

Genetically and Functionally Distinct Immunoglobulin Heavy Chain Locus Duplication in Bats

Author list

Taylor Pursell^{1,2}

Ashley Reers³

Artem Mikelov¹

Prasanti Kotagiri⁴

James A. Ellison⁵

Christina L. Hutson⁵

Scott D. Boyd^{1‡}

Hannah K. Frank^{3‡} (Hannah and Scott equal contribution)

¹Department of Pathology, Stanford University, Stanford, CA 94305

²Department of Microbiology & Immunology, Stanford University, Stanford, CA 94305

³Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, LA 70118

⁴Department of Immunology and Pathology, Monash University, Melbourne, Australia.

⁵Poxvirus and Rabies Branch, Division of High-Consequence Pathogens and Pathology, Centers for Disease Control and Prevention, Atlanta, GA 30329

[‡]Equal contribution

Corresponding authors hkfrank@tulane.edu; publications_scott_boyd@stanford.edu

Abstract

The genetic locus encoding immunoglobulin heavy chains (IgH) is critical for vertebrate humoral immune responses and diverse antibody repertoires. Immunoglobulin and T cell receptor loci of most bat species have not been annotated, despite the recurrent role of bats as viral reservoirs and sources of zoonotic pathogens. We investigated the genetic structure and function of IgH loci across the largest bat family, Vespertilionidae, focusing on big brown bats (*Eptesicus fuscus*). We discovered that *E. fuscus* and ten other species within Vespertilionidae have two complete, functional, and distinct immunoglobulin heavy chain loci on separate chromosomes. This locus organization is previously unknown in mammals, but is reminiscent of more limited duplicated loci in teleost fish. Single cell transcriptomic data validate functional rearrangement and expression of immunoglobulin heavy chains of both loci in the expressed repertoire of *Eptesicus fuscus*, with maintenance of allelic exclusion, bias of usage toward the smaller and more compact IgH locus, and evidence of differential selection of antigen-experienced B cells and plasma cells varying by IgH locus use. This represents a unique mechanism for mammalian humoral immunity and may contribute to bat resistance to viral pathogenesis.

Main

Bats are of immunological interest due to their unique relationship with infectious diseases. While bats host more zoonotic viruses per species than any other mammalian order¹, they often do not experience overt pathology². Multiple adaptations in their innate immunity which impart tolerance to these viruses have been characterized^{3–6}, but bat adaptive immunity remains understudied. Bats mount humoral responses to a range of viruses as evidenced by serological studies, and serology is widely used to estimate viral prevalence in bat species across the globe^{7–11}, but the genetic and molecular basis for the formation and evolution of bat B cell receptor repertoires and antibody responses is largely uncharacterized.

The genes encoding the heavy and light chains of mammalian immunoglobulin (Ig) proteins are known to be generated via DNA recombination at a single heavy chain locus and one of the two light chain loci in most mammals, designated kappa and lambda. Each of these loci contain arrays of variable (IGHV/IGKV/IGLV), diversity (IGHD), joining (IGHJ/IGKJ/IGLJ), and constant (IGHC/IGKC/IGLC) genes. The number, diversity and organization of these genes contributes to the scale of the resulting immunoglobulin repertoire and varies across species. The phylogenetic order comprising the bats, Chiroptera, contains over 1,400 species with diverse biology and modes of living, and is second only to rodents in mammalian species richness. Despite this diversity, studies characterizing immunoglobulin genes in bats have been limited to a few species within the less-speciose suborder Yinpterochiroptera (e.g. *Pteropus* spp.^{12,13}, *Rousettus aegyptiacus*¹⁴, *Rhinolophus* spp.¹⁵), leaving the vast majority of species unstudied.

We conducted comprehensive germline genetic annotation and evolutionary analysis of Ig loci in a sampling of vespertilionid bats, and single-cell transcriptome studies of B cells and plasma cells of the big brown bat (*Eptesicus fuscus*), a common North American species and

target of infectious disease research, specifically rabies virus for which it is a main reservoir species^{16–18}. We report that vespertilionid bats have dual, functional, and distinct IGH loci located on separate chromosomes in their genomes. Our findings identify a novel organization of immunoglobulin genes in mammals and terrestrial vertebrates, and provide initial functional characterization of the impact of these loci on the immune system biology of a pivotal zoonotic disease reservoir.

Two complete, functional heavy chain genomic loci in *E. fuscus*

We identified two immunoglobulin heavy chain loci in a recent high-quality, long-read genome assembly of *E. fuscus* (NCBI genome assembly DD_ASM_mEF_20220401). Rigorous determination of locus organization was enabled by long-read DNA sequencing permitting full assembly of the complex, repetitive loci in their correct chromosomal localization. The smaller locus (IGH locus A, A-IGH) spans 272 kilobases on chromosome 5; the other (IGH locus B, B-IGH) spans 918 kilobases on the telomeric end of chromosome 24. We annotated the immunoglobulin germline genes of these two parallel immunoglobulin heavy chain loci (**Fig. 1**). The overall structure and orientation of both loci are similar to humans with sequential arrays of variable (V_H), diversity (D), and joining (J_H) gene segments followed by multi-exon constant region genes (C_H). No flanking genes appear to have been transposed as A-IGH is flanked by TMEM212 and CRIP1, both genes known to flank the IGH locus in other mammals, while B-IGH is flanked by MXL.

Each locus has a single, functional copy of each of the five major mammalian constant region isotype genes, IGHM, IGHD, IGHG, IGHE, and IGHA (**Fig. 1A-B; Supplementary Table 1 - 2**). The order of the IGHCs is consistent with other mammalian species, with IGHM immediately downstream of the IGHJ gene cluster followed by IGHD, IGHG, IGHE, and IGHA.

Preliminary annotations suggest that the exonal structure of each gene is conserved within all genes except for IgD. Consistent with previous work¹⁹, the CH2 exon from IGHD is absent from both gene copies and the hinge H2 exon is fused to the CH3 exon similar to IGHD structures observed in rodents²⁰. Additionally, we identified a transmembrane domain previously thought missing in the IGHD gene annotation.

Recombination and isotype switching is initiated by cytidine deamination mediated by the activation induced cytidine deaminase enzyme (AID) which preferentially uses the 5'-AGCT-3' motif as a hotspot²⁰⁻²². Increased density of canonical mammalian switch motifs immediately upstream of the CH1 domains for each IGHM, IGHG, IGHE, and IGHA in both loci (**Fig. 1A-B**) supports functionality and conserved mechanisms for AID-dependent class switching as observed in other mammalian species. Similar to humans, there is no switch region observed between genes coding for IgM and IgD, consistent with alternative splicing for coexpression of these isotypes²³⁻²⁵. Interestingly, B-IGH has a 17.5 kb region with increased switch motif density between B-IGHG and B-IGHE that is not observed in A-IGH which might affect the rate and efficiency of class-switching at this locus.

The IGHJ gene cluster is composed of 6 functional genes on A-IGH and 5 functional genes and one ORF on B-IGH (**Supplementary Table 3**). The canonical di-glycine bulge WGXXG motif is encoded in all IGHJ segments and a 23-base pair spacer is observed in the recombination signal sequence (RSS) upstream of each J_H. All inter-chromosome homologs, except the ORF, share at least 90% nucleotide identity (**Extended Data Fig. 1A**).

Diversity genes were identified by homology as well as RSS searching in the intergenomic space between the most 3' IGHV and the most 5' IGHJ. We identified 10 IGHDs on A-IGH and 9 on B-IGH (**Supplementary Table 4**). The IGHDs are more divergent than

IGHJs between the loci with only two gene pairs sharing >90% homology (**Extended Data Fig. 1B**). Thus with the duplication, the effective germline IGHD diversity is only slightly smaller than that of the human locus which has 27 germline IGHDs. Notably, the germline IGHDs in A-IGH are longer (mean_A=18 nt, mode_A= 19 nt) than those in B-IGH (mean_B = 16.67 nt, mode_B = 10 nt), potentially contributing to the longer complementarity determining region 3 heavy chain (CDR-H3) loops of sequences originating from this locus, as discussed further below.

While A-IGH encodes just 33 IGHV genes (27 functional, 5 pseudogenes, and a single ORF) (**Supplementary Table 5; Fig. 1A**), the larger B-IGH locus encodes 99 germline IGHVs, of which 48 are functional, 46 are pseudogenes, and 5 are ORFs (**Supplementary Table 6; Fig. 1B**). Interestingly, the percent of pseudogenes is much lower in A-IGH (15%) compared to B-IGH (46.5%) (**Fig. 1A-B**). Both loci contain functional genes from the three major human IGHV clans but differ in the proportions of genes from each of the seven human V_H families (**Fig. 1A-C**). As observed with the IGHJ genes, there are highly similar IGHVs between the loci including three genes that have over 90% nucleotide identity (**Extended Data Fig. 1C**). Within clan III (containing the IGHV3 gene family), we observe expansion of functional genes in B-IGH compared to A-IGH. Proportionally, A-IGH is enriched for functional IGHV1 family genes (**Fig. 1A**) while B-IGH contains the only functional IGHV7 and IGHV6 family genes (**Fig. 1B**). These patterns between the loci are consistent with a model in which a full ancestral IGH locus duplicated, followed by separate gene retention, losses and duplications at each locus.

IGH duplication is common in vespertilionid bats

Comparison of the *E. fuscus* IGHM regions with other bat species revealed that this IGH locus duplication is present across many vespertilionid species (though not all, e.g., *Antrozous pallidus*, data not shown). Each IGH locus forms a separate, well supported clade. Phylogenetic

relationships within a given IGHM locus are consistent with the phylogenetic relationships between the species themselves, supporting a single duplication of the IGH locus in the vespertilionid ancestor or very early in the family's history (**Fig. 2**). All IGHM genes (except *Myotis brandtii* NW005353568, which had too much missing data) have CH1, CH2, CH3, CH4-S and M domains without stop codons. Additionally, all cysteine residues present in the human IGHM are also present in the bats, suggesting the same disulfide bonds that stabilize the structure of the immunoglobulin in humans are present in bats.

