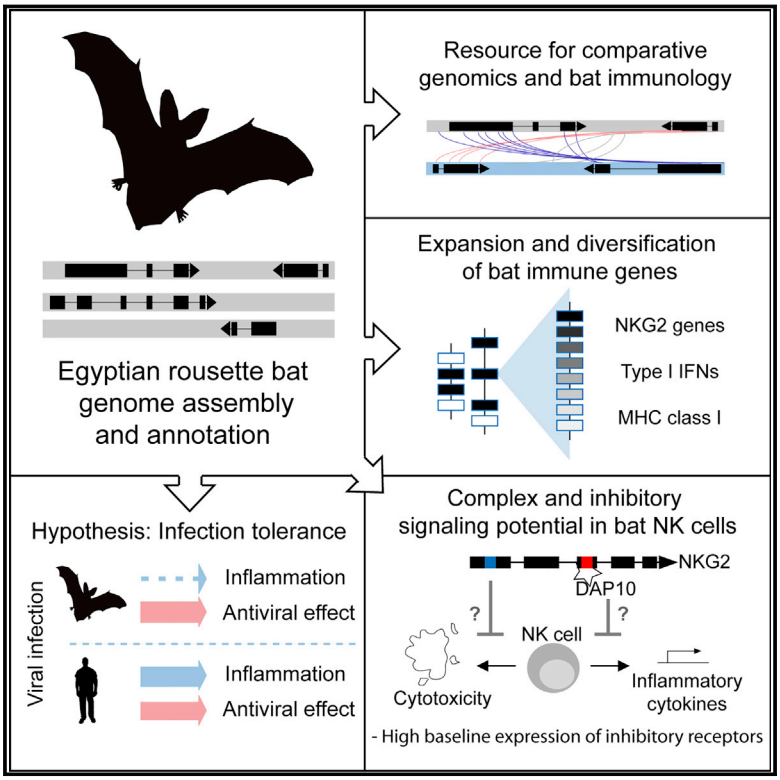
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Article

Cell

The Egyptian Rousette Genome Reveals Unexpected Features of Bat Antiviral Immunity

Graphical Abstract



Highlights

- High-contiguity genome assembly for *Rousettus aegyptiacus*, a host of Marburg virus
- Expansion of NK cell receptors, MHC class I genes, and type I interferons
- Genetic signatures of unique signaling in NK cell receptors in multiple bats
- Enhanced infection tolerance may be an antiviral defense strategy in bats

Authors

Stephanie S. Pavlovich, Sean P. Lovett, Galina Koroleva, ..., Mariano Sanchez-Lockhart, Thomas B. Kepler, Gustavo Palacios

Correspondence

tbkepler@bu.edu (T.B.K.), gustavo.f.palacios.ctr@mail.mil (G.P.)

In Brief

The genome of the Egyptian fruit bat reveals how its immune defenses allow tolerance of pathogenic viruses.

Pavlovich et al., 2018, Cell *173*, 1098–1110
May 17, 2018
https://doi.org/10.1016/j.cell.2018.03.070

CellPress

Cell

Article

The Egyptian Rousette Genome Reveals Unexpected Features of Bat Antiviral Immunity

Stephanie S. Pavlovich,[1] Sean P. Lovett,[2] Galina Koroleva,[2] Jonathan C. Guito,[5] Catherine E. Arnold,[2] Elyse R. Nagle,[2] Kirsten Kulcsar,[2] Albert Lee,[3] Françoise Thibaud-Nissen,[4] Adam J. Hume,[1,8] Elke Mühlberger,[1,8] Luke S. Uebelhoer,[5] Jonathan S. Towner,[5] Raul Rabadan,[3] Mariano Sanchez-Lockhart,[2,6] Thomas B. Kepler,[1,7,8,9,*] and Gustavo Palacios[2,9,10,*]

[1]Department of Microbiology, Boston University School of Medicine, Boston, MA 02118, USA
[2]Center for Genome Sciences, United States Army Research Institute of Infectious Diseases (USAMRIID), Frederick, MD 21702, USA
[3]Departments of Systems Biology and Biomedical Informatics, Columbia University, New York, NY 10032, USA
[4]National Center for Biotechnology Information, National Library of Medicine, NIH, Bethesda, MD 20892, USA
[5]Viral Special Pathogens Branch, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA
[6]Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE 68198, USA
[7]Department of Mathematics and Statistics, Boston University, Boston, MA 02215, USA
[8]National Emerging Infectious Diseases Laboratory, Boston University, Boston, MA 02118, USA
[9]These authors contributed equally
[10]Lead Contact
*Correspondence: tbkepler@bu.edu (T.B.K.), gustavo.f.palacios.ctr@mail.mil (G.P.)
https://doi.org/10.1016/j.cell.2018.03.070

SUMMARY

Bats harbor many viruses asymptomatically, including several notorious for causing extreme virulence in humans. To identify differences between antiviral mechanisms in humans and bats, we sequenced, assembled, and analyzed the genome of *Rousettus aegyptiacus*, a natural reservoir of Marburg virus and the only known reservoir for any filovirus. We found an expanded and diversified KLRC/KLRD family of natural killer cell receptors, MHC class I genes, and type I interferons, which dramatically differ from their functional counterparts in other mammals. Such concerted evolution of key components of bat immunity is strongly suggestive of novel modes of antiviral defense. An evaluation of the theoretical function of these genes suggests that an inhibitory immune state may exist in bats. Based on our findings, we hypothesize that tolerance of viral infection, rather than enhanced potency of antiviral defenses, may be a key mechanism by which bats asymptomatically host viruses that are pathogenic in humans.

INTRODUCTION

Bats, members of the large, diverse order *Chiroptera*, appear to harbor significantly more zoonotic viruses than other mammals and do so without overt pathology (Calisher et al., 2006; Olival et al., 2017). Such asymptomatic infection is especially noteworthy in the case of human pathogens such as henipaviruses (Nipah and Hendra viruses), coronaviruses (severe acute respiratory syndrome [SARS] and Middle East respiratory syndrome [MERS] coronaviruses), and filoviruses (Marburg virus [MARV]), which cause severe, and often lethal, systemic disease in humans and non-human primates (Calisher et al., 2006; Smith and Wang, 2013). This stark difference between bats and primates has motivated efforts to deeply characterize the genes involved in the immune system of bats and understand the antiviral immune mechanisms used to control viral infection.

Genomic analyses of immune genes in bats have produced conflicting and surprising observations. The most thoroughly studied bat genome is that of *Pteropus alecto* (Zhang et al., 2013), a reservoir host of Hendra virus. Additional bat genomes have been studied to a more limited extent (Seim et al., 2013; Zhang et al., 2013; Parker et al., 2013). The most notable findings from these studies involve two large classes of immune genes: natural killer (NK) cell receptors and type I interferons (IFNs). Multiple studies have reported the absence of canonical NK cell receptors in bat genomes (Shaw et al., 2012; Zhang et al., 2013; Lee et al., 2015), although a few receptors were identified in the *P. alecto* transcriptome (Papenfuss et al., 2012). A few studies suggest that significant differences exist in type I IFNs between bats and humans (Zhang et al., 2017; Zhou et al., 2016; Kepler et al., 2010; De La Cruz-Rivera et al., 2018), although the precise nature of the differences remains unclear. For example, the type I IFN locus has contracted in *Pteropus alecto* (Zhou et al., 2016), but expanded in *Pteropus vampyrus* and *Myotis lucifugus* (Kepler et al., 2010).

While these genome projects provide important insights into the unique biology of bats, the bat genomes currently available were generated with low-coverage sequencing or with only short-read next-generation sequencing (NGS) technologies (Zhang et al., 2013; Seim et al., 2013). These sequencing strategies impact the overall contiguity of genome assemblies and limit the ability to resolve repetitive genome loci where important immune gene loci reside. To overcome this limitation, we used a hybrid strategy combining both short- and long-read NGS technologies to generate a high-quality annotated genome for the Egyptian rousette bat (*Rousettus aegyptiacus*),

Table 1. Contiguity and Coverage among Bat Genomes

Species (Assembly Name)	Scaffold N50 (Kb)	Contig N50 (Kb)	Total Length (Gb)	Total Gap Length (Mb)	Coverage/Sequencing Technology
Rousettus aegyptiacus (Raegyp2.0)	2,007.2	1,489.0	1.910	0.482	169.2× Illumina HiSeq 2500 and Pacific Biosciences RS II[a]
Pteropus vampyrus (Pvam_2.0)	5,954.0	21.9	2.198	181.040	188.0× Illumina
Pteropus alecto (ASM32557v1)	15,954.8	31.8	1.986	41.334	110× Illumina HiSeq 2000
Myotis lucifugus (Myoluc2.0)	4,293.3	64.3	2.035	68.155	7× Sanger
Myotis brandtii (ASM32734v1)	3,225.8	23.3	2.107	125.473	120× Illumina HiSeq 2000
Myotis davidii (ASM32734v1)	3,454.5	15.2	2.060	181.338	110× Illumina HiSeq 2000
Miniopterus natalensis (Mnat.v1)	4,315.2	29.8	1.803	68.170	77.0× Illumina HiSeq
Eptesicus fuscus (EptFus1.0)	13,455.9	21.4	2.027	215.248	84× Illumina HiSeq
Megaderma lyra (ASM46534v1)	16.9	7.0	1.736	20.607	18.0× Illumina HiSeq
Eidolon helvum (ASM46528v1)	27.7	12.7	1.838	7.320	18.0× Illumina HiSeq
Rhinolophus ferrumequinum (ASM46549v1)	21.2	11.7	1.926	4.731	17.0× Illumina HiSeq
Rhinolophus sinicus (ASM188883v1)	3,754.4	37.8	2.073	58.015	146.44× Illumina HiSeq
Pteronotus parnellii (ASM46540v1)	22.7	9.5	1.960	13.7	17.0× Illumina HiSeq
Hipposideros armiger (ASM189008v1)	2,328.2	39.9	2.237	281.734	218.6× Illumina HiSeq

The contiguity statistics (scaffold and contig N50, total length, and total gap length) are reported for each bat genome available in the NCBI GenBank database (accessed on 8/8/17), along with the GenBank Assembly Name and the given coverage. Kb, kilobases; Mb, megabases; Gb, gigabases. See also Figure S1, Tables S1 and S2, and STAR Methods.

[a]Approximately 145× coverage of Illumina HiSeq 2500 data and 24× coverage of Pacific Biosciences RS II data.

an asymptomatic host of MARV (Towner et al., 2009). Here, we use this genome, the most contiguous bat genome available, to understand the evolution of immune genes and gene families in bats, and describe several observations relevant to defense against viruses.

We report an unusual expansion of the KLRC (NKG2) and KLRD (CD94) gene families in *R. aegyptiacus* relative to other species and show genomic evidence of unique features and expression of these receptors that may result in a net inhibitory balance within bat NK cells. The expansion of NK cell receptors is matched by an expansion of potential MHC class I ligands, which are distributed both within and, surprisingly, outside the canonical MHC loci. We also observe that the type I IFN locus is considerably expanded and diversified in *R. aegyptiacus*. The IFN-ω subfamily contributes most to this expansion, and members of this subfamily are induced after viral infection and show antiviral activity. All these features strengthen the notion of the unique biology of bats and suggest the existence of a distinct immunomodulatory mechanism used to control viral infection.

RESULTS

The Genome Assembly Has High Contiguity and Completeness

We used a hybrid paired-end short-read plus long-read strategy to generate 720 Gb of short-read data and 34.9 Gb of long-read data, for an approximate total coverage of 169×. We generated a draft genome (Raegyp2.0) comprising 1.91 Gb of sequence represented in 2,490 scaffolds (N50 = 2.007 Mb; NG50 = 1.811 Mb), which is ~90% of the estimated genome size. Of 248 core eukaryotic genes, 88.31% were complete in the genome, while 92.34% were at least partially represented, suggesting that gene information is covered well. The genome size and N50 are comparable to the assembled genomes of other bats. However, Raegyp2.0 has a larger contig N50 and a lower total gap length than any other available bat genome, making it the most contiguous bat genome to date (Table 1; Figures S1C and S1D). Whole-genome annotation was performed via the NCBI eukaryotic annotation pipeline using all *R. aegyptiacus*-specific transcript and RNA sequencing (RNA-seq) data in the NCBI databases, as well as proteins of other well-characterized species (Thibaud-Nissen et al., 2013). Similar to the number estimated by transcriptomic analysis (Lee et al., 2015; Hölzer et al., 2016) and the numbers in other bats (Table S2), 19,668 of the annotated genes in Raegyp2.0 are protein-coding genes with high support from transcriptomic or protein data.

Several Gene Families Have Expanded Significantly during the Evolution of *Chiroptera*

To gain insight into the evolutionary relationship of *R. aegyptiacus* to other bats and to incidental hosts of MARV, we inferred homologous protein groups among *R. aegyptiacus* and 14 other mammals and constructed a maximum-likelihood phylogenetic tree of all 15 species using single-copy orthologous genes (Figure 1). As previously established, the closest taxon to bats among the taxa included in the analysis is the horse (*Equus caballus*) (Seim et al., 2013; Zhang et al., 2013). We performed gene family expansion and contraction analysis on inferred gene families and observed many more expanded families in microbats than megabats, similar to what has previously been shown (Zhang et al., 2013; Tsagkogeorga et al., 2017) (Figure 1). 55 gene families are significantly expanded in *R. aegyptiacus* compared to the megabat ancestor (Figure 1; Tables S3 and S4).

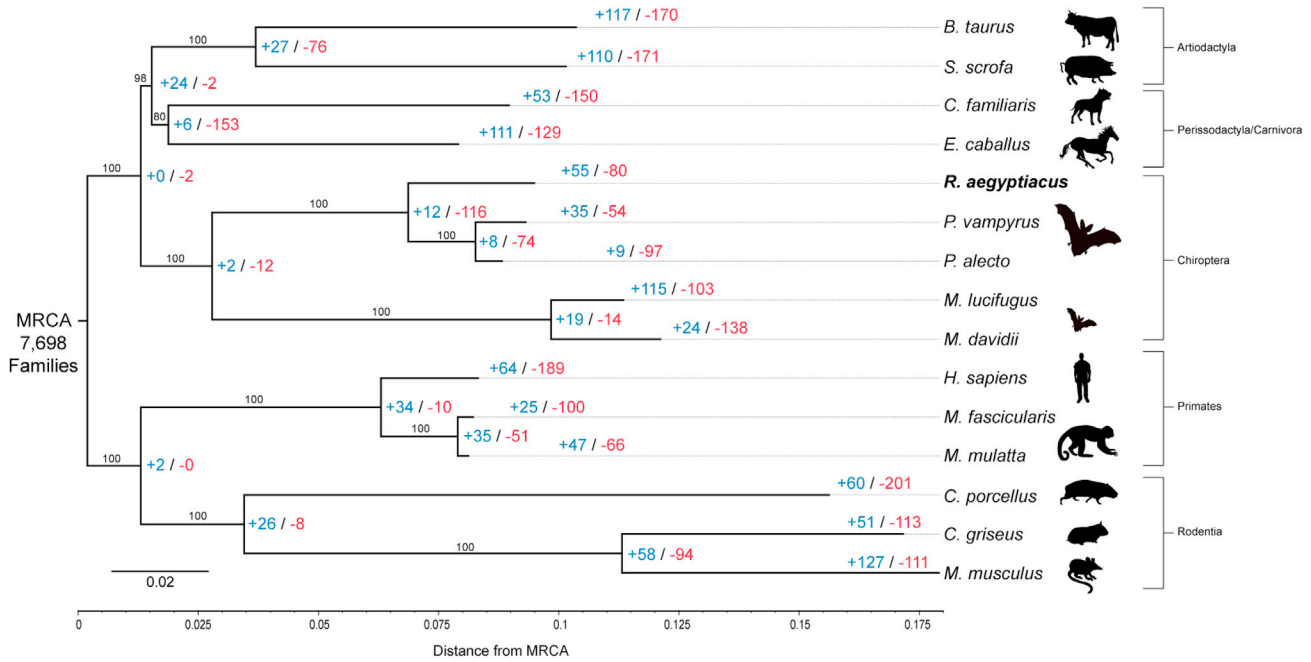


Figure 1. Gene Family Expansion and Contraction across a Phylogenetic Tree of 15 Mammalian Species

A maximum likelihood tree based on 2,400 orthologous proteins was generated and used to infer expansion and contraction of 7,698 gene families. The number of expanded and contracted gene families is in blue and red, respectively. Numbers in black are the bootstrap evidence for partitions based on 1,000 bootstrap replicates. Images used under a creative commons license. MRCA, most recent common ancestor.

See also Tables S3 and S4, Data S1, and STAR Methods.

Several Key Immune Genes Have Experienced Positive Selection

To investigate whether evolution at the gene level could contribute to the unique phenotype of *R. aegyptiacus* compared to other mammals, we studied disease-relevant immune genes for selection pressures (Figure S2; Table S5). Multiple innate immune response genes experience positive selection pressures along the *R. aegyptiacus* branch, including ISG15, an interferon-stimulated gene, IFNAR1, a subunit of the type I IFN receptor, and SIKE1, a negative regulator of the interferon response (Table 2). Several more innate immune genes experience relaxed purifying selection, including JAK2 and STAT3, components of the JAK/STAT signaling cascade, DDX58 (RIG-I), a sensor of viral double-stranded RNA, and TLR8, a pathogen-sensing molecule (Table 2). These results are consistent with those from similar analyses in *P. alecto* and *M. davidii* (Zhang et al., 2013).

Natural Killer Cell Receptors in *R. aegyptiacus* Have Unusual Origins and Unique Interaction Motifs

In mammals, NK cell receptors are encoded in two distinct gene complexes: the natural killer complex (NKC) contains killer lectin-like receptors (KLRs) such as CD94 (KLRD), NKG2 (KLRC), and Ly49, while the leukocyte receptor complex (LRC) contains immunoglobulin superfamily proteins such as the ILT/LIR family and the killer cell immunoglobulin-like receptors (KIRs). Both complexes vary in gene content among species, and encode both inhibitory and activating receptors. We did not find functional KIRs in the LRC, but consistent with what was found in *P. alecto* (Papenfuss et al., 2012), we identified 14 NKG2A/B-like genes (referred to as NKG2-1 through NKG2-14, including four pseudogenes), one NKG2D-like gene, one NKG2C-like gene, and 5 CD94-like genes (Table S3).

