

# Novel de Novo Genome of *Cynopterus brachyotis* Reveals Evolutionarily Abrupt Shifts in Gene Family Composition across Fruit Bats

Balaji Chattopadhyay<sup>1,\*</sup>, Kritika M. Garg<sup>1</sup>, Rajasri Ray<sup>2,3</sup>, Ian H. Mendenhall<sup>4</sup>, and Frank E. Rheindt<sup>1,\*</sup> 

<sup>1</sup>Department of Biological Sciences, National University of Singapore, Singapore

<sup>2</sup>Center for Ecological Sciences, Indian Institute of Science, Bangalore, Karnataka, India

<sup>3</sup>Centre for Studies in Ethnobiology, Biodiversity and Sustainability (CEiBa), Mokdumpur, Malda, West Bengal, India

<sup>4</sup>Programme in Emerging Infectious Diseases, Duke-NUS Medical School, Singapore

\*Corresponding authors: E-mails: balaji@nus.edu.sg; dbrsfe@nus.edu.sg.

Accepted: February 12, 2020

Data deposition: The assembled genome of *Cynopterus brachyotis* has been deposited at GenBank under the accession GCA\_009793145.1.

## Abstract

Major novel physiological or phenotypic adaptations often require accompanying modifications at the genic level. Conversely, the detection of considerable contractions and/or expansions of gene families can be an indicator of fundamental but unrecognized physiological change. We sequenced a novel fruit bat genome (*Cynopterus brachyotis*) and adopted a comparative approach to reconstruct the evolution of fruit bats, mapping contractions and expansions of gene families along their evolutionary history. Despite a radical change in life history as compared with other bats (e.g., loss of echolocation, large size, and frugivory), fruit bats have undergone surprisingly limited change in their genic composition, perhaps apart from a potentially novel gene family expansion relating to telomere protection and longevity. In sharp contrast, within fruit bats, the new *Cynopterus* genome bears the signal of unusual gene loss and gene family contraction, despite its similar morphology and lifestyle to two other major fruit bat lineages. Most missing genes are regulatory, immune-related, and olfactory in nature, illustrating the diversity of genomic strategies employed by bats to contend with responses to viral infection and olfactory requirements. Our results underscore that significant fluctuations in gene family composition are not always associated with obvious examples of novel physiological and phenotypic adaptations but may often relate to less-obvious shifts in immune strategies.

**Key words:** lesser short-nosed fruit bat, gene family evolution, immunity, olfactory, histones.

## Introduction

The genomic era affords us new opportunities to link fundamental physiological innovations to underlying genomic correlates. Whole-genome sequencing of a diverse array of nonmodel species in the past decade has led to great insights into genomic contingents of phenotypic adaptation (Kim et al. 2011; Axelsson et al. 2013; Castoe et al. 2013; Qu et al. 2013; Alberto et al. 2018). Comparative genomic analyses have revealed the importance of gene family fluctuations as an evolutionary mechanism affecting species biology (Sharma et al. 2018).

Among mammals, bats are unique in many aspects (Altringham 1999; Zubaid et al. 2006; Kunz and Parsons

2009). Their volant nature, nocturnal behavior, ubiquitous distribution, occupancy of numerous habitats, long lifespan, diverse dietary adaptations, vast array of social and mating systems in addition to the ability of ~85% of bat species to echolocate have made this group of animals an interesting system for scientific investigations to discover the genomic underpinnings of physiological idiosyncrasies (Altringham 1999; Zubaid et al. 2006; Kunz and Parsons 2009). Large shifts in the gene family evolution of bats have been linked to physiological innovations. Bats possess the smallest mammalian genomes (Smith et al. 2013; Kapusta et al. 2017) with a high turnover and loss of genes related to immunity, regulation, metabolism, and responses to stimuli (Smith et al.

2013; Zhang et al. 2013; Ahn et al. 2016; Tsagkogeorga et al. 2017). For example, some immune genes are known to have contracted, whereas others have expanded in the common ancestor of bats (Smith et al. 2013; Zhang et al. 2013; Ahn et al. 2016). Olfactory genes have also contracted in bats as compared with other mammals (Hayden et al. 2014; Tsagkogeorga et al. 2017). Furthermore, the demonstration of pseudogenization of genes involved in rhinolophid bat vision suggests a trade-off among various sensory modalities (Dong et al. 2017; Tsagkogeorga et al. 2017).

Bats are characterized by an underappreciated wealth of physiological diversity. Old World fruit bats differ from insectivorous bats in various life history traits such as echolocation, diet, vision, and olfaction (Altringham 1999; Kunz and Parsons 2009). Comparison between echolocating and non-echolocating fruit bats has demonstrated a significant expansion of olfactory genes and contraction of genes related to immunity and pathogen recognition within fruit bat lineages (Tsagkogeorga et al. 2017). Numerous outbreaks of zoonotic diseases in recent decades were traced to bat-borne viruses, and subsequent major scientific research initiatives have revealed that many bat species act as a reservoir for major viral pathogens while being largely immune to infections. Understanding the immunity of bats to a spectrum of viral pathogenic infections has recently become an urgent focus. Although bats show an overall contraction of the immune system (Ng et al. 2016; Zhou et al. 2016) and a more primitive mammalian organization of immune regions in the genome (Ng et al. 2016), our present understanding remains in its infancy. Even within bats, and specifically fruit bats, viral load may differ significantly among species groups and genera (Schountz 2014; Laing et al. 2018). However, we do not know the physiological manifestations of such variability and its genomic origins.

In this study, we sequenced a genome of the common Southeast Asian fruit bat *Cynopterus brachyotis* from Singapore to understand in greater detail the evolution of gene families within Old World fruit bats (family Pteropodidae). Many Paleotropical fruit bats, including the one sequenced in this study, are characterized by complex behavior (McCracken and Wilkinson 2000; Campbell 2008; Chattopadhyay et al. 2011; Garg et al. 2012), live in close proximity to humans, and are major reservoirs of viral pathogens implicated in zoonotic disease outbreaks (Schountz 2014; Mani et al. 2017; Laing et al. 2018). For example, *Cynopterus* fruit bats including our study species are a natural reservoir of the Nipah virus (Chong et al. 2009). In addition, the genus *Cynopterus* is part of a major fruit bat subfamily (Cynopterinae) that has so far been omitted from whole-genome bat research and therefore constitutes an important gap. Whole-genome information from this genus in combination with a comparative genomic approach can inform our knowledge about the genomic contingents of important physiological innovations and contribute to our understanding

of functional diversity in bats and its genomic underpinnings. We performed comparative genomic analyses across major bat lineages as well as other mammalian lineages to investigate if Old World fruit bats in general, and the genus *Cynopterus* in particular, show unique signatures of gene family evolution in conjunction with immune system and sensory capabilities. Our conclusions provide novel insights into gene family evolution in fruit bats with regards to immune function, olfaction, and longevity.

## Materials and Methods

### DNA Extraction and Whole-Genome Sequencing

We collected tissue samples from one adult male of *C. brachyotis* from Singapore (National University of Singapore IACUC protocol B16-0159 and National Parks Board Singapore permit NP/RP14-109). The Qiagen DNeasy Blood and Tissue kit (QIAGEN, Germany) was used to extract genomic DNA following the manufacturer's instructions. Whole-genome and mate pair libraries were prepared by AITbiotech Singapore for the following insert sizes: 650 bp, 2 kb, 8 kb, and 12 kb. The whole-genome libraries were sequenced as 250-bp paired-end runs on two lanes of HiSeq2500, whereas all the mate pair libraries were pooled and run on a single lane of HiSeq4000 producing 150-bp paired-end reads.

### Data Processing

We obtained a total of 401 million reads from the whole-genome libraries and 301 million reads from the mate pair libraries and checked data quality in FastQC 0.11.7 (Andrews 2010). *Platanus\_trim* and *Platanus\_internal\_trim* were used to remove adapters, trim low-quality reads (PHRED score < 15 as suggested in the online documentation), and remove short reads (<25 bp) from whole-genome and mate pair libraries. These scripts are part of the PLATANUS 1.2.4 genome assembler (Kajitani et al. 2014). We then removed PCR duplicates using FastUnique (Xu et al. 2012). We further performed k-mer correction on reads prior to genome assembly using a k-mer value of 23 (as suggested by the software developer) in SOAPec\_v2.01 (Luo et al. 2012).

### Genome Size Estimate and Coverage

Genome size was estimated using k-mer analysis. We used Jellyfish 2.2.6 (Marçais and Kingsford 2011) to generate a k-mer histogram and, based on the frequency of k-mers, estimated the genome size as the ratio of  $k\_num/k\_depth$ , where  $k\_num$  is the total number of k-mers and  $k\_depth$  is the frequency of the most common k-mer (supplementary fig. S1, Supplementary Material online). We then inferred genomic coverage based on the estimated genome size. Genomic coverage is defined as the product of the number of reads and read length divided by the estimated genome size.

## Genome Assembly

We assembled the nuclear genome using three different assemblers for a total of four de novo genome assemblies. First, we utilized SOAPdenovo2 (Luo et al. 2012) for two genome assemblies at different k-mer values (see below). Second, we used CLC workbench 9.5 (<https://www.qiagen-bioinformatics.com/>; last accessed February 28, 2020) to assemble contigs and then SOAPdenovo to assemble scaffolds. Third, we employed PLATANUS assembler specifically designed for assembling heterozygous genomes (Kajitani et al. 2014).

