*Integrative Zoology* 2024; 0: 1–19 doi: 10.1111/1749-4877.12915

**ORIGINAL ARTICLE**

# Comparative analysis of chromosome-level genomes provides insights into chromosomal evolution in Chiroptera

Zerong WANG,[1] Shilin TIAN,[1,2] Jiaxin PANG,[1] Xiangyi ZHANG,[1] Xiangyu HAO,[3] Libiao ZHANG[4] and Huabin ZHAO[1]

[1]Key Laboratory of Biodiversity and Environment on the Qinghai-Tibetan Plateau of the Ministry of Education, Frontier Science Center for Immunology and Metabolism, College of Life Sciences, Wuhan University, Wuhan, China, [2]Novogene Bioinformatics Institute, Beijing, China, [3]College of Plant Protection, Northwest A&F University, Yangling, China and [4]Institute of Zoology, Guangdong Academy of Sciences, Guangzhou, China

## Abstract

Chiroptera (bats) presents a fascinating model due to its remarkable variation in chromosome numbers, which range from 14 to 62. This astonishing diversity makes bats an excellent subject for studying chromosome evolution. The black-bearded tomb bat (*Taphozous melanopogon*) occupies a pivotal phylogenetic position within Chiroptera, emphasizing its crucial role in the systematic examination of bat chromosome evolution. In this study, we present the first chromosome-level genome of *T. melanopogon* within the family Emballonuridae. Together with previously published genomes, we construct a strongly supported phylogenetic tree of bats, which supports that Emballonuridae forms a basal group within Yangochiroptera. Furthermore, we reconstruct ancestral karyotypes at key nodes along the bat phylogeny and conduct a synteny analysis among the genomes of 12 bat species. Our findings identified evolutionary breakpoint regions (EBRs) that are of particular interest. Notably, some bat genomes exhibit an enrichment of genes related to host defense against microbial pathogens within EBRs. Remarkably, one species possesses multiple copies of some $\beta$-defensin genes, while six other species have experienced the loss of some $\beta$-defensin genes due to EBRs. Furthermore, some olfactory receptor genes are located in EBRs of 12 species, 4 of which have a significant enrichment in sensory perception of smell. Together, our comparative genomic analysis underscores the potential link between chromosome rearrangements and the adaptation of bats to defend against microbial pathogens.

**Key words:** bats, chromosomal evolution, chromosome-level genome, evolutionary breakpoint region

## INTRODUCTION

Chromosome rearrangements are a hallmark of chromosome evolution that include duplications, deletions, fusions, fissions, translocations, and inversions (Morin *et al.* 2017). As advances in the development of whole-genome sequencing technology, the recognition of chromosome rearrangements can be performed by comparative genomic study, which shows a significant advantage over fluorescence *in situ* hybridization (FISH)-based cytogenetics technology and could precisely identify homologous synteny blocks (HSBs) and evolutionary breakpoint

*Correspondence*: Huabin Zhao, College of Life Sciences, Wuhan University, Wuhan 430072, China. Email: huabinzhao@whu.edu.cn

regions (EBRs) (Kim *et al.* 2017). The associations of EBRs or HSBs with sequence features, gene functions, and regulatory elements indicate functional effects resulting from chromosome evolution (Larkin *et al.* 2009; Farré *et al.* 2016). Previous studies have characterized associations between EBRs and gene functions, which identified genes located in EBRs that are enriched in functional categories of taste and smell (Groenen *et al.* 2012; Fan *et al.* 2019), immune responses and forebrain development (Ullastres *et al.* 2014; Farré *et al.* 2016). In comparison, HSBs usually accumulate developmental and housekeeping genes and conserved regulatory elements (Farré *et al.* 2016; Damas *et al.* 2017).

The mammalian order Chiroptera consists of over 1400 species that are geographically widespread, accounting for approximately a fifth of mammalian biodiversity (Fenton & Simmons 2015). Chiroptera contains two suborders, Yinpterochiroptera and Yangochiroptera, which diverged some 60 million years ago (Ma) (Teeling *et al.* 2005; Hao *et al.* 2024). Bats play a vital role in ecosystems by preying on agricultural insect pests, pollinating plants, and dispersing seeds (Boyles *et al.* 2011; Kunz *et al.* 2011; Frick *et al.* 2020). Particularly, they have peculiar mammalian adaptations, including powered flight, sophisticated echolocation, unique immunity, and extraordinary longevity (Teeling *et al.* 2018; Ramírez-Fráncel *et al.* 2022). The lack of whole genomic data for representative species makes the inference of Yangochiroptera higher-level relationships less robust, which is specifically reflected in the phylogenetic position of the superfamily Emballonuroidea. Some studies determined the basal position of Yangochiroptera (Teeling *et al.* 2018; Hao *et al.* 2024), whereas some agreed with the sister-group relationships with the superfamily Noctilionoidea (Meredith *et al.* 2011; Amador *et al.* 2018; Álvarez-Carretero *et al.* 2022). The bat family Emballonuridae, members of which are widely distributed in tropical and subtropical regions (Wilson & Reeder 2005), shows a key phylogenetic location within Yangochiroptera (Teeling *et al.* 2005; Meredith *et al.* 2011; Wu *et al.* 2023; Hao *et al.* 2024). The black-bearded tomb bat (*Taphozous melanopogon*), as a representative species of Emballonuridae, is thus crucial to study the origin and evolution of Chiroptera.

As chromosome-level genome assemblies of bats have been continuously reported, we have opportunities to explore the molecular basis of bats' unique adaptations in terms of chromosome evolution. Reconstruction of ancestral karyotypes is one of the fundamental targets of evolutionary biology and is of significance for understanding chromosome evolution. It identifies gross changes that shaped extant genomes and the time of change occurrence, suggesting the evolutionary history of species and clades (Damas *et al.* 2018). In addition, ancestral chromosome reconstruction permits the identification of novel chromosome rearrangements and EBRs (Farré *et al.* 2019). The first ancestral karyotype reconstruction was determined using chromosome banding patterns, zoo-FISH, and genetic maps (Ijdo *et al.* 1991; Murphy *et al.* 2003; Richard *et al.* 2003). Since gene homology-based ancestral karyotype reconstruction algorithms were applied, it expanded the range of species selection for ancestral karyotype reconstruction, but they are appropriated for chromosome-level genome assemblies and have limited suitability for fragmented genome assemblies (Kim *et al.* 2017).

Here, we present a chromosome-level genome assembly of *T. melanopogon*, which is a representative species of Emballonuridae. We reconstructed a phylogenetic tree of bats based on 11 669 orthologs. Together with 10 other chromosome-level bat genomes, we performed chromosomal evolutionary analyses in Chiroptera. First, we reconstruct the ancestral karyotype of three nodes in bat phylogeny. These reconstructions trace dynamic changes of bat chromosomes and indicate potential chromosome rearrangement events resulting from chromosome fusions and fissions during Chiroptera evolution. Next, we detect interchromosomal rearrangement events and define HSBs and EBRs among 12 modern bat species. In addition, we performed Gene Ontology (GO) functional enrichment analyses for the genes located in EBRs of the 12 bat genomes linking chromosome rearrangements to the evolution of $\beta$-defensin and olfactory receptor (OR) genes.

## MATERIALS AND METHODS

### Sample collection and DNA extraction

The liver tissue of one male *T. melanopogon* was collected and stored in liquid nitrogen before DNA extraction. Field sampling of this bat individual was approved by the Institute of Zoology Guangdong Academy of Sciences (Permit ID: GIABR2020810). Genomic DNA was extracted by the Qiagen tissue kit (Catalog no. 13323). Genomic DNA quality and concentration were measured using 0.75% agarose gel electrophoresis and Qubit fluorimeter (Invitrogen).

## Library construction and genome sequencing

The genomic DNA, which met the required qualifications, underwent size selection using BluePippin (Sage Science) and was processed following the Ligation Sequencing Kit 1D (SQK-LSK109) protocol. The quality of the DNA was assessed using a Qubit. In more detail, DNA fragments were first repaired using NEBNext FFPE Repair Mix from New England Biolabs. Following the repair and 3′-adenylation steps using the NEBNext End repair/dA-tailing Module reagents from New England Biolabs, the Oxford Nanopore sequencing adapters were ligated using the NEBNext Quick Ligation Module (E6056) also from New England Biolabs. Finally, the library was sequenced on four flow cells utilizing the PromethION DNA sequencer by Oxford Nanopore for 48 h.

Base calling was conducted using Guppy (version 2.0.8) with default parameters, and the reads were filtered to remove sequences with low-quality bases and adapter remnants, controlled by the mean_qscore_template being greater than or equal to 7. Next, we utilized Nextdeonovo (https://github.com/Nextomics/NextDenovo) to perform correction with specific parameters, including seed_cutoff = 25k. Subsequently, the corrected reads were assembled using smartdenovo (https://github.com/ruanjue/smartdenovo) with the following parameters: −k 21, −J 3000, and –e dom.

To further enhance the accuracy of the assembly, Illumina reads were aligned to the genome assembly using BWA (Li & Durbin 2010), and two rounds of consensus correction were performed through NextPolish (https://github.com/Nextomics/NextPolish). Following these steps, benchmarking universal single-copy orthologs (BUSCO) was employed to evaluate the completeness of the genome assembly against the mammalia_odb9 database. Last, we conducted a GC-depth analysis using minimap2 (Li 2018) or mapping Nanopore reads to the genome to assess the decontamination procedure.

## Hi-C library construction and sequencing

The liver tissue cells were initially fixed with formaldehyde, and the DNA was digested using the restriction endonuclease DpnII. Subsequently, the 5′ overhangs of the resulting fragments were repaired and labeled with biotinylated nucleotides. Following this, the fragments were ligated, and the DNA underwent purification to remove biotin residues located at non-ligated fragment ends. The ends of the fragments, which were sheared by sonication, were further repaired using DNA polymerase. A-tails were added to the fragment ends, and Illumina paired-end sequencing adapters were then ligated. The final Hi-C sequencing library was subjected to PCR amplification and subsequently sequenced on an Illumina NovaSeq platform, producing 2 × 150-bp paired-end reads.

