Unraveling molecular mechanisms of immunity and cancer-resistance using the genomes of the Neotropical bats *Artibeus jamaicensis* and *Pteronotus mesoamericanus*

Armin Scheben[1], Olivia Mendivil Ramos[2], Melissa Kramer[2], Sara Goodwin[2], Sara Oppenheim[3], Daniel J Becker[4], Michael C Schatz[1,5], Nancy B Simmons[6], Adam Siepel[1*], W Richard McCombie[2*]

[1] Cold Spring Harbor Laboratory, Simons Center for Quantitative Biology, Cold Spring Harbor, NY
[2] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
[3] American Museum of Natural History, Sackler Institute for Comparative Genomics, New York, NY
[4] University of Oklahoma, Department of Biology, Norman, OK
[5] Johns Hopkins University, Departments of Computer Science and Biology, Baltimore, MD
[6] American Museum of Natural History, Department of Mammalogy, Division of Vertebrate Zoology, New York, NY

*To whom correspondence should be addressed: asiepel@cshl.edu, mccombie@cshl.edu

Abstract

Bats are exceptional among mammals for harbouring diverse pathogens and for their robust immune systems. In addition, bats are unusually long-lived and show low rates of cancer. Contiguous and complete reference genomes are needed to determine the genetic basis of these adaptations and establish bats as models for research into mammalian health. Here we sequenced and analysed the genomes of the Jamaican fruit bat (*Artibeus jamaicensis*) and the Mesoamerican mustached bat (*Pteronotus mesoamericanus*). We sequenced these two species using a mix of Illumina and Oxford Nanopore Technologies (ONT), assembling draft genomes with some of the highest contig N50s (28-29Mb) of bat genomes to date. Work is in progress to increase the base-level accuracies of these genomes. We conducted gene annotation and identified a set of 10,928 orthologs from bats and mammalian outgroups including humans, rodents, horses, pigs, and dogs. To detect positively selected genes as well as lineage-specific gene gains and losses, we carried out comprehensive branch-site likelihood ratio tests and gene family size analyses. Our analysis found signatures of rapid evolution in the innate immune response genes of bats, and evidence of past infections with diverse viral clades in *Artibeus jamaicensis* and *Pteronotus mesoamericanus*. We additionally found evidence of positive selection of tumor suppressors, which may play a role in the low cancer rates, in the most recent common ancestor of bats. These new genomic resources enable insights into the extraordinary adaptations of bats, with implications for mammalian evolutionary studies and public health.

Introduction

Bats are the second largest group of mammals and are known for a wide variety of unique adaptations, including powered flight [1], laryngeal echolocation [2,3], numerous dietary specializations [4,5], unusual longevity [6], and low rates of cancer [7]. Bats are also likely asymptomatic carriers of diverse viruses [8,9], and have played a role in outbreaks of the emerging zoonotic diseases Marburg virus [10], Nipah virus [11], and SARS (severe acute respiratory syndrome) [12] either through direct human contact or intermediate hosts. It has been suggested that their tolerance of many viral infections stems from features of the bat antiviral immune response [13]. This array of adaptations makes bats an intriguing group of animals in which to investigate genotype-to-phenotype relationships, many of which may have implications for human health. For example, by better understanding the mechanisms of the bat immune system that may allow them to tolerate viral infections [7], we can help prevent zoonotic outbreaks. Comparative genomic analyses of bats and cancer-susceptible mammals can also potentially allow for powerful inferences about the causes of cancer. Overall, the genomic study of non-model mammalian species such as bats can complement research on the widespread mouse models by shedding light on novel adaptive mechanisms to overcome diseases that impact humans.

In the past decade, the genomes of at least 41 bat species have been sequenced and assembled (**Supplementary Table 1**). A key result of this genomic research on bats is recognition that there are substantial differences in the immune system within bats and between bats and other mammals [13–17]. For instance, bats have lost the PYHIN gene family involved in inflammatory response [15], while also experiencing positive selection of pro-inflammatory transcription factor nuclear factor-kappaB (NF-κB) regulators [15,16]. Gene duplications of natural killer receptors [13] and antiviral genes [16] have also been found. Possibly as a result of a history of coevolution with viruses in their evolutionary past [16,18], bats have apparently thus evolved innovations in antiviral defence and inflammatory response. Although little is known about bat-specific genetic mechanisms of cancer resistance, previous studies have implicated genetic modifications in telomere maintenance genes [19], tumor suppressors (including *TP53* and *BRCA2*) and DNA damage checkpoint-DNA repair pathway genes [15] as well as growth hormone [20]. Other studies of bats have detected gene losses and positive selection associated with dietary specializations [17,21–23] and echolocation [24,25].

Except for two recently published studies [13,16], most previous genomic investigations in bats have relied on short-read assemblies based on Illumina DNA sequencing technology. Such assemblies benefit from the high accuracy of short reads, but are generally limited by having relatively poor contiguity, and typically consist of tens of thousands of scaffolds. As a result, repetitive regions in these assemblies may be collapsed, leading to incomplete reconstructions of genome content and architecture. Incomplete assemblies with low contiguity hamper the detection of genomic diversity including structural variants, which may play an important role in trait evolution. Therefore, reference genomes of consistent quality and completeness are important for further unravelling the genotypes underlying evolutionary adaptations in bats [26]. Additionally, genomic research has often focused on Old World bats, with a limited sampling of Neotropical bats that have adapted to different ecological niches and pathogen pressures.

To address these deficiencies, we sequenced and assembled the genomes of two Neotropical bats, the Jamaican fruit bat (*Artibeus jamaicensis*) and the Mesoamerican mustached bat (*Pteronotus mesoamericanus*) using a mix of short and long read sequencing. *Artibeus jamaicensis* is one of the most common bats in the tropical Americas, with microbial studies indicating that it is an important host for an array of viruses [27]. Its distant relative, *Pteronotus mesoamericanus*, occurs in a more restricted range between Mexico and Panama [28]. *P. mesoamericanus* is a member of a small clade – the *Pteronotus parnellii* group – that is unique among Neotropical bats in that it uses high-duty cycle echolocation [29], which likely evolved to improve the ability to hunt fluttering prey [30]. *Artibeus jamaicensis* and *Pteronotus mesoamericanus* both belong to the Noctilionoidea, a Neotropical superfamily that comprises ~16% of global bat diversity (www.batnames.org). Within this superfamily, the fruit-eating *Artibeus jamaicensis* represents the species-rich Phyllostomidae (>200 species), which evolved from an insect-eating ancestor and radiated into a wide range of ecological niches with specialized diets including consumption of fruit, nectar, insects, vertebrates, and blood. The insect-eating *Pteronotus mesoamericanus* represents the less diverse Mormoopidae (18 species), which diverged from the phyllostomids ~35 million years ago [31]. *P. mesoamericanus* is therefore a useful outgroup to study the phyllostomid radiation. Our long read genome assemblies enable a comprehensive comparison of *P. mesoamericanus* and *Artibeus jamaicensis* with one another and with previously assembled bat genomes, to help gain insights into antiviral immune responses, cancer resistance and diverse feeding strategies.

Results and discussion

Genome assembly and annotation

We generated over 131x coverage of the *Artibeus jamaicensis* genome using Oxford Nanopore Technologies (ONT) long sequencing reads, with 38x coverage from >30kb fragments; and 156x coverage of the *Pteronotus mesoamericanus* genome, with 37x coverage from >30kb fragments. To correct ONT-associated single base sequencing errors, we also generated 30x coverage with short Illumina reads for both bats. We performed a k-mer analysis using Genomescope2 [32] (**Supplementary Figure 1**) to predict the size of the genome (1.98 Gb for *Artibeus jamaicensis* and 2.02 Gb for *Pteronotus mesoamericanus*) and assess the amount of heterozygosity (0.8% *Artibeus jamaicensis* and 0.36% *Pteronotus mesoamericanus*). Heterozygosity in *Artibeus jamaicensis* is roughly twice as high as in *Pteronotus mesoamericanus* and *Myotis brandtii* [20], possibly reflecting the larger population size of the widespread *Artibeus jamaicensis*. We generated long-read assemblies using wtdbg2 [33] followed by genome polishing with short reads. The resulting reference-quality assemblies have contig N50 values of 28-29Mb, which are the second and third highest values for any bat genomes to date (**Figure 1**, **Supplementary Table 1**). Our maximum contig size is >100Mb, the largest for any bat genome yet assembled. We confirmed the contiguity of our assemblies by comparison with the chromosome-level assemblies of *Phyllostomus discolor* and *Rhinolophus ferrumequinum* using Assemblytics (**Supplementary Figure 2**). In comparison to previous short-read assemblies in bats, our assemblies show dramatically higher contiguities (**Figure 1**, **Table 1**) and have fewer gaps.

Based on a BUSCO assessment of our draft assemblies, the gene sets in both bats are fairly complete at 86.5% in *Pteronotus mesoamericanus* and 80.3% in *Artibeus jamaicensis*. In comparison, recently published high-quality bat genomes showed somewhat higher BUSCO scores of 90.6-93.0%, whereas an earlier Illumina-based draft assembly of *A. jamaicensis* showed 69.8% (**Supplementary Figure 3**). Further inspection showed that our BUSCO scores are likely reduced due to limitations in base-level accuracy associated with ONT sequencing, which we expect can be improved substantially by further genome polishing[34]. Work is in progress to optimize our polishing approach to further increase base-level accuracy and gene completeness.

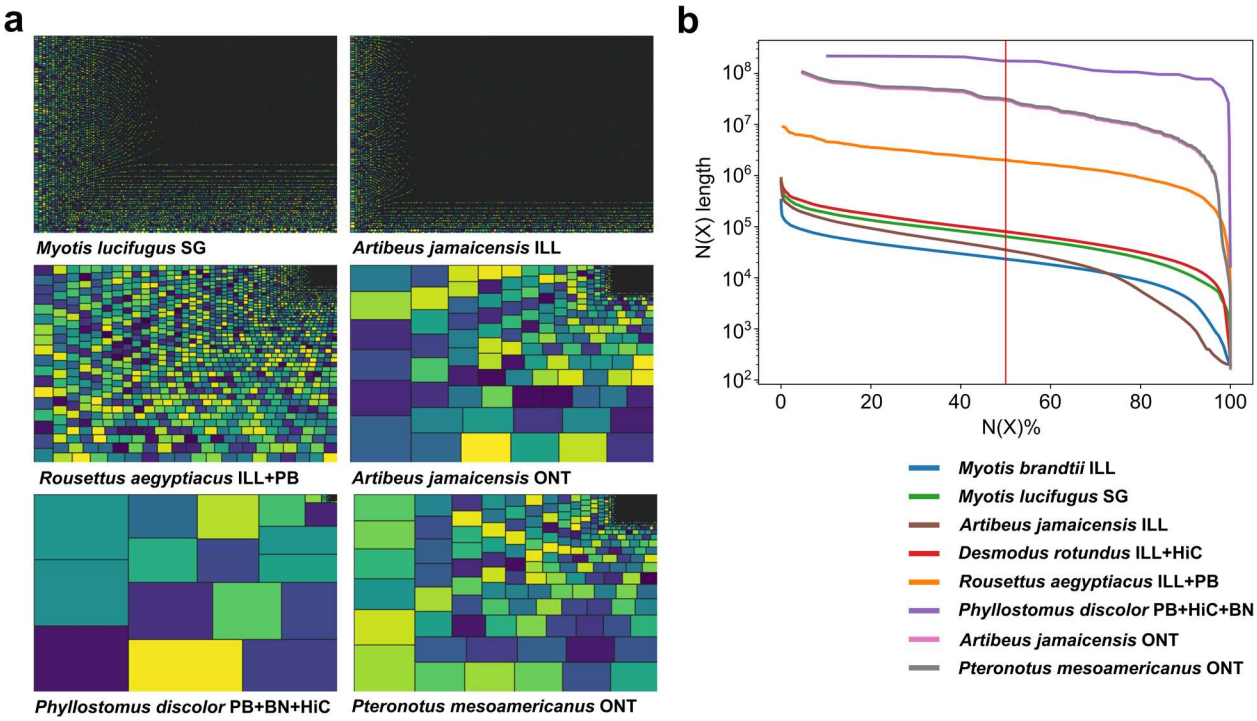


Figure 1. Assembly quality statistics for several bat genomes. a) Six treemaps [35] comparing bat assemblies based on different technologies. The whole rectangle area denotes the whole genome and each interior rectangle represents the relative size of each contig (or chromosome, in the case of *Phyllostomus discolor)*. SG denotes Sanger sequencing, ONT denotes Oxford Nanopore Technologies, ILL denotes Illumina, BN denotes BioNano and HiC denotes HiC-sequencing. b) Cumulative sum of the length of the ordered contigs (or chromosome, in the case of *P. discolor*) versus the cumulative sum of the total proportion of the genome. The vertical red line indicates the N50 metric. The curve for *Pteronotus mesoamericanus* has been jittered upwards to prevent it fully overlapping the curve for *Artibeus jamaicensis*.

Table 1. Contiguity measures of the unscaffolded *Artibeus jamaicensis* and *Pteronotus mesoamericanus* assemblies compared to scaffolded bat genomes assembled using different technologies. The Illumina assembly of *Artibeus jamaicensis* is unpublished but available from GenBank via accession GCA_004027435.1.

