

SCIENTIFIC REPORTS



OPEN

Phylogenomic analyses of bat subordinal relationships based on transcriptome data

Ming Lei & Dong Dong

Received: 05 January 2016

Accepted: 24 May 2016

Published: 13 June 2016

Bats, order Chiroptera, are one of the largest monophyletic clades in mammals. Based on morphology and behaviour bats were once differentiated into two suborders Megachiroptera and Microchiroptera. Recently, researchers proposed alternative views of chiropteran classification (suborders Yinpterochiroptera and Yangochiroptera) based on morphological, molecular and fossil evidence. Since genome-scale data can significantly increase the number of informative characters for analysis, transcriptome RNA-seq data for 12 bat taxa were generated in an attempt to resolve bat subordinal relationships at the genome level. Phylogenetic reconstructions were conducted using up to 1470 orthologous genes and 634,288 aligned sites. We found strong support for the Yinpterochiroptera-Yangochiroptera classification. Next, we built expression distance matrices for each species and reconstructed gene expression trees. The tree is highly consistent with sequence-based phylogeny. We also examined the influence of taxa sampling on the performance of phylogenetic methods, and found that the topology is robust to sampling. Relaxed molecular clock estimates the divergence between Yinpterochiroptera and Yangochiroptera around 63 million years ago. The most recent common ancestor of Yinpterochiroptera, corresponding to the split between Rhinolophoidea and Pteropodidae (Old World Fruit bats), is estimated to have occurred 60 million years ago. Our work provided a valuable resource to further explore the evolutionary relationship within bats.

Bats belong to the order Chiroptera, one of the largest monophyletic clades in mammals. They constitute ~20% of living mammalian species, arranged in 20 families¹. Their wings make bats distinct among mammals². Living bats had been placed in one of two suborders based on morphology and behaviour³. All bats that produce echolocation calls in their larynges were placed in the suborder Microchiroptera⁴. All other bats were placed in the suborder Megachiroptera (Old World fruit bats, non-echolocating bats). However, a reconsideration of morphological, behavioural and molecular evidence demonstrates that there are two suborders of bats, Yinpterochiroptera and Yangochiroptera that do not coincide with the previous subordinal classification^{5,6}. The two new suborders are strongly supported by statistical tests. Phylogenomic analysis based on genome sequencing data support the classification of living bats in Yinpterochiroptera and Yangochiroptera⁷.

Recently, O'Leary *et al.* claimed that living echolocating bats were monophyletic⁸. They based this on morphological data set and published molecular sequence data. Although the bat genome data set is rich in the number of loci, it is not comprehensive in taxon sampling, an important component for accurately estimating phylogeny⁹. We investigated the evolutionary relationships of bats based on more taxa at the genome level. Because regulatory changes affecting gene expression might explain many or even most phenotypic differences between species¹⁰, we made between-species comparisons at sequence and expression levels. We generated transcriptome data for 12 bat taxa and used data from two published bat genomes¹¹. The bat transcriptome data we present is largely expanded the coverage across the bat clade. After evaluating the influence of taxa sampling on the performance of phylogenetic methods, we found strong support for the Yinpterochiroptera-Yangochiroptera classification. Furthermore, the expression-based tree is consistent with sequence-based phylogeny. These results provided a phylogenetic framework and timescale with which to interpret the evolution of bats.

Laboratory of Molecular Ecology and Evolution, Institute of Estuarine and Coastal Research, East China Normal University, Shanghai, 200062, China. Correspondence and requests for materials should be addressed to D.D. (email: ddong.ecnu@gmail.com)

Species	Family	Genus	No. reads	No. contigs	N50 (bp)	Mean (bp)
<i>Cynopterus sphinx</i>	Pteropodidae	Cynopterus	34084746	82348	1722	940
<i>Rousettus leschenaultii</i>	Pteropodidae	Rousettus	36608760	96717	1924	993
<i>Aselliscus stoliczkanus</i>	Hipposideridae	Aselliscus	59114136	229392	2912	1294
<i>Hipposideros pratti</i>	Hipposideridae	Hipposideros	65019524	254130	2895	1192
<i>Rhinolophus pusillus</i>	Rhinolophidae	Rhinolophus	60320924	237672	2819	1192
<i>Megaderma lyra</i>	Megadermatidae	Megaderma	57258600	154630	2117	1008
<i>Rhinopoma hardwickei</i>	Rhinopomatidae	Rhinopoma	61182790	163706	2929	1309
<i>Taphozous melanopogon</i>	Emballonuridae	Taphozous	55846398	123339	1666	831
<i>Tadarida teniotis</i>	Molossidae	Tadarida	54692960	153967	2683	1144
<i>Murina leucogaster</i>	Vespertilionidae	Murina	40919120	200255	2574	1119
<i>Myotis ricketti</i>	Vespertilionidae	Myotis	38235688	82834	1745	945
<i>Scotophilus kuhlii</i>	Vespertilionidae	Scotophilus	49022368	214832	2291	1040

Table 1. Global statistic of *de novo* brain transcriptome assembles.

