EVOLUTIONARY BIOLOGY

Comparative analyses of bat genomes identify distinct evolution of immunity in Old World fruit bats

Shilin Tian[1]†, Jiaming Zeng[1]†, Hengwu Jiao[1], Dejing Zhang[2], Libiao Zhang[3], Cao-qi Lei[1], Stephen J. Rossiter[4], Huabin Zhao[1]*

Bats have been identified as natural reservoir hosts of several zoonotic viruses, prompting suggestions that they have unique immunological adaptations. Among bats, Old World fruit bats (Pteropodidae) have been linked to multiple spillovers. To test for lineage-specific molecular adaptations in these bats, we developed a new assembly pipeline to generate a reference-quality genome of the fruit bat *Cynopterus sphinx* and used this in comparative analyses of 12 bat species, including six pteropodids. Our results reveal that immunity-related genes have higher evolutionary rates in pteropodids than in other bats. Several lineage-specific genetic changes were shared across pteropodids, including the loss of *NLRP1*, duplications of *PGLYRP1* and *C5AR2*, and amino acid replacements in *MyD88*. We introduced *MyD88* transgenes containing Pteropodidae-specific residues into bat and human cell lines and found evidence of dampened inflammatory responses. By uncovering distinct immune adaptations, our results could help explain why pteropodids are frequently identified as viral hosts.

INTRODUCTION

Bats are widely thought to be the natural reservoir hosts of several viruses of public health concern, including Nipah (*1*), Hendra (*2*), Ebola (*3*), Marburg (*4*), Middle East respiratory syndrome (MERS) (*5*), severe acute respiratory syndrome (SARS) (*6*), and, most recently, SARS coronavirus 2 (SARS-CoV-2)—the cause of coronavirus disease 2019 (*7*). Most of these cases involve either Old World fruit bats (Pteropodidae) or horseshoe bats (Rhinolophidae), two major clades that are broadly codistributed and that belong to the suborder Yinpterochiroptera. Observations that, when infected with viruses, these bats frequently show no clinical symptoms, whereas other mammals develop severe disease, have prompted suggestions that bats have unique immune adaptations to cope with viruses (*8*).

In support of specialized immunity in bats, studies of inoculated captive individuals (*1*, *3*, *9*) and bat-derived cells (*10*) have reported evidence of dampened inflammatory and innate immune responses. The first of these, down-regulated inflammation, is thought to limit tissue damage—a common pathology linked to viral infection in other species (*11*)—and may have evolved as an adaptation to flight (*12*). Molecular studies have also provided insights, shedding light on the potential genetic bases of bat immunity. For example, reduced activation of the stimulator of interferon genes (STING)–dependent pathway has been attributed to specific amino acid residues in some bats (*13*), while constitutive expression of interferon-α (IFN-α) in some pteropodids may help maintain a state of preparedness to control viral replication (*14*).

Comparative genomic approaches have further improved our understanding of bat immune adaptations, including differences among lineages (*12*, *15*–*20*). For example, bats show putative adaptive amino acid variation in the ligand-binding ectodomain of the Toll-like receptor (TLR) gene *TLR8*, which may confer tolerance to specific viruses. Other discoveries include positive selection on the inflammasome *NLRP3* in the genus *Pteropus* and the earlier loss of the *PYHIN* gene family, which functions in inflammasome formation (*12*). Consistent with these results, Pavlovich *et al.* (*21*) compared the genome of the fruit bat *Rousettus aegyptiacus* to other bat genomes and speculated that positive selection on immune genes, alongside expansions of natural killer cell receptors, major histocompatibility complex class I genes, and type I IFNs, were likely to enhance viral tolerance rather than increase antiviral potency. More recently, comparisons of reference-quality genomes have added to earlier findings, uncovering amino acid substitutions in other inflammation-related loci and the putatively adaptive expansion of the *APOBEC3* gene family (*15*, *22*).

Despite substantial advances in the genomics and mechanisms of bat immunity, few studies have focused on specific clades of bats implicated in hosting zoonotic viruses. Here, we apply methods in comparative genomics with functional assays to test for immune-related adaptations in the family Pteropodidae, members of which have been reported to harbor several viruses from the families Paramxyoviridae (*23*, *24*) and Filoviridae (*25*–*27*). To this end, we developed a new assembly pipeline and used this to generate a near-complete genome of the short-nosed fruit bat *Cynopterus sphinx*. This species is widely distributed across South and Southeast Asia and is reported to host several β-coronaviruses (*28*, *29*)—a group that includes MERS-CoV, SARS-CoV, and SARS-CoV-2 (*30*). By combining our newly generated genome with high-quality genome assemblies from five other pteropodid bats (*15*, *19*), related bats, and mammalian outgroups, we seek to determine whether pteropodid-specific modifications can explain variation in viral load (*31*).

[1]College of Life Sciences, Frontier Science Center for Immunology and Metabolism, Hubei Key Laboratory of Cell Homeostasis, Wuhan University, Wuhan 430072, China. [2]Novogene Bioinformatics Institute, Beijing 100015, China. [3]Institute of Zoology, Guangdong Academy of Sciences, Guangzhou 510260, China. [4]School of Biological and Chemical Sciences, Queen Mary University of London, London, UK.
*Corresponding author. Email: huabinzhao@whu.edu.cn
†These authors contributed equally to this work.

RESULTS AND DISCUSSION

A new assembly pipeline generates a reference-quality *C. sphinx* genome

We developed a new assembly pipeline to assemble a reference-quality genome of the short-nosed fruit bat *C. sphinx* (Fig. 1A and fig. S1), using a combination of 104.17-fold Nanopore reads, 60.79-Gb Illumina paired-end data, and 574.14 million Hi-C (high-throughput chromosome conformation capture) interaction pairs (tables S1 and S2). A preliminary genome survey based on *k*-mer analysis with Illumina paired-end data showed that *C. sphinx* had a smaller genome size (1.91 Gb) and a higher degree of heterozygosity (0.79%) than reported for previously published bat genomes (fig. S2 and tables S3 and S4) (*12*, *16*, *17*, *21*). We first assembled and corrected a set of 370 contigs (hereafter termed Contig v1) with an N50 size of 66.91 Mb (table S5). Second, using Contig v1 as a reference, we calculated the Hi-C interaction frequency using an agglomerative hierarchical clustering algorithm (table S6) and combined this information with the karyotype data identification results for *C. sphinx* to give 17 linked contig groups (fig. S3) (*32*). Third, in a deviation from common assembly strategies, we realigned all high-quality Nanopore reads to our set of contigs and then extracted optimal mapped reads (99.77%) of each contig group to perform a local assembly (table S7). Fourth, we merged all assemblies to obtain a set of 207 contigs (termed as Contig v2) with a 26% increase in the N50 value (84.24 Mb), exceeding that of previous published chromosome-level bat genomes and equating to a genome size of 1.86 Gb (Fig. 1B and table S8). Last, we used Hi-C interaction pairs to anchor 43 of these contigs (1.82 Gb) onto 17 chromosomes (table S9).

Typical genome assembly procedures based on third-generation sequencing data use Hi-C as a final step to generate interaction pairs that link contigs to chromosomes. In contrast, we used Hi-C interaction pairs to cluster Nanopore long sequences with potential link relationships in advance, thereby avoiding any erroneous overlap caused by long-distance repetitive sequences when constructing the string graph during assembly (*33*). In doing so, we obtained four gapless chromosome sequences, the longest of which was 158.00 Mb (fig. S4). Compared to the standard approach, our pipeline improved assembly quality based on genome completeness (98.64% versus 99.04%, respectively) as measured by the proportion of short-reads that map to the assembly: genome coverage (99.37% versus 99.75%) and coverage of BUSCO (Benchmarking Universal Single-Copy Orthologs) (*34*) genes for mammals (96.44% versus 97.27%) (table S10). In addition, the assembly quality value (QV) was estimated as 41.05 by an approach of reference-free and *k*-mer–based approach, exceeding the Vertebrate Genome Project standard of QV40 (*35*, *36*). Together, these results indicate that our new assembly pipeline has the potential to improve upon current approaches in producing genome assemblies with higher accuracy, completeness, and contiguity.

Transposable elements (TEs) comprised 26.77% of the *C. sphinx* genome, which is comparable to estimates for five other pteropodids (25.50 to 27.72%) but is lower than estimates for five other bat genomes (30.59 to 38.99%) (fig. S5 and table S11). Despite this, we found that one of the active TEs, SINE/5S rRNA [5S ribosomal RNA (rRNA)–derived short interspersed nuclear element (SINE)] , was more abundant in pteropodids than that in other bats (fig. S6 and table S12). Lower than 10% divergence rate in comparison to the Repbase (*37*) indicated that the SINE/5S rRNA had experienced considerable recent TE activity (fig. S7). We observed frequent insertions in coding genes, then performed Gene Ontology (GO) functional enrichment analysis for the set of open reading frames that contained rRNA-derived SINE/5S elements, and found that these were significantly enriched for the GO terms related to visual function, auditory function, feeding, limb development, antiviral function, and immune function (table S13). Thus, active SINE/5S rRNA may play a role in generating genetic and epigenetic variation (*38*), possibly via modulating the expression of genes that underpin key traits in pteropodids, although experimental validation is needed.

Pteropodids exhibit high genomic synteny and X chromosome conservation

We undertook comparative genomic analysis using our dataset of 18 focal mammals (data S1). For this, we performed gene prediction

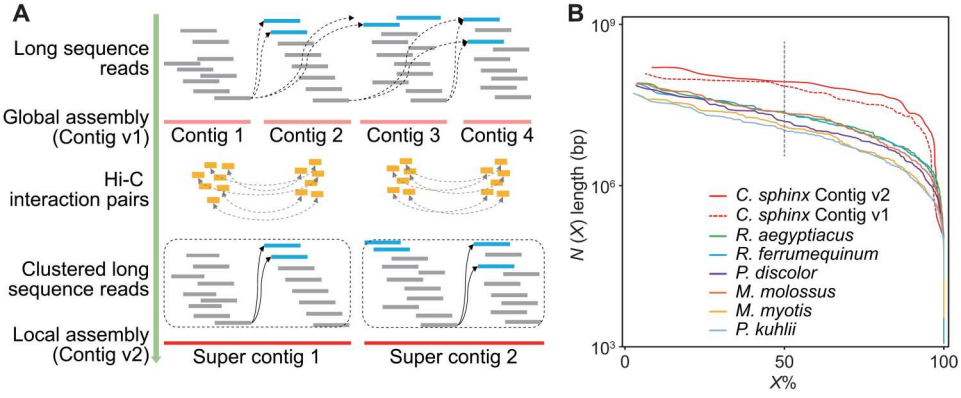

Fig. 1. Improved assembly quality by a new assembly pipeline. (**A**) New assembly pipeline used to newly assemble a reference-quality genome of *C. sphinx*. Application of Hi-C technology is to remove the assembly noise due to the linkage of nonproximal repeats with low frequency. Blue blocks represent local Nanopore reads in repetitive regions of the genome, gray blocks indicate other Nanopore reads, and yellow blocks denote Hi-C interaction pairs. (**B**) An increase in contig N50 of the newly assembled *C. sphinx* genome as compared to six other chromosome-level bat genomes. We accumulated the contigs from long to short; when the accumulated length of contigs reaches *X*% of the total genome length, the length of the corresponding contig is *N* (*X*). The vertical dashed line in gray corresponds to values of contig N50 of each genome assembly.

for our new *C. sphinx* genome as well as for the six chromosome-level bat genomes (*15*) and the Old World fruit bats *Cynopterus brachyotis* (*19*) and *Macroglossus sobrinus* [National Center for Biotechnology Information (NCBI) accession no. GCA_004027375.1] (data S1 and tables S14 to S16). Combined with the genes from seven other chromosome-level mammalian genomes, we obtained 7139 one-to-one orthologs to reconstruct a maximum likelihood tree of 18 focal mammals. This tree was consistent with the established phylogenetic relationships (*39*) and recovered the suborders Yinpterochiroptera and Yangochiroptera with an estimated divergence of 65.5 million years ago (Ma ago) (Fig. 2A). Next, we performed whole-genome synteny analysis by reciprocally aligning 17 mammalian genomes to *C. sphinx* genome. We found that the syntenic ratio decreased as a function of evolutionary distance from *C. sphinx* (Fig. 2A and table S17), and we observed a higher level of conserved synteny within the Pteropodidae, with a genome syntenic ratio of >93% for each species (table S17). These results showed an overall strong synteny despite known variation in chromosome number among different pteropodid species (table S17).

Among the results for chromosome-level genomes, we found that the large interchromosomal rearrangement events (≥1 Mb in length) between *C. sphinx* and other bats primarily occurred on the top 12 longest chromosomes in *C. sphinx*, with the exception of chromosome LG10 (Fig. 2B and data S2). We further found that chromosome LG10 sequences of *C. sphinx* were a unique match to the X chromosome sequences of each non-bat mammal with an average syntenic rate of 67.12% (ranging from 51.84% in mouse to 74.05% in horse) (Fig. 2A), suggesting that the chromosome LG10 is the X chromosome in *C. sphinx*. Notably, in the chromosome-level genomes of all bats, except *Pipistrellus kuhlii*, there was one chromosome that had good synteny (ranging from 69.76% in *Molossus molossus* to 87.41% in *R. aegyptiacus*) with the X chromosome of *C. sphinx* (Fig. 2). We suspected that the X chromosome sequences of *P. kuhlii* are scattered into multiple scaffolds, and, thus, we used our assembly pipeline to recluster and anchor chromosome sequences of *P. kuhlii* using Hi-C interaction pairs, resulting in a 101.77-Mb consecutive X chromosome sequence (fig. S8 and table S18). Collectively, our results unambiguously showed that the X chromosomes of bats were highly conserved (Fig. 2 and data S2), as in other mammals.

