

Spring 5-9-2020

Fundamental differences in the B cell locus, V(D)J repertoire, and Marburg virus epitope recognition of the Egyptian rousette bat that may allow for asymptomatic presentation

Maggie Linn Bartlett
University of Nebraska Medical Center

Tell us how you used this information in this [short survey](#).

Follow this and additional works at: <https://digitalcommons.unmc.edu/etd>



Part of the [Immunology and Infectious Disease Commons](#)

Recommended Citation

Bartlett, Maggie Linn, "Fundamental differences in the B cell locus, V(D)J repertoire, and Marburg virus epitope recognition of the Egyptian rousette bat that may allow for asymptomatic presentation" (2020). *Theses & Dissertations*. 432.

<https://digitalcommons.unmc.edu/etd/432>

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@UNMC. It has been accepted for inclusion in Theses & Dissertations by an authorized administrator of DigitalCommons@UNMC. For more information, please contact digitalcommons@unmc.edu.

**Fundamental differences in the B cell locus, V(D)J repertoire,
and *Marburgvirus* epitope recognition of the Egyptian rousette
bat that may allow for asymptomatic presentation**

By

Maggie L. Bartlett

A DISSERTATION

Presented to the Faculty of the University of Nebraska Graduate College in Partial
Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Immunology, Pathology, & Infectious Diseases

Graduate Program

Under the Supervision of Professor Mariano Sanchez-Lockhart

April, 2020

Supervisory Committee:

Ken Bayles, Ph.D. Tammy Kielian, Ph.D.

Paul Davis, Ph.D. Gustavo Palacios, Ph.D.

Abstract**Fundamental differences in the B cell locus, V(D)J repertoire, and *Marburgvirus* epitope recognition of the Egyptian rousette bat that may allow for asymptomatic presentation**

Maggie L. Bartlett, Ph.D.

University of Nebraska, 2020

Supervisor: Mariano Sanchez-Lockhart, PhD.

Marburg virus (MARV) causes a hemorrhagic fever in humans but is asymptomatic in a known reservoir, the Egyptian rousette bat (*Rousettus aegyptiacus*, ERB). Understanding the mechanisms that drive these different outcomes could potentially advance the development of therapeutics. The immunoglobulin (Ig) response to MARV infection in ERBs is known to serve a role in protection. The Ig germline encodes Variable (VH), Diversity (DH), and Joining (JH) genes that then recombine (V(D)J) to make up the binding site of an Ig. Understanding the gene composition of the Ig germline is critical to defining the potential B cell repertoire. We hypothesize that the B cell repertoire contributes to the ability of ERBs to overcome MARV infection. However, to study V(D)J rearrangement in ERBs, an accurate annotation of the Ig loci is needed. We constructed one contiguous locus which allowed us to unequivocally define genes and expansions. We described all VH, DH, and JH genes which were only partially annotated or missing entirely in the previous assembly and found an expansion of IGHV genes associated with protection from various pathogens in humans. Strikingly, we found the presence of two functional IgE genes, something not found in any other mammal to date. This is the first complete IGH locus for a bat species and one of only a handful for mammalian species. We conducted *in silico* functional prediction assays for the Ig isotypes and subclasses to evaluate their theoretical functions and examined components of innate immunity that can

interact with Igs. We characterized the B cell receptor (BCR) repertoire of ERBs and defined their B cell diversity mechanism to be somatic hypermutation, which is elevated relative to humans. Finally, we assessed the epitope recognition profile of ERBs to MARV and identified a novel region of recognition that may prevent viral escape which could ultimately protect ERBs from MARV. These unique features of ERB immunoglobulins may contribute to the asymptomatic presentation of MARV and substantially advance the field.

Table of Contents

Abstract	i
Table of Contents	iii
Acknowledgements	vi
Dedication	viii
Disclaimers	ix
List of Figures	x
List of Tables	xi
List of Abbreviations	xii
Chapter 1: Introduction	1
Zoonotic Diseases	1
Bats as viral reservoirs	2
Distinct viruses associated with specific bats	2
The Egyptian rousette bat as a viral reservoir	4
Filoviruses	5
Marburgvirus and significance	8
Marburg virus pathogenesis and immune responses in humans	10
Marburg virus pathogenesis and immune responses in ERBs	11
Antiviral immune response overview	13
Innate	13
Adaptive	14
Resistance and tolerance	15
Hypothesis	17
Overview of techniques	19
Chapter 2: Novel insights into the humoral response of Egyptian rousette bats based on immunoglobulin heavy chain locus	22
Introduction	22
Methods	26
BAC identification and purification	26
PacBio BAC sequencing	27
BAC Assembly	28
Annotation of constant genes	30
Annotation of IGHV, IGHD, and IGHJ genes	33

Gene ontology for function prediction	36
Tissue expression and visualization	36
Positive selection analysis	36
Results	38
Sequencing and assembly of locus	38
Annotation and description of IGHV, IGHD and IGHJ genes	40
Annotation and description of IGHC genes	46
IGHC genes evolutionary relationship	51
Functional characteristics of the ERB IGHC genes.....	51
Tissue expression of immunoglobulin constant genes	54
Gene Ontology suggests IgE and IgG subclasses serve unique functions.....	56
Identification and expression of ERB FcRs	58
Complement proteins and expression levels	61
Identification and expression of pentraxins	61
Positive selection in IGHG, IGHE, and FcRs.....	63
Discussion	65
An expanded repertoire of IGHV known to respond to pathogens	65
Fc functions vary between subclasses while FcRs have conserved motifs.....	73
Summary	74
Chapter 3: Mechanisms of diversity in the V(D)J repertoire of Egyptian rousette bats	76
Introduction	76
Methods	81
Animals	81
Identification of IGHV, IGHD and IGHJ genes.....	82
Somatic hypermutation calculation	82
Results	84
Combinatorial diversity potential and usage in ERBs	84
Usage of V(D)J genes in ERBs	85
CDR3 composition in ERB repertoire.....	91
Somatic hypermutation in ERBs.....	95
Discussion	98
Lower combinatorial diversity in ERBs than humans.....	98

Evidence of higher somatic hypermutation in ERB IgM than humans	99
Summary	100
Chapter 4: Profile of <i>Marburgvirus</i> epitopes ERB antibodies recognize	102
Introduction	102
Methods	104
Results	107
Profile of all MARV Epitopes ERB Antibodies Recognize	107
Profile of all MARV Epitopes NHP Antibodies Recognize	111
Comparison of MARV recognition in NHP and ERB samples	114
Discussion	119
Chapter 5: Conclusion	122
Summary findings	122
Future Directions	123
Call for better characterization of ERBs as a model and review of what is known	123
Effector functions	127
MARV-specific V(D)J response	133
Validation of novel epitopes and endosome entrapment mechanism	136
Concluding hypotheses and approaches	139
Beneficial infection hypothesis	139
Affinity maturation hypothesis	141
Summary	142
References	143

Acknowledgements

I would like to thank my mentor, Dr. Mariano Sanchez-Lockhart. Mariano's guidance was indispensable to my development as a scientist. His passion and persistence fostered an environment in which I could develop hypotheses and test them even in the face of what at times seemed like unescapable obstacles. Mentorship es muy importante.

A special thanks is due to Dr. Ken Bayles, who not only served on my committee but allowed me the privilege of working with him during the completion of my first year courses before transitioning to the Sanchez-Lockhart lab, and continued to be a source of wisdom and inspiration throughout my doctoral work. I also want to thank my other committee members, Dr. Tammy Kielian, Dr. Paul Davis, and Dr. Gustavo Palacios. My interactions were positive and fruitful, and they continually challenged me and demanded intellectual reasoning and rigor. I am grateful and indebted to each one of them for their roles in helping me become not only a competent experimentalist, but also an effective implementer of molecular biology, immunology, and computational analyses.

I would also like to thank the entire Center for Genome Sciences as well as the Bayles lab, both past and present. Their fervor and supportive attitude helped me to persist and exceed. I would like to specifically thank Dr. Peter Larson. He is a high-caliber geneticist capable of tremendous tenacity and ingenuity, of which I was fortunate to have as a mentor, collaborator, and friend. I would further like to thank Dr. Katie Caviness, Dr. Kirsten Kulscar, and Dr. Brett Bietzel, Dr. Joshua Richardson, Major Christine Hulseberg for their insight and advice throughout my graduate career. I would also like to thank my student, Jacob Martens, who revitalized my passion for leading and contributed significantly during his tenure with us. A special thanks is due to Dr. Farooq Nasar who helped guide me to confidence and was a wealth of advice and mentorship. I am deeply

grateful to Dr. Darci Smith for her mentorship and for co-leading a phenomenal mentorship circle with me. Thank you to all the scientists in our community that allowed me to tap into their knowledge and supported me along the way, including Dr. Julie Dyll, Dr. Melek Sunay, Dr. Louise Pitt, Dr. Amy Shurtleff, Dr. Aura Garrison, Dr. Mike Wiley, Dr. Chris Cooper, Jennifer Endres, Karla Garcia, Sean Lovett, and many more. None of my work would have been possible without our incredible collaborators, Dr. Jon Towner and Dr. Amy Schuh for whom I owe a great debt for trusting a graduate student with precious bat samples.

Thank you Maddie and Mollie Bartlett for being there for me to practice speeches, thank you Dad for reminding me to see the sun on the horizon every day, and thank you Mom for always believing in me. Endless thanks are due Austin Sanford, Matt Martens, and Tom Schulze for making me feel like a part of the UNMC community no matter where I was physically. I will forever be grateful to Elyse Abbott for her dedication to outstanding science and immeasurable contribution to every aspect of my graduate journey, both in and out of lab. Thanks to both Elyse and Matt Abbott for the countless hours of joy that helped me push through the hardest days and believing in me when I needed it most.

So many were essential to my success, none more so than Alex Chitty, who perhaps could have written most of this dissertation himself from the innumerable discussions shared. Thank you for being a constant source of support and optimism throughout this tremendous experience. There is an Icelandic saying that translates to “the north winds made the Vikings”, and I believe earning a Ph.D. is a measure of not only surviving the journey, but thriving on the challenges. I am indebted to the breath of warriors who helped me earn mine, and can’t thank any one of them enough.

Thank you.

Dedication

To all the warriors who made me the scientist I am, to Della who I still see in every rainbow, and to Alex who made the journey an adventure.

Disclaimers

This manuscript is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 1757346 and in part by an appointment to the Student Research Participation Program at the United States Army Medical Research Institute of Infectious Disease and administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and USAMRMC. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the U.S. Army.

List of Figures

Figure 1 : Viremia and antibody profiles in ERBs post MARV infection.	18
Figure 2: Raegypt2.0 assembly of IGH locus incomplete.....	21
Figure 3 : Assembly and co-localization of BACs.....	29
Figure 4 : IGHC locus representation	42
Figure 5 : Genome annotation of V(D)J genes.....	43
Figure 6 : Length distribution of IGHV, IGHD, and IGHJ genes.	44
Figure 7 : V/J gene counts from transcriptomic data.....	45
Figure 8 : IGHE gene alignments with reads.....	48
Figure 9 : IGHG gene alignments with reads.	49
Figure 10 : IGHC gene motifs and phylogenies	52
Figure 11: Tissue expression of ERB Igs.....	55
Figure 12 : Predicted functions of ERB Igs	57
Figure 13 : FcR comparison between ERB and other mammals and expression.....	60
Figure 14 : Expression of complement proteins and pentraxins.....	62
Figure 15: Illustration typical V(D)J repertoire protocol.	79
Figure 16 : Framework regions and complementarity determining region map.	80
Figure 17 : SHM calculations schematic.....	83
Figure 18 : V(D)J gene usage across four ERBs for IgM and IgG.....	87
Figure 19 : IGHJ gene alignment of WGXX motif.	88
Figure 20 : Representative V-J usage chord diagram and Heatmap.....	90
Figure 21 : CDR3 amino acid composition of ERB IgM and IgG	92
Figure 22 : CDR3 mutations across VH families.	96
Figure 23 : Percent difference from transcripts to reference VH genes in ERBs (RA) and humans (HS).	97
Figure 24: Domain Programmable Array overview.....	103
Figure 25 : Construction of the DPA phage oligo library.	105
Figure 26 : ERB response to MARV after a single, or double challenge.	108
Figure 27 : ERB responses to MARV after a double challenge.....	109
Figure 28 : ELISA titer for ERBs	110
Figure 29 : MARV DPA results from vaccinated NHPs.	112
Figure 30 : MARV ELISA for vaccinated NHPs.....	113
Figure 31 : Western blot for MARV recognition in ERBs and NHPs.....	115
Figure 32 : Comparison of clusters between ERBs and NHPs.....	117
Figure 33 : Structure of recognition site on surface of MARV GP.....	121
Figure 34 : In vitro ADCC assay.	129
Figure 35 : In vitro CML assay.....	132
Figure 36 : Libra-seq for antigen-specific VDJ-VJ paired sequenced.....	135
Figure 37 : Endosomal entrapment illustration.....	138

List of Tables

Table 1 : Taxonomy of the viral family Filoviridae and associated bat species.....	7
Table 2 : Outbreaks of MVD since identification	9
Table 3 : IGHV genes present on Raegypt2.0	31
Table 4 : IgE and IgG characteristics	35
Table 5 : Protein sequences used to identify immunoglobulins by Mass Spectrometry. .	50
Table 6 : FcRs and associated functions in humans, mice, and ERBs.....	59
Table 7 : Positive selection of IGHG and IGHE residues.....	64
Table 8 : IGHV genes with known activities against viruses.....	68
Table 9 : VH families between species	89
Table 10 : Amino acid composition of CDR3s in ERBs	93
Table 11 : Gap lengths of CDR3 from ERBs.....	94
Table 12 : Epitopes found in literature compared to peptides in ERB serum.....	116
Table 13 : Oligos shared and unique between ERB and NHPs.	118

List of Abbreviations

7-AAD	7-aminoactinomycin D
ADCC	Antibody dependent cellular cytotoxicity
ADCP	Antibody dependent cellular phagocytosis
ADE	Antibody-dependent enhancement
ALT	Alanine aminotransferase
APC	Antigen presenting cell
BAC	Bacterial artificial chromosome
BCR	B cell receptor
BDBV	Bundibugyo virus
BM	Bone marrow
BMDC	Bone marrow derived dendritic cell
bnAbs	Broadly neutralizing antibodies
BOMV	Bombali virus
C3	Complement component 3
C1qA	Complement Component 1, Q Subcomponent, Alpha Polypeptide
C1qB	Complement Component 1, Q Subcomponent, Beta Polypeptide
C1qC	Complement Component 1, Q Subcomponent, Gamma Polypeptide
CCL8	C-C Motif Chemokine Ligand 8
CD	Combinatorial diversity
CDS	Coding sequences
CH	Constant segment, heavy chain
cDNA	Complementary DNA
CDR	Complementarity determining region
CFSE	Carboxyfluorosuccinimide ester
CML	Complement mediated lysis
CoV	Coronavirus
CRP	C-reactive protein
CXCL2	C-X-C Motif Chemokine Ligand 2
DAP10	DNAX-activation protein 10
DAP12	DNAX-activation protein 12

DC	Dendritic cell
DC-SIGN integrin	Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-
DH	Diversity heavy segment
DNA	Deoxyribonucleic acid
DPA	Domain Programmable Array
DRC	Democratic Republic of the Congo
dTNT	Terminal deoxynucleotidy transferase
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
EBOV	Ebola virus
EBV	Epstein Barr virus
ELISA	Enzyme-linked immunosorbent assay
ERB	Egyptian rousette bat
EVD	Ebola virus disease
Fab	Fragment antibody binding
Fc	Fragment crystallizable
Fc α R	Fc alpha receptor
Fc γ R	Fc gamma receptor
Fc ϵ R	Fc epsilon receptor
Fc μ R	Fc mu receptor
FWR	Framework region
GO	Gene ontology
GP	Glycoprotein
HeV	Hendra virus
HIV	Human immunodeficiency virus
HSV	Herpes Simplex virus
HRP	Horse radish peroxidase
HUJV	Huángjiāo virus
IFN	Interferon
IFN α	Interferon alpha
IFN ω	Interferon omega
IFN γ	Interferon gamma

Ig	Immunoglobulin
IgA	Immunoglobulin alpha
IgD	Immunoglobulin delta
IgE	Immunoglobulin epsilon
IgG	Immunoglobulin gamma
IGH	Immunoglobulin heavy
IGHA	Immunoglobulin heavy alpha
IGHE	Immunoglobulin heavy epsilon
IGHG	Immunoglobulin heavy gamma
IGHM	Immunoglobulin heavy mu
IGL	Immunoglobulin light
IGLK	Immunoglobulin light kappa
IPLL	Immunoglobulin light lambda
IgM	Immunoglobulin mu
IgSeq	Immunoglobulin sequencing
IL4	Interleukin 4
IL10	Interleukin 10
IL13	Interleukin 13
ISG	Interferon stimulated gene
ITAM	Immunoreceptor Tyrosine-based Activation Motif
ITIM	Immunoreceptor Tyrosine-based Inhibitory Motif
JAK/STAT	Janus kinase/signal transducers and activators of transcription
JH	Joining heavy segment
JL	Joining light segment
KIR	Killer cell immunoglobulin receptor
KLR	Killer cell lectin-like receptor
L	L gene
LN	Lymph node
LLOV	Lloviu virus
mAb	Monoclonal antibody
MARV	Marburg virus

MBL	Mannose-binding lectin
MCM	Medical countermeasures
MDDCs	Myeloid derived dendritic cells
MERS	Middle East Respiratory Syndrome
MHC	Major histocompatibility complex
MLAV	Měnglà virus
MLD	Mucin-like domain
MPTX	Mucosal pentraxin
MVD	Marburg virus disease
nAb	Neutralizing antibody
NHP	Non-human primate
NiV	Nipah virus
NK	Natural Killer cell
NKG2	NK cell receptor G2
NLR	NOD-like receptor
NP	Nucleoprotein
NPC1	Niemann-Pick C1
NPTX1	Neuronal pentraxin 1
NPTX2	Neuronal pentraxin 2
NPTXR	Neuronal pentraxin receptor
RAVN	Ravn virus
RESV	Reston virus
RNA	Ribonucleic acid
PBMC	Peripheral blood mononuclear cell
PBST	Phosphate buffered saline with tween
PCR	Polymerase chain reaction
PDB	Protein databank
PEDS	Porcine epidemic diarrhea
PKR	Protein kinase R
PTX3	Pentraxin 3
PTX4	Pentraxin 4

RLR	RIG-I-like receptor
SADS	Severe acute diarrhea syndrome
SAP	Serum amyloid P component
SARS	Severe acute respiratory syndrome
SGC	Somatic gene conversion
SHM	Somatic hypermutation
SOSV	Sosuga virus
Ssp	Subspecies
SUDV	Sudan virus
TAFV	Taï Forest virus
TLR	Toll-like receptor
TNF α	Tumor necrosis factor alpha
V(D)J	Recombination of one VH, DH, JH
VH	Variable heavy segment
VL	Variable light segment
VLP	Virus-like particle
VP24	Viral protein 24
VP35	Viral protein 35
VP40	Viral protein 40
VSV	Vesicular stomatitis virus
XILV	Xīlǎng virus

University of Nebraska Medical Center

Omaha, Nebraska

April, 2020

Unique Features of Immunity in Egyptian Rousette Bats

Maggie L. Bartlett, B.S.

University of Nebraska Medical Center, 2020

Advisor: Mariano Sanchez-Lockhart, Ph.D

Chapter 1: Introduction

Zoonotic Diseases

Zoonoses are ailments that are transmitted from animals to humans or vice versa [1]. Zoonotic diseases have a significant effect on public health and studies show that approximately 75% of all emerging infectious diseases in humans originate from animals [1]. These include outbreaks caused by filoviruses, lyssaviruses, alphaviruses, arenaviruses, influenza viruses, flaviviruses, coronaviruses, *Salmonella* spp., *Brucella* spp., *Leptospira* spp., *Trypanosoma* spp., *Borrelia* spp., *Rickettsia* spp., *Mycobacterium* spp. and many more [2]. These can be transmitted by consumption of contaminated food and water or direct or indirect contact with infected animals, insects, or humans. As zoonoses constitute 58-61% of all communicable diseases worldwide understanding host-pathogen interaction in reservoir hosts is paramount to identifying novel mechanisms of control [3]. As we increase our encroachment on habitats of previously undisturbed species, we also increase our risk to potential spillover events. The incidence and prevalence of zoonotic diseases have been on the rise since the 1940s, likely due to increased travel, urbanization, demographic changes, land encroachment, agricultural practices, and lifestyle changes [4]. Studies across the 27 terrestrial mammalian orders for zoonotic potential suggested that zoonoses were correlated to species richness in an order [4]. While zoonotic diseases are found in 21/27 mammalian orders to date, the most species-rich orders have the greatest number of unique pathogens. Rodentia is the most abundant and species-rich among mammals of which 244/2220 species (10.7%) carry 85 known unique zoonotic diseases [5]. The second most diverse order is Chiroptera (bats) and accounts for ~20% of classified mammals [6]. This order includes 108/1100 (9.8%) zoonotic hosts with a total of 27 unique zoonoses described thus far [5].

Bats as viral reservoirs

Within both Chiropteran suborders, Yinpterochiroptera and Yangochiroptera, it is well established bats host more zoonotic viruses per species than other studied mammals with some bats identified as the natural reservoirs of zoonotic pathogens [7-9]. Bats are uniquely positioned to readily spread diseases, as they are the only mammals capable of flight, are highly social, and long-lived [8]. These features mean bats cross continental boundaries more easily than other terrestrial mammals. Further, it is known that bats often co-habitat with other bat species and migrate to and from distant caves, bringing the viruses they harbor with them [10]. Bats also aggregate in groups and pass pathogens back and forth through play biting and exchange of bodily fluids, an ideal environment for a pathogen to spread for which co-housing experiments have confirmed [11]. In addition, bats can live decades compared to rodents who live years. Bat longevity paired with the seasonal dynamics of viral infections within colonies maintains viruses in nature as colony seropositivity is linked to biannual birthing circles [10, 12-14]. As zoonoses account for more than half of all emerging infectious diseases, understanding host-pathogen interactions is paramount to preventing zoonotic events where viruses jump from animals to other animals or to humans. Thus, it is crucial to world health efforts to understand host-pathogen interactions within reservoir hosts and specifically, identify how bats successfully withstand viral infections that are pathogenic in humans.

Distinct viruses associated with specific bats

There is growing evidence to support bats as viral reservoirs, and specific bat species are associated with distinct viruses. Flying foxes, of the Yinpterochiroptera suborder, include 60 known species and are considered the reservoirs for the paramyxoviruses, Nipah virus (NiV) and Hendra virus (HeV) [15, 16]. . Paramyxoviruses are negative-stranded RNA viruses with two surface glycoproteins, an attachment protein,

and a fusion protein. More than 100 paramyxoviruses have been identified in bats and rodents worldwide [17-19]. The *Paramyxoviridae* family includes some of the most significant livestock and human viruses such as mumps-, measles-, distemper-, parainfluenza-, Newcastle disease-, respiratory syncytial virus and metapneumoviruses [20]. A novel paramyxovirus, Sosuga virus (SOSV), was isolated from ERBs after an incidental human case, highlighting the potential for other viral spillovers beyond what is currently known [21].

Chinese horseshoe bats, of the Yangochiroptera suborder, are the reservoir for Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) [22]. The Coronavirus (CoV) responsible for the outbreak that started in Wuhan, China in 2019 is also believed to have originated from a bat strain. The novel CoV, named SARS-CoV2, is the causative agent of Coronavirus Disease 2019 (COVID19) and has spread worldwide affecting several countries including the US [23]. Even though an intermediate host is suspected (currently thought to be the pangolin), the most likely origin based on sequence homology is a bat- underscoring the importance of understanding host-pathogen interactions particularly in bats [24]. To date numerous CoV have emerged that cause severe disease including: (SARS), Middle East respiratory syndrome (MERS), porcine epidemic diarrhea (PED) and severe acute diarrhea syndrome (SADS). Similar to filoviruses and paramyxoviruses, there is no evidence that CoVs cause disease in bats studied [25].

Although there is mounting evidence of distinct viruses associated with only particular species of bats, increased surveillance suggests these pathogens may be more widespread in different bat species than previously anticipated [26]. This is supported by the numerous novel viruses that have been collected from various bat species including members of the viral families: *Astroviridae*, *Bunyaviridae*, *Circoviridae*, *Parvoviridae*, *Partitviridae*, *Coronaviridae*, *Picobirnaviridae*, *Adenoviridae*, *Herpesviridae*,

Papillomaviridae, *Phenuiviridae*, *Picornaviridae*, *Orthomyxoviridae*, and *Filoviridae* [27]. To this end, what is found for one bat species and virus may not be applicable to all bats and all viruses as there are clear associations and host-pathogen interactions that have yet to be explained.

However, a majority of studies on bat antiviral immunity have been conducted on only two species, ERBs and the closely related *Pteropus alecto*. While both are reservoirs to RNA viruses (MARV, RAVV, SOSV and HeV, Broorne viruses, respectively), the viruses as well as the mechanisms to control infection appear to have been achieved through distinct pathways [8]. Most notably through the difference in IFN induction post infection which appears to not only be upregulated post HeV and NiV infection, but constitutively expressed in *P.alecto* [28]. Post MARV infection in ERB cells, very few IFN genes were induced, although mild upregulation of putative unannotated antiviral paralogs were detected, as well as elevated basal expression of key antiviral genes in uninfected cells [29]. These data support better characterization of bat immunity and call attention to the importance of not attributing a finding in one bat to all bats.

The Egyptian rousette bat as a viral reservoir

The Egyptian rousette bat (ERB, *Rousettus aegyptiacus*), a member of the Yinpterochiroptera suborder, is a known reservoir for the filoviruses Marburg virus (MARV), Ravn virus (RAVV) and the paramyxovirus, SOSV [21, 30]. Interestingly, ERBs display no disease manifestation upon infection with MARV, RAVV or SOSV. However, humans and non-human primates (NHPs) present with a severe hemorrhagic fever and lethality up to 90% to MARV and RAVV [31], and severe disease to SOSV [21]. Understanding how this reservoir host controls these viruses could advance therapeutics for these viruses for which no FDA treatments currently exist. As SOSV was recently identified and only has one known human case, my work focused on filoviruses and ERBs.

Filoviruses

Filoviruses are non-segmented, negative-strand, RNA viruses from the family of *Filoviridae* which includes *Ebolaviruses*, *Marburgviruses*, and *Cuevaviruses*. *Ebolaviruses* (Zaire virus, Sudan virus, Bundibugyo virus, Taï forest virus, Bombali virus, and Reston virus) differ by 2.7-5.2% at a nucleotide level while *Marburgviruses* (Ravn virus (RAVV) and Marburg virus (MARV)) differ by 20% at a nucleotide level [32]. Further, some species are represented by a single lineage, and new genres are being identified with some frequency [33-36]. As increased surveillance and genome sequencing occurs, more are likely to be discovered like the recent discovery of the Dianiovirus family member, Měnglà virus (MLAV) [37]. To date, only four *Ebolaviruses* (Zaire virus, Sudan virus, Bundibugyo virus, Taï forest virus) and the two *Marburgviruses* are known to cause disease in humans. *Reston ebolavirus* is known to cause disease in NHPs and pigs, but not in humans. The field recently proposed a new taxonomy to integrate the novel filoviruses recently discovered that have no known human cases (Table 1) [33].

With an average genome of 19kb, all filoviruses to date have encoded seven proteins: nucleoprotein (NP), glycoprotein (GP), RNA-dependent RNA polymerase (L), viral proteins (VP) 24, 30, 35, and 40 (VP24, VP30, VP35, and VP40) [38]. MARV was the first filovirus discovered after laboratory workers in Marburg, Germany were exposed to the virus after contact with infected cells from African green monkeys (*Chlorocebus aethiops*) [39]. Small outbreaks of MARV and RAVV occurred between 1975-1997 with links to bat caves [39], and the second largest scale outbreak is thought to have originated from a gold mining cave [40]. The following three outbreaks and two single human cases in Uganda were associated with Kitaka and Python cave, respectively [40]. To confirm these associations, samples were taken from bats in both caves and not only were both populations seropositive for virus, fully infectious virus was isolated [30]. Infectious MARV

has now been isolated from 21 bats which represent the only successful filovirus isolations from any bat species to date [30, 41]. Follow-up experimental inoculation studies confirmed the susceptibility of ERBs to MARV/RAVV including viremia, widespread dissemination in tissues, viral shedding in urine, feces, and saliva [42, 43]. In addition, bat-to-bat transmission has been shown experimentally as well as long-term immunity to reinfection after challenge [11].

Table 1 : Taxonomy of the viral family *Filoviridae* and associated bat species

Genus	Species	Virus	Abbreviation	Bat associated
<i>Cuevavirus</i>	<i>Lloviu cuevavirus</i>	Lloviu virus	LLOV	<i>Miniopterus schreibersii</i> ²
<i>Dianloviurs</i>	<i>Mengla dianlovirus</i>	Měnglà virus	MLAV	
<i>Ebolavirus</i>	<i>Bombali ebolavirus</i>	Bombali virus	BOMV	
	<i>Bundibugyo ebolavirus</i>	Bundibugyo virus	BDBV	
	<i>Reston ebolavirus</i>	Reston virus	RESTV	<i>Cynopterus sphinx</i> , <i>Hipposideros pomona</i> , <i>Miniopterus schreibersii</i> , <i>Myotis pilosus</i> , <i>Pipistrellus</i> , <i>Rousettus</i> <i>amplexicaudatus</i> , <i>Rousettus leschenaultii</i>
	<i>Sudan ebolavirus</i>	Sudan virus	SUDV	
	<i>Tai Forest ebolavirus</i>	Taī Forest virus	TAFV	
	<i>Zaire ebolavirus</i>	Ebola virus	EBOV	<i>Eidolon helvum</i> , <i>Epomops franqueti</i> ² , <i>Epomophorus</i> <i>gambianus</i> , <i>Hypsignathus</i> <i>monstrosus</i> ² , <i>Micropteropus pusillus</i> , <i>Myonycteris torquata</i> ² , <i>Rousettus aegyptiacus</i> , <i>Rousettus leschenaultii</i>
<i>Marburgvirus</i>	<i>Marburg marburgvirus</i>	Marburg virus	MARV	<i>Rousettus aegyptiacus</i> ¹ , <i>Epomops franqueti</i> , <i>Hypsignathus</i> <i>monstrosus</i> , <i>Miniopterus</i> <i>inflatus</i> ² , <i>Rhinolophus</i> <i>eloquens</i> ²
		Ravn virus	RAVV	<i>Rousettus aegyptiacus</i> ¹
<i>Striavirus</i>	<i>Xilang striavirus</i>	Xīlǎng virus	XILV	
<i>Thamnovirus</i>	<i>Huangjiao thamnovirus</i>	Huángjiāo virus	HUJV	

Colored viruses associated with human diseases. All bats were positive by antibody titer, reservoir species indicated¹ and bats with additional PCR confirmation². [33-35, 37].

Marburgvirus and significance

The Ebola virus disease (EVD) outbreak in Western Africa (2013-2016) that caused more than 11,323 deaths accelerated the urgency and emphasized the significant need for effective filovirus treatments [44]. In fact, one EVD outbreak in the Democratic Republic of the Congo (DRC) is ongoing and as of January 29th of 2020 had 66% fatalities and 3,304 confirmed cases [45, 46]. While Marburg virus disease (MVD) outbreaks have not reached the same proportions of the EVD outbreaks, the 1998-2000 DRC MVD outbreak had a case-fatality rate of 83% (128 deaths), and the 2004-2005 Angola of 90% (227 deaths), underscoring the potential risk MARV poses to public health [39] (Table 2). Further, a strain similar to Angola MARV was recently identified as circulating in ERBs in Sierra Leone posing the potential for new spillover events [47]. Although vaccines, monoclonal antibodies (mAbs), and antiviral drugs are under development for filoviruses [48-54], there are no FDA approved medical countermeasures (MCMs) for MARV. However, similar to EBOV many different approaches are being pursued, although the work is significantly behind for MARV. Both preventative and post-exposure treatments have been explored for MARV including lipid nanoparticle delivery of anti-MARV NP-targeting small interfering RNA (siRNA), vaccination with recombinant vesicular stomatitis virus-GP, mAbs, phosphorodiamidate morpholino oligomers and broad-spectrum nucleoside analogs [55-58]. However, these are only shown to be effective in NHPs if treatment is initiated within 48-72 hours of exposure and have not been tested against more than one strain of MARV in most cases. As early diagnostics are still limited for MARV this timeline is not feasible in an outbreak scenario [59]. This emphasizes the significant need for early detection as well as novel approaches to treatment to minimize and control future outbreaks.

Table 2 : Outbreaks of MVD since identification

Year(s)	Country	Cases	Deaths (%)
1967	Germany/Yugoslavia	31	7 (23%)
1975	South Africa	3	1 (33%)
1980	Kenya	2	1 (50%)
1987	Kenya	1	1 (100%)
1990	Russia	1	1 (100%)
1998-2000	DRC	154	128 (83%)
2004-2005	Angola	252	227 (90%)
2007	Uganda	4	1 (25%)
2008	USA*	1	0 (0)
2008	Netherlands*	1	1 (100%)
2012	Uganda	15	4 (27%)
2014	Uganda	1	1 (100%)
2017	Uganda	3	3 (100%)
* Case exported from Uganda			

Outbreaks of MVD. Adapted from CDC report retrieved from:
<https://www.cdc.gov/vhf/marburg/resources/outbreak-table.html>, [60-63].

Marburg virus pathogenesis and immune responses in humans

In humans, initial exposure to MARV is thought to occur through a mucosal route from contact with infected bodily fluids. Virions infect macrophages, monocytes, and dendritic cells via micropinocytosis and are transported to secondary lymph tissues. MARV replication in macrophages and monocytes has been shown to occur as early as 24-48 hours in guinea pigs and cynomolgus macaques, respectively [39]. The virally infected cells travel to secondary sites including the lymph nodes, spleen and liver where most severe necrotic lesions occur [39]. Although lymphocytes have been shown to not be susceptible to MARV infection, massive bystander apoptosis of lymphocytes is a MVD hallmark and impedes the patient's ability to overcome disease [64-66]. The mechanisms for this are not fully delineated but it is thought that cytokine secretion from MARV-infected cells are a major factor including TNF α [67], supported by improved survival in guinea pigs treated with anti- TNF α antibodies [68]. While MARV-infected human sera is not readily available to assess, levels of cytokines and chemokines following infection in NHPs, guinea pigs, and mice have been well documented [64, 69-71]. In late infection stages MARV particles from nearly every organ of NHPs can be isolated, however despite this widespread viral dissemination only minor inflammation is observed in infected tissues and organs [64, 66, 72]. Unfortunately, these data are not consistent between models so drawing correlates to what might be seen in humans remains unclear [39]. Instead I will focus on the findings from human outbreaks and human primary culture in response to MARV. Clinical symptoms appear in three phases. The first phase of symptoms (malaise, headache, fever) appear three to nine days post exposure and are not indicative of a particular disease [73]. Symptoms worsen during the second phase of infection including diarrhea, nausea, vomiting, high fever, and a maculopapular rash presents typically within a week. Liver enzymes are increased during this period and lymphopenia and

thrombocytopenia are typically observed. The third phase of infection consists of severe hemorrhage and organ failure or convalescence. At autopsy, large necrotic areas were present in almost every organ except the lungs, skeletal muscles and skeleton with very large necrotic areas in the testicles and ovaries [39, 74]. Human monocytes and macrophages produce pro-inflammatory cytokines interleukin-1 β (IL-1 β), IL6, IL8 after infection which are thought to subsequently lead to vascular permeability and coagulation abnormalities [67, 75]. Moreover, MARV proteins VP24 and VP35 inhibit type I interferon (IFN) which allows for uncontrolled viral replication [76]. Hemorrhage and death appear to occur as a complication of the immune response rather than the virus infection itself [62, 77]. Patients who do not succumb are shown to generate a CD4+ T cell response with limited CD8+ T cells [78]. The humoral response has a role in protection which is best demonstrated by the presence of protective antibodies that can be transferred to naïve or sick individuals (humans or NHPs) resulting in prevention of diseases or better survival rates [79-82]. These findings support the notion that the humoral response is an effective MARV defense. However, it is well known that antibody titers alone do not correlate with survival outcome [83]. This observation suggests that antibody function, not just quantity, is a critical correlate of protection.

Marburg virus pathogenesis and immune responses in ERBs

For ERBs, it is thought infection occurs through either the mucosal route or through play-biting [11, 84]. It is proposed ERBs control MARV through a differentially regulated innate response that lacks the inflammation seen in humans, coupled with a protective adaptive response that resolves infection [85]. ERBs do not upregulate IFN or IFN stimulated genes (ISGs) post MARV infection *in vitro* [29] or *in vivo* (manuscript submitted, Nature Communications) despite an expansion of these genes. These findings emphasize the importance of combining transcriptomic, genomic, and functional studies as some

expansions in the genome that were thought to likely be induced by infection were shown to later be irrelevant to MARV infection such as the expansion of IFN genes [29, 85].

A recent study sought to define the clinical and histopathologic effects in bats post MARV challenge as little was known compared to humans. No animals became febrile or lost weight while white blood cell counts increased in infected ERBs [86]. A blood chemistry panel revealed mock and infected individuals only differed in the liver enzyme alanine aminotransferase (ALT), however, the increase was mild compared to the elevated amounts seen in humans post infection. No significant gross lesions were found, although some subjects had liver lesions between days 6-8 post infection which resolved and coincided with the ability to detect MARV antigen in blood and tissues examined [86].

Not only are ERBs known to produce antibodies to MARV, but maternal antibodies provide passive immunity to juvenile bats [87]. Further, infected ERBs shed virus for three weeks post infection and can transmit the virus for up to four months after initial infection [87]. Viral maintenance in ERB colonies is suspected to occur through a decrease in antibody levels over time which leads to an increase in viral load from residual virus, resulting in shedding and subsequent infection of other ERBs [11, 87, 88]. The importance of the Ig response in control of MARV infection in ERBs is underscored by the decrease in viremia that corresponds with an increase in antibody titer (Figure 1) [11]. Despite knowing Igs appear to be important post MARV challenge for subsequent exposures, nothing is known about how ERBs form their B cell receptor (BCR) repertoire and what that repertoire recognizes.

