Developmental and Comparative Immunology 72 (2017) 57–68

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

journal homepage: www.elsevier.com/locate/dci

The Florida manatee (*Trichechus manatus latirostris*) immunoglobulin heavy chain suggests the importance of clan III variable segments in repertoire diversity

Breanna Breaux [a], Thaddeus C. Deiss [a], Patricia L. Chen [a], Maria Paula Cruz-Schneider [b], Leonardo Sena [c], Margaret E. Hunter [d], Robert K. Bonde [d], Michael F. Criscitiello [a, e, *]

[a] *Comparative Immunogenetics Laboratory, Department of Veterinary Pathobiology, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX 77843, USA*
[b] *Laboratory of DNA Polymorphism, Federal University of Pará, Belém, Pará, Brazil*
[c] *Laboratory of Medical and Human Genetics, Federal University of Pará, Belém, Pará, Brazil*
[d] *Wetland and Aquatic Research Center, U.S. Geological Survey, 7920 NW 71st Street, Gainesville, FL 32653, USA*
[e] *Department of Microbial Pathogenesis and Immunology, College of Medicine, Texas A&M Health Science Center, Texas A&M University, College Station, TX 77843, USA*

ARTICLE INFO

Article history:
Received 21 October 2016
Received in revised form
23 January 2017
Accepted 23 January 2017
Available online 25 January 2017

Keywords:
Manatee
Immunoglobulin
Clan
Genomic
Mutation

ABSTRACT

Manatees are a vulnerable, charismatic sentinel species from the evolutionarily divergent Afrotheria. Manatee health and resistance to infectious disease is of great concern to conservation groups, but little is known about their immune system. To develop manatee-specific tools for monitoring health, we first must have a general knowledge of how the immunoglobulin heavy (IgH) chain locus is organized and transcriptionally expressed. Using the genomic scaffolds of the Florida manatee (*Trichechus manatus latirostris*), we characterized the potential IgH segmental diversity and constant region isotypic diversity and performed the first Afrotherian repertoire analysis. The Florida manatee has low V(D)J combinatorial diversity (3744 potential combinations) and few constant region isotypes. They also lack clan III V segments, which may have caused reduced VH segment numbers. However, we found productive somatic hypermutation concentrated in the complementarity determining regions. In conclusion, manatees have limited IGHV clan and combinatorial diversity. This suggests that clan III V segments are essential for maintaining IgH locus diversity.

1. Introduction

Manatees are largely tropical marine mammals within Superorder Afrotheria, which includes aardvarks, elephants, and tenrecs (Kellogg et al., 2007). The Order Sirenia includes three manatee species, the West Indian (*Trichechus manatus*), the Amazonian (*T. inunguis*), and the African (*T. senegalensis*) manatee that inhabit the Caribbean and Atlantic coasts, the Amazon River drainage, and the West African coasts, respectively (Domning, 1982). The two American species are of particular interest due to a history of overhunting, which caused several bottlenecks and their current vulnerable status (O'Shea, 1988). The West Indian manatee also has significant population fragmentation due to habitat destruction (Vianna et al., 2006). To date, two subspecies have been identified: the Florida manatee (*T. m. latirostris*) and the Antillean manatee (*T. m. manatus*). Because these species maintain a wide distribution across varied aquatic environments, have a history of bottlenecks and founder events, and serve a role as sentinel species within coastal environments (Bonde et al., 2004), characterizing their adaptive immune receptors is essential to supplement conservation efforts. Some earlier studies have quantified the robustness of the manatee immune system (Bossart, 1995; McGee, 2012; Sweat et al., 2005; Walsh et al. 2003, 2005), but the specific antigen binding diversity of the adaptive immune receptors has not been described. Manatees make an interesting non-model species to study because

* Corresponding author. Texas A&M University, Mailstop 4467, College Station, TX 77843, USA.
E-mail addresses: bbreaux@cvm.tamu.edu (B. Breaux), tdeiss@cvm.tamu.edu (T.C. Deiss), pchen@cvm.tamu.edu (P.L. Chen), paula@ufpa.br (M.P. Cruz-Schneider), lsena@ufpa.br (L. Sena), mhunter@usgs.gov (M.E. Hunter), rbonde@usgs.gov (R.K. Bonde), mcriscitiello@cvm.tamu.edu (M.F. Criscitiello).

http://dx.doi.org/10.1016/j.dci.2017.01.022
0145-305X/

there are little data on the lymphocyte antigen receptors of the Afrotherian branch of mammalian evolution (Guo et al., 2011).

Florida manatees have two significant disease associated agents identified in their populations (*Brevetoxicosis* (red tide) and papilloma virus) (Bossart et al., 1998, 2002; Walsh et al., 2015). The brevetoxins produced by *Karenia brevis* cause disease in two ways. There are acute neurotoxic effects from the brevetoxin when they bind to sodium channels, which activates them and disrupts the normal salt gradient required for neuronal action potentials (Wang, D. Z. 2008). The more fatal aspects of the disease stem from the chronic immunosuppressive effects when the brevetoxins are internalized by leukocytes because it initiates apoptosis, which releases inflammatory mediators that lead to edema and pulmonary hemorrhage (Bossart et al., 1998). Additionally, several parasites have been identified in the Florida populations, including trematodes (*Chiorchis fabaceus, Chiochis groschafti, Pulmonicola cochleotrema, Moniligerum blairi,* and *Nudacotyle undicola*), nematodes (*Heterocheilus tunicatus*) and coccidians (*Eimeria manatus* and *Eimeria nodulosa*), but their impact on population health is not yet understood (Bando et al., 2014). Little is known about the manatee immunoglobulin (Ig) repertoire, with the exception of serological analysis that determined baseline levels of circulating IgG in *T. manatus* (McGee, 2012). Since the *T. m. latirostris* genome is the only assembled manatee genome currently available, we characterized the genomic organization of this manatee species and evaluated the gene expression to assess the potential repertoire diversity.

Immunoglobulins are both the secreted and membrane bound receptors of B cells that can opsonize and neutralize extracellular antigens and activate the complement cascade. The adaptive immune receptor loci are exceptional in their organization and rearrangement mechanism which creates the diversity needed to recognize a wide variety of antigens. The Immunoglobulin heavy chain (IgH) locus contains a unique set of several variable (V), diversity (D), and joining (J) segments (Tonegawa, 1983). During development, each B cell recombines one of each segment at the DNA level so that each cell produces one receptor, allowing for the specific proliferation of cells and receptors that recognize the invading antigen (Nossal and Lederberg, 1958).

To increase the diversity of these receptors beyond the combination of different segments, there are two mechanisms driven by the enzyme activation-induced cytosine deaminase (AID) in most vertebrates: somatic hypermutation (SHM) and class-switch recombination (CSR). These mechanisms usually proceed after antigen recognition, except in sheep where mutation occurs independent of external antigen exposure (Reynaud et al., 1995). AID also mediates another uncommon mechanism of diversification called gene conversion, seen in rabbits and chickens. In SHM, deaminating a cytosine results in repair mechanisms that can cause a point mutation. CSR occurs when there are two double stranded breaks in the switch regions of two constant regions, which are resolved by bringing the two breaks together and deleting the intervening sequence. This allows for isotype switching which gives the receptors different functions.

Mammalian V segments can be separated into three clans based on sequence similarity primarily in the framework 1 and 3 regions (Kirkham et al., 1992). In most mammals, there are germline V segment representatives for each clan, but some species lack V segments from one or two clans. However, clan III is usually retained, such as the platypus that only has clan III V segments (Zhao et al., 2009). Clan III is also the most homologous to ancestral V segments in fish (Andersson, 1995; Zhao et al., 2009). However, there are three mammalian species that do not have clan III (cow, sheep, and horse), where it appears to be coupled with low numbers of V segments (Dufour et al., 1996; Niku et al., 2012; Walther et al., 2015). The mechanism and selection pressure behind the maintenance of these clans is unclear, but there do seem to be functional differences between the three clans (Schroeder and Wang, 1990; Tutter and Riblet, 1989).

While the immunoglobulin heavy chain locus structure is similar across jawed vertebrates, there are differences in the potential repertoire diversity and the range of functions of the available constant regions. By comparing the manatee segmental diversity and expression with other species, we can see which segments are important both within and between species and determine how well equipped manatees are for defense against infectious disease in their environment.

2. Materials and methods

2.1. Sample collection

Blood was collected into EDTA-containing vacutainer tubes from the flipper of *T. manatus latirostris* during wild capture health assessments in Crystal River, Florida in November of 2014. The blood was processed at the site of collection using the LeukoLock Total RNA Isolation System (Life Technologies, Carlsbad CA) to preserve RNA from peripheral blood leukocytes. The filters were transported to Texas A&M University at room temperature and then stored at −20 °C.

2.2. Total RNA isolation and cDNA synthesis

RNA was isolated from the filter-bound leukocytes using the LeukoLock Total RNA Isolation System, following the manufacturer's instructions. The quantity and quality of the RNA samples were assessed using a NanoDrop 2000c spectrophotometer (Thermofisher, Waltham, MA) and stored at −80 °C. The 5′ RACE cDNA libraries were prepared from the leukocyte RNA using the GeneRacer kit (Life Technologies) with GeneRacer oligo dT and random primers in equal ratios.

2.3. Identification of IgH genes in the genome

The Igμ constant region for the manatee was annotated by aligning the Asian elephant (*Elephas maximus*) Igμ transcript against the *T. m. latirostris* genome (Broad v1.0/triMan1). Once the scaffold containing the Igμ constant region was identified, BLASTx was used to search for potential upstream V segments and downstream constant regions. The Recombination Signal Sequence Site (http://www.itb.cnr.it/rss/index.html) was used to predict recombination signal sequences (RSSs) to identify D and J segments. D segments were predicted based on two opposite facing 12-spacer RSSs located between the most 3′ V segment and most 5′ J segment. J segments were predicted based on a reverse orientation 23 RSS that was followed by either a FGXG or a WGXG motif. Genomic V and D segments were separated into families based on 70% nucleotide identity (Brodeur and Riblet, 1984). Both V and D segment families were ordered by the number of segments within the family, starting with the largest.

2.4. IgH RACE PCR, cloning, and sanger sequencing

The 5′ RACE products were amplified by PCR using the 5′ Generacer adaptor forward primer and a *T. manatus* Igμ constant region reverse primer (Table S1). The 3′ RACE products were amplified by PCR using several forward primers specific for the framework regions of different V and the 3′ Generacer reverse primer. The amplicons were purified from a 0.8% agarose gel after electrophoresis in tris/acetic acid/EDTA (TAE) buffer, cloned into pCR II vector

with the TOPO TA Cloning Kit (Life Technologies), and transformed into chemically competent TOP10 *Escherichia coli* cells (Invitrogen, Carlsbad, CA). Colonies were picked based on blue/white screening produced by X-Gal (Sigma-Aldrich, Saint Louis MO). The plasmid DNA was purified using the Zyppy Plasmid Miniprep kit (Zymo Research Corporation, Irvine CA) and digested with *Eco*RI (Promega, Madison, WI) to identify clones with inserts. Products for sequencing were amplified using either M13 forward or reverse primers, purified using ABI BigDye X Terminator Purification Kit (Life Technologies), and sequenced by the Gene Technologies Lab in the Department of Biology at Texas A&M University.

2.5. PacBio SMRT sequencing

PacBio Single Molecule Real Time (SMRT) sequencing provides long read length and circular consensus sequences to provide high accuracy, which makes it ideal to cover the entire V(D)J rearrangement and identify SHM.

Transcripts were amplified from one individual *T. m. latirostris* RACE library. Primers used for PacBio sequencing were tagged with a 16 bp barcode, then 5′ RACE PCR was performed and the amplicons extracted from a 0.8% agarose gel after electrophoresis in tris/acetic acid/EDTA (TAE) buffer using the Purelink Quick Gel Extraction Kit (Invitrogen). Amplicon DNA was quantified using Qubit (Invitrogen). The DNA samples were sequenced by the Duke University Genome Sequencing Center (Durham, NC) on a PacBio RS II third-generation sequencer.

2.6. Sequence analysis

V and J segments in transcripts that were at least 90% identical to the genomic sequence were considered to have originated from that genomic segment. Sequences were aligned to the genomic D segment sequence to determine the transcript sequence that was contributed by a D segment. Individual identification of D segments in sequences is difficult due to high similarity between the D segments and a high propensity for mutation in the Complementarity Determining Region 3 (CDR3) region. D segment usage was determined by aligning the CDR3 region of each transcript to each genomic D segment. Each alignment was scored manually under these conditions: a single match was +1 point, the match extension bonus was +0.1 for each consecutive match, a mismatch was −0.5 points, and the mismatch extension cost was an added −0.5 for each consecutive mismatch. Scored regions did not cross over gaps in the alignment. The best scored alignment determined the D segment sequence used in each transcript.

CDR1 length was calculated as the number of amino acids between (not including) position 26 and 39, given that position 23 is the first conserved cysteine and position 41 is the tryptophan of the WYRQ motif (Lefranc, 1999). CDR2 length was calculated as the number of amino acids between (not including) position 55 and 66, given that position 104 is the second conserved cysteine in the YXC motif (Lefranc, 1999). CDR3 length was calculated as the number of amino acids between (not including) the second conserved cysteine of the V segment and the phenylalanine/tryptophan of the F/WGXG motif of the J segment (Lefranc, 1999).

2.7. Sequence alignment and tree building

Amino acid alignments were performed using the Geneious (version 9.1.5) software ClustalW alignment tool, employing gap opening penalties of 17 and gap extension penalties of 6.0 for pairwise alignments. The alignments were used to create the two neighbor-joining trees using the Tamura-Nei distance model (Tamura and Nei, 1993) with 1000 bootstrap iterations.

Both trees were also constructed using RaxML to create maximum likelihood trees with the CAT BLOSUM62 protein model and bootstrapping algorithm in Geneious version 9.1.5 (http://www.geneious.com) (Kearse et al., 2012). The tree topologies were similar with both methods, so the maximum likelihood trees were not included.

3. Results

3.1. Genomic organization of the T. m. latirostris IgH locus

3.1.1. IgH locus

The IgH locus is located on scaffold 53 of the *T. m. latirostris* genome assembly (Broad v1.0/triMan1). The locus is in the reverse orientation relative to the scaffold and spans 3,765,922 bp from the most 5′ V segment on the scaffold, which is a pseudogene, to the transmembrane tail of Igα, the most 3′ constant region (Fig. 1). There are many sequence gaps within the locus. The total locus length may be underestimated because the scaffold ends before a non-Ig gene is identified. The scaffolds that contained VIPR2 and Zfp386, which flank the 5′ end of the IgH locus in other species, did not contain any IGHV-like sequences and no other V segments were located on other scaffolds. Therefore, we were unable to identify the 5′ boundary of the locus. It is possible that the manatee IgH locus is at the end of the chromosome similar to the human IgH locus and therefore does not have a flanking non-Ig gene, so the beginning of the genomic scaffold could be the beginning of the locus. Additionally, the region of the *T. m. latirostris* locus encoded by this scaffold is larger than the human (1.25 Mb), mouse (2.9 Mb) and elephant (2.97 Mb) loci, so it is reasonable to hypothesize that the beginning of the scaffold could be the beginning of the locus. However, the chromosomal location of the locus has not yet been identified.

3.1.2. Constant region genes

The *T. m. latirostris* genome assembly codes for the five IgH constant region isotypes found in most mammals: Igμ, Igδ, Igγ, Igε, and Igα, yet only contains one sub-class of each isotype (Fig. 2).

The *T. m. latirostris* Igμ gene is the most 5′ constant region and is made up of four Ig domain exons and two transmembrane exons. The secretory protein is 445 amino acids long. The Igμ gene contains all conserved cysteines for inter- and intra-domain disulfide bonds. It also contains seven potential *N*-linked glycosylation sites, with the one in the C4 domain being unique to manatees and elephants (Fig. S1).

There is one 273 bp Igδ constant domain fragment encoded in the *T. m. latirostris* genome 5690 bp after the cytoplasmic tail exon of Igμ. The elephant also has a similar Igδ fragment (Guo et al., 2011). When aligned to Igδ sequence from other species, this exon is most similar to the C3 domain (Fig. S2). In the *T. m. latirostris* genome, there are no sequence gaps between the Igμ cytoplasmic tail and this domain, the end of the exon is truncated, there is no transmembrane region 3′ of the exon, and this exon has not been found in any sequence using 3′ RACE PCR. Therefore, the *T. manatus latirostris* δ exon is most likely a pseudogene remnant.

The *T. m. latirostris* genome encodes one Igγ constant region consisting of three Ig domain exons with a 14 amino acid hinge region between the C1 and C2 domains; the secretory protein is 330 amino acids long. There are two transmembrane region exons. Between the Igδ exon and the Igγ C1 domain, there are four sequence gaps that are 1–2 kb in length each and one that is 7.7 kb long. There are no gaps between Igγ and Igε. It is possible that other subclasses may be found in these sequence gaps, however we have only identified transcriptional evidence for the one described Igγ subclass. The amino acids involved in human FcRn binding are

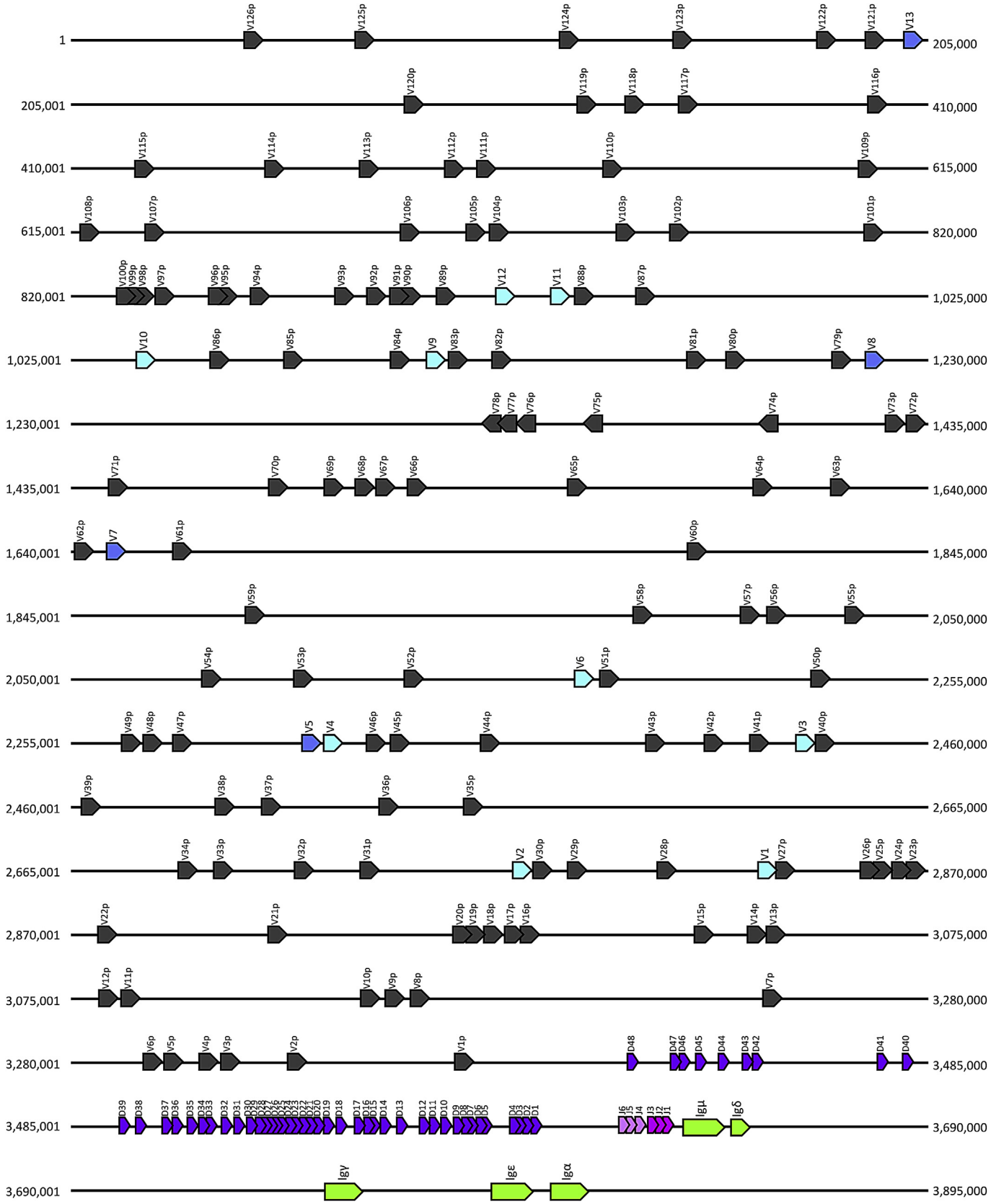


Fig. 1. Genomic organization of IgH locus: The *T. m. latirostris* IgH locus is localized to scaffold 53 and has a length of 3,765,922 bp from the most 5′ V segment to the most 3′ exon of Igα. The numbers along each line indicate nucleotide position in the scaffold. The color coding is as follows: gray blocks are pseudogene V segments, light blue blocks are potentially functional V segments, medium blue blocks are expressed V segments, purple blocks are D segments, light pink blocks are potentially functional J segments, medium pink blocks are expressed J segments, and green blocks are constant regions. Segments are numbered in chronological order beginning with the most 3′ segment. Transcriptional orientation is indicated by the direction of the point of the block. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

