013) was used for Bayesian phylogenetic inference based on the mixture model (CAT-GTR model). Two independent chains were run, and a consensus tree was obtained when the discrepancy of bipartition frequencies (maxdiff value) between the two chains was smaller than 0.3. Alternative topologies were tested using the four-cluster likelihood mapping (FcLM) (Strimmer & von Haeseler 1997), Kishino-Hasegawa (KH) (Kishino & Hasegawa 1989), Shimodaira-Hasegawa (SH) (Shimodaira & Hasegawa 1999; Goldman et al. 2000), and approximately unbiased (AU) (Shimodaira 2002) methods, respectively.

Divergence times of all species were estimated using the MCMCtree program in the PAML package (Yang 2007). Baseml program was used for the overall estimation of the substitution rate of the dataset (GTR+G model), and MCMCtree was then used to estimate the gradient and Hessian of the likelihood values, branch lengths, and divergence times. We used seven fossil constraints (two in Yinpterochiroptera and five in Yangochiroptera, see Table S3, Supporting Information, in detail) in our molecular dating analyses, which were taken from Meredith *et al.* (2011).

### **RESULTS AND DISCUSSION**

In this study, we used four mitochondrial genes and five nuclear genes to provide a robust family-level phylogenetic tree of bats based on the integrated datasets from Amador *et al.* (2018) and our newly extracted or assembled genes with a lower level of missing data (19.5% missing genes).

Phylogenetic analyses using both ML (IQ-TREE) and BI (MrBayes) methods vielded nearly identical familylevel relationships for the same sub-datasets, and topologies based on different datasets were distinct (Dataset S2, Supporting Information). The familial relationships inferred from the PR and AA matrices were the same except for some conflicts within superfamilies Rhinolophoidea and Emballonuroidea (Trees 1-4 in Dataset S2, Supporting Information). As a result, we only presented two main topologies (PR and P12R matrix) inferred from different analyses in Fig. 2a. Several phylogenetic inconsistencies were observed in the two trees, most notably concerning the position of the superfamily Emballonuroidea (including three families: Emballonuridae, Nycteridae, and Myzopodidae) (Fig. 2a). The topologies of the PR matrix inferred by both IQ-TREE and MrBayes strongly supported the basal position of Emballonuroidea within the suborder Yangochiroptera (BPP = 1) (Fig. 2a; Trees 1 and 2 in Dataset S2, Supporting Information). Conversely, the P12R matrix supported a sister relationship between the two superfamilies Emballonuroidea and Noctilionoidea, the latter of which included six families: Phyllostomidae, Mormoopidae, Noctilionidae, Furipteridae, Thyropteridae, and Mystacinidae (Fig. 2a; Trees 5 and 6 in Dataset S2, Supporting Information). Neither topology was explicitly rejected when using topology tests based on both PR and P12R matrices (Table S4, Supporting Information). It is noteworthy that the P12R\_ML tree contained some unsolved nodes, but a more robust family-level topology, indicated by higher support values, was obtained through the utilization of the PR matrix compared to the P12R matrix (Fig. 2a).

Furthermore, the relationships among families, such as the three families Rhinonycteridae, Rhinolophidae, and Hipposideridae within the superfamily Rhinolophoidea, as well as the three families Myzopodidae, Nycteridae, and Emballonuridae within the superfamily Emballonuroidea, remain equivocal (Fig. 2a–c). Various FcLM analyses preferred the sister-group relationship between Hipposideridae and Rhinolophidae (Hypothesis A3), whereas the KH/SH/AU tests supported two other hypotheses A1 and A2 (Fig. 2d,f,g; Table 1). For the

<sup>© 2023</sup> International Society of Zoological Sciences, Institute of Zoology/ Chinese Academy of Sciences and John Wiley & Sons Australia, Ltd.