Pteropodidae-specific sequences are enriched with endogenous retroviral elements

We identified and analyzed Pteropodidae-specific sequences. To accomplish this, we first identified 1573.16-Mb (84.71% of the *C. sphinx* genome) and 1073.91-Mb (57.83% of the *C. sphinx* genome) core syntenic blocks (shared in all bat species) across the genomes of 6 pteropodids and all 11 bats, respectively (table S19 and fig. S9). Next, we detected an average of 26.09-Mb Pteropodidae-specific sequences (ranging from 6.33 to 52.50 Mb) that were present in a given pteropodid genome but absent in other bat genomes, harboring an average of 417 genes in each pteropodid that were overlapped with at least 50% of the length of a coding sequence (table S20). We subsequently found that a total of 2014 (80.50%) Pteropodidae-specific genes showed poor homology with other bats that failed to meet the criterion of 80% of the length being covered with 50% identity (fig. S9). A total of 92.45% of the Pteropodidae-specific genes generally exhibited known functions when compared to public databases (table S21). Noticeably, we observed that 23.68% (153 of 646; $P = 1.47 \times 10^{-70}$, Fisher′s exact test) and 37.25% (76 of 204; $P = 3.86 \times 10^{-76}$, Fisher′s exact test) of Pteropodidae-specific genes carried the fragments of endogenous retroviral elements in the chromosome-level genomes of *C. sphinx* and *R. aegyptiacus*, respectively (tables S22 and S23). We found that most of the retroviruses in both pteropodids belonged to gammaretroviruses, followed by betaretroviruses (tables S24), as reported in other bats elsewhere (*15*, *40*). We further predicted the 5′ and 3′ long terminal repeat (LTR) sequences of each retroviral gene and estimated their divergence rate to date the retrovirus infection. Our analysis revealed that both pteropodids underwent two ancient retroviral integration burst events—one was around 1 to 3 Ma ago and the other was around 10 to 12 Ma ago (fig. S10). Because bats can better tolerate and survive viral infections than most other mammals (*8*), we speculate that more frequent virus gene integration into host genomes could be one of the important factors contributing to genomic diversity in pteropodids.

Segmental duplications contribute to *PGLYRP1* duplication in pteropodids

Segmental duplications (SDs) are large (>1 kb in length) and high-identical (>90% sequence identity) copies of genomic regions, which have important implications for gene innovations in genomes (*41*). Distinct properties of SDs could reveal copy number variation in genomes. We identified an average of 34.38 Mb of SDs in these chromosome-level genomes of bats, containing an average of 611 genes (table S25). Compared with the five non-pteropodid bats, we found coembedding of 24 human homologs unique in the SD regions of these two pteropodid genomes (*C. sphinx* and *R. aegyptiacus*), with 20 (83.33%) occurred gene duplication in at least one genome (table S26). Among these 20 genes, *PGLYRP1*—a member of the peptidoglycan recognition protein family—was found to be duplicated in pteropodids (table S26). This locus is generally conserved across mammals as a single copy (*42*), and its protein product is released by activated neutrophils and recognizes infected microorganisms by binding to peptidoglycan, exerting bacteriostatic and bactericidal activities (*42*). Hence, the detected copy number variation could help to attenuate inflammation.

Gene family analysis identifies *NLRP1* loss and *C5AR2* duplication in pteropodids

To assess gene family evolution in the Pteropodidae, we used OrthoMCL (*43*) to assign genes into gene families and classified these as core (shared across all species), partially shared, or species-specific. By comparing each pairwise combination of genomes, we found that the number of all gene families approached an asymptote, suggesting good representation of our selected genomes for the Pteropodidae (Fig. 3A). We found that core gene families accounted for an average of 67.68% of the total number of gene families, partially shared gene families accounted for an average of 25.58%, and species-specific gene families accounted for 6.75% (table S27).

We also tested for gene family expansion and contraction in the Pteropodidae. For this, we analyzed a phylogenetic tree of 18 mammals using the software Café (*44*), with the ancestral branch leading to Pteropodidae defined as the foreground. We found that 14 gene families underwent expansion and 11 underwent

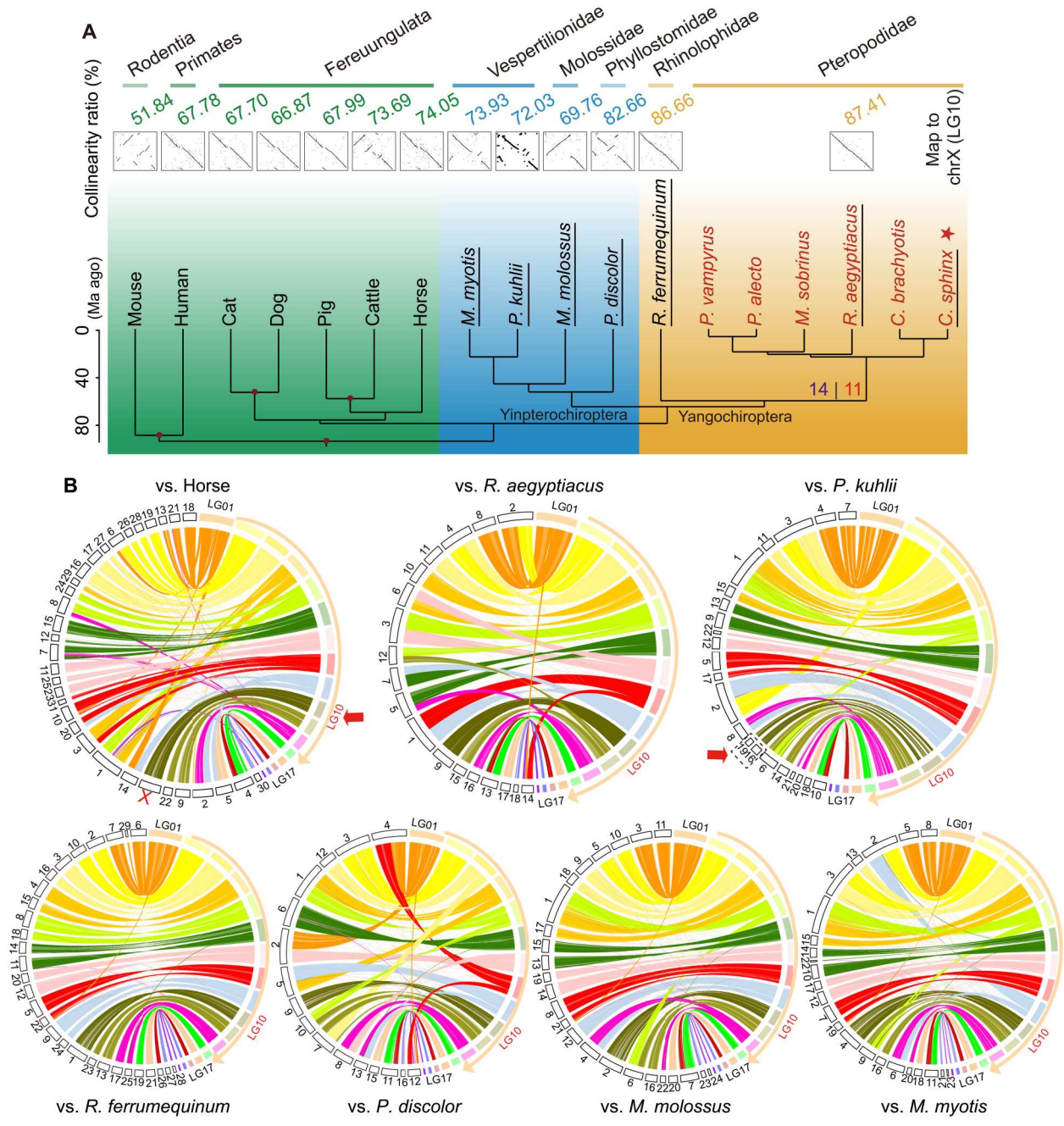

Fig. 2. Phylogenetic tree and genome synteny. (**A**) A maximum likelihood phylogenetic tree of 18 mammals. The newly assembled *C. sphinx* genome is indicated by a red star, and all seven bat species with chromosome-level genomes are underlined. Numbers on the ancestral branch leading to Pteropodidae denote expanded (purple) and contracted (red) gene families. Branch lengths are scaled by millions of years. The dot plots of syntenic sequences of chromosome X are shown on the top; the X sequences from seven non-bat mammals and six chromosome-level bats are aligned to X sequence of *C. sphinx*, and high-stringency alignments are shown by black dots. (**B**) Circos plots showing genome synteny. The chromosome-level genomes of six bats and horse were aligned to *C. sphinx* genome. Syntenic blocks are linked between genomes in Circos plots. Chromosome order of *C. sphinx* is indicated in the rightmost Circos plot, and LG10 is the X chromosome of *C. sphinx*.

contraction on the ancestral branch of Pteropodidae (Fig. 2A). At least one member in most expanded gene families (11 of 14) was located within the SD regions of *C. sphinx* and *R. aegyptiacus* genomes (fig. S11), again suggesting that SDs could be associated with gene innovations in genomes. GO enrichment analyses of expanded genes, based on human homologs, revealed significant terms relating to immunity, including "complement activation" (GO:0030449), as well as to multiple terms associated with olfactory receptors (e.g., GO:0004984 and GO:0050911), supporting previous findings (*45*) (table S28). GO analysis of contracted genes revealed 44 significant terms, of which 33 were related to immunity, including "innate immune response" (GO:0045087) (table S28).

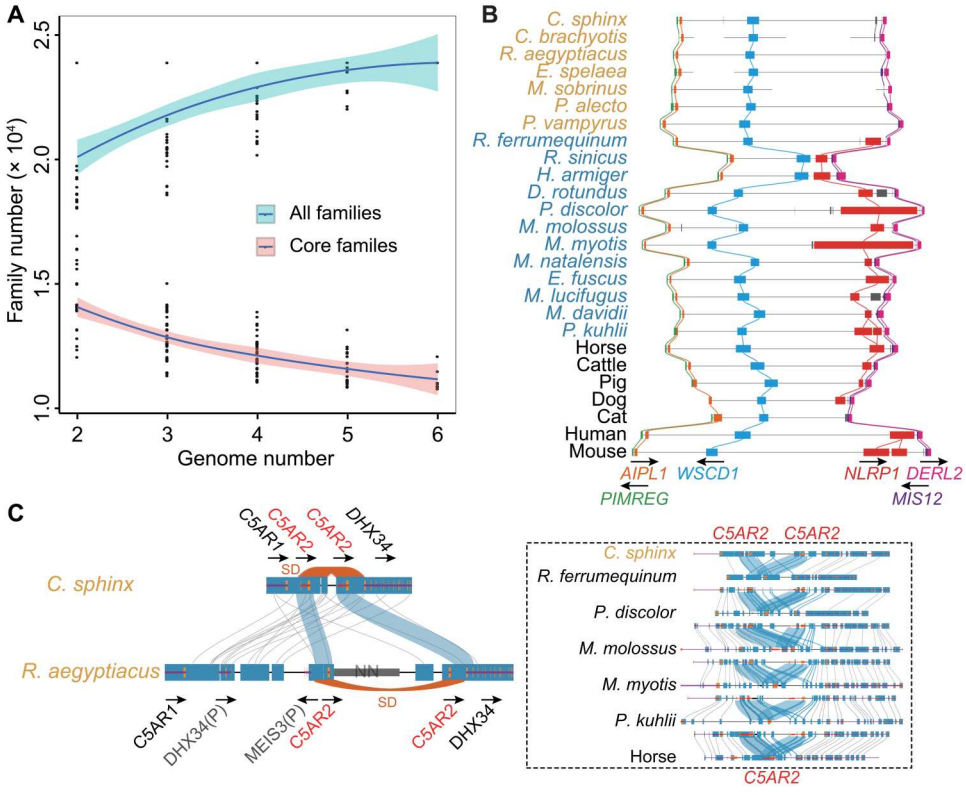

Fig. 3. Evolution of gene families. (**A**) Increasing trend of all gene families and decreasing trend of core gene families as number of pteropodid genomes increases. (**B**) Loss of the key inflammasome *NLRP1* gene in Pteropodidae. The orthologous relationships of *NLRP1* and its surrounding genes were shown. Species names in the left side indicate pteropodids (in yellow), other bats (in blue), and non-bat mammals (in black). Colored bars represent five genes surrounding *NLRP1* that have syntenic relationships, and unique genes in a species are indicated by gray bars. The black arrows above the gene name indicate the direction of a gene from 5′ to 3′. (**C**) Duplication of *C5AR2* in pteropodids. Left: Two copies of *C5AR2* are shown in two pteropodids, with the thick orange lines indicating SDs; the black arrows indicate the direction of a gene from 5′ to 3′. Right: *C. sphinx* with two *C5AR2* copies was compared to each of non-pteropodid bats with one *C5AR2* copy, with the horse as an example of non-bat mammals.