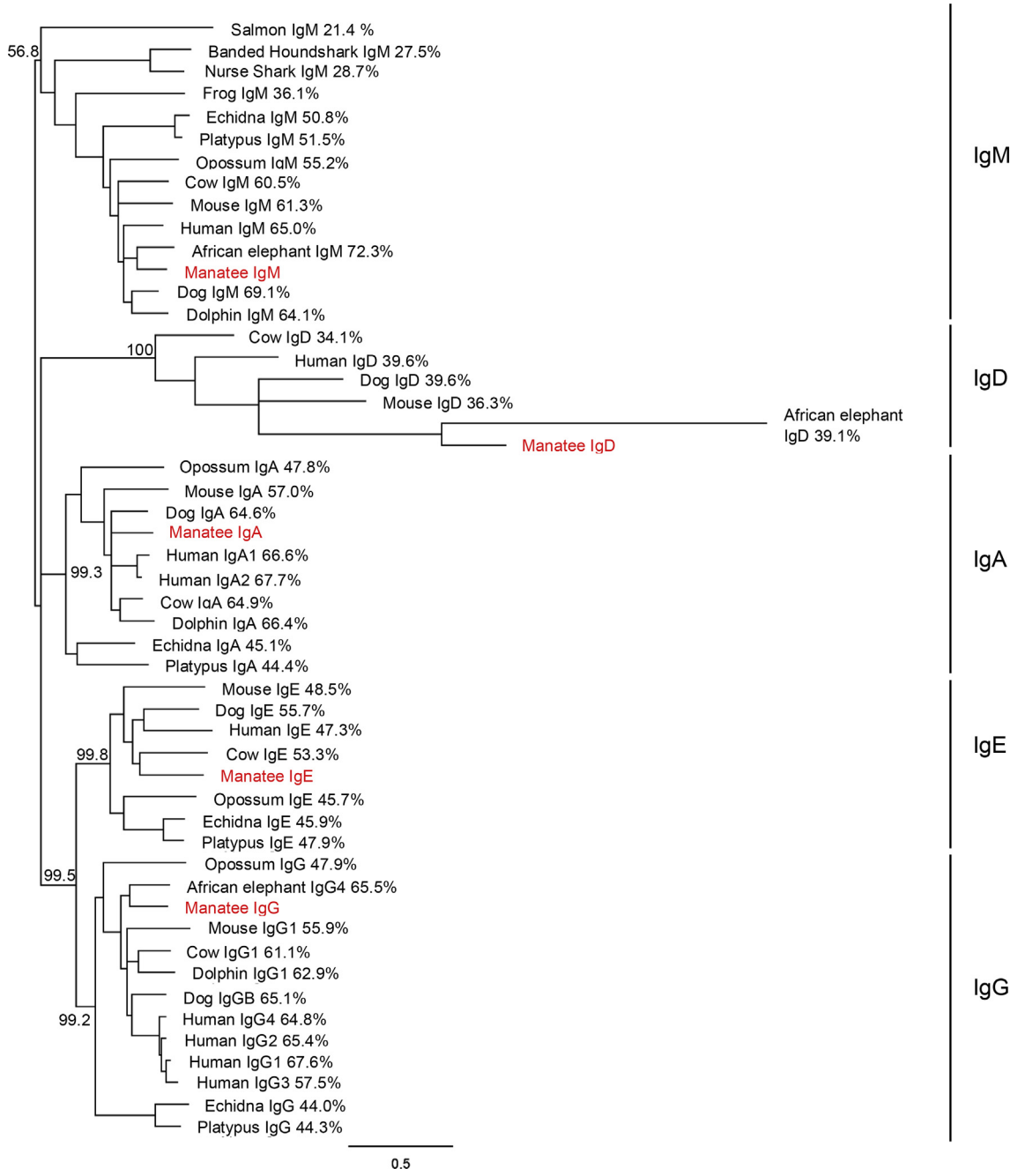


Fig. 2. ***T. m. latirostris*** **has one constant region gene for each isotype**. Neighbor-joining tree generated from an amino acid alignment of selected sequences from GenBank. The *T. m. latirostris* sequences are highlighted in red. Bootstrap values at each major branch point are based on 1000 iterations, and sequence distance is indicated by the branch length according to the scale below the tree. The amino acid percent identity to the respective *T. m. latirostris* isotype is indicated at the end of each label. The tree was constructed using all the domains of each constant region with the exception of the elephant and manatee IgD. Both species only have one nonfunctional exon remnant to include while the other species have the complete 2–3 exons. Accession numbers are provided in Table S4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

conserved in the *T. m. latirostris* Igγ, but there are differences at the human Igγ1 FcγRI and C1q binding motifs (Fig. S3). Compared to the four human subclasses, the *T. m. latirostris* Igγ is structurally similar to Igγ2, with four cysteines in the hinge region and the inter-chain disulfide bond cysteine near the beginning of C1. It is also similar to human Igγ2 in disruption of the C1q, FcγRI, and FcγRIII binding motifs; Igγ1 and Igγ3 have a leucine at position 235 in the C2 domain, while Igγ2 has an alanine that lowers its C1q binding activity and eliminates FcγRI and FcγRIII binding ability. The *T. m. latirostris* Igγ has a proline at this position, therefore it is likely it has lower complement activation ability and no FcγRI and FcγRIII binding activity (Fig. S3). However, it has the most nucleotide similarity to human Igγ1 and does not have the deletion that Igγ2 has at position 237 (Fig. S3). The amino acids that are involved in FcRn binding are highly conserved in eutherian mammals, but are less conserved in monotremes and marsupials.

Igε has four Ig domain exons and 2 transmembrane exons. The secretory protein is comprised of 426 amino acids. The *T. m. latirostris* Igε is very similar to other species when considering conserved cysteines, but the glycosylation sites throughout the

protein are variable in number and location between the species examined. In *T. m. latirostris*, there are six glycosylation sites in total: two in the C1 domain, two in C2, two in C3, and none in C4. The domain distribution is relatively similar to other species compared except for the C4 domain; all others except the primate species had at least one glycosylation site. The high-affinity FcεRI receptor binds several loops in the C3 domain that are more conserved than other regions of the domain: the Cε2-Cε3 linker, BC loop, DE loop, and FG loop (Fig. S4). There was only one conserved glycosylation site throughout all species, and it was located within the DE loop.

Igα is encoded by 3 Ig domain exons with the 7 amino acid hinge region encoded at the 5′ end of the C2 exon and one transmembrane exon. The secretory protein is 339 amino acids long. The residues involved in binding pIgR, FcαRI, and the J-chain are relatively conserved across all species when compared (Fig. S5) The *T. m. latirostris* Igα has features of both Igα1 and Igα2 in human. It is similar to human Igα1 because it has a cysteine in the C1 domain in the position that is involved in a disulfide bridge between the heavy and light chains. This bridge is not present in human Igα2. However, the length of the human Igα2 hinge region and the presence of glycosylation sites in C1 is similar to *T. m. latirostris*. There are also unique features of the manatee Igα: it has the most *N*-glycosylation sites (seven) of the species compared with a unique glycosylation site in both the C2 and C3 domains, and there is an additional cysteine in C1 between two highly conserved cysteines that is not observed in other species. There are no gaps between the Igε cytoplasmic tail exon and the Igα C1 exon and we identified only one subclass expressed.

3.1.3. VH gene segments

A total of 139 V heavy chain (VH) segments was identified on scaffold 53 of the *T. m. latirostris* genome assembly. Of these V segments, 118 were designated as pseudogenes (due to the presence of stop codons or frame shifts) and eight were designated as open reading frames (ORFs; due to a missing splice site, a missing motif, a truncation or insertion, or no detectable RSS) (Table S2). This leaves 13 potentially functional V segments encoded on this genomic scaffold, resulting in a locus with one of the fewest number of functional V segments and the greatest number of pseudogene V segments observed in mammals (Fig. 3, Table S3). There are gaps in the sequence among the V segments, and the 5′ boundary of the locus has not been confirmed, so this may be an underestimation. There is potential for more functional V segments to exist, but no other V segments have been identified on other scaffolds.

The functional regions encoded by the V segments were characterized under the international immunogenetics (IMGT) information numbering system (Lefranc, 1999). The length of the framework (FW) 1 region is 26 amino acids, FW 2 is 17 amino acids, and FW 3 is 39 amino acids for all V segments. Complementarity determining region 1 (CDR1) ranged from 8 to 10 amino acids with an average of 8.77, and CDR2 ranged from 6 to 8 amino acids with an average of 6.23.

The functional VH segments were separated into seven families based on 70% nucleotide identity, and the family size ranged from one to five segments (Fig. 4). When the pseudogene and ORF V segments were included, 44 families were identified (data not shown). Compared to human and mouse sequences representative of the three VH clans defined by the IMGT, the 13 functional VH segments cluster with clan I and II (3 and 10 segments, respectively) (Fig. 5). No functional VH segments belong to clan III.

3.1.4. DH gene segments

Between the most 3′ V segment and the most 5′ J segment, 48 functional D segments were identified that have 12-spaced RSSs on both sides. These segments range from 10 to 41 bp in length. The D segments were aligned and separated into 10 families based on 70% nucleotide identity; their size ranged from one to twelve segments (Fig. S6).

3.1.5. JH gene segments

Six functional J segments have been identified with 23-spaced RSS and a WGxG or WDxG motif. There is one ORF J segment that has a functional RSS and WGXG motif, but no splice site on the 3′ end; this segment is 5′ to the functional J segments. The predicted sequence length of the functional J segments range from 15 to 20 amino acids. There is a moderate level of sequence identity between the J segments, the highest is 82.2% between IGHJ2 and IGHJ3, and the rest range from 75.5% and 54.9% (Fig. S7).

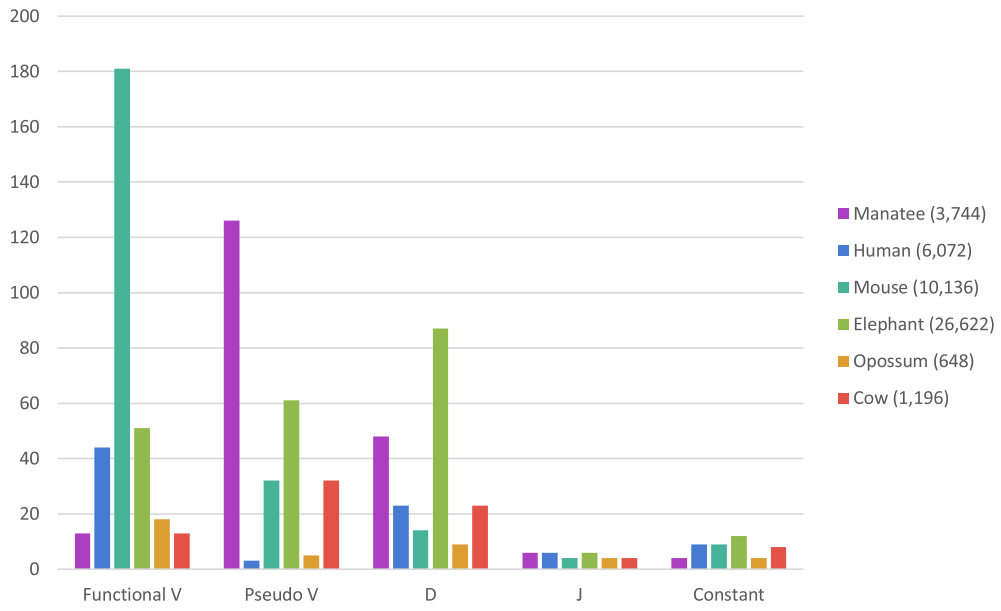


Fig. 3. *T. m. latirostris* has limited IgH diversity compared to other species. Graphical representation of the varied segmental diversity between species. The y-axis represents the number of segments or genes. In the legend, the number in parentheses indicates the potential number of segment combinations possible.

```
                                                CDR1                                                CDR2
Family 1
IGHV6    GVLSQVQLQESGPGLVKPSDSLTLTCTVFGFSITSSSC-CHWFCQPLGKDIEWIGEICYN-GSTYYSPSLKSCSTVYRDTAKNQIFLQQSSMTTKNTTMYYCAR
IGHV5    ................N..QT.S...A.S.V.....YNYWS.IR.STE.GLQ...Y.G.S-..N.N.....R.SIS...S..EFS..L..V..DD.AV.....
IGHV4    .I.......V..Q.MS.SLQT.S.S.S.SE.C..T.GHYWS.IH..ME.GL..VDY...D-.N...N..F.N.CSI.....P..FS..L....AEEIA..F...
IGHV2    ................S.QTIS.S.A.PRY....GYG-WS.IR.STE.GLQ.T.Y.S.S-..N.N.....R.SIS...S.YEFS..L.....DD.VV.....
IGHV1    .....I..Y.........QT.S...AFS.....T.GYCWS.M..STEEGLQ...V....-.GSG.N.FF....SIS...S..EFS.LL..V..D.KAV....K
Family 2
IGHV13   .A................QT.S...A.S.L.SS.Y--GVG.VR..P..GL..V.A.D.D-.D...N.T...RVSIT...S.S.VY.KLN.VNSED.AV....G
IGHV8    .A................QT.S...A.S...LS..--AVG.VR..P..GL..V.A.AGS-..A..N.T...RVSITK.NS.S.VYFKLN.VNSED.AT.....
Family 3
IGHV9    ..EVKDHIVQW.Y.T.VS.QK.R.A.ATYTLPVSE..--MS.IRLHP..GL..VSV.SA.-.N.N...N.Q.RVSIS...RE..V...L.TV.AEDPGV...TQ
IGHV7    ..EAKDHIVQW.E.T.VS.QK.S.A.ATYTLPVSE..--VS.IRLRP..GL..V.V.WA.-.N.N...N.Q.RVSIS...G...V...L.TV.AED.G....TQ
Family 4
IGHV10   WA....T.KK...R.M...QT.S....FSE.PLST.NIGVA.IH..P.NAL..LAH.SW.-ENKN.N......L.ISK..S...V..RMT..YPVDKDT.....
Family 5
IGHV11   .SQ.....VQ..AEVR..RAAMKVS..GTRYIF..YN--I.CVQ.TP.QGLD..EQ.NPRTWR.S.EQIFQGRV.TT.N.FLSTAYMKL..LNSDDM.V...V.
Family 6
IGHV12   AAH.....VQ..AEVKRLAE.VKIS.RAS.YTF..YA--M..VW.TPERGLQF..W.NT.T.KPK.YQGFTERYVFSM.SLVSTAY..I.G.KSED.AT.....
Family 7
IGHV3    ..CA....VK..AEVKR.GQ..RI..ETS.Y.F..YW--IS.VR.MP..GL....M.YPGDSD.R....FQ.QL.IS..NSISTTY..W..LKATD........
```

Fig. 4. Genomic VH segment alignment. The 13 functional VH segments were aligned using the ClustalW program in the Geneious software package. VH segments were separated into families based on 70% nucleotide identity. CDR regions are indicated by black boxes and were determined using the IMGT numbering system. Conserved cysteines are highlighted in yellow. Dots represent identity to the top sequence, which is used as a reference for all segments. Dashes represent gaps introduced to make the alignment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. IgH expression

IgH transcripts were sequenced from a peripheral blood leukocyte cDNA library of one *T. m. latirostris* using 5′ and 3′ RACE to identify the extent of mutation within one individual. Seventy-three unique transcripts (544 bp in length) were sequenced using 5′ RACE and eight unique transcripts (1269–1508 bp in length) using 3′ RACE. Sequences that had frame shifts in the CDR3 were not included. Of the 13 potentially functional genomic V segments, only four were expressed (IGHV5, IGHV7, IGHV8, and IGHV13) (Fig. S8). Considering all 73 5′ RACE sequences, which did not use a V specific primer, 95% of the V segments had at least 90% identity to IGHV8.

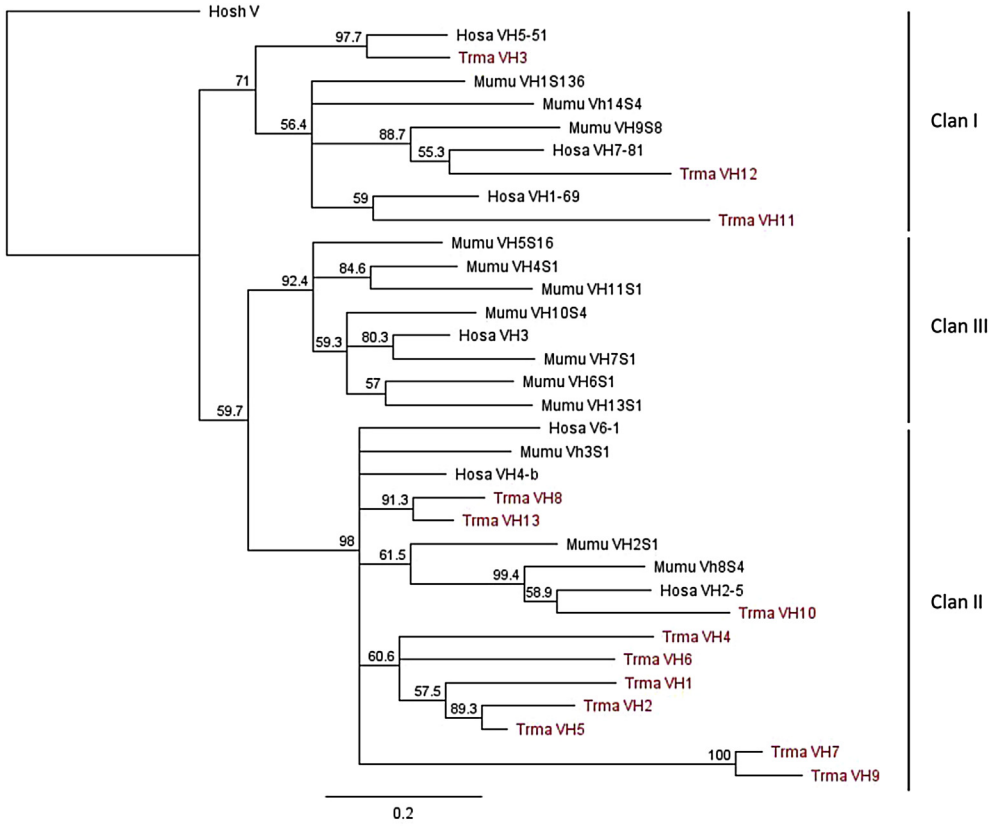


Fig. 5. *T. m. latirostris* lacks clan III V segments. A neighbor-joining tree of the 13 functional *T. m. latirostris* VH (Trma) segments and the human (Hosa) and mouse (Mumu) VH segments that are representative of the three mammalian V segment clans. The horned shark (Hosh) VH segment is used as an outgroup. Bootstrap values at each branch point are based on 1000 iterations. The *T. m. latirostris* VH segments are highlighted in red. Clan boundaries are indicated by the vertical lines to the left of the tree. Accession numbers are provided in Table S4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Eight of the ten D segment families were found in the expressed transcripts (Fig. 6). Family 1 was the most expressed (49%), but the most expressed individual D segment was IGHD47, which is in family 4. There was no apparent organization of family distribution or segment expression for the IGHD region.