To ensure the quality of the data, adapter sequences and low-quality paired-end reads were filtered out using fastp (v0.12.6) with default parameters (Chen *et al.* 2018). The clean Hi-C reads were primarily aligned to the draft genome assembly using Bowtie 2 (v2.2.3) with the following parameters: –very-sensitive −L 30, and only unique mapped paired-end reads were retained for further analysis (Langmead & Salzberg 2012). Subsequently, HiC-Pro (v2.7.8) was employed to filter out invalid reads, including self-circle, dangling-end, and discarded pairs from the pool of unique mapped read pairs (Servant *et al.* 2015).

## Chromosomal-level genome assembly using Hi-C data

Contigs were clustered, ordered, and oriented to form chromosomes utilizing the LACHESIS tool (Burton *et al.* 2013), with specific parameters set as follows: CLUSTER MIN RE SITES = 100; CLUSTER MAX LINK DENSITY = 2.5; CLUSTER NONINFORMATIVE RATIO = 1.4; ORDER MIN N RES IN TRUNK = 60; and ORDER MIN N RES IN SHREDS = 60. To assess the quality of the genome assembly at the chromosomal level, a genome-wide Hi-C heatmap was generated and visualized using ggplot2 in the R package.

## Genome annotation

Gene prediction was approached through three distinct methods: ab initio, homology-based, and transcript-based. For ab initio prediction, several tools were employed, including Program to Assemble Spliced Alignment (PASA) (Haas *et al.* 2003), Augustus (Stanke & Waack 2003), Supplemental Nutrition Assistance Program (SNAP) (Korf 2004), GlimmerHMM (Majoros *et al.* 2004), GeneID (Guigó 1998), and GeneScan (Burge & Karlin 1997). Briefly, assembled transcripts were aligned against the genome assembly, and gene structure models, which served as gene model training sets, were refined through PASA. Subsequently, Augustus, SNAP, and GlimmerHMM were applied to predict genes based on these training sets. Gene models were generated using GeneID and GeneScan. Homology-based prediction involved aligning protein sequences from a reference

protein set of bats, horses, humans, and mice to the target genome using TBLASTN with an e-value of $1 \times 10^{-5}$. GeneWise (Birney *et al.* 2004) was utilized to predict gene models by incorporating the aligned sequences and their corresponding query proteins as input files. In the transcript-based approach, transcriptome data from six different bats (*Cynopterus sphinx*, *Myotis myotis*, *Molossus molossus*, *Phyllostomus discolor*, *Rhinolophus ferrumequinum*, and *Rousettus aegyptiacus*) were downloaded from NCBI. These RNA-seq data were then aligned to the genome using Tophat (Trapnell *et al.* 2009) to identify exon–intron splice junctions. The final gene models were predicted using Cufflinks (Trapnell *et al.* 2012). Finally, all the gene models predicted by the aforementioned three methods were integrated using EVidenceModeler (Haas *et al.* 2008). Gene function information was obtained through homology searches in public gene databases, which included SwissProt (Apweiler *et al.* 2004), Pfam (Finn *et al.* 2016), NCBI Refseq (NR) (Marchler-Bauer *et al.* 2010), GO (The Gene Ontology Consortium 2017), and Kyoto Encyclopedia of Genes and Genomes (Kanehisa *et al.* 2014).

Transposable elements were identified through both homology-based and ab initio methods. In the homology-based prediction, we utilized RepeatMasker (Tarailo-Graovac & Chen 2009) to map the genome against the Repbase (Jurka *et al.* 2005) TE library, applying default parameters for the analysis. For the ab initio prediction, we employed LTR FINDER (Xu & Wang 2007), PILER (Edgar & Myers 2005), and RepeatScout (Price *et al.* 2005) with their default parameters to construct a reference repeat library. Subsequently, RepeatMasker was employed to align the genome against this newly created library. In addition to these methods, tandem repeats were also identified using the Tandem Repeats Finder package (Benson 1999), with the following parameters: 2 7 7 80 10 50 2000 −d –h.

## Phylogenetic reconstruction and divergence time estimation

To elucidate the evolutionary history of *T. melanopogon*, we expanded our analysis to include the genomes of 17 bat species and 7 outgroup species (horse, cattle, pig, dog, cat, mouse, and human). Single-copy orthologous genes were identified using OrthoFinder (Emms & Kelly 2015). Coding sequences were aligned with the L-INS-I strategy in MAFFT (Katoh & Standley 2013), and conserved sites were extracted using Gblocks (Talavera & Castresana 2007). We created a super-matrix dataset with 11 669 loci and 12 041 540 sites, excluding the third codon positions. Phylogenetic trees were reconstructed using both concatenated approach (IQ-TREE) and the coalescent method (ASTRAL). IQ-TREE (Nguyen *et al.* 2015) was used for concatenated super-matrix dataset, and the best nucleotide substitution models for maximum likelihood (ML) analyses were determined using ModelFinder (Kalyaanamoorthy *et al.* 2017). ML analyses involved the ultrafast bootstrap algorithm with 10 000 replicates. For the coalescent tree, all 11 669 individual gene trees were constructed using IQ-TREE mentioned above, and then all gene trees were collected to construct the species tree by ASTRAL (Zhang *et al.* 2018). Additionally, the estimation of divergence times among bats was carried out using the MCMCtree program within the PAML package (Yang 2007). Fossil constraints for these estimations were drawn from Meredith *et al.* (2011) and TimeTree, as detailed in Table S62, Supporting Information.

## Reconstruction of ancestral karyotypes

To reconstruct the ancestral karyotypes of Yinpterochiroptera, Yangochiroptera, and Chiroptera, we employed 12 chromosomal-level genome assemblies, which included 11 bat genomes along with the human genome used as an outgroup. Our approach consisted of three steps. First, we employed BLASTP to identify orthologous genes between pairs of species. Subsequently, we constructed conserved synteny blocks containing a minimum of five consecutive orthologous genes using MCScanX (v1.1) (Wang *et al.* 2012). Finally, for the identification of contiguous ancestral regions (CARs), we utilized ANGES (v1.01) (Chauve & Tannier 2008), applying identical parameters, except for adjustments made to the target reconstruction node. Fission and fusion events were determined by comparing the distribution of chiropteran ancestral chromosomes in the CARs (or chromosomes) of Yinpterochiroptera, Yangochiroptera, and *T. melanopogon*. A fission event is indicated if a CAR is formed from one or more segments of ancestral chromosomes. Conversely, a fusion event is indicated if a CAR is formed from two or more ancestral chromosomes.

## EBR detection among bat genomes

The analytical workflow to detect EBRs among bat genomes was adapted from the study by Fan *et al.* (2019). Initially, protein-coding genes were obtained through annotation for *C. sphinx*, *Hipposideros armiger*, *Rhinolophus sinicus*, *Megaderma lyra*, *Pteronotus davyi*, and *T.*

*melanopogon*. For *R. aegyptiacus*, *R. ferrumequinum*, *P. discolor*, *Desmodus rotundus*, *M. molossus*, and *M. myotis*, gene data were sourced from the NCBI database. Subsequently, we utilized SyMAP software (Soderlund *et al.* 2011) to construct large-scale HSBs, using both genome sequences and orthologous protein-coding genes as input data. The local synteny blocks were defined as 20 bp or longer exact matches between two genomes. EBRs were defined as the intervals between two consecutive HSBs.

### Functional enrichment analysis of genes located in EBRs

In this analysis, the genes situated within EBRs were aligned with human protein sequences via BLASTP, with the best hit establishing the corresponding orthologous genes. Subsequently, these orthologous genes were converted into their corresponding human orthologs, complete with gene IDs for reference. Finally, we performed GO term enrichment analysis (Ashburner *et al.* 2000) using the clusterProfiler package (Yu *et al.* 2012). Any enrichment with an adjusted *P*-value of less than 0.05 was deemed significant.

### Identification of bat OR genes located in EBRs

We constructed an ML tree using IQ-TREE (Nguyen *et al.* 2015) to identify the family of OR genes within EBRs in 12 bats. The input sequence file included human OR genes downloaded from Human Olfactory Data Explorer (HORDE) database (https://genome.weizmann.ac.il/horde) and 12 bats OR genes within EBRs.

### Identification of bat $\beta$-defensins

We initiated our analysis by downloading intact $\beta$-defensin protein sequences from the human database in Uniprot. These sequences served as queries in TBLASTN searches conducted against each of the 12 genome sequences. Our aim was to identify $\beta$-defensin genes, and we used a stringent e-value threshold of $1 \times 10^{-5}$ for this purpose. Subsequently, non-overlapping blast hits, extended by 10 000 base pairs in both directions, were extracted to be used as inputs for Genewise (Birney *et al.* 2004). This step was crucial for predicting the gene structures of the $\beta$-defensins. All the $\beta$-defensin sequences obtained through homologous annotation were aligned, and a neighbor-joining tree was constructed. Based on this analysis, $\beta$-defensin genes were categorized into intact genes, pseudogenes, and partial genes.

## RESULTS

### Genome assembly

A total of 152.8-Gb high-quality clean reads were generated by the Oxford Nanopore PromethION DNA sequencer (Table S1, Supporting Information). After these reads were assembled and corrected, we obtained the *T. melanopogon* genome of 2158.2 megabases (Mb) with 178 contigs and a contig N50 of 47.1 Mb (Table S2, Supporting Information). The average GC content of genome assembly was 41.27%, indicating a qualified decontamination procedure (Figs S1,S2, Supporting Information). Furthermore, using high-throughput chromosome conformation capture (Hi-C) data, these assembled contigs were successfully anchored into 21 chromosomes with the scaffold N50 of 126.1 Mb. (Figs S1,S3 and Table S3, Supporting Information). Particularly, during the alignments, approximately 98.42% of the Hi-C reads were mapped to the genome, 70% of which were uniquely mapped (Table S4, Supporting Information). A total of 415 339 123 valid read pairs were detected by HiC-Pro pipeline, accounting for 80.86% of the unique mapped read pairs (Table S5, Supporting Information). Finally, BUSCO results revealed that 95.13% of complete BUSCOs were detected and only 2.56% of BUSCO genes were missed, indicating a high level of completeness of the genome assembly (Table S6, Supporting Information).

### Repeat analysis and genome annotation

A total of 798.41-Mb repeat sequences were identified in the *T. melanopogon* genome, which accounted for 36.99% of the genome assembly (Table S7, Supporting Information). These repeat sequences belonged to four major repeat classes: long interspersed nuclear elements (30.80%), long terminal repeats (LTRs, 8.79%), DNA transposons (3.14%), and short interspersed nuclear elements (1.50%) (Table S7, Supporting Information).