As part of one contracted gene family, we detected the common loss of the key inflammasome *NLRP1* gene in all six pteropodids examined, which was present in all other bats and mammals tested (Fig. 3B and fig. S12). This finding was also corroborated by a recent study, in which *NLRP1* gene was found to be absent from genomes of three pteropodids studied (i.e., *R. aegyptiacus*, *Pteropus* alecto, and *Pteropus vampyrus*) (*46*). *NLRP1* belongs to the NLR (nucleotide-binding domain leucine-rich repeat) and was the first protein shown to form an inflammasome (*47*). We compared homologous LINE-1 (L1) insertions upstream and downstream of *NLRP1* (or its inferred site of deletion) across bats and found that the content, as measured by both number of insertions and cumulative L1 length, was lower in pteropodids (fig. S13). Thus, it is plausible that the loss of *NLRP1*, and any associated reduction in inflammatory response, is a consequence of an L1 insertion-mediated deletion, as previously suggested in humans (*48*). In addition, a gene family found to have undergone expansion in pteropodids was seen to contain a tandem duplicate of the complement receptor *C5AR2* gene (Fig. 3C and fig. S14). *C5AR2* is a receptor of the complement anaphylatoxin C5a, which is known to regulate TLR4-mediated *NLRP3* inflammasome activation (*49*), and has previously been suggested to perform an anti-inflammatory function for regulating macrophage *IL1*β release (*49*–*51*). In contrast to a single copy of *C5AR2* seen in the five non-pteropodid bats, two copies were identified in pteropodids, a consequence of SD evolution (Fig. 3C and table S26). This duplication might have, thus, enhanced the ability of pteropodids to regulate macrophage *IL1*β release and so suppress inflammatory responses following viral infection, although experimental evidence is needed to confirm this.

Immune genes show higher evolutionary rates across pteropodids

We analyzed 7,139 gene trees based on one-to-one orthologs, each constrained to the reconstructed phylogeny (Fig. 2A). By applying branch tests and branch-site tests in phylogenetic analysis of maximum likelihood (PAML) (*52*) to the ancestral branch of Pteropodidae, we identified 316 rapidly evolving genes (REGs) and 241 positively selected genes (PSGs), respectively. Of these respective sets of genes, 35 (11%) and 46 (19%) loci have known roles in immunity and/or antiviral responses (tables S29 and S30), supporting previous reports of adaptations in bat immune systems (*53*). Consistent with this result, we also performed separate functional gene enrichment analyses for PSGs and REGs and found that both gene sets were significantly enriched for the GO terms "viral process" (GO:0016032, 14 PSGs and 18 REGs) and "immune system process" (GO:0002376, 25 PSGs and 32 REGs) (data S3 and S4).

In addition, four PSGs (*TAX1BP1*, *TRAF3IP1*, *cGAS*, and *ACOD1*) and five REGs (*TLR2*, *MyD88, TICAM1*, *TRAF3*, and *XRCC6*) were associated with the significantly enriched GO term "regulation of IFN I production" (GO:0032479) (data S3 and S4), three PSGs (*PECAM1*, *STAM*, and *CCR2*) and two REGs (*AKR1B1* and *EPOR*) were associated with the significantly enriched term "Janus kinase–signal transducers and activators of transcription (JAK-STAT) cascade" (GO:0007259), and two PSGs (*NUP188* and *TPR*) were associated with the significantly enriched Reactome term "IFN-stimulated gene 15 (ISG15) antiviral mechanism" (R-HSA-1169408). Together, these enriched GO and Reactome terms imply that pteropodids have evolved diverse adaptations to counter viral infections, encompassing pattern recognition and IFN-mediated control of viral replication (Fig. 4A) (*54*). Specifically, the binding of IFNs to IFN-α/β receptors on infected and adjacent cells triggers the JAK-STAT signaling pathway, which leads to phosphorylation of STAT proteins. Subsequently, phosphorylated STAT proteins induce the expression of ISGs that conjugate some viral proteins (*8*, *55*), inhibiting virus proliferation and release.

Other notable PSGs include *IGF2R*, *MEFV*, and *cGAS* (Fig. 4A). *IGF2R* encodes the insulin-like growth factor 2 receptor (IGF2R), which is a member of the collectin family of molecules that plays a role in the first line of host defense by recognizing oligomannose molecules on the surface of bacteria and viruses; these infectious agents are then neutralized by complement-mediated cell lysis or are attacked by phagocytes (*56*–*58*). *MEFV* encodes the inflammasome recognition receptor pyrin and acts as an autophagy receptor for the degradation of several inflammasome components such as *NLRP3* (Fig. 4A), hence for preventing excessive *IL1*β- and *IL18*-mediated inflammation in pteropodids (*59*, *60*). This PSG likely provides additional evidence for the previously demonstrated low level of NLRP3 inflammasome activation in bats (*10*). *cGAS* encodes cyclic guanosine 5′-monophosphate–adenosine 5′-monophosphate (cGAMP) synthase (cGAS), which binds to microbial

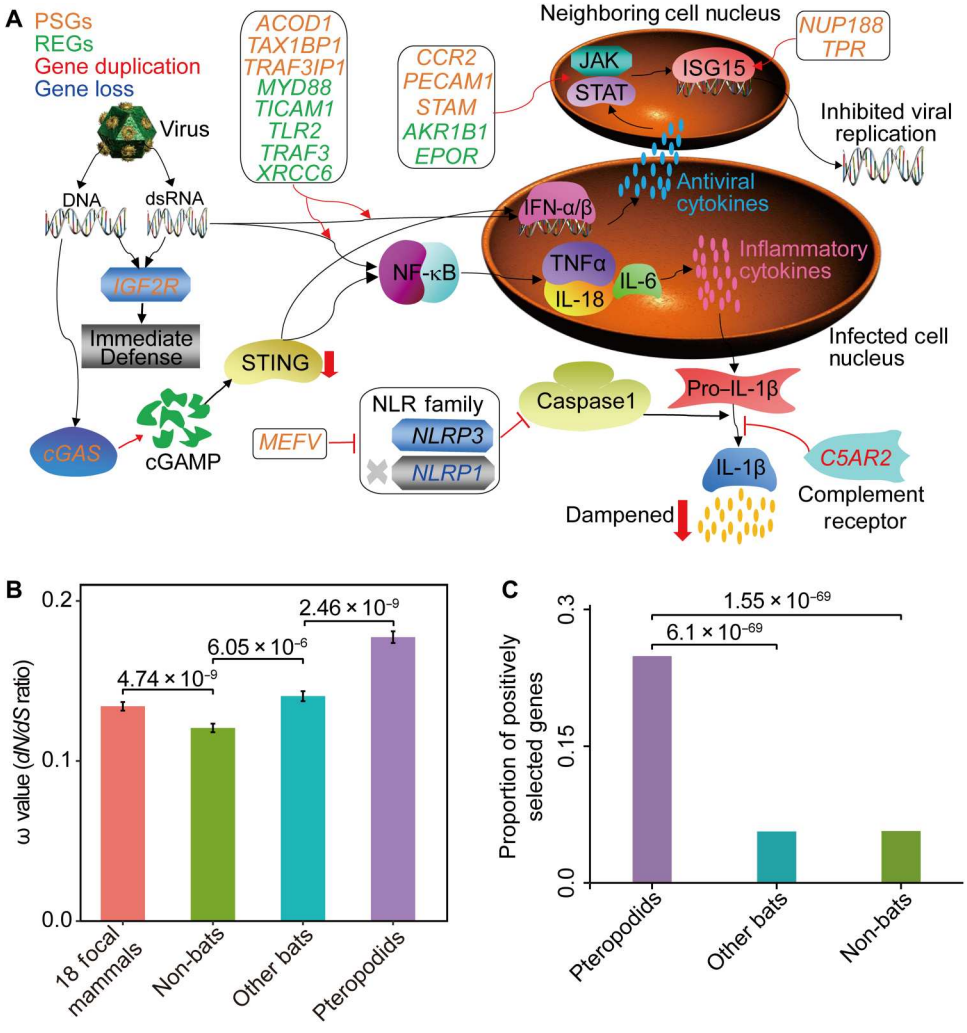

Fig. 4. Pteropodid-specific molecular adaptations of immune genes. (**A**) Schematic diagram showing pteropodid-specific PSGs (in orange), REGs (in green), gene duplication (in red), and gene loss (in blue). The thick red arrows indicate dampened expression. dsRNA, double-stranded RNA; NF-κB, nuclear factor κB. (**B**) Histogram plot showing the highest ω (means ± SE) in pteropodids as compared to other bats or mammals. Statistical significances were determined by Student's *t* test. (**C**) Histogram plot showing more genes undergoing positive selection in pteropodids than other bats or non-bat mammals. Statistical significances were determined by Fisher's exact test.

DNA and is a key component of the cGAS-STING pathway (Fig. 4A), forming part of the innate immune response (*61*). Of 16 sites found to be under positive selection in cGAS (bayes empirical bayes probability > 0.95), 14 were located in the C-terminal containing the catalytic domain (table S31). Mutations in the C terminus can disrupt the cGAMP production (*62*), and it is noteworthy that three such residues (K282, K355, and K414) have previously been shown to affect the DNA binding activity of cGAS, with the first two potentially reducing enzyme activity by more than 40% (*62*, *63*), and cGAS acetylation of K414 resulting in enzymatic inactivation (*64*). Reduced cGAS activity leads to lower levels of cGAMP, leading to dampened activation of the cGAS-STING pathway (*65*). Our finding that putative molecular adaptations could alter DNA binding thus provides further insights into possible routes by which pteropodids might attenuate inflammatory responses to viruses.

Previous studies have also shown that a region of the N terminus of cGAS, spanning residues 120 to 160, is critical for antiviral innate immune activation (*66*). In addition to containing only two sites under selection, it did show numerous amino acid replacements in pteropodids. To test whether these pteropodid-specific replacements are concentrated in this genic region, we aligned orthologs for 11 bats (6 pteropodids and 5 other bats) and compared these to other mammals. Of 35 amino acid replacements observed in at least one pteropodid but not in other bats (figs. S15 and S16), nine were located in this stretch of amino acids, revealing a disproportionately high concentration compared to whole gene ($P < 0.001$, chi-square test). To verify this pattern, we examined an expanded set of 41 bat sequences (11 pteropodids and 30 other bats) and confirmed that the nine replacements are unique to members of Pteropodidae (fig. S17). Specific substitutions of interest include P142R and P142H, representing a physicochemical shift to positively charged residues. A previous study on the N-terminal α-helical extension of the human cGAS has shown that introducing residues with negative and neutral charges reduced the detection of cytosolic DNA (*67*). Thus, we hypothesize that the molecular changes of cGAS observed in our study might alter protein-DNA interactions to enhance sensing of viral DNA in pteropodids, which should be tested in the future.

To test whether Pteropodidae shows higher rates of molecular adaptation in immune-related genes than other bats and other mammals, we examined 2162 orthologs (1203 innate immunity and 1040 adaptive immunity, with 81 shared) that belong to the "immune system process" and "viral process" categories in the KOBAS 3.0 database. We first compared gene-wise ω values estimated separately under the one-ratio model (model 0) in PAML (*52*) for three groups of taxa (pteropodids, other bats, and non-bats) and found that the mean ω was significantly higher in pteropodids than in the other bats ($P = 2.46 \times 10^{-9}$, Student's *t* test) (Fig. 4B). We then performed an explicit site model test of positive selection for each locus by comparing model M8a, in which ω = 1, to model M8, in which ω > 1 (*68*). After correcting for the false discovery rate (FDR), we found that more genes showed evidence of being under positive selection in Pteropodidae ($n = 538$) than in other bats ($n = 122$) ($P = 6.1 \times 10^{-69}$, Fisher's exact test) (Fig. 4C), with only 44 shared genes (2.04% of 2162 genes).

MyD88 mutations inhibit TLR2-dependent inflammatory activation in pteropodids

Despite growing numbers of reports of putative molecular adaptations underlying immunity in bats, few studies have experimentally validated the phenotypic impacts of bat-specific variants (*12*, *18*, *69*). In a notable exception, Xie *et al.* (*13*) showed that dampened IFN activation in bats was attributable to the replacement of a conserved S358 residue in the bat STING protein, a key component of several DNA sensing pathways. STING has a role in regulating the NLRP3 inflammasome, and Ahn *et al.* (*10*) independently used cell-based assays to show that major bat isoforms of *NLRP3* are associated with reduced induction by TLR stimuli. We screened alignments of 2162 orthologs aforementioned and identified 628 loci that contained fixed amino acid differences between pteropodids and others bats, wherein at least one fixed difference was associated with a shift in physicochemical property. The mean ω of these genes showed significantly higher than those of other 1534 immune genes ($P < 0.001$, Student's *t* test) (fig. S18), indicating that higher rates of molecular adaptation are more likely to lead to changes in the physicochemical pro4perties of amino acids. To identify candidates for functional assays in our study, we further screened out 36 loci (16 and 23, respectively) in the PSGs and/or REGs (fig. S19). Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of these loci showed that three genes (*MyD88*, *TICAM1*, and *IRAK4*) are involved in the KEGG pathway "TLR signaling pathway" (hsa04620) (table S32). For this reason, we focused on myeloid differentiation primary response 88 (MyD88), a key component of the MyD88 pathway and a canonical adaptor molecule in TLR signaling. Most TLRs have been shown to depend on MyD88 for activation, leading to the subsequent downstream release of inflammatory cytokines (type I IFN, *TNF*α, *IL6*, and *IL8*) as part of the antiviral innate immune response (*70*). Note that one of these TLRs (*30*), TLR2, was identified as an REG (table S30) and could sense diverse microbial ligands after infection (*71*). Previous work on human sequences identified four residues in MyD88 (P169, S170, A208, and Q237) that interface with TLR2 in a TLR2-MyD88 complex (Fig. 5A) (*72*). Of these, three showed specific substitutions in pteropodids that are predicted to alter physicochemical properties; specifically, positive charge (S170R), hydrophilic affinity (A208T), and negative charge (Q237E) (Fig. 5B), and one was specific to all bats but was not seen in non-bats (P169S). Using a coimmunoprecipitation (co-IP) approach, we found that pteropodid-specific changes in MyD88 reduced its binding affinity with TLR2 when coexpressed in bat *P. alecto* kidney-derived PaKi cells (Fig. 5C) (*73*).