For the assembly based exclusively on SOAPdenovo, we used two types of assemblers included in the program, named 63-mer and 127-mer, to generate contigs. The 127-mer assembler requires more memory to assemble contigs and is recommended for long read data (Luo et al. 2012). We varied the k-mer length from 45 to 127 to generate contigs from the whole-genome libraries and set the merge level (-M) to two. We used default settings for mapping and scaffolding in SOAPdenovo and GapCloser (Luo et al. 2012) to close any gaps generated during the scaffolding process. The overlap parameter (-p) was set to 31 for all our analyses.

For the next assembly pipeline, CLC workbench was run to generate contigs using default settings. CLC uses a range of word sizes (12–64) and bubble sizes to generate de Bruijn graphs. We discarded any contigs <1,000 bp. Based on the best word size estimated in CLC, we employed SOAPdenovo to generate scaffolds and GapCloser to generate final assemblies using the previously mentioned settings.

For the final comparative assembly, we ran PLATANUS, which is especially adept at assembling highly heterozygous diploid genomes from high-density data. All steps, including generating contigs, scaffolds, and gap closing, were performed using default settings. All four assemblies were carried out on a dedicated Linux server system with 1 TB RAM and 64 cores.

We compared all four genome assemblies using QUAST 4.6 (Gurevich et al. 2013) in terms of N50, number of scaffolds, and length of the longest contig for all four assemblies. We discarded any scaffold <1,000 bp in length. For our best assembly, we used BUSCO 2.0 (Simão et al. 2015; Waterhouse et al. 2018) to test for completeness. This method relies on a defined set of ultraconserved eukaryotic protein families to quantitatively measure the quality of genome assemblies. We also used the Laurasiatheria ortho database version 9 (Zdobnov et al. 2017), which consists of 6,253 single copy genes, to test the quality of our genome assembly. Additionally, we tested for any strong effect of the quality of genome assembly on BUSCO results. For this purpose, we performed BUSCO analysis for the seven other bat genomes used for comparative analysis in this study (large flying fox, *Pteropus vampyrus*; Egyptian fruit bat, *Rousettus aegyptiacus*; great leaf-nosed bat, *Hipposideros armiger*; Chinese rufous horseshoe bat, *Rhinolophus sinicus*; little brown bat, *Myotis*

*lucifugus*; big brown bat, *Eptesicus fuscus*; and Natal long-fingered bat, *Miniopterus natalensis*).

## Estimating Genome Heterozygosity

We mapped the filtered reads to the assembled genome using the BWA-MEM algorithm within the BWA 0.7.17-r1188 package (Li and Durbin 2009) and used SAMTOOLS 0.1.19 (Li et al. 2009) mpileup for variant calling (SNP calling and InDels). We used a minimum coverage of 10× and a maximum coverage of 100× for variant calling to avoid low-quality SNPs. We further filtered any variant with a Phred score of <20 to avoid errors due to low base quality.

## Repeat Masking

RepeatMasker 4.0.7 (Smit et al. 2015) was used to identify and mask repeat regions within the genome. We first ran RepeatModeler 1.0.10 (Smit and Hubley 2015) to generate a custom repeats library for the *C. brachyotis* genome. RepeatModeler uses RECON 1.08 (Bao and Eddy 2002) and RepeatScout 1.0.5 (Price et al. 2005) to identify repeat element boundaries and family relationships from sequence data to generate custom repeat libraries. The custom library was used with RepeatMasker along with RM BLAST to identify and mask repeat regions within the genome.

## Gene Annotation and Gene Function Assignment

We used AUGUSTUS 3.2.2 (Stanke et al. 2006) to predict genes using the repeat-masked genome. AUGUSTUS is an accurate ab initio gene prediction tool for eukaryotic genomes (Stanke et al. 2006). We applied two different approaches to annotate the genome: firstly, a human gene set as a training data set to identify genes as suggested by the authors, and secondly cDNA hints from transcriptome data of the closely related *Cynopterus sphinx* (GAOV01.1, Dong et al. 2013) to help in gene prediction. We verified both annotations using InterProScan 5 within Blast2GO (Götz et al. 2008) to identify homologous proteins across databases. For a better annotation, we additionally compared our predicted proteins to human, little brown bat (*M. lucifugus*), large flying fox (*P. vampyrus*), and Egyptian fruit bat (*R. aegyptiacus*) proteomes in OrthoVenn (Wang et al. 2015) using a 0.00001 E-value cutoff in protein similarity comparisons and setting the inflation value for generating orthologous clusters to 1.5.

## Mitochondrial Genome Assembly and Identification of Mitochondrial Lineages

We assembled the mitochondrial genome using NOVOPlasty 2.6.3 (Dierckxsens et al. 2017), a de novo assembler for organelle genomes using whole-genome data. We used cleaned reads (after removing PCR duplicates and k-mer corrected) from a single lane for the assembly of the mitogenome. To aid in assembly, we used the already assembled

mitogenome of *C. brachyotis* (NC\_026465) (Yoon et al. 2016). We further annotated the mitogenome using MITOS Web Server (Bernt et al. 2013) applying default settings and the vertebrate mitochondrial genetic code.

We then isolated the *cyt b* sequence from the mitogenome and aligned it to other reference *cyt b* sequences from GenBank to identify the mitochondrial affinity of the sequenced individual (see [Supplementary Material](#) online).

### Identification of Orthologous Genes

We used OrthoFinder 1.1.4 (Emms and Kelly 2015) to identify orthologous protein sequences across various mammalian genomes using default settings (including *C. brachyotis* generated in this study; human, *Homo sapiens*; rhesus macaque, *Macaca mulatta*; mouse, *Mus musculus*; rat, *Rattus norvegicus*; dog, *Canis lupus familiaris*; cat, *Felis catus*; cow, *Bos taurus*; horse, *Equus caballus*; rhinoceros, *Ceratotherium simum simum*; pig, *Sus scrofa*; bottleneck dolphin, *Tursiops truncatus*; big brown bat, *E. fuscus*; great leaf-nosed bat, *H. armiger*; Chinese rufous horseshoe bat, *Rhi. sinicus*; little brown bat, *M. lucifugus*; large flying fox, *P. vampyrus*; Egyptian fruit bat, *R. aegyptiacus*; and Natal long-fingered bat, *Min. natalensis*). OrthoFinder uses MCL 12.135 (Enright et al. 2002) and BLAST 2.2.28+ (Altschul et al. 1990) to identify orthologs across various genomes. We downloaded the protein sequences from Ensembl release 88 and used the longest protein isoform when alternate isoform information was available.

### Phylogenomic Reconstruction and Divergence Dating

We performed phylogenomic reconstructions (using both concatenation and species tree approaches) of major mammalian groups, concentrating on chiropteran lineages. The phylogenomic reconstructions were used for downstream analyses of divergence time estimation along the lineage leading to *C. brachyotis* and for tracing gene family expansions and contractions on the basis of the *C. brachyotis* genome. For both approaches, we used two different data sources: nucleotide sequences (single copy cDNA: 4,326 loci for concatenated tree and 3,194 loci for species tree approach) and protein sequences (1,342 loci for concatenated tree and 298 loci for species tree approach) (see [Supplementary Material](#) online). For all four phylogenomic reconstructions (protein concatenated trees, protein species trees, DNA concatenated trees, and DNA species trees), we used r8s 1.81 (Sanderson 2003) to obtain estimates of divergence times on the basis of three calibrations (see [Supplementary Material](#) online).

### Gene Family Expansions and Contractions

In order to understand gene family expansions and contractions, we used the analytical approach implemented in Computational Analysis of gene Family Evolution (CAFE)

4.0.1 (Han et al. 2013). CAFE employs a random birth–death process to model changes in gene family size while accounting for phylogenetic relationships (Han et al. 2013). CAFE analysis was performed to test for gene family expansions and contractions in *C. brachyotis* with respect to the most recent common ancestor (MRCA) of fruit bats (including lesser short-nosed fruit bat, *C. brachyotis*; large flying fox, *P. vampyrus*; and Egyptian fruit bat, *R. aegyptiacus*) and the MRCA of all bats considered in this study. We used three different ultrametric trees (concatenation-based nucleotide and protein trees and species tree based on nucleotide sequences) obtained for the 19 genomes compared in this study along with the gene counts for each family obtained from OrthoFinder. For estimating the birth–death parameter ( $\lambda$ ), we used gene families with < 100 gene copies for any species to avoid bias in estimates.

The birth–death parameter ( $\lambda$ ) can vary across different branches of the tree. To test if allowing for multiple  $\lambda$  values is significantly better than a global  $\lambda$  model, we performed 100 simulations using the *genfamily* command option in CAFE. Based on the observed  $\lambda$ , we simulated 100 gene count data sets and estimated the likelihood ratio of a global  $\lambda$  versus multiple  $\lambda$  values.

The number of genes within a family can vary due to errors in genome assembly and annotation. To account for this error, we used the *cafferror.py* script provided along with CAFE. This script assumes a single birth–death parameter along all branches of the tree and iteratively searches across a priori defined error distributions. The error distribution with the highest probability for the given data is used for further analysis. For all analyses, we used a *P* value threshold of 0.01 to identify significant gene family expansions and contractions.