Remarkably, six of the ten putatively functional NKG2A/B-like genes simultaneously encode activating and inhibitory interaction motifs. Of the remaining genes, three encode only inhibitory motifs, and only one gene encodes an activating motif alone (Figures 2A and 2B). No other potential NK cell receptor genes were found. While inhibitory NKG2 receptors signal via ITIMs in their cytoplasmic domains, activating CD94/NKG2 receptors transmit signals via a positively charged residue in their transmembrane domains. This residue recruits adaptor molecules that contain activating motifs, with a lysine residue associated with DAP12 recruitment, and an arginine residue associated with DAP10 recruitment. *R. aegyptiacus* NKG2 proteins with putative activating function show a strong preference for arginine at this location, suggesting that these receptors favor an association with DAP10 (Figure 2B).

Multiple diverse NKG2 and CD94 genes presumably allow substantial combinatorial diversity among heterodimeric NKG2/CD94 receptors (Figures 2A, 3, and S3C). Four of the five CD94 genes in *R. aegyptiacus* lack two cysteines at position 58 and 59 that are highly conserved in multiple mammal species. These cysteines participate in the disulfide-mediated heterodimeric interaction between CD94 and NKG2 proteins in humans (Kaiser et al., 2008; Petrie et al., 2008), suggesting that these CD94 molecules might interact with their NKG2 co-receptors in

Table 2. Positive and Relaxed Purifying Selection in *R. aegyptiacus* Immune Genes

Symbol	Gene	ω_0	ω_1	FDR
SIKE1	Suppressor of IKBKE 1	0.117	1.107[b]	1.85E−09
ISG15	Interferon stimulated gene 15 ubiquitin-like modifier	0.151	1.841[b]	3.01E−08
PFN1	Profilin 1	0.161	2.003[b]	1.58E−05
CD48	CD48 molecule	0.638	2.074[b]	1.62E−04
IL1RL1	Interleukin 1 receptor like 1	0.417	1.246[b]	5.83E−04
NQO1	NAD(P)H quinone dehydrogenase 1	0.204	1.210[c]	6.23E−04
IL17A	Interleukin 17A	0.316	1.146[b]	2.22E−03
LEP	Leptin	0.360	2.289[a]	2.42E−03
OSMR	Oncostatin M receptor	0.534	1.053[b]	6.85E−03
TNFRSF1A	TNF receptor superfamily member 1A	0.346	1.133[a]	1.58E−02
IFNAR1	Interferon alpha and beta receptor subunit 1	0.527	1.193[c]	4.64E−02
DDX58	Dexd/h-box helicase 58	0.339	0.586[b]	5.77E−03
TLR8	Toll like receptor 8	0.346	0.608[b]	1.82E−03
NOD2	Nucleotide binding oligomerization domain containing 2	0.150	0.267[b]	1.02E−03
LTBR	Lymphotoxin beta receptor	0.314	0.803[b]	5.24E−04
NFKB2	Nuclear factor kappa B subunit 2	0.083	0.324[b]	6.67E−12
REL	REL proto-oncogene, NF-κB subunit	0.225	0.424[b]	1.11E−02
RELA	RELA proto-oncogene, NF-κB subunit	0.093	0.424[b]	4.10E−11
RELB	RELB proto-oncogene, NF-κB subunit	0.087	0.545[c]	8.31E−15
RNASEL	Ribonuclease L	0.538	0.906[b]	5.77E−03
JAK2	Janus kinase 2	0.090	0.303[c]	1.60E−03
STAT3	Signal transducer and activator of transcription 3	0.002	0.016[d]	4.72E−02

Evolution rates were estimated under several models, and the best fitting model was chosen. ω_0 refers to the background rate of evolution and always includes non-bat species, but also includes some bat species under certain models. $\omega 1$ refers to the rate of evolution of the *R. aegyptiacus* gene but also includes other bat species under certain models. FDR, false discovery rate. See Figure S2 and STAR Methods for full description of the models. See also Table S5.

[a] ω_1 estimate under Model 1a.
[b] ω_1 estimate under Model 1b.
[c] ω_1 estimate under Model 1c.
[d] ω_1 estimate under Model 3.

an alternate way. Mouse and rat CD94 are capable of direct association with DAP12 or DAP10 via a lysine residue in their transmembrane domains (Koch et al., 2013), but the CD94 sequences from all bats we examined have no lysine residues in their transmembrane regions, so are unlikely to associate directly with DAP proteins (Figure S4C).

To determine whether CD94 and NKG2 genes are expressed, we queried our published transcriptomic data (Lee et al., 2015) and found that the majority of these genes are expressed at low levels in peripheral blood mononuclear cells and secondary lymphoid organs (Figures 2C and 3B), similar to human baseline expression. However, two receptors with inhibitory signaling motifs—NKG2-13 and NKG2-14—along with CD94-1, are expressed at higher levels in the same tissues, suggesting that inhibitory signaling dominates in uninfected bats.

To determine whether NKG2/CD94 receptors are diversified across bats, we examined the NKC in additional bats, and found that all bats we studied have multiple NKG2-like genes, although some were not originally classified as C-type lectins because of missing exons; *P. vampyrus* and *P. alecto* also have multiple CD94-like genes (Figures 2, 3, and S3). Each bat has one CD94 gene with canonical cysteine residues at position 58 and 59, (Figures 3A and S4A) and like *R. aegyptiacus*, multiple bats have CD94 genes without these residues, suggesting that they may have alternative functions. Phylogenetic analysis shows that NKG2 genes have undergone considerable diversification before and after the speciation of megabats (Figures 2D and S5). As with *R. aegyptiacus*, the putative functional and truncated NKG2 genes in both megabats (*P. alecto* and *P. vampyrus*) and microbats (*M. lucifugus* and *M. davidii*) contain ITIMs or both ITIMs and a positively charged transmembrane residue, suggesting that both inhibitory and activating signaling in these receptors is common among all bats in our analysis (Figure 2B).

The MHC Class I Genes Are Located Both within and outside of the Canonical MHC Locus

While KIRs interact with classical MHC class I molecules (cMHCs), NKG2A/B/C and F receptors interact with HLA-E, a non-classical MHC class I molecule (ncMHC) that displays nonamers derived from the signal peptides of cMHCs. This mechanism is thought to allow NK cells to monitor the expression of cMHCs and deliver cytotoxic hits to cells lacking such expression (Yokoyama and Plougastel, 2003). We hypothesized that

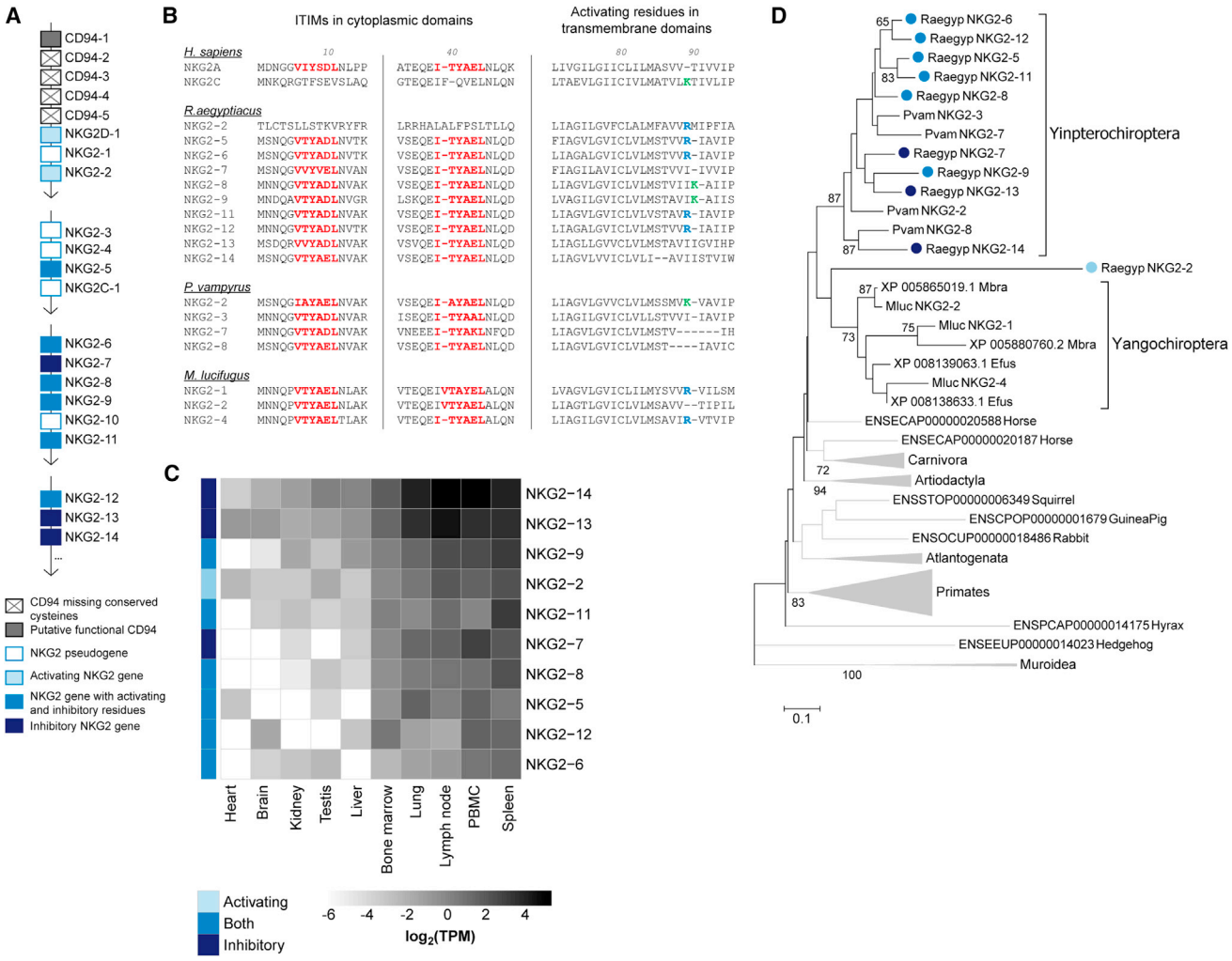


Figure 2. Expansion of the NKG2 Genes in *R. aegyptiacus*

(A) CD94 and NKG2 genes in the natural killer complex in Raegyp2.0. Each arrow designates a scaffold sequence in the Raegyp2.0 genome (see STAR Methods for accessions). Not pictured are pseudogenes and non-coding genes. The ellipse indicates the presence of additional non-NKG2 genes on the same scaffold.

(B) Multiple sequence alignments showing activating and inhibitory signaling motifs in NKG2 genes in humans and three bats. There were no putative functional bat NKG2 genes identified in in *P. alecto* or *M. davidii* except NKG2-D. ITIM (immunoreceptor tyrosine-based inhibition motif) residues are in red, and the signal anchor residue lysine (K) or arginine (R) are in green and blue respectively. Dashes represent gaps in the alignment.

(C) Expression of putative functional NKG2 genes in transcriptomic data from 10 tissues in a *R. aegyptiacus* bat. Rows are ordered by highest average expression of transcripts across all tissues for a given gene. Expression is reported in $\log_2$(TPM), where TPM refers to transcripts per million. Data analyzed from Lee et al. (2015).

(D) Maximum likelihood phylogenetic tree of bat NKG2 proteins and homologs in other species. NKG2 proteins from *R. aegyptiacus* are colored by predicted function. Bootstrap evidence (percentage of 500 bootstrap replicates) is labeled on branches if 65 or over.

See also Figures S3, S5, and S7 and STAR Methods.

the expansion of the NKG2 receptor family would be matched by an expansion of cMHCs or ncMHCs. In humans, MHC class I genes are located in three areas referred to as the α, κ, and β duplication blocks, which are separated by framework regions containing non-HLA genes (Kulski et al., 2002). HLA-E and its functional equivalent in mice, H2-Qa1, are located in the κ block. Raegyp2.0 appears to lack the α and κ blocks (Figure 4A), as do *P. alecto* and *E. fuscus* (Ng et al., 2016).

Consistent with an expansion, we find twelve MHC class I-like genes and seven pseudogenes in Raegyp2.0 (Figures 4A and 4B), none of which is discernible as the functional equivalent of HLA-E. Nonamers inferred bioinformatically from *R. aegyptiacus* MHC class I signal peptides are much less diverse than those observed in human or mouse, a pattern observed also in the gray mouse lemur (Figure 4E). Only two MHC class I genes and two pseudogenes are located in the β block (Figure 4A). Two MHC class I genes in Raegyp2.0 were found in genomic contexts apparently outside of the MHC class I region (Figure 4B). Eight genes and five pseudogenes were identified on scaffolds that could not be further localized in the

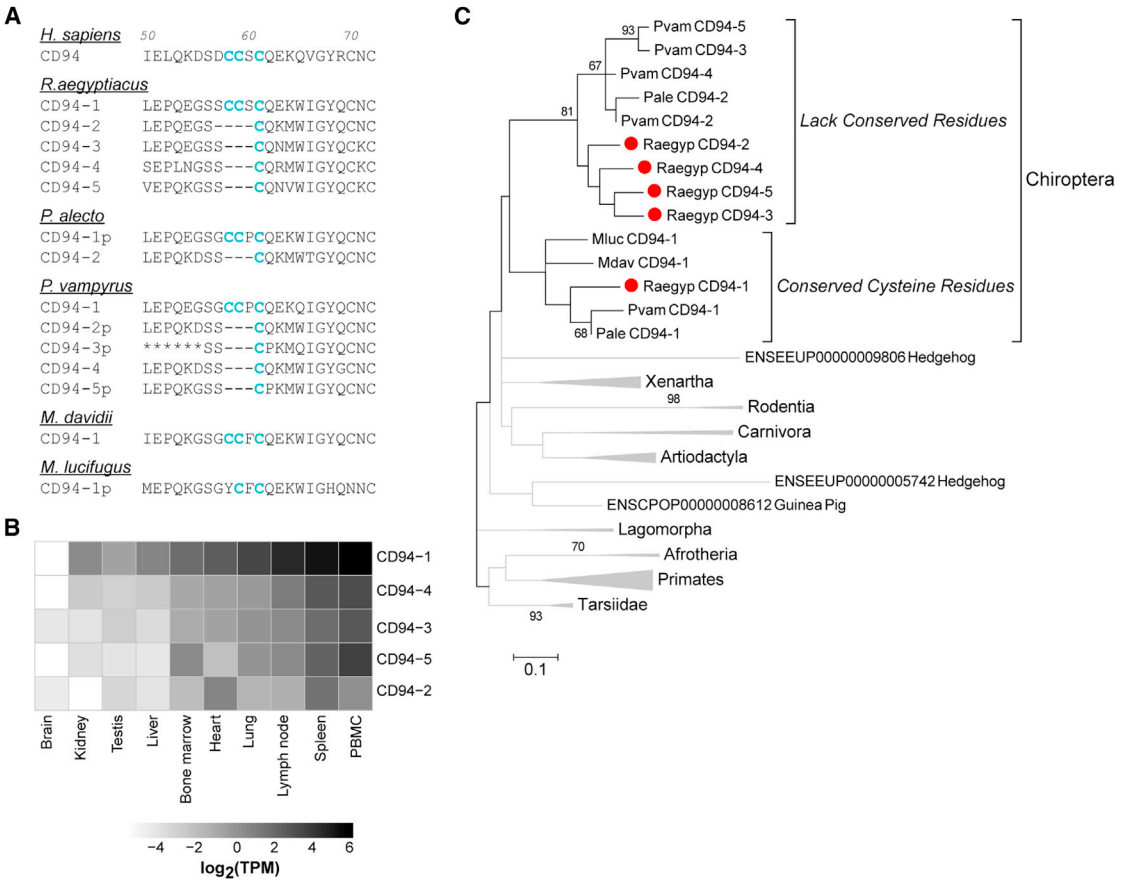


Figure 3. Expression and Diversity of CD94 in *R. aegyptiacus*

(A) Multiple sequence alignments showing conserved cysteine residues in CD94 genes in humans and five bats. Pseudogenes are indicated with the letter p in protein name. Asterisks indicate missing residues from a partial *P. vampyrus* CD94.

(B) Expression of CD94 genes in transcriptomic data from ten tissues in an Egyptian rousette bat. Rows are ordered by highest average expression of transcripts across all tissues for a given gene. Expression is reported in $\log_2$(TPM), where TPM refers to transcripts per million. Data analyzed from Lee et al. (2015).

(C) Maximum likelihood phylogenetic tree of bat CD94 proteins and homologs in other species. CD94 proteins from *R. aegyptiacus* are marked by red dots. Bootstrap evidence (percentage of 500 bootstrap replicates) is labeled on branches if over 65.

See also Figures S3–S5 and STAR Methods.

genome. These could potentially be allelic variants or additional genomic contexts for MHC-I genes. Examination of additional bat genomes also shows MHC class I genes outside the canonical MHC class I region, suggesting that dispersion of class I genes is not an assembly artifact but is common to many bats (data not shown). Many of the MHC class I genes in Raegyp2.0 are expressed across a wide range of tissues, including genes located outside the canonical MHC locus (Figure 4C) like MHC-11 and -12, suggesting that they may function in the canonical self-detection role of cMHCs.

Unlike NKG2A/CD94 receptors, NKG2D forms homodimers that bind a number of MHC class Ib ligands, including MICA and MICB, and members of the ULBP family, which are upregulated in cells during infection and stress (Molfetta et al., 2016). MICA and MICB were also not found in the MHC-I locus, although a candidate MICB ortholog was found outside of the MHC loci (Figure 4B). *R. aegyptiacus* and other bats appear to have two additional groups of NKG2D ligands, which are closest to ULBPs in humans (Figure S6). Further investigation is needed to determine whether these genes functionally resemble ULBP or MIC family members.

Type I Interferons Are Expanded in *R. aegyptiacus*

We estimate 20 type I IFN genes in the megachiropteran ancestor and find 46 putative functional genes in Raegyp2.0, including 12 IFN-α genes, one of each of the IFN-β, -ε, and -κ genes, nine IFN-δ genes, and 22 IFN-ω genes (Figure 5; Table S3). The greatest expansion occurred in the IFN-ω subfamily, which has only one copy in humans.