	Illumina		**Illumina+ HiC**	**Illumina + PacBio**	**PacBio + HiC + BioNano**	**ONT**	
Species	*Leptonycteris yerbabuenae*	*Artibeus jamaicensis*	*Desmodus rotundus*	*Rousettus aegyptiacus*	*Phyllostomus discolor*	*Artibeus jamaicensis*	*Pteronotus mesoamericanus*
# Scaffolds (Scaf)	34,419	794,376	29,801	2,490	141	-	-
# Contigs (Ctg)	121,525	798,700	84,109	3,049	964	1,086	3,238
Scaf Mb	2,054	2,425	2,064	1,910	2,118	-	-
Ctg Mb	2,049	2,424	2,018	1,910	2,095	2,037	2,100
Gap %	0.22	0.02	2.23	0.03	1.08	0.00	0.01
Scaf N50	14.74	0.04	26.87	2.01	110.24	-	-
Scaf L50	38	16,717	26	297	6	6	6
Ctg N50	0.04	0.03	0.08	1.49	6.89	29.04	28.29
Ctg L50	16,248	18,729	6,951	393	94	22	22
Scaf N90	0.48	0.00	9.95	0.55	31.41	-	-
Scaf L90	276	132,515	73	978	16	-	-
Ctg N90	0.01	0.00	0.02	0.40	1.26	5.13	3.99
Ctg L90	57,040	137,099	27,361	1,312	351	89	99
Scaf max	70.82	0.88	73.32	9.00	210.69	-	-
Ctg max	0.35	0.56	0.76	7.81	24.91	82.60	100.93
GC avg	0.42	0.42	0.42	0.40	0.42	0.41	0.42
GC std	0.08	0.09	0.09	0.07	0.08	0.12	0.08
Reference	[23]	-	[17]	[13]	[16]	This study	This study

We applied the MAKER3 pipeline [36] to annotate the two genomes using short read RNA-sequencing (RNA-seq) data (**Supplementary Table 2**), protein evidence, and de novo gene predictions, identifying 23,145 genes in *Pteronotus mesoamericanus* and 27,600 genes in *Artibeus jamaicensis* (**Supplementary Table 3**). BUSCO analysis based on the final protein sequences showed reduced completeness compared to the genome-based analysis, with 74.3% in *A. jamaicensis* and 79.3% in *Pteronotus mesoamericanus*. This reduction reflects differences between the BUSCO analysis and the annotation approach, with the annotation possibly being more sensitive to base-level inaccuracies. For the annotation, we were also limited by using publicly available RNA-seq from different *Artibeus jamaicensis* individuals and a different *Pteronotus* species. A total of 20,252 genes in *P. mesoamericanus* genes 20,178 in *Artibeus jamaicensis* were annotated with a PFAM domain or a PANTHER gene family using Interproscan. These numbers of genes with PFAM or PANTHER annotation are closer to the expected number of ~20,000 genes in bats. The additional gene models generated therefore likely reflect partial or incorrect gene models. We expect our work towards the improvement of the base-level accuracy of the assembly will also lead to substantial improvements in the annotation.

Repeat analysis

With a typical size of ~2 Gb (**Supplementary Table 1**), bat genomes are substantially smaller than most other mammalian genomes, which average ~3.5 Gb in size [37]. This genomic reduction has been linked to increased rates of genetic sequence loss [38] and lower repeat content [16,20]. Interestingly, it has also been suggested that rapid expansions of repeats in bat genomes may have facilitated genetic sequence loss via nonallelic homologous recombination [38]. We found a total of 38.6% of the *A. jamaicensis* and 36.7% of the *Pteronotus mesoamericanus* genome consisted of repeats, with long interspersed nuclear elements (LINEs) making up the largest repeat class (**Figure 2a**) as has been found in other bat genomes [16]. Within the four noctilionid bats that we compared, we found highly homogenous landscapes of repeat classes. We also analysed repeats in two previously published genomes of *Artibeus jamaicensis*, which were generated using short reads. Here we found substantial deviations from the landscapes observed based on our assembly, and from those found in the other bats. For the unpublished Illumina assembly of *A. jamaicensis* (GenBank accession GCA_004027435.1), we found that the repetitive content (40.23%) and total genome size (2.42 Gb) was unexpectedly elevated relative to our assembly and other bat assemblies. In contrast, a recently published genome of *A. jamaicensis* [23] based on a consensus of Illumina reads aligned to *Leptonycteris yerbabuenae* showed a substantial decrease across all repeat categories, with only 30.24% of the genome identified as repetitive. Although both previously generated Illumina-based *Artibeus jamaicensis* genomes show similar relative proportions of repeat classes compared to our novel assembly, they exhibit large differences in the absolute numbers of repetitive bases of different repeat classes and families (**Figure 2a**). This observation suggests that Illumina-based assemblies can under- or over-estimate repeat content by substantial margins.

Bats exhibit a wide range of repeat families with different recent accumulations, amounting to 0.38-4.2% of the genome [39]. We found recent repeat accumulations within these expectations, observing 0.54% in *Pteronotus mesoamericanus* and 0.45% in *Artibeus jamaicensis*. The composition of Class I and Class II transposon repeat classes within recently diverged elements

differed broadly between species (**Figure 2b**, **Supplementary Table 4**), despite relatively close phylogenetic relationships. Class I transposons (LINEs, SINEs and LTRs) are prevalent in mammalian genomes [40], but DNA-mediated Class II transposons (e.g., hAT and piggyBac elements) are rarer and have been inactive in most mammals during the last 40 million years [41]. Vespertilionid bats are a notable exception, with substantial activity of Class II transposons in the recent past [16,42].

We found a recent expansion of the yangochiropteran tRNA-derived SINEs Ves40 and Ves41 in *A. jamaicensis*. Smaller recent expansion of the Ves elements Ves30,Ves40, Ves43,Ves42 and Ves41 occur in all of the noctilionid bats analysed here. These elements may be noctilionid-specific, as they do not occur in vespertilionid bats [43]. *Pteronotus mesoamericanus* experienced a major recent expansion of lineage-specific Gypsy LTRs, which makes up over 60% of recently accumulated transposons in this species. Although Class II transposons such as hAT and piggyBac were not found in previous studies of *Artibeus* and *Pteronotus* using PCR detection [44], we found recent activity of piggyBac elements in *P. mesoamericanus* and *Phyllostomus discolor*. PiggyBac elements have also been found to be active in the little brown bat (*Myotis lucifugus*) [42] as well as in primates in the past 40myr [45]. In *Artibeus jamaicensis*, we detected a recent lineage-specific expansion of a Class II Harbinger element. Additionally, we detected recent accumulations of hAT elements in *A. jamaicensis* and common vampire bat (*Desmodus rotundus*).

In rare cases, the detected transposons may have played a role in the evolution of noctilionid bat genomes by inserting into or near functional genomic regions of the host, or by transforming into novel host genes [46]. The dynamic nature of the noctilionid repeat landscape reflects a general trend in bats to show recent expansions of repeat classes that have been recently inactive in most other mammals [16].

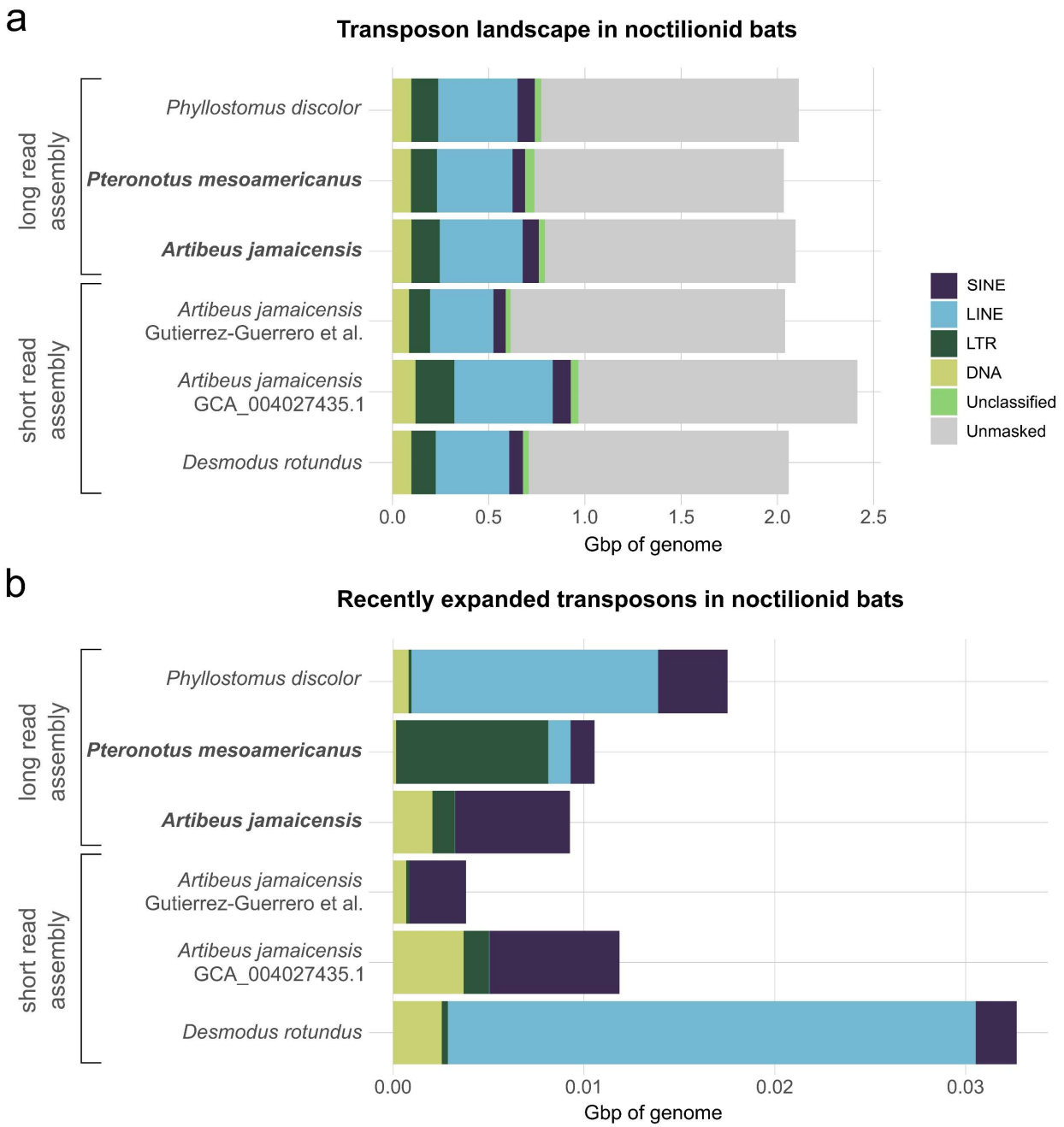


Figure 2. RepeatMasker analysis of repeat classes in noctilionid genomes. Newly assembled genomes presented here are shown in bold. We provide the accession of the unpublished Illumina-based *Artibeus jamaicensis* assembly; the other Illumina-based *A. jamaicensis* consensus sequence was published by Gutiérrez-Guerrero et al. [23]. a) Fraction of genome covered by all repeat classes. b) Fraction of genome covered by recently diverged sequences within each repeat class.

Mining of endogenous viral elements

Retroviruses integrate into host genomes as part of their life cycle, and retroviral sequences are frequently found as endogenous viral elements (EVEs) in mammalian genomes [47,48]. In some cases, EVEs gain gene-regulatory functions in mammalian host genomes [49]. Non-retroviral elements may also be integrated into host genomes and can play a role in host evolution [47]. Importantly, EVEs can provide evidence of past interactions between hosts and viruses enabling insights into viral diversity and host range in hosts of potential zoonotic pathogens such as bats.

We found that retroviruses were the most frequent EVEs in the noctilionids, but we also detected a diverse set of non-retroviral EVEs (**Supplementary Tables 5-8**). The most frequent non-retroviral families that we detected were Herpesviridae, Flaviviridae, Bornaviridae, and Parvoviridae (**Supplementary Table 5**). Non-retroviral EVEs in other mammals are also commonly derived from the Bornaviridae and Parvoviridae families [50]. The non-retroviral genera occurring most frequently were *Carbovirus*, *Rhadinovirus,* and *Pestivirus* (**Supplementary Table 6**).

Pestivirus is a genus of flaviviruses that infect and cause economically important diseases in domestic animals, especially swine and ruminants [51]. Novel pestiviruses have been identified in bat species (particularly *Rhinolophus affinis*) from China [52,53]. However, we are unaware of prior pestivirus detections in Neotropical bat species, and sequences were identified within all four of the analysed Neotropical bat genomes. Pestiviruses thus may have a broader host range across bats. In particular, detection of these viruses in the blood-feeding *Desmodus rotundus* is also important from the perspective of contemporary cross-species transmission, as this bat regularly feeds on domestic animals such as pigs and cattle [54,55].