To better understand the humoral responses in ERBs, we sought to define their germline Ig genes as well as associated molecules like FcR, complement, and pentraxins. Previous work had highlighted the significant differences at a genome level between other facets of the immune response in ERBs [85]. Further, these data drew attention to the

potential of ERBs to rapidly create protective antibodies, or have a population of natural IgM capable of viral control while humans require appropriate supportive care to survive long enough to produce protective antibodies. Thus we also sought to define the BCR repertoire and antigen recognition of ERBs.

Antiviral immune response overview

Innate

The immune system is comprised of innate and adaptive immunity [89]. The innate immune system recognizes viral infections through pattern-recognition receptors (PRRs) capable of detecting genomic DNA, single stranded or double stranded RNA, and RNA with a 5'-triphosphate [90]. Three classes have been shown to be associated with virus-specific recognition: Toll-like receptors (TLRs), retinoic acid-inducible gene I (RIG-I)-like receptors (RLRs), and NOD-like receptors (NLRs) [91]. These receptors are present in various cell types from both the innate and the adaptive arms of the immune system, including epithelial cells, NK cells, myeloid lineage cells, DCs, B cells, and T cells [91]. Activation of these receptors leads to induction of type I IFNs that trigger the Janus kinase/signal transducers and activators of transcription (JAK/STAT) signaling cascade which upregulates hundreds of genes to ultimately render cells resistant to infection or better able to block viral replication [92, 93]. In human cells infected with MARV, VP35 inhibits IFN production at numerous steps of the signaling pathway downstream of RLRs [94]. In addition to suppression of type I IFNs, VP35 inhibits the IFN-induced protein kinase R (PKR) pathway in humans, but not in ERB cells [95].

Another mediator of the innate antiviral response are pentraxins, which can stimulate the complement pathway resulting in virus neutralization [96, 97]. The complement system is a significant component of innate immunity and is comprised of soluble factors and cell surface receptors that sense and respond to invading pathogens

[98]. The complement system is thought to link innate and adaptive responses by a myriad of mechanisms including enhancement of humoral immunity, regulation of Ig effector mechanisms, and modulation of T cell function [99]. There are three activation pathways (classical, alternative, and lectin) that converge on complement component 3 (C3). Numerous viruses have evolved mechanisms to evade complement, supporting its protective qualities. Viruses can be directly inactivated by complement after being bound by mannose binding lectin (MBL), which has been shown effective in HIV, EBOV, MARV, and SARS-CoV [100].

Adaptive

During this primary innate immune response, antigen presenting cells (APCs) such as DCs and macrophages recognize pathogens with PRRs and present antigen to T cells to produce cytokines which promote B cell maturation and antibody production. These B cells then can undergo somatic hypermutation (SHM) and clonal selection to become long-lived plasma cells to produce antigen-specific immunoglobulins (Igs) in the secondary immune response. Igs are heterodimeric proteins comprised of two heavy (IGH) and two light (IGL) chains. They have a fragment antigen-binding (Fab) region and fragment crystallizable (Fc) region that confers effector function [101]. The Fab domain is further composed of a Variable (VH) gene, Diversity (DH) gene, and Joining (JH) gene from the heavy chain and a Variable (VL) gene and Joining (JL) gene from the light chain. The breadth and specificity of the B cell response is defined initially by the germline composition of these genes. Further mechanisms of diversity occur, including SHM, through affinity maturation to produce high-affinity Igs [102]. Five isotypes exist of Igs in most mammals (IgM, IgG, IgA, IgE, and IgD) that differ in the Fc region of the molecule. IgM/IgD are associated with development and are the first expressed on B cells as they leave the bone marrow. Plasmablasts that only express IgD are linked to respiratory

commensals and pathogens, as well as contradictory reports claiming roles in tolerogenic pathways or protection from tolerance [103, 104]. IgA is the main isotype found in mucosal surfaces including gut, respiratory tract and urogenital tract [105], while IgG is the major serum isotype and can cross the placental barrier to provide fetal immunity [106]. Antigen-specific Igs elicit protection through activation of the complement system, pathogen neutralization, phagocytosis of apoptotic cells, induced cytotoxicity, and priming of immune cells [107]. Natural Igs have gained increased attention with protective roles shown in viral, bacterial, fungal, and parasitic infections [108]. However, these Igs are considered limited in potential breadth by the low amount of SHM detected [109]. Natural Igs have been implicated as potent inducers of complement [110], and multiple studies have shown effective viral control can be achieved through natural Igs and complement [111-113]. In addition to their role of binding antigen, Igs regulate the immune response by interaction with Fc receptors (FcRs) [107] which is discussed in greater detail in Chapter 2. T cells are also known to play a role in the antiviral immune response by killing virus-infected cells [114] as well as helping promote B cell responses [115].

Resistance and tolerance

Resistance and tolerance are often described as two distinct defense strategies to infectious diseases [116]. While resistance is thought to be dependent on the ability of a host to kill an invading pathogen, tolerance is recognized for its role in protecting hosts by promoting host health. It should be noted that disease tolerance is not the same as immunological tolerance in which self-reactive T cells are eliminated, although in certain diseases immune tolerance can be considered a mechanism of disease tolerance [116].

Mechanisms that fall under disease tolerance include host encoded processes that support the resolution of infection or prevent the onset of immunopathology in host tissues that are an inevitable consequence of common inflammatory stimuli [116]. Our studies as

well as others of Chiropteran genomes revealed numerous gene expansions and theoretical features based on sequences of genes that suggest tolerance in bats to viral infection [85, 117]. In ERBs, we had observed an expansion of IFN omega (IFN ω) genes, natural killer (NK) cell receptor genes with both inhibitory and activating domains, and a large number of major histocompatibility complex (MHC) genes encoded outside the canonical locus [85]. While the expansions of IFN ω may aid resistance mechanisms, as well as other uncharacterized antiviral genes in ERBs, the expansions in NK receptors and MHC are hypothesized to aid in tolerance by requiring higher activation thus producing less inflammation [29, 85, 118-120]. These findings include the finding that the main NK cell receptor in ERBs is NKG2/CD94, which is associated with inhibitory responses in other species [121, 122]. The expansion of CD94 genes suggests the possible combinatorial diversity of heterodimeric receptors is large, and since signal transmission occurs via a conserved residues, additional receptor diversity would likely be associated with additional capacity to bind ligands or interact differently to the same ligands [85]. This greater ligand binding diversity could provide better MHC class I recognition, possibly distinguishing between a pathogen's mimics as well as ability to recognize self [123]. This study also found that the NK cell receptors possessed inhibitory and activating domains which is unusual in mammals [85]. This paired with an arginine residue that preferentially recruits DAP10 rather than DAP12 could suggest the activating receptors are less potent cytokine inducers, as DAP12 is more efficient in cytokine production induction than DAP10 [85, 124].

These data support a disease-tolerance model in ERBs. We sought to explore this further by defining additional features of ERB immunity that could support or refute disease tolerance. Our findings add to the growing body of evidence that ERBs have encoded processes that likely contribute to the lack of immunopathology observed, as well as

highlight new features never before described in a species that may add in a myriad of pathogen responses.

Hypothesis

Multiple immune-related gene families that have undergone expansion in ERBs compared to humans have been identified including NK cell receptors, MHC class I genes, and IFN genes [85]. These findings suggested ERBs may have a tolerance for viral infections resulting in a less-inflammatory response. We aimed to shift the focus to the adaptive humoral immune response as most of the effort to dissect the antiviral mechanism(s) was previously focused on innate immunity [29, 117, 125], in part because of a lack of reagents and protocols. Here we sought to examine three facets of the ERB Ig response: 1) characterization of ERB loci associated with the humoral immune response, 2) definition of the B cell receptor repertoire for gene usage and diversity mechanisms, and 3) mechanistic identification of antigen recognition of ERB antibodies to MARV. Differences at the germline for Ig genes could prime the response toward recognition of specific pathogens, while diversity mechanisms and gene usage define antigen recognition breadth and capacity, and epitope recognition could support recognition of novel regions not detected in other species. Together we anticipated these findings would support or refute our previous hypothesis that ERBs have evolved tolerogenic mechanisms that contribute to an asymptomatic presentation through a less inflammatory response.

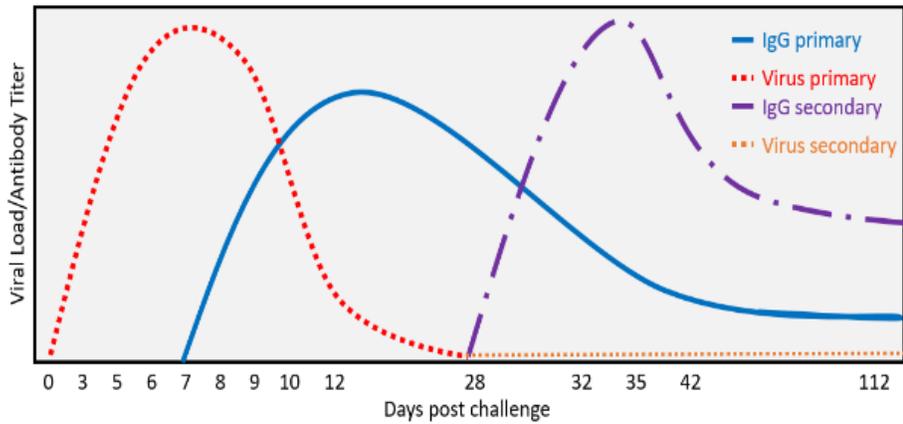


Figure 1 : Viremia and antibody profiles in ERBs post MARV infection.

Illustration of viral load and antibody titer over 112 days post MARV infection. Viremia decrease coincides with peak antibody titer and virus is not detectable upon secondary exposure, while a robust IgG titer is observed.

Overview of techniques

In order to determine the V(D)J usage, B cell diversity mechanism(s), potential diversity, and the repertoire to specific pathogens like MARV, we first needed to reconstruct the IGH locus which contains the germline genes. The previous assembly, Raegypt2.0, despite being high quality had IGH genes on 21 different scaffolds and potentially five IgE genes as well as a novel pattern of D-J-D-J genes (Figure 2). To ensure there were not due to assembly errors from the high repetition in this region, sequencing reads that span these novel findings was critical. We used a combination of Bacterial artificial chromosomes [126] and Bionano optical mapping [127] to assemble the locus which is described in Chapter 2. This locus revealed unique features including an increase in copy number of IGHV family members associated with viral protection in humans, a distinctive organization of the IGHD and IGHJ genes within the locus, an expansion of *IGHE* genes not before reported in any other species, and distinct *in silico* putative functions for the IgE and IgG subclasses. We confirmed these findings with transcriptomics and proteomics, defined expression and location of each Ig across multiple tissues in ERBs. Additionally, we also characterized the FcR genes present and theoretical functionality in the ERB genome as well as other immune mediators such as complement proteins and pentraxins.

With an accurate and annotated IGH locus we were able to characterize Ig transcripts and study the V(D)J rearrangement process in ERBs which was previously unknown and requires reference genes to compare to. V(D)J recombination, gene usages and somatic hypermutation rate observed in these rearrangement are described in Chapter 2. B cells were sorted from peripheral blood mononuclear cells (PBMCs), then RNA was extracted, cDNA generated, ERB Ig-specific primers used to enrich for isotype-specific transcripts, then libraries prepared and quantitated before sequencing and

analysis. Analysis required filtering of reads, consensus of mate-pairs, and annotation using our newly defined germline reference set. We found that the ERB has higher somatic hypermutation in IgM than humans, a larger range of V-D and D-J gaps, and overall less combinatorial diversity than humans. We observed preferential usage for certain IGHV, IGHD and IGHJ genes.

The counterpart to the V(D)J region of an antibody is the epitope recognized. Thus the region of MARV ERBs recognize was of interest. In Chapter 3 I discuss the use of a phage display technology known as Domain Programmable Array (DPA) to quantitatively characterize the epitope recognition profile of ERB Igs for MARV [128]. Previously developed and tested on NHP sera for EBOV [128], I modified and executed this technique for multiple bat species including the Schreiber's Bent-Winged Insectivorous bat [36] and ERBs (In preparation). We found that regions from MARV GP that ERBs is different than in vaccinated NHPs. More importantly, these unique epitopes coincide with the GP1/GP2 cleavage site, which may contribute to ERBs ability to control MARV.

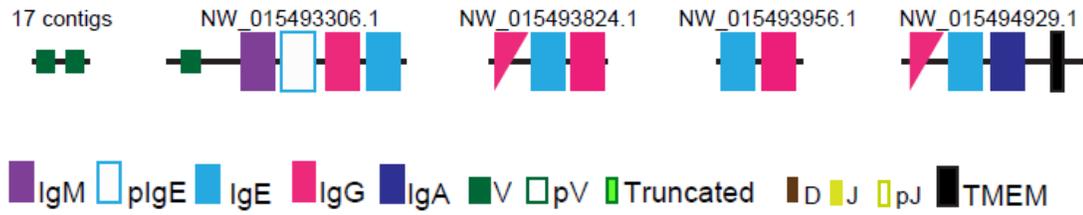


Figure 2: Raegypt2.0 assembly of IGH locus incomplete.

Original assembly placed IGH genes on 21 different scaffolds. Due to the high repetition, these were unable to be assembled using current short-read assemblers.

**Unpublished. Submitted to Cell.*

Chapter 2: Novel insights into the humoral response of Egyptian rousette bats based on immunoglobulin heavy chain locus

Introduction

Bats (Chiroptera) have been shown to host more zoonotic viruses per species than any other mammal studied and several bat species were identified as the natural reservoirs of dangerous zoonotic human pathogens [7-9]. Thus, understanding how bats successfully withstand viral infections that are highly pathogenic and/or lethal in humans has become a significant area of research. Recently, we and others has sought to address this question by examining the genomes of relevant reservoir bat species to uncover factors that could contribute to natural resistance to viral infections [29, 85, 129]. In some bat species, like *Pteropus alecto*, a potent innate response was proposed as one successful antiviral mechanisms [130, 131], while in other bat species, like *Rousettus aegyptiacus*, tolerance was proposed as the strategy to control viral infection with minimal inflammation [29, 85]. Moreover, several groups have made important observations regarding the varied antiviral mechanism(s) different bat species employ, mainly on the innate immunity space (reviewed in [117]).

A member of the Yinpterochiroptera suborder, the Egyptian rousette bat is a known reservoir for MARV, RAVV, and SOSV [21, 30]. While ERBs display no overt disease manifestation upon infection with MARV, humans and several non-human primate (NHP) species present with a severe hemorrhagic fever and high lethality [31]. We previously reported multiple immune-related gene families that have undergone expansion in ERBs compared to humans including natural killer cell receptors, MHC class I genes, and interferon genes [85]. The evaluation of the theoretical function of these expanded genes suggests that a tolerogenic immune state may exist in ERBs, which allows these animals

to asymptotically control viral infections. Here, we expand our previous work by examining the ERB genome loci associated with the humoral immune response and investigate whether their putative functionality may also contribute to a tolerogenic immune state.

Little is known about the generation of specific antiviral immunoglobulins (Igs) in ERBs. However, when exposed to MARV, ERBs mount a humoral response that can promote protection [88, 132] and appears not to be mediated by virus neutralization [133]. These observations suggest that other humoral immune mechanisms of viral control for MARV could be favored in ERBs. Igs can interact with other immune mediators like Fc receptors (FcRs) and/or complement to mediate viral clearance and opsonization. The ability of the Igs to interact efficiently with those mediators is dictated by the Ig isotype/subclass. Most mammalian Igs arise from the tetramerization of two identical Ig heavy chains (IGH) with two identical Ig light chains (IGL) [134]. The IGH locus comprises numerous variable (IGHV), diversity (IGHD), and joining (IGHJ) genes, as well as constant (IGHC) genes, that recombine to form the IGH chain of an Ig. The genetic diversity and copy number of IGHV, IGHD, and IGHJ genes contributes to antigen recognition breadth and varies widely between species [135]. The V(D)J genes of the IGH and the associated IGL comprise the fragment antigen binding (Fab) region and is responsible for epitope recognition. The effector function of the Ig is mediated by the fragment crystalline (Fc) region encoded by IGHC genes. In mammals, IGHC genes are represented by five isotypes; IgD, IgM, IgG, IgA, and IgE, however, the number and type of isotypes/subclasses are not constant across species [136]. For example, humans and mice have one functional copy of IgM, IgD and IgE and four IgGs. In contrast, mice have a single IgA while humans have two. Moreover, despite both species containing the same number of IgG genes, mouse and human IgG subclasses are not functionally equivalent

[137]. The European rabbit is a striking example of the Ig diversification having one copy of IgM, IgG and IgE and 13 IgA genes [138]. Within Chiroptera there is transcriptomic evidence of IgM, IgA, IgE and IgG expression, and IgD is only found in Yangochiroptera but not Yinpterochiroptera [139-142]. IgG subclass number can range from only one copy detected (*Carollia perspicillata*) to up to five (*Myotis lucifugus*) [141]. However, given the complexity and repetitiveness of the IGH locus, no accurate description, gene organization, or annotation has been completed for any bat species.

The Ig Fc region functionally links the humoral adaptive immunity mediated by B cells with several innate components of immunity like natural killer (NK) cells, monocytes, macrophages, granulocytes, as well as with complement (reviewed in [101]. Downstream immune processes are primarily mediated by protein-protein interactions of the Ig Fc region with FcRs or complement [136, 143]. FcRs modulate the immune response by delivering activation or inhibition signals to effector cells [136, 144]. FcRs have characteristic expression patterns among the various leukocyte populations and can bind specific Ig isotype(s) with varied affinities depending on the amino acid sequence and N-glycosylation status of the Fc region [145]. The complex evolutionary relationship between FcRs and Ig isotypes/subclasses has resulted in varied FcRs presence/absence among different vertebrate lineages. Tetraploidization and chromosome doubling has been the driving force for Ig-domain containing protein expansion and evolution, including FcRs [146]. Regional duplications were also involved in shaping the FcR repertoire, like the duplication of the low affinity IgG receptor resulting in FcγRIIC and FcγRIIIB in humans. Another example of the plasticity of the FcR repertoire in different animals is the high affinity Fc receptor for IgA (FcαRI) found in several placental mammals but not in mouse [147].

Complement proteins participate in important Ig effector functions, bridging innate

and adaptive immune systems (reviewed in [98]). C1q interaction with Fc triggers a cascade of enzymatic reactions resulting in an activation of processes leading to opsonization, inflammation, or cell lysis. Pentraxins are a key component of humoral innate immunity as they can recognize microbial moieties to mediate opsonophagocytosis and inflammation [148]. Pentraxins are well conserved in diverse species ranging from arachnids to insects and mammals, and evolutionarily predate the establishment of the adaptive humoral immune system. Recently, pentraxins were also shown to activate complement by the classical pathway and interact with FcRs expressed on neutrophils and monocytes to mediate phagocytosis [149].

Innate immune responses are the first barrier of defense when a microorganism enters the body until the adaptive immune responses establishes long-term protection. To understand how bats are asymptomatic to viral pathogens that cause severe disease in humans, a complete description of the bat's immune system is needed, including the innate, adaptive and interphase between both immune systems. To characterize the humoral immune components in ERBs, we used a combination of molecular biology and sequencing technologies that allowed us to describe and annotate one contiguous ERB IGH locus. Completion of the locus revealed unique features including an increase in copy number of IGHV family members known to act in viral protection in humans; an expansion of IgE genes with different tissue expression pattern and theoretical functions; and distinctive *in silico* putative functions and structural characteristics of the four IgGs found. The description of the FcR gene repertoire and its theoretical characteristics allowed us to observe that ERB lack the expansion of FcR for IgG (FcγR) observed in primates; thus, lacking FcRs with ITAM motif embedded in the cytosolic tails. Moreover, the high affinity ERB FcγR (FcγRI) present a structure liked to decrease affinity for IgGs. Finally, we observed the complete absence of functional short pentraxins, proteins that participate in

the acute phase of resistance against pathogens. Together, our data reinforce the idea that ERBs might establish a tolerogenic state to surmount pathogenic infections. Our observations highlight the novelties and uniqueness of the Chiropteran genomes [150, 151] and provide a strong argument for the continued characterization of relevant reservoir species.

Methods

BAC identification and purification

Bacterial artificial chromosome (BAC) libraries were created from ERB liver tissue (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 2). 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). BAClg4 was sequenced as described except we used a SMRT Cell 1M v3 LR Tray and sequenced with 1200 minute movies. Upon completion of sequencing fastq files from each individual BAC were downloaded from SMRT Link and assembled using the long read assembler Canu [152] and an in house script for BAClg1, BAClg2, BAClg3. For BAClg4 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. 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. Resequencing also indicates the average read depth of each assembly. BAClg1 was sequenced with an average coverage of 524x and a concordance of 99.99%, BAClg2 at 501x average coverage and a concordance of 99.98%, BAClg3 at 485x average coverage and a concordance of 99.98%, and BAClg4 at 370x average coverage and a concordance of 99.99%.

BAC Assembly

Upon completion of sequencing fastq files from each individual BAC were downloaded from SMRT Link and assembled using the long read assembler Canu [152] for BACs BAClg₁, BAClg₂, and BAClg₃. For BAClg₄ 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. BACs were polished using reads that were used for their original assembly. Completed BAC sequences were uploaded to Geneious. BAClg₁₋₃ were assembled using the 'De Novo Assembly' function. BAClg_{1/4} were assembled using the 'Map to Reference' function using BAClg₁ as a reference sequence. The final assembly is represented as complete sequences from BAClg₃, BAClg₂, and a partial sequence of BAClg₁ (Figure 3). The near perfect alignment amongst these three BACs likely represents a single haplotype. For the overlap sequence between BAClg₁ and BAClg₄ all BAClg₁ sequence (Figure 3, dotted line) was replaced with BAClg₄ 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.

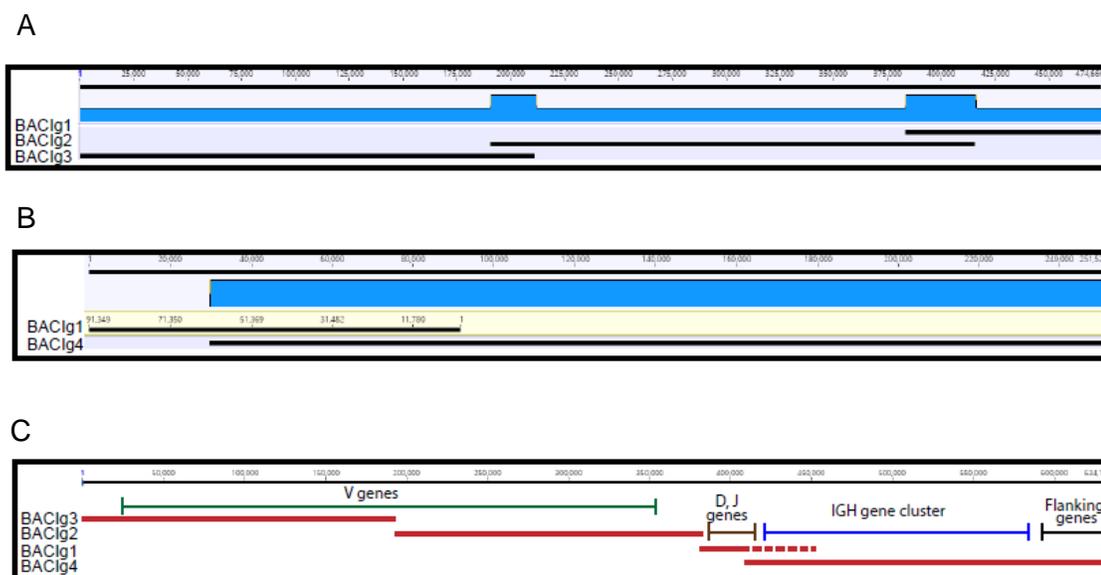


Figure 3 : Assembly and co-localization of BACs.

(A) Geneious 'De Novo Assembly' assembly of BAC1g1-3. (B) Geneious 'Map To Reference' assembly of BAC1g1 set as reference sequence and BAC1g4. (C) Organization of BACs used in the final sequence assembly and relative location of V, D, J, IGH, and flanking genes in the final assembly. The dotted line indicates the overlap sequence between BAC1g1 and BAC1g4. Only sequence from BAC1g4 was used in the final assembly.

**Unpublished. Submitted to Cell.*

Annotation of constant genes

All annotated IGH constant genes were downloaded from Raegyp2.0 using the NCBI genome browser (Table 3). 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 (Table 3); 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).

Table 3 : IGHV genes present on Raegypt2.0

V gene Annotation Raegyp2.0	
Scaffold	NCBI
NW_015493306.1	1
NW_015493547.1	6
NW_015493575.1	5
NW_015493590.1	8
NW_015493682.1	5
NW_015493749.1	5
NW_015493769.1	10
NW_015493770.1	2
NW_015493822.1	5
NW_015493830.1	2
NW_015493970.1	2
NW_015494112.1	2
NW_015494214.1	1
NW_015494338.1	1
NW_015494352.1	1
NW_015494409.1	1
NW_015494440.1	1
	58

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 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 RaegyplGH3.0 would yield accurate results. Sequences (Table 3) were uploaded to Geneious and aligned to RaegyplGH3.0 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 RaegyplGH3.0 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. plgE1, 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 RaegyplGH3.0 would yield accurate results. Sequences (Table 3) were uploaded to Geneious and aligned to RaegyplGH3.0 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 RaegyplGH3.0 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.

IgM/IgA annotations

NCBI annotated a single IgA and *IGHM* in Raegypt2.0. Both sequences contained missannotated exons upstream of the CH1 exon, however both CDS contained complete open reading frames with no premature stop codons. Sequences (Table 3) were uploaded to Geneious and aligned to RaegyplGH3.0 using the 'Annotate from function' with a mapping requirement of 75% identity. Annotated exons for each IgE gene were inspected in the context of RaegyplGH3.0 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 IgA contained complete open reading frames.

Annotation of IGHV, IGHD, and IGHJ genes

Raegypt2.0 scaffolds and the final RaegyplGH3.0 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.

Table 4 : IgE and IgG characteristics

	Tissue expression	Transcriptomic expression	Proteomic peptide	# of C in the hinge	CXXC motif	LLGG motif	N-Gly (Asn ²⁹⁷)	Extra N-Gly
IgE1	Bl, Sp, LN, Te	Yes	Shared peptide	NA	NA	NA	NA	NA
IgE2	Bl, Sp, LN, BM, Lu, Te	Yes		NA	NA	NA	NA	NA
IgG1	Bl, Sp, LN, BM, Lu, Te, Kd, He	Yes	Unique peptide	6	CPRC	LLGG	Present	At the hinge
IgG2	Bl, Sp, LN, BM, Lu, Te, Br, Kd,	Yes	Unique peptide	1	--RS	LPGG	Present	At CH2
IgG3	Bl, Sp, Te, Br	Yes	Undetected	1	--RS	LPGG	Present	At CH2
IgG4	Bl, Sp, LN, BM, Lu, Te, Kd, He, Li	Yes	Unique peptide	2	--RC	LPGG	Present	No

Bl: blood, Sp: spleen, LN: lymph node, BM: bone marrow, Lu: lung, Te: testes, Br: brain, Kd: kidney, He: heart and Lv: liver. BOLD FONT: >0.6 normalized log reads, REGULAR FONT: <0.6 normalized log reads. IGHG gene features reported but not applicable (NA) to IGHE genes.

Gene ontology for function prediction

COFACTOR is an *in silico* protein function prediction algorithm that employs three complementary approaches to surmise protein features [153]. 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 (protein database) 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

Reads from ERB tissues previously published [154] were aligned to genes of interest using the pseudo aligner Kalisto [155]. Transcripts per million were log transformed and normalized to *GAPDH*. Heatmaps were created using the R package pheatmap [156].

Positive selection analysis

Nucleotide sequences were codon aligned using muscle in Geneious. Alignments were then trimmed to the first in-common, in-frame, codon to maintain open-reading frames for each species. Final stop codons were removed and edges trimmed to a multiple of three. Alignments were exported as phy files. Double spaces were added between sequence names and sequences, and all dashes replaced with N. The alignment was uploaded into EasyCodeML along with a Newick tree based on individual alignments. Each sample was run on the branch-site model to detect individual residues and branches under selection based on LRT values.

Animals

There are challenges to studying the humoral response of ERBs to MARV. First, in contrast to mice, these animals are not commercially available and are not inbred, nor do genetically modified strains exist. There are currently only three institutes in the world, the CDC [157], the National Institute for Communicable Diseases of the National Health Laboratory Service (South Africa), and the Friedrich-Loeffler-Institut (Germany) with ERB colonies for research. For the proposed work we are continuing to collaborate with the CDC, which houses their colony of ERBs in a biosafety level 3 (BSL-3) facility. In addition, all MARV work has to be done at biosafety level 4 (BSL-4). The biocontainment burden lengthens the amount of time assays take, and requires significant funding for those trained to perform the work. Further, conducting animal studies in biocontainment severely restricts the number of subjects, volume of samples that can be obtained, and ability to repeat these assays independently. To ameliorate some of these restrictions, we have tested the inactivation of sera from biocontainment and found no detectable difference in activity post inactivation. We have also collaborated with the Lube Conservatory to receive peripheral blood mononuclear cells (PBMCs) from ERBs in captivity for cell purification. These collaborations ensured a source of cells without the biocontainment constraints, which will allowed us to delegate studies between our teams and minimized the biocontainment work necessary. RNA was isolated from three healthy ERBs maintained at the CDC. Research was conducted under an IACUC approved protocol and complied with the Animal Welfare Act, PHS Policy, and other Federal regulations and statues relating to animal experiments. The work was conducted at an accredited facility by the Association for Assessment and Accreditation of Laboratory Animal Care, International and adhered to the principles in the Guide for the Care and Use of Laboratory Animals, National Research Council, 2011.

Results

Sequencing and assembly of locus

Our first goal was to obtain an accurate gene organization and annotation of the ERB IGH locus. The high-quality assembly ERB genome (Raegyp2.0) [85] contained annotated IGHC genes on four contigs: NW_015493306.1, NW_015494929.1, NW_015493824.1, and NW_015493956.1 (Figure 2). Contig NW_015494929.1 (~1.36 Mb) contains a single copy of an IgA gene (*IGHA*) upstream *TMEM121* which also flanks *IGHA* in the human IGH locus, suggesting human and ERB loci are orthologous and that this contig represents the 3' end of the locus [158]. The 5' portion of the ERB IGH locus was found on Contig NW_015493306.1 (~144kb) containing a single IGHV gene and an IgM gene (*IGHM*). We reasoned that these two contigs represent the boundaries of the IGHC genes and predict that it is distal on an ERB chromosome orthologous to human chromosome 14q [159]. Several complete or partial IgE (*IGHE*) and IgG (*IGHG*) genes were found distributed among the four contigs, one of those labeled as a pseudo IgE gene (Figure 2). Annotated IGHV genes were present on 17 contigs (Table 3) and no IGHD or IGHJ genes were annotated.

Despite the increasing quality of newly published mammalian draft genomes the repetitive structure of some genomic regions continues to hinder contiguous and complete genome assemblies [160]. To overcome this limitation, we utilized bacterial artificial chromosomes [126]. Noteworthy, the ERB tissue used for the BAC library generation was of different origin (see methods) than that used to construct the original Raegyp2.0 genome [85]. Pools of BACs were screened by PCR and we identified four BACs, BACIg1 to BACIg4. BACs were sequenced using Pacific Biosciences long read sequencing, assembled, and compared to the Raegyp2.0 genome by using NCBI's Basic Local Alignment Search Tool (BLAST). All four BACs mapped to portions of the four contigs

above as well as the 17 contigs containing IGHV genes. We assembled BACIg1, BACIg2 and BACIg3 resulting in a contig of 474,666 bp (Figure 3). BACIg1/BACIg2 overlap (~32kb) is 100% identical and BACIg2/BACIg3 overlap (~20kb) contains only two single nucleotide deletions (present in BACIg2). These data strongly suggest that these three BACs represent the same haplotype. BACIg4 overlapped with BACIg1 (~61kb) but had a pairwise identity of 96.5% (Figure 3), suggesting that BACIg4 represents the alternative haplotype. Due to the relatively low pairwise identity at the BACIg1/BACIg4 overlap, a consensus sequence was not created at the overlap, and the final IGH assembly, named RaegyplGH3.0, contains the entirety of BACIg4 sequence and the consensus sequence of the assembled BACIg1-3 (Figure 3). At one end, RaegyplGH3.0 (634,142 nucleotides) extends into what is likely a sub-telomeric repeat region while at the other, to genes normally flanking the IGH locus in other mammals including TMEM121, CRIP1, TEDC1 and a portion of CRIP2, and thus expectedly contains the complete IGHV, IGHD, IGHJ, and IGHC gene repertoire. RaegyplGH3.0 links together Raegy2.0 contigs NW_015493306.1 and NW_015494929.1 and adds ~54kb of intervening sequence. Bionano Optical Mapping data demonstrated that contigs NW_015493306.1 and NW_015494929.1 are separated by ~56kb of sequence (data not shown). The other two Raegy2.0 contigs (NW_015493824.1, and NW_015493956.1) that contain IGH genes likely represent missassembled portions of the newly sequenced 54kb intervening sequence.

Annotation and description of IGHV, IGHD and IGHJ genes

In Raegyp2.0 58 IGHV genes present on 17 contigs were annotated by NCBI's pipeline (Table 3). Within RaegypIGH3.0 we were able to annotate 66 IGHV genes using IMGT/LIGMotif (Figure 4) [134]. Of those 66 IGHV genes, ten are pseudo genes and one (IGHV6-1.1) appears to be truncated. All IGHV genes have common conserved amino acids (C23, W41, C104) as identified by IMGT/LIGMotif. Similar to what has been observed in other pteropid bats, IGHV genes represent all three V clans (based on human subgroupings: clan I: IGHV1/IGHV5/IGHV7; clan II: IGHV2/IGHV4/IGHV6; and clan III: IGHV3) and include the families IGHV1, IGHV3, IGHV4, IGHV6 and IGHV7 (Figure 5) [139, 161]. They are on average 301 nt +/- 31 nt long (excluding the truncated IGHV6-1.1) which is similar to what is reported in other mammals (Figure 6) [134, 161]. Additionally, 65 IGHV genes contained downstream recombination signal sequences [162] were identified by IMGT/LIGMotif, with only one pseudogene (VH1-2.1) lacking one (Figure 5).

No IGHD or IGHJ genes were annotated in Raegyp2.0 by NCBI. We examined RaegypIGH3.0 using IMGT/LIGMotif and identified eight IGHD and nine IGHJ genes (Figure 4) which included IGHD genes comprising the IGHD1, IGHD2, IGHD3, and IGHD6 families. ERB IGHD genes are 17-37 nucleotides in length similar to human IGHD genes, which vary from 11-37 nucleotides in length (Figure 6) [161]. All IGHD genes contained upstream and downstream RSS sequences (Figure 5). The nine IGHJ genes identified comprise the IGHJ2 (two genes), IGHJ4 (six genes), and IGHJ5 (one gene) families (Figure 5). ERB IGHJ genes range between 40-58 nucleotides in length, slightly smaller on average than what is reported in humans (48-63 nucleotides) (Figure 6) [161]. All IGHJ contain functional RSS with the exception of the two IGHJ2 genes (Figure 5). Both IGHJ2 genes contain impaired heptamers, 5'-GAGCGGG-3' instead of the conserved 5'-GAGCGTG-3', observed in all human IGHJ gene heptamers. Moreover, all IGHJ genes from the IGHJ4 and IGHJ5 families retained the highly conserved WGXXG amino acid motif,

while IGHJ2.1 (WCRG) and IGHJ2.2 (WGEA) do not (Figure 5) [163]. Thus, it is likely that the IGHJ2 genes are not functional and not utilized in recombination. As has been observed with other mammals, we identified variation in the N-terminus of IGHJ genes and high conservation in the C terminus. One intriguing observation was the IGHD-IGHJ-IGHD-IGHJ organization within the IGH locus (Figure 4). This observation is similar to a duplication event previously described in the bovine (*Bos taurus*) draft genome [164].

After annotating the IGHV, IGHD and IGHJ genes, we wanted to determine if these genes are expressed. Mining previously published transcriptomic data derived from ERB lymph node to identify V(D)J containing transcripts by BLAST against RaegyplGH3.0. Recovered reads were annotated by IMGT [85]. The annotated reads were then remapped to RaegyplGH3.0 to confirm location and expression of annotated genes. We were able to validate the expression of the majority of IGHV genes annotated as functional. Very limited transcriptome support was detected for any of pseudogenes and the truncated *IGHV6-1.1*. We also identified transcriptomic evidence for all IGHJ genes (Figure 7). Notably, >30 fold fewer transcripts mapped to the two IGHJ2, and IGHJ4.1 genes which is consistent with our supposition that the IGHJ genes in the first cluster might not incorporate in mature Igs (Figure 4, Figure 7). As IGHD genes are small and difficult to identify by BLAST, we used reads that were annotated with an IGHJ gene and annotated them with IMGT-HighVQuest. This method was able to further confirm all eight IGHD genes.

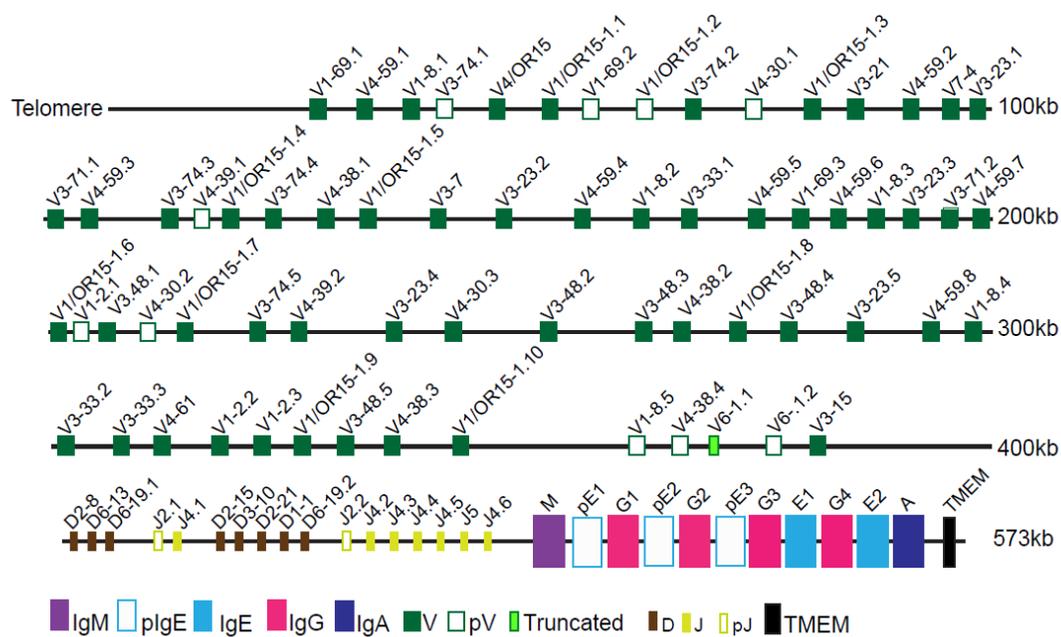


Figure 4 : IGHC locus representation

Complete assembly of the IGH locus (RaegyIGH3.0). Genes are placed relative to their location within the entire assembly and the locus spans ~573kb.