Results

Transcriptome sequencing and assembly. In this work, we sequenced cDNAs from 12 bat species and generated 610 million raw reads (Table 1). After sequencing reads filter steps, a total of 595 million clean reads was obtained. Next, the clean reads were *de novo* assembled using trinity package¹², and a summary of the assembly statistics is shown in Table 1. We excluded all contigs less than 200 bp from further analysis, and finally obtained a total of 1,993,822 contigs (82348–254130 per sample). Next, we performed redundancy reduction on the raw assemblies, processing them to identify candidate open reading frames (ORF) within the transcripts. We chose the longest ORF was chosen and selected one peptide per putative unigene. At last, 32,227–51,271 we retained peptide sequences per taxon

Phylogenomics analyses and molecular dating. Along with 18 previously published mammalian genome sequences, a total of 1470 1:1 orthologous was obtained across 30 species (14 bat species and 16 other mammals). We performed multiple sequence alignments and our aligned supermatrix included 634,288 amino acids. We first used concatenated nucleotide and protein sequences using maximum likelihood to reconstruct the phylogeny. As with previous works¹³, Yinpterochiroptera and Yangochiroptera received 100% bootstrap support based on the nucleotide and protein supermatrices (Fig. 1). The rhinolophoid bats are sister group to the Old World fruit bats within Yinpterochiroptera. This result is also strongly supported by coalescent analyses (Fig. S1). All relationships within Yinpterochiroptera and Yangochiroptera were congruent with previous study¹³ with 100% bootstrap support (Fig. 1). Next, we measured the gene expression phylogeny within bats. We built expression distance matrices for each species and reconstructed gene expression trees. As shown in Fig. 2, the gene expression-based tree is highly consistent with the sequence-based phylogeny.

We obtained a disagreement using nucleotide and amino acid sequences when addressing the position of bats within the superorder Laurasiatheria. The nucleotide tree recovered Pegasoferae group (Chiroptera + Perissodactyla with 95% bootstrap value support), whereas amino acid tree supported that bats are a sister group to Fereuungulata group (Carnivores + Perissodactyla + Cetariodactyla with 100% bootstrap value support) (Fig. S2). Previous works have published eight proposed higher clades within Laurasiatheria (Fig. 3). To dissect the phylogenetic signal, we measured the relative support of each locus for the evolutionary relationships of bats. The approximately unbiased (AU) test statistics analyses of eight potential topologies suggested that all Microchiroptera-Megachiroptera topologies were significantly rejected (P -value < 0.05, Table 2), and four potential Yinpterochiroptera-Yangochirptera topologies could not be significantly rejected.

To evaluate the influence of the taxa sampling on phylogenetic reconstruction, analyses were performed from different subsets of taxa. We constructed concatenated trees for different taxa sets that include at least two Old World fruit bat, two rhinolophoid bats and two Yangochiroptera. Phylogeny analyses assigned high support (bootstrap value >90%) based on different sampling datasets. As shown in Fig. 4, both the results of concatenation and coalescence analyses give consistent phylogenetic estimation of relationships and support the Yinpterochiroptera-Yangochirptera topologies (concatenated method: 2% for nucleotide and 15.5% for amino acids; coalescent method: 5% for nucleotide and 12% for amino acids, respectively).

Results of divergence time estimation carried out under the auto-correlated models of molecular clock relaxation are shown in Table 3. The results varied depending on the models and datasets used, but were nevertheless consistent between PhyloBayes and MCMCTree approaches. The result of MCMCTree WAG + G model is consistent with previous findings¹³ and values from TimeTree database¹⁴. We considered that the estimation of divergence time obtained from this model are the most reliable. We estimated the origin of Chiroptera at 63 Myr (million years ago), following the Cretaceous-Tertiary boundary. The divergence of Pteropodidae and Rhinolophoidea was estimated to be 60 Myr, earlier than previous suggestions¹³. We estimated the most recent common ancestor of Rhinolophidae and Hipposideridae at 40 Myr.