To test the impacts of four of these putative adaptive residues in MyD88 on the production of inflammatory cytokines, we expressed human and pteropodid transgenes containing these four residue mutations in the cotransfected human PEAKrapid and bat PaKi cells, respectively. Synthetic MyD88 mutants differed with respect to whether they contained the four mutations of human-specific residues ("pMyD88-MUT") or none (pMyD88-WT) in pteropodids, and the four mutations of pteropodid-specific residues ("hMyD88-MUT") or none (hMyD88-WT) in humans. We cotransfected PaKi cells with a eukaryotic expression plasmid containing the TLR2 gene alongside a second plasmid containing the pMyD88-MUT, the pMyD88-WT, or an empty plasmid and cotransfected PEAKrapid cells with the hMyD88-MUT and hMyD88-WT. Following exposure of the PaKi and PEAKrapid

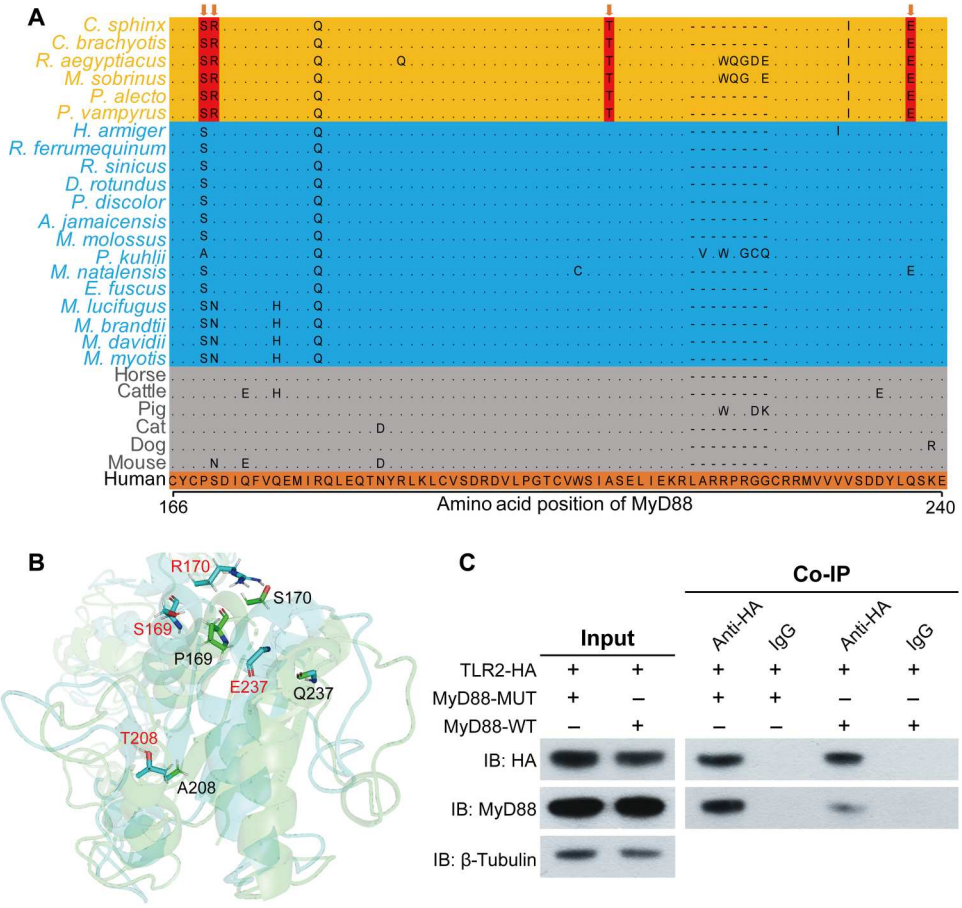

Fig. 5. Evidence of pteropodid-specific changes in MyD88 for reducing its binding affinity with TLR2. (**A**) Alignment of MyD88 protein sequences from pteropodids, other bats, and non-bat mammals. Dots indicate identical amino acids to the final sequence (human), and dashes denote alignment gaps. Four interfacing residues between the TLR2-MyD88 complexes are shaded in red. (**B**) Comparison of MyD88 protein structural models between pteropodids (blue) and humans (green). Four mutated residues of MyD88 in pteropodids are shown in red, whereas those in humans are shown in black. Structural models of MyD88 proteins were predicted by I-TASSER v.5.1 (*120*). (**C**) Immunoprecipitates and total lysates from bat Paki cells cotransfected with two combinations of plasmids: *P. alecto* TLR2/MyD88-MUT and *P. alecto* TLR2/MyD88-WT. Pteropodid-specific changes in MyD88 were shown to reduce its binding affinity with TLR2. β-Tubulin is a housekeeping gene used as an internal control; HA denotes the hemagglutinin tag that is a 9–amino acid–long peptide corresponding to residues 98 to 106 of the human influenza HA molecule. Anti-HA, HA-tag antibody; IgG, nonspecific immunoglobulin G antibody; input, total lysates for immunoblotting; IB, immunoblotting.

cells to the TLR2 agonist Pam_3CSK_4, we conducted an immunofluorescence experiment to confirm the cellular localization consistency of the overexpressed exogenous MyD88 and TLR2 proteins (fig. S20). Our prior experiments (see Materials and Methods) and Western blot analysis also showed that exogenous MyD88 had an absolute advantage in the competitive binding of the TLR2 protein compared to endogenous MyD88 (fig. S21). This ensured that our understanding of MyD88-dependent signaling would not be skewed by overexpression of endogenous MyD88.

Subsequently, we first determined expression changes of four key genes (i.e., *NFκB1*, *TAK1*, *IRAK4*, and *TRAF6*) in the TLR2 pathway (*74*). We observed the three key genes (*NFκB1*, *TAK1*, and *IRAK4*) to show similar levels of expression before and after stimulating by Pam_3CSK_4, whereas *TRAF6* expression was not detected (fig. S22). This observation indicates that stimulation by Pam_3CSK_4 would not change the expression of signaling molecules in TLR2 signaling pathway, suggesting that expression changes of inflammatory genes in the TLR2-mediated downstream signaling pathway cannot be attributed to expression changes of signaling molecules. We next observed a marked increase in expression levels of three inflammatory cytokines (*IL8*, *TNF*α, and *IL6*) in bat cells harboring MyD88 transgenes in contrast to empty plasmids cells, and interleukin-18 (IL-8) and tumor necrosis factor–α (TNFα) remain the same patterns of expression changes in the human cells (Fig. 6A), indicating that MyD88 transgenes could work in the TLR2-dependent pathway in both bat Paki cells and human PEAKrapid cells. We also conducted an additional experiment to serve as a negative test to determine whether MyD88 mutants affect other TLRs, such as TLR8, which has been reported to evolve species-specific adaptations in bats (*69*). We used the TLR8 agonist R848 to stimulate the cotransfected PaKi cells with a plasmid containing the TLR8 gene alongside a second plasmid containing the pMyD88-MUT, the pMyD88-WT, or an empty plasmid. The results showed no significant difference in the expression level of inflammatory cytokines between pMyD88-MUT and pMyD88-WT cells (fig. S23), indicating that MyD88 with pteropodid-specific

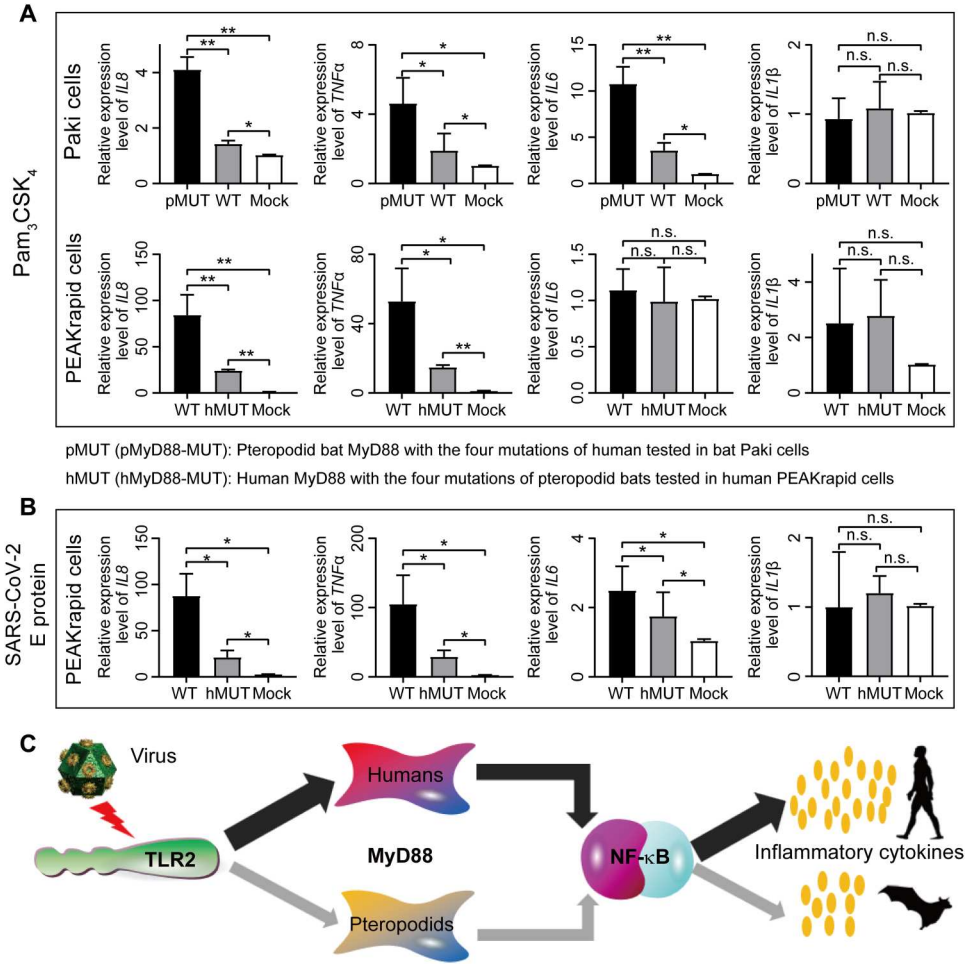

Fig. 6. Functional evidence of MyD88 for inhibiting excessive inflammatory activation in pteropodids. (**A**) Comparisons of relative expression levels of inflammatory cytokines between MyD88-MUT (MUT) variants and MyD88-WT (WT) in PaKi and PEAKrapid cells, respectively, when stimulated by the TLR2 agonist Pam_3CSK_4. Mutants contained the four mutations of human-specific residues (pMyD88-MUT) in PaKi cells and the four mutations of pteropodid-specific residues (hMyD88-MUT) in PEAKrapid cells. Expression levels in PaKi and PEAKrapid cells with the empty plasmid (mock) were used for normalization, respectively. The data of three independent experiments are expressed as the means ± SE by Student's *t* test (**$0.001 \leq P < 0.01$; *$0.01 \leq P < 0.05$; n.s., not significant). (**B**) Comparisons of relative expression levels of inflammatory cytokines between hMyD88-MUT variants and hMyD88-WT cells of PEAKrapid, when stimulated by the E protein of SARS-CoV-2. Expression levels in PEAKrapid cells with the empty plasmid (mock; as control) were used for normalization. Data for three independent experiments expressed as the means ± SE with significance estimated by the Student's *t* test (*$0.01 \leq P < 0.05$). (**C**) Schematic diagram showing TLR2-MyD88 signaling pathway.

residues did not alter TLR8-mediated inflammatory response in pteropodids. Notably, expression levels of MyD88 transgenes with pteropodid-specific residues in both bat Paki and human PEAKrapid cells were significantly lower than those with human residues (Fig. 6A). Therefore, MyD88 with pteropodid-specific residues could dampen TLR2-dependent inflammatory response in both human and bat cells.

To evaluate the effect of MyD88 mutants in the TLR2 pathway during virus infection, we stimulated the human PEAKrapid cells with E protein of SARS-CoV-2, which was proved to induce TLR2-dependent inflammation (*30*). We observed a marked decrease in levels of inflammatory cytokines (i.e., *IL8*, *IL6*, and *TNFα*) in hMyD88-MUT (human MyD88 with four pteropodid-specific residues) in contrast to in hMyD88-WT (human MyD88) cells (Fig. 6B). Thus, MyD88 proteins with pteropodid-specific residues can dampen inflammatory response during virus infection.

Together, proteins with pteropodid-specific replacements were associated with a dampened inflammatory response, consistent with an attenuated ability of TLR2 to bind MyD88, leading to a reduced induction of inflammatory cytokines (Fig. 6C). Such a result is consistent with data from STING (*13*) and highlights the diversity of pathways by which bats dampen their immune and inflammatory responses to microbial infection.

In summary, we developed a new assembly pipeline to obtain a near-complete genome of *C. sphinx*, a widespread species of Old World fruit bat (family Pteropodidae). Using this assembly as a reference for comparative genome analyses, we show that pteropodids have undergone pervasive molecular evolutionary changes in immune genes. In particular, we uncover genetic evidence of molecular adaptations in members of the inflammatory signaling pathway and show that pteropodid-specific changes in the adapter molecule MyD88 reduce its binding affinity with TLR2, consistent

with attenuated downstream activation. These results add to our understanding of the specific derived mechanisms that have enabled Old World fruit bats to act as natural reservoir hosts for viruses that cause disease in other species. In addition, we also demonstrate the potential for the short-nosed fruit bat as a model for future studies of immunity; thus, its reference-quality genome will help establish captive breeding colonies for in vivo bat immunological studies (*75*). Future scrutiny of immunological functions in Old World fruit bats may help explain why they show distinct evolution of immunity compared to other bats.

