Received: 18 February 2020 | Revised: 22 June 2020 | Accepted: 30 June 2020

DOI: 10.1111/mec.15542



ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

# Molecular adaptation and convergent evolution of frugivory in Old World and neotropical fruit bats

Kai Wang[1,2] | Shilin Tian[1,3] | Jorge Galindo-González[4] | Liliana M. Dávalos[5] | Yuzhi Zhang[1] | Huabin Zhao[1,6]

[1]Department of Ecology, Tibetan Centre for Ecology and Conservation at WHU-TU, Hubei Key Laboratory of Cell Homeostasis, College of Life Sciences, Wuhan University, Wuhan, China

[2]The State Key Laboratory of Biocatalysis and Enzyme Engineering of China, College of Life Sciences, Hubei University, Wuhan, China

[3]Novogene Bioinformatics Institute, Beijing, China

[4]Biotechnology and Applied Ecology Institute (INBIOTECA), Universidad Veracruzana, Xalapa, Veracruz, Mexico

[5]Department of Ecology and Evolution and Center for Inter-Disciplinary Environmental Research, State University of New York at Stony Brook, Stony Brook, NY, USA

[6]College of Science, Tibet University, Lhasa, China

**Correspondence**
Huabin Zhao, Department of Ecology, College of Life Sciences, Wuhan University, 299 Bayi Road, Wuhan, Hubei 430072, China.
Email: huabinzhao@whu.edu.cn

**Funding Information**
This work was supported by the National Natural Science Foundation of China (31722051 and 31672272) and Natural Science Foundation of the Hubei Province (2019CFA075). L.M.D. was supported, in part by United States National Science Foundation awards DEB 1442142 and 1456455.

## Abstract

Although cases of independent adaptation to the same dietary niche have been documented in mammalian ecology, the molecular correlates of such shifts are seldom known. Here, we used genomewide analyses of molecular evolution to examine two lineages of bats that, from an insectivorous ancestor, have both independently evolved obligate frugivory: the Old World family Pteropodidae and the neotropical subfamily Stenodermatinae. New genome assemblies from two neotropical fruit bats (*Artibeus jamaicensis* and *Sturnira hondurensis*) provide a framework for comparisons with Old World fruit bats. Comparative genomics of 10 bat species encompassing dietary diversity across the phylogeny revealed convergent molecular signatures of frugivory in both multigene family evolution and single-copy genes. Evidence for convergent molecular adaptations associated with frugivorous diets includes the composition of three subfamilies of olfactory receptor genes, losses of three bitter taste receptor genes, losses of two digestive enzyme genes and convergent amino acid substitutions in several metabolic genes. By identifying suites of adaptations associated with the convergent evolution of frugivory, our analyses both reveal the extent of molecular mechanisms under selection in dietary shifts and will facilitate future studies of molecular ecology in mammals.

KEYWORDS

adaptation, bats, convergence, diet, genome, pseudogenization

## 1 | INTRODUCTION

Finding, obtaining and digesting food places some of the most powerful selective demands in all organisms, resulting in a range of morphological and physiological adaptations often characterizing vertebrate adaptive radiations (Martin & Richards, 2019). Although the molecular basis of dietary variation is seldom known, recent advances in comparative genomics have revealed unique signals of molecular dietary adaptations in insects (Duncan et al., 2014), hummingbirds (Baldwin et al., 2014) and snakes (Vonk et al., 2013). Yet, since single instances of shifts into a new niche often correlate to other traits besides diet, the occupation of similar niches emerging repeatedly from independent ancestors provides a uniquely powerful test of adaptation to a particular lifestyle (e.g. high-altitude adaptation) (Lim, Witt, Graham, & Davalos, 2019), or diet (Hu et al., 2017).

Living mammals have extraordinarily diverse diets, encompassing all available dietary niches on Earth (Wilson & Reeder, 2005). Among all mammalian orders, Chiroptera (bats) has the widest variety of diets, which include insects and other arthropods, mammals, birds, reptiles, amphibians, fish, blood, carrion, fruit, flowers, nectar, pollen and foliage (Altringham, 1996). Based on palaeontological and phylogenetic analyses, ancestral bats are inferred to have been insectivorous (Gunnell & Simmons, 2005; Jiao et al., 2019; Simmons, Seymour, Habersetzer, & Gunnell, 2008), whereas modern bats show independent origins of frugivory, nectarivory, carnivory and omnivory that are unparalleled in other mammalian clades (Neuweiler, 2000). Although approximately 70% of all bats are insectivorous, two lineages that first diverged around 64 million years ago (Teeling et al., 2005) have independently evolved obligate frugivory. One is the family Pteropodidae (also known as Old World fruit bats, within the suborder Yinpterochiroptera), and the other is the subfamily Stenodermatinae (referred to hereafter as neotropical fruit bats, within the family Phyllostomidae and in the other suborder Yangochiroptera; Voigt, Zubaid, Kunz, & Kingston, 2011). The former lineage occurs only in Africa and Eurasia, the Old World, whereas the latter occurs only in the Americas. Thus, the two lineages represent a rare but classic example of convergent evolution of frugivory.

Dietary changes generate strong selective demands on sensory perception, digestion and metabolism, because sensory perception is partially dedicated to selecting and ingesting food, and digestion and metabolism are required for the conversion of food to energy and the transport of nutrients to cells. Identifying genetic signatures associated with dietary changes can provide valuable insights into understanding animal evolution and diet-related diseases in humans. For example, the bitter taste receptor gene numbers are generally shaped by diet across vertebrates (Hong & Zhao, 2014; Li & Zhang, 2014; Wang & Zhao, 2015), suggesting that the sense of taste must play an important role in selecting their dietary components. Similarly, the evolution of digestive enzyme genes, such as the chitinase genes (*CHIA*s), the trehalase gene (*Treh*) and the amylase gene (*AMY*), is largely correlated with dietary changes in mammals (Emerling, Delsuc, & Nachman, 2018; Jiao et al., 2019; Pajic et al., 2019). The metabolic enzyme alanine–glyoxylate aminotransferase (AGT) appears to have undergone molecular adaptations to diet in birds and mammals, with a tendency for AGT to target mitochondria in carnivorous species, and peroxisomes in herbivorous species (Birdsey et al., 2005; Wang et al., 2019). In addition, convergent losses of *Paraoxonase 1* in all marine mammals may have linked to changes in lipid metabolism from land to water (Meyer et al., 2018).

In this study, we identified molecular signatures of convergent evolution underlying obligate frugivory in Old World and neotropical fruit bats by comparing published Old World bat genomes with two novel genome sequences of neotropical fruit bats. Our analyses examined all the olfactory receptor genes, bitter taste receptor genes, digestive enzyme genes and genes with convergent amino acid substitutions to reveal molecular adaptations contributing to convergent evolution of frugivory.

## 2 | MATERIALS AND METHODS

### 2.1 | Illumina paired-end and mate-pair sequencing

The Jamaican fruit-eating bat, *Artibeus jamaicensis*, was sampled in Morelia, Michoacán, Mexico; the yellow-shouldered bat, *Sturnira hondurensis*, was collected in the central coast, Veracruz, Mexico. The collection permits were issued by the General Directorate of Wildlife SEMARNAT Mexico (SGPA/DGVS/08711/15 and SGPA/DGVS/13168/14). Both species belong to the subfamily Stenodermatinae (Figure 1), whose species are predominantly or obligate frugivores (Altringham, 1996). Although *A. jamaicensis* feeds on some insects (around 10% in their diet), this species is predominantly frugivorous (Wilman et al., 2014). For convenience, we classified *A. jamaicensis* into the category: obligate frugivore. All experiments on the bats were approved by the Ethics Committee of Biotechnology and Applied Ecology Institute and Wuhan University and conformed to the rules and guidelines on animal experimentation in Mexico and China. Genomic DNA (1.5 μg per sample) was extracted from the liver of one individual from two neotropical fruit bat species (*A. jamaicensis* and *S. hondurensis*). Sequencing libraries were generated using a Truseq Nano DNA HT Sample Preparation Kit (Illumina) following the manufacturer's instructions. Index codes were added to identify sequences from each sample. Briefly, the DNA sample was fragmented by sonication to short-insert (180 and 500 bp) for paired-end and long-insert (2 and 5 kb) for mate-pair sequencing. Next, the DNA fragments were end-polished, A-tailed and ligated with the full-length adapters for Illumina sequencing with further PCR amplification. Finally, PCR products were purified (AMPure XP system, Beckman Coulter), and libraries were analysed for size distribution using an Agilent 2100 Bioanalyzer and quantified using real-time PCR. We generated 247.95 and 206.73 Gb raw data using the Illumina Hiseq platform for the two species, respectively (Table S1).

### 2.2 | 10X Genomics library construction, sequencing and extending scaffold

DNA sample preparation, indexing and barcoding were done using the GemCode Instrument from 10X Genomics. About 1 ng input DNA (50 kb in length) was used for the GEM reaction procedure during PCR, and 16-bp barcodes were introduced into droplets. The droplets were then fractured following the purification of the intermediate DNA library. Next, DNA was sheared into 500-bp fragments for constructing libraries, which were subsequently sequenced on the NovaSeq instrument. After sequencing, we used the supernova-2.0.0 software to tackle the FASTQ files containing the barcoded reads and build a graph-based assembly to produce a FASTA file suiTable for downstream processing and analysis. In total, we generated 230.33 and 172.60 Gb raw data for the two bat genomes, respectively (Table S1).

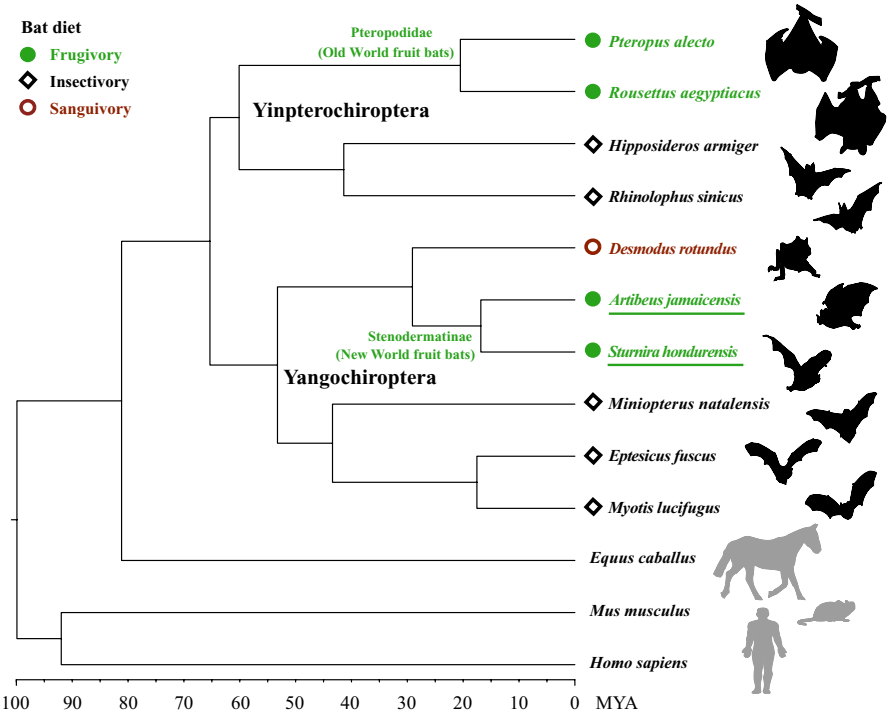



**FIGURE 1** Phylogenetic tree of the 10 bat species in this study. The tree was reconstructed using maximum-likelihood method under the GTR + GAMMA model based on the concatenated alignments of all the 6,332 single-copy orthologs. The bootstrap values of all nodes are 100%. Frugivorous bats are indicated by green circles, insectivorous bats were denoted by black diamonds, and the red circle represents the sanguivorous bat. The two species with newly sequenced genomes are underlined. Branch length represents millions of years and is drawn to the scale. New World fruit bats are also known as neotropical fruit bats. [Colour figure can be viewed at wileyonlinelibrary.com]

## 2.3 | Sequence quality checking and filtering

We used strict filters to avoid reads with artificial biases. First, we removed low-quality paired reads that mainly resulted from base-calling duplicates and adaptor contamination (i.e. with ≥10% unidentified nucleotides "N"; >10 nucleotides aligned to the adaptors, allowing ≤10% mismatches; >50% nucleotides having phred quality <5; and putative PCR duplicates generated in the library construction process). Ultimately, we generated 802.78 Gb of high-quality data (451.89 Gb for *A. jamaicensis* and 350.89 Gb for *S. hondurensis*; Table S1). There were 92.18% (*A. jamaicensis*) and 89.65% (*S. hondurensis*) nucleotides with phred quality ≥Q30 (with an accuracy of 99.9%) in the two genomes (Table S1).