**Unpublished. Submitted to Cell.*

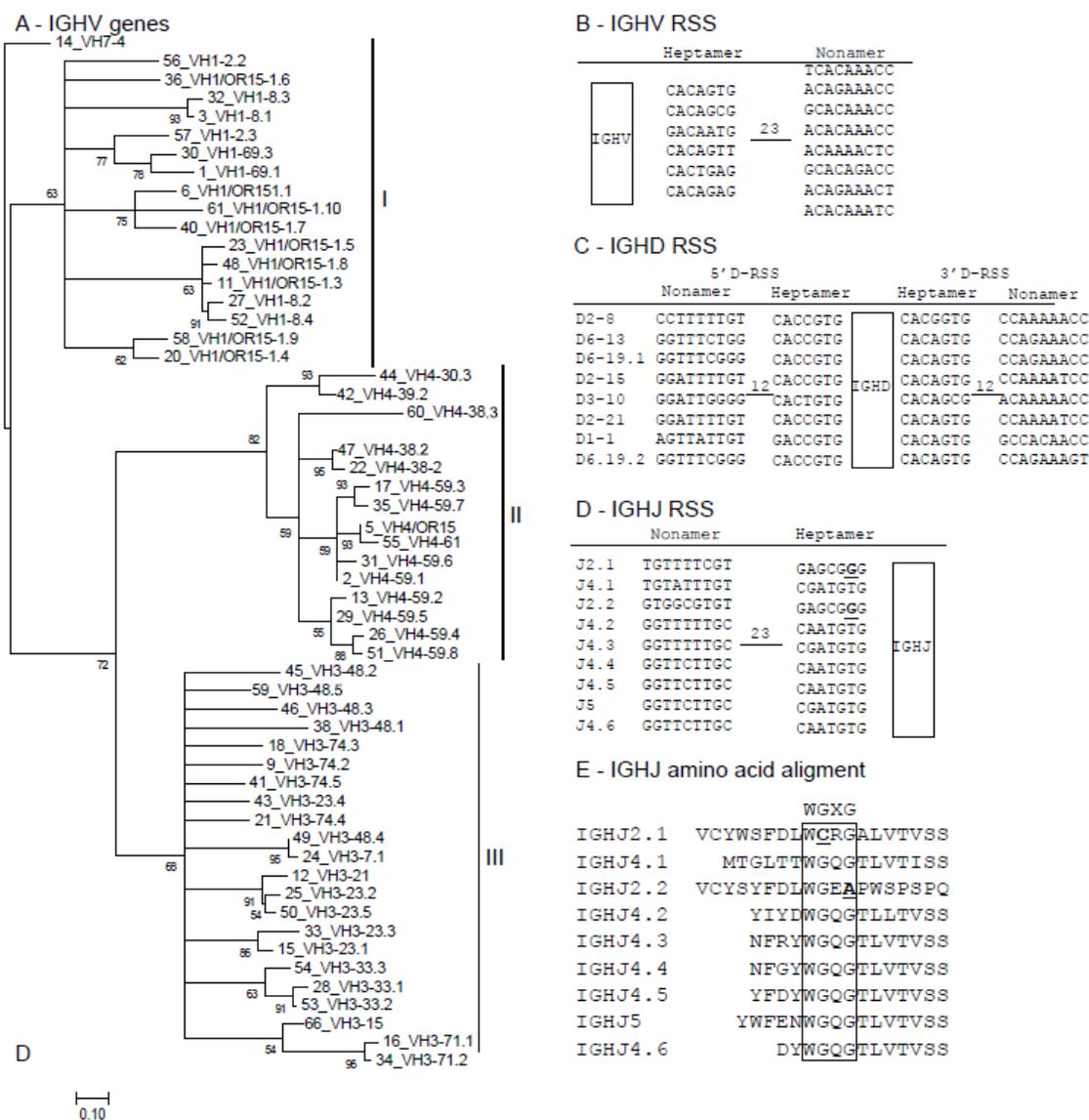


Figure 5 : Genome annotation of V(D)J genes.

(A) IGHV phylogeny of functional genes. Amino acids were aligned with ClustalW and then a tree created in Geneious using Bayes. Genes cluster within defined clans I, II and III. (B) IGHV recombination signal sequences. (C and D) IGHD recombination signal sequences. (E) IGHJ recombination signal sequences. (F) Amino acid alignment of IGHJ genes with conserved WGXX motif (boxed) and residues that could impair functional products (underlined) indicated.

*Unpublished. Submitted to Cell.

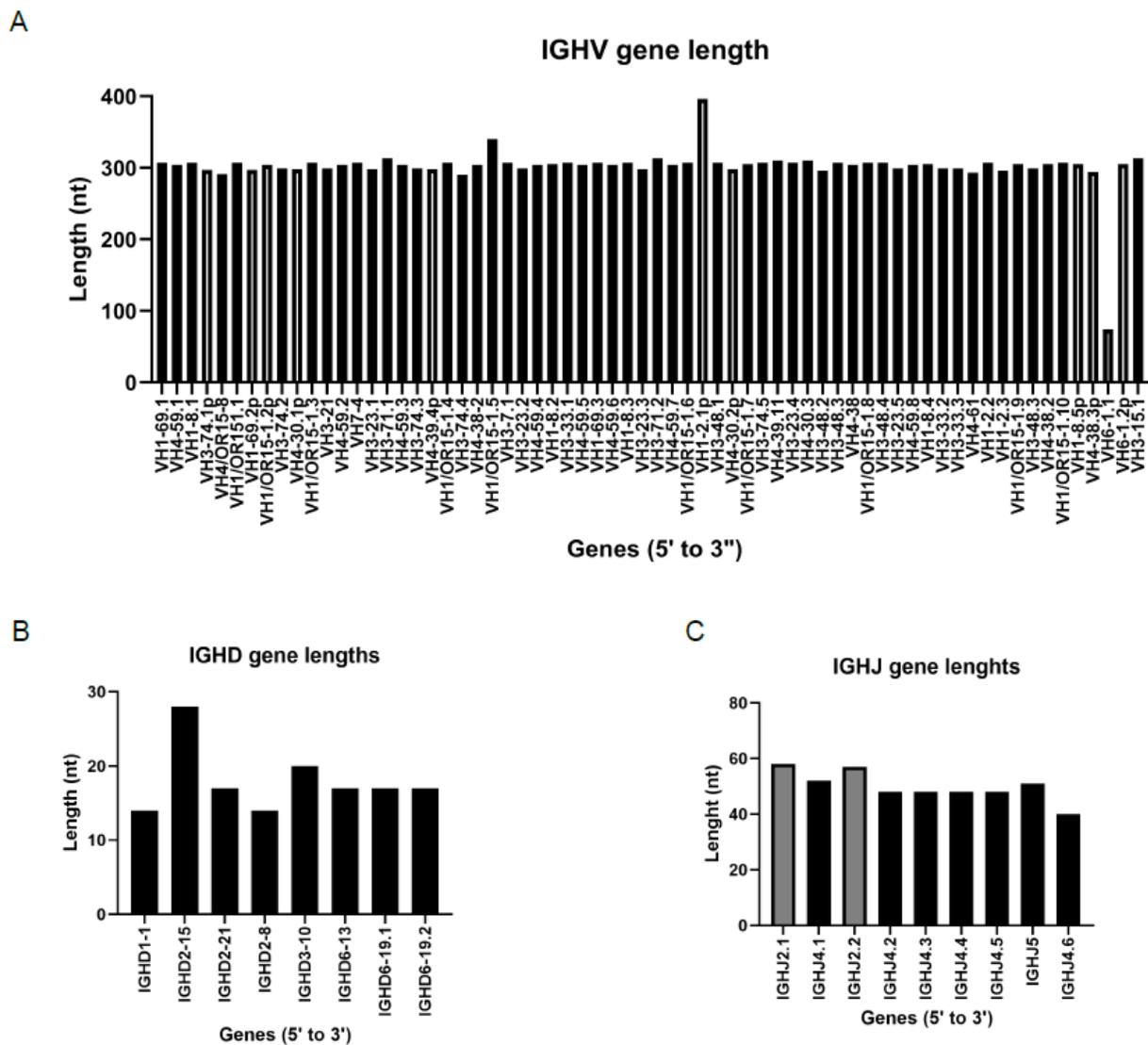


Figure 6 : Length distribution of IGHV, IGH D, and IGH J genes.

Nucleotide lengths for ERB IGHV genes (A), IGH D (B) and IGH J (C) contained on the IGH locus (RaegypIGH3.0), ordered from 5' to 3'. Pseudo genes shaded gray, functional in black.

**Unpublished. Submitted to Cell.*

A: IGHV genes

IGHV genes	Count(A)	Count(B)
1_VH1-69.1	112.5	50
2_VH4-59.1	231.7	119
3_VH1-8.1	19.4	7
4_VH3-74.1 (ψ)	34.4	6
5_VH4/OR15-8	257.2	182
6_VH1/OR151.1	439.4	411
7_VH1-69.2 (ψ)	217.1	95
8_VH1/OR15-1.2 (ψ)	54.3	34
9_VH3-74.2	222	201
10_VH4-30.1 (ψ)	6.3	0
11_VH1/OR15-1.3	45.8	8
12_VH3-21	773.7	624
13_VH4-59.2	198.3	78
14_VH7-4	695.1	673
15_VH3-23.1	65.4	0
16_VH3-71.1	20	14
17_VH4-59.3	101.2	27
18_VH3-74.3 (ψ)	284.9	217
19_VH4-39.4 (ψ)	29.1	1
20_VH1/OR15-1.4	261.1	201
21_VH3-74.4	572.6	428
22_VH4-38-2	22.6	5
23_VH1/OR15-1.5	73.4	0
24_VH3-7.1	519.5	256
25_VH3-23.2	394.9	298
26_VH4-59.4	105.3	0
27_VH1-8.2	68.7	0
28_VH3-33.1	178.6	30
29_VH4-59.5	177.3	59
30_VH1-69.3	361	278
31_VH4-59.6	125.4	30
32_VH1-8.3	49	10
33_VH3-23.3	101.2	31

B: IGHV genes

IGHV genes	Count(A)	Count(B)
34_VH3-71.2	38.2	26
35_VH4-59.7	73	2
36_VH1/OR15-1.6	100.9	40
37_VH1-2.1 (ψ)	1.4	1
38_VH3-48.1	116.8	49
39_VH4-30.2 (ψ)	6.3	0
40_VH1/OR15-1.7	249	229
41_VH3-74.5	664.4	432
42_VH4-39.11	122	94
43_VH3-23.4	96.6	65
44_VH4-30.3	99	99
45_VH3-48.2	365.2	185
46_VH3-48.3	543.4	481
47_VH4-38	22.3	6
48_VH1/OR15-1.8	73.4	0
49_VH3-48.4	272.9	10
50_VH3-23.5	767.6	592
51_VH4-59.8	105.3	0
52_VH1-8.4	68.7	0
53_VH3-33.2	329.3	121
54_VH3-33.3	613.2	307
55_VH4-61	132.9	106
56_VH1-2.2	185.9	144
57_VH1-2.3	1003.7	824
58_VH1/OR15-1.9	116.4	105
59_VH3-48.3	235.1	156
60_VH4-38.2	19.6	14
61_VH1/OR15-1.10	187.5	165
62_VH1-8.5 (ψ)	33.6	12
63_VH4-38.3 (ψ)	45	45
64_VH6-1.1	N/A	N/A
65_VH6-1.2 (ψ)	N/A	N/A
66_VH3-15.1	N/A	N/A

IGHJ genes	Count(A)	Count(B)
1_JH2.1	5	5
2_JH4.1	5	5
3_JH2.2	7	7
4_JH4.2	276.7	276
5_JH4.3	1102.1	942
6_JH4.4	718.2	596
7_JH4.5	1083.2	938
8_JH5	721.8	619
9_JH4.6	809.4	703

C: representative example (IGHJ4.6)

CCAAGAGCCAGGTTTTCTTGACACTGAACTCTGTGACCAGCGAGGACAC
 GGCCGTATATTACTGTGCAAGAGAGTTACGGATCCCTGGGGCCAGGG
GACCCCTGGTCACCGTCTCCACAGATC

Figure 7 : V/J gene counts from transcriptomic data.

ERB IGHV and IGHJ genes were BLAST against transcriptomic reads from ERB lymph node. A) IGHV genes and B) IGHJ genes where count A is the sum of the count of reads and if a read mapped to multiple genes is split between all hits and where count B is unique reads that only map to that gene. C) Representative example of a read which a IGHD (highlighted) and IGHJ (underlined) gene identified.

**Unpublished. Submitted to Cell.*

Annotation and description of IGHC genes

To identify RaegypIGH3.0 IGHC genes, we downloaded NCBI annotated IGHC sequences from Raegyp2.0 and performed a BLAST search against RaegypIGH3.0. The annotation of each gene was inspected to ensure that exon/intron rules were followed. The most 5' IGHC gene of the locus is a single *IGHM* followed by a repeating structure of *IGHE* and *IGHG*, and ends with a single *IGHA* (Figure 4). Like some pteropid bats, we were unable to identify an *IGHD* gene in ERBs [139]. In contrast to all other mammals studied thus far we identified five *IGHE* genes in RaegypIGH3.0 [165]. The first three *IGHE* genes contain indels that result in the introduction of stop codons in all three forward reading frames. Thus these three genes are likely unprocessed pseudogenes (Figure 4). The last two *IGHE* genes contain complete open reading frames and thus are putatively functional; making ERBs the only species identified to date that contain more than one putatively functional *IGHE* gene. Similar to what has been observed in mice and humans, the ERB contains four IgG genes and all four contain complete open reading frames and are putatively functional.

Given the quality of the assembly and the fact that open reading frames for all IgG, IgA, IgM genes, and two of five IgE genes were complete and required no additional sequence manipulation on our behalf we are confident in the accuracy of our annotation. IgE and IgG genes are numbered based on their position in a 5' to 3' orientation with the telomeric end representing the 5' end (Figure 4). To determine the sequence identity among the IgE and IgG genes we performed nucleotide and amino acid alignments using ClustalW (Figure 8, Figure 9). At the amino acid level the pairwise identity between the four IgG genes is 89%, and between *IGHE1* and *IGHE2* is 94.4% while at the nucleotide level they had pairwise identities of 91.9% and 97.8%, respectively. Consistent with

previous observation in other species, the majority of the *IGHG* sequence heterogeneity is clustered at the hinge and CH2 domains [162].

We used our previously published transcriptomic data to verify the expression of the different IGHCs found in RaegyplGH3.0 [154]. Raw reads were mapped by BLAST against the coding sequence of each IGHC gene. Unique reads specific to a single isotype or subclass were identified. Finally, we also performed mass spectrometry on protein A/G purified total sera collected from wild-caught ERBs to verify levels of expression (Table 5). Using a custom database of tryptic peptides based of on the coding sequences defined here, we were able to detect unique peptides corresponding to IgG1, IgG2, IgG4, IgM and IgA. Remarkably, we were also able to detect an IgE peptide (shared by IgE1 and IgE2), but not IgG3 in serum. These data demonstrate that mRNAs derived from the identified genes are translated and confirms the presence of multiple IgG subclasses, IgA, IgM, and IgE.

IgE alignments to select reads

```

IgE1  CTCACCTGTCTGGTCCAGAACTTCTTCCCCGTGGACACCTCATTTGCAGTGGCTGCGGAAT
1     -----
2     -----
3     -----CCGTGGACACCTCATTTGCAGTGGCTGCGGAAT

IgE1  GACGCCCTGGTCCAGACGGACCAGCAGGCCACCACGCAGCCCCCTCAAGGCCAACAGCTCC
1     -----
2     ----CCCTGGTCCAGACGGACCAGCAGGCCACCACGCAGCCCCCTCAAGGCCAACAGCTCC
3     GACGCCCTGGTCCAGACGGACCAGCAGGCCACCACGCAGCCCCCTCAAGGCCAACAGCTCC

IgE1  AGCCCCGCCTCCTTTGTCTTCAGCCGCCTGGAGGTCCACCGGGCGGGCTGGGAGAAGAGG
1     -----AGGTCCACCGGGCGGGCTGGGAGAAGAGG
2     AGCCCCGCCTCCTTTGTCTTCAGCCGCCTGGAGGTCCACCGGGCGGGCTGGGAGAAGAGG
3     AGCCCCGCCTCCTTTGTCTTCAGCCGCCTGGAG-----

IgE1  CACAAGTTCACCTGCCAAGTGGTCCACAAAGGCGCTGCCCGGCTTCAGGACCCCAAGAAA
1     CACAAGTTCACCTGCCAAGTGGTCCACAAAGGCGCTGCCCGGCTTCAGGACCCCAAGAAA
2     CACAAG-----
3     -----

=====

IgE2  CTGCCGGTGGACGCCAACGACTGGATCGAGGGTGTGACCTACCGGTGCGAGGTGTCCAC
1     -----
2     -----TACCGGTGCGAGGTGTCCAC

IgE2  CCGCACCTGCCAGGGCCATTGCGCGCACCATCGCCAAGTCCCCCGGTAAGCGCGCCGCC
1     -----CCAGGGCCATTGCGCGCACCATCGCCAAGTCCCCCGGTAAGCGCGCCGCC
2     CCGCACCTGCCAGGGCCATTGCGCGCACCATCGCCAAGTCCCCCGGTAAGCGCGCCGCC

IgE2  CCCGAGGTCTACGTGTTCCCACCACCCGAGGAGGGCCAGGGCACCAAGGACACGGTCACC
1     CCCGAGGTCTACGTGTTCCCACCACCCGAGGAGGGCCAGGGCACCAAGGACACGGTCACC
2     CCCGAGGTCTACGTGTTCCCACCACCCGAGGAGGGCCAGGGCAC-----

IgE2  CTCACCTGTCTGGTCCAGAACTTCTTCCCCGTGGACATCTCAGTGCAGTGGCTGCGGAAT
1     CTCACCTGTCTGGT-----
2     -----

=====

```

Figure 8 : IGHE gene alignments with reads.

Alignments of reads from RNA-seq data sets demonstrating the association of private SNPs with specific immunoglobulins. **Yellow** highlight denotes the unique SNP in identified in the read(s) (numbered rows) and the immunoglobulin (top row). **Red** highlight denotes SNPs present in the reads that are not present in the immunoglobulin.

**Unpublished. Submitted to Cell.*

IgG alignments to select reads

```

IgG1  TACCTTGTGACCGAGAAACGGCAGCTGCTCTTGCTGCAGCGCCTGCCCTCGCTGCCA
1  TACCTTGTGACCGAGAAACGGCAGCTGCTCTTGCTGCAGCGCCTGCCCTCGCTGCCA
2  TACCTTGTGACCGAGAAACGGCAGCTGCTCTTGCTGCAGCGCCTGCCCTCGCTGCCA
3  -----AAACGGCAGCTGCTCTTGCTGCAGCGCCTGCCCTCGCTGCCA

IgG1  GCTCCCGAGCTCCTGGGAGGACCCTCGGTCTTCATCTTCCCCCGAAACCAAGGACACC
1  GCTCCCGAGCTCCTGGGAGGACCCTCGGTCTTCATCTTCCCCC-----
2  GCTCCCGAGCTCCTGGGAGGACCCTCGGTCTTCATCTTCCCCCGAAACCAAG-----
3  GCTCCCGAGCTCCTGGGAGGACCCTCGGTCTTCATCTTCCCCCGAAACCAAGGACACC

IgG1  CTCATGATTTCCGGGAAGCCCTCGGTCAAGTGTGGTGGTGGACGTGAGCCAGGACGAC
1  -----
2  -----
3  CTCATGATTTCCGGGAAGCCCT-----

=====

IgG2  ACATGTGCGGTGACGCACGAGGCCTTGCCCAATTACCACATGGGAAATCCATCTCCATT
1  -----ACATGGGAAATCCATCTCCATT

IgG2  AGTCCGAGATTCCTTGACGAGACCTGTGTGGAGGCCAGGACGGGGAGCTGGACGGG
1  AGTCCGAGATTCCTTGACGAGACCTGTGTGGAGGCCAGGACGGGGAGCTGGACGGG

IgG2  CTGTGGACCACCATCTCCATCTTCATCACCCCTCTCCTGCTCAGCGTGTGTACAGTGC
1  CTGTGGACCACCATCTCCATCTTCATCACCCCTCTCCTGCTCAGCGTGTGTACAGTGC

IgG2  ACCGTACCCCTCTCAAGGTGAAGTGGATTTACTCTTCGTGGTAGACCTGAAGCGGACC
1  -----

=====

IgG3  TTCATCTTCCCCCGAAATAAAGGACGCCCTCATGATTACCGGAAGCCACAGTCAAG
1  --CATCTTCCCCCGAAATAAAGGACGCCCTCATGATTACCGGAAGCCACAGTCAAG
2  -----
3  -----

IgG3  TGTGTGGTGGTGGATGTGATCGAGGACGTCAAATTCAGCTGGTATGTGGATGACAAATGAG
1  TGTGTGGTGGTGGATGTGATCGAGG-----
2  -----ATGTGATCGAGGACGTCAAATTCAGCTGGTATGTGGATGACAAATGAG
3  -----

IgG3  TTGGACTCGGCCGAGACAAAAACCAGAGATACAGGAGAAAGACACCTACCGCATCGTC
1  -----
2  TTGGACTCGGCCGAGACAAAAACCAGAGATACAGGAGAAAGACACCTACCG-----
3  -----GGAGAAAGACACCTACCGCATCGTC

IgG3  AGCCCCCTAGAGATCAATCACCGGACTGGCTAAATGGCAAAGAGTTCAAGTGCAAGGTT
1  -----
2  -----
3  AGCCCCCTAGAGATCAATCACCGGACTGGCTAAATGGCAAAGAGTTCAAGTGCAAGGTT

=====

IgG4  TGGAAAGTCCAGCAGCAAGGCCCTGCGGCCCCCATCGAGAAGAGCATCAAAAAGGCCACA
1  -----CA
2  -----
3  -----

IgG4  GGACCGATCCAGGAGCCGACGATATATGTTCTGGCTCCACACCCGGATGAGCTGGCCAAG
1  GGACCGATCCAGGAGCCGACGATATATGTTCTGGCTCCACACCCGGATGAGCTGGCCAAG
2  ---CCGATCCAGGAGCCGACGATATATGTTCTGGCTCCACACCCGGATGAGCTGGCCAAG
3  -----

IgG4  GACACGGTCAGTGTGACCTGCCTGGTCAAAGACTTCTCCCGTCTGACATCACGTGGAG
1  GACACGGTCAGTGTGACCTGCCTGGTCAAAGACTTCTCCCGTCTGACATCACGTGGAG
2  GACACGGTCAGTGTGACCTGCCTGGTCAAAGACTTCTCCCGTCTGACATCACGTGGAG
3  -----G

IgG4  TGGGAGCAATAGGCAGCCAGAGCCAGAGGCCAAGTACAGCAGCACCCCGCCAGCTG
1  TGG-----
2  TGGGAGAG-----
3  TGGGAGCAATAGGCAGCCAGAGCCAGAGGCCAAGTACAGCAGCACCCCGCCAGCTG

```

Figure 9 : *IGHG* gene alignments with reads.

Alignments of reads from RNA-seq data sets demonstrating the association of private SNPs with specific immunoglobulins. **Yellow** highlight denotes the unique SNP in identified in the read(s) (numbered rows) and the immunoglobulin (top row). **Red** highlight denotes SNPs present in the reads that are not present in the immunoglobulin.

**Unpublished. Submitted to Cell.*

Table 5 : Protein sequences used to identify immunoglobulins by Mass Spectrometry.

SEQUENCE	IMMUNOGLOBULIN
WLHGNQELSR	IgA
VNYLGTIIEDHAR	IgE
PSVTCVVVDVSQDDTDVK	IgG1
IVSTLEINHR	IgG2
ATGPIQEPQVYVLAPHPDELAK	IgG4
GFSPSDVQVQWMQR	IgM

**Unpublished.*

IGHC genes evolutionary relationship

As expected, ERBs IGHG genes are most closely related to members of Yinpterochiroptera than of Yangochiroptera. Our data bolster previous reports of IGHG clustering within Chiropteran suborders (Figure 10) [166]. While ERB *IGHM* is basal to *Cynopterus sphinx* and both pteropid bats included in our analysis, *Cynopterus sphinx* *IGHA* and *IGHE* are basal to ERB and both pteropid bats *IGHA* and *IGHE*, respectively. The relationship of the four ERB IGHGs is unresolved, however, our analyses clearly demonstrate they cluster within the Yinpterochiroptera clade. All ERB IGHGs and IGHEs appear to cluster together, demonstrating that the tandem duplication observed occurred after the demarcation of the species. In summary, our molecular data recapitulate other taxonomic data demonstrating ERB clusters within Yinpterochiroptera [166].

Functional characteristics of the ERB IGHG genes.

IGHG motifs associated with Ig structure and function had been widely described in other mammals (Table 4). Ig cysteines are involved in intra- and inter-molecular disulfide bonds and are crucial for the Ig structure. All the ERB IGHG genes contained at least one cysteine in the hinge region that would potentially allow for IGH:IGH disulfide bonding to produce canonical dimeric receptors (Figure 10). The hinge domain of *IGHG* is known to be variable at both inter- and intra-specific levels, and crucial to form disulfide bonds between IGHs [167]. ERB *IGHG1* contains a single 5'-CPRCP-3' motif, which is present, repeated 4 times, in human *IGHG3* (Figure 10). Remarkably, neither ERB *IGHG2*, *IGHG3* nor *IGHG4* have the canonical CXXC domain at the hinge that was associated with the ability to form inter-molecular disulfide bonds between both IGHGs. Moreover, ERB *IGHG2* and *IGHG3* lack both cysteines at that domain, the first cysteine is missing and the second (Cys109) is replaced by a serine.

N-glycosylation modulates Ig effector functions, including its interaction with FcRs and complement. ERB *IGHM* contains all the N-glycosylation sites homologous to human *IGHM* and an additional two putative sites (Figure 10) [168]. Human *IGHE* contains seven sites that can be subject to N-glycosylation. Both ERB *IGHE1* and *IGHE2* share three of those sites (Figure 10) [168]. However, four N-glycosylation sites present in humans are missing in ERBs while both ERB *IGHE1* and *IGHE2* contain two additional putative N-glycosylation sites compared to human. Human *IGHA2* contains four N-glycosylation sites of which only two are conserved in ERB *IGHA* (Figure 10). Three other sites not homologous to human are present in ERB *IGHA*, although the second N-glycosylation site in ERB *IGHA* is relatively close to the first site in human *IGHA2* and may potentially perform a similar function [169]. Finally, all four human IGHGs contain a single N-glycosylation site at Asn297 [170], all four ERB IGHGs share that conserved site but ERB *IGHG2* and *IGHG3* have one additional putative site (Figure 10). Remarkably, the *IGHG1* of ERBs contains the longest hinge domain (23 amino acids) and a putative N-glycosylation site (IGG1 hinge: PVKYTCDRENGSCSCCSACPRCP) not present in the other ERB or human IGHGs.

The region of the IGHC responsible for making contact with the FcγRI had been very well characterized [171-173]. Beside the already mentioned N-glycosylation site at Asn²⁹⁷ three critical contacts have been identified: 1) the lower hinge region (Leu²³³Leu²³⁴Gly²³⁵Gly²³⁶); 2) the CH2 BC loop Asp²⁶⁵; and 3) the CH2 FG loop (Ala³²⁷Leu³²⁸Pro³²⁹Ala³³⁰Pro³³¹). All ERB Igs have the conserved Asp²⁶⁵ and the CH2 FG loop, however, Leu²³³Leu²³⁴ are only present in ERB IgG1 and not in the other subclasses, where the first leucine is conserved but the second is substituted by a proline (Leu²³³Pro²³⁴Gly²³⁵Gly²³⁶). The mutation of these leucines into alanines (LALA mutation) is associated with the inability of the IgG to bind to FcRs [174].

Tissue expression of immunoglobulin constant genes

We determined the tissue-specific expression profiles of the identified expanded IGH subclasses. Using RNA-seq transcriptomic data, reads were first pseudo aligned using Kalisto to obtain transcripts per million (TPM), log(2) transformed, and subsequently normalized to Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) to obtain level of expression of each IGHC per tissue (Figure 11). *IGHM* is expressed in all tissues and most highly expressed in the bone marrow, lymph node, blood, spleen, and lung. The highest expression of *IGHA* is observed in the lungs but low relative expression was identified across all tissues in a manner consistent with observations in other mammals [175]. Similar to other mammals, each subclass of *IGHG* had differential expression with one subclass being predominant in most sites examined [162]. We observed *IGHG1* generally had the highest expression followed by *IGHG4*, *IGHG2*, and finally *IGHG3* (lowest expressed in all tissues). While *IGHG1* is the most highly expressed in lymph node, spleen, and lung, it was not found in the brain or liver. *IGHG2* expression was found to be highest in the spleen with no expression in the heart and liver. *IGHG3* expression was highest in testes although it was lower than all other IGHGs examined and absent in bone marrow, heart, kidney, lymph node, liver and lung, which agrees with our inability to detect protein of this isotype in serum. In contrast, *IGHG4* was found in all tissues except brain and was the highest expressed in spleen. Of the two functional *IGHE* genes, expression of *IGHE2* was present in lymph node, blood, spleen, and testes whereas *IGHE1* was expressed in those tissues as well as bone marrow and lung. The differential expression of both IgE genes are highly indicative of a differentiation of function of these genes.

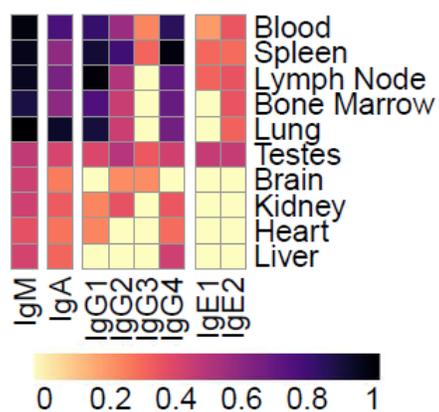


Figure 11: Tissue expression of ERB Igs

(A) Heatmap of log transformed reads normalized to GAPDH within each tissue: blood, spleen, lymph node, bone marrow, lung testes, brain, kidney, heart, liver.

**Unpublished. Submitted to Cell.*

Gene Ontology suggests IgE and IgG subclasses serve unique functions

Expansion in *IGHG* and *IGHE* genes and the differential tissue expression suggest a diversification of functions between subclasses. We utilized the COFACTOR [153], an algorithm built to predict functions based on conserved motifs, to determine putative functional differences between IGHC sequences. COFACTOR create homology models and the GO knowledgebase to predict functional differences.

Due to the complexity and diversity of IgG across mammals we restricted our analysis to just human and ERBs (Figure 12). According to this analysis, the four ERB IgGs likely perform different effector functions than their human counterparts with no clear statistically significant functional equivalency between IgGs from both species. Further, similar to how the genes cluster phylogenetically (Figure 10), the Igs cluster within species suggesting no clear similarity in predicted functions overall. Strikingly, no cellular component (CP) Plasma Membrane Part GO term was identified for ERB IgG1 and IgG3, in sharp contrast with all the human IgGs and ERB IgG2 and IgG4.

IGHE of all mammals examined were predicted to exhibit immune effector process and classical complement activation functions. Both *IGHE1* and *IGHE2* are predicted to be important for the response to other organisms, a key feature of *IGHE* in other mammals [165]. But ERB *IGHE1* and *IGHE2* also exhibit differences that support a possible diversification of function (Figure 12). While *IGHE1* functionality appears to more closely resemble human, mouse, horse, dog, cow, and pig; *IGHE2* predicted function better mirrors the *Myotis lucifugus* (Figure 12). The cellular component prediction suggests that *IGHE2* is secreted while *IGHE1* is more likely to be membrane bound. Although these predictions require further validation, they offer insight into the potential functions of these novel immunoglobulins. Finally, we examined ERB *IGHA* and *IGHM*, and found that their predicted functions are broadly similar to what has been described in human homologues.

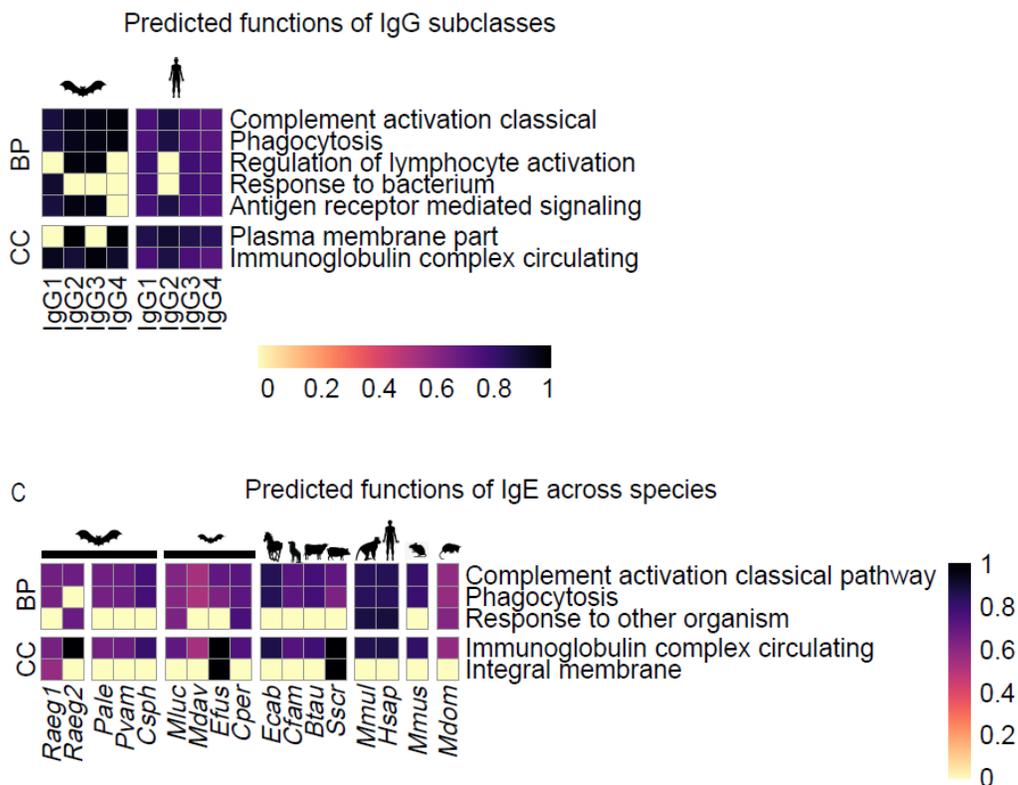


Figure 12 : Predicted functions of ERB Igs

A) Correlated gene ontologies between Human and ERBs *IGHG* subclasses based on confidence score. Biological process (BP): GO:0006958, GO:0006910/GO:0006911, GO:0051707 and cellular component (CC): GO:0044425, GO:0044459, GO:0042571 shown. B) Correlated gene ontologies for *IGHG* genes between species: ERB (Raeg1, Raeg2), Pale (*Pteropus alecto*), Pvam (*Pteropus vampyrus*), Csph (*Cynopterus sphinx*), Mluc (*Myotis lucifugus*), 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 domestica*). Biological process (BP): GO:0006958, GO:0006910, GO:0051249, GO:0009617, GO:005085 and cellular component (CC): GO:0016021, GO:0042571 shown.

*Unpublished. Submitted to Cell.

Identification and expression of ERB FcRs

Given that ERB IgG does not appear to mediate virus neutralization on MARV infection [133] we investigated functions associated to the Fc region of the Ig mediated by FcR interactions or proteins from the complement system. FcRs can be expressed either at the cell membrane or intracellularly. Their number and composition varies between species [176]. Indeed, the mouse and human genomes contain 11 and 16 FcRs, respectively [136]. We identified 13 FcRs annotated in Raegyp2.0, the orthologs for human FCεRI, FCεRII, FCγRI, FCγRIIB, FCγRIIA, FCμR, FCαμR, FcRn, FCRL4, FCRL5, PIGR and TRIM21, and a receptor with no-ortholog annotated as FCγRII-like (

Table 6). Fc signaling associated chains were also found in Raegyp2.0 (FCεR1β and FCεR1γ). However, we did not find orthologs for FCαR, FCγRIIA, FCγRIIC, and FCγRIIIB. Among ERB FCγRs, one is the canonical inhibitor FCγRIIB containing the Immunoreceptor Tyrosine-based Inhibition Motifs (ITIM) at the cytosolic tail (Figure 13). In contrast to human FCγRIIA and FCγRIIC, no ERB FcγRs contain Immunoreceptor Tyrosine-based Activation Motifs (ITAM) motif embedded in their cytoplasmic tails. Interestingly, within ERB FCγRI and the FCγRII-like an Asn³⁰⁶ has reverted to an Asp³⁰⁶ (Figure 13). This substitution would still allow for association with the signaling γ chain but this association would be required for cell surface expression [177]. Moreover, the ERB FCγRI contains a longer FG loop associated with lower affinity for IgG [173] (Figure 13).

We also determined their specific expression profiles across ERB tissues. All FcRs were found expressed in at least one tissue; FcRn, as in humans, was the most ubiquitously expressed (Figure 13) [178]. Detailed studies of the interaction between these receptors and the corresponding ERB Igs, examination of their expression at the single-cell level, and determination of the mechanism to mediate function will be the focus of future studies.