Given previous reports of constitutive IFN expression in bats, we sought to determine whether the expanded IFN genes may be constitutively expressed and found limited baseline expression of these genes (Table S6). To determine whether these IFN-ωs may be induced by viral infection, we infected immortalized *R. aegyptiacus* cells (RoNi) with the Cantell strain of Sendai virus, a known inducer of IFNs in other species. We observed

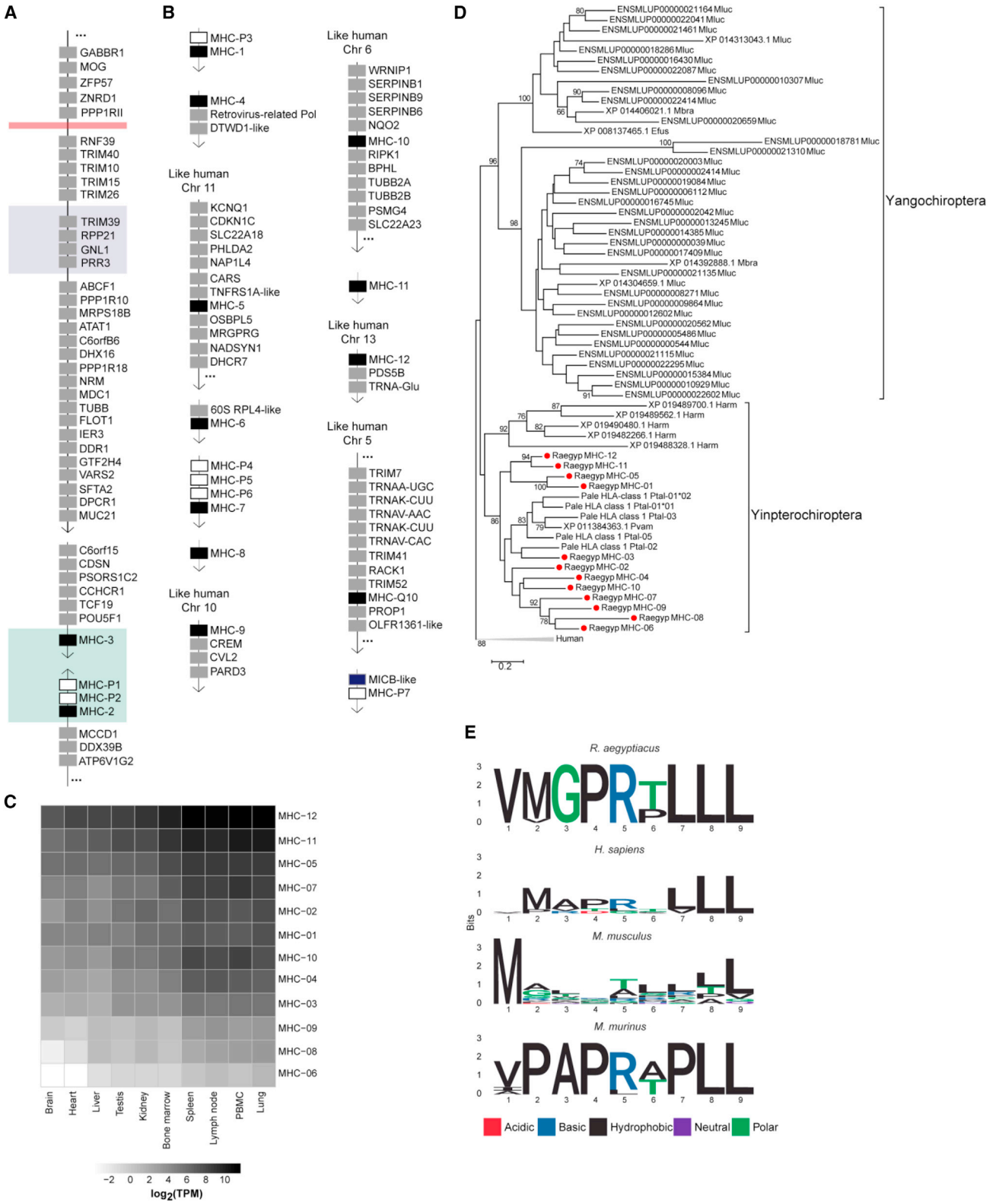


Figure 4. Characterization of the MHC Class I Region in Raegyp2.0

(A and B) Locus maps of (A) the MHC class I region and (B) MHC class I genes outside the canonical class I region in Raegyp2.0. Each arrow designates a scaffold sequence in the Raegyp2.0 genome (see STAR Methods for accessions). Not pictured are non-MHC pseudogenes and non-coding genes. The ellipse indicates

(legend continued on next page)

induction of IFN-ω transcripts (Figure S7C), although at relatively low levels compared to IFN-β transcripts. To examine whether IFN-ω proteins retain the canonical antiviral function of type I IFNs, we synthesized recombinant Egyptian rousette IFN-β and IFN-ω4, and an unrelated protein of similar size, and used these proteins in an antiviral assay in RoNi cells. Consistent with a bona-fide antiviral function, treatment of RoNi cells with IFN-ω4 blocks infection with vesicular stomatitis virus encoding eGFP (VSV-eGFP) (Figure S7).

DISCUSSION

Few viruses are known to cause acute disease in bats, including those that cause profound, often lethal, disease in humans. The precise mechanism(s) of this viral resistance is not known. One hypothesis that has been gaining acceptance is that bats present especially potent innate antiviral defenses compared to primates, controlling viral replication early in infection, and as a result, developing effective adaptive immune responses (Baker et al., 2013). Our results suggest a different hypothesis that is consistent with the idea that bat antiviral mechanisms are different in essential ways from those of other mammals, but is additionally associated with enhanced infection tolerance rather than enhanced defense. This hypothesis is supported by infection studies of MARV transmission among *R. aegyptiacus* bats, in which bats that are ''naturally'' infected by other experimentally inoculated bats appear to have a protracted incubation period (Amman et al., 2015; Schuh et al., 2017). Further, these studies showed that infected bats can remain viremic and shed infectious virus for extended periods of time (up to 3 weeks after infection) before eventually clearing the virus (Data S1). Despite prolonged infection, limited inflammation is observed in even the most highly infected tissues, similar to what was previously observed in infected wild-caught bats (Jones et al., 2015; Towner et al., 2009).

Type I IFNs

Bats may tolerate viral infections to a greater extent by minimizing the proinflammatory effectors that promote damage to the host in many viral infections. Type I IFNs are induced very early in viral infection and act by inducing effectors encoded by interferon stimulated genes (ISGs). Different IFN subtypes specifically interact with the common IFN receptor inducing a distinct spectra of ISGs with different antiviral potencies (Hoffmann et al., 2015). The magnitude and nature of the IFN response determine whether the resulting effects on the host are harmful or beneficial (Malireddi and Kanneganti, 2013). For example, dysregulation of the type I IFN response has been implicated in the pathogenesis of multiple emerging viruses, including MARV (Liu et al., 2017; Connor et al., 2015).

Although IFN gene families differ substantially across mammals (Secombes and Zou, 2017), the extensive expansion of IFN-ω genes in *R. aegyptiacus* to almost two dozen genes is striking. This expansion is dramatically different from what was observed in *P. alecto* (Zhou et al., 2016), but is consistent with the expansion in *P. vampyrus* (Kepler et al., 2010). In our hands, at least one of the IFN-ω members, IFN-ω4, has a marked antiviral effect against VSV-eGFP infection in RoNi cells (Figure S7). However, the effect observed in this *in vitro* system, was less potent than that of IFN-β. Thus, IFN-ω4 most likely induces a unique ISG pattern and may be more effective against different viruses, or may induce lower levels of effectors in a more regulated response. Consistent with this latter scenario, Banerjee et al. (2017) have shown that poly I:C treatment induces type I IFNs in both human and *Eptesicus fuscus* bat cells, but bat cells express much lower levels of inflammatory mediators. A wider variety of signaling mediators may provide *R. aegyptiacus* greater flexibility to develop a more nuanced antiviral response. It is also possible that different IFN-ωs may complement and/or synergize with each other.

We observed IFN-ω gene induction in RoNi cells after Sendai virus infection, albeit at lower levels than IFNβ or IFN-α genes (Figure S7). Given that Sendai virus is known to be a potent inducer of IFN-α and -β in particular, it is possible that different viruses or other stimuli may preferentially induce IFN-ω genes. Unlike *P. alecto*, *R. aegyptiacus* shows no evidence of constitutive IFN expression (Table S6). Both bats are reservoirs for different emerging viruses of high virulence to humans, so it will be of great interest to determine whether these apparent differences in IFN expression represent distinct mechanisms of viral control by both bats species.

NK Cell Receptors

NK cells are an important component of innate antiviral responses, and have been associated with survival of Ebola virus infection (Liu et al., 2017). The unique organization, structure, and increased signaling complexity of the NK cell receptors in multiple bats point to adaptations that are also consistent with the infection tolerance hypothesis. Our findings suggest that NKG2/CD94 receptors, which are more associated with an inhibitory response in other species, serve as the primary NK cell receptors in bats. Consistent with this, all but one of the NKG2A/B-like genes in *R. aegyptiacus* have inhibitory motifs at the cytosolic tail. The expansion of CD94 genes makes the

the presence of additional genes on the same scaffold. Black, MHC class I genes; gray, non-MHC genes; dark blue, MICB; unfilled boxes, MHC class I pseudogenes. The α, κ, and β class I duplication blocks are shown in red, purple, and green, respectively.

(C) Expression of MHC class I genes in transcriptomic data from 10 tissues in an Egyptian rousette bat. Rows are ordered by highest average expression of transcripts across all tissues for a given gene. Expression is reported in $\log_2$(TPM), where TPM refers to transcripts per million. Data analyzed from Lee et al. (2015).

(D) Maximum likelihood phylogenetic tree of bat MHC class I proteins (*R. aegyptiacus* proteins in red) with human MHC class I proteins as an outgroup. Bootstrap evidence (percentage of 500 bootstrap replicates) is labeled on branches if 65 or over.

(E) Sequence logo plots showing the sequence diversity of predicted nonamer peptides derived from the signal sequences of MHC class I genes from human (*H. sapiens*), mouse (*M. musculus*), *R. aegyptiacus*, and the gray mouse lemur (*M. murinus*). The y axis shows information content in bits, and the x axis shows position in the nonamer.

See also STAR Methods.

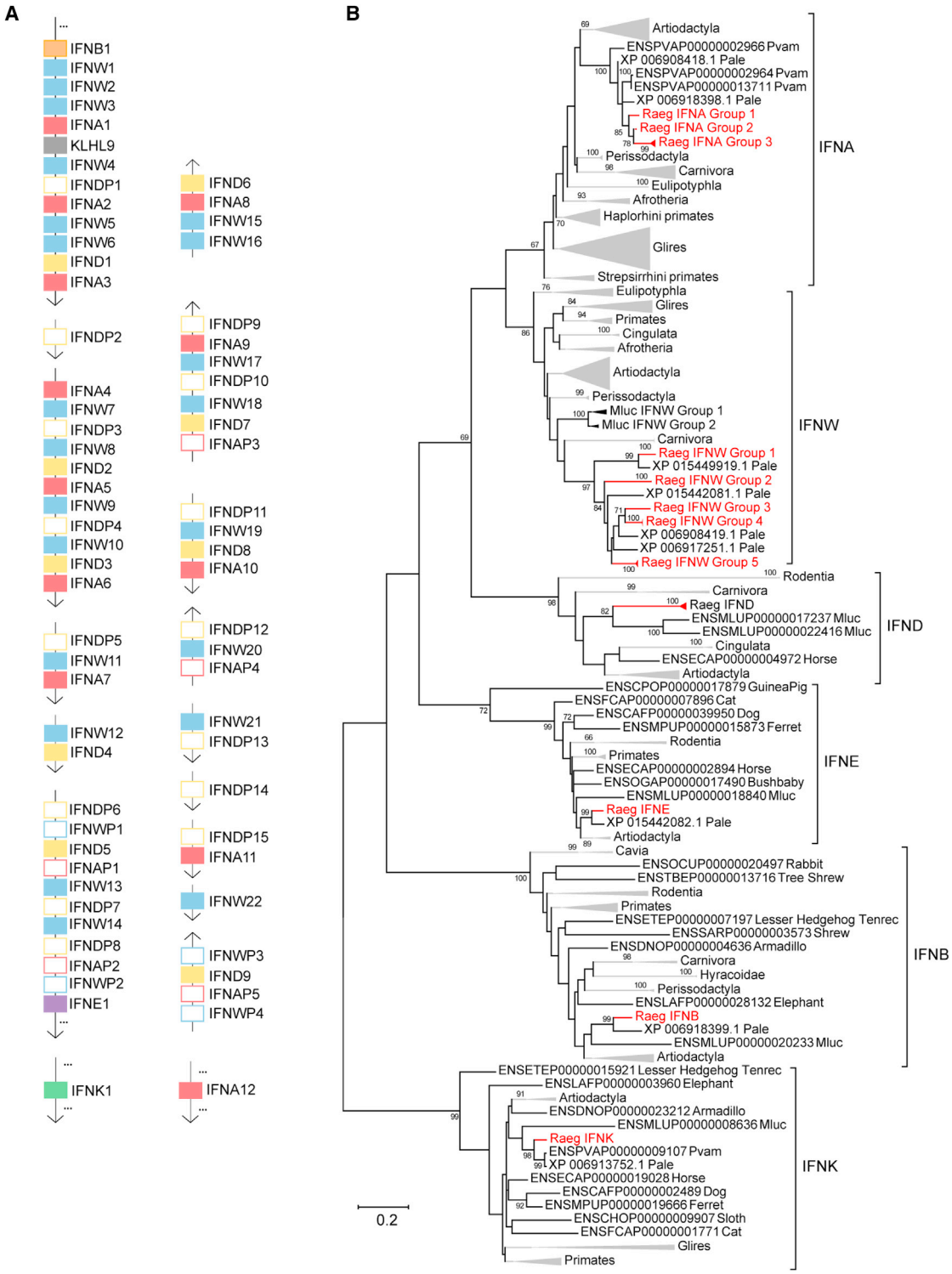


Figure 5. Diversity of the Type I Interferons in Raegyp2.0

(A) Locus map of the type I IFNs in Raegyp2.0. Each arrow designates a scaffold sequence in the Raegyp2.0 genome (see STAR Methods for scaffold accessions). Unfilled boxes indicate pseudogenes. Orange, IFNβ; blue, IFNω; red, IFNα; yellow, IFNδ; purple, IFNε; green, IFNκ. The single non-IFN gene within the locus (KLHL9) is in gray. Not pictured are non-coding genes. The ellipse indicates the presence of additional non-IFN genes on the same scaffold.

(B) Maximum likelihood phylogenetic tree of bat type I IFN proteins and homologs in other species. *R. aegyptiacus* proteins are marked in red, with groups of closely related proteins collapsed. Bootstrap evidence (percentage of 500 bootstrap replicates) is labeled on branches if 65 or over.

See also Figure S6, Table S6, and STAR Methods.

possible combinatorial diversity of heterodimeric receptors very large, as previously demonstrated for prosimians (Averdam et al., 2009). Because signal transmission occurs via a conserved set of residues, the additional receptor diversity is likely to be associated with the capacity to bind additional ligands or differentially interact with the same ligands.

Greater diversity in ligand binding would provide better recognition of MHC class I alleles to recognize self for NK cell tuning and licensing, and also to distinguish a variety of pathogen mimics of MHC class I molecules (Parham and Moffett, 2013). The KLRD variants missing conserved cysteines add an additional level of complexity to this interaction that needs further characterization once tools for isolating NK cells in bats are available. In other mammals, KLRC genes are also expressed on T cells, and NKG2A has been shown to control the level of T cell activation during viral infection, preventing excessive activation and immunopathology in mice (Rapaport et al., 2015). The high baseline expression of inhibitory KLRC genes may indicate an immune-inhibitory state associated with both NK cells and T cells in *R. aegyptiacus* bats in the absence of infection, although this remains to be confirmed with functional studies.

A further striking feature of the NK cell receptor genes in *R. aegyptiacus* and the other bats we studied is the signaling mode they are predicted to use. In genes with potential for activating signaling, an arginine residue is preferred (Figure 2B), suggesting DAP10, rather than DAP12, recruitment (Koch et al., 2013). DAP12 has been shown to be more efficient in activating cytokine production than DAP10 (Lanier 2009), therefore a preference for DAP10 could mean that activating NK receptors in bats have adapted to be less potent inducers of cytokines, and thus less inflammatory.

NK cell receptors that possess both inhibitory and activating domains are unusual across mammals. The only known genes with both functions are the single gene KIR2DL4 in humans and the NKG2 genes in lemurs (Parham and Moffett, 2013). KIR2DL4 activation promotes robust cytokine secretion but not cytotoxicity (Kikuchi-Maki et al., 2005). It is possible that the *R. aegyptiacus* KLRC genes mimic the signaling of human KIR2DL4, although parallels between both receptors are difficult to make without functional assays. However, our genomic evidence suggests that NK cell receptors in bats are uniquely regulated, especially given the multiple changes in the signaling potential of these receptors, which are highly conserved in eukaryotes ranging from *Drosophila* to humans.

MHC Class I Genes

The potential extended capacity of the KLRC/KLRD system to bind diverse ligands is accompanied by expansion of MHC class I genes outside of the canonical MHC locus, although there is no obvious bat ortholog of HLA-E. A similar expansion and absence of a clear HLA-E ortholog has been observed in prosimians (Averdam et al., 2009). If one of the MHC class I genes in Raegyp2.0 is functionally similar to HLA-E, the high similarity of predicted class I nonamers (Figure 4E) suggests that presenting a peptide from one MHC class I molecule would likely be interchangeable with presenting a peptide from another. Thus, expression of MHC class I molecules would have to be dramatically decreased in order for "missing-self" detection to occur. While functional orthologs cannot be inferred without functional assays, the large number of MHC genes encoded outside the canonical locus in the bats we studied (Figure 4B) may be suitable ligands for the expanded KLR genes. Distribution of the MHC genes across the genome might serve as a mechanism to generate redundancy as different KLRC receptors might interact with distinct MHC class I genes. This could potentially result in a higher activation threshold for NK cells.