Discussion

The development of next-generation sequencing technologies have generated many more genome sequences. This has allowed reconstruction of phylogenetic trees based on genome scale data providing a powerful approach to

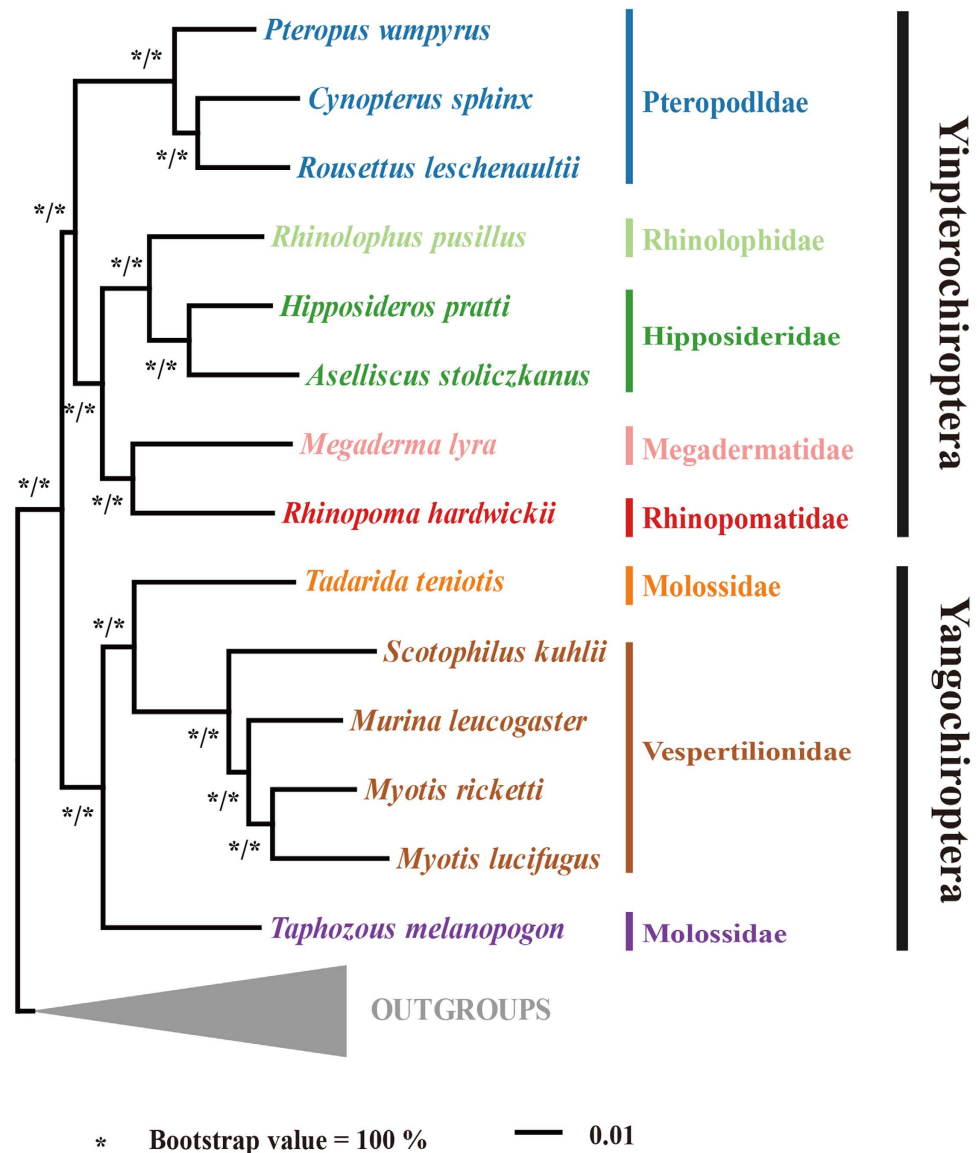


Figure 1. Maximum likelihood phylogenetic tree for nucleotide and amino acid datasets with bootstrap support values (1000 replicates) under partition model. The asterisks indicate 100/100 bootstrap support for nucleotide and amino acid datasets respectively.

resolve the evolutionary relationships^{15,16}. Transcriptome data are often used in the phylogenomic analyses studies for those non-model organisms without genome information^{17–19}. RNA-seq technology is a high-efficiency way to obtain full-length coding sequence at a lower cost^{12,20}. In this study, we sequenced large-scale whole brain transcriptome data for 12 bats and presented a large-scale, phylogenomic perspective to resolving the backbone phylogeny of bats using a larger taxon set.

Variation in gene expression and protein sequence can both influence phenotype²¹, and a better understanding of the evolutionary relationship between gene expression and protein sequence may provide great insights into the processes that ultimately contribute to phenotypic diversification. Both expression and sequence-based phylogeny support the Yinpterochiroptera-Yangochiroptera subdivision, while rhinolophoid bats and Old World fruit bats form a monophyletic group. Studies have proved that taxon sampling is an important way for accurately assessing phylogenies^{22,23}. Further improved taxon sampling gives a consistent phylogenetic estimation of relationship, and only few misleading phylogenies were generated. One caveat of our work is that no bats belonging to Noctilionoidea were included. Further analyses with the brain transcriptome data of bats belong to Noctilionoidea is needed.

The conflict of ‘species tree’ and ‘gene tree’ challenges the traditional methodology of molecular phylogeny. Between-genes phylogenetic incongruences can arise for several reasons, involving convergent evolution in Microchiroptera. The homoplastic signal in morphology within echolocating bats can also be reflected from molecular evidences. For example, the ‘hearing gene’ *Prestin* has been shown to have undergone sequence convergence among echolocating mammals. Recent genome analysis demonstrated that adaptive convergences are

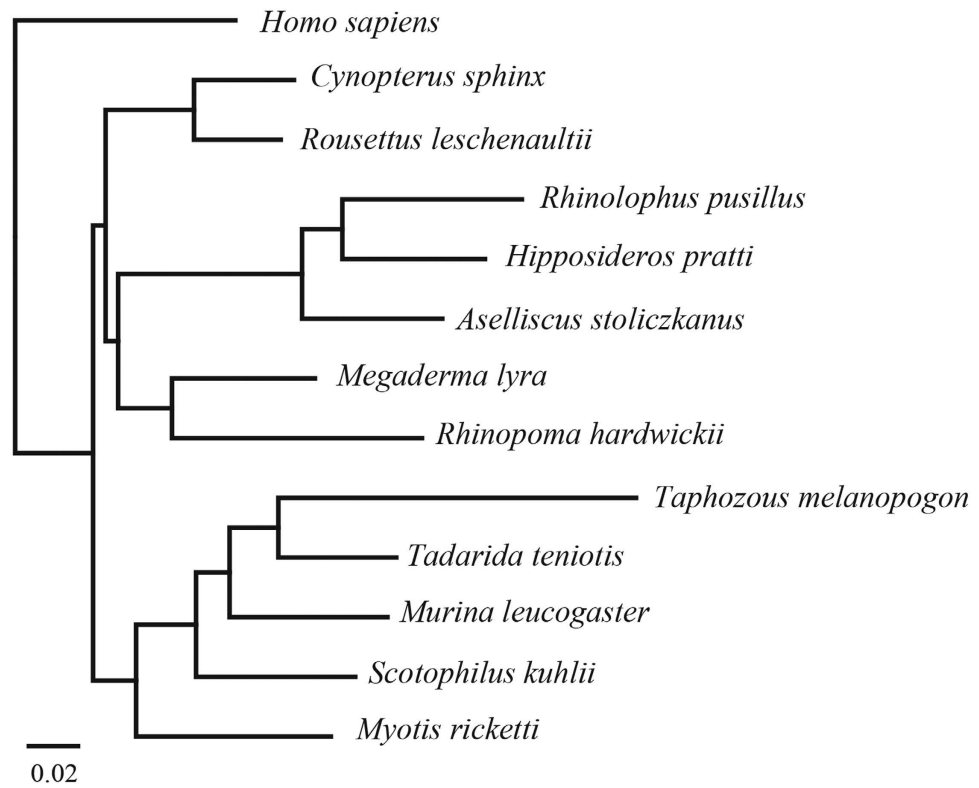


Figure 2. Bat gene expression phylogeny. Neighbor-joining tree based on pairwise distance matrices for brain transcriptome.