Table 6 : FcRs and associated functions in humans, mice, and ERBs

Receptor	Hsap	Mmus	Raeg	Function	Signaling
IgA/IgM					
FcαR	Yes	No	No	Activation	γ chain
FcαμR	Yes	Yes	Yes	Phago/endocytosis	
FcμR	Yes	Yes	Yes	Endocytosis	
IgG					
FcγRIa	Yes	Yes	Yes	Activating	γ chain (ITAMi)
FcγRIIa	Yes	No	No	Activating	ITAM (ITAMi)
FcγRIIb	Yes	Yes	Yes	Inhibitory	ITIM
FcγRIIc	Yes	No	No	Activating	ITAM
FcγRII-like	N/A	N/A	Yes	Unknown	Unknown (γ chain?)
FcγRIIIa	Yes	Yes	Yes	Activating	γ chain
FcγRIIIb	Yes	No	No	Decoy/activation	GPI
FcγRIV	No	Yes	No	Activating	γ chain
IgE					
FcεRIa	Yes	Yes	Yes	Activation	β and γ chain
FcεRII	Yes	Yes	Yes	IgE regulation	
FcR-like					
FcR L1	Yes	Yes	Yes		(ITAM) ₂
FcR L2	Yes	No	No		ITAM + (ITIM) ₂
FcR L3	Yes	No	Yes		ITAM + ITIM
FcR L4	Yes	No	Yes		(ITIM) ₂
FcR L5	Yes	Yes	Yes	Inhibitory	ITAM + (ITIM) ₂
FcR L6	Yes	Yes	No		ITIM
FcR LA	Yes	Yes	Yes		Intracellular
FcR LB	Yes	Yes	Yes		Intracellular
Others					
FcRn	Yes	Yes	Yes	Recycling/transport	
PIGR	Yes	Yes	Yes	Transport	
TRIM21	Yes	Yes	Yes	Activation/proteasome	
Associated/signaling molecules					
FcγBP	Yes	Yes	Yes		
FcεRIβ	Yes	Yes	Yes	Signaling molecule	ITAM
FcεRIγ	Yes	Yes	Yes	Signaling molecule	ITAM

Summary of FcRs in human, mouse, and ERBs with known functions and associated motifs. Motif presence in ERBs examined manually after alignment of amino acid sequences in Geneious R11.

**Unpublished. Submitted to Cell.*

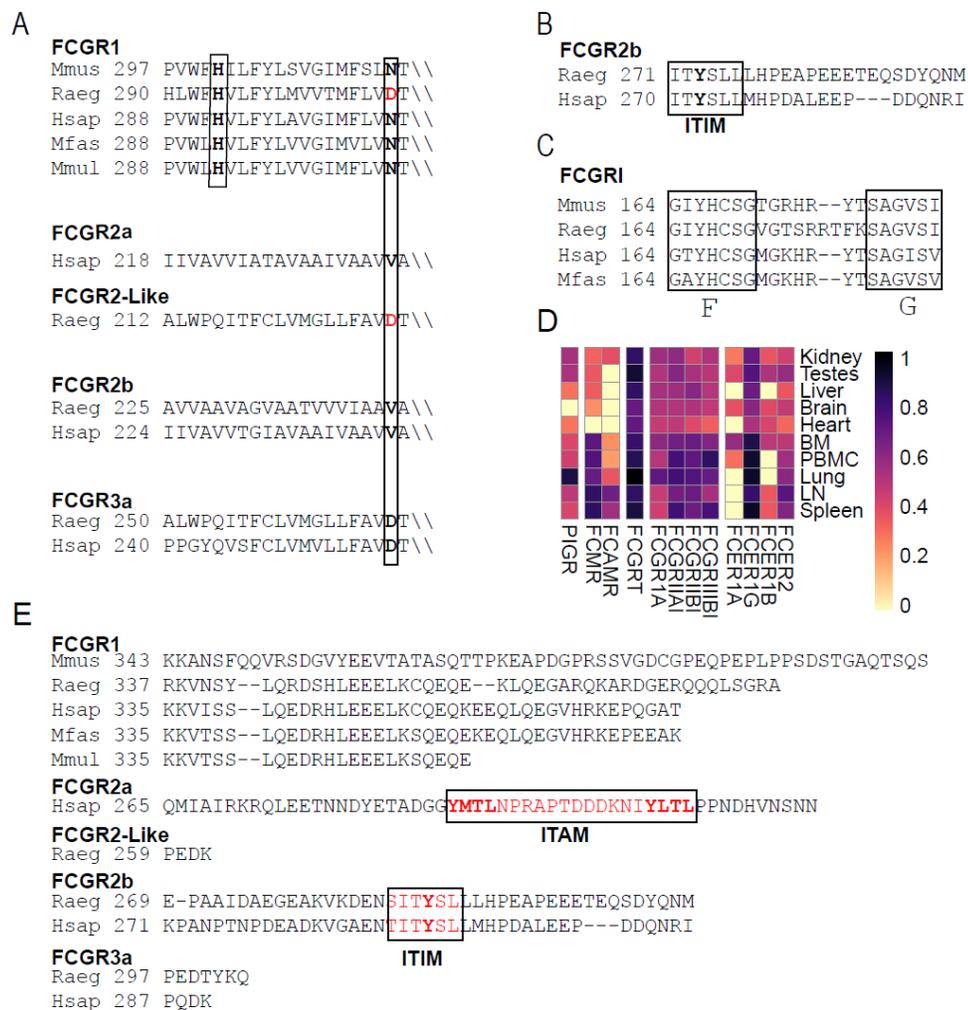


Figure 13 : FcR comparison between ERB and other mammals and expression.

A) Alignments of FcR cytosolic domains. ITIM and ITAM motifs are shown. Mmus (*Mus musculus*), Raeg (*Rousettus aegyptiacus*), Hsap (*Homo sapiens*), Mfas (*Macaca fascicularis*) and Mmul (*Macaca mulatta*). B) FcR transmembrane residues for γ -chain association and surface expression. C) Alignment of F and G loop of FcRs showing additional residues in the ERB loop. D) Tissue expression of FcRs in ERBs normalized to GAPDH for each tissue.

*Unpublished. Submitted to Cell.

Complement proteins and expression levels

Complement interacts with Ig FcR and participates in several of the Ig effector functions (reviewed in [98]). The three subunits of the C1q complex, C1qA, C1qB and C1qC, are present in Raegypt2.0 and are putatively functional. Complement proteins are primarily synthesized in the liver but a variety of other cells can also produce complement proteins. Interestingly, C1qA, C1qB, C1qC and C3 are ubiquitously expressed in ERB tissues supporting the important role these key immune mediators have bridging innate and adaptive immunity (Figure 14). While C3 is mainly expressed at the Liver and lung, the C1q proteins are mainly expressed at Spleen, lymph nodes, Liver and lung.

Identification and expression of pentraxins

Beyond Igs, many humoral components of innate immunity can interact with complement proteins and FcRs. We examined Raegypt2.0 and found no evidence of functional short pentraxins (C reactive protein (*CRP*) and serum amyloid P component (*SAP*)). Only a pseudogene *SAP* gene was annotated containing numerous premature stop codons. We inspected annotated *SAP* sequences from other published bat genomes and found they too contained numerous stop codons, suggesting that functional *SAP* was inactivated well before the emergence of extant bat species (data not shown). Although no short pentraxins were detected, we were able to identify several long pentraxins: pentraxin 3 (*PTX3*) and *PTX4*, neuronal pentraxin 1 (*NPTX1*) and *NPTX2* as well as the neuronal pentraxin receptor (*NPTXR*). We also identified the mucosal pentraxin (*MPTX*) gene which is annotated as a pseudogene in ERBs and humans. We were also interested in *DC-SIGN* as it is of the same receptor class as *FCεRII* and confirmed a human orthologue present in Raegypt2.0. *NPTX1*, *NPTX2*, and *NPTXR* were found to be highest expressed in the brain, while *PTX3* was found highest in the lung, *PTX4* highest in testes, and *DC-SIGN* in lymph node (Figure 14).

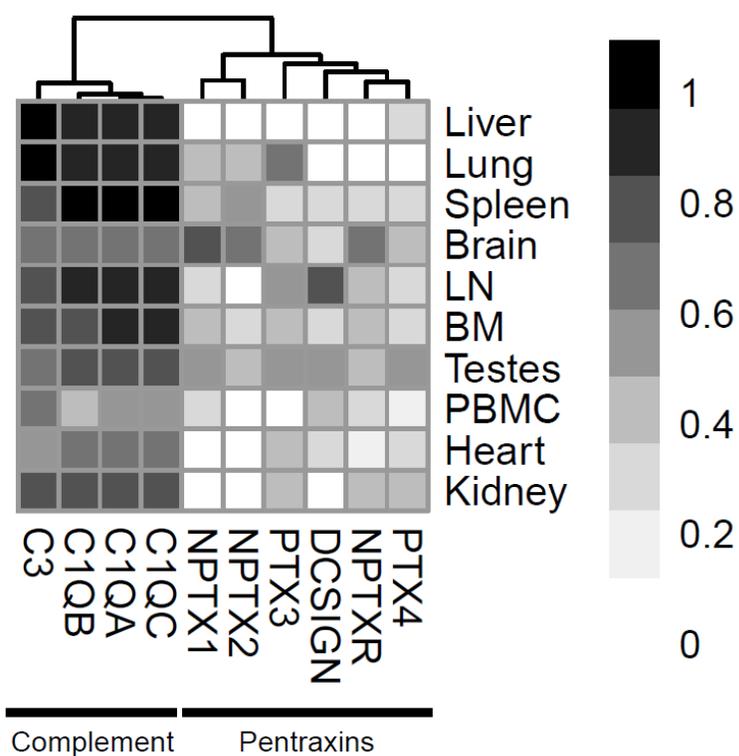


Figure 14 : Expression of complement proteins and pentraxins.

Heatmap of normalized log transformed reads for pentraxins and complement proteins C3, C1QB, C1QA and C1QC. Reads normalized to GAPDH for: liver, lung, spleen, brain, lymph node [179], bone marrow (BM), testes, peripheral blood mononuclear cells (PBMCs), heart and kidney.

**Unpublished. Submitted to Cell.*

Positive selection in IGHG, IGHE, and FcRs

We examined IGHC and FcR genes of ERBs to determine whether gene level evolution could contribute to unique functionalities. We first wanted to ascertain if receptors or IGHC genes were selected at a gene-level using the branch model [180]. We found *IGHG4* to be positively selected ($w=1.24044$, $LRT=0.0491$) while *IGHG1*, *IGHG2*, *IGHG3*, *IGHE1*, and *IGHE2* under neutral selection (Table 7). We further found no evidence of branch level positive selection in FcRs.

In addition to whole gene expansions, we were interested in positively selected residues in ERBs. To this end, we used the branch site and site models to identify any potentially selected for residues in FcRs and IGHC genes. No residues were found to be under significant positive selection in the FcRs, however, we did find *IGHE2* to have a positively selected site under the branch site model ($LRT=0.01157$), namely Ala³. We also found *IGHG4*, under the branch site model ($LRT= 0.000003176$), to have a positively selected Lys³⁵⁴ at a site that in all other species examined is a serine. These residues and their importance to ERBs is of interest for future studies.

Table 7 : Positive selection of *IGHG* and *IGHE* residues.

Gene	Omega	LRT
<i>IGHG1</i>	0.09324	0.040542015
<i>IGHG2</i>	0.30076	0.185150251
<i>IGHG3</i>	0.30063	0.675926128
<i>IGHG4</i>	1.24044	0.049072014
<i>IGHE1</i>	0.85691	0.862195588
<i>IGHE2</i>	0.85592	0.916705202

**Unpublished. Submitted to Cell.*

Discussion

Bats asymptotically harbor viruses that can cause severe disease in humans [8, 9, 140]. Even when some bat's antiviral mechanism(s) have been described (reviewed in [117]), a comprehensive understanding of the strategy used by bats to overcome these pathogens remains elusive. It has been proposed that some bat species have a potent innate antiviral response compared to primates, which allows for early control of virus replication and ultimately results in development of an effective adaptive immune response [131]. Alternatively, for other bat species, like ERBs, it was proposed an enhanced tolerance of infection rather than enhanced viral defense [85]. This model of viral tolerance is supported by studies of ERBs that result in protracted incubation and sustained viremia for up to three weeks after infection with limited inflammation post MARV infection [11, 43]. Here, we dissect components of the humoral immune response of ERBs that support this latter option resulting in an enhanced tolerance to viruses.

An expanded repertoire of IGHV known to respond to pathogens

So far, most of the efforts to understand the bat anti-viral response were focused on their innate immunity, mostly due to lack of adequate bat specific reagents to address humoral and cellular responses of the adaptive immunity. ERBs mount humoral responses to MARV infection and become refractory to future exposure to the virus [132]. Moreover, maternal immunity can be transmitted to newborns by breastfeeding [132], [87]. Surprisingly, anti-viral mechanism(s) of protection of ERB humoral immunity seems not to be mediated by viral neutralization [133]. To be able to better characterize the potential B cell repertoire diversity and begins to understand the adaptive humoral responses in ERBs, we first described and annotated the gene germline organization of the IGH locus resulting, to our knowledge, in the most complete description of any bat species to date.

We found representatives of all three mammalian IGHV clans (Figure 5), consistent with findings in pteropid bats [129]. Within these gene clusters, we found that the IGHV germline of ERBs contains numerous gene expansions relative to human orthologs (Figure 4). Advances in high-throughput DNA sequencing are enabling the characterization of the complex B cell repertoire for many animal species uncovering the existence of common signatures in the response to pathogens. The approach has shown that different individuals can respond to a given pathogen antigen with V(D)J rearrangements that share amino acid signatures and are encoded by common IGHV genes [181, 182]. We observed that several of the ERB IGHV genes present at the IGH locus, are associated with protective antibody responses against viral pathogens in humans (Table 8). For example, two human IGHV genes associated with MARV protection, IGHV4-39 and IGHV4-59 [79], are represented in the ERB IGH locus with one and eight genes copies, respectively. IGHV4-39 is the IGHV germline for mAb MR191, an antibody that competes with NPC1 for binding of GP and provides protection to MARV in NHPs [183]. Human IGHV3-23 was associated with a protective role against Ebola virus (EBOV) as it is part of the ADI-15878 V(D)J, a pan-filoviral antibody that targets the fusion loop of GP2 [184] as well as 2.1.1D7 and 2.1.7G7, neutralizing antibodies that target the GP1/2 interface [185]. ERBs have five functional copies of IGHV3-23 in the IGH locus. Similarly, human IGHV3.15 was also associated with protective mAbs against EBOV; ERBs contain a single copy of this gene [182]. The ERB also has two functional copies of IGHV1-69. Usage of IGHV1-69 had been associated with MERS-CoV [186, 187] Hendra and Nipah viruses [187], EBOV [188], and influenza virus [189]. Similarly, ERBs contain IGHV3-21 which has also been demonstrated to confer protection against dengue [190] and EBOV [185]. IGHV3-74, which is also expanded in the ERB IGH locus, has been shown to be associated with several viral pathogens, such as Andes Hantavirus [191], H5N6 avian influenza virus [192], influenza vaccination [193] and Epstein-Barr virus [194].

On the other hand, IGHV3-13, which has been associated with protective antibodies against EBOV in humans [185, 188], was not found in the ERB IGH locus. The observation that several of the IGHV gene associated with protection against diverse viral pathogens are present and some expanded in the ERB IGH locus, suggests that these animals might be equipped with an IGHV germline prone to generate antibodies that can bind to the viruses they host.

The pattern and location of ERB IGHD and IGHJ genes were found to be more similar to the particular organization observed in cattle than the canonical organization of most mammals [195]. The IGHV-IGHD-IGHJ-IGHD-IGHJ-IGHC organization at the ERB IGH locus is certainly intriguing. However, the most 5' cluster of IGHJ genes is composed by IGHJ2.1 that has a non-functional RSS (Figure 5) and no transcriptomic support; and IGHJ4.1 that was the least expressed IGHJ gene in ERBs (Figure 7), which suggests this IGHJ cluster might not be utilized in the V(D)J rearrangement.

An additional finding at the ERB IGH locus was the expansion of the IGHJ4 gene to six copies that varied by only a few amino acids in length and nucleotide composition. Even though in humans IGHJ4 is known to be the most commonly used J segment, the human IGH locus only contains one copy of this gene [196]. The duplication of the IGHJ4 genes might reflect its important role in V(D)J recombination in ERBs. The contribution of the IGHJ4 gene to protection has not yet been elucidated in other animals [197] so, the functional role of the different IGHJ4 copies in ERBs would need to be further addressed.

Table 8 : IGHV genes with known activities against viruses

IGHV	Functional	Pseudogenes	Truncated	Found in human responses	mAbs
VH1/OR15-1	9	1			
VH1-2	2	1			
VH1-8	4	1		EBOV	2.1.6C11
VH1-69	2	1		Influenza, MERS-CoV, Hendra and Nipah virus	m336, m101, m102
VH3-7	1			MARV	MR238
VH3-15	1			EBOV	5T0180, 1T0027, 3T0265, 2.1.1D5
VH3-21	1			EBOV	9.6.3A6
VH3-23	5			EBOV	9.6.3D6, 2.1.1D7, 2.1.1G7
VH3-33	3				
VH3-48	5			EBOV	2.10.1E6
VH3-71	2				
VH3-74	4	1			
VH4/OR15-8	1				
VH4-30	1	2			
VH4-38	2	1			
VH4-38-2	1				
VH4-39	1	1		MARV	MR191, MR201, MR209, MR232
VH4-59	8			MARV	MR111, MR114, MR186, MR241
VH4-61	1			MARV	MR208
VH6-1		1	1		
VH7-4	1				

**Unpublished. Submitted to Cell.*

Expansion of the *IGHE* - *IGHG* tandem repeat resulted in two functional IgEs and four functional IgGs in ERBs

To our knowledge, the ERB is the first species identified thus far to contain more than one functional *IGHE*. Although humans contain two copies of *IGHE*, one of the copies is a non-functional pseudogene [198]. The tightly regulated *IGHE* copy number in the mammalian IGH locus might reflect a balance of a cost/benefit between IgE function and expression level. It is known that although only one copy is present in humans, individuals with increased IgE expression can suffer conditions such as hyper-IgE syndrome [199]. The three IgE pseudogenes observed in ERBs might be a direct reflection of tightly regulated *IGHE* copy number. It is possible that as *IGHE-IGHG* tandem gene duplications occur, selective pressure inactivates the newly duplicated *IGHE*. Maintenance of multiple functional IgG genes is common and it could be that the *IGHE-IGHG* cassette contains sequences suited for unequal crossing over [200]. Regardless of the mechanism, two *IGHEs* remain functional in ERBs. The more likely explanation for retaining two functional *IGHEs*, as opposed to other mammals, is the diversification of the IgE functional response.

Gene duplications are not uncommon in the IGH locus, and occurred when IgY underwent a duplication ~220-300 million years ago gave rise to two distinct Ig isotypes, IgG and IgE, each performing a part of the functionality of the ancestral IgY. Additionally, isotype duplications result in variable subclass number which have been documented in several mammals. Thus, it is possible that a duplication of the *IGHEs* in ERBs could allow for such a diversification in which each copy performs a specialized function(s) that was previously fulfilled by a single IgE. An alternative but not mutually exclusive hypothesis is that the expansion and diversification of *IGHEs* results in tissue specific expression. We tried to discern between these two hypotheses by evaluating *IGHE1* and *IGHE2* tissue expression, by comparison of the primary sequence of the two functional ERB IgEs along

the mouse and human IgE, and by performing a structure-based multiple-level protein function predictions (using COFACTOR and GO terms). *IGHE2* is the only IgE gene 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 11). The tissue expression pattern of both IgE genes leads us to hypothesize that *IGHE2* could have diversified to perform an important role in the lung microenvironment, either complementing or augmenting IgA function, and/or act as a sensing antibody to modulate the IgG response in the lungs as is seen with many virus infections [201]. This notion was further reinforced by GO term analysis, which revealed that ERB IgE subclasses vary in predicted function (Figure 11). These include significant differences in prediction of the phagocytosis function. While *IGHE1* appears to mimic most of the other IgE species described to date, the function is missing in *IGHE2*. Conversely, unique GO terms, e.g. “response to other organism”, are associated with *IGHE2* but not indicated by *IGHE1*. Moreover, a residue known to be important for CD23 (FcεRII) binding, glu414 [202], 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.

In humans and mice, IgE class switching typically requires antigen, the presence of cytokines (IL4 and IL13), and the CD40-CD40 ligand T cell interaction [203]. However, another form of IgE, known as natural IgE, is produced without the need for antigen presentation or T cell co-stimulation [201]. In humans, natural IgE does not undergo somatic hypermutation and is directed against self-antigens, environmental toxins, and parasites. Natural IgE's specificity is dictated by the discrete combinatorial diversity in the V(D)J repertoire. Due to the expansion of pan-viral IGHV genes it is possible that ERBs natural IgE is capable of recognizing a broad range of pathogens without requiring co-stimulation, allowing for an innate-like adaptive response in which IgE rapidly detects a wide breadth of viral antigens [204]. Further support for IgE sensing of viral epitopes is

derived from the historical toxin hypothesis in which IgE acts as a sensor for detecting miniscule amounts of a particular protein; acting as a surveillance mechanism for the immune system [201]. Others have suggested that this could in turn enhance other antibody responses, such as IgG, stimulating a complex and effective adaptive response. Together, our data highlights the unique features of the ERB immune system and supports a diversification in IgE function. IgE1 resembling the function in other mammals IgE functions while IgE2 may have evolved a different function in the lung tissue, where it is primarily expressed.

ERB contains four IgG subclasses, identical to the number of IgGs found in human and mouse. Despite human and mouse containing the same number of functional IgGs, they are not functionally equivalent [137]. Thus, we do not anticipate that the ERB IgGs will share direct functional equivalence with any individual mouse/human IgG. Transcriptomic data suggests ERB IgG1 is the highest expressed subclass (Figure 11), especially in the periphery, lymphoid tissues and lung. ERB IgG4 is also principally expressed in periphery and spleen, but also in lymph nodes, bone marrow and lung. ERB IgG2 expression appears to be largely restricted to spleen, while IgG3 is very lowly expressed, which is congruent with our inability to detect IgG3 via proteomic analysis. Structure-based protein function predictions assigns each of the subclasses a distinct specialized function that could not be directly correlated with either a human or mouse IgG subclass. However, this prediction should be carefully interpreted after acquisition of functional data. For example, while ERB IgG2 and IgG3 are 99.8% identical at the amino acid level, differing by only one amino acid at the CH2 domain (IgG2: Thr¹⁸², IgG3: Ala¹⁸²), their GO terms predicted differences are significant (Figure 12). The predicted functional differences are likely explained by this single amino acid substitution (Thr¹⁸²Ala).

There are other structural features of interest in ERB IgGs. ERB IgG1 contains the longest hinge and a single CPRCP domain which is also present, although expanded, in

human IgG3. The long hinge of human IgG3 provides more flexibility to the two Fabs, but also reduces the half-life of the immunoglobulin due to increased susceptibility to protease degradation. Moreover, ERB IgG1 has a putative N-glycosylation site at the hinge (PVKYTCDRENGSCSCCSACPRCP), not observed in any of the other ERB nor the human IgG subclasses.

Another structural feature of ERB IgGs is the lack of the CXXC hinge domain in ERB IgG2, IgG3, and IgG4. ERB IgG4 contains a --RC in that region, while ERB IgG2 and IgG3 contain a –RS (Figure 10). This region of the hinge was associated with the ability of human IgGs to form disulfide bonds between both IGHs chains. Human IgG4, containing a CPSC domain in that region, was shown to undergo Fab-arm exchange and can exist as “half molecules”, where the IGH:IGL dimer is associated by non-covalent interaction with the other IGH:IGL dimer [205]. Because of this human IgG4 does not crosslink antigen and demonstrates bispecificity, most likely due to interchange of “half molecules” between different IgG4 [206]. The fact that ERB IgG2, IgG3 and IgG4 lack the hinge CXXC domain suggests three out of the four ERB IgGs might lack inter-chain disulfide bonds between IGHs, and potentially form bispecific half molecules, similar to human IgG4. However, it is important to note that all ERB IgGs have at least one cysteine at the hinge in a conserved domain: “PV(K/E)YTCDR(E/G)”. For ERB IgG2 and IgG3, this is the only cysteine, and its role in inter- or intra-disulfide bonds would need to be elucidated. This hypothesis, if confirmed experimentally, might have a great impact in how ERB humoral immunity works and affects their interaction with FcRs and complement. Human IgG4 contributes to anti-inflammatory properties, limiting its ability to form immunocomplexes and activate complement. If this is the case for all the ERB IgGs, except IgG1, another mechanism of anti-inflammation might be present in this animal species to control infection with low inflammation.

Fc functions vary between subclasses while FcRs have conserved motifs

Considering the lack of antibody-mediated neutralization observed *in vitro* [133], we decided to evaluate the key counterparts to Igs, the FcRs and humoral proteins that bind and enable different antibody-mediated biological functions. ERB FcRs resemble those found in mice, lacking the expansion observed in primates and specifically in humans [143], [145]. Consequently, while the inhibitory receptor FcγRIIB is well conserved, ERBs do not contain the expanded human activation FcRs (FcγRIIA and FcγRIIC) with Immunoreceptor Tyrosine-based Activation Motif (ITAM) at their cytosolic tail. Therefore, ERBs FcRs can only transduce and activator signal through the accessory signaling molecules (γ and β chains). ERBs also lack the GPI linked FcγRIIIB highly expressed in human neutrophils [207]. Moreover, the high affinity FcγRI expressed in ERBs has a particular feature in its trans-membrane sequence (Figure 13): While human, cynomolgus, and rhesus macaques as well as mouse FcγRI have a conserved TM motif with an Asn³⁰⁶ (in human: LAVGIMFLVNTVL), ERB FcγRI TM motifs contains an aspartic acid in that position (LMVV³⁰⁶TMFLVDTVF) which resembles the human FcγRIIIA (Figure 13) and FcεRI motif. This asparagine amino acid is crucial for association with the signaling γ chain and also renders FcR membrane expression independent of γ chain expression [208]. While Asp³⁰⁶ allows for interaction with the signaling γ chain, results in receptor membrane expression dependent on γ chain expression [209].

In humans, NHPs, and mice, FcγRI is the high affinity IgG receptor [210]. The D2 domain FG-loop of FcγR in human FcγRI (171MGKHRY176) is one amino acid shorter than the low affinity FcγRII and FcγRIII equivalent regions. An insertion of a valine in this structure (e.g. FcγRIII: 171MVGKHRY177) is sufficient to decrease 15 fold the FcγRI affinity for IgG [173]. A statistical analysis of this loop in different species showed that the length of the high affinity FcγR is six amino acids and that additional amino acids

drastically decreases affinity. Strikingly, ERB FcγRI has eight amino acids at the D2 domain FG-loop (171VGTSRRTF178) instead of the six observed in human, NHP and mouse FcγRI (Figure 13). Moreover, although ERB IgGs conserve the N-glycosylation site at Asn²⁹⁷; ERB IgG2, IgG3 and IgG4 had mutated one of the three critical binding contacts with FcγRI. While all the ERB IgGs retained the CH₂ FG loop Asp²⁶⁵ and the CH₂ FG loop, the lower hinge region (Leu²³³Leu²³⁴Gly²³⁵Gly²³⁶) is only conserved in ERB IgG1, but not in the other subclasses. IgG2, IgG3 and IgG4 have the second leucine of this motif substituted by a proline (Leu²³³Pro²³⁴Gly²³⁵Gly²³⁶). The mutation of the Leu²³⁴ is associated with the inability of the IgG to bind FcγRs [174]. Even though the mutation of the second leucine in ERBs was not functionally tested, based on data from mice and human, this modification is expected to have a direct impact on the ability of these three ERB IgG subclasses to interact with the FcγRI. This Leu²³⁴Pro substitution, together with the ERB FcγRI D2 domain FG-loop structure, strongly suggests that the FcγR affinity for the different ERB IgGs will be impacted. The threshold of interaction/activation of the ERBs IgGs/FcγRs are predicted to be lower than the human and mouse orthologues and is of high interest for further study. Since FcγRI has been implicated in inflammation in human and mouse [211, 212] the distinct features observed in ERB FcγRI might imply inflammation regulated by this receptor requires a higher threshold of activation.

Summary

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. In this study we comprehensively characterized the ERB IGH locus including 66 IGHV genes, 8 IGHD genes, 9 IGHJ genes, five IgE genes (two functional and three pseudo), four IgG genes (all functional), one IgM gene (functional), and one IgA gene (functional). We identified differential tissue expression of ERB Igs, potential functional differences,

and alterations in effector mediating proteins, suggesting that the functional IgE genes might be performing distinct function in different tissues. We identified features at the hinge domain of IgGs that might affect the inter-chain disulfide bonds, generating “half molecules” similar to human IgG4. We also identified the lack of the FcγR expansion observed in primates and features of FcγRI that suggest a dependency on the γ chain for expression and a longer FG-loop that would decrease the affinity for IgGs. Moreover, we found that ERBs lack functional short pentraxins in clear contrast with humans and mice. All of these findings support previous claims that ERB is biased to tolerance to viral infection with reduced inflammation.

Chapter 3: Mechanisms of diversity in the V(D)J repertoire of Egyptian rousette bats

Introduction

The immune system's ability to respond with high-affinity antibodies to a wide assortment of antigens is hinged on the generation of a diverse B cell repertoire dictated by the IGH germline and further antigen-driven selection and affinity maturation. The ability to recombine genes to create a great number of diverse binding units able to recognize different epitopes and elicit a potentially protective response to pathogens is paramount to adaptive humoral protection. Assembly of B cell receptors (BCRs) is achieved through recombination of germline-encoded gene segments: variable heavy (VH), diversity heavy (DH), joining heavy (JH) and constant heavy (CH) for the Ig heavy chains and variable light (VL) and joining light (JL) for Ig light chains. Our approach focused only on the heavy chain recombination as IGH IgSeq is enough to define and characterize the BCR repertoire and study the mechanisms of V(D)J recombination [213-217]. The portion of the antibody encoded by V(D)J is part of the Fab domain and is responsible for antigen binding. We were also interested in determining the mechanism(s) of B cell diversity which requires comparing BCR repertoire sequencing (IgSeq) data to reference genes (IGH germline).

Three mechanisms are used to varying degrees in different species to generate antibody diversity: 1) combinatorial diversity (CD), 2) somatic hypermutation (SHM), and 3) somatic gene conversion (SGC) [218]. Potential combinations of V(D)J for the heavy chain and VJ for the light chain define CD through recombination. SHM is the process in which single nucleotide polymorphisms are introduced and alter antibody specificity through affinity maturation while SGC is the process in which portions of the VH gene are homologously recombined with upstream VH regions to modify receptors. Both SGC and

SHM affect the complementarity determining regions (CDR) of the Fab domain which binds antigen. While they can occur across both framework regions (FWR) and CDRs, mutations are best tolerated in CDRs as FWR confer important structural features. In particular, the CDR3 region is considered integral to specificity as it has the additional diversity garnered by the imprecise combination of V(D)J which includes additions between VD and DJ recombinations (Figure 16).

Co-evolution with pathogens plays a critical role in shaping antibody repertoire diversity and specificity. At the germline, positive selection on germline-encoded VH segments is thought to result in preservation of VH segments utilized in protection against commonly encountered pathogens over a long period of host-pathogen evolutionary history [219]. Bats are natural reservoirs for a variety of RNA viruses and account for a majority of emerging diseases [220]. Despite having active viral infections, bats rarely display clinical symptoms [221]. Negative strand RNA viruses MARV, RAVV, and SOSV present asymptotically in ERBs [222] but cause a severe disease in primates that can be up to 90% lethal for MARV [132]. ERBs have been shown to produce a protective antibody response when exposed to the MARV, RAVV, and SOSV [21, 132], however, nothing is known about BCR generation in ERBs or their diversification mechanisms. To this end, I developed a molecular and bioinformatics protocol to recover, sequence, and analyze BCRs from ERB IgM (to represent naïve B cells) and IgG (to represent antigen experienced B cells more prone to SHM or SGC) (Figure 15). Here, I report the first molecular characterization of BCRs from ERBs.

In this study, we present evidence based on the genes identified in RaegyplGH3.0 that the potential CD of the V(D)J genes in ERBs is about half of what is possible in the human genome (Larson et al 2020, submitted). We also found ERBs IgM had elevated SHM when compared with human IgM. Interestingly, ERB and human IgG SHM rates were

not significantly different suggesting that the higher SHM in IgM is not due to missing germline genes or alleles as a reference. In Yangochiroptera, transcriptionally based studies indicated potentially 236 VH3 genes alone in the germline [223]. When calculating SHM, the authors were not able to distinguish between no SHM or so high SHM that the genes were indistinguishable from the described germline genes as they did not have a complete IGH locus to compare back to. They report some of these may be alleles, as was seen when mice were originally estimated to have more than 1000 VH genes which is now known to be a significant over estimate (BALB/c has only 101 functional VH genes) [223]. Based on our findings, it is possible that the transcripts from this order of bats also displays higher SHM and could account for the over estimation of germline genes. To confirm this hypothesis, a contiguous IGH locus for *Myotis lucifugus* would be required. The data presented here fits with what is currently known for other Yinpterochiroptera bats. The average number of IGHV genes described within this family of bats is ~60 genes, which we also report for ERBs and additionally placed them all on one locus. My research advanced the field as I describe the mechanism of BCR diversification which is a first in any Yinpterochiroptera species in addition to defining V(D)J combinations of another pteropid species [139]. This represents the first comprehensive analysis of any bat species BCR repertoire in that we describe potential CD and identify the likely mechanism of additional diversity. These data support unique features of adaptive humoral immunity in ERBs and could in part explain the ability of this species to overcome lethal pathogenic human viruses which could ultimately aid in the development of novel therapeutics.

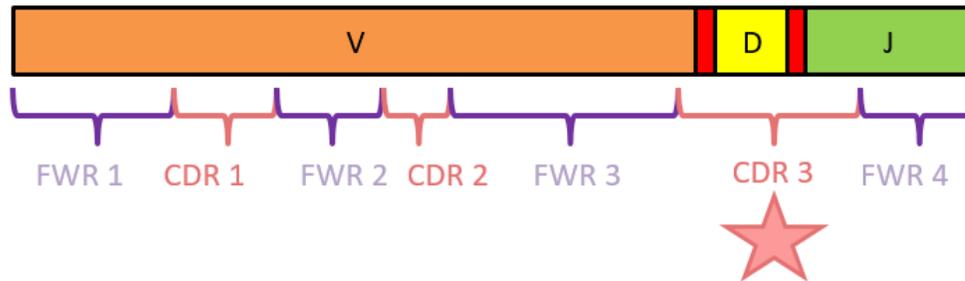


Figure 16 : Framework regions and complementarity determining region map.

The V(D)J region can be further described by framework regions 1-4 (FWR1-4) and complementarity determining regions 1-3 (CDR1-3) that play structural or binding roles, respectively. CDR3 includes gap regions (red) from the imprecise process of recombination that adds diversity in these junctional sites.

Methods

Animals

RNA was isolated from PBMCs from four healthy ERBs maintained at the CDC. Research was conducted under an IACUC approved protocol and complied with the Animal Welfare Act, PHS Policy, and other Federal regulations and statues relating to animal experiments. The work was conducted at an accredited facility by the Association for Assessment and Accreditation of Laboratory Animal Care, International and adhered to the principles in the Guide for the Care and Use of Laboratory Animals, National Research Council, 2011.

Generation of RNA transcripts from ERB

RNA from ERBs was provided from Jon Towner at the CDC. RNA was isolated from 2×10^6 PBMCs from individual bats using RNazol and ethanol. Quality was assessed with RNA tapestation. cDNA was generated using 5'RACE and superscript II under standard kit conditions. cDNA was cleaned up with 0.7x volume AMPure beads and underwent target enrichment PCR with primers specific to the 5'RACE and either the IgM or IgG constant exon 1. Samples were then library prepped on the Apollo system at 520bp. Samples were collected and cloned using a TOPO II blunt vector. Colonies that grew on Kanamycin plates were grown again overnight at 37°C shaking and gDNA was isolated. Insert size was confirmed via EcoR1 digest on a 2% gel. Products with appropriate size were sent for Sanger sequencing. Once libraries were confirmed to have rescued the amplicons of interest using IMG_T HighVQuest and manual annotation methods, libraries were sequenced on the Illumina MiSeq using a 600 cycle kit (2x301 cycles) following Illumina protocols. Samples were pooled and diluted to 2nM after qPCR and denatured with 0.1N NaOH along with 30% 2nM Phix spike-in. Paired reads were merged with FLASH and used to validate gene annotation. Sequences were submitted to IMG_T for

annotation, then re-annotated post germline definition and collapsed by UMI and 98% CDR3 for transcriptional analysis [224].

Identification of IGHV, IGHD and IGHJ genes

IgM sequences were filtered to remove reads that lacked both the constant primer and a 5'RACE, All VDJ transcripts passing filters were submitted to IMGT HighVQuest for annotation [225]. Gene nucleotide sequences were extracted and mapped against the BCR back to confirm functionality of annotated genes and identify likely locations of IGHD and IGHJ genes. If a transcript mapped to the BAC using the map to reference feature of Geneious, the region was manually inspected for recombination signal sequences, open reading frame, and conserved motifs. If all were found, the gene was annotated manually based on which transcripts were mapping to that gene. If multiple transcripts with different annotations mapped to one gene, the more prevalent annotation was kept and confirmed via phylogeny with other ERB and human genes.

Somatic hypermutation calculation

As most current somatic hypermutation analysis programs utilize human motifs in the determination of mutation rate, a novel method of analysis was required. The percent of SHM across ERB VH genes was quantified using an original bioinformatics pipeline implemented using BioPython, R, and the NCBI BLAST command line module. To recover V gene sequences, IgM transcripts from ERBs were sequenced and annotated using IMGT [225]. Annotated VH gene sequences were then aligned to the germline Ig genes using the BLAST algorithm. Both the number of nucleotide differences per transcript VH gene and the location of these differences were cataloged by SHMscript.py. The overall mutation percentage per VH family as well as across CDR and frame work regions (FWR) were visualized using R packages. Diversity metrics were assessed using IMGT output and VDJtools pipeline [224].