We further performed gene ontology (GO) functional enrichment analysis (Boyle et al. 2004) on select gene families in *C. brachyotis*, and in the MRCA of fruit bats and bats. We used the GoTermFinder tool available from Princeton University (<https://go.princeton.edu/cgi-bin/GOTermFinder>; last accessed February 28, 2020) for enrichment analysis. For each gene family with significant variation in gene copy number, we used a representative human protein sequence to perform gene enrichment analysis. Whenever a human protein sequence was not available, we used either mouse, rat, or rhesus macaque protein sequences for analyses, keeping to animal groups with the best-annotated genomes available. The human annotation served as a reference data set for comparison. We performed tests for enrichment of GO terms and corrected for multiple testing. We further used REVIGO (Supek et al. 2011) to summarize GO term enrichment presentation. REVIGO uses semantic similarity measures to cluster and remove redundant GO terms and to visualize long lists of GO terms that can be difficult to interpret.

For all gene families that showed significant fluctuations in *C. brachyotis*, we tested for gene duplication and loss in the ETE Toolkit v3.1.1 (Huerta-Cepas et al. 2016) by using the



species overlap method (Huerta-Cepas et al. 2007) and the strict tree reconciliation algorithm (Page and Charleston 1997). The species overlap algorithm searches for overlap of taxa on either side of a node within a gene tree to discover duplication events and hence does not require a species tree. On the other hand, the species reconciliation algorithm compares a gene tree with a species tree to identify historical events such as expansions and contractions. Hence to test for contraction events, we implemented only the species reconciliation algorithm. To implement the species overlap algorithm, we first aligned each gene family tree using MAFFT v7.310 (Katoh and Standley 2013), followed by phylogenetic reconstruction in RAxML 8.2 (Stamatakis 2014) using a GTR+Gamma model of evolution with 100 rapid bootstraps. We used midpoint rooting to root each gene tree. For the gene/species tree reconciliation algorithm, we compared each of these gene family trees with the concatenated tree that was generated from RAxML as mentioned in the previous sections. Both analyses were performed in the PhyloTree module using default parameters within the ETE toolkit (see IPython notebook for the code).

### Demographic History and Paleoclimatic Habitat Reconstruction

Using the pairwise sequentially Markovian coalescent (PSMC) approach (Li and Durbin 2011), we reconstructed the population history of *C. brachyotis* and *P. vampyrus* (see [Supplementary Material](#) online). We further reconstructed the potential distribution of both species across four time periods (current, mid-Holocene, last glacial maximum, and last interglacial) approximately during the last 100,000 years (see [Supplementary Material](#) online).

## Results

### Genome Assembly

We retained 345 million reads from the whole-genome libraries and 65 million reads from the mate pair libraries ([supplementary table S1, Supplementary Material](#) online) after cleanup steps. The four different assemblies varied in quality ([table 1](#)). The PLATANUS assembly returned the highest contig length and scaffold N50 values as well as the lowest number of scaffolds, performing significantly better than the other assembly pipelines ([table 1](#)). All pipelines returned similar GC percentages for *C. brachyotis* ([table 1](#)). We performed all downstream analyses on the basis of the PLATANUS assembly. Based on k-mer analysis ([supplementary fig. S1, Supplementary Material](#) online), the estimated size of our *C. brachyotis* genome was 1.72 Gb, with an expected coverage at 108 $\times$ .

We performed BUSCO analysis to test for completeness of the assembled genome and compared our genome as well as seven other bat genomes with the Laurasiatheria ortho-

database version 9. We identified 88.6% of single copy orthologs (79.9% with a complete match and 8.7% with partial matches) for the *C. brachyotis* genome, suggesting a good assembly. When compared with other genomes analyzed in our study, we observed that genome completeness in BUSCO was not tightly linked to genome coverage as all study species were characterized by a genome completeness between ~87% and 92%, irrespective of genome coverage which varied from 7 $\times$  to 218.6 $\times$  ([supplementary table S2, Supplementary Material](#) online), suggesting that not all genes within the Laurasiatheria database are present in bats.

A total of 8,805,734 heterozygous sites were detected within the *C. brachyotis* genome. We repeat-masked 24.49% of the genome using species-specific libraries and observed that long-interspersed nuclear elements (LINEs) formed the most commonly occurring repeat elements (10.18% of the genome) in the *C. brachyotis* genome ([table 2](#)).

The assembled mitogenome was 16,637 bp in length and AT rich (58.20%). Similar to other mammals, we identified 13 protein-coding genes, 2 ribosomal RNA genes, and 22 transfer RNA genes ([supplementary fig. S2, Supplementary Material](#) online). Phylogenetic reconstruction based on the mitochondrial cytochrome *b* gene (*cyt b*) confirmed that the sampled individual belongs to the Sunda lineage of *C. brachyotis*, following Campbell et al. (2004) and Chattopadhyay et al. (2016) ([supplementary fig. S3, Supplementary Material](#) online).

### Gene Annotation and Ortholog Identification

We identified 21,822 and 23,727 genes using the human gene set and cDNA hints from transcriptome data of *C. sphinx*, respectively. The slight disparity in these numbers suggests an improvement in gene annotation when the transcriptome of a closely related species was included. For both runs, at least 89% of proteins had known homologs in protein databases ([table 3](#)). As the use of cDNA hints provided better annotation, we employed this set of proteins for further analysis. Comparison of our annotated proteins with those of humans (*Homo sapiens*), little brown bat (*M. lucifugus*), large flying fox (*P. vampyrus*), and Egyptian fruit bat (*R. aegyptiacus*) identified 143 gene families unique to the *C. brachyotis* genome ([supplementary fig. S4 and table S3A, Supplementary Material](#) online). However, 68 out of these 143 gene families had previously been annotated in organisms other than the aforementioned four species (see [supplementary table S3A, Supplementary Material](#) online). Interestingly, we found 2,284 gene families which were simultaneously present in humans, little brown bats, large flying foxes, and Egyptian fruit bats but absent in the *C. brachyotis* genome ([supplementary fig. S4 and table S3B, Supplementary Material](#) online). We further identified 17,529 orthologous gene families across all 19 genomes analyzed (including *C. brachyotis* generated in

**Table 1**

Comparison of Four Genome Assemblies Using QUASt

| Parameter                | 63-mer SOAP Assembly | 127-mer SOAP Assembly | CLC and SOAP Hybrid Assembly | PLATANUS Assembly |
|--------------------------|----------------------|-----------------------|------------------------------|-------------------|
| Number of scaffolds      | 365,196              | 720,696               | 298,912                      | 48,012            |
| N50                      | 17.08 kb             | 3.52 kb               | 16.46 kb                     | 251.28 kb         |
| N75                      | 5.39 kb              | 1.91 kb               | 5.23 kb                      | 109.30 kb         |
| L50                      | 28,676               | 126,156               | 22,166                       | 1,873             |
| L75                      | 82,804               | 335,486               | 80,691                       | 4,545             |
| Length of largest contig | 1.33 Mb              | 473.87 kb             | 837.12 kb                    | 4.47 Mb           |
| GC content               | 39.53%               | 39.43%                | 39.38%                       | 38.98%            |

NOTE.—All statistics are based on scaffolds  $\geq 1,000$  bp. Abbreviations: kb, kilobase (=1,000 bp); Mb, megabase (=1,000,000 bp).**Table 2**Percentages of Different Repeat Elements in the *Cynopterus brachyotis* Genome

| Repeat Element         | Percentage of the Genome |
|------------------------|--------------------------|
| SINEs                  | 0.93                     |
| LINEs                  | 10.18                    |
| LTR elements           | 0.68                     |
| DNA elements           | 0.68                     |
| Unclassified           | 9.37                     |
| Simple repeats         | 2.4                      |
| Low complexity repeats | 0.27                     |

NOTE.—SINE, short-interspersed nuclear elements; LINE, long-interspersed nuclear elements; LTR, long-terminal repeats.

**Table 3**

Details of Gene Annotation from AUGUSTUS

| Gene Annotation  | Number of Genes Identified | Number of Genes with InterProScan ID | Number of Genes with Gene Ontology ID |
|--|----------------------------|--------------------------------------|---------------------------------------|
| Without cDNA hints   | 21,822                     | 19,371                               | 11,636                                |
| Using cDNA hints from the <i>Cynopterus sphinx</i> transcriptome | 23,727                     | 22,070                               | 14,390                                |

this study; human, *Homo sapiens*; rhesus macaque, *Mac. mulatta*; mouse, *Mus musculus*; rat, *Rat. norvegicus*; dog, *Can. lupus familiaris*; cat, *F. catus*; cow, *B. taurus*; horse, *Equ. caballus*; rhinoceros, *Cer. simum simum*; pig, *S. scrofa*; bottleneck dolphin, *T. truncatus*; big brown bat, *E. fuscus*; great leaf-nosed bat, *H. armiger*; Chinese rufous horseshoe bat, *Rhi. sinicus*; little brown bat, *M. lucifugus*; large flying fox, *P. vampyrus*; Egyptian fruit bat, *R. aegyptiacus*; and Natal long-fingered bat, *Min. natalensis*). The number of gene copies across families varied from 1 to 871.