Of the six potentially functional genomic J segments, three were found to be expressed (IGHJ1, IGHJ2, and IGHJ3) (Fig. S8). We did not include the sequence with the IGHJ2 segment in the figure because the IGHV8 segment was truncated during sequencing and the CDR3 had a frame shift. Considering 81 5′ and 3′ RACE sequences, 92% of the J segments had at least 95% identity to IGHJ3.

CDR3 length was measured according to IMGT numbering: the exclusive amino acids between the last cysteine of the V segments and the WGxG motif of the J segment (Lefranc, 1999). The CDR3 lengths of the expressed transcripts ranged from 5 to 23 amino acids, with an average of 13.4 amino acids.

Mutation rates were based on disagreements to the genomic V segment to which the transcript had the greatest (and at least 90%) identity. Four transcripts had no mutations. While the V segment CDRs contain only 16.4–18.0% of the nucleotides, they held on average 45.8% of the mutations. Of the CDR mutations, more tended to be nonsynonymous than synonymous ($p = 2.2e\text{-}16$) (Fig. 7).

4. Discussion

This study is the first to characterize the IgH genomic organization and expression in the Order Sirenia and the first repertoire analysis at the mRNA expression level from the early Afrotherian

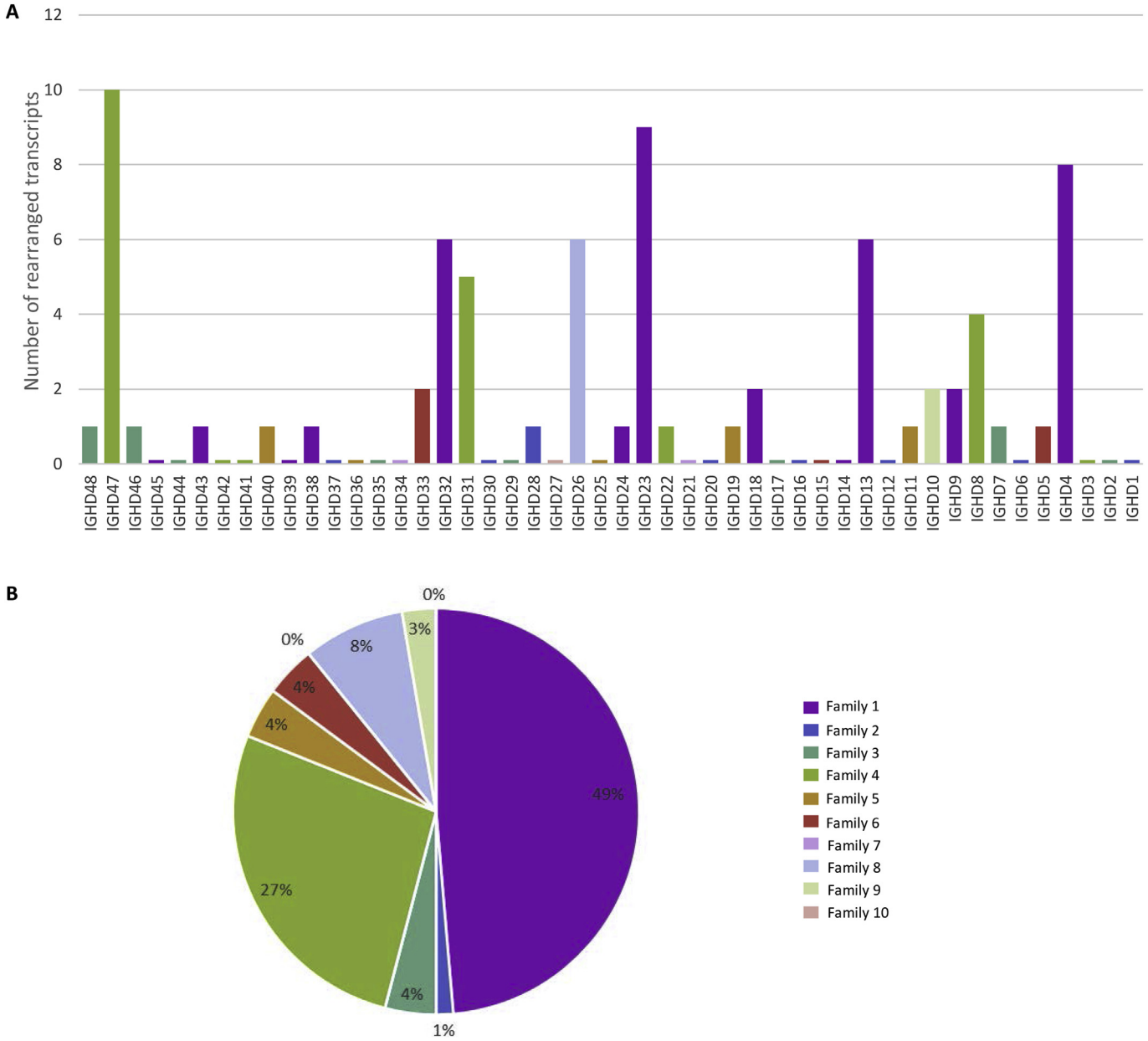


Fig. 6. DH segments show expression bias for Family 1. Graphical representation of DH segment usage in expressed transcripts. A) Frequency of individual DH segment usage and DH family genomic organization. DH segments are ordered as they appear in the *T. m. latirostris* genome. Each segment bar is colored according to the family it belongs to. B) DH segment expression by family. Percentages represent the proportion of the total number of expressed transcripts that used a DH segment from that family. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

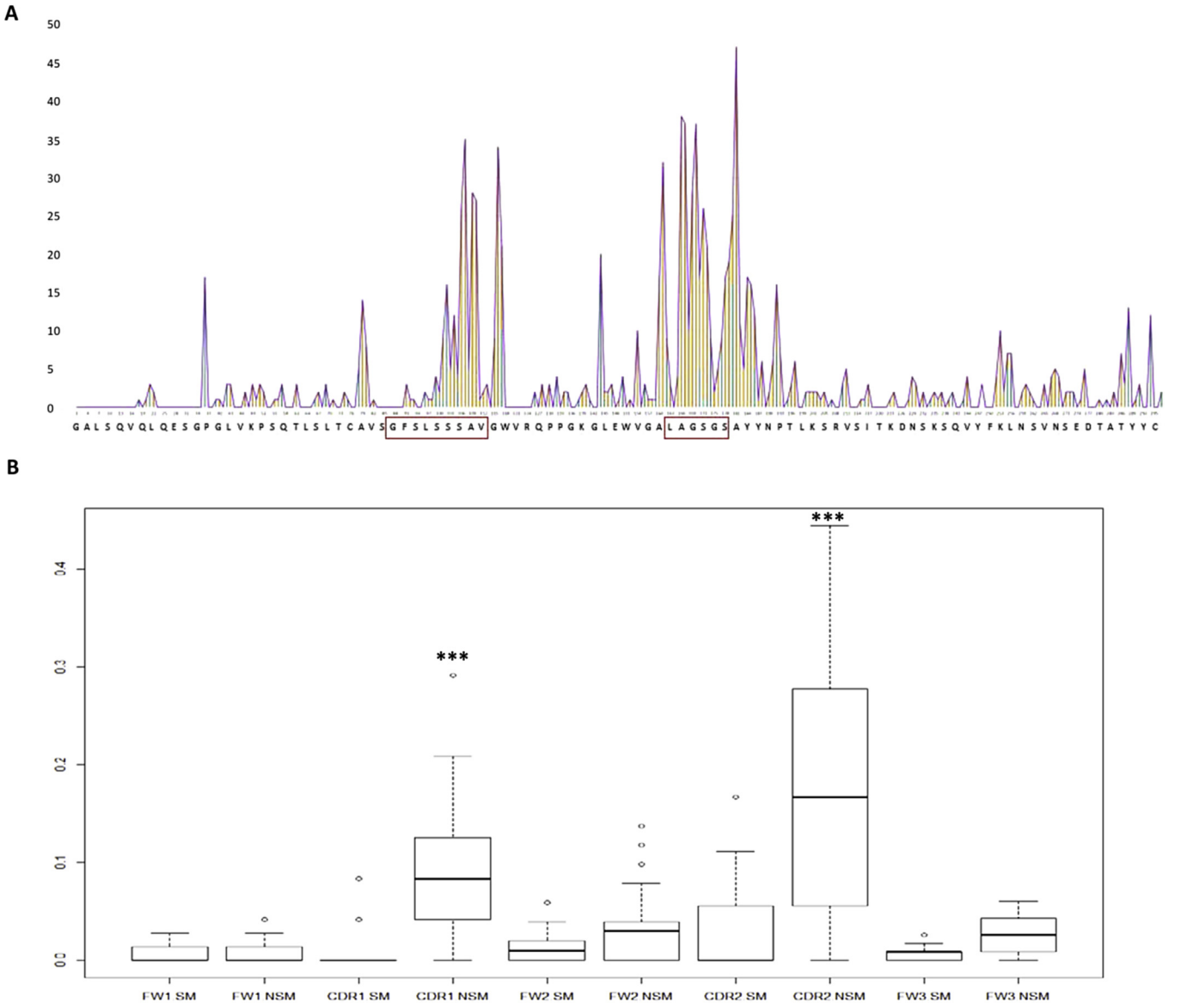


Fig. 7. Mutation rates are higher in the CDRs. Depiction of the trends of somatic hypermutation in expressed transcripts. A) Frequency of mutations at individual nucleotide positions in rearranged transcripts containing the IGHV8 segment. The pink line represents the total number of mutations at each nucleotide position across all transcripts sequenced, the vertical lines represent the frequency of synonymous mutations (green) to nonsynonymous mutations (yellow) at each position. The amino acid translation of the genomic IGHV8 segment is below the graph, which the CDRs boxed in red. B) A boxplot representation of the frequency of synonymous mutations (SM) and nonsynonymous mutations (NSM) between the framework (FW) and complementarity determining regions (CDRs). The number of mutations is normalized by the total number of nucleotides within each region. Three stars (***) represents a p value less than 0.001 using an ANOVA test. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

bifurcation of eutherian mammals. While the 5′ terminus of the locus has not been identified, there is little doubt that *T. m. latirostris* IgH locus has a translocon organization because no other genomic scaffolds contained V-like segments.

4.1. IgH constant regions

Igδ is found in most mammals, some reptiles, and teleost fish and varies in size, domain structure and genomic organization across species (Edholm et al., 2011). The number of domains can range from two to ten (Mushinski et al., 1980; White et al., 1985; Zhao et al., 2009). In some species, such as pig and catfish, the C1 domain is similar to Igμ and is coupled with the remaining Igδ-like domains (Wilson et al., 1997). Some species, such as the African elephant (*L. africana*), only have one domain encoded in the genome, which is thought to be a non-functional remnant lost to that species (Guo et al., 2011). And finally, some species, such as the opossum, do not have a delta gene in the IgH locus (Wang, X. et al., 2009). The *T. m. latirostris* genome has a similar C3 Igδ remnant found in the African elephant (*L. africana*) genome (Fig. S2).

Most mammalian species have multiple subclasses of Igγ with high sequence identity but have varying function due to differences in hinge flexibility and Fc receptor affinity (Liu and May 2012, Pincetic et al., 2014). The number of subclasses varies among species: humans and mice have four, elephants (*L. africana*) have nine (Guo et al., 2011), and pigs (*Sus scrofa*) have eleven (Butler et al., 2009). However, the *T. manatus latirostris* genome codes for only one Igγ constant region, similar to the opossum (*Monodelphis domesticus*) (Wang, X. et al., 2009). There is also only one Igα subclass that possesses features of both human Igα1 and Igα2. The

limited constant region diversity may be an artifact of the genome assembly because there are gaps, so there is a possibility that there are more Igγ and Igα subclasses. However, we have only found evidence for the one encoded in the genome.

Igε is exclusive to all three mammalian lineages and is involved in the immune response against parasites and nematodes (Vernersson et al., 2004). Several parasitic worms are prevalent in the *T. m. latirostris* population, so this isotype is critical for this species (Bando et al., 2014). The unique glycosylation site distribution of *T. m. latirostris* may impact the strength of mast cell binding and could lead to a more or less effective response against parasitic worms (Crispin, 2015).

Of all studied eutherian mammals, *T. m. latirostris* may have the fewest constant regions because they have only one copy of each isotype represented in the current genomic assembly. Functionally, they are tied with *M. domesticus* for least number of isotypes among studied mammals because the Igδ is a pseudogene remnant and is not expressed. Although they have few isotypes compared to other species, more studies could focus on the expressed proteins, specifically for IgG and IgA. It is possible that the *T. m. latirostris*-specific amino acid substitutions at receptor binding sites may not ablate function as is observed in the human-specific substitutions or the receptors have congruent mutations to retain function.

4.2. IgH segmental diversity

Compared to most mammals, *T. m. latirostris* has a limited repertoire based on their low segmental diversity from 13 functional V segments, 48 D segments, and six J segments. There are 3744 segmental combinations possible, which is lower than human, mice, and elephant, but higher than opossum and cow (Fig. 3). This may be an underestimation due to the incomplete assembly of the locus. The V segments group into seven families: one family of five segments, two families of two segments, and four families of one segment (Fig. 4). A majority of the segments (10) group to clan II, only three group to clan I, and none group to clan III (Fig. 5).

From the 13 functional V segments, only four have been identified in expressed transcripts in one sample. The four segments belong to three different families: IGHV5 from family 1, IGHV13 and IGHV8 from family 2, and IGHV7 from family 3. These four V segments all belong to clan II. Of the 73 5′ RACE sequences, 95% of the transcripts used the IGHV8 segment. There seems to be preferential expression for this segment, however these sequences are from one individual and may be an expression bias unique to this individual. Other V segments may also be expressed in other individuals or at lower expression levels that were not captured in this sequencing. In the future, we plan to analyze the expressed repertoire of more individuals from Florida, as well as from Antillean and Amazonian manatee populations. Comparing synonymous and non-synonymous mutations across the different framework and CDR regions, there is a highly significant selection bias for non-synonymous mutations in the CDR regions (Fig. 7). This shows that somatic hypermutation successfully selects for mutations that will affect antigen binding diversity in the genomically encoded CDR1 and CDR2 and against mutations that will change the framework structure.

While the V segment diversity is limited, the 48 D segments provide an increased opportunity for CDR3 diversity. The *T. m. latirostris* IgH locus is far from having the most D segments (*L. africana* has the most at 87 D segments (Guo et al., 2011)), but they are above average compared to many mammals including humans, mice, cow, and opossum (Fig. 3). *T. m. latirostris* has a slightly longer maximum D segment length at 41 bp than both human and elephant at 37 bp (Corbett et al., 1997; Guo et al., 2011). There are also ten D segment families compared to seven for both human and elephant. Within each family, there is not only high similarity in coding sequence and in the conserved heptamer and nonamer sequences of the RSS, but also in the 12 bp spacers (Fig. S6).

Although the D segments do not appear to be organized by family in the genome, there does appear to be expression bias of certain families (Fig. 6). Family 1 has the highest probability (25%) to be used because it has the most segments, but it was expressed nearly twice that percentage (49%). Surprisingly, the two families with the second most segments, family 2 and family 3, had low expression (1% and 4%, respectively). The two families that were not expressed were family 7, which has two segments, and family 10, which has one segment. Therefore, it is possible they are used but not represented in this data set because their probability of use is low. While there is one family that dominates expression, there is still widespread D segment usage, which is important in creating a diverse CDR3 repertoire. The high number of D segments, the high diversity of those segments, and the broad expression of those segments leads to increased antigen-binding diversity in the CDR3, whereas the diversity in CDR1 and CDR2 is limited to the few VH segments and SHM.

4.3. Pattern associated with the absence of clan III

The lack of clan III V segments may explain the low numbers of V segments and the high number of pseudogenes in *T. m. latirostris* because the clan III V segments have features that indicate their potential role in maintaining germline V segment diversity. Clan III contains the only mammalian IGHV segments that will group with IGHV segments from bony fish and reptiles, so they may have stemmed from a common origin and maintained to fulfill a specific and essential role (Andersson, 1995). Furthermore, it is suggested that clan III V segments play a noncoding role in targeting recombination due to the high level of nucleotide conservation in the framework 1 and 3 regions, the presence of the Chi recombination octamer, and their preferential use in rearrangements in human and mouse (Schroeder and Wang, 1990; Tutter and Riblet, 1989). Also, the regulatory elements that determine chromatin states are organized by clan throughout the mouse IgH locus, so transcriptional regulation is clan-specific (Bolland et al., 2016). The absence of clan III in *T. m. latirostris* may have reduced the levels of recombination, allowing the V segments to pseudogenize without much consequence. This could explain the large disparity between the number of functional V segments and pseudogenes (Fig. 3).

There are only three other species studied that lack clan III V segments and they also have lower numbers of functional V segments: cows have 13 (Niku et al., 2012), sheep have approximately 10 (Dufour et al., 1996), and horses have 14 (Walther et al., 2015). However, cows and sheep have other mechanisms to create diversity in their IgH repertoires. In cows, approximately 10% of the Ig repertoire has an ultralong CDR3 due to unusually long D segments (148 bp) (Wang, F. et al., 2013). These ultralong immunoglobulins have various disulfide bridge configurations and are highly mutable, which increases their antigen binding diversity without needing segmental diversity (de Los Rios et al., 2015). In sheep, Ig SHM is not restricted to post-antigen exposure, so their naïve repertoire is not limited to what is encoded in the genome (Reynaud et al., 1995).

T. m. latirostris is the first species independent of the ungulate lineage that has evidence of this loss of all clan III V segments. The lack of clan III most likely had a common origin in cows and sheep because of their evolutionary proximity, but it evolved independently in horse and *T. m. latirostris*. It is possible that *T. m. latirostris* uses diversifying mechanisms similar to cows and sheep, such as pre-antigen exposure SHM, but has not yet been determined. It is

unclear whether similar evolutionary mechanisms caused this convergent evolution because there are significantly more pseudogenes in *T. m. latirostris* than the other three species. However, the loss of clan III consistently correlates with considerably lower numbers of V segments compared to other eutherians. This pattern reveals that clan III may have a more critical role in maintaining V segment diversity than previously thought.

This study is the first characterization of the *T. m. latirostris* IgH locus. While the segmental diversity is limited both in the number of genomically encoded segments and number of expressed segments, somatic hypermutation is still an efficient source of CDR-targeted diversity. This limited segmental diversity could be due to the lack of clan III V segments that are thought to be important for recombination regulation in other species. Despite their limited segmental diversity, manatees are theorized to have a strong immune system when they are not under cold stress (Bonde et al., 2004). Therefore, the lack of Ig diversity may indicate a stronger dependence of cell-mediated immunity through the T cell receptor. The two primary disease causing agents in *T. m. latirostris* (chronic red tide exposure in leukocytes and papilloma virus) are intracellular, which typically require a cytotoxic T cell response (Bossart et al., 1998). The lack of diversity may also be mitigated through the gamma/delta T cell receptor, which is not MHC restricted and therefore can act similarly to a membrane-bound antibody.

Understanding the genomic organization and expression profile of the immunoglobulin heavy chain locus for the Florida manatee is important because the manatee represents three unique groups. First, Afrotheria represents an important radiation in mammalian evolution because the most basal split within eutherian mammals may be between Afrotheria and all other placental mammals (Foley et al., 2016; Murphy et al., 2001, 2004; Springer et al., 2007). Second, marine mammals are not only unique in their adaptations to the aquatic environment, but they also have microbiota that are distinct from both terrestrial mammals and the water they inhabit, which could impact segmental selection (Bik et al., 2016). Finally, the absence of clan III V segments is not only rare, but seems to have a big impact on the maintenance of V segment diversity. The results of this study not only fill the gap of knowledge in manatee immunity, but they question the evolutionary mechanisms that maintain diversity at this complex locus.