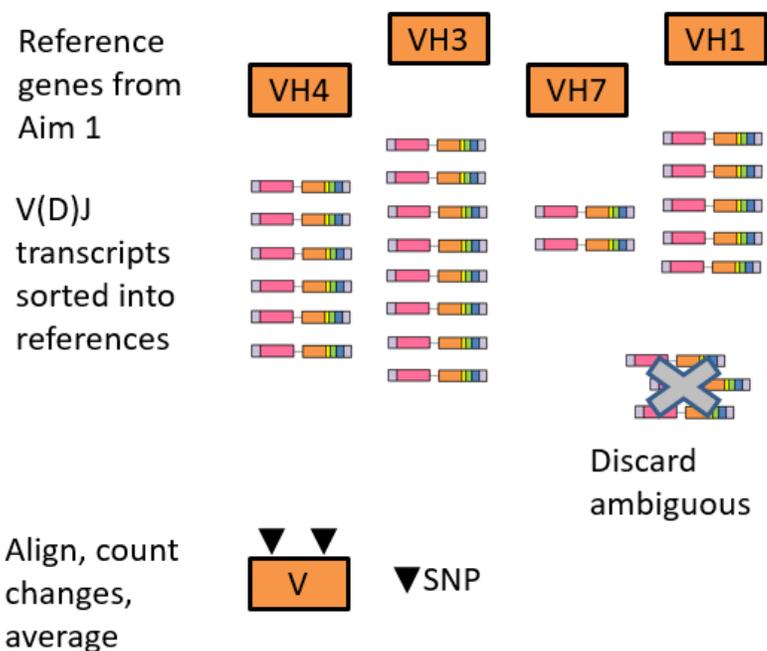


Figure 17 : SHM calculations schematic

V(D)J reads were aligned to a set of reference genes from RaegyptBCR3.0. Individual bins were maintained for each VH family and then the number of changes relative to the reference gene counted and averaged. Sequences that could not be placed with one of the references were discarded, and accounted for less than 1% of total reads.

Results

Combinatorial diversity potential and usage in ERBs

The antigen binding capacity of an antibody is initially set by the recombination of a V(D)J for the heavy chain and V(J) for the light chain. Initial diversity can be assessed by comparing the unique number of combinations a germline repertoire could form. To this end we focused on determining the composition of both genes in the new assembly for the IGH, RaegyptBCR3.0, and the previous assembly for IGL, Raegypt2.0. We recently generated one contiguous IGH locus for ERBs and meticulously annotated 66 VH, 8 DH, and 9 JH genes (Larson et al 2020, submitted). Based on these results, the IGH V(D)J CD capacity in ERBs is 4,032 ($56 \times 8 \times 9$) as ten VH genes are pseudo genes and would not produce functional BCRs. In humans, there are ~50 functional VH genes, approximately 27 DH genes, and six JH genes which could create ~12,150 unique combinations [226]. Suggesting that the base CD in ERBs is nearly a third of what is possible for humans. However- without proper annotation of the Ig light [227] chains VL and JL genes, the actual CD cannot be ascertained. The IGL loci are also highly repetitive and difficult to assemble, with the additional hurdle of being spread across not one locus but multiple loci. In humans these genes fall into two chains on two separate loci, kappa (IGLK) and Lambda (IGLL) [228]. Although the IGL loci remain uncharacterized, Raegypt2.0 has 158 IGLLV and 2 IGLKV genes. As humans have 71 IGLLV and 76 IGLKV genes, the diversity by combination then becomes 645,230 for ERBs and $\sim 3.5 \times 10^6$ for human- a substantial difference [218].

The relationships between the IGL isotypes across evolutionary time are less well understood than for IGH isotypes [229]. Raegypt2.0 has 29 putative IGLL constant genes (IGLLC) and no IGLK constant genes (IGLKC). IGLKC is missing in birds and potential Yangochiroptera, so to further confirm or refute its absence, IGLKC was ultimately identified by BLAST using another Yinpterochiroptera species IGLK (*Pteropus vampyrus*,

LOC105302304). We identified NW_01549288.1 (74240-74748) as the ERB IGLKC gene with 92% identity and 84% query cover, with 67% identity to the human ortholog of IGLKC. Despite a large potential IGLLC repertoire, IGLLC7 was the most common in transcriptomic data as well as proteomics (data not shown) [154]. However, these loci face the same issues as the IGH locus and reports based on published scaffolds alone could lead to artificially high counts. To this end, ERBs will be examined further to define the light loci and functional repertoire in another study.

Usage of V(D)J genes in ERBs

We wanted to assess the actual CD of IGH genes in ERBs so we developed a method to be able to sequence Ig transcripts containing IGH V(D)J sequences. The BCR repertoire sequencing was successfully applied to other mammals to recover of V(D)J transcripts [214, 230, 231]. We adapted this method, but specifically targeted ERB Igs (IgM and IgG). The rationale to focus on the BCRs of ERBs' IgM and IgG was to compare a more naïve subset (IgM, closest to the germline) to an antigen-experienced subset (IgG, produced after one or more rounds of class switching). Since few reagents exist for ERBs, this approach allowed us to lay the groundwork for future studies and comprehensively describe the B cell repertoire in ERBs. As ERB germline sequences are not currently in the IMGT database, we further conformed our data to the recent annotation of ERB germline V(D)J genes (Larson et al 2020, submitted). The most prevalent IGHV family used in V(D)J rearrangement was VH3, with approximately three genes contributing more than 10% of the V(D)J rearrangement in each for each individual bat (Figure 18). The next most predominate family was VH1, then VH4, and finally the single member family VH7 (Figure 18A). Variation between individuals was observed as was anticipated, but overall trends for family dominance were preserved as is seen in humans [217]. As each studied species appears to have a different composition of VH families as well as predominate VH family usage, these features are key in dissecting differences in response (Table 9). We

found that V(D)J rearrangements containing VH3 were predominate in ERBs, which is common in mammals but the extent to which the dominant family contributes varies between species which contributes to unique repertoires. (Table 9). Even though differences between individuals are observed, the overall expression allowed me to evaluate the most commonly expressed DH genes (Figure 18B). Based on the finding that ERBs have an expansion of JH4 genes, we wanted to determine if these genes were expressed differently. We observed preferential usage of four of the five JH4 genes in ERBs, suggestive of a strong selection of this gene family, with JH4.5 responsible for highest expression between subjects and isotypes (Figure 18C). While majority of transcripts were found to contain a JH4 gene similar to what is observed in humans, the usage differed between these highly similar genes varied greatly. In fact, one of these expansions, IGHJ4.1, contributes little to the repertoire, despite only varying by a few nucleotides (Figure 18C). Interestingly, the least used JH4 gene has two mutations in the FWR4 region, one of which changes the codon (V17I) (Figure 19). The second least used JH gene, JH4.2, also has a mutation in the FWR4 region (V15L). These mutations do not compromise the predicted open reading frame, but may confer a disadvantage to the fully assembled antibody which leads to decreased overall usage in JH4.1 and JH4.2. Multiple JH4 genes have not been reported in any other species, and overall the combinatorial diversity is significantly less than what is reported for Yangochiptera. This data supports the large differences in these species and that features prevalent in Yangochiptera may not be prevalent in Yinpterochiroptera [139].

We found that all but one functional VH gene was used (VH6), all eight DH genes, and seven of nine JH genes- however usage varied between individuals as was anticipated. Our recent work on RaegyplGH3.0 indicated impaired heptamers for the two JH genes not seen, supporting their lack of expression. This decrease in usable segments makes the actual CD closer to ~3,472 which is almost a third to the CD in humans.

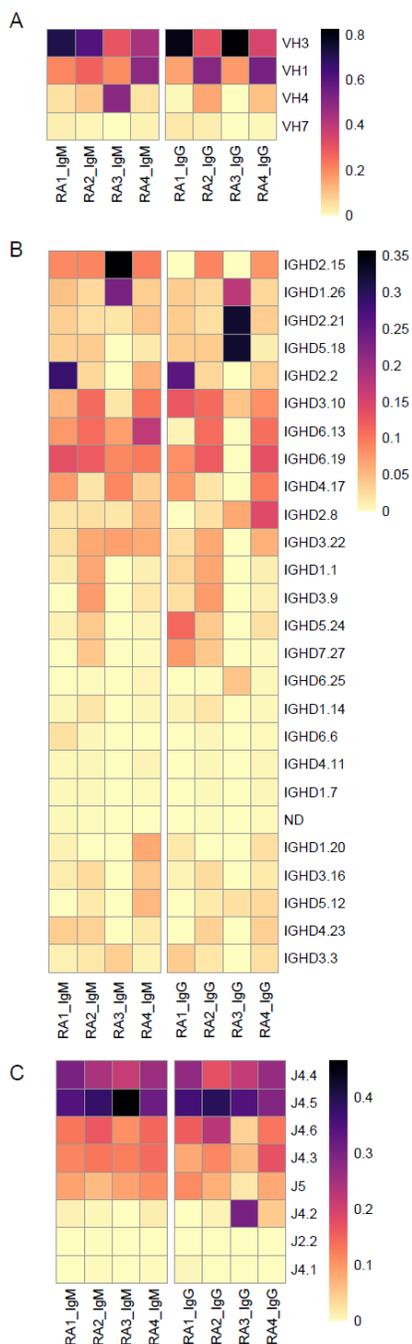


Figure 18 : V(D)J gene usage across four ERBs for IgM and IgG.

V(D)J sequences were annotated using IMGT High-VQuest then reannotated with an in house script to follow with the new annotations from RaegyptBCR3.0 not yet in the database. For four ERBs IgM and IgG are shown for A) IGHV, B) IGHD, and C) IGHJ.

**Unpublished. Preparing for Journal of Immunology.*

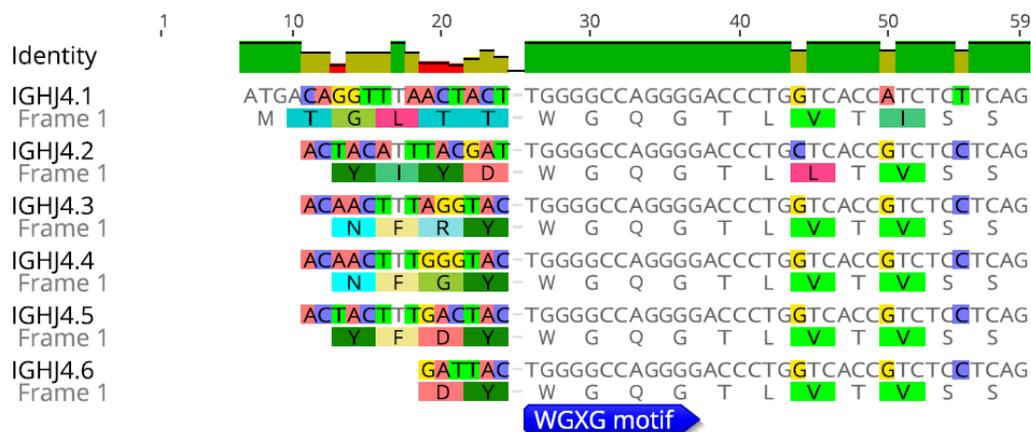


Figure 19 : IGHJ gene alignment of WGXXG motif.

Alignment in Geneious of germline IGHJ genes from ERBs. The framework region in JH genes begins with a structurally conserved WGXXG motif present in all functional JH genes. Variation is observed in the 5' end that falls within CDR3 and the least used JH4 gene, JH4.1, has a mutation that changes the conserved 3' valine.

**Unpublished. Preparing for Journal of Immunology.*

Table 9 : VH families between species

Species	Functional VH Families (#)	VH Families Transcribed	Highest transcribed
<i>B. taurus</i>	1	VH1	VH1
<i>S. scrofa</i>	1	VH1	VH1
<i>C. familiaris</i>	3	VH1, VH3, VH4	VH3
<i>E. caballus</i>	3	VH1- VH3	VH1
<i>R. aegyptiacus</i>	4	VH1, VH3, VH4, VH7	VH3
<i>P. alecto</i>	5	VH1-VH5	VH1
<i>M. lucifugus</i>	5	VH1- VH5	VH3
<i>H. sapien</i>	7	VH1-VH7	VH3
<i>M. mulatta</i>	7	VH1-VH7	VH3
<i>C. porcellus</i>	3	VH1- VH3	VH3
<i>M. musculus</i>	14	VH1-VH14	VH1

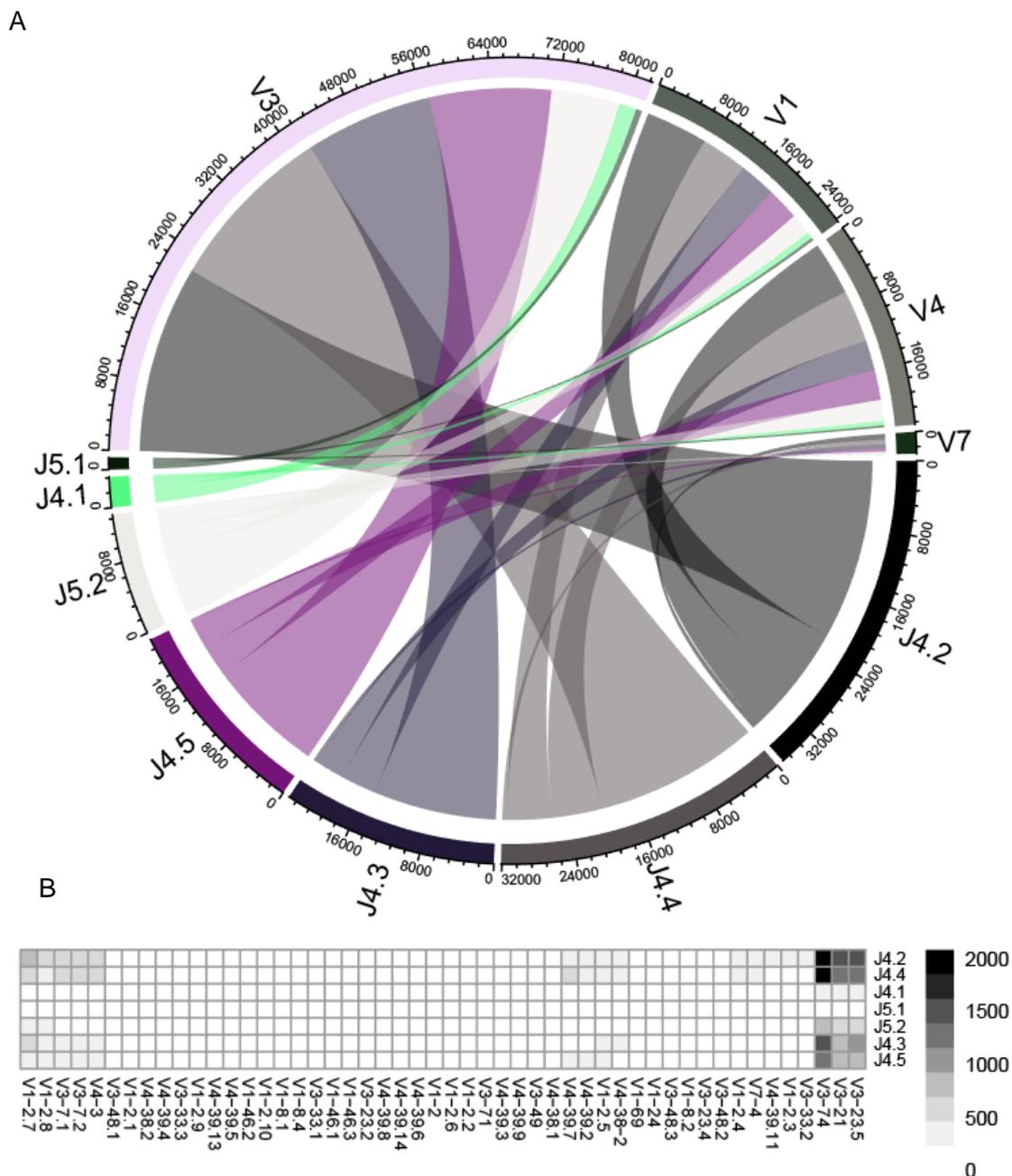


Figure 20 : Representative V-J usage chord diagram and Heatmap.

A) Chord diagram illustrates which V genes pair with which J genes in repertoire. B) Heatmap of individual VH genes pairing with individual JH genes. The most common combination is that of V3-74 and J4.2.

**Unpublished. Preparing for Journal of Immunology.*

CDR3 composition in ERB repertoire

As the CDR3 is known to play a critical role in antigen binding, so much so that the exact same antibody with only changes to CDR3 recognize different antigens [232]. We sought to characterize common features of this region of ERB antibodies as it was previously uncharacterized. The mean length of IgM CDR3s was 48.5+/-2.6 nt while the mean length of IgG CDR3s was 45.4+/-12.8 nt, similar to the mean sizes observed in humans [196]. The CDR3s were high in aromatic residues like tyrosine and tryptophan which are indicative of broad binding capacity (Figure 21). This capacity could be beneficial for antigen recognition but would also pose risk for autoimmunity from self-reactivity [233]. Overall, the most abundant amino acids used in ERB IgM CDR3s were tyrosine (Y, 12.33%), glycine (G, 10.96%), and tryptophan (W, 10.59%) while IgG CDR3s were highest in tyrosine (Y, 12.85%), tryptophan (W 9.51%), and alanine (A 9.17%). Overall there was not a significant difference between ERB CDR3 amino acid composition between IgM (representing a more naïve B cell population) and IgG (representing antigen experienced B cells) (Table 10).

An important facet of CDR3 composition comes from the junctional diversity regions between VD (Gap 1) and DJ (Gap 2) (Table 11). The imprecise mechanism of recombination allows for greater diversity and is primarily mediated by Palindromic (P) and Non-templated (N) additions. Based on IMGT annotations, we found ERB IgM Gap 1 and Gap 2 to be 8.70 nt and 9.40 nt while ERB IgG Gap 1 and Gap 2 were 7.99 nt and 10.21 nt, respectively (Table 11). While the average gap length is similar to humans, the range appears to be larger for ERBs in both IgM and IgG which could offer additional diversity in the potential BCR repertoire of ERBs.

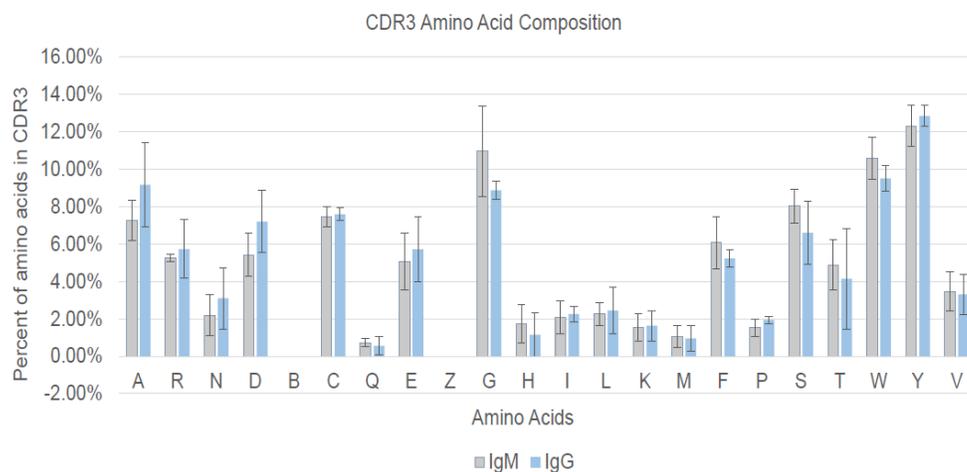


Figure 21 : CDR3 amino acid composition of ERB IgM and IgG

IMGT annotated CDR3 regions were examined for amino acid combination between IgG and IgM of ERBs. Average of four ERBs shown with standard deviation error bars.

**Unpublished. Preparing for Journal of Immunology.*

Table 10 : Amino acid composition of CDR3s in ERBs

AA	IgM	IgG
A	7.27%	9.17%
R	5.26%	5.74%
N	2.20%	3.10%
D	5.43%	7.20%
B	0.00%	0.00%
C	7.46%	7.59%
Q	0.73%	0.56%
E	5.06%	5.73%
Z	0.00%	0.00%
G	10.96%	8.87%
H	1.75%	1.15%
I	2.10%	2.26%
L	2.26%	2.45%
K	1.53%	1.63%
M	1.04%	0.95%
F	6.09%	5.24%
P	1.53%	1.93%
S	8.04%	6.61%
T	4.90%	4.14%
W	10.59%	9.51%
Y	12.33%	12.85%
V	3.47%	3.32%

Table 11 : Gap lengths of CDR3 from ERBs.

	IgM	IgG
Gap 1	8.7+/-1.56	7.99+/-2.45
Gap 2	9.4+/-1.53	10.21+/-1.51

Somatic hypermutation in ERBs

Although some have speculated the major BCR diversity mechanism in Yangochiroptera is either extraordinarily high SHM or SGC, it remained undetermined for any bat species [223]. To verify or refute the hypothesis that this could be a general trait for Chiroptera, we studied the BCR diversity mechanisms in ERBs. We compared repertoire VH genes to germline VH genes annotated from Raegypt3.0BCR for single nucleotide changes (SHM) as well as stretches of changes indicative of homologous recombination with upstream VH genes (SGC). Our analysis supports SHM as the major diversity mechanism for ERBs, similar to other mammalian species. These alignments and subsequent calculations followed the expected mammalian pattern of higher mutations in CDRs than FWRs (Figure 22). Overall, IgM VH genes varied by 5.21 \pm 2.08% to reference genes whereas IgG VH genes varied 11.13 \pm 1.17% to reference genes. In humans, IgM variation between transcripts and genomic genes is ~3% while IgG is around ~8% [234]. To verify our methods, we repeated this process on IgM and IgG from human and found 2.59 \pm 1.37% and 10.87 \pm 3.54% (data not shown). A student T-test confirmed variation in VH genes of the repertoire to VH reference genes of IgM was significantly elevated in ERBs at a $p < 0.001$ (Figure 23).

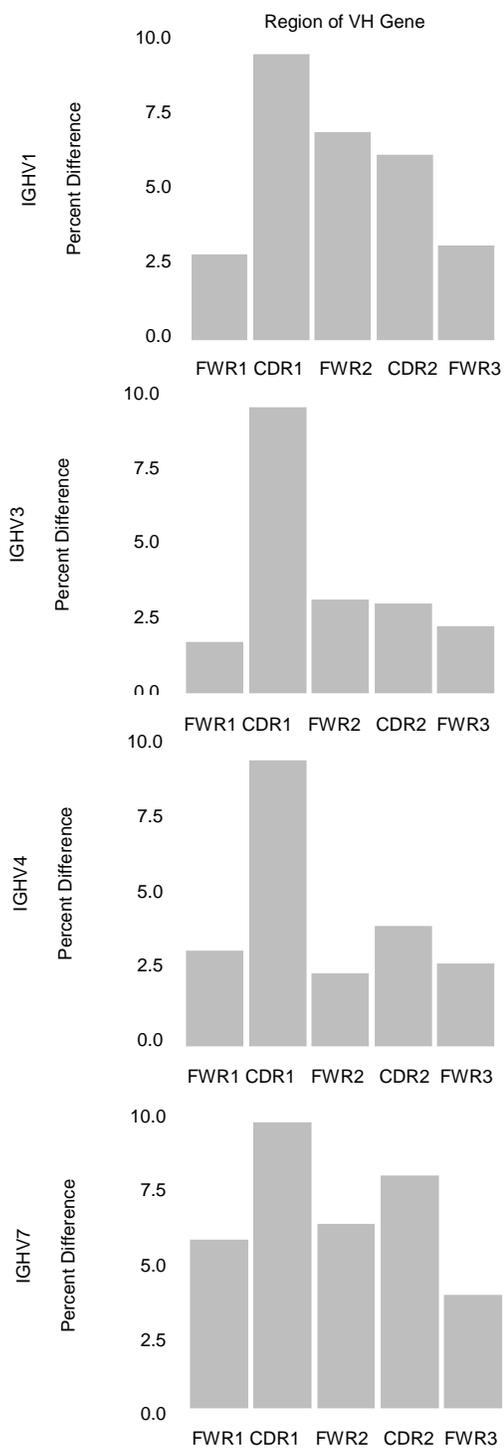


Figure 22 : CDR3 mutations across VH families.

Location of mutations identified by SHM script in FWRs and CDRs of ERB IgSeq for each VH family.

**Unpublished. Preparing for Journal of Immunology.*

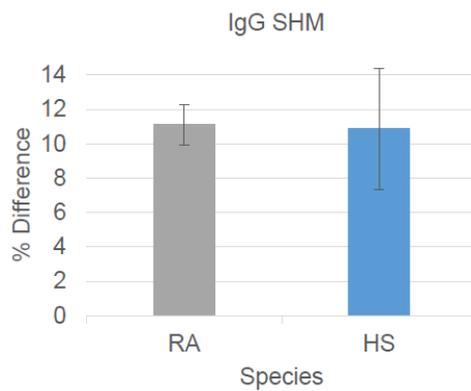
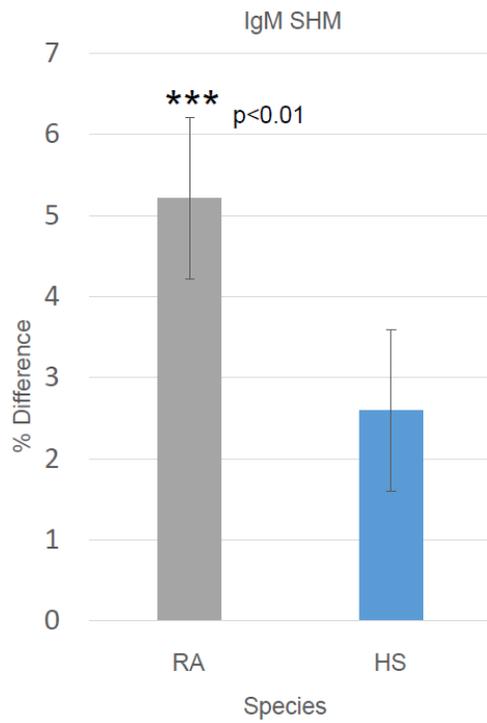


Figure 23 : Percent difference from transcripts to reference VH genes in ERBs (RA) and humans (HS).

Percent difference relative to reference genes for IgM (top) and IgG (bottom). Significance tested by student T-test.

**Unpublished. Preparing for Journal of Immunology.*

Discussion

Lower combinatorial diversity in ERBs than humans

The ability for ERBs to overcome MARV, RAVV, and SOSV infections without symptoms may rely in part on a protective antibody response. A species' antibody repertoire is built upon the genes encoded in the germline and so the number of unique combinations possible is a critical first determinant of overall diversity. The IGH locus of ERBs contains 66 VH, 8 DH, and 9 JH genes (Larson et al 2020, submitted). This plus the comparable IGL genes suggest a potential for less than a third of the combinations for human V(D)Js. While the number, composition, and usage of these genes varies between each studied species, particular genes have been linked with a protective response to specific pathogens in different species [235]. The presence (and expansion) of these “protective” genes in the IGH germline would be beneficial for the animal frequently exposed to those potential pathogens. Moreover, a wider and more diverse B cell repertoire would be favored, as it would increase the chances of having a beneficial BCR within the B cell population [236]. Beyond the composition of these genes in the genome, we wanted to know the usage and applied IgSeq to recover and examine V(D)J transcripts. We found that majority of the functional genes were used between the individual animals tested, similar to what is seen in other mammals [216].

CDR3 composition supports polyreactive antibodies

To examine if ERBs are capable of a more broad antigen binding that could compensate for diminished CD, we characterized the amino acid composition of the CDR3 region. The CDR3 of antibodies is the junction between the IGHV, IGHD and IGHJ genes, contains the VD and DJ gaps, and is regarded as the most important region for antigen binding and recognition [236]. We found evidence that supports polyreactive antibodies which have CDR3 regions high in aromatic residues like tyrosine and tryptophan. These

residues interact with diverse antigens by forming a “sticky” antigen-binding site [237, 238]. The discrete differences in CDR3 composition between species may indicate a bat species specific preference in CDR3 amino acid composition [139]. For example, differences exist between germline VH segments (VH1, VH3, VH4, VH7 in ERB vs VH1, VH2, VH3, VH4, VH5 in *P.alecto*) that skew amino acid preference and would imply differences seen between ERB and other Yinpterochiroptera. This difference could also be heavily influenced by antigen exposure- since the ERB examined come from colony at the CDC kept in biocontainment while the *P.alecto* studied were wild-caught animals, (which have an unknown history of viral exposure). Although broad antigen permissibility is often a hallmark of autoimmune diseases, a decrease in arginines in the CDR3 may counter-act the broad capacity of the binding pocket as arginine is correlated to autoreactive antibodies in humans [236].

The imprecise recombination that results in P and N additions are mediated by terminal deoxynucleotidy transferase (dTNT). P nucleotides occur because Artemis cleavage is often asymmetrical, resulting in a short and long strand that requires extension. N nucleotides are random additions of 2-20 bp. This leads to over 66% of JH region rearrangements resulting in nonfunctional antibodies [239]. While we found that ERB gaps were similar to humans on average, there was a wider range in gap length which could afford more diversity in final rearrangements.

Evidence of higher somatic hypermutation in ERB IgM than humans

The BCR diversity mechanism in ERBs was unknown. We compared repertoire VH genes to germline genes which supports SHM as the major diversity mechanism. The finding that ERBs have elevated SHM in IgM could mean ERBs are capable of producing more broadly reactive antibodies. It has been well documented that high levels of SHM and long CDR3s are common for broadly neutralizing antibodies (bnAbs) [240, 241], however, very low SHM has also been associated with bnAbs highlighting the gradient of

possibilities in the antibody response [185]. Although it has been shown that ERBs don't generate neutralizing IgG antibodies to MARV [133], they may in fact be generating neutralizing IgM, IgA or IgE antibodies which are important for mucosal exposure to viruses like influenza [242]. Higher SHM could also support that ERB antibodies are able to tune a response quickly without multiple rounds of affinity maturation [243]. Indeed, cattle which have a propensity to generate rapid broadly neutralizing antibodies are known to have longer CDR3s and different patterns of SHM [244]. Although ERBs appear to have less CD than humans, they may compensate for limited CD diversity through other diversity mechanisms as is suggested in chickens, which also have diminished capacity for CD relative to human [236]. It is possible that IgM plays a larger role in the ERB response to pathogens, and thus the higher SHM allows for a wider breadth of response than what is seen in humans. This mechanism may allow for a rapid, pan-pathogen response that protects ERBs from overt symptomology sooner than waiting for class switching to IgG.

Summary

Bats represent a diverse and species-rich group of mammals that have evolved distinct characteristics since their divergence from other mammals. Despite the role they play in viral spillover to other species and the apparent lack of overt symptoms to pathogens that cause human pathology, little is known about their anti-viral response and in particular their BCR repertoires. To identify genes involved in the BCR repertoire of ERBs, we developed and demonstrated the first ERB IgSeq methodology from bench to analysis. Understanding the immune response of ERBs to MARV and other pathogens could aid in development of effective medical counter measures but was previously hampered by lack of gene annotations, reagents, and protocols. We have defined the base repertoire of ERB V(D)J's as well as identified the likely mechanisms that contribute

to receptor diversity. Interestingly, SHM varied between gene families with the largest gene family being closest to the overall mutation rate, similar to what is seen in humans while families like VH7 saw very little SHM. In conclusion, using high-throughput sequencing of ERB BCRs we were able to describe the overall repertoire of this important reservoir host. The identification of sticky amino acids in the CDR3s suggest a broadly-binding capacity of ERB BCRs not previously reported. This combined with the decreased potential heavy CD leads us to suspect a broad-response although it is not improbable that an expansive IGLV repertoire might rescue the CD overall. These data support the hypothesis that ERBs may be protected from MARV and other viral infections by a broad humoral response that elicits protection without overt inflammation seen in other mammals. This may be in part due to the expansion of novel Ig isotypes described previously, as well as the potential natural antibody repertoire. To our knowledge, this study provides the first description of ERB BCR repertoire and the first evidence of a B cell diversity mechanism in a Yinpterochiroptera species. This study lays the groundwork for future studies to decipher the humoral response of ERBs to MARV, RAVV, and SOSV and provides key insight into the generation of the antibody response by ERBs.

Chapter 4: Profile of *Marburgvirus* epitopes ERB antibodies recognize

Introduction

Monoclonal antibodies have proven successful as therapeutics in recent EBOV outbreaks [52]. As a strong humoral response is linked to vaccine efficacy for filoviruses, and convalescent serum from EBOV and MARV vaccinated NHPs provides protection by passive transfer, discovery of novel protective antibodies is a bright course of pursuit [80]. While no FDA-approved treatment currently exists for MARV, many mAb treatments are being used for EBOV [245]. However, recent findings suggest EBOV immunodominant epitopes alone might not be adequate vaccine targets to control viral infection. This is also observed in other viruses, such as influenza virus and HIV, in which the immunodominant antigenic target reduces the efficacy of vaccination while other antigenic regions elicit broad-spectrum neutralizing antibodies [128]. Recently, we assessed the polyclonal response to EBOV at an epitope level using a technique known as Domain Programmable Arrays (DPA), a phage display, high-throughput serological platform for epitope mapping (Figure 24). This technique takes advantage of the classical phage display system, a comprehensive detection array of synthetic oligos, and a bioinformatics metrology to comprehensively identify domain variants within epitopes [128]. We further validated this technique in another bat species, Schreiber's Bent-Winged Insectivorous bats, for their response to a lesser known filovirus, Lloviu virus (LLOV) [36].

To this end, applying this technique to the reservoir host of MARV might identify unique regions of recognition that primate antibodies do not detect. If so, those epitopes could offer significant protection and in theory be generated and given as therapeutics. We have optimized the DPA technique for use with ERB sera and tested multiple naïve and MARV challenged ERBs as well as naïve and MARV GP vaccinated NHPs. We found that ERBs recognize a wider breadth of MARV GP, which crosses the GP1/GP2 cleavage site and could potentially offer protection by controlling viral spread.

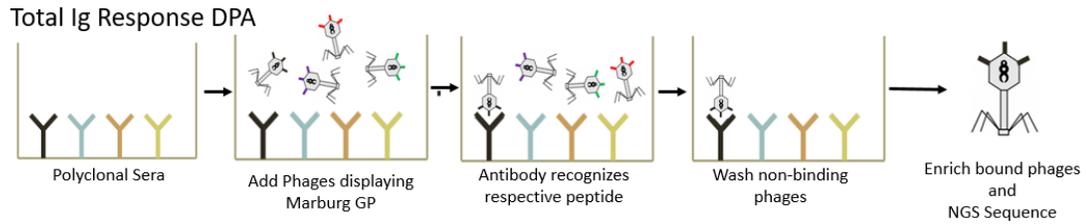


Figure 24: Domain Programmable Array overview.

The DPA assay is set up similar to an ELISA in which sera is added to a 96-well plate and incubated with phages displaying filovirus oligos that span the entire coding region. Bound phages are subsequently recovered and sequenced and a bioinformatics pipeline used to normalize the data.

Methods

Animals

Serum was isolated from ERBs on day 56 post two sequential challenges with MARV (days 0 and day 28) by our CDC collaborators. Research was conducted under an IACUC approved protocol and complied with the Animal Welfare Act, PHS Policy, and other Federal regulations and statues relating to animal experiments. The work was conducted at an accredited facility by the Association for Assessment and Accreditation of Laboratory Animal Care, International and adhered to the principles in the Guide for the Care and Use of Laboratory Animals, National Research Council, 2011.

Domain programmable array screening

Protein-coding domain sequences from all filovirus strains deposited into GenBank as of December 2016 (EBOV, SUDV, TAFV, BDBV, RESTV, MARV, RAVV, and LLOV) were retrieved, and coding sequences split into 30 amino acid (aa) peptides that tiled every seven aa (Figure 25). Identical sequences were eliminated and the remaining 12,735 codon-optimized peptides were synthesized on high-density microarrays, cloned into a T7 phage display system, and deep-sequenced to confirm successful display. A 96-well plate was blocked with 10%BSA in 0.1N NaHCO₃ for two hours at room temperature. Then sera was added at a 1:500 dilution in PBS in triplicate and incubated at 4°C overnight. Wells were washed with TBST eight times and 100uL/well 2M urea was added to recover phages. The recovered phages were lysed and cloned oligonucleotides amplified via PCR (T7 forward primer: GGA GCT GTC GTA TTC CAG TC / T7 reverse primer: CCC CTC AAG ACC CGT TTA GAG GCC C) using triplicate primers with a 5' staggered end. The PCR protocol followed T7 Clontech specifications. Triplicates were pooled and cleaned using Zymo-5 column clean up beads and 70% ethanol. Samples were resuspended in 20uL Illumina buffer and size checked on the DNA tapestation.

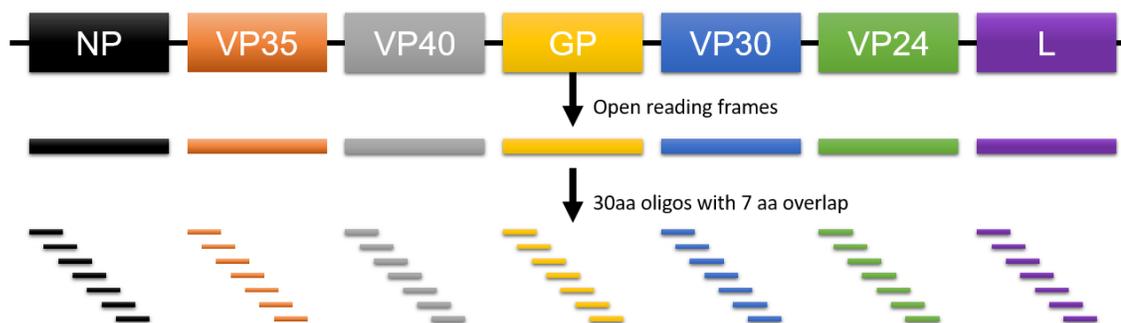


Figure 25 : Construction of the DPA phage oligo library.

All open reading frames were pulled from NCBI for current filoviruses. Sequences were then chopped into 30 amino acid sequences tiling the ORFs by seven amino acids to ensure full coverage.

Sequencing

DNA was quantified by Nanodrop and 100 ng used for an Illumina 220bp library preparation on the Apollo 324 (Wafergen) system using the Prep X kit (Wafergen complete IIMN DNA library kit). Libraries were then enriched using the library amplification kit (KAPA) and a post-enrichment clean-up done on the Apollo 324. Libraries were then pooled to 2nM and sequenced on a MiSeq DNA sequencer instrument using a 600 cycle kit (2x250) with 20% PhiX.

Bioinformatics

An in-house developed pipeline was developed previously that performs the following: Array Description, Input Randomization, Removal of Duplicates and Read Mate Correction, Read Cleaning, and Expression Analysis. The DPA analysis pipeline can be downloaded at https://github.com/kygarcia/DPA_Analysis_Pipeline. The reporting module creates an HTML and EXCEL based output that was reworked for this project to highlight strain of interest (MARV) and fix minor display issues. Each library was run in triplicate and T-tests and Chia Square performed to remove peptides that fail either if $P > 0.05$.

Western Blot to Confirm Immunodominant epitopes

Non-overlapping peptides were selected from DPA results and synthesized including a glycine linker and biotin. Peptides were diluted to 1mg/mL water immediately before use and stored at -20C. Peptides were used as antigen in the WES (Protein Simple) system. Assay conditions were done following the manufacture guidelines (Protein Simple). Biotinylated peptides were mixed with 5X Fluorescent Standard, DTT, and 1X Sample Buffer (WES Reagents) for a final concentration of 0.8 μ g of protein, and loaded and run in the capillary system. Antigen was detected using the serum diluted at 1/100, 1/500 or 1/1000 in PBS, and further developed using the isotype specific secondary

antibodies. Anti-rabbit HRP Antibody and the provided peroxidase and Luminol-S reagent were used to develop the assay. Results were analyzed using the Protein Simple software Compass. Western-blotting was not performed quantitatively, and instead scored only as “positive” or “negative.” An anti-biotin HRP antibody (CAT) was used as a positive control for peptide sizing, and negative controls included a well without antigen, a well without primary, and a well without secondary.

Results

Profile of all MARV Epitopes ERB Antibodies Recognize

Four ERBs that were challenged with MARV on day 0 and day 28, and five ERBs that were mock challenged with media alone were all bled on day 56. Serum was bound to polystyrene plates and then the phage library added and incubated. After multiple washes the bound fraction was recovered and sequenced. We found that the antigen recognition breadth was increased in ERBs after one challenge of MARV and even more so after a second dose of MARV (Figure 26). We noted that this individual ERB's recognition crossed the GP1/GP2 cleavage site so we repeated with different animals to confirm. We found that there was a conserved public response recognizing much of MARV GP which indeed did cross the GP1/GP2 cleavage site (Figure 27). We detected recognition at the GP1/GP2 cleavage site consistently in ERBs. However, we did not find clusters in all four ERBs despite repeating the experiment and altering conditions. This could be due to a lack of seroconversion in this subject, so to confirm exposure we conducted an ELISA. We found that all challenged ERBs had positive titers while all control animals had negative titers (Figure 28). Unfortunately, this was the last of this experimental sample and could not be explored further.

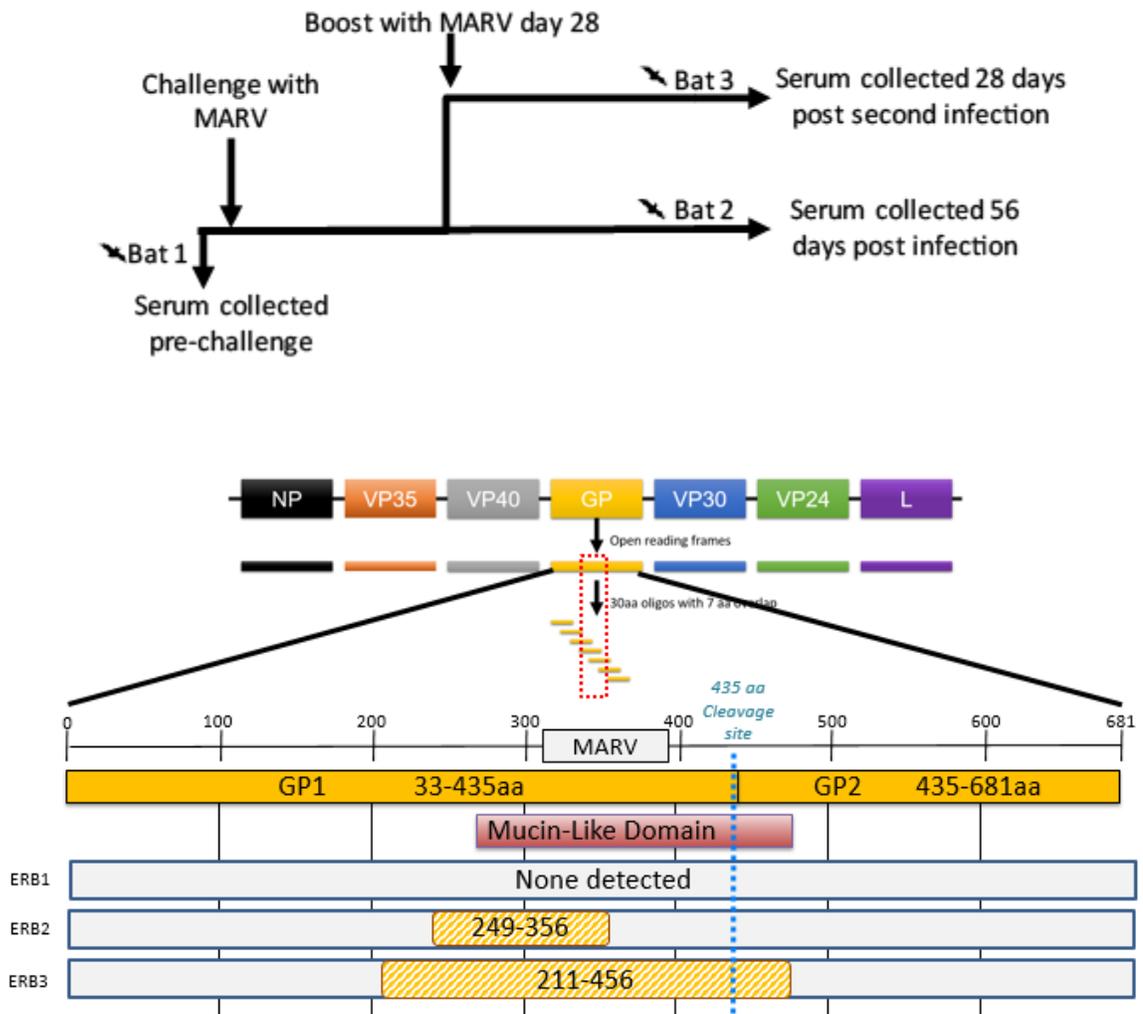


Figure 26 : ERB response to MARV after a single, or double challenge.

Three sera samples were received from the CDC. One which had not been challenged, one that had been challenged once, and one that had been challenged twice [246]. All were analyzed for their recognition of MARV (bottom). Bars represent span of oligo recognition.

**Unpublished. Preparing for Viruses.*

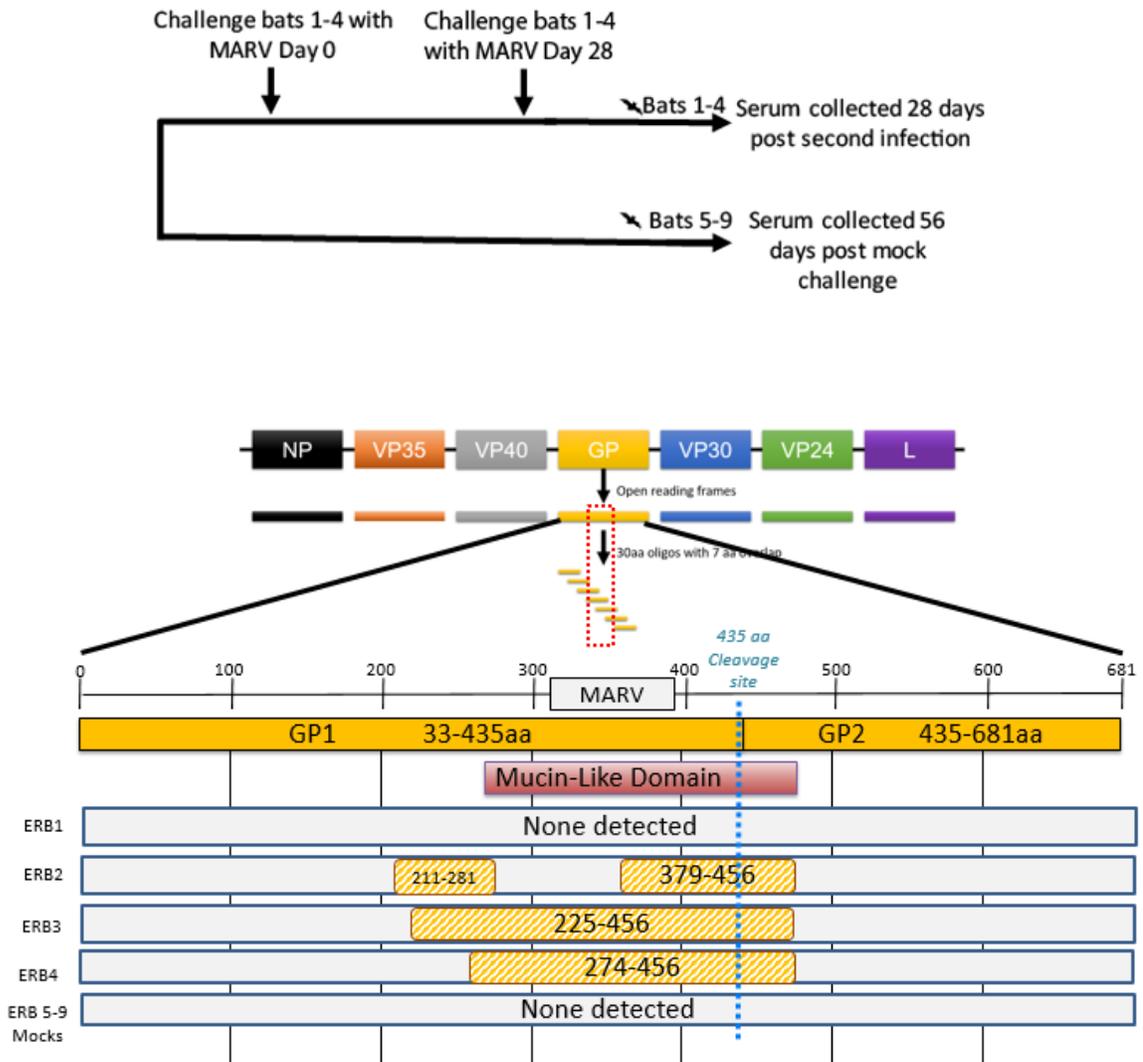


Figure 27 : ERB responses to MARV after a double challenge.

Nine sera samples were received from the CDC. Five that had not been challenged and four that had been challenged twice [246]. All were analyzed for their recognition of MARV (bottom). Bars represent span of oligo recognition.

**Unpublished. Preparing for Viruses.*

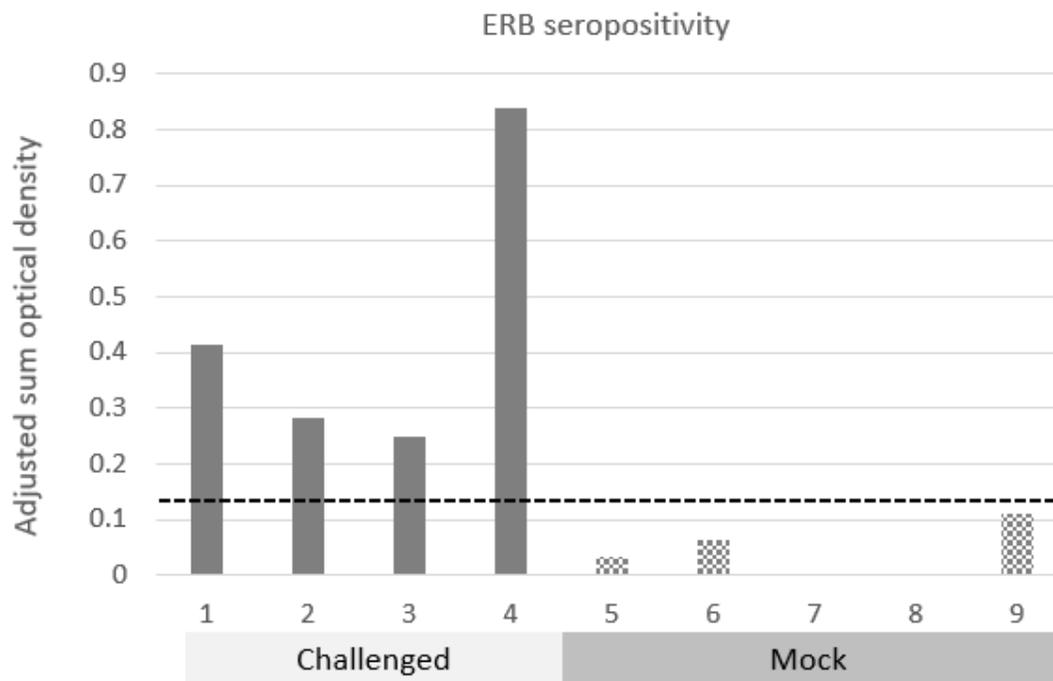


Figure 28 : ELISA titer for ERBs

Nine ERBs, four challenged with MARV on day 0 and day 28 and five challenged with media were examined for recognition of GP day 56 post first challenge. IgG antibodies were detected by ELISA using a purified recombinant Glycoprotein of the Angola Strain of Marburg virus. Antibody levels expressed as adjusted sum OD values while dotted line represented the threshold of the assay. MARV seropositive >0.15.

**Unpublished. Preparing for Viruses.*

Profile of all MARV Epitopes NHP Antibodies Recognize

We repeated the procedure with serum from seven NHPs, six that were vaccinated with MARV-GP twice and one that was mock challenged with the vehicle alone. We found that three of the six vaccinated animals mounted detectable epitope recognition via DPA (Figure 29). Further, one of the animals that mounted a response perished. We found no recognition of the GP1/GP2 cleavage site in NHPs and few overlaps in response. These results make comparisons to ERBs difficult as the vaccine does not appear to be entirely effective, nor are the correlates of protection known. We performed an ELISA to ensure seroconversion and found that NHPs 2-7 had positive titers despite one succumbing to MARV and two others not having detectable epitopes by DPA (Figure 30).

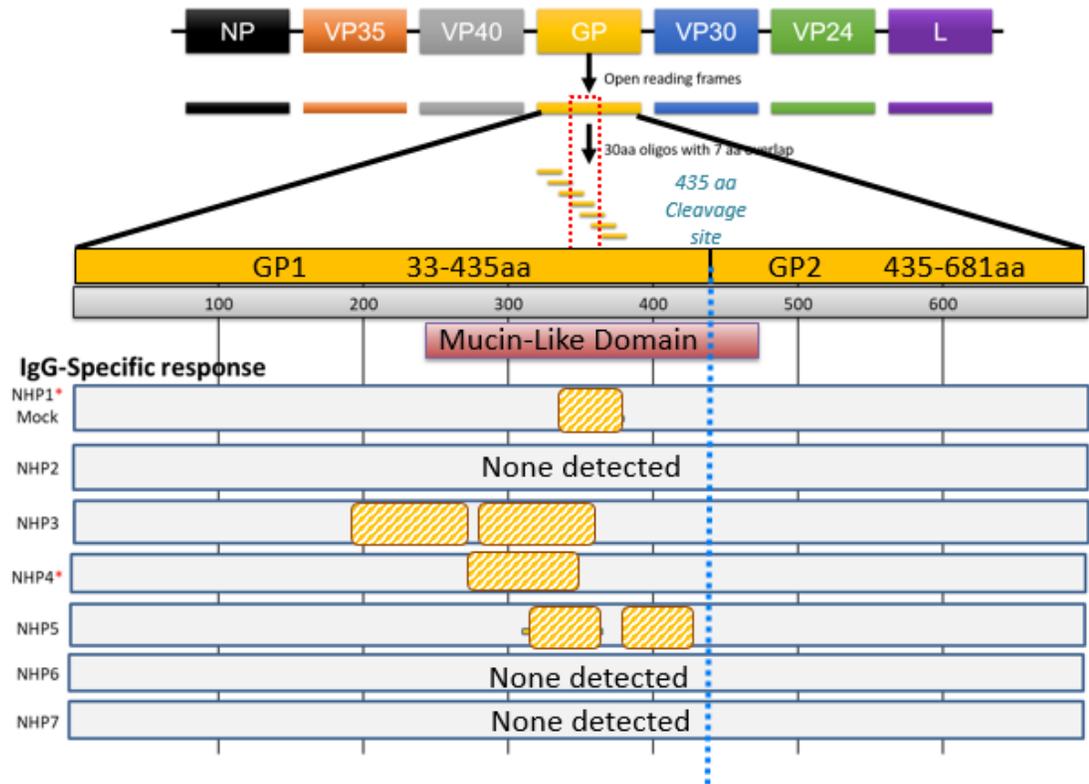


Figure 29 : MARV DPA results from vaccinated NHPs.

DPA of seven NHPs vaccinated with a MARV-GP VLP. NHP1 was mock treated while NHPs 2-7 were vaccinated. Sera taken day 84 post vaccination, animals subsequently challenged. (*) Indicates animal perished. Bars represent span of oligo recognition.

**Unpublished.*

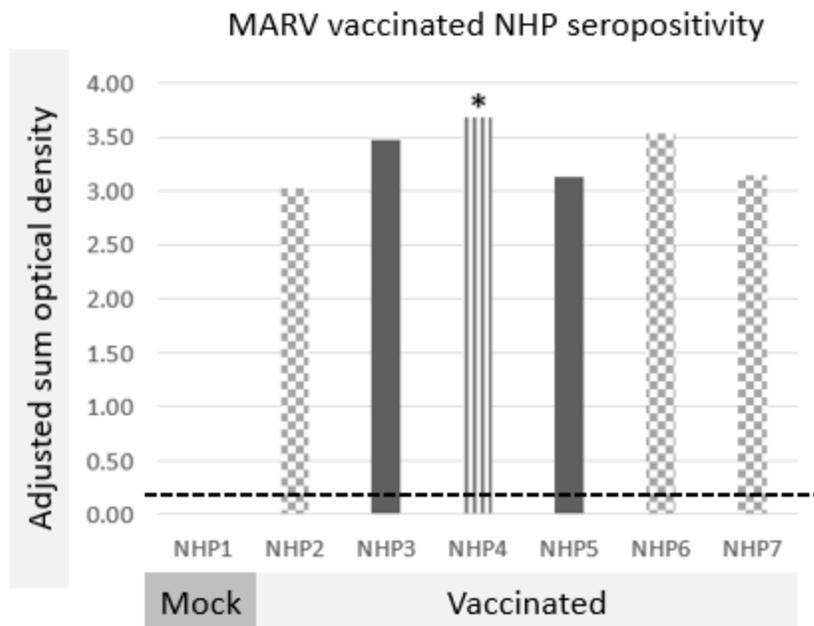


Figure 30 : MARV ELISA for vaccinated NHPs.

DPA of seven NHPs vaccinated with a MARV-GP VLP. NHP1 was mock treated while NHPs 2-7 were vaccinated. Sera taken day 84 post vaccination, animals subsequently challenged. IgG antibodies were detected by ELISA using a purified recombinant Glycoprotein of the Angola Strain of Marburg virus. Antibody levels expressed as adjusted sum OD values while dotted line represented the threshold of the assay. MARV seropositive >0.15. (*) Indicates animal perished. Subjects without results by DPA checkered.

**Unpublished.*

Comparison of MARV recognition in NHP and ERB samples

We were also interested in how the response differed to GP between ERBs and NHPs so we conducted a series of western blots. While the response appeared to be more heterogeneous in NHPs, the response in ERBs was highly conserved between individuals using both a commercial lysate and total virus lysate (Figure 31). We were able to identify epitopes previously reported to be important for the MARV response in humans, although not all (Table 12) [247]. Some of these epitopes were used to purify mAbs which were subsequently tested for neutralization capacity and protection. While many were good neutralizers, they did not protect guinea pigs suggesting that more than neutralization is required [248]. Further, many mAbs target the mucin-like domain (MLD) which has been shown to have a role in antibody-dependent enhancement (ADE) of viral infection, dependent on cross-linking of virus-antibody complexes through FcR interaction [249]. Intriguingly we found an additional 139 oligos in ERBs to NHPs, with ERBs having more oligos overall as well as to the major antigenic proteins (Table 13, Figure 32). The recognition of MARV GP extends well beyond the MLD in ERBs (Figure 27).

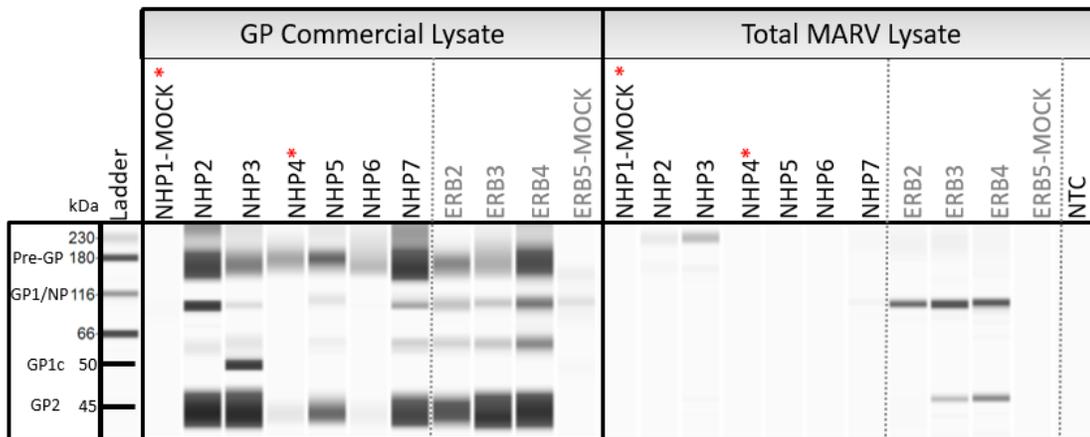


Figure 31 : Western blot for MARV recognition in ERBs and NHPs.

Western blot results of NHP and ERB sera reactivity to commercial GP or total virus lysate. NHP1 was mock treated while NHPs 2-7 were vaccinated. Sera from NHPs taken day 84 post vaccination, animals subsequently challenged. (*) Indicates animal perished. ERB 2-4 were challenged with MARV twice, sera taken 56 days post first challenge. One ERB sera sample from an unchallenged ERBs.

**Unpublished.*

Table 12 : Epitopes found in literature compared to peptides in ERB serum

EID	Epitopes in literature	Detected in ERB?
7673	DAVTE	Yes
156545	PTTTVPNTTNKYSTSPS	Yes
156547	TAPENEQTSAPSKTTLL	Yes
181232	NKYSTSPSPTPNSTAQHLVY	Yes
156548	TNKYSTSPSPTPNSTAQHLVY	No
7674	DAVTELDKN	No
150883	SDDEDLATS	No

Epitopes pulled from IEDB [247].

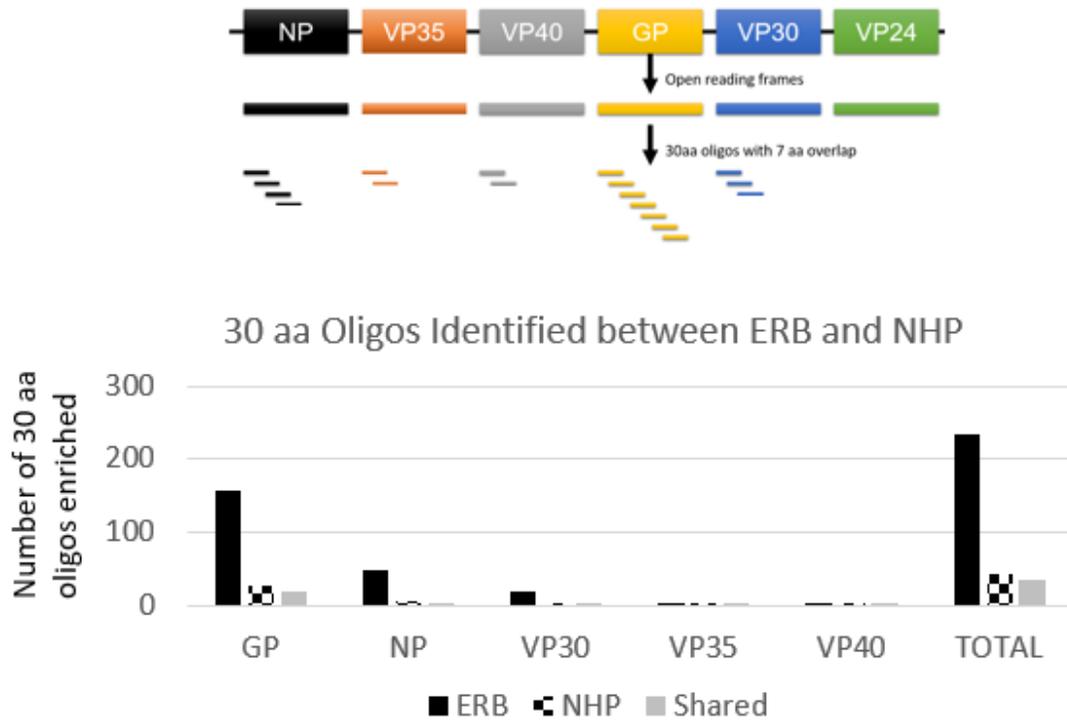


Figure 32 : Comparison of clusters between ERBs and NHPs

Oligos detected by DPA for GP, NP, VP30, VP35, and VP40 of MARV for ERB (blue), NHP (orange), and shared (gray).

**Unpublished.*

Table 13 : Oligos shared and unique between ERB and NHPs.

Protein	ERB	NHP	Shared
GP	158	27	19
NP	48	7	5
VP30	21	3	4
VP35	5	3	4
VP40	3	4	5
TOTAL	235	44	37

Oligos identified using DPA to compare recognition of MARV proteins between NHP and ERB sera. Shared oligos are those that match 100% between species.

**Unpublished.*

Discussion

Phage display technology has been proven an effective technique to resolve total linear epitope recognition in NHPs and Schreiber's Bent-Winged bats to filoviruses [36, 250]. While ERBs are known to produce a robust antibody response to MARV, what epitopes they recognize had remained elusive [87, 132]. To clarify this, we applied the DPA technique to screen all oligos from filoviruses. We identified which regions ERBs recognize of MARV and found fundamental differences to what is seen in primates.

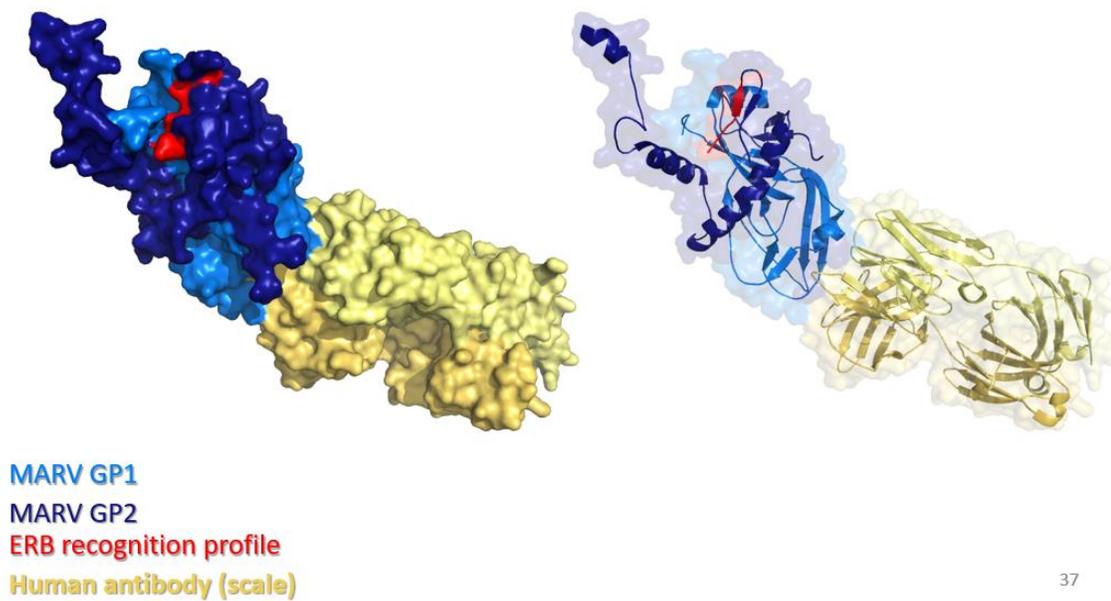
While humans and primates are known to produce an antibody response to MARV, the mechanism(s) in which they protect survivors remains unknown. However, much work has been done to characterize panels of neutralizing antibodies from MARV survivors [79]. The antibodies primarily targeted what is thought to be the NPC1 receptor-binding domain of MARV GP which include Q128 and N129. Other antibodies that are successful in humans target the wing domain of MARV, which is not present in EBOV (435aa-500aa) [79]. Intriguingly, a few mouse antibodies have been recovered that bind synthetic MARV peptides from 401aa-420aa, 421aa-435aa, and 411aa-430aa but only antibodies to 411aa-430aa prevent viral cleavage and ultimately escape [251].

We found that ERBs recognize a larger portion of MARV GP than primates, almost the entire region predicted to be antigenic. This recognition breadth is greater than what is seen in primates by hundreds of base pairs in either direction and spans important regions such as the GP1/GP2 cleavage site (435aa). This presents with two possibilities, ERBs antibodies bind and prevent escape of the endosome by occluding the cleavage site or antibodies bind and prevent entry of cells as the entry receptor is still not well characterized for MARV (Figure 33). We not only identified what was antigenic, but what regions are not antigenic as we found no recognition in any bats examined in the first and last hundred amino acids. As the immune response is complex, likely a combination of

antibodies affords ideal protection and reduces the chances of viral escape.

While DPA allowed us to examine recognition of all MARV proteins, we found the most public response to be in GP with little, if any, shared epitopes identified in the other proteins suggesting a GP-dominated response similar to what is reported for humans and filoviruses [252]. Interestingly, in humans who are symptomatic the response is primarily targeted against GP while in asymptomatic individuals the primary response is to VP40 [253]. Additionally, the region of GP1 recognized is larger in asymptomatic EBOV patients than those who succumb suggesting broader recognition may afford protection and contribute to the asymptomatic presentation [253].

Phage display is a versatile technology that allows for presentation of many oligopeptides for screening antibodies, a prime place to start when characterizing the recognition of the B and T cell response [254, 255]. However, while we characterized the linear epitopes using DPA, a short coming of this approach is the inability to detect discontinuous epitopes. The polyclonal response to a pathogen includes multiple specificities primarily comprised of structural discontinuous epitopes [256]. Nor does DPA detect non-contact residues which can confer important conformational changes affecting antigen binding. Further, DPA identifies regions of recognition but does not directly identify the functional epitope, which consists of three to five hot spot residues critical to binding [256].



37

Figure 33 : Structure of recognition site on surface of MARV GP.

The crystal structure of MARV GP (blue) bound by a human antibody (yellow) for scale, PDB: 5MOL). ERB recognition profile (red) is surface accessible.

**Unpublished. Preparing for Viruses.*

Chapter 5: Conclusion

Summary findings

In my dissertation I have discussed three integral components of the B cell response in Egyptian rousette bats and how these features may contribute to the ability of this host to withstand MARV infection, with implications for how ERBs handle viral infections generally. In my first chapter I presented the complete IGH locus for ERBs. The challenge with loci like these are the high repetition and diversity between species. We painstakingly and meticulously curated sequences to bridge all previously unassembled regions and aided by optical mapping constructed a final contiguous locus- never before done for another bat species and rarely achieved in general due to the complexity and ambiguity. Further, the annotation of germline genes was imperfect due to species divergence and so each gene was manually assessed and annotated to ensure the highest quality and confidence in RaegypIGH3.0. In the second chapter I addressed the lack of B cell repertoire data for ERBs. I developed a protocol for IgSeq in ERBs and analyzed the results using agnostic software to avoid species-biases. I found increased SHM in IgM and defined this as a major mechanism of secondary diversity in ERBs, highlighted the ~66% decrease in combinatorial diversity in ERBs, and identified unique features of ERB CDR3s. In my third chapter I sought to define the antigen recognition of ERB Igs to MARV, using a phage display technology developed for filoviruses. We found a unique recognition of MARV glycoprotein and hypothesize the potential to control infection by preventing endosomal escape. The references we created and tools we developed will contribute to the growing field of bat research and aid researchers in the field to ask more targeted questions about B cell biology and how it relates to MARV and other infections within ERBs.

Future Directions

Call for better characterization of ERBs as a model and review of what is known

Despite the advancements made with this dissertation, how ERBs overcome MARV remains unknown. However, much has been learned about the ERB immune system, V(D)J repertoire, and recognition of MARV epitopes compared to primates that may contribute to how ERBs respond to MARV and pathogens in general. The work here adds to the diverse knowledge already accumulated for the bat anti-viral immune response [26, 29, 36, 85, 86, 125, 129, 133, 139, 141, 154, 223, 257-261]. We recently reported a number of genomic differences between ERBs and humans that supported disease tolerance [85]. However, an important factor to consider is that gene regulation, particularly during infections, can critically alter infection outcome and therefore expression or kinetic validation is paramount to the relevance of these genomic differences. Moreover, absence of genes or gene families does not preclude other proteins from compensating or replacing function. Thus while gene analysis provides context for hypotheses, they on their own require additional studies to distinguish between tolerance and resistance. This has been addressed with both *in vitro* and *in vivo* transcriptomic experiments, however, both have potential pitfalls to consider. Cell culture assays have been used to determine permissibility to viral replication and for testing of receptor signaling pathways, however, some studies are done in immortalized cells that act differently from primary cultures and are well known to have additional changes in cell localization, cell recruitment, and cell to cell interactions than in a whole animal [262]. Transcriptome studies on *in vivo* experiments come with the enormous hurdle of having to house and breed bat colonies and the added difficulty of determining route, dose, of challenge to establish infection [43]. Fortunately, research on ERBs is aided establishment of a breeding colony in containment at the Center for Disease Control and Prevention, creation of a high-quality genome and transcriptome, and generation of ERB-specific

protocols and reagents.

Much work has been done to characterize the disease course in ERBs to MARV infection. Viremia is known to peak on days 5-6 post infection, typically at lower levels than in humans, followed by oral shedding that peaked days 7-8 post-infection, and finally dissemination to tissues including spleen, liver, kidney, and salivary glands [42, 43, 263]. Viral replication is not associated with increased white blood cell counts, or any signs of clinical infection like changes in weight or temperature. Further, there is a lack of evidence for inflammatory infiltrates in infected tissues [43]. Viremia is cleared by day 13 and shedding ends by day 19. An interesting co-housing experiment resulted in MARV transmission to uninfected bats 4-7 months after experimental infection supporting the potential for long latent periods without detectable viral replication [11]. Immunity is known to be established because upon secondary challenge of previously MARV-infected ERBs there is no detectable viral replication or shedding [132]. However, no long term study has been done to confirm viral persistence nor are the mechanisms of persistence for filoviruses in humans well understood.

Unlike in other pteropus bats, ERBs have been shown to have no constitutive expression of type I IFNs [125] but that IFN is induced in cell lines with Sendai virus suggesting that responses to one virus in ERBs is not applicable to all [85]. It also appears that type I genes are expanded in ERBs relative to the best studied closest relative, *Pteropus alecto* [85]. This recent genome revealed substantial variations in the NK cell receptor repertoire, namely the lack of functional killer cell immunoglobulin receptors (KIRs) and all killer lectin-like receptors (KLRs) encoding either 1) activating and inhibitory interaction motifs, or 2) inhibitory interaction motifs only [85]. This highlights the need for continued characterization of Chiropteran species and underscores the importance of not simplifying findings from one species of bat to another- even within the same suborder.

This highlights the need for more standardized methods for assessing novel

viruses and host-pathogen relationships. For example, not all MARV strains are used in every study nor are consistent cell lines utilized making comparative analysis difficult. For example, ERB bone marrow derived dendritic cells (BMDCs) are susceptible to MARV infection and active viral transcription occurs up to three days post infection with MARV371 (ERB derived strain), but to lesser levels than described for MARV-Musoke (human derived strain) in human myeloid derived dendritic cells (MDDCs) [259]. This could be due to difference in virus isolates (bat versus human) or cell preparations (BMDCs versus MDDCs), or perhaps due to induction of strong antiviral responses in bat DCs. Infection of BMDCs resulted in significant upregulation of IFN-related genes with antiviral properties and significant downregulation of genes and pathways for DC maturation and adaptive responses [259]. One gene, *CCL8* which encodes a chemokine responsible for recruiting leukocytes to infection was downregulated in bat cells whereas it is known to be upregulated throughout disease in NHPs. Similarly, *CXCL2*, a pro-inflammatory chemokine, is upregulated in mice that succumb to EBOV but downregulated in ERB BMDCs. These studies are important in elucidating the differences between human and bat responses, but highlight a major flaw in transcriptomic comparisons between species. Indeed, even between mice, NHPs and humans significant differences exist in the transcriptome landscape [264, 265]. This is likely only exacerbated in a species such as ERBs which are much further away in evolutionary time, requiring a more discerning characterization of cells and base levels. Further, biologically relevant changes are still debated in the field. A two-fold increase or decrease does not necessarily correlate to more or less importance in outcome and in fact, fold changes of <2 can be important biologically but are often removed from transcriptomic analysis [266].

While genomic and transcriptomic analyses are necessary starting points in novel species characterization, much can be missed since current pipelines are inherently biased based on what is known. In our annotation of this locus and others we found

unannotated genes that were previously reported as missing, unannotated genes that are more divergent from the closest relative sequenced, misannotated genes that contained extraneous exons and more. As sequencing of new species increases, there is a need for more meticulous annotation and validation outside of what is known for human. This is exacerbated by the incomplete understanding of human genetics as new species are inevitably compared to the human reference, which remains imperfect and still contains hundreds of gaps [267-269].