**Figure 2** Inconsistencies of phylogenetic inferences based on different datasets. (a) Comparing phylogenetic topologies between PR matrix (MrBayes) (left panel) and P12R matrix (IQ-TREE) (right panel). Lineages with major phylogenetic conflicts are indicated by dashed lines. Green, yellow, red, and black squares at each node represent 95–100%, 90–95%, 75–90%, and 0–75% of the node support value (the ultrafast bootstrap value for maximum likelihood tree and the Bayesian posterior probability for Bayesian inference tree), respectively. (b,c) Partially highlighted images showing phylogenetic inconsistencies. (d,e) Three alternative phylogenetic hypotheses for topology test using the four-cluster likelihood mapping method. (f–i) Four-cluster likelihood mapping of three alternative hypotheses based on PR and P12R matrices.

superfamily Emballonuroidea, all topology tests placed Myzopodidae at the basal position of Emballonuroidea (Fig. 2e,h,i; Table 1). In addition, the results of the P3R sub-dataset showed an obvious phylogenetic error, as the monophyly of Yinpterochiroptera is not recovered (Trees 7-8, Dataset S2, Supporting Information). One possible explanation is that the third codon positions may have valuable phylogenetic information when inferring relationships within genera or closely related genera (Lartillot et al. 2013), but they may not suit the family-level inference. Moreover, these fast-evolving codon positions may have excessive variations in lineages with rapid adaptive radiations such as bats, and the resulting substitution saturations may bias phylogenetic inference (Yang 1996; Breinholt & Kawahara 2013). Another possibility is that the sequences of the P3R subset lack enough resolutions as it is almost one-third the size of the entire PR dataset. Notably, in the subsequent PhyloBayes analyses utilizing the mixture heterogeneity model that incorporates variations across sites in the amino acid replacement process, we observed that the CAT + GTR model did not yield satisfactory results for any of the sub-datasets, although the maxdiff values indicated convergence of the trees based on the PR matrix (maxdiff = 0.157) and P12R matrix (maxdiff = 0.127). This could be explained by the CAT + GTR model that does not fit well on these small datasets due to the lack of sufficient information on the sequences (e.g. the number of parsimony-informative sites is small) (Lartillot et al. 2009). Nonetheless, all our analyses suggest that the family Myzopodidae should be placed into the superfamily Emballonuroidea, which is the sister group to Noctilionoidea. Given the highest support

17494877, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/1749-4877.1272 by Wuhan University, Wiley Online Library on [27/10/023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons



**Figure 3** Time-calibrated phylogenetic tree of family-level relationships of bats. The topology was inferred by MrBayes based on the PR matrix. Multiple branches in each family are collapsed to display the family-level relationships. Bold numbers on the nodes indicate the molecular dates in millions of years. Values in the intervals and the blue bars represent the 95% credibility interval of divergence time estimates. Numbers in parentheses after each family name indicate the number of species selected in this study for each family. The most recent common ancestors of Yinpterochiroptera and Yangochiroptera are marked with a red dot, respectively.

Matrix	Hypothesis	logL	deltaL	bp-RELL	<i>Р-</i> КН	P-SH	P-WKH	P-WSH	c-ELW	P-AU
PR	A1	-263734.18	2.722	0.212 +	0.275	0.391 +	0.275 +	0.435 +	0.22 +	0.35 +
	A2	-263731.46	0	0.642 +	0.726 +	1 +	0.726 +	0.817 +	0.62 +	0.745 +
	A3	-263734.79	3.331	0.146 +	0.233 +	0.331 +	0.233 +	0.382 +	0.16 +	0.247 +
P12R	A1	-137990.02	0	0.459 +	0.513 +	1 +	0.513 +	0.652 +	0.444 +	0.559 +
	A2	-137991.82	1.805	0.114 +	0.271 +	0.451 +	0.271 +	0.454 +	0.14 +	0.205 +
	A3	-137990.12	0.099	0.427 +	0.487 +	0.634 +	0.487 +	0.635 +	0.416 +	0.574 +
PR	B1	-263733.41	2.366	0.383 +	0.397 +	0.507 +	0.397 +	0.539 +	0.384 +	0.418 +
	B2	-263741.31	10.262	0.0031 -	0.0628 +	0.104 +	0.0628 +	0.122 +	0.00503 -	0.00429 -
	B3	-263731.04	0	0.614 +	0.603 +	1 +	0.603 +	0.719 +	0.611 +	0.615 +
P12R	B1	-137994.61	6.275	0.219 +	0.224 +	0.268 +	0.224 +	0.355 +	0.222 +	0.247 +
	B2	-137998.62	10.29	0.0188 -	0.0776 +	0.105 +	0.0776 +	0.156 +	0.0231 -	0.0465 -
	В3	-137988.33	0	0.762 +	0.776 +	1 +	0.776 +	0.847 +	0.755 +	0.818 +

Table 1 Topology tests of different alternative topologies with PR and P12R matrices