Acknowledgments

We thank Jeannine Ott for critical review of the manuscript, and Alex Bass for annotation of D segment use. This work was funded by the Texas A&M-CAPES Collaborative Research Grant Program from an award (2014-028) to MFC, MPC-S, and LS. Collection of samples were obtained from wild manatees in conjunction with health assessments conducted in Crystal River, Florida. The sample collection and research was carried out under U.S. Geological Survey (USGS) Institutional Animal Care and Use Committee (IACUC) guidelines and under the authority granted to the USGS by the U.S. Fish and Wildlife Service Federal wildlife research permit #MA791721-4. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.dci.2017.01.022.

References

Andersson, E., 1995. The Evolutionary Study of the Immunoglobulin Heavy Chain Genes of a Bony Fish, Rainbow Trout (Oncorhynchus mykiss) (Doctoral Dissertation). Semantic Scholar. (ISSN 0346–6612).

Bando, M., Larkin, I.V., Wright, S.D., Greiner, E.C., 2014. Diagnostic stages of the parasites of the Florida manatee, *Trichechus manatus latirostris*. J. Parasitol. 100 (1), 133–138.

Bik, E.M., Costello, E.K., Switzer, A.D., Callahan, B.J., Holmes, S.P., Wells, R.S., Carlin, K.P., Jensen, E.D., Venn-Watson, S., Relman, D.A., 2016. Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea. Nat. Commun. 7, 10516.

Bolland, D.J., Koohy, H., Wood, A.L., Matheson, L.S., Krueger, F., Stubbington, M.J., Baizan-Edge, A., Chovanec, P., Stubbs, B.A., Tabbada, K., Andrews, S.R., Spivakov, M., Corcoran, A.E., 2016. Two mutually exclusive local chromatin states drive efficient v(d)j recombination. Cell Rep. 15 (11), 2475–2487.

Bonde, R., Aguirre, A.A., Powell, J., 2004. Manatees as sentinels of marine ecosystem health: are they the 2000-pound canaries? EcoHealth 1 (3).

Bossart, G.D., 1995. Immunocytes of the Atlantic Bottlenose Dolphin (Tursiops truncatus) and West Indian Manatee (Trichechus manatus latirostris) : Morphologic Characterizations and Correlations between Healthy and Disease States Under Free-ranging and Captive Conditions (Doctoral Dissertation). FIU Electronic Theses and Dissertations. (1772).

Bossart, G.D., Baden, D.G., Ewing, R.Y., Roberts, B., Wright, S.D., 1998. Brevetoxicosis in manatees (*Trichechus manatus latirostris*) from the 1996 epizootic: gross, histologic, and immunohistochemical features. Toxicol. Pathol. 26 (2), 276–282.

Bossart, G.D., Ewing, R.Y., Lowe, M., Sweat, M., Decker, S.J., Walsh, C.J., Ghim, S.J., Jenson, A.B., 2002. Viral papillomatosis in Florida manatees (*Trichechus manatus latirostris*). Exp. Mol. Pathol. 72 (1), 37–48.

Brodeur, P.H., Riblet, R., 1984. The immunoglobulin heavy chain variable region (igh-v) locus in the mouse i. One hundred igh-v genes comprise seven families of homologous genes. Eur. J. Imunology. 14, 922–930.

Butler, J.E., Wertz, N., Deschacht, N., Kacskovics, I., 2009. Porcine igg: structure, genetics, and evolution. Immunogenetics 61 (3), 209–230.

Corbett, S.J., Tomlinson, I.M., Sonnhammer, E.L.L., Buck, D., Winter, G., 1997. Sequence of the human immunoglobulin diversity (d) segment locus: a systematic analysis provides no evidence for the use of dir segments, inverted d segments, “minor” d segments or d-d recombination. J. Mol. Biol. 270 (4), 587–597.

Crispin, M., 2015. Breaking the allergic response by disrupting antibody glycosylation. J. Exp. Med. 212 (4), 433.

de Los Rios, M., Criscitiello, M.F., Smider, V.V., 2015. Structural and genetic diversity in antibody repertoires from diverse species. Curr. Opin. Struct. Biol. 33, 27–41.

Domning, D.P., 1982. Evolution of manatees: a speculative history. J. Paleontol. 56 (3), 599–619.

Dufour, V., Malinge, S., Nau, F., 1996. The sheep ig variable region repertoire consists of a single vh family. J. Immunol. 156, 2163–2170.

Edholm, E.S., Bengten, E., Wilson, M., 2011. Insights into the function of igd. Dev. Comp. Immunol. 35 (12), 1309–1316.

Foley, N.M., Springer, M.S., Teeling, E.C., 2016. Mammal madness: is the mammal tree of life not yet resolved? Philos. Trans. R. Soc. -l. B Biol. Sci. 371 (1699).

Guo, Y., Bao, Y., Wang, H., Hu, X., Zhao, Z., Li, N., Zhao, Y., 2011. A preliminary analysis of the immunoglobulin genes in the African elephant (*Loxodonta africana*). PLoS One. 6 (2), e16889.

Kearse, M.,R., Moir, A., Wilson, S., Stones-Havas, C.M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., Drummond, A., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28 (12), 1647–1649.

Kellogg, M.E., Burkett, S., Dennis, T.R., Stone, G., Gray, B.A., McGuire, P.M., Zori, R.T., Stanyon, R., 2007. Chromosome painting in the manatee supports Afrotheria and Paenungulata. BMC Evol. Biol. 7, 6.

Kirkham, P.M., Mortari, F., Newton, J.A., Schroeder Jr., H.W., 1992. Immunoglobulin vh clan and family identity predicts variable domain structure and may influence antigen binding. EMBO 11 (2), 603–609.

Lefranc, M.P., 1999. Immunologist 7, 132–136.

Liu, H., May, K., 2012. Disulfide bond structures of igg molecules: structural variations, chemical modifications and possible impacts to stability and biological function. MAbs 4 (1), 17–23.

Mcgee, J.L., 2012. Immunological Investigations in the West Indian Manatee (Trichechus manatus) and Asian Elephant (Elephas maximus) (Doctoral Dissertation). University of Florida Digital Collections. (UFE0044127:00001).

Murphy, W.J., Eizirik, E., O'Brien, S.J., Madsen, O., Scally, M., Douady, C.J., Teeling, E., Ryder, O.A., Stanhope, M.J., de Jong, W.W., Springer, M.S., 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science 294, 2348–2351.

Murphy, W.J., Pevzner, P.A., O'Brien, S.J., 2004. Mammalian phylogenomics comes of age. Trends Genet. 20 (12), 631–639.

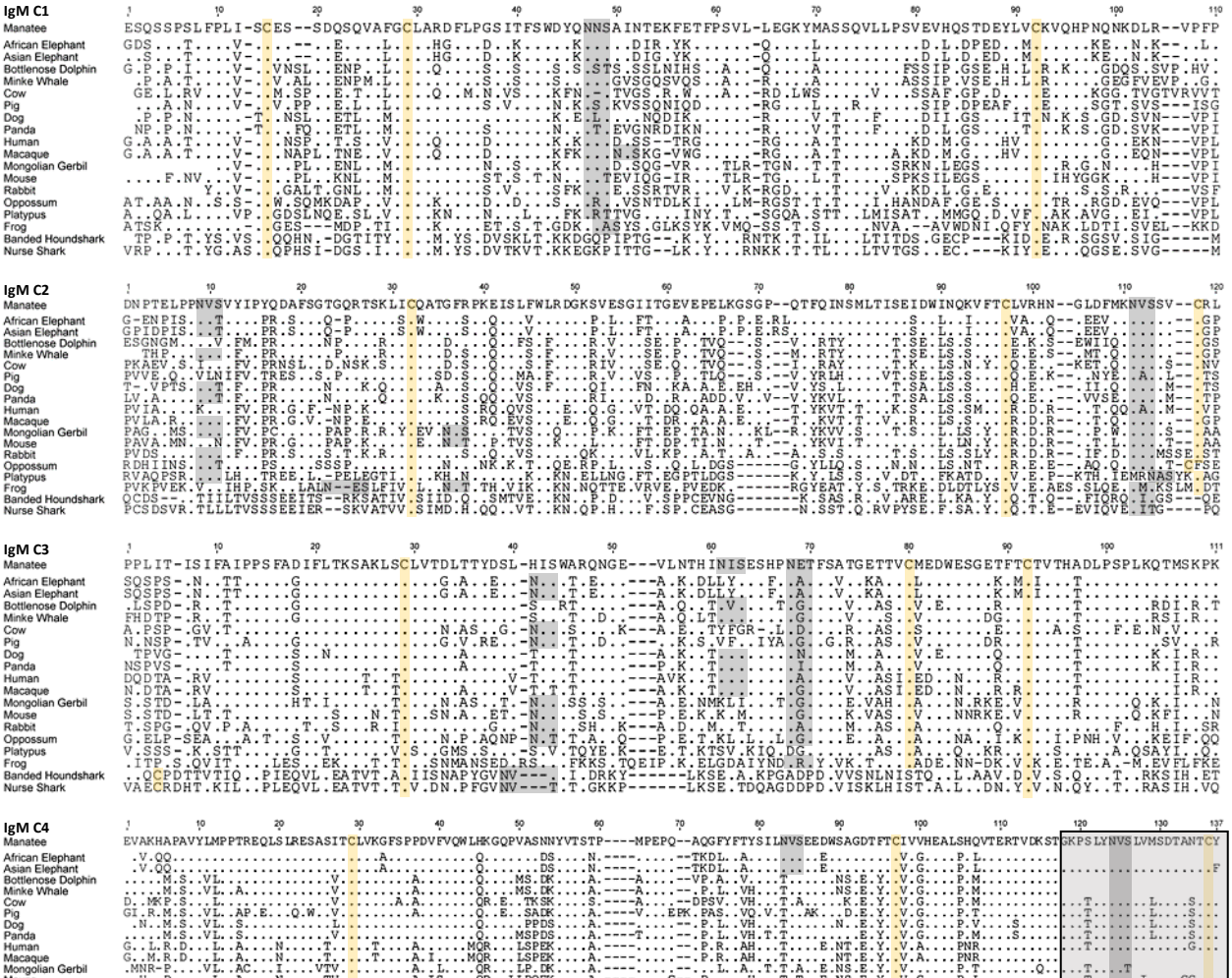
Mushinski, J.F., Blattner, F.R., Owens, J.D., Finkelman, F.D., Kessler, S.W., Fitzmaurice, L., Potter, M., Tucker, P.W., 1980. Mouse immunoglobulin d: construction and characterization of a cloned delta chain cdna. Proc. Natl. Acad. Sci. U. S. A. 77 (12), 7405–7409.

Niku, M., Liljavirta, J., Durkin, K., Schroderus, E., Iivanainen, A., 2012. The bovine genomic DNA sequence data reveal three ighv subgroups, only one of which is functionally expressed. Dev. Comp. Immunol. 37 (3–4), 457–461.

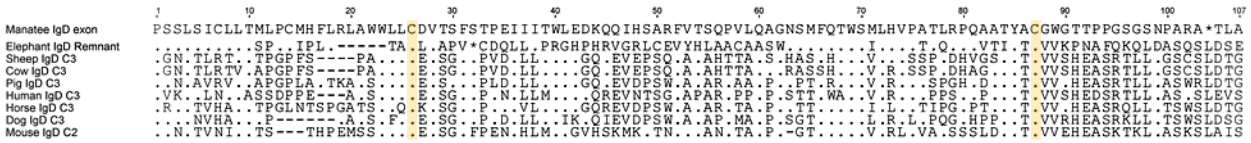
Nossal, G.J.V., Lederberg, J., 1958. Antibody production by single cells. Nature 181 (4620), 1419–1420.

O'Shea, T.J., 1988. The past, present, and future of manatees in the southeastern

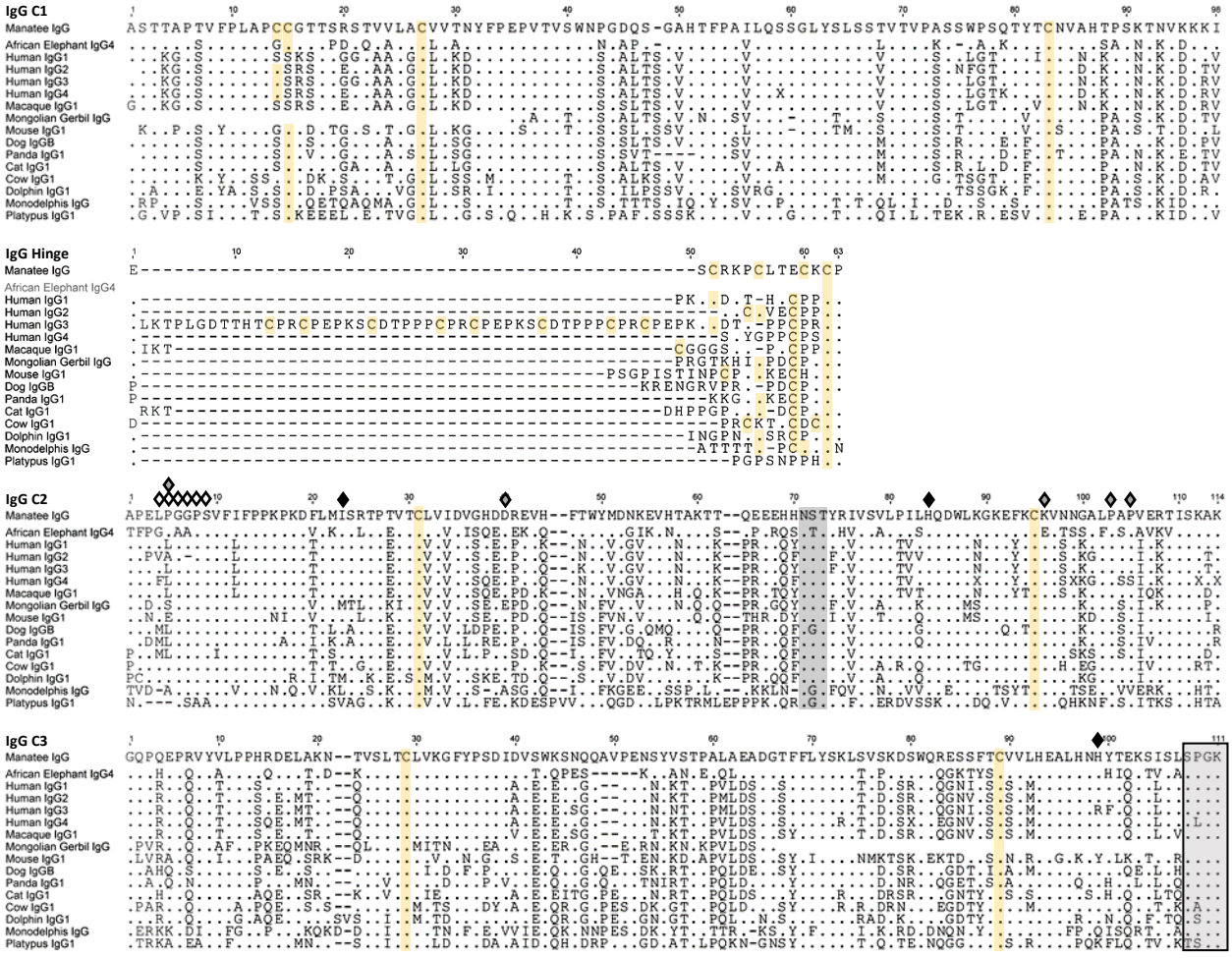
United States: realities, misunderstandings, and enigmas. In: Proceedings of the Third Southeastern Nongame and Endangered Wildlife Symposium, pp. 184–203.
Pincetic, A., Bournazos, S., DiLillo, D.J., Maamary, J., Wang, T.T., Dahan, R., Fiebiger, B.M., Ravetch, J.V., 2014. Type i and type ii fc receptors regulate innate and adaptive immunity. Nat. Immunol. 15 (8), 707–716.
Reynaud, C., Garicia, C., Hein, W.R., Weill, J., 1995. Hypermutation generating the sheep immunoglobulin repertoire is an antigen-independant process. Cell 80, 115–125.
Schroeder Jr., H.W., Wang, J.Y., 1990. Preferential utilization of conserved immunoglobulin heavy chain variable gene segments during human fetal life. Proc. Natl. Acad. Sci. U. S. A. 87, 6146–6150.
Springer, M.S., Burk-Herrick, A., Meredith, R.W., Eizirik, E., Teeling, E., O'Brien, S.J., Murphy, W.J., 2007. The adequacy of morphology for reconstructing the early history of placental mammals. Syst. Biol. 56 (4), 673–684.
Sweat, J.M., Johnson, C.M., Marikar, Y., Gibbs, E.P., 2005. Characterization of surface interleukin-2 receptor expression on gated populations of peripheral blood mononuclear cells from manatees, *Trichechus manatus latirostris*. Vet. Immunol. Immunopathol. 108 (3–4), 269–283.
Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10 (3), 512–526.
Tonegawa, S., 1983. Somatic generation of antibody diversity. Nature 302 (14), 575–581.
Tutter, A., Riblet, R., 1989. Conservation of an immunoglobulin variable-region gene family indicates a specific, noncoding function. PNAS 86, 7460–7464.
Vernersson, M., Aveskogh, M., Hellman, L., 2004. Cloning of ige from the Echidna (*Tachyglossus aculeatus*) and a comparative analysis of ε chains from all three extant mammalian lineages. Dev. Comp. Immunol. 28 (1), 61–75.
Vianna, J.A., Bonde, R.K., Caballero, S., Giraldo, J.P., Lima, R.P., Clark, A., Marmontel, M., Morales-Vela, B., De Souza, M.J., Parr, L., Rodriguez-Lopez, M.A., Mignucci-Giannoni, A.A., Powell, J.A., Santos, F.R., 2006. Phylogeography, phylogeny and hybridization in Trichechid sirenians: implications for manatee conservation. Mol. Ecol. 15 (2), 433–447.
Walsh, C.J., Butawan, M., Yordy, J., Ball, R., Flewelling, L., de Wit, M., Bonde, R.K., 2015. Sublethal red tide toxin exposure in free-ranging manatees (*Trichechus manatus*) affects the immune system through reduced lymphocyte proliferation responses, inflammation, and oxidative stress. Aquat. Toxicol. 161, 73–84.
Walsh, C.J., Luer, C.A., Noyes, D.R., 2005. Effects of environmental stressors on lymphocyte proliferation in Florida manatees, *Trichechus manatus latirostris*. Vet. Immunol. Immunopathol. 103 (3–4), 247–256.
Walsh, C.J., Romano, T.R., Stott, J.L., Manire, C.A., 2003. Diagnostic Indicators of Manatee Immune Function. Mote Marine Laboratory Technical Report, p. 903.
Walther, S., Rusitzka, T.V., Diesterbeck, U.S., Czerny, C.P., 2015. Equine immunoglobulins and organization of immunoglobulin genes. Dev. Comp. Immunol. 53 (2), 303–319.
Wang, D.Z., 2008. Neurotoxins from marine dinoflagellates: a brief review. Mar. Drugs. 6 (2), 349–371.
Wang, F., Ekiert, D.C., Ahmad, I., Yu, W., Zhang, Y., Bazirgan, O., Torkamani, A., Raudsepp, T., Mwangi, W., Criscitiello, M.F., Wilson, I.A., Schultz, P.G., Smider, V.V., 2013. Reshaping antibody diversity. Cell 153 (6), 1379–1393.
Wang, X., Olp, J.J., Miller, R.D., 2009. On the genomics of immunoglobulins in the gray, short-tailed opossum *Monodelphis domestica*. Immunogenetics 61 (8), 581–596.
White, M.B., Shen, A.L., Word, C.J., Tucker, P.W., Blattner, F.R., 1985. Human immunoglobulin d: genomic sequence of the delta heavy chain. Science 228, 733–737.
Wilson, M., Bengten, E., Miller, N.W., Clem, L.W., Pasquier, L.D., Warr, G.W., 1997. A novel chimeric ig heavy chain from a teleost fish shares similarities to igd. Proc. Natl. Acad. Sci. U. S. A. 94, 4593–4597.
Zhao, Y., Cui, H., Whittington, C.M., Wei, Z., Zhang, X., Zhang, Z., Yu, L., Ren, L., Hu, X., Zhang, Y., Hellman, L., Belov, K., Li, N., Hammarstrom, L., 2009. *Ornithorhynchus anatinus* (platypus) links the evolution of immunoglobulin genes in eutherian mammals and nonmammalian tetrapods. J. Immunol. 183 (5), 3285–3293.



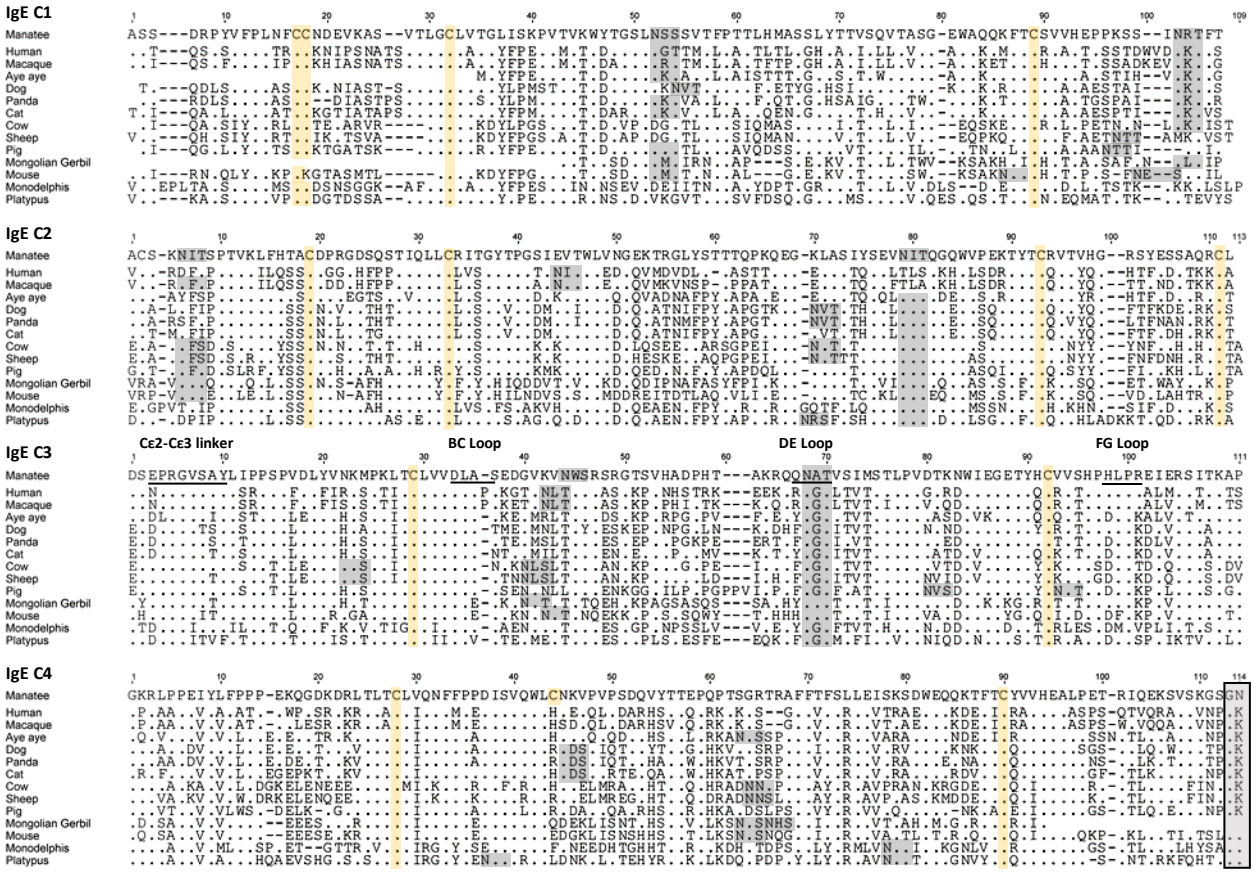
Supplemental Figure 1. Comparative Alignment of Igμ Constant Region. Amino acid alignment of Igμ constant region sequences from different species. Cysteine residues are highlighted in yellow and N-linked glycosylation sites are highlighted in grey. The grey box in the C4 domain indicated the secretory tail portion of the exon. Dots represent similarity to the *T. m. latirostris* sequence and dashes represent gaps introduced to make the alignment. Accession numbers are provided in Table S3.



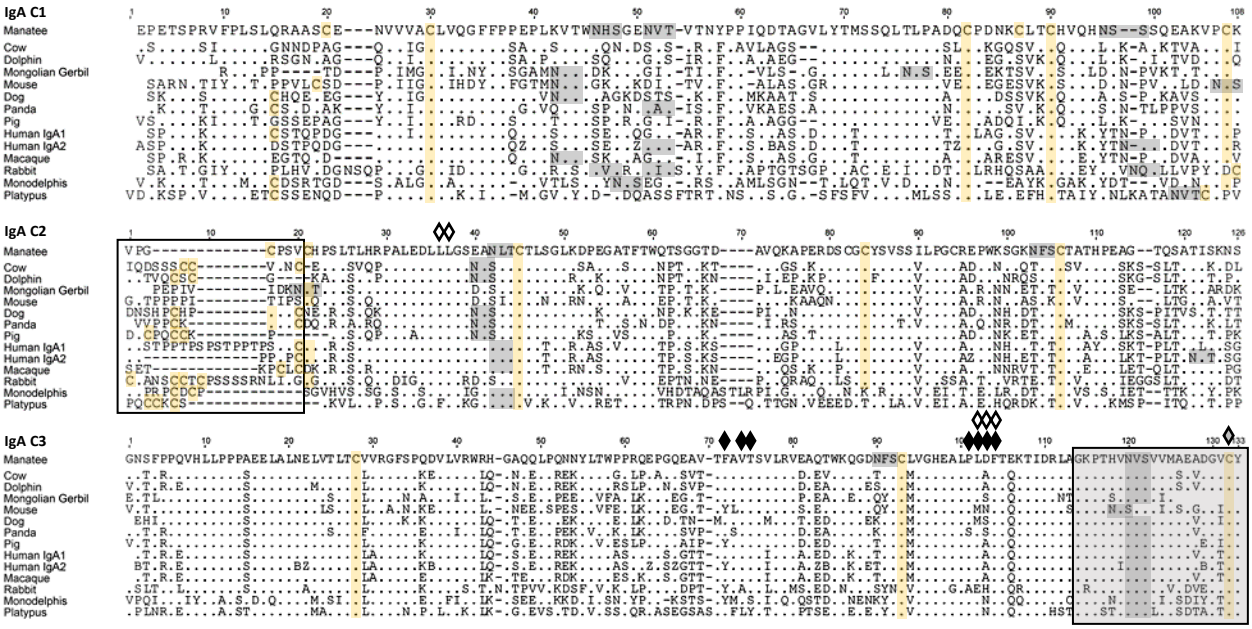
Supplemental Figure 2. Comparative Alignment of Igδ C3 domain. Amino acid alignment of the *T. m. latirostris* Igδ-like exon with the C3 Igδ exon in other species. The mouse Igδ only has two constant domains, so it is most similar to the C2 domain. Cysteines are highlighted in yellow. Dots represent similarity to the *T. m. latirostris* sequence and dashes represent gaps introduced to make the alignment. Accession numbers are provided in Table S3.



Supplemental Figure 3. Comparative Alignment of Igγ Constant Region. Amino acid alignment of Igγ constant region sequences from different species. Cysteines are highlighted in yellow and N-linked glycosylation sites are highlighted in grey. White diamonds represent amino acids important for FcγRI binding, grey diamonds represent amino acids important for C1q binding (Morgan et al. 1995), and black diamonds represent amino acids important of FcRn binding (West and Bjorkman 2000). The grey box in the C3 domain indicates the secretory tail portion of the exon. Dots represent similarity to the *T. m. latirostris* sequence and dashes represent gaps introduced to make the alignment. Accession numbers are provided in Table S3.



Supplemental Figure 4. Comparative Alignment of Igε Constant Region. Amino acid alignment of Igε constant region sequence from different species. Cysteine residues are highlighted in yellow and N-linked glycosylation sites are highlighted in grey. The amino acids important for FcεRIα binding are underlined and their region indicated above (Garman et al. 2000). Dots represent similarity to the *T. m. latirostris* sequence and dashes represent gaps introduced to make the alignment. Accession numbers are provided in Table S3.

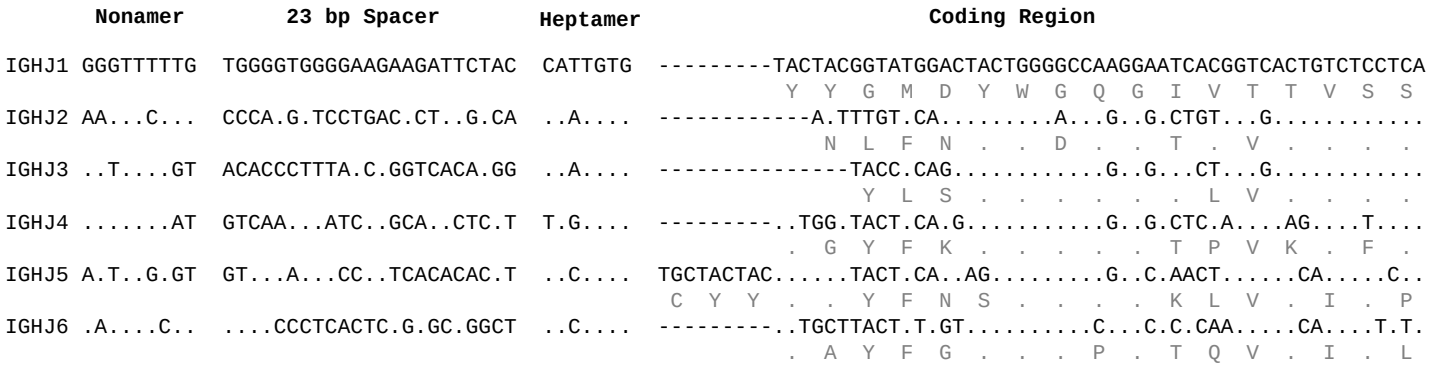


Supplemental Figure 5. Comparative Alignment of Igα Constant Region. Amino acid alignment of Igα constant region sequence from different species. Cysteine residues are highlighted in yellow and N-linked glycosylation sites are highlighted in grey. White diamonds represent amino acids that are important for FcαRI binding (Pleass et al. 1999), grey diamonds represent amino acids important for J chain binding (Krugmann et al. 1997), and black diamonds represent amino acids important for pIgR binding (Lewis et al. 2005). The white box in the C2 domain indicates the hinge region portion of the exon. The grey box in the C3 domain indicates the secretory tail portion of the exon. Dots represent similarity to the *T. m. latirostris* sequence and dashes represent gaps introduced to make the alignment. Accession numbers are provided in Table S3

	Nonamer	12 bp Spacer	Heptamer	Coding Region	Heptamer	12 bp Spacer	Nonamer
Family 1							
IGHD13	GGATTGTGT	GAGGGCATGTGT	CACCGTG	ACAATATTGTAGTGGTGCTTACTGCTACTAC----------	CACAGTG	ACATAGCCACTA	CACAAAAGC
IGHD38	A.G.....	....T.....	..TT...	......C.....C.C..T...T.ATG.----------	.....	...CGT.AC..G	..G.....
IGHD9	..TA.....	.G..C.....	...T...	........ACCT.T.GAG.....A.AGA.----------	.....	.....ATG.G	.C.....
IGHD43	.....	.G..A.....	...T...	.....C.A...GAG......TGGA.----------	.....	.TGC....CTGC	.CA...GC.
IGHD45	.....	.G.....	...A...	.....C...T.GAG...T...A.....----------	..T....	..TC.A.AC..G	.C.....
IGHD18	.....	.G.....	....A..	.....C...T...GCAC..AT......----------	.....	.TG....AC..G	.C.....
IGHD23	.....	...CAA.....	.....	........CA......AG.....A.----------	.....	...C..TGG.AG	.C......T
IGHD4	.....	.....	.....	........A.....AG...T..GA.----------	.....	...C...AC..G	.C.....
IGHD14	.....	.G.....CA.	...T...	......C..C.----------GAC.----------	..G....	...C..TGTAC.	TC.....
IGHD39	........A	.G.....C..	...T...	..G......C.C...G.GG...A..A.A.----------	.....	.A.G..T.CAC.	TC....G..
IGHD24	.....	.G.....G..	....A..	.....C.....C.G.....AGA.----------	..T....	...C..TGCTCT	TT.....
IGHD32	A.....	.G.....G....	...T...	.....AC...T.G.G...T.ATGGA.----------	.....	...C...ACA.G	.T.....
Family 2							
IGHD6	CA..AT...	CTA..TTCC...	...T...	G-GCGG.-------.GAG.AGTA.TAGT.----------	.....	G...G.TTC..G	TT.....T.
IGHD28	CAG.AT...	CT...TTCC...	...T...	G-GTGGG-------.GAGCA.TA.TAGT.----------	..AG...	....G...C..G	TT.....T.
IGHD16	CA..AT...	CT..ATCCC...	...T...	G-GTGG.-------.AAG.AGTAATAGT.----------	.....	G.....C..G	TT.....TA
IGHD1	CA..AT...	CT...TCCC...	...T...	G-GTGG.-------.GTG.AGTA.TA.T.----------	.....T	G...G...T...	TT.....TG
IGHD20	CA.CAT...	CT.A.TCCC...	...T...	G-.TGG.-------.GAA.AGTA.TAGT.G.----------	.....	CA..G...CT..	TT.....TA
IGHD37	CA..AT...	CT...TCCC...	...T...	GG.GGG.-------.GAG.AGTA.TATT.C.----------	.....	CA..G...C..G	TT.....TA
IGHD12	CAG.A....	CT..A.CCC...	...T...	G-GTGG.-------.GAG.GGTA.TAGT.----------	.....	CA..GA..C..G	TT.....TA
IGHD30	CA..AT...	TT...TCCT...	...T...	G-.TG.G-------.GAG.AGTATTAGT.----------	...G...	G...G..TC..G	.TT....T.
Family 3							
IGHD2	ACT..T...	C.A..ACCC.CC	..TT...	TA.C...A----------ACTAC----------	..GGT..	...G.C...GC.	GCA....TT
IGHD7	ACT..T...	C.A..ACCC.CC	G.TT...	TA.C...A----------ACTAC----------	..G....	...G.C.TGGC.	GCA....T.
IGHD17	A.T..T...	C.A..ACCC.CC	T.TT...	.A.C...A----------ACTAC----------	....T..	.G.G.C..GGA.	GCAC...T.
IGHD35	A.T..T...	C.A..ACCC.CC	T.TT...	.A.C...A----------ACTAC----------	....A..	.G.G.C..GGC.	GCA....CT
IGHD29	A.T..T...	C.A..AC.C.CC	T.TT...	TA.C...A----------ACTAC----------	..T....	.G.G.C...G..	GCA...GCT
IGHD44	.CT..T...	C.A..ACGC.TC	A.TT...	TA.C...A----------ACTAC----------	......A	TT.C.C..GGC.	GGA....C.
IGHD48	ACT..T...	C.A..GTCA.CC	..TT...	TA.CA..A----------ACTA.----------	....A..	...GTCT.TGC.	GCA....C.
IGHD46	.AT..T...	C.AT.AC.C.CC	T.TT...	TA.TG..A----------ACTAC----------	..G....	.G...C...GC.	ACA....C.
Family 4							
IGHD3	......G.G	TCAA.....CA.	..TT...	GT.T..CCT.GCCA..T..GGT.....TGG.----------	....ACC	T..C.ATGC..T	GCA....C.
IGHD47	.A....G.G	.CA...G...A.	...T...	TT.TG..--------...TA.GGT.AT..TGGA----------	CC	T..C.ATGC..T	GCA....C.
IGHD8	......G.G	ACA.ATG..CAC	...T...	GT.T..CA..TA....A..ACT..G..TG..----------	.....CT	T..C.ATACA.T	GCA....C.
IGHD22	......G.G	.CA..TG..CAC	.T.T...	GT.T..C...GA.A..TA.GGT.....TGG.----------	.....CT	T..C.ATGC..T	GTA....C.
IGHD31	..T...A.G	.CA.ATG..CAC	...TC..	GT.T..C...GA....T..GGT.....TGG.----------	..T..CC	T..C.ATGC..T	GCA....C.
IGHD41	......G.G	.CA.......AC	...T...	GT.T..C...GA.A..T..AGT..T..TGC.----------	CC	T..C.ATGC..T	GCA....C.
IGHD42	......G.G	.CA......CAC	...T...	GT.T..C...GA.A..TA.A....T..TGT.----------	....TCC	T..C.AAGC..T	GCA....C.
Family 5							
IGHD19	.....C...	.CT.C.C.....	..AG...	G----------GTA.GGTG.GAG.----------	...T...	TG.A..A.TG.G	.CA...CTG
IGHD36	.....C...	.CT.C.C.....	..AG...	G----------GTA.GGTG.GAG.----------	...T...	TG.AT.ATTT.G	.CA....CA
IGHD11	.....C...	.CT.C.C.....	..AG...	G----------ATA.GGTGAGAGG.----------	...T...	TG.A..A.CA.G	.CA....CT
IGHD40	.....C...	.CT.C.C.....	..AG...	G----------GTA.GTTGAAAG.----------	.....	TG.G..A.CA.G	.CA....CT
IGHD25	.....C...	.CTAC.C.....	..AG...	G----------GTA.AGTA.AAG..GA----------	...T...	CA.A..ATCAC.	TCA..T.CT
Family 6							
IGHD5	A....TG..	AT.A.GT.C.AA	TG.T...	.T.C..CA.AG--------GT-ACTAC----------	...T...	.TTG.CTTC.C.	ACA..C.C.