MATERIALS AND METHODS

Genome sequencing

Genomic DNA was extracted from the liver tissue of the greater short-nosed fruit bat (*C. sphinx*) by a QIAGEN Genomic DNA extraction kit (catalog no. 13323) according to the standard operating procedure provided by the manufacturer. Field sampling of one individual of this fruit bat was approved by the Institute of Zoology Guangdong Academy of Sciences (GIABR2020810). To obtain high-quality DNA, we strictly performed the following four steps: (i) We ensured that there were no observable pollutants in the sample; (ii) degraded samples were excluded by 0.75% agarose electrophoresis; (iii) DNA purity detection was performed using a NanoDrop One UV-Vis spectrophotometer (Thermo Fisher Scientific, USA) to ensure that the optical density at 260/280 nm (OD260/280) ratio ranged from 1.8 to 2.0 and that the OD260/230 ratio was between 1.9 and 2.4; and (iv) we ensured that the concentration of DNA was more than 40 ng/μl using a Qubit 3.0 fluorometer (Invitrogen, USA). The high-quality DNA extracts were used for constructing a Nanopore library and an Illumina short-insert library [500 base pairs (bp)]. For Nanopore sequencing, the Nanopore library was first added to the flow cell via the sample port in a dropwise fashion. Then, the flow cell was transferred into a Nanopore PromethION sequencer (Oxford Nanopore Technologies, UK) for real-time single-molecule sequencing. Last, the high-quality Nanopore reads were obtained after filtering the low-quality reads with less than Q7 (Phred quality score). For Illumina pair-end sequencing, a total of 1.5 μg of genomic DNA was used as input material to construct one short-insert library using the TruSeq Nano DNA HT Sample Preparation Kit (Illumina, San Diego, CA, USA) according to the manufacturer′s instructions. We generated raw sequences based on the Illumina NovaSeq platform, which were subsequently preformed quality control to remove low-quality reads by the fastp software (*76*).

Hi-C libraries were constructed from liver cells of an individual of *C. sphinx*. First, the cells were fixed with formaldehyde and lysed them. Second, the cross-linked DNA was digested with Dpn II, and the sticky ends were biotinylated and proximity-ligated to form chimeric junctions. Third, 300- to 500-bp DNA fragments were enriched after physical shearing. The chimeric fragments representing the original cross-linked long-distance physical interactions were processed into paired-end sequencing libraries. Last, the Hi-C libraries were sequenced on the Illumina NovaSeq platform and generated the paired-end reads. The quality of Hi-C raw data was initially checked using the fastp software (*76*). The low-quality Illumina reads that contained base-calling duplicates and adaptor contamination were removed on the basis of the following parameters: reads with ≥10% unidentified nucleotides; >10 nt aligned to the adaptor, allowing ≤10% mismatches; >40% bases having a Phred quality of <15; and putative polymerase chain reaction (PCR) duplicates generated in the library construction process.

Genome size estimation

The genome size was estimated on the basis of *k*-mer spectrum analysis. Specifically, the high-quality paired-end reads were counted by 17-mers using Jellyfish (v2.3.0) (*77*). The *k*-mers with a low frequency of 3 were possibly caused by sequencing errors and were removed. Last, we obtained the total *k*-mers and their coverage depth, and the total *k*-mers divided by their coverage depth were the value of the genome size.

A new assembly pipeline

Long-read sequencing technology typically results in a better and more reliable assembly, notably by preventing errors due to duplicated or repetitive sequences in the same region (*78*), but it cannot handle long-distance sequence repetition (*33*). Hi-C technology can recognize the stronger interaction of the proximal region over the weaker interaction of long-distance genomic regions, thereby removing the noise due to the linkage of long-distance repeats with a low contact frequency and effectively reducing the chimeric linkage between homologous sequences in the genome (*79*). Therefore, we can optimize the assembly pipeline to obtain a more complete genome assembly by removing the noise due to the linkage of nonproximal repeats with low frequency. We designed a new assembly pipeline to obtain an extremely complete genome by taking advantage of this feature of Hi-C reads. The pipeline consists of four main steps:

1) Assemble and correct contigs: We used the high-quality Nanopore sequences to assemble initial contigs of *C. sphinx* by applying a "correct-then-assemble" strategy from the package NextDenovo v2.4.0 (https://github.com/Nextomics/NextDenovo) with the parameters of "read_cutoff = 1k, seed_cutoff = 32k, blocksize = 3g." Subsequently, the initial contigs were corrected using the package NextPolish v1.3.1 (*80*) with the Illumina paired-end reads and Nanopore sequences with the "best" algorithm module. On the basis of the above analytical strategies, we obtained all contigs, namely, the Contig v1 assembly genome.

2) Cluster contigs by Hi-C interaction pair: The Hi-C read pairs were mapped to Contig v1 using the Bowtie2 software (*81*) with a single-ended model. We obtained 402,709,409 (70.14%) unique mapped pairs and then discarded the invalid self-ligated and unligated fragments using the HiCUP pipeline (version 0.8.0) (*82*). Last, we used the valid interaction pairs to calculate the linkage frequency among all contigs using an agglomerative hierarchical clustering algorithm (*79*). We clustered the linked contigs based on the linkage suggested by Hi-C signal density, presumably along the same chromosome according to a preset number of partitions.

3) Classify the Nanopore reads to perform local assembly: We realigned the Nanopore reads to Contig v1 using the Minimap2 package (*83*). After removing the suboptimal alignment reads, we extracted the mapped reads of each linked contig group. Subsequently, we performed local assembly for each classified mapped read. Compared to global assembly, the strategy could avoid the false overlap relationships induced by repetitive sequences of the genome when constructing the string graph during assembly (*33*). We set an appropriate parameter "seed_cutoff" for each local assembly according to the calculated results by the command

"seq_stat." Then, the assembly and correction method in this step was similar to that mentioned in the step (1). Last, we obtained a second set of all contigs, namely, Contig v2.

4) Anchor contigs onto chromosomes: Chromosome-scale genome was anchored by linkage information, restriction enzyme site, and string graph formulation using the algorithm ALLHiC (*79*). The placement and orientation errors exhibiting obvious discrete chromatin interaction patterns were manually adjusted.

Genome assessment

A variety of methods were applied to assess the quality of the assembled genome. To validate the single-base accuracy, the package Merqury was used for assessing the quality of genome assemblies using a reference-free and *k*-mer–based approach (*35*). To assess the completeness of the genome, we estimated the realignment ratio and coverage rate by realigning the Illumina pairs to the assembly genome with Burrows-Wheeler Aligner software (*84*), and we also performed BUSCO analysis to assess the completeness of the genome by searching against 4104 mammalian BUSCO genes (version 5.2.1) (*34*).

TE annotation

To avoid systematic biases related to the use of different methods for the annotation of TEs in previously published bat genomes, we predicted the TEs for all bat genomes using the same method. We identified TEs by a combination of homology searching and ab initio prediction. For homology-based prediction, we used RepeatMasker (*85*) and RepeatProteinMask to search against the Repbase TE library with default parameters. For ab initio prediction, we first built a reference repeat library based on the results from LTR FINDER (*86*), PILER (*87*), RepeatScout (*88*), and RepeatModeler with default parameters and then used RepeatMasker to search against this library. In addition, the package Tandem Repeats Finder (*89*) was used to identify tandem repeats in the genome with the parameters "2 7 7 80 10 50 2000 -d –h".

Annotation of gene structure

Gene prediction was conducted through a combination of homology- and ab initio–based methods, aided by evidence of transcription. For homology-based prediction, we created two uniform bat reference protein sets that included nonredundant proteins from 13 released bat genomes as shown in data S1. Specifically, one representative Yangochiroptera protein set consisted of protein sets from eight bats: *Myotis lucifugus* (*90*), *Myotis brandtii*, *Eptesicus fuscus*, *Miniopterus natalensis* (*91*), *Myotis davidii*, *Sturnira honurensis*, *Desmodus* rotundus, and *Artibeus jamaicensis*. The other set representing the suborder Yinpterochiroptera was composed of protein sets from five bats: *R. aegyptiacus*, *P. alecto*, *P. vampyrus*, *Hipposideros* armiger, and *Rhinolophus sinicus*. For each uniform protein set, all-against-all BLASTP (*92*) was applied to determine the similarities in bat proteins with an *E* value of 1×10^{-7} and to build a homologous gene family using the OrthoMCL method (*43*). Each homologous gene family had more than one gene. To build a uniform bat reference protein set, we selected one gene from each homologous gene family based on the contig N50 value. In addition, the lineage-specific genes of each bat genome were retained in the uniform bat reference gene set. Last, two uniform bat protein sets with 34,870 proteins from the suborder Yangochiroptera proteins and 26,897 proteins from the suborder Yinpterochiroptera were used for homology prediction. Subsequently, protein repertoires from the bat uniform reference protein set, horse, human, and mouse were used as queries to search against the target genome using TBLASTN (*93*). The BLAST hits were conjoined by the program Solar (Sorting Out Local Alignment Results). Then, the exact gene structure of the corresponding genomic region in each BLAST hit was defined for gene models using the GeneWise pipeline (*94*).

For transcript prediction, the RNA sequencing (RNA-seq) data of six bats (*C. sphinx*, *Myotis myotis*, *M. molossus*, *Phyllostomus discolor*, *Rhinolophus ferrumequinum*, and *R. aegyptiacus*) were downloaded from NCBI; the accession numbers were available in table S14. The RNA-seq data were aligned to the genome using TopHat (*95*) to identify putative exonic regions and splicing junctions. Then, gene models (Cufflinks-set) were assembled from mapped reads by Cufflinks (*96*).

For ab initio prediction, the assembled transcripts from the RNA-seq data of *C. sphinx* brain tissue were aligned against the *C. sphinx* genome using the Program to Assemble Spliced Alignment (PASA) (*97*). The transcript alignments were assembled into gene structure models by PASA and denoted as the gene model training set for the Augustus (*98*), SNAP (*99*), and GlimmerHMM (*100*) pipelines. On the basis of those training sets, we performed ab initio prediction of coding regions in the repeat-masked genome using Augustus, GlimmerHMM, and SNAP. In addition, GeneID (*101*) and GeneScan (*102*) were used to generate directly predicted gene models in the genome.

After the above methods were completed, the generated gene models were integrated by EVidenceModeler. The weights of each evidence were set as follows: Homology-set = Cufflinks-set > Augustus > GeneID = SNAP = GlimmerHMM = GeneScan. Last, the gene models were updated by PASA2 to generate untranslated regions and alternative splicing variation information.

Annotation of gene function

Gene functions were annotated by searching functional domains, motifs, and possible biological processes of genes in known databases, including Swiss-Prot (*103*), Pfam (*104*), NR database (from NCBI), GO (*105*), and KEGG (*106*).

Identification of one-to-one orthologous genes

The 18 focal mammals shown in data S1 were used to identify one-to-one orthologous genes. For genes with alternative splicing sites, we chose the longest translation to represent each gene and filtered genes with fewer than 50 codons. To build a graph of protein-coding genes, all-against-all BLASTP was used to determine similarities between all proteins in the 18 focal mammals with an *E* value of 1×10^{-7}. To identify one-to-one orthologs, we extracted alignment pairs from any pair of genomes and used a maximum of five hits per protein sequence as the input for the MCScanX algorithm (*107*), which was used to detect high-confidence syntenic blocks of coding genes and identify orthologous gene pairs. We retained only one-to-one orthologs among the two mammalian species. After integrating a matrix of orthologs for the 18 focal mammals, we ensured that all species were present in each orthologous cluster. Subsequently, multiple sequence alignment for these one-to-one orthologs was aligned by PRANK (v.170427) (*108*) combined with the package Gblocks (v0.91b) (*109*) to minimize

the impact of multiple sequence alignment errors and divergent regions. Final alignments shorter than 150 nt were discarded.

Phylogeny construction and divergence time estimation

We ran MODELTEST (*110*) using the codon alignments of one-to-one orthologs and found that the general time-reversible (GTR) substitution model best fit the observed data. On this basis, we constructed a phylogenetic tree of the 18 focal mammals by applying the maximum likelihood method in the package RaxML (*111*) with the best-fitting substitution model "GTR + GAMMAX" and 1000 bootstrap replicates. Subsequently, divergence time was estimated using the MCMCTree program as implemented in the PAML package (*52*). There were two fossil constraints incorporated into the analyses: (i) The divergence time of 61.5 to 100.5 Ma ago is for human and mouse (*112*), and (ii) the divergence time of 95.3 to 113 Ma ago is for human and horse (*112*). Two additional calibration times (i.e., the divergence time of 54 Ma ago is between pig and cattle, and that of 62 Ma ago is between dog and cat) were derived from the TimeTree database (*113*).

Genome synteny and large interchromosomal rearrangement events

We performed reciprocal alignments of 17 repeat-masked genomes, using the *C. sphinx* genome as the reference, by applying the LASTZ program (*114*) with the parameters T = 2 (no transition), Y (ydrop) = 15,000, L (gappedthresh) = 3000, and K (hspthresh) = 4500. The raw alignments were combined into larger syntenic blocks using the ChainNet algorithm. The best hit at every location on the chromosomes was chosen by the utility "axtBest" based on a dynamic programming algorithm, with the same substitution matrix adopted for alignment. Subsequently, we screened out syntenic blocks between *C. sphinx* and other query genomes with an identity of ≥90% and a length of ≥1 kb. Last, these syntenic blocks were conjoined by the program Solar with a maximum gap length of 50 kb.