Conserved and divergent constant region features related to Fc receptor binding

The key protein features of Ig heavy chain constant regions responsible for contacting Fc receptors to initiate effector functions are well described^{26,27}. To understand potential for locus specific effector functions of IGHG, we looked for conservation of features that mediate Fc receptor binding: (1) the conserved N-glycosylation site at Asn²⁹⁷; (2) CH2 region residues near the lower hinge found in IgG1, IgG3 and IgG4 (Leu²³⁴ Leu²³⁵ Gly²³⁶ Gly²³⁷); (3) CH2 BC loop Asp²⁶⁵; and (4) CH2 FG loop (Ala³²⁷ Leu³²⁸ Pro³²⁹ Ala³³⁰ Pro³³¹). The IGHGs from both *E. fuscus* loci contain the conserved N-glycosylation site, conserved Asp²⁶⁵, and the canonical CH2 FG loop, only with an Iso³²⁸ instead of a Leu³²⁸. The cytosolic tails are also highly similar to humans containing the canonical motifs (DYXNM and SSVV²³ [SSVA in *E. fuscus*]) for intracellular signaling.

The hinge length and sequence play an important role in the function of different IgG subtypes in humans. We found a single hinge exon (H) for each IGHG similar to IgG1, IgG2, and IgG4 in humans, but the sequence differs between A-IGHG and B-IGHG and between *E. fuscus* and humans. Both bat IGHG H exons encode only two cysteines similar to IGHG4 H in humans, however the lengths of the bat IGHG H regions are more similar to IGHG1 H and IGHG3 H2-

H4 genes. A-IGHG H possesses the run of TTPPC found in the IGHG3 H2-4 genes, while B-IGHG H has TTTPC which is more similar to the TTHTC of IGHG3 H1 gene. Additionally, in the human CH2/ lower hinge region, the flexibility of residues Pro²³²-Pro²³⁸ is critical for Fc receptor binding²⁶; both A-IGHG and B-IGHG encode a Cys²³² instead of a Pro²³² and a Thr²³⁴ instead of a Leu²³⁴, suggesting a modified interaction compared to human constant region binding to Fc receptors.

***E. fuscus* has retained lambda and lost kappa light chains**

Antibodies are made up of heavy and light chains. In mammals, light chains are of two types, kappa (κ) or lambda (λ). Salmon and rainbow trout, both species of fish that have a duplicated IGH locus, also express multiple types of light chains^{28,29}. In contrast, birds have a single light chain, lambda (λ)³⁰⁻³² which is postulated to be the result of genome size restraints linked to their high metabolic rate and adaptation to flight. In light of the IGH duplication, we asked whether *E. fuscus* has two light chain loci, as do other mammals and fish, or a single light chain locus, similar to birds. We identified a single light chain lambda locus (**Fig. 3A; Supplementary Table 7-9**). Neither the NCBI eukaryotic genome annotation pipeline³³, IgDetective³⁴ nor initial BLAST³⁵ searching detected any kappa variable, joining, or constant-like genes in the current assembly. In humans, the IGK locus is on chromosome 2 near the centromere, flanked on the telomeric side by RPIA in reverse orientation and EIF2AK3 in forward orientation. In *E. fuscus*, these genes are found in close proximity on chromosome 9. A focused search for IGK-like genes from 3.5Mb upstream of RPIA to 1Mb downstream of EIF2AK3 revealed three IGKJ but only one was preceded by two overlapping RSS sequences, one with a 12bp spacer and the other with a 23bp spacer (**Fig. 3B**). However, no IGKC was found in between the IGKJ and RPIA genes, nor were any IGKV genes found in the 5Mb region, suggesting loss of the IGK locus in the

evolutionary history of *E. fuscus*. We additionally searched single cell data presented below for evidence of a second light chain but found only IgL CDR3 in 89.8% of B lymphocytes and 96.7% of the cells for which we have complete, productive IGH, with no evidence of kappa light chains. Together, these data support that *E. fuscus* has no functional or expressed kappa light chains, consistent with previous work³⁶.

The *E. fuscus* lambda locus on chromosome 23 (contig NC_072495.1) spans only 506 kilobases despite encoding 127 variable (V λ , Table 7) genes, 12 joining (J λ , Table 8), and 12 constant (C λ , Table 9) genes. The locus is compact compared to the human kappa locus that encodes 116 genes but spans 1.9 megabases (Mb). The organization of J λ and C λ genes mirrors other species, such that each C λ gene is preceded by one J λ gene, forming 12 J-C gene clusters (**Fig. 3A**). Each J λ gene is flanked on the 5' end by an RSS with a 12-base spacer and the V λ genes with a 3' RSS with a 23-base spacer as is observed in other species. Notably, the locus encodes nine functional V λ genes downstream of the J λ - C λ gene cassettes in an inverted orientation (**Fig. 3A**). This organization is also seen in the equine lambda locus³⁷.

We found V λ genes from all five human clans and all V families except V λ 9 in *E. fuscus* (**Fig. 3C, 3E; Supplementary Table 7**). Of the 127 V λ genes, 118 are functional, 4 are open reading frames, and 5 are pseudogenes (**Fig. 3D**), nearly double the number of functional genes found in the human lambda and kappa light chain loci combined. In analyzing the rearranged repertoire, all functional V λ genes, three ORFs, and a pseudogene, IGLV1-102^P, were identified in productive rearrangements. Sequences using the inverted V genes comprised 6% of the total repertoire, indicating that functional rearrangement does occur despite the non-canonical orientation.

Distinct B cell phenotypes within *E. fuscus* spleen

Different B cells subsets serve different functions during an immune response, as do the antibodies they produce. To investigate *E. fuscus* B cell and plasma cell phenotypes, we performed single cell RNA-seq on cryopreserved splenocytes from four individual bats using the 10x platform. After pre-processing, batch correction, and removal of low quality cells, 62,747 cells were analyzed. Using canonical transcriptional markers present in the current genome annotation, we identified clusters corresponding major immune cell types (**Fig. 4A**). We identified two B lymphocyte lineage clusters, one enriched for B cell markers (i.e., *MS4A1* and *CD19*) associated with naive and memory B cells while the other was enriched for plasma cell markers (i.e., *XBPI*, *MZB1*, *PRDM1*, *SDC1* and *JCHAIN*) (**Fig. 4B**).

In humans, there are well-defined subsets of B cells found in the spleen with different activation states as well as stages of development³⁸. Many subsets described in humans have relatively similar counterparts in mice³⁹, but are less well understood in other mammals. Previous studies have identified B lymphocytes in scRNAseq analysis in various bat species^{40–42} but functional B cell subsets have not been fully defined, in part due to a lack of Ig gene annotations. Therefore, we undertook a focused analysis of B lymphocyte lineage subsets in *E. fuscus*. We identified a cluster analogous to germinal center (GC) cells that express canonical human markers such as *BCL6*, *BACH2* and *AICDA* (**Fig. 4D**). Naive, transitional, and memory B cells showed less distinct cluster separation. This is consistent with other single cell studies in various bat species^{40–42} in which clusters were mixed or defined by non-canonical genes of unknown cell phenotype specificity. We found clusters enriched for memory-associated genes (e.g., *NOTCH2*⁴³, *VIM*, *ZBTB32*⁴⁴, *CD27*), and others enriched (i.e., $\geq 50\%$) for B cells with low somatic hypermutation (SHM) frequencies ($< 1.5\%$), and expression of IGHM/IGHD isotypes presumed to be naive B cells (**Extended Data Fig. S2**).

Further genome annotation curation and experimental validation will be needed to fully define the gene expression patterns of naive and memory B cell populations, but we adopted a functional classification of naive B cells as those expressing IGHM or IGHD without mutation (SHM <1.5%). We grouped the remaining somatically mutated cells into non-class switched B cells (non-csB-cells) or class switched B cells (csB-cells) based on their expression of IGHM/D or other isotypes, respectively.

In contrast, plasma cell phenotype clusters were relatively homogeneous with elevated expression of canonical markers (i.e., *IRF4*, *MZB1*, *JCHAIN*, *XBPI*) except for a cluster of IgA class-switched plasma cells. This cluster is enriched for *RUNX2*, a transcription factor shown to induce IgA class switching in mature B cells during BCR signaling^{45–48}. We divided the plasma cells into class-switched antibody secreting cells (csASC) and non-class switched antibody secreting cells (non-csASC).

Rearrangement and usage of the dual IGH loci in *E. fuscus*

Having determined that numerous vespertilionid bats have dual heavy chain loci with different sizes and gene segment composition, we sought to answer key questions about the use of these loci in B cells and plasma cells, specifically: 1) whether there is allelic exclusion preventing each B cell from making two different heavy chains; 2) whether inter-locus V-D-J rearrangements occur; 3) how rearranged VDJ products of the two loci differ in gene segment and CDR-H3 characteristics; 4) whether locus usage is correlated with the B cell or plasma cell phenotype; 5) whether class switching or SHM processes differ between the two loci; and 6) whether there is evidence of selection favoring particular IGHV genes in antigen-experienced B cells and plasma cells.

Strong allelic exclusion of the dual IGH loci and evidence for A-IGH rearranging before B-IGH

Allelic exclusion is the process that ensures each B cell expresses only one heavy and one light chain sequence despite having two chromosomal copies of each locus, and two different light chain loci in most mammals. It is considered to be an important mechanism for controlling autoreactivity that could result if B cells expressed multiple antigen receptors specifically self-reactive and pathogen-specific receptors on the same cell which would allow for maturation to occur resulting in a cell where response to a pathogen could trigger autoimmunity. We hypothesized that allelic exclusion in bats with dual IGH loci would ensure that only a single IGH locus would be functionally rearranged to give an in-frame expressed protein product in each B cell or plasma cell. Using custom primers, we generated enriched and sequenced BCR libraries for single cells and assembled IGH rearrangements with CDR-H3s for 9,922 (68%) of the 14,618 B lymphocytes identified in the gene expression dataset. After removing low confidence assemblies, we found only 198 cells appearing to have two expressed in-frame CDR-H3 transcripts, or 2.0%. Given potential artifacts of single cell sequencing such as incomplete removal of doublet events or RNA from lysed cells associating with the surface of other B cells, we interpret these findings as most consistent with allelic exclusion acting at the dual IGH loci.