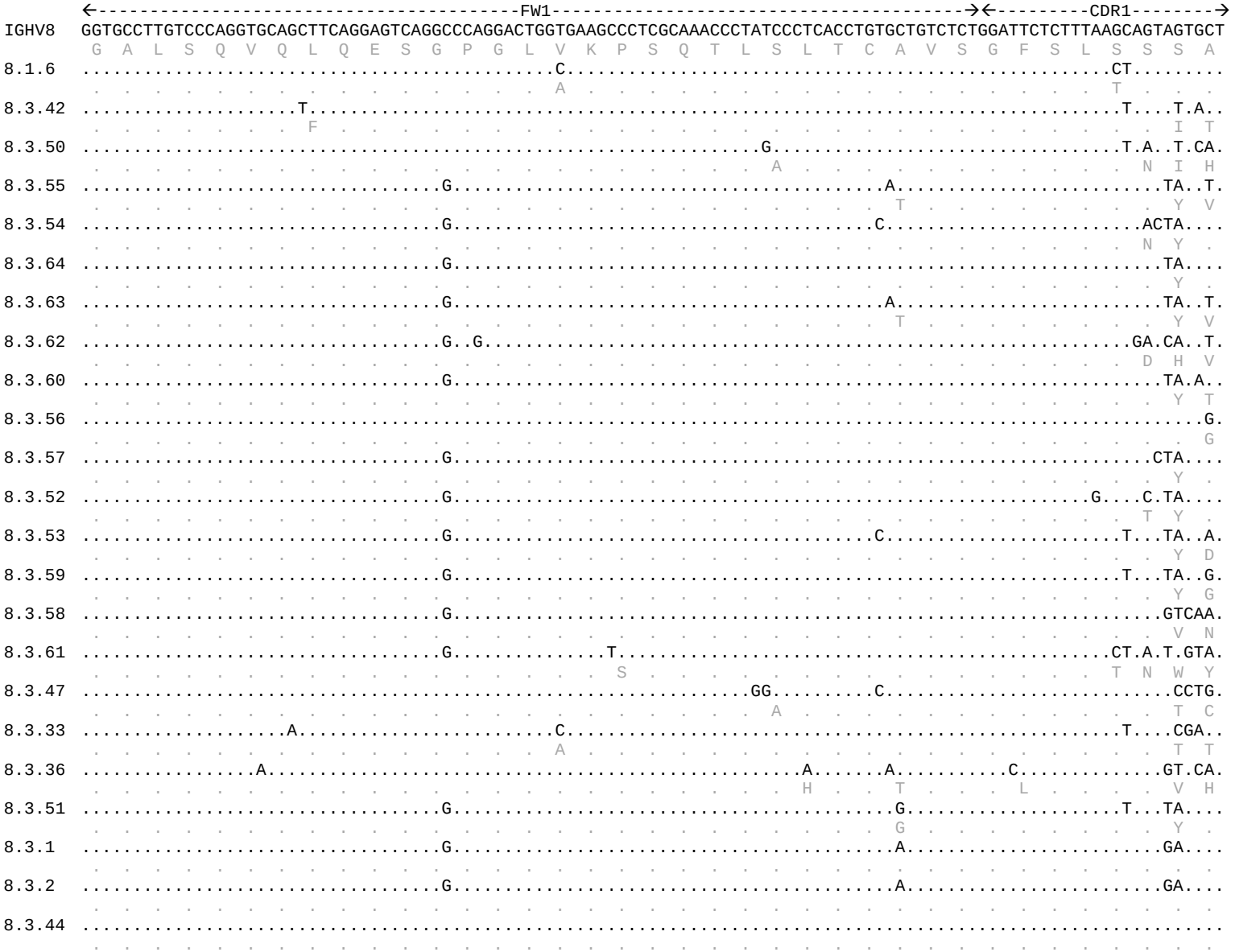
```
IGHD33   A....T...  .T.A.GC.C.AA  TGT....  ...CG.CA..G--------GTTACTAC--------------  ...T...  G.TG.CTTC.A.  ACA..TGC.
IGHD15   AA...T...  AT.A.GC.C.AA  TG..A..  ...G...A..G--------GTTGCTAC--------------  ...T..A  G.CG.CTTC.C.  GCA..T.C.
Family 7
IGHD21   A.T.CT...  C.CA.AGCA.CA  AT.T...  GT.GAT.ATG.-------AA..CTAGGTG..----------  .......  G..CCA..CA..  GCAGC..T.
IGHD34   A.T.CT...  C.CA.AGCA.CA  AT.A...  GT.GAT.AC..--T.ACTAG.T.ATGGTA..----------  ......T  G..C.AA.TA..  GCAGC..AA
Family 8
IGHD26   A.TA.T...  .CA..TCCA.AA  ...A...  GTT..GGATGG--------G..------------------  .....C.  ...C.CGTCTC.  ..T....CT
Family 9
IGHD10   .....AG.G  .CA.AGG..CAC  .T.T...  GT.T..CAC.TA..C.A..GGT..G..TG..CACAGCCTCA  ..AT.CC  TTGC.AA.C.AT  .C.T.C.C.
Family 10
IGHD27   A....A.TC  A.A.AGCCC.C.  ..TG...  GGGT..-----------------C.CC--------------  ..T....  G.TC..G.CAC.  AC.......
```

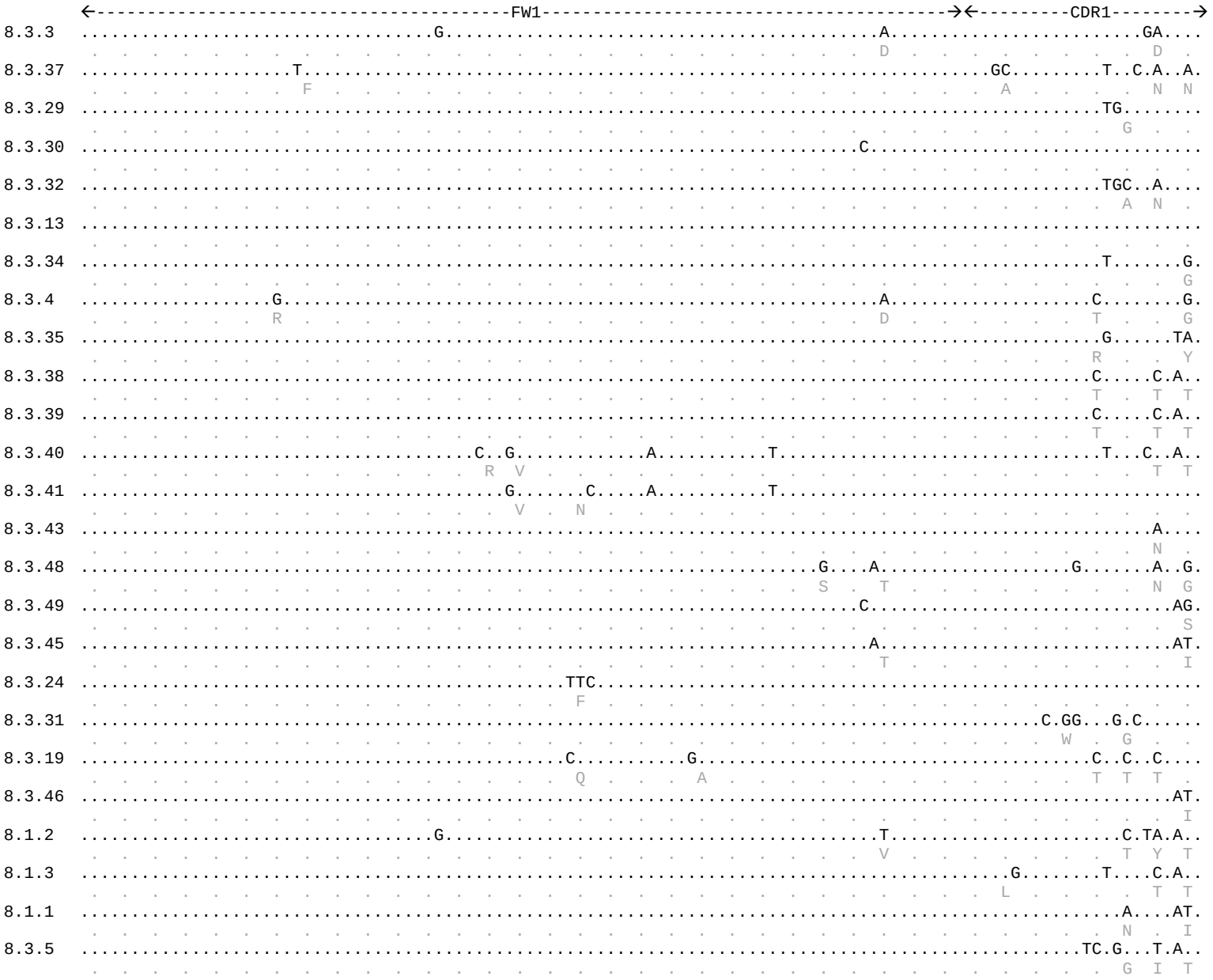
Supplemental Figure 6. DH Families. Nucleotide alignment of the 48 *T. m. latirostris* genomic D segments into 10 families. Families were determined based on 70% nucleotide identity and ordered from largest to smallest. Dots represent similarity to the top D segment sequence, IGHD13, and dashes represent gaps introduced to make the alignment. The RSSs are included and separated from the coding region on both sides.

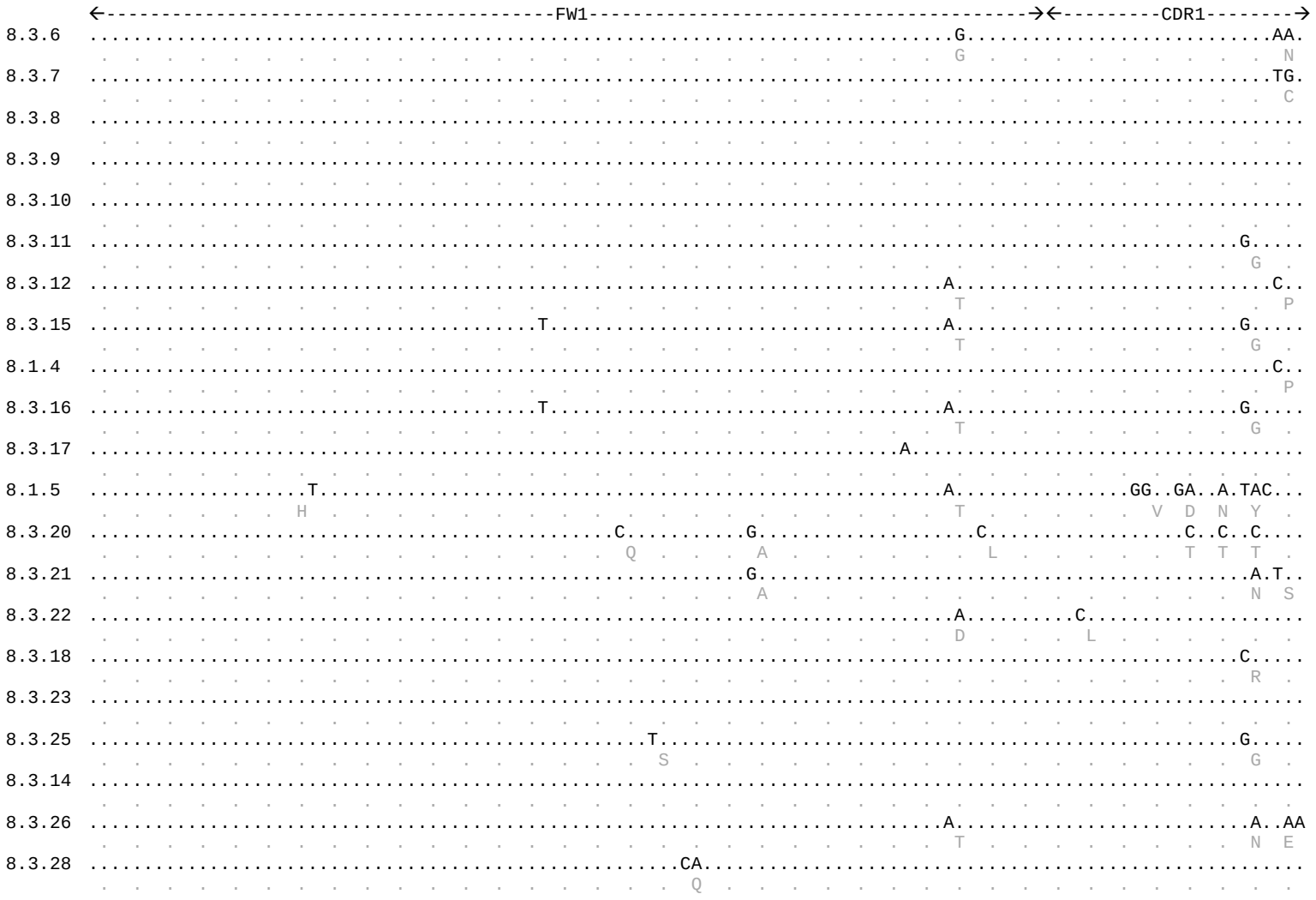


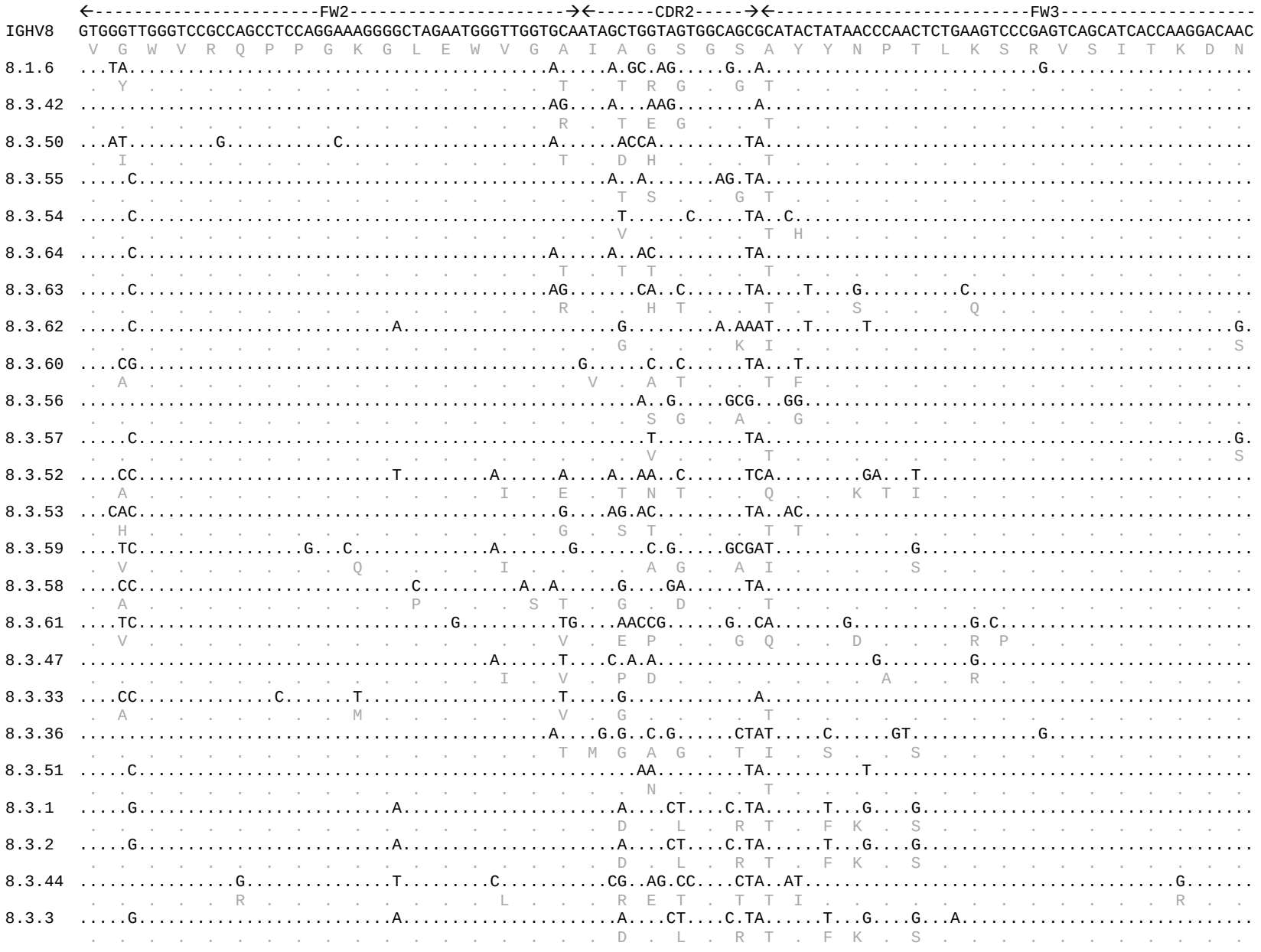
Supplemental Figure 7. JH Alignment. Nucleotide alignment of the six *T. m. latirostris* genomic J segments. Dots represent similarity to the top J segment sequence, IGHJ1, and dashes represent gaps introduced to make the alignment. The 5' RSS is included and separated from the coding sequence. The amino acid translation is included below the coding regions in grey.

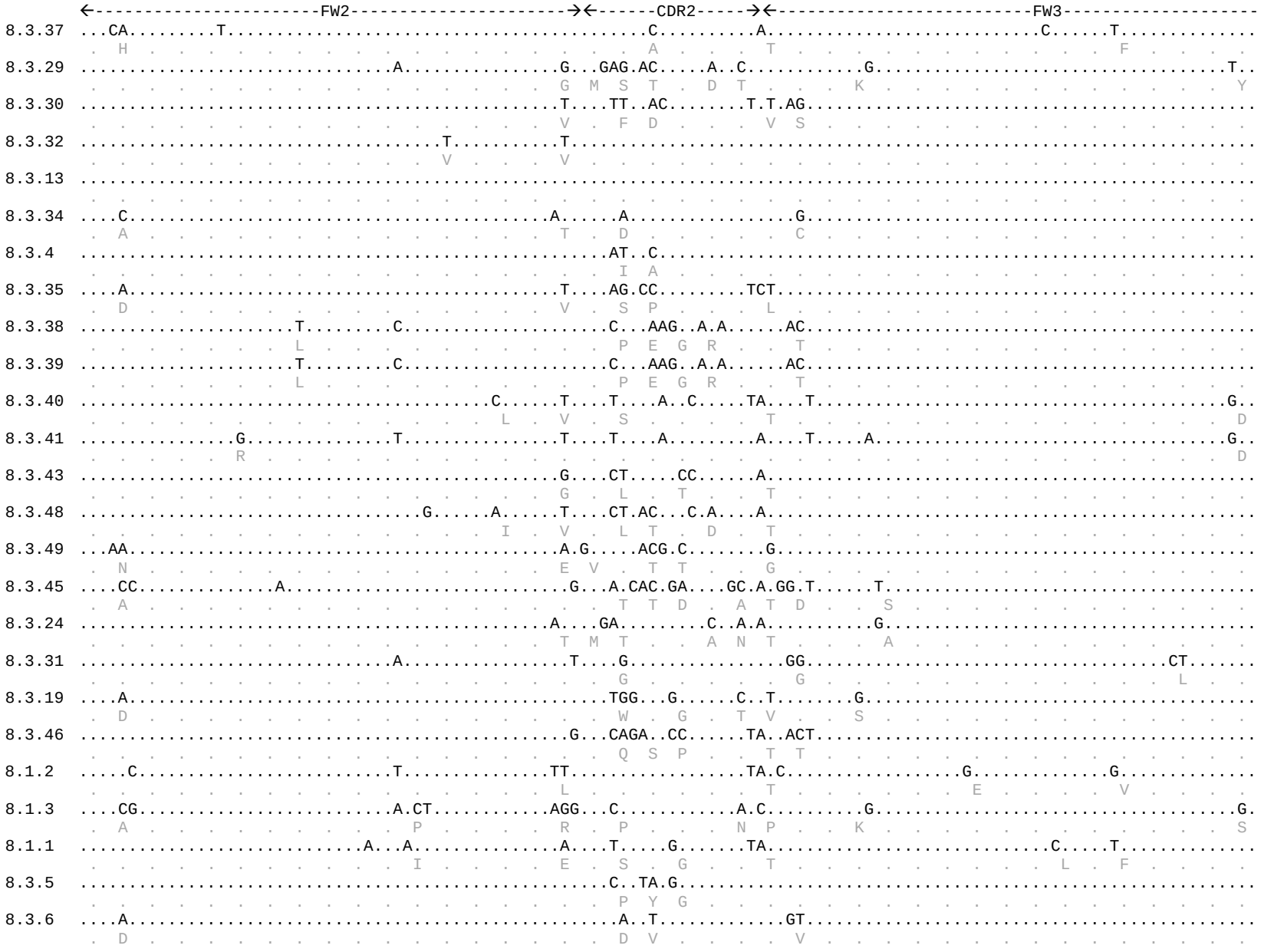
IGHV8 Sequences

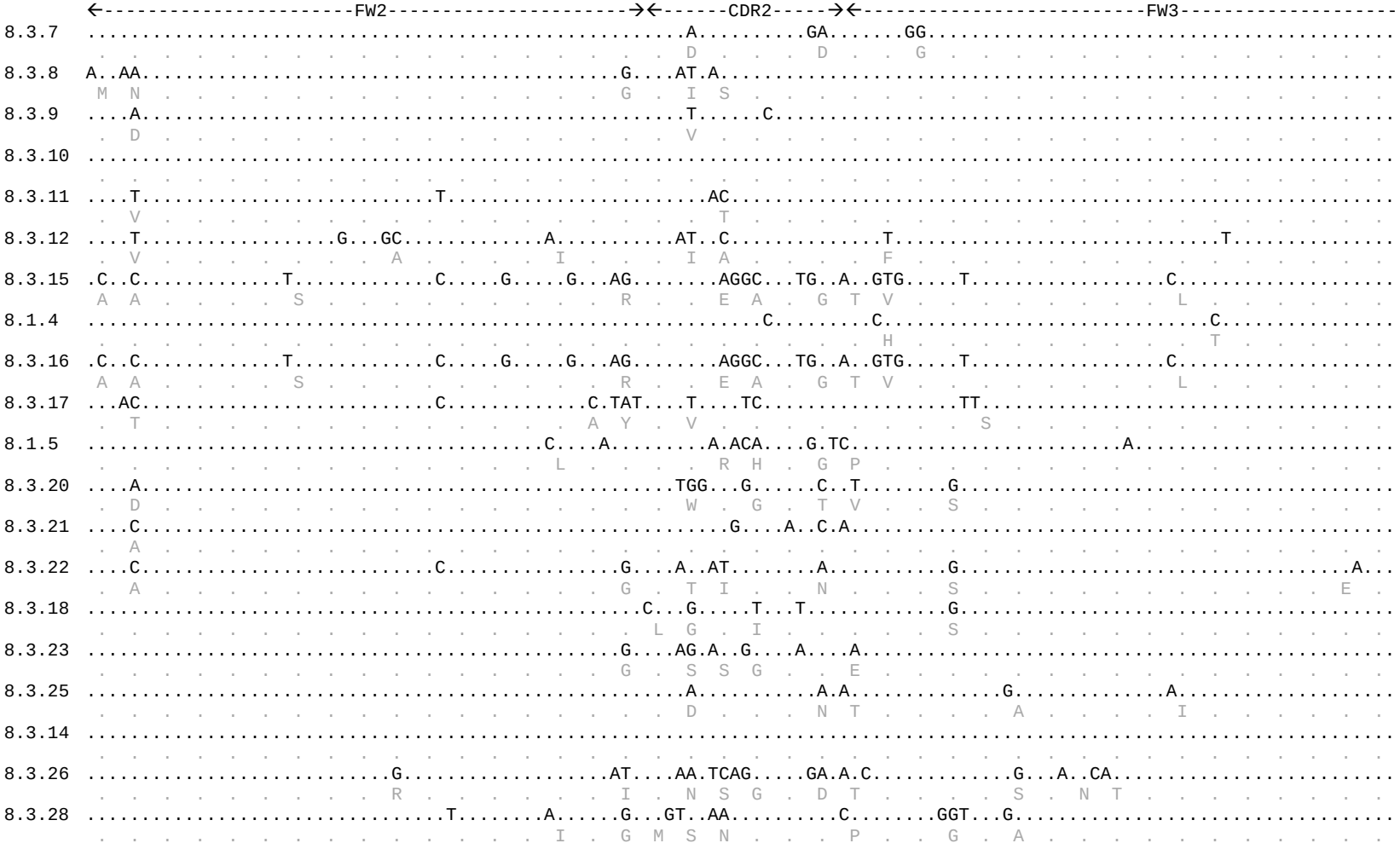



FW1
CDR1
8.3.3
8.3.37
8.3.29
8.3.30
8.3.32
8.3.13
8.3.34
8.3.4
8.3.35
8.3.38
8.3.39
8.3.40
8.3.41
8.3.43
8.3.48
8.3.49
8.3.45
8.3.24
8.3.31
8.3.19
8.3.46
8.1.2
8.1.3
8.1.1
8.3.5