An interchromosomal rearrangement event was defined when at least two syntenic blocks from different chromosomes in the query genome were found adjacent to each other on the same chromosome in the *C. sphinx* genome. A large interchromosomal rearrangement event was defined as the length of the continuous syntenic block reached 1 Mb. The large interchromosomal rearrangement events in *C. sphinx* are showed in Fig. 2B and data S2.

Dating retrovirus infection

First, we extracted the sequences from 153 *C. sphinx* and 76 *R. aegyptiacus* retroviral genes, along with their upstream and downstream 5-kb regions. To identify LTR retrotransposons (LTR-RTs) within these sequences, we used LTRharvest (v.1.5.8) (*115*) with the parameters "-motif tgca -motifmis 1" and LTR FINDER (*86*) with the parameters "-minlenltr 100 -maxlenltr 5000 -mindistltr 1000 -maxdistltr 20000." Next, we used LTRdigest (v.1.5.8) (*116*) to annotate the primer-binding site motifs. We then identified domains by searching for retrotransposons containing primer-binding site motifs, using HMM profiles collated by the Gypsy Database (http://gydb.org/). We used LTRdigest (v.1.5.8) to annotate the internal features of LTR-RTs. Last, we retained intact LTR-RTs to estimate their insertion time (T).

Second, on the basis of divergence times of the phylogenetic tree of 18 focal mammals (Fig. 2A), we calculated the substitution rate (r) using the program baseml in the PAML package under the GTR model. We selected the fourfold degenerate codon positions that form 7139 one-to-one orthologs of the 18 focal mammals for this analysis. As a result, we estimated the neutral substitution rate (r) per year to be 2.51×10^{-9} for *C. sphinx* and 2.99×10^{-9} for *R. aegyptiacus*.

Last, we calculated the divergence rate (D) of the 5he (enc LTR sequences of LTR-RTs using the MUSCLE alignments and the distmat program in EMBOSS (*117*). Upon integration into the genome, the 5he on intLTR sequences are identical. Independent mutations occur in each LTR-RT over time based on the host mutation rate. Therefore, we estimated the insertion time (T) of each LTR-RT in *C. sphinx* and *R. aegyptiacus* by applying the formula $T = D/2r$.

SD analysis

We performed self-genome alignments using the repeat-masked genome sequences with the package LASTZ. The maximum simultaneous gap allowed was 100 bp. After alignment, we removed overlapping fragments and obtained a nonredundant set. The genomic regions with alignment lengths larger than 1 kb with an identity greater than 90% were considered SDs.

Gene family clustering

Using the OrthoMCL method (*43*) with the parameter "-inflation 1.5," we conjoined fragmental alignments that are used by all-against-all BLASTP to cluster gene pairs. Consequently, we obtained gene families for the 18 mammalian species. See also the above method "Identification of one-to-one orthologous genes."

Expansion and contraction analysis

We determined the expansion and contraction of gene families by comparing the cluster size differences between the ancestor and each of the 18 mammals using the Café program (*44*), which is based on a probabilistic graphical model. Using conditional likelihoods as the test statistics, we calculated the corresponding P values for each lineage. A P value of 0.05 was used to identify families that were significantly expanded in Pteropodidae.

Identification of PSGs and REGs

We identified putative PSGs and REGs based on the 7139 orthologous genes using the CodeML program in the PAML package (version 4.9) (*52*). The Pteropodidae branch was set to the foreground branch. For PSGs, the branch-site model A was treated as an alternative model that assumes that the foreground branch is under positive selection, whereas a null model allows sites to undergo purifying selection or evolve neutrally. For REGs, we applied the branch model; specifically, the one-ratio model was taken as the null model that assumes the same evolutionary rate for all branches, and the two-ratio model was used as an alternative model that allows a different evolutionary rate for the foreground branch. The likelihood ratio test was used to detect the difference between the two nested models. P values were computed on the basis of chi-square statistics and corrected for multiple testing by the FDR method.

Functional enrichment analysis

The GO, KEGG pathway, and Reactome enrichment analyses for gene sets were all implemented by KOBAS 3.0 software based on human homologs (*118*, *119*). The binomial distribution test was

used for statistical significance, and the *P* values were corrected for multiple testing by the Benjamini-Hochberg method.

Identification of immune genes

We selected 3444 genes (including 3039 involved in immune system processes and 1055 viral processes) from the KOBAS 3.0 database (*118*). The immune system–related genes included 1203 innate immunity genes and 1040 adaptive immunity genes. The query sequences of each gene in each species were blasted (BLASTX) against the 3444 proteins with an *E* value of 1×10^{-10} and an identity of 0.25, and then the best-hit query sequences with more than 50 amino acids in length were identified. In total, 2162 orthologs were retrieved from the 18 species, namely, immunity-related genes.

Physicochemical dissimilarity analysis of amino acids

We coded all amino acids based on four physicochemical properties: hydrophobic, neutral, positively charged, and negatively charged. For each ortholog, we assessed the physicochemical dissimilarity due to amino acid replacements between pteropodids and other bats using the human amino acid as a reference. We first identified fixed amino acid sites in six pteropodids and five other bats. Then, we selected the amino acid sites with differences in physicochemical codes between pteropodids and other bats, as well as pteropodids and humans. Last, we annotated these amino acid sites into genes.

Plasmids and cell cultures

Human PEAKRapid and bat *P. alecto* kidney-derived PaKi cells from our laboratory were cultured in Dulbecco's modified Eagle's medium (Gibco Inc.) supplemented with 10% fetal bovine serum (Wisent Inc.), 2.0 mM L-glutamine, sodium pyruvate (110 mg/liter), and D-glucose (4.5 g/liter). Cells were incubated at 37°C in 5% CO_2. The pcDNA3.1-TLR2, pcDNA3.1-TLR8, and pcDNA3.1-MYD88-WT plasmids were synthesized by Sangon Biotech (Shanghai) Co. Ltd. The empty plasmid was from our laboratory. The mutated form of the pcDNA3.1-MYD88-WT plasmid, pcDNA3.1-MYD88-MUT, which results from four mutations (P169S, S170N, A208T, and Q237E), was obtained using site-directed mutagenesis with the following primers: (i) P169S, 5′-TTC-ATC-TGC-TAC-TGC-TCT-AGC-GAC-ATT-CAG-T-3′ (forward) and 5′-ACT-GAA-TGT-CGC-TAG-AGC-AGT-AGC-AGA-TGA-A-3′ (reverse); (ii) S170N, 5′-TCT-GCT-ACT-GCC-CTA-ACG-ACA-TTC-AGT-TCG-TG-3′ (forward) and 5′-CAC-GAA-CTG-AAT-GTC-GTT-AGG-GCA-GTA-GCA-GA-3′ (reverse); (iii) A208T, 5′-TGC-GTG-TGG-AGC-ATC-ACT-AGC-GAG-CTG-ATC-G-3′ (forward) and 5′-CGA-TCA-GCT-CGC-TAG-TGA-TGC-TCC-ACA-CGC-A-3′ (reverse); and (iv) Q237E, 5′-AGC-GAC-GAC-TAC-CTG-GAG-AGC-AAG-GAG-TGC-G-3′ (forward) and 5′-CGC-ACT-CCT-TGC-TCT-CCA-GGT-AGT-CGT-CGC-T-3′ (reverse). All plasmids were verified by sequencing at Sangon Biotech (Shanghai) Co. Ltd.

Transfection, cell stimulation, and quantitative real-time PCR

Before transient transfection, cells were cultured overnight in 12-well plates. Subsequently, PEAKRapid and PaKi cells were cotransfected with the following the combinations of plasmids: pcDNA3.1(+)-TLR2/pcDNA3.1(+)-MyD88-WT, pcDNA3.1(+)-TLR2/pcDNA3.1(+)-MyD88-MUT, and pcDNA3.1(+)-TLR2/empty plasmid, using the Lipofectamine 2000 (Thermo Fisher Scientific) transfection protocol as recommended by the manufacturer and then cultured for 24 hours. We first conducted prior experiments to optimize the concentrations of TLR2 and MyD88 plasmids and the stimuli. As a result, PEAKRapid cells and PaKi cells were stimulated with TLR2 agonist Pam_3CSK_4 (5 and 10 μg/ml; Target Molecule Corp.) for 24 hours, respectively. In addition, PEAKRapid cells were also stimulated by E protein (5 μg/ml) of SARS-CoV-2 for 24 hours. Prior experiments showed that bat cells expressed no endogenous MyD88 mRNA, whereas the mRNA expression level of exogenous MyD88 in human cells was over 223 times higher than that of endogenous MyD88 (MyD88-WT, 223.85 times; MyD88-MUT, 263.89 times). We also performed an immunofluorescence experiment to demonstrate the cellular localization consistency of overexpressed exogenous MyD88 and TLR2 proteins between PaKi and PEAKrapid cells. Last, we cotransfected PaKi cells with the following the combinations of plasmids: pcDNA3.1(+)-TLR8/pcDNA3.1(+)-MyD88-WT, pcDNA3.1(+)-TLR8/pcDNA3.1(+)-MyD88-MUT, and pcDNA3.1(+)-TLR8/empty plasmid and then were stimulated TLR8 agonist R848 (10 μg/ml).

Total RNA was extracted using EASYspin kits (Aidlab Biotechnologies Co. Ltd.) according to the manufacturer's protocols. cDNA was prepared from 1 μg of total RNA using the reverse transcription system from the NovoScript Plus All-in-One Strand cDNA Synthesis SuperMix from Novoprotein Scientific Co. Ltd. The expression levels of multiple inflammatory cytokines (i.e., *IL8*, *IL6*, *TNF*α, and IL1β) were normalized to those of glyceraldehyde-3-phosphate dehydrogenase, which is a housekeeping gene used as an endogenous control. All primers were synthesized by Sangon Biotech (Shanghai) Co. Ltd.

We performed quantitative real-time PCR (qRT-PCR) on the ABI7500 Real-Time PCR Detection System (Applied Biosystems) using NovoStart SYBR qPCR SuperMix Plus (Novoprotein Scientific, Shanghai, China) following the manufacturer's protocols. The qRT-PCR measurements were evaluated in five replicates. The fold change in the expression level was determined using the $2^{-\Delta\Delta Ct}$ method. Data were analyzed by GraphPad Prism software 7.0 (GraphPad, USA), and $P < 0.05$ was considered statistically significant.

Co-IP assay

We used the immunoprecipitation kit with protein A + G magnetic beads (Beyotime Biotechnology, Shanghai, China) to perform a co-IP assay. First, plasmids expressing the hemagglutinin (HA)–tagged *P. alecto* TLR2 were cotransfected with either *P. alecto* MyD88-MUT or *P. alecto* MyD88-WT into PaKi cells for 48 hours. Second, cell lysates were generated by lysing cells in lysis buffer containing protease inhibitor cocktail and were centrifuged at 12,000 rpm under 4°C for 5 min. Third, we collected the supernatant and incubated it with 12 μl of protein A + G magnetic beads bound to the antibody. Fourth, following overnight incubation, we washed the beads with lysis buffer three times and then collected by magnetic separation. Last, we harvested samples after incubating in 1× SDS loading buffer at 100°C for 5 min. Electrophoresis was conducted to separate proteins in a 4 to 12% polyacrylamide gel. Proteins were transferred onto polyvinylidene difluoride membranes and blocked with 5% skim milk for 1 hour at 37°C. All the primary antibodies were incubated at 37°C for 1.5 hours, while secondary antibodies with horseradish peroxidase were incubated for 1

hour at 37°C. Bound antibodies were visualized using Western blots.

Supplementary Materials

This PDF file includes:
Figs. S1 to S23
Tables S1 to S32
Legends for data S1 to S4

Other Supplementary Material for this manuscript includes the following:
Data S1 to S4