Hypotheses A1, A2, A3, B1, B2, and B3 are also shown in Fig. 2. Assessment of conflicting tree topologies using Kishino–Hasegawa (KH), Shimodaira–Hasegawa (SH), and approximately unbiased (AU) (Shimodaira 2002) tests. deltaL, logL difference from the maximal logL in the set; bp-RELL, bootstrap proportion using the RELL method (Kishino *et al.* 1990); *P*-KH, *P*-value of KH test; *P*-SH, *P*-value of SH test; *P*-WKH, *P*-value of weighted KH test; *P*-WSH, *P*-value of weighted SH test; c-ELW, expected likelihood weight (Strimmer & Rambaut 2002); *P*-AU, *P*-value of AU test. Plus signs (+) and minus signs (–) denote the 95% confidence sets and significant exclusion, respectively. All tests performed 10 000 resamplings using the RELL method.

values at the nodes compared to other trees (Dataset S2, Supporting Information), we have chosen a single tree (PR matrix\_BI, Fig. 2a) as the optimal phylogenetic hypothesis.

Twenty-one lineages in the phylogenetic tree correspond to 21 bat families (Fig. 3). Almost all inter-familial nodes in the phylogenetic tree are strongly supported. Compared to the results of Amador et al. (2018) with most unresolved nodes, our integrative analyses obtained a more robust phylogenetic framework with improved phylogenetic signals and a lower level of missing data (Fig. 3). Compared to the results of many studies (Meredith et al. 2011; Amador et al. 2018; Álvarez-Carretero et al. 2022), our results do not agree with their placement of Emballonuroidea. Conversely, we are generally more supportive of putting this superfamily as the basal lineage of Yangochiroptera, which is a sister group to the remaining Yangochiroptera taxa sensu Teeling et al. (2018). In addition, we contend that the placement of Myzopodidae as the sister group to Nycteridae (Amador et al. 2018) (Fig. 1c) lacks sufficient resolution, which has been significantly excluded in our topology tests (Fig. 2h,i; Table 1). Our results instead support that Myzopodidae should be placed as a basal lineage of Emballonuroidea and a sister to the clade consisting of Nycteridae and Emballonuridae (BPP = 0.953, Fig. 3). These phylogenetic inconsistencies may be attributed to distinct levels of missing data between molecular datasets. In addition, our results are consistent with those obtained from the available whole genomic data (Foley et al. 2023), although they cannot cover all 21 families due to the lack of genome data for some families. However, in either this study or previous studies, it is difficult to avoid phylogenetic errors (for instance, incomplete lineage sorting or potential long branch attraction) based on current concatenated super-matrix (or multiple sub-matrices). Thus, major challenges remain for accurate phylogenetic reconstructions using concatenated few-gene datasets. Despite the Bat1K Project being launched in 2018 (Teeling et al. 2018), phase 1 of the project has not been completed so far, that is, sequencing representative species in each of the 21 bat families. The scarcity of genomic data for key taxa remains the biggest issue, which makes the real evolutionary histories of some taxa confusing.

The dated phylogenetic tree of Chiroptera based on the PR matrix using seven fossil calibration points is shown in Fig. 3 and Fig. S2, Supporting Information. Divergence time estimates showed that bats diverged approximately 61.4 Ma (65.7–56.5 Ma, 95% highest posterior density [HPD]) at Paleocene after the Cretaceous-Paleogene (K-Pg, 66 Ma) boundary. The divergence times are similar to the estimates of Amador *et al.* (2018) and 121/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley com/doi/10.1111/1749-4877.1272 by Wham University, Wiley Online Library on [27/10/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Álvarez-Carretero *et al.* (2022), but slightly more recent than those of Teeling *et al.* (2005) and Meredith *et al.* (2011). Moreover, Yinpterochiroptera diversified at 54.8 Ma (60.2–49.1 Ma, 95% HPD), whereas Yangochiroptera diversified at 58.9 Ma (61.4–54.1 Ma, 95% HPD). Together, this study proposes a robust molecular phylogeny of all 21 families in bats with a lower level of missing data and provides divergence times for each bat family. The puzzling relationships within the suborder Yangochiroptera call for in-depth phylogenetic studies of these bats in the future.

# ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (32270436).

# **CONFLICT OF INTEREST**

The authors declare no conflict of interest.