### Phylogenetic Relationships and Divergence Dating

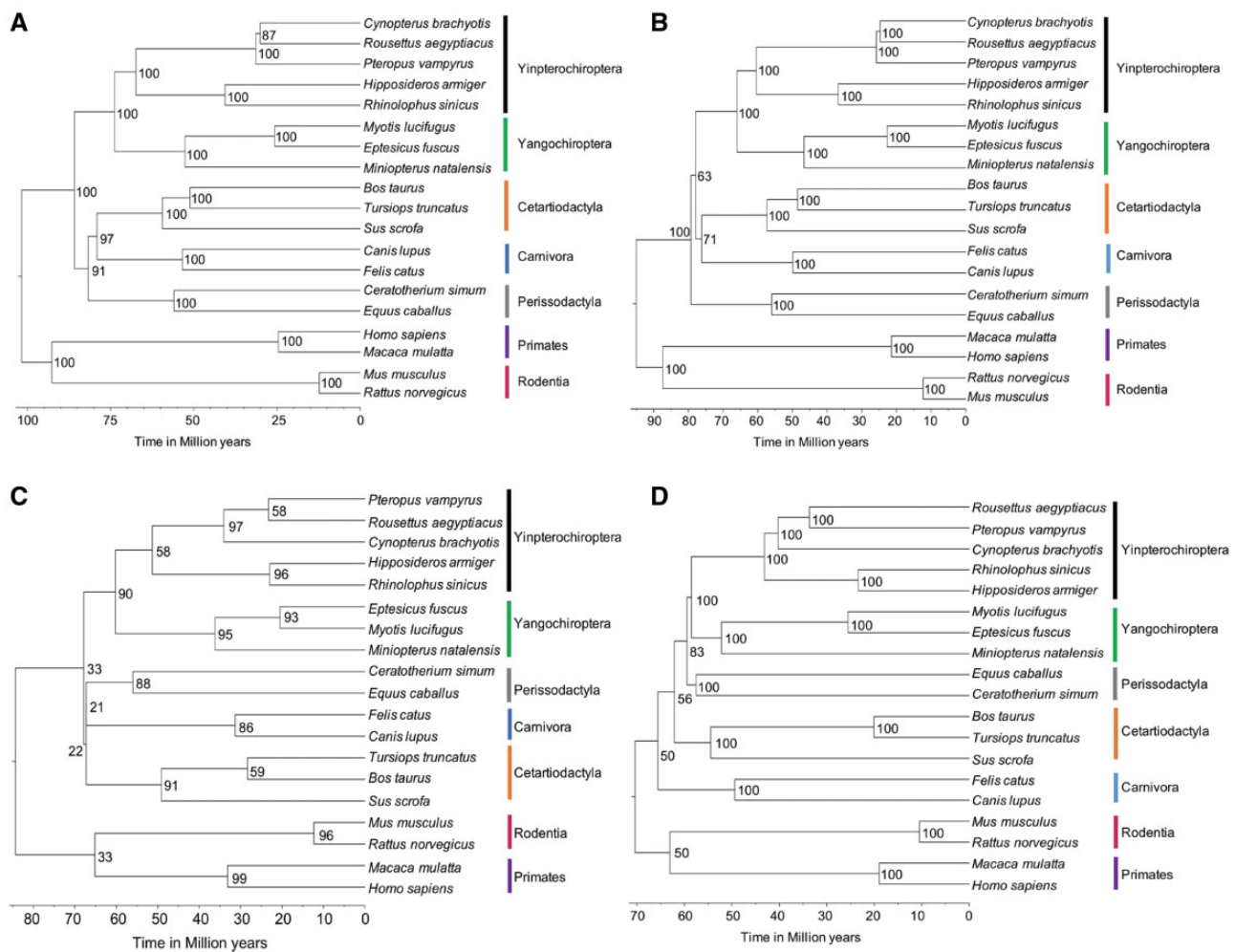
We used both concatenation and species tree approaches to understand the relationships among major mammalian lineages (all species used for the identification of orthologous

genes were also included in phylogenomic reconstructions, but specifically concentrated on relationships among bat lineages in general and fruit bats in particular. We utilized both protein sequences as well as nucleotide sequences for phylogenomic reconstructions.

Phylogenetic relationships of protein-based and DNA-based data sets were largely congruent in the concatenation-based reconstructions (fig. 1A and B). The species tree reconstructions did not provide as much resolution as concatenation-based trees (fig. 1C and D), but were congruent across well-supported branches with the exception of the placement of *C. brachyotis*, which emerged as sister to *R. aegyptiacus* in the concatenated DNA sequence analysis (fig. 1B) but as basal to *Rousettus* and *Pteropus* in the DNA-based species tree (fig. 1D), whereas protein-based results were generally poorly supported. All other phylogenetic relationships within bats were stable across trees, with Yinpterochiroptera and Yangochiroptera forming separate clades (fig. 1). Basal relationships among bats, Cetartiodactyla, Carnivora, and Perissodactyla were generally not well supported (fig. 1), consistent with many earlier phylogenomic mammalian data sets (Lindblad-Toh et al. 2011; Tsagkogeorga et al. 2013; Foley et al. 2016; Lei and Dong 2016). The initial diversification of bats ranged from 58.51 to 73.82 Ma, in agreement with the most rigorous published mammalian family dating study (Liu et al. 2017), and for Old World fruit bats from 25.76 to 40.27 Ma depending on the starting phylogenetic tree (table 4).

### Gene Family Expansions and Contractions

For all CAFE analyses, we observed a better model fit when allowing for multiple  $\lambda$  (birth–death rate) values as compared with a single  $\lambda$  model ( $P$  value  $< 0.001$ ). Three different values of  $\lambda$  were applied, one for all bats, the second one for primates and rodents, and the third for a clade consisting of ungulates, carnivores, and cetaceans. This latter clade did not emerge in all our phylogenomic analyses (fig. 1) but has been corroborated by other genomic studies with a wider general mammalian taxon sampling (Tsagkogeorga et al. 2013; Liu et al. 2017). We obtained estimates from three trees



**Fig. 1.**—Phylogenetic reconstructions showing evolutionary relationships of *Cynopterus brachyotis* with other mammalian taxa included in this study. (A) Maximum likelihood tree generated using concatenated data in RAxML based on 473,499 amino acids, (B) maximum likelihood tree generated using concatenated data in RAxML based on 9,353,867 bp of DNA sequence, (C) species tree reconstruction in MP-EST based on 298 single copy protein sequences, and (D) species tree reconstruction in MP-EST based on 3,194 DNA loci. Nodal values represent bootstrap support. Time of divergence is denoted in millions of years by a scale bar below the tree.

**Table 4**

Point Estimate of Age of the Most Recent Common Ancestor Computed by r8s for the Four Different Starting Phylogenetic Trees

| Most Recent Common Ancestor of | Tree Topology                         |                                   |  |                                    |
|--------------------------------|---------------------------------------|-----------------------------------|--|------------------------------------|
|                                | RAxML Tree Based on Protein Sequences | RAxML Tree Based on DNA Sequences | MP-EST Tree Based on Protein Sequences | MP-EST Tree Based on DNA Sequences |
| Fruit bats                     | 31.42                                 | 25.76                             | 34.05                                  | 40.27                              |
| Bats                           | 73.82                                 | 65.98                             | 60.20                                  | 58.51                              |

NOTE.—Ages are given in millions of years.

(concatenated protein tree, concatenated DNA sequence tree, and DNA sequence-based species tree) as these three trees had produced the highest overall resolution.

Concatenated trees resulted in the highest inference of significant expansion or contraction events of gene families along the nodes examined (table 5). However, there was significant discrepancy in the identification of gene family fluctuations, specifically within bats, between concatenated trees and species trees as the CAFE analysis based on the species tree did not return significant contractions or expansions of gene families for the MRCA of bats in general and fruit bats in particular, and only revealed significant gene family fluctuations in the MRCA of *C. brachyotis* (table 5). After correcting for errors in genome assembly and annotation, concatenated trees revealed 196–207 gene families with a significant expansion and 15–17 gene families with a significant contraction in the bat ancestor as compared with other mammals, and 1–3 gene families with a significant expansion in the fruit bat ancestor as compared with microbats (table 5). We

**Table 5**

Number of Gene Families Exhibiting Significant Expansion or Contraction in *Cynopterus brachyotis*, Fruit Bats, and Bats for the Three Different Phylogenies Tested

| Phylogenetic Tree                  | <i>C. brachyotis</i>              |                                     | Fruit Bats                        |                                     | Bats                              |                                     |
|------------------------------------|-----------------------------------|-------------------------------------|-----------------------------------|-------------------------------------|-----------------------------------|-------------------------------------|
|                                    | Number of Gene Families Expanding | Number of Gene Families Contracting | Number of Gene Families Expanding | Number of Gene Families Contracting | Number of Gene Families Expanding | Number of Gene Families Contracting |
| RAxML tree using protein sequences | 19                                | 32                                  | 1                                 | 0                                   | 196                               | 17                                  |
| RAxML tree using DNA sequences     | 19                                | 27                                  | 3                                 | 0                                   | 207                               | 15                                  |
| MP-EST tree using DNA sequences    | 14                                | 17                                  | 0                                 | 0                                   | 0                                 | 0                                   |

observed a significant expansion in 14–19 gene families and significant contraction in 17–32 gene families in *C. brachyotis* (tables 5–6B). The difference in the number of gene families exhibiting significant changes among phylogenomic reconstructions could be attributed to branch length differences (fig. 1). In case of species trees, the relationship between Chiroptera, Cetartiodactyla, Carnivora, and Perissodactyla is not well resolved (fig. 1D), resulting in poor resolution and hence a lack of pronounced gene family contraction or expansion.

