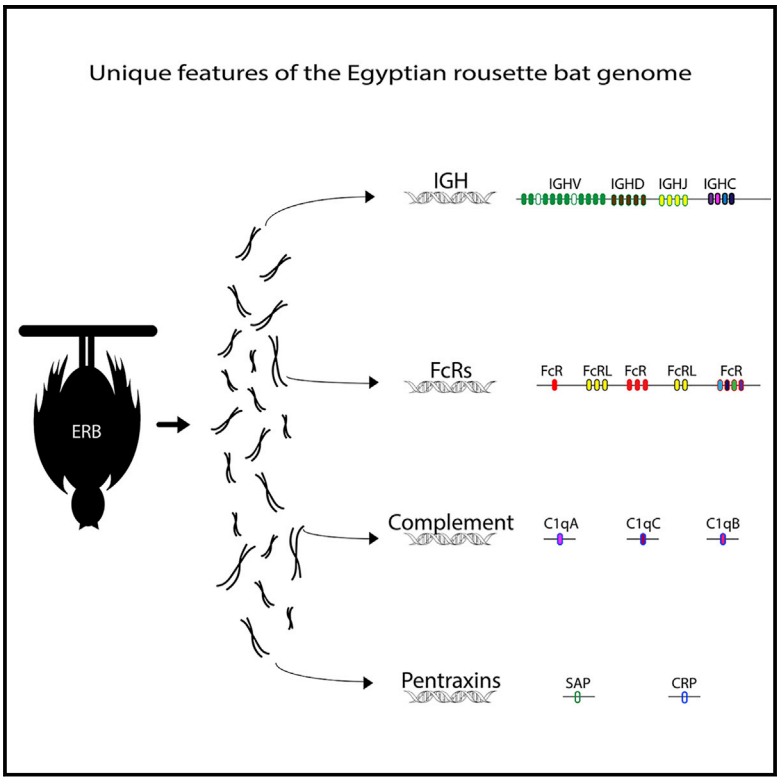
Cell Reports

Genomic features of humoral immunity support tolerance model in Egyptian rousette bats

Graphical abstract



Highlights

- Immunoglobulin heavy-chain locus in *Rousettus aegyptiacus* presents unique features
- Two functionally diverse IGHE and four functional IGHG genes are present
- IGHC/FCγR gene sequences suggest higher activation and inflammation threshold
- Absence of short pentraxins suggests evolutionary pressure to reduce inflammation

Authors

Peter A. Larson, Maggie L. Bartlett, Karla Garcia, ..., Jeffrey Kugelman, Gustavo Palacios, Mariano Sanchez-Lockhart

Correspondence

gustavo.f.palacios.civ@mail.mil (G.P.), mariano.sanchez-lockhart.civ@mail.mil (M.S.-L.)

In brief

The sequencing, assembly, and annotation of the immunoglobulin heavy (IGH) chain locus by Larson et al. reveals features of the Egyptian rousette bat (ERB) humoral immune response. IGH-specific features, in conjunction with those of Fc receptors and pentraxin genes, may explain reduced inflammation and asymptomatic viral infections in ERBs.

Larson et al., 2021, Cell Reports *35*, 109140
May 18, 2021
https://doi.org/10.1016/j.celrep.2021.109140

Cell Reports

CellPress
OPEN ACCESS

Report

Genomic features of humoral immunity support tolerance model in Egyptian rousette bats

Peter A. Larson,[1,5] Maggie L. Bartlett,[1,2,5] Karla Garcia,[1,2] Joseph Chitty,[1] Anne Balkema-Buschmann,[3] Jonathan Towner,[4] Jeffrey Kugelman,[1] Gustavo Palacios,[1,*] and Mariano Sanchez-Lockhart[1,2,6,*]

[1]Center for Genome Sciences, US Army Medical Research Institute of Infectious Diseases, Frederick, MD 21702, USA
[2]Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE 68198, USA
[3]Friedrich-Loeffler-Institut, 17493 Greifswald-Insel Riems, Germany
[4]Viral Special Pathogens Branch, Centers for Disease Control and Prevention, Atlanta, GA 30329, USA
[5]These authors contributed equally
[6]Lead contact
*Correspondence: gustavo.f.palacios.civ@mail.mil (G.P.), mariano.sanchez-lockhart.civ@mail.mil (M.S.-L.)
https://doi.org/10.1016/j.celrep.2021.109140

SUMMARY

Bats asymptomatically harbor many viruses that can cause severe human diseases. The Egyptian rousette bat (ERB) is the only known reservoir for *Marburgviruses* and Sosuga virus, making it an exceptional animal model to study antiviral mechanisms in an asymptomatic host. With this goal in mind, we constructed and annotated the immunoglobulin heavy chain locus, finding an expansion on immunoglobulin variable genes associated with protective human antibodies to different viruses. We also annotated two functional and distinct immunoglobulin epsilon genes and four distinctive functional immunoglobulin gamma genes. We described the Fc receptor repertoire in ERBs, including features that may affect activation potential, and discovered the lack of evolutionary conserved short pentraxins. These findings reinforce the hypothesis that a differential threshold of regulation and/or absence of key immune mediators may promote tolerance and decrease inflammation in ERBs.

INTRODUCTION

Bats (order Chiroptera) have been shown to host more zoonotic viruses per species than any other mammalian order (Wynne and Wang, 2013) and are reservoirs of several human pathogenic zoonotic viruses, yet show no overt sign of disease (Calisher et al., 2006; Olival et al., 2017; Smith and Wang, 2013). Bat genomes have been sequenced to uncover factors that contribute to virus-induced disease resistance (Jebb et al., 2020; Papenfuss et al., 2012; Pavlovich et al., 2018). For some species, a potent innate response was proposed (Zhang et al., 2013; Zhou et al., 2016), while in others, tolerance was suggested (Arnold et al., 2018; Pavlovich et al., 2018). The Egyptian rousette bat (ERB, *Rousettus aegyptiacus*) is a known reservoir for Marburg virus (MARV), Ravn virus (RAVV), and Sosuga virus (SOSV) (Amman et al., 2015a; Amman et al., 2015b; Towner et al., 2009). While ERBs display no overt disease manifestation upon MARV infection, some primate species present with hemorrhagic fever and high lethality (Towner et al., 2006). We previously reported expansions of immune-related gene families in ERBs compared to humans (Pavlovich et al., 2018). Evaluation of the theoretical function of these genes suggested that a tolerogenic immune state may exist in ERBs.

When exposed to MARV, ERBs mount a detectable humoral response; however, *in vitro* experiments were not able to detect neutralizing activity (Schuh et al., 2017b, 2019; Storm et al., 2018), suggesting other mechanisms to mediate viral clearance. Most mammalian immunoglobulins (Igs) arise through the tetramerization of two identical Ig heavy chain (IGH) and Ig light chain (IGL) proteins (Lefranc, 2014). The typical IGH locus comprises numerous variable (IGHV), diversity (IGHD), joining (IGHJ) and constant (IGHC) genes that recombine to form the IGH. The genetic diversity and copy number of IGHV, IGHD, and IGHJ genes contributes to antigen recognition breadth and varies between species (Watson and Breden, 2012), while effector functions are mediated by the fragment crystalline (Fc) region encoded by IGHC genes. In mammals, IGHC genes are represented by five isotypes that vary in number and presence (Akula et al., 2014). Within Chiroptera, there is transcriptomic evidence of IgM, IgA, IgE, and IgG expression (Baker et al., 2010; Butler et al., 2011; Schountz et al., 2017), and IgG subclass number ranges from 1 (*Carollia perspicillata*) to 5 (*Myotis lucifugus*) (Butler et al., 2011). However, given the complexity and repetitiveness of the IGH locus, no complete description of gene organization exists for any bat species.

The Ig Fc region functionally links adaptive immunity with several innate mediators (Akula et al., 2014; Schroeder and Cavacini, 2010). Fc receptors (FcRs) have characteristic expression patterns among leukocytes and can bind specific Ig isotype(s) with varied affinities, resulting in different effector outcomes (Bournazos and Ravetch, 2017). The evolutionary relationship between FcRs and Igs has resulted in differential FcRs presence/absence among vertebrate genomes (Zucchetti et al.,

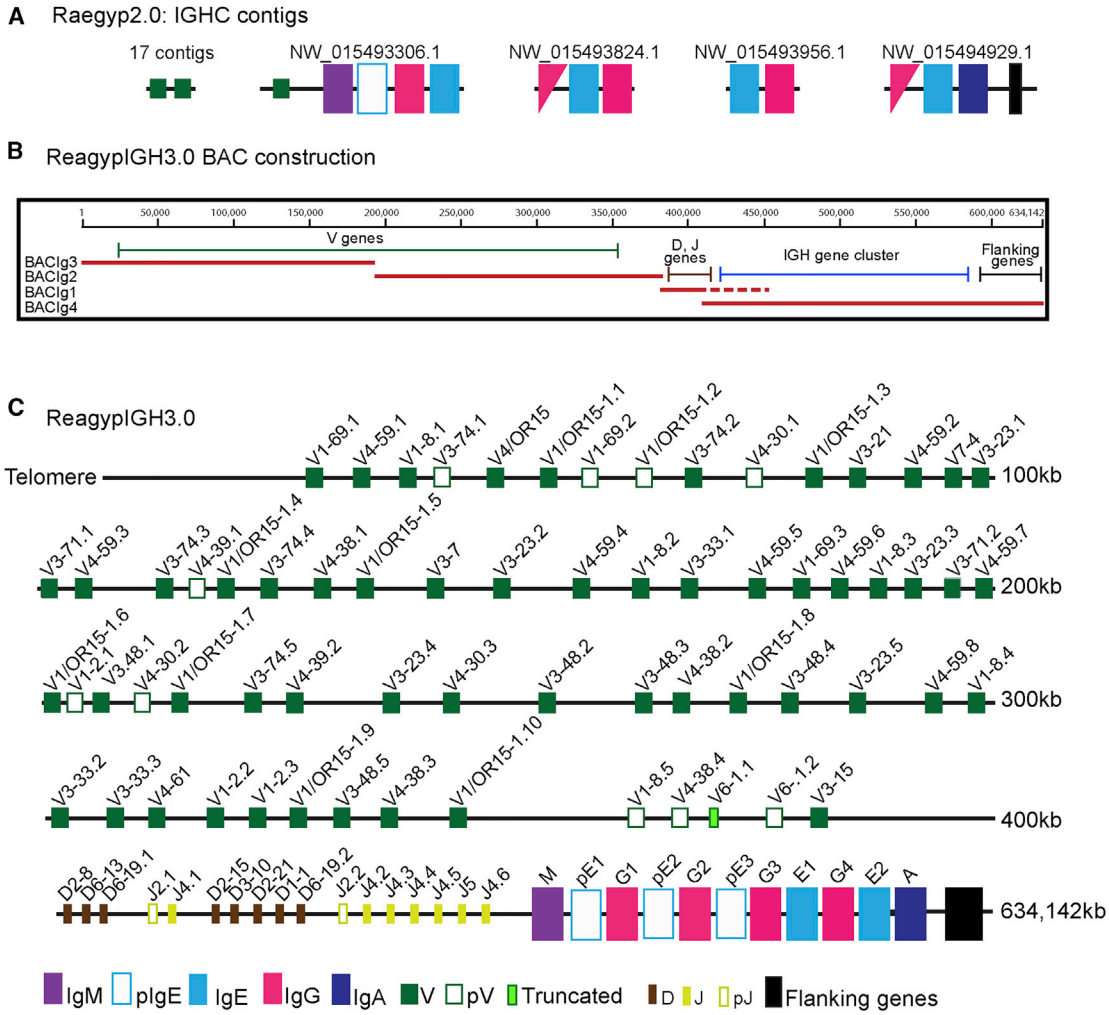


Figure 1. Gene organization of ERB IGH locus

(A) Representation of contigs containing IGH genes from Raegyp2.0. Triangles represent truncated genes, empty boxes represent pseudogenes, and green boxes represent IGHV genes.