# REFERENCES

- Agnarsson I, May-Collado LJ (2008). The phylogeny of Cetartiodactyla: The importance of dense taxon sampling, missing data, and the remarkable promise of cytochrome b to provide reliable species-level phylogenies. *Molecular Phylogenetics and Evolution* **48**, 964– 85.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215, 403–10.
- Álvarez-Carretero S, Tamuri AU, Battini M *et al.* (2022). A species-level timeline of mammal evolution integrating phylogenomic data. *Nature* **602**, 263–67.
- Amador LI, Arévalo RLM, Almeida FC, Catalano SA, Giannini NP (2018). Bat systematics in the light of unconstrained analyses of a comprehensive molecular supermatrix. *Journal of Mammalian Evolution* 25, 37– 70.
- Birney E, Clamp M, Durbin R (2004). GeneWise and Genomewise. *Genome Research* 14, 988–95.
- Breinholt JW, Kawahara AY (2013). Phylotranscriptomics: Saturated third codon positions radically influence the estimation of trees based on next-gen data. *Genome Biology and Evolution* **5**, 2082–92.
- Castresana J (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**, 540–552.

- Foley NM, Mason VC, Harris AJ *et al.* (2023). A genomic timescale for placental mammal evolution. *Science* **380**, eabl8189.
- Foley NM, Thong VD, Soisook P *et al.* (2015). How and why overcome the impediments to resolution: Lessons from rhinolophid and hipposiderid bats. *Molecular Biology and Evolution* **32**, 313–33.
- Goldman N, Anderson JP, Rodrigo AG (2000). Likelihood-based tests of topologies in phylogenetics. *Systematic Biology* **49**, 652–70.
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 35, 518–22.
- Hoofer SR, Van Den Bussche RA (2003). Molecular phylogenetics of the chiropteran family Vespertilionidae. *Acta Chiropterologica* **5**(suppl), 1–63.
- Jin JJ, Yu WB, Yang JB *et al.* (2020). GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* **21**, 241.
- Katoh K, Misawa K, Kuma K, Miyata T (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059–66.
- Katoh K, Standley DM (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772–80.
- Kishino H, Hasegawa M (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *Journal of Molecular Evolution* **29**, 170–79.
- Kishino H, Miyata T, Hasegawa M (1990). Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *Journal of Molecular Evolution* **31**, 151–60.
- Kück P, Meid SA, Groß C, Wägele JW, Misof B (2014). AliGROOVE–visualization of heterogeneous sequence divergence within multiple sequence alignments and detection of inflated branch support. *BMC Bioinformatics* 15, 294.
- Lack JB, Roehrs ZP, Stanley CE Jr, Ruedi M, Van Den Bussche RA (2010). Molecular phylogenetics of *My*otis indicate familial-level divergence for the genus *Cistugo* (Chiroptera). Journal of Mammalogy **91**, 976– 92.
- Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B (2017). PartitionFinder 2: New methods for

<sup>© 2023</sup> International Society of Zoological Sciences, Institute of Zoology/ Chinese Academy of Sciences and John Wiley & Sons Australia, Ltd.

selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* **34**, 772–73.

- Lartillot N, Lepage T, Blanquart S (2009). PhyloBayes
  3: A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25, 2286–88.
- Lartillot N, Rodrigue N, Stubbs D, Richer J (2013). PhyloBayes MPI: Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Systematic Biology* **62**, 611–15.
- 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. *Systematic Biology* **58**, 130–45.
- Lu Q, Jiao H, Wang Y, Norbu N, Zhao H (2021). Molecular evolution and deorphanization of bitter taste receptors in a vampire bat. *Integrative Zoology* **16**, 659–69.
- Meredith RW, Janečka JE, Gatesy J *et al.* (2011). Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* **334**, 521–24.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**, 268–74.
- Philippe H, Snell EA, Bapteste E, Lopez P, Holland PW, Casane D (2004). Phylogenomics of eukaryotes: Impact of missing data on large alignments. *Molecular Biology and Evolution* **21**, 1740–52.
- Ramírez-Fráncel LA, García-Herrera LV, Losada-Prado S *et al.* (2021). Bats and their vital ecosystem services: A global review. *Integrative Zoology* **17**, 2–23.
- Ronquist F, Teslenko M, Van Der Mark P *et al.* (2012).
  MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**, 539–42.
- Sharnuud R, Ameca EI (2023). Taxonomy, distribution, and contemporary exposure of terrestrial mammals to floods and human pressure across different areas for biodiversity conservation in China. *Integrative Zoology*, https://doi.org/10.1111/1749-4877.12753
- Shimodaira H (2002). An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* **51**, 492–508.
- Shimodaira H, Hasegawa M (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16**, 1114.