Another common pitfall of comparisons is the presumption that a gene with a known function in humans will have a similar function in a new species despite high sequence variation. In the case of IFNs, there is remarkable variability across species in sequence, number, regulation, and expression [270]. It was recently demonstrated that ERB IFN gamma (IFN γ) shares limited sequence similarity to human counterparts [125]. This group hypothesized that how filoviruses antagonize the innate immune response in bats might not mimic human. They found that type 1 IFNs induced an immune response in human and bat cells but bat type II IFNs only stimulated rousette cells suggesting they function differently to human IFNs. Beyond the well characterized IFN α and IFN δ subtypes, recent works have shown significant expansions in the number of IFN ω genes of bats, pigs, and cattle which are poorly understood [271]. This underscores the need for extensive research to characterize this model and to avoid attributing known phenomena in the immune response in humans, mammals, or even other bats, to ERBs.

To this end, I propose the following in continuation of the findings in this dissertation: 1) A systematic characterization of ERB Igs, FcRs and functions, 2) an in-depth characterization of the V(D)J response to MARV in ERBs, and 3) validation of MARV endosomal entrapment. These experiments would further the field significantly and better define features of this model organism.

Effector functions

We identified numerous differences in ERBs Igs and FcRs which demand further investigation. As there is significant differences between even mice and humans which are much closer evolutionarily, the FcRs and how they pair are likely unique in ERBs to both humans and mice. An interesting next step would be to determine the combinations in which these pair, and what cells these receptors are expressed on as well as what mechanisms are induced. The absence of *in vitro* MARV neutralization activity in more than 50 ERB sera indicates that clearance of MARV by humoral response in ERBs relies on other antibody-mediated functions and/or innate immune response [133]. Three mechanisms of Ig mediated immunity: antibody-dependent cellular cytotoxicity (ADCC), antibody-dependent cellular phagocytosis (ADCP), and complement mediated lysis (CML).

The binding specificities for each ERB FcR are not known. As FcRs play important roles in antibody effector function, understanding FcR specificity is critical to studying these processes. As no bat-specific reagents to explore this exist, we could create a series of hybrid antibodies (hAb) with the Fab domain of the monoclonal anti-MARV antibody, FILORAB3, and the Fc domain of each ERB Ig. A molecular tag would be added so we could detect if the chimeras bind antigen, allowing us to dissect the binding ability of each ERB FcR to each Ig by way of fluorescence similar to what has been shown in mice (Figure 34) [81]. To validate this interaction, both the FcR and the Fc would be modified to ablate binding (FcS will be aglycosylated and FcRs will be synthesized with mutations in residues known to be crucial for Fc/FcR interactions in humans along the second domain of the receptor) and subjected to the same *in vitro* and immunoprecipitation assays. Antibody dilution buffer can be used as a negative control to account for background and the Fab portion of polyclonal sera as an experimental negative control as it will lack the Fc domain and should not bind FcRs.

A major mode of action of many therapeutic EBOV mAbs is ADCC [272]. This mechanism relies on the ability for FcRs on effector cells such as NK cells to recognize and bind the Fc portion of Igs. Despite numerous attempts to purify NK cells from ERB PBMCs, no reagents exist to select for and culture this cell population. Although there are commercially available human and mouse NK cell lines, the genes associated with NK cell receptors vary significantly between humans, mice and ERBs, leading us to suspect these cells will not be ideal alternatives as they would lack the ERB FcRs. Instead, I would transfect ERB FcRs into a mammalian NK cell line (NK-92) and test the ability of ERB sera to induce ADCC. MARV GP can be transfected into target bat cells and then incubated with sera from MARV infected ERBs [273] (Figure 34). ADCC can be measured by apoptosis in wells with FcRs and Ig compared to Ig alone, NK cells alone, bat cells alone, and Triton-x treated cells (apoptosis positive control). As a negative control of FcR function, we can aglycosylate sera previous to addition of NK cell surrogates to ablate ADCC activity [274].

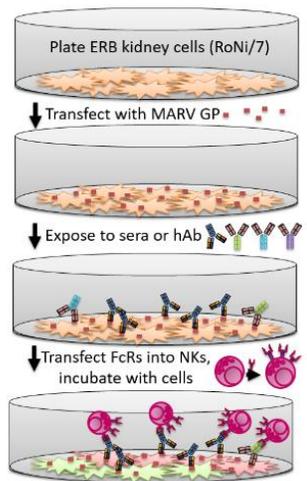


Figure 34 : *In vitro* ADCC assay.

ADCC relies on the ability for FcRs on effector cells such as NK cells to recognize and bind the Fc portion of Igs. MARV GP will be transfected into target bat cells derived from adult ERB kidneys (RoNi/7) and then incubated with sera from MARV infected ERBs. Additionally this can be repeated in containment with virus (not depicted). Following incubation, cells will be washed to remove any unbound antibodies and then incubated with NK cells expressing FcRs, ADCC will be measured by apoptosis in wells with FcRs and Igs.

Our findings in Chapter 1 have indicated that ADCP is a potential function of multiple ERB Igs. Additionally, this mechanism has been identified as an important mode of action for filovirus mAb therapeutics [54]. Phagocytosis removes infected cells and can be mediated by macrophages, monocytes, dendritic cells, or neutrophils. To test the ADCP function of ERB serum we could label target cells (primary culture of monocytes) with a pH sensitive dye that fluoresces only in mature phagosomes and quantify with cell imaging [275]. Primary bat monocytes can then be incubated with a bat cell line transfected with MARV GP after incubation with serum from convalescent ERBs [276]. Beads can be coated in engineered hAbs and will only be engulfed if the FcR on ERB macrophages binds that isotypes Fc domain. As a negative control we will use aglycosylated hAbs which would not bind FcRs but would detect GP [274]. Phagocytosis can be measured by the amount of fluorescence detected that should only be observed if the opsonized cell/bead is engulfed by the macrophage.

Another mechanism predicted by the work outlined in Chapter 1 for ERB Igs is CML. For CML, apoptosis can be measured with various *in vitro* assays (Figure 35). We could examine this mechanism using fluorescence dyes that do not require cell lysis and provides a direct measurement of cytotoxicity rather than common indirect indicators such as release of ATP or activity of lactate dehydrogenase. Antibodies that recognize viral antigens on infected cells can initiate activation of the complement cascade by binding complement via their Fc region to trigger cell death. To determine whether MARV challenged ERBs generate antibodies capable of inducing CML, a bat cell line will be transfected with MARV GP and incubated with ERB sera and horse complement (closest related species with commercially available reagents) [129]. Cultures can be incubated with fluorescent markers and assessed on a flow cytometer. Spontaneous lysis can be measured in wells with complement but no antibodies, and total lysis can be determined in cells incubated with 1% Triton X-100 [277]. These experiments can identify the amount

CML contributes to control of MARV by ERB sera.

While complement is considered a major component of innate immunity, it links the innate and adaptive responses through a variety of mechanisms including enhancement of humoral immunity, regulating antibody effector mechanisms, and T cell function modification. Mannose-binding lectin (MBL) activates the lectin complement pathway and has been shown to bind EBOV effectively, but not MARV, resulting in strong neutralization of EBOV and minimal neutralization of MARV [96]. We anticipate seeing some CML as it has been shown to be an important mediator in multiple viral infections, but as this pathway induces an inflammatory in other mammals (a response not documented in ERBs during MARV infection,) CML may not be the primary mechanism of protection [99].

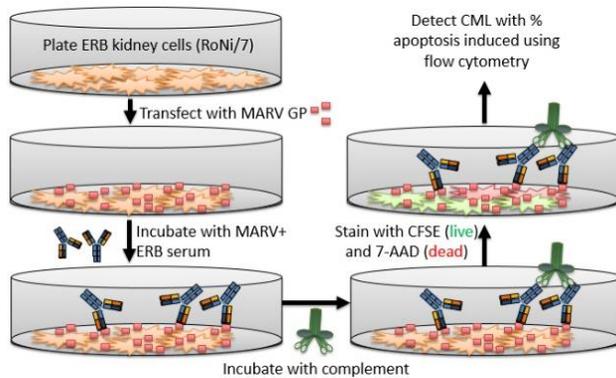


Figure 35 : *In vitro* CML assay.

CML will be measured by the amount of apoptosis induced after incubation with antibodies and complement. This method utilizes carboxyfluorosuccinimide ester (CFSE), a green fluorescent probe which labels live cells and 7-aminoactinomycin D (7-AAD), a red fluorescent probe which labels non-viable cells. This does not require cell lysis and provides a direct measurement of cytotoxicity rather than common indirect indicators such as release of ATP or activity of lactate dehydrogenase. Antibodies that recognize viral antigens on infected cells can initiate activation of the complement cascade by binding complement via their Fc region to trigger cell death. Assay can be performed in containment with fully virulent virus as well (not shown).

MARV-specific V(D)J response

Now that we have laid the groundwork to examine MARV-specific V(D)J repertoire, the next logical step would be to create fluorescently tagged MARV proteins to pull down specific B cell populations and interrogate their complement of B cell receptors. Recently, researchers from Vanderbilt University have developed a high-throughput, unbiased methodology to screen the antibody repertoire of single B cells called LIBRA-seq (Figure 36) [278]. This method uses the 10x Genomics Chromium Single Cell Immune Profiling Solution to recover DNA-barcoded antigens and B cell receptor sequences simultaneously to overcome the limitations of fluorescence based screening methods. To date this has been shown to work effectively in human HIV samples and identified a diverse new panel of antibodies [278]. This process may encounter significant challenges as purification would need to be tested and employed in containment. However, once it is validated this could be used for all high-consequence pathogen samples for novel antibody discovery which would be a significant benefit to the field.

This method combines barcoding and antigen specificity. A library of MARV antigens would be labeled with the same fluorophore to enable sorting of MARV-positive B cells by fluorescence activated sorting (FACS). Individual B cells will be sorted using droplet microfluidics and then BCR transcripts tagged with a common cell barcode attached to a bead to ensure direct mapping of BCR sequence to antigen detected (Figure 36). Single-cell suspensions will be loaded onto the Chromium Controller microfluidics device (10x Genomics) and processed under the B cell Single Cell V(D)J solution to manufacturer's suggestions to capture 4,000 B cells per 1/8 10 cassette. Libraries will be prepared following the CITE-seq protocol, adaptor removed by bead cleanup, and quantified by qPCR. Samples will be sequenced on the NovaSeq 6000 at 2.5% of a flow cell. Reads will be processed with the LIBRA-seq pipeline, which takes paired-end FASTQ

files and annotates reads for each barcode, UMI, and antigen. V(D)J will be annotated with our custom germline reference RaegyptBCR3.0.

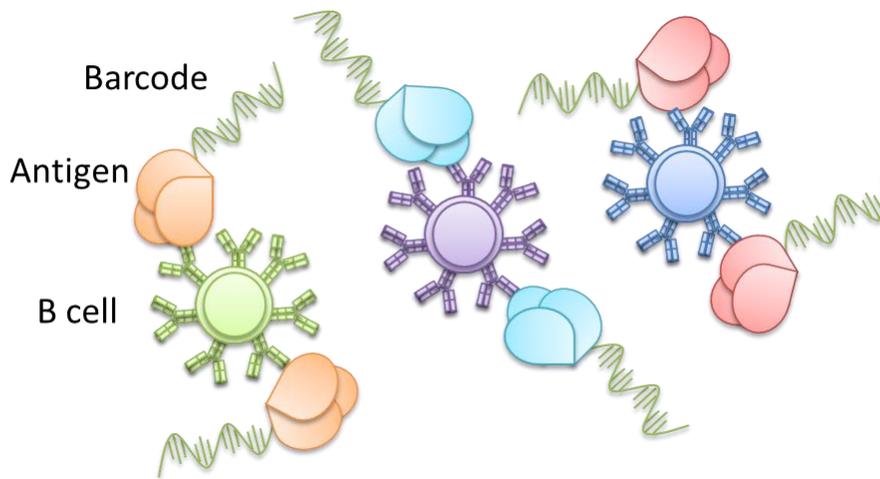


Figure 36 : Libra-seq for antigen-specific VDJ-VJ paired sequenced.

PBMCs from humans are mixed with DNA-barcoded antigens and purified based on each antigen by flow cytometry. Single cell sequencing and functional assays can be done on individual clones.

High-Throughput Mapping of B Cell Receptor Sequences to Antigen Specificity [278].

Validation of novel epitopes and endosome entrapment mechanism

Finally, I identified regions of MARV GP ERBs recognize that may be protective. To test if these epitopes are protective, we could modify GP to ablate the binding site. If the epitope was protective for ERBs then the ablation will lead to disease presentation. However, as this is a dual-use research concern, instead I recommend purifying bat antibodies that recognize the epitope by immunoprecipitation. Once we identify which sequence generates the protective antibody, we can synthesize it and test it in NHPs. Additionally, the mechanism in which ERBs are protected is unknown. Although has been shown that ERB IgG doesn't neutralize MARV, the location of epitopes I identified suggest a difficulty to escape the endosome (Figure 37). I would utilize fluorescently labeled MARV GP virus-like particles (VLPs) to monitor where these antibodies bind and control virus [279] If it is indeed endosomal entrapment, we should be able to follow in real-time the VLPs and the endosome would fluoresce after the cell is infected and diminish after acidification and degradation of the particle. Further, the ablated binding mAbs we created can be used to show the VLPs can escape the endosome and infect more cells to serve as a positive control for mAb efficacy.

Recombinant vesicular stomatitis virus (VSV) expressing MARV glycoprotein (rVSV-GP-MARV) will be stained with a fluorophore and incubated with a bat cell line such as RoNi/7, fixed and examined after two hours. As the GTPase Rab7 specifically associated with late endosomes, we will use this as a marker for the compartment and identify co-localization of internalized virions [280]. The plasma membrane will be incubated with different fluorophore and external viral particles detected with the mouse monoclonal specific for MARV GP. The GP antibodies will be detected with a secondary goat anti-mouse antibody. After washing with PBS cells will be mounted to glass slides and fluorescence monitored on an epifluorescent microscope.

The endosomal proteases cathepsin B and cathepsin L are required for MARV to

enter cytoplasm and infect host cells [281]. Another experiment to test the mechanism of these antibodies would be to incubate MARV-GP with ERB antibodies, followed by washing and incubation with cathepsins. Proteins can be visualized on a western blot to detect successful or inhibited cleavage. The negative control would be incubated with PBS instead of sera, and the positive control would be a monoclonal antibody built to the GP1/GP2 cleavage interface.

Although we made significant strides toward identifying the regions of MARV ERBs can recognize, the discrete epitopes remain uncharacterized. While western blots with NHP serum posed minor challenges to achieve satisfactory results, ERB serum and the few reagents that exist had unavoidable background that made quantitation out of the question and detection difficult. Currently, the only antibody for ERB Ig was created in a *Yangochiroptera* species that is reported to have some cross-reactivity but in our hands has proven unreliable. To this end, I would next make multiple ERB Ig-specific monoclonal antibodies in a mouse background. While this is an expensive and timely, it would push the field forward and make ERBs easier to use as a model organism.

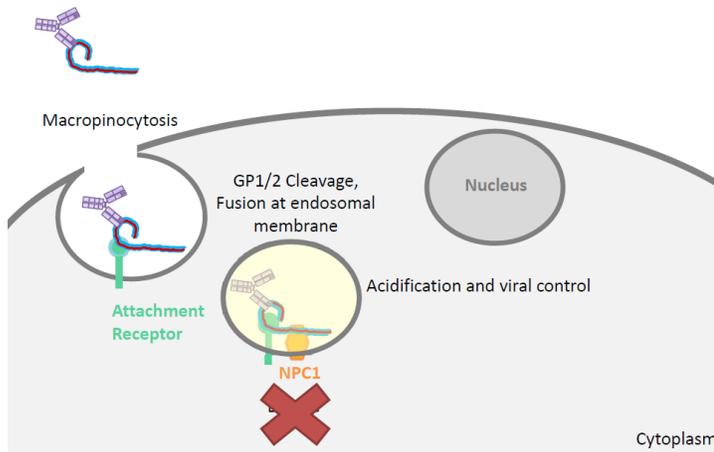


Figure 37 : Endosomal entrapment illustration.

Hypothesis that if ERB antibodies bind GP1/GP2 cleavage site than viruses won't be able to escape endosome and will be subsequently degraded by acidification.

Concluding hypotheses and approaches

Beneficial infection hypothesis

While infection with MARV in humans and primates results in a potentially lethal hemorrhagic fever, ERBs are resistant to these consequences. However, it is well known that numerous viruses present asymptomatically in humans as well. For example, influenza virus infections are associated with a wide range of manifestation from asymptomatic to fatal [282]. Herpes Simplex Virus-2 (HSV-2) is widespread in part because individuals can shed without symptoms [283]. Epstein Barr virus is (EBV) results in numerous asymptomatic cases because of an effective host immune response [284]. EBV is able to persist for a lifetime in the host in a subset of memory B cells where the virus maintains a latent state to remain hidden from the immune system. Control of EBV is thought to occur through the combination of CD8+ T cells, and NK cells to a lesser extent. As EBV periodically re-emerges and can be shed, the virus persists at a population level as well as individually. While the tropism for MARV in humans is defined, the tissues that MARV targets preferentially in ERBs is not yet determined. Further, the prevalence of MARV in wild adult ERBs (7-8 months of age) is consistent year round (2.4% PCR, 21.5% seroprevalence) [11]. If the virus persists in ERBs lifelong, it is possible these sporadic shedding events allow for maintenance within the colony to continually expose new juveniles. In EBV and other cytomegaloviruses, transmission first occurs in early childhood through genital excretions at birth, breast milk, or saliva from a family member [285]. Studies assessing EBV have shown that humans infected with EBV before age two had lower risk of IgE sensitization, decreasing allergies [285]. Although this mechanism is not well understood in humans, it offers an interesting avenue of thought for ERBs and age of exposure to MARV. This raises the question of what benefit MARV might serve ERBs. As it is well maintained within colonies without obvious detriments, it could potentially serve

a beneficial purpose. Perhaps infection as juveniles helps stack the immune system for a decrease in IgE sensitivity, allowing for the handling of more IgE overall.

One fascinating finding of this work was the expansion of IgE genes. While IgE was the last of the five human antibody classes to be discovered and is commonly associated today with allergic disease manifestation, its role in mammalian evolution appears to be primarily focused on defense against parasites and venoms [286]. Indeed, while IgE induced allergic inflammation can be deleterious, when the target is not an autoantigen but a response to helminths/ectoparasite bites- the development of effector mechanisms help clear parasites. Further, chronic infection turns on downregulation of inflammation and tissue damage is in part mediated by regulatory T cells and IL10 secretion [287]. In allergic disorders these down regulatory mechanisms do not fully develop or are lost by the overwhelmed inflammatory response [287]. As it is well known that ERBs host numerous ectoparasites, they may have evolved to have a less-inflammatory IgE response as a result [288]. Further it has been posited for dogs that the heightened parasite burden they endure correlates with elevated overall IgE serum levels [289]. Based on the work presented here, I hypothesize that MARV infection in ERBs helps prime the immune system for recurrent stimulation of the IgE response. This is based on the principle that co-habitation with ectoparasities has led to a less-inflammatory response with consistently stimulated IgE in bats [290].

To determine if MARV has a beneficial impact on the immune response to parasites, and if time of exposure is relevant I would conduct an *in vivo* double challenge study. I would challenge 16 juveniles and 16 adult ERBs with MARV, then challenge groups of four with common ectoparasites for ERBs or PBS and monitor for differences in outcome including weight, fever, blood chemistries, and antibody titers. An additional 12 ERBs I would not challenge with MARV but would challenge with ectoparasites alone. If MARV infection plays a role in decreasing IgE sensitization response to ectoparasites, I

would anticipate to observe symptoms to infection in the MARV-naïve animals. However, it should be noted that ectoparasite challenge studies are not currently well defined and would require significant troubleshooting to develop and validate.

Affinity maturation hypothesis

Another potential reason ERBs maintained two IgE genes is to aid affinity maturation for the IgE response by allowing sequential class switching and accumulation of SHMs. It is thought that increased SHM leads to antibodies with increased specificity, and thus more rounds of affinity maturation leads to highly specialized antibodies. The isolation of the first HIV-1 broadly neutralizing antibodies revealed mutation frequencies as high as 30%, well above the 4-6% frequency typically of anti-pathogen IgG [291]. In mice, it is known that high affinity IgE is produced through sequential class switching ($\mu \rightarrow \gamma \rightarrow \epsilon$) in which the intermediary phase of IgG is necessary for the IgE response through inherited SHM from IgG1 [292]. In contrast, low affinity is produced through direct class switching ($\mu \rightarrow \epsilon$) and has much less SHMs. Further, mice without IgG1 cannot produce high affinity IgE even after multiple immunizations. To this end, ERBs may have maintained two IgE genes to undergo sequential class switching from the IgG response to IgE and then again to produce highly curated responses by allowing for the accumulation of SHMs over an additional round of switching ($\mu \rightarrow \gamma \rightarrow \epsilon \rightarrow \gamma \rightarrow \epsilon$). This can be tested by examining the V(D)J repertoire of IgE+ B cells after sequential challenges with a series of antigens. Plasma cells would be sorted and purified for antigen-specific cells and bat-specific antibodies for IgG and IgE. I would expose groups of ten bats on day 0, day 28, and take blood on days 1, 3, 14, 29, 30, 42, 56. IgE+ B cells would be assessed on days: 1, 3, 29, 30 and IgG+ B cells on day 14, 28, 42, 56. Primers unique to each isotype and subclass will be used to examine the potential for sequential class switching and accumulation of SHMs. Alternatively, an organoid culture of ERB lymph nodes could

be established and stimulated with common immune mediators and test the production of antibodies over sequential additions of stimulant and analyzed in a similar way as plasma cells above [293].

Summary

Zoonoses cause more than half of all emerging diseases and bats host more zoonotic viruses than any other studied mammal. Increasing encroachment and exposure of humans to new species poses the risk of spillover events with the potential to cause severe disease such as the 2004 Marburg virus outbreak, the 2013-2016 Ebola virus outbreak, and the ongoing SARS-CoV-2 pandemic. Understanding the immunological mechanisms that protect reservoir hosts can offer insights into development of better medical counter measures. Further, as bats and the ERB in particular is able to harbor multiple human pathogenic viruses, ERBs may utilize pan-viral mechanisms of control that would be highly beneficial to potentially combating disease X. The work here contributes substantially to the field and revealed numerous features of ERB immunoglobulins that support the model of disease tolerance. These immunoglobulin characteristics and overall tolerogenic state may contribute to the asymptomatic presentation of ERBs to multiple viruses.

References

1. Gebreyes, W.A., et al., *The global one health paradigm: challenges and opportunities for tackling infectious diseases at the human, animal, and environment interface in low-resource settings*. PLoS neglected tropical diseases, 2014. **8**(11): p. e3257-e3257.
2. Plowright, R.K., et al., *Pathways to zoonotic spillover*. Nature reviews. Microbiology, 2017. **15**(8): p. 502-510.
3. Woolhouse, M.E.J. and S. Gowtage-Sequeria, *Host range and emerging and reemerging pathogens*. Emerging infectious diseases, 2005. **11**(12): p. 1842-1847.
4. Han, B.A., A.M. Kramer, and J.M. Drake, *Global Patterns of Zoonotic Disease in Mammals*. Trends in parasitology, 2016. **32**(7): p. 565-577.
5. Han, B.A., et al., *Rodent reservoirs of future zoonotic diseases*. Proc Natl Acad Sci U S A, 2015. **112**(22): p. 7039-44.
6. Teeling, E.C., et al., *A molecular phylogeny for bats illuminates biogeography and the fossil record*. Science, 2005. **307**(5709): p. 580-4.
7. Olival, K.J., et al., *Host and viral traits predict zoonotic spillover from mammals*. Nature, 2017. **546**(7660): p. 646-650.
8. Calisher, C.H., et al., *Bats: important reservoir hosts of emerging viruses*. Clin Microbiol Rev, 2006. **19**(3): p. 531-45.
9. Smith, I. and L.F. Wang, *Bats and their virome: an important source of emerging viruses capable of infecting humans*. Curr Opin Virol, 2013. **3**(1): p. 84-91.
10. Amman, B.R., et al., *Seasonal pulses of Marburg virus circulation in juvenile Rousettus aegyptiacus bats coincide with periods of increased risk of human infection*. PLoS Pathog, 2012. **8**(10): p. e1002877.
11. Schuh, A.J., et al., *Modelling filovirus maintenance in nature by experimental transmission of Marburg virus between Egyptian rousette bats*. Nat Commun, 2017. **8**: p. 14446.
12. Páez, D.J., et al., *Conditions affecting the timing and magnitude of Hendra virus shedding across pteropodid bat populations in Australia*. Epidemiology and infection, 2017. **145**(15): p. 3143-3153.
13. Plowright, R.K., et al., *Transmission or Within-Host Dynamics Driving Pulses of Zoonotic Viruses in Reservoir–Host Populations*. PLOS Neglected Tropical Diseases, 2016. **10**(8): p. e0004796.
14. Peel, A.J., et al., *Support for viral persistence in bats from age-specific serology and models of maternal immunity*. Sci Rep, 2018. **8**(1): p. 3859.
15. Sharma, V., et al., *Emerging trends of Nipah virus: A review*. Rev Med Virol, 2019. **29**(1): p. e2010.
16. Mortlock, M., et al., *Novel Paramyxoviruses in Bats from Sub-Saharan Africa, 2007-2012*. Emerging infectious diseases, 2015. **21**(10): p. 1840-1843.
17. Lau, S.K., et al., *Identification and complete genome analysis of three novel paramyxoviruses, Tuhoko virus 1, 2 and 3, in fruit bats from China*. Virology, 2010. **404**(1): p. 106-16.
18. Iorio, R.M. and P.J. Mahon, *Paramyxoviruses: different receptors - different mechanisms of fusion*. Trends in microbiology, 2008. **16**(4): p. 135-137.
19. Baker, K.S., et al., *Novel, potentially zoonotic paramyxoviruses from the African straw-colored fruit bat Eidolon helvum*. J Virol, 2013. **87**(3): p. 1348-58.
20. Drexler, J.F., et al., *Bats host major mammalian paramyxoviruses*. Nat Commun, 2012. **3**: p. 796.

21. Amman, B.R., et al., *A Recently Discovered Pathogenic Paramyxovirus, Sosuga Virus, is Present in Rousettus aegyptiacus Fruit Bats at Multiple Locations in Uganda*. J Wildl Dis, 2015. **51**(3): p. 774-9.
22. Menachery, V.D., et al., *A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence*. Nat Med, 2015. **21**(12): p. 1508-13.
23. Malik, Y.S., et al., *Emerging novel coronavirus (2019-nCoV)-current scenario, evolutionary perspective based on genome analysis and recent developments*. Vet Q, 2020. **40**(1): p. 68-76.
24. Ji, W., et al., *Cross-species transmission of the newly identified coronavirus 2019-nCoV*. J Med Virol, 2020. **92**(4): p. 433-440.
25. Banerjee, A., et al., *Bats and Coronaviruses*. Viruses, 2019. **11**(1): p. 41.
26. Schulz, J.E., et al., *Serological Evidence for Henipa-like and Filo-like Viruses in Trinidad Bats*. J Infect Dis, 2020.
27. Wang, L.F. and D.E. Anderson, *Viruses in bats and potential spillover to animals and humans*. Curr Opin Virol, 2019. **34**: p. 79-89.
28. Glennon, N.B., et al., *Transcriptome Profiling of the Virus-Induced Innate Immune Response in Pteropus vampyrus and Its Attenuation by Nipah Virus Interferon Antagonist Functions*. J Virol, 2015. **89**(15): p. 7550-66.
29. Arnold, C.E., et al., *Transcriptomics Reveal Antiviral Gene Induction in the Egyptian Rousette Bat Is Antagonized In Vitro by Marburg Virus Infection*. Viruses, 2018. **10**(11).
30. Towner, J.S., et al., *Isolation of genetically diverse Marburg viruses from Egyptian fruit bats*. PLoS Pathog, 2009. **5**(7): p. e1000536.
31. Towner, J.S., et al., *Marburgvirus genomics and association with a large hemorrhagic fever outbreak in Angola*. J Virol, 2006. **80**(13): p. 6497-516.
32. Carroll, S.A., et al., *Molecular evolution of viruses of the family Filoviridae based on 97 whole-genome sequences*. J Virol, 2013. **87**(5): p. 2608-16.
33. Kuhn, J.H., et al., *New filovirus disease classification and nomenclature*. Nature reviews. Microbiology, 2019. **17**(5): p. 261-263.
34. Negredo, A., et al., *Discovery of an ebolavirus-like filovirus in europe*. PLoS Pathog, 2011. **7**(10): p. e1002304.
35. Olival, K.J. and D.T. Hayman, *Filoviruses in bats: current knowledge and future directions*. Viruses, 2014. **6**(4): p. 1759-88.
36. Ramirez de Arellano, E., et al., *First Evidence of Antibodies Against Lloviu Virus in Schreiber's Bent-Winged Insectivorous Bats Demonstrate a Wide Circulation of the Virus in Spain*. Viruses, 2019. **11**(4).
37. Yang, X.L., et al., *Author Correction: Characterization of a filovirus (Mengla virus) from Rousettus bats in China*. Nat Microbiol, 2019. **4**(3): p. 543.
38. Olival, K.J. and D.T.S. Hayman, *Filoviruses in bats: current knowledge and future directions*. Viruses, 2014. **6**(4): p. 1759-1788.
39. Brauburger, K., et al., *Forty-five years of Marburg virus research*. Viruses, 2012. **4**(10): p. 1878-1927.
40. Bausch, D.G., et al., *Risk factors for Marburg hemorrhagic fever, Democratic Republic of the Congo*. Emerging infectious diseases, 2003. **9**(12): p. 1531-1537.
41. Amman, B.R., et al., *Marburgvirus resurgence in Kitaka Mine bat population after extermination attempts, Uganda*. Emerg Infect Dis, 2014. **20**(10): p. 1761-4.
42. Paweska, J.T., et al., *Virological and serological findings in Rousettus aegyptiacus experimentally inoculated with vero cells-adapted hogan strain of Marburg virus*. PLoS One, 2012. **7**(9): p. e45479.

43. Jones, M.E., et al., *Experimental Inoculation of Egyptian Rousette Bats (Rousettus aegyptiacus) with Viruses of the Ebolavirus and Marburgvirus Genera*. *Viruses*, 2015. **7**(7): p. 3420-42.
44. Coltart, C.E., et al., *The Ebola outbreak, 2013-2016: old lessons for new epidemics*. *Philos Trans R Soc Lond B Biol Sci*, 2017. **372**(1721).
45. *Outbreak of Ebola virus disease in the Democratic Republic of the Congo, April-May, 2018: an epidemiological study*. *Lancet*, 2018. **392**(10143): p. 213-221.
46. Mbala-Kingebeni, P., et al., *2018 Ebola virus disease outbreak in Equateur Province, Democratic Republic of the Congo: a retrospective genomic characterisation*. *Lancet Infect Dis*, 2019.
47. Amman, B.R., et al., *Isolation of Angola-like Marburg virus from Egyptian rousette bats from West Africa*. *Nat Commun*, 2020. **11**(1): p. 510.
48. Henao-Restrepo, A.M., et al., *Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola Ca Suffit!)*. *Lancet*, 2017. **389**(10068): p. 505-518.
49. Matassov, D., et al., *Single-Dose Trivalent VesiculoVax Vaccine Protects Macaques from Lethal Ebolavirus and Marburgvirus Challenge*. *J Virol*, 2018. **92**(3).
50. Davey, R.T., Jr., et al., *A Randomized, Controlled Trial of ZMapp for Ebola Virus Infection*. *N Engl J Med*, 2016. **375**(15): p. 1448-1456.
51. *Public health round-up*. *Bull World Health Organ*, 2019. **97**(1): p. 4-5.
52. Dhama, K., et al., *Advances in Designing and Developing Vaccines, Drugs, and Therapies to Counter Ebola Virus*. *Front Immunol*, 2018. **9**: p. 1803.
53. Misasi, J., et al., *Structural and molecular basis for Ebola virus neutralization by protective human antibodies*. *Science*, 2016. **351**(6279): p. 1343-6.
54. Saphire, E.O., et al., *Systematic Analysis of Monoclonal Antibodies against Ebola Virus GP Defines Features that Contribute to Protection*. *Cell*, 2018. **174**(4): p. 938-952.e13.
55. Dye, J.M., et al., *Postexposure antibody prophylaxis protects nonhuman primates from filovirus disease*. *Proc Natl Acad Sci U S A*, 2012. **109**(13): p. 5034-9.
56. Thi, E.P., et al., *Marburg virus infection in nonhuman primates: Therapeutic treatment by lipid-encapsulated siRNA*. *Science translational medicine*, 2014. **6**(250): p. 250ra116-250ra116.
57. Warren, T.K., et al., *Advanced antisense therapies for postexposure protection against lethal filovirus infections*. *Nat Med*, 2010. **16**(9): p. 991-4.
58. Warren, T.K., et al., *Protection against filovirus diseases by a novel broad-spectrum nucleoside analogue BCX4430*. *Nature*, 2014. **508**(7496): p. 402-5.
59. Emperador, D.M., et al., *Diagnostics for filovirus detection: impact of recent outbreaks on the diagnostic landscape*. *BMJ Glob Health*, 2019. **4**(Suppl 2): p. e001112.
60. Siya, A., et al., *Lowland grazing and Marburg virus disease (MVD) outbreak in Kween district, Eastern Uganda*. *BMC public health*, 2019. **19**(1): p. 136-136.
61. Feldmann, H., W. Slenczka, and H.D. Klenk, *Emerging and reemerging of filoviruses*. *Arch Virol Suppl*, 1996. **11**: p. 77-100.
62. Nikiforov, V.V., et al., *[A case of a laboratory infection with Marburg fever]*. *Zh Mikrobiol Epidemiol Immunobiol*, 1994(3): p. 104-6.
63. Languon, S. and O. Quaye, *Filovirus Disease Outbreaks: A Chronological Overview*. *Virology (Auckl)*, 2019. **10**: p. 1178122x19849927.

64. Alves, D.A., et al., *Aerosol exposure to the angola strain of marburg virus causes lethal viral hemorrhagic Fever in cynomolgus macaques*. *Vet Pathol*, 2010. **47**(5): p. 831-51.
65. Geisbert, T.W., et al., *Apoptosis induced in vitro and in vivo during infection by Ebola and Marburg viruses*. *Lab Invest*, 2000. **80**(2): p. 171-86.
66. Hensley, L.E., et al., *Pathogenesis of Marburg hemorrhagic fever in cynomolgus macaques*. *J Infect Dis*, 2011. **204 Suppl 3**: p. S1021-31.
67. Stroher, U., et al., *Infection and activation of monocytes by Marburg and Ebola viruses*. *J Virol*, 2001. **75**(22): p. 11025-33.
68. Ignat'ev, G.M., et al., *[Effects of tumor necrosis factor antiserum of the course of Marburg hemorrhagic fever]*. *Vestn Ross Akad Med Nauk*, 1998(3): p. 35-8.
69. Geisbert, T.W., et al., *Marburg virus Angola infection of rhesus macaques: pathogenesis and treatment with recombinant nematode anticoagulant protein c2*. *J Infect Dis*, 2007. **196 Suppl 2**: p. S372-81.
70. Warfield, K.L., et al., *Development and characterization of a mouse model for Marburg hemorrhagic fever*. *J Virol*, 2009. **83**(13): p. 6404-15.
71. Fritz, E.A., et al., *Cellular immune response to Marburg virus infection in cynomolgus macaques*. *Viral Immunol*, 2008. **21**(3): p. 355-63.
72. Martini, G.A., *Marburg virus disease*. *Postgrad Med J*, 1973. **49**(574): p. 542-6.
73. Glaze, E.R., et al., *A Comparison of the Pathogenesis of Marburg Virus Disease in Humans and Nonhuman Primates and Evaluation of the Suitability of These Animal Models for Predicting Clinical Efficacy under the 'Animal Rule'*. *Comparative medicine*, 2015. **65**(3): p. 241-259.
74. Bechtelsheimer, H., G. Korb, and P. Gedigk, *The morphology and pathogenesis of "Marburg virus" hepatitis*. *Hum Pathol*, 1972. **3**(2): p. 255-64.
75. Feldmann, H., et al., *Filovirus-induced endothelial leakage triggered by infected monocytes/macrophages*. *J Virol*, 1996. **70**(4): p. 2208-14.
76. Mehedi, M., et al., *Clinical aspects of Marburg hemorrhagic fever*. *Future Virol*, 2011. **6**(9): p. 1091-1106.
77. Olejnik, J., E. Mühlberger, and A.J. Hume, *Recent advances in marburgvirus research*. *F1000Research*, 2019. **8**: p. F1000 Faculty Rev-704.
78. Stonier, S.W., et al., *Marburg virus survivor immune responses are Th1 skewed with limited neutralizing antibody responses*. *J Exp Med*, 2017. **214**(9): p. 2563-2572.
79. Flyak, A.I., et al., *Mechanism of human antibody-mediated neutralization of Marburg virus*. *Cell*, 2015. **160**(5): p. 893-903.
80. Froude, J.W., et al., *Generation and characterization of protective antibodies to Marburg virus*. *MAbs*, 2017. **9**(4): p. 696-703.
81. Keshwara, R., et al., *A Recombinant Rabies Virus Expressing the Marburg Virus Glycoprotein Is Dependent upon Antibody-Mediated Cellular Cytotoxicity for Protection against Marburg Virus Disease in a Murine Model*. *J Virol*, 2019. **93**(6).
82. Mire, C.E., et al., *Therapeutic treatment of Marburg and Ravn virus infection in nonhuman primates with a human monoclonal antibody*. *Sci Transl Med*, 2017. **9**(384).
83. Bradfute, S.B. and S. Bavari, *Correlates of immunity to filovirus infection*. *Viruses*, 2011. **3**(7): p. 982-1000.
84. Amman, B.R., et al., *Oral shedding of Marburg virus in experimentally infected Egyptian fruit bats (*Rousettus aegyptiacus*)*. *J Wildl Dis*, 2015. **51**(1): p. 113-24.
85. Pavlovich, S.S., et al., *The Egyptian Roussette Genome Reveals Unexpected Features of Bat Antiviral Immunity*. *Cell*, 2018. **173**(5): p. 1098-1110.e18.