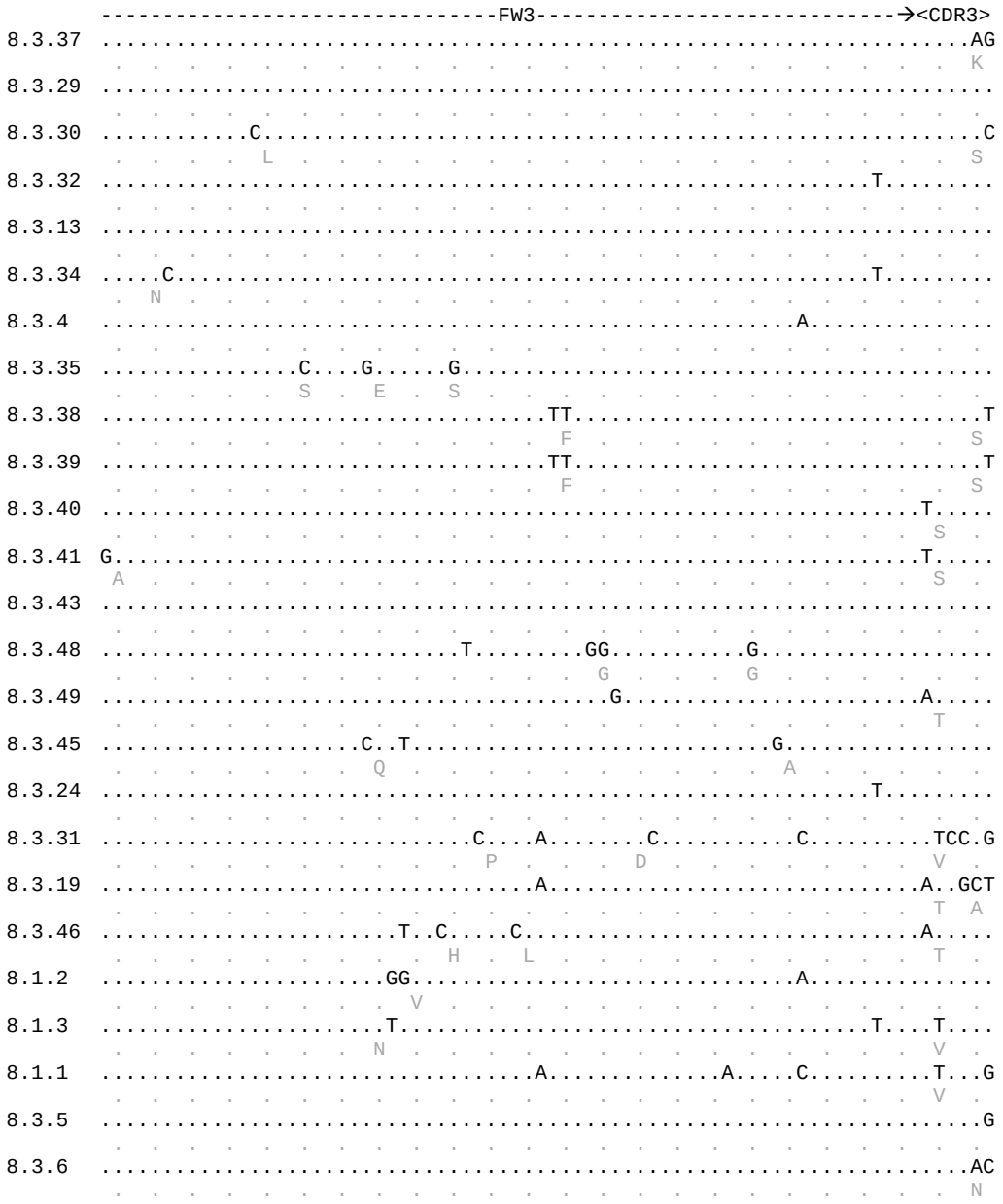

FW1
CDR1
8.3.6
8.3.7
8.3.8
8.3.9
8.3.10
8.3.11
8.3.12
8.3.15
8.1.4
8.3.16
8.3.17
8.1.5
8.3.20
8.3.21
8.3.22
8.3.18
8.3.23
8.3.25
8.3.14
8.3.26
8.3.28

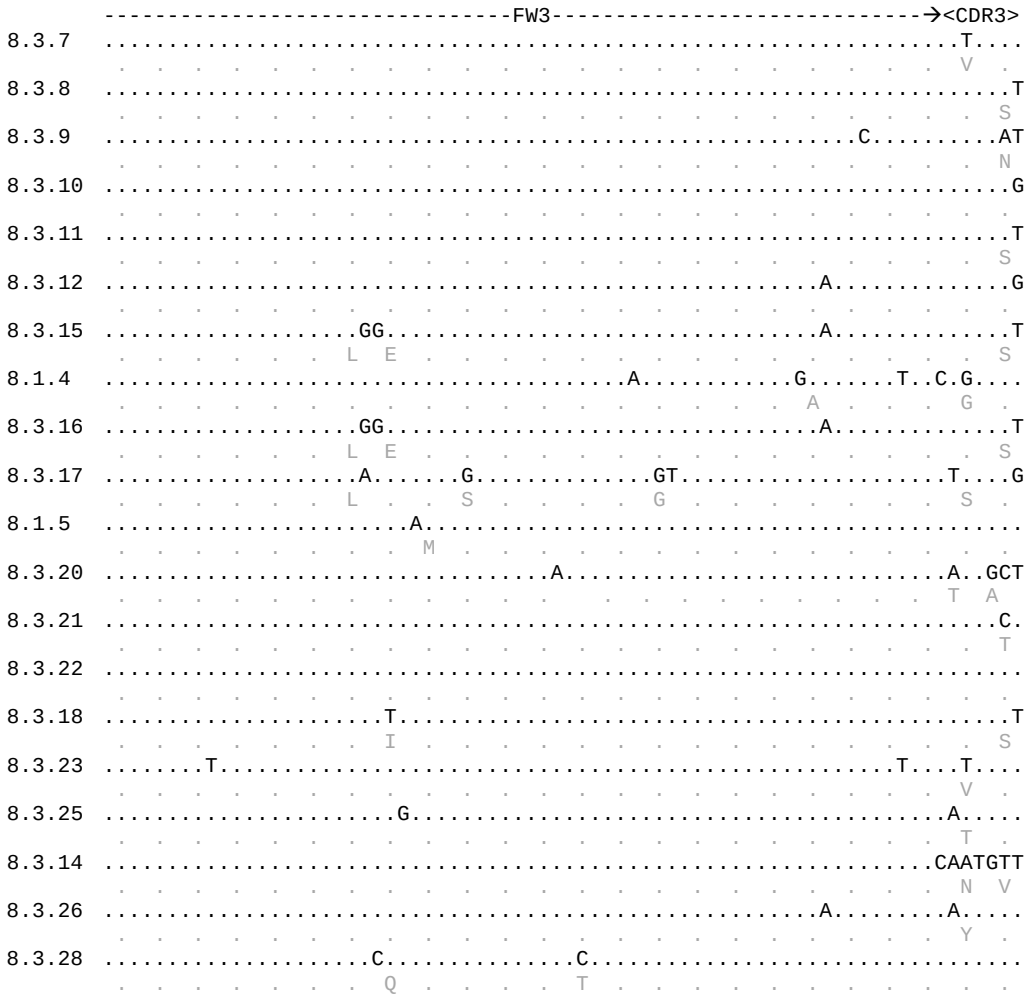

←----------FW2----------→←------CDR2-----→←----------FW3----------
IGHV8 GTGGGTTGGGTCCGCCAGCCTCCAGGAAAGGGGCTAGAATGGGTTGGTGCAATAGCTGGTAGTGGCAGCGCATACTATAACCCAACTCTGAAGTCCCGAGTCAGCATCACCAAGGACAAC
V G W V R Q P P G K G L E W V G A I A G S G S A Y Y N P T L K S R V S I T K D N
8.1.6 ...TA.....A.....A.GC.AG.....G..A.....G.....
. Y . . . . . T . T R G . G T . . . . . .
8.3.42 .....AG....A...AAG.....A.....
. . . . . R . T E G . . T . . . . . .
8.3.50 ...AT.....G.....C.....A.....ACCA.....TA.....
. I . . . . . T . D H . . . T . . . . . .
8.3.55 .....C.....A..A.....AG.TA.....
. . . . . T S . . G T . . . . . .
8.3.54 .....C.....T.....C.....TA..C.....
. . . . . V . . . . T H . . . . . .
8.3.64 .....C.....A.....A..AC.....TA.....
. . . . . T . T T . . T . . . . . .
8.3.63 .....C.....AG.....CA..C.....TA....T....G.....C.....
. . . . . R . . H T . . T . . S . . . Q . . . . . .
8.3.62 .....C.....A.....G.....A.AAAT...T.....T.....G.
. . . . . G . . . K I . . . . . S
8.3.60 ....CG.....G.....C..C.....TA...T.....
. A . . . . . V . A T . . T F . . . . . .
8.3.56 .....A..G.....GCG...GG.....
. . . . . S G . A . G . . . . . .
8.3.57 .....C.....T.....TA.....G.
. . . . . V . . . T . . . . . S
8.3.52 ....CC.....T.....A.....A....A..AA..C.....TCA.....GA...T.....
. A . . . . . I . E . T N T . . Q . . K T I . . . . . .
8.3.53 ...CAC.....G....AG.AC.....TA..AC.....
. H . . . . . G . S T . . . T T . . . . . .
8.3.59 ....TC.....G...C.....A.....G.....C.G.....GCGAT.....G.....
. V . . . . . Q . . . I . . . . A G . A I . . . . . S . . . . . .
8.3.58 ....CC.....C.....A..A.....G....GA.....TA.....
. A . . . . . P . . . S T . G . D . . T . . . . . .
8.3.61 ....TC.....G.....TG....AACCG.....G..CA.....G.....G.C.....
. V . . . . . V . E P . . G Q . . D . . . R P . . . . . .
8.3.47 .....A.....T....C.A.A.....G.....G.....
. . . . . I . V . P D . . . . . A . . R . . . . . .
8.3.33 ....CC.....C.....T.....T.....G.....A.....
. A . . . . . M . . . . . V . G . . . T . . . . . .
8.3.36 .....A....G.G..C.G.....CTAT.....C.....GT.....G.....
. . . . . T M G A G . T I . S . . S . . . . . .
8.3.51 .....C.....AA.....TA.....T.....
. . . . . N . . . T . . . . . .
8.3.1 .....G.....A.....A....CT....C.TA.....T...G....G.....
. . . . . D . L . R T . F K . S . . . . . .
8.3.2 .....G.....A.....A....CT....C.TA.....T...G....G.....
. . . . . D . L . R T . F K . S . . . . . .
8.3.44 .....G.....T.....C.....CG..AG.CC....CTA..AT.....G.....
. . . . . R . . . . . L . . . R E T . T T I . . . . . R . .
8.3.3 .....G.....A.....A....CT....C.TA.....T...G....G...A.....
. . . . . D . L . R T . F K . S . . . . . .


←----------FW2----------→←---CDR2---→←----------FW3----------
8.3.37 ...CA.....T.....C.....A.....C.....T.....
8.3.29 .....A.....G...GAG.AC.....A..C.....G.....T..
8.3.30 .....T....TT..AC.....T.T.AG.....
8.3.32 .....T.....T.....
8.3.13 .....
8.3.34 ....C.....A.....A.....G.....
8.3.4 .....AT..C.....
8.3.35 ....A.....T....AG.CC.....TCT.....
8.3.38 .....T.....C.....C...AAG..A.A.....AC.....
8.3.39 .....T.....C.....C...AAG..A.A.....AC.....
8.3.40 .....C.....T....T...A..C.....TA...T.....G..
8.3.41 .....G.....T.....T....T...A.....A....T.....A.....G..
8.3.43 .....G....CT.....CC.....A.....
8.3.48 .....G.....A.....T....CT.AC...C.A....A.....
8.3.49 ...AA.....A.G.....ACG.C.....G.....
8.3.45 ....CC.....A.....G...A.CAC.GA....GC.A.GG.T.....T.....
8.3.24 .....A....GA.....C..A.A.....G.....
8.3.31 .....A.....T....G.....GG.....CT.....
8.3.19 ....A.....TGG...G.....C..T.....G.....
8.3.46 .....G...CAGA..CC.....TA..ACT.....
8.1.2 .....C.....T.....TT.....TA.C.....G.....G.....
8.1.3 ....CG.....A.CT.....AGG...C.....A.C.....G.....G.
8.1.1 .....A...A.....A....T.....G.....TA.....C.....T.....
8.3.5 .....C..TA.G.....
8.3.6 ....A.....A..T.....GT.....

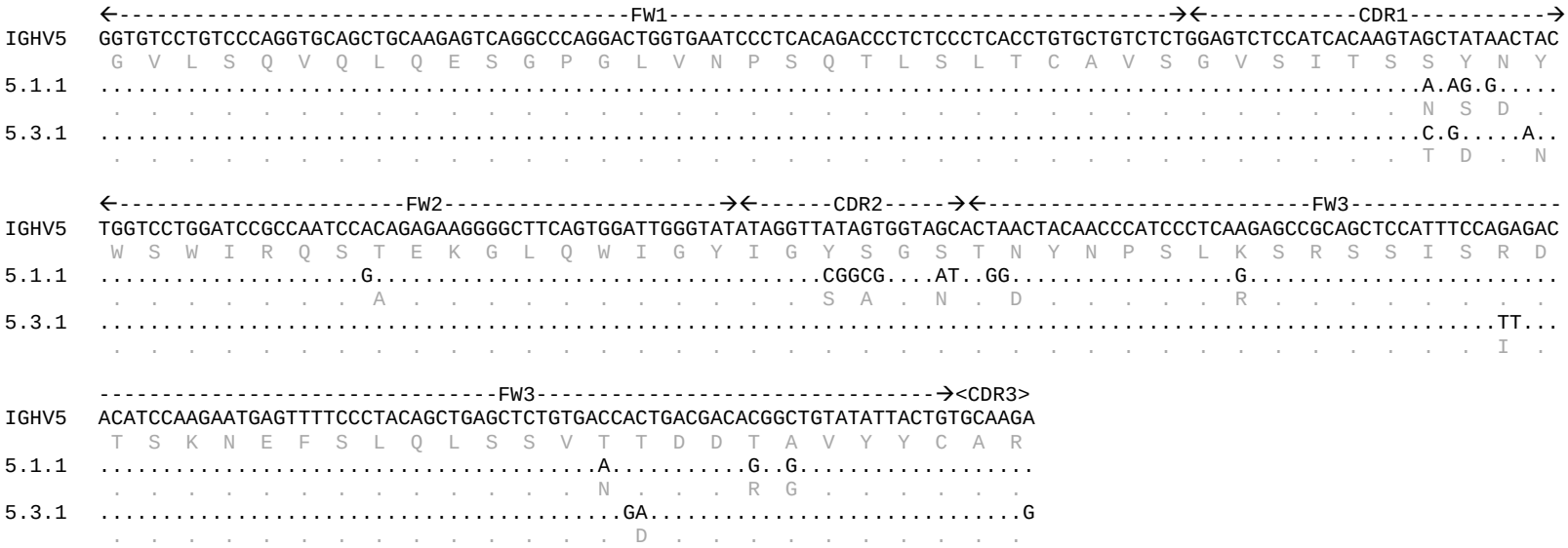

←FW2→←CDR2→←FW3→
8.3.7
8.3.8
8.3.9
8.3.10
8.3.11
8.3.12
8.3.15
8.1.4
8.3.16
8.3.17
8.1.5
8.3.20
8.3.21
8.3.22
8.3.18
8.3.23
8.3.25
8.3.14
8.3.26
8.3.28

```
        --------------------------------FW3----------------------------->
IGHV8   TCGAAGAGCCAAGTTTATTTCAAACTGAACTCTGTGAATTCTGAGGACACGGCCACGTATTACTGTGCAAGA
         S  K  S  Q  V  Y  F  K  L  N  S  V  N  S  E  D  T  A  T  Y  Y  C  A  R
8.1.6   ...G.......G.........................G..................A.........A....G
         .  E  .  .  .  .  .  .  .  .  .  .  S  .  .  .  .  .  .  .  .  .  T  R
8.3.42  ...........................................................A....G
         .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  T  R
8.3.50  ....G..........................C..............................A...AC
         .  R  .  .  .  .  .  .  .  .  .  L  .  .  .  .  .  .  .  .  .  T  N
8.3.55  ...........................................................
         .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
8.3.54  ............A..............................................
         .  .  .  .  .  I  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
8.3.64  ........T.............CT.....................................T
         .  .  .  .  .  .  .  .  N  .  .  .  .  .  .  .  .  .  .  .  .  .  S
8.3.63  ...........................................................G....
         .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  G  .
8.3.62  ...G............C.....................................A.T...
         .  E  .  .  .  S  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  T  .
8.3.60  ........................G..............................GG............
         .  .  .  .  .  .  .  .  .  D  .  .  .  .  .  .  .  .  .  W  .  .  .
8.3.56  ............A..............................C...........A.........G....
         .  .  .  .  .  I  .  .  .  .  .  .  .  .  D  .  .  .  .  .  .  G  .
8.3.57  ................C.....................................................TCT
         .  .  .  .  .  .  S  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  S
8.3.52  .........................................C.....................ATG.TG
         .  .  .  .  .  .  .  .  .  .  .  .  .  .  D  .  .  .  .  .  .  M  D
8.3.53  ............C...........................................................A....C
         .  .  .  .  .  L  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  T  S
8.3.59  ..............................A.............................T....T..CC
         .  .  .  .  .  .  .  .  .  .  .  .  M  .  .  .  .  .  .  .  .  .  V  T
8.3.58  ...........................................................A.........A...CT
         .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  T  T
8.3.61  ..C.....................................................................CT
         .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  T
8.3.47  ...........................................................C.............T.
         .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  I
8.3.33  ....................GG.....................................T........T
         .  .  .  .  .  .  .  G  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  S
8.3.36  ........................G.....................................CCTC
         .  .  .  .  .  .  .  .  .  S  .  .  .  .  .  .  .  .  .  .  .  .  L
8.3.51  .....................................................................A.
         .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  K
8.3.1   ......................G............G.................CG......T.......AG
         .  .  .  .  .  .  .  .  .  .  .  .  D  .  .  .  .  .  R  .  .  .  K
8.3.2   ......................G............G.................CG......T.......AG
         .  .  .  .  .  .  .  .  .  .  .  .  D  .  .  .  .  .  R  .  .  .  K
8.3.44  .......A............G.........................G.....................T
         .  .  N  .  .  .  .  E  .  .  .  .  .  .  .  .  A  .  .  .  .  .  S
8.3.3   ......................G............G.................CG......T.......AG
         .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  K
```