The combination of ab initio-based, homologue-based, and transcript-based methods predicted 20 821 protein-coding genes in the new genome, with an average gene length, average exon length and average intron length of 37 973, 175, and 4313 bp, respectively (Table S8, Supporting Information). A total of 18 913 genes, which accounted for 90.84% of the predicted genes, were annotated with putative functions (Table S9, Supporting Information).

## Phylogenetic relationships of bats

Currently, two main hypotheses were proposed in terms of the phylogeny of bats, with some controversies in Yangochiroptera in previous studies: (1) The superfamily Emballonuroidea was placed as the basal lineages of Yangochiroptera, being a sister group to other yangochiropteran taxa (Noctilionoidea and Vespertilionoidea; Fig. 1a) (Teeling *et al.* 2018); (2) Emballonuroidea unites Noctilionoidea as a clade, which is the sister to Vespertilionoidea (Fig. 1b) (Meredith *et al.* 2011; Amador *et al.* 2018; Álvarez-Carretero *et al.* 2022).

To address the above controversy, we undertook an analysis reconstructing the phylogeny and timeline of the Chiroptera. This endeavor involved the utilization of a dataset consisting of 17 bat species and 7 outgroups, including horse, cattle, pig, dog, cat, mouse, and human, based on 11 669 single-copy orthologous genes (Fig. 1c). Both concatenated method and coalescent method showed the same topology in bat species with high strong support (bootstrap values of all nodes for the ML method = 100) and favored the hypothesis that the superfamily Emballonuroidea is the basal lineages of Yangochiroptera (Fig. 1c; Fig. S4, Supporting Information).

The molecular dating analyses showed that the divergence times of bats, Yinpterochiroptera, and Yangochiroptera are approximately 65.3 Ma (60.3–70.1 Ma, 95% highest posterior density [HPD]), 59.3 Ma (55.7–61.1 Ma, 95% HPD), and 56.6 Ma (51.8–61.2 Ma, 95% HPD), respectively, which are similar with previous analyses (Teeling *et al.* 2005; Teeling *et al.* 2018).

## Reconstruction of ancestral karyotypes

We inferred ancestral Chiroptera karyotype (ACK), ancestral Yinpterochiroptera karyotype (AYIK), and ancestral Yangochiroptera karyotype (AYAK) by comparing modern species and reconstructing the order of ancestral genes (also referred to as protogenes) within contiguous ancestral regions (CARs; also referred to as protochromosomes) (Fig. 2). The ACK was refined with 19 CARs (18 autosomes plus X chromosome) and 14 673 ordered protogenes (Table S10, Supporting Information). The AYIK was refined with 22 CARs (21 autosomes plus X chromosome) and 11902 ordered protogenes (Table S11, Supporting Information). The AYAK was refined with 23 CARs (22 autosomes plus X chromosome) and 13 753 ordered protogenes (Table S12, Supporting Information). The AYIK evolved from the ACK by at least six fission and seven fusion events (Fig. 2). The AYAK differed from that of ACK by at least nine fission and six fusion events (Fig. 2). The chromosomes of *T. melanopogon* evolved from that of AYAK by at least one fission and eight fusion events (Fig. 2).

As far as the chiropteran ancestral chromosomes are concerned, yinpterochiropteran ancestor had 12 chromosomes rearranged (CAR1-CAR4, CAR7-CAR8, CAR10, CAR12-CAR13, and CAR15-CAR17) and yangochiropteran ancestor had 11 chromosomes impacted by rearrangements (CAR1-CAR4, CAR6-CAR7, CAR10, CAR13, CAR15, CAR17, and CAR19; Fig. 2). As a result, only three chiropteran ancestral chromosomes (CAR5, CAR9, and CAR14) and X chromosome were intact in three reconstructed ancestors, suggesting that they have been conserved during 65 millions of Chiroptera evolution (Hao *et al.* 2024) (Fig. 2). In addition, CAR18 was only distributed in AYIK and its descendant modern species (*R. sinicus*, *R. ferrumequinum*, *R. aegyptiacus*, and *C. sphinx*). Likewise, CAR19 was only distributed in AYAK and its descendant modern species (*P. discolor* and *P. davyi*; Fig. 2).

## Chromosome rearrangements among bat genomes

Advances in whole-genome sequencing have made it possible to identify chromosome rearrangements at high resolution. Due to strong collinearity, interchromosomal rearrangements could be identified with unknown orientations of scaffolds. Specifically, when the *T. melanopogon* genome was aligned to *C. sphinx*, *R. aegyptiacus*, *H. armiger*, *R. sinicus*, *R. ferrumequinum*, *M. lyra*, *P. discolor*, *D. rotundus, P. davyi*, *M. molossus*, and *M. myotis* genomes (Fig. 3), a range of 8–30 chromosome fission events were detected (Tables S13,S14, Supporting Information).

Furthermore, we used the same approach to align each bat genome to other bat genomes. When *C. sphinx* genome was mapped to the 11 remaining bat genomes, we identified a range of 12–30 chromosome fission events (Fig. S5 and Table S15, Supporting Information). When *R. aegyptiacus* was aligned to 11 remaining bat genomes, we identified a range of 9–24 chromosome fission events (Fig. S6 and Table S16, Supporting Information). Regarding *H. armiger*, we identified 10–28 chromosome fission events (Fig. S7 and Table S17, Supporting Information). For *R. sinicus*, we identified 13–25 chromosome fission events (Fig. S8 and Table S18, Supporting Information). As to *R. ferrumequinum*, we identified 3–17 chromosome fission events (Fig. S9 and Table S19, Supporting Information). With regard to *M. lyra*, we identified 21–31 chro-

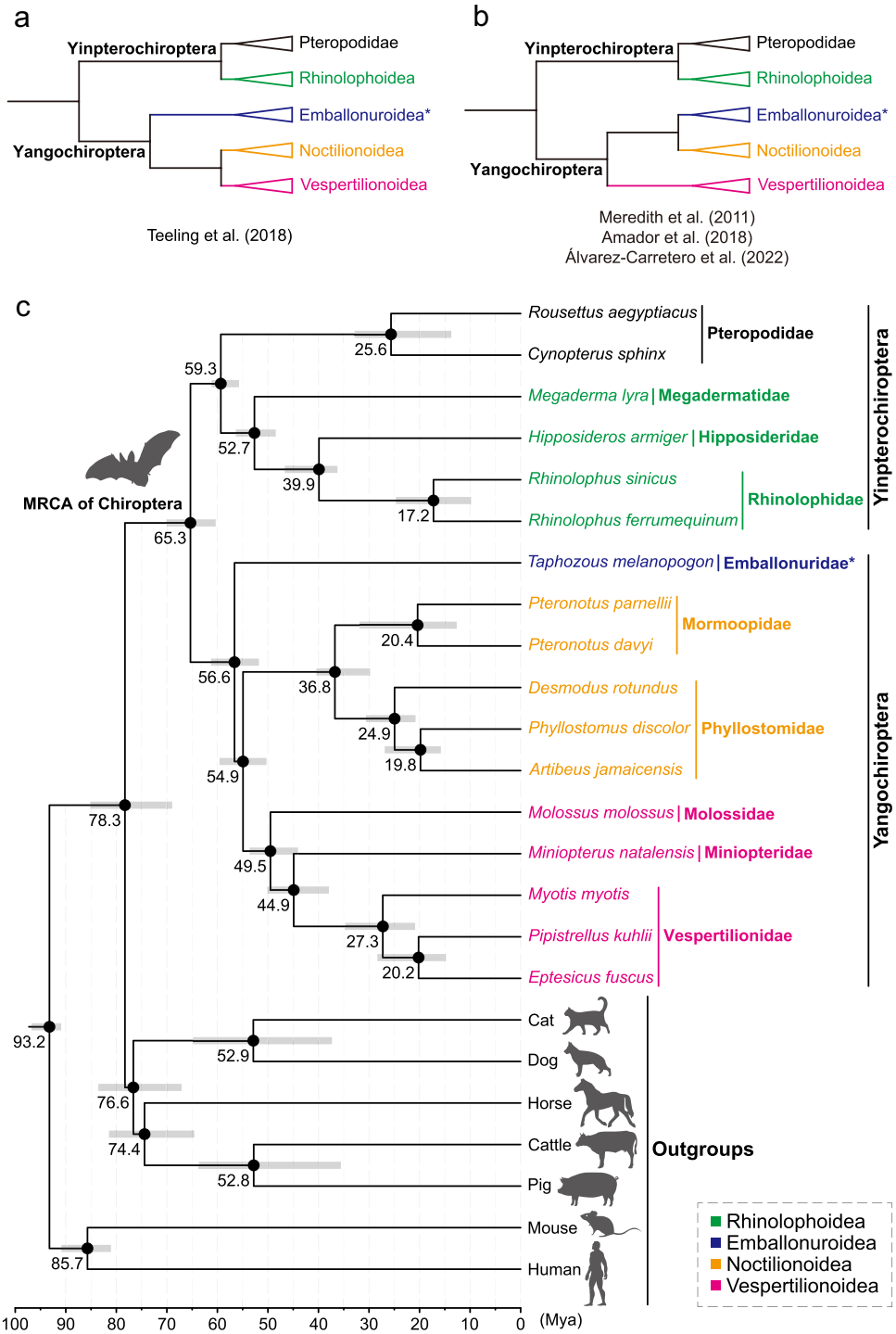


**Figure 1** Phylogenetic relationships among bats and comparison of previous studies' phylogenies. (a) Phylogenetic tree from Teeling *et al.* (2018). (b) Phylogenetic tree from Meredith *et al.* (2011), Amador *et al.* (2018), and Álvarez-Carretero *et al.* (2022). (c) Phylogenetic tree reconstructed using the maximum likelihood (ML) method in this study. The black dot on each node represents the support value of 100. The family Emballonuridae is indicated with an asterisk. MRCA, most recent common ancestor.