86. Jones, M.E.B., et al., *Clinical, Histopathologic, and Immunohistochemical Characterization of Experimental Marburg Virus Infection in A Natural Reservoir Host, the Egyptian Rousette Bat (Rousettus aegyptiacus)*. *Viruses*, 2019. **11**(3).
87. Storm, N., et al., *Antibody Responses to Marburg Virus in Egyptian Rousette Bats and Their Role in Protection against Infection*. *Viruses*, 2018. **10**(2): p. 73.
88. Subudhi, S., N. Rapin, and V. Misra, *Immune System Modulation and Viral Persistence in Bats: Understanding Viral Spillover*. *Viruses*, 2019. **11**(2): p. 192.
89. Bayne, C.J., *Origins and evolutionary relationships between the innate and adaptive arms of immune systems*. *Integr Comp Biol*, 2003. **43**(2): p. 293-9.
90. Takeuchi, O. and S. Akira, *Innate immunity to virus infection*. *Immunological reviews*, 2009. **227**(1): p. 75-86.
91. Michallet, M.C., et al., *Innate receptors for adaptive immunity*. *Curr Opin Microbiol*, 2013. **16**(3): p. 296-302.
92. Messaoudi, I., G.K. Amarasinghe, and C.F. Basler, *Filovirus pathogenesis and immune evasion: insights from Ebola virus and Marburg virus*. *Nat Rev Microbiol*, 2015. **13**(11): p. 663-76.
93. Harrison, D.A., *The Jak/STAT pathway*. *Cold Spring Harbor perspectives in biology*, 2012. **4**(3): p. a011205.
94. Ramanan, P., et al., *Structural basis for Marburg virus VP35-mediated immune evasion mechanisms*. *Proc Natl Acad Sci U S A*, 2012. **109**(50): p. 20661-6.
95. Hume, A. and E. Mühlberger, *Marburg Virus Viral Protein 35 Inhibits Protein Kinase R Activation in a Cell Type-Specific Manner*. *The Journal of infectious diseases*, 2018. **218**(suppl_5): p. S403-S408.
96. Ji, X., et al., *Mannose-binding lectin binds to Ebola and Marburg envelope glycoproteins, resulting in blocking of virus interaction with DC-SIGN and complement-mediated virus neutralization*. *J Gen Virol*, 2005. **86**(Pt 9): p. 2535-42.
97. Shishido, S.N., et al., *Humoral innate immune response and disease*. *Clinical immunology (Orlando, Fla.)*, 2012. **144**(2): p. 142-158.
98. Merle, N.S., et al., *Complement System Part II: Role in Immunity*. *Front Immunol*, 2015. **6**: p. 257.
99. Stoermer, K.A. and T.E. Morrison, *Complement and viral pathogenesis*. *Virology*, 2011. **411**(2): p. 362-373.
100. Stoermer, K.A. and T.E. Morrison, *Complement and viral pathogenesis*. *Virology*, 2011. **411**(2): p. 362-73.
101. Schroeder, H.W., Jr. and L. Cavacini, *Structure and function of immunoglobulins*. *The Journal of allergy and clinical immunology*, 2010. **125**(2 Suppl 2): p. S41-S52.
102. Doria-Rose, N.A. and M.G. Joyce, *Strategies to guide the antibody affinity maturation process*. *Current opinion in virology*, 2015. **11**: p. 137-147.
103. Chen, K. and A. Cerutti, *New insights into the enigma of immunoglobulin D*. *Immunol Rev*, 2010. **237**(1): p. 160-79.
104. Chen, K., et al., *Immunoglobulin D enhances immune surveillance by activating antimicrobial, proinflammatory and B cell-stimulating programs in basophils*. *Nat Immunol*, 2009. **10**(8): p. 889-98.
105. Woof, J.M. and M.A. Kerr, *The function of immunoglobulin A in immunity*. *J Pathol*, 2006. **208**(2): p. 270-82.
106. Schur, P.H., *IgG subclasses--a review*. *Ann Allergy*, 1987. **58**(2): p. 89-96, 99.
107. Forthal, D.N., *Functions of Antibodies*. *Microbiology spectrum*, 2014. **2**(4): p. 1-17.
108. Zhou, Z.H., et al., *The broad antibacterial activity of the natural antibody repertoire is due to polyreactive antibodies*. *Cell Host Microbe*, 2007. **1**(1): p. 51-61.

109. Panda, S. and J.L. Ding, *Natural antibodies bridge innate and adaptive immunity*. J Immunol, 2015. **194**(1): p. 13-20.
110. Ehrenstein, M.R. and C.A. Notley, *The importance of natural IgM: scavenger, protector and regulator*. Nat Rev Immunol, 2010. **10**(11): p. 778-86.
111. Beebe, D.P. and N.R. Cooper, *Neutralization of vesicular stomatitis virus (VSV) by human complement requires a natural IgM antibody present in human serum*. J Immunol, 1981. **126**(4): p. 1562-8.
112. Jayasekera, J.P., E.A. Moseman, and M.C. Carroll, *Natural antibody and complement mediate neutralization of influenza virus in the absence of prior immunity*. J Virol, 2007. **81**(7): p. 3487-94.
113. Moulton, E.A., J.P. Atkinson, and R.M. Buller, *Surviving mousepox infection requires the complement system*. PLoS Pathog, 2008. **4**(12): p. e1000249.
114. Rosendahl Huber, S., et al., *T cell responses to viral infections - opportunities for Peptide vaccination*. Frontiers in immunology, 2014. **5**: p. 171-171.
115. Swain, S.L., K.K. McKinstry, and T.M. Strutt, *Expanding roles for CD4⁺ T cells in immunity to viruses*. Nature reviews. Immunology, 2012. **12**(2): p. 136-148.
116. McCarville, J.L. and J.S. Ayres, *Disease tolerance: concept and mechanisms*. Current opinion in immunology, 2018. **50**: p. 88-93.
117. Mandl, J.N., et al., *Going to Bat(s) for Studies of Disease Tolerance*. Front Immunol, 2018. **9**: p. 2112.
118. Orr, M.T. and L.L. Lanier, *Natural killer cell education and tolerance*. Cell, 2010. **142**(6): p. 847-856.
119. Ong, E.Z., K.R. Chan, and E.E. Ooi, *Viral Manipulation of Host Inhibitory Receptor Signaling for Immune Evasion*. PLoS pathogens, 2016. **12**(9): p. e1005776-e1005776.
120. Li, S.F., et al., *Interferon-omega: Current status in clinical applications*. Int Immunopharmacol, 2017. **52**: p. 253-260.
121. Gunturi, A., R.E. Berg, and J. Forman, *The role of CD94/NKG2 in innate and adaptive immunity*. Immunol Res, 2004. **30**(1): p. 29-34.
122. Wada, H., et al., *The inhibitory NK cell receptor CD94/NKG2A and the activating receptor CD94/NKG2C bind the top of HLA-E through mostly shared but partly distinct sets of HLA-E residues*. Eur J Immunol, 2004. **34**(1): p. 81-90.
123. Parham, P. and A. Moffett, *Variable NK cell receptors and their MHC class I ligands in immunity, reproduction and human evolution*. Nat Rev Immunol, 2013. **13**(2): p. 133-44.
124. Lanier, L.L., *DAP10- and DAP12-associated receptors in innate immunity*. Immunol Rev, 2009. **227**(1): p. 150-60.
125. Kuzmin, I.V., et al., *Innate Immune Responses of Bat and Human Cells to Filoviruses: Commonalities and Distinctions*. J Virol, 2017. **91**(8).
126. Breton, C.S., et al., *A novel anti-CD19 monoclonal antibody (GBR 401) with high killing activity against B cell malignancies*. J Hematol Oncol, 2014. **7**: p. 33.
127. Chan, S., et al., *Structural Variation Detection and Analysis Using Bionano Optical Mapping*. Methods Mol Biol, 2018. **1833**: p. 193-203.
128. Sanchez-Lockhart, M., et al., *Qualitative Profiling of the Humoral Immune Response Elicited by rVSV-DeltaG-EBOV-GP Using a Systems Serology Assay, Domain Programmable Arrays*. Cell Rep, 2018. **24**(4): p. 1050-1059.e5.
129. Papenfuss, A.T., et al., *The immune gene repertoire of an important viral reservoir, the Australian black flying fox*. BMC Genomics, 2012. **13**: p. 261.
130. Zhang, C., P.L. Freddolino, and Y. Zhang, *COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information*. Nucleic acids research, 2017. **45**(W1): p. W291-W299.

131. Zhou, P., et al., *Unlocking bat immunology: establishment of Pteropus alecto bone marrow-derived dendritic cells and macrophages*. Scientific Reports, 2016. **6**: p. 38597.
132. Schuh, A.J., et al., *Egyptian rousette bats maintain long-term protective immunity against Marburg virus infection despite diminished antibody levels*. Sci Rep, 2017. **7**(1): p. 8763.
133. Schuh, A.J., et al., *Antibody-Mediated Virus Neutralization Is Not a Universal Mechanism of Marburg, Ebola, or Sosuga Virus Clearance in Egyptian Rousette Bats*. J Infect Dis, 2019. **219**(11): p. 1716-1721.
134. Lefranc, M.P., et al., *IMGT(R), the international ImMunoGeneTics information system(R) 25 years on*. Nucleic Acids Res, 2015. **43**(Database issue): p. D413-22.
135. Watson, C.T. and F. Breden, *The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease*. Genes Immun, 2012. **13**(5): p. 363-73.
136. Akula, S., S. Mohammadamin, and L. Hellman, *Fc receptors for immunoglobulins and their appearance during vertebrate evolution*. PLoS One, 2014. **9**(5): p. e96903.
137. Collins, A.M., *IgG subclass co-expression brings harmony to the quartet model of murine IgG function*. Immunol Cell Biol, 2016. **94**(10): p. 949-954.
138. Burnett, R.C., et al., *The IgA heavy-chain gene family in rabbit: cloning and sequence analysis of 13 C alpha genes*. Embo j, 1989. **8**(13): p. 4041-7.
139. Baker, M.L., M. Tachedjian, and L.F. Wang, *Immunoglobulin heavy chain diversity in Pteropid bats: evidence for a diverse and highly specific antigen binding repertoire*. Immunogenetics, 2010. **62**(3): p. 173-84.
140. Wynne, J.W. and L.-F. Wang, *Bats and viruses: friend or foe?* PLoS pathogens, 2013. **9**(10): p. e1003651-e1003651.
141. Butler, J.E., et al., *The two suborders of chiropterans have the canonical heavy-chain immunoglobulin (Ig) gene repertoire of eutherian mammals*. Dev Comp Immunol, 2011. **35**(3): p. 273-84.
142. Schountz, T., *Immunology of bats and their viruses: challenges and opportunities*. Viruses, 2014. **6**(12): p. 4880-901.
143. Bruhns, P. and F. Jonsson, *Mouse and human FcR effector functions*. Immunol Rev, 2015. **268**(1): p. 25-51.
144. Hogarth, P.M., *Fc Receptors: Introduction*. Immunol Rev, 2015. **268**(1): p. 1-5.
145. Bournazos, S. and J.V. Ravetch, *Diversification of IgG effector functions*. Int Immunol, 2017. **29**(7): p. 303-310.
146. Zucchetti, I., et al., *Origin and evolution of the vertebrate leukocyte receptors: the lesson from tunicates*. Immunogenetics, 2009. **61**(6): p. 463-81.
147. Reljic, R., *In search of the elusive mouse macrophage Fc-alpha receptor*. Immunology letters, 2006. **107**: p. 80-1.
148. Ma, Y.J. and P. Garred, *Pentraxins in Complement Activation and Regulation*. Front Immunol, 2018. **9**: p. 3046.
149. Lu, J., et al., *Pentraxins and Fc Receptor-Mediated Immune Responses*. Front Immunol, 2018. **9**: p. 2607.
150. Huang, Z., et al., *Longitudinal comparative transcriptomics reveals unique mechanisms underlying extended healthspan in bats*. Nat Ecol Evol, 2019. **3**(7): p. 1110-1120.
151. Sotero-Caio, C.G., R.J. Baker, and M. Volleth, *Chromosomal Evolution in Chiroptera*. Genes (Basel), 2017. **8**(10).
152. Koren, S., et al., *Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation*. Genome Res, 2017. **27**(5): p. 722-736.

153. Zhang, C., P.L. Freddolino, and Y. Zhang, *COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information*. *Nucleic Acids Res*, 2017. **45**(W1): p. W291-w299.
154. Lee, A.K., et al., *De novo transcriptome reconstruction and annotation of the Egyptian rousette bat*. *BMC Genomics*, 2015. **16**: p. 1033.
155. Bray, N.L., et al., *Near-optimal probabilistic RNA-seq quantification*. *Nat Biotechnol*, 2016. **34**(5): p. 525-7.
156. Kolde, R., *pheatmap: Pretty heatmaps [Software]*. 2015.
157. Pinheiro, A., et al., *Identification of a new European rabbit IgA with a serine-rich hinge region*. *PLoS One*, 2018. **13**(8): p. e0201567.
158. Gertz, E.M., et al., *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*, 2013. **65**(10): p. 749-762.
159. McBride, O.W., et al., *Localization of human variable and constant region immunoglobulin heavy chain genes on subtelomeric band q32 of chromosome 14*. *Nucleic acids research*, 1982. **10**(24): p. 8155-8170.
160. Alhakami, H., H. Mirebrahim, and S. Lonardi, *A comparative evaluation of genome assembly reconciliation tools*. *Genome Biol*, 2017. **18**(1): p. 93.
161. Lefranc, M.P., et al., *IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains*. *Dev Comp Immunol*, 2005. **29**(3): p. 185-203.
162. Vidarsson, G., G. Dekkers, and T. Rispen, *IgG subclasses and allotypes: from structure to effector functions*. *Front Immunol*, 2014. **5**: p. 520.
163. Tsakou, E., et al., *Partial versus productive immunoglobulin heavy locus rearrangements in chronic lymphocytic leukemia: implications for B-cell receptor stereotypy*. *Mol Med*, 2012. **18**: p. 138-45.
164. Rosse, I.C., et al., *Whole genome sequencing of Guzera cattle reveals genetic variants in candidate genes for production, disease resistance, and heat tolerance*. *Mamm Genome*, 2017. **28**(1-2): p. 66-80.
165. Hellman, L.T., et al., *Tracing the Origins of IgE, Mast Cells, and Allergies by Studies of Wild Animals*. *Front Immunol*, 2017. **8**: p. 1749.
166. Agnarsson, I., et al., *A time-calibrated species-level phylogeny of bats (Chiroptera, Mammalia)*. *PLoS Curr*, 2011. **3**: p. Rrn1212.
167. Dard, P., et al., *The IGHG3 gene shows a structural polymorphism characterized by different hinge lengths: sequence of a new 2-exon hinge gene*. *Hum Genet*, 1997. **99**(1): p. 138-41.
168. Arnold, J.N., et al., *The glycosylation of human serum IgD and IgE and the accessibility of identified oligomannose structures for interaction with mannan-binding lectin*. *J Immunol*, 2004. **173**(11): p. 6831-40.
169. Yoo, E.M., et al., *Differences in N-glycan structures found on recombinant IgA1 and IgA2 produced in murine myeloma and CHO cell lines*. *MAbs*, 2010. **2**(3): p. 320-34.
170. Plomp, R., et al., *Subclass-specific IgG glycosylation is associated with markers of inflammation and metabolic health*. *Sci Rep*, 2017. **7**(1): p. 12325.
171. Kiyoshi, M., et al., *Structural basis for binding of human IgG1 to its high-affinity human receptor FcγRI*. *Nat Commun*, 2015. **6**: p. 6866.
172. Lu, J., et al., *Structure of FcγRI in complex with Fc reveals the importance of glycan recognition for high-affinity IgG binding*. *Proceedings of the National Academy of Sciences*, 2015. **112**(3): p. 833-838.

173. Lu, J., et al., *Crystal structure of Fcγ receptor I and its implication in high affinity gamma-immunoglobulin binding*. J Biol Chem, 2011. **286**(47): p. 40608-13.
174. Arduin, E., et al., *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, 2015. **63**(2): p. 456-63.
175. Cerutti, A., et al., *Regulation of mucosal IgA responses: lessons from primary immunodeficiencies*. Annals of the New York Academy of Sciences, 2011. **1238**(1): p. 132-144.
176. Koenderman, L., *Inside-Out Control of Fc-Receptors*. Front Immunol, 2019. **10**: p. 544.
177. Ernst, L.K., A.M. Duchemin, and C.L. Anderson, *Association of the high-affinity receptor for IgG (Fc gamma RI) with the gamma subunit of the IgE receptor*. Proceedings of the National Academy of Sciences, 1993. **90**(13): p. 6023-6027.
178. Pyzik, M., et al., *The Neonatal Fc Receptor (FcRn): A Misnomer?* Front Immunol, 2019. **10**: p. 1540.
179. Yang, Z., et al., *Novel fluorinated pyrrolomycins as potent anti-staphylococcal biofilm agents: Design, synthesis, pharmacokinetics and antibacterial activities*. Eur J Med Chem, 2016. **124**: p. 129-137.
180. Gao, F., et al., *EasyCodeML: A visual tool for analysis of selection using CodeML*. Ecol Evol, 2019. **9**(7): p. 3891-3898.
181. Watson, C.T., J. Glanville, and W.A. Marasco, *The Individual and Population Genetics of Antibody Immunity*. Trends Immunol, 2017. **38**(7): p. 459-470.
182. Cohen-Dvashi, H., et al., *Structural Basis for a Convergent Immune Response against Ebola Virus*. Cell Host Microbe, 2020.
183. King, L.B., et al., *The Marburgvirus-Neutralizing Human Monoclonal Antibody MR191 Targets a Conserved Site to Block Virus Receptor Binding*. Cell Host Microbe, 2018. **23**(1): p. 101-109 e4.
184. West, B.R., et al., *Structural Basis of Pan-Ebolavirus Neutralization by a Human Antibody against a Conserved, yet Cryptic Epitope*. MBio, 2018. **9**(5).
185. Davis, C.W., et al., *Longitudinal Analysis of the Human B Cell Response to Ebola Virus Infection*. Cell, 2019. **177**(6): p. 1566-1582.e17.
186. Ying, T., et al., *Junctional and allele-specific residues are critical for MERS-CoV neutralization by an exceptionally potent germline-like antibody*. Nat Commun, 2015. **6**: p. 8223.
187. Ying, T., et al., *Exceptionally potent neutralization of Middle East respiratory syndrome coronavirus by human monoclonal antibodies*. J Virol, 2014. **88**(14): p. 7796-805.
188. Corti, D., et al., *Protective monotherapy against lethal Ebola virus infection by a potently neutralizing antibody*. Science, 2016. **351**(6279): p. 4.
189. Wrammert, J., et al., *Broadly cross-reactive antibodies dominate the human B cell response against 2009 pandemic H1N1 influenza virus infection*. J Exp Med, 2011. **208**(1): p. 181-93.
190. Hu, D., et al., *A broadly neutralizing germline-like human monoclonal antibody against dengue virus envelope domain III*. PLoS Pathog, 2019. **15**(6): p. e1007836.
191. Garrido, J.L., et al., *Two recombinant human monoclonal antibodies that protect against lethal Andes hantavirus infection in vivo*. Sci Transl Med, 2018. **10**(468).
192. Peng, W., et al., *Profiling the TRB and IGH repertoire of patients with H5N6 Avian Influenza Virus Infection by high-throughput sequencing*. Sci Rep, 2019. **9**(1): p. 7429.

193. Jackson, K.J., et al., *Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements*. Cell Host Microbe, 2014. **16**(1): p. 105-14.
194. Lai, M., et al., *Signatures of B-cell receptor diversity in B lymphocytes following Epstein-Barr virus transformation*. Physiol Genomics, 2019. **51**(6): p. 197-207.
195. Ma, L., et al., *Internal Duplications of DH, JH, and C Region Genes Create an Unusual IgH Gene Locus in Cattle*. J Immunol, 2016. **196**(10): p. 4358-66.
196. Shi, B., et al., *Comparative analysis of human and mouse immunoglobulin variable heavy regions from IMGT/LIGM-DB with IMGT/HighV-QUEST*. Theor Biol Med Model, 2014. **11**: p. 30.
197. Arnaout, R., et al., *High-resolution description of antibody heavy-chain repertoires in humans*. PLoS One, 2011. **6**(8): p. e22365.
198. Max, E.E., et al., *Duplication and deletion in the human immunoglobulin epsilon genes*. Cell, 1982. **29**(2): p. 691-9.
199. Oettgen, H.C., *Fifty years later: Emerging functions of IgE antibodies in host defense, immune regulation, and allergic diseases*. J Allergy Clin Immunol, 2016. **137**(6): p. 1631-1645.
200. Reams, A.B. and J.R. Roth, *Mechanisms of gene duplication and amplification*. Cold Spring Harb Perspect Biol, 2015. **7**(2): p. a016592.
201. Kelly, B.T. and M.H. Grayson, *Immunoglobulin E, what is it good for?* Ann Allergy Asthma Immunol, 2016. **116**(3): p. 183-7.
202. Sutton, B.J. and A.M. Davies, *Structure and dynamics of IgE-receptor interactions: FcepsilonRI and CD23/FcepsilonRII*. Immunol Rev, 2015. **268**(1): p. 222-35.
203. Yu, K. and M.R. Lieber, *Current insights into the mechanism of mammalian immunoglobulin class switch recombination*. Crit Rev Biochem Mol Biol, 2019. **54**(4): p. 333-351.
204. Lawrence, M.G., et al., *Half-life of IgE in serum and skin: Consequences for anti-IgE therapy in patients with allergic disease*. J Allergy Clin Immunol, 2017. **139**(2): p. 422-428.e4.
205. Bloom, J.W., et al., *Intrachain disulfide bond in the core hinge region of human IgG4*. Protein Sci, 1997. **6**(2): p. 407-15.
206. Schuurman, J., et al., *Normal human immunoglobulin G4 is bispecific: it has two different antigen-combining sites*. Immunology, 1999. **97**(4): p. 693-8.
207. Meknache, N., et al., *Human basophils express the glycosylphosphatidylinositol-anchored low-affinity IgG receptor FcgammaRIIB (CD16B)*. J Immunol, 2009. **182**(4): p. 2542-50.
208. Suzuki, K., et al., *The Fc receptor (FcR) gamma subunit is essential for IgE-binding activity of cell-surface expressed chimeric receptor molecules constructed from human high-affinity IgE receptor (Fc epsilon RI) alpha and FcR gamma subunits*. Mol Immunol, 1998. **35**(5): p. 259-70.
209. Chenoweth, A.M., et al., *The high-affinity receptor for IgG, FcgammaRI, of humans and non-human primates*. Immunol Rev, 2015. **268**(1): p. 175-91.
210. Allen, J.M. and B. Seed, *Isolation and expression of functional high-affinity Fc receptor complementary DNAs*. Science, 1989. **243**(4889): p. 378-81.
211. Barnes, N., et al., *FcgammaRI-deficient mice show multiple alterations to inflammatory and immune responses*. Immunity, 2002. **16**(3): p. 379-89.
212. Mancardi, D.A., et al., *The high-affinity human IgG receptor FcgammaRI (CD64) promotes IgG-mediated inflammation, anaphylaxis, and antitumor immunotherapy*. Blood, 2013. **121**(9): p. 1563-73.

213. Krasnov, A., S.M. Jorgensen, and S. Afanasyev, *Ig-seq: Deep sequencing of the variable region of Atlantic salmon IgM heavy chain transcripts*. Mol Immunol, 2017. **88**: p. 99-105.
214. Lopez-Santibanez-Jacome, L., S.E. Avendano-Vazquez, and C.F. Flores-Jasso, *The Pipeline Repertoire for Ig-Seq Analysis*. Front Immunol, 2019. **10**: p. 899.
215. Kreer, C., et al., *Exploiting B Cell Receptor Analyses to Inform on HIV-1 Vaccination Strategies*. Vaccines (Basel), 2020. **8**(1).
216. Galson, J.D., et al., *In-Depth Assessment of Within-Individual and Inter-Individual Variation in the B Cell Receptor Repertoire*. Front Immunol, 2015. **6**: p. 531.
217. Hong, B., et al., *In-Depth Analysis of Human Neonatal and Adult IgM Antibody Repertoires*. Front Immunol, 2018. **9**: p. 128.
218. Janeway CA Jr, T.P., Walport M, et al. , *Immunobiology: The Immune System in Health and Disease*. 5th ed. 2001, New York: Garland Science.
219. Flajnik, M.F. and M. Kasahara, *Origin and evolution of the adaptive immune system: genetic events and selective pressures*. Nature reviews. Genetics, 2010. **11**(1): p. 47-59.
220. Lei, M. and D. Dong, *Phylogenomic analyses of bat subordinal relationships based on transcriptome data*. Sci Rep, 2016. **6**: p. 27726.
221. Rodhain, F., *[Bats and Viruses: complex relationships]*. Bull Soc Pathol Exot, 2015. **108**(4): p. 272-89.
222. Chen, D., et al., *Inebilizumab, a B Cell-Depleting Anti-CD19 Antibody for the Treatment of Autoimmune Neurological Diseases: Insights from Preclinical Studies*. J Clin Med, 2016. **5**(12).
223. Bratsch, S., et al., *The little brown bat, M. lucifugus, displays a highly diverse V H, D H and J H repertoire but little evidence of somatic hypermutation*. Dev Comp Immunol, 2011. **35**(4): p. 421-30.
224. Shugay, M., et al., *VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires*. PLOS Computational Biology, 2015. **11**(11): p. e1004503.
225. Alamyar, E., et al., *IMGT((R)) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS*. Methods Mol Biol, 2012. **882**: p. 569-604.
226. Volpe, J.M. and T.B. Kepler, *Large-scale analysis of human heavy chain V(D)J recombination patterns*. Immunome research, 2008. **4**: p. 3-3.
227. Perez-Sautu, U., et al., *Target-independent high-throughput sequencing methods provide evidence that already known human viral pathogens play a main role in respiratory infections with unexplained etiology*. Emerg Microbes Infect, 2019. **8**(1): p. 1054-1065.
228. Watson, C.T., et al., *Sequencing of the human IG light chain loci from a hydatidiform mole BAC library reveals locus-specific signatures of genetic diversity*. Genes and immunity, 2015. **16**(1): p. 24-34.
229. Townsend, C.L., et al., *Significant Differences in Physicochemical Properties of Human Immunoglobulin Kappa and Lambda CDR3 Regions*. Frontiers in immunology, 2016. **7**: p. 388-388.
230. Kovaltsuk, A., et al., *How B-Cell Receptor Repertoire Sequencing Can Be Enriched with Structural Antibody Data*. Front Immunol, 2017. **8**: p. 1753.
231. Sinkora, M., J. Sinkorova, and J.E. Butler, *B cell development and VDJ rearrangement in the fetal pig*. Vet Immunol Immunopathol, 2002. **87**(3-4): p. 341-6.
232. Sela-Culang, I., V. Kunik, and Y. Ofran, *The structural basis of antibody-antigen recognition*. Front Immunol, 2013. **4**: p. 302.

233. Verkoczy, L. and M. Diaz, *Autoreactivity in HIV-1 broadly neutralizing antibodies: implications for their function and induction by vaccination*. *Curr Opin HIV AIDS*, 2014. **9**(3): p. 224-34.
234. Kitaura, K., et al., *Different Somatic Hypermutation Levels among Antibody Subclasses Disclosed by a New Next-Generation Sequencing-Based Antibody Repertoire Analysis*. *Front Immunol*, 2017. **8**: p. 389.
235. Chen, F., et al., *VH1-69 antiviral broadly neutralizing antibodies: genetics, structures, and relevance to rational vaccine design*. *Curr Opin Virol*, 2019. **34**: p. 149-159.
236. Wu, L., et al., *Fundamental characteristics of the immunoglobulin VH repertoire of chickens in comparison with those of humans, mice, and camelids*. *J Immunol*, 2012. **188**(1): p. 322-33.
237. Padlan, E.A., *Anatomy of the antibody molecule*. *Mol Immunol*, 1994. **31**(3): p. 169-217.
238. Mian, I.S., A.R. Bradwell, and A.J. Olson, *Structure, function and properties of antibody binding sites*. *J Mol Biol*, 1991. **217**(1): p. 133-51.
239. Mak, T.W. and M.E. Saunders, *8 - The Immunoglobulin Genes*, in *The Immune Response*, T.W. Mak and M.E. Saunders, Editors. 2006, Academic Press: Burlington. p. 179-208.
240. Burton, D.R. and L. Hangartner, *Broadly Neutralizing Antibodies to HIV and Their Role in Vaccine Design*. *Annu Rev Immunol*, 2016. **34**: p. 635-59.
241. Sok, D., et al., *The effects of somatic hypermutation on neutralization and binding in the PGT121 family of broadly neutralizing HIV antibodies*. *PLoS Pathog*, 2013. **9**(11): p. e1003754.
242. Corti, D., et al., *Tackling influenza with broadly neutralizing antibodies*. *Curr Opin Virol*, 2017. **24**: p. 60-69.
243. Bowers, P.M., W.J. Boyle, and R. Damoiseaux, *The Use of Somatic Hypermutation for the Affinity Maturation of Therapeutic Antibodies*. *Methods Mol Biol*, 2018. **1827**: p. 479-489.
244. Kaushik, A.K., et al., *Somatic hypermutations and isotype restricted exceptionally long CDR3H contribute to antibody diversification in cattle*. *Vet Immunol Immunopathol*, 2009. **127**(1-2): p. 106-13.
245. Gilchuk, P., et al., *Analysis of a Therapeutic Antibody Cocktail Reveals Determinants for Cooperative and Broad Ebolavirus Neutralization*. *Immunity*, 2020. **52**(2): p. 388-403.e12.
246. Kuzmin, I.V., et al., *Innate Immune Responses of Bat and Human Cells to Filoviruses: Commonalities and Distinctions*. *Journal of Virology*, 2017. **91**(8): p. e02471-16.
247. Fleri, W., et al., *The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design*. *Front Immunol*, 2017. **8**: p. 278.
248. Hevey, M., D. Negley, and A. Schmaljohn, *Characterization of monoclonal antibodies to Marburg virus (strain Musoke) glycoprotein and identification of two protective epitopes*. *Virology*, 2003. **314**(1): p. 350-7.
249. Nakayama, E., et al., *Antibody-dependent enhancement of Marburg virus infection*. *J Infect Dis*, 2011. **204 Suppl 3**: p. S978-85.
250. Mitchell, D.A.J., et al., *Epitope mapping of Ebola virus dominant and subdominant glycoprotein epitopes facilitates construction of an epitope-based DNA vaccine able to focus the antibody response in mice*. *Hum Vaccin Immunother*, 2017. **13**(12): p. 2883-2893.

251. Kajihara, M., et al., *Novel mutations in Marburg virus glycoprotein associated with viral evasion from antibody mediated immune pressure*. J Gen Virol, 2013. **94**(Pt 4): p. 876-83.
252. Williamson, L.E., et al., *Early Human B Cell Response to Ebola Virus in Four U.S. Survivors of Infection*. J Virol, 2019. **93**(8).
253. Becquart, P., et al., *Identification of Continuous Human B-Cell Epitopes in the VP35, VP40, Nucleoprotein and Glycoprotein of Ebola Virus*. PLOS ONE, 2014. **9**(6): p. e96360.
254. Moreira, G., V. Fuhner, and M. Hust, *Epitope Mapping by Phage Display*. Methods Mol Biol, 2018. **1701**: p. 497-518.
255. Wu, C.-H., et al., *Advancement and applications of peptide phage display technology in biomedical science*. Journal of Biomedical Science, 2016. **23**(1): p. 8.
256. Sivalingam, G.N. and A.J. Shepherd, *An analysis of B-cell epitope discontinuity*. Mol Immunol, 2012. **51**(3-4): p. 304-9.
257. Ahn, M., et al., *Unique Loss of the PYHIN Gene Family in Bats Amongst Mammals: Implications for Inflammasome Sensing*. Sci Rep, 2016. **6**: p. 21722.
258. Kading, R.C., et al., *Neutralizing antibodies against flaviviruses, Babanki virus, and Rift Valley fever virus in Ugandan bats*. Infect Ecol Epidemiol, 2018. **8**(1): p. 1439215.
259. Prescott, J., et al., *Rousette Bat Dendritic Cells Overcome Marburg Virus-Mediated Antiviral Responses by Upregulation of Interferon-Related Genes While Downregulating Proinflammatory Disease Mediators*. mSphere, 2019. **4**(6).
260. Storm, N., et al., *Antibody Responses to Marburg Virus in Egyptian Rousette Bats and Their Role in Protection against Infection*. Viruses, 2018. **10**(2).
261. Zhou, P., et al., *Unlocking bat immunology: establishment of Pteropus alecto bone marrow-derived dendritic cells and macrophages*. Sci Rep, 2016. **6**: p. 38597.
262. Kaur, G. and J.M. Dufour, *Cell lines: Valuable tools or useless artifacts*. Spermatogenesis, 2012. **2**(1): p. 1-5.
263. Paweska, J.T., et al., *Lack of Marburg Virus Transmission From Experimentally Infected to Susceptible In-Contact Egyptian Fruit Bats*. J Infect Dis, 2015. **212 Suppl 2**: p. S109-18.
264. Lin, S., et al., *Comparison of the transcriptional landscapes between human and mouse tissues*. Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(48): p. 17224-17229.
265. Pipes, L., et al., *The non-human primate reference transcriptome resource (NHPRT) for comparative functional genomics*. Nucleic acids research, 2013. **41**(Database issue): p. D906-D914.
266. St Laurent, G., et al., *On the importance of small changes in RNA expression*. Methods, 2013. **63**(1): p. 18-24.
267. Huddleston, J. and E.E. Eichler, *An Incomplete Understanding of Human Genetic Variation*. Genetics, 2016. **202**(4): p. 1251-1254.
268. Guo, Y., et al., *Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis*. Genomics, 2017. **109**(2): p. 83-90.
269. Ballouz, S., A. Dobin, and J.A. Gillis, *Is it time to change the reference genome?* Genome Biol, 2019. **20**(1): p. 159.
270. Shaw, A.E., et al., *Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses*. PLoS Biol, 2017. **15**(12): p. e2004086.

271. Shields, L.E., et al., *Cross-Species Genome-Wide Analysis Reveals Molecular and Functional Diversity of the Unconventional Interferon-omega Subtype*. *Front Immunol*, 2019. **10**: p. 1431.
272. Suzuki, M., C. Kato, and A. Kato, *Therapeutic antibodies: their mechanisms of action and the pathological findings they induce in toxicity studies*. *Journal of toxicologic pathology*, 2015. **28**(3): p. 133-139.
273. Miller, A.S., M.L. Tejada, and H. Gazzano-Santoro, *Methods for measuring antibody-dependent cell-mediated cytotoxicity in vitro*. *Methods Mol Biol*, 2014. **1134**: p. 59-65.
274. Wang, X., M. Mathieu, and R.J. Brezski, *IgG Fc engineering to modulate antibody effector functions*. *Protein Cell*, 2018. **9**(1): p. 63-73.
275. Kamen, L., et al., *A novel method for determining antibody-dependent cellular phagocytosis*. *J Immunol Methods*, 2019. **468**: p. 55-60.
276. Ackerman, M.E., et al., *A robust, high-throughput assay to determine the phagocytic activity of clinical antibody samples*. *J Immunol Methods*, 2011. **366**(1-2): p. 8-19.
277. Warfield, K.L., et al., *Ebola virus-like particle-based vaccine protects nonhuman primates against lethal Ebola virus challenge*. *J Infect Dis*, 2007. **196** **Suppl 2**: p. S430-7.
278. Setliff, I., et al., *High-Throughput Mapping of B Cell Receptor Sequences to Antigen Specificity*. *Cell*, 2019. **179**(7): p. 1636-1646.e15.
279. Carette, J.E., et al., *Ebola virus entry requires the cholesterol transporter Niemann-Pick C1*. *Nature*, 2011. **477**(7364): p. 340-3.
280. Nanbo, A., et al., *Ebolavirus is internalized into host cells via macropinocytosis in a viral glycoprotein-dependent manner*. *PLoS Pathog*, 2010. **6**(9): p. e1001121.
281. Staring, J., M. Raaben, and T.R. Brummelkamp, *Viral escape from endosomes and host detection at a glance*. *J Cell Sci*, 2018. **131**(15).
282. Ip, D.K., et al., *Viral Shedding and Transmission Potential of Asymptomatic and Paucisymptomatic Influenza Virus Infections in the Community*. *Clin Infect Dis*, 2017. **64**(6): p. 736-742.
283. Johnston, C. and L. Corey, *Current Concepts for Genital Herpes Simplex Virus Infection: Diagnostics and Pathogenesis of Genital Tract Shedding*. *Clin Microbiol Rev*, 2016. **29**(1): p. 149-61.
284. Hatton, O.L., et al., *The interplay between Epstein-Barr virus and B lymphocytes: implications for infection, immunity, and disease*. *Immunol Res*, 2014. **58**(2-3): p. 268-76.
285. Saghafian-Hedengren, S., et al., *Early-life EBV infection protects against persistent IgE sensitization*. *J Allergy Clin Immunol*, 2010. **125**(2): p. 433-8.
286. Sutton, B.J., et al., *IgE Antibodies: From Structure to Function and Clinical Translation*. *Antibodies (Basel)*, 2019. **8**(1).
287. Galli, S.J., M. Tsai, and A.M. Piliponsky, *The development of allergic inflammation*. *Nature*, 2008. **454**(7203): p. 445-54.
288. Szentivanyi, T., P. Christe, and O. Glaziot, *Bat Flies and Their Microparasites: Current Knowledge and Distribution*. *Front Vet Sci*, 2019. **6**: p. 115.
289. Owczarek-Lipska, M., et al., *Two loci on chromosome 5 are associated with serum IgE levels in Labrador retrievers*. *PLoS One*, 2012. **7**(6): p. e39176.
290. Haelewaters, D., T. Hiller, and C.W. Dick, *Bats, Bat Flies, and Fungi: A Case of Hyperparasitism*. *Trends Parasitol*, 2018. **34**(9): p. 784-799.
291. Kelsoe, G. and B.F. Haynes, *What Are the Primary Limitations in B-Cell Affinity Maturation, and How Much Affinity Maturation Can We Drive with Vaccination?*

- Breaking through Immunity's Glass Ceiling*. Cold Spring Harb Perspect Biol, 2018. **10**(5).
292. Xiong, H., et al., *Sequential class switching is required for the generation of high affinity IgE antibodies*. J Exp Med, 2012. **209**(2): p. 353-64.
293. Purwada, A. and A. Singh, *Immuno-engineered organoids for regulating the kinetics of B-cell development and antibody production*. Nat Protoc, 2017. **12**(1): p. 168-182.