(B) Organization of BACs used in the final sequence assembly and relative location of IGHV, IGHD, IGHJ, IGHC, and flanking genes in the final assembly. The dotted line indicates the overlap sequence between $BACIg_1$ and $BACIg_4$. Only sequence from $BACIg_4$ was used in the final assembly. See also STAR Methods.

(C) J (yellow), IGHM (purple), IGHG (pink), IGHE (blue), IGHA (navy), pseudogenes (empty), truncated IGHV (light green), and TMEM121 (black).

2009). The complement system also bridges innate and adaptive immunity (Merle et al., 2015a, 2015b). C1q interaction with Fc can trigger opsonization, inflammation, and cell lysis. Pentraxins are conserved from arthropods to mammals and can recognize microbial moieties to mediate opsonophagocytosis, inflammation, and activate the classical complement pathway (Bottazzi et al., 2010; Lu et al., 2018).

To begin to understand ERB innate and adaptive immunity, we constructed one contiguous IGH locus and described the Ig genes. We found an expansion of the IgE genes resulting in two putatively functional genes and three pseudogenes. We also described four functional IgG, one IgA, and one IgM genes. We described and evaluated the FcR gene repertoire and proteins of the complement system. We also observed the complete absence of functional short pentraxins. Our observations emphasize the uniqueness of Chiropteran genomes and highlight the importance of the continued characterization of reservoir species (Jebb et al., 2020; Sotero-Caio et al., 2017).

RESULTS

Sequencing and assembly of ERB IGH locus

The ERB genome (Raegyp2.0) (Pavlovich et al., 2018) contains annotated IGHC genes on four contigs (Figure 1A). Contig NW_015494929.1 contains a single copy of an IgA gene (*IGHA*) upstream of *TMEM121*, which also flanks *IGHA* in the human IGH locus (Gertz et al., 2013). The 5′ portion of the ERB IGH locus is on contig NW_015493306.1 containing a single IGHV gene and an IgM gene (*IGHM*). These two contigs likely represent the boundaries of the IGHC locus, orthologous to human chromosome 14q (McBride et al., 1982). Several complete or partial IgE (*IGHE*) and IgG (*IGHG*) genes were distributed among four

contigs, with one labeled as a pseudo IgE gene. Annotated IGHV genes were present on 17 contigs (see Method details). No IgD, IGHD, or IGHJ genes were annotated.

To overcome the discontinuous assemblies of repetitive gene structures in mammalian genomes (Alhakami et al., 2017; Lander et al., 2001), we used bacterial artificial chromosomes (BACs) to resolve the ERB IGH locus. We identified, sequenced, and assembled 4 BACs that mapped to portions of the 4 contigs above, as well as the 17 contigs containing IGHV genes (Figure 1B). Assembling the 4 BACs generated the final IGH locus (RaegypIGH3.0) (Figure 1C). RaegypIGH3.0 links contigs NW_015493306.1 and NW_015494929.1 and adds ~54 kb of intervening sequence. Bionano Optical Mapping data demonstrated that contigs NW_015493306.1 and NW_015494929.1 are separated by ~56 kb of sequence (data not shown). The other 2 contigs (NW_015493824.1 and NW_015493956.1) likely represent misassembled portions of the 54-kb intervening sequence. At one end, RaegypIGH3.0 (634,142 nucleotides) extends into what is likely a sub-telomeric repeat, and at the other, extends through genes flanking the IGH locus in other mammals, including *TMEM121*, *CRIP1*, *TEDC1*, and a portion of *CRIP2*.

Annotation and description of IGHV, IGHD, and IGHJ genes

In Raegyp2.0, 58 IGHV genes on 17 contigs were annotated by NCBI. Using IMGT/LIGMotif (Lefranc et al., 2015), we annotated 66 IGHV genes on RaegypIGH3.0 (Figure 1C). Of those, 11 are predicted non-functional (10 pseudogenes and 1 truncated) (Data S1). Similar to pteropid bats, ERB IGHV functional genes represent all 3 clans and include representatives of families IGHV1 (17), IGHV3 (22), IGHV4 (15), and IGHV7 (1) (Figure 2A; Data S1; Baker et al., 2010; Elemento and Lefranc, 2003). ERB IGHVs are similar in length to other mammals (Data S1; Lefranc et al., 2005) and contain downstream recombination signal sequences (RSSs), with the exception of pseudogene VH1-2.1 (Figure 2B). Intriguingly, we observed that several of the ERB IGHV genes present, and in some cases, expanded, at the IGH locus were associated with protective responses against viruses in humans (Table S1).

Using IMGT/LIGMotif, we annotated 8 IGHD and 9 IGHJ genes in RaegypIGH3.0. ERB IGHD genes comprise the IGHD1, IGHD2, IGHD3, and IGHD6 families, are similar in length to the human genes (Data S1; Lefranc et al., 2005), and contain flanking RSSs (Figure 2C). The IGHJ genes identified make up the IGHJ2, IGHJ4, and IGHJ5 families, are slightly smaller than human genes (Data S1; Lefranc et al., 2005), and contain functional RSSs, with the exception of the 2 IGHJ2 genes, which lack the conserved 5′-GAGCGTG-3′ observed in human IGHJ gene heptamers (Figure 2D). All IGHJ4 and IGHJ5 genes retain the highly conserved WGXG amino acid (aa) motif, while IGHJ2 genes do not (Figure 2E), suggesting they are not functional (Tsakou et al., 2012). One intriguing observation was the IGHD-IGHJ-IGHD-IGHJ organization within the IGH locus.

We determined the IGHV, IGHD, and IGHJ gene expression pattern by mining previously published transcriptomic data (Lee et al., 2015). Recovered reads were annotated by IMGT (Lefranc et al., 2015) and remapped to RaegpyIGH3.0 to confirm location and expression of annotated genes. The expression of the majority of the functional IGHV genes was confirmed. Limited transcriptome support was detected for the pseudogenes and the truncated *IGHV6-1.1*. Transcripts mapped to all IGHJ genes (data not shown), however, >30-fold fewer to the two *IGHJ2* and *IGHJ4.1*, suggesting that these genes are likely not used. *IGHJ2.1* and *IGHJ4.1* are the genes in the first IGHJ cluster in the locus. Due to short IGHD length, we annotated IGHD using IMGT-HighVQuest in reads in which an IGHJ gene was found. This method was able to further confirm the functionality and expression of all eight IGHD genes.

Annotation and description of IGHC genes

To identify IGHC genes, we BLAST NCBI annotated Raegyp2.0 IGHC sequences against RaegypIGH3.0. Like other pteropid bats, we were unable to identify an *IGHD* (Baker et al., 2010). We identified five *IGHE*s in RaegypIGH3.0. The first three *IGHE*s contain indels resulting in premature stop codons and are likely unprocessed pseudogenes (Data S1). Two *IGHE*s contain complete open reading frames (ORFs) and are putatively functional. This finding makes ERBs the only mammalian species identified that contains more than one putatively functional *IGHE* (Sun et al., 2012). We identified four *IGHG*s containing complete ORFs and these are likely functional (Data S1). *IGHE*s and *IGHG*s are numbered based on their 5′–3′ position, with the telomeric end representing the 5′ end (Figure 1C). We determined sequence identity among *IGHE*s and *IGHG*s using ClustalW (Data S1). Consistent with previous observations, the majority of the *IGHG* sequence heterogeneity is clustered at the hinge and CH2 domains (Vidarsson et al., 2014).

Transcriptomic data (Lee et al., 2015) was used to verify IGHC gene expression. Using BLAST, we queried the coding sequences of each *IGHC* and mapped reads to unique Ig isotype or subclass sequence-specific motifs, further validating our assembly and annotation (Data S2). Finally, we performed mass spectrometry on protein A/G purified total sera collected from wild-caught ERBs to verify protein expression (data not shown; Table S2). We detected unique peptides to IgG1, IgG2, IgG4, IgM, and IgA, and a shared IgE peptide, demonstrating that mRNAs derived from the identified genes are translated.

IGHC evolutionary relationship

Our data bolster previous reports of IGHC genes clustering within Chiropteran suborders (Figure S1A; Agnarsson et al., 2011). Our molecular data demonstrate that multiple characteristics of the IHC loci unique to ERB occurred after speciation and recapitulate other taxonomic data demonstrating ERB clusters within *Yinpterochiroptera*.

Functional characteristics of the ERB IGHC genes

IGHC motifs associated with Ig structure/function have been widely described for mammals (Table S2). The hinge domain of IGHG is variable at both inter- and intra-species levels and crucial for forming disulfide bonds between IGHs (Dard et al., 1997). ERB *IGHG1* contains a single 5′-CPRCP-3′ motif, which is repeated 4 times in human *IGHG3* (Figure S1E). Neither ERB *IGHG2*, *IGHG3*, nor *IGHG4* contain the canonical CXXC domain at the hinge. ERB *IGHG2* and *IGHG3* lack both cysteines in that domain, with the second Cys^{109} replaced by a serine (Figure S1E; Table S2). Nevertheless, we cannot rule out dimerization, since