Moreover, we found evidence of avian viruses (*Alpharetrovirus* and *Avipoxvirus*) that are rarely found in mammals. Alpharetroviruses have been detected in bats previously [56], but their presence in the genomes of all four Neotropical bat species here suggests a broader host range among bats. *Avipoxvirus* is a genus of poxviruses that are typically associated with mortality in songbirds and thought to be primarily transmitted by mosquito vectors [57]. Members of this viral genus have been previously identified in diverse bat species in China [58]. However, identification of avipoxvirus sequences in all four Neotropical bat genomes again suggests a broader geographic and evolutionary range of these viruses in bats. Similarly, other viruses detected here in the noctilionids suggest infection with a broad range of viruses during bat evolution in the Neotropics.

Positive selection analysis

Similarly to other mammals, most genes in bats have predominantly experienced purifying (negative) selection, but certain classes of genes show signatures of rapid evolution driven by positive selection. We carried out branch-site likelihood ratio tests for positive selection [59,60] on 10,928 single-copy orthologs detected in bats and a set of outgroup mammals (**Figure 3**). We tested for positive selection on the branch leading to all bats, as well as the branches leading to the nested groups Yangochiroptera, Noctilionoidea, and Phyllostomidae. In the following, we refer

to the most recent common ancestor of bats as the bat ancestor and the incident branch as the bat ancestral branch. We also tested each of the three terminal branches to *Artibeus jamaicensis*, *Pteronotus mesoamericanus* and the naked mole rat. We included the naked mole rat because earlier research had identified convergently evolved genes in bats and the naked mole rat, which may be associated with the heightened longevity of these animals [61].

We found 225 genes positively selected on the bat ancestral branch (**Supplementary Table 9**). These genes were enriched for immune-related functions, particularly innate immune response (GO:0045087, p=0.0007), as well as for lipid metabolism and reproductive functions (**Supplementary Table 10**). In the Noctilionoidea and Phyllostomidae, immune system gene groups also dominated the enrichment analysis. While these broader enriched categories are frequently enriched in studies of positive selection in mammals [62–65], we focused on positively selected genes that may be associated with bat-specific adaptations. Based on gene ontologies, we identified orthologs associated with immunity, DNA repair, telomere maintenance, aging, reproduction, thermoregulation, hearing and digestion.

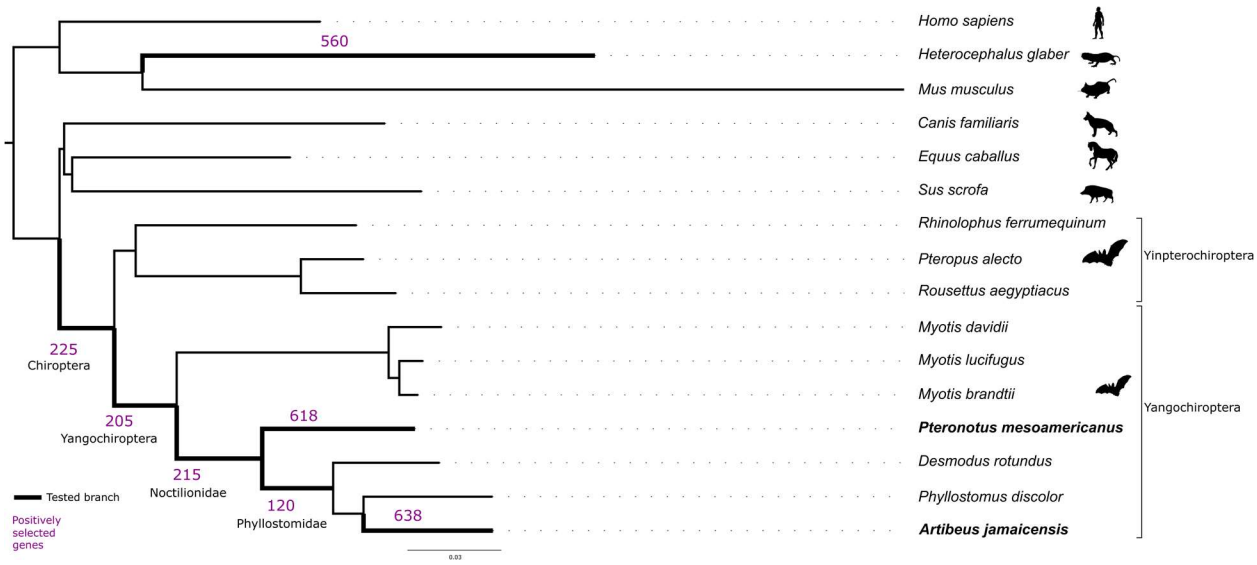


Figure 3. Maximum likelihood phylogeny based on codon-site partitioned analysis of 3,048 gene alignments. The newly sequenced bat species *Pteronotus mesoamericanus* and *Artibeus jamaicensis* are shown in bold. Branches show the number of positively selected genes detected by aBS-REL, corrected for within-gene multiple testing (p<0.05).

Selection of immune system genes in the bat ancestor

Immune system genes involved in cell signalling and inflammatory response were prominent among enriched groups. These genes include the key tissue damage sensor *NLRP3*, which was also found to be under selection in an earlier study using just two bats [15]. The general trend of rapidly evolving genes involved in inflammation is consistent with recent studies in bats [16,66]. Not all of the positive selection of immune-related genes identified in the bat ancestor are likely to be unique to bats. However, positive selection of inflammation-associated genes in bats is particularly interesting because a unique feature of the bat immune system may be increased tolerance to pathogens due to dampened inflammatory response [13].

Genes that play a role in antiviral defence and susceptibility were also found to be under selection. For instance, the interferon-induced *IFIT2* gene inhibits replication of viruses, including coronaviruses, flaviviruses, and poxviruses [67]. The positively selected phospholipase-encoding gene *PLA2G2D* has anti-inflammatory functions, and impacts susceptibility to SARS coronavirus and influenza virus infection in mice [68]. The host susceptibility gene *DPP4* was also positively selected. The *Artibeus jamaicensis DDP4* receptor, for instance, interacts with Middle East Respiratory Syndrome coronavirus [69], supporting its potential importance in this species. Like the currently widely discussed *ACE2* receptor, *DPP4* is also a target for coronaviruses and both genes are rapidly evolving in bats [69–71], possibly as a result of coevolution with viruses.

Selection of cancer-associated genes in the bat ancestor

Genetic modifications in *Myotis* and *Eptesicus* bats have previously been linked to cancer resistance and the longevity of these bats [20,72], though comparative genomic studies have generally not focused on cancer-resistance. We identified five positively selected cancer-associated genes (**Supplementary Table 9**), three of which (*PALB2* [73], *LATS2* [74], *SPDL1* [75]) are broadly involved in tumor suppression. We did not find parallel positive selection of genes associated with cancer, telomere maintenance or DNA repair in the bat ancestor and the naked mole rat. Although the identified positively selected genes are involved in different specific tumor suppressive mechanisms, they may play a role in the low rates of cancer observed in bats. Our findings also intersect with previous research that found positive selection of DNA repair genes, including *BRCA2*, which is recruited by *PALB2*, in the bat ancestor [15,39] and in individual bats [15,25]. Intriguingly, it was hypothesized that selection of these genes resulted from a need to reduce the negative effects of reactive oxygen species generated as a consequence of flight [15]. *PALB2* may therefore also have experienced positive selection for this reason. However, similarly to other genes involved in DNA repair, *PALB2* is also under positive selection in other mammals [76], suggesting it does not play an important role only in bats. Large-scale scans for positive selection across mammals may help resolve the different roles of tumor suppressor genes in cancer-susceptible and cancer-resistant lineages, and the evolutionary history of these genes.

Gene family expansion and contraction analysis

We identified one expanded gene family and 9 contracted gene families in the bat ancestor. The expanded gene family was associated with translation (**Supplementary Table 11**), consisting of ribosomal genes known to be expanded in bats [23]. Gene losses were driven by olfactory gene families, confirming earlier research finding losses of olfactory receptors in bats [20,25]. Losses of olfactory receptors may be linked to the emphasis of vision and echolocation in bat evolution. A recent study found a minor expansion of antiviral APOBEC3 genes in bats [16]. We found this family to be among the rapidly evolving gene families based on our CAFE analysis, but did not detect a significant expansion in the bat ancestor, likely due to our different outgroup and ingroup sampling.

We further investigated changes in gene families in the noctilionid bat superfamily, including *Pteronotus mesoamericanus* and the phyllostomids *Artibeus jamaicensis*, *Desmodus rotundus* and *Phyllostomus discolor*. Our analysis revealed an expansion of the interferon-induced transmembrane (IFITM) gene family (PTHR13999, p=2.18e-05) in the phyllostomids (**Figure 4**). IFITM genes are potent antiviral factors as they can prevent infection before a virus passes the lipid bilayer of the cell [77,78]. These genes belong to a group of host restriction factors, that include members of the APOBEC3 family, protecting cells from viral invasion. Of the five human IFITM proteins, three have antiviral activity (*IFITM1*, *IFITM2*, *and IFITM3*) while the other two (*IFITM5* and *IFITM10*) have no known role in immunity [79]. In humans, the IFITM locus consists of a ~22kb region on chromosome 11, comprising the three immune-related IFITM (IR-IFITM) genes and *IFITM5*. The *IFITM10* gene is ~1.4Mb distant from the other four IFITM genes. This *IFITM5*–IR-IFITM–*IFITM10* structure of the IFITM locus appears conserved across most mammals [80], with minor changes in the copy number of the IR-IFITM genes. Many species, including humans, also have additional copies of IR-IFITM genes in genomic locations distant from the IFITM locus. Here, we found that the IFITM locus in the three analysed phyllostomids has retained the structure of the human IFITM locus but with a substantial expansion of IR-IFITM genes (**Figure 4**), likely via tandem duplication. The phyllostomids that we investigated have 9-13 apparently functional IR-IFITM genes in regions of ~69kb to ~162kb, flanked by *IFITM5* and *IFITM10*.

Dynamic evolution of immune-related genes has played an important role in animal evolution [62,81–83], helping animals adapt to diverse pathogens. The expansion of IR-IFITM genes may reflect an evolutionary response to pathogen pressure in phyllostomid bats. Like the *IFIT2* gene, which we found to be positively selected in the bat ancestor, IR-IFITM genes belong to the group of interferon-stimulated genes that form a first line of defence against infections, particularly from viruses [84,85]. Experiments with cultured cells have shown the activity of IR-IFITMs against numerous viruses associated with bats, including dengue virus, ebola virus and SARS coronavirus [79]. Four recently sequenced additional phyllostomid species [23] do not show the same expansion, suggesting losses or independent gains within this group. Although the incompleteness of the protein annotations in these species (**Supplementary Table 12**) may have hampered the detection of further IR-IFITM genes, this suggests that losses or independent gains of IR-IFITM genes may have occurred within phyllostomids, possibly linked to different pathogen pressures. The IFITM locus may also be challenging to assemble using short reads, as suggested

by its incomplete assembly in the bat *Myotis brandtii*, underlining the value of long read sequencing to capture potentially adaptive structural variation.

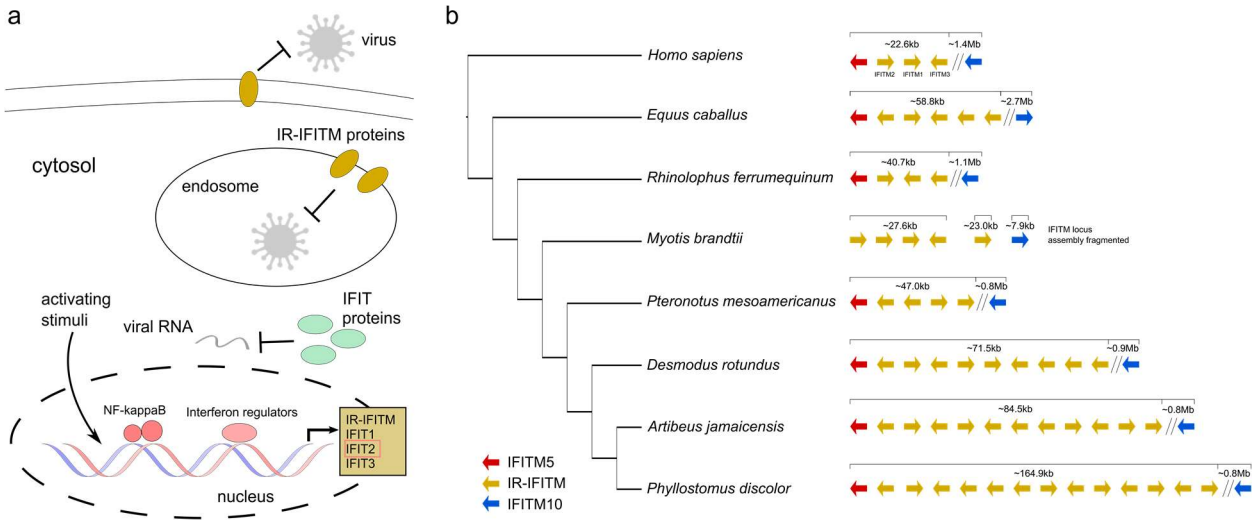


Figure 4. Gene family expansion associated with the bat antiviral immune response. a) Model of cellular processes involving IFITM and IFIT antiviral interferon-induced gene families under positive selection (highlighted with red box) or experiencing expansion. The model is based on a review of IFIT and IFITM proteins [86] and studies on IFITM interactions with RNA viruses [77,87–89]. b) Phylogeny of sampled noctilionid bats and selected outgroups, showing gene copy number at the IFITM locus. Immune-related IFITM (IR-IFITM) genes are shown in yellow.

We also detected substantial expansions and losses of immune-related gene families in individual noctilionoid species. *Pteronotus mesoamericanus* showed significant losses in the sialic acid immunoglobulin-lectin (Siglec) family (PTHR12035, p=7.02e-05), with the lowest number of these genes among the bats analysed here. Underlining the rapid evolution of this gene family found in earlier studies of mammals [62], we also detected positive selection of *SIGLEC1* in the bat ancestor. Siglecs are immune-modulatory receptors and their rapid evolution is likely driven by 'Red Queen' dynamics between hosts and pathogens [90]. In *Phyllostomus discolor*, we observed a major expansion of killer lectin-like receptors (PTHR22800, p=5.51e-05). This gene family, which plays a role in T cell activation during viral infection, also shows a substantial expansion in the phylogenetically distant *Rousettus aegyptiacus* [13]. Furthermore, it has been suggested that natural killer receptors in bats increase tolerance to viruses by less strongly inducing cytokines and therefore lowering inflammatory responses [13]. Overall our findings suggest that gene duplication and loss have played an important role in the evolution of the bat immune system.

Evolution of diet-related genes was previously shown to be particularly important in phyllostomid bats [23], which are adapted to a wide variety of dietary niches in the Neotropics. Here we find key differences between the fruit-eating *Artibeus jamaicensis* and the insect-eating *Pteronotus mesoamericanus* in losses of two well-characterised genes related to insect consumption (trehalase and chitinase). Trehalase encodes for an enzyme crucial for digestion of trehalose, the primary sugar in insect blood [91]. Trehalase can be duplicated in insects [92] but was found to be a

single-copy gene in bats and other mammals [22]. We found evidence for independent deletions of the trehalase gene linked to dietary specialization in noctilionid bat genomes (**Figure 5**). Additionally, we observed a likely ancestrally duplicated trehalase pseudogene that was present in all four analysed genomes. Compared to the functional trehalase gene with fifteen exons, only nine exons can be detected in the pseudogene (exons 5-8 and 11-15). Furthermore the pseudogene appears to have an internal inversion impacting exons 5,6,7 and 8. Our synteny analysis of the trehalase locus suggests that that the ancestrally functional trehalase gene was deleted in *Artibeus jamaicensis* and *Desmodus rotundus*, and only the ancestral pseudogene with the inversion remains in their respective genomes (**Figure 5**). The insect-eating *Pteronotus mesoamericanus* and the omnivorous *Phyllostomus discolor* have each maintained a functional trehalase gene as well as the neighboring pseudogene. The majority of phyllostomids sequenced to date do not have a functional trehalase gene [22,23], suggesting multiple independent loss events. Such losses of the trehalase gene are linked to dietary shifts across mammals [22]. These gene losses can occur via small deactivating mutations or larger structural variations such as deletions. Our findings support the idea that structural variation can underpin trait evolution.

A similarly important gene for insect-eating is chitinase, of which there are five ancestral copies in placental mammals [21]. Chitinase is expressed in the gastrointestinal tract and encodes for an enzyme capable of digesting chitin, a major component of insect exoskeletons [93]. In the noctilionid bats investigated here, we confirm earlier findings that chitinase copy number is linked to diet. The insect-eating *Pteronotus mesoamericanus* shows the highest copy number (5), while the omnivore *Phyllostomus discolor* has four copies, and the fruit-eating *Artibeus jamaicensis* as well as the blood-feeding *D*esmodus *rotundus* each have only two copies. The chitinase genes annotated in *Artibeus jamaicensis* are not supported by aligned RNA-seq data and may be pseudogenes. A sixth chitinase copy in *Pteronotus mesoamericanus* shows fewer annotated exons, suggesting that this may also be a pseudogene. Pseudogenization of additional chitinase copies is common, though these copies can also be neofunctionalised to gain immune-related functions [21].

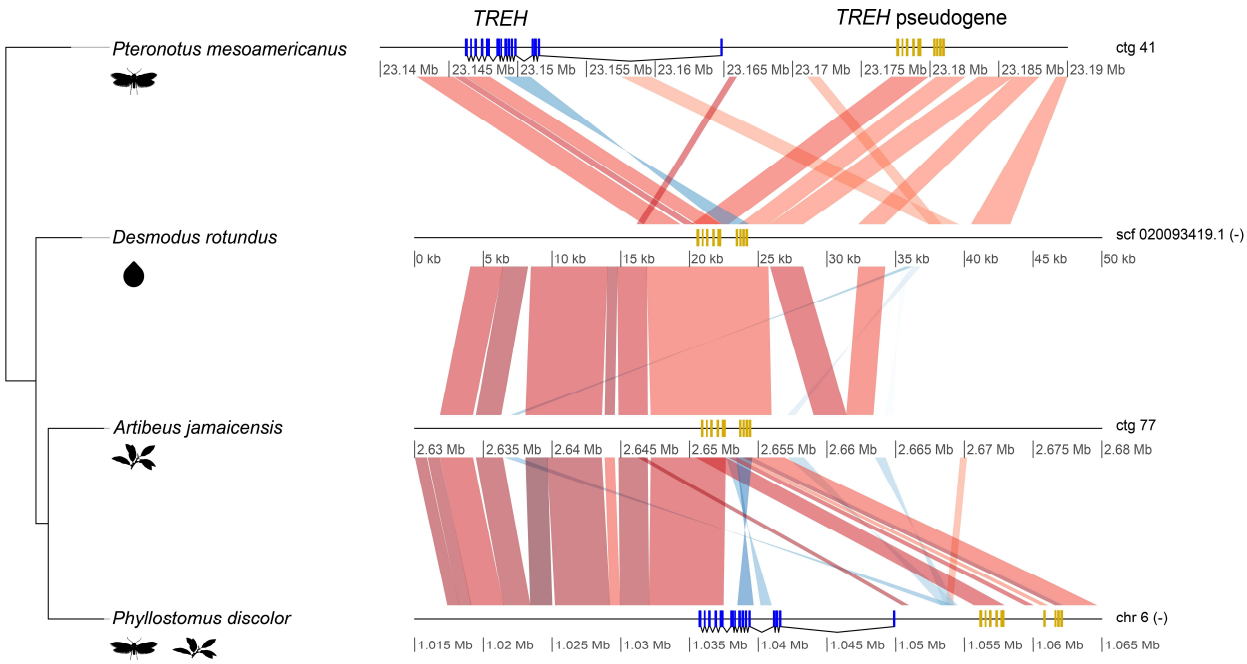


Figure 5. Synteny plot of the trehalase locus for four noctilionid bats with different dietary specializations. Diets are indicated by symbols showing, from top to bottom, insect-eating, blood-feeding, fruit-eating and omnivory. The synteny plot shows inverted alignments in blue.

Conclusions

We generated high-quality reference genomes for the Neotropical bats *Pteronotus mesoamericanus* and *Artibeus jamaicensis* and used them to investigate patterns of genomic evolution in Neotropical bats and their relatives. Using these reference genomes, and the genomes of eight additional diverse bat species, we detected genetic changes associated with the bat immune system. Specifically, we found adaptations in interferon-induced genes, including tandem duplications of antiviral IFITM genes in phyllostomid bats and positive selection of the *IFIT2* gene in the bat ancestor. These genetic changes may be associated with the high tolerance of bats to viral infection. In line with these immune-related adaptations, we also identified evidence of the occurrence of EVEs known from Europe and Asia in Neotropical bats, expanding the known host and geographic range of these viruses. Furthermore, we found that gene losses via deletion have played a role in dietary specialization in phyllostomid bats.

We also found tumor suppressors undergoing rapid evolution in the bat lineage. Further investigation of the evolution of cancer-associated genes in bats may reveal whether genetic changes in tumor suppressors contributed to the cancer resistance of these long-lived mammals. An intriguing possibility is that the rapid evolution of the bat immune system has also played a role in reducing cancer rates because immune-related genes can play a role in cancer surveillance [94] and tumor suppression [95–97]. This possibility could be investigated further if bats are developed as a system to study mammalian health.

The genomes of *Pteronotus mesoamericanus* and *Artibeus jamaicensis* generated in this study represent a substantial addition to the growing genomic resources available for bats [26]. We demonstrate that ONT long read sequencing can produce highly contiguous mammalian assemblies with larger contig sizes than PacBio sequencing approaches. Work is underway to further improve the base-level accuracy of these assemblies. The genomes and annotations of *Pteronotus mesoamericanus* and *Artibeus jamaicensis* can be mined and integrated into comparative genomic studies to identify protein-coding changes, structural variation and gene regulatory mechanisms that explain exceptional bat adaptations including tolerance of viruses and cancer resistance. The data generated here will help us explain bat evolution and gain insights to improve human health.

Acknowledgements

We thank the Cancer Center at Cold Spring Harbor Laboratory (Grant 5P30CA045508) and Director David Tuveson for support. We are also grateful for the support of the Simons Center for Quantitative Biology at Cold Spring Harbor Laboratory. We would further like to acknowledge funding support from the Cold Spring Harbor Laboratory and Northwell Health Affiliation for purchase of the ONT PromethION sequencer used in this study. Finally, we thank Brock Fenton, Neil Duncan, the staff at Lamanai, and many other colleagues who assisted with the fieldwork necessary for this study.

Materials and methods

Sample acquisition

Fresh liver samples from a single individual of *Artibeus jamaicensis* (AMNH 279493, male) and one *Pteronotus mesoamericanus* (AMNH 279536) were collected by NBS in April 2017 at Lamanai Archaeological Reserve in Orange Walk District, Belize (17.75117°N, 88.65446°W). Sampling followed best practices for humane capture, handling, and euthanasia of live mammals outlined by the American Society of Mammalogists [98]. All work was conducted with permission of the Belize Forest Department under permit number WL/2/1/17(19) with IACUC approval from the American Museum of Natural History (AMNHIACUC-20170403) and University of Georgia (A2014 04-016-Y3-A5). Bats were captured in ground-level mist nets and placed in individual cloth bags for transport to the Lamanai Field Research Center. After identification, the bats were euthanized using isoflurane, and the liver was removed immediately after death. Samples were placed in multiple individual 2ml cryotubes and flash-frozen by placement in a liquid nitrogen dry shipper. The cold chain was maintained through shipment to the AMNH, storage in the AMNH Ambrose Monell Cryo Collection, and subsequent sample processing and transfers.

Genome sequencing and assembly

Approximately 40 mg of liver tissue from each bat was received at CSHL and stored at -80C. The liver samples were crushed with a micro pestle and mixed with 10 ml of TLB buffer and 50µl of RNase A immediately before use. After a one hour incubation at 37C, 50 µll of Proteinase K was added, followed by incubation at 50C for 3 hours with hourly inversion mixing. After addition of 10ml of a phenol chloroform / isoamyl alcohol mixture, the sample was rocked for 10 minutes and then centrifuged at 4500RPM for 10 minutes. The top aqueous layer was retained and an equal volume of chloroform / isoamyl alcohol was added, and rocking and centrifugation was repeated as above. The top layer was transferred to a fresh tube with 4ml of 5M ammonium acetate. Following the addition of 30ml of ice cold 100% ethanol, the sample was rocked for 10 minutes. The visible DNA was then hooked out with a glass pipet and placed in a 1.5ml tube. The sample was washed once with 100% ethanol and centrifuged for 5 minutes at 10000RPM. The ethanol was removed and any remaining ethanol was evaporated on a 37C heat block for 10 minutes. The final DNA was resuspended in 10mM Tris HCl pH 8.5 and stored overnight at 4C. DNA was then sheared to ~50-75 using a Diagnode Megarupter following manufacturer's recommendations. DNA was further enriched for long fragments via the Circulomics small read eliminator XL kit, which iteratively degrades short fragments. DNA was prepared for Nanopore sequencing using the ONT 1D sequencing by ligation kit (SQK-LSK109). Briefly, 1-1.5ug of fragmented DNA was repaired with the NEB FFPE repair kit, followed by end repair and A-tailing with the NEB Ultra II end-prep kit. After an Ampure clean up step, prepared fragments were ligated to ONT specific adapters via the NEB blunt/TA master mix kit.

The library underwent a final clean up and was loaded onto a PromethION PRO0002 flowcell per manufacturer's instructions. The flowcells were sequenced with standard parameters for 3 days. Basecalling was performed in real time with Guppy 3.2. Illumina short read libraries were prepared with the Illumina TruSeq DNA kit, targeting a 550bp insert size with PCR enrichment. Libraries were sequenced at the New York Genome Center, on a NovaSeq S4 flowcell in paired end 150bp format to ~30x genome coverage. Nanopore reads were filtered for minimum length of 10 and minimum 85% accuracy using filtLong (http//github.com/rrwick/Filtlong). Reads were assembled using wtbg2 [33]. A consensus was generated with wtpoa-cns, and then polished using both the long Nanopore reads and the Illumina short reads. First, the long Nanopore reads were aligned to the assembled contigs using minimap2 [99]. The corrected consensus was then derived with wtpoa-cns. Secondly, bwa [100] was used to align the Illumina short read data to the long read polished assembly, and the short read consensus was generated with wtpoa-cns. Assemblies were then assessed using Assemblytics [101] to compare assembly contiguity by contig alignment to published bat reference genomes, as well as with Benchmarking Sets of Universal Single-Copy Orthologs 4.0.5 (BUSCO) [102] for mammals (odb10) to assess completeness of conserved orthologs in bats. Additionally, we used python scripts to compute the cumulative sum of contig length (N(X) length) versus the cumulative sum of N(X)% of the total genome. Also the python module squarify was used to compute tree maps of the genome assembly, in which each area denotes the size of each contig and its contribution to the whole assembly. Further assembly statistics were calculated using BBTools 38.86 (http//sourceforge.net/projects/bbmap/).

Gene annotation

Public RNA-seq data from SRA were downloaded for *Artibeus jamaicensis* and *Pteronotus parnelli* (**Supplementary Table 2**) and proteins for human (GCF_000001405.39), mouse (GCF_000001635.26) and seven bat species (*Myotis brandtii*, GCF_000412655.1; *M. davidii*, GCF_000327345.1; *M. lucifugus*, GCF_000147115.1;*Phyllostomus discolor*, GCF_004126475.1; *Pteropus alecto*, GCF_000325575.1; *Rhinolophus ferumequinum*, GCF_004115265.1; *Rousettus aegyptiacus*, GCF_001466805.2) were downloaded from RefSeq. RNA-seq reads were aligned to the new reference genomes using HISAT 2.2.0[103] with the parameters "--no-mixed --no-discordant --downstream-transcriptome-assembly" and transcripts were assembled using StringTie 2.1.1 [104]. To reduce potential loss of transcripts due to limitations of short-read alignment, transcriptomes were also de novo assembled with Trinity 2.9.1. PASA 2.4.1 [105] was used to generate the final set of transcripts based on alignment with GMAP [106] and BLAT [107] using minimum thresholds of 90% of transcript length aligned at 90% identity. We used the transcripts and proteins sequences as evidence for the MAKER3 annotation pipeline [1] with the ab initio gene predictors SNAP 2006.07.28 [108] and Augustus 3.3.3 [109]. The completeness of the protein set was assessed using BUSCO. To detect fragmented genes, we used BLASTP one-to-one alignments to detect directly neighboring genes that partially aligned to non-overlapping regions of the same single gene in a different species. With a bespoke script fix_split_genes.py, we merged partial gene models to more complete models. Using this method, we improved the status of 250 BUSCO groups in *Artibeus jamaicensis* and 142 in *Pteronotus mesoamericanus*. Eggnog 1.0.3 as well as Interproscan 5.45-80 with the PANTHER 14.1 database were used to functionally annotate the final protein sets.

Repeat analysis

Repeat masking was carried out using an iterative masking and a de novo repeat detection approach. To allow comparison of the repeat landscape within Noctilionoidea, the same repeat masking was carried out for *Desmodus rotundus* and *Phyllostomus discolor*. After masking repeats with Repeatmasker 4.0.9 [110] and the combined RepBase-20181026 and Dfam-3.1 repeat databases, novel repeats were detected in the masked genomes with RepeatModeler 2.0.1. The consensus de novo repeats longer than 100 bp were then concatenated with a vertebrate repeat library including novel bat repeats [16] and clustered with CD-HIT 4.81 [111]. All clustered novel sequences with >80% sequence similarity across >80% of the length of the clustered sequences were excluded [112]. Novel repeats were then aligned to the nt database using BLAST and all repeats matching annotated transcripts were removed. A further BLAST analysis was used to exclude repeats with fewer than 10 alignments to the reference genome from which they were derived. Transposable elements were classified and assigned families using TEclass [113] and DeepTE [114]. The consensus de novo repeats were then concatenated with the vertebrate repeat library and a final masking of the genome was carried out with Repeatmasker using the sensitive setting. The recently diverged repeat landscape was analysed using the RepeatMasker script calcDivergenceFromAlign.pl with correction of substitution rates based on the Kimura 2-Parameter model. Transposons were considered recently diverged at 6.6% divergence from the consensus[16], which approximates an insertion less than 30 mya, assuming a mammalian substitution rate of 2.2×10^{-9} [115].

Mining of endogenous viral elements

Bats are likely important reservoir hosts for viruses, possibly due to their highly effective immune response. We therefore investigated our genomes for endogenous viral elements that provide evidence of past infections. We aligned the complete RVDB-prot viral protein database [2] to our genomes using BLAST with an e-value threshold of 1e-5. Proteins with non-standard amino acid codes were excluded. To confirm these initial results, overlapping matches were combined to perform a query against the RefSeq non-redundant protein database using PLAST [3] with an e-value threshold of 1e-5. TaxonKit [4] 0.6.0 was used to filter matches to non-viral sequences.

Identification of orthologous gene groups

We analysed bat genes for lineage-specific signals of positive selection and gene duplications based on clustered ortholog groups. Coding sequences for the seven bats used for annotation and *Desmodus rotundus* as well as six outgroup mammals (human, mouse, naked mole rat, dog, pig, and horse) were downloaded from RefSeq. Sequences that did not show a nucleotide number that was a multiple of three or that contained internal stop codons were discarded. The longest isoform for each gene was retained and translated to an amino acid sequence using biopython. All proteins were clustered with the proteins of our bats using OrthoFinder [116]. Genes were then aligned using PRANK v.170427 [117] with the species tree provided to guide alignment. Species-specific indels in coding sequences are rare and the ONT assemblies were found to have some indel errors. We therefore masked a single codon to either side of these indels in *Artibeus jamaicensis* and *Pteronotus mesoamericanus*. When indels occurred within 10 bases of exon junctions, two adjacent codons were masked. A set of single copy orthologs was extracted from the OrthoFinder results, allowing up to 4 species to be missing and requiring *Artibeus jamaicensis* or *Pteronotus mesoamericanus* to be represented. Cancer-associated ortholog clusters were identified based on the COSMIC catalogue [5]. Furthermore, we used gene ontology biological processes to assign genes to the functional categories immunity, DNA repair, telomere maintenance, aging, reproduction, thermoregulation, hearing and digestion.

Positive selection analysis

We aimed to determine whether each gene was positively selected in one or more of seven groups (the bat ancestor, Yangochiroptera, Noctilionoidea, Phyllostomidae, naked mole rat, *P. mesoamericanus* and *A. jamaicensis*). To detect selection in each gene, we used PAML 4.9 [6] to apply Test 2 [7], a branch-site test of positive selection. This test compares the alternative model where some branches are under positive selection and thus exhibit sites with omega>1 with the corresponding null model where omega is fixed as 1. Comparing the likelihood of these models can determine whether a lineage-specific group of codons in the alignment is experiencing significant positive selection. We computed p-values according to a $\chi 2$ distribution with one degree of freedom [8]. In addition, we also detected positive selection using the adaptive branch-site random effects model (aBS-REL) method implemented in HyPhy [9,10]. Unlike the branch-site test implemented in PAML, aBS-REL allows rate variation among the background branches, which can reduce false positive errors [10]. The p-values calculated by PAML and HyPhy were adjusted for multiple testing of seven branches per gene using the false discovery rate (FDR) method

implemented in base R. We did not correct for multiple testing of the 10,928 genes, allowing us to also consider genes experiencing low to moderate selection [81]. We used the maximum likelihood species tree inferred using 3,048 orthologs (**see Methods section 'Gene family expansion analysis'**) for the positive selection scans. In this tree, the Chiroptera and Perissodactyla are sister to the Carnivora and Cetartiodactyla. However, the relationships between laurasiatherian mammals are difficult to resolve and remain controversial [118–120]. To help rule out an impact of a misspecified phylogeny on the analysis of positive selection, we therefore carried out additional HyPhy analyses using a relatively widely proposed alternate phylogeny with bats as the sister group to a clade including Cetartiodactyla, Perissodactyla, Carnivora [16]. Genes with FDR-corrected $p<0.05$ using both trees were retained.

Gene family expansion analysis

Alignments of 3,048 single copy genes that were present in all species were concatenated into a single alignment, which was divided into three partitions corresponding to codon positions. A maximum likelihood phylogeny was inferred using RAxML 8.2.12 [121] rapid bootstrapping under the GTR+G+I model with 100 bootstraps. For the following gene family expansion analysis we used the topology derived from our maximum likelihood phylogeny but with the sister taxon of bats determined as *(Canis familiaris,(Equus caballus, Sus scrofa))*. This topology follows recent phylogenetic work to resolve the controversial relationship between bats and other laurasiatherian mammals [16], and was therefore used for the following gene family expansion analysis. A dated phylogeny was then generated using MCMCtree in PAML. Node ages were calibrated based on timetree [11] ages for Euarchontoglires, bats, Yangochiroptera and Yinpterochiroptera. Convergence was assessed based on analysis of two replicate runs with tracer [12]. The dated phylogeny was used to calculate gene family expansions and contractions with Cafe 4.21 [13] based on a p-value threshold of 0.05. Orthofinder orthogroups were collapsed into 7,381 PANTHER gene families based on BiomaRt [122] data. For all 19,274 orthogroups, a representative gene was selected to obtain a PANTHER assignment. For 2,154 orthogroups, where no human or mouse gene was represented, we selected a gene for sequence-based annotation using interproscan and eggnog. As Cafe assumes all genes were present in the common ancestor of all analysed species, orthogroups represented in less than two bats and two outgroup mammals were excluded.

The IFITM gene family was analysed for pseudogenes in *A. jamaicensis* and *P. mesoamericanus* using PseudoPipe [123] with the human IFITM1, IFITM2 and IFITM3 genes as query proteins. Genes lacking introns or showing premature stop codons were considered pseudogenes. The presence of the CD225 domain in all IFITM proteins was confirmed using pfam_scan.pl 1.6 [124] with the Pfam 33.1 database release. A synteny plot for the trehalase gene was generated using BLAST alignment with default parameters of contigs containing the trehalase locus and genoPlotR [125]. Alignments under 500 base pairs or with bitscores lower than 200 were excluded. The *Phyllostomus discolor* trehalase protein sequence was aligned using BLAST with "-word_size 8" to the reference genomes of all noctilionid bats to detect trehalase genes and pseudogenes and provide an estimation of exon boundaries.

Enrichment analysis

GO, Reactome and KEGG annotations for genes were obtained via BiomaRt. A total of 18,452 orthogroups were annotated with at least one feature and 17,840 were assigned GO features. We carried out enrichment analysis on groups of genes that were positively selected or that showed gene family expansions. Enrichment analysis was performed using topGO, with the elim algorithm and the Fisher test ($p<0.01$). All genes tested for selection were used as the background. The elim algorithm is a conservative approach that processes the GO graph from the bottom up, to eliminate higher order terms that would otherwise appear enriched due to correlation with more specific terms. GO terms with fewer than 10 genes annotated were excluded.

Supplementary materials

Supplementary tables

Supplementary Table 1. Publicly available bat genome assemblies included in GenBank. When multiple assemblies for a single species where available, the assembly with the highest quality assembly based on annotation, contig N50 and size is shown.

Species name	Accession	Annotation	Assembly level	Contig N50 (kbp)	Size (Gbp)	Year published
Myotis lucifugus	GCF_000147115.1	yes	Scaffold	64	2.035	2010
Eptesicus fuscus	GCF_000308155.1	yes	Scaffold	21	2.027	2012
Myotis davidii	GCF_000327345.1	yes	Scaffold	15	2.060	2012
Eidolon helvum	GCA_000465285.1	no	Scaffold	13	1.838	2013
Myotis brandtii	GCF_000412655.1	yes	Scaffold	23	2.107	2013
Pteronotus parnellii	GCA_000465405.1	no	Scaffold	10	1.960	2013
Pteropus alecto	GCF_000325575.1	yes	Scaffold	32	1.986	2013
Pteropus vampyrus	GCF_000151845.1	yes	Scaffold	22	2.198	2014
Rhinolophus sinicus (contaminated assembly)	GCA_001888835.1	no	Scaffold	38	2.073	2016
Hipposideros armiger	GCF_001890085.1	yes	Scaffold	40	2.237	2016
Miniopterus natalensis	GCA_004027415.1	yes	Scaffold	30	1.803	2016
Desmodus rotundus	GCF_002940915.1	yes	Scaffold	80	2.064	2018

Eonycteris spelaea	GCA_003508835.1	no	Contig	8,003	1.967	2018
Anoura caudifer	GCA_004027475.1	no	Scaffold	143	2.207	2019
Antrozous pallidus	GCA_007922775.1	no	Scaffold	13	2.597	2019
Artibeus jamaicensis	GCA_004027435.1	no	Scaffold	32	2.425	2019
Carollia perspicillata	GCA_004027735.1	no	Scaffold	10	2.689	2019
Craseonycteris thonglongyai	GCA_004027555.1	no	Scaffold	23	2.272	2019
Cynopterus brachyotis	GCA_009793145.1	no	Scaffold	17	1.759	2019
Hipposideros galeritus	GCA_004027415.1	no	Scaffold	34	2.441	2019
Lasiurus borealis	GCA_004026805.1	no	Scaffold	35	2.858	2019
Macroglossus sobrinus	GCA_004027375.1	no	Scaffold	338	1.898	2019
Macrotus californicus	GCA_007922815.1	no	Scaffold	16	2.161	2019
Megaderma lyra	GCA_004026885.1	no	Scaffold	72	2.621	2019
Micronycteris hirsuta	GCA_004026765.1	no	Scaffold	63	2.315	2019
Miniopterus schreibersii	GCA_004026525.1	no	Scaffold	83	1.776	2019
Mormoops blainvillei	GCA_004026545.1	no	Scaffold	143	2.112	2019
Murina aurata feae	GCA_004026665.1	no	Scaffold	23	2.332	2019
Noctilio leporinus	GCA_004026585.1	no	Scaffold	136	2.099	2019
Nycticeius humeralis	GCA_007922795.1	no	Scaffold	14	2.780	2019
Phyllostomus discolor	GCF_004126475.1	yes	Chromosome	6,893	2.118	2019
Pipistrellus pipistrellus	GCA_004026625.1	no	Scaffold	27	2.091	2019
Rhinolophus ferrumequinum	GCF_004115265.1	yes	Chromosome	31,900	2.076	2019
Tadarida brasiliensis	GCA_004025005.1	no	Scaffold	23	2.710	2019
Tonatia saurophila	GCA_004024845.1	no	Scaffold	142	2.106	2019
Aeorestes cinereus	GCA_011751065.1	no	Scaffold	121	2.149	2020
Molossus molossus	GCA_014108415.1	yes	Scaffold	22,175	2.316	2020
Myotis myotis	GCA_014108235.1	yes	Scaffold	12,511	2.003	2020
Pipistrellus kuhlii	GCA_014108245.1	yes	Scaffold	10,592	1.776	2020
Pteropus giganteus	GCA_902729225.1	no	Scaffold	256	1.985	2020
Rousettus aegyptiacus	GCA_014176215.1	yes	Scaffold	21,999	1.894	2020

Supplementary Table 2. Public RNA sequencing data used for gene annotation of *Artibeus jamaicensis* and *Pteronotus mesoamericanus*. Sequencing reads from the close relative *P. parnelli* were aligned to the *P. mesoamericanus* genome.

Species	SRA ID	Tissue	Mapped	Mapped percent	Properly paired	Properly paired percent
A. jamaicensis	SRR10059484	vomeronasal	68,871,802	85.05	56,566,330	76.71
A. jamaicensis	SRR2982097	spleen	44,832,500	92.4	41,874,452	89.11
A. jamaicensis	SRR2982098	spleen	44,953,626	90.91	41,533,668	86.99
A. jamaicensis	SRR2982101	kidney	84,076,271	92.31	78,804,850	89.74
A. jamaicensis	SRR2982102	kidney	57,901,531	91.02	54,321,912	88.11
A. jamaicensis	SRR2982105	liver	45,417,522	87.27	42,163,408	83.64
A. jamaicensis	SRR2982106	liver	49,635,258	87.77	46,038,798	84.17
A. jamaicensis	SRR539297	spleen	117,742,040	90.35	106,807,924	85.37
A. jamaicensis	SRR7910151	liver	54,375,933	82.02	46,859,458	72.65
A. jamaicensis	SRR7910153	liver	54,618,199	86.48	50,170,036	81.33
A. jamaicensis	SRR7910154	liver	31,860,421	78.71	26,934,452	68.09
P. parnelli	SRR7522707	Cricothyroid muscle	38,414,181	74.21	26,395,946	58.32
P. parnelli	SRR7522708	Cricothyroid muscle	45,422,544	76.28	30,639,926	59.72
P. parnelli	SRR7522709	Cricothyroid muscle	44,760,826	78.16	30,972,832	62.69
P. parnelli	SRR7522710	Breast muscle	34,752,586	72.8	22,915,242	55.85
P. parnelli	SRR7522711	Breast muscle	42,782,814	73.67	31,122,386	59.04
P. parnelli	SRR7522712	Breast muscle	69,921,019	73.76	46,520,120	57.29

Supplementary Table 3. Gene annotation statistics. Annotation edit distance (AED) is calculated by the MAKER3 annotation pipeline.

	Pteronotus mesoamericanus	*Artibeus jamaicensis*
Total proteins	23,145	27,600
Proteins with PFAM domain	17,065 (74%)	16,851 (59%)
Proteins with PANTHER annotation	20,069 (87%)	20,000 (72%)
Average AED quality score (0-1, where 0 is best)	0.26	0.30
Median AED quality score	0.20	0.25
Average protein length	452.66	389.98
Median protein length	318	251
Detected adjacent split genes (gene number after merging)	998 (599)	1,244 (474)

Supplementary Table 4. Recent accumulations of transposable elements identified by RepeatMasker in six noctilionid reference genomes. Base pairs counted are from non-overlapping elements with <6.6% divergence from the consensus sequence as determined by Kimura 2-Parameter corrected rate calculated by calcDivergenceFromAlign.pl. doi:10.6084/m9.figshare.12934652

Supplementary Table 5. Number of confident (<1e-5) unique blastx alignments of putatative endogenous viral elements to viral families in four noctiolionid bat genomes. doi:10.6084/m9.figshare.12934652

Supplementary Table 6. Number of confident (<1e-5) unique blastx alignments of putatative endogenous viral elements to viral genera in four noctiolionid bat genomes. doi:10.6084/m9.figshare.12934652

Supplementary Table 7. BLASTX alignment results (<1e-5) for putative non-retroviral endogenous viral elements of four noctilionid bats. doi:10.6084/m9.figshare.12934652

Supplementary Table 8. BLASTX alignment results (<1e-5) for putative retroviral endogenous viral elements of four noctilionid bats. doi:10.6084/m9.figshare.12934652

Supplementary Table 9. Positively selected genes in the bat ancestor. Raw and within-gene FDR-corrected p-values are provided for the HyPhy aBS-REL test as well as the likelihood ratio test for PAML Test 2 and the null hypothesis Test 2 (fixed). Two phylogenetic trees are used, with alternative relationships between laurasiatherian mammals: Tree 1 (Chiroptera + Perissodactyla are sister to the Carnivora + Cetartiodactyla) and Tree 2 (bats as the sister group to a the clade (Cetartiodactyla+Perissodactyla) + Carnivora). Functional groups are defined based on GO annotations for the categories immunity, DNA repair, telomere maintenance, aging, reproduction, thermoregulation, hearing and digestion; genes not belonging to these categories are classified as NA. doi:10.6084/m9.figshare.12934652

Supplementary Table 10. TopGO gene ontology enrichment analysis of positively selected genes in the bat ancestor. The 'elim' algorithm was used and GO enrichments with p<0.01 and >=10 annotated genes were retained.

GO ID	Term	Annotated genes	Significant genes	Expected genes	p-value
GO:0045087	innate immune response	782	25	9.42	0.00072
GO:0007252	I-kappaB phosphorylation	18	3	0.22	0.00123
GO:0050777	negative regulation of immune response	138	7	1.66	0.00142
GO:0002374	cytokine secretion involved in immune response	19	3	0.23	0.00145
GO:0033344	cholesterol efflux	42	4	0.51	0.00160
GO:0071677	positive regulation of mononuclear cell migration	20	3	0.24	0.00169
GO:0016042	lipid catabolic process	322	11	3.88	0.00185
GO:0046718	viral entry into host cell	107	6	1.29	0.00187
GO:0030097	hemopoiesis	855	20	10.3	0.00348
GO:0030195	negative regulation of blood coagulation	52	4	0.63	0.00352
GO:0002443	leukocyte mediated immunity	683	17	8.23	0.00377
GO:0001960	negative regulation of cytokine-mediated signaling pathway	55	4	0.66	0.00432
GO:0045332	phospholipid translocation	28	3	0.34	0.00452
GO:0007276	gamete generation	689	20	8.3	0.00478
GO:0050728	negative regulation of inflammatory response	135	6	1.63	0.00590
GO:0034372	very-low-density lipoprotein particle remodeling	10	2	0.12	0.00610
GO:0034378	chylomicron assembly	10	2	0.12	0.00610
GO:0060068	vagina development	10	2	0.12	0.00610
GO:0032621	interleukin-18 production	10	2	0.12	0.00610
GO:1900102	negative regulation of endoplasmic reticulum unfolded protein response	10	2	0.12	0.00610
GO:1903431	positive regulation of cell maturation	10	2	0.12	0.00610

GO:0006721	terpenoid metabolic process	97	5	1.17	0.00636
GO:0002701	negative regulation of production of molecular mediator of immune response	32	3	0.39	0.00661
GO:0048278	vesicle docking	62	4	0.75	0.00661
GO:0055081	anion homeostasis	62	4	0.75	0.00661
GO:0007178	transmembrane receptor protein serine/threonine kinase signaling pathway	333	10	4.01	0.00724
GO:0019532	oxalate transport	11	2	0.13	0.00739
GO:0050713	negative regulation of interleukin-1 beta production	11	2	0.13	0.00739
GO:0006956	complement activation	65	4	0.78	0.00781
GO:0060338	regulation of type I interferon-mediated signaling pathway	34	3	0.41	0.00783
GO:0006954	inflammatory response	709	20	8.54	0.00810
GO:0006897	endocytosis	679	16	8.18	0.00814
GO:0002274	myeloid leukocyte activation	565	14	6.81	0.00862
GO:0034380	high-density lipoprotein particle assembly	12	2	0.14	0.00880
GO:0035641	locomotory exploration behavior	12	2	0.14	0.00880
GO:0051775	response to redox state	12	2	0.14	0.00880
GO:0006887	exocytosis	807	18	9.72	0.00883
GO:0007286	spermatid development	150	6	1.81	0.00969
GO:0002755	MyD88-dependent toll-like receptor signaling pathway	37	3	0.45	0.00990

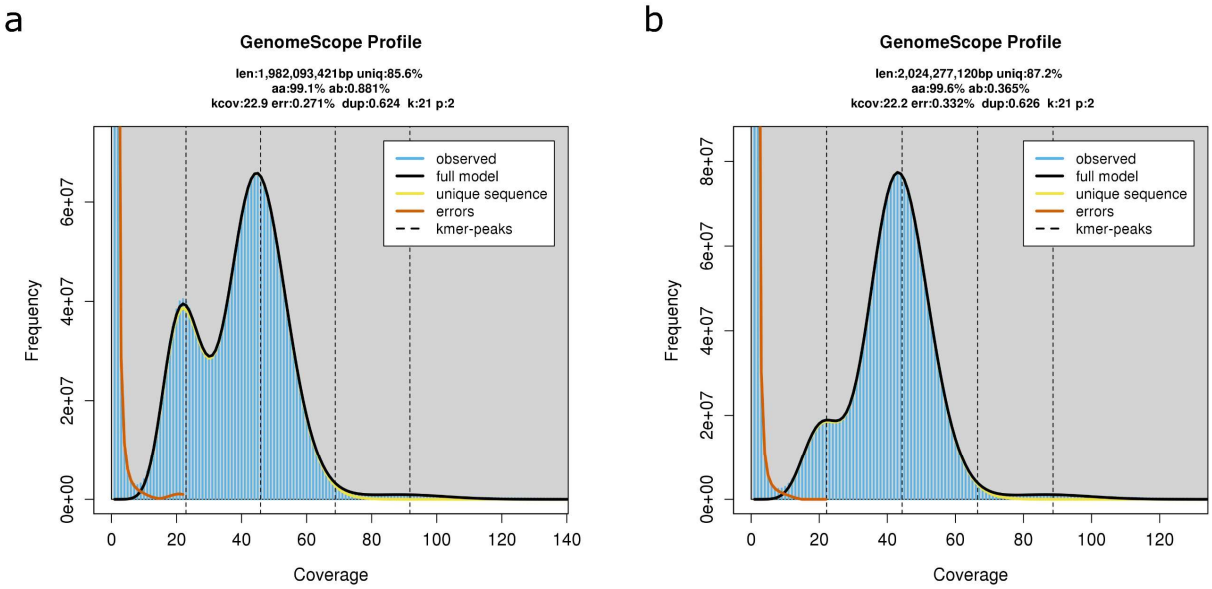
Supplementary Table 11. Gene family expansion and contraction analysis in the bat ancestor.

PANTHER family	Gene gain or loss	PANTHER family description	Viterbi p-value
PTHR45696	+2	60S acidic ribosomal protein P1	0.00105861
PTHR24300	-4	Cytochrome P450 508A4-related	0.000417925
PTHR21859	-3	Acrosome-specific protein, spermatogenesis-associated protein 31C1-related	4.62E-05
PTHR19368	-2	XLR/SCP3/FAM9, protein FAM9C	0.000709781
PTHR24118	-2	Pote ankyrin domain family member B-related	0.00921962
PTHR23262	-6	Keratin-associated protein 9-2-related	7.80E-07
PTHR24062	-4	Vomeronasal type-1 receptor 5-related	7.98E-06
PTHR43447	-3	Alpha-amylase 1-related	2.30E-05
PTHR43313	-2	Short-chain dehydrogenase/reductase family 9C,17-beta-hydroxysteroid dehydrogenase type 6	0.00750727
PTHR24242	-10	G-protein coupled receptor, olfactory receptor 11H1-related	3.92E-11

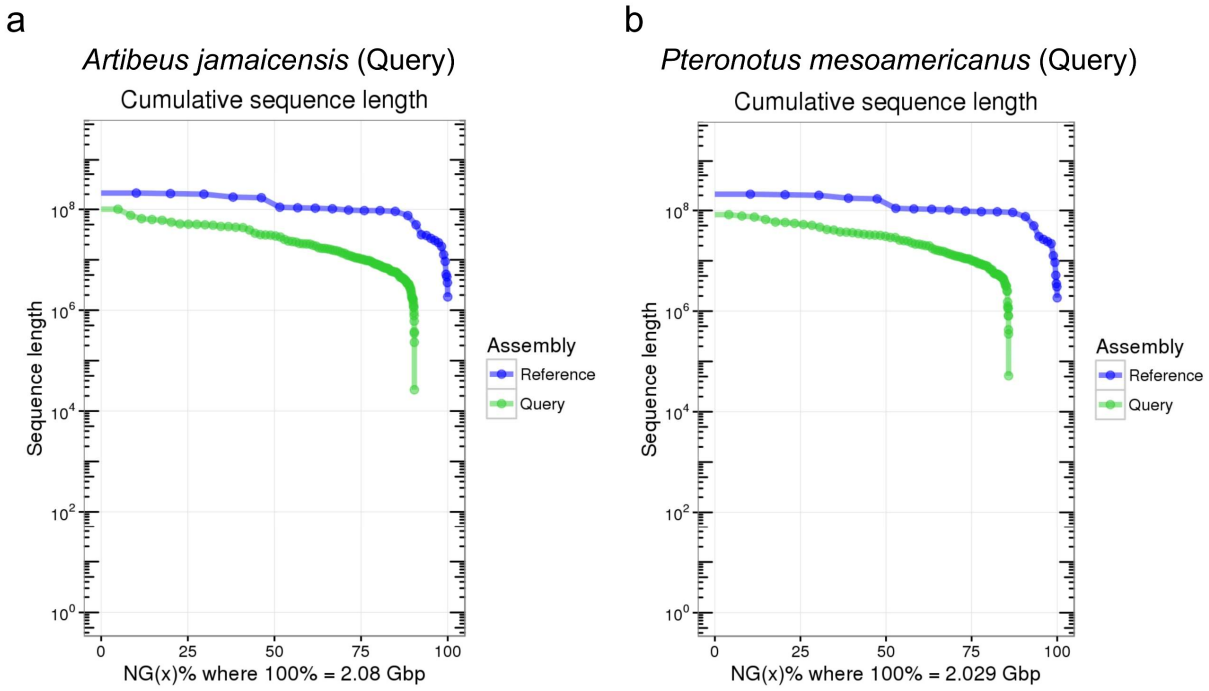
Supplementary Table 12. Protein-based BUSCO analysis of published protein sets for four phyllostomid bats (Gutiérrez-Guerrero et al., 2020). The mammalian BUSCO set (odb10) was used.

Species	Total BUSCO proteins	Complete	Complete Single copy	Complete Duplicated	Fragmented	Missing
Leptonycteris nivalis	*9226*	*58.9%*	*58.6%*	*0.3%*	*10.6%*	*30.5%*
Leptonycteris yerbabuenae	*9226*	*44.2%*	*43.9%*	*0.3%*	*12.2%*	*43.6%*
Macrotus waterhousii	*9226*	*58.4%*	*58.1%*	*0.3%*	*10.6%*	*31.0%*
Musonycteris harrisoni	*9226*	*59.7%*	*59.5%*	*0.2%*	*10.2%*	*30.1%*

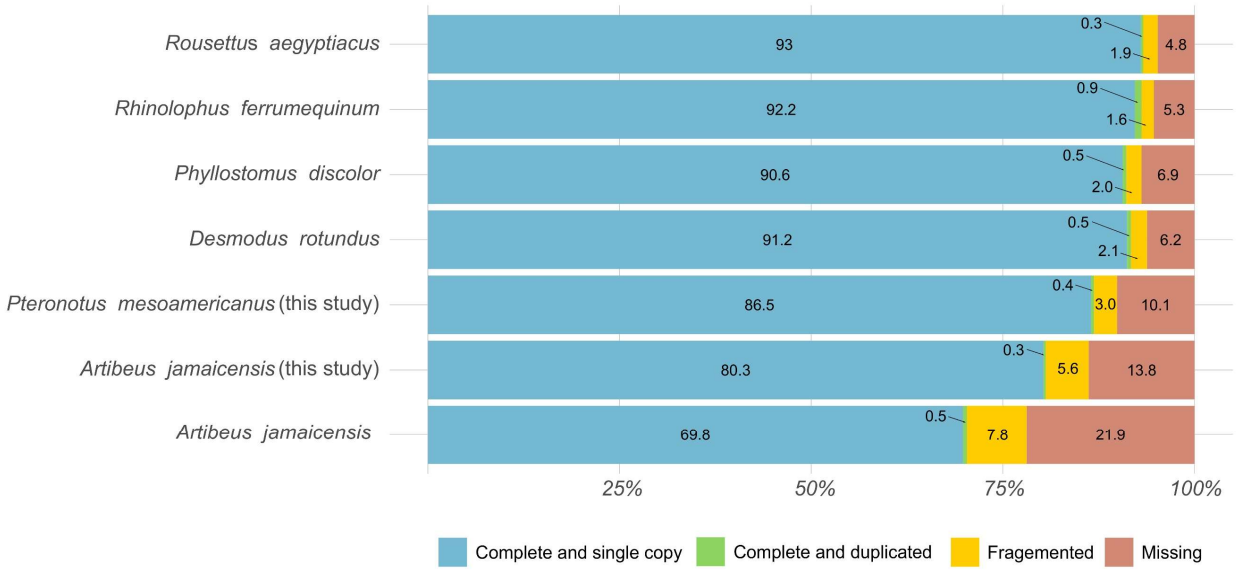
Supplementary figures



Supplementary Figure 1. Genomescope2 plots of k-mer analyses of the Illumina reads from a) *Artibeus jamaicensis* and b) *Pteronotus mesoamericanus*.



Supplementary Figure 2. Assemblytics plots of cumulative ordered sequence length versus NG(X)% cumulative percentage of genome. Comparison of the chromosome-level *Phyllostomus discolor* assembly (blue, reference) with the *Artibeus jamaicensis* assembly (green, query). b) Comparison of the *Phyllostomus discolor* assembly (blue, reference) with the *Pteronotus mesoamericanus* assembly (green, query).



Supplementary Figure 3. BUSCO analysis on genome assembly completeness comparing different publicly available bat genome sequences. The additional *Artibeus jamaicensis* genome is GenBank accession GCA_004027435.1.

References

1. Simmons, N. B., Seymour, K. L., Habersetzer, J. & Gunnell, G. F. Primitive Early Eocene bat from Wyoming and the evolution of flight and echolocation. *Nature* **451**, 818–821 (2008).
2. Simmons, N. B. & Geisler, J. H. Phylogenetic relationships of *Icaronycteris*, *Archaeonycteris*, *Hassianycteris*, and *Palaeochiropteryx* to extant bat lineages, with comments on the evolution of echolocation and foraging strategies in Microchiroptera. *Bull. Am. Nat. Hist. Mus.*, pp. 1–182 (1998).
3. Moss, C. F. & Surlykke, A. Auditory scene analysis by echolocation in bats. *J. Acoust. Soc. Am.* **110**, 2207–2226 (2001).
4. Monteiro, L. R. & Nogueira, M. R. Evolutionary patterns and processes in the radiation of phyllostomid bats. *BMC Evol. Biol.* **11**, 137 (2011).
5. Santana, S. E. & Cheung, E. Go big or go fish: morphological specializations in carnivorous bats. *Proc. R. Soc. B Biol. Sci.* **283**, 20160615 (2016).
6. Wilkinson, G. S. & Adams, D. M. Recurrent evolution of extreme longevity in bats. *Biol. Lett.* **15**, 20180860 (2019).
7. Wang, L.-F., Walker, P. J. & Poon, L. L. M. Mass extinctions, biodiversity and mitochondrial function: are bats 'special' as reservoirs for emerging viruses? *Curr. Opin. Virol.* **1**, 649–657 (2011).
8. Olival, K. J. *et al.* Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017).
9. Calisher, C. H., Childs, J. E., Field, H. E., Holmes, K. V. & Schountz, T. Bats: Important reservoir hosts of emerging viruses. *Clin. Microbiol. Rev.* **19**, 531–545 (2006).
10. Amman, B. R. *et al.* Seasonal pulses of Marburg virus circulation in juvenile *Rousettus aegyptiacus* bats coincide with periods of increased risk of human infection. *PLoS Pathog.* **8**, e1002877 (2012).
11. Pulliam, J. R. C. *et al.* Agricultural intensification, priming for persistence and the emergence of Nipah virus: a lethal bat-borne zoonosis. *J. R. Soc. Interface* **9**, 89–101 (2012).
12. Li, W. *et al.* Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676–679 (2005).
13. Pavlovich, S. S. *et al.* The Egyptian Rousette genome reveals unexpected features of bat antiviral immunity. *Cell* **173**, 1098-1110.e18 (2018).
14. Escalera-Zamudio, M. *et al.* The evolution of bat nucleic acid-sensing Toll-like receptors. *Mol. Ecol.* **24**, 5899–5909 (2015).
15. Zhang, G. *et al.* Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science* **339**, 456–460 (2013).
16. Jebb, D. *et al.* Six reference-quality genomes reveal evolution of bat adaptations. *Nature* **583**, 578–584 (2020).
17. Zepeda Mendoza, M. L. *et al.* Hologenomic adaptations underlying the evolution of sanguivory in the common vampire bat. *Nat. Ecol. Evol.* **2**, 659–668 (2018).
18. Taylor, D. J., Dittmar, K., Ballinger, M. J. & Bruenn, J. A. Evolutionary maintenance of filovirus-like genes in bat genomes. *BMC Evol. Biol.* **11**, 336 (2011).
19. Foley, N. M. *et al.* Growing old, yet staying young: The role of telomeres in bats' exceptional longevity. *Sci. Adv.* **4**, eaao0926 (2018).
20. Seim, I. *et al.* Genome analysis reveals insights into physiology and longevity of the Brandt's bat *Myotis brandtii*. *Nat. Commun.* **4**, 2212 (2013).
21. Emerling, C. A., Delsuc, F. & Nachman, M. W. Chitinase genes (CHIAs) provide genomic footprints of a post-Cretaceous dietary radiation in placental mammals. *Sci. Adv.* **4**, eaar6478 (2018).
22. Jiao, H. *et al.* Trehalase gene as a molecular signature of dietary diversification in mammals. *Mol. Biol. Evol.* **36**, 2171–2183 (2019).

23. Gutiérrez-Guerrero, Y. T. *et al.* Genomic consequences of dietary diversification and parallel evolution due to nectarivory in leaf-nosed bats. *GigaScience* **9**, (2020).
24. Parker, J. *et al.* Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* **502**, 228–231 (2013).
25. Dong, D. *et al.* The genomes of two bat species with long constant frequency echolocation calls. *Mol. Biol. Evol.* **34**, 20–34 (2017).
26. Teeling, E. C. *et al.* Bat biology, genomes, and the Bat1K project: To generate chromosome-level genomes for all living bat species. *Annu. Rev. Anim. Biosci.* **6**, 23–46 (2018).
27. Shaw, T. I. *et al.* Transcriptome sequencing and annotation for the Jamaican fruit bat (*Artibeus jamaicensis*). *PLoS One* **7**, e48472 (2012).
28. Herd, R. *Pteronotus parnellii*. in *Mammalian Species* vol. 209 1–5 (1983).
29. Clare, E. L. *et al.* Diversification and reproductive isolation: cryptic species in the only New World high-duty cycle bat*, Pteronotus parnellii*. *BMC Evol. Biol.* **13**, 26 (2013).
30. Fenton, M. B., Faure, P. A. & Ratcliffe, J. M. Evolution of high duty cycle echolocation in bats. *J. Exp. Biol.* **215**, 2935–2944 (2012).
31. Rojas, D., Warsi, O. M. & Dávalos, L. M. Bats (Chiroptera: Noctilionoidea) challenge a recent origin of extant neotropical diversity. *Syst. Biol.* **65**, 432–448 (2016).
32. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432 (2020).
33. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155–158 (2020).
34. Wang, W. *et al.* The draft nuclear genome assembly of *Eucalyptus pauciflora*: a pipeline for comparing de novo assemblies. *GigaScience* **9**, (2020).
35. Bederson, B. B., Shneiderman, B. & Wattenberg, M. Ordered and quantum treemaps: Making effective use of 2D space to display hierarchies. in *The Craft of Information Visualization* (eds. Bederson, B. B. & Shneiderman, B.) 257–278 (2003). doi:10.1016/B978-155860915-0/50033-0.
36. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
37. Evans, B. J., Upham, N. S., Golding, G. B., Ojeda, R. A. & Ojeda, A. A. Evolution of the largest mammalian genome. *Genome Biol. Evol.* **9**, 1711–1724 (2017).
38. Kapusta, A., Suh, A. & Feschotte, C. Dynamics of genome size evolution in birds and mammals. *Proc. Natl. Acad. Sci.* **114**, E1460–E1469 (2017).
39. Jebb, D. *et al.* Six new reference-quality bat genomes illuminate the molecular basis and evolution of bat adaptations. *bioRxiv* 836874 (2019) doi:10.1101/836874.
40. Sotero-Caio, C. G., Platt, R. N., Suh, A. & Ray, D. A. Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol. Evol.* **9**, 161–177 (2017).
41. Pace, J. K. & Feschotte, C. The evolutionary history of human DNA transposons: Evidence for intense activity in the primate lineage. *Genome Res.* **17**, 422–432 (2007).
42. Mitra, R. *et al.* Functional characterization of piggyBat from the bat *Myotis lucifugus* unveils an active mammalian DNA transposon. *Proc. Natl. Acad. Sci.* **110**, 234–239 (2013).
43. Ray, D. A. *et al.* Differential SINE evolution in vesper and non-vesper bats. *Mob. DNA* **6**, 10 (2015).
44. Ray, D. A. *et al.* Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res.* **18**, 717–728 (2008).
45. Pagan, H. J. T., Smith, J. D., Hubley, R. M. & Ray, D. A. PiggyBac-ing on a primate genome: novel elements, recent activity and horizontal transfer. *Genome Biol. Evol.* **2**, 293–303 (2010).