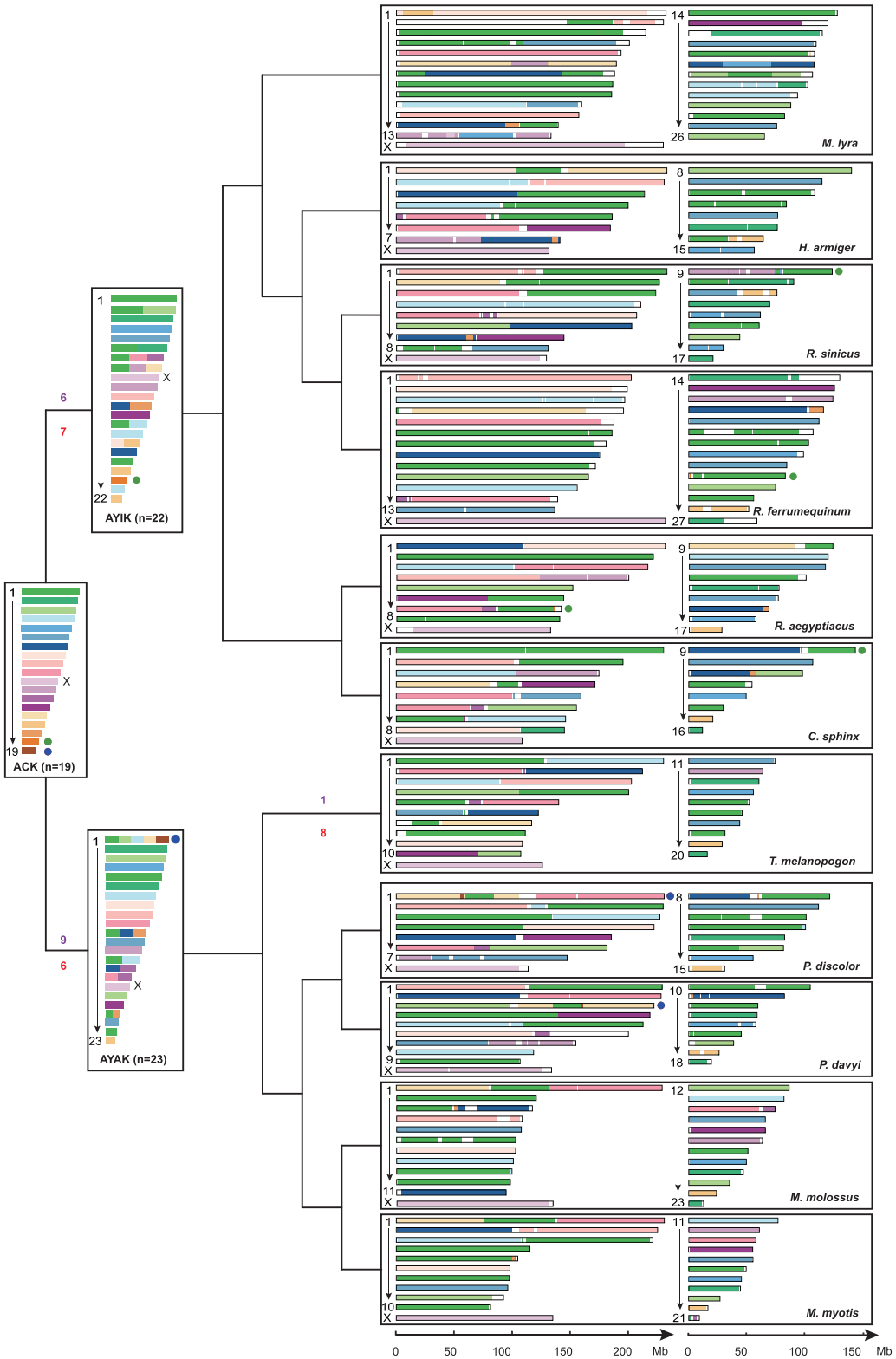


**Figure 2** Reconstruction of proto-chromosomes for Chiroptera genome evolution. We reconstructed the ancestral karyotypes of bats, which included the Ancestral Chiroptera Karyotype (ACK), Ancestral Yinpterochiroptera Karyotype (AYIK), and Ancestral

mosome fission events (Fig. S10 and Table S20, Supporting Information). For *P. discolor*, we identified 9–35 chromosome fission events (Fig. S11 and Table S21, Supporting Information). For *D. rotundus*, we identified 10–34 chromosome fission events (Fig. S12 and Table S22, Supporting Information). Focusing on *P. davyi*, we identified 5–29 chromosome fission events (Fig. S13 and Table S23, Supporting Information). Concentrating on *M. molossus*, we identified 1–23 chromosome fission events (Fig. S14 and Table S24, Supporting Information). For *M. myotis*, we identified 3–26 chromosome fission events (Fig. S15 and Table S25, Supporting Information).

Notably, *M. lyra* exhibits a higher number of chromosome rearrangements compared to other bats, not only due to its larger chromosome count but also implying the possibility of lineage-specific chromosome evolution.

Yangochiroptera Karyotype (AYAK). We inferred that the most recent common ancestor of bats had a $2n = 38$ ancestral karyotype. Color-coded blocks correspond to their chromosomal origins in the chiropteran ancestor. CAR18 is highlighted with a green circle, while CAR19 is marked with a blue circle. The numbers of estimated rearrangements on the three branches are indicated, with purple numbers representing fissions and red numbers indicating fusions. CAR, contiguous ancestral region.

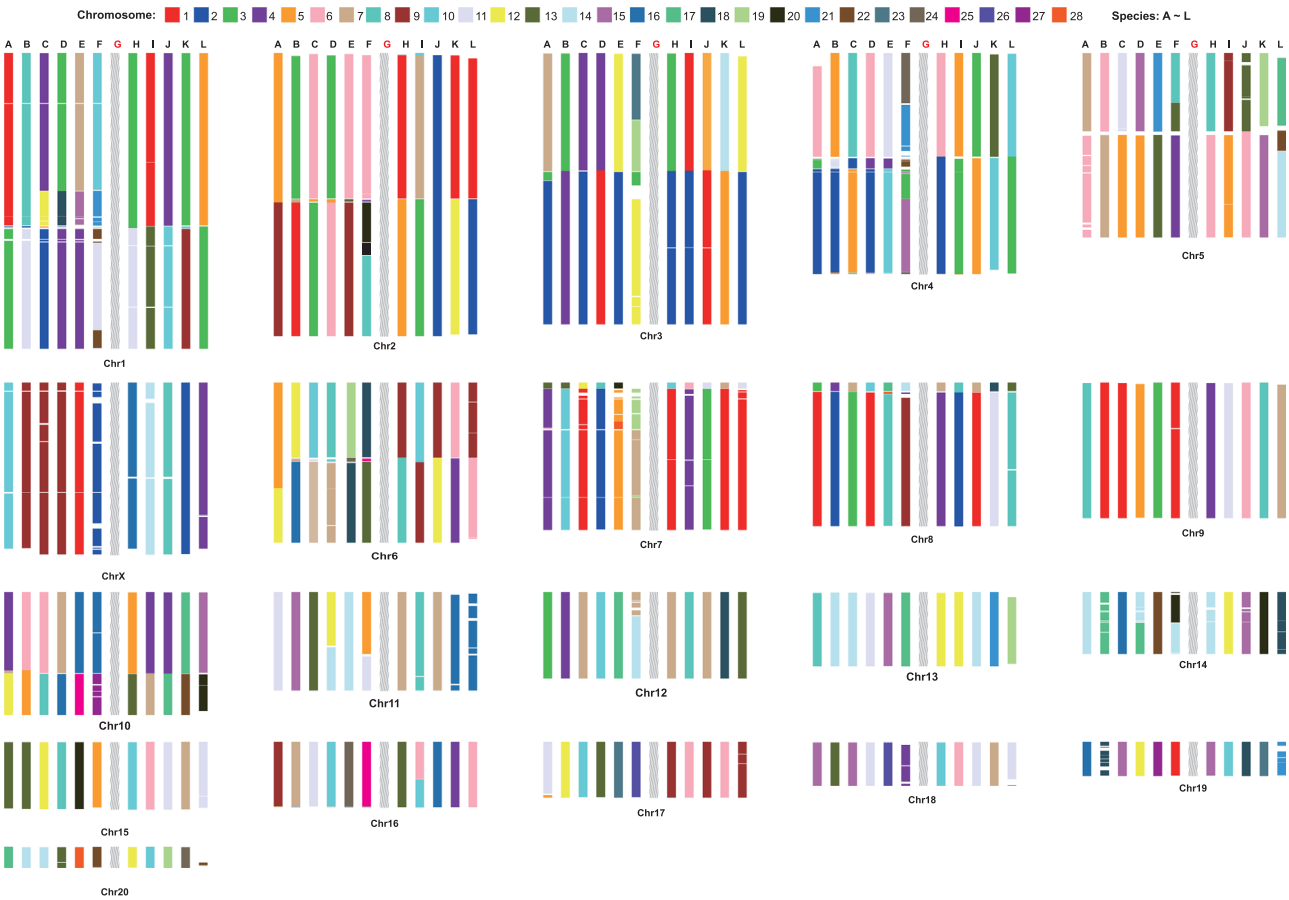


**Figure 3** Diagram depicting chromosomal rearrangements in the genome of *Taphozous melanopogon*. Colored blocks illustrate the range of chromosome identity, spanning from 1 to 28 (Table S13, Supporting Information). Letters from A to L represent the following bat species: *Cynopterus sphinx*, *Rousettus aegyptiacus*, *Hipposideros armiger*, *Rhinolophus sinicus*, *Rhinolophus ferrumequinum*, *Megaderma lyra*, *T. melanopogon*, *Phyllostomus discolor*, *Desmodus rotundus*, *Pteronotus davyi*, *Molossus molossus*, and *Myotis myotis*, respectively. Among them, the letter G corresponds to *T. melanopogon*, which is highlighted in red. *T. melanopogon* serves as the reference point for comparison with the 11 other bat species. Regions that match *T. melanopogon* are known as homologous synteny blocks (HSBs), while the interval regions between two adjacent HSBs are referred to as evolutionary breakpoint regions (EBRs) specific to *T. melanopogon*.

## Identification and analysis of EBRs among bat genomes

To detect potential EBRs, we identify large-scale HSBs among 12 chromosome-level bat genomes by pairwise comparisons. A total of 57, 65, 55, 59, 53, 95, 45, 48, 43, 33, and 59 large-scale HSBs were identified after aligning *T. melanopogon* genome to *C. sphinx*, *R. aegyptiacus*, *H. armiger*, *R. sinicus*, *R. ferrumequinum*, *M. lyra*, *P. discolor*, *D. rotundus*, *P. davyi*, *M. molossus*, and *M. myotis* genomes, respectively (Table S26, Supporting Information).