To address IGH locus usage in B cells of different cellular phenotype, and to assess whether one locus is rearranged before the other in the formation of the naive B cell repertoire, we assembled full length BCR heavy sequences from 5,010 single B cells or plasma cells. Both A-IGH and B-IGH are used with productive rearrangements in all phenotypic cell subsets (**Fig. 4E**). Clusters 0, 1, 2, 5, and 10 contain predominantly naive B cells expressing IgM, while IgG dominates the GC and ASC clusters (**Fig. 4F; Extended Data Fig. 2**). We assume that B cell

precursor IGH V-D-J rearrangement attempts are successful in forming an in-frame product one-third of the time because the V and J genes must be in the same reading frame in the final rearrangement, and exonuclease chewback of V, D and J segment ends together with random base addition by the terminal deoxynucleotidyl transferase at the V-D and D-J junctions can alter the J reading frame relative to the V reading frame. Therefore, sequential attempts to rearrange both copies of A-IGH and then B-IGH would result in a ratio of A-IGH: B-IGH usage in naive cells of approximately 2.25, whereas random selection of loci to rearrange would result in a 1:1 ratio. The mean ratio of A-IGH to B-IGH locus usage we observed in naive cells is 3.41 (**Fig. 4G**). This bias supports a model of ordered rearrangement and allelic exclusion such that A-IGH rearrangements are attempted first, and only if no productive V-D-J rearrangement is obtained is rearrangement attempted at B-IGH, resulting in a larger percentage of naive cells with A-IGH BCRs compared to B-IGH. Other possible explanations for this ratio of A-IGH to B-IGH could include strong negative selection against B cells expressing the B-IGH locus in the developing B cell.

No evidence of cross loci rearrangement.

We assessed whether VDJ rearrangement occurs between loci by determining the chromosome of origin for each BCR by looking for agreement between the origin of the V gene and the J gene. The origin of the C-gene was not used due to lack of polymorphisms upstream of the enrichment primers retained after library preparation. The D gene origin was also not considered due to the greater uncertainties in reliably identifying sequences derived from D genes in rearrangements, and the level of homology across chromosomes for certain D-genes. Over 95% of sequences had a V gene and a J-gene from the same locus. We found similar patterns and no evidence of cross-loci rearrangement in the sequenced bulk BCR repertoires (*data not shown*).

Although we cannot fully rule out the possibility of inter-locus recombination, it is likely that the apparent locus mismatches are due to errors in germline gene calling related to the high homology between certain V_H, D, and J_H genes between the loci (**Extended Data Fig. 1**), and our limited knowledge of allelic variation between *E. fuscus* individual animals. We therefore used the V gene locus alignment to define the origin of each IGH sequence, as the V region is longer than the J region, and therefore far more likely to have the correct locus call.

Differences in VDJ rearrangements from the dual IGH loci

As discussed above, the germline repertoire is largely similar across different loci but exhibits notable differences that likely contribute to the diversity of expressed repertoires. We hypothesized that rearranged VDJs from different IGH loci exhibit distinct V gene segment usage and complementarity-determining region 3 (CDR-H3) features. To capture a representative sample, expressed immunoglobulin rearrangements were enriched for IgM and IgG isotypes using 5'-RACE cDNA from flash frozen splenic tissue of six animals yielding a total of 307,312 productive BCR sequences for the following analysis.

Genes from the VH3 family are the most common in both the IgM and IgG repertoires for both loci (**Extended Data Fig. 3A-B**). However, for A-IGH, VH3 family gene usage significantly decreases in IgG compared to IgM ($p < 0.05$ by clone-wise non-paired Wilcoxon rank-sum test), while VH1 family gene usage significantly increases ($p < 0.05$ by clone-wise non-paired Wilcoxon rank-sum test). For B-IGH, there is almost no VH1 usage and heavy usage of VH7 family genes which significantly increases in IgG clones compared to IgM ($p < 0.05$ by clone-wise non-paired Wilcoxon rank-sum test). These data demonstrate differences in the V gene usage during generation of the naive repertoire. Furthermore, differences in IGHG gene usage support locus specific antigen-driven selection.

Junctional diversity was assessed by comparing the length V-D (N1) and D-J (N2) regions between chromosomes for IGHM associated sequences. The N1 region is significantly shorter for A-IGH ($\text{mean}_{\text{A-IGH}} = 3.1$ nucleotides) compared to B-IGH ($\text{mean}_{\text{B-IGH}} = 3.4$ nucleotides; $p < 0.05$ by clone-wise non-paired Wilcoxon rank-sum test) while N2 region is not significantly different ($\text{mean}_{\text{A-IGH}} = 1.76$ nucleotides; $\text{mean}_{\text{B-IGH}} = 1.78$ nucleotides; $p = 0.5$ by clone-wise non-paired Wilcoxon rank-sum test; **Extended Data Fig. 3C**). Overall, *E. fuscus* shows fewer N nucleotide additions compared to humans that average 7.7 and 6.5 nucleotides N1 and N2 respectively⁴⁹. These limited nucleotide additions likely function to complete partially encoded codons rather than introducing new non-templated amino acids as observed in humans. Thus, the junctional diversity of these bats appears to be constrained compared to humans⁴⁹ leading to a more germline-focused IGHM repertoire.

The length of CDR-H3s from A-IGH is significantly longer than B-IGH for both IGHM ($\text{mean}_{\text{A-IGH}} = 10.2$ amino acids, $\text{mean}_{\text{B-IGH}} = 9.3$ amino acids not including the initial cysteine and terminal tryptophan residues; $p < 0.01$ by clone-wise non-paired Wilcoxon rank-sum test) and IGHG ($\text{mean}_{\text{A-IGH}} = 10.6$ amino acids, $\text{mean}_{\text{B-IGH}} = 9.1$ amino acids not including the initial cysteine and terminal tryptophan residues; $p < 0.01$ by clone-wise non-paired Wilcoxon rank-sum test; **Extended Data Fig. 3D**), owing primarily to the longer IGHDs encoded by the A-IGH locus.

Evidence of selection pressure on IGH locus usage in B cell subsets

To assess for selection pressures affecting A-IGH or B-IGH usage in different B cell subsets, we compared the ratio of locus usage across the major B cell and plasma cell phenotypic groups, compared to naive B cells. The A-IGH:B-IGH ratio in non-csB cells is lower compared to naive B cells and it remains relatively unchanged in csB cells (**Fig. 4G**). The A-IGH bias is only lost in

non-csASC, mean = 1.26, which is significantly lower than in naive cells ($p = 0.05$ by non-paired Wilcoxon rank; **Fig. 4G**). A Chi-square test comparing the frequency of A-IGH and B-IGH usage across the different B cell subsets showed a highly significant difference ($p < 0.001$), indicating a strong association between B cell phenotype and locus usage.

Decreased class switching in the B-IGH locus

Notably, IgA csB cells and csASCs disproportionately used the A-IGH locus, indicating that this locus may have particular importance for mucosal humoral responses dependent on IgA (**Fig. 4F**). Further, ASCs using B-IGH were more likely to be non-class switched, expressing IgM, compared to those using A-IGH (**Fig. 4F**). Decreased class switching to IgA in the B-IGH locus could be related to the unusual, oversized switch region between IGHG and IGHE compared to A-IGH (**Fig. 1B**), but we observed relatively few IgE-expressing B cells or ASCs. Therefore, despite the greater diversity of IGHV genes in the B-IGH locus, it is more often used in non-csASCs such as those that are described in other mammals as being part of primary and extrafollicular humoral responses^{50–52}. Overall, these data support a model in which the B-IGH locus is less efficient at class-switching, potentially due to decreased recruitment or accessibility of AID, or that B cells using the B-IGH locus are less likely to enter immunological niches promoting class switching such as the germinal center.

Decreased somatic hypermutation in the B-IGH locus.

To further evaluate functional differences between the two IGH loci, we measured the frequency of SHM changes in VDJ rearrangements. In mammals, SHM is due to focused recruitment of the AID enzyme to the IG loci, followed by DNA repair processes triggered by initial cytidine deamination, and is the basis for affinity maturation in which B cells expressing higher affinity for antigen preferentially expand within the stimulated clone. To look for potential SHM

differences between loci, we analyzed IGH sequences with complete VDJ assemblies (5,010 of the single B lymphocytes, 34% of total B cells identified, 50% of those with a CDR3). Rates of somatic hypermutation ranged up to 22.7%. As expected, we observed the lowest mean SHM in IgM sequences (1.8%) and higher mean rates in IgA and IgG (4.8% and 4.1%, respectively) in the B cell clusters. The SHM rates were not significantly different in ASCs compared to non-naive B cells for all isotypes except IgM ($p < 0.01$, clone-wise non-paired Wilcoxon rank-sum test).

For A-IGH, class-switched cells had significantly higher levels of SHM compared to non-class-switched cells within ASCs ($p < 0.001$, clone-wise Kruskal-Wallis rank sum test followed by Dunn's multiple comparisons) (**Fig. 5A, left panel**). This is consistent with what is observed in humans and suggests similar mechanisms likely result in class switching early in activation resulting in low mutation IgM. However, this is not observed when comparing cs to non-cs B cells for A-IGH ($p = 0.50$ by clone-wise Kruskal-Wallis rank sum test followed by Dunn's multiple comparisons). For B-IGH, however, the SHM of non-class switched cells is higher than class switched B-cells and there is no difference in ASCs (**Fig. 5A, right panel**). This supports the hypothesis that B-IGH cells remain non-class switched through multiple rounds of SHM compared to class switched cells. Alternatively, once class switching has occurred, the recruitment of AID is impaired. There is no significant difference in SHM rate of csB-cells compared to csASCs within either locus. The level of SHM in non-csB-cells is significantly higher than in non-csASCs for both B-IGH and A-IGH ($p < 0.001$ by clone-wise Kruskal-Wallis rank sum test followed by Dunn's multiple comparisons) conforming to canonical trends in SHM.

The SHM rates in csB-cells are significantly higher in A-IGH sequences compared to B-IGH ($p < 0.001$ by clone-wise non-paired Wilcoxon rank-sum test) as well as in csASCs ($p < 0.001$ by clone-wise non-paired Wilcoxon rank-sum test), but there was no significant difference in non-class switched cells or GCs (**Fig. 4B**). These findings and statistical significances were confirmed by permutation tests (**Extended Data Fig. 4**). Together, these results support the hypothesis that the ability of AID to access each locus is different.