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. D. J. Middleton, C. J. Morrissy, B. M. van der Heide, G. M. Russell, M. A. Braun, H. A. Westbury, K. Halpin, P. W. Daniels, Experimental Nipah virus infection in pteropid bats (*Pteropus poliocephalus*). *J. Comp. Pathol.* **136**, 266–272 (2007).
2. K. Halpin, A. D. Hyatt, R. Fogarty, D. Middleton, J. Bingham, J. H. Epstein, S. A. Rahman, T. Hughes, C. Smith, H. E. Field, P. Daszak; Henipavirus Ecology Research Group, Pteropid bats are confirmed as the reservoir hosts of henipaviruses: A comprehensive experimental study of virus transmission. *Am. J. Trop. Med. Hyg.* **85**, 946–951 (2011).
3. J. T. Paweska, N. Storm, A. A. Grobbelaar, W. Markotter, A. Kemp, P. J. van Vuren, Experimental inoculation of Egyptian fruit bats (*Rousettus aegyptiacus*) with Ebola virus. *Viruses* **8**, 29 (2016).
4. A. J. Schuh, B. R. Amman, T. K. Sealy, J. R. Spengler, S. T. Nichol, J. S. Towner, Egyptian rousette bats maintain long-term protective immunity against Marburg virus infection despite diminished antibody levels. *Sci. Rep.* **7**, 8763 (2017).
5. V. J. Munster, D. R. Adney, N. van Doremalen, V. R. Brown, K. L. Miazgowicz, S. Milne-Price, T. Bushmaker, R. Rosenke, D. Scott, A. Hawkinson, E. de Wit, T. Schountz, R. A. Bowen, Replication and shedding of MERS-CoV in Jamaican fruit bats (*Artibeus jamaicensis*). *Sci. Rep.* **6**, 21878 (2016).
6. W. Li, Z. Shi, M. Yu, W. Ren, C. Smith, J. H. Epstein, H. Wang, G. Crameri, Z. Hu, H. Zhang, J. Zhang, J. McEachern, H. Field, P. Daszak, B. T. Eaton, S. Zhang, L.-F. Wang, Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676–679 (2005).
7. P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, Z.-L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
8. A. Banerjee, M. L. Baker, K. Kulcsar, V. Misra, R. Plowright, K. Mossman, Novel insights into immune systems of bats. *Front. Immunol.* **11**, 26 (2020).
9. J. C. Guito, J. B. Prescott, C. E. Arnold, B. R. Amman, A. J. Schuh, J. R. Spengler, T. K. Sealy, J. R. Harmon, J. A. D. Coleman-McCray, K. A. Kulcsar, E. R. Nagle, R. Kumar, G. F. Palacios, M. Sanchez-Lockhart, J. S. Towner, Asymptomatic infection of Marburg virus reservoir bats is explained by a strategy of immunoprotective disease tolerance. *Curr. Biol.* **31**, 257–270.e5 (2021).
10. M. Ahn, D. E. Anderson, Q. Zhang, C. W. Tan, B. L. Lim, K. Luko, M. Wen, W. N. Chia, S. Mani, L. C. Wang, J. H. J. Ng, R. M. Sobota, C.-A. Dutertre, F. Ginhoux, Z.-L. Shi, A. T. Irving, L.-F. Wang, Dampened NLRP3-mediated inflammation in bats and implications for a special viral reservoir host. *Nat. Microbiol.* **4**, 789–799 (2019).
11. J. Gu, C. Korteweg, Pathology and pathogenesis of severe acute respiratory syndrome. *Am. J. Pathol.* **170**, 1136–1147 (2007).
12. G. Zhang, C. Cowled, Z. Shi, Z. Huang, K. A. Bishop-Lilly, X. Fang, J. W. Wynne, Z. Xiong, M. L. Baker, W. Zhao, M. Tachedjian, Y. Zhu, P. Zhou, X. Jiang, J. Ng, L. Yang, L. Wu, J. Xiao, Y. Feng, Y. Chen, X. Sun, Y. Zhang, G. A. Marsh, G. Crameri, C. C. Broder, K. G. Frey, L.-F. Wang, J. Wang, Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science* **339**, 456–460 (2013).
13. J. Xie, Y. Li, X. Shen, G. Goh, Y. Zhu, J. Cui, L.-F. Wang, Z.-L. Shi, P. Zhou, Dampened STING-dependent interferon activation in bats. *Cell Host Microbe* **23**, 297–301.e4 (2018).
14. P. Zhou, M. Tachedjian, J. W. Wynne, V. Boyd, J. Cui, I. Smith, C. Cowled, J. H. J. Ng, L. Mok, W. P. Michalski, I. H. Mendenhall, G. Tachedjian, L.-F. Wang, M. L. Baker, Contraction of the type I IFN locus and unusual constitutive expression of IFN-α in bats. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2696–2701 (2016).
15. D. Jebb, Z. Huang, M. Pippel, G. M. Hughes, K. Lavrichenko, P. Devanna, S. Winkler, L. S. Jermiin, E. C. Skirmuntt, A. Katzourakis, L. Burkitt-Gray, D. A. Ray, K. A. M. Sullivan, J. G. Roscito, B. M. Kirilenko, L. M. Dávalos, A. P. Corthals, M. L. Power, G. Jones, R. D. Ransome, D. K. N. Dechmann, A. G. Locatelli, S. J. Puechmaille, O. Fedrigo, E. D. Jarvis, M. Hiller, S. C. Vernes, E. W. Myers, E. C. Teeling, Six reference-quality genomes reveal evolution of bat adaptations. *Nature* **583**, 578–584 (2020).
16. I. Seim, X. Fang, Z. Xiong, A. V. Lobanov, Z. Huang, S. Ma, Y. Feng, A. A. Turanov, Y. Zhu, T. L. Lenz, M. V. Gerashchenko, D. Fan, S. Hee Yim, X. Yao, D. Jordan, Y. Xiong, Y. Ma, A. N. Lyapunov, G. Chen, O. I. Kulakova, Y. Sun, S.-G. Lee, R. T. Bronson, A. A. Moskalev, S. R. Sunyaev, G. Zhang, A. Krogh, J. Wang, V. N. Gladyshev, Genome analysis reveals insights into physiology and longevity of the Brandt's bat *Myotis brandtii*. *Nat. Commun.* **4**, 2212 (2013).
17. K. Wang, S. Tian, J. Galindo-González, L. M. Dávalos, Y. Zhang, H. Zhao, Molecular adaptation and convergent evolution of frugivory in Old World and neotropical fruit bats. *Mol. Ecol.* **29**, 4366–4381 (2020).
18. M. L. Zepeda Mendoza, Z. Xiong, M. Escalera-Zamudio, A. K. Runge, J. Thézé, D. Streicker, H. K. Frank, E. Loza-Rubio, S. Liu, O. A. Ryder, J. A. Samaniego Castruita, A. Katzourakis, G. Pacheco, B. Taboada, U. Löber, O. G. Pybus, Y. Li, E. Rojas-Anaya, K. Bohmann, A. Carmona Baez, C. F. Arias, S. Liu, A. D. Greenwood, M. F. Bertelsen, N. E. White, M. Bunce, G. Zhang, T. Sicheritz-Pontén, M. P. T. Gilbert, Hologenomic adaptations underlying the evolution of sanguivory in the common vampire bat. *Nat. Ecol. Evol.* **2**, 659–668 (2018).
19. B. Chattopadhyay, K. M. Garg, R. Ray, I. H. Mendenhall, F. E. Rheindt, Novel de novo genome of *Cynopterus brachyotis* reveals evolutionarily abrupt shifts in gene family composition across fruit bats. *Genome Biol. Evol.* **12**, 259–272 (2020).
20. D. Dong, M. Lei, P. Hua, Y.-H. Pan, S. Mu, G. Zheng, E. Pang, K. Lin, S. Zhang, The genomes of two bat species with long constant frequency echolocation calls. *Mol. Biol. Evol.* **34**, 20–34 (2017).
21. S. S. Pavlovich, S. P. Lovett, G. Koroleva, J. C. Guito, C. E. Arnold, E. R. Nagle, K. Kulcsar, A. Lee, F. Thibaud-Nissen, A. J. Hume, E. Mühlberger, L. S. Uebelhoer, J. S. Towner, R. Rabadan, M. Sanchez-Lockhart, T. B. Kepler, G. Palacios, The egyptian rousette genome reveals unexpected features of bat antiviral immunity. *Cell* **173**, 1098–1110.e18 (2018).
22. D. D. Moreno Santillan, T. M. Lama, Y. T. G. Guerrero, A. M. Brown, P. Donat, H. Zhao, S. J. Rossiter, L. R. Yohe, J. H. Potter, E. C. Teeling, S. C. Vernes, K. T. J. Davies, E. Myers, G. M. Hughes, Z. Huang, F. Hoffmann, A. P. Corthals, D. A. Ray, L. M. Dávalos, Large-scale genome sampling reveals unique immunity and metabolic adaptations in bats. *Mol. Ecol.* **30**, 6449–6467 (2021).
23. K. Halpin, P. L. Young, H. E. Field, J. S. Mackenzie, Isolation of Hendra virus from pteropid bats: A natural reservoir of Hendra virus. *J. Gen. Virol.* **81**, 1927–1932 (2000).
24. K. B. Chua, C. Lek Koh, P. S. Hooi, K. F. Wee, J. H. Khong, B. H. Chua, Y. P. Chan, M. E. Lim, S. K. Lam, Isolation of Nipah virus from Malaysian island flying-foxes. *Microbes Infect.* **4**, 145–151 (2002).
25. E. M. Leroy, B. Kumulungui, X. Pourrut, P. Rouquet, A. Hassanin, P. Yaba, A. Délicat, J. T. Paweska, J.-P. Gonzalez, R. Swanepoel, Fruit bats as reservoirs of Ebola virus. *Nature* **438**, 575–576 (2005).
26. J. S. Towner, X. Pourrut, C. G. Albariño, C. N. Nkogue, B. H. Bird, G. Grard, T. G. Ksiazek, J.-P. Gonzalez, S. T. Nichol, E. M. Leroy, Marburg virus infection detected in a common African bat. *PLOS ONE* **2**, e764 (2007).
27. E. D. Laing, I. H. Mendenhall, M. Linster, D. H. W. Low, Y. Chen, L. Yan, S. L. Sterling, S. Borthwick, E. S. Neves, J. S. L. Lim, M. Skiles, B. P. Y.-H. Lee, L.-F. Wang, C. C. Broder, G. J. D. Smith, Serologic evidence of fruit bat exposure to filoviruses, Singapore, 2011-2016. *Emerg. Infect. Dis.* **24**, 114–117 (2018).
28. Y. Han, J. Du, H. Su, J. Zhang, G. Zhu, S. Zhang, Z. Wu, Q. Jin, Identification of diverse bat alphacoronaviruses and betacoronaviruses in China provides new insights into the evolution and origin of coronavirus-related diseases. *Front. Microbiol.* **10**, 1900 (2019).
29. S. Wacharapluesadee, P. Duengkae, A. Rodpan, T. Kaewpom, P. Maneeorn, B. Kanchanasaka, S. Yingsakmongkon, N. Sittidetboripat, C. Chareesaen, N. Khlangsap, A. Pidthong, K. Leadprathom, S. Ghai, J. H. Epstein, P. Daszak, K. J. Olival, P. J. Blair, M. V. Callahan, T. Hemachudha, Diversity of coronavirus in bats from Eastern Thailand. *Virol. J.* **12**, 57 (2015).
30. M. Zheng, R. Karki, E. P. Williams, D. Yang, E. Fitzpatrick, P. Vogel, C. B. Jonsson, T.-D. Kanneganti, TLR2 senses the SARS-CoV-2 envelope protein to produce inflammatory cytokines. *Nat. Immunol.* **22**, 829–838 (2021).
31. T. Schountz, Immunology of bats and their viruses: Challenges and opportunities. *Viruses* **6**, 4880–4901 (2014).
32. E. Capanna, M. V. Civitelli, Chromosomal mechanisms in the evolution of Chiropteran karyotype chromosomal tables of Chiroptera. *Caryologia* **23**, 79–111 (1970).
33. E. W. Myers, The fragment assembly string graph. *Bioinformatics* **21 Suppl 2**, ii79-85 (2005).