Based on the identified large-scale HSBs, we estimated the number and distribution of EBRs among these 12 genomes. The alignments between *T. melanopogon* and each of 11 other genomes (*C. sphinx*, *R. aegyptiacus*, *H. armiger*, *R. sinicus*, *R. ferrumequinum*, *M. lyra*, *P. discolor*, *D. rotundus*, *P. davyi*, *M. molossus*, and *M. myotis*) revealed 32, 40, 32, 36, 31, 69, 14, 21, 16, 10, and 34 EBRs, respectively (Table S27, Supporting Information). Upon consolidating identical EBRs, we identified a total of 144 EBRs spanning 120.38 Mb within the *T. melanopogon* genome (Tables S28,S29, Supporting Information).

We also identified large-scale HSBs and EBRs in *C. sphinx*, *R. aegyptiacus*, *H. armiger*, *R. sinicus*, *R. ferrumequinum*, *M. lyra*, *P. discolor*, *D. rotundus*, *P. davyi*, *M. molossus*, and *M. myotis* (Tables S30–S51, Supporting Information). The detailed information of each bat EBRs are shown in Tables S28–S29, Supporting Information.

## Functional enrichment of genes located in EBRs

We identified a total of 1083, 543, 949, 734, 555, 1160, 1097, 1147, 589, 328, 407, and 622 genes in EBRs of *C. sphinx*, *R. aegyptiacus*, *H. armiger*, *R. sinicus*, *R. ferrumequinum*, *M. lyra*, *P. discolor*, *D. rotundus*, *P. davyi*, *M. molossus*, *M. myotis*, and *T. melanopogon*, respectively (Table S29, Supporting Information). Previous studies have shown that EBRs are commonly located in high-density regions of genomes, where the clustering of genes is associated with adaptive phenotypes (Larkin *et al.* 2009; Ullastres *et al.* 2014; Farré *et al.* 2016, 2019). We performed GO functional enrichment analyses for the genes located in EBRs of the 12 bat genomes. Ten of the 12 species were found to show functional enrichments, 6 species of which have significant enrichments in GO terms such as viral transcription (GO:0019083) and/or viral gene expression (GO:0019080), and 2 species of which have significant enrichments in defense response to gram-positive bacteria (GO:0050 830) (Tables S29,S52–S61, Supporting Information). In addition, 4 of the 12 species have significant enrichments in GO terms such as detection of chemical stimulus involved in sensory perception of smell (GO:0050911) and sensory perception of smell (GO:0007608) (Tables S52–S61, Supporting Information). All of the five functionally enriched terms have been detected in *T. melanopogon* (Table S52, Supporting Information).

We identified a total of 17, 47,108, 66, 18, 7, 21, 71, 29, 19, 26, and 54 OR genes in EBRs of *C. sphinx*, *R. aegyptiacus*, *H. armiger*, *R. sinicus*, *R. ferrumequinum*, *M. lyra*, *P. discolor*, *D. rotundus*, *P. davyi*, *M. molossus*, *M. myotis*, and *T. melanopogon*, respectively (Table S29, Supporting Information). These OR genes are classified in 16 OR families (Fig. S16, Supporting Information).

Defensins possess potent antibacterial, antiviral, and antifungal activities, which could be subdivided into $\alpha$, $\beta$, and $\theta$ defensins (Holly *et al.* 2017). We identified 32 types of $\beta$-defensin genes in 12 bats, with a varied number ranging from 24 to 40, 9 of which are located in EBRs: *DEFB1*, *DEFB103–DEFB109*, and *DEFB130* (Fig. 4; Table S29, Supporting Information). Among these genes, seven $\beta$-defensin genes (*DEFB1*, *DEFB103–DEFB108*) were identified as located in *C. sphinx* EBRs; three $\beta$-defensin genes (*DEFB104*, *DEFB106*, and *DEFB107*) were identified as located in *R. aegyptiacus* EBRs; eight $\beta$-defensin genes (*DEFB1*, *DEFB103–DEFB106*, *DEFB108–DEFB109*, and *DEFB130*) were identified as located in *H. armiger* and *P. discolor* EBRs; nine $\beta$-defensin genes (*DEFB1*, *DEFB103–DEFB109*, and *DEFB130*) were identified as located in *R. sinicus*, *R. ferrumequinum*, and *P. davyi* EBRs; seven $\beta$-defensin genes (*DEFB1*, *DEFB103–DEFB106*, *DEFB109*, and *DEFB130*) were identified as located in *M. lyra* EBRs; eight $\beta$-defensin genes (*DEFB103–DEFB109* and *DEFB130*) were identified as located in *D. rotundus* EBRs; four $\beta$-defensin genes (*DEFB103*, *DEFB105–DEFB107*) were identified as located in *M. molossus* EBRs; three $\beta$-defensin genes (*DEFB108*, *DEFB109*, and *DEFB130*) were identified as located in *M. myotis* EBRs; and eight $\beta$-defensin genes (*DEFB1*, *DEFB103*, *DEFB105–DEFB109*, and *DEFB130*) were identified as located in *T. melanopogon* EBRs (Fig. 4; Table S29, Supporting Information).

Specifically, *DEFB105* and *DEFB107*, situated within EBRs of *D. rotundus* experienced gene duplications, resulting in a higher copy number compared to other bats (Fig. 4). *DEFB105* is found in five copies, while *DEFB107* is present in two copies in *D. rotundus* (Fig. 4). In addition, *DEFB1* of *M. myotis*, *DEFB104* of *M. molossus*, *DEFB105* of *R. aegyptiacus*, and *DEFB107* of *M.*

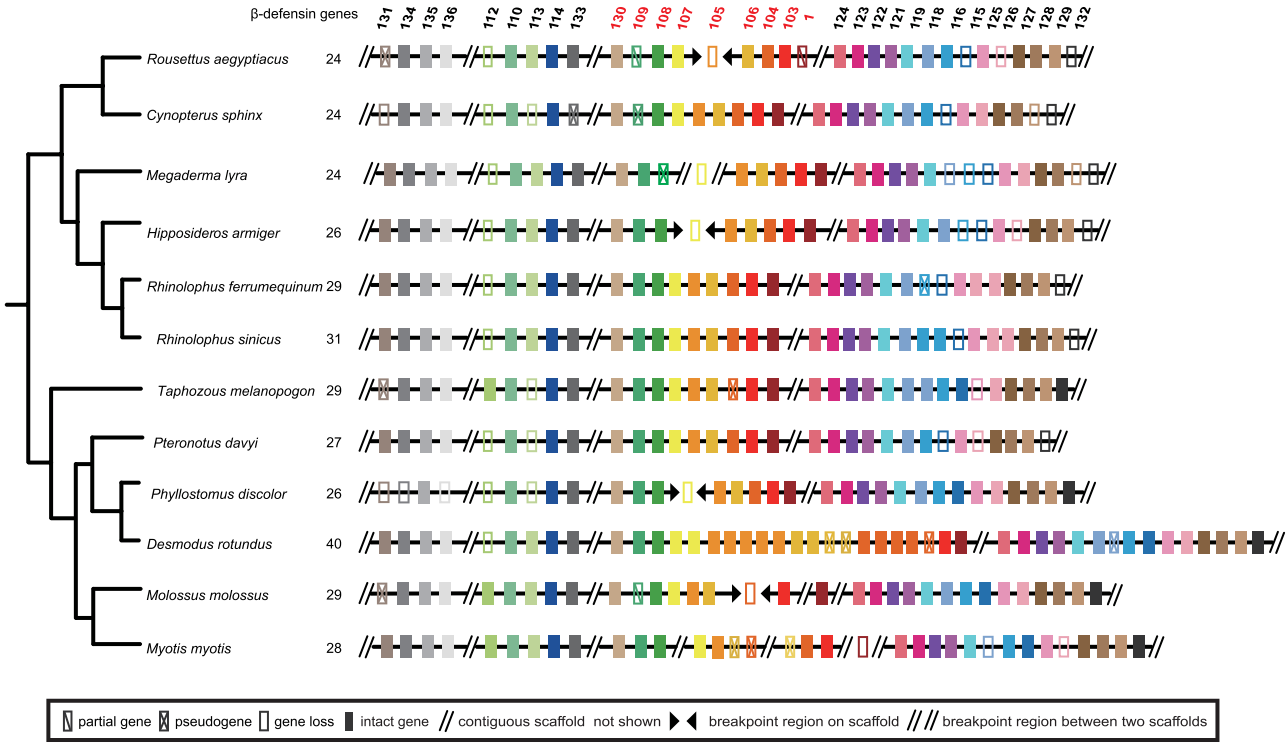


**Figure 4** Nine $\beta$-defensin genes located within evolutionary breakpoint regions (EBRs). A total of 32 $\beta$-defensin genes have been identified in bats, each designated with Arabic numerals following "DEFB" (e.g. *DEFB131*). The numbers following each species' Latin name indicate the total count of $\beta$-defensin genes within that specific species. Gene names excluding "DEFB" are displayed above the colored blocks, with nine $\beta$-defensin genes situated in EBRs highlighted in red. *DEFB104* in *Molossus molossus*, *DEFB105* in *Rousettus aegyptiacus*, *DEFB107* in *Hipposideros armiger* and *Phyllostomus discolor* are lost due to EBRs positioned on a single scaffold. Additionally, *DEFB1* in *Myotis myotis* and *DEFB107* in *Megaderma lyra* are absent because of EBRs found between two separate scaffolds.

*lyra*, *H. armiger*, and *P. discolor*, are lost resulting from evolutionary chromosome breakages (Fig. 4).