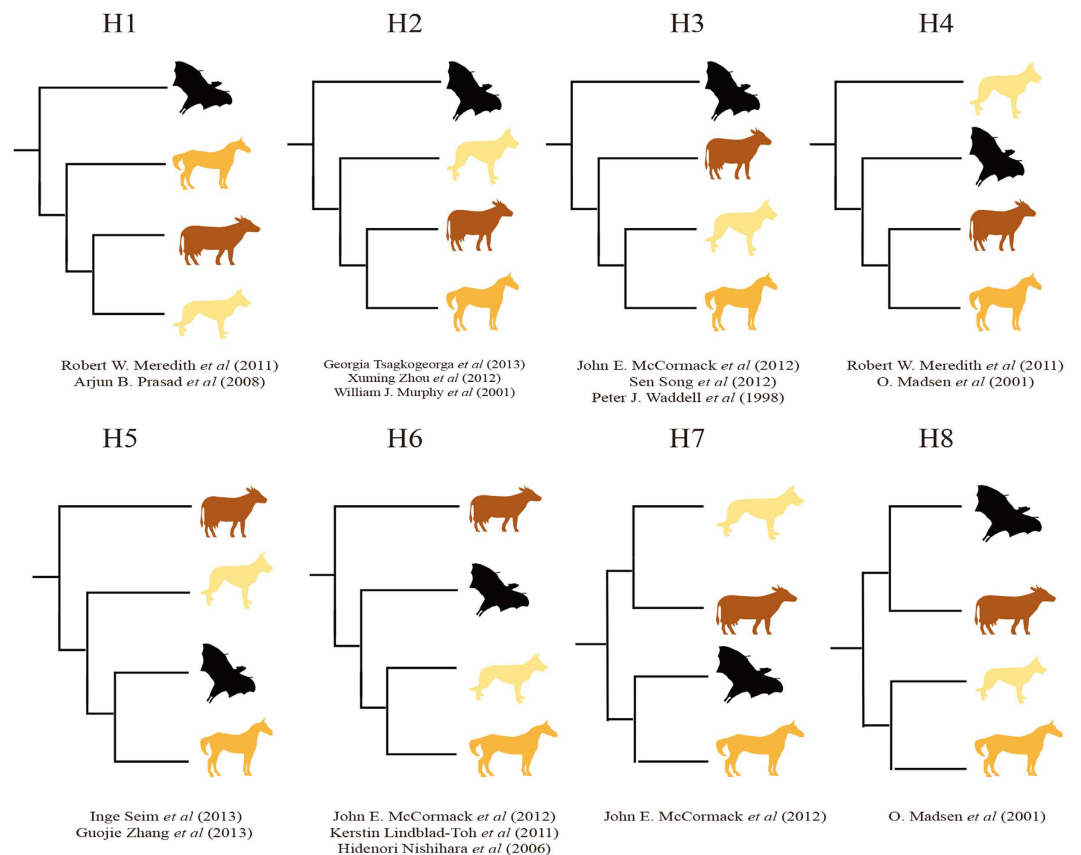


Figure 3. Eight proposed species tree topologies differing in the position of bats within mammals.

Bat subdivision	AU P value (nucleotide)	AU P value (amino acids)	Laurasiathia topology
Yinpterochiroptera–Yangochiroptera	0.815	0.443	(Eulipotyphyla, ((Cetartiodactyla, Carnivora), (Perissodactyla, Chiroptera)))
	0.316	0.241	(Eulipotyphyla, (Cetartiodactyla, (Chiroptera, (Perissodactyla, Carnivora))))
	0.184	0.768	(Eulipotyphyla, (Chiroptera, (Perissodactyla, (Cetartiodactyla, Carnivora))))
	0.071	0.109	(Eulipotyphyla, (Carnivora, (Chiroptera, (Perissodactyla, Cetartiodactyla))))
	0.029	0.451	(Eulipotyphyla, (Chiroptera, (Carnivora, (Perissodactyla, Cetartiodactyla))))
	0.009	0.002	(Eulipotyphyla, (Cetartiodactyla, (Carnivora, (Perissodactyla, Chiroptera))))
	0.008	0.272	(Eulipotyphyla, (Chiroptera, (Cetartiodactyla, (Carnivora, Perissodactyla))))
	0.000	0.000	((Eulipotyphyla, Chiroptera), Cetartiodactyla, (Carnivora, Perissodactyla))
Microchiroptera–Megachiroptera	0.031	0.000	(Eulipotyphyla, (Chiroptera, (Perissodactyla, (Cetartiodactyla, Carnivora))))
	0.008	0.000	((Eulipotyphyla, Chiroptera), Cetartiodactyla, (Carnivora, Perissodactyla))
	0.001	0.001	(Eulipotyphyla, ((Cetartiodactyla, Carnivora), (Perissodactyla, Chiroptera)))
	0.001	0.000	(Eulipotyphyla, (Chiroptera, (Carnivora, (Perissodactyla, Cetartiodactyla))))
	0.000	0.001	(Eulipotyphyla, (Cetartiodactyla, (Carnivora, (Perissodactyla, Chiroptera))))
	0.000	0.000	(Eulipotyphyla, (Cetartiodactyla, (Chiroptera, (Perissodactyla, Carnivora))))
	0.000	0.000	(Eulipotyphyla, (Carnivora, (Chiroptera, (Perissodactyla, Cetartiodactyla))))
	0.000	0.000	(Eulipotyphyla, (Chiroptera, (Cetartiodactyla, (Carnivora, Perissodactyla))))

Table 2. Approximately Unbiased test for 16 hypothetical species topologies.

widespread at both the molecular and morphological level²⁴. Although sequence convergence is traditionally considered to be rare, our data provided a large resource to decipher genome-wide sequence convergence within bats.

Taken together, our work generated large-scale transcriptomes of bats and analyzed bat subordinal relationships at genome-wide scale. With increased taxa sampling and sufficient numbers of loci, we can obtain a more reliable phylogeny. These data also provided valuable information for further researches related to molecular evolutionary analyses.

Materials and Methods

Ethics Statement. All animal experiments were carried out in strict accordance with the regulations for the use of laboratory animals (Decree No. 2, State Science and Technology Commission, People's Republic of China, November 14, 1988) and were approved by the Animal Ethics Committee of East China Normal University (ID no: 20090219 and 20101002). No endangered or region-protected animal species were included, and no specific permission was required.

Sample collection and transcriptome sequencing. We collected adult specimens of 12 bat species covering 8 of the 18 extant chiropteran families, namely, three from the family Vespertilionidae (*Murina leucogaster*, *Myotis ricketti* and *Scotophilus kuhlii*), two from the family Pteropodidae (*Cynopterus sphinx* and *Rousettus leschenaultii*), two from the family Hipposideridae (*Aselliscus stoliczkanus* and *Hipposideros pratti*), one from the family Rhinolophidae (*Rhinolophus pusillus*), one from the family Megadermatidae (*Megaderma lyra*), one from the family Rhinopomatidae (*Rhinopoma hardwicki*), one from the family Emballonuridae (*Taphozous melanopogon*), and one from the family Molossididae (*Tadarida teniotis*) (Table S1). These bats were euthanized by halothane hyperanesthesia followed by thoracotomy. Efforts were devoted to minimize animal suffering. The whole brain tissues of these individuals were placed on ice immediately after sacrifice. All brain tissues were flash frozen in liquid nitrogen and kept at -80°C freezer until processed for total RNA isolation.

Total RNA was extracted from brain tissue using Trizol (Life Technologies Corp) according to manufacturer's protocols. RNA concentration was determined using a NanoDrop spectrophotometer, and RNA quality was assessed by Agilent Bioanalyzer. New sequence data of each bat brain were generated using the Illumina HiSeq2500 platform. Raw reads were deposited into the Short Read Archive (SRA) database of NCBI under the accession no. SRP062200.

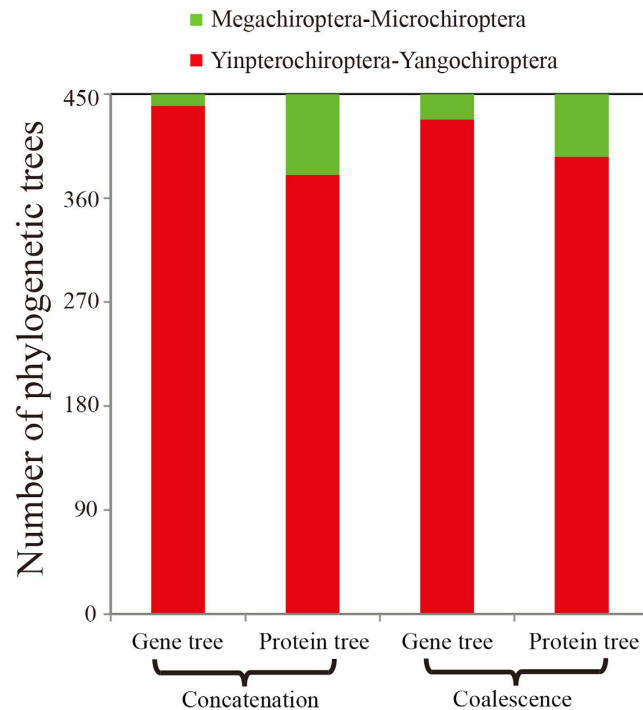


Figure 4. Influence of taxon sampling on the performance of phylogenetic trees. Both concatenated and coalescent analyses yield consistent phylogeny of Yinpterochiroptera-Yangochiroptera classification.

	Amino acids			Nucleotides			
	MCMCTree WAG + G (Million Years)	Phylobayes LG + G (Million Years)	Phylobayes CAT-GTR + G (Million Years)	MCMCTree GTR + G (Million Years)	Phylobayes GTR + G (Million Years)	Phylobayes CAT-GTR + G (Million Years)	TimeTree Mean/Median (Million Years)
MRCAs of Rhinolophidae + Hipposideridae	40 [26,46]	39 [29,49]	42 [34,51]	37 [27,47]	44 [34,54]	38 [28,49]	41.6/40.0
Molossidae/Vespertilionidae	48 [36,59]	53 [43,62]	49 [40,56]	49 [37,59]	57 [47,66]	46 [35,56]	45.5/48.0
Rhinolophidae + Hipposideridae/Megadermatidae + Rhinopomatidae	52 [37,60]	53 [43,63]	59 [50,67]	52 [40,62]	59 [49,69]	54 [42,65]	52.5/52.0
MRCAs of Yangochiroptera	54 [41,66]	59 [49,68]	57 [48,65]	56 [43,67]	70 [60,78]	61 [49,71]	55.3/55.0
Pteropodidae/Rhinolophoidea	60 [45,71]	64 [54,73]	63 [54,70]	62 [49,73]	70 [60,78]	61 [49,71]	58.5/58.0
MRCAs of Yinpterochiroptera and Yangochiroptera	63 [48,75]	67 [57,76]	64 [55,72]	67 [52,78]	73 [64,81]	63 [51,73]	62.6/63.9

Table 3. Results of Bayesian relaxed molecular clock analyses.