34. F. A. Simao, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
35. A. Rhie, B. P. Walenz, S. Koren, A. M. Phillippy, Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).
36. A reference standard for genome biology. *Nat. Biotechnol.* **36**, 1121 (2018).
37. W. Bao, K. K. Kojima, O. Kohany, Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
38. J. Jurka, V. V. Kapitonov, O. Kohany, M. V. Jurka, Repetitive sequences in complex genomes: Structure and evolution. *Annu. Rev. Genomics Hum. Genet.* **8**, 241–259 (2007).
39. E. C. Teeling, M. S. Springer, O. Madsen, P. Bates, S. J. O'Brien, W. J. Murphy, A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* **307**, 580–584 (2005).
40. E. C. Skirmuntt, A. Katzourakis, The evolution of endogenous retroviral envelope genes in bats and their potential contribution to host biology. *Virus Res.* **270**, 197645 (2019).
41. R. V. Samonte, E. E. Eichler, Segmental duplications and the evolution of the primate genome. *Nat. Rev. Genet.* **3**, 65–72 (2002).
42. A. Mantovani, M. A. Cassatella, C. Costantini, S. Jaillon, Neutrophils in the activation and regulation of innate and adaptive immunity. *Nat. Rev. Immunol.* **11**, 519–531 (2011).
43. L. Li, C. J. Stoeckert Jr., D. S. Roos, OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
44. T. De Bie, N. Cristianini, J. P. Demuth, M. W. Hahn, CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).
45. G. Tsagkogeorga, S. Muller, C. Dessimoz, S. J. Rossiter, Comparative genomics reveals contraction in olfactory receptor genes in bats. *Sci. Rep.* **7**, 259 (2017).
46. B. V. Tsu, C. Beierschmitt, A. P. Ryan, R. Agarwal, P. S. Mitchell, M. D. Daugherty, Diverse viral proteases activate the NLRP1 inflammasome. *eLife* **10**, e60609 (2021).
47. F. Martinon, K. Burns, J. Tschopp, The inflammasome: A molecular platform triggering activation of inflammatory caspases and processing of proIL-β. *Mol. Cell* **10**, 417–426 (2002).
48. K. Han, S. K. Sen, J. Wang, P. A. Callinan, J. Lee, R. Cordaux, P. Liang, M. A. Batzer, Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. *Nucleic Acids Res.* **33**, 4040–4052 (2005).
49. M. D. Haggadone, J. J. Grailer, F. Fattahi, F. S. Zetoune, P. A. Ward, Bidirectional crosstalk between C5a receptors and the NLRP3 inflammasome in macrophages and monocytes. *Mediators Inflamm.* **2016**, 1340156 (2016).
50. N. P. Gerard, B. Lu, P. Liu, S. Craig, Y. Fujiwara, S. Okinaga, C. Gerard, An anti-inflammatory function for the complement anaphylatoxin C5a-binding protein, C5L2. *J. Biol. Chem.* **280**, 39677–39680 (2005).
51. A.-M. Scola, K.-O. Johswich, B. P. Morgan, A. Klos, P. N. Monk, The human complement fragment receptor, C5L2, is a recycling decoy receptor. *Mol. Immunol.* **46**, 1149–1162 (2009).
52. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
53. M. Baker, T. Schountz, L.-F. Wang, Antiviral immune responses of bats: A review. *Zoonoses Public Health* **60**, 104–116 (2013).
54. T. Omatsu, E.-J. Bak, Y. Ishii, S. Kyuwa, Y. Tohya, H. Akashi, Y. Yoshikawa, Induction and sequencing of Rousette bat interferon alpha and beta genes. *Vet. Immunol. Immunopathol.* **124**, 169–176 (2008).
55. N. Grandvaux, B. R. tenOever, M. J. Servant, J. Hiscott, The interferon antiviral response: From viral invasion to evasion. *Curr. Opin. Infect. Dis.* **15**, 259–267 (2002).
56. O. C. Probst, E. Karayel, N. Schida, E. Nimmerfall, E. Hehenberger, V. Puxbaum, L. Mach, The mannose 6-phosphate-binding sites of M6P/IGF2R determine its capacity to suppress matrix invasion by squamous cell carcinoma cells. *Biochem. J.* **451**, 91–99 (2013).
57. R. Roy, M. Touaibia, Application of multivalent mannosylated dendrimers in glycobiology. *Comprehensive Glycoscience*, 821–870 (2007).
58. R. A. B. Ezekowitz, Ante-antibody immunity. *Curr. Biol.* **1**, 60–62 (1991).
59. P. Broz, V. M. Dixit, Inflammasomes: Mechanism of assembly, regulation and signalling. *Nat. Rev. Immunol.* **16**, 407–420 (2016).
60. T. Kimura, A. Jain, S. W. Choi, M. A. Mandell, K. Schroder, T. Johansen, V. Deretic, TRIM-mediated precision autophagy targets cytoplasmic regulators of innate immunity. *J. Cell Biol.* **210**, 973–989 (2015).
61. D. Gao, J. Wu, Y.-T. Wu, F. du, C. Aroh, N. Yan, L. Sun, Z. J. Chen, Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses. *Science* **341**, 903–906 (2013).
62. W. Xie, L. Lama, C. Adura, D. Tomita, J. F. Glickman, T. Tuschl, D. J. Patel, Human cGAS catalytic domain has an additional DNA-binding interface that enhances enzymatic activity and liquid-phase condensation. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 11946–11955 (2019).
63. J. A. Boyer, C. J. Spangler, J. D. Strauss, A. P. Cesmat, P. Liu, R. K. McGinty, Q. Zhang, Structural basis of nucleosome-dependent cGAS inhibition. *Science* **370**, 450–454 (2020).
64. J. Dai, Y.-J. Huang, X. He, M. Zhao, X. Wang, Z. S. Liu, W. Xue, H. Cai, X.-Y. Zhan, S.-Y. Huang, K. He, H. Wang, N. Wang, Z. Sang, T. Li, Q.-Y. Han, J. Mao, X. Diao, N. Song, Y. Chen, W.-H. Li, J.-H. Man, A.-L. Li, T. Zhou, Z.-G. Liu, X.-M. Zhang, T. Li, Acetylation blocks cGAS activity and inhibits self-DNA-induced autoimmunity. *Cell* **176**, 1447–1460.e14 (2019).
65. Z. Cheng, T. Dai, X. He, Z. Zhang, F. Xie, S. Wang, L. Zhang, F. Zhou, The interactions between cGAS-STING pathway and pathogens. *Signal Transduct. Target. Ther.* **5**, 91 (2020).
66. Y. Wang, X. Ning, P. Gao, S. Wu, M. Sha, M. Lv, X. Zhou, J. Gao, R. Fang, G. Meng, X. Su, Z. Jiang, Inflammasome activation triggers caspase-1-mediated cleavage of cGAS to regulate responses to DNA virus infection. *Immunity* **46**, 393–404 (2017).
67. P. J. Kranzusch, A. S.-Y. Lee, J. M. Berger, J. A. Doudna, Structure of human cGAS reveals a conserved family of second-messenger enzymes in innate immunity. *Cell Rep.* **3**, 1362–1368 (2013).
68. W. J. Swanson, R. Nielsen, Q. Yang, Pervasive adaptive evolution in mammalian fertilization proteins. *Mol. Biol. Evol.* **20**, 18–20 (2003).
69. J. Schad, C. C. Voigt, Adaptive evolution of virus-sensing Toll-like receptor 8 in bats. *Immunogenetics* **68**, 783–795 (2016).
70. T. Sheahan, T. E. Morrison, W. Funkhouser, S. Uematsu, S. Akira, R. S. Baric, M. T. Heise, MyD88 is required for protection from lethal infection with a mouse-adapted SARS-CoV. *PLOS Pathog.* **4**, e1000240 (2008).
71. L. Oliveira-Nascimento, P. Massari, L. M. Wetzler, The role of TLR2 in infection and immunity. *Front. Immunol.* **3**, 79 (2012).
72. S. Basith, B. Manavalan, R. G. Govindaraj, S. Choi, In silico approach to inhibition of signaling pathways of Toll-like receptors 2 and 4 by ST2L. *PLOS ONE* **6**, e23989 (2011).
73. J.-S. Lin, E.-M. Lai, in *Bacterial Protein Secretion Systems*. (Springer, 2017), pp. 211–219.
74. M. Naghib, A. Kariminik, M. Kazemi Arababadi, TLR2, as a pathogen recognition receptor, plays critical roles in hepatitis B outcome. *Viral Immunol.* **35**, 15–23 (2022).
75. L.-F. Wang, A. M. Gamage, W. O. Y. Chan, M. Hiller, E. C. Teeling, Decoding bat immunity: The need for a coordinated research approach. *Nat. Rev. Immunol.* **21**, 269–271 (2021).
76. S. Chen, Y. Zhou, Y. Chen, J. Gu, Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
77. G. Marcais, C. Kingsford, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
78. J. Huddleston, S. Ranade, M. Malig, F. Antonacci, M. Chaisson, L. Hon, P. H. Sudmant, T. A. Graves, C. Alkan, M. Y. Dennis, R. K. Wilson, S. W. Turner, J. Korlach, E. E. Eichler, Reconstructing complex regions of genomes using long-read sequencing technology. *Genome Res.* **24**, 688–696 (2014).
79. X. Zhang, S. Zhang, Q. Zhao, R. Ming, H. Tang, Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nat. Plants* **5**, 833–845 (2019).
80. J. Hu, J. Fan, Z. Sun, S. Liu, NextPolish: A fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**, 2253–2255 (2020).
81. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
82. S. Wingett, P. Ewels, M. Furlan-Magaril, T. Nagano, S. Schoenfelder, P. Fraser, S. Andrews, HiCUP: Pipeline for mapping and processing Hi-C data. *F1000Res* **4**, 1310 (2015).
83. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
84. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
85. C. M. Bergman, H. Quesneville, Discovering and detecting transposable elements in genome sequences. *Brief. Bioinform.* **8**, 382–392 (2007).
86. Z. Xu, H. Wang, LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
87. R. C. Edgar, E. W. Myers, PILER: Identification and classification of genomic repeats. *Bioinformatics* **21**, i152–i158 (2005).
88. A. L. Price, N. C. Jones, P. A. Pevzner, De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).
89. G. Benson, Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
90. K. Lindblad-Toh, M. Garber, O. Zuk, M. F. Lin, B. J. Parker, S. Washietl, P. Kheradpour, J. Ernst, G. Jordan, E. Mauceli, L. D. Ward, C. B. Lowe, A. K. Holloway, M. Clamp, S. Gnerre, J. Alföldi, K. Beal, J. Chang, H. Clawson, J. Cuff, F. di Palma, S. Fitzgerald, P. Flicek, M. Guttman, M. J. Hubisz, D. B. Jaffe, I. Jungreis, W. J. Kent, D. Kostka, M. Lara, A. L. Martins, T. Massingham, I. Moltke, B. J. Raney, M. D. Rasmussen, J. Robinson, A. Stark, A. J. Vilella, J. Wen, X. Xie, M. C. Zody; Broad Institute Sequencing Platform and Whole Genome Assembly Team, J. Baldwin, T. Bloom, C. W. Chin, D. Heiman, R. Nicol, C. Nusbaum, S. Young, J. Wilkinson, K. C. Worley, C. L. Kovar, D. M. Muzny, R. . A. Gibbs; Baylor College of Medicine Human Genome

Sequencing Center Sequencing Team, A. Cree, H. H. Dihn, G. Fowler, S. Jhangiani, V. Joshi, S. Lee, L. R. Lewis, L. V. Nazareth, G. Okwuonu, J. Santibanez, W. C. Warren, E. R. Mardis, G. M. Weinstock, R. K. Wilson; Genome Institute at Washington University, K. Delehaunty, D. Dooling, C. Fronik, L. Fulton, B. Fulton, T. Graves, P. Minx, E. Sodergren, E. Birney, E. H. Margulies, J. Herrero, E. D. Green, D. Haussler, A. Siepel, N. Goldman, K. S. Pollard, J. S. Pedersen, E. S. Lander, M. Kellis, A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).

91. W. L. Eckalbar, S. A. Schlebusch, M. K. Mason, Z. Gill, A. V. Parker, B. M. Booker, S. Nishizaki, C. Muswamba-Nday, E. Terhune, K. A. Nevonen, N. Makki, T. Friedrich, J. E. VanderMeer, K. S. Pollard, L. Carbone, J. D. Wall, N. Illing, N. Ahituv, Transcriptomic and epigenomic characterization of the developing bat wing. *Nat. Genet.* **48**, 528–536 (2016).
92. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
93. D. W. Mount, Using the basic local alignment search tool (BLAST). *CSH Protoc.* **2007**, pdb.top17 (2007).
94. E. Birney, M. Clamp, R. Durbin, GeneWise and Genomewise. *Genome Res.* **14**, 988–995 (2004).
95. D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S. L. Salzberg, TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
96. C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, L. Pachter, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
97. B. J. Haas, A. L. Delcher, S. M. Mount, J. R. Wortman, R. K. Smith Jr., L. I. Hannick, R. Maiti, C. M. Ronning, D. B. Rusch, C. D. Town, S. L. Salzberg, O. White, Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
98. M. Stanke, S. Waack, Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19 Suppl 2**, ii215–ii225 (2003).
99. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
100. W. H. Majoros, M. Pertea, S. L. Salzberg, TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
101. R. Guigo, Assembling genes from predicted exons in linear time with dynamic programming. *J. Comput. Biol.* **5**, 681–702 (1998).
102. C. Burge, S. Karlin, Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
103. T. UniProt Consortium, UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* **46**, 2699 (2018).
104. R. D. Finn, P. Coggill, R. Y. Eberhardt, S. R. Eddy, J. Mistry, A. L. Mitchell, S. C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G. A. Salazar, J. Tate, A. Bateman, The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **44**, D279–D285 (2016).
105. The Gene Ontology Consortium, Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **45**, D331–D338 (2017).
106. M. Kanehisa, S. Goto, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Res.* **42**, D199–D205 (2014).
107. Y. Wang, H. Tang, J. D. DeBarry, X. Tan, J. Li, X. Wang, T. H. Lee, H. Jin, B. Marler, H. Guo, J. C. Kissinger, A. H. Paterson, MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
108. A. Veidenberg, A. Medlar, A. Loytynoja, Wasabi: An integrated platform for evolutionary sequence analysis and data visualization. *Mol. Biol. Evol.* **33**, 1126–1130 (2016).
109. G. Talavera, J. Castresana, Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).
110. D. Posada, K. A. Crandall, MODELTEST: Testing the model of DNA substitution. *Bioinformatics* **14**, 817–818 (1998).
111. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
112. M. J. Benton, P. C. Donoghue, Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* **24**, 26–53 (2007).
113. S. Kumar, G. Stecher, M. Suleski, S. B. Hedges, TimeTree: A resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
114. R. S. Harris, *Improved Pairwise Alignment of Genomic DNA*. (The Pennsylvania State University, 2007).
115. D. Ellinghaus, S. Kurtz, U. Willhoeft, LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18 (2008).
116. S. Steinbiss, U. Willhoeft, G. Gremme, S. Kurtz, Fine-grained annotation and classification of de novo predicted LTR retrotransposons. *Nucleic Acids Res.* **37**, 7002–7013 (2009).
117. P. Rice, I. Longden, A. Bleasby, EMBOSS: The European molecular biology open software suite. *Trends Genet.* **16**, 276–277 (2000).
118. C. Xie, X. Mao, J. Huang, Y. Ding, J. Wu, S. Dong, L. Kong, G. Gao, C.-Y. Li, L. Wei, KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **39**, W316–W322 (2011).
119. X. Mao, T. Cai, J. G. Olyarchuk, L. Wei, Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* **21**, 3787–3793 (2005).
120. J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, The I-TASSER Suite: Protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).

Acknowledgments: We thank B. Zhong for advice on the TLR2 and TLR8 experiments. **Funding:** This study was supported in part by the National Key Research and Development Program of China (2021YFF0702004), Fundamental Research Funds for the Central Universities (2042021kf0217), National Natural Science Foundation of China (31722051), Natural Science Foundation of the Hubei Province (2019CFA075) to H.Z. **Author contributions:** H.Z. conceived and designed research. S.T. designed and performed analyses. J.Z. performed functional experiments. H.J. and L.Z. collected the samples. D.Z. performed genome synteny analysis. S.T., C.-q.L., S.J.R., and H.Z. analyzed data, discussed results, and wrote the manuscript. All authors have read and approved the paper. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The *C. sphinx* whole-genome shotgun project has been deposited at the BIG Data Center (http://bigd.big.ac.cn), Beijing Institute of Genomics, Chinese Academy of Sciences, under accession code PRJCA003799. The assembled genome in this paper has been deposited into Genome Warehouse (GWH) under accession code GWHBHQA00000000. The genome resequencing reads have been deposited into the Genome Sequence Archive (GSA) under accession code CRA003492. In addition, all these data have also been deposited into the NCBI Sequence Read Archive and GenBank under the BioProject accession number PRJNA812996.

Submitted 17 May 2022
Accepted 31 March 2023
Published 5 May 2023
10.1126/sciadv.add0141