Taken together, these results show that multiple bats, including *R. aegyptiacus*, have expanded and diversified numerous antiviral loci, and potentially developed unique adaptations in NK cell receptor signaling, and type I IFN responses. Our findings are consistent with the hypothesis that certain key components of the immune system in bats have coevolved with viruses toward a state of respective tolerance and avirulence, although tolerance is likely not the only mechanism at play. For example, in addition to enhanced flexibility, the expansion of type I IFNs may also point to enhanced potency of antiviral defenses. Recent studies have shown that IFN stimulation in bats induces some ISGs that are not induced by IFN in humans, and that the response kinetics may vary as well (De La Cruz-Rivera et al., 2018). Adaptations in potency are also indicated by observations of lower viral loads in *R. aegyptiacus* bats compared to humans (Amman et al., 2015; Jones et al., 2015). Finally, even with mutual disarmament, the host must be alert to viruses that may spontaneously revert to an antagonistic phenotype. In either case, definitive tests of these hypotheses await the development of further experimental reagents for cytometry and biochemical intervention—reagents that are being developed now with information made available by the completed genome project.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Source organism
  - Cell lines
- METHOD DETAILS
  - Nucleic acid extraction and sequencing
  - Genome Size Estimation and Heterozygosity analysis
  - Genome Assembly
  - Genome Annotation
  - Designating gene families
  - Identification of repetitive content
  - Heterozygosity analysis
  - Endogenous viral content
  - Species tree generation
  - Gene family expansion and contraction analysis
  - Annotation of genes
  - Phylogenetic analysis
  - Transcriptome analysis
  - Sendai virus infection of RoNi cells
  - Production of recombinant 6xHis-tagged IFNs

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, six tables, and one data file and can be found with this article online at https://doi.org/10.1016/j.cell.2018.03.070.

ACKNOWLEDGMENTS

We thank John H. Connor for his generous contribution of VSV-eGFP, Kaitlin Sawatzki for her help with Figure 1 and members of the Kepler lab for their helpful discussion. We gratefully acknowledge Ayan K. Chakrabarti, who performed the RNA extraction for the interferon induction experiment. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the U.S. Army or the Centers for Disease Control and Prevention. Research was conducted under an IACUC-approved protocol in compliance with the Animal Welfare Act, PHS Policy, and other federal statutes and regulations relating to animals and experiments involving animals. The facility where this research was conducted is accredited by the Association for Assessment and Accreditation of Laboratory Animal Care, International and adheres to principles stated in the Guide for the Care and Use of Laboratory Animals, National Research Council, 2011. This work has been funded by the Defense Threat Reduction Agency, Department of Defense (HDTRA1-14-1-0016 to G.P. and J.S.T.) and NIH NIAID (training grant T32AI007309-26 to S.S.P., awarded to David H. Sherr). Part of the work was funded by NIH NIGMS (5R01GM117591 and 1R01GM109018 to R.R. and A.L.). Part of this project was also supported by the Intramural Research Program of the NIH, National Library of Medicine (F.T.-N.). The *Rousettus aegyptiacus* cell line RoNi/7.1 was generated by M.A. Müller and C. Drosten, Charité-Universitätsmedizin Berlin, Germany with funds from the EU-FP7 ANTIGONE (no. 278976) framework and the German Research Council (DR 772/10-2).

AUTHOR CONTRIBUTIONS

Conceptualization, G.P. and J.S.T.; Methodology, S.S.P., S.P.L., T.B.K., A.J.H., E.M., and E.R.N.; Formal analysis, S.S.P., S.P.L., F.T.-N., A.L., K.K., and C.E.A.; Investigation, S.S.P., S.P.L., J.C.G., G.P., and C.E.A.; Resources, L.S.U., J.S.T., G.K., and E.M.; Data Curation, F.T.-N., A.L., and R.R.; Writing – Original Draft, S.S.P., S.P.L., G.P., T.B.K., and M.S.-L.; Writing – Review & Editing, S.S.P., S.P.L., G.P., T.B.K., M.S.-L., J.S.T., and F.T.-N.; Supervision, T.B.K., G.P., M.S.-L., J.S.T., and E.M.; Funding Acquisition, G.P. and J.S.T.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 20, 2017
Revised: January 22, 2018
Accepted: March 27, 2018
Published: April 26, 2018

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Amman, B.R., Jones, M.E.B., Sealy, T.K., Uebelhoer, L.S., Schuh, A.J., Bird, B.H., Coleman-McCray, J.D., Martin, B.E., Nichol, S.T., and Towner, J.S. (2015). Oral shedding of Marburg virus in experimentally infected Egyptian fruit bats (Rousettus aegyptiacus). J. Wildl. Dis. *51*, 113–124.

Averdam, A., Petersen, B., Rosner, C., Neff, J., Roos, C., Eberle, M., Aujard, F., Münch, C., Schempp, W., Carrington, M., et al. (2009). A novel system of polymorphic and diverse NK cell receptors in primates. PLoS Genet. *5*, e1000688.

Bailly-Bechet, M., Haudry, A., and Lerat, E. (2014). "One code to find them all": a perl tool to conveniently parse RepeatMasker output files. Mob. DNA *5*, 13.

Baker, M.L., Schountz, T., and Wang, L.F. (2013). Antiviral immune responses of bats: a review. Zoonoses Public Health *60*, 104–116.

Banerjee, A., Rapin, N., Bollinger, T., and Misra, V. (2017). Lack of inflammatory gene expression in bats: a unique role for a transcription repressor. Sci. Rep. *7*, 2232.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Series B Stat. Methodol. *57*, 289–300.

Bielawski, J.P., and Yang, Z. (2005). Maximum likelihood methods for detecting adaptive protein evolution. In Statistical Methods in Molecular Evolution, R. Nielsen, ed. (New York: Springer-Verlag), pp. 103–124.

Biesold, S.E., Ritz, D., Gloza-Rausch, F., Wollny, R., Drexler, J.F., Corman, V.M., Kalko, E.K.V., Oppong, S., Drosten, C., and Müller, M.A. (2011). Type I interferon reaction to viral infection in interferon-competent, immortalized cell lines from the African fruit bat Eidolon helvum. PLoS ONE *6*, e28131.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. Bioinformatics *27*, 578–579.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

Broad Institute. Picard. http://broadinstitute.github.io/picard/.

Calisher, C.H., Childs, J.E., Field, H.E., Holmes, K.V., and Schountz, T. (2006). Bats: important reservoir hosts of emerging viruses. Clin. Microbiol. Rev. *19*, 531–545.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.

Connor, J.H., Yen, J., Caballero, I.S., Garamszegi, S., Malhotra, S., Lin, K., Hensley, L., and Goff, A.J. (2015). Transcriptional Profiling of the Immune Response to Marburg Virus Infection. J. Virol. *89*, 9865–9874.

De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics *22*, 1269–1271.

De La Cruz-Rivera, P.C., Kanchwala, M., Liang, H., Kumar, A., Wang, L.F., Xing, C., and Schoggins, J.W. (2018). The IFN response in bats displays distinctive IFN-stimulated gene expression kinetics with atypical RNASEL induction. J. Immunol. *200*, 209–217.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Fischer, S., Brunk, B.P., Chen, F., Gao, X., Harb, O.S., Iodice, J.B., Shanmugam, D., Roos, D.S., and Stoeckert, C.J., Jr. (2011). Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. Curr. Protoc. Bioinformatics *35*, 6.12.1–6.12.19.

Gifford, R.J. DIGS for EVEs. https://github.com/giffordlabcvr/DIGS-for-EVEs.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics *29*, 1072–1075.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. *41*, 95–98.

Hoffmann, H.-H., Schneider, W.M., and Rice, C.M. (2015). Interferons and viruses: an evolutionary arms race of molecular interactions. Trends Immunol. *36*, 124–138.

Hölzer, M., Krähling, V., Amman, F., Barth, E., Bernhart, S.H., Carmelo, V.A., Collatz, M., Doose, G., Eggenhofer, F., Ewald, J., et al. (2016). Differential

transcriptional responses to Ebola and Marburg virus infection in bat and human cells. Sci. Rep. *6*, 34589.

Jones, M.E., Schuh, A.J., Amman, B.R., Sealy, T.K., Zaki, S.R., Nichol, S.T., and Towner, J.S. (2015). Experimental inoculation of Egyptian Rousette bats (Rousettus aegyptiacus) with viruses of the Ebolavirus and Marburgvirus Genera. Viruses *7*, 3420–3442.

Kaiser, B.K., Pizarro, J.C., Kerns, J., and Strong, R.K. (2008). Structural basis for NKG2A/CD94 recognition of HLA-E. Proc. Natl. Acad. Sci. USA *105*, 6696–6701.

Kapustin, Y., Souvorov, A., Tatusova, T., and Lipman, D. (2008). Splign: algorithms for computing spliced alignments with identification of paralogs. Biol. Direct *3*, 20.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

Kepler, T.B., Sample, C., Hudak, K., Roach, J., Haines, A., Walsh, A., and Ramsburg, E.A. (2010). Chiropteran types I and II interferon genes inferred from genome sequencing traces by a statistical gene-family assembler. BMC Genomics *11*, 444.

Kikuchi-Maki, A., Catina, T.L., and Campbell, K.S. (2005). Cutting edge: KIR2DL4 transduces signals into human NK cells through association with the Fc receptor γ protein. J. Immunol. *174*, 3859–3863.

Koch, J., Steinle, A., Watzl, C., and Mandelboim, O. (2013). Activating natural cytotoxicity receptors of natural killer cells in cancer and infection. Trends Immunol. *34*, 182–191.

Kolde, R. (2012). Pheatmap: pretty heatmaps. https://cran.r-project.org/web/packages/pheatmap/.

Kulski, J.K., Shiina, T., Anzai, T., Kohara, S., and Inoko, H. (2002). Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. Immunol. Rev. *190*, 95–122.

Kwiecinski, G.G., and Griffiths, T.A. (1999). Rousettus egyptiacus. Mamm. Species *611*, 1–9.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lanier, L.L. (2009). DAP10- and DAP12-associated receptors in innate immunity. Immunol. Rev. *227*, 150–160.

Lee, A.K., Kulcsar, K.A., Elliott, O., Khiabanian, H., Nagle, E.R., Jones, M.E., Amman, B.R., Sanchez-Lockhart, M., Towner, J.S., Palacios, G., and Rabadan, R. (2015). De novo transcriptome reconstruction and annotation of the Egyptian rousette bat. BMC Genomics *16*, 1033.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, arXiv:1303.3997 [q-bio.GN].

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., Huang, Q., Cai, Q., Li, B., Bai, Y., et al. (2010). The sequence and de novo assembly of the giant panda genome. Nature *463*, 311–317.

Liu, X., Speranza, E., Muñoz-Fontela, C., Haldenby, S., Rickett, N.Y., Garcia-Dorival, I., Fang, Y., Hall, Y., Zekeng, E.G., Lüdtke, A., et al. (2017). Transcriptomic signatures differentiate survival from fatal outcomes in humans infected with Ebola virus. Genome Biol. *18*, 4.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. *25*, 955–964.

Maldonado, E., Almeida, D., Escalona, T., Khan, I., Vasconcelos, V., and Antunes, A. (2016). LMAP: lightweight multigene analyses in PAML. BMC Bioinformatics *17*, 354.

Malireddi, R.K.S., and Kanneganti, T.-D. (2013). Role of type I interferons in inflammasome activation, cell death, and disease during microbial infection. Front. Cell. Infect. Microbiol. *3*, 77.

Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics *27*, 764–770.

Mi, H., Poudel, S., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2016). PANTHER version 10: expanded protein families and functions, and analysis tools. Nucleic Acids Res. *44* (D1), D336–D342.

Molfetta, R., Quatrini, L., Zitti, B., Capuano, C., Galandrini, R., Santoni, A., and Paolini, R. (2016). Regulation of NKG2D expression and signaling by endocytosis. Trends Immunol. *37*, 790–802.

Morgulis, A., Gertz, E.M., Schäffer, A.A., and Agarwala, R. (2006). WindowMasker: window-based masker for sequenced genomes. Bioinformatics *22*, 134–141.

Mudunuri, U., Che, A., Yi, M., and Stephens, R.M. (2009). bioDBnet: the biological database network. Bioinformatics *25*, 555–556.

National Center for Biotechnology Information. Gnomon - the NCBI eukaryotic gene prediction tool. https://www.ncbi.nlm.nih.gov/genome/annotation_euk/gnomon/.

NCBI Eukaryotic Annotation Group (2016). NCBI Rousettus aegyptiacus Annotation Release 100. https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Rousettus_aegyptiacus/100/.

Ng, J.H., Tachedjian, M., Deakin, J., Wynne, J.W., Cui, J., Haring, V., Broz, I., Chen, H., Belov, K., Wang, L.F., and Baker, M.L. (2016). Evolution and comparative analysis of the bat MHC-I region. Sci. Rep. *6*, 21256.

O'Callaghan, C.A. (2000). Molecular basis of human natural killer cell recognition of HLA-E (human leucocyte antigen-E) and its relevance to clearance of pathogen-infected and tumour cells. Clin. Sci. (Lond.) *99*, 9–17.

Olival, K.J., Hosseini, P.R., Zambrana-Torrelio, C., Ross, N., Bogich, T.L., and Daszak, P. (2017). Host and viral traits predict zoonotic spillover from mammals. Nature *546*, 646–650.

Papenfuss, A.T., Baker, M.L., Feng, Z.P., Tachedjian, M., Crameri, G., Cowled, C., Ng, J., Janardhana, V., Field, H.E., and Wang, L.F. (2012). The immune gene repertoire of an important viral reservoir, the Australian black flying fox. BMC Genomics *13*, 261.

Parham, P., and Moffett, A. (2013). Variable NK cell receptors and their MHC class I ligands in immunity, reproduction and human evolution. Nat. Rev. Immunol. *13*, 133–144.

Parker, J., Tsagkogeorga, G., Cotton, J.A., Liu, Y., Provero, P., Stupka, E., and Rossiter, S.J. (2013). Genome-wide signatures of convergent evolution in echolocating mammals. Nature *502*, 228–231.

Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics *23*, 1061–1067.

Petrie, E.J., Clements, C.S., Lin, J., Sullivan, L.C., Johnson, D., Huyton, T., Heroux, A., Hoare, H.L., Beddoe, T., Reid, H.H., et al. (2008). CD94-NKG2A recognition of human leukocyte antigen (HLA)-E bound to an HLA class I leader sequence. J. Exp. Med. *205*, 725–735.

Rapaport, A.S., Schriewer, J., Gilfillan, S., Hembrador, E., Crump, R., Plougastel, B.F., Wang, Y., Le Friec, G., Gao, J., Cella, M., et al. (2015). The inhibitory receptor NKG2A sustains virus-specific $CD8^+$ T cells in response to a lethal poxvirus infection. Immunity *43*, 1112–1124.

Salmela, L., and Rivals, E. (2014). LoRDEC: accurate and efficient long read error correction. Bioinformatics *30*, 3506–3514.

Schuh, A.J., Amman, B.R., Jones, M.E., Sealy, T.K., Uebelhoer, L.S., Spengler, J.R., Martin, B.E., Coleman-McCray, J.A., Nichol, S.T., and Towner, J.S. (2017). Modelling filovirus maintenance in nature by experimental transmission of Marburg virus between Egyptian rousette bats. Nat. Commun. *8*, 14446.

Secombes, C.J., and Zou, J. (2017). Evolution of interferons and interferon receptors. Front. Immunol. *8*, 209.

Seim, I., Fang, X., Xiong, Z., Lobanov, A.V., Huang, Z., Ma, S., Feng, Y., Turanov, A.A., Zhu, Y., Lenz, T.L., et al. (2013). Genome analysis reveals insights into physiology and longevity of the Brandt's bat Myotis brandtii. Nat. Commun. *4*, 2212.

Shaw, T.I., Srivastava, A., Chou, W.C., Liu, L., Hawkinson, A., Glenn, T.C., Adams, R., and Schountz, T. (2012). Transcriptome sequencing and annotation for the Jamaican fruit bat (Artibeus jamaicensis). PLoS ONE *7*, e48472.

Smit, A.F.A., Hubley, R., and Green, P. RepeatMasker Open-4.0. http://www.repeatmasker.org.

Smith, I., and Wang, L.F. (2013). Bats and their virome: an important source of emerging viruses capable of infecting humans. Curr. Opin. Virol. *3*, 84–91.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. *30*, 2725–2729.

Thibaud-Nissen, F., Souvorov, A., Murphy, T., DiCuccio, M., and Kitts, P. (2013). Eukaryotic genome annotation pipeline. In The NCBI Handbook (Bethesda: National Center for Biotechnology Information [US]).

Towner, J.S., Amman, B.R., Sealy, T.K., Carroll, S.A., Comer, J.A., Kemp, A., Swanepoel, R., Paddock, C.D., Balinandi, S., Khristova, M.L., et al. (2009). Isolation of genetically diverse Marburg viruses from Egyptian fruit bats. PLoS Pathog. *5*, e1000536.

Tsagkogeorga, G., Müller, S., Dessimoz, C., and Rossiter, S.J. (2017). Comparative genomics reveals contraction in olfactory receptor genes in bats. Sci. Rep. *7*, 259.

UniProt Consortium (2015). UniProt: a hub for protein information. Nucleic Acids Res. *43*, D204–D212.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE *9*, e112963.

Warren, R.L., Yang, C., Vandervalk, B.P., Behsaz, B., Lagman, A., Jones, S.J., and Birol, I. (2015). LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. Gigascience *4*, 35.

Whitlow, Z.W., Connor, J.H., and Lyles, D.S. (2006). Preferential translation of vesicular stomatitis virus mRNAs is conferred by transcription from the viral genome. J. Virol. *80*, 11733–11742.

Whitlow, Z.W., Connor, J.H., and Lyles, D.S. (2008). New mRNAs are preferentially translated during vesicular stomatitis virus infection. J. Virol. *82*, 2286–2294.

Xue, W., Li, J.T., Zhu, Y.P., Hou, G.Y., Kong, X.F., Kuang, Y.Y., and Sun, X.W. (2013). L_RNA_scaffolder: scaffolding genomes with transcripts. BMC Genomics *14*, 604.

Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. *15*, 568–573.

Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., et al. (2016). Ensembl 2016. Nucleic Acids Res. *44* (D1), D710–D716.

Ye, C., Ma, Z.S., Cannon, C.H., Pop, M., and Yu, D.W. (2012). Exploiting sparseness in de novo genome assembly. BMC Bioinformatics *13* (*Suppl 6*), S1.

Ye, C., Hill, C.M., Wu, S., Ruan, J., and Ma, Z.S. (2016). DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. Sci. Rep. *6*, 31900.

Yokoyama, W.M., and Plougastel, B.F.M. (2003). Immune functions encoded by the natural killer gene complex. Nat. Rev. Immunol. *3*, 304–316.

Zhang, G., Cowled, C., Shi, Z., Huang, Z., Bishop-Lilly, K.A., Fang, X., Wynne, J.W., Xiong, Z., Baker, M.L., Zhao, W., et al. (2013). Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. Science *339*, 456–460.

Zhang, Q., Zeng, L.P., Zhou, P., Irving, A.T., Li, S., Shi, Z.L., and Wang, L.F. (2017). IFNAR2-dependent gene expression profile induced by IFN-α in Pteropus alecto bat cells and impact of IFNAR2 knockout on virus infection. PLoS ONE *12*, e0182866.

Zhou, P., Tachedjian, M., Wynne, J.W., Boyd, V., Cui, J., Smith, I., Cowled, C., Ng, J.H., Mok, L., Michalski, W.P., et al. (2016). Contraction of the type I IFN locus and unusual constitutive expression of IFN-α in bats. Proc. Natl. Acad. Sci. USA *113*, 2696–2701.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
Sendai virus, Cantell strain	Charles River Laboratories	Cat# 10100774
Vesicular stomatitis virus-eGFP	Whitlow et al., 2006, 2008; (Laboratory of John Connor, Boston University School of Medicine)	N/A
Biological Samples		
Healthy, wild-caught, older juvenile male Egyptian Rousette bat (*R. aegyptiacus*) – liver and spleen tissue	This paper	https://www.ncbi.nlm.nih.gov/biosample/SAMN04287759
Chemicals, Peptides, and Recombinant Proteins		
Recombinant 6xHis-IFN-β1 (R. aegyptiacus protein)	This paper (Blue Heron Biotech)	N/A
Recombinant 6xHis-IFN-ω4 (R. aegyptiacus protein)	This paper (Blue Heron Biotech)	N/A
Recombinant 6xHis-PA-D1 (Domain 1 of *B. anthracis* PA protein)	This paper (Blue Heron Biotech)	N/A
Human: FreeStyle 293-F Cells	Thermo Fisher Scientific	Cat# R79007
FreeStyle MAX Reagent	Thermo Fisher Scientific	Cat# 16447500
OptiPRO SFM (1X)	Thermo Fisher Scientific	Cat# 12309-050
Lysis/Binding Solution Concentrate	Thermo Fisher Scientific	Cat# 4462362
Critical Commercial Assays		
Capturem His-tagged purification maxiprep kit	Clontech, Takara Bio	Cat# 635713
Vivaspin 2 Protein Concentrators, MWCO 10000	GE Life Sciences	Cat# 28932247
UltraClean Blood DNA Isolation Kit (customized protocol)	MO BIO Laboratories	Cat# 12000-100
Prep X Complete ILMN DNA Library Kit	WaferGen Biosystems	Cat# 640101
KAPA Library Amplification kit	KAPA Biosystems	Cat# KK2621
KAPA Library Quantification kit	KAPA Biosystems	Cat# KK4824
TruSeq PE Cluster Kit v3-cBot-HS	Illumina	Cat# PE-401-3001
TruSeq SBS Kit – HS (200 cycle)	Illumina	Cat# FC-401-3001
MagMAX-96 Total RNA Isolation Kit	Ambion	Cat# AM1830
SMRTbell Template Prep Kit 1.0	Pacific Biosciences	Cat# 100-259-100
DNA/Polymerase Binding Kit P4	Pacific Biosciences	Cat# 100-236-500
DNA/Polymerase Binding Kit P5	Pacific Biosciences	Cat# 100-256-000
DNA Sequencing Reagent 2.0	Pacific Biosciences	Cat# 100-216-400
DNA Sequencing Reagent 3.0	Pacific Biosciences	Cat# 100-254-800
SMRT Cell 8Pack V3	Pacific Biosciences	Cat# 100-171-800
Bioanalyzer DNA 12000 Kit	Agilent	Cat# 5067-1508
TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Human/Mouse/Rat High Throughput (96 samples, 96 indexes)	Illumina	Cat# 20020597
HiSeq PE Cluster Kit V4 - cBot	Illumina	Cat# PE-401-4001
HiSeq SBS Kit V4 250 cycle kit	Illumina	Cat# FC-401-4003

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
High Sensitivity D1000 ScreenTape	Agilent	Cat# 5067-5582
High Sensitivity D1000 Reagents	Agilent	Cat# 5067-5583
Deposited Data		
Raegyp2.0 genome sequence	This paper	GenBank: GCA_001466805.2; RefSeq: GCF_001466805.2; GenBank: LOCP00000000.2
R. aegyptiacus transcriptome raw and analyzed data	Lee et al., 2015	GenBank: GECF00000000.1; SRA Project: SRP066106
Sendai virus-infected RoNi cell transcriptome data	This paper	NCBI GEO: GSE108941
Ensembl database release 90	Yates et al., 2016	http://aug2017.archive.ensembl.org/index.html
RefSeq proteins – Pteropus vampyrus	Baylor College of Medicine; NCBI RefSeq	RefSeq: GCF_000151845.1
RefSeq proteins – Pteropus alecto	Zhang et al., 2013; NCBI RefSeq	RefSeq: GCF_000325575.1
RefSeq proteins – Myotis davidii	Zhang et al., 2013; NCBI RefSeq	RefSeq: GCF_000327345.1
RefSeq proteins – Myotis lucifugus	Broad Institute; NCBI RefSeq	RefSeq: GCF_000147115.1
RefSeq proteins – Macaca fascicularis	Washington University; NCBI RefSeq	RefSeq: GCF_000364345.1
RefSeq proteins – Macaca mulatta	Baylor College of Medicine Genome Sequencing Center; NCBI RefSeq	RefSeq: GCF_000772875.2
RefSeq proteins – Cavia porcellus	The Genome Sequencing Platform, The Genome Assembly Team; NCBI RefSeq	RefSeq: GCF_000151735.1
RefSeq proteins – *Cricetulus griseus*	Beijing Genomics Institute; NCBI RefSeq	RefSeq: GCF_000223135.1
RefSeq proteins – Equus caballus	The Genome Assembly Team; NCBI RefSeq	RefSeq: GCF_000002305.2
RefSeq proteins – *Mus musculus*	Genome Reference Consortium; NCBI RefSeq	RefSeq: GCF_000001635.24
RefSeq proteins – Sus scrofa	The Swine Genome Sequencing Consortium (SGSC); NCBI RefSeq	RefSeq: GCF_000003025.5
RefSeq proteins – Bos taurus	Cattle Genome Sequencing International Consortium; NCBI RefSeq	RefSeq: GCF_000003205.7
RefSeq proteins – Canis familiaris	Dog Genome Sequencing Consortium; NCBI RefSeq	RefSeq: GCF_000002285.3
RefSeq proteins – Homo sapiens	Genome Reference Consortium; NCBI RefSeq	RefSeq: GCF_000001405.33
RefSeq proteins – Rousettus aegyptiacus	This paper	RefSeq: GCF_001466805.2
Experimental Models: Cell Lines		
R. aegyptiacus: kidney fibroblast cells (RoNi)	Biesold et al., 2011	N/A
Recombinant DNA		
pCAGGS/6xHis–IFN-ω4	This paper	N/A
pCAGGS/6xHis–IFN-β1	This paper	N/A
pET22b/6xHis-PA-D1	This paper	N/A
Software and Algorithms		
Jellyfish	Marçais and Kingsford, 2011	https://github.com/gmarcais/Jellyfish/releases
SparseAssembler	Ye et al., 2012	https://sourceforge.net/projects/sparseassembler/
LoRDEC v0.5	Salmela and Rivals, 2014	https://gite.lirmm.fr/lordec/lordec-releases/wikis/home
DBG2OLC	Ye et al., 2016	https://github.com/yechengxi/DBG2OLC
LINKS v1.5.1	Warren et al., 2015	https://github.com/warrenlr/LINKS

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
L_RNA_Scaffolder (downloaded 9/28/2015)	Xue et al., 2013	http://www.fishbrowser.org/software/L_RNA_scaffolder/index.php/Home/Index/downloads.html
SSPACE v3.0	Boetzer et al., 2011	https://github.com/nsoranzo/sspace_basic
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Pilon	Broad Institute; Walker et al., 2014	https://github.com/broadinstitute/pilon/releases/
Quiver	Pacific Biosciences	https://github.com/PacificBiosciences/GenomicConsensus
QUAST v3.0	Gurevich et al., 2013	http://bioinf.spbau.ru/quast
CEGMA v2.5	Parra et al., 2007	http://korflab.ucdavis.edu/datasets/cegma/
bwa-mem v0.7.10	Li, 2013	https://sourceforge.net/projects/bio-bwa/files/
SAMtools v0.1.18; v1.3	Li et al., 2009	http://samtools.sourceforge.net/; http://www.htslib.org/
NCBI Eukaryotic Genome Annotation Pipeline; Gnomon tRNAscan-SE	Thibaud-Nissen et al., 2013; National Center for Biotechnology Information n.d.; Lowe and Eddy, 1997	https://www.ncbi.nlm.nih.gov/genome/annotation_euk/; https://www.ncbi.nlm.nih.gov/genome/annotation_euk/gnomon/
WindowMasker	Morgulis et al., 2006	ftp://ftp.ncbi.nih.gov/toolbox/ncbi_tools++/CURRENT/
Splign; ProSplign	Kapustin et al., 2008	https://www.ncbi.nlm.nih.gov/sutils/splign/splign.cgi
OrthoMCL v2.0.9	Fischer et al., 2011	http://orthomcl.org/common/downloads/software/v2.0/
bioDB	Mudunuri et al., 2009	https://biodbnet-abcc.ncifcrf.gov/db/db2db.php
RepeatMasker	Smit et al. n.d.	http://www.repeatmasker.org/RMDownload.html
"One code to find them all" script	Bailly-Bechet et al., 2014	http://doua.prabi.fr/software/one-code-to-find-them-all
PicardTools v1.131	Broad Institute	http://broadinstitute.github.io/picard/
BCFTools v1.3.1	Li et al., 2009	http://www.htslib.org/
Mafft v7.305b	Katoh and Standley, 2013	https://mafft.cbrc.jp/alignment/software/
trimAL v1.3; v1.4	Capella-Gutiérrez et al., 2009	http://trimal.cgenomics.org/downloads
RAxML v8.2.9	Stamatakis, 2014	https://github.com/stamatak/standard-RAxML
CAFE v3.1	De Bie et al., 2006	https://github.com/hahnlab/CAFE
Mega 6.0	Tamura et al., 2013	https://www.megasoftware.net/
MUSCLE v3.8.31	Edgar, 2004	http://www.drive5.com/muscle/
kallisto v.0.43.0	Bray et al., 2016	https://pachterlab.github.io/kallisto/download
pheatmap	Kolde 2012	https://cran.r-project.org/web/packages/pheatmap/
PAML v4.9b	Bielawski and Yang, 2005; Yang 1998	http://abacus.gene.ucl.ac.uk/software.html
LMAP v1.0.0	Maldonado et al., 2016	http://lmapaml.sourceforge.net/
BioEdit v7.0.0	Hall, 1999	http://www.mbio.ncsu.edu/BioEdit/bioedit.html
Trimmomatic-0.33	Bolger et al., 2014	http://www.usadellab.org/cms/?page=trimmomatic

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DIGS for EVE, EVE library v0.1	Gifford n.d.	https://github.com/giffordlabcvr/DIGS-for-EVEs
Other		
PANTHER database	Mi et al., 2016	http://www.pantherdb.org/
Ensembl release 90	Yates et al., 2016	http://www.ensembl.org/index.html?redirect=no

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Gustavo Palacios (gustavo.f.palacios.ctr@mail.mil).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Source organism

Genomic DNA was isolated from the liver and spleen tissue of a wild-caught healthy older juvenile male Egyptian rousette bat captured at Python Cave in Uganda. Research was conducted under an IACUC approved protocol in compliance with the Animal Welfare Act, PHS Policy, and other Federal statutes and regulations relating to animals and experiments involving animals. The facility where this research was conducted is accredited by the Association for Assessment and Accreditation of Laboratory Animal Care, International and adheres to principles stated in the Guide for the Care and Use of Laboratory Animals, National Research Council, 2011.

Cell lines

Rousettus aegyptiacus immortalized fibroblasts (RoNi) were originally generated in the work described in Biesold et al. (2011). Briefly, kidney tissue from a wild-caught sub-adult *Egyptian rousette* bat was cultured and immortalized via lentiviral transduction of the SV40 large T antigen. The sex of the bat was not recorded at the time of capture. Cell cultures were genotyped and tested for mycoplasma, SV5, filoviurses and lyssaviruses by RT-PCR. Cells were cultured at 37°C at 5% CO_2 in DMEM (Dulbecco's Modified Eagles Medium) with 4.5g L glucose supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin 100x conentrate, 1% L-Glutamine 200mM, 1% Sodium Pyruvate 100mM, and 1% MEM nonessential amino acids 100x concentrate.

METHOD DETAILS

Nucleic acid extraction and sequencing

We used the UltraClean Blood DNA Isolation kit (MO BIO Laboratories, Carlsbad, CA) with a customized protocol for tissue (available upon request) to isolate genomic DNA from liver and spleen tissue of a healthy older juvenile male *Rousettus aegyptiacus* bat captured at Python Cave in Uganda. For sequencing on the Illumina platform, 1 ug of genomic DNA was sheared to 400 bp using Covaris LE220 Focused-ultrasonicator (Covaris, Woburn, MA). End repair, A-tailing and ligation of adapters were performed on the Apollo 324 automated system, using Prep X Complete ILMN DNA Library Kit (WaferGen Biosystems, Fremont, CA). KAPA Library Amplification Kit with 10 cycles of PCR was used for library enrichment. Libraries were quantified by qPCR using KAPA Library Quantification Kit (KAPA Biosystems, Wilmington, MA). Each library was loaded on 8 lanes of the high output flow cell. Cluster amplification was performed on the cBot with the TruSeq PE Cluster Kit v3-cBot-HS (Illumina, San Diego, CA). Clustered flow cell was sequenced on the HiSeq 2500 instrument with the TruSeq SBS Kit –HS (Illumina). 720 Gb of 2x101bp paired-end data were produced (approximate coverage of 145x).

For sequencing on the PacBio platform, genomic DNA was sheared to 20kb average size using g-TUBE (Covaris). After DNA damage repair and ends repair, hairpin adapters were ligated to form a SMRTbell template. ExoIII and ExoVII treatment was used to remove failed ligation products. Size selection was performed on Blue Pippin system (Sage Sciences, Beverly, MA) using 0.75% dye-free agarose gel cassette, marker S1 and Hi-Pass protocol; low cut was set on 4000 bp. Final library assessment was obtained by Qubit dsDNA BR assay and Agilent 2100 Bioanalyzer DNA 12000 chip. To obtain longer reads, libraries were sequenced with P5-C3 chemistry. Annealing of sequencing primer and binding polymerase P5 to the SMRTbell template was performed according to PacBio calculator. The polymerase-template complexes were bound to MagBeads, loaded onto SMRTcells (SMRT Cell 8 pack V3) at final concentration 180 pM, and sequenced with 180 min movies on PacBio RS II instrument (Pacific Biosciences, Menlo Park, CA).

Genome Size Estimation and Heterozygosity analysis

Genome size was estimated by k-mer analysis. K-mers from 4 lanes of the Illumina dataset were counted using Jellyfish (Marçais and Kingsford, 2011). Excluding rare 25-mers that occur at low depth, the most frequently occurring depth was 57x (Figure S1). The presence of multiple peaks suggests that this genome has a high degree of heterozygosity. Peak k-mer frequency (M), real sequencing depth (N), read length (L), k-mer length (K), total bases (T), and genome size (G) are related by the following formulas (Li et al., 2010):

$$M = N * (L - K + 1)/L$$

$$G = T/N$$

Using this method, we estimate a complete genome size of 2.08 Gb, which is very close to previous estimates of 2.11 Gb based on nuclear densitometry (Kwiecinski and Griffiths, 1999) (Figure S1A).

Genome Assembly

Illumina reads were assembled separately from the PacBio reads with SparseAssembler, a short read assembler that exploits high coverage to construct a modified de Bruijn graph (DBG) (Ye et al., 2012). Because raw Pacific Biosciences (PacBio) long sequencing reads can be error-prone, we corrected potential sequencing errors in these reads with ~34x of high-accuracy Illumina reads using LoRDEC v0.5 (long read de Bruijn graph error correction) (Salmela and Rivals, 2014). Contiguity of the short-read assembly was improved by incorporating long-read data using DBG2OLC (Ye et al., 2016), which anchors the short-read contigs generated with SparseAssembler to the long PacBio reads.

Multiple approaches were used to scaffold the assembly. First, we used PacBio reads with the long-read scaffolding program LINKS v1.5.1, which uses a paired k-mer approach to scaffold assemblies with long reads even if they have already been used in the assembly (Warren et al., 2015). LINKS was used iteratively until no improvement was observed in contiguity, which was after seven iterations. Second, we made use of transcriptome data recently published (Lee et al., 2015) to scaffold our assembly with L_RNA_Scaffolder (downloaded 9/28/2015) (Xue et al., 2013). Last, we reincorporated the paired-end information from our Illumina data for scaffolding with SSPACE v3.0 (Boetzer et al., 2011). As a post-processing step, we aligned the Illumina reads to the assembly using Bowtie2 (Langmead and Salzberg, 2012) and ran the variant calling program Pilon (Walker et al., 2014) on the resulting alignment to correct potential mis-assemblies, consensus calling errors, or homopolymer indel errors. We also tested the use of the Quiver algorithm from Pacific Biosciences as a post-processing step prior to pilon, but found that it reintroduced insertion and deletion errors that led to poor gene models. The genome-wide average GC content is about 40%, and intra-scaffold GC content ranges from 31.1% to 71.4%. Raegyp2.0 has a higher degree of heterozygosity (0.53%, Figure S1B) than has been reported for other bat genomes (Zhang et al., 2013).

To check for mis-assemblies, we used the gene coverage program CEGMA v2.5 with mammalian settings to identify the presence and coverage of highly conserved eukaryotic genes in the genome (Parra et al., 2007). We also remapped ~54x of short paired reads onto the assembly with bwa-mem v0.7.10 and assessed the number of reads that map appropriately (in the correct orientation and with the expected insert size) with SAMtools flagstat (SAMtools v0.1.18) (Li 2013; Li et al., 2009). 98.41% of reads align, and 91.19% of reads map appropriately, indicating no major misassemblies.

To assess the quality of the Raegyp2.0 assembly, we used a statistical analysis program, QUAST v3.0 (Gurevich et al., 2013), to look at baseline statistics, including the total assembly length, the percentage of estimated genome size covered, and contiguity statistics. The Nx and NGx plots (generated using QUAST) in Figure S1 show the shortest sequence for which the total length of all sequences of its length or longer make up "x" percentage of the total assembly size (Nx) or the total estimated reference genome size (NGx). For example, In a length-sorted list of sequences, the N50 refers to the length of the smallest sequence for which the sum of the length of all sequences of the same or longer length constitute 50% of the total assembly length. The scaffold N50 refers to the N50 of all the scaffolds in the genome, while the contig N50 refers to the N50 of all the contigs in the genome. To compare genome contiguity statistics across all available bat genome projects (Table 1), we accessed the appropriate GenBank Genome page for each species and reported contiguity statistics as listed (accessed on 8/8/17).

Genome Annotation

The assembly was submitted to GenBank (accession GCA_001466805.2) and annotated using the NCBI eukaryotic genome annotation pipeline (Thibaud-Nissen et al., 2013). A total of 26.25% of the genomic sequence was identified as repetitive by the *de novo* repeat finder WindowMasker (Morgulis et al., 2006) and was masked for the purpose of aligning evidence and predicting genes. Transcripts and proteins available in GenBank and known RefSeq transcripts and proteins for bats and human were aligned to the genome with Splign (Kapustin et al., 2008) and ProSplign, along with model RefSeq proteins previously annotated on the *Pteropus alecto* and *Myotis brandtii* genomes. In addition, over 2 billion RNA-Seq reads derived from 12 different *Rousettus aegyptiacus* tissues and available in SRA were aligned. Model precursors were created by Gnomon (National Center for Biotechnology Information n.d.), a gene calling algorithm trained on *Pteropus alecto*, by collapsing overlapping alignments with compatible splice

patterns. In a second step also run by Gnomon, model precursors with high coding propensity were extended or joined if missing start or stop codon using an HMM model. The resulting models were evaluated and filtered based on multiple criteria, including supporting evidence, conflicts with models on the other genomic strand, Blast hits to UniProtKB/SwissProt if over 50% *ab initio* sequence, and number of exons (single-exon non-coding RNA were eliminated). Models in the final set were assigned a function by orthology calculation to human, or if no ortholog could be calculated, by homology to proteins in UniProtKB/SwissProt. Finally, models were assigned GeneIDs and RefSeq model transcript and protein accession (with XM_, XP_, XR_ prefixes) and loaded to the Nucleotide, Protein and Gene databases as part of NCBI *Rousettus aegyptiacus* Annotation Release 100. The genome contains 36.4% repetitive content, similar to that found in other bats (Table S1). 19,668 of the annotated genes were protein-coding genes and 2,380 were non-coding genes. 2,198 coding sequences were corrected for premature stop codons, small internal gaps, or frameshifts based on aligning evidence (NCBI Eukaryotic Annotation Group, 2016). When present, the corrected models were used for downstream analysis. 5,958 long non-coding RNAs were predicted with full support from transcript data and 340 tRNA models were predicted with tRNAscan-SE (Lowe and Eddy, 1997). 96.03% of predicted mRNAs were fully supported by transcriptomic or protein data. A total of 16,254 of predicted protein-coding genes (83%) had at least one protein aligning to a UniProtKB/SwissProt protein for over 95% of its length, which was higher than for any other bat annotated by the NCBI pipeline at the time. In addition, the number of the UniProtKB/SwissProt hits covered over 95% of their length by a predicted protein was similarly high (82%), indicating that the predicted proteins on Raegyp2.0 represent full-length models. Of the 19,668 protein-coding genes annotated in the genome, 317 genes have no associated gene labels or homolog in open reading frames from genomes of other species in RefSeq, and may represent novel *R. aegyptiacus*-specific genes. However, the median length of the protein product of these genes (257 amino acids) is significantly lower than the median length of all annotated proteins (498 amino acids). Thus, it remains a possibility that these genes have been previously identified in other species but are partial or poor models in Raegyp2.0 because of misannotation or local misassembly.

Designating gene families

We inferred homologous protein groups among 15 species of mammals using a similarity-based approach within the OrthoMCL pipeline (Fischer et al., 2011). The following species were included in the analysis: *Rousettus aegyptiacus* (Egyptian rousette bat), *Pteropus vampyrus* (large flying fox), *Pteropus alecto* (black flying fox), *Myotis davidii, Myotis lucifugus* (little brown bat), *Homo sapiens, Macaca fascicularis* (crab-eating macaque), *Macaca mulatta* (rhesus macaque), *Mus musculus* (mouse), *Cavia porcellus* (guinea pig), *Cricetulus griseus* (Chinese hamster), *Sus scrofa* (pig), *Bos taurus* (cow), *Equus caballus* (horse), *Canis familiaris* (dog). Protein data were downloaded from RefSeq and filtered to include only the longest protein product of a gene for use in the pipeline. Briefly, OrthoMCL gathers all-against-all blastp hits into reciprocal best hits (between species) and reciprocal better hits (within species). These hits are converted into a graph network describing likely orthologous or paralogous relationships among proteins, and MCL clustering is performed to group proteins into families. Using OrthoMCL v2.0.9, we obtained 19,310 groups of proteins and 7,041 single-copy orthologous groups.

We annotated all OrthoMCL groups with functional family designations from the PANTHER database (Mi et al., 2016). We first labeled each group with a representative RefSeq protein accession from that group, using a protein from a well-curated genome (human, mouse, macaque) whenever possible. We mapped all RefSeq accessions directly to PANTHER family IDs or to UniProtKB accessions for subsequent mapping to PANTHER IDs using bioDB, PantherDB, and/or the UniProtKB accession mapping feature (Mi et al., 2016; Mudunuri et al., 2009; UniProt Consortium, 2015). All OrthoMCL groups with the same PANTHER family ID were collapsed into new gene families, which produced a total of 9,555 gene families (2,400 single-copy orthologous gene families). Some families remain unlabeled after this process because proteins from relatively new genomes may be missing in PANTHER. Of note, 3,450 OrthoMCL groups did not map to PANTHER families and were designated unlabeled gene families for downstream analysis. Using this process, upon examining the type I IFN gene family, we observed that many genes were not accounted for despite being present in the NCBI annotation. We discovered that these genes were labeled as singletons or in families with only *R. aegyptiacus* proteins, and therefore would not be classified as a PANTHER family since no *R. aegyptiacus* proteins were in the PANTHER database at the time of analysis. We manually corrected the numbers of genes in this family based on NCBI annotations and yielding 9,550 families, 3,445 of which were unlabeled.

Identification of repetitive content

Repetitive content was identified based on homology to the RepBase database using RepeatMasker (Smit et al. n.d.). ‘‘One code to find them all’’ (Bailly-Bechet et al., 2014) was used to resolve nested repeats and provide family-level quantitative information.

Heterozygosity analysis

All short reads were mapped to the genome using Bowtie2 (Langmead and Salzberg, 2012). Duplicate reads were marked and removed with PicardTools v1.131 (Broad Institute n.d.). Variants were called using Samtools v1.3 (Li et al., 2009). BCFTools v1.3.1 was used to filter variant calls in regions with parameters –g3 –G10 –e ‘%QUAL < 20 || (RPB < 0.1 && %QUAL < 30 || (DP < 30) || (DP > 250) || (MQ < 20)’.

Endogenous viral content

All genomes were screened for endogenous viral elements (EVEs) with DIGS for EVEs (Gifford n.d.) using the EVEs v0.1 reference library composed of EVEs derived from viruses in the Parvoviridae, Bornaviridae, Hepadnaviridae, Circoviridae, Filoviridae, and Bunyaviridae families.

Species tree generation

We extracted all 2,400 single-copy orthologous proteins (inferred by methods described above) and performed multiple sequence alignments of each group with Mafft v7.305b (Katoh and Standley, 2013). All alignments were trimmed with trimAL v1.3 (Capella-Gutiérrez et al., 2009) using the -automated1 parameter, and concatenated into a super-protein for each species. After concatenation, the super-proteins were re-trimmed with trimAL (-automated1 parameter) and used to generate a maximum likelihood species tree with RAxML v8.2.9 under a JTT + Γ substitution model with empirical base frequencies (Stamatakis, 2014). 1,000 bootstrap replicates were used to assess branch reliability.

Gene family expansion and contraction analysis

Gene families from the proteomes of 15 mammals (as designated in above methods) were used as input for the CAFE v3.1 (Computational Analysis of gene Family Evolution) algorithm (De Bie et al., 2006) to estimate a value for lambda (the birth and death rate for a given gene per million years) by maximum likelihood. Since CAFE assumes that all input gene families have at least one member in the most recent common ancestor of all species, gene families were filtered to exclude single lineage families. Only families with at least one gene in both the Laurasiatheria superorder (bats, ungulates, carnivores) and the Euarchontoglires superorder (primates, rodents) were included. We manually corrected type I IFN gene family sizes after observing that several genes that were annotated as type I IFNs by the NCBI did not appear in our inferred gene families. Further exploration revealed that these additional IFNs get classified as single gene families or as unlabeled gene families because they consist of only bat IFNs. The resulting families were further filtered to exclude eight families with large ranges in size (> 100) across the tree, leaving 7,698 families for expansion and contraction analysis. CAFE was used to obtain a maximum likelihood estimate of a global birth and death rate parameter λ (lambda; rate of gain/loss per gene per million years) across the species tree (created by above methods) of 0.0169284. The size of each family at each ancestral node was estimated and used to obtain a family-wise p value to indicate non-random expansion or contraction for each family, as well as significant expansion or contraction across all branches of the species tree. Tables S3 and S4 contain all the gene families with expansions and contractions in *R. aegyptiacus* prior to and post IFN-correction (described above), respectively.

Annotation of genes

We used annotations derived from the NCBI pipeline to extract and characterize NK cell receptor genes and MHC genes. These annotations were supplemented with homology-based predictions from the gene family analysis described above, and manual correction after examination in BioEdit v7.0.0 (Hall, 1999). For example, out of the 12 genes classified as C-type lectins in our gene family analysis, one gene is missing exons and is considered a pseudogene. The NCBI gene database contains another gene annotated as an NKG2A-like gene, but was classified as a singleton in our gene family analysis because of missing exons. Additionally, one protein sequence identified as an NK cell receptor-like C-type lectin by our gene family analysis remained uncharacterized by the NCBI pipeline. Upon manual reannotation, we discovered that the uncharacterized protein sequence was actually a misannotation of two NKG2A-like genes—one complete pseudogene and one partial gene. Thus, this analysis is able to detect homologs and is robust to potential misannotation issues.

To annotate the type I IFN genes, we used IFN sequences inferred from *P. vampyrus* sequencing traces from Kepler et al. (2010) as queries in a blastn search with the following parameters: -task blastn -evalue 0.05. Resulting hits were screened based on the lowest e-value, and the sequences were extracted with blastdbcmd and examined in BioEdit v7.0.0 for coding potential (Altschul et al., 1990; Hall, 1999).

Gene annotations are shown in Figures 2, 4, 5, and S3. The scaffolds containing the NKC locus in Figure 2A are in order from top to bottom: NW_15493182.1, NW_15493451.1, NW_15493213.1, and NW_15494625.1.

The scaffolds containing the MHC class I locus and extra-locus genes in Figure 4 are, in order from top to bottom: panel A. NW_015494903.1, NW_015493289.1, and NW_015494931.1; panel B, column 1: NW_015493957.1, NW_015493337.1, NW_015493066.1, NW_015493167.1, NW_015493330.1, NW_015493360.1, and NW_015494802.1; B, column 2: NW_015494846.1, NW_015493471.1, NW_015492968.1, NW_015494660.1, and NW_015493352.1.

The scaffolds containing the type I IFN locus in Figure 5A are, in order from top to bottom – column 1: NW_015494712.1, NW_015494244.1, NW_015493479.1, NW_015493859.1, NW_015494258.1, NW_015492835., NW_015494622.1; column 2: NW_015493694.1, NW_015493581.1, NW_015493794.1, NW_015493974.1, NW_015494085.1, NW_015494371.1, NW_015494373.1, NW_015494147.1, NW_015494111.1, NW_015494299.1.

The scaffolds containing the NKC loci in Figure S3B are, in order from top to bottom: *P. vampyrus* - NW_011888897.1, NW_011889578.1, NW_011889241.1, NW_011889581.1, NW_011889318.1; *P. alecto* - NW_006431924.1, NW_006436696.1, NW_006429163.1, NW_006432008.1; *M. davidii* - NW_006299270.1, NW_006295002.1, NW_006281977.1, NW_006289839.1; *M. lucifugus* - NW_005871058.1, NW_005873184.1, NW_005874133.1.

To extract MHC class I putative nonamers, proteins derived from MHC class I genes (or the closest mouse homologs to HLA-A) were aligned, and nonamers determined based on location of known nonamer sequences (O'Callaghan, 2000; Kaiser et al., 2008).

NKG2D ligands were collected from the NCBI Gene database or from UniProt (UniProt Consortium, 2015). For bat proteins, all putative functional, in-frame, non-partial genes were identified based on their annotation as NKG2D ligand-like by the NCBI Eukaryotic Annotation pipeline. In the phylogenetic tree displaying NKG2D ligands (Figure S7), abbreviations are as follows: Hsap (*H. sapiens*), Raeg (*R. aegyptiacus*), Pvam (*P. vampyrus*), Efus (*E. fuscus*), Pale (*P. alecto*), Mluc (*M. lucifugus*), Mbra (*M. brandtii*), Mdav (*M. davidii*), Harm (*H. armiger*), Mnat (*M. natalensis*), Rsin (*R. sinicus*), Mmus (*M. musculus*), Rnor (*R. norvegicus*).

Phylogenetic analysis

Gene sequences were retrieved from Ensembl release 90 (Yates et al., 2016) by looking up the human ortholog (horse ortholog for IFND) and retrieving every annotated placental mammal ortholog (except NKG2D ligands as noted above). To this, we added a curated set of bat orthologs retrieved from GenBank. The protein sequences were then aligned in Mega 6.0 (Tamura et al., 2013) with MUSCLE (Edgar, 2004) using a maximum of 20 iterations. Mega 6.0 was then used to estimate the most likely model, which in all cases was JTT+Γ. This model was used to generate phylogenies for each gene. Sites were not considered for phylogenetic analysis if a deletion was present in > 5% of sequences (10% for NKG2A). Bootstrap confidence values were determined using 500 replicates.

Transcriptome analysis

Reads from the Lee et al. (2015) dataset (SRA project: SRP066106) (Lee et al., 2015) were pseudoaligned to the set of CDSs annotated in the Raegyp_2.0 assembly supplemented with manually annotated type I IFN genes and NKG2 genes using kallisto v0.43.0 (Bray et al., 2016). Gene-level transcript counts were calculated by summing the transcript per million (TPM) values for all transcripts of a given gene. The resulting TPM values were log2 transformed and heatmaps were generated using pheatmap (Kolde, 2012). For Table S6, TPM values of IFNs were normalized by dividing by GAPDH TPM in the same tissue.

Sendai virus infection of RoNi cells

RoNi cells in 10% DMEM (GIBCO) were seeded at $5x10^4$ in 24-well plates (Corning), then infected in triplicate by Sendai virus (SeV, Cantell strain, Charles River Laboratories, Wilmington, MA) at an MOI of 1, or mock infected. Fresh media was exchanged following a 1hr adsorption and plates were incubated at 37°C. Cells were harvested in 1x RNA Lysis/Binding Solution Concentrate (Thermo Fisher Scientific, Waltham, MA) at 3hr, 8hr and 24hr post-infection, followed by magnetic bead purification and TURBO DNase treatment using the MagMAX-96 Total RNA Isolation Kit (Ambion, Thermo Fisher Scientific) according to manufacturer guidelines. RNA extractions were performed on a MagMAX Express 96 Magnetic Particle Processor (Applied Biosystems). Purified total RNA samples were verified by NanoDrop spectrophotometer (Thermo Fisher Scientific) and stored at −80°C prior to use. RNA libraries were generated using the TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Human/Mouse/Rat High-Throughput kit (Illumina). The completed libraries were screened for quality with the High Sensitivity D1000 Screentape and Reagents on the Tapestation 2200 (Agilent) and the Library Qauntification kit (KAPA Biosystems). RNA-sequencing was performed using a dual-index paired end (2x125 bp) format on Illumina HiSeq 2500 with the Hiseq SBS Kit v4 (250 cycles) and the HiSeq PE Cluster Kit v4 (Illumina). Raw reads were cleaned with Trimmomatic v0.33 (Bolger et al., 2014), aligned to the *R. aegyptiacus* transcriptome (from NCBI, supplemented with manually annotated type I IFN genes and NKG2 genes) and quantified in TPM using kallisto v0.43.0 as described above (Bray et al., 2016).

Production of recombinant 6xHis-tagged IFNs

Human FreeStyle 293F cells (Thermo Fisher Scientific; see Key Resources Table) were maintained in FreeStyle 293 Expression Medium (Thermo Fisher Scientific) at 37°C at 8% CO_2 with continuous shaking at 135 rpm. 293F cells at a density of $1x10^6$/mL were transfected with plasmids encoding pCAGGS/6xHis–IFN-β1, and –IFN-ω4 (Blue Heron Biotech, Bothell, WA) using FreeStyle MAX reagent diluted in OptiPRO SFM (Thermo Fisher Scientific) according to the manufacturer's protocol (https://www.thermofisher.com/us/en/home/references/protocols/cell-culture/transfection-protocol/freestyle-max-reagent.html#2). 3 days post-transfection, clarified media (centrifuged at 4°C for 1 hr at 4000rpm) was collected and frozen at −20°C till purification. BL21(DE) cells were transformed with pET22b/6xHis-PA-D1 and streaked on LB agar plates containing 50ug/mL ampicillin. Resulting colonies were grown in 4mL of 2xYt media (ampicillin 50ug/mL) at 37°C till a 600nm OD of 0.6 was reached. 3 mL of starter culture was added to 97 mL of 2xYt media (ampicillin 50ug/mL) and grown till a 600nm OD > 1. Cells were stimulated with IPTG (0.5mM) for 3 hr at 37°C, and then spun down. Cell pellets were frozen at −20°C before protein purification. Proteins were purified via a Capturem His-tagged purification maxiprep kit (Clontech, Takara Bio, Mountain View, CA) and dialyzed into sterile 1x PBS with a Vivaspin 2 protein concentration column (MWCO 10kDa; GE Life Sciences). Proteins were quantified by western blot for the 6x-His tag, and by 280nm absorbance measured on a NanoDrop spectrophotometer. Proteins were stored at −20°C; immediately before use, proteins were diluted in PBS and sterile filtered with a 0.2 μm pore filter.