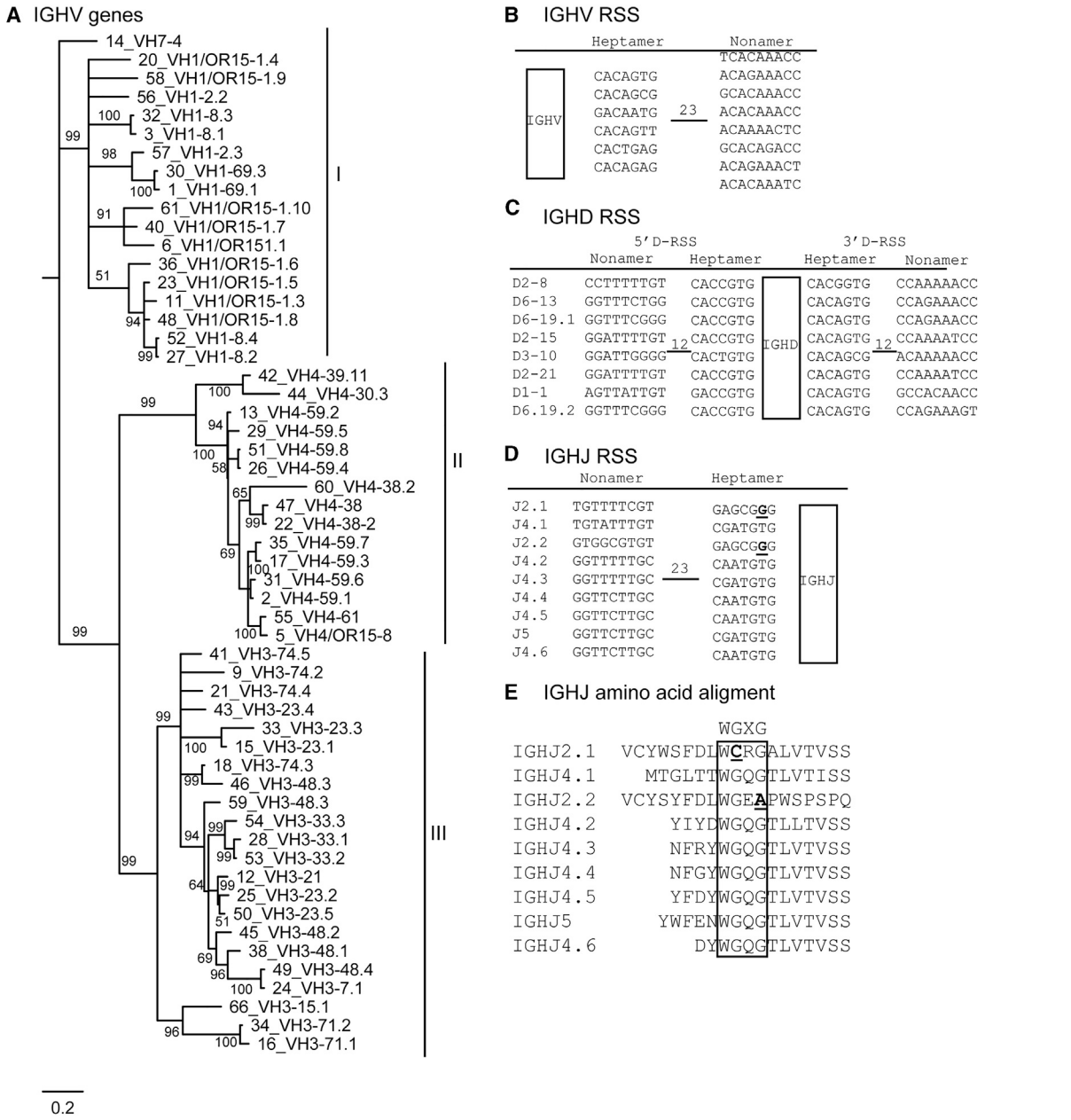


Figure 2. Annotation and description of IGHV, IGHD, and IGHJ genes

(A) Phylogeny of functional non-truncated IGHV genes. Amino acids were aligned with ClustalW Bayesian analysis to generate the tree; bootstrap values shown. Genes cluster within defined clans I, II, and III.

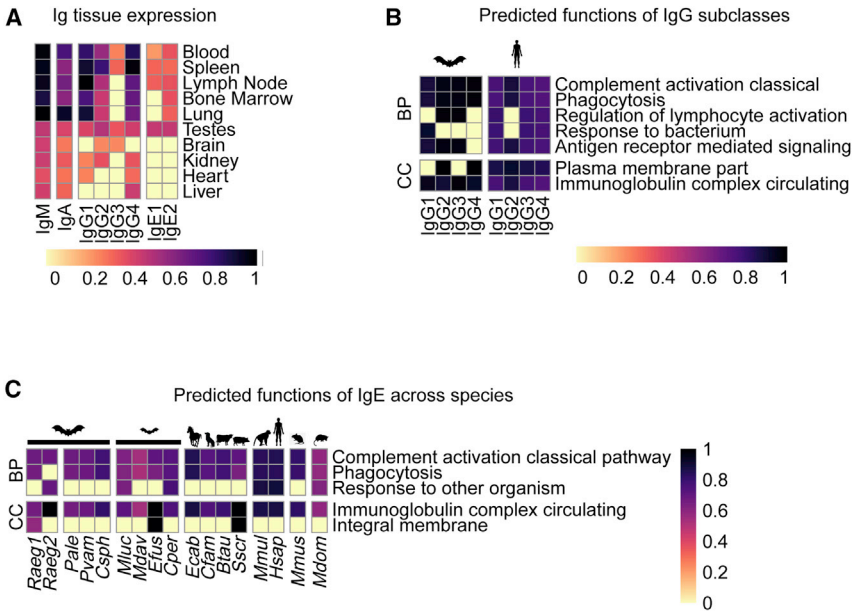
(B–D) IGHV, IGHD, and IGHJ recombination signal sequences for all genes, respectively.

(E) Amino acid alignment of IGHJ genes with conserved WGXG motif (boxed) and residues that could impair functional products (underlined) indicated.

all IGHG genes contain at least one cysteine in the hinge that could allow disulfide bonding.

N-glycosylation modulates Ig effector functions. We compared the N-glycosylation pattern between ERB and human IGH and summarized them in Figure S1. In brief, ERB *IGHE1* and *IGHE2* have the highly conserved N-glycosylation site (Asn^{287}) in the CH3 domain associated with protein structure and folding (Figure S1C; Arnold et al., 2004; Vernersson et al., 2004). All human and ERB *IGHG*s contain the conserved, single N-glycosylation site at Asn^{297} (Figure S1E; Plomp et al., 2017). In addition, the ERB *IGHG1* contains the longest hinge domain and a putative N-glycosylation site not present in the other ERB or human *IGHG*s.

The region of the IGHC responsible for making contact with the FcγRI has been well characterized (Kiyoshi et al., 2015; Lu et al., 2015). In addition to the N-glycosylation site at Asn^{297}, 3 critical contacts have been identified: (1) the lower hinge region



ERB (Raeg1, Raeg2), Pale (Pteropus alecto), Pvam (Pteropus vampyrus), Csph (Cynopterus sphinx), Mluc (Myotis lucifigus), Mdav (Myotis davidii), Efus (Eptesicus fuscus), Cper (Carollia perspicilata), Ecab (Equus caballus), Cfam (Canine familiaris), Btau (Bos taurus), Sscr (Sus scrofa), Mmul (Macaca mulatta), Hsap (Homo sapiens), Mmus (Mus musculus), Mdom (Monodelphis

Figure 3. Tissue expression and predicted function of ERB Igs

(A) Expression of immunoglobulins in transcriptomic data from 10 tissues in ERBs. Rows are ordered by highest average expression. Expression is reported as log2(TPM) (transcripts per million) and was normalized to glyceraldehyde 3-phosphate dehydrogenase (GAPDH).

(B) Correlated gene ontologies between human and ERBs IGHG subclasses based on confidence score. Biological process (BP): GO: 0006958, GO: 0006910/GO: 0006911, GO: 0051707. Cellular component (CC): GO: 0044425, GO: 0044459, GO: 0042571.

(C) Correlated gene ontologies for IGHE genes between species. BP: GO: 0006958, GO: 0006910, GO: 0051249, GO: 0009617, GO: 005085. CC: GO: 0016021, GO: 0042571.

(Leu^{233}Leu^{234}Gly^{235}Gly^{236}); (2) the CHγ2 BC loop Asp^{265}; and (3) the CHγ2 FG loop (Ala^{327}Leu^{328}Pro^{329}Ala^{330}Pro^{331}). All ERB IgGs contain the conserved Asp^{265} and the CHγ2 FG loop. However, Leu^{233}Leu ^{234} is only present in ERB *IGHG1*, while the other subclasses present a Leu^{233}Pro^{234}. The leucine to alanine mutation (LALA mutation) is associated with the inability of the human IgG to bind to FcγRs (Arduin et al., 2015; Wines et al., 2000).

Ig cytosolic tails (CTs) are evolutionarily conserved among mammals and interact with the α and β B cell receptor heterodimer to modulate and enhance signaling (Table S3; Chen et al., 2015). Two motifs are associated with Ig CT signaling: (1) a tyrosine at a DYXNM motif and (2) a SSVV domain (Chen et al., 2015). When comparing both functional *IGHE*s, we observed an unusual internal deletion of the IgE1 transmembrane (TM) and CT region (Figure S1C). Human IgE and ERB IgE2 present the DYXNM motif 19 aas downstream of the TM (DYANV for ERB IgE2); however, the deletion on ERB IgE1 positions the motif 5 aas closer to the plasma membrane and may affect the function.

Tissue expression of Ig constant genes

We used transcriptomic data again to determine the level of expression of each IGHC per tissue (Figure 3A). *IGHM* is expressed in all tissues and most highly expressed in bone marrow, lymph node (LN), blood, spleen, and lung. Consistent with other mammals, the highest expression of *IGHA* is observed in lungs (Cerutti et al., 2011). Each *IGHG* subclass showed differential expression (Vidarsson et al., 2014). We observed that *IGHG*1 had the highest expression in most tissues, followed by *IGHG*4, *IGHG*2, and lastly *IGHG*3. Expression of *IGHE*2 was present in LN, blood, spleen, and testes, whereas *IGHE*1 was expressed in those tissues as well as bone marrow and lung. The differential expression of *IGHE* genes is indicative of a differentiation of function.

Gene Ontology (GO) suggests IgE and IgG subclasses serve unique functions

COFACTOR (Roy et al., 2012) was used to determine putative functional differences of *IGHGs* and *IGHEs*. The four ERB *IGHGs* likely perform different effector functions from their human counterparts, with no clear statistically significant functional equivalency (Figure 3B). Both ERB *IGHE1* and *IGHE2* are predicted to be important for immune response to other organisms, a key feature of *IGHE* in other mammals (Hellman et al., 2017). However, ERB *IGHE*1 and *IGHE2* also exhibit differences that support a possible diversification of function (Figure 3C), although these predictions require further functional validation. Finally, we examined ERB *IGHA* and *IGHM* and found that their predicted functions are broadly similar to their human homologs (Data S3).

Identification and expression of ERB FcRs

The number and composition of FcRs vary between species (Akula et al., 2014; Koenderman, 2019). We identified 13 FcRs in Raegyp2.0, the orthologs for human *FCERI*, *FCERII*, *FCGRI*, *FCGRIIB*, *FCGRIIIA*, *FCMR*, *FCAMR*, *FCGRT*, *FCRL4*, *FCRL5*, *PIGR*, and *TRIM21*, and a receptor with no-ortholog annotated as *FCGRII*-like. Fc signaling-associated chains were also found in Raegyp2.0 (FCεR1β and FCεR1γ). No orthologs were found for human *FCαR*, *FCGRIIA*, *FCGRIIC*, and *FCGRIIIB*. Among the four ERB FCGRs, we identified the canonical inhibitor *FCGRIIB* containing the immunoreceptor tyrosine-based inhibition motifs (ITIMs) at the CT, but none contain immunoreceptor tyrosine-based activation motifs (ITAMs) (Figure 4A). Furthermore, in the activating ERB *FCGRI* and *FCGRII*-like, Asn^{306} reverted to Asp^{306} (Figure 4B). This substitution still allows for association with the signaling γ chain, but it is reliant on co-expression for cell surface expression (Ernst et al., 1993; Scholl and Geha, 1993). Moreover, the ERB *FCGRI* contains a longer FG loop associated with lower affinity for IgG (Figure 4C; Lu