Greater evidence for selection pressure on SHM changes in A-IGH

Antigen experience shapes the B cell receptor repertoire through selective expansion of B cells with high-affinity receptors. Positive selection for improved antigen binding occurs mainly in the complementary-determining regions (CDRs) of an antibody, while negative selection against destabilizing mutations acts most strongly on the intervening structural framework regions (FWRs). Selection scores (sigma, Σ) are higher in complementary determining region (CDR) compared to the framework regions for both A-IGH and B-IGH (**Fig. 4C**). This supports similar mechanisms of affinity maturation in both loci of *E. fuscus* as observed in other species in which the antigen contacting regions are under positive selective pressure while other regions are not. However, positive selection ($\Sigma > 0$) only reached significance for the CDRs of A-IGH, but not B-IGH (**Fig. 4C**). This is likely due to the overall lower levels of SHM in B-IGH sequences resulting in the inability to detect positive selection above background without the aid of a specialized model to detect mutational hot and cold zones as is used in humans. Furthermore, this evidence of positive selection is only found in IgG and IgA, but not IgM for A-IGH sequences (**Fig. 4D**). Lastly, the only B-IGH sequences that demonstrate positive selection are associated with GC B cells. As stated above, this is likely due to the overall lower levels of somatic hypermutation in B-IGH sequences, except GC B cells (**Fig. 4B**). One possible

explanation for these data would be if A-IGH more easily engages AID compared to B-IGH. The data could also suggest a propensity for B cells using A-IGH to engage in follicular responses while B cells using B-IGH may participate more often in extrafollicular responses.

A novel strategy for mammalian humoral immunity

Our analysis of high-quality long-read chromosome level assemblies for *E. fuscus* and other vespertilionid bats identified novel IGH locus duplications, highlighting the limited extent to which these complex, repetitive loci have been explored in mammals and other vertebrates. These bats share many common features of humoral immunity with other mammals such as allelic exclusion, somatic hypermutation, evidence of antigen-driven selection, and class switching, but also provide evidence for unique functional specialization of the two IGH loci.

The A-IGH and B-IGH germline loci differ notably in size, gene segment complexity, D gene lengths, and class switch region distributions, but both have similar constant region composition and can produce IgM, IgD, IgG, IgE and IgA. VDJ rearrangements from the shorter, less V gene rich A-IGH locus are more than 3-fold more common than rearrangements from B-IGH in naive B cells, supporting a model in which rearrangement of A-IGH is attempted at both chromosome copies in developing B cell precursors prior to attempting to rearrange the B-IGH locus. A-IGH rearrangements have longer CDR-H3s than those of B-IGH, adding diversity for this key protein loop that is often involved in antigen binding. The greater diversity of B-IGH V segments encoding CDR-H1 and CDR-H2 loops is initially deployed in a smaller fraction of the naive B cell repertoire in these bats.

Both IGH loci are used in antigen-experienced B cells and plasma cells whose BCRs contain SHM changes with or without class-switching, but A-IGH locus rearrangements have consistently higher SHM and are more often class switched. These data could indicate that B

cells or plasma cells expressing A-IGH rearrangements are more likely to engage in germinal center reactions in the bats, or could point to an overall deficiency in the ability of the B-IGH locus to recruit the AID enzyme complexes required for both SHM and class-switching. Functional consequences of these divergent loci are likely to be significant; for example, in these bats most IgA, the key antibody isotype for mucosal immunity, derives from the A-IGH locus with its more limited repertoire of IGHV genes. Nonetheless, usage of the B-IGH locus increases in antigen-experienced B cells and plasma cells, suggesting that the greater diversity of IGHV gene segments in this locus provide greater possibilities for recognition of foreign antigens and potential pathogens. We speculate that in these bats the fraction of their secreted antibody populations derived from B-IGH, characterized by less-frequent class-switching and lower SHM, but higher gene segment diversity, could provide a broad, lower-affinity arm to their humoral immune responses, complemented by A-IGH-derived antibody responses that have higher SHM frequencies, greater evidence of antigen-driven selection for increased affinity, and higher rates of class-switching to IgG or IgA for potentially altered effector functions.

The combination of these humoral immunity types could contribute, together with the innate immune antiviral adaptations reported in many bat species^{53,54}, to their greater resistance to viral pathogens and their roles as vectors for asymptomatic or less-symptomatic transmission of zoonotic diseases to humans. We note that the duplication and retention of two IGH loci is particularly striking in bats, because flying species typically have genomes of reduced size⁵⁵; consistent with this, bats have smaller genomes than most other mammals⁵⁶. Thus, there may be a significant selective advantage to having two functional IGH loci that outweighs the pressures to reduce genome size.

The finding of two functional IGH loci in vespertilionid bats is the most striking example of emerging themes in bat immunity research: that bats share the same molecular components of immunity with their mammalian relatives but show bat-specific adaptations; and that gene family evolution is critical for bat immune adaptation^{57,58}. It also emphasizes the differences between bat species, as we found no evidence of the IGH duplication outside of the Vespertilionidae, the same family in which recent work has found a duplication of an important antiviral protein, protein kinase R⁵⁹.

Further testing of the functional roles of the A-IGH and B-IGH loci will benefit from studies of vaccination and infectious challenges in vespertilionid bats. Analysis of the dual IGH loci should provide a unique internally-controlled opportunity to elucidate the fundamental genetic and epigenetic mechanisms governing the formation and evolution of B cell receptor repertoires in intact primary B cells.

Acknowledgments:

We would like to thank Yana Safonova for insightful discussions regarding the presented research. We are grateful to the following organizations for funding the work: National Science Foundation IOS (2032157; HKF), National Institutes of Health NRSA T32 (T32OD011121, TP), National Institutes of Health NIAID (5 R21 AI169548; HKF), Life Sciences Research Foundation Fellowship (HKF), Open Philanthropy Project (HKF), Stanford Woods Institute for the Environment Environmental Venture Program (SDB; HKF), Stanford Center for Computational, Evolutionary and Human Genomics Postdoctoral Fellowship (HKF), National Institutes of Health grants ‘Molecular and Cellular Immunobiology’ (5 T32 AI07290; HKF), Stanford School of Medicine Dean’s Postdoctoral Fellowship (HKF), funds from the Stanford Pathology Department and an endowment from the Crown foundation (SDB). This work was

supported by funding to The Viral Emergence Research Initiative (viralemergence.org) from the U.S. National Science Foundation, including NSF BII 2021909 and NSF BII 2213854 (HKF). We thank the skilled staff of the Stanford Clinical Genomics Laboratory for assistance with sequencing experiments. The authors would also like to thank the invaluable contributions and animal care assistance supported by members of the Poxvirus and Rabies Branch and the Comparative Medicine Branch at CDC, Atlanta, GA.

Author contributions Conceptualization: TP, SDB, HKF; Methodology: TP, AR, AM, PK, JAE, SDB, HKF; Software: TP, AM, PK; Formal analysis: TP, AR, HKF; Samples: JAE, CLH; Data Curation: TP, AR, HKF; Writing-Original Draft: TP, SDB, HKF; Writing- Review & Editing: TP, AR, AM, PK, JAE, CLH, SDB, HKF; Visualization: TP, AM; Supervision: SDB, HKF; Funding acquisition: JAE, SDB, HKF.

Competing interest

The authors declare no competing interests

Disclaimer:

Use of trade names and commercial sources are for identification only and do not imply endorsement by the US Department of Health and Human Services or the Centers for Disease Control and Prevention. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of their institutions.

Additional information

Supplementary information is available for this paper.

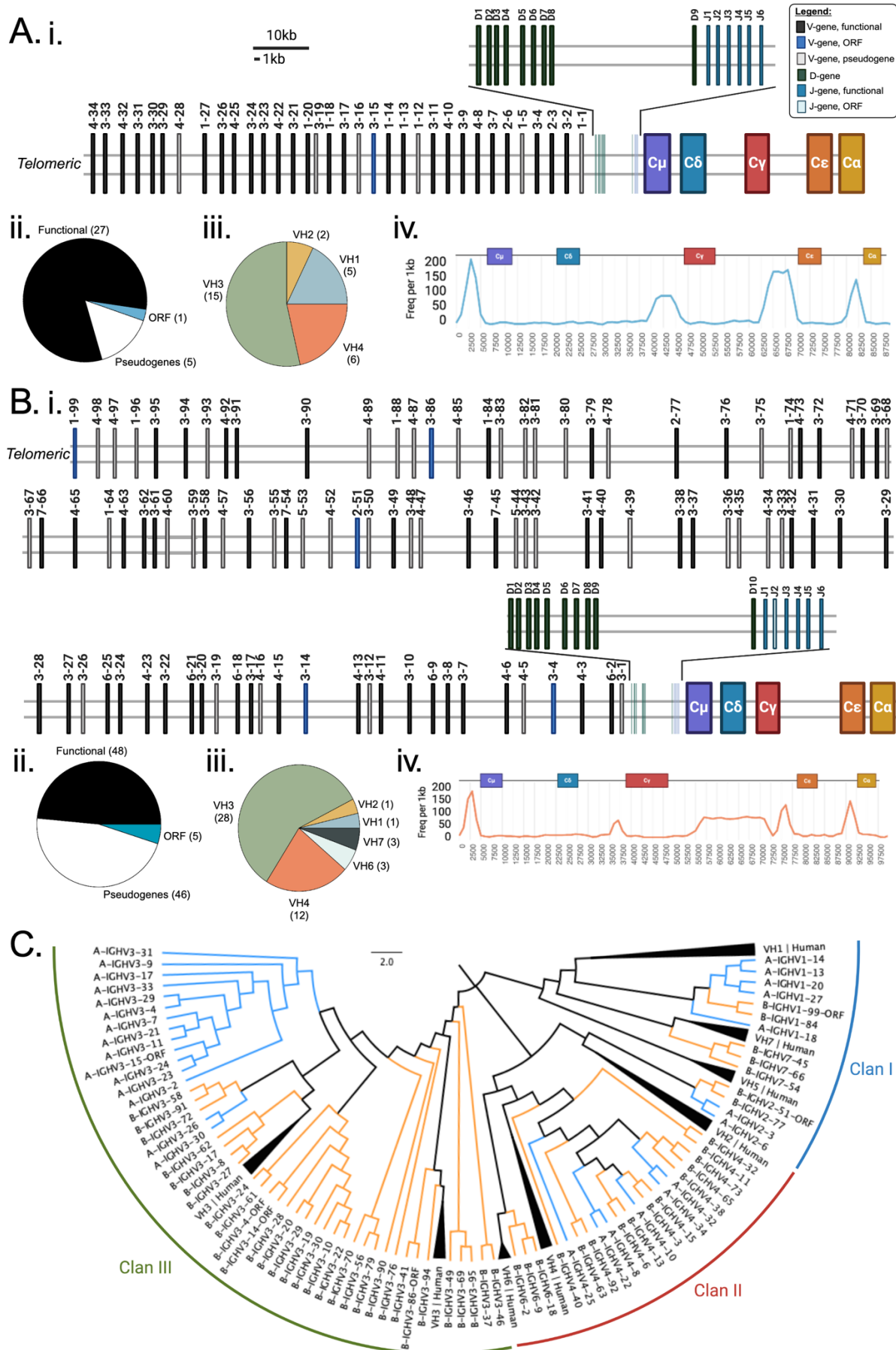
Correspondence and requests for materials should be addressed to H.K. Frank and Scott Boyd

Data availability

IG gene annotations can be found in Supplementary Tables 1-9. All raw sequencing reads have been deposited in NCBI's Short Read Archive (BioProject #####). The GenBank IDs for all genome assemblies used for phylogenetic analysis listed in Supplementary Table 1. Genome assembly and non-IG gene annotations for *E. fuscus* are available through GenBank (<https://www.ncbi.nlm.nih.gov/datasets/genome/>) under the GenBank ID provided in **Methods**

Table 1.

Photo credits (Fig. 2): A. "brown long-eared bat (*Plecotus auritus*)" https://www.flickr.com/photos/wolf_359/123404678/ by wolf_359 is used under a CC BY-NC-SA 2.0 license. B. "Big brown Bat (*Eptesicus fuscus*)" <https://www.inaturalist.org/observations/145464659> by misspt is used under a CC BY-NC 4.0 license. C. "North American Hoary Bat (*Lasiurus cinereus* ssp. *cinereus*)" <https://www.inaturalist.org/photos/18602149> by juancruzado is used under a CC BY-SA 4.0 license. We removed the background from all photos; licensors make no endorsements.



511 **Figure 1 | Genomic annotations and analysis for both IGH loci.** Genomic annotations and
 512 analysis for A-IGH locus contained on contig NC_072477.1 (A) and B-IGH contained on contig
 513 NC_072496.1 (B). For each locus, we present (i) a schematic of IG gene organization with an
 514 inset for diversity (D) and joining (J) gene, pie charts of (ii) functionality and (iii) gene family
 515 distribution for functional variable (V) genes, and (iv) graph of AID hotspot densities per
 516 kilobase in the constant gene encoding regions. Nearest neighbor joining tree of functional and
 517 ORF V genes for A-IGH and B-IGH with human V genes for reference (C). Created with
 518 BioRender.com

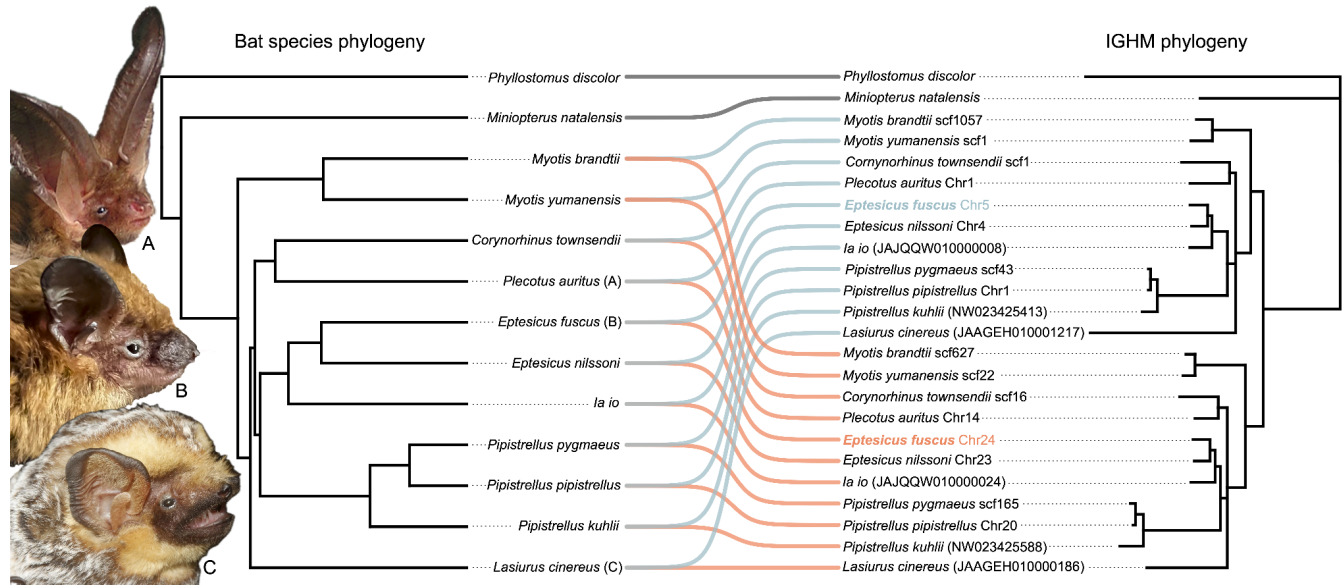


Figure 2 | IGHM duplication phylogeny mirrors species phylogeny. The phylogenetic hypothesis on the left is the inferred relationship between the species⁶⁰. The phylogenetic hypothesis to the right is the inferred relationships between IGHM loci. Identifiers of the IGHM phylogenetic hypothesis refer to the chromosome or scaffold on which each IGHM is found. Lines connect the species to its IGHM sequences with blue lines going to A-IGH-like loci and orange lines going to B-IGH-like loci. Bat profiles are A. *Plecotus auritus*; B. *Eptesicus fuscus*; C. *Lasiurus cinereus*. Picture credits are in the acknowledgments.

528 **Figure 3 | Annotation of single lambda-like light chain locus .** (A) Schematic of IG gene
 529 organization for lambda locus contained on contig NC_072495.1 (B) Schematic of vestigial IG
 530 kappa locus on chromosome 9 (contig NC_072481.1) with the human locus for comparison. Pie
 531 chart of (C) functionality and (D) family distribution for lambda variable genes. (E) Nearest
 532 neighbor joining tree of functional and open reading frame (ORF) V genes with human V genes
 533 for reference and human clans. Created with BioRender.com

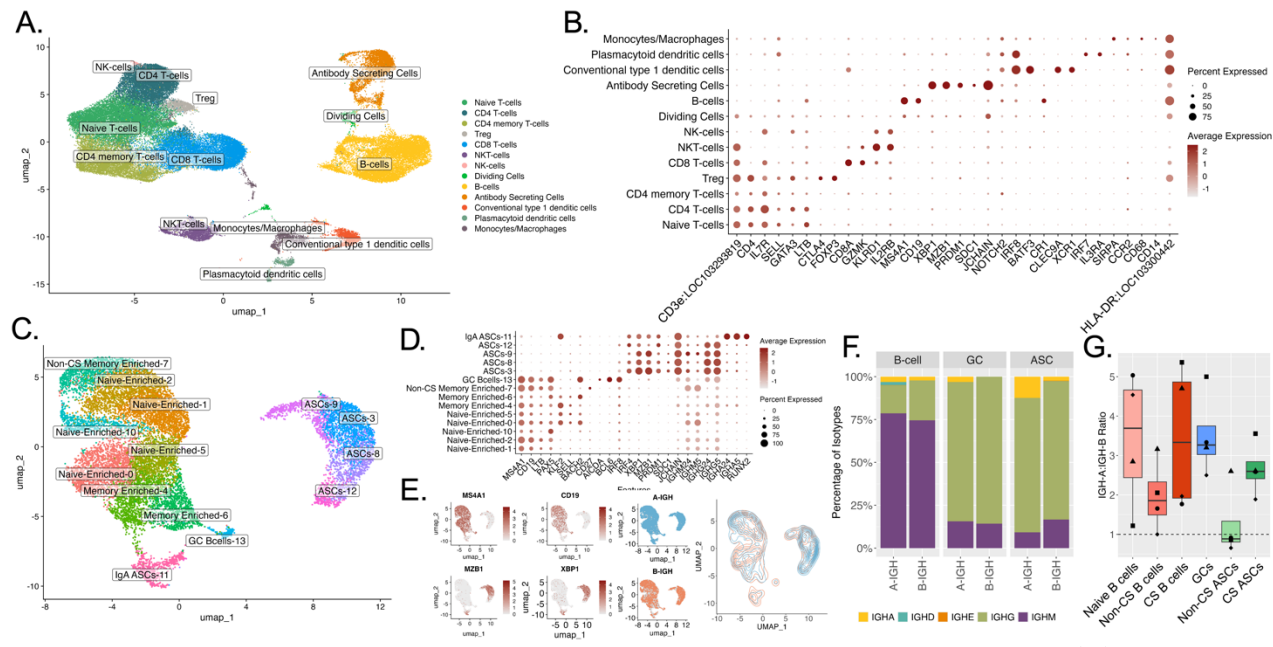


Figure 4 | Splenic single cell transcriptomics reveal distinct B cell phenotypes. (A) UMAP

representation of 62,747 single cell transcriptomes from four *E. fuscus* bats' ($n = 4$) spleens. Cell

clusters were annotated and grouped according to major cell type classification based on

canonical marker genes. (B) Dot plot of the main marker genes used to define each cell type. Dot

size is proportional to the percentage of cells with detectable expression of the indicated gene

and dot color is indicative of the average expression value for the indicated gene, scaled across

all identified clusters. (C) UMAP representation of subclusters of 14,618 cells within the B

lymphocyte compartment after doublet exclusion. (D) Dot plot of the canonical B lymphocyte

lineage and activation marker genes used for phenotype classification. (E) Feature plot of log-

normalized expression of marker genes used to distinguish B cells from antibody secreting cell

phenotype cells. UMAP highlighting 5,010 single cells with full length IGH B cell receptor

(BCR) assemblies from A-IGH (blue) and B-IGH (orange). Density plot of cell with A-IGH

(blue lines) and B-IGH (orange lines). (F) Stacked bar plot of the percent of each isotype of

BCRs from A-IGH and B-IGH for B cells, germinal center (GC), and antibody secreting cells

(ASCs). (G) Box plot of ratio of A-IGH:B-IGH for each cell phenotype. Line represents the

550 mean and ratio for each individual ($n = 4$) plotted with different shapes. Dashed line at ratio of 1.