## 2.4 | Estimation of genome size using the K-mer method

The genome size was estimated based on the k-mer spectrum: $G = (K_{total} - K_{error})/D$, where $G$ is the genome size, $K_{total}$ is the total count of k-mers, $K_{error}$ is the total count of low-frequency (≤3) k-mers that were likely caused by sequencing errors, and $D$ is the k-mer depth. We used Jellyfish (v2.1.3) (Marcais & Kingsford, 2011) to generate 17-mer information based on the high-quality short-insert reads (180 and 500 bp). Therefore, the estimated sizes of two bat genomes were 2.19 Gb for *A. jamaicensis* and 2.37 Gb for *S. hondurensis*.

## 2.5 | De novo assembly

The high-quality paired-end reads of 180-bp, 500-bp, 2-kb and 5-kb DNA libraries were first assembled using the program ALLPATHS-LG (Butler et al., 2008) and then gap-filled with the package Gapcloser (version 1.12; Luo et al., 2012), and finally, the scaffolds were improved using the fragScaff program (Adey et al., 2014). First, the primary genome assembly was generated with the program ALLPATHS-LG. We built unipaths using a minimum overlap of $K = 96$. The short-insert reads (180 and 500 bp) were first error-corrected and then merged into single "super-reads." These were then formed into an initial unipath graph. This graph is expected to be highly accurate, although it has gaps arising from biases in the Illumina data. Some of these gaps in the unipath graph were filled using the jumping reads (2 and 5 kb), ignoring their pairing. We used the unipath graph to first determine the fragment size distribution of the jumping pairs reads and then estimate the distances between unipaths. Second, we used the jumping links to connect the unipaths into a primary graph. Illumina's short-insert paired-end reads and 10X Genomics linked reads were used to perform gap-filling with the package Gapcloser (version 1.12). Finally, the resulting scaffolds from the previous step were further connected to super-scaffolds by 10X Genomics linked-read data using the fragScaff software (Adey et al., 2014).

## 2.6 | Assessment of genome assembly quality

To validate the single-base accuracy of the genome assemblies, we realigned the high-quality reads of the short-insert (500 bp) DNA libraries to the assemblies with BWA (Li & Durbin, 2009) and found that more than 97.41% and 96.54% of the two bat genomes had a coverage depth ≥10, respectively. We then performed variant calling with SAMtools (Xavier et al., 2011) and obtained heterozygous single nucleotide polymorphisms.

## 2.7 | Repeat annotation

We identified the repetitive sequences for the assembled genome using a combination of homology searching and ab initio prediction.

### 2.7.1 | Homology-based prediction

RepeatMasker (version open-4.0.5) and RepeatProteinMask (http://www.RepeatMasker.org) were used to search against the Repbase TE library (RM database vision 20140131; Bergman & Quesneville, 2007).

### 2.7.2 | Ab initio prediction

We built a de novo repeat library for each genome using package RepeatModeler (version 1.73), which uses two core programs BuildDatabase and RepeatModeler to annotate the TE families. We then found that 28.06% of the *A. jamaicensis* genome and 32.69% of the *S. hondurensis* genome was composed of repetitive elements.

## 2.8 | Protein-coding gene prediction

We performed gene prediction through a combination of homology-based, transcriptome-based and ab initio prediction methods for the two genome assemblies.

### 2.8.1 | Homology-based prediction

The protein repertoires of vertebrates including *Desmodus rotundus* (GCF_002940915.1), *Eptesicus fuscus* (GCF_000308155.1), *Equus caballus* (GCF_000002305.2), *Mus musculus* (GCF_000001635.26) and *Homo sapiens* (GCF_000001405.38) were used as queries to search against the newly assembled genomes using the genBlastG program (v1.0.138), with default parameters (She, Chu, Wang, Pei, & Chen, 2009). Genomic regions with hits (and their flanking 1-kb regions in both directions) were conjoined using an in-house Perl script. Then, each retrieved genomic sequence was aligned against the matching proteins using Genewise (Birney, Clamp, & Durbin, 2004) to define gene models. Homology predictions were denoted as the Homology-set.

### 2.8.2 | Transcriptome-based prediction

From the National Center for Biotechnology Information (NCBI) database, we downloaded about 14.2 Gb of RNA-seq data from *A. jamaicensis* (accession no: SRR539297), which was then assembled with Trinity (Grabherr et al., 2011). These assembled sequences were aligned against the newly assembled genome using the Program to Assemble Spliced Alignment (PASA; Haas et al., 2003). Valid transcript alignments were clustered based on genome mapping locations and assembled into gene structures. Gene models created by PASA were denoted as the PASA-T-set (PASA Trinity set). In addition, RNA-seq reads were directly mapped to the assembled genome using TopHat (Kim et al., 2013) to identify putative exon regions and splice junctions. Cufflinks (Trapnell et al., 2012) were used to assemble the mapped reads into gene models (Cufflinks-set).

### 2.8.3 | Ab initio prediction

Augustus (Stanke & Waack, 2003), GlimmerHMM (Majoros, Pertea, & Salzberg, 2004) and SNAP (Korf, 2004) were also used to predict coding regions in the repeat-masked genomes. They were all trained by the PASA-T-set gene model. The resulting gene models were integrated using EvidenceModeler (Haas et al., 2008). Weights for each type of evidence were set as follows: PASA-T-set > Homology-set > Cufflinks-set > Augustus>SNAP = GlimmerHMM. The gene models were further updated by PASA2 to generate untranslated regions and alternative splicing variation information.

Three prediction methods (homology-based, transcriptome-based and ab initio) were used to annotate the *A. jamaicensis* genome, which identified 22,079 protein-coding genes, with a mean of 8.15 exons per gene. For the *S. hondurensis* genome, we only used two methods (homology-based and ab initio prediction) due to the absence of RNA-seq data and genetic material for RNA sequencing and identified 20,813 protein-coding genes, with a mean of 7.88 exons per gene. Approximately 90.46% (*A. jamaicensis*) and 87.06% (*S. hondurensis*) of these genes were functionally annotated to gene function databases.

## 2.9 | Phylogenetic analysis and species divergence time estimation

There were a total of 15 published Chiroptera genomes in NCBI when we started this project (7 August 2018). After excluding low-quality genomes (N50 of contigs < 20kb), we selected one representative species (with the longest contig N50) for each genus. In addition to the two neotropical fruit bats sequenced in this study, eight additional bat species and three outgroup species (horse: GCF_000002305.2, mouse: GCF_000001635.26 and human: GCF_000001405.38) with published genomes were selected to identify orthologs. We retained the longest transcripts for each gene and filtered out genes with less than 50 amino acids. For each transcript, the open reading frame (ORF) was checked by homologous protein sequences from a nonredundant (nr) protein database. All filtered protein sequences of the 13 mammals were clustered into orthologous groups using OrthoMCL (v2.0.9) with the default parameters (Li, Stoeckert, & Roos, 2003). A total of 6,332 single-copy orthologs were retrieved for phylogenetic tree construction. First, multiple sequence alignments for each single-copy orthologous gene were generated using MUSCLE, with the default

parameters (Edgar, 2004). Next, the alignments of each gene were concatenated to a super alignment matrix. We then constructed a phylogenetic tree under the maximum-likelihood framework with the package RaxML with 1,000 bootstrap replicates (Stamatakis, 2014), using the best-fitting substitution model "GTR + GAMMA" which was determined by jModelTest2 according to the Bayesian information criterion (Darriba, Taboada, Doallo, & Posada, 2012). Divergence times between species were estimated based on the concatenated alignments of fourfold degenerate sites from all the 6,332 single-copy orthologs using MCMCtree in PAML (Yang, 1997) with the "independent rates" option and the best-fitting substitution model "GTR + G" which was determined by jModelTest2. A Markov chain Monte Carlo analysis was run for 20,000 generations, using a burn-in of 10,000 iterations. There were four fossil constraints incorporated in the analyses: (a) divergence time for human and mouse, which is about 61.5–100.5 Mya; (b) divergence time for human and horse, which is about 95.3–113 Mya; (c) divergence time between the common vampire bat and the two neotropical fruit bats, the maximum of which is 34 Mya; and (d) divergence time for *Rhinolophus sinicus* and *Hipposideros armiger*, which is about 37–55 Mya (Benton & Donoghue, 2007; Teeling et al., 2005). Additionally, a mean prior of 65 MYA for the base of Chiroptera was also incorporated (Figure S1).

## 2.10 | Identification of *Tas2r* and *OR* genes

We identified *OR* and *Tas2r* genes from each of the 10 bats based on homology searches using published full-length *Tas2r* and *OR* gene sequences of other vertebrates as queries. This method was described in detail in an earlier study (Wang & Zhao, 2015). All confirmed *OR* and *Tas2r* genes were classified into three categories: intact genes (at least 270 codons, with an intact ORF and a complete coding region), partial genes (at least 100 codons, with an intact but truncated ORF due to incomplete genome sequencing) and pseudogenes (at least 300 nucleotides, with a disrupted ORF interrupted by nonsense or frameshifting mutations). For both *OR* and *Tas2r* genes, all intact genes and all partial genes longer than 700 nucleotides were considered to be putatively functional. Subsequently, all putatively functional *OR* genes were assigned to subfamilies using OR family Assigner, ORA (v1.9) (Hayden et al., 2010). In addition, aiming to validate our gene identification approach, we also identified all the *OR* and *Tas2r* genes in other 13 newly released genomes (N50 of contigs > 20kb), including frugivorous bats, partially frugivorous bats and nonfrugivorous bats, as of 14 March 2020 (Table S2).

## 2.11 | Phylogenetic analysis of *Tas2r* genes

Coding sequences of 174 putatively functional *Tas2r* genes from the 10 bat species, as well as a mouse *V1R* gene sequence, were aligned and manually checked using MUSCLE (Edgar, 2004). Next, we employed jModelTest2 to determine the best-fitting substitution model according to the Bayesian information criterion (Darriba et al., 2012). Finally, we used the resulting alignment to reconstruct a Bayesian inference (BI) phylogenetic tree using the default parameters in MrBayes (version 3.1.2), with 5 million generations (Yang & Rannala, 1997). A BI phylogenetic tree of 472 putatively functional *Tas2r* genes from the 10 bat genomes and 13 additional bat genomes was also conducted using the same method.

## 2.12 | Phylogenetic principal component analysis

Phylogenetic principal component analysis (PCA) was conducted using *phyl.pca* from the *phytools* R package v0.6–60 (Revell, 2012). The resulting tree (Figure S1) was used with normalized *OR* gene frequencies of subfamilies as the input for the phylogenetic PCA. Of note, normalized *OR* gene frequencies of subfamilies are the ratios of the putatively functional *OR* gene numbers in each subfamily to the putatively functional *OR* gene numbers in one species multiplied by 100 (Hayden et al., 2014). For the 13 additional bat species, phylogenetic relationships and branch lengths were taken from a previous study (Teeling et al., 2005) and the TimeTree database (http://www.timetree.org/). The PCA algorithm was based on the covariance matrix of the data. PC1 and PC2 of functional *OR* gene frequencies in fruit bats and nonfrugivorous bats were compared across all examined bats while accounting for phylogeny. Phylogenetic generalized least-squares (PGLS) analyses based on the phylogeny described earlier were conducted using the PGLS component of the *caper* (v1.0.1) R package (Orme, Freckleton, Thomas, & Petzoldt, 2013). The *p*-value for each comparison of PCs was generated from *t* test, and the Benjamini–Hochberg method was used to correct for multiple testing.