46. Joly-Lopez, Z. & Bureau, T. E. Exaptation of transposable element coding sequences. *Curr. Opin. Genet. Dev.* **49**, 34–42 (2018).
47. Johnson, W. E. Origins and evolutionary consequences of ancient endogenous retroviruses. *Nat. Rev. Microbiol.* **17**, 355-370 (2019).
48. Farkašová, H. *et al.* Discovery of an endogenous Deltaretrovirus in the genome of long-fingered bats (Chiroptera: Miniopteridae). *Proc. Natl. Acad. Sci.* **114**, 3145–3150 (2017).
49. Sakashita, A. *et al.* Endogenous retroviruses drive species-specific germline transcriptomes in mammals. *Nat. Struct. Mol. Biol.* 1–11 (2020) doi:10.1038/s41594-020-0487-4.
50. Katzourakis, A. & Gifford, R. J. Endogenous viral elements in animal genomes. *PLoS* Genet. 6, e1001191 (2010).
51. Bauermann, F. V., Ridpath, J. F., Weiblen, R. & Flores, E. F. HoBi-like viruses: an emerging group of pestiviruses. *J. Vet. Diagn. Invest.* **25**, 6–15 (2013).
52. Wu, Z. *et al.* Discovery of Diverse Rodent and Bat Pestiviruses With Distinct Genomic and Phylogenetic Characteristics in Several Chinese Provinces. *Front. Microbiol.* **9**, 2562 (2018).
53. Wu, Z. *et al.* Virome analysis for identification of novel mammalian viruses in bat species from Chinese provinces. *J. Virol.* **86**, 10999–11012 (2012).
54. Bobrowiec, P. E. D., Lemes, M. R. & Gribel, R. Prey preference of the common vampire bat (*Desmodus rotundus*, Chiroptera) using molecular analysis. *J. Mammal.* **96**, 54–63 (2015).
55. Becker, D. J. *et al.* Livestock abundance predicts vampire bat demography, immune profiles and bacterial infection risk. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **373**, (2018).
56. Zheng, X.-Y. *et al.* Viral metagenomics of six bat species in close contact with humans in southern China. *Arch. Virol.* **163**, 73–88 (2018).
57. van Riper, C., van Riper, S. G. & Hansen, W. R. Epizootiology and Effect of Avian Pox on Hawaiian Forest Birds. *The Auk* **119**, 929–942 (2002).
58. Yuan, L. *et al.* Evidence for retrovirus and paramyxovirus infection of multiple bat species in China. *Viruses* **6**, 2138–2154 (2014).
59. Yang, Z. & Nielsen, R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**, 908–917 (2002).
60. Zhang, J., Nielsen, R. & Yang, Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* **22**, 2472–2479 (2005).
61. Lambert, M. J. & Portfors, C. V. Adaptive sequence convergence of the tumor suppressor *ADAMTS9* between small-bodied mammals displaying exceptional longevity. *Aging* **9**, 573–581 (2017).
62. Kosiol, C. *et al.* Patterns of positive selection in six mammalian genomes. *PLoS Genet.* **4**, (2008).
63. Nielsen, R. *et al.* A Scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**, e170 (2005).
64. Torgerson, D. G., Kulathinal, R. J. & Singh, R. S. Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. *Mol. Biol. Evol.* **19**, 1973–1980 (2002).
65. Rhesus Macaque Genome Sequencing and Analysis Consortium *et al.* Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234 (2007).
66. Hawkins, J. A. *et al.* A metaanalysis of bat phylogenetics and positive selection based on genomes and transcriptomes from 18 species. *Proc. Natl. Acad. Sci.* **116**, 11351–11360 (2019).
67. Mears, H. V. & Sweeney, T. R. Better together: the role of IFIT protein–protein interactions in the antiviral response. *J. Gen. Virol.* **99**, 1463–1477 (2018).

68. Vijay, R. *et al.* Critical role of phospholipase A2 group IID in age-related susceptibility to severe acute respiratory syndrome–CoV infection. *J. Exp. Med.* **212**, 1851–1868 (2015).
69. Letko, M. *et al.* Adaptive Evolution of MERS-CoV to Species Variation in DPP4. *Cell Rep.* **24**, 1730–1737 (2018).
70. Frank, H. K., Enard, D. & Boyd, S. D. Exceptional diversity and selection pressure on SARS-CoV and SARS-CoV-2 host receptor in bats compared to other mammals. *bioRxiv* 2020.04.20.051656 (2020) doi:10.1101/2020.04.20.051656.
71. Cui, J., Eden, J.-S., Holmes, E. C. & Wang, L.-F. Adaptive evolution of bat dipeptidyl peptidase 4 (*dpp4*): implications for the origin and emergence of Middle East respiratory syndrome coronavirus. *Virol. J.* **10**, 304 (2013).
72. Tollis, M., Schiffman, J. D. & Boddy, A. M. Evolution of cancer suppression as revealed by mammalian comparative genomics. *Curr. Opin. Genet. Dev.* **42**, 40–47 (2017).
73. Park, J.-Y. *et al.* Breast cancer-associated missense mutants of the *PALB2* WD40 domain, which directly binds *RAD51C*, *RAD51* and *BRCA2*, disrupt DNA repair. *Oncogene* **33**, 4803–4812 (2014).
74. Furth, N. & Aylon, Y. The *LATS1* and *LATS2* tumor suppressors: beyond the Hippo pathway. *Cell Death Differ.* **24**, 1488–1501 (2017).
75. Kodama, K. *et al. MRTFB* suppresses colorectal cancer development through regulating *SPDL1* and *MCAM*. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 23625–23635 (2019).
76. Vicens, A. & Posada, D. Selective pressures on human cancer genes along the evolution of mammals. *Genes* **9**, 582 (2018).
77. Desai, T. M. *et al.* IFITM3 restricts influenza A virus entry by blocking the formation of fusion pores following virus-endosome hemifusion. *PLoS Pathog.* **10**, e1004048 (2014).
78. Perreira, J. M., Chin, C. R., Feeley, E. M. & Brass, A. L. IFITMs restrict the replication of multiple pathogenic viruses. *J. Mol. Biol.* **425**, 4937–4955 (2013).
79. Bailey, C. C., Zhong, G., Huang, I.-C. & Farzan, M. IFITM-family proteins: The cell's first line of antiviral defense. *Annu. Rev. Virol.* **1**, 261–283 (2014).
80. Zhang, Z., Liu, J., Li, M., Yang, H. & Zhang, C. Evolutionary dynamics of the interferon-induced transmembrane gene family in vertebrates. *PLoS One* **7**, e49265 (2012).
81. Shultz, A. J. & Sackton, T. B. Immune genes are hotspots of shared positive selection across birds and mammals. *eLife* **8**, e41815 (2019).
82. Ito, J., Gifford, R. J. & Sato, K. Retroviruses drive the rapid evolution of mammalian APOBEC3 genes. *Proc. Natl. Acad. Sci.* **117**, 610–618 (2020).
83. Münk, C., Willemsen, A. & Bravo, I. G. An ancient history of gene duplications, fusions and losses in the evolution of APOBEC3 mutators in mammals. *BMC Evol. Biol.* **12**, 71 (2012).
84. Yan, N. & Chen, Z. J. Intrinsic antiviral immunity. *Nat. Immunol.* **13**, 214–222 (2012).
85. Huang, I.-C. *et al.* Distinct patterns of IFITM-mediated restriction of filoviruses, SARS coronavirus, and influenza A virus. *PLoS Pathog.* **7**, e1001258 (2011).
86. Diamond, M. S. & Farzan, M. The broad-spectrum antiviral functions of IFIT and IFITM proteins. *Nat. Rev. Immunol.* **13**, 46–57 (2013).
87. Brass, A. L. *et al.* The IFITM proteins mediate cellular resistance to influenza a H1N1 virus, West Nile virus, and dengue virus. *Cell* **139**, 1243–1254 (2009).
88. Compton, A. A. *et al.* IFITM Proteins incorporated into HIV-1 virions impair viral fusion and spread. *Cell Host Microbe* **16**, 736–747 (2014).
89. Foster, T. L. *et al.* Resistance of transmitted founder HIV-1 to IFITM-mediated restriction. *Cell Host Microbe* **20**, 429–442 (2016).
90. Läubli, H. & Varki, A. Sialic acid–binding immunoglobulin-like lectins (Siglecs) detect self-associated molecular patterns to regulate immune responses. *Cell. Mol. Life Sci.* **77**, 593–605 (2020).

91. Shukla, E., Thorat, L. J., Nath, B. B. & Gaikwad, S. M. Insect trehalase: Physiological significance and potential applications. *Glycobiology* **25**, 357–367 (2015).
92. Zhou, Y. *et al.* Duplication and diversification of trehalase confers evolutionary advantages on lepidopteran insects. *Mol. Ecol.* **28**, 5282–5298 (2019).
93. Merzendorfer, H. & Zimoch, L. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *J. Exp. Biol.* **206**, 4393–4412 (2003).
94. Ostrand-Rosenberg, S. Immune surveillance: a balance between protumor and antitumor immunity. *Curr. Opin. Genet. Dev.* **18**, 11–18 (2008).
95. Li, D. *et al.* KLF4-mediated negative regulation of IFITM3 expression plays a critical role in colon cancer pathogenesis. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **17**, 3558–3568 (2011).
96. Siegrist, F., Ebeling, M. & Certa, U. The small interferon-induced transmembrane genes and proteins. *J. Interferon Cytokine Res. Off. J. Int. Soc. Interferon Cytokine Res.* **31**, 183–197 (2011).
97. Alteber, Z. *et al.* The anti-inflammatory IFITM genes ameliorate colitis and partially protect from tumorigenesis by changing immunity and microbiota. *Immunol. Cell Biol.* **96**, 284–297 (2018).
98. Sikes, R. S. 2016 Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. *J. Mammal.* **97**, 663–688 (2016).
99. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
100. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
101. Nattestad, M. & Schatz, M. C. Assemblytics: a web analytics tool for the detection of variants from an assembly. *Bioinformatics* **32**, 3021–3023 (2016).
102. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
103. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
104. Kovaka, S. *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, (2019).
105. Haas, B. J. *et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
106. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).
107. Kent, W. J. BLAT—The BLAST-Like Alignment Tool. *Genome Res.* **12**, 656–664 (2002).
108. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
109. Stanke, M. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
110. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinforma.* **Chapter 4**, Unit 4.10 (2009).
111. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
112. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982 (2007).
113. Abrusán, G., Grundmann, N., DeMester, L. & Makalowski, W. TEclass—a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics* **25**, 1329–1330 (2009).

114. Yan, H., Bombarely, A. & Li, S. DeepTE: a computational method for de novo classification of transposons with convolutional neural network. *Bioinformatics* doi:10.1093/bioinformatics/btaa519.
115. Kumar, S. & Subramanian, S. Mutation rates in mammalian genomes. *Proc. Natl. Acad. Sci.* **99**, 803–808 (2002).
116. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
117. Löytynoja, A. Phylogeny-aware alignment with PRANK. *Methods Mol. Biol. Clifton NJ* **1079**, 155–170 (2014).
118. Tsagkogeorga, G., Parker, J., Stupka, E., Cotton, J. A. & Rossiter, S. J. Phylogenomic Analyses Elucidate the Evolutionary Relationships of Bats. *Curr. Biol.* **23**, 2262–2267 (2013).
119. Doronina, L. *et al.* Speciation network in Laurasiatheria: retrophylogenomic signals. *Genome Res.* **27**, 997–1003 (2017).
120. Nishihara, H., Hasegawa, M. & Okada, N. Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. *Proc. Natl. Acad. Sci.* **103**, 9929–9934 (2006).
121. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
122. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005).
123. Zhang, Z. *et al.* PseudoPipe: an automated pseudogene identification pipeline. *Bioinformatics* **22**, 1437–1439 (2006).
124. J, M., A, B. & Rd, F. Predicting active site residue annotations in the Pfam database. *BMC Bioinformatics* **8**, 298–298 (2007).
125. Guy, L., Roat Kultima, J. & Andersson, S. G. E. genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* **26**, 2334–2335 (2010).