Immune-related gene families and olfactory receptors seemed to undergo a significant contraction in bats in general, and in *C. brachyotis* in particular (table 6B and supplementary tables S4 and S5, Supplementary Material online), whereas fruit bats were characterized by a significant expansion of aging-related genes regulating telomerase activity (supplementary table S6, Supplementary Material online). The number of aging-related genes within fruit bats varied between 1 and 16, whereas one to five copies of the genes were observed within insectivorous bats (supplementary fig. S5, Supplementary Material online). Testing for gene duplication using the ETE toolkit, we found evidence for multiple duplication events.

Across bats, we detected a significant enrichment of GO terms related to neuron development and cellular activities such as cellular component organization or organelle organization in expanding gene families (fig. 2A and supplementary table S7, Supplementary Material online), whereas contracting gene families were significantly enriched for olfactory receptors (fig. 2B and supplementary table S7, Supplementary Material online). No expanding or contracting gene family showed significant GO enrichment within fruit bats. In *C. brachyotis*, there was significant enrichment of cell adhesion molecules and structural molecules in expanding gene families (supplementary table S8, Supplementary Material online). At the same time, *C. brachyotis* was affected by contractions of gene families that showed a significant enrichment for olfactory receptors, nucleosome, DNA packaging, and protein–DNA complex as well as signaling molecules (supplementary table S8, Supplementary Material online).

Interestingly, for all gene families showing significant expansions in *C. brachyotis* in CAFE (table 6A), both species overlap methods and species reconciliation methods equally revealed multiple episodes of expansion events. The same agreement among methods applied to contracting gene families in *C. brachyotis* (see IPython notebook for examples).

## Discussion

Bats display a remarkable repertoire and diversity of sensory modalities and a greater range of physiological and ecological specializations than any other mammals (Teeling et al. 2018). In the present study, we sequenced the first genome of a *Cynopterus* fruit bat to add to our knowledge of links between functional and genic diversity across bats. A rigorous comparison across four types of genome assembly demonstrated that pipelines specifically designed for heterozygous genomes considerably improve assembly quality (table 1). The genome size estimate (~1.7 Gb) for *C. brachyotis* is comparable to other bat genomes and, as expected, falls in the lower spectrum of genome size in mammals (Kapusta et al. 2017; Teeling et al. 2018; Wen et al. 2018). Just like birds, bats have a small, streamlined genome, which has been linked to a reduction of redundancy to facilitate flight (Kapusta et al. 2017; Teeling et al. 2018).

### Bat Genome Evolution Reflects Ecological Release in the Early Paleogene

We dated the MRCA of bats to around the Cretaceous–Tertiary boundary (based on concatenation) or a few million years afterward (based on species tree methods; table 4), in good agreement with other recent studies (Teeling 2005; Lei and DoNg 2016; Bhak et al. 2017; Dong et al. 2017; Liu et al. 2017). This timing places the beginnings of bats into the early Paleogene, a time when the Earth had just passed through the K-Pg Boundary mass extinction crisis, with surviving lineages undergoing explosive radiations against the background of the ecological release exerted by vacant niches in a



**Table 6**

Comparison of Number of Gene Copies in *Cynopterus brachyotis*, Fruit Bats (excluding *C. brachyotis*), Bats (excluding *C. brachyotis*), and Mammals (excluding bats) for (A) Gene Families Exhibiting Expansion within *C. brachyotis* and (B) Gene Families Exhibiting Contraction within *C. brachyotis*

| Gene Family ID | Example Protein within the Gene Family            | Number of Genes Identified in <i>C. brachyotis</i> | Average Number of Genes Identified across Fruit Bats Excluding <i>C. brachyotis</i> | Average Number of Genes Identified across Bats Excluding <i>C. brachyotis</i> | Average Number of Genes Identified across Mammals Excluding Bats |
|----------------|---|--|---|---|--|
| <b>(A)</b>     |   |  |   |   |  |
| OG0000313      | Microtubule-actin crosslinking factor 1           | 12   | 2   | 8.71  | 1.73   |
| OG0000353      | Ribosomal protein L23a                            | 9  | 1.5   | 2   | 5.91   |
| OG0000372      | RNA-binding motif protein 23                      | 14   | 5   | 7.43  | 1.82   |
| OG0000374      | LDL receptor-related protein 1B                   | 14   | 4.5   | 4.86  | 3.45   |
| OG0000648      | Ferritin light chain                              | 7  | 1   | 3.29  | 3.64   |
| OG0001063      | Hemicentin 1                                      | 10   | 2.5   | 2.71  | 2.64   |
| OG0001804      | Proline-rich coiled-coil 2C                       | 9  | 1   | 3.86  | 1  |
| OG0002597      | ATP-binding cassette, subfamily D (ALD), member 4 | 8  | 1.5   | 2.86  | 1  |
| OG0006526      | Sodium voltage-gated channel alpha subunit 7      | 8  | 0   | 0.86  | 1  |
| OG0007974      | Tubulin alpha 4a                                  | 6  | 0   | 1.14  | 0.82   |
| <b>(B)</b>     |   |  |   |   |  |
| OG0000007      | Olfactory receptor family 6 subfamily C member 74 | 0  | 21  | 23.14   | 34   |
| OG0000011      | Histone cluster 1 H2B family member a             | 2  | 19  | 17.14   | 21.91  |
| OG0000012      | Histone cluster 1 H2A family member a             | 2  | 18.5  | 16.71   | 19.64  |
| OG0000020      | Major histocompatibility complex, class I, A      | 6  | 29.5  | 27.43   | 4.27   |
| OG0000024      | Lymphoid enhancer-binding factor 1                | 2  | 8.5   | 9.57  | 14.09  |
| OG0000025      | Histone cluster 2 H3 family member d              | 0  | 5   | 14.43   | 11   |
| OG0000027      | Leukocyte immunoglobulin-like receptor B5         | 3  | 27  | 24.29   | 4.18   |
| OG0000029      | IKAROS family zinc finger 3                       | 2  | 9   | 8.43  | 14.18  |
| OG0000043      | Olfactory receptor family 4 subfamily F member 16 | 0  | 7.5   | 5.71  | 13.45  |
| OG0000080      | Olfactory receptor family 8 subfamily K member 1  | 0  | 8.5   | 6.14  | 9.91   |
| OG0000110      | Olfactory receptor family 2 subfamily A member 4  | 0  | 8   | 5.43  | 8.73   |
| OG0000156      | Abl interactor 1                                  | 1  | 12  | 12.86   | 2.36   |
| OG0000185      | Killer cell lectin-like receptor C4               | 0  | 21  | 10.86   | 3.18   |

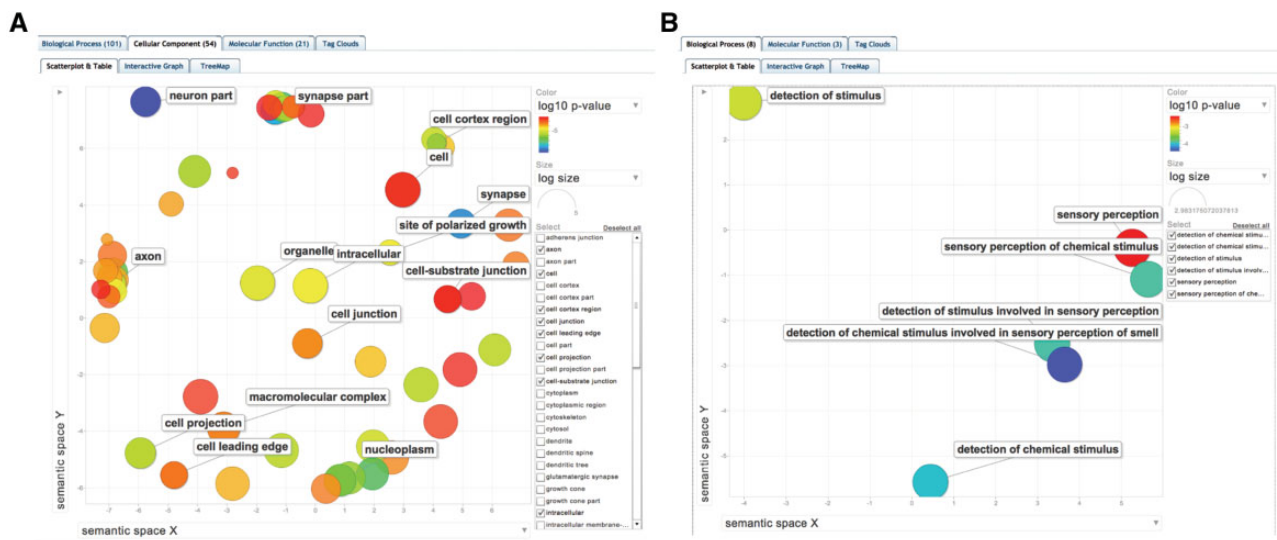
depauperate landscape. The explosive diversification of other notable vertebrate lineages, such as Neoaves (Jarvis et al. 2014), roughly coincides with this scenario.

Ecological release during this time would have been the trigger for bats to come up with novel adaptations and specializations to colonize diverse environments and adapt to new sensory niches requiring genomic modifications. Consequently, we observed an expansion of gene families involved in cellular processes and neuron development (fig. 2A and supplementary tables S5 and S7, Supplementary Material online). The nervous system of bats plays an important role especially in those species that rely on echolocation to navigate, identify prey, and communicate (Altringham 1999), and has undergone numerous modifications to accommodate flight and echolocation (Covey 2005), which are likely reflected in our detection of a significant expansion and enrichment of genes related with the nervous system (fig. 2A and supplementary tables S5 and S7, Supplementary Material online).

We detected patterns of gene family evolution in bats that were in close agreement with previous comparative genomic enquiries (Zhang et al. 2013; Dong et al. 2017; Tsagkogeorga et al. 2017). For example, we noted an expansion of gene families involved in metabolic regulation, cellular organization, and development (fig. 2A and supplementary tables S5 and S7, Supplementary Material online) as well as a considerable decline in olfactory receptors (fig. 2B and supplementary tables S5 and S7, Supplementary Material online).

#### Fruit Bats' Remarkable Shift in Sensory and Metabolic Evolution Is Not Closely Mirrored in Their Genomes

Our reconstructions place the division of fruit bats (family Pteropodidae) from other bats ~5–12 Myr after the emergence of bats (fig. 1), a timing that is largely in agreement with Liu et al. (2017), who used a wider taxon sampling for bats along with a more extensive calibration regime. Because of their great phenotypic, ecological, and physiological



**FIG. 2.**—Scatter plot of GO terms based on semantic similarity identified in bats: (A) exhibiting significant expansion and (B) exhibiting significant contraction.

distinctions, fruit bats are known as an ancient lineage within bats characterized by exceptional trait evolution. Yet, despite the confirmed old age of fruit bats within the bat radiation, we uncovered a surprisingly limited scope of gene family expansions or contractions. This is unexpected, given that most fruit bats lack echolocation, are significantly larger than microbats and differ in diet and other sensory abilities.

Interestingly, one of the only significant gene family expansions we did detect in fruit bats relates to loci coding for the protection of telomerases (*POT1*) (supplementary fig. S5 and table S6, Supplementary Material online). These genes protect the ends of chromosomes by regulating telomere length (Loayza and De Lange 2003). Telomeres have been implicated in longevity in mammals (Morgan et al. 2013) including bats (Foley et al. 2018). The three fruit bats examined in this study are all long-lived species (*P. vampyrus*, 15 years; *R. aegyptiacus*, ~22 years; and *C. brachyotis*, 20–30 years) (<https://animaldiversity.org/>) in the wild, although some microbats are known to live even longer (Foley et al. 2018). In the microbat genus *Myotis*, for instance, telomeres do not shorten in size with age and the genes *ATM* and *SETX* which repair and prevent DNA damage may be responsible for protecting telomeres (Foley et al. 2018). Our study improves upon this information by recovering a gene—*POT1*—which is directly responsible for the protection of telomeres in Old World fruit bats, thereby providing evidence of a link between longevity in fruit bats and the possible genes responsible. This expansion signal is mainly driven by *R. aegyptiacus* and *C. brachyotis*, and future studies with high coverage genomes and a greater taxonomic depth might provide a better resolution in this aspect.

### Fruit Bats Likely Underwent a Rapid Radiation

Our comparative genomic approach contrasts gene family evolution among three fruit bat subfamilies. However, our phylogenomic reconstruction (fig. 1) of the relationships of these three subfamilies (Cynopterinae represented by *C. brachyotis*; Rousettinae represented by *R. aegyptiacus*; and Pteropodinae represented by *P. vampyrus*) has added to previous conflicting results in the literature (Teeling 2005; Almeida et al. 2011; Lei and DoNg 2016). Concatenated transcriptomic analyses previously indicated a sister relationship between *Cynopterus* and *Rousettus* (Lei and DoNg 2016), as corroborated by our concatenation-based trees (fig. 1A and B). However, our species tree reconstructions returned *Cynopterus* as basal to a monophyletic *Pteropus*–*Rousettus* clade with partially high support (fig. 1C and D), in agreement with Teeling (2005). The diversification of these three fruit bat subfamilies likely occurred during an Oligocene rapid radiation (fig. 1 and table 4) accompanied by possible incomplete lineage sorting, rendering the exact resolution of the sequence of divergence events difficult. Insect bat lineages characterized by rapid radiation dynamics reveal a similar evolutionary footprint (Platt et al. 2018). Future divergence dating with more Old World fruit bat genomes across all subfamilies should help obtain a finer resolution of the timing of this explosive radiation. Although the exact basal topology of Old World fruit bats must remain contentious for now, recent genome studies demonstrated that species tree methods—as compared with concatenation methods—are uniquely suited in correctly retrieving the phylogenomic information content across large numbers of unlinked loci (Jarvis et al. 2014; Liu et al. 2017).

### Contraction of Regulatory, Immune, and Sensory Genes in *C. brachyotis*

*Cynopterus brachyotis* is characterized by unusual shifts in gene family evolution. It has undergone a significant decline in gene families coding for histone proteins (table 6B and supplementary tables S4 and S8, Supplementary Material online) that bind to DNA, help in packaging DNA (Wang et al. 2008), and regulate gene expression. This result emerged on the basis of all three starting tree topologies (supplementary tables S4 and S8, Supplementary Material online). For example, in the histone cluster 1 H2A gene family, humans have 25 genes coding for H2A, whereas *C. brachyotis* has only 2 (table 6B). Although this trend might be an effective mechanism to reduce genome size and genomic redundancy, such gene attrition can also indicate a possible role in the active regulation of gene expression. Multiple unique or novel genes in the *C. brachyotis* genome regulate deacetylation of histone molecules (supplementary table S3, Supplementary Material online), thus helping in regulating gene expression. These two trends combined indicate a possibility of gene regulation through the evolution of histone genes in *C. brachyotis*. This finding may have a bearing on the diversification of bat lineages as regulation of gene expression is reported to be an important process influencing adaptation through gene–phenotype connections (Teeling et al. 2018 and references therein).

We observed an overall contraction of immune gene families coding for both innate and adaptive immunity in bats, with an even further reduction trend in *C. brachyotis* (table 6B and supplementary tables S4 and S5, Supplementary Material online). Bats are known reservoirs of many viruses, but infections are rarely pathogenic (Beltz 2017; Pavlovich et al. 2018; Teeling et al. 2018), possibly on account of immune response regulation, specifically the suppression of inflammatory responses, thereby reducing the subsequent pathology resulting from viral infections (Amman et al. 2015; Jones et al. 2015; Banerjee et al. 2017; Beltz 2017; Schuh et al. 2017). By regulating the natural killer cell pathway, bats may be able to modulate the inflammatory response (Pavlovich et al. 2018 and references therein). In *C. brachyotis*, we observed a decline in gene families coding for natural killer cell receptors and other receptors that interact with MHC class I molecules along with a decline in the number of gene families coding for MHC class I molecules (table 6B and supplementary table S4, Supplementary Material online). *Cynopterus brachyotis* therefore constitutes the most extreme example to date for immune gene contractions accompanied by nonpathogenicity of viral infections.

Interestingly, genomic studies have provided evidence of both expansion and contraction of immune-related genes in different species of bats (Shaw et al. 2012; Zhang et al. 2013; Zhou et al. 2016; Pavlovich et al. 2018), suggesting that bats may have evolved diverse strategies to avert viral infections. For example, *R. aegyptiacus* is characterized by an expansion

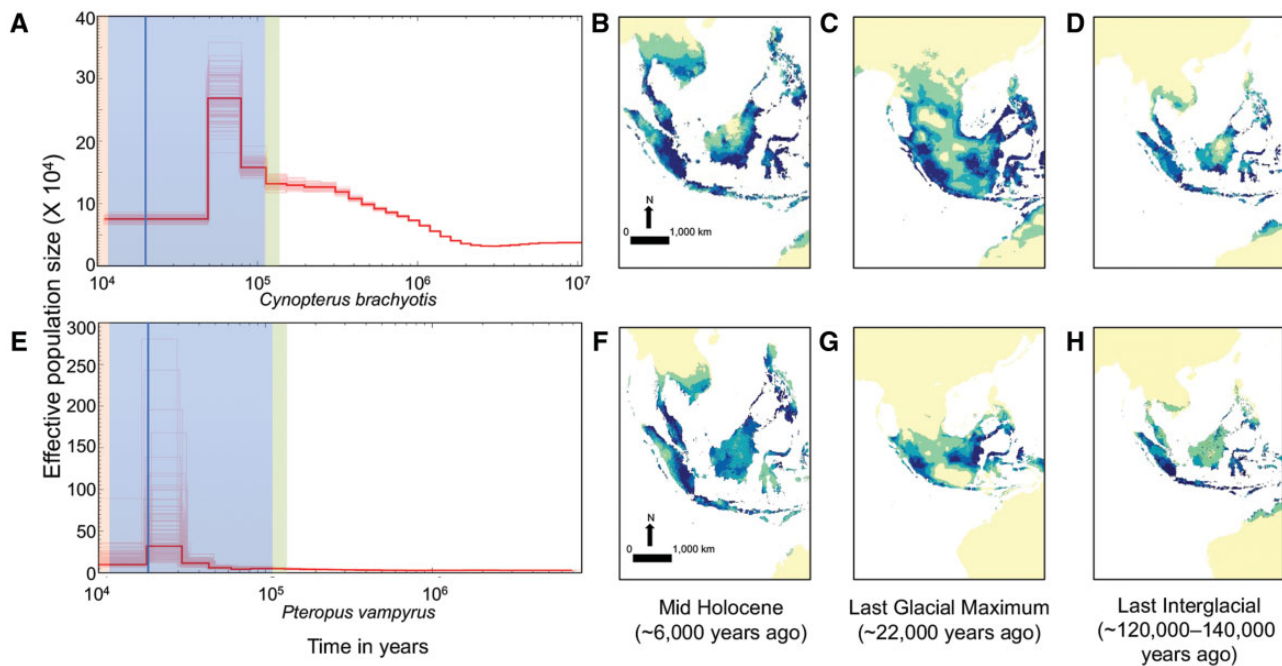
of MHC class I molecules and inhibitory natural killer cell receptors (Pavlovich et al. 2018), whereas in *Pteropus alecto* and *E. fuscus* the MHC class I complex has contracted and lacks the  $\alpha$  or  $\kappa$  duplication blocks (Ng et al. 2016). In *C. brachyotis*, we observed a contraction in the interferon  $\alpha$  gene family which provides a first line of defense for viral infections (supplementary table S4, Supplementary Material online), a trend also observed in *P. alecto* (Zhou et al. 2016).

Another novel finding was the contraction in olfactory receptor families 2, 4, 5, 6, 8, and 13 in the *C. brachyotis* genome when compared with other bats (supplementary table S4, Supplementary Material online). Other Old World fruit bats are associated with olfactory receptor families 2 and 13 (Hayden et al. 2014), which show a decline in *C. brachyotis* (supplementary table S4, Supplementary Material online). On the other hand, the olfactory receptor gene families (1, 3, and 7) associated with frugivory (Hayden et al. 2014) did not undergo any contraction in *C. brachyotis* when compared with *P. vampyrus* and *R. aegyptiacus*. Olfaction in fruit bats serves a dual purpose, that of identification of ripe fruits and of pheromones. In a comparative analysis of olfactory receptor genes across mammals, Hayden et al. (2014) documented a remarkable diversity in bat olfactory receptor genes closely linked to ecological specialization. Our results confirm that this diversity of olfactory strategies extends well below the subfamily level in fruit bats.

When compared with three other bat genomes and the human genome, we identified 143 genes unique to *C. brachyotis* (supplementary fig. S4 and table S3, Supplementary Material online) as well as 2,284 genes simultaneously present in all other bats and humans but absent in *C. brachyotis* (supplementary fig. S4 and table S3, Supplementary Material online). Although differences in genome assembly quality may play a role in generating gaps in gene coverage, the unusually high number of genes missing only in *Cynopterus* advocates a biological explanation (supplementary fig. S4, Supplementary Material online). Some of the genes missing in *Cynopterus* coincide with the gene families shown to have contracted in this genus (e.g., genes coding for histones, olfaction, and immunity; table 6B and supplementary table S3B, Supplementary Material online). BUSCO analysis of single copy orthologs suggests a similar percentage of single copy conserved genes identified across bats irrespective of the genome coverage (supplementary table S2, Supplementary Material online). Slight deviations in our conclusions on gene family fluctuations with those of other studies are likely attributable to our more comprehensive taxon sampling in comparison with Dong et al. (2017) and our practice of only considering statistically significant fluctuations in contrast to Tsagkogeorga et al. (2017).

### Genomic Correlates of Macroevolutionary Change

The unusual pattern of gene family evolution in *Cynopterus* when compared with the other two fruit bat genomes



**FIG. 3.**—(A and E) Quaternary fluctuations in effective population size in *Cynopterus brachyotis* and *Pteropus vampyrus* based on complete sequence data (dark red), with bootstraps depicted in light red, assuming a generation time of 8 years and a mutation rate of  $2.2 \times 10^{-9}$  per base pair per year. Colored highlights refer to: early Holocene (light orange; 10,000–12,000 years ago), last glacial period (light blue; ~12,000–110,000 years ago), last glacial maximum (dark blue line; ~22,000 years ago), and last interglacial (light green; 110,000–130,000 years ago). (B–G) Ecological niche models of *C. brachyotis* and *P. vampyrus* for different time periods: (B and F) mid-Holocene (~6,000 years ago), (C and G) last glacial maximum (~22,000 years ago), and (D and H) last interglacial period (~120,000–140,000 years ago). The following colors designate the probability of presence in ecological niche model maps: (0–0.1) pale yellow, (0.1–0.3) light green, (0.3–0.5) pale blue, (0.5–0.7) light blue, and (0.7–1) dark blue.

(*Rousettus* and *Pteropus*) is surprising, considering that all three fruit bats share a frugivorous diet, the loss (or rudimentary use) of echolocation, and good eye-sight and olfactory capabilities, while having diversified during similar evolutionary times on the occasion of a rapid radiation (see above). Despite their similar phenotype and sensory abilities, fruit bats seem to have evolved different genomic solutions to physiological and environmental challenges. Immune-related gene families seem to be most affected by significant contractions and expansions across fruit bats (Zhang et al. 2013; Ng et al. 2016; Zhou et al. 2016; Pavlovich et al. 2018), and are a likely trigger for such genomic change in the absence of phenotypic and physiological differences: bats' exposure, susceptibility, and nonpathogenicity vis-à-vis viral infections may have generated a rich genomic landscape of response mechanisms, and whereas these different responses would be readily detectable as gene family expansions or contractions, they would not be phenotypically obvious. PSMC analysis and ancestral habitat reconstructions of two of the three fruit bats in our study (*Cynopterus* and *Pteropus*) indicate that they have largely had different responses to Quaternary climatic oscillations, with *Pteropus* showing signs of historically greater levels of genetic diversity and also suitable habitat during the peak of the last glaciation,

whereas *Cynopterus* underwent larger declines in genetic diversity but major increments in suitable habitat, especially during the coldest parts of the most recent glaciation (fig. 3). These patterns are indicative of different tolerance thresholds to temperatures, precipitation, but perhaps also to shifting pathogenic environments. The sampling of additional fruit bat genomes is a high priority to shed light on the unique evolutionary trajectories of this fascinating animal lineage, with potential implications for our understanding about mammalian viral response evolution.

## Conclusions

In this study, we generated a good-quality genome of the paleotropical fruit bat *C. brachyotis* and performed comparative analyses to understand genomic contingents of physiological trait evolution. Our observations revealed that Old World fruit bats underwent major shifts in their sensory and metabolic capabilities, but exhibit less-significant signatures of change in genic composition in their genomes. Within our panel of Old World fruit bats, we discovered a hitherto unknown signal of gene family expansion directly linked to telomerase protection and longevity. Among the three paleotropical fruit bats with similar lifestyles analyzed in this



study, *C. brachyotis* revealed a unique incidence of gene family loss specifically with regards to regulation, immunity, and olfaction, suggesting that fruit bats employ diverse strategies for responses to viral infections and olfaction.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

The authors thank Ben Fulton for help with CAFE analyses. This work was supported by South East Asian Biodiversity Genomics (SEABIG) (Grant Numbers WBS R-154-000-648-646 and WBS R-154-000-648-733).

## Author Contributions

B.C., K.M.G., and F.E.R. conceived and designed the study. I.H.M. contributed sample material. K.M.G. and B.C. performed lab work. B.C. and K.M.G. performed all genomic analyses. R.R. performed ecological niche modeling. All analyses were performed with input from F.E.R. B.C., K.M.G., and F.E.R. wrote the article with contributions from R.R. and I.H.M.

## Literature Cited

- Ahn M, Cui J, Irving AT, Wang L-F. 2016. Unique loss of the PYHIN gene family in bats amongst mammals: implications for inflammasome sensing. *Sci Rep.* 6(1):21722.
- Alberto FJ, et al. 2018. Convergent genomic signatures of domestication in sheep and goats. *Nat Commun.* 9(1):813.
- Almeida FC, Giannini NP, DeSalle R, Simmons NB. 2011. Evolutionary relationships of the old world fruit bats (Chiroptera, Pteropodidae): another star phylogeny? *BMC Evol Biol.* 11:281.
- Altringham JD. 1999. *Bats: biology and behavior*. New York: Oxford University Press.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215(3):403–410.
- Amman BR, et al. 2015. Oral shedding of Marburg virus in experimentally infected Egyptian fruit bats (*Rousettus aegyptiacus*). *J Wildl Dis.* 51(1):113–124.
- Axelsson E, et al. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495(7441):360–364.
- Banerjee A, Rapin N, Bollinger T, Misra V. 2017. Lack of inflammatory gene expression in bats: a unique role for a transcription repressor. *Sci Rep.* 7(1):2232.
- Bao Z, Eddy SR. 2002. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* 12(8):1269–1276.
- Beltz LA. 2017. *Bats and human health: Ebola, SARS, rabies and beyond*. New Jersey: John Wiley & Sons.
- Bernt M, et al. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol.* 69(2):313–319.
- Bhak Y, et al. 2017. *Myotis rufoniger* genome sequence and analyses: *M. rufoniger's* genomic feature and the decreasing effective population size of *Myotis* bats. *PLoS One* 12(7):e0180418.
- Boyle EI, et al. 2004. GO:: termFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* 20(18):3710–3715.
- Campbell P. 2008. The relationship between roosting ecology and degree of polygyny in harem-forming bats: perspectives from *Cynopterus*. *J Mammal.* 89(6):1351–1360.
- Campbell P, Schneider CJ, Adnan AM, Zubaid A, Kunz TH. 2004. Phylogeny and phylogeography of Old World fruit bats in the *Cynopterus brachyotis* complex. *Mol Phylogenet Evol.* 33(3):764–781.
- Castoe TA, et al. 2013. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc Natl Acad Sci U S A.* 110(51):20645–20650.
- Chattopadhyay B, et al. 2016. Genome-wide data reveal cryptic diversity and genetic introgression in an oriental cynopterine fruit bat radiation. *BMC Evol Biol.* 16(1):41.
- Chattopadhyay B, Garg KM, Doss PS, Ramakrishnan U, Kandula S. 2011. Molecular genetic perspective of group-living in a polygynous fruit bat, *Cynopterus sphinx*. *Mammal Biol.* 76(3):290–294.
- Chong HT, Abdullah S, Tan CT. 2009. Nipha virus and bats. *Neurol Asia.* 14:73–76.
- Covey E. 2005. Neurobiological specializations in echolocating bats. *Anat Rec.* 287:1103–1116.
- Dierckxens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45(4):e18.
- Dong D, et al. 2017. The genomes of two bat species with long constant frequency echolocation calls. *Mol Biol Evol.* 34(1):20–34.
- Dong D, Lei M, Liu Y, Zhang S. 2013. Comparative inner ear transcriptome analysis between the Rickett's big-footed bats (*Myotis ricketti*) and the greater short-nosed fruit bats (*Cynopterus sphinx*). *BMC Genomics* 23:916.
- Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16(1):157.
- Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30(7):1575–1584.
- Foley NM, et al. 2018. Growing old, yet staying young: the role of telomeres in bats' exceptional longevity. *Sci Adv.* 4(2):eaao0926.
- Foley NM, Springer MS, Teeling EC. 2016. Mammal madness: is the mammal tree of life not yet resolved? *Philos Trans R Soc B.* 371(1699):20150140.
- Garg KM, et al. 2012. Promiscuous mating in the harem-roosting fruit bat, *Cynopterus sphinx*. *Mol Ecol.* 21(16):4093–4105.
- Götz S, et al. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36(10):3420–3435.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29(8):1072–1075.
- Han MV, Thomas GW, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol.* 30(8):1987–1997.
- Hayden S, et al. 2014. A cluster of olfactory receptor genes linked to frugivory in bats. *Mol Biol Evol.* 31(4):917–927.
- Huerta-Cepas J, Dopanzo H, Dopanzo J, Gabaldón T. 2007. The human phylome. *Genome Biol.* 8(6):R109.
- Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol.* 33(6):1635–1638.
- Jarvis ED, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346(6215):1320–1331.
- Jones M, et al. 2015. Experimental inoculation of Egyptian rousette bats (*Rousettus aegyptiacus*) with viruses of the Ebolavirus and Marburgvirus genera. *Viruses* 7(7):3420–3442.
- Kajitani R, et al. 2014. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* 24(8):1384–1395.

- Kapusta A, Suh A, Feschotte C. 2017. Dynamics of genome size evolution in birds and mammals. *Proc Natl Acad Sci U S A*. 114(8):E1460–E1469.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.
- Kim EB, et al. 2011. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature* 479(7372):223–227.
- Kunz TH, Parsons S. 2009. Ecological and behavioral methods for the study of bats. 2nd ed. Baltimore: Johns Hopkins University Press.
- Laing ED, et al. 2018. Serologic evidence of fruit bat exposure to filoviruses, Singapore, 2011–2016. *Emerg Infect Dis*. 24:122.
- Lei M, Dong D. 2016. Phylogenomic analyses of bat subordinal relationships based on transcriptome data. *Sci Rep*. 6(1):27726.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493–496.
- Li H, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
- Lindblad-Toh K, et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478(7370):476–482.
- Liu L, et al. 2017. Genomic evidence reveals a radiation of placental mammals uninterrupted by the KPg boundary. *Proc Natl Acad Sci U S A*. 114(35):E7282–E7290.
- Loayza D, De Lange T. 2003. POT1 as a terminal transducer of TRF1 telomere length control. *Nature* 423(6943):1013–1018.
- Luo R, et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1(1):18.
- Mani R, et al. 2017. Serological evidence of lyssavirus infection among bats in Nagaland, a north-eastern state in India. *Epidemiol Infect*. 145(8):1635–1641.
- Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27(6):764–770.
- McCracken GF, Wilkinson GS. 2000. Bat mating systems. In: Crichton EG, Krutzsch PH, editors. *Reproductive biology of bats*. SanDiego: Academic Press. p. 321–362.
- Morgan CC, et al. 2013. Molecular adaptation of telomere associated genes in mammals. *BMC Evol Biol*. 13(1):251.
- Ng JH, et al. 2016. Evolution and comparative analysis of the bat MHC-I region. *Sci Rep*. 6(1):21256.
- Page RD, Charleston MA. 1997. From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem. *Mol Phylogenet Evol*. 7(2):231–240.
- Pavlovich SS, et al. 2018. The Egyptian rousette genome reveals unexpected features of bat antiviral immunity. *Cell* 173(5):1098–1110.
- Platt RN, et al. 2018. Conflicting evolutionary histories of the mitochondrial and nuclear genomes in new world *Myotis* bats. *Syst Biol*. 67(2):236–249.
- Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. *Bioinformatics* 21(Suppl 1):i351–i358.
- Qu Y, et al. 2013. Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. *Nat Commun*. 4(1):2071.
- Smit A, Hubley R. 2015. RepeatModeler Open-1.0. [Internet]. 2008–2015. Available from: <http://www.repeatmasker.org/>. Accessed February 28, 2020.
- Smit A, Hubley R, Green P. 2015. RepeatMasker Open-4.0. [Internet]. 2013–2015. Available from: <http://www.repeatmasker.org/>. Accessed February 28, 2020.
- Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19(2):301–302.
- Schountz T. 2014. Immunology of bats and their viruses: challenges and opportunities. *Viruses* 6(12):4880–4901.
- Schuh AJ, et al. 2017. Modelling filovirus maintenance in nature by experimental transmission of Marburg virus between Egyptian rousette bats. *Nat Commun*. 8(1):14446.
- Sharma V, et al. 2018. A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nat Commun*. 9(1):1215.
- Shaw TI, et al. 2012. Transcriptome sequencing and annotation for the Jamaican fruit bat (*Artibeus jamaicensis*). *PLoS One* 7(11):e48472.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212.
- Smith JD, Bickham JW, Gregory TR. 2013. Patterns of genome size diversity in bats (order Chiroptera). *Genome* 56(8):457–472.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Stanke M, et al. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res*. 34(Web Server):W435–W439.
- Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6(7):e21800.
- Teeling EC. 2005. A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* 307(5709):580–584.
- Teeling EC, et al. 2018. Bat biology, genomes, and the Bat1K project: to generate chromosome-level genomes for all living bat species. *Annu Rev Anim Biosci*. 6(1):23–46.
- Tsagkogeorga G, Müller S, Dessimoz C, Rossiter SJ. 2017. Comparative genomics reveals contraction in olfactory receptor genes in bats. *Sci Rep*. 7(1):259.
- Tsagkogeorga G, Parker J, Stupka E, Cotton JA, Rossiter SJ. 2013. Phylogenomic analyses elucidate the evolutionary relationships of bats. *Curr Biol*. 23(22):2262–2267.
- Wang Z, et al. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet*. 40(7):897.
- Wang Y, Coleman-Derr D, Chen G, Gu YQ. 2015. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res*. 43(W1):W78–W84.
- Waterhouse RM, et al. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol*. 35(3):543–548.
- Wen M, et al. 2018. Exploring the genome and transcriptome of the cave nectar bat *Eonycteris spelaea* with PacBio long-read sequencing. *Gigascience* 7(10):gij116.
- Xu H, et al. 2012. FastUniq: a fast de novo duplicates removal tool for paired short reads. *PLoS One* 7(12):e52249.
- Yoon KB, Kim JY, Park YC. 2016. Characteristics of complete mitogenome of the lesser short-nosed fruit bat *Cynopterus brachyotis* (Chiroptera: Pteropodidae) in Malaysia. *Mitochondrial DNA A*. 27:2091–2092.
- Zdobnov EM, et al. 2017. OrthoDB v9. 1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res*. 45(D1):D744–D749.
- Zhang G, et al. 2013. Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science* 339(6118):456–460.
- Zhou P, et al. 2016. Contraction of the type I IFN locus and unusual constitutive expression of IFN- $\alpha$  in bats. *Proc Natl Acad Sci U S A*. 113(10):2696–2701.
- Zubaid A, McCracken GM, McCracken GF, Kunz T. 2006. *Functional and evolutionary ecology of bats*. New York: Oxford University Press.

Associate editor: Federico Hoffmann