Quality control and *De novo* transcriptome assembly. For each paired-end library, we first removed the adapters of raw reads. Then, the DynamicTrim Perl script in the SolexQA package²⁵ was used to trim the poor quality positions of reads with parameters setting: ‘-b -h 15’. Next, the Trinity (version: trinityrnaseq_r20140413) software was used to *de novo* assemble the transcriptome of each tissue with default parameters¹². The program was run on 64-bit Linux system (Red Hat 6.0) with 256 internal memory. TransDecoder, a program nested in the trinity package, was then used to identify the candidate coding sequence (CDS) from the contigs. At last, the CD-Hit program was used to reduce sequence redundancy of coding sequences with at least 95% global similarity²⁶. All final CDSs with a length more than 200 bp were used for further analyses. Assembled contigs were deposited into transcriptome shotgun assembly (TSA) database of NCBI under the accession no. SRP062200.

Phylogenomic Analyses. Except for the 12 newly sequenced bat brain transcriptome data, we also downloaded the coding sequences of other 18 mammalian species (large flying fox, little brown bat, horse, rhinoceros, cow, pig, dolphin, dog, cat, hedgehog, shrew, mouse, rat, human, chimpanzee, elephant, armadillo, opossum) from Ensembl database (Table S1). Only sequences of coding regions with the length larger than 400 bp were retained for further analysis. To obtain the orthologous genes among all the species, all-against-all reciprocal blastp search was employed. For each orthologous gene, we extracted protein sequences and their corresponding coding sequences. Multiple alignments were performed using MAFFT software with default settings²⁷.

We generated two sequence super-matrices by concatenating aligned nucleotide and protein sequences of orthologous genes separately. For each matrix, we conducted the maximum likelihood analyses with model

partitioning. The nucleotides substitution model $GTR + \Gamma_4 + I$ was selected based on the result of ProTest with bootstrap analyses were replicated for 100 times using RAXML program (version: 8.0.20)²⁸.

We calculated standard RPKM expression values (that were then log₂-transformed) for the orthologous genes. We constructed expression tree using the neighbor-joining approach, based on pairwise distance matrices between samples. The distance between samples was computed as $1 - \rho$, where ρ is Spearman's correlation coefficient.

To compare the alternative topologies, approximately unbiased (AU) test statistic was computed using CONSEL package²⁹. Cumulative scores of support for eight previously published species trees were calculated by counting the number of loci supporting the phylogeny based on the AU *P-values* at a critical value $\alpha = 0.05$. In addition, we randomly sampled one species for each order within Laurasiatherian (Perissodactyla, Cetartiodactyla, Carnivora, Eulipotyphyla) and combined other species to generate nucleotide and protein sequences super-matrices. We repeatedly performed the same phylogenetic tree reconstruction and AU-test analyses for these matrices.

Coalescent-based analyses were performed using ASTRAL method³⁰ with the neighbor-joining algorithm on a matrix of ranks of taxon pairs in the gene trees under a $GTR + \Gamma$ model. This analysis were replicated for 100 time using Phybase³¹.

Taxon sampling. To evaluate the influence of taxa sampling on the phylogeny, several analyses were performed from different subset of taxa using concatenated genes. Because basal lineages of major groups are crucial for phylogeny reconstruction, at least two Old World fruit bats, two rhinolophoid bats and two Yangochiroptera from the taxon set were included for each sampling analysis.

Molecular dating analyses. The divergence time was estimated using the Bayesian relaxed molecular clock approaches implemented in PhyloBayes, MCMCTree nested in PAML package³². With PhyloBayes approach, the CAT + GTR + G4 mixture model (for amino acid and nucleotide sequences), LG + G (amino acid sequences) and GTR + G (nucleotide sequences) models were employed. With MCMCTree, the standard WAG + G (amino acid sequences) and GRT + G (nucleotide sequences) models were used. For the PhyloBayes, all analyses were performed by running two independent MCMC chains from a random tree for 20,000 cycles, sampling posterior rates and dates every 10 cycles until 2,000 points were obtained. Posterior estimation of divergence times were estimated from the last 1800 samples of each chain after discarding the initial 10% burn-in periods within each MCMC run. For the MCMCTree, the first 1,000,000 replicates were discarded as burn-in, and the MCMC was run for 100,000,000 replicates, with the sampling frequency of 100 iterations.