## DISCUSSION

The phylogenetic position of Emballonuroidea has long been a topic of debate (Teeling *et al.* 2005; Meredith *et al.* 2011; Amador *et al.* 2018; Teeling *et al.* 2018; Álvarez-Carretero *et al.* 2022). This ongoing debate can be attributed to the lack of extensive genomic datasets encompassing representative species within this superfamily. Previous phylogenetic studies, while striving to achieve comprehensive species sampling, produced somewhat less robust relationships, likely owing to the substantial amounts of missing data (Philippe *et al.* 2004). To tackle this challenge, we have provided the first chromosome-level genome for a member of the superfamily Emballonuroidea and have reconstructed the evolutionary history of bats using data from 11 669 orthologs. In our analysis, ML method placed *T. melanopogon* at the base of the Yangochiroptera clade. To minimize the potential effects of incomplete lineage sorting, we conducted a more detailed examination of the topology using coalescent methods, particularly ASTRAL, which yielded consistent results. A recent family-level phylogenetic study with a more extensive taxon sampling based on nine genes has also recovered the basal position of Emballonuroidea, which is consistent with our results (Hao *et al.* 2024). Nevertheless, our analysis might still benefit from the inclusion of other emballonurid species to mitigate the potential issue of long-branch attraction and obtain a more robust evolutionary framework in the future.

Reconstructing ancestral karyotypes plays a crucial role in enhancing our understanding of genome evolution. By deciphering the chromosome organization of common ancestors within the vertebrate phylogeny, we can detect chromosome rearrangements and establish their connections between phenotypic evolution and speciation

(Deakin & Ezaz 2014; O'Connor *et al.* 2018; Sacerdot *et al.* 2018). Damas *et al.* reconstructed ancestral karyotypes of 14 important nodes within the avian phylogeny and described the rates of chromosome rearrangements and the distribution of EBRs of avian ancestors (Damas *et al.* 2018). They also reconstructed ancestral karyotypes of 16 ancestral nodes along mammalian phylogeny (Damas *et al.* 2022). Previous studies utilizing comparative chromosome mapping have reconstructed ancestral karyotypes of ruminants and carnivores; these studies have shown that the genome organization of different taxa evolved at varying rates and involved different types of rearrangements (Kulemzina *et al.* 2011; Ruiz-Herrera *et al.* 2012; Beklemisheva *et al.* 2016). In this study, we reconstructed ancestral karyotypes of three nodes within the bat phylogeny. The reconstruction of ancestral karyotypes is dependent on the input tree file. Based on our reconstructed bat phylogeny, the ACK likely had 18 pairs of autosomes plus X chromosome, which is similar to the ancestral eutherian karyotype (19 pairs of autosomes plus X chromosome). The reconstructed karyotypes have provided valuable insights into the patterns of structural evolution resulting from chromosome fusions and fissions in bats chromosomes over a span of approximately 61 million years of evolution (Hao *et al.* 2024). By tracing these differences, we have a better opportunity to understand the evolutionary dynamics and genomic changes that have occurred in bat lineages.

The number of EBRs identified in each bat indicated that *R. ferrumequinum* has relatively small number of EBRs (Table S29, Supporting Information), which may be related to the high chromosome number. However, it is not applied to *M. lyra* ($2n = 54$), implying lineage-specific chromosome evolution in *M. lyra*.

Chromosome breakage is nonrandom and EBRs appear to be hotpots of evolutionary activity (Damas *et al.* 2021). EBRs tend to locate in regions with a higher density of zinc-finger genes, and genes located in EBRs could be related to lineage-specific adaptive phenotypes (Larkin *et al.* 2009; Ullastres *et al.* 2014; Farré *et al.* 2016, 2019). Some studies have identified associations between gene functions and EBRs. For example, Rhesus macaque EBRs are in gene-rich regions being enriched in GO terms related to immune system (Ullastres *et al.* 2014). An eariler study identified 124 EBRs in the cattle lineage (Bovine Genome Sequencing and Analysis Consortium *et al.* 2009). These EBRs were found to contain genes that are associated with immune response. Larkin *et al.* identified a total of 1064 EBRs through pairwise comparisons of 10 amniote genomes; the genes for these EBRs were enriched in relation to the inflammatory response (Larkin *et al.* 2009). In our study, we identified EBRs of 12 bats (Table S29, Supporting Information). An enrichment analysis showed that genes located in EBRs are enriched in GO terms associated with the defense against microbial pathogens (viral transcription, viral gene expression, and defense response to gram-positive bacterium), which may be related to bats' ability to carry multiple viruses (Letko *et al.* 2020; Zhao *et al.* 2022, 2023; Wang *et al.* 2024), indicating that chromosome rearrangements have the potential to impact the defense against microbial pathogen.

In addition, we found for some species of Chiroptera, genes located in EBRs are significantly enriched for functions in sensory perception of smell (GO:0007608) (Tables S52–S61, Supporting Information). Most of the species with this enrichment in OR genes are insectivorous, including *T. melanopogon, H. armiger*, and *R. sinicus*. Fan *et al.* also found that some genes located in carnivoran genome EBRs were functionally enriched the sensory perception of olfaction (Fan *et al.* 2019). These results suggest that evolution of olfactory system could be generally affected by chromosome rearrangement events.

Among those genes located in EBRs of bat genomes, we detected nine $\beta$-defensin genes (*DEFB1*, *DEFB103–DEFB109*, and *DEFB130*). *DEFB105* and *DEFB107*, located in EBRs of *D. rotundus*, exhibit a higher copy number compared to other bats. These genes may be expressed in the testis, as shown in primates (Semple *et al.* 2003), suggesting potential functional implications for the host defense of male reproductive tract. This highlights the importance of structural genomic changes in the evolutionary and adaptive processes related to microbial pathogen defense (Semple *et al.* 2003).

Previous studies showed that five $\beta$-defensin genes (*DEFB118–DEFB119* and *DEFB121–DEFB123*) located in macaque EBRs (Ullastres *et al.* 2014). The observed differences in the categories of $\beta$-defensin genes between macaque and Chiroptera EBRs suggest that different taxa have selectively focused on lineage-specific $\beta$-defensin genes during the course of chromosome evolution.

## CONCLUSIONS

The study of chromosome evolution has seen renewed interest due to the rapid growth of chromosome-level genome assemblies, particularly in vertebrates. In this work, we presented a chromosome-level genome for *T. melanopogon* and reconstructed the phylogenetic relationships among bats, suggesting that Emballonuridae represents the basal lineage of Yangochiroptera. Based on these phylogenetic relationships, we reconstructed

the karyotypes of the Chiroptera ancestor, Yinpterochiroptera ancestor, and Yangochiroptera ancestor, providing valuable new insights into the evolutionary history and patterns of chromosome evolution in Chiroptera. Our analysis also involved synteny analysis among 12 chromosome-level Chiroptera genomes, which allowed us to identify HSBs and EBRs. Additionally, functional enrichment analysis of EBR genes in the 12 bat genomes indicated that events associated with chromosome rearrangement may impact the defense against microbial pathogens. This finding establishes a potential connection between chromosome evolution and the evolution of functional genes in the context of microbial pathogen defense.

## ACKNOWLEDGMENTS

We thank all personnel for their assistance in collecting the sample. This study was supported by the National Key Research and Development Program of China (2021YFF0702004), the National Natural Science Foundation of China (32270436), Fundamental Research Funds for the Central Universities (2042022dx0003), and the Natural Science Foundation of Hubei Province (2023AFA015).

## ETHICS STATEMENT

*Taphozous melanopogon* sequenced in this study were sampled in the field (Permit ID: GIABR2020810). No specific permits were required, and no endangered or protected species were involved.

## CONFLICT OF INTEREST STATEMENT

None of the authors have any competing interests.

## DATA AVAILABILITY STATEMENT

The assembled *T. melanopogon* genome and whole-genome sequencing data have been deposited into the National Center for Biotechnology Information (NCBI) under the BioProject accession number PRJNA1025362.

## REFERENCES

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.

Apweiler R, Bairoch A, Wu CH *et al*. (2004). UniProt: the universal protein knowledgebase. *Nucleic Acids Research* **32**, D115–19.

Ashburner M, Ball CA, Blake JA *et al*. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics* **25**, 25–29.

Beklemisheva VR, Perelman PL, Lemskaya NA *et al*. (2016). The ancestral carnivore karyotype as substantiated by comparative chromosome painting of three pinnipeds, the walrus, the steller sea lion and the Baikal seal (Pinnipedia, Carnivora). *PLoS ONE* **11**, e0147647.

Benson G (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research* **27**, 573–80.

Birney E, Clamp M, Durbin R (2004). Genewise and genomewise. *Genome Research* **14**, 988–95.

Bovine Genome Sequencing and Analysis Consortium, Elsik CG, Tellam RL *et al*. (2009). The genome sequence of taurine cattle: A window to ruminant biology and evolution. *Science* **324**, 522–28.

Boyles JG, Cryan PM, McCracken GF, Kunz TH (2011). Economic importance of bats in agriculture. *Science* **332**, 41–42.

Burge C, Karlin S (1997). Prediction of complete gene structures in human genomic DNA. *Journal of Molecular Biology* **268**, 78–94.

Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nature Biotechnology* **31**, 1119–25.

Chauve C, Tannier E (2008). A methodological framework for the reconstruction of contiguous regions of ancestral genomes and its application to mammalian genomes. *PLoS Computational Biology* **4**, e1000234.

Chen S, Zhou Y, Chen Y, Gu J (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–90.

Damas J, Corbo M, Kim J *et al*. (2022). Evolution of the ancestral mammalian karyotype and syntenic regions. *PNAS* **119**, e2209139119.

Damas J, Corbo M, Lewin HA (2021). Vertebrate chromosome evolution. *Annual Review of Animal Biosciences* **9**, 1–27.

Damas J, Kim J, Farré M, Griffin DK, Larkin DM (2018). Reconstruction of avian ancestral karyotypes reveals

differences in the evolutionary history of macro- and microchromosomes. *Genome Biology* **19**, 155.

Damas J, O'Connor R, Farré M *et al*. (2017). Upgrading short-read animal genome assemblies to chromosome level using comparative genomics and a universal probe set. *Genome Research* **27**, 875–84.

Deakin JE, Ezaz T (2014). Tracing the evolution of amniote chromosomes. *Chromosoma* **123**, 201–16.

Edgar RC, Myers EW (2005). PILER: Identification and classification of genomic repeats. *Bioinformatics* **21**, i152–58.

Emms DM, Kelly S (2015). OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**, 157.

Fan H, Wu Q, Wei F, Yang F, Ng BL, Hu Y (2019). Chromosome-level genome assembly for giant panda provides novel insights into Carnivora chromosome evolution. *Genome Biology* **20**, 267.