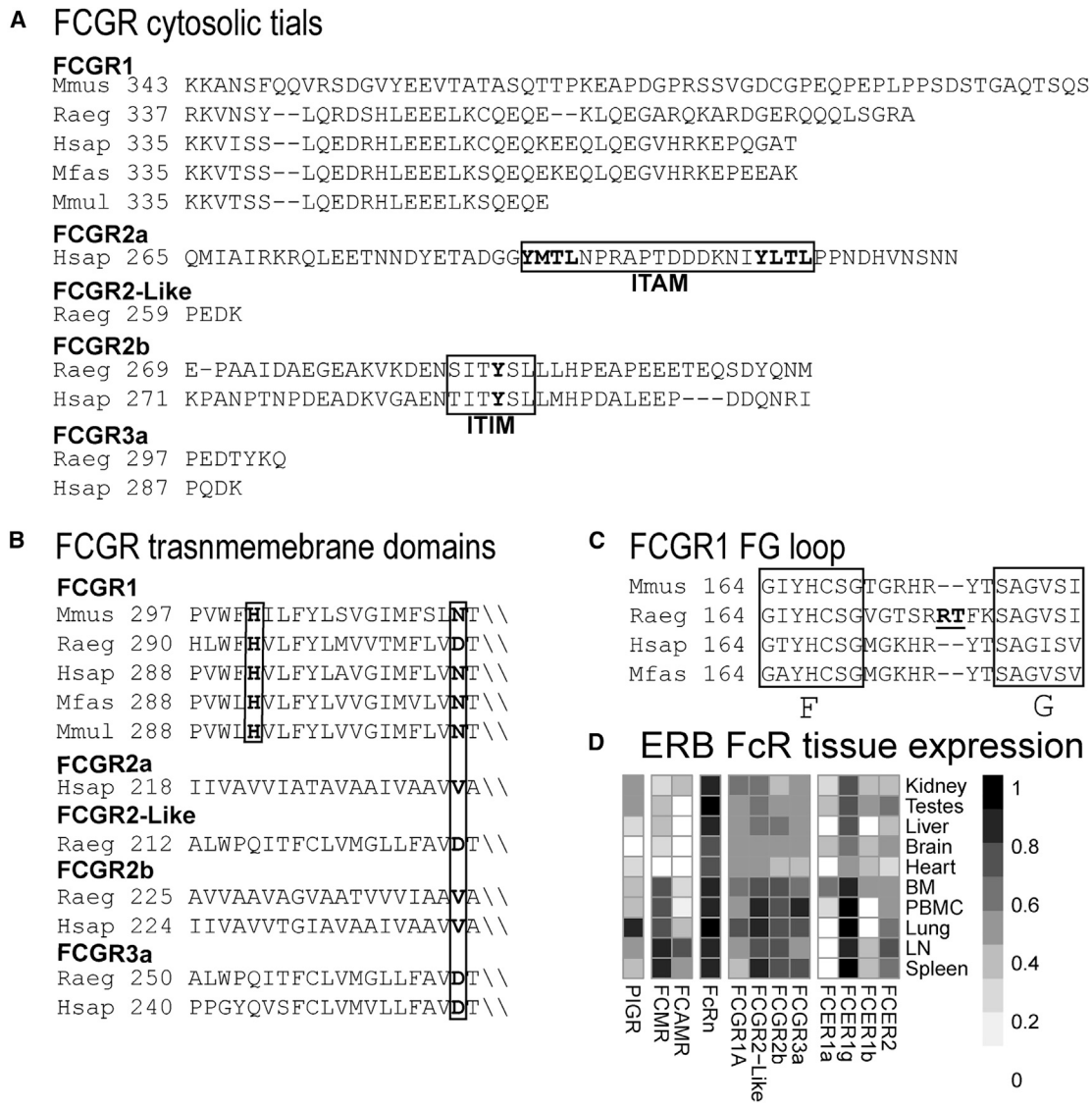


Figure 4. FcR comparison between ERB and other mammals

(A) Alignments of FcR cytosolic domains using ClustalW. ITIM and ITAM motifs are boxed.

(B) FcR transmembrane residues for γ-chain association and surface expression. Residues important for interaction boxed and bolded.

(C) Alignment of F and G loop of FcRs showing additional residues in the ERB loop.

(D) FcR expression in transcriptomic data from 10 ERB tissues. Rows ordered by highest average expression, which is reported as log2(TPM) and is normalized to GAPDH.

et al., 2011). The specific expression profiles of ERB FcRs show that all are expressed in at least one tissue. Similar to humans, *FCGRT* was the most ubiquitously expressed (Figure 4D; Pyzik et al., 2019).

Complement proteins and expression levels

Complement interacts with Ig Fc and participates in several of the Ig effector functions (Merle et al., 2015a, 2015b). The three subunits of the C1q complex, C1qA, C1qB, and C1qC, in Raegyp2.0 are putatively functional. C1qA, C1qB, C1qC, and C3 are ubiquitously expressed in ERB tissues (Figure S2). While C3 is mainly expressed at the liver and lung, the C1q proteins are mainly expressed in the spleen, LNs, liver, and lung.

Identification and expression of ERB pentraxins

No evidence of functional short pentraxins (C-reactive protein [CRP] and serum amyloid P component [SAP]) was found in Raegyp2.0. A *SAP* pseudogene was identified but contained

numerous premature stop codons. At published bat genomes (*Pteropus alecto*, *Desmodus rotundus*, *Phyllostomus discolor*, *Miniopterus natalensis*), annotated *SAP* sequences also contained numerous stop codons, suggesting the gene inactivation well before the emergence of the genus (data not shown). Long pentraxins were annotated in Raegyp2.0: pentraxin 3 (*PTX3*) and *PTX4*, neuronal pentraxin 1 (*NP1*) and *NP2*, as well as the neuronal pentraxin receptor (*NPTXR*). We also identified *DC-SIGN* and the mucosal pentraxin (*MPTX*), which is annotated as a pseudogene in ERBs and humans. *NPTXR1*, *NPTX2*, and *NPTXR* expression was highest in brain, *PTX3* in lung, *PTX4* in testes, and *DC-SIGN* in LN (Figure S2).

DISCUSSION

Although some bat antiviral mechanism(s) have been described (Baker et al., 2013), a comprehensive understanding of bat immunity remains elusive. It has been proposed that some bat species evolved a potent innate antiviral response, allowing for early control of virus replication (Zhang et al., 2013; Zhou et al., 2016). For other species, like ERBs, it was proposed that disease tolerance rather than enhanced viral defense plays a primary role (Ahn et al., 2019; Arnold et al., 2018; Jebb et al., 2020; Mandl et al., 2018; Pavlovich et al., 2018). This model of viral tolerance is supported by experimental studies in ERBs demonstrating limited inflammation, protracted incubation, and sustained viremia/shedding for up to 3 weeks post-MARV infection (Amman et al., 2015a; Schuh et al., 2017a). ERBs mount humoral responses to MARV infection and become refractory to future infection (Schuh et al., 2017b); however, no neutralizing activity was detected when tested *in vitro* (Schuh et al., 2019). Here, we investigated components of the humoral immune response of ERBs and evaluated whether the features found would support the enhanced tolerance model.

Distinct features found on the IGH locus

The IGHV germline of ERBs contains numerous gene expansions relative to human orthologs (Figure 1). We observed that several of the ERB IGHV genes present, and, in some cases, expanded, were associated with V(D)J rearrangement signatures for specific pathogens (Table S1; Cohen-Dvashi et al., 2020; Watson et al., 2017), suggesting that the ERB may be equipped with an IGHV repertoire prone to generating antibodies that can bind to zoonotic viruses. The expansion of the IGHJ4 genes in ERBs may reflect its important role in V(D)J recombination; however, its impact in B cell diversity and antibody specificity is still unknown (Arnaout et al., 2011). Despite humans containing only a single IGHJ4, it is the most commonly used J segment (Shi et al., 2020).

Since the duplication of the ancient IgY gene resulting in IgG and IgE, these two Ig isotypes have diversified their functions (Hellman et al., 2017). While IgGs function mainly via complement activation, immune complex clearance, and antibody-dependent cellular functions, IgEs primarily function via interaction with the FcεR on mast cells and basophils. It is possible that the *IGHE-IGHG* cassette contains sequences suited for unequal crossing over (Reams and Roth, 2015) generating the *IGHE-IGHG* tandem gene duplications observed in the locus. However, selective pressure may inactivate the newly duplicated *IGHE.* The tightly regulated *IGHE* copy number may reflect a balance of a cost-benefit between IgE function and expression level (Oettgen, 2016). The three IgE pseudogenes observed in the ERB IGH locus may be a direct reflection of tightly regulated *IGHE* copy number. To our knowledge, the ERB is the first mammalian species demonstrated to contain more than one functional *IGHE*. Although humans contain two copies of *IGHE*, one is a non-functional pseudogene (Max et al., 1982). Diversification of function between IgE1 and IgE2 in ERBs likely explains the retention of two functional IgE copies. We found that *IGHE2* is expressed in lung and bone marrow, while both *IGHE1* and *IGHE2* transcripts are detected in the periphery (blood) and secondary lymphoid organs (LN and spleen) (Figure 3A). Therefore, IgE2 may perform an important role in the lung microenvironment, acting as a sensing antibody to modulate the IgG response, similar to that seen with many other virus infections (Kelly and Grayson, 2016). Using GO term analysis, we observed that *IGHE1* appears to mimic the majority of described *IGHE*s, while a distinct predicted functionality was observed for *IGHE2* (Figure 3C). Moreover, a residue known to be important for CD23 (FcεRII) binding, Glu^{414} (Sutton and Davies, 2015), is present in *IGHE2* but not *IGHE1*, which may suggest less effective or ablated binding capacity, further supporting a diversification of IgE function in ERBs. Finally, in other mammals, the α-β B cell antigen receptor (BCR) heterodimer interacts with the DYXN domain at the Ig CT, phosphorylating the tyrosine and amplifying the B cell signaling. The predicted displacement of the IgE1 DYXN domain due to the 5-aa deletion at the TM-CT interface may affect the efficiency of phosphorylation of the tyrosine and the ability of IgE1 to enhance the α-β BCR heterodimer signaling. Functional data are required to determine the impact of the TM-CT deletion on IgE1 function.

ERBs, humans, and mice contain four IgG subclasses; however, they are not functionally equivalent across species (Collins, 2016). The structure-based protein function predictions assigned each of the ERB IgG subclasses a specialized function that could not be directly correlated with either a human or mouse IgG subclass. Transcriptomic data suggests that ERB IgG1 is the highest expressed subclass (Figure 3A). ERB IgG1 contains the longest hinge and a single CPRCP domain that is also present, although expanded, in human IgG3. The long hinge of human IgG3 likely provides more flexibility to the two Fabs, and reduces the half-life of the Ig due to increased susceptibility to protease degradation. If this is also the case for ERB IgG1, it may explain the rapid decline in MARV antibody response *in vivo* (Schuh et al., 2017b). Moreover, ERB IgG1 has a putative N-glycosylation site at the hinge, not observed in any of the other ERB nor the human IgG subclasses, which may have a profound impact on the IgG structure and dimerization kinetics. ERB IgG2, IgG3, and IgG4 lack the CXXC hinge domain (Figure S3E), associated with the ability of human IgGs to form disulfide bonds between IGH chains. Interestingly, human IgG4 contains a CPSC motif in that region and was shown to undergo Fab-arm exchange to create ''half molecules,'' in which the IGH:IGL dimers are associated by non-covalent interaction (Bloom et al., 1997). Thus, human IgG4 does not crosslink antigen and may be bispecificity (Schuurman et al., 1999; van der Zee et al., 1986). The fact that ERB IgG2, IgG3, and IgG4 lack the hinge CXXC domain

suggests these IgGs may lack inter-chain disulfide bonds between IGHs, and potentially form bispecific half-molecules. This hypothesis, if confirmed experimentally, would likely affect how ERB humoral immunity functions. Human IgG4 contributes to anti-inflammatory properties, limiting its ability to form immunocomplexes and activate complement, and it is also involved in the regulation of IgE function (James and Till, 2016). If recapitulated for ERB IgGs, except IgG1, then this would represent another mechanism by which ERBs effectively control infection while minimizing inflammation.