- Simmons MP (2012). Radical instability and spurious branch support by likelihood when applied to matrices with non-random distributions of missing data. *Molecular Phylogenetics and Evolution* **62**, 472–84.
- Strimmer K, Rambaut A (2002). Inferring confidence sets of possibly misspecified gene trees. *Proceedings of the Royal Society of London: Biological Sciences* **269**, 137–42.
- Strimmer K, von Haeseler A (1997). Likelihoodmapping: A simple method to visualize phylogenetic content of a sequence alignment. *PNAS* **94**, 6815– 6819.
- Teeling EC, Scally M, Kao DJ, Romagnoli ML, Springer MS, Stanhope MJ (2000). Molecular evidence regarding the origin of echolocation and flight in bats. *Nature* 403, 188–92.
- Teeling EC, Springer MS, Madsen O, Bates P, O'brien SJ, Murphy WJ (2005). A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* 307, 580–84.
- Teeling EC, Vernes SC, Dávalos LM *et al.* (2018). Bat biology, genomes, and the Bat1K project: To generate chromosome-level genomes for all living bat species. *Annual Review of Animal Biosciences* **6**, 23–46.
- Van Den Bussche RA, Hoofer SR (2004). Phylogenetic relationships among recent chiropteran families and the importance of choosing appropriate out-group taxa. *Journal of Mammalogy* **85**, 321–30.
- Wiens JJ (2006). Missing data and the design of phylogenetic analyses. *Journal of Biomedical Informatics* **39**, 34–42.
- Wiens JJ, Morrill MC (2011). Missing data in phylogenetic analysis: Reconciling results from simulations and empirical data. *Systematic Biology* **60**, 719–31.
- Wilson DE, Mittermeier RA (2020). Handbook of the Mammals of the World, *Volume 9:* Bats. Lynx Edicions, Barcelona, Spain.
- Xi Z, Liu L, Davis CC (2016). The impact of missing data on species tree estimation. *Molecular Biology and Evolution* **33**, 838–60.
- Xiang CY, Gao F, Jakovlić I *et al.* (2023). Using Phylo-Suite for molecular phylogeny and tree-based analyses. *iMeta* **2**, e87.
- Yang Z (1996). Among-site rate variation and its impact on phylogenetic analyses. *Trends in Ecology & Evolution* **11**, 367–72.
- Yang Z (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24, 1586–91.

<sup>© 2023</sup> International Society of Zoological Sciences, Institute of Zoology/ Chinese Academy of Sciences and John Wiley & Sons Australia, Ltd.

Zhang D, Gao F, Jakovlić I *et al.* (2020). PhyloSuite: An integrated and scalable desktop platform for stream-linedmolecular sequence data management and evolutionary phylogenetics studies. *Molecular Ecology Resources* **20**, 348–55.

### SUPPLEMENTARY MATERIALS

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

**Figure S1** Heterogeneous sequence divergences of different sub-datasets. The mean similarity score between alignments is represented by a colored square based on AliGROOVE scores, ranging from -1 to +1, which indicate full random similarity (i.e. heterogeneity, red coloring) to nonrandom similarity (blue coloring).

**Figure S2** Time-calibrated phylogenetic tree based on the topology obtained from PR matrix (BI). Bold numbers indicate the molecular dates in millions of years. Values in the intervals and the blue bars represent the 95% credibility interval of divergence time estimates.

**Table S1** GenBank accession numbers of sequences of154 species used in this study

 Table S2 Best partitioning schemes and models based

 on different datasets for ML and BI analyses

**Table S3** Seven fossil constraints used in the estimation of divergence time sensu Meredith *et al.* (2011) and its references

**Table S4** Topology test of different two topologies (H1and H2) with the matrix PR and P12R

**Dataset S1** Newly extracted/assembled supplementary sequences in this study

Dataset S2 Supplementary tree files in Newick format

#### Cite this article as:

Hao X, Lu Q, Zhao H (2023). A molecular phylogeny for all 21 families within Chiroptera (bats). *Integrative Zoology* **00**, 1–10. https://doi.org/10.1111/1749-4877.12772