551 Created with BioRender.com

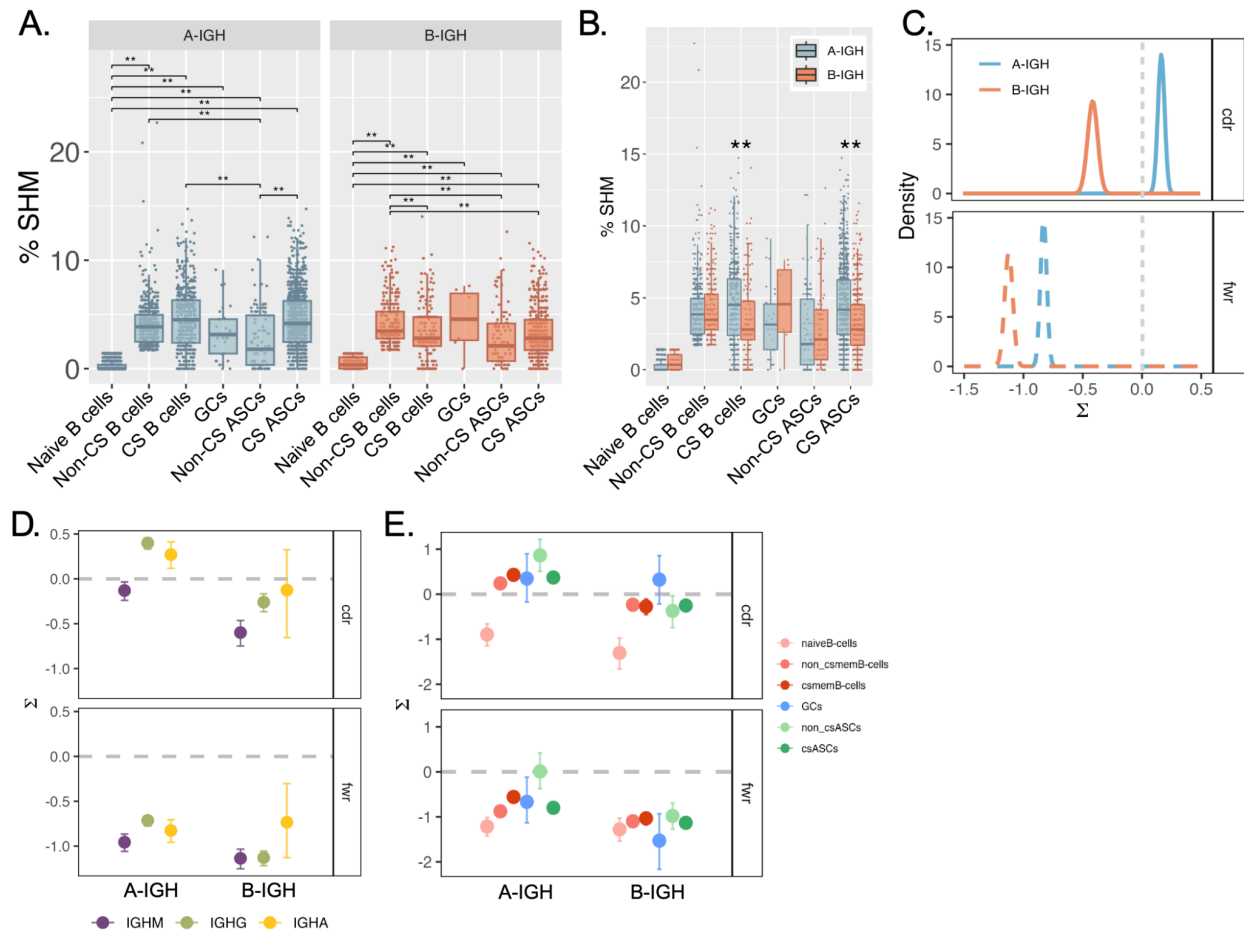
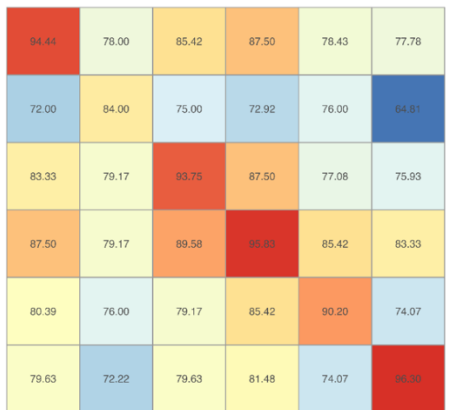


Figure 5 | Levels of SHM vary between and within IGH loci. (A) Distribution of percent nucleotide SHM of IGHV segment for each clone grouped by phenotype and locus. (B) Distribution of percent nucleotide SHM of IGHV segment for A-IGH (blue) and B-IGH (orange) clones grouped by cell phenotype. (C) Density plot of selection scores (sigma, Σ) of the CDRs (solid line, top panel) and FWR (dashed line, bottom panel) for IGHV segments from A-IGH (blue) and B-IGH (orange) clones. Threshold for positive selection gray dashed line ($\Sigma = 0$). (D) Plot of mean with confidence interval for the selection scores of CDR (top panel) and FWR (bottom panel) grouped by locus for IgM (purple), IgG (green), and IgA (yellow). (E) Plot of mean with confidence interval for selection scores of CDR (top panel) and FWR (bottom panel) grouped by locus for each cell phenotype. Comparisons between groups were performed with the

563 Wilcoxon rank-sum test. $**p < 0.05$. Graphs generated in R and figures created with
564 BioRender.com

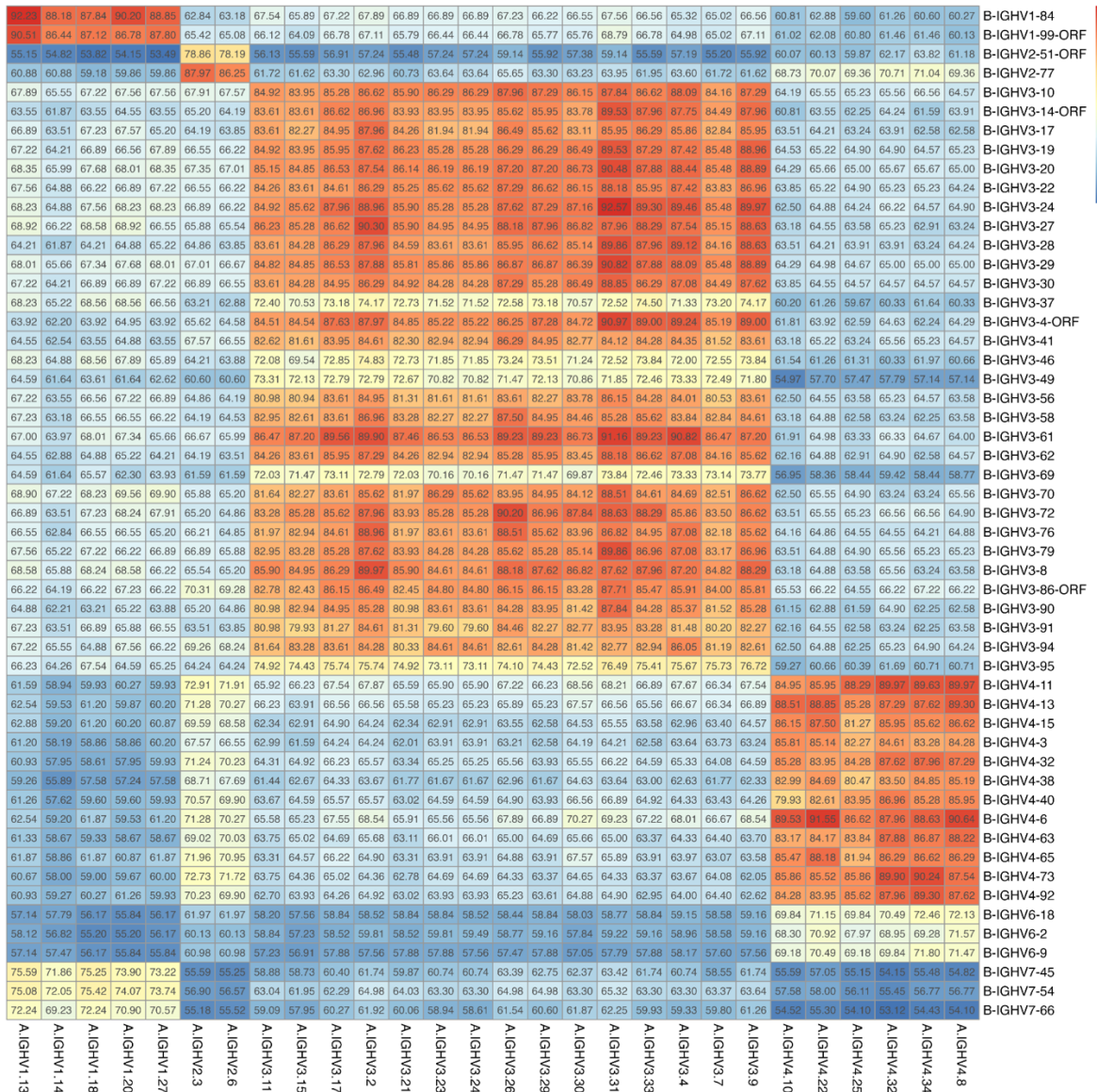
A.



B.

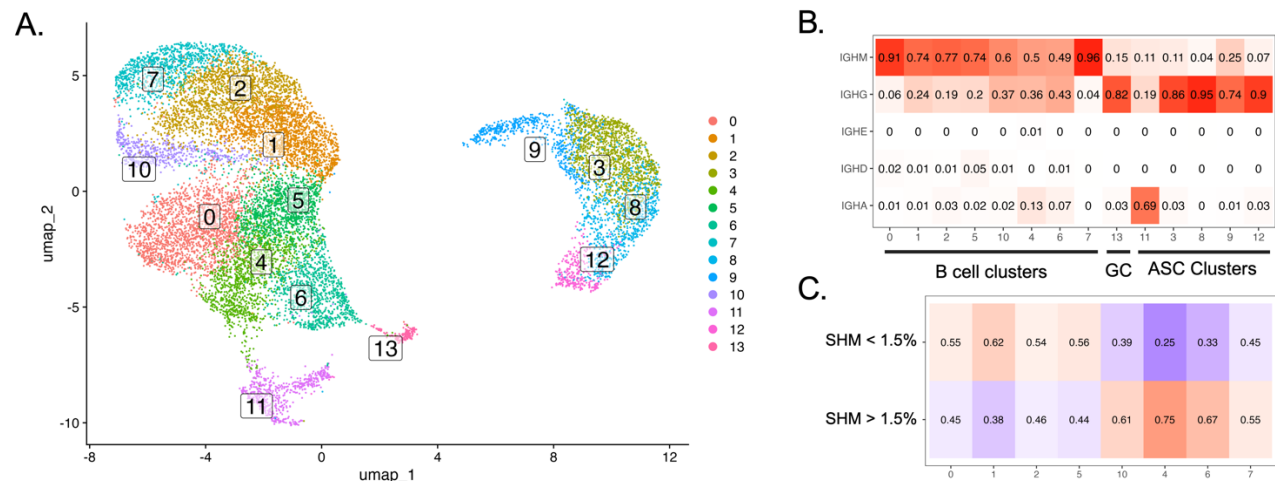


C.