FcR repertoire, characteristics, and theoretical functions

ERB FcRs resemble those found in mice, lacking the expansion observed in primates and specifically in humans (Bruhns and Jönsson, 2015; Nimmerjahn and Ravetch, 2006). While the inhibitory receptor *FcγRIIB* is well conserved, ERBs do not contain the expanded human activation FcRs (*FcγRIIA* and *FCγIIC*) or the glycophosphatidylinositol (GPI)-linked *FcγRIIIB* highly expressed in human neutrophils (Treffers et al., 2019). The ERBs FcγRs found can only transduce an activating signal through the accessory signaling molecules (γ and β chains). Moreover, the high-affinity ERB *FcγRI* has a particular feature in its TM sequence (Figures 4A and 4B). While human, cynomolgus, rhesus macaques, and mouse FcγRIs have a conserved TM motif with an Asn^{306} (in human: LAVGIMFLV**N**TVL), the ERB FcγRI TM motif contains an aspartic acid in that position (LMVVTMFLV**D**TVF), a change that closely resembles the human FcγRIIIA and FcεRI motifs (Figure 4B). Asn^{306} has been shown to be crucial for the association with γ chain signaling and rendering FcR membrane expression independent of γ chain expression (Ernst et al., 1993; Scholl and Geha, 1993; Stockner et al., 2004). Therefore, Asp^{306} likely mandates interaction with the signaling γ chain, resulting in receptor membrane expression dependent on γ chain expression (Chenoweth et al., 2015).

FcγRI is the high-affinity IgG receptor (Allen and Seed, 1989). The D2 domain FG loop in human FcγRI (^{171}MGKHRY^{176}) is 1 aa shorter than the low-affinity FcγRII and FcγRIII equivalent regions. A valine insertion in this structure (e.g., FcγRIII: ^{171}M**V**GKHRY^{177}) is sufficient to decrease FcγRI affinity 15-fold for IgG (Lu et al., 2011). A statistical analysis of this loop in different species showed that additional aas drastically decrease affinity for the IgG (Lu et al., 2011). ERB *FcγRI* has 8 aas at the D2 domain FG loop (^{171}VGTSRRTF^{178}) instead of the six observed in human, non-human primate (NHP), and mouse (Figure 4C). Moreover, although all ERB IgGs conserve the N-glycosylation site at Asn^{297}, the CHγ2 BC loop Asp^{265}, and the CHγ2 FG loop, ERB IgG2, IgG3, and IgG4 contain a leucine-proline substitution ($Leu^{233}Pro^{234}Gly^{235}Gly^{236}$), associated with the inability of the IgG to bind FcγRs (Arduin et al., 2015; Wines et al., 2000). The L234P substitution, together with the ERB FcγRI D2 domain FG loop structure, strongly suggests that the FcγR affinity for the different ERB IgGs may be negatively affected. Since FcγRI has been implicated in inflammation in human and mouse (Barnes et al., 2002; Mancardi et al., 2013), the distinct features observed in ERB FcγRI may imply that inflammation regulated by this receptor requires a higher threshold of activation.

ERBs lack the short pentraxins involved in the acute-phase responses

The fact that we could not identify functional short pentraxins is without known precedent in mammals. Pentraxins are a superfamily of conserved proteins that play an important role as humoral components of innate immunity that stimulate the complement system (Bottazzi et al., 2010). Pentraxins appeared early in the evolution of metazoans and have been found in vertebrates, non-vertebrate chordates, arthropods, and mollusks. The fact that these different phyla have preserved pentraxins throughout evolution is indicative of their important role in immunity. Within this superfamily, the short pentraxins (CRP and SAP) are the prototypic acute-phase response proteins (Lu et al., 2018). They are produced in the liver and promote complement activation, phagocytosis, and inflammation. Pentraxins interact with different ligands, including complement and FcRs. To our knowledge, the complete absence of both types of short pentraxins in a mammal has never been reported and may be indicative of the evolutionary pressure to reduce acute inflammation.

Overall, we have identified a suite of features specific to ERBs that may aid in their ability to overcome infection, reduce inflammation, and remain largely asymptomatic to acute infection with MARV. Although the genetic and predictive observations presented here need to be confirmed by functional assays before we can arrive at definitive conclusions, it is striking to identify so many distinctive genomic features in ERB immunity. While there is more to learn about the bat immune system, all of our findings support previous claims that ERBs are biased to tolerate viral infection with reduced inflammation.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - BAC identification and growth
  - PacBio BAC sequencing
  - Assembly of PacBio BAC reads
  - Geneious sequence manipulation and assemblies
  - Annotations of constant genes
  - IgE Annotations
  - IgG Annotations
  - IGHM/A Annotation
  - Multiple sequence alignments and phylogenetic analyses
  - Annotation of IGHV, IGHD, and IGHJ genes
  - Gene Ontology for function prediction
  - Tissue expression and visualization
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.109140.

ACKNOWLEDGMENTS

This work has been funded by the DTRA (HDTRA1-14-1-0016), the NSF (grant no. 1757346), and the ORISE. Any opinions, findings, and conclusions are those of the author(s) and do not necessarily reflect the views of the NSF, CDC, or USAMRIID.

AUTHOR CONTRIBUTIONS

Conceptualization, P.A.L., M.L.B., J.T., G.P., and M.S.-L.; data curation, P.A.L., M.L.B., J.C., and K.G.; formal analysis, visualization and writing – original draft, P.A.L., M.L.B., G.P., and M.S.-L.; funding acquisition, J.T. and G.P.; investigation, P.A.L., M.L.B., K.G., G.P., and M.S.-L.; methodology, P.A.L., M.L.B., and K.G.; resources, A.B.-B. and J.T.; software, K.G.; supervision, J.T., J.K., G.P., and M.S.-L.; writing – review & editing, P.A.L., M.L.B., A.B.-B., J.T., J.K., G.P., and M.S.-L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper received support from a program designed to increase minority representation in science.

Received: March 30, 2020
Revised: October 8, 2020
Accepted: April 26, 2021
Published: May 18, 2021

SUPPORTING CITATIONS

The following references appear in the supplemental information: Arnold et al. (2005); Corti, et al. (2016); Davis et al. (2019); Flyak et al. (2015); Garrido et al. (2018); Hu et al. (2019); Jackson et al. (2014); King et al. (2018); Lai et al. (2019); Peng et al. (2019); West et al. (2018); Wrammert et al. (2011); Ying et al. (2014); Yoo et al. (2010).

REFERENCES

Agnarsson, I., Zambrana-Torrelio, C.M., Flores-Saldana, N.P., and May-Collado, L.J. (2011). A time-calibrated species-level phylogeny of bats (Chiroptera, Mammalia). PLoS Curr. *3*, RRN1212.

Ahn, M., Anderson, D.E., Zhang, Q., Tan, C.W., Lim, B.L., Luko, K., Wen, M., Chia, W.N., Mani, S., Wang, L.C., et al. (2019). Dampened NLRP3-mediated inflammation in bats and implications for a special viral reservoir host. Nat. Microbiol. *4*, 789–799.

Akula, S., Mohammadamin, S., and Hellman, L. (2014). Fc receptors for immunoglobulins and their appearance during vertebrate evolution. PLoS ONE *9*, e96903.

Alhakami, H., Mirebrahim, H., and Lonardi, S. (2017). A comparative evaluation of genome assembly reconciliation tools. Genome Biol. *18*, 93.

Allen, J.M., and Seed, B. (1989). Isolation and expression of functional high-affinity Fc receptor complementary DNAs. Science *243*, 378–381.

Amman, B.R., Jones, M.E., 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. (2015a). Oral shedding of Marburg virus in experimentally infected Egyptian fruit bats (Rousettus aegyptiacus). J. Wildl. Dis. *51*, 113–124.

Amman, B.R., Albariño, C.G., Bird, B.H., Nyakarahuka, L., Sealy, T.K., Balinandi, S., Schuh, A.J., Campbell, S.M., Ströher, U., Jones, M.E., et al. (2015b). A Recently Discovered Pathogenic Paramyxovirus, Sosuga Virus, is Present in Rousettus aegyptiacus Fruit Bats at Multiple Locations in Uganda. J. Wildl. Dis. *51*, 774–779.

Arduin, E., Arora, S., Bamert, P.R., Kuiper, T., Popp, S., Geisse, S., Grau, R., Calzascia, T., Zenke, G., and Kovarik, J. (2015). Highly reduced binding to high and low affinity mouse Fc gamma receptors by L234A/L235A and N297A Fc mutations engineered into mouse IgG2a. Mol. Immunol. *63*, 456–463.

Arnaout, R., Lee, W., Cahill, P., Honan, T., Sparrow, T., Weiand, M., Nusbaum, C., Rajewsky, K., and Koralov, S.B. (2011). High-resolution description of antibody heavy-chain repertoires in humans. PLoS ONE *6*, e22365.

Arnold, J.N., Radcliffe, C.M., Wormald, M.R., Royle, L., Harvey, D.J., Crispin, M., Dwek, R.A., Sim, R.B., and Rudd, P.M. (2004). The glycosylation of human serum IgD and IgE and the accessibility of identified oligomannose structures for interaction with mannan-binding lectin. J. Immunol. *173*, 6831–6840.

Arnold, J.N., Wormald, M.R., Suter, D.M., Radcliffe, C.M., Harvey, D.J., Dwek, R.A., Rudd, P.M., and Sim, R.B. (2005). Human serum IgM glycosylation: identification of glycoforms that can bind to mannan-binding lectin. J. Biol. Chem. *280*, 29080–29087.

Arnold, C.E., Guito, J.C., Altamura, L.A., Lovett, S.P., Nagle, E.R., Palacios, G.F., Sanchez-Lockhart, M., and Towner, J.S. (2018). Transcriptomics Reveal Antiviral Gene Induction in the Egyptian Rousette Bat Is Antagonized In Vitro by Marburg Virus Infection. Viruses *10*, 607.

Baker, M.L., Tachedjian, M., and Wang, L.F. (2010). Immunoglobulin heavy chain diversity in Pteropid bats: evidence for a diverse and highly specific antigen binding repertoire. Immunogenetics *62*, 173–184.

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

Barnes, N., Gavin, A.L., Tan, P.S., Mottram, P., Koentgen, F., and Hogarth, P.M. (2002). FcgammaRI-deficient mice show multiple alterations to inflammatory and immune responses. Immunity *16*, 379–389.

Bloom, J.W., Madanat, M.S., Marriott, D., Wong, T., and Chan, S.Y. (1997). Intrachain disulfide bond in the core hinge region of human IgG4. Protein Sci. *6*, 407–415.

Bottazzi, B., Doni, A., Garlanda, C., and Mantovani, A. (2010). An integrated view of humoral innate immunity: pentraxins as a paradigm. Annu. Rev. Immunol. *28*, 157–183.

Bournazos, S., and Ravetch, J.V. (2017). Diversification of IgG effector functions. Int. Immunol. *29*, 303–310.

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

Bruhns, P., and Jönsson, F. (2015). Mouse and human FcR effector functions. Immunol. Rev. *268*, 25–51.

Butler, J.E., Wertz, N., Zhao, Y., Zhang, S., Bao, Y., Bratsch, S., Kunz, T.H., Whitaker, J.O., Jr., and Schountz, T. (2011). The two suborders of chiropterans have the canonical heavy-chain immunoglobulin (Ig) gene repertoire of eutherian mammals. Dev. Comp. Immunol. *35*, 273–284.

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.

Cerutti, A., Chen, K., and Chorny, A. (2011). Immunoglobulin responses at the mucosal interface. Annu. Rev. Immunol. *29*, 273–293.