Farré M, Kim J, Proskuryakova AA *et al*. (2019). Evolution of gene regulation in ruminants differs between evolutionary breakpoint regions and homologous synteny blocks. *Genome Research* **29**, 576–89.

Farré M, Narayan J, Slavov GT *et al*. (2016). Novel insights into chromosome evolution in birds, archosaurs, and reptiles. *Genome Biology and Evolution* **8**, 2442–51.

Fenton MB, Simmons NB (2015). *Bats, a World of Science and Mystery*. University of Chicago Press, Chicago, IL.

Finn RD, Coggill P, Eberhardt RY *et al*. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research* **44**, D279–85.

Frick WF, Kingston T, Flanders J (2020). A review of the major threats and challenges to global bat conservation. *Annals of the New York Academy of Sciences* **1469**, 5–25.

Groenen MA, Archibald AL, Uenishi H *et al*. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**, 393–98.

Guigó R (1998). Assembling genes from predicted exons in linear time with dynamic programming. *Journal of Computational Biology* **5**, 681–702.

Haas BJ, Delcher AL, Mount SM *et al*. (2003). Improving the arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research* **31**, 5654–66.

Haas BJ, Salzberg SL, Zhu W *et al*. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biology* **9**, R7.

Hao X, Lu Q, Zhao H (2024). A molecular phylogeny for all 21 families within Chiroptera (bats). *Integrative Zoology* **19**, 989–98.

Holly MK, Diaz K, Smith JG (2017). Defensins in viral infection and pathogenesis. *Annual Review of Virology* **4**, 369–91.

Ijdo JW, Baldini A, Ward DC, Reeders ST, Wells RA (1991). Origin of human chromosome 2: An ancestral telomere-telomere fusion. *PNAS* **88**, 9051–55.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005). Repbase update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* **110**, 462–67.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**, 587–89.

Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research* **42**, D199–D205.

Katoh K, Standley DM (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772–80.

Kim J, Farré M, Auvil L *et al*. (2017). Reconstruction and evolutionary history of eutherian chromosomes. *PNAS* **114**, E5379–88.

Korf I (2004). Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59.

Kulemzina AI, Yang F, Trifonov VA, Ryder OA, Ferguson-Smith MA, Graphodatsky AS (2011). Chromosome painting in Tragulidae facilitates the reconstruction of Ruminantia ancestral karyotype. *Chromosome Research* **19**, 531–39.

Kunz TH, Braun de Torrez E, Bauer D, Lobova T, Fleming TH (2011). Ecosystem services provided by bats. *Annals of the New York Academy of Sciences* **1223**, 1–38.

Langmead B, Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–59.

Larkin DM, Pape G, Donthu R, Auvil L, Welge M, Lewin HA (2009). Breakpoint regions and homologous

synteny blocks in chromosomes have different evolutionary histories. *Genome Research* **19**, 770–77.

Letko M, Seifert SN, Olival KJ, Plowright RK, Munster VJ (2020). Bat-borne virus diversity, spillover and emergence. *Nature Reviews Microbiology* **18**, 461–71.

Li H (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–100.

Li H, Durbin R (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–95.

Majoros WH, Pertea M, Salzberg SL (2004). TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–79.

Marchler-Bauer A, Lu S, Anderson JB *et al*. (2010). CDD: A conserved domain database for the functional annotation of proteins. *Nucleic Acids Research* **39**, D225–29.

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.

Morin SJ, Eccles J, Iturriaga A, Zimmerman RS (2017). Translocations, inversions and other chromosome rearrangements. *Fertility and Sterility* **107**, 19–26.

Murphy WJ, Bourque G, Tesler G, Pevzner P, O'Brien SJ (2003). Reconstructing the genomic architecture of mammalian ancestors using multispecies comparative maps. *Human Genomics* **1**, 30–40.

Nguyen LT, 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.

O'Connor RE, Romanov MN, Kiazim LG *et al*. (2018). Reconstruction of the diapsid ancestral genome permits chromosome evolution tracing in avian and non-avian dinosaurs. *Nature Communications* **9**, 1883.

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.

Price AL, Jones NC, Pevzner PA (2005). De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–58.

Ramírez-Fráncel LA, García-Herrera LV, Losada-Prado S *et al*. (2022). Bats and their vital ecosystem services: a global review. *Integrative Zoology* **17**, 2–23.

Richard F, Lombard M, Dutrillaux B (2003). Reconstruction of the ancestral karyotype of eutherian mammals. *Chromosome Research* **11**, 605–18.

Ruiz-Herrera A, Farré M, Robinson TJ (2012). Molecular cytogenetic and genomic insights into chromosomal evolution. *Heredity* **108**, 28–36.

Sacerdot C, Louis A, Bon C, Berthelot C, Crollius HR (2018). Chromosome evolution at the origin of the ancestral vertebrate genome. *Genome Biology* **19**, 166.

Semple CAM, Rolfe M, Dorin JR (2003). Duplication and selection in the evolution of primate $\beta$-defensin genes. *Genome Biology* **4**, R31.

Servant N, Varoquaux N, Lajoie BR *et al*. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biology* **16**, 259.

Soderlund C, Bomhoff M, Nelson WM (2011). SyMAP v3.4: A turnkey synteny system with application to plant genomes. *Nucleic Acids Research* **39**, e68.

Stanke M, Waack S (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**, ii215–25.

Talavera G, Castresana J (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* **56**, 564–77.

Tarailo-Graovac M, Chen N (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics* **5**, 4–10.

Teeling EC, Springer MS, Madsen O, Bates P, O'Brien S J, 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.

The Gene Ontology Consortium (2017). Expansion of the gene ontology knowledgebase and resources. *Nucleic Acids Research* **45**, D331–38.

Trapnell C, Pachter L, Salzberg SL (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–11.

Trapnell C, Roberts A, Goff L *et al*. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* **7**, 562–78.

Ullastres A, Farré M, Capilla L, Ruiz-Herrera A (2014). Unraveling the effect of genomic structural changes in the rhesus macaque—Implications for the adaptive role of inversions. *BMC Genomics* **15**, 530.

Wang X, Jia JK, Wang Q *et al*. (2024). Myotis bat STING attenuates aging-related inflammation in female mice. *Zoological research* **45**, 961–71.

Wang Y, Tang H, Debarry JD *et al*. (2012). MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* **40**, e49.

Wilson DE, Reeder DM (2005). *Mammal Species of the World: A Taxonomic and Geographic Reference*. Johns Hopkins University Press, Baltimore, MD.

Wu J, Zhang L, Shen C, Sin SYW, Lei C, Zhao H (2023). Comparative transcriptome analysis reveals molecular adaptations underlying distinct immunity and inverted resting posture in bats. *Integrative Zoology* **18**, 493–505.

Xu Z, Wang H (2007). LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research* **35**, W265–68.

Yang Z (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**, 1586–91.

Yu G, Wang LG, Han Y, He QY (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omics* **16**, 284–87.

Zhang C, Rabiee M, Sayyari E, Mirarab S (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC bioinformatics* **19**, 153.

Zhao K, Zhang W, Li B *et al*. (2022). Ecological study of cave nectar bats reveals low risk of direct transmission of bat viruses to humans. *Zoological Research* **43**, 514–22.

Zhao R, Niu S, Han P *et al*. (2023). Cross-species recognition of bat coronavirus RsYN04 and cross-reaction of SARS-CoV-2 antibodies against the virus. *Zoological Research* **44**, 1015–25.

## SUPPLEMENTARY MATERIALS

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

**Figure S1** Genome characteristics of *Taphozous melanopogon*. (a) Ideogram of 21 chromosomes, with each tick representing 10 megabases (Mb). (b) Gene density, expressed as the number of genes per 100 kilobases, ranging from 1 to 19. (c) GC content within a 100-kilobase (kb) window, with values ranging from 31% to 65%. (d) Percentage of repeat sequence coverage in 100-kb windows, encompassing a range from 3% to 94%.

**Figure S2** GC-depth distribution of genome. X-axis indicates GC contents and Y-axis shows depths with a window size of 50 kb.

**Figure S3** The heatmap depicting all-by-all interactions among 21 chromosomes (resolution = 100 kb). Colour bar shows interaction degrees ranging from white (low) to red (high).

**Figure S4** The reconstructed phylogenetic tree of bats using the coalescent method.

**Figure S5** The diagram of chromosomal rearrangements in *C. sphinx* genome. Colored blocks show chromosome identity range from 1 to 28 (Supplementary Materials: Table S13). Letter from A to L indicates species of *C. sphinx*, *R. aegyptiacus*, *H. armiger*, *R. sinicus*, *R. ferrumequinum*, *M. lyra*, *P. discolor*, *D. rotundus*, *P. davyi*, *M. molossus*, *M. myotis* and *T. melanopogon*, respectively. HSBs and *C. sphinx* lineage-specific EBRs compared with other 11 bat species are shown.

**Figure S6** The diagram of chromosomal rearrangements in *R. aegyptiacus* genome. Colored blocks show chromosome identity range from 1 to 28 (Supplementary Materials: Table S13). Letter from A to L indicates species of *R. aegyptiacus*, *C. sphinx*, *H. armiger*, *R. sinicus*, *R. ferrumequinum*, *M. lyra*, *P. discolor*, *D. rotundus*, *P. davyi*, *M. molossus*, *M. myotis* and *T. melanopogon*, respectively. HSBs and *R. aegyptiacus* lineage-specific EBRs compared with other 11 bat species are shown.

**Figure S7** The diagram of chromosomal rearrangements in *H. armiger* genome. Colored blocks show chromosome identity range from 1 to 28 (Supplementary Materials: Table S13). Letter from A to L indicates species of *H. armiger*, *C. sphinx*, *R. aegyptiacus*, *R. sinicus*, *R. ferrumequinum*, *M. lyra*, *P. discolor*, *D. rotundus*, *P. davyi*, *M. molossus*, *M. myotis* and *T. melanopogon*, respectively. HSBs and *H. armiger* lineage-specific EBRs compared with other 11 bat species are shown.

**Figure S8** The diagram of chromosomal rearrangements in *R. sinicus* genome. Colored blocks show chromosome identity range from 1 to 28 (Supplementary Materials: Table S13). Letter from A to L indicates species of *R. sinicus*, *C. sphinx*, *R. aegyptiacus*, *H. armiger*, *R. ferrumequinum*, *M. lyra*, *P. discolor*, *D. rotundus*, *P. davyi*, *M. molossus*, *M. myotis* and *T. melanopogon*, respectively. HSBs and *R. sinicus* lineage-specific EBRs compared with other 11 bat species are shown.