RoNi cell antiviral assay

RoNi cells in 10% FBS media were seeded at $3x10^4$ in 96-well plates, and treated with serial dilutions of recombinant 6xHis-tagged IFN-β1, recombinant 6xHis-tagged IFN-ω4, or an unrelated recombinant 6xHis-tagged protein (rPA-D1) for 4 hr at 37°C. The interferon-containing media was removed, and cells were infected with vesicular stomatitis virus encoding eGFP (VSV-eGFP; kindly provided by John Connor, Boston University School of Medicine; Whitlow et al., 2006, 2008) at an MOI of 0.05 in 2% FBS RoNi media, and imaged for GFP expression 1 day post infection (multiple viral replication cycles).

Selection analysis

Among many others, we included genes involved in dendritic cell maturation, induction, and signaling of type I IFNs, and cytokine responses, since dysfunctions of these pathways have all been reported as contributors to the pathogenesis of filoviruses in primates (Connor et al., 2015; Liu et al., 2017). We also included genes in DNA damage response pathways, since adaptations in DNA damage responses during the selection for flight in bats have been hypothesized to influence immune responses in bats (for a full list of genes studied, see Table S5). All genes of interest (immune genes, echolocation genes, flight genes) were downloaded from RefSeq in GenBank format and coding sequences for each gene were extracted using a python script. For each member of an ortholog group, the longest isoform was selected. Amino acid sequences of ortholog groups were aligned with Muscle v3.8.31 (Edgar, 2004). The resulting alignments were then trimmed with trimAl v1.4 (Capella-Gutiérrez et al., 2009) to retain only high-quality aligned regions and alignment columns containing gaps. We used codeml within the PAML v4.9b suite of phylogenetic analysis tools to estimate ω (omega), the nonsynonymous/synonymous nucleotide substitution rate ratio, and κ (kappa), the transition/transversion rate ratio with the F3 × 4 codon frequency matrix (Bielawski and Yang, 2005; Yang, 1998). Seven separate models were used (Figure S2) and maximum likelihood scores estimated for each. The models shown in Figure S2 describe the hierarchy of models tested for each gene included in the analysis. Likelihood ratio tests were performed, progressively allowing more degrees of freedom. Model 0 assumes one rate of evolution in all branches. The fit of three models (1a, 1b, and 1c) are separately compared to Model 0 via a likelihood ratio test, and the best fitting model (highest likelihood of the three) is then compared to the next model in the hierarchy (e.g., Model 1a is compared to 2a, Model 1c is compared to both 2a and 2b), and the best fitting model is then compared to Model 3. If the next model in the hierarchy does not fit the data better than the previous model, the previous model is taken to be the most likely.

Two-ratio unconstrained models were first tested against the one-ratio null model to test for differential selective pressure. Any gene for which the null model was rejected proceeded to testing of the most likely two-ratio model against all three-ratio models in which it was nested using the same procedure, and, finally, any gene for which the two-ratio model was rejected against any three-ratio model proceeded to testing the most likely three-ratio model against the four-ratio model. FDR was controlled at each level using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). Models showing $\omega_1 > 1$ along any branch were then compared against two-ratio constrained models with ω_1 fixed at 1 to test for neutral evolution. Branches showing significantly higher than average, but not greater than 1, ω_1 may have positively selected sites. However, relaxation of purifying selection along these branches cannot be ruled out using these models. Analyses were run with multiple starting values of both κ (2,2.5,4) and ω (0.1,1,2) to provide a check against local optimas. These analyses were automated using LMAP v1.0.0 (Maldonado et al., 2016).

QUANTIFICATION AND STATISTICAL ANALYSIS

Significance of expanded and contracted gene families was determined as follows: CAFE v3.1 provides a family-wide p value for gene families that evolved differently than expected, and additionally provides a Viterbi p value that assigns a p value to the contribution of each species to the family-wide evolution across the tree (De Bie et al., 2006). The Benjamini-Hochberg procedure was applied to family-wide p values for the gene families that evolved differently than expected. The p values were ranked from lowest to highest with identical p values assigned the same rank, and the rank for the next non-identical p value incremented by the total size of the identical group. The rank of each p value was divided by the total number of hypotheses (7,698, for each gene family) and multiplied by 0.05 to obtain the adjusted p value with a false-discovery rate of 0.05. For families with adjusted p values < 0.05, the Viterbi p values were used to determine whether the expansion or contraction in a specific lineage was significant. Families with expansions and contractions in Figure 1 have family-wide adjusted p values less than 0.05 and Viterbi p values less than 0.05.

Significance of differential expression between Sendai virus-infected samples and mock-infected samples in Figure S6C was determined via an unpaired two tailed t test with TPM values from three biological replicates, with adjusted p values considered significant if less than 0.05.

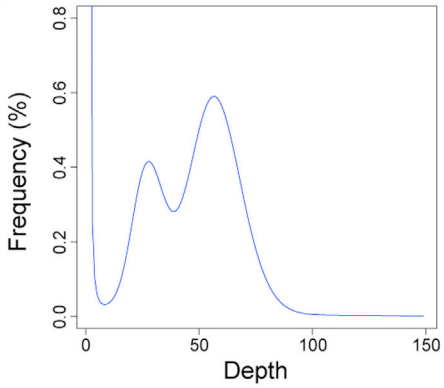
Meta-analysis of viremia and oral shedding of MARV-infected *R. aegyptiacus* bats from two previous studies (Amman et al., 2015; Schuh et al., 2017) was conducted by averaging raw TCID50 equivalent units per mL values across both experiments, with values matched to respective time point post-infection up to 28 days. The standard deviation of the averages was taken for each time point against the number of MARV-positive bat samples, except at time points with fewer than two MARV-positive bats. Linear values were then converted into log form, and trendlines of the average log values from each dataset across time were included using the moving average option with a periodicity of 2.

DATA AND SOFTWARE AVAILABILITY

The accession number for the *Rousettus aegyptiacus* genome Raegyp2.0 reported in this paper is GenBank: GCA_001466805.2 and RefSeq: GCF_001466805.2. The accession number for the raw and analyzed SeV-infected RoNi cell transcriptome data reported in this paper is GEO: GSE108941. The accession number for the whole genome shotgun sequencing project reported in this paper is GenBank: LOCP00000000.2.

The images of bats, cow, mouse, horse, monkey, and dog are used under a creative commons license (CC BY 4.0, Clipart by AnimalsClipart.com). The image of a pig is used under a creative commons license (CC BY 3.0, Anysnapshot.com). The images of hamster and guinea pig are used under a creative commons license (CC BY 4.0, Author: Bob Comix). The image of a human is in the public domain (CC0 1.0). These images can be found at the following websites: http://anysnapshot.com/pig-graphics/, http://animalsclipart.com/bats-flying-silhouette/, http://animalsclipart.com/gallery/mouses/page/2/, http://animalsclipart.com/gallery/cows/page/2/, http://animalsclipart.com/gallery/dogs/page/3/, https://commons.wikimedia.org/wiki/File:Typical_ASL_Signing_Space_representation.svg, http://animalsclipart.com/gallery/horses/page/2/, http://www.supercoloring.com/silhouettes/hamster, http://www.supercoloring.com/silhouettes/guinea-pig, http://animalsclipart.com/gallery/monkeys/page/2/

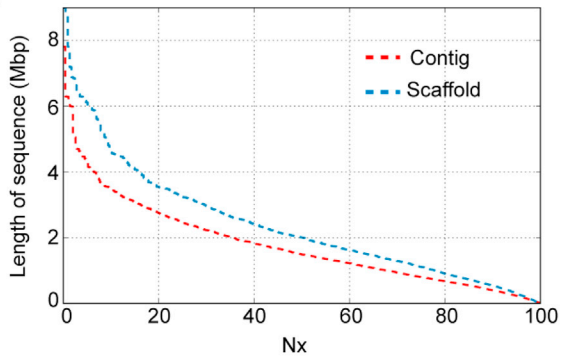
A



B

	SNPs		Indels		
	Number of sites	Percent of genome	Number of sites	Percent of genome	Total percent of genome
R. aegyptiacus	10,065,960	0.527	1,002,056	0.052	0.579
P. alecto[a]	8,790,573	0.453	1,312,432	0.068	0.521
M. davidii[a]	5,140,505	0.279	956,196	0.052	0.331
M. brandtii[b]	7,312,602	0.371	1,271,833	0.065	0.436

C



D

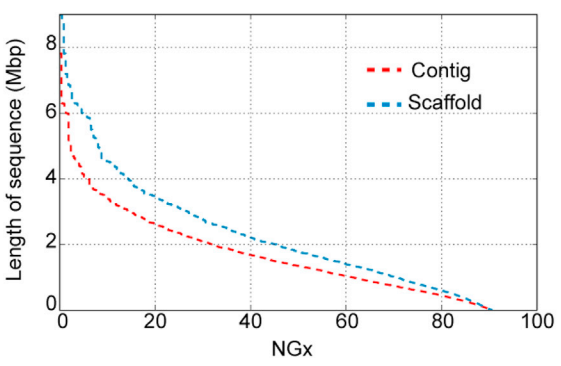


Figure S1. Genome Characteristics of Raegyp2.0, Related to Table 1

(A) K-mer frequency distribution in Raegyp2.0. The percentage (frequency, y axis) of all 25-mers that are present a given number of times (depth, x axis) in the *Rousettus aegyptiacus* genome sequence.

(B) Heterozygosity of Raegyp2.0. SNPs: Single-nucleotide polymorphisms; indels: insertions or deletions. Scaffold and contig.

(C and D) (C) Nx and (D) NGx plots in megabases (Mbp) for Raegyp2.0 with an estimated reference genome size of 2.11 Gb. See STAR Methods for detailed description of Nx and NGx plots.

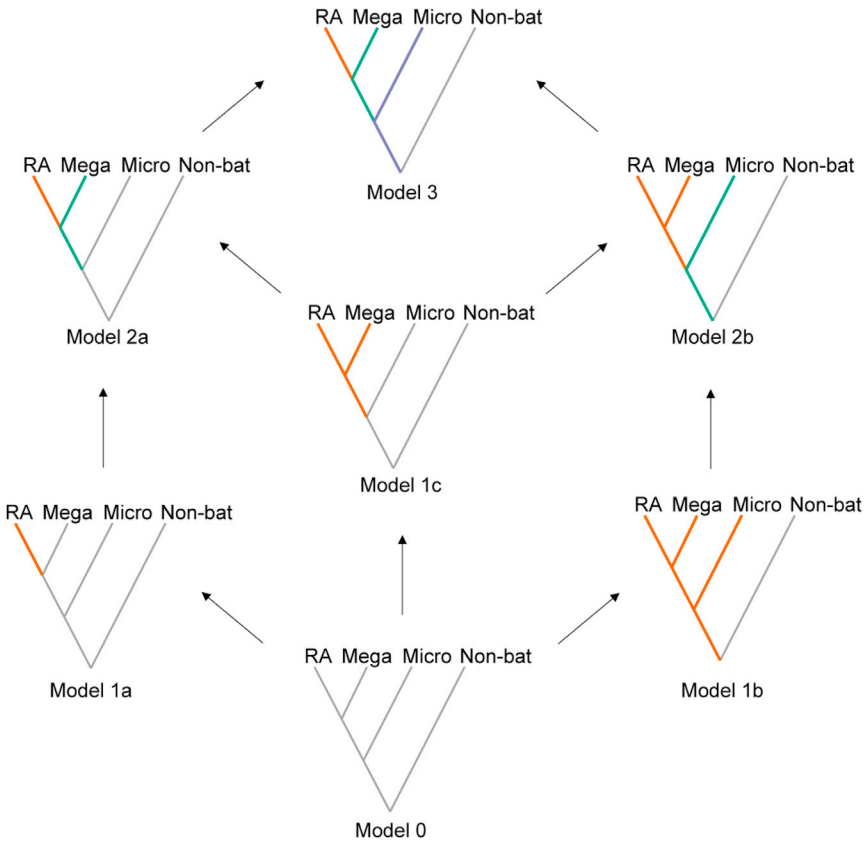


Figure S2. Model Selection Hierarchy for Positive and Purifying Selection Analysis, Related to Table 2 and STAR Methods

RA = *R. aegyptiacus*, Mega = megabats other than *R. aegyptiacus*, Micro = microbats, Non-bat = all species in non-bat branches (human, crab-eating macaque, rhesus macaque, mouse, dog, cow, pig, guinea pig, hamster, and horse). Arrows indicate nested models (e.g., an arrow pointing from Model 1a to 2a means that Model 1a is nested in Model 2a). For each applicable model, color indicates which branches were used to estimate which evolution rate (orange - ω_1; green - ω_2; purple - ω_3; gray - ω_0, i.e., the background rate of evolution). Model 0 was the best-fitting for 330 genes, Model 1a for 5 genes, Model 1b for 65 genes, Model 1c for 23 genes, Model 2a for 0 genes, model 2b for 32 genes, and Model 3 for 1 gene.

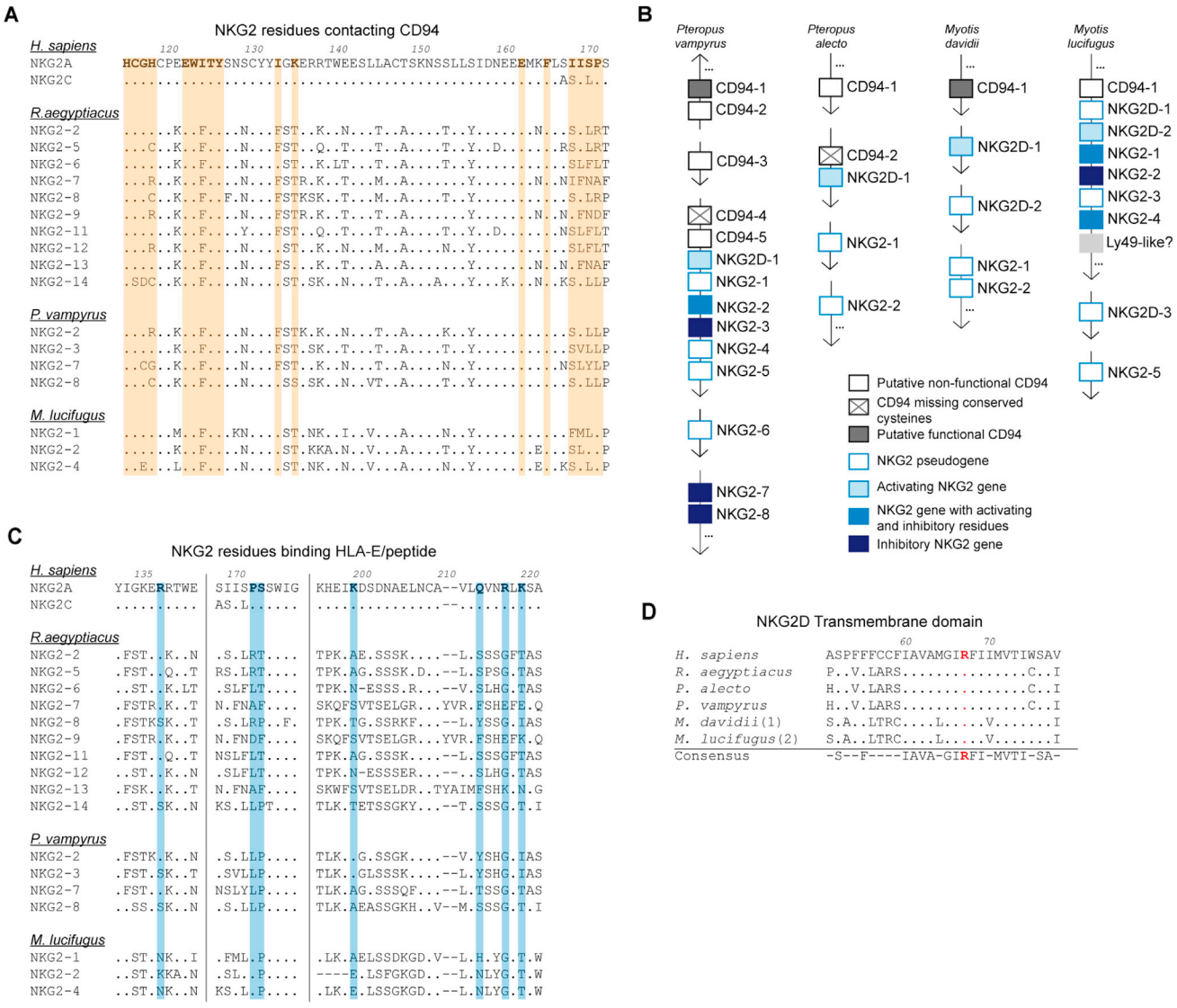


Figure S3. Multiple Sequence Alignments and Locus Maps of NKG2 Proteins in Bats, Related to Figures 2 and 3

Dots in alignments represent identity to the human protein sequence.

(A and B) Alignment of human NKG2A and C and putative functional bat NKG2 protein sequences. (A) shows the NKG2 residues that are known to contact CD94 in the human NKG2A protein (in orange), and (B) shows the NKG2 residues that are known to contact HLA-E/peptide in human NKG2A protein (in blue) (Kaiser et al., 2008).

(C) Locus maps of the NKG2 and CD94 genes in the natural killer complex. Each arrow designates a scaffold sequence in corresponding bat genome (see STAR Methods for accessions). Not pictured are unrelated pseudogenes and non-coding genes. The ellipse indicates the presence of additional non-NKG2 genes on the same scaffold.