Chen, X., Li, G., Wan, Z., Liu, C., Zeng, Y., and Liu, W. (2015). How B cells remember? A sophisticated cytoplasmic tail of mIgG is pivotal for the enhanced transmembrane signaling of IgG-switched memory B cells. Prog. Biophys. Mol. Biol. *118*, 89–94.

Chenoweth, A.M., Trist, H.M., Tan, P.S., Wines, B.D., and Hogarth, P.M. (2015). The high-affinity receptor for IgG, FcγRI, of humans and non-human primates. Immunol. Rev. *268*, 175–191.

Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., et al. (2013). Nonhybrid,

finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods *10*, 563–569.

Cohen-Dvashi, H., Zehner, M., Ehrhardt, S., Katz, M., Elad, N., Klein, F., and Diskin, R. (2020). Structural Basis for a Convergent Immune Response against Ebola Virus. Cell Host Microbe *27*, 418–427.e4.

Collins, A.M. (2016). IgG subclass co-expression brings harmony to the quartet model of murine IgG function. Immunol. Cell Biol. *94*, 949–954.

Corti, D., Misasi, J., Mulangu, S., Stanley, D.A., Kanekiyo, M., Wollen, S., Ploquin, A., Doria-Rose, N.A., Staupe, R.P., Bailey, M., et al. (2016). Protective monotherapy against lethal Ebola virus infection by a potently neutralizing antibody. Science *351*, 1339–1342.

Dard, P., Huck, S., Frippiat, J.P., Lefranc, G., Langaney, A., Lefranc, M.P., and Sanchez-Mazas, A. (1997). The IGHG3 gene shows a structural polymorphism characterized by different hinge lengths: sequence of a new 2-exon hinge gene. Hum. Genet. *99*, 138–141.

Davis, C.W., Jackson, K.J.L., McElroy, A.K., Halfmann, P., Huang, J., Chennareddy, C., Piper, A.E., Leung, Y., Albarino, C.G., Crozier, I., et al. (2019). Longitudinal Analysis of the Human B Cell Response to Ebola Virus Infection. Cell *177*, 1566–1582.e17.

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

Elemento, O., and Lefranc, M.P. (2003). IMGT/PhyloGene: an on-line tool for comparative analysis of immunoglobulin and T cell receptor genes. Dev. Comp. Immunol. *27*, 763–779.

Ernst, L.K., Duchemin, A.M., and Anderson, C.L. (1993). Association of the high-affinity receptor for IgG (Fc gamma RI) with the gamma subunit of the IgE receptor. Proc. Natl. Acad. Sci. USA *90*, 6023–6027.

Flyak, A.I., Ilinykh, P.A., Murin, C.D., Garron, T., Shen, X., Fusco, M.L., Hashiguchi, T., Bornholdt, Z.A., Slaughter, J.C., Sapparapu, G., et al. (2015). Mechanism of human antibody-mediated neutralization of Marburg virus. Cell *160*, 893–903.

Garrido, J.L., Prescott, J., Calvo, M., Bravo, F., Alvarez, R., Salas, A., Riquelme, R., Rioseco, M.L., Williamson, B.N., Haddock, E., et al. (2018). Two recombinant human monoclonal antibodies that protect against lethal Andes hantavirus infection in vivo. Sci. Transl. Med. *10*, eaat6420.

Gertz, E.M., Schäffer, A.A., Agarwala, R., Bonnet-Garnier, A., Rogel-Gaillard, C., Hayes, H., and Mage, R.G. (2013). Accuracy and coverage assessment of Oryctolagus cuniculus (rabbit) genes encoding immunoglobulins in the whole genome sequence assembly (OryCun2.0) and localization of the IGH locus to chromosome 20. Immunogenetics *65*, 749–762.

Hellman, L.T., Akula, S., Thorpe, M., and Fu, Z. (2017). Tracing the Origins of IgE, Mast Cells, and Allergies by Studies of Wild Animals. Front. Immunol. *8*, 1749.

Hu, D., Zhu, Z., Li, S., Deng, Y., Wu, Y., Zhang, N., Puri, V., Wang, C., Zou, P., Lei, C., et al. (2019). A broadly neutralizing germline-like human monoclonal antibody against dengue virus envelope domain III. PLoS Pathog. *15*, e1007836.

Jackson, K.J., Liu, Y., Roskin, K.M., Glanville, J., Hoh, R.A., Seo, K., Marshall, E.L., Gurley, T.C., Moody, M.A., Haynes, B.F., et al. (2014). Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements. Cell Host Microbe *16*, 105–114.

James, L.K., and Till, S.J. (2016). Potential Mechanisms for IgG4 Inhibition of Immediate Hypersensitivity Reactions. Curr. Allergy Asthma Rep. *16*, 23.

Jebb, D., Huang, Z., Pippel, M., Hughes, G.M., Lavrichenko, K., Devanna, P., Winkler, S., Jermiin, L.S., Skirmuntt, E.C., Katzourakis, A., et al. (2020). Six reference-quality genomes reveal evolution of bat adaptations. Nature *583*, 578–584.

Kelly, B.T., and Grayson, M.H. (2016). Immunoglobulin E, what is it good for? Ann. Allergy Asthma Immunol. *116*, 183–187.

King, L.B., Fusco, M.L., Flyak, A.I., Ilinykh, P.A., Huang, K., Gunn, B., Kirchdoerfer, R.N., Hastie, K.M., Sangha, A.K., Meiler, J., et al. (2018). The Marburg-virus-Neutralizing Human Monoclonal Antibody MR191 Targets a Conserved Site to Block Virus Receptor Binding. Cell Host Microbe *23*, 101–109.e4.

Kiyoshi, M., Caaveiro, J.M., Kawai, T., Tashiro, S., Ide, T., Asaoka, Y., Hatayama, K., and Tsumoto, K. (2015). Structural basis for binding of human IgG1 to its high-affinity human receptor FcγRI. Nat. Commun. *6*, 6866.

Koenderman, L. (2019). Inside-Out Control of Fc-Receptors. Front. Immunol. *10*, 544.

Kolde, R., Laur, S., Adler, P., and Vilo, J. (2012). Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics *28*, 573–580.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res. *27*, 722–736.

Lai, M., Wang, Q., Lu, Y., Xu, X., Xia, Y., Tu, M., Liu, Y., Zhang, Q., Peng, Y., and Zheng, X. (2019). Signatures of B-cell receptor diversity in B lymphocytes following Epstein-Barr virus transformation. Physiol. Genomics *51*, 197–207.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al.; International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

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.

Lefranc, M.P. (2014). Immunoglobulins: 25 years of immunoinformatics and IMGT-ONTOLOGY. Biomolecules *4*, 1102–1139.

Lefranc, M.P., Giudicelli, V., Ginestoux, C., Bodmer, J., Müller, W., Bontrop, R., Lemaitre, M., Malik, A., Barbié, V., and Chaume, D. (1999). IMGT, the international ImMunoGeneTics database. Nucleic Acids Res *27*, 209–212.

Lefranc, M.P., Pommié, C., Kaas, Q., Duprat, E., Bosc, N., Guiraudou, D., Jean, C., Ruiz, M., Da Piédade, I., Rouard, M., et al. (2005). IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. Dev. Comp. Immunol. *29*, 185–203.

Lefranc, M.P., Giudicelli, V., Duroux, P., Jabado-Michaloud, J., Folch, G., Aouinti, S., Carillon, E., Duvergey, H., Houles, A., Paysan-Lafosse, T., et al. (2015). IMGT®, the international ImMunoGeneTics information system® 25 years on. Nucleic Acids Res. *43*, D413–D422.

Lu, J., Ellsworth, J.L., Hamacher, N., Oak, S.W., and Sun, P.D. (2011). Crystal structure of Fcγ receptor I and its implication in high affinity γ-immunoglobulin binding. J. Biol. Chem. *286*, 40608–40613.

Lu, J., Chu, J., Zou, Z., Hamacher, N.B., Rixon, M.W., and Sun, P.D. (2015). Structure of FcγRI in complex with Fc reveals the importance of glycan recognition for high-affinity IgG binding. Proc. Natl. Acad. Sci. USA *112*, 833–838.

Lu, J., Mold, C., Du Clos, T.W., and Sun, P.D. (2018). Pentraxins and Fc Receptor-Mediated Immune Responses. Front. Immunol. *9*, 2607.

Mancardi, D.A., Albanesi, M., Jönsson, F., Iannascoli, B., Van Rooijen, N., Kang, X., England, P., Daëron, M., and Bruhns, P. (2013). The high-affinity human IgG receptor FcγRI (CD64) promotes IgG-mediated inflammation, anaphylaxis, and antitumor immunotherapy. Blood *121*, 1563–1573.

Mandl, J.N., Schneider, C., Schneider, D.S., and Baker, M.L. (2018). Going to Bat(s) for Studies of Disease Tolerance. Front. Immunol. *9*, 2112.

Max, E.E., Battey, J., Ney, R., Kirsch, I.R., and Leder, P. (1982). Duplication and deletion in the human immunoglobulin epsilon genes. Cell *29*, 691–699.

McBride, O.W., Battey, J., Hollis, G.F., Swan, D.C., Siebenlist, U., and Leder, P. (1982). Localization of human variable and constant region immunoglobulin heavy chain genes on subtelomeric band q32 of chromosome 14. Nucleic Acids Res. *10*, 8155–8170.

Merle, N.S., Church, S.E., Fremeaux-Bacchi, V., and Roumenina, L.T. (2015a). Complement System Part I - Molecular Mechanisms of Activation and Regulation. Front. Immunol. *6*, 262.

Merle, N.S., Noe, R., Halbwachs-Mecarelli, L., Fremeaux-Bacchi, V., and Roumenina, L.T. (2015b). Complement System Part II: Role in Immunity. Front. Immunol. *6*, 257.

Nimmerjahn, F., and Ravetch, J.V. (2006). Fcgamma receptors: old friends and new family members. Immunity *24*, 19–28.

Oettgen, H.C. (2016). Fifty years later: emerging functions of IgE antibodies in host defense, immune regulation, and allergic diseases. J. Allergy Clin. Immunol. *137*, 1631–1645.

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.

Pavlovich, S.S., Lovett, S.P., Koroleva, G., Guito, J.C., Arnold, C.E., Nagle, E.R., Kulcsar, K., Lee, A., Thibaud-Nissen, F., Hume, A.J., et al. (2018). The Egyptian Rousette Genome Reveals Unexpected Features of Bat Antiviral Immunity. Cell *173*, 1098–1110.e18.

Peng, W., Liu, S., Meng, J., Huang, J., Huang, J., Tang, D., and Dai, Y. (2019). Profiling the TRB and IGH repertoire of patients with H5N6 Avian Influenza Virus Infection by high-throughput sequencing. Sci. Rep. *9*, 7429.

Plomp, R., Ruhaak, L.R., Uh, H.W., Reiding, K.R., Selman, M., Houwing-Duistermaat, J.J., Slagboom, P.E., Beekman, M., and Wuhrer, M. (2017). Subclass-specific IgG glycosylation is associated with markers of inflammation and metabolic health. Sci. Rep. *7*, 12325.

Pyzik, M., Sand, K.M.K., Hubbard, J.J., Andersen, J.T., Sandlie, I., and Blumberg, R.S. (2019). The Neonatal Fc Receptor (FcRn): A Misnomer? Front. Immunol. *10*, 1540.