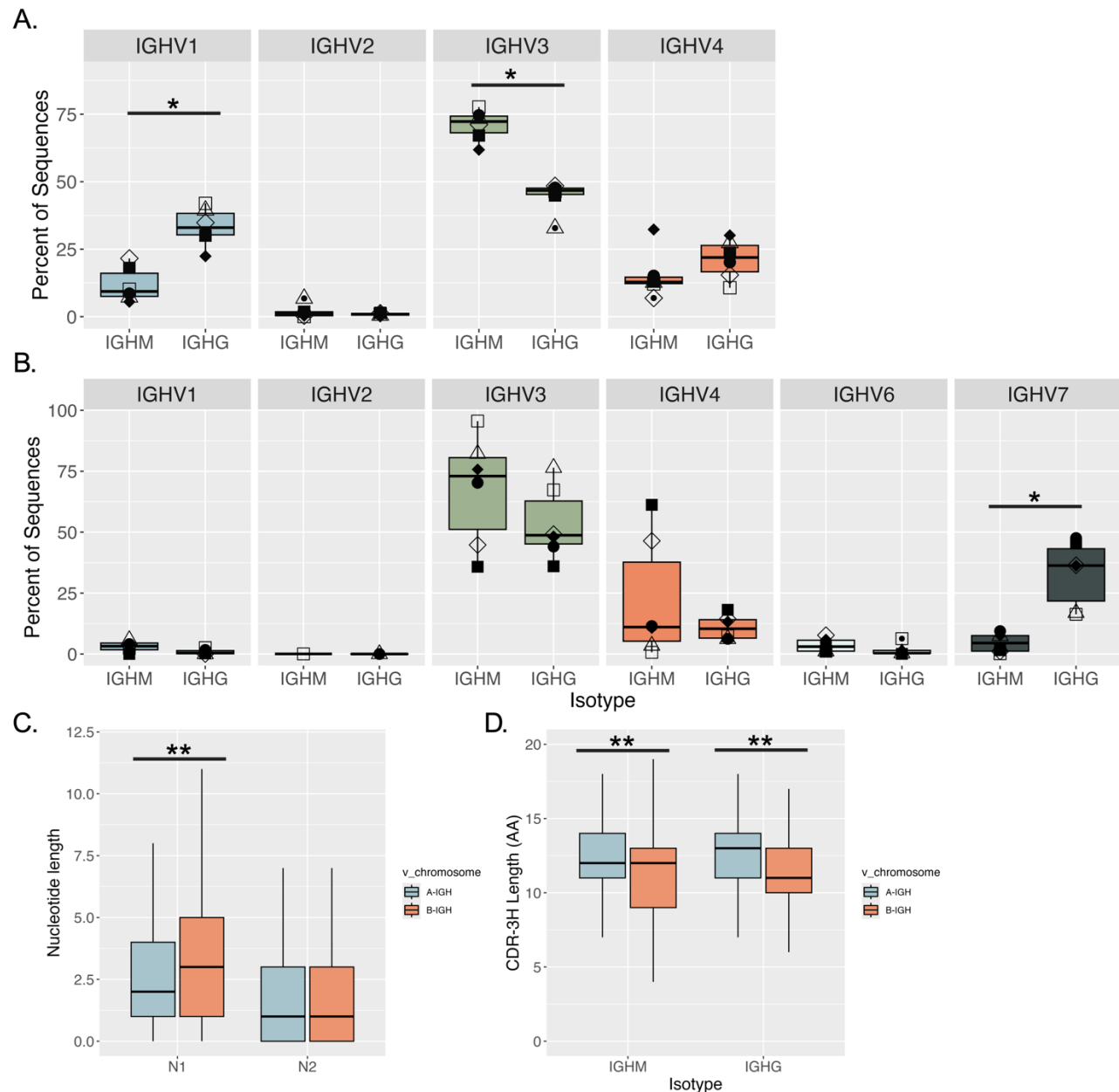


566 **Extended Data Figure 1 | Levels of inter-locus homology varies across germline genes.**

567 Heatmaps of the distance matrix of (A) joining genes, (B) diversity, and (C) variable genes from
568 A-IGH compared to B-IGH where the color and number indicate the percent nucleotide identity
569 for the given pair of V genes. MUSCLE alignment with iterations (n = 10) clustered with
570 neighbor joining and CLUSTALW sequence weighting scheme. Created with BioRender.com



Extended Data Figure 2 | Enrichment of low mutation IGM/D expressing cells in distinct clusters. (A) UMAP of B lymphocyte lineage subclusters. (B) Heatmap of proportion of cells expressing each isotype within each subcluster; square color intensity and label based on the proportion. (C) Heatmap of proportion of IGHM expressing cells with low SHM (<1.5%) compared to high SHM (>1.5%) within the B cell subclusters; square color intensity and label based on the proportion. Created with BioRender.com



Extended Data Figure 3 | Analysis of Bulk IG repertoires. VH-family percentages for (A) A-

IGH and (B) B-IGH within IGHM and IGHG from productive rearrangements found in bulk

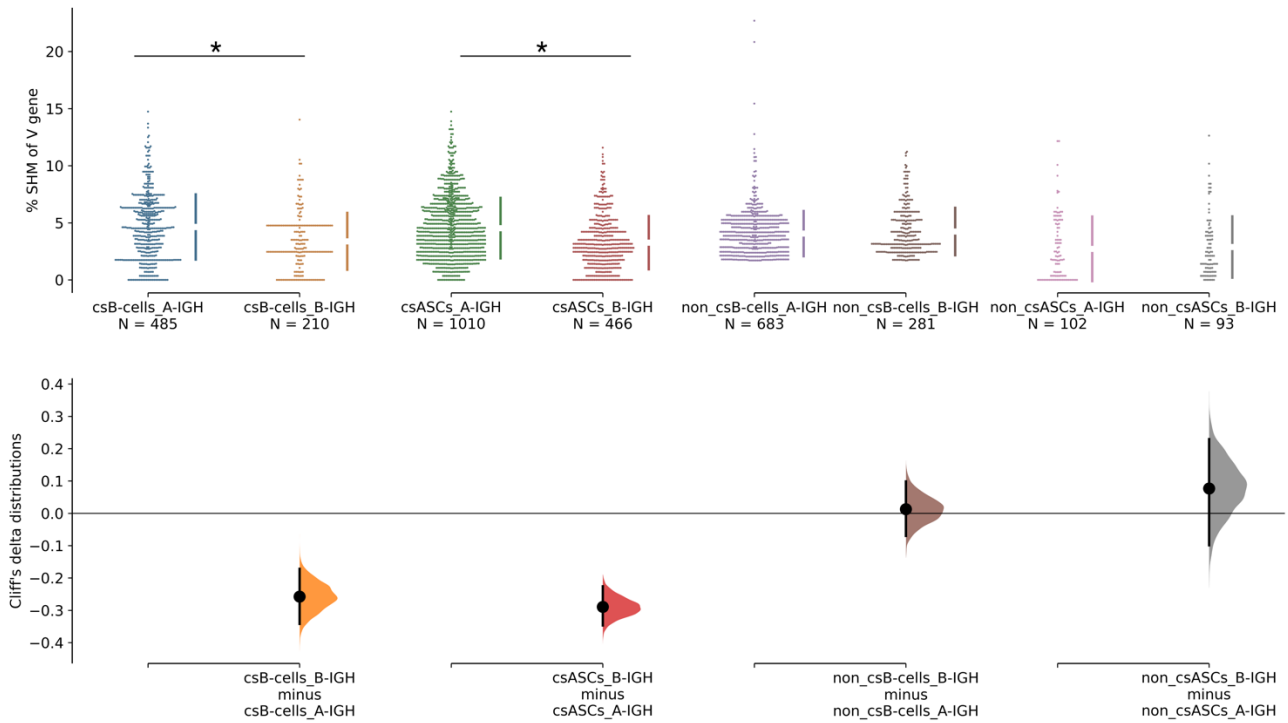
repertoire data where each point represents an individual bat ($n = 6$). (C) Box-and-whisker plot

of length of V-D (N1 nucleotides) and D-J junctions (N2 nucleotides) for IGHM for A-IGH

(blue) and B-IGH (orange). The horizontal line represents the median and whiskers span the

range of the data excluding the outliers (C) Box-and-whisker plot of CDR-3H amino acid lengths

585 including cysteine and tryptophan for IGHM and IGHG for A-IGH (blue) and B-IGH (orange).
586 The horizontal line represents the median and whiskers span the range of the data excluding the
587 outliers. * $p < 0.005$, * * $p < 0.000001$, defined by clone wise unpaired Wilcoxon rank sum test.
588 Graphs generated in R. Post-processing and figures created with BioRender.com



Extended Data Figure 4 | Permutation analysis of SHM. Top plot contains swarm plot of

percent SHM of V gene sequences grouped by cell type comparing A-IGH (left) and B-IGH

(right). Bottom plot is a main effect size plot non-paired Cliff's delta with its 95% confidence

interval for permutations of the non-parametric comparisons between A-IGH and B-IGH for

each cell type. * $p < 0.001$

Main References

1. Wynne, J. W. & Wang, L.-F. Bats and viruses: friend or foe? *PLoS Pathog.* **9**, e1003651 (2013).
2. Schountz, T., Baker, M. L., Butler, J. & Munster, V. Immunological Control of Viral Infections in Bats and the Emergence of Viruses Highly Pathogenic to Humans. *Front. Immunol.* **8**, 1098 (2017).
3. Banerjee, A. *et al.* Novel Insights Into Immune Systems of Bats. *Front. Immunol.* **11**, 26 (2020).
4. Fu, F. *et al.* Bat STING drives IFN-beta production in anti-RNA virus innate immune response. *Front. Microbiol.* **14**, 1232314 (2023).
5. Ahn, M. *et al.* Dampened NLRP3-mediated inflammation in bats and implications for a special viral reservoir host. *Nat. Microbiol.* **4**, 789–799 (2019).
6. Hayman, D. T. S. Bat tolerance to viral infections. *Nature microbiology* vol. 4 728–729 (2019).
7. Barr, J. *et al.* Detection of filovirus-reactive antibodies in Australian bat species. *J. Gen. Virol.* **103**, (2022).
8. Seetahal, J. F. R. *et al.* The Serological Prevalence of Rabies Virus-Neutralizing Antibodies in the Bat Population on the Caribbean Island of Trinidad. *Viruses* **12**, (2020).
9. Pourrut, X. *et al.* Large serological survey showing cocirculation of Ebola and Marburg viruses in Gabonese bat populations, and a high seroprevalence of both viruses in *Rousettus aegyptiacus*. *BMC Infect. Dis.* **9**, 159 (2009).
10. Kading, R. C. *et al.* Exposure of Egyptian Rousette Bats (*Rousettus aegyptiacus*) and a Little Free-Tailed Bat (*Chaerephon pumilus*) to Alphaviruses in Uganda. *Diseases* **10**,

(2022).