(D) Alignment of the transmembrane domain of putative functional NKG2D proteins in humans and bats. Dashes indicate positions of diversity in the consensus sequence. The conserved arginine residue that serves as a signal anchor is shown in red. NKG2D-1 is shown for *M. davidii*, and NKG2D-2 is shown for *M. lucifugus*.

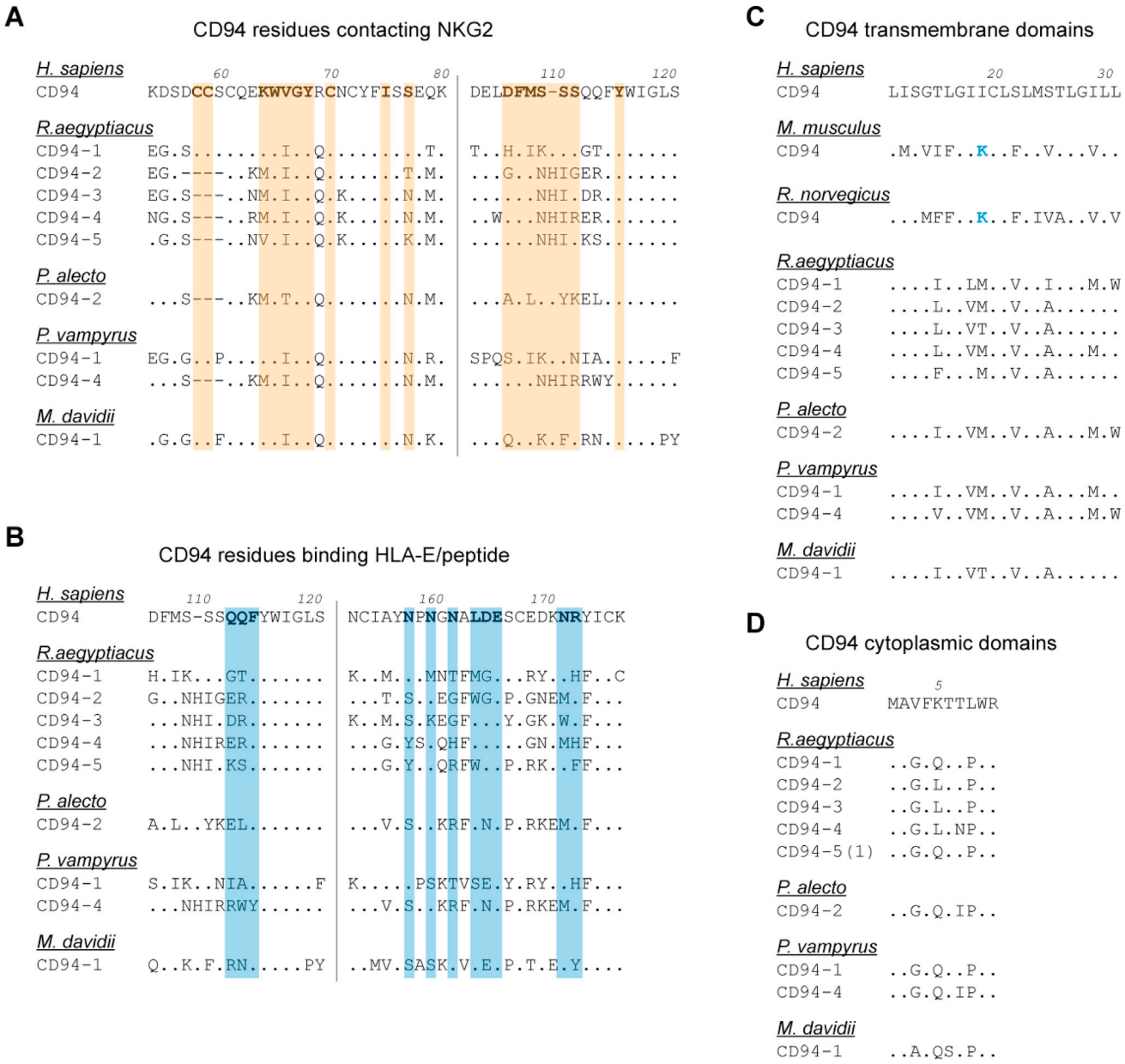


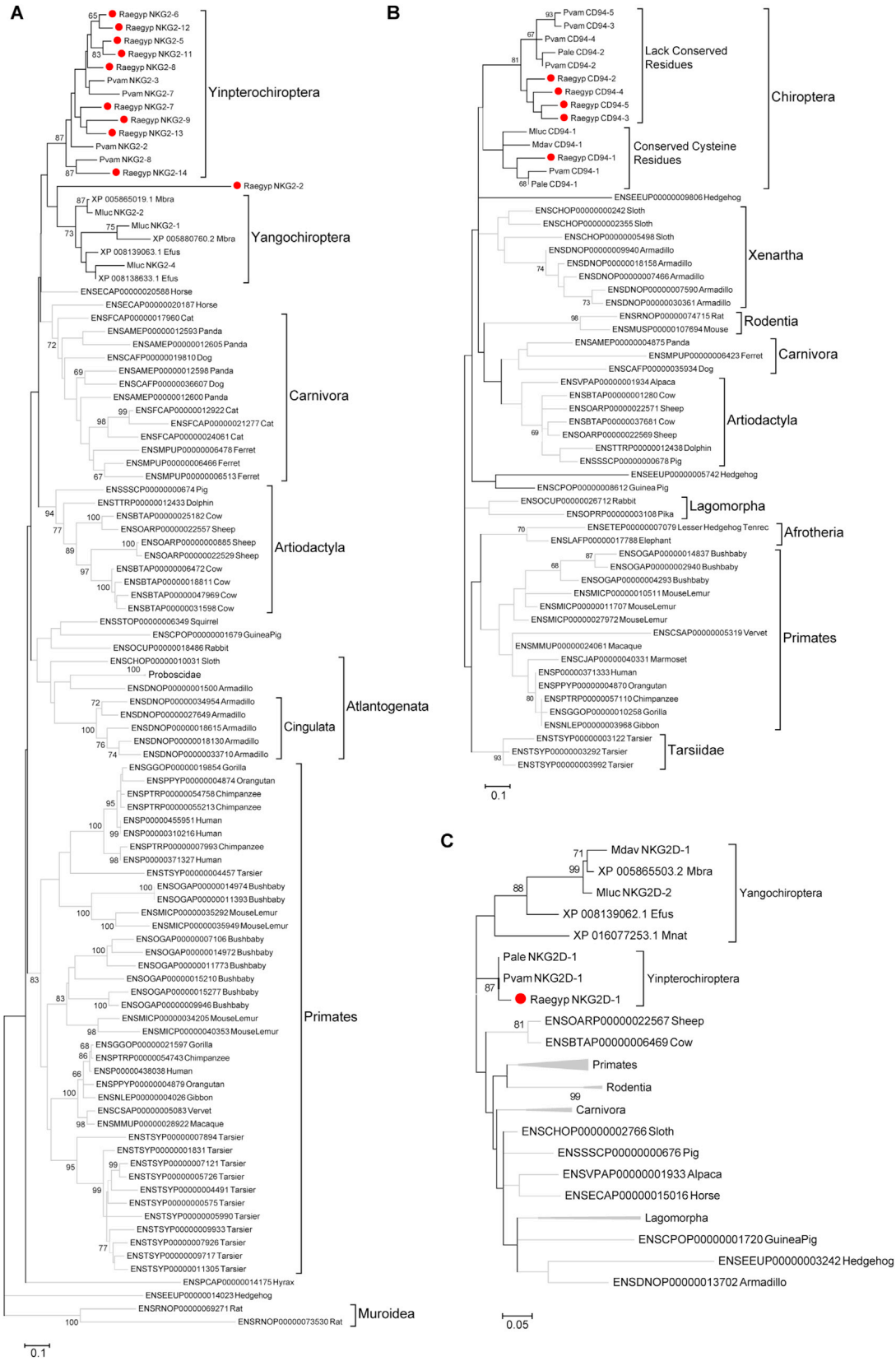
Figure S4. Multiple Sequence Alignments of CD94 Proteins in Bats, Related to Figure 3

Dots in alignments represent identity to the human protein sequence, while dashes represent gaps in the alignment.

(A and B) Alignment of human CD94 and putative functional bat CD94 protein sequences. (A) shows the CD94 residues that are known to contact NKG2A in the human CD94 protein (in orange), and (B) shows the CD94 residues that are known to contact HLA-E/peptide in human CD94 protein (in blue) (Kaiser et al., 2008).

(C) Alignment of the transmembrane domain of CD94 in human, mouse, rat, and bats. The lysine (K) residue that serves as a signal anchor for DAP10 and DAP12 in rodents is shown in blue. This residue is not conserved in bat CD94 proteins.

(D) Alignment of the cytoplasmic domain of CD94 in human and bats.



(legend on next page)

Figure S5. Related to Figures 2 and 3

(A–C) Expanded maximum likelihood phylogenies of (A) NKG2, (B) CD94 proteins, and (C) NKG2D proteins. *R. aegyptiacus* proteins are marked by red dots. Bootstrap evidence (percentage of 500 bootstrap replicates) is labeled on branches if over 65.

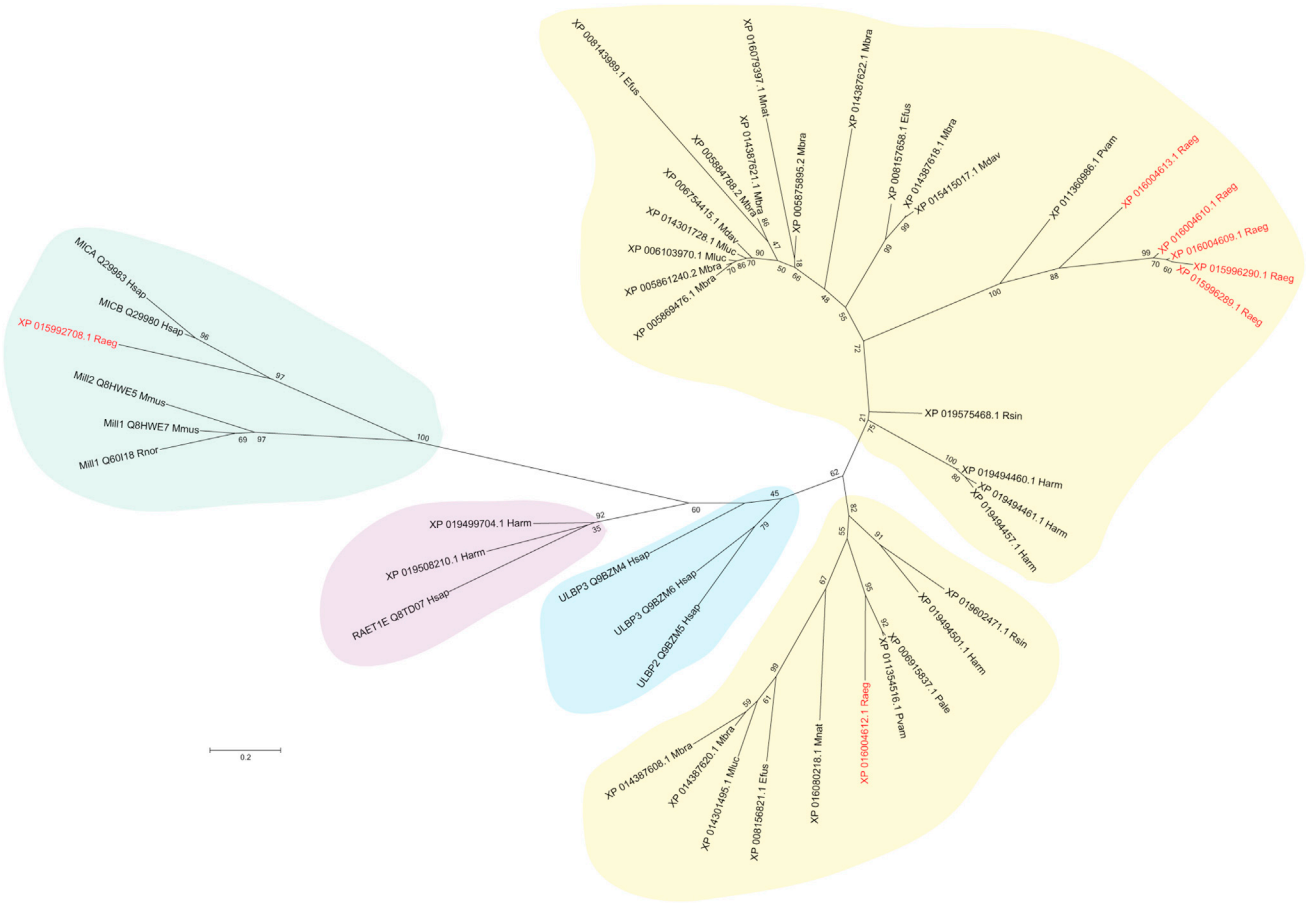


Figure S6. NKG2D Ligands in Bats, Related to Figure 2

Maximum likelihood phylogenetic tree of bat and human NKG2D ligands or NKG2D ligand-like proteins. *R. aegyptiacus* sequences are shown in red. Bootstrap evidence (percentage of 500 bootstrap replicates) is labeled on branches if over 65. Green – MILL and MIC-like proteins, purple – RAET1E-like proteins, blue – human ULBP proteins, yellow – two groups of bat proteins. See STAR Methods for abbreviations.

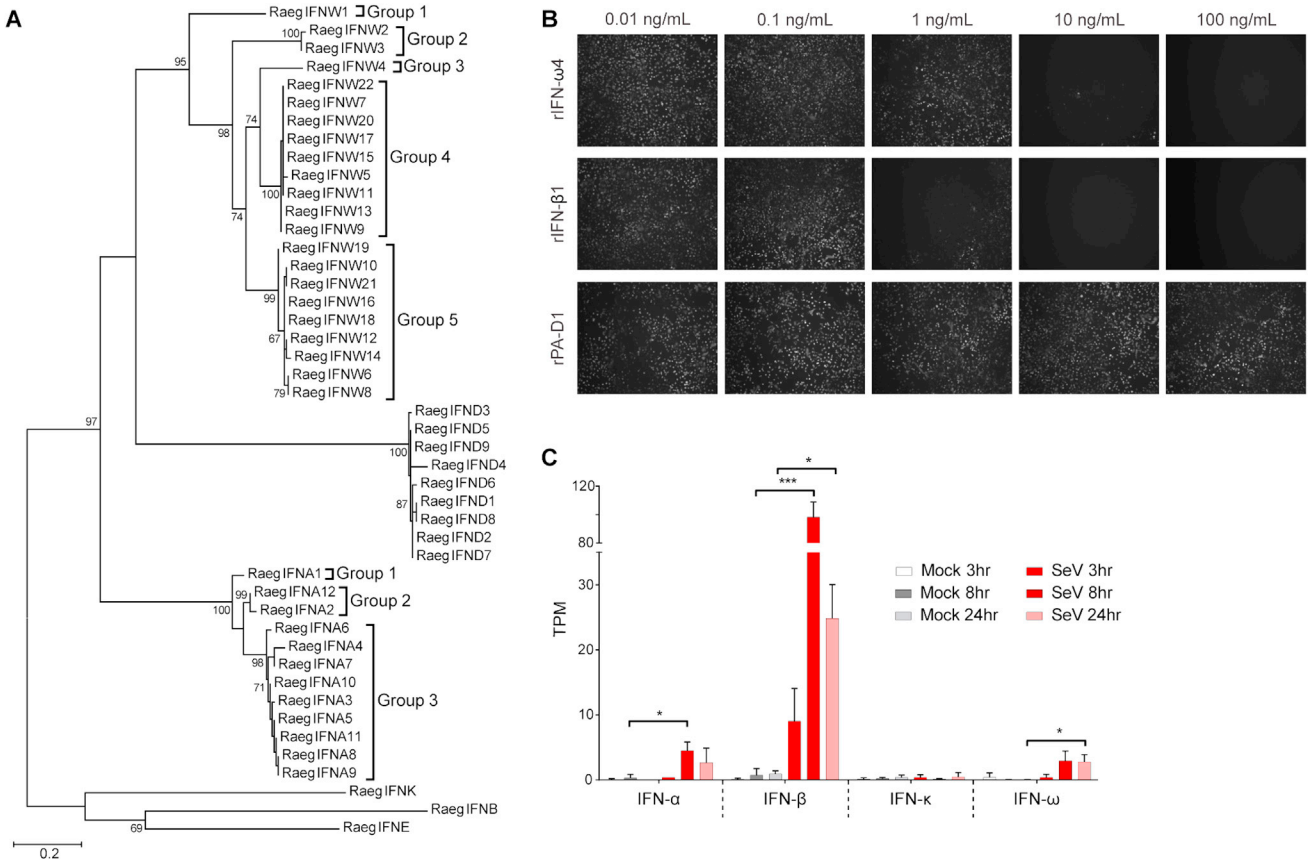


Figure S7. Diversity and Expression of Type I IFN Genes in *R. aegyptiacus*, Related to Figure 5

(A) Phylogeny of *R. aegyptiacus* type I IFN proteins. Maximum likelihood phylogenetic tree of bat type I IFN proteins. Bootstrap evidence (percentage of 500 bootstrap replicates) is labeled on branches if over 65.

(B) Antiviral effect of recombinant *R. aegyptiacus* IFN-ω4. RoNi cells were treated with recombinant IFN-ω4 (rIFN-ω4), rIFN-β1, or an unrelated protein (rPA-D1) for 4 hr, infected with VSV-eGFP at an MOI of 0.05, and imaged for eGFP expression 1 day post infection. Higher concentrations of recombinant IFN-ω4 inhibit viral replication as demonstrated by the absence of eGFP expression in cells after multiple viral replication cycles. Brightness was increased by 20% on all images.

(C) Sendai virus (SeV) infection of RoNi cells elicits an IFN response, including IFN-ω. RoNi cell monolayers were infected with SeV strain Cantell at an MOI of 1.0 or mock infected, and harvested for total RNA extraction and sequencing at 3, 8, and 24 hr. Sequencing data were quantified by IFN subtype in transcripts per million (TPM). Values plotted are the mean ± standard deviation of three replicates for each time point. IFN-ε and IFN-δ were not expressed. Adj. p values from unpaired t test between SeV and mock: * < 0.05, ** < 0.005, *** < 0.0005. See STAR Methods.