Reams, A.B., and Roth, J.R. (2015). Mechanisms of gene duplication and amplification. Cold Spring Harb. Perspect. Biol. *7*, a016592.

Roy, A., Yang, J., and Zhang, Y. (2012). COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. Nucleic Acids Res. *40*, W471–W477.

Scholl, P.R., and Geha, R.S. (1993). Physical association between the high-affinity IgG receptor (Fc gamma RI) and the gamma subunit of the high-affinity IgE receptor (Fc epsilon RI gamma). Proc. Natl. Acad. Sci. USA *90*, 8847–8850.

Schountz, T., Baker, M.L., Butler, J., and Munster, V. (2017). Immunological Control of Viral Infections in Bats and the Emergence of Viruses Highly Pathogenic to Humans. Front. Immunol. *8*, 1098.

Schroeder, H.W., Jr., and Cavacini, L. (2010). Structure and function of immunoglobulins. J. Allergy Clin. Immunol. *125* (*2, Suppl 2*), S41–S52.

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. (2017a). Modelling filovirus maintenance in nature by experimental transmission of Marburg virus between Egyptian rousette bats. Nat. Commun. *8*, 14446.

Schuh, A.J., Amman, B.R., Sealy, T.K., Spengler, J.R., Nichol, S.T., and Towner, J.S. (2017b). Egyptian rousette bats maintain long-term protective immunity against Marburg virus infection despite diminished antibody levels. Sci. Rep. *7*, 8763.

Schuh, A.J., Amman, B.R., Sealy, T.K., Kainulainen, M.H., Chakrabarti, A.K., Guerrero, L.W., Nichol, S.T., Albarino, C.G., and Towner, J.S. (2019). Antibody-Mediated Virus Neutralization Is Not a Universal Mechanism of Marburg, Ebola, or Sosuga Virus Clearance in Egyptian Rousette Bats. J. Infect. Dis. *219*, 1716–1721.

Schuurman, J., Van Ree, R., Perdok, G.J., Van Doorn, H.R., Tan, K.Y., and Aalberse, R.C. (1999). Normal human immunoglobulin G4 is bispecific: it has two different antigen-combining sites. Immunology *97*, 693–698.

Shi, B., Dong, X., Ma, Q., Sun, S., Ma, L., Yu, J., Wang, X., Pan, J., He, X., Su, D., et al. (2020). The usage of human IGHJ genes follows a particular non-random selection: the recombination signal sequence may affect the usage of human IGHJ genes. Front. Genet. *11*, 254413.

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.

Sotero-Caio, C.G., Baker, R.J., and Volleth, M. (2017). Chromosomal Evolution in Chiroptera. Genes (Basel) *8*, 272.

Stockner, T., Ash, W.L., MacCallum, J.L., and Tieleman, D.P. (2004). Direct simulation of transmembrane helix association: role of asparagines. Biophys. J. *87*, 1650–1656.

Storm, N., Jansen Van Vuren, P., Markotter, W., and Paweska, J.T. (2018). Antibody Responses to Marburg Virus in Egyptian Rousette Bats and Their Role in Protection against Infection. Viruses *10*, 73.

Sun, Y., Liu, Z., Ren, L., Wei, Z., Wang, P., Li, N., and Zhao, Y. (2012). Immunoglobulin genes and diversity: what we have learned from domestic animals. J. Anim. Sci. Biotechnol. *3*, 18.

Sutton, B.J., and Davies, A.M. (2015). Structure and dynamics of IgE-receptor interactions: FcεRI and CD23/FcεRII. Immunol. Rev. *268*, 222–235.

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.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.

Towner, J.S., Khristova, M.L., Sealy, T.K., Vincent, M.J., Erickson, B.R., Bawiec, D.A., Hartman, A.L., Comer, J.A., Zaki, S.R., Ströher, U., et al. (2006). Marburgvirus genomics and association with a large hemorrhagic fever outbreak in Angola. J. Virol. *80*, 6497–6516.

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.

Treffers, L.W., van Houdt, M., Bruggeman, C.W., Heineke, M.H., Zhao, X.W., van der Heijden, J., Nagelkerke, S.Q., Verkuijlen, P.J.J.H., Geissler, J., Lissenberg-Thunnissen, S., et al. (2019). FcγRIIIb Restricts Antibody-Dependent Destruction of Cancer Cells by Human Neutrophils. Front. Immunol. *9*, 3124.

Tsakou, E., Agathangelidis, A., Boudjoghra, M., Raff, T., Dagklis, A., Chatzouli, M., Smilevska, T., Bourikas, G., Merle-Beral, H., Manioudaki-Kavallieratou, E., et al. (2012). Partial versus productive immunoglobulin heavy locus rearrangements in chronic lymphocytic leukemia: implications for B-cell receptor stereotypy. Mol. Med. *18*, 138–145.

van der Zee, J.S., van Swieten, P., and Aalberse, R.C. (1986). Serologic aspects of IgG4 antibodies. II. IgG4 antibodies form small, nonprecipitating immune complexes due to functional monovalency. J. Immunol. *137*, 3566–3571.

Vernersson, M., Aveskogh, M., and Hellman, L. (2004). Cloning of IgE from the echidna (Tachyglossus aculeatus) and a comparative analysis of epsilon chains from all three extant mammalian lineages. Dev. Comp. Immunol. *28*, 61–75.

Vidarsson, G., Dekkers, G., and Rispens, T. (2014). IgG subclasses and allotypes: from structure to effector functions. Front. Immunol. *5*, 520.

Watson, C.T., and Breden, F. (2012). The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease. Genes Immun. *13*, 363–373.

Watson, C.T., Glanville, J., and Marasco, W.A. (2017). The Individual and Population Genetics of Antibody Immunity. Trends Immunol. *38*, 459–470.

West, B.R., Moyer, C.L., King, L.B., Fusco, M.L., Milligan, J.C., Hui, S., and Saphire, E.O. (2018). Structural Basis of Pan-Ebolavirus Neutralization by a Human Antibody against a Conserved, yet Cryptic Epitope. MBio *9*, e01674-18.

Wines, B.D., Powell, M.S., Parren, P.W., Barnes, N., and Hogarth, P.M. (2000). The IgG Fc contains distinct Fc receptor (FcR) binding sites: the leukocyte receptors Fc gamma RI and Fc gamma RIIa bind to a region in the Fc distinct from that recognized by neonatal FcR and protein A. J. Immunol. *164*, 5313–5318.

Wrammert, J., Koutsonanos, D., Li, G.M., Edupuganti, S., Sui, J., Morrissey, M., McCausland, M., Skountzou, I., Hornig, M., Lipkin, W.I., et al. (2011). Broadly cross-reactive antibodies dominate the human B cell response against 2009 pandemic H1N1 influenza virus infection. J. Exp. Med. *208*, 181–193.

Wynne, J.W., and Wang, L.F. (2013). Bats and viruses: friend or foe? PLoS Pathog. *9*, e1003651.

Ying, T., Du, L., Ju, T.W., Prabakaran, P., Lau, C.C., Lu, L., Liu, Q., Wang, L., Feng, Y., Wang, Y., et al. (2014). Exceptionally potent neutralization of Middle East respiratory syndrome coronavirus by human monoclonal antibodies. J. Virol. *88*, 7796–7805.

Yoo, E.M., Yu, L.J., Wims, L.A., Goldberg, D., and Morrison, S.L. (2010). Differences in N-glycan structures found on recombinant IgA1 and IgA2 produced in murine myeloma and CHO cell lines. MAbs *2*, 320–334.

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, C., Freddolino, P.L., and Zhang, Y. (2017). COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. Nucleic Acids Res. *45* (*W1*), W291–W299.

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.

Zucchetti, I., De Santis, R., Grusea, S., Pontarotti, P., and Du Pasquier, L. (2009). Origin and evolution of the vertebrate leukocyte receptors: the lesson from tunicates. Immunogenetics *61*, 463–481.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
pCC1BAC epicenter vector in Invitrogen DH10b Phage Resistant	AMplicon Express	N/A
Biological samples		
Healthy, wild-caught, male Egyptian Rousette bat (*R. aegyptiacus*) – liver tissue	Friedrich-Loeffler-Institut, Germany	N/A
Chemicals, peptides, and recombinant proteins		
Chloramphenicol	RPI	C61000-25.0
Critical commercial assays		
Plasmid Maxi kit	QIAGEN	Cat# 12162
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
Blue Pippin system 0.75% dye-free agarose gel cassette, marker S1 and High-pass weight kit	Sage Science	BLF-7510
Qubit dsDNA BR Assay kit	ThermoFisher Scientific	Cat# Q32853
Deposited data		
RaegypIGH3.0 genome sequence	This paper	GenBank: BankIt2442758 Seq0, MW800879
Raegyp2.0 genome sequence	Pavlovich et al., 2018	GenBank: GCA_001466805.2; WGS Project: LOCP02
R. aegyptiacus transcriptome	Lee et al., 2015	GenBank: GECF00000000.1; SRA Project: SRP066106
RefSeq proteins – Homo sapiens	Genome Reference Consortium; NCBI RefSeq	RefSeq: GCF_000001405.33
RefSeq proteins – Rousettus aegyptiacus	Pavlovich et al., 2018	RefSeq: GCF_001466805.2
RefSeq proteins – *Mus musculus*	Genome Reference Consortium; NCBI RefSeq	RefSeq: GCF_000001635.24
Oligonucleotides		
BAC Forward Primer Pair 1: CACCCGGCATAAAATAAAACAC	This paper	N/A
BAC Reverse Primer Pair 1: ACAGGTCTCGTCCAAGAGAATC	This paper	N/A
BAC Forward Primer Pair 2: CTCTTGAGTTTGGCAAATTGTTC	This paper	N/A
BAC Reverse Primer Pair 2: TGAGGACTCAGCCTTACAGTTTC	This paper	N/A
BAC Forward Primer Pair 3: CAGATGGACTACCCCTTTATTCC/	This paper	N/A
BAC Reverse Primer Pair 3: GGCCTTGTTTGTTTATGTCGTAG	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BAC Forward Primer Pair 4: GGTTCTCATTGTCAGAGTTGGTC	This paper	N/A
BAC Reverse Primer Pair 4: TACGTTCCTAGCACTTTTGAAGC	This paper	N/A
BAC Forward Primer Pair 5: ATGTTCCCAAAATACAACAGTGG	This paper	N/A
BAC Reverse Primer Pair 5: AGGTAGTTAAACCTGGGAGCTTG	This paper	N/A
Software and algorithms		
Geneious	Geneious 11.1.1	https://www.geneious.com/
ClustalW	Thompson et al., 1994	https://www.ebi.ac.uk/Tools/msa/clustalo/
IMGT LigMotif	(Lefranc et al., 1999)	http://www.imgt.org/ligmotif/
PicardTools v1.131	Broad Institute	http://broadinstitute.github.io/picard/
Canu	Koren et al., 2017	https://canu.readthedocs.io/en/latest/
SMRTLink	Chin et al., 2013	https://www.pacb.com/support/software-downloads/
Mega 6.0	Tamura et al., 2013	https://www.megasoftware.net/
HGAP	Chin et al., 2013	https://github.com/PacificBiosciences/Bioinformatics-Training/wiki/HGAP
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 et al., 2012	https://cran.r-project.org/web/packages/pheatmap/
COFACTOR	Zhang et al., 2017	https://zhanglab.dcmb.med.umich.edu/COFACTOR/
Other		

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Mariano Sanchez-Lockhart (mariano.sanchez-lockhart.civ@mail.mil).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The annotated sequence of RaegypIGH3.0 (variable, diveristy, joining and constant genes) is available at GenBank accession number: BankIt2442758 Seq0, MW800879

This study generate a unique script to parse sequenced BACs from PacBio reads and is available upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Genomic DNA was isolated from the liver tissue of a wild-caught healthy male Egyptian rousette bat from a captive colony at Friedrich-Loeffler-Institut, Germany. This individual bat was a descendant of the bats brought to Europe in the 1960s. 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.

METHOD DETAILS

BAC identification and growth

The *Rousettus aegyptiacus* genome (Raegyp2.0; GCA_001466805.2) has been archived since the inception of this work. As such when performing BLAST the contigs indicated below will have to be manually entered when performing search functions. Bacterial artificial chromosome (BAC) libraries were created from ERB liver tissue (male, 13 months old, captive colony at Friedrich-Loeffler-Institut, Germany from ancestor individuals that were brought to Europe in the 1960s) supplied to a commercial vendor (AMplicon Express). To isolate BACs of interest we screened BAC libraries using a polymerase chain reaction (PCR) scheme. Forward and reverse primer pairs unique to single contigs that would result in ~1100-1600bp amplicons were created. Primers were developed to amplify regions at distal ends of contigs NW_015493306.1, NW_015494929.1, NW_015493824.1, and NW_015493956.1 (Figure 1A). Pools of BACs were screened using these primer pairs resulting in the identification of a single BAC. Due to the overlap of some BACs a single primer pair could identify more than one BAC clone, in these cases both BACs were used.

Once identified, BACs were picked from 384-well plates and grown in 250mL of 2X YT broth in the presence of 12.5ug/mL Chloramphenicol for 18-24 hours. BACs were isolated from cultures using the QIAGEN Plasmid Maxi kit following the manufacturer's recommendations. Purified nucleic acids were then subject to Pacific Biosciences library preparation and sequencing.

PacBio BAC sequencing

Isolated BAC DNA was sheared to ~20kb average size using needle shearing. After shearing, DNA damage repair and end repair was performed, followed by ligation of hairpin adapters resulting in a SMRTBell template. SMRTBell templates were subject to ExoIII and ExoVII treatment to remove unligated 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. Individual BACs were barcoded using 7bp barcodes and sequenced using the Sequel sequencing kit v2.1. Annealing of sequencing primer and binding polymerase 2.0 to the SMRTbell template was performed according to PacBio calculator and polymerase/template complexes were loaded onto SMRT cells (SMRT Cell 1M v3 Tray) via diffusion at a final concentration of 8pM. Libraries were sequenced with 600 min movies on PacBio Sequel instrument (Pacific Biosciences, Menlo Park, CA). BACIg4 was sequenced as described except we used a SMRT Cell 1M v3 LR Tray and sequenced with 1200 minute movies.

Assembly of PacBio BAC reads

Upon completion of sequencing fastq files from each individual BAC were downloaded from SMRT Link and assembled using the long read assembler Canu v1.8 (Koren et al., 2017) and an in house script for BACIg1, BACIg2, BACIg3. For BACIg4 we used the HGAP4 assembler which is embedded within the SMRT Link suite using default settings except the "Genome Length" was set to 250kb and the 'Minimum Subread length' was set to 8kb. Each assembled BAC was subsequently uploaded to SMRT Link and subject to 'polishing' using the embedded 'Resequencing' program. After assembly all sequences were subject to the 'Resequencing' function of SMRTLink which uses all of the reads from a sequencing run to correct errors present in the final assembly. Sequences were subject to 'Resequencing' until they reached a sequence concordance of 99.95% or greater. Resequencing also indicates the average read depth of each assembly. BACIg1 was sequenced with an average coverage of 524x and a concordance of 99.99%, BACIg2 at 501x average coverage and a concordance of 99.98%, BACIg3 at 485x average coverage and a concordance of 99.98%, and BACIg4 at 370x average coverage and a concordance of 99.99%.

Geneious sequence manipulation and assemblies

Upon completion of sequencing fastq files from each individual BAC were downloaded from SMRT Link and assembled using the long read assembler Canu (Koren et al., 2017) for BACs BACIg1, BACIg2, and BACIg3. For BACIg4 we used the HGAP4 assembler which is embedded within the SMRT Link suite. Each assembled BAC was subsequently uploaded to SMRT Link and subject to 'polishing' using the embedded 'Resequencing' program. Completed BAC sequences were uploaded to Geneious. BACIg1-3 were assembled using the '*De Novo* Assembly' function. BACIg3 and BACIg2 overlapped by ~18,000bp and BACIg2 and BACIg1 overlapped by ~50,000bp. BACIg1/4 were assembled using the 'Map to Reference' function using BACIg1 as a reference sequence. The final assembly is represented as complete sequences from BACIg3, BACIg2, and a partial sequence of BACIg1 (Figure 1B). The near perfect alignment (99.99% pairwise identity) among these three BACs likely represents a single haplotype. For the overlap sequence between BACIg1 and BACIg4 (~60,000bp, 96.5% pairwise identity) all BACIg1 sequence (Figure 1B, dotted line) was replaced with BACIg4 sequence. These two BACs likely represent alternative haplotypes and thus the completed ERB IGH locus is a hybrid where all constant genes are representative of one haplotype and all V(D)J genes are representative of the alternative haplotype.

Annotations of constant genes

All annotated IGH constant genes were downloaded from Raegyp2.0 using the NCBI genome browser. Coding sequences (CDS) were extracted and mapped against the completed IGH locus using the annotate function of Geneious. Complete coding sequences were renamed for ease of discussion; for example, the annotated IgE on contig NW_0145493306.1 is labeled as 'Ig epsilon chain C region-like' and bears the locus tag 'LOC107506273' and was subsequently renamed to 'IgE_3306' (bearing the gene name and final four numbers of the contig designation number).

IgE Annotations

NCBI annotated four genes as IgE in Raegyp2.0 and all sequences contained one, or two misannotated exons upstream of the canonical constant heavy 1 (CH1) exon. These misannotated exons were discarded when determining IGH constant gene annotation. Only IgE_4929 CDS contained a complete open reading frame with no premature stop codons, however, it did not contain the final two exons constituting the transmembrane (TM) domains. IgE_3306/3956/3824 contained only minor mutations (< 3 indels per sequence) that resulted in premature stop codons, however, they did contain annotated TM domain exons. Given these minor changes we were confident that the usage of all four sequences for annotating the CDS in RaegypIGH3.0 would yield accurate results. Sequences were uploaded to Geneious and aligned to RaegypIGH3.0 using the 'Annotate from function' with a mapping requirement of 75% identity. Five IgE genes were identified. Annotated exons for each IgE gene were inspected in the context of RaegypIGH3.0 to ensure that exon/intron sequence rules were maintained. No adjustments to exon/intron lengths were necessary. Exon sequences from each gene were then extracted and concatenated to form the final CDS. pIgE1, 2, 3, all contained indels introducing stop codons into the sequence, whereas the CDS for IgE1, 2 were complete and contained no premature stops.

IgG Annotations

NCBI annotated five genes as IgG in Raegyp2.0. Two annotated IgG (IgG_3824_TM, IgG_4929) were only partial genes as they resided at the end of contigs. Complete IgG genes all contained missannotated exons upstream of the CH1 exon. These misannotated exons were discarded when determining IGH constant gene annotation. Only IgG_3306 CDS contained a complete open reading frame with no premature stop codons. The remaining IgGs contained minor mutations (2-6 indels per sequence) that resulted in premature stop codons. Given these minor changes we were confident that the usage of all five sequences for annotating the CDS in RaegypIGH3.0 would yield accurate results. Sequences were uploaded to Geneious and aligned to RaegypIGH3.0 the 'Annotate from function' with a mapping requirement of 75% identity. Four IgG genes were identified. Annotated exons for each IgG gene were inspected in the context of RaegypIGH3.0 to ensure that exon/intron sequence rules were maintained. No adjustments to exon/intron lengths were necessary. Exon sequences from each gene were then extracted and concatenated to form the final CDS. All four IgG genes contained complete open reading frames.

IGHM/A Annotation

NCBI annotated a single IgA and *IGHM* in Raegyp2.0. Both sequences contained missannotated exons upstream of the CH1 exon, however both CDS contained complete open reading frames with no premature stop codons. Sequences were uploaded to Geneious and aligned to RaegypIGH3.0 using the 'Annotate from function' with a mapping requirement of 75% identity. Annotated exons for each IgE gene were inspected in the context of RaegypIGH3.0 to ensure that exon/intron sequence rules were maintained. No adjustments to exon/intron lengths were necessary. Exon sequences from each gene were then extracted and concatenated to form the final CDS. Both *IGHM* and *IGHA* contained complete open reading frames.

Multiple sequence alignments and phylogenetic analyses

Annotated ERB sequences and those downloaded from NCBI were aligned using either ClustalW (nucleotides) or MUSCLE (amino acids). A multi-species alignment of all ERB coding IGH amino acid sequences as well as the indicated bat species (Figure S3) were aligned with all coding human IGH sequences using MUSCLE. A maximum parsimony phylogenetic tree was inferred from that alignment using MrBayes with the human IGHM set as the outgroup.

Annotation of IGHV, IGHD, and IGHJ genes

Raegyp2.0 scaffolds containing IGHV genes (NW_015493306.1, NW_015493547.1, NW_015493575.1, NW_015493590.1, NW_015493682.1, NW_015493749.1, NW_015493769.1, NW_015493770.1, NW_015493822.1, NW_015493830.1, NW_015493970.1, NW_015494112.1, NW_015494214.1, NW_015494338.1, NW_015494352.1, NW_015494409.1, NW_015494440.1) and the final RaegypIGH3.0 assembly were annotated using IMGT-LIGMotif using all databases. Annotations were manually inspected to confirm RSS elements reported. Finally, sequences were extracted and BLAST against the transcriptomic read set described above to confirm unique expression. Reads that mapped to an IGHV or IGHJ gene were then annotated with IMGT-HighVQuest to confirm IGHD annotations, as they may be missed in BLAST due to their small size.

Gene Ontology for function prediction

COFACTOR is an *in silico* protein function prediction algorithm that employs three complementary approaches to surmise protein features (Zhang et al., 2017). The protein sequence is analyzed through three pipelines: 1) structure-function database (BioLiP) where gene ontology GO is predicted by structure, enzyme commission prediction, and ligand binding site prediction, 2) Sequence-function database (Uni-Prot-GOA) where GO is predicted by sequence, 3) protein-protein interaction function database (STRING) where GO is predicted by PPI. COFACTOR requires a pdb structure and so we generated homology models using Phyre2. For IgA, IgM, and IgE composite models were used and for all IgGs the template with the highest confidence (in this case all 100%), and the highest percent identity was selected.

Tissue expression and visualization

Transcriptomic reads from ERBs (Lee et al., 2015) were aligned to genes of interest using the pseudo aligner Kalisto v0.43.0 (Bray et al., 2016). Transcript counts at a gene-level were calculatied by summing the transcript per million (TPM) values for all transcripts of a particular gene. The resulting TPMs were log transformed and normalized to *GAPDH*. Heatmaps were created using the R package pheatmap (Kolde et al., 2012).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical significance reported in figures and legends when appropriate. R-package pheatmap was used for transcript expression and functional prediction scores. Gene ontology (GO) terms listed in figure legends and confident predictions with Cscorego $\geq$ 0.5displayed. Data shown as Cscorego 0-1 where 1 is high confidence and 0 is no confidence.