11. Tsuda, S. *et al.* Genomic and serological detection of bat coronavirus from bats in the Philippines. *Arch. Virol.* **157**, 2349–2355 (2012).
12. Baker, M. L., Tachedjian, M. & Wang, L.-F. Immunoglobulin heavy chain diversity in Pteropid bats: evidence for a diverse and highly specific antigen binding repertoire. *Immunogenetics* **62**, 173–184 (2010).
13. Papenfuss, A. T. *et al.* The immune gene repertoire of an important viral reservoir, the Australian black flying fox. *BMC Genomics* **13**, 261 (2012).
14. Larson, P. A. *et al.* Genomic features of humoral immunity support tolerance model in Egyptian rousette bats. *Cell Rep.* **35**, 109140 (2021).
15. Ma, L. *et al.* Landscape of IGH germline genes of Chiroptera and the pattern of *Rhinolophus affinis* bat IGH CDR3 repertoire. *Microbiol. Spectr.* **12**, e0376223 (2024).
16. Krebs, J. W., Wheeling, J. T. & Childs, J. E. Rabies surveillance in the United States during 2002. *J. Am. Vet. Med. Assoc.* **223**, 1736–1748 (2003).
17. Turmelle, A. S., Jackson, F. R., Green, D., McCracken, G. F. & Rupprecht, C. E. Host immunity to repeated rabies virus infection in big brown bats. *J. Gen. Virol.* **91**, 2360–2366 (2010).
18. Davis, A. D., Gordy, P. A. & Bowen, R. A. Unique characteristics of bat rabies viruses in big brown bats (*Eptesicus fuscus*). *Arch. Virol.* **158**, 809–820 (2013).
19. 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.* **35**, 273–284 (2011).
20. Rogers, K. A., Richardson, J. P., Scinicariello, F. & Attanasio, R. Molecular

characterization of immunoglobulin D in mammals: immunoglobulin heavy constant delta genes in dogs, chimpanzees and four old world monkey species. *Immunology* **118**, 88–100 (2006).

21. Hackney, J. A. *et al.* DNA targets of AID evolutionary link between antibody somatic hypermutation and class switch recombination. *Adv. Immunol.* **101**, 163–189 (2009).
22. Han, L., Masani, S. & Yu, K. Overlapping activation-induced cytidine deaminase hotspot motifs in Ig class-switch recombination. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11584–11589 (2011).
23. Chen, K. & Cerutti, A. New insights into the enigma of immunoglobulin D. *Immunol. Rev.* **237**, 160–179 (2010).
24. Arpin, C. *et al.* The normal counterpart of IgD myeloma cells in germinal center displays extensively mutated IgVH gene, Cmu-Cdelta switch, and lambda light chain expression. *J. Exp. Med.* **187**, 1169–1178 (1998).
25. Kluin, P. M. *et al.* IgD class switching: identification of a novel recombination site in neoplastic and normal B cells. *Eur. J. Immunol.* **25**, 3504–3508 (1995).
26. Kiyoshi, M. *et al.* Structural basis for binding of human IgG1 to its high-affinity human receptor FcγRI. *Nat. Commun.* **6**, 6866 (2015).
27. Lu, J. *et al.* Structure of FcγRI in complex with Fc reveals the importance of glycan recognition for high-affinity IgG binding. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 833–838 (2015).
28. Rego, K., Bengtén, E., Wilson, M., Hansen, J. D. & Bromage, E. S. Characterization of immunoglobulin light chain utilization and variable family diversity in rainbow trout. *Dev. Comp. Immunol.* **104**, 103566 (2020).

29. Solem, S. T. & Jørgensen T, T. Characterisation of immunoglobulin light chain cDNAs of the Atlantic salmon, *Salmo salar* L.; evidence for three IgL isotypes. *Dev. Comp. Immunol.* **26**, 635–647 (2002).
30. Ratcliffe, M. J. Antibodies, immunoglobulin genes and the bursa of Fabricius in chicken B cell development. *Dev. Comp. Immunol.* **30**, 101–118 (2006).
31. Bao, Y. *et al.* The immunoglobulin light chain locus of the turkey, *Meleagris gallopavo*. *Vet. Immunol. Immunopathol.* **147**, 44–50 (2012).
32. Huang, T. *et al.* Analysis of immunoglobulin transcripts in the ostrich *Struthio camelus*, a primitive avian species. *PLoS One* **7**, e34346 (2012).
33. Thibaud-Nissen, F. *et al.* P8008 The NCBI Eukaryotic Genome Annotation Pipeline. *J. Anim. Sci.* **94**, 184–184 (2016).
34. Sirupurapu, V., Safonova, Y. & Pevzner, P. A. Gene prediction in the immunoglobulin loci. *Genome Res.* **32**, 1152–1169 (2022).
35. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
36. Lee, W. T. *et al.* Identification of secreted and membrane-bound bat immunoglobulin using a Microchiropteran-specific mouse monoclonal antibody. *Dev. Comp. Immunol.* **65**, 114–123 (2016).
37. Sun, Y. *et al.* A comprehensive analysis of germline and expressed immunoglobulin repertoire in the horse. *Dev. Comp. Immunol.* **34**, 1009–1020 (2010).
38. King, H. W. *et al.* Single-cell analysis of human B cell maturation predicts how antibody class switching shapes selection dynamics. *Sci Immunol* **6**, (2021).
39. Gros, P. & Casanova, J.-L. Reconciling Mouse and Human Immunology at the Altar of

Genetics. *Annu. Rev. Immunol.* **41**, 39–71 (2023).

40. Friedrichs, V. *et al.* Landscape and age dynamics of immune cells in the Egyptian rousette bat. *Cell Rep.* **40**, 111305 (2022).

41. Kessler, S. *et al.* Deciphering bat influenza H18N11 infection dynamics in male Jamaican fruit bats on a single-cell level. *Nat. Commun.* **15**, 4500 (2024).

42. Gamage, A. M. *et al.* Single-cell transcriptome analysis of the in vivo response to viral infection in the cave nectar bat *Eonycteris spelaea*. *Immunity* **55**, 2187–2205.e5 (2022).

43. Milpied, P. *et al.* Human germinal center transcriptional programs are de-synchronized in B cell lymphoma. *Nat. Immunol.* **19**, 1013–1024 (2018).

44. Jash, A. *et al.* ZBTB32 Restricts the Duration of Memory B Cell Recall Responses. *J. Immunol.* **197**, 1159–1168 (2016).

45. Zhang, Y. & Derynck, R. Transcriptional regulation of the transforming growth factor-beta-inducible mouse germ line Ig alpha constant region gene by functional cooperation of Smad, CREB, and AML family members. *J. Biol. Chem.* **275**, 16979–16985 (2000).

46. Watanabe, K. *et al.* Requirement for Runx proteins in IgA class switching acting downstream of TGF-beta 1 and retinoic acid signaling. *J. Immunol.* **184**, 2785–2792 (2010).

47. Stavnezer, J. & Kang, J. The surprising discovery that TGF beta specifically induces the IgA class switch. *J. Immunol.* **182**, 5–7 (2009).

48. Sugai, M., Watanabe, K., Nambu, Y., Hayashi, T. & Shimizu, A. Functions of Runx in IgA class switch recombination. *J. Cell. Biochem.* **112**, 409–414 (2011).

49. Jackson, K. J. L., Gaëta, B. A. & Collins, A. M. Identifying highly mutated IGHD genes in the junctions of rearranged human immunoglobulin heavy chain genes. *J. Immunol. Methods* **324**, 26–37 (2007).

50. Chan, T. D. *et al.* Antigen affinity controls rapid T-dependent antibody production by driving the expansion rather than the differentiation or extrafollicular migration of early plasmablasts. *J. Immunol.* **183**, 3139–3149 (2009).
51. Paus, D. *et al.* Antigen recognition strength regulates the choice between extrafollicular plasma cell and germinal center B cell differentiation. *J. Exp. Med.* **203**, 1081–1091 (2006).
52. Taylor, J. J., Pape, K. A., Steach, H. R. & Jenkins, M. K. Humoral immunity. Apoptosis and antigen affinity limit effector cell differentiation of a single naïve B cell. *Science* **347**, 784–787 (2015).
53. Zhou, P. *et al.* Contraction of the type I IFN locus and unusual constitutive expression of IFN- α in bats. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 2696–2701 (2016).
54. Pavlovich, S. S. *et al.* The Egyptian Rousette Genome Reveals Unexpected Features of Bat Antiviral Immunity. *Cell* **173**, 1098–1110.e18 (2018).
55. Hughes, A. L. & Hughes, M. K. Small genomes for better flyers. *Nature* **377**, 391 (1995).
56. Smith, J. D., Bickham, J. W. & Gregory, T. R. Patterns of genome size diversity in bats (order Chiroptera). *Genome* **56**, 457–472 (2013).
57. Moreno Santillán, D. D. *et al.* Large-scale genome sampling reveals unique immunity and metabolic adaptations in bats. *Mol. Ecol.* **30**, 6449–6467 (2021).
58. Jebb, D. *et al.* Six reference-quality genomes reveal evolution of bat adaptations. *Nature* **583**, 578–584 (2020).
59. Jacquet, S. *et al.* Adaptive duplication and genetic diversification of protein kinase R contribute to the specificity of bat-virus interactions. *Sci Adv* **8**, eadd7540 (2022).
60. Shi, J. J. & Rabosky, D. L. Speciation dynamics during the global radiation of extant bats. *Evolution* **69**, 1528–1545 (2015).