References

- Wilson, D. E. & Reeder, D. M. *Mammal species of the world : a taxonomic and geographic reference*. 3rd edn, (Johns Hopkins University Press, 2005).
- Altringham, J. D. *Bats: biology and behaviour*. (Oxford University Press, 1996).
- Gunnell, G. & Simmons, N. Fossil Evidence and the Origin of Bats. *J Mammal Evol* **12**, 209–246, doi: 10.1007/s10914-005-6945-2 (2005).
- Hill, J. E. & Smith, J. D. *Bats : a natural history*. 1st edn, (University of Texas Press, 1984).
- Springer, M. S., Teeling, E. C., Madsen, O., Stanhope, M. J. & de Jong, W. W. Integrated fossil and molecular data reconstruct bat echolocation. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 6241–6246, doi: 10.1073/pnas.111551998 (2001).
- Teeling, E. C. *et al.* Microbat paraphyly and the convergent evolution of a key innovation in Old World rhinolophoid microbats. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 1431–1436, doi: 10.1073/pnas.022477199 (2002).
- Tsagkogeorga, G., Parker, J., Stupka, E., Cotton, J. A. & Rossiter, S. J. Phylogenomic analyses elucidate the evolutionary relationships of bats. *Current biology : CB* **23**, 2262–2267, doi: 10.1016/j.cub.2013.09.014 (2013).
- O'Leary, M. A. *et al.* The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science* **339**, 662–667, doi: 10.1126/science.1229237 (2013).
- Pollock, D. D., Zwickl, D. J., McGuire, J. A. & Hillis, D. M. Increased taxon sampling is advantageous for phylogenetic inference. *Systematic biology* **51**, 664–671, doi: 10.1080/10635150290102357 (2002).
- King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
- Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482, doi: 10.1038/nature10530 (2011).
- Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **29**, 644–652, doi: 10.1038/nbt.1883 (2011).
- Teeling, E. C. *et al.* A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* **307**, 580–584, doi: 10.1126/science.1105113 (2005).
- Kumar, S. & Hedges, S. B. TimeTree2: species divergence times on the iPhone. *Bioinformatics* **27**, 2023–2024, doi: 10.1093/bioinformatics/btr315 (2011).
- Rannala, B. & Yang, Z. Phylogenetic inference using whole genomes. *Annual review of genomics and human genetics* **9**, 217–231, doi: 10.1146/annurev.genom.9.081307.164407 (2008).
- Song, S., Liu, L., Edwards, S. V. & Wu, S. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 14942–14947, doi: 10.1073/pnas.1211733109 (2012).
- Smith, S. A. *et al.* Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature* **480**, 364–367, doi: 10.1038/nature10526 (2011).
- Cannon, J. T. *et al.* Phylogenomic resolution of the hemichordate and echinoderm clade. *Current biology : CB* **24**, 2827–2832, doi: 10.1016/j.cub.2014.10.016 (2014).
- Wickett, N. J. *et al.* Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E4859–E4868, doi: 10.1073/pnas.1323926111 (2014).
- Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics* **10**, 57–63, doi: 10.1038/nrg2484 (2009).
- Carroll, S. B. Evolution at two levels: on genes and form. *PLoS biology* **3**, e245, doi: 10.1371/journal.pbio.0030245 (2005).

22. Teeling, E. C. *et al.* Molecular evidence regarding the origin of echolocation and flight in bats. *Nature* **403**, 188–192, doi: 10.1038/35003188 (2000).
23. Simmons, N. B. & Geisler, J. H. Phylogenetic relationships of Icaronycteris, Archaeonycteris, Hassianycteris, and Palaeochiropteryx to extant bat lineages, with comments on the evolution of echolocation and foraging strategies in Microchiroptera. *B Am Mus Nat Hist*, 4–182 (1998).
24. Parker, J. *et al.* Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* **502**, 228–231, doi: 10.1038/nature12511 (2013).
25. Cox, M. P., Peterson, D. A. & Biggs, P. J. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC bioinformatics* **11**, 485, doi: 10.1186/1471-2105-11-485 (2010).
26. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659, doi: 10.1093/bioinformatics/btl158 (2006).
27. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**, 772–780, doi: 10.1093/molbev/mst010 (2013).
28. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313, doi: 10.1093/bioinformatics/btu033 (2014).
29. Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Systematic biology* **51**, 492–508, doi: 10.1080/10635150290069913 (2002).
30. Mirarab, S. *et al.* ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**, i541–548, doi: 10.1093/bioinformatics/btu462 (2014).
31. Liu, L. & Yu, L. Phybase: an R package for species tree analysis. *Bioinformatics* **26**, 962–963, doi: 10.1093/bioinformatics/btq062 (2010).
32. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* **24**, 1586–1591, doi: 10.1093/molbev/msm088 (2007).

Acknowledgements

This project is supported by Key Construction Program of the National ‘985’ project of East China Normal University to Dong Dong (79633006).

Author Contributions

D.D. designed the study. M.L. carried out the data analysis. M.L. and D.D. wrote the manuscript. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Lei, M. and Dong, D. Phylogenomic analyses of bat subordinal relationships based on transcriptome data. *Sci. Rep.* **6**, 27726; doi: 10.1038/srep27726 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>