## 2.13 | Identification of genes involved in digestion and metabolism

First, we downloaded five full-length protein sequences of *CHIA*s in *Tarsius syrichta* from a published study (Emerling et al., 2018). We also downloaded from NCBI one representative full-length gene sequence in *Homo sapiens* for each of the remaining digestive or metabolic genes in this study. Second, we conducted a homology search using the downloaded protein sequences as queries for each gene, using the default parameters in the genBlastG program (v1.0.138; She et al., 2009). For each genomic region that had blast hits, we only retained the genomic sequence of the longest hit and its flanking 1-kb regions in both directions. Third, the retrieved genomic sequences were annotated by genewise (wise2-4–1) using the respective query sequence for each gene as a reference (Birney et al., 2004). Finally, we used the resulting coding sequences (CDS) to search against the nonredundant (nr) protein database based on BLASTX (Altschul, Gish, Miller, Myers, & Lipman, 1990). Only CDSs with the best hits of genes of interest were retained for subsequent analysis. In addition, alignments of each retrieved gene were checked manually, and *CHIA* genes were identified based on phylogenetic and synteny analyses.

### 2.14 | Identification of genes with convergent amino acid substitutions

Here, convergent sites refer to both previously defined "parallel" and "convergent" sites (Zhang & Kumar, 1997). Based on the phylogenetic tree shown in Figure S2, amino acid sequences at each node for all the 6,332 single-copy orthologs were deduced under the "aaml" model using CODEML in PAML (Yang, 1997). In this study, convergent amino acid sites between the two divergent lineages of fruit bats are defined as parallel or convergent changes that have resulted in the same amino acids at the respective ancestral nodes of the two Old World fruit bats and the two neotropical fruit bats. Specifically, convergent amino acid sites must meet the following criteria: (a) $O_1 = N_1$; (b) $O_1 \neq O_2$; (c) $N_1 \neq N_2$, where $O_1$, $O_2$, $N_1$ and $N_2$ represent ancestral amino acids for the interior nodes in Figure S2. The impact on the biological function of a protein for each convergent amino acid site was predicted by PROVEAN v1.1.5 (Choi & Chan, 2015). Comparisons of observed convergent sites versus random expectation under the JTT-$f_{gene}$ amino acid substitution models for each gene were performed using scripts provided in an earlier study (Zou & Zhang, 2015). A Poisson test was used to determine whether the observed number of convergent sites of each gene was significantly larger than random expectation, and the Benjamini–Hochberg method was used to correct for multiple testing ($q < 0.05$).

## 3 | RESULTS

### 3.1 | Genome sequencing of two bats and phylogenetic tree reconstruction

To enable genomic comparisons between Old World and neotropical fruit bats, we sequenced the genomes of two neotropical fruit bats: the Jamaican fruit-eating bat (*A. jamaicensis*) and the yellow-shouldered bat (*S. hondurensis*), which are analysed as biological replicates in this study. A total of 201.74-fold coverage for *A. jamaicensis* and 158.77-fold coverage for *S. hondurensis* data were generated (Table S1). Finally, 2.22 Gb (*A. jamaicensis*) and 2.11 Gb (*S. hondurensis*) genome assemblies were generated. The contig and scaffold N50 values for *A. jamaicensis* were 41.70 kb and 2.56 Mb, in order, and 99.90 kb and 10.02 Mb for *S. hondurensis* (Table 1). Both genomes displayed a relatively high level of heterozygosity—0.573% for *A. jamaicensis* and 0.416% for *S. hondurensis* (Table S3).

To assess the completeness of the two genomes, we performed Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis by searching against the mammalian BUSCO gene groups (version 3.0; Simao, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015). Overall, more than 91.8% (*A. jamaicensis*) and 91.5% (*S. hondurensis*) complete sequences of the 4,104 mammalian BUSCOs were identified in the genomes (Table 1 and Table S4). We also used the core eukaryotic genes mapping approach (CEGMA) pipeline to evaluate the completeness (Parra, Bradnam, & Korf, 2007). This approach identified 98.79% (*A. jamaicensis*) and 95.56% (*S. hondurensis*) of the 248 complete core eukaryotic genes (CEGs) for the two genome assemblies (Table 1). Finally, a total of 22,079 (*A. jamaicensis*) and 20,813 (*S. hondurensis*) protein-coding genes were annotated for the genomes (**Table 1**).

**TABLE 1** Summary of the genome assemblies for *A. jamaicensis* and *S. hondurensis* newly generated in this study

| Genomic features | *Artibeus jamaicensis* | *Sturnira hondurensis* |
|---|---|---|
| Assembled genome size (Gb) | 2.22 | 2.11 |
| Contig N50 (Kb) | 41.7 | 99.9 |
| Scaffold N50 (Mb) | 2.56 | 10.02 |
| GC content (%) | 41.92 | 42.3 |
| Repeat rate (%) | 28.06 | 32.69 |
| Predicted protein-coding genes | 22,079 | 20,813 |
| Average gene length (bp) | 29,436.16 | 22,882.57 |
| Average CDS length (bp) | 1,445.98 | 1,572.26 |
| Average exon number per gene | 8.15 | 7.88 |
| Percentage of 4,104 mammalian BUSCOs (%) | 91.8 | 91.5 |
| Percentage of 248 ultra-conserved CEGs (%) | 98.79 | 95.56 |

For comparative genomics, we obtained a data set consisting of 13 species that represent diverse lineages of bats with a variety of diets and three outgroup species (horse, mouse and human), including two Old World fruit bats (*Pteropus alecto*, *Rousettus aegyptiacus*), two neotropical fruit bats (*A. jamaicensis*, *S. hondurensis*), one horseshoe bat (*Rhinolophus sinicus*), one roundleaf bat (*Hipposideros armiger*), one vampire bat (*D. rotundus*), one long-fingered bat (*Miniopterus natalensis*) and two vesper bats (*E. fuscus*, *Myotis lucifugus*; Figure 1 and Table S2). While the two neotropical fruit bats were newly sequenced in this study, the genome sequences for other bat species were previously published (Teeling et al., 2018). We also assessed the eight published bat genomes used in this study by mammalian BUSCO gene groups, and the result showed that more than 91% complete sequences were identified in all bat genomes (Table S4). We used these data to reconstruct a phylogenetic tree with the maximum-likelihood method based on the concatenated alignment of 6,332 single-copy orthologs from the 10 bat and three outgroup species (human, mouse and horse) (Figure 1). The bootstrap values of all nodes are 100%, and the topology agrees with that from earlier studies (Figure 1; Miller-Butterworth et al., 2007; Teeling et al., 2005). To minimize systematic errors such as saturation in substitutions and bias in base frequencies (Davalos & Perkins, 2008), we repeated the phylogenetic analysis using the same method based on the concatenated alignment of single-copy orthologs with third-codon positions removed, which resulted in an identical phylogenetic tree with all nodes receiving 100% bootstrap support.

### 3.2 | Evolution of sensory genes linked to frugivory

To investigate how dietary changes shaped the evolution of sensory genes in the two lineages of frugivorous bats, we characterized

olfactory receptor and bitter taste receptor gene repertoires from the 10 bat genomes (Figure 1). Using published intact *OR* gene sequences from human, rat, macaque, cow, dog, platypus and opossum as queries, we identified all *OR* genes in the whole-genome assemblies of the 10 bat species. A total of 7,209 *OR* genes were retrieved from the 10 bat genomes, with a range of 261–635 putatively functional (all intact genes and all partial genes with the length longer than 700 bp) *OR* genes for each species (Table S5). All identified *OR* genes are provided in Dataset S1 (Dryad https://doi.org/10.5061/dryad.vt4b8gtpn). Next, all putatively functional *OR* genes were assigned to 13 subfamilies (Table S6), as suggested by phylogenetic analyses in a previous study (Hayden et al., 2010).

To compare *OR* gene subfamily distribution patterns between frugivorous and nonfrugivorous bats, we conducted phylogenetic PCA based on the species tree of the 10 bats (Figure 1 and Figure S1). To minimize the impact of different genome quality, we did not use the absolute *OR* gene numbers for each subfamily, but instead used the *OR* gene frequencies (i.e. the ratios of the putatively functional *OR* gene numbers in each subfamily to the putatively functional *OR* gene numbers in one species) as input in phylogenetic PCA (Table S7). The first two principal components (PCs) explained 54.04% and 28.82% of the variance in *OR* subfamily gene number frequencies, respectively (Figure 2a). This analysis showed that the gene frequencies of *OR* subfamilies in frugivorous bats were significantly different from those in their nonfrugivorous relatives (FDR-adjusted *P*-value of PC1 = 0.33, *t*-value=−1.25, *df* = 8; FDR-adjusted *P*-value of PC2 = 0.03, *t*-value=−0.76, *df* = 10; Table S8).

To test the influence of different genome assembly methods and parameters, we identified all *OR* genes from other versions of genome assemblies for the same two fruit bat species (RouAeg_v1_BIUU for *R. aegyptiacus* and ArtJam_v1_BIUU for *A. jamaicensis*) available in the National Center for Biotechnology Information (NCBI) database (Table S2). After adding the data from the two additional genome assemblies (RoaegV2 and ArjamV2), we repeated the phylogenetic PCA based on the *OR* gene frequencies (Figure 2b). The repeated analysis showed the two additional assemblies of fruit bat species were closely clustered with other fruit bat species (Figure 2b), and the PCs of *OR* subfamily gene frequencies between frugivorous and nonfrugivorous bats remain significantly different (FDR-adjusted *P*-value of PC1 < 0.001, *t*-value = −6.08, *df* = 10; FDR-adjusted *p*-value of PC2 = 0.53, *t*-value=−0.76, *df* = 10; Table S8), suggesting that the influence of different genome assembly methods and parameters is minimal in the analysis. Both PCAs showed that the first two PCs explained the majority of the variance (>80%) in *OR* gene frequency for each subfamily, indicating that the subfamilies *OR4*, *OR5/8/9* and *OR2/13* may be linked to frugivory in bats (Figure 2a,b). To test the validity of our method, we repeated the PCA using their data and obtained a generally identical result compared to a previous study which used a set of degenerate primers to amplify OR sequences using PCR (Hayden et al., 2014; Figure S3a). Furthermore, we conducted an additional PCA using the combined data of functional *OR* gene frequencies of each subfamily from this and that study (Hayden et al., 2014); Figure S3b) and identified a significant difference among PCs of *OR* subfamily gene frequencies between the previous (based on PCR) and our data (based on whole-genome sequences) (FDR-adjusted *p*-value of PC1 = 0.002, *t*-value = 3.76, *df* = 35; FDR-adjusted *p*-value of PC2 = 0.05, *t*-value=−2.20, *df* = 35; Table S8), indicating that different methods of *OR* sequence sampling may influence the result.

To add more data points for further validation of our analysis, we additionally identified all the *OR* genes in 13 recently published genomes of frugivorous and nonfrugivorous bats; phylogenetic PCAs based on functional *OR* gene frequencies of all the 23 bat genomes were repeated. Results also showed that the subfamilies *OR4*, *OR5/8/9* and *OR2/13* are likely linked to frugivory in bats (FDR-adjusted *p*-value of PC1 = 0.04, *t*-value = −2.18, *df* = 21; FDR-adjusted *p*-value of PC2 = 0.67, *t*-value = −0.43, *df* = 21; Figure S4 and Table S8). Unexpectedly, the clustering of two nonfrugivorous bat species (*Craseonycteris thonglongyai* and *Megaderma lyra*) also showed increases in subfamilies *OR4*, *OR5/8/9* and *OR2/13* (Figure S4).

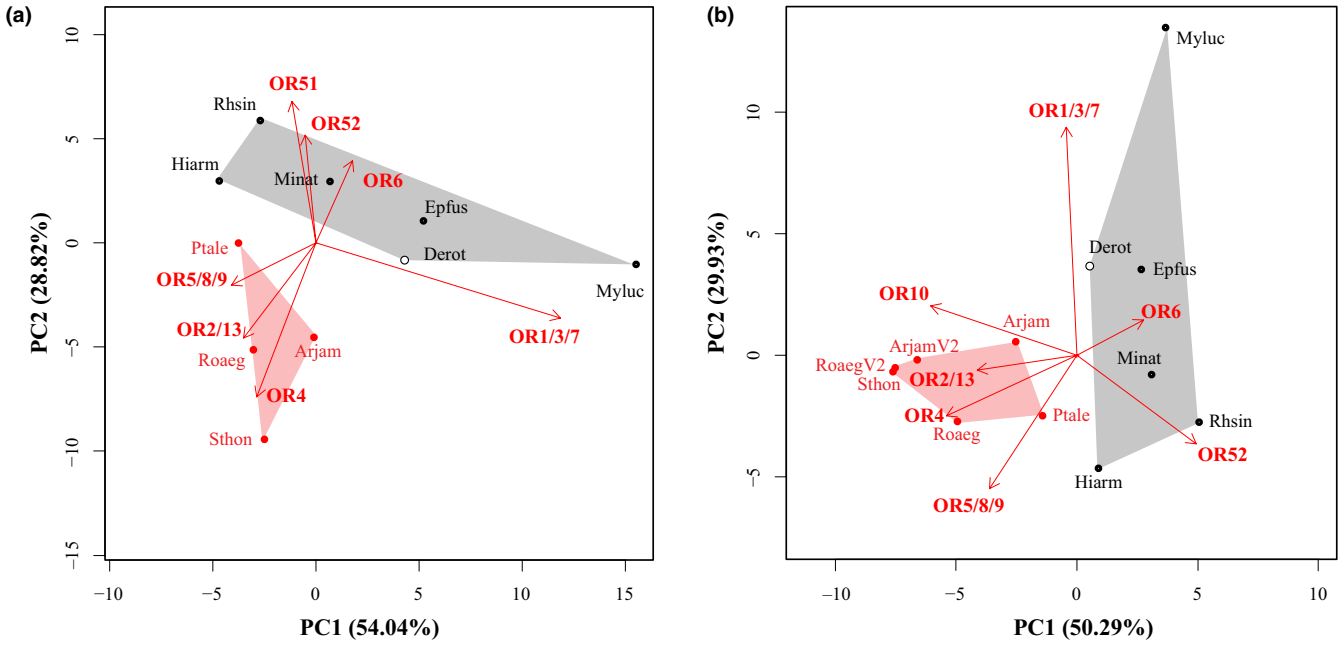
Owing to the conservation of *Tas1r1* genes in bats (Zhao et al., 2010), we only identified all *Tas2r* genes in the 10 bat genomes based on a homology search using published full-length Tas2r proteins from human, mouse, chicken, zebrafish, lizard and frog as queries. In total, we retrieved 8–29 intact genes, 0–4 partial genes and 6–15 pseudogenes from the 10 bat genomes (Table S9 and Dataset S2, Dryad https://doi.org/10.5061/dryad.vt4b8gtpn). Notably, the common vampire bat (*D. rotundus*) was found to possess only eight putatively functional (intact and partial) *Tas2r* genes, which is much fewer than in other bats (Table S9).

Compared to the vampire bat, which has only eight putatively functional *Tas2r* genes, we found that the repertoires of putatively functional *Tas2r* genes were larger in both the insectivorous and frugivorous bats examined, with 17–30 and 14–18 genes detected, respectively (Table S9). Furthermore, we reconstructed a gene tree with the Bayesian approach based on an alignment of all the putatively functional gene sequences from the 10 bats, using the mouse *V1r1* gene (GenBank: NM_001166728) as the outgroup. The tree showed that *Tas2r* genes were generally clustered into 18 clades, and the nomenclature was followed as proposed in previous studies (Hayakawa, Suzuki-Hashido, Matsui, & Go, 2014; Jiao, Wang, Zhang, Jiang, & Zhao, 2018; Figure 2c and Figure S5). Almost all of these clades contained *Tas2r* genes from both insectivorous and frugivorous bats, but four clades did not. With one exception, three clades (*Tas2r11*, *Tas2r18* and *Tas2r67*) contained genes from at least four of the five insectivorous bats, but none from frugivorous bats (Figure 2c). The only exception is *Tas2r40*, which includes seven genes from frugivorous bats and one gene from the sanguivorous bat (Figure 2c). In addition, we also reconstructed a Bayesian gene tree using the same method based on all the 472 putatively functional *Tas2r* genes from the 10 and additional 13 recently published bat genomes (Figure S6). *Tas2r* genes of obligate frugivorous bats were also absent in the three clades (*Tas2r11*, *Tas2r18* and *Tas2r67*) (Figure S6).

### 3.3 | Loss of digestive enzymes in two lineages of fruit bats

To ascertain whether any digestive enzyme genes are specifically lost in both Old World and neotropical fruit bats, we examined the whole-genome assemblies of the 10 bats and two partially

frugivorous phyllostomids (*Tonatia saurophila* with about 10% of fruit/nectar in its diet and *Phyllostomus discolour* with a 70% fraction but recently diverged from its animalivorous ancestor) to identify digestive enzymes (Figure 3). There were 16 digestive enzyme genes that have been well characterized in previous studies, including carbohydrases, proteases, lipases or esterases, chitinases,



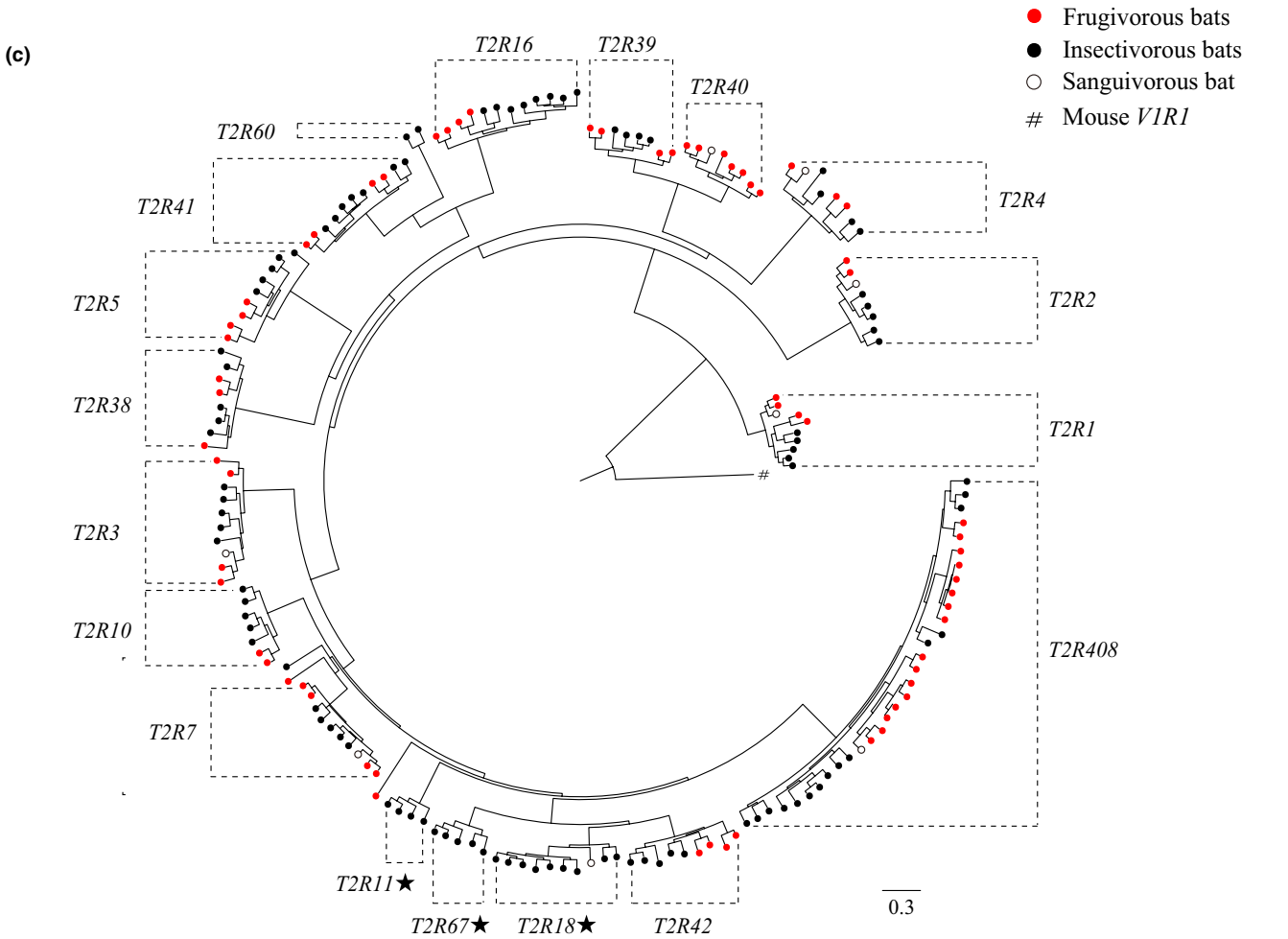



**FIGURE 2** PCA of *OR* gene frequencies of (a) 10 bat genomes and (b) 10 bat genomes and additional genome versions of two frugivorous bats, and (c) the *Tas2r* gene tree. (a) PCA showing PC1 and PC2 for the 10 bat species. PCs were generated from the ratios of the putatively functional *OR* gene numbers in each subfamily to the putatively functional *OR* gene numbers in one species. Bats of different diet types are marked with various colours: frugivorous (red circle), insectivorous (black circle) and sanguivorous (hollow circle). (b) PCA showing PC1 and PC2 for the 10 bat species and two additional frugivorous species (RoaegV2 and ArjamV2). (C) Evolutionary relationships of all 174 functional *Tas2r* genes from the 10 bats. The tree was reconstructed using the Bayesian approach under the GTR + I+G model. Branch lengths are drawn to the scale. Note that *Tas2r11*, *Tas2r18* and *Tas2r67* are specifically lost in obligate frugivorous bats. Abbreviations: Arjam, *Artibeus jamaicensis*; Sthon, *Sturnira hondurensis*; Derot, *Desmodus rotundus*; Ptale, *Pteropus alecto*; Roaeg, *Rousettus aegyptiacus*; Hiarm, *Hipposideros armiger*; Rhsin, *Rhinolophus sinicus*; Minat, *Miniopterus natalensis*; Epfus, *Eptesicus fuscus*; Myluc, *Myotis lucifugus* [Colour figure can be viewed at wileyonlinelibrary.com]

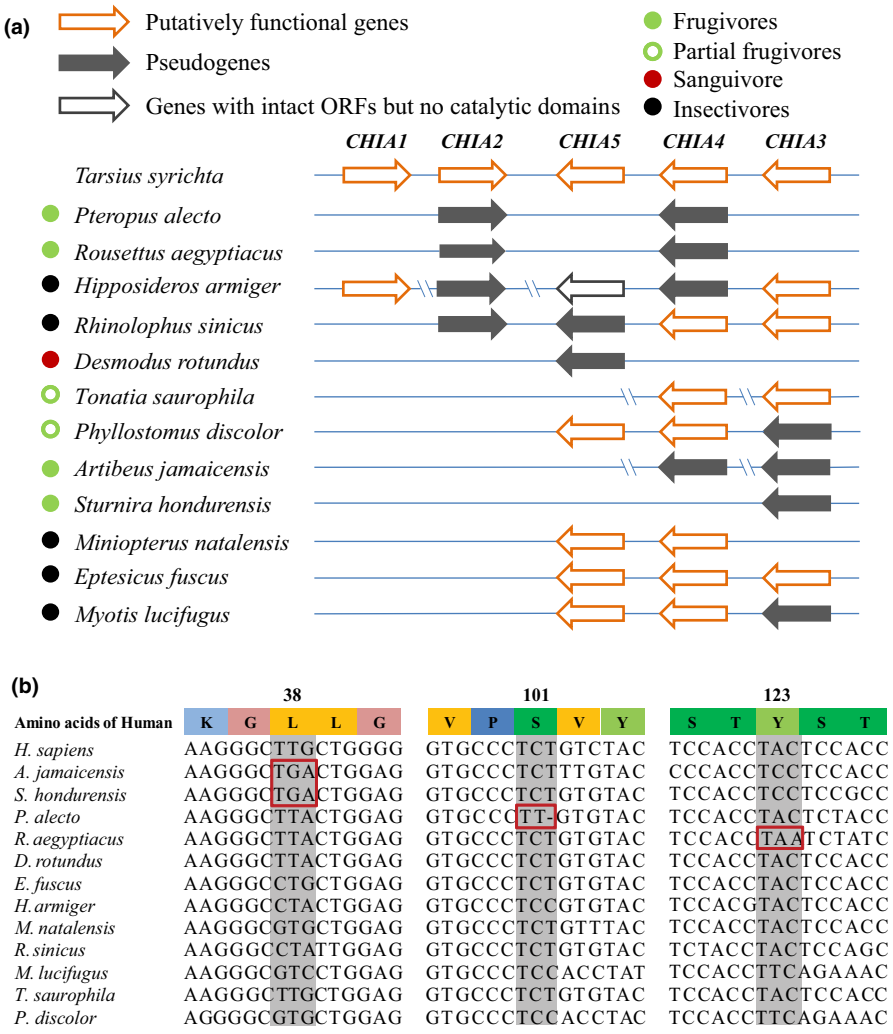



**FIGURE 3** Convergent gene loss of digestive enzymes in Old World and neotropical fruit bats. (a) No putatively functional *CHIA* genes were found in New and Old World fruit bats or in the common vampire bat (*Desmodus rotundus*). Species with different diet types are marked with different colours. Yellow arrows represent putatively functional *CHIA* genes, whereas grey arrows indicate nonfunctional genes. Double slashes indicate that genes were retrieved from different scaffolds. (b) Representative nonsense or frame-shifting mutations in the *PGC* genes from frugivorous bats [Colour figure can be viewed at wileyonlinelibrary.com]

and lysozymes (Chen & Zhao, 2019; Karasov, del Rio, & Caviedes-Vidal, 2011). Intriguingly, we found that acidic mammalian chitinase (*CHIA*; Enzyme Commission number: 3.2.1.14) and progastricsin (*PGC*; Enzyme Commission number: 3.4.23.3) were lost in both lineages of Old World and neotropical fruit bats, but not in insectivorous bats or partially frugivorous bats (Figure 3).

*CHIA* is expressed in the mammalian gastrointestinal tract and participates in degrading chitin from insects (Strobel, Roswag, Becker, Trenczek, & Encarnacao, 2013). There are five paralogs of *CHIA* in mammals, which were inherited from the last common ancestor of placental mammals (Emerling et al., 2018). It has been hypothesized that adaptation to noninsectivorous diets would lead to their loss (Jeuniaux, 1971). Indeed, all *CHIA* genes were lost in Old World fruit bats, but not in insectivorous bats (Emerling et al., 2018). Here, we take advantage of the two newly sequenced genomes to investigate whether there was a convergent loss of *CHIA* genes in neotropical fruit bats. Our results showed that there were one to five copies of *CHIA*s in each of the 12 bat species (Dataset S3). Only genes that met the following two criteria were considered to be putatively functional: (a) an intact ORF without premature stop codons or frameshift mutations and (b) the presence of a canonical catalytic domain (DXXDXDXE; Olland et al., 2009). For instance, the *CHIA5* of *H. armiger* is considered to be nonfunctional due to an absence of the canonical catalytic domain, although its ORF is intact. The results

showed that all insectivores and partially frugivorous bats possess 2–3 putatively functional *CHIA* genes, and the remaining *CHIA*s are pseudogenes or absent from the genomes. In contrast, all *CHIA*s retrieved from obligate frugivorous bats and the common vampire bat are pseudogenes (Figure 3a).

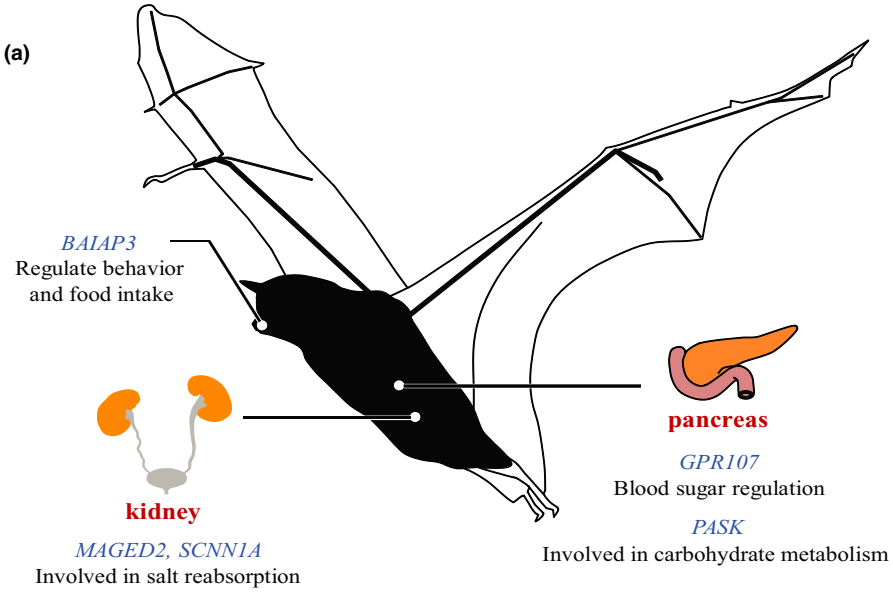
Gastricsin (or pepsinogen C), encoded by the progastricsin (*PGC*) gene, is an enzyme present in the gastric juice of most vertebrates and is responsible for degrading dietary proteins in the stomach (Taggart et al., 1989). In the 12 bat genome sequences, we found 1–2 copies of *PGC*s (Dataset S3). At least one putatively functional *PGC* was retrieved from all nonfrugivorous bats except *M. lucifugus*, whereas all *PGC*s of obligate frugivorous bats were pseudogenized. Notably, there is a premature stop codon in the *PGC* shared by the two neotropical fruit bat species (Figure 3b), indicating that the common ancestor of these two species had already lost *PGC*s. Since it was unexpected to find a single 1-bp deletion in the ORF of the *M. lucifugus PGC* gene, we examined the homologous genomic region of two additional *Myotis* species with relatively high-quality genomes (*M. brandtii* and *M. david*). However, we did not find any deletions in either species (Dataset S3). Thus, the 1-bp frameshift deletion of the *M. lucifugus PGC* is likely to be an artefact of sequencing, or an isolated case in *M. lucifugus*, uncommon in nonfrugivorous bats. All *PGC*s were lost in both lineages of obligate frugivorous bats, probably due to selection relaxing as their ancestral diet shifted to less proteins.

All the pseudogenes of *CHIA* and *PGC* from obligate frugivorous bats contain two or more nonsense or frame-shifting mutations except *CHIA3* (only one frame-shifting mutation) and *PGC1* (only one nonsense mutation) in *S. hondurensis* (Dataset S3). In order to reduce the impact of assembly or sequencing errors, the raw DNA sequencing reads were mapped to the genome sequence of *S. hondurensis* by SAMtools (Li, 2011). According to quality scores generated from SAMtools, we found that all the high-quality reads supported the frame-shifting mutation of *CHIA3* and the nonsense mutation of *PGC1* in *S. hondurensis*. This analysis indicates that pseudogenization events of the *CHIA* and *PGC* genes in frugivorous bats are not likely the result of assembly or sequencing errors.

### 3.4 | Convergence of metabolic genes between two lineages of fruit bats

To identify convergent amino acid substitutions between the common ancestor of the two Old World fruit bats and the common ancestor of the two neotropical fruit bats, we analysed the 6,332 single-copy orthologous protein-coding genes of the 10 bat species using a published method for detecting convergent and parallel evolution (Zhang & Kumar, 1997). We identified 878 genes that contain at least one convergent amino acid substitution and then discarded those without convergent substitutions that are potentially "deleterious/functionally important" (PROVEAN score <= −1.3). Thus, a total of 305 genes were retained. To reduce the impact of neutral evolution, we compared the observed number of convergent substitutions with the random expectation for each gene under the JTT-$f_{gene}$ amino acid substitution models (Zou & Zhang, 2015). Only genes with a significantly greater number of observed convergent substitutions were considered to be truly under convergent evolution (Benjamini–Hochberg FDR-adjusted $p < .05$; Thomas & Hahn, 2015). A total of 36 candidate genes were retained (Table S10). In addition, we used the PCOC method to predict posterior probabilities of convergent evolution for each convergent site in the 36 candidate genes under the "PCOC" model (Rey, Gueguen, Semon, & Boussau, 2018), which defines convergent shifts according to amino acid preferences but not convergent substitutions to the exact same amino acids. For the 107 convergent substitutions in these 36 genes, 99% substitutions (106/107) had a posterior probability above 0.95, and 91% substitutions (97/107) obtained a posterior probability greater than 0.99 (Table S11). To remove false positives, candidate genes containing one or more convergent substitutions with a posterior probability lower than 0.99 were discarded. Finally, we obtained 29 genes that are putatively under convergent evolution between the two lineages of frugivorous bats.

Among these convergent genes, we found several metabolic genes that may be specifically linked to frugivory, including *GPR107*, *PASK*, *BAIAP3*, *MAGED2* and *SCNN1A* (Figure 4a). The posterior probabilities under the PCOC model for all the 14 convergent sites from the five metabolic genes were above 0.99 (Table S11). Notably, all the 14 convergent substitutions are specific to the four frugivorous species, and seven of the 14 sites are conserved in nonfrugivorous bats (Table S12). GPR107 is a G protein-coupled receptor and was proposed to be a receptor for neuronostatin (NST), a peptide produced from the somatostatin preprohormone (Yosten, Redlinger, & Samson, 2012). Colocalization of GPR107 and NST on pancreatic α-cells has been demonstrated in mouse and human, and interaction of GPR107 and NST may be involved in blood sugar regulation in response to low glucose by stimulating glucagon release (Elrick et al., 2016), which would help maintain glucose homeostasis when frugivorous bats are fasting (Protzek & Pinheiro, 2011b). PASK (PAS kinase) is a member of the nutrient-regulated family of protein kinases, which may be involved in carbohydrate metabolism by regulating glucose-stimulated insulin production in mammalian pancreas and glucagon secretion (An et al., 2006; Wilson et al., 2005; Xavier et al., 2011). Brain-specific angiogenesis inhibitor I-associated protein 3 (BAIAP3), a C2 domain-containing Munc13 protein, functions in the process of endosome-to-Golgi retrograde transport, and the cellular content of insulin and prohormone convertase 2 was reduced dramatically in pancreatic β cell lines with BAIAP3 knockdown (Zhang et al., 2017). Moreover, BAIAP3 was suggested to play a role in regulating behaviour and food intake in human and mouse by controlling calcium-stimulated exocytosis of hormones such as insulin and other neurotransmitters (Lauridsen et al., 2011; Mariman, Bouwman, Aller, van Baak, & Wang, 2015; Wojcik et al., 2013). In addition, *MAGED2* encodes the protein melanoma-associated antigen D2, which is involved in salt reabsorption by affecting the expression and function of two cotransporters of sodium chloride (NKCC2 and NCC), two key components in the distal renal tubule for salt



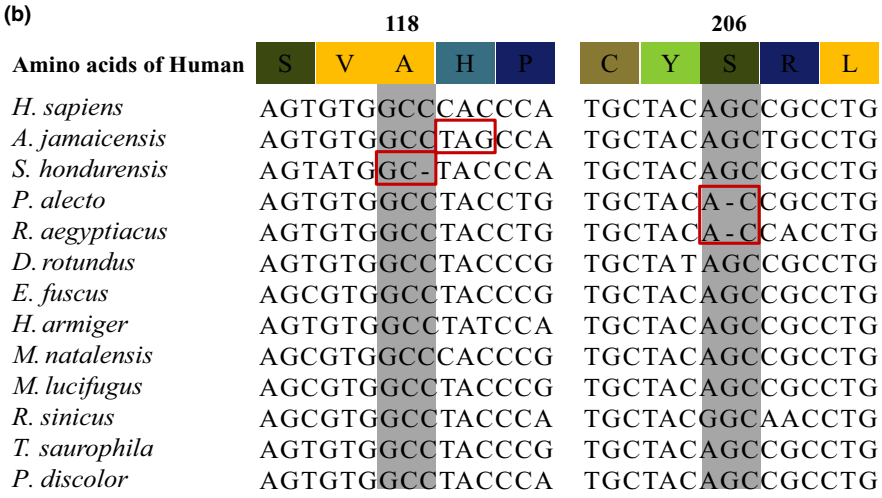




**FIGURE 4** Examples of metabolic genes showing signatures of convergent evolution. (a) Five genes with convergent substitutions were related to metabolism. (2) Examples of nonsense or frame-shifting mutations in the *FFAR3* genes from the two lineages of frugivorous bats [Colour figure can be viewed at wileyonlinelibrary.com]

reabsorption (Laghmani et al., 2016). *SCNN1A* encodes the α-subunit of epithelial sodium channel (ENaC), which is essential for electrolyte and blood pressure homeostasis by regulating the reabsorption of sodium in kidney, lung, colon and eccrine sweat glands (Hobbs et al., 2013; Salih et al., 2017; Voilley et al., 1994).

It should be noted that a previous study has hypothesized that losses of genes involved in metabolic pathways may have contributed to adaptations to a frugivorous diet in Old World fruit bats (Sharma et al., 2018). For example, genes involved in insulin metabolism and signalling (*FFAR3* and *FAM3B*), renal transporter (*SLC22A12*, *SLC2A9* and *SLC22A6*) and renal ammonium secreting transporter (*RHBG*) may be relevant in the dietary transition. Here, we tested those hypotheses by examining the same genes, with the aim of identifying convergent gene losses in neotropical fruit bats. Our results showed that the *FFAR3* gene was pseudogenized in both lineages of obligate frugivorous bats (Figure 4b and Dataset S3), but remained putatively functional in nonfrugivorous bats and partially frugivorous bats examined in this study. As *FFAR3* was suggested to inhibit glucose-stimulated insulin secretion, the loss of this gene may help fruit bats to secrete more insulin for glucose digestion (Sharma et al., 2018). However, *FAM3B* was pseudogenized only in *P. alecto*, and there were no nonsense or frame-shifting mutations found in the ORFs of the other three frugivorous bats in this study. In addition, although the four renal genes (*SLC22A12*, *SLC2A9*, *SLC22A6* and *RHBG*) were all lost in the two Old World fruit bats, we found that all of these genes were complete and intact in the two neotropical fruit bat species.

## 4 | DISCUSSION

In this study, we report two novel genome assemblies of neotropical fruit bats (the Jamaican fruit-eating bat, *A. jamaicensis*, and the yellow-shouldered bat, *S. hondurensis*), which together provide a framework for comparative studies with Old World fruit bats. While we did not generate chromosome-level genomes such as those of six

newly sequenced bat genomes (Jebb et al., 2019), BUSCO assessments showed that our newly sequenced bat genomes are comparable to previously published ones (Table S4). Comparative genomics of 10 bat species with diverse diets across the phylogeny revealed a number of convergent genomic signatures of obligate frugivory, including genes involved in sensory perception, digestion and metabolism. Using the classic model of independently evolved obligate frugivory in Old World and neotropical fruit bats, we explored molecular adaptations contributing to convergent evolution of obligate frugivory.

Because olfactory receptor protein diversity determines the range of molecular ligands that can be recognized (Malnic, Hirono, Sato, & Buck, 1999), *OR* gene repertoires play a central role in the recognition of volatile organic compounds such as those emitted by ripe fruit. Behavioural studies have shown that olfaction could help bats locate fruits (Kalko & Condon, 1998), and morphological studies show remarkable differences in olfactory bulb size and olfactory epithelium thickness between frugivorous bats and nonfrugivorous bats (Reep & Bhatnagar, 2000), but prior analyses of *OR* gene repertoires (Hayden et al., 2014) were potentially limited in their *OR* gene sampling (Yohe et al., 2020). To overcome the limitations of prior analyses, we instead analysed the *OR* subgenome based on new and previously published genomes, identifying three subfamilies of olfactory receptor genes (*OR4*, *OR5/8/9* and *OR2/13*) with links to frugivory.

Our findings corroborate the links between *OR1/3/7* and *OR2/13* and frugivory (Hayden et al., 2014). However, the previous study, reliant on PCR, was based on representative but not comprehensive sampling of *OR* genes (Hayden et al., 2014). Indeed, Sanger-sequenced amplicons based on degenerate primers were able to retrieve less than 33% of *OR* genes from the common vampire bat genome (Yohe et al., 2020); thus, *OR* gene data identified from whole-genome sequences encompasses a less biased, more representative sample of *OR* gene diversity.

We also implemented a series of robustness checks to ensure shifts in gene family evolution were not caused by assembly artefacts or within-population variation. *OR* gene number varies across different genome assemblies or different individuals, even for the same species (Hasin et al., 2008; Nozawa, Kawahara, & Nei, 2007; Young et al., 2008). To minimize the impact of gene copy number variants on our analyses, we (a) did not use raw *OR* gene numbers, but instead used *OR* gene frequencies for each subfamily, which could reduce the impact of different genome quality; (b) identified all *OR* genes from other available versions of published genome assemblies for two fruit bat species to test the influence of different genome assembly methods and parameters; and (c) added 13 more data points to validate our results. Thus, our findings are unlikely to be the result of artefacts and are instead linked to lineage-specific traits. However, the unexpected clustering of two nonfrugivorous bat species (*C. thonglongyai* and *M. lyra*) suggests the three subfamilies we identify here *OR4*, *OR5/8/9* and *OR2/13* may have played roles unrelated to frugivory in some nonfrugivorous species, requiring further studies in the future.

We found three bitter taste receptor genes (*Tas2r11*, *Tas2r18* and *Tas2r67*) were all lost in obligate frugivorous bats. Intriguingly, there is one gene in the clade *Tas2r67* from each of the four partially frugivorous bats (*T. saurophila*, *M. hirsuta*, *A. caudifer* and *P. discolour*; Figure S6). These four species all include insects in their diets (Table S2); the functional *Tas2r67* may be required for them because of their partially insectivorous diets. This finding suggests *Tas2r11*, *Tas2r18* and *Tas2r67* may be critical for insectivorous bats but is no longer necessary for obligate frugivorous bats, possibly because of dietary shifts from their insectivorous ancestors (Emerling et al., 2018; Jiao et al., 2019). Indeed, bitter taste reception could help identify and avoid poisonous substances that typically taste bitter, such as plant secondary metabolites and insect defensive secretions (John & Hankins, 1975). A previous study examined two prevalent plant secondary metabolites (alkaloids and cyanogenic glycosides) in wild fruits and showed that of the 75 fruit species examined, only two had a high concentration of alkaloids, and three contained cyanogenic glycosides (Bondjengo et al., 2017). Hence, due to the low proportion of bitter-tasting substances in fruits, selection on the three *Tas2r* genes may have relaxed, and the genes were eventually lost in frugivorous bats. This finding agrees with the hypothesis that bitter substances in diet shaped the evolution of *Tas2r* gene repertoires in vertebrates (Li & Zhang, 2014; Wang & Zhao, 2015). Notably, the common vampire bat (*D. rotundus*) was found to possess only eight putatively functional (intact and partial) *Tas2r* genes, much fewer than among other bats (Table S9). This finding is consistent with the hypothesis that the narrow diet of vampire bats resulted in their inferior bitter taste reception (Hong & Zhao, 2014).

Digestive enzymes play important roles in breaking down nutrients for energy production (Luca, Perry, & Di Rienzo, 2010); thus, evolutionary patterns of digestive enzymes may be distinct in mammals with different diets. We found that losses of digestive enzyme genes (*CHIA*s and *PGC*s) in the obligate frugivorous bats from both the Old and neotropical lineages. A recent study examining the genome data of 107 extant placental mammals showed both a positive correlation between the number of functional *CHIA*s and the percentage of invertebrates in diet, and all *CHIA* genes were lost in Old World fruit bats (Emerling et al., 2018). In this study, we found all *CHIA* genes were also lost in neotropical obligate frugivorous bats and the common vampire bat, supporting the hypothesis that *CHIA* genes are lost as animals shift to noninsectivorous dietary niches (Emerling et al., 2018). In addition to *CHIA*s and *PGC*s, another digestive enzyme gene, encoding trehalase, has been inactivated or pseudogenized in noninsectivorous bats (Jiao et al., 2019). In contrast to gene losses, a previous study showed that activities of intestinal sucrase and maltase have been increased during the dietary shift from insectivory to nectarivory or frugivory in bats of family Phyllostomidae (Schondube, Herrera, & Martinez del Rio, 2001). In birds, diets shaped the evolution of digestive enzymes in forms of gene loss and duplication, functional divergence and positive selection (Chen & Zhao, 2019). Thus, molecular adaptations of digestive enzymes to different diets can be found in different evolutionary patterns.

Frugivorous bats are able to catabolize exogenous sugars directly and fuel their metabolism rapidly (Amitai et al., 2010); molecular adaptations to carbohydrate metabolism are thus expected. For example, the frugivorous bat *Artibeus lituratus* exhibits high insulin sensitivity and elevated glucose tolerance and is able to maintain normoglycaemia after several days of starvation (Protzek & Pinheiro, 2011a). In this study, convergent amino acid substitutions in the two lineages of fruit bats were identified in several genes involved in carbohydrate metabolism, such as *GPR107*, *PASK* and *BAIAP3*. GPR107 is involved in blood sugar regulation in response to low glucose (Elrick et al., 2016), while PASK and BAIAP3 were suggested to regulate insulin production (An et al., 2006; Wilson et al., 2005; Xavier et al., 2011; Zhang et al., 2017). These genes related to carbohydrate metabolism may have contributed to adaptation to rapid and efficient digestion of sugars for frugivorous bats. In addition, we found two renal genes (*MAGED2* and *SCNN1A*) involved in salt reabsorption have undergone convergent evolution. Indeed, a previous study showed adaptations of renal functions along with dietary shifts from insectivorous to frugivory or nectarivory in bats, because frugivores and nectarivores show reduced relative medullary thickness of kidneys, a putative indicator for estimating the ability to produce concentrated urine (Schondube et al., 2001). Although our analyses provided evidence for convergent evolution in all five genes, our conclusions are mainly based on bioinformatic inference. Future experimental assays can validate the functional changes of these convergent amino acid substitutions. In addition, five metabolic genes assumed to be associated with herbivory (*FAM3B*, *SLC22A12*, *SLC2A9*, *SLC22A6* and *RHBG*)—previously found to be lost in Old World fruit bats—remained complete and intact in neotropical fruit bats. This last finding suggests the two lineages of fruit bats might use different mechanisms to cope with similar metabolic challenges.

In conclusion, our genomewide analyses of molecular evolution have revealed several convergent molecular signatures underlying evolutionary adaptations to obligate frugivory in two distantly related lineages of frugivorous bats. We identified three subfamilies of olfactory receptor genes, losses of three bitter taste receptor genes, losses of two digestive enzyme genes and convergent amino acid substitutions in several metabolic genes that may have linked to frugivory. Future in-depth functional research of convergent amino acid substitutions identified in this study is needed for a better understanding of metabolic adaptations to obligate frugivory. Our study provides an excellent model to explore the molecular adaptations contributing to convergent evolution of obligate frugivory in mammals.

## 5 | Data and materials availability

All data needed to evaluate the conclusions in the manuscript are present in the paper and/or the Supporting Information Materials. Dataset S1 and Dataset S2 are available on Dryad (https://doi.org/10.5061/dryad.vt4b8gtpn). Additional data related to this manuscript may be requested from the authors. *A. jamaicensis* and *S. hondurensis* genomes have been deposited at GenBank under the Accession nos. VSFN00000000 and VSFL00000000. Short-read data have been deposited into the Short Read Archive under Accession nos. PRJNA554589 and PRJNA554734. The whole-genome sequence data of *A. jamaicensis* and *S. hondurensis* have also been deposited in the Genome Warehouse in BIG Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under Accession nos GWHAAYX00000000 and GWHAAZA00000000 that are publicly accessible at http://bigd.big.ac.cn/gwh. Short-read data have also been deposited into the Genome Sequence Archive in BIG Data Center under Accession nos CRA001876 and CRA001916.

## ACKNOWLEDGEMENTS

We are grateful to D.H. Zou, H. Jiao, J.M. Xia and Q. Wang for technical assistance and valuable comments. We thank R.A. Saldaña-Vázquez and G. Vázquez-Domínguez for their help in collecting the samples.

## CONFLICT OF INTERESTS

The authors declare that they have no competing interests.

## AUTHOR CONTRIBUTIONS

H.Z. conceived the study; K.W., S.T., Y.Z. and H.Z. collected and analysed the data; H.Z., K.W. and S.T. wrote the manuscript; J.G-G provided the samples and revised the manuscript; and L.M.D. commented and revised the manuscript.

## ORCID

*Huabin Zhao* https://orcid.org/0000-0002-7848-6392

## REFERENCES

Adey, A., Kitzman, J. O., Burton, J. N., Daza, R., Kumar, A., Christiansen, L., … Shendure, J. (2014). In vitro, long-range sequence information for de novo genome assembly via transposase contiguity. *Genome Research*, *24*(12), 2041–2049. https://doi.org/10.1101/gr.178319.114

Altringham, J. D. (1996). *Bats: Biology and Behavior*. Oxford, UK: Oxford University Press.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Amitai, O., Holtze, S., Barkan, S., Amichai, E., Korine, C., Pinshow, B., & Voigt, C. C. (2010). Fruit bats (Pteropodidae) fuel their metabolism rapidly and directly with exogenous sugars. *Journal of Experimental Biology*, *213*(15), 2693–2699. https://doi.org/10.1242/jeb.043505

An, R., Xavier, G. D., Hao, H. X., Semplici, F., Rutter, J., & Rutter, G. A. (2006). Regulation by Per-Arnt-Sim (PAS) kinase of pancreatic duodenal homeobox-1 nuclear import in pancreatic beta-cells. *Biochemical Society Transactions*, *34*, 791–793. https://doi.org/10.1042/BST0340791

Baldwin, M. W., Toda, Y., Nakagita, T., O'Connell, M. J., Klasing, K. C., Misaka, T., … Liberles, S. D. (2014). Evolution of sweet taste perception in hummingbirds by transformation of the ancestral umami receptor. *Science*, *345*(6199), 929–933. https://doi.org/10.1126/science.1255097345/6199/929

Benton, M. J., & Donoghue, P. C. J. (2007). Paleontological evidence to date the tree of life. *Molecular Biology and Evolution*, *24*(1), 26–53. https://doi.org/10.1093/molbev/msl150

Bergman, C. M., & Quesneville, H. (2007). Discovering and detecting transposable elements in genome sequences. *Briefings in Bioinformatics*, *8*(6), 382–392. https://doi.org/10.1093/bib/bbm048

Birdsey, G. M., Lewin, J., Holbrook, J. D., Simpson, V. R., Cunningham, A. A., & Danpure, C. J. (2005). A comparative analysis of the evolutionary relationship between diet and enzyme targeting in bats, marsupials and other mammals. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1565), 833–840. https://doi.org/10.1098/rspb.2004.3011

Birney, E., Clamp, M., & Durbin, R. (2004). GeneWise and genomewise. *Genome Research*, *14*(5), 988–995. https://doi.org/10.1101/gr.1865504

Bondjengo, N., Kitengie, G., Musibono, D., Lubini, C., Hohmann, G., & Fruth, B. (2017). Presence of alkaloids and cyanogenic glycosides in fruits consumed by sympatric bonobos and the Nkundo people at LuiKotale/Salonga National Park, Democratic Republic of Congo and its relationship to food choice. *African Primates*, *12*, 9–22.

Butler, J., MacCallum, I., Kleber, M., Shlyakhter, I. A., Belmonte, M. K., Lander, E. S., … Jaffe, D. B. (2008). ALLPATHS: De novo assembly of whole-genome shotgun microreads. *Genome Research*, *18*(5), 810–820. https://doi.org/10.1101/gr.7337908

Chen, Y. H., & Zhao, H. (2019). Evolution of digestive enzymes and dietary diversification in birds. *Peerj*, *7*, e6840. https://doi.org/10.7717/peerj.6840

Choi, Y., & Chan, A. P. (2015). PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, *31*(16), 2745–2747. https://doi.org/10.1093/bioinformatics/btv195

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*., *9*(8), 772. https://doi.org/10.1038/nmeth.2109

Davalos, L. A., & Perkins, S. L. (2008). Saturation and base composition bias explain phylogenomic conflict in Plasmodium. *Genomics*, *91*(5), 433–442. https://doi.org/10.1016/j.ygeno.2008.01.006

Duncan, R. P., Husnik, F., Van Leuven, J. T., Gilbert, D. G., Davalos, L. M., McCutcheon, J. P., & Wilson, A. C. C. (2014). Dynamic recruitment of amino acid transporters to the insect/symbiont interface. *Molecular Ecology*, *23*(6), 1608–1623. https://doi.org/10.1111/mec.12627

Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, *5*, 113. https://doi.org/10.1186/1471-2105-5-1131471-2105-5-113

Elrick, M. M., Samson, W. K., Corbett, J. A., Salvatori, A. S., Stein, L. M., Kolar, G. R., … Yosten, G. L. C. (2016). Neuronostatin acts via GPR107 to increase cAMP-independent PKA phosphorylation and proglucagon mRNA accumulation in pancreatic alpha-cells. *American Journal of Physiology-Regulatory Integrative and Comparative Physiology*, *310*(2), R143–R155. https://doi.org/10.1152/ajpregu.00369.2014

Emerling, C. A., Delsuc, F., & Nachman, M. W. (2018). Chitinase genes (*CHIA*s) provide genomic footprints of a post-Cretaceous dietary radiation in placental mammals. *Science Advances*, *4*(5), eaar6478. https://doi.org/10.1126/sciadv.aar6478

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Gunnell, G. F., & Simmons, N. B. (2005). Fossil evidence and the origin of bats. *Journal of Mammalian Evolution*, *12*(1–2), 209–246. https://doi.org/10.1007/s10914-005-6945-2

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith, R. K. Jr, Hannick, L. I., … Town, C. D. (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research*, *31*(19), 5654–5666. https://doi.org/10.1093/nar/gkg770

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., … Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biology*, *9*(1), R7. https://doi.org/10.1186/gb-2008-9-1-r7

Hasin, Y., Olender, T., Khen, M., Gonzaga-Jauregui, C., Kim, P. M., Urban, A. E., … Korbel, J. O. (2008). High-resolution copy-number variation map reflects human olfactory receptor diversity and evolution. *PLoS Genetics*, *4*(11), e1000249. https://doi.org/10.1371/journal.pgen.1000249

Hayakawa, T., Suzuki-Hashido, N., Matsui, A., & Go, Y. (2014). Frequent expansions of the bitter taste receptor gene repertoire during evolution of mammals in the Euarchontoglires clade. *Molecular Biology and Evolution*, *31*(8), 2018–2031. https://doi.org/10.1093/molbev/msu144

Hayden, S., Bekaert, M., Crider, T. A., Mariani, S., Murphy, W. J., & Teeling, E. C. (2010). Ecological adaptation determines functional mammalian olfactory subgenomes. *Genome Research*, *20*(1), 1–9. https://doi.org/10.1101/gr.099416.109

Hayden, S., Bekaert, M., Goodbla, A., Murphy, W. J., Davalos, L. M., & Teeling, E. C. (2014). A cluster of olfactory receptor genes linked to frugivory in bats. *Molecular Biology and Evolution*, *31*(4), 917–927. https://doi.org/10.1093/molbev/msu043

Hobbs, C. A., Blanchard, M. G., Alijevic, O., Da Tan, C., Kellenberger, S., Bencharit, S., … Tarran, R. (2013). Identification of the SPLUNC1 ENaC-inhibitory domain yields novel strategies to treat sodium hyperabsorption in cystic fibrosis airway epithelial cultures. *AJP Lung Cellular and Molecular Physiology*, *305*(12), L990–L1001. https://doi.org/10.1152/ajplung.00103.2013

Hong, W., & Zhao, H. (2014). Vampire bats exhibit evolutionary reduction of bitter taste receptor genes common to other bats. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1788), 20141079. https://doi.org/10.1098/rspb.2014.1079

Hu, Y., Wu, Q., Ma, S., Ma, T., Shan, L., Wang, X., … Wei, F. (2017). Comparative genomics reveals convergent evolution between the bamboo-eating giant and red pandas. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(5), 1081–1086. https://doi.org/10.1073/pnas.1613870114

Jebb, D., Huang, Z., Pippel, M., Hughes, G. M., Devanna, P., Myers, E., & Teeling, E. C. (2019). Six new reference-quality bat genomes illuminate the molecular basis and evolution of bat adaptations. *bioRxiv*, 836874.

Jeuniaux, C. (1971). On some biochemical aspects of regressive evolution in animals. In E. Schoffeniels (Ed.) *Biochemical evolution and the origin of life*. Amsterdam, The Netherlands: North-Holland.

Jiao, H., Wang, Y., Zhang, L., Jiang, P., & Zhao, H. (2018). Lineage-specific duplication and adaptive evolution of bitter taste receptor genes in bats. *Molecular Ecology*, *27*(22), 4475–4488. https://doi.org/10.1111/mec.14873

Jiao, H., Zhang, L., Xie, H. W., Simmons, N. B., Liu, H., & Zhao, H. (2019). Trehalase gene as a molecular signature of dietary diversification in mammals. *Molecular Biology and Evolution*, *36*(10), 2171–2183. https://doi.org/10.1093/molbev/msz127

John, G., & Hankins, W. (1975). The evolution of bitter and the acquisition of toxiphobia. In D. A. Denton & J. P. Coghlan (Eds.), *Olfaction and Taste V, Proceedings of the 5th International Symposium in Melbourne, Australia*. New York, NY: Academic Press.

Kalko, E. K. V., & Condon, M. A. (1998). Echolocation, olfaction and fruit display: How bats find fruit of Flagellichorous cucurbits. *Functional Ecology*, *12*(3), 364–372. https://doi.org/10.1046/j.1365-2435.1998.00198.x

Karasov, W. H., del Rio, C. M., & Caviedes-Vidal, E. (2011). Ecological physiology of diet and digestive systems. *Annual Review of*

*Physiology*, *73*, 69–93. https://doi.org/10.1146/annurev-physiol-012110-142152

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, *14*(4), R36. https://doi.org/10.1186/gb-2013-14-4-r36

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, *5*, 59. https://doi.org/10.1186/1471-2105-5-59

Laghmani, K., Beck, B. B., Yang, S. S., Seaayfan, E., Wenzel, A., Reusch, B., … Komhoff, M. (2016). Polyhydramnios, transient antenatal Bartter's syndrome, and MAGED2 mutations. *New England Journal of Medicine*, *374*(19), 1853–1863. https://doi.org/10.1056/NEJMoa1507629

Lauridsen, J. B., Johansen, J. L., Rekling, J. C., Thirstrup, K., Moerk, A., & Sager, T. N. (2011). Regulation of the Bcas1 and Baiap3 transcripts in the subthalamic nucleus in mice recovering from MPTP toxicity. *Neuroscience Research*, *70*(3), 269–276. https://doi.org/10.1016/j.neures.2011.03.011

Li, D., & Zhang, J. (2014). Diet shapes the evolution of the vertebrate bitter taste receptor gene repertoire. *Molecular Biology and Evolution*, *31*(2), 303–309. https://doi.org/10.1093/molbev/mst219

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research*, *13*(9), 2178–2189. https://doi.org/10.1101/gr.1224503

Lim, M. C. W., Witt, C. C., Graham, C. H., & Davalos, L. M. (2019). Parallel molecular evolution in pathways, genes, and sites in high-elevation hummingbirds revealed by comparative transcriptomics. *Genome Biology and Evolution*, *11*(6), 1552–1572. https://doi.org/10.1093/gbe/evz101

Luca, F., Perry, G. H., & Di Rienzo, A. (2010). Evolutionary adaptations to dietary changes. *Annual Review of Nutrition*, *30*, 291–314. https://doi.org/10.1146/annurev-nutr-080508-141048

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., … Wang, J. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience*, *1*(1), 18. https://doi.org/10.1186/2047-217X-1-18

Majoros, W. H., Pertea, M., & Salzberg, S. L. (2004). TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. *Bioinformatics*, *20*(16), 2878–2879. https://doi.org/10.1093/bioinformatics/bth315

Malnic, B., Hirono, J., Sato, T., & Buck, L. B. (1999). Combinatorial receptor codes for odors. *Cell*, *96*(5), 713–723. https://doi.org/10.1016/s0092-8674(00)80581-4

Marcais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, *27*(6), 764–770. https://doi.org/10.1093/bioinformatics/btr011

Mariman, E. C. M., Bouwman, F. G., Aller, E. E. J. G., van Baak, M. A., & Wang, P. (2015). Extreme obesity is associated with variation in genes related to the circadian rhythm of food intake and hypothalamic signaling. *Physiological Genomics*, *47*(6), 225–231. https://doi.org/10.1152/physiolgenomics.00006.2015

Martin, C. H., & Richards, E. J. (2019). The paradox behind the pattern of rapid adaptive radiation: How can the speciation process sustain itself through an early burst? *Annual Review of Ecology, Evolution, and Systematics*, *50*, 569–593. https://doi.org/10.1146/annurev-ecolsys-110617-062443

Meyer, W. K., Jamison, J., Richter, R., Woods, S. E., Partha, R., Kowalczyk, A., … Clark, N. L. (2018). Ancient convergent losses of Paraoxonase 1 yield potential risks for modern marine mammals. *Science*, *361*(6402), 591–594. https://doi.org/10.1126/science.aap7714

Miller-Butterworth, C. M., Murphy, W. J., O'Brien, S. J., Jacobs, D. S., Springer, M. S., & Teeling, E. C. (2007). A family matter: Conclusive resolution of the taxonomic position of the long-fingered bats, Miniopterus. *Molecular Biology and Evolution*, *24*(7), 1553–1561. https://doi.org/10.1093/molbev/msm076

Neuweiler, G. (2000). *The biology of bats*. Oxford, UK: University Press.

Nozawa, M., Kawahara, Y., & Nei, M. (2007). Genomic drift and copy number variation of sensory receptor genes in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(51), 20421–20426. https://doi.org/10.1073/pnas.0709956104

Olland, A. M., Strand, J., Presman, E., Czerwinski, R., Joseph-McCarthy, D., Krykbaev, R., … Mosyak, L. (2009). Triad of polar residues implicated in pH specificity of acidic mammalian chitinase. *Protein Science*, *18*(3), 569–578. https://doi.org/10.1002/pro.63

Orme, D., Freckleton, R., Thomas, G., & Petzoldt, T. (2013). The caper package: Comparative analysis of phylogenetics and evolution in R. *R Package Version*, *5*(2), 1–36.

Pajic, P., Pavlidis, P., Dean, K., Neznanova, L., Romano, R. A., Garneau, D., … Gokcumen, O. (2019). Independent amylase gene copy number bursts correlate with dietary preferences in mammals. *Elife*, *8*, e44628. https://doi.org/10.7554/eLife.44628

Parra, G., Bradnam, K., & Korf, I. (2007). CEGMA: A pipeline to accurately annotate core genes in eukaryotic genornes. *Bioinformatics*, *23*(9), 1061–1067. https://doi.org/10.1093/bioinformatics/btm071

Protzek, A. O. P., & Pinheiro, E. C. (2011a). Metabolic adaptations in frugivorous bats. In J. L. Zupan & S. L. Mlakar (Eds.), *Bats: Biology, behavior and conservation* (pp. 299–313). Hauppauge, NY: Nova Science.

Protzek, A. O. P., Pinheiro, E., & d. C. , (2011b). Metabolic adaptations in frugivorous bats. In *Bats: Biology, behavior and conservation*. Hauppauge, NY: Nova Science Publishers Inc.

Reep, R. L., & Bhatnagar, K. P. (2000). Brain ontogeny and ecomorphology in bats. In R. A. Adams, & S. C. Pedersen (Eds.), *Ontogeny, functional ecology and evolution of bats* (pp. 93–136). Cambridge, UK: Cambridge University Press.

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Rey, C., Gueguen, L., Semon, M., & Boussau, B. (2018). Accurate detection of convergent amino-acid evolution with PCOC. *Molecular Biology and Evolution*, *35*(9), 2296–2306. https://doi.org/10.1093/molbev/msy114

Salih, M., Gautschi, I., van Bemmelen, M. X., Di Benedetto, M., Brooks, A. S., Lugtenberg, D., … Hoorn, E. J. (2017). A missense mutation in the extracellular domain of alpha ENaC causes liddle syndrome. *Journal of the American Society of Nephrology*, *28*(11), 3290–3298. https://doi.org/10.1681/ASN.2016111163

Schondube, J. E., Herrera, M. L., & Martinez del Rio, C. (2001). Diet and the evolution of digestion and renal function in phyllostomid bats. *Zoology*, *104*(1), 59–73. https://doi.org/10.1078/0944-2006-00007

Sharma, V., Hecker, N., Roscito, J. G., Foerster, L., Langer, B. E., & Hiller, M. (2018). A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nature Communications*, *9*(1), 1215. https://doi.org/10.1038/s41467-018-03667-1

She, R., Chu, J. S., Wang, K., Pei, J., & Chen, N. (2009). GenBlastA: Enabling BLAST to identify homologous gene sequences. *Genome Research*, *19*(1), 143–149. https://doi.org/10.1101/gr.082081.108

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Simmons, N. B., Seymour, K. L., Habersetzer, J., & Gunnell, G. F. (2008). Primitive Early Eocene bat from Wyoming and the evolution of flight

and echolocation. *Nature*, *451*(7180), 818–821. nature06549 [pii]. 10.1038/nature06549

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stanke, M., & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*, *19*, ii215-ii225. https://doi.org/10.1093/bioinformatics/btg1080

Strobel, S., Roswag, A., Becker, N. I., Trenczek, T. E., & Encarnacao, J. A. (2013). Insectivorous bats digest chitin in the stomach using acidic mammalian chitinase. *PLoS One*, *8*(9), e72770. https://doi.org/10.1371/journal.pone.0072770

Taggart, R. T., Cass, L. G., Mohandas, T. K., Derby, P., Barr, P. J., Pals, G., & Bell, G. I. (1989). Human Pepsinogen C (Progastricsin). Isolation of cDNA clones, localization to chromosome 6, and sequence homology with Pepsinogen A. *Journal of Biological Chemistry.*, *264*(1), 375–379.

Teeling, E. C., Springer, M. S., Madsen, O., Bates, P., O'Brien, S. J., & Murphy, W. J. (2005). A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science*, *307*(5709), 580–584. https://doi.org/10.1126/science.1105113

Teeling, E. C., Vernes, S. C., Davalos, L. M., Ray, D. A., Gilbert, M. T. P., Myers, E. & Bat1K-Consortium (2018). Bat biology, genomes, and the Bat1K project: To generate chromosome-level genomes for all living bat species. *Annu Rev Anim Biosci*, *6*, 23–46. https://doi.org/10.1146/annurev-animal-022516-022811

Thomas, G. W. C., & Hahn, M. W. (2015). Determining the null model for detecting adaptive convergence from genomic data: A case study using echolocating mammals. *Molecular Biology and Evolution*, *32*(5), 1232–1236. https://doi.org/10.1093/molbev/msv013

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., … Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, *7*(3), 562–578. https://doi.org/10.1038/nprot.2012.016

Voigt, C. C., Zubaid, A., Kunz, T. H., & Kingston, T. (2011). Sources of assimilated proteins in Old and New World phytophagous bats. *Biotropica*, *43*(1), 108–113. https://doi.org/10.1111/j.1744-7429.2010.00632.x

Voilley, N., Lingueglia, E., Champigny, G., Mattéi, M. G., Waldmann, R., Lazdunski, M., & Barbry, P. (1994). The lung amiloride-sensitive Na+ channel: Biophysical properties, pharmacology, ontogenesis, and molecular cloning. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(1), 247–251. https://doi.org/10.1073/pnas.91.1.247

Vonk, F. J., Casewell, N. R., Henkel, C. V., Heimberg, A. M., Jansen, H. J., McCleary, R. J., … Richardson, M. K. (2013). The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(51), 20651–20656. https://doi.org/10.1073/pnas.1314702110

Wang, B. J., Xia, J. M., Wang, Q., Yu, J. L., Song, Z., & Zhao, H. (2019). Diet and adaptive evolution of alanine-glyoxylate aminotransferase mitochondrial targeting in birds. *Molecular Biology and Evolution*, *37*(3), 786–798. doi:10.1093/molbev/msz266

Wang, K., & Zhao, H. (2015). Birds generally carry a small repertoire of bitter taste receptor genes. *Genome Biology and Evolution*, *7*(9), 2705–2715. https://doi.org/10.1093/gbe/evv180

Wilman, H., Belmaker, J., Simpson, J., de la Rosa, C., Rivadeneira, M. M., & Jetz, W. (2014). EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology*, *95*(7), 2027–2027. https://doi.org/10.1890/13-1917.1

Wilson, D. E., & Reeder, D. M. (2005). *Mammal species of the world. A taxonomic and geographic reference* (3rd edn). Baltimore, MD: Johns Hopkins University Press.

Wilson, W. A., Skurat, A. V., Probst, B., de Paoli-Roach, A., Roach, P. J., & Rutter, J. (2005). Control of mammalian glycogen synthase by PAS kinase. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16596–16601. https://doi.org/10.1073/pnas.0508481102

Wojcik, S. M., Tantra, M., Stepniak, B., Man, K. N. M., Muller-Ribbe, K., Begemann, M., … Ehrenreich, H. (2013). Genetic markers of a Munc13 protein family member, BAIAP3, are gender specifically associated with anxiety and benzodiazepine abuse in mice and humans. *Molecular Medicine*, *19*, 135–148. https://doi.org/10.2119/molmed.2013.00033

da Xavier, G. D., Farhan, H., Kim, H., Caxaria, S., Johnson, P., Hughes, S., … Rutter, G. A. (2011). Per-arnt-sim (PAS) domain-containing protein kinase is downregulated in human islets in type 2 diabetes and regulates glucagon secretion. *Diabetologia*, *54*(4), 819–827. https://doi.org/10.1007/s00125-010-2010-7

Yang, Z. (1997). PAML: A program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences*, *13*(5), 555–556. https://doi.org/10.1093/bioinformatics/13.5.555

Yang, Z., & Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: A Markov Chain Monte Carlo method. *Molecular Biology and Evolution*, *14*(7), 717–724. https://doi.org/10.1093/oxfordjournals.molbev.a025811

Yohe, L. R., Davies, K. T. J., Simmons, N. B., Sears, K. E., Dumont, E. R., Rossiter, S. J., & Dávalos, L. M. (2020). Evaluating the performance of targeted sequence capture, RNA-seq, and degenerate-primer PCR cloning for sequencing the largest mammalian multigene family. *Molecular Ecology Resources*, *20*(1), 140–153. https://doi.org/10.1111/1755-0998.13093

Yosten, G. L. C., Redlinger, L. J., & Samson, W. K. (2012). Evidence for an interaction of neuronostatin with the orphan G protein-coupled receptor, GPR107. *American Journal of Physiology-Regulatory Integrative and Comparative Physiology*, *303*(9), R941–R949. https://doi.org/10.1152/ajpregu.00336.2012

Young, J. M., Endicott, R. M., Parghi, S. S., Walker, M., Kidd, J. M., & Trask, B. J. (2008). Extensive copy-number variation of the human olfactory receptor gene family. *American Journal of Human Genetics*, *83*(2), 228–242. https://doi.org/10.1016/j.ajhg.2008.07.005

Zhang, J., & Kumar, S. (1997). Detection of convergent and parallel evolution at the amino acid sequence level. *Molecular Biology and Evolution*, *14*(5), 527–536. https://doi.org/10.1093/oxfordjournals.molbev.a025789

Zhang, X. M., Jiang, S., Mitok, K. A., Li, L. J., Attie, A. D., & Martin, T. F. J. (2017). BAIAP3, a C2 domain-containing Munc13 protein, controls the fate of dense-core vesicles in neuroendocrine cells. *Journal of Cell Biology*, *216*(7), 2151–2166. https://doi.org/10.1083/jcb.201702099

Zhao, H., Zhou, Y., Pinto, C. M., Charles-Dominique, P., Galindo-Gonzalez, J., Zhang, S., & Zhang, J. (2010). Evolution of the sweet taste receptor gene *Tas1r2* in bats. *Molecular Biology and Evolution*, *27*(11), 2642–2650. https://doi.org/10.1093/molbev/msq152

Zou, Z., & Zhang, J. (2015). Are convergent and parallel amino acid substitutions in protein evolution more prevalent than neutral expectations? *Molecular Biology and Evolution*, *32*(8), 2085–2096. https://doi.org/10.1093/molbev/msv091

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Wang K, Tian S, Galindo-González J, Dávalos LM, Zhang Y, Zhao H. Molecular adaptation and convergent evolution of frugivory in Old World and neotropical fruit bats. *Mol Ecol*. 2020;29:4366–4381. https://doi.org/10.1111/mec.15542