**Figure S9** The diagram of chromosomal rearrangements in *R. ferrumequinum* genome. Colored blocks show

chromosome identity range from 1 to 28 (Supplementary Materials: Table S13). Letter from A to L indicates species of *R. ferrumequinum*, *C. sphinx*, *R. aegyptiacus*, *H. armiger*, *R. sinicus*, *M. lyra*, *P. discolor, D. rotundus*, *P. davyi*, *M. molossus*, *M. myotis* and *T. melanopogon*, respectively. HSBs and *R. ferrumequinum* lineage-specific EBRs compared with other 11 bat species are shown.

**Figure S10** The diagram of chromosomal rearrangements in *M. lyra* genome. Colored blocks show chromosome identity range from 1 to 28 (Supplementary Materials: Table S13). Letter from A to L indicates species of *M. lyra*, *C. sphinx*, *R. aegyptiacus*, *H. armiger*, *R. sinicus*, *R. ferrumequinum*, *P. discolor*, *D. rotundus*, *P. davyi*, *M. molossus*, *M. myotis* and *T. melanopogon*, respectively. HSBs and *M. lyra* lineage-specific EBRs compared with other 11 bat species are shown.

**Figure S11** The diagram of chromosomal rearrangements in *P. discolor* genome. Colored blocks show chromosome identity range from 1 to 28 (Supplementary Materials: Table S13). Letter from A to L indicates species of *P. discolor*, *C. sphinx*, *R. aegyptiacus*, *H. armiger*, *R. sinicus*, *R. ferrumequinum*, *M. lyra*, *D. rotundus*, *P. davyi*, *M. molossus*, *M. myotis* and *T. melanopogon*, respectively. HSBs and *P. discolor* lineage-specific EBRs compared with other 11 bat species are shown.

**Figure S12** The diagram of chromosomal rearrangements in *D. rotundus* genome. Colored blocks show chromosome identity range from 1 to 28 (Supplementary Materials: Table S13). Letter from A to L indicates species of *D. rotundus*, *C. sphinx*, *R. aegyptiacus*, *H. armiger*, *R. sinicus*, *R. ferrumequinum*, *M. lyra*, *P. discolor*, *P. davyi*, *M. molossus*, *M. myotis* and *T. melanopogon*, respectively. HSBs and *D. rotundus* lineage-specific EBRs compared with other 11 bat species are shown.

**Figure S13** The diagram of chromosomal rearrangements in *P. davyi* genome. Colored blocks show chromosome identity range from 1 to 28 (Supplementary Materials: Table S13). Letter from A to L indicates species of *P. davyi*, *C. sphinx*, *R. aegyptiacus*, *H. armiger*, *R. sinicus*, *R. ferrumequinum*, *M. lyra*, *P. discolor*, *D. rotundus*, *M. molossus*, *M. myotis* and *T. melanopogon*, respectively. HSBs and *P. davyi* lineage-specific EBRs compared with other 11 bat species are shown.

**Figure S14** The diagram of chromosomal rearrangements in *M. molossus* genome. Colored blocks show chromosome identity range from 1 to 28 (Supplementary Materials: Table S13). Letter from A to L indicates species of *M. molossus*, *C. sphinx*, *R. aegyptiacus*, *H. armiger*, *R. sinicus*, *R. ferrumequinum*, *M. lyra*, *P. discolor*, *D. rotundus*, *P. davyi*, *M. myotis* and *T. melanopogon*, respectively. HSBs and *M. molossus* lineage-specific EBRs compared with other 11 bat species are shown.

**Figure S15** The diagram of chromosomal rearrangements in *M. myotis* genome. Colored blocks show chromosome identity range from 1 to 28 (Supplementary Materials: Table S13). Letter from A to L indicates species of *M. myotis*, *C. sphinx*, *R. aegyptiacus*, *H. armiger*, *R. sinicus*, *R. ferrumequinum*, *M. lyra*, *P. discolor*, *D. rotundus*, *P. davyi*, *M. molossus* and *T. melanopogon*, respectively. HSBs and *M. myotis* lineage-specific EBRs compared with other 11 bat species are shown.

**Figure S16** The phylogenetic tree of OR genes within EBRs in 12 bats using the ML method.

**Table S1** Statistics of the Nanopore subreads for genome assembly

**Table S2** Statistics of the genome assembly before Hi-C correction

**Table S3** Statistics of the Hi-C assembly of the genome

**Table S4** The alignment of the Hi-C clean data to the genome assembly

**Table S5** Statistics of different types of read pairs produced by Hi-C

**Table S6** BUSCO assessment of the genome assembly

**Table S7** Statistics of the repeat sequences in the genome assembly

**Table S8** Statistics of predicted protein-coding genes in the genome

**Table S9** Summary of functional annotations for predicted genes

**Table S10** The ordered ancestral genes of ACK and related modern species

**Table S11** The ordered ancestral genes of AYIK and related modern species

**Table S12** The ordered ancestral genes of AYAK and related modern species

**Table S13** Chromosome identity and corresponding scaffold for 12 bat genomes

**Table S14** The chromosome fission events identified in 11 bat genomes relative to *T. melanopogon*

**Table S15** The chromosome fission events identified in 11 bat genomes relative to *C. sphinx*

**Table S16** The chromosome fission events identified in 11 bat genomes relative to *R. egyptiacus*

**Table S17** The chromosome fission events identified in 11 bat genomes relative to *H. armiger*

**Table S18** The chromosome fission events identified in 11 bat genomes relative to *R. sinicus*

**Table S19** The chromosome fission events identified in 11 bat genomes relative to *R. ferrumequinum*

**Table S20** The chromosome fission events identified in 11 bat genomes relative to *M. lyra*

**Table S21** The chromosome fission events identified in 11 bat genomes relative to *P. discolor*

**Table S22** The chromosome fission events identified in 11 bat genomes relative to *D. rotundus*

**Table S23** The chromosome fission events identified in 11 bat genomes relative to *P. davyi*

**Table S24** The chromosome fission events identified in 11 bat genomes relative to *M. molossus*

**Table S25** The chromosome fission events identified in 11 bat genomes relative to *M. myotis*

**Table S26** The detailed information of HSBs in *T. melanopogon* genome

**Table S27** The detailed information of identified EBRs caused by inter-/intra-chromosome events in *T. melanopogon* genome

**Table S28** The detailed position of identified EBRs in *T. melanopogon*, *C. sphinx*, *R. egyptiacus*, *H. armiger*, *R. sinicus*, *R. ferrumequinum*, *M. lyra*, *P. discolor*, *D. rotundus*, *P. davyi*, *M. molossus* and *M. myotis* genomes

**Table S29** The information of EBRs and genes located in EBRs in 12 Chiroptera genomes

**Table S30** The detailed information of HSBs in *C. sphinx* genome

**Table S31** The detailed information of HSBs in *R. egyptiacus* genome

**Table S32** The detailed information of HSBs in *H. armiger* genome

**Table S33** The detailed information of HSBs in *R. sinicus* genome

**Table S34** The detailed information of HSBs in *R. ferrumequinum* genome

**Table S35** The detailed information of HSBs in *M. lyra* genome

**Table S36** The detailed information of HSBs in *P. discolor* genome

**Table S37** The detailed information of HSBs in *D. rotundus* genome

**Table S38** The detailed information of HSBs in *P. davyi* genome

**Table S39** The detailed information of HSBs in *M. molossus* genome

**Table S40** The detailed information of HSBs in *M. myotis* genome

**Table S41** The detailed information of identified EBRs caused by inter-/intra-chromosome events in *C. sphinx* genome

**Table S42** The detailed information of identified EBRs caused by inter-/intra-chromosome events in *R. egyptiacus* genome

**Table S43** The detailed information of identified EBRs caused by inter-/intra-chromosome events in *H. armiger* genome

**Table S44** The detailed information of identified EBRs caused by inter-/intra-chromosome events in *R. sinicus* genome

**Table S45** The detailed information of identified EBRs caused by inter-/intra-chromosome events in *R. ferrumequinum* genome

**Table S46** The detailed information of identified EBRs caused by inter-/intra-chromosome events in *M. lyra* genome

**Table S47** The detailed information of identified EBRs caused by inter-/intra-chromosome events in *P. discolor* genome

**Table S48** The detailed information of identified EBRs caused by inter-/intra-chromosome events in *D. rotundus* genome

**Table S49** The detailed information of identified EBRs caused by inter-/intra-chromosome events in *P. davyi* genome

**Table S50** The detailed information of identified EBRs caused by inter-/intra-chromosome events in *M. molossus* genome

**Table S51** The detailed information of identified EBRs caused by inter-/intra-chromosome events in *M. myotis* genome

**Table S52** Significantly enriched gene ontology (GO) biological process terms for the genes in EBRs of *T. melanopogon* genome

**Table S53** Significantly enriched gene ontology (GO) biological process terms for the genes in EBRs of *C. sphinx* genome

**Table S54** Significantly enriched gene ontology (GO) biological process terms for the genes in EBRs of *R. egyptiacus* genome

**Table S55** Significantly enriched gene ontology (GO) biological process terms for the genes in EBRs of *H. armiger* genome

**Table S56** Significantly enriched gene ontology (GO) biological process terms for the genes in EBRs of *R. sinicus* genome

**Table S57** Significantly enriched gene ontology (GO) biological process terms for the genes in EBRs of *R. ferrumequinum* genome

**Table S58** Significantly enriched gene ontology (GO) biological process terms for the genes in EBRs of *P. discolor* genome

**Table S59** Significantly enriched gene ontology (GO) biological process terms for the genes in EBRs of *D. rotundus* genome

**Table S60** Significantly enriched gene ontology (GO) biological process terms for the genes in EBRs of *P. davyi* genome

**Table S61** Significantly enriched gene ontology (GO) biological process terms for the genes in EBRs of *M. myotis* genome

**Table S62** Six fossil constraints used in this study

**Cite this article as:**

Wang Z, Tian S, Pang J *et al* (2024). Comparative analysis of chromosome-level genomes provides insights into chromosomal evolution in Chiroptera. *Integrative Zoology* **00**, 1–19. https://doi.org/10.1111/1749-4877.12915