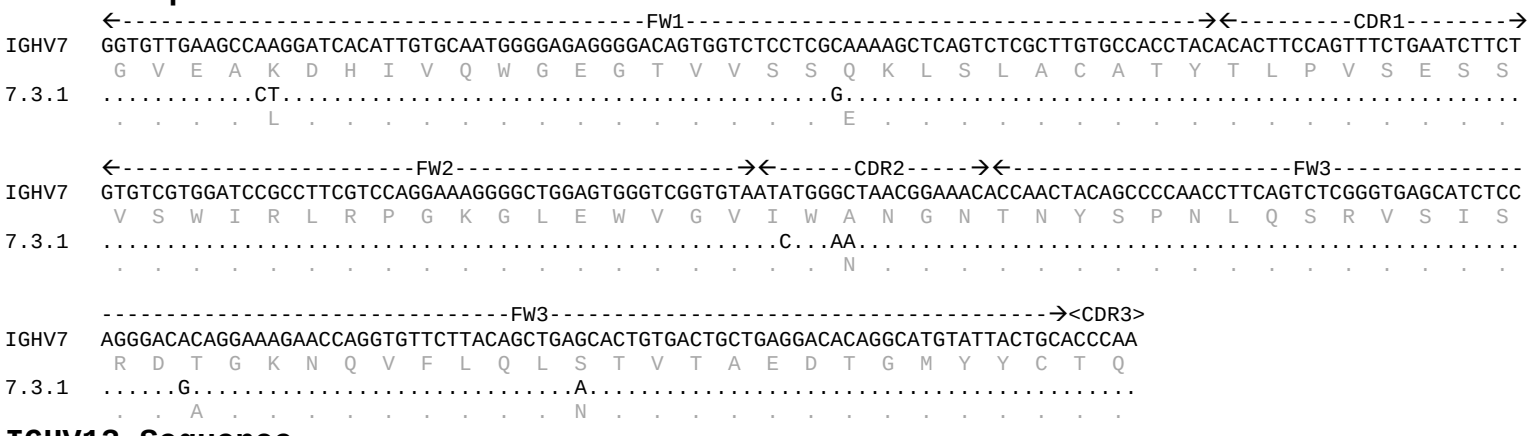

FW3
<CDR3>
8.3.37
8.3.29
8.3.30
8.3.32
8.3.13
8.3.34
8.3.4
8.3.35
8.3.38
8.3.39
8.3.40
8.3.41
8.3.43
8.3.48
8.3.49
8.3.45
8.3.24
8.3.31
8.3.19
8.3.46
8.1.2
8.1.3
8.1.1
8.3.5
8.3.6


FW3
<CDR3>
8.3.7
8.3.8
8.3.9
8.3.10
8.3.11
8.3.12
8.3.15
8.1.4
8.3.16
8.3.17
8.1.5
8.3.20
8.3.21
8.3.22
8.3.18
8.3.23
8.3.25
8.3.14
8.3.26
8.3.28

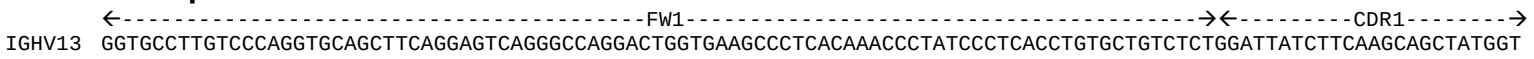
IGHV5 Sequences

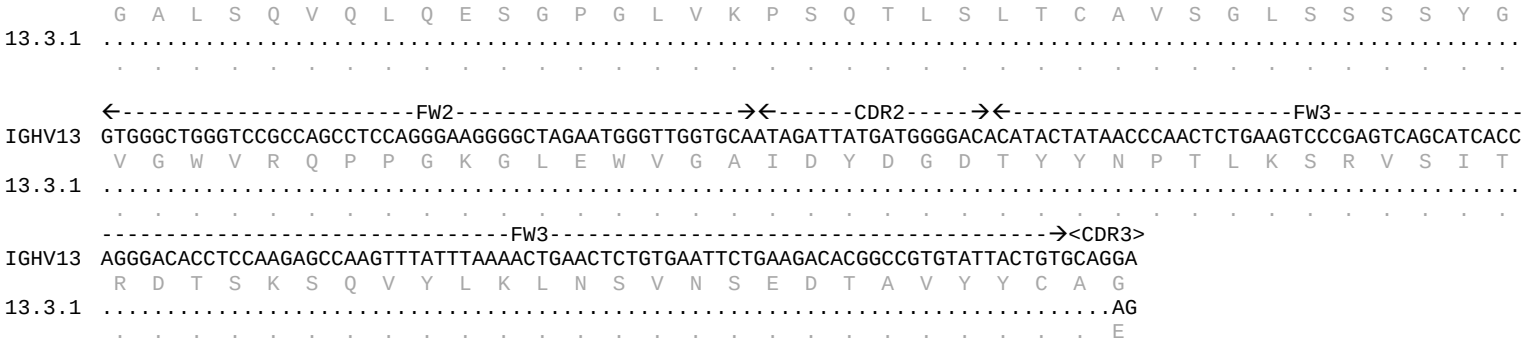


IGHV7 Sequence



IGHV13 Sequence





Supplemental Figure 8. All Expressed IgH Transcripts by V Segment. Nucleotide alignment of the V segment sequence of *T. m. latirostris* expressed transcripts to the genomic V segment they originated from (based on at least 90% identity). Framework (FW) and Complementarity Determining Regions (CDR) are indicated at the top of the alignments. Dots represent similarity to the genomic sequence for each V segment section. The amino acid translation is included below each sequence in grey. Sequence ID numbers are organized as: genomic V segment. genomic J segment. transcript number.

Table S1. Primers and PCR Conditions

	Forward Primer	Reverse Primer	Polymerase	Initial	Denature	Annealing	Elongation	# Cycles	Final
Sanger 5' RACE (IgM)	GGACACTGACATGGACTGAAGGAGTA	GATGGAGCCAGGCAGAAAGT	Phusion	98°C for 0:30	98°C for 0:10	65°C for 0:30	72°C for 0:30	35	72°C for 5:00
Sanger 3' RACE (IGHV7)	CCAGGAAAGGGGCTGGAGTG	CGTTACGTAGCGTATCGTTGACAGC	Hot Start	95°C for 15:00	95°C for 0:30	60°C for 0:30	72°C for 1:30	35	72°C for 5:00
Sanger 3’ RACE (IGHV8)	GGTGCCTTGTCCCAGGTGCA	CGTTACGTAGCGTATCGTTGACAGC	Hot Start	95°C for 15:00	95°C for 0:30	60°C for 0:30	72°C for 1:30	35	72°C for 5:00
Sanger 3’ RACE (IGHV11)	GTCCAGTTAGTGCAGTCTGG	CGTTACGTAGCGTATCGTTGACAGC	Hot Start	95°C for 15:00	95°C for 0:30	56°C for 0:30	72°C for 1:30	35	72°C for 5:00
PacBio 5' RACE (IgM)	TCAGACGATGCGTCATGGACACTGACATGGACTGAAGGAGTA	CACGATAGTCGCTATGGATGGAGCCAGGCAGAAAGT	Phusion	98°C for 0:30	98°C for 0:10	72°C for 0:30	72°C for 0:30	35	72°C for 5:00

Table S2. Pseudogene Determination.

Locus	FS/Stop	Splice site	No RSS	Missing Motif	Truncation/ Insertion
IGHV1p			X	X	
IGHV2p	X	X			
IGHV3p	X				
IGHV4p	X				
IGHV5p	X		X		
IGHV6p	X				
IGHV7p	X			X	
IGHV8p					X
IGHV9p	X				
IGHV10p	X				
IGHV11p	X				
IGHV12p	X				
IGHV13p	X		X	X	
IGHV14p	X				
IGHV15p	X	X			
IGHV16p	X		X		
IGHV17p	X		X	X	
IGHV18p	X	X		X	
IGHV19p	X		X		
IGHV20p	X		X		
IGHV21p	X				
IGHV22p	X		X		
IGHV23p	X		X		X
IGHV24p	X			X	
IGHV25p		X	X	X	X
IGHV26p	X				
IGHV27p	X				
IGHV28p	X			X	
IGHV29p	X		X		
IGHV30p	X			X	
IGHV31p	X			X	
IGHV32p	X			X	
IGHV33p	X				
IGHV34p			X	X	
IGHV35p	X				
IGHV36p	X			X	
IGHV37p	X	X	X		
IGHV38p	X	X	X		
IGHV39p	X				
IGHV40p	X	X	X	X	X
IGHV41p	X		X	X	
IGHV42p	X		X		
IGHV43p	X			X	
IGHV44p	X			X	
IGHV45p	X			X	
IGHV46p	X		X	X	
IGHV47p	X				
IGHV48p	X		X	X	
IGHV49p	X				
IGHV50p	X				
IGHV51p	X			X	
IGHV52p	X		X		
IGHV53p	X		X	X	
IGHV54p	X		X	X	
IGHV55p	X			X	
IGHV56p	X				
IGHV57p	X				
IGHV58p				X	
IGHV59p	X		X		
IGHV60p	X				
IGHV61p	X		X		
IGHV62p	X		X	X	
IGHV63p	X	X		X	

Locus	FS/Stop	Splice site	No RSS	Missing Motif	Truncation/ Insertion
IGHV64p	X		X	X	X
IGHV65p	X				
IGHV66p	X		X		
IGHV67p	X			X	
IGHV68p	X			X	
IGHV69p				X	
IGHV70p	X		X	X	
IGHV71p		X	X	X	X
IGHV72p	X		X	X	
IGHV73p	X			X	X
IGHV74p	X	X	X	X	
IGHV75p	X			X	
IGHV76p	X			X	
IGHV77p	X			X	
IGHV78p	X		X		
IGHV79p	X				
IGHV80p	X				
IGHV81p	X				
IGHV82p	X		X		
IGHV83p	X				
IGHV84p	X		X	X	
IGHV85p	X			X	
IGHV86p	X		X		
IGHV87p			X	X	
IGHV88p	X	X	X	X	
IGHV89p	X		X	X	
IGHV90p	X		X		
IGHV91p	X			X	
IGHV92p	X		X	X	
IGHV93p	X				
IGHV94p	X	X	X	X	
IGHV95p	X		X	X	
IGHV96p	X		X	X	
IGHV97p	X			X	
IGHV98p	X	X		X	
IGHV99p	X				
IGHV100p	X			X	
IGHV101p	X				
IGHV102p	X				
IGHV103p	X				
IGHV104p	X		X	X	
IGHV105p	X			X	
IGHV106p	X	X		X	
IGHV107p	X				
IGHV108p	X		X		
IGHV109p	X			X	
IGHV110p	X	X		X	
IGHV111p	X		X	X	
IGHV112p	X		X		
IGHV113p	X	X		X	
IGHV114p	X		X	X	
IGHV115p	X	X			
IGHV116p	X				
IGHV117p	X			X	
IGHV118p	X	X	X	X	
IGHV119p	X				
IGHV120p	X		X	X	
IGHV121p	X				
IGHV122p	X			X	
IGHV123p	X		X		
IGHV124p	X			X	
IGHV125p	X				
IGHV126p	X				

Supplemental Table 3: V Segment Genomic Locations

Gene	Scaffold	Orientation	Start	Stop
IgHV1	JH594659	-	14,507,938	14,507,630
IgHV2	JH594659	-	14,565,935	14,565,630
IgHV3	JH594659	-	14,909,083	14,908,778
IgHV4	JH594659	-	15,024,261	15,023,953
IgHV5	JH594659	-	15,027,007	15,026,699
IgHV6	JH594659	-	15,166,421	15,166,113
IgHV7	JH594659	-	15,689,591	15,689,289
IgHV8	JH594659	-	16,122,497	16,122,195
IgHV9	JH594659	-	16,227,019	16,226,789
IgHV10	JH594659	-	16,296,569	16,296,261
IgHV11	JH594659	-	16,401,914	16,401,609
IgHV12	JH594659	-	16,415,891	16,415,586
IgHV13	JH594659	-	17,140,251	17,139,949

Table S4. Accession Numbers

Figure	Species	Gene	Accession #
2	*Bos taurus*	IgM	AAC98391
	Bos taurus	IgD	AF411240
	Bos taurus	IgG1	AAB37381
	Bos taurus	IgE	AAB09546
	Bos taurus	IgA	AAC98391
	Canis lupus familiaris	IgM	ABN11172-AB11175
	Canis lupus familiaris	IgD	DQ297185
	Canis lupus familiaris	IgGB	AAL35302
	Canis lupus familiaris	IgE	AAA56797
	Canis lupus familiaris	IgA	AAA56796
	Ginglymostoma cirratum	IgM	ABW84256.1
	Homo sapiens	IgM	P01871.3
	Homo sapiens	IgD	P01880.2
	Homo sapiens	IgG1	CAC20454.1
	Homo sapiens	IgG2	CAC20455.1
	Homo sapiens	IgG3	CAC20456.1
	Homo sapiens	IgG4	CAC20457.1
	Homo sapiens	IgE	AAB59424
	Homo sapiens	IgA1	AAT74070.1
	Homo sapiens	IgA2	A2HU
	Loxodonta africana	IgM	XP_010596164
	Loxodonta africana	IgD	NA
	Loxodonta africana	IgG4	NA
	Monodelphis domestica	IgM	AAD24482.1
	Monodelphis domestica	IgG	AAC79675
	Monodelphis domestica	IgE	AAC79674.1
	Monodelphis domestica	IgA	AAC48835
	Mus musculus domesticus	IgM	X03690
	Mus musculus domesticus	IgD	AAB59654.1
	Mus musculus domesticus	IgG	AAB59658.1
	Mus musculus domesticus	IgE	BAQ55489.1
	Mus musculus domesticus	IgA	AAB59662.1
	Ornithorhynchus anatinus	IgM	AAO37747
	Ornithorhynchus anatinus	IgG	AAL17703.1
	Ornithorhynchus anatinus	IgE	AAL17702.1
	Ornithorhynchus anatinus	IgA	AAL17700.1
	Salmo trutta	IgM	AAF69489
	Tachyglossus aculeatus	IgM	AAN33013
	Tachyglossus aculeatus	IgG	AAM61760
	Tachyglossus aculeatus	IgE	AAM45140
	Tachyglossus aculeatus	IgA	AAN33012
	Tursiops truncatus	IgM	AF306861
	Tursiops truncatus	IgG1	AAT65196
	Tursiops truncatus	IgA	AAT65195
	Xenopus laevis	IgM	C31933
4	*Heterodontus francisci*	V(H)F101	X13449
	Homo sapiens	V6-1	Z14223
	Homo sapiens	VH1-69	X92298
	Homo sapiens	VH2-5	X62111
	Homo sapiens	VH3-72	X92206
	Homo sapiens	VH4-b	X92289
	Homo sapiens	VH5-51	M99686
	Homo sapiens	VH7-81	AB019437
	Mus musculus domesticus	VH1S136	AF304557
	Mus musculus domesticus	VH2S1	AJ851868
	Mus musculus domesticus	VH3S1	K01569
	Mus musculus domesticus	VH4S1	AC079273
	Mus musculus domesticus	VH5S16	AJ851868
	Mus musculus domesticus	VH6S1	X03398
	Mus musculus domesticus	VH7S1	AJ851868
	Mus musculus domesticus	VH8S4	AC073939
	Mus musculus domesticus	VH9S8	L14368
	Mus musculus domesticus	VH10S4	M21470
	Mus musculus domesticus	VH11S1	AJ851868
	Mus musculus domesticus	VH14S4	X55934
S1	*Ailuropoda melanoleuca*	IgM	AAX73309.1
	Balaenoptera acutorostrata scammoni	IgM	XP_007176340
	Bos taurus	IgM	AAB09544
	Canis lupus familiaris	IgM	ABN11172-AB11175
	Elephas maximus	IgM	AIA77883
	Ginglymostoma cirratum	IgM	ABW84256.1
	Homo sapiens	IgM	X67301
	Loxodonta africana	IgM	XM_010597862.1
	Macaca mulatta	IgM	AF046784
	Meriones unguiculatus	IgM	BAL70398.1
	Monodelphis domestica	IgM	AAD24482.1
	Mus musculus domesticus	IgM	X03690
	Ornithorhynchus anatinus	IgM	AAO37747
	Oryctolagus cuniculus	IgM	AAA64251
	Sus scrofa	IgM	BAM66301.1
	Triakis scyllium	IgM	AB557736
	Tursiops truncatus	IgM	AF306861
	Xenopus laevis	IgM	C31933
S2	*Bos taurus*	IgD	AAN07165
	Equus caballus	IgD	AAU09793.1
	Homo sapiens	IgD	P01880.2
	Loxodonta africana	IgD	NA
	Mus musculus domesticus	IgD	AAB59652.1
	Ovis aries	IgD	AAN07166
	Sus scrofa	IgD	AAN03672
S3	*Ailuropoda melanoleuca*	IgG1	AAX73307.1
	Bos taurus	IgG1	AAB37381.2
	Canis lupus familiaris	IgGB	AAL35302
	Felis catus	IgG1	BAA32230.1
	Homo sapiens	IgG1	CAC20454.1
	Homo sapiens	IgG2	CAC20455.1
	Homo sapiens	IgG3	CAC20456.1
	Homo sapiens	IgG4	CAC20457.1
	Loxodonta africana	IgG4	NA
	Macaca mulatta	IgG1	ADX62854
	Meriones unguiculatus	IgG	BAK23457
	Monodelphis domestica	IgG	AAC79675
	Mus musculus domesticus	IgG1	AAB59658.1
	Ornithorhynchus anatinus	IgG1	AAL17703.1
	Tursiops truncatus	IgG1	AAT65196.1
S4	*Ailuropoda melanoleuca*	IgE	AAX73306
	Bos taurus	IgE	AAB09546
	Canis lupus familiaris	IgE	AAA56797
	Daubentonia madagascariensis	IgE	AEG78116
	Felis catus	IgE	AAF43901.1
	Homo sapiens	IgE	AAB59424
	Macaca mulatta	IgE	EHH28231
	Meriones unguiculatus	IgE	BAL70400.1
	Monodelphis domestica	IgE	AAC79674.1
	Mus musculus domesticus	IgE	BAQ55489.1
	Ornithorhynchus anatinus	IgE	AAL17702.1
	Ovis aries	IgE	AAA51378
	Sus scrofa	IgE	BAM66316.1
S5	*Ailuropoda melanoleuca*	IgA	AAX73304.1
	Bos taurus	IgA	AAC98391
	Canis lupus familiaris	IgA	AAA56796
	Homo sapiens	IgA1	AAT74070.1
	Homo sapiens	IgA2	A2HU
	Macaca mulatta	IgA	CAA37742.1
	Meriones unguiculatus	IgA	BAL70401.1
	Monodelphis domestica	IgA	AAC48835
	Mus musculus domesticus	IgA	AAB59662.1
	Ornithorhynchus anatinus	IgA	AAL17700.1
	Oryctolagus cuniculus	IgA	S09271
	Sus scrofa	IgA	ADD51207.1
	Tursiops truncatus	IgA	AAT65195