Developmental and Comparative Immunology 85 (2018) 71–85

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

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

The Florida manatee (*Trichechus manatus latirostris*) T cell receptor loci exhibit V subgroup synteny and chain-specific evolution

Breanna Breaux [a], Margaret E. Hunter [b], Maria Paula Cruz-Schneider [c], Leonardo Sena [d], Robert K. Bonde [b], 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] Sirenia Project, Wetland and Aquatic Research Center, U.S. Geological Survey, 7920 NW 71st Street, Gainesville, FL 32653, USA
[c] Laboratory of DNA Polymorphism, Federal University of Pará, Belém, Pará, Brazil
[d] Laboratory of Medical and Human Genetics, Federal University of Pará, Belém, Pará, Brazil
[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 7 March 2018
Received in revised form
6 April 2018
Accepted 6 April 2018
Available online 9 April 2018

Keywords:
Evolution
Manatee
T cell receptor
Synteny blocks
VHD

ABSTRACT

The Florida manatee (*Trichechus manatus latirostris*) has limited diversity in the immunoglobulin heavy chain. We therefore investigated the antigen receptor loci of the other arm of the adaptive immune system: the T cell receptor. Manatees are the first species from Afrotheria, a basal eutherian superorder, to have an in-depth characterization of all T cell receptor loci. By annotating the genome and expressed transcripts, we found that each chain has distinct features that correlates to their individual functions. The genomic organization also plays a role in modulating sequence conservation between species. There were extensive V subgroup synteny blocks in the TRA and TRB loci between *T. m. latirostris* and human. Increased genomic locus complexity correlated to increased locus synteny. We also identified evidence for a VHD pseudogene for the first time in a eutherian mammal. These findings emphasize the value of including species within this basal eutherian radiation in comparative studies.

1. Introduction

T cells are the central component of the adaptive immune system that is present in all studied jawed vertebrates. They are responsible for novel antigen recognition and the subsequent activation of various immune cells, such as B cells, macrophages, and neutrophils. They can also kill infected cells directly or suppress self-reactive cells within an individual. T cells can collectively distinguish between different pathogens through the highly diverse T cell receptor (TR).

Each TR chain is made of two immunoglobulin-like domains: one constant region and one variable region. The variable region is unique to each T cell and is determined by V(D)J recombination. The TR loci are comprised of several variable (V), diversity (D), and joining (J) segments followed by the constant region. In each individual lymphocyte's DNA, one V, D, and J segment recombine to encode the variable region. The three segments together form three complementarity determining regions (CDRs), which are distal loops that interact with the antigen: the V segment encodes CDR1 and CDR2, and the CDR3 spans the 3′ region of the V, the D, and the 5′ region of the J segment. This creates combinatorial diversity between the segments.

There are four TR chains that are conserved in all jawed vertebrates: α, β, γ, and δ. The four chains form two heterodimers: αβ receptors and γδ receptors. The β and δ chains serve as the heavy chains (with D segments) and the α and γ chains serve as the light chains (without D segments). Three sets of discrete genomic loci encode the four canonical TR genes. The TRD genes are nested within the TRA genes in one locus. There is a collective set of V segments that can rearrange solely with TRA or TRD genes or with both; the D and J segments are chain specific. The TRB genes and the TRG genes are each encoded in independent loci, however the TRG genes can be encoded in one or two genomic loci depending on the species (Vaccarelli et al., 2008). The TR chain diversity and

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

https://doi.org/10.1016/j.dci.2018.04.007
0145-305X/

genomic organization are more conserved across jawed vertebrates than the immunoglobulin (Ig) genes. However, the level of sequence conservation across mammals is less resolved in the TR genes (Parra et al., 2008; Su et al., 1999). The current literature does not include the basal radiation of mammalian evolution between Boreoeutheria (primates, rodents, carnivores, and ruminants) and Marsupialia. It is therefore important to characterize the TR of species in clades that are not currently represented to identify evolutionary patterns.

Manatees are herbivorous aquatic mammals that inhabit warm waters of varying salinity. Of the three species, the West Indian (*Trichechus manatus*) and African (*T. senegalensis*) manatee are found in fresh, brackish, and salt water, while the Amazonian manatee (*T. inunguis*) is strictly fresh water (Domning, 1982). All three species have a vulnerable conservation status due to a history of overhunting that persists illegally today and ongoing habitat destruction (O'Shea, 1988). The Florida manatee (*T. m. latirostris*) is a subspecies of the West Indian manatee that mainly inhabits the coasts of Florida and has low mtDNA and microsatellite diversity (Tucker et al., 2012; Vianna et al., 2006). However, *T. m. latirostris* is predicted to have a strong immune system. While the evidence for this claim are mostly anecdotal, *T. m. latirostris* deaths are rarely contributed to infectious diseases in the absence of red tide toxin or cold stress induced immunosuppression (Walsh et al. 2005, 2015). It is also the only species within the order Sirenia to have a genome assembly available, so we focused on this subspecies for characterization of their TR loci and expressed transcripts.

Manatees belong to the order Afrotheria along with aardvarks, tenrecs, elephant shrews, golden moles, and their two closest relatives: the elephant and rock hyrax (Kellogg et al., 2007). While the placental mammalian phylogeny is not fully resolved, Afrotheria is most often included in the most basal split, either individually or as a sister taxa to Xenarthra in the clade Atlantogenata (Foley et al., 2016). This places manatees in an interesting evolutionary position between Marsupialia and Boreoeutheria. They are therefore crucial in understanding the evolution of complex genes such as the TR across eutherian mammals.

2. Methods

2.1. Sample collection

Blood was collected from *T. m. latirostris* (n = 4) during wild capture health assessments in Crystal River, Florida from the flipper in an EDTA-containing vacutainer tube. The blood was processed at the site of collection using the LeukoLock Total RNA Isolation System (Life Technologies, Carlsbad CA) to capture the peripheral blood leukocytes (PBLs). The filters were transported to Texas A&M University at room temperature then stored at −20 °C.

2.2. Total RNA isolation and cDNA synthesis

RNA was isolated from the filter-bound PBLs using the LeukoLock Total RNA Isolation System Kit, following the manufactures instructions. The quantity and quality of the RNA samples were assessed by NanoDrop 2000c spectrophotometer. The 5' RACE cDNA libraries were prepared from the leukocyte RNA using Life Technologies GeneRacer kit with GeneRacer oligo dT and random primers.

2.3. Primer design

Constant regions for TCR α, β, and δ were identified on the *T. manatus latirostris* genome (Broad v1.0/triMan1) by using BLAT to align the human nucleotide sequence to the genomic scaffolds. The TCR γ constant region was not identifiable on the assembled scaffolds, but the transmembrane region was. Reverse primers for the respective constant or transmembrane region for each chain were designed using the Geneious primer design function. The forward primer used was the Generacer 5' Primer.

2.4. TCR RACE PCR, cloning, and sequencing

To validate the primers, 5′ RACE products were amplified by standard PCR using the 5' GeneRacer forward primer and chain specific constant region/transmembrane region primers (Table S1). The amplicons were purified from a 0.8% agarose gel after electrophoresis in tris/acetic acid/EDTA (TAE) buffer, ligated 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 selected based on blue/white screening produced by X-Gal (Sigma-Aldrich, Saint Louis MO). The plasmid DNA was purified using Zyppy Plasmid Miniprep kit (Zymo Research Corporation, Irvine CA) and was digested with *Eco*RI (Promega, Madison, WI) to validate insert size. 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 sample the full TR repertoire.

Transcripts were amplified from four individual *T. m. latirostris* RACE libraries. 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. Sequences are available in GenBank (GenBank: MG989752-MG990740).

2.6. Sequence analysis

Genomic loci were identified by aligning human TR constant regions to the *T. m. latirostris* genome (Broad v1.0/triMan1). V segments were then identified using BLAST to identify Ig domains (Table S2). BLAST hits were visually inspected for signs of functionality: splice sites in the correct frame, lack of stop codons in the coding frame, conserved motifs, proper length of the framework and complementarity determining regions, and the conserved recombination signal sequence (RSS) downstream of the YXC motif. J segments were identified by having a proper sized RSS in the reverse orientation followed by an FGXG motif and 3' splice site in the proper frame. D segments were identified by having two opposite facing RSSs of the proper size for each chain within approximately 100 bp of each other and identification in expressed transcripts. A threshold was applied of four contiguous nucleotides with identity to the genomic D segment for assignment.

The V segments for each chain were assigned to subgroups based on 75% nucleotide identity. Subgroups were numbered by the order in the genome from 5′ to 3'. The number before the dash is based on the order of where the first member of that subgroup occurs, and the number after the dash is based on the order that each member of that subgroup occurs. Subgroup identity between

species was also determined by 75% identity.

Sequence alignments were made in Geneious using the ClustalW plug-in. IMGT V segment alignments were made using the IMGT/V-QUEST using the *Homo sapiens* TR V segment database. Phylogenetic trees were constructed in MEGA7 using the Maximum-likelihood method based on the General Time Reversible model with 100 bootstrap replications.

2.7. Statistical analysis

To compare N and P nucleotide addition on each side of the TRD and TRB D segments, a two-sided T-test was performed in R Studio. CDR length differences between species were compared using ANOVA and Tukey-HSD in R Studio.

3. Results

3.1. TRAD locus

3.1.1. Genomic organization

The *T. m. latirostris* TRAD locus spans several genomic scaffolds (Fig. 1). The 5′ boundary (olfactory receptor loci) is located on scaffold00273 and the 3′ boundary (DAD1) is located on scaffold00268 in the reverse orientation. TRADV segments are also found on scaffold00745, scaffold01496, scaffold03036, and contig164553; only pseudogenes are found on scaffold01603 and scaffold01076. The locus included in the genomic scaffold assembly spans 2,011,757 bp from the leader peptide of the most 5' V segment to the cytoplasmic tail exon of the TRA constant region (TRAC). This is likely an underestimation due to sequence gaps between the scaffolds, but is already larger than the human (*Homo sapiens*; 1 Mb), mouse (*Mus musculus*; 1.65 Mb), and opossum (*Monodelphis domesticus*; 1.3 Mb) TRAD loci, but smaller than the cattle (*Bos taurus*; 2.4 Mb) locus. The *T. m. latirostris* TRAD locus has the typical organization with the TRDD, J, C, and inverted V segment nested between the pool of shared TRADV segments and the TRAJ segments and C region.

3.1.2. TRADV segments

The TRAD locus assembled on the genomic scaffolds contained 112 V segments: 50 functional and 62 pseudogenes or ORFs. The TRAV34-1 segment appeared to have a frameshift in the genomic scaffolds, but was found in expression analyses in all four individuals with one nucleotide difference that determined that it was functional. Similar to other species, some V segments in the TRAD locus expressed with either the TRA or TRD constant region, while some are expressed with both. Of the 50 functional V segments, the *T. m. latirostris* TRAD locus contained 41 TRAV, 6 TRADV, and 3 TRDV (Fig. S1, Fig. S2). The A/D classification of the V segment subgroups was determined by expression data only and not based on sequence similarity to other αδ V segments in other species. We did not use the comparative approach because three TRAD expressed V segments were most similar to the human TRDV1 subgroup and one TRD expressed V segment (TRDV1-4) was not at least 70% identical to any human or mouse subgroup. Therefore, we could not confidently attribute classification based on comparative sequence identity. The expressed transcripts contained 25 V segments expressed only with the TRAC, three V segments only expressed with the TRDC, and five V segments expressed with both the TRAC and TRDC. Concerning the three TRD-only expressed V segments: TRAV15-4 was classified as a TRADV because it belonged to a subgroup of three other TRADV segments; TRDV1-4 had one other functional V segment and two pseudogenes in its subgroup that was not expressed, but were classified as TRDV segments; TRDV5-1 was the sole member of its subgroup. There were three pseudogenes (subgroup TRD2, 3, and 4) downstream of TRAV42-1 and upstream of the TRDD1 that were most similar to TRDV in other species, so due to their locus position and inability to be expressed, they were considered to be TRDV segments. The remaining 16 functional V segments were classified as TRAV segments, but are subject to change with the inclusion of additional expression data.

A feature of the TRAD locus seen in amphibians, birds, and monotremes is the presence of an Ig heavy chain V (VH) segment near the TRD-specific gene segments. This VH segment is expressed with the TRDC, and is therefore called VHD. This feature is not seen in teleost fish, marsupials or eutherians. However, there was one pseudogene in the *T. m. latirostris* TRAD locus downstream of the TRDV3 pseudogenes and upstream of the TRDD1 that was not similar to TRA or TRD V segments in other species. Although it appears highly frameshifted throughout the potential V region, it was most similar to immunoglobulin heavy chain V segments (IgHV), as well as the platypus (*Ornithorhynchus anatinus*) VHD segment (Fig. 2). The highest similarity to a *T. m. latirostris* TRADV was 37.90%, but the highest similarity to a *T. m. latirostris* IgHV was 51.29%. This V segment was considered a VHD pseudogene.

The CDR region lengths of the TRAV and TRDV segments are summarized in Table 1. The two functional members of the TRDV1 subgroup did not produce results for the V-QUEST analysis for either TRD or TRA. Their numbering was inferred based on the FR3 numbering of the human and mouse TRDV segments that are not the conserved inverted TRDV (Fig. S2).

The functional V segments of the TRAD locus were separated into 32 subgroups based on 75% nucleotide identity (Figs. S1 and S2). The inclusion of pseudogenes (excluding the VHD) increased the number to 47 subgroups, which is equal to opossum (47), but more than human (44), mouse (28) and cattle (46). Thirty-six of the *T. m. latirostris* TRADV subgroups grouped with 34 of the 44 human subgroups (77.0%; Fig. 3). Ten *T. m. latirostris* subgroups were not represented in the human repertoire. Six of the 44 human subgroups were not represented in any of the four mammalian species compared (13.6%).

When comparing the locus order of the shared TRADV subgroups between human and *T. m. latirostris*, certain patterns or synteny blocks were revealed. There were five TRAD synteny blocks (SBα) that ranged from two to fourteen TRADV segments (Fig. 4). These blocks indicated regions where the shared subgroups between the two species maintained the same order in the locus. Each SB occurred once in human, but occurred more than once in the *T. m. latirostris* locus. The multiple duplications of the synteny blocks in *T. m. latirostris* generally follow the same order as the human synteny blocks (with the exception of the first copies of SB2α, SB1α, and SB5α).

There was also locus synteny between individual TRADV segment subgroups that had similar locus positions and/or functions between species (human, mouse, cattle, and *T. m. latirostris*). The most 5′ TRAV segment in the locus that is interleaved in the olfactory genes in all three species grouped with the human subgroup HuTRAV1, which is also the V segment used by mucosal-associated invariant T cells (Figs. 3 and 4). The most 3′ TRAV segment before the TRDV segments shares identity between human and *T. m. latirostris*, but not cattle and mouse. The inverted TRDV segment 3' of the δ constant region in all three species shared identity with the human subgroup HuTRDV3. Finally, there were *T. m. latirostris* TRAV segments that had identity with the TRAV segments used by the invariant NKT cells in human (HuTRAV10), mouse, and cattle. The human locus had one, mouse had two, cattle had four, and *T. m. latirostris* had two functional TRAV segments (TRAV6-1 and TRAV6-2) in this interspecies subgroup. We captured one transcript containing TRAV6-2 in our sequencing, and it did not

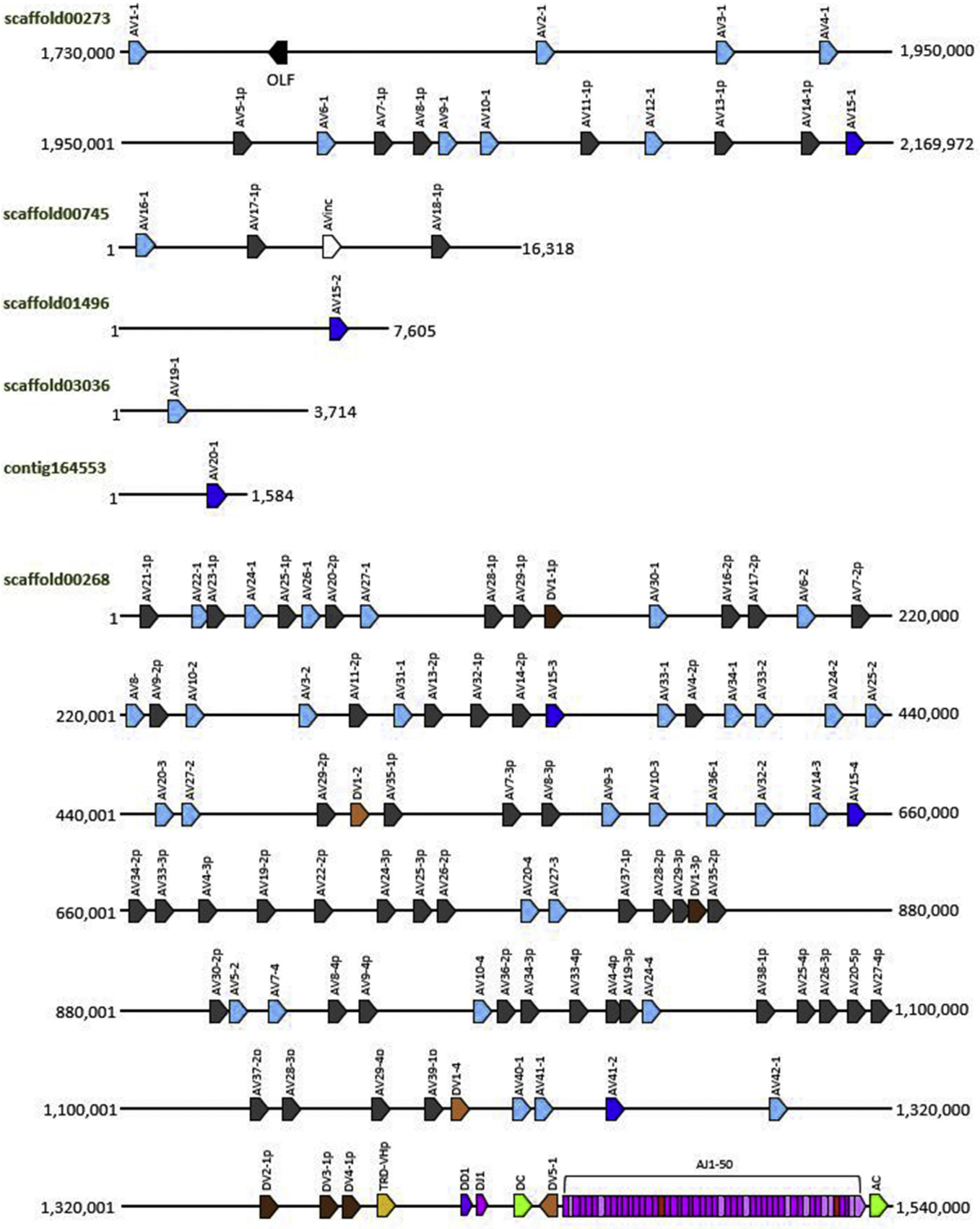


Fig. 1. The TRAD locus is the most complex TR locus. The *T. m. latirostris* TRAD locus span six genomic scaffolds. Scaffold00273 contains the beginning of the locus, and scaffold00268 contains the end of the locus. The remaining four scaffold locations are unknown, but are placed in the middle in descending order by length. The numbers along each line indicate nucleotide position in the scaffold. The color coding is as follows: medium blue are functional TRAV segments, dark blue are expressed TRADV segments, grey are

contain any N and P nucleotides between the V and J segment (TRAJ42). However, this J segment does not contain the DRG motif seen in human and mouse NKT cell receptor J segments that rearrange with these V segments (Fig. S3).

Compared to other species, *T. m. latirostris* has similar numbers of V segments in the TRAD locus (Fig. 5). Even with the conservative assignment of TRADV and TRDV segments, the proportions of both subsections are similar to both human and mouse. Cattle have exceptionally high V segment numbers compared to other mammals. *Trichechus manatus latirostris* is unique in their high proportion of pseudogenes (55.0% of total V segments), with the second highest proportion in sheep (*Ovis aries*; 34.8%).

3.1.3. TRDD segment

The *T. m. latirostris* TRD locus contained one typical D segment of 14 bp between a 5′ reverse 12-spacer RSS and a 3′ forward 23-spacer RSS. While most species have few TRDD segments, *T. m. latirostris* is the only species to date that has one (Fig. 5). This D segment was identified in the majority (117/130) of TRD expressed transcripts with a four contiguous nucleotide identity threshold (Fig. S4). Only one sequence did not show at least a three contiguous nucleotide identity. There are significantly more N and P nucleotide additions 5′ to the D segment than 3′ to the D segment ($p = 7.741e\text{-}15$). This one D segment is located very close to the TRDJ segment, so there may be another D segment 5′ of the identified segment with a less conserved RSS that could not be identified by the program used. There are approximately 22 kb between the most 3′ V segment and the TRDD1 segment. The potential rearrangement of two D segments (D-D fusion) would explain the 5′ skew for the N and P additions and this phenomenon occurs in other species. The CDR3 regions 5′ of the TRDD1 sequence in expressed transcripts were aligned to the genomic region between the most 3' V segment and TRDJ1 to identify potential TRDD segments with cryptic RSSs. However, the sequences did not group to any particular region, so we could not identify any additional TRDD segments.

3.1.4. TRADJ segments

There were 50 TRAJ segments, two of which were pseudogenes (Fig. S3). The functional J segments ranged from 18 to 22 amino acids. The nucleotide similarity between the TRAJ segments ranged from 29.5 to 71.9% with an average of 49.3%. The other mammalian species compared had roughly 50 TRAJ segments except mouse, which has 38 (Fig. 5).

There was one TRDJ segment identified in the genome that was 15 amino acids long. It was used in all expressed TRD sequences, so it is likely the only TRDJ segment. Similar to the TRDD segments, there are few TRDJ segments in the mammalian species compared, but *T. m. latirostris* is the only species with one (Fig. 5).

3.1.5. Expression profile

From 5668 total sequences, we obtained 164 unique functionally rearranged transcripts from four individuals' PBLs using a TRA constant region primer. These transcripts encoded 30 of the 50 functional TRADV segments and 38 of the 48 functional TRAJ segments. There was one dominantly expressed TRADV segment, TRAV3-2, which was expressed in 39.0% of the transcripts; the other 29 TRADV segment's frequency ranged from 9.8 to 0.6% (Fig. 6). The TRAJ segment expression was more evenly distributed with the most frequent TRAJ segment, TRAJ28, expressed in 7.3% of the transcripts. The CDR3 length ranged from 8 to 20 amino acids with an average of 12.2 (Table 1).

From 156 total sequences, we obtained 130 unique functionally rearranged transcripts from four individuals' PBLs using a TRD constant region primer. These transcripts encoded 8 of the 50 functional TRADV segments, five of which were also expressed with the TRA constant region, and the one genomic J segment was found in all transcripts. Two TRDV segments dominated expression, TRDV5-1 and TRAV15-4, with frequencies of 36.2% and 26.9%, respectively (Fig. 6). Both of these V segments were only expressed with the TRD constant region. The CDR3 length ranged from 8 to 24 amino acids with an average of 16.2 (Table 1).

3.2. TRB locus

3.2.1. Genomic organization

The TRB locus is located on scaffold00093 and spans 433,758 bp from the most 5′ V segment leader peptide to the leader of the inverted V segment 3′ of the constant region (Fig. 7). The 5′ end of the locus is flanked by trypsinogen genes (TRY) and monooxygenase DBH-like 2 (DBHL), and the 3' end is flanked by ephrin type-b receptor 6 (EPHB6). There are several sequence gaps throughout the locus. The *T. m. latirostris* TRB locus is similar in size to the opossum locus (400 kb) and smaller than both human and mouse (approximately 620 kb and 700 kb, respectively).

3.2.2. TRBV segments

The *T. m. latirostris* TRB locus contained 70 V segments: 40 were functional, 30 were pseudogene or ORFs. The CDR region lengths of the TRBV segments are summarized in Table 1. The functional TRBV segments were separated into 19 subgroups based on 75% nucleotide identity (Fig. S5). The inclusion of pseudogenes created 24 subgroups, which is fewer than human (33), mouse (31), cattle (31) and opossum (28). Eleven of the 24 *T. m. latirostris* subgroups (42.3%) shared at least 75% nucleotide identity to human TRBV subgroups, leaving 14 *T. m. latirostris* subgroups not represented in the human repertoire (Fig. 3). Seven of the 30 human subgroups were not represented in any of the four mammalian species compared (23.3%).

When comparing the locus order of the shared TRBV subgroups between human and *T. m. latirostris*, synteny blocks were revealed similar to the TRA locus. There were three beta synteny blocks (SBβ) that ranged from two to nine TRBV segments that appear in different copy numbers in each species (Fig. 4). The multiple copies of the synteny blocks group together in the locus. Interestingly, the human and *T. m. latirostris* TRB loci both have regions of multiple copies of three V segments, but the subgroups involved are not the same. The only positional locus synteny in the TRB locus is with the inverted HuTRBV30 downstream of the D-J-C clusters, which showed identity to the mouse and cattle TRBV segments of that position, but not to *T. m. latirostris*.

Compared to other mammalian species, *T. m. latirostris* had an intermediate number of V segments (Fig. 5). There was a broad range in the number of TRB V segments between species. Similar to TRA, cattle have an exceptionally high number of V segments (79), followed by human (48) and *T. m. latirostris* (40) with nearly half the amount of cattle. Mouse (22), opossum (27), and dog (*Canis familiaris*; 20) have nearly half the amount of V segments of *T. m. latirostris*. While pseudogene numbers are typically high for TRB, the proportion of functional to pseudogene V segments is high in *T. m. latirostris* (42.9%), surpassed only by dog (45.9%).

pseudogene TRAV segments, medium orange are functional TRDV segments, dark orange are pseudogene TRDV segments, yellow are pseudogene VHD segments, purple are TRDD segments, pink are TRADJ segments, red are pseudogene TRADJ segments, and green are TRADC regions. V segments are numbered by subgroup and chronological order within the locus. The points on the pentagons indicate transcriptional orientation. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

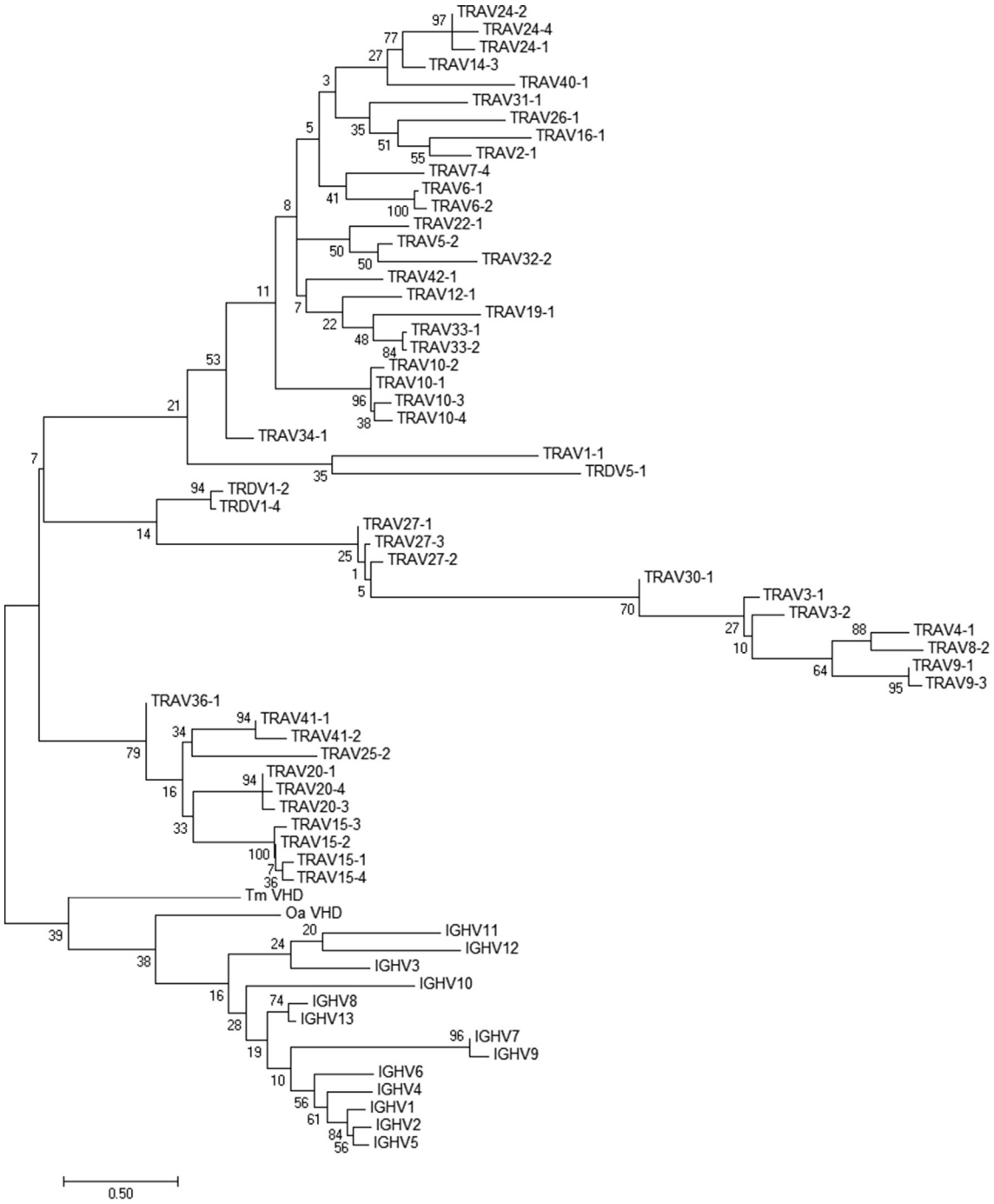


Fig. 2. The VHD pseudogene in the TRAD locus is most similar to IGHV segments and the platypus VHD segment. Maximum-likelihood tree of the *T. m. latirostris* TRAD and IGH V segments with the *T. m. latirostris* and platypus (*Ornithorhynchus anatinus*) VHD segment. The VHD segment branches are highlighted in red and identified with the first letters of their genus and species (Tm and Oa). The branch lengths are measured by the number of substitutions per site, as indicated by the scale.

Table 1
All TR chain characteristics.

	#V	# Vψ	CDR1	CDR2	CDR3	Subgroups	#D	#J	J AA	#C
TCRA	51	57	5–8	4–8	8–20	44	0	50	18–22	1
TCRB	40	30	5–6	5–7	8–19	24	2	10	15–17	2
TCRG	14	2	5–8	8	5–21	4	0	7	16–18	5
TCRD	3	5	7–8	5–6	8–24	4	1	1	15	1

3.2.3. TRBD segments

The two D-J-C clusters each have one TRBD segment that share 69.2% nucleotide identity: TRBD1 is 13 nucleotides long, and TRBD2 is 15 nucleotides long. Each of the TRBD segments were flanked by a 5′ reverse orientation 12-spacer RSS and a 3' forward orientation 23-spacer RSS. All species compared shared the one TRBD segment per D-J-C cluster arrangement, however cattle had three clusters and opossum had four.

A

Human	Mouse	Cow	Manatee	Opossum
TCRAV1	■	■	■	
TCRAV2		■		
TCRAV3		■		
TCRAV4	■	■		
TCRAV5		■	■	
TCRAV6				
TCRAV7				
TCRAV8		■	■	
TCRAV9	■	■	■	
TCRAV10	■	■	■	
TCRAV11		■	■	
TCRAV12			■	
TCRAV13	■	■	■	
TCRAV14		■	■	
TCRAV15				
TCRAV16		■	■	
TCRAV17	■	■	■	
TCRAV18		■	■	
TCRAV19	■	■	■	
TCRAV20		■	■	
TCRAV21		■	■	
TCRAV22		■	■	
TCRAV23	■	■	■	
TCRAV24		■	■	
TCRAV25		■	■	
TCRAV26		■	■	
TCRAV27		■	■	
TCRAV28		■	■	
TCRAV29		■	■	
TCRAV30				
TCRAV31			■	
TCRAV32			■	
TCRAV33			■	
TCRAV34			■	
TCRAV35		■	■	
TCRAV36		■	■	
TCRAV37			■	
TCRAV38	■	■	■	■
TCRAV39		■		
TCRAV40				
TCRAV41			■	
TCRDV1		■	■	
TCRDV2				
TCRDV3	■	■	■	

B

Human	Mouse	Cow	Manatee	Opossum
TRBV1				
TRBV2				
TRBV3	■	■		
TRBV4		■		
TRBV5		■		
TRBV6		■		
TRBV7		■	■	
TRBV8				
TRBV9		■	■	
TRBV10	■	■	■	
TRBV11	■	■	■	
TRBV12	■	■		
TRBV13		■		
TRBV14		■		
TRBV15	■	■	■	
TRBV16		■	■	
TRBV17				
TRBV18		■		
TRBV19	■	■	■	
TRBV20		■		
TRBV21		■	■	
TRBV22				
TRBV23	■			
TRBV24		■	■	
TRBV25		■	■	
TRBV26				
TRBV27				
TRBV28	■	■		
TRBV29	■	■	■	
TRBV30	■	■		

C

Human	Mouse	Cow	Manatee	Opossum
TRGV1				
TRGV2				
TRGV3		■	■	
TRGV4				

Fig. 3. Subgroup conservation across eutherian mammals varies by locus. This matrix indicates 75% nucleotide identity to human TRADV (**A**), TRBV (**B**), and TRGV (**C**) subgroups for mouse, bovine, manatee, and opossum. A red cell denotes 75% identity. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.2.4. *TRBJ segments*

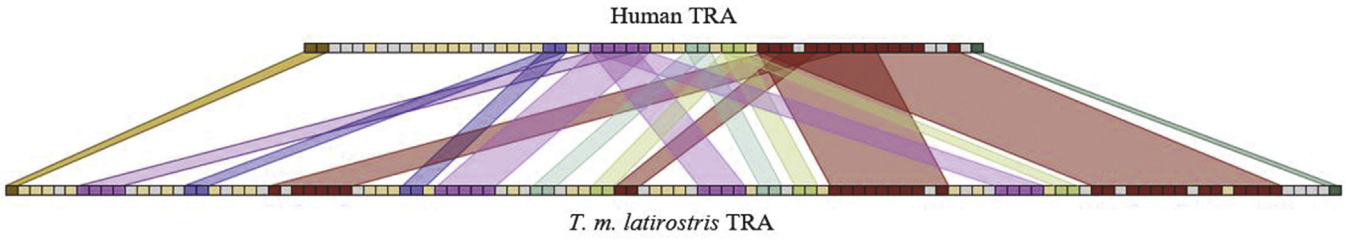
The TRBJ segments were split between two D-J-C clusters: six in cluster 1, and four in cluster 2. The lengths ranged from 14 to 17 amino acids (Fig. S6). The similarity between the J segments ranged from 71.4% to 42.2% with an average of 57.3% overall. The average percent identity within clusters was not significantly different from identity between clusters ($p = 0.1216$). The number of J segments is relatively similar between species, with the exception of cattle and opossum that have more than two D-J-C clusters (Fig. 5).

3.2.5. *Expression profile*

From 4095 total sequences, we obtained 490 unique functionally rearranged sequences between the four individuals' PBLs using a TRB constant region reverse primer. These transcripts encoded 29 of the 40 functional V segments and all 10 of the functional J segments. The two most dominantly expressed V segments were TRBV13-2 at 13.1% and TRBV20-1 at 14.9% (Fig. 8). Two J segments dominated expression: TRBJ1-2 with 25.5% and TRBJ 2-1 with 19.0%. Considering only the sequences that were assigned one D segment, 80.7% used TRBD1 and 19.3% used TRBD2 (Fig. S7). There were 163 of the sequences that did not pass the threshold of five contiguous nucleotides similar to either D segment, and could therefore not be assigned a D segment with confidence. The confirmed D-D fusions made up 6.3% of the total sequences (Fig. 9). The CDR3 length ranged from 8 to 19 amino acids with an average of 12.6 amino acids (Table 1).

Both of the D segments can rearrange with J segments from both clusters (Fig. S7). Of the 238 transcripts that contained TCRBD1, 71 (29.8%) rearranged with cluster 2 J segments. Of the 58 transcripts that contained TCRBD2, 12 (20.7%) rearranged with cluster 1 J

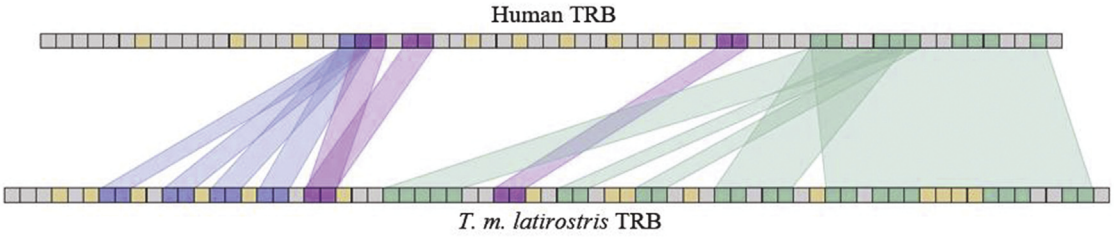
A



TRA	Human	*T. m. latirostris*
Shared	77.19 %	78.57 %
Syntenic	49.12 %	56.25 %
Not syntenic, shared	55.17 %	51.02 %
Not shared	22.81 %	21.43 %

	Human Subgroups	*T. m. latirostris* Subgroups
SBα1	HsTRA13,14	TmTRA16,17
SBα2	HsTRA12,8,16,17,18	TmTRA7,8,9,10,3
SBα3	HsTRA22,23	TmTRA32,14
SBα4	HsTRA24,25	TmTRA34,33
SBα5	HsTRA27,28,29,31,32,33, 26,34,35,36,37,38,38,41	TmTRA19,22,24,25,26,20, 27,37,29,39,40,41,41,42

B



TRB	Human	*T. m. latirostris*
Shared	38.46 %	71.43 %
Syntenic	24.62 %	50.00 %
Not syntenic, shared	18.37 %	42.86 %
Not shared	61.54 %	28.57 %

	Human Subgroups	*T. m. latirostris* Subgroups
SBβ1	HsTRA9,10	TmTRA4,5
SBβ2	HsTRA10,11	TmTRA5,9
SBβ3	HsTRA15,16,19,20,21,24, 25,26,29	TmTRA12,13,14,15,16,18, 19,20,23,23

Fig. 4. TRA has more locus synteny between human and *T. m. latirostris* than TRB in the V array. A cartoon representation of the synteny blocks identified between the human and *T. m. latirostris* TR loci. Each square represents one functional or pseudogene TRADV segment. The synteny blocks (SB) are represented by their color indicated in the table below each locus pair. TRADV segments that have at least 75% identity between human a *T. m. latirostris*, but are not within synteny blocks are colored yellow, and TRADV segments that do not share identity between species are in grey. Intergenic sequence is not included. A) There are five synteny blocks between human and *T. m. latirostris* TRA loci. The two locus position synteny groups are: human subgroup TRAV1 and *T. m. latirostris* subgroup TRAV1 (light brown), and human subgroup TRDV3 and *T. m. latirostris* subgroup TRDV4 (dark green). The table indicates the percentage of TRADV segment subgroups that are shared between human and *T. m. latirostris* (all colored squares), TRADV segments within synteny blocks (all colored squares except yellow), TRADV segments not within synteny blocks but shared between human and *T. m. latirostris* (only yellow squares), and TRADV segments not shared (grey squares). B) There are three synteny blocks between human and *T. m. latirostris* TRB loci. There are no locus position synteny groups in TRB. The table indicates the percentage of TRBV segment subgroups that are shared between human and *T. m. latirostris*, TRBV segments within synteny blocks, TRBV segments not within synteny blocks but shared between human and *T. m. latirostris*, and TRBV segments not shared. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

segments. The D-D fusions could rearrange with both J clusters, regardless the order of the TRBD segments (Fig. 9). There were 20 more transcripts containing evidence for D-D fusions, but they did not pass the threshold requirements.

3.3. TRG locus

3.3.1. Genomic organization

The *T. m. latirostris* genome assembly only covers a few regions of the TRG locus, so the genomic organization is largely undetermined (Fig. 10). The 5′ end of the locus is located on scaffold00114, identified by the presence of amphiphysin, which flanks the TRG locus in other mammals. This scaffold is 9,676,305 bp long; however, the region 3' of amphiphysin is only 50,580 bp and contains one functional TRGV segment and one pseudogene TRGV. Another functional TRGV segment is located on contig164298, which is 1722 bp long. Two scaffolds contain one TRGJ segment each: scaffold01250 at 7715 bp and scaffold01597 at 6351 bp. Finally, while no constant regions were identified on a genomic scaffold, a TRG transmembrane region was identified on scaffold03157, which is 3598 bp long.

3.3.2. TRG constant regions

Five Ig domains and five connecting peptides were identified in expressed transcripts using a reverse primer specific for the transmembrane region identified in the genomic assembly (Fig. S8). The five Ig domains are all 109 amino acids long and share 92.0–99.1% nucleotide identity. One of the Ig domains, TRGC4, has two splice variants that are 91 and 65 amino acids long due to alternative splice sites that truncate the domain at the 5' end identified in functionally rearranged transcripts. The 91 amino acid splice variant was seen in two individuals and in 2.66% of the total transcripts, while the 65 amino acid splice variant was seen in one individual and in 0.52% of the total transcripts.

The relationship between Ig domains and connecting peptides is not 1:1; some connecting peptides were expressed with more than one Ig domain and some Ig domains with more than one connecting peptide (Fig. 11). The connecting peptides range in size

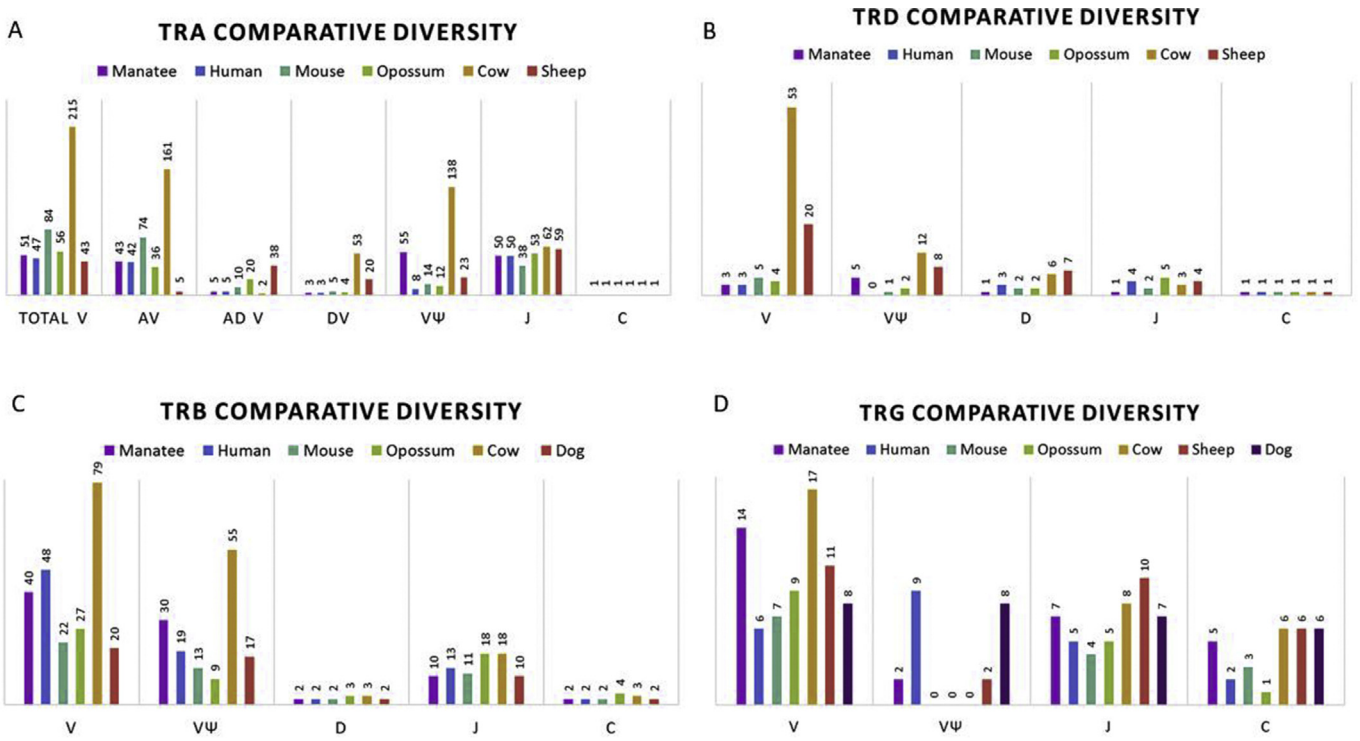


Fig. 5. ***T. m. latirostris*** **segmental diversity is average for TRA and TRB, low for TRD and high for TRG compared to select other mammals.** Graphical representation of the varied segmental diversity between species. The y-axis represents the number of genes, and the numerical values are indicated above each bar. The genes of the TRAD locus were analyzed as a collective of all V segments broken down into their subcategories with the TRA non-V genes and as TRD only.

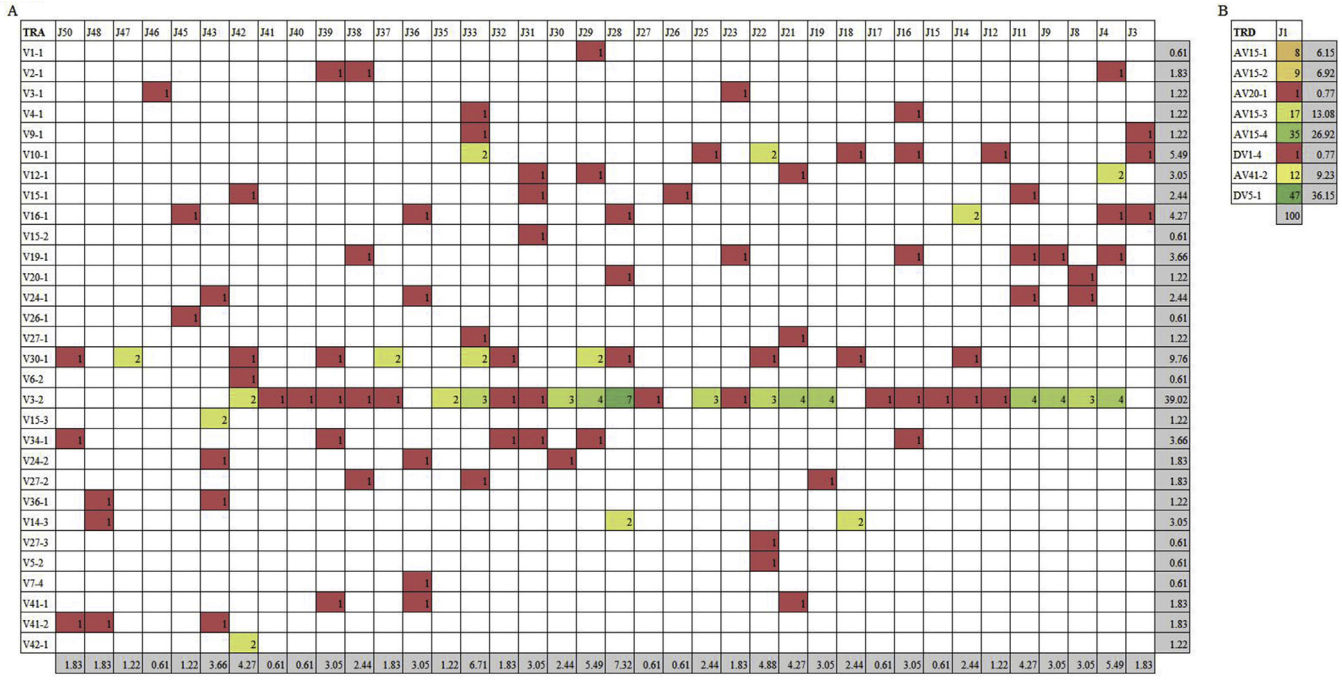


Fig. 6. TRA has high combinatorial diversity while TRD has low combinatorial diversity. The expressed A) TRAV and B)TRDV segments are indicated in the rows and the expressed TRJ segments are indicated in the columns. The percent of the total number of expressed transcripts that each segment was used is indicated at the end of each row or column. The number of expressed transcripts each V/J combination was observed is indicated in the cells. The cells are colored on a scale based on their values: the lowest values are in red, the midpoint values are in yellow, and the highest values are in green. Empty cells indicate combinations that were not observed. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

from 26 to 59 amino acids. The connecting peptide that splices with TRGC4 (TRGCP4) also has a splice variant that is 11 amino acids longer than the more frequently used variant (11.66% vs 43.15%). Both connecting peptide variants are expressed with all three

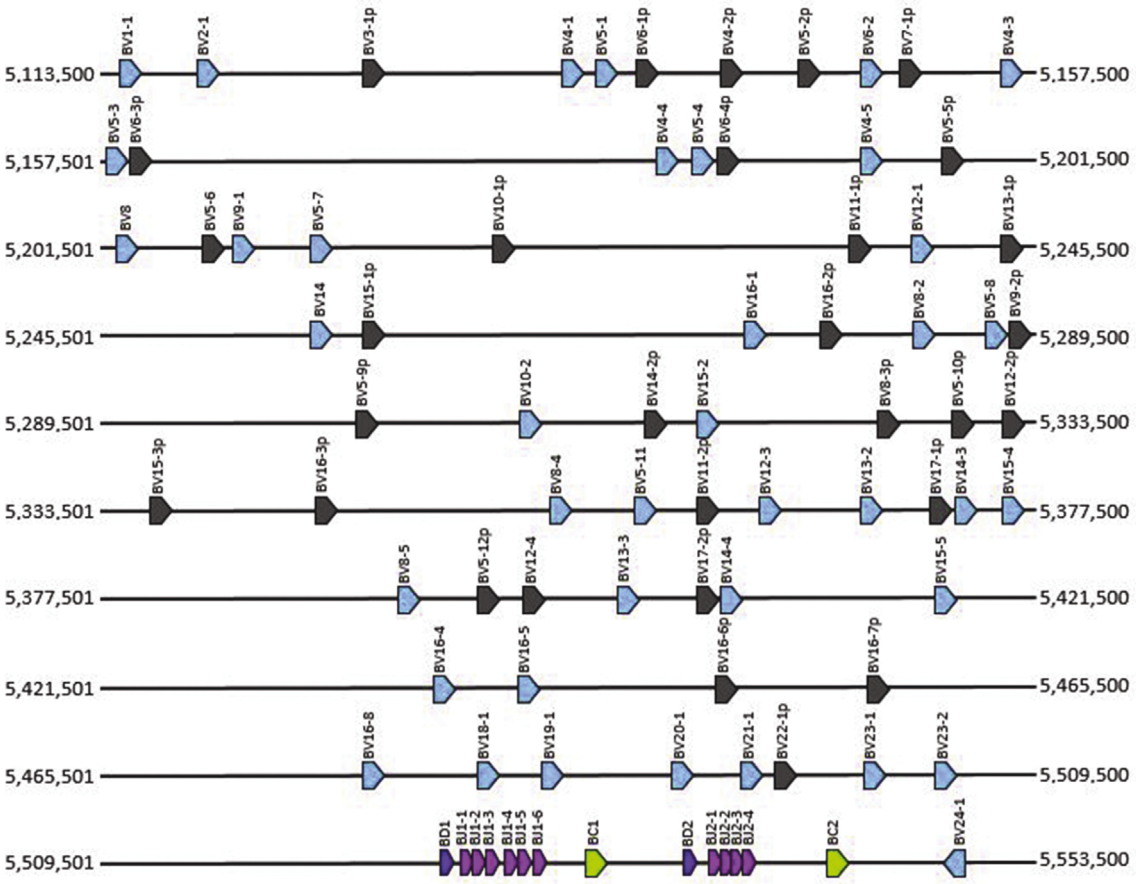


Fig. 7. The TRB locus is moderately complex. The *T. m. latirostris* TRB locus is located on genomic scaffold00093. The numbers along each line indicate nucleotide position in the scaffold. The color coding is as follows: medium blue blocks are functional TRBV segments, grey are pseudogene TRBV segments, purple are TRBD segments, pink are TRBV J segments, and green are TRBC regions. V segments are numbered by subgroup and chronological order within the locus. The points on the blocks indicate transcriptional orientation. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

TRGC4 splice variants. The transmembrane region sequence for all the constant regions are identical, but because some of the connecting peptide is encoded in the same exon as the transmembrane region, the transmembrane region captured by the genomic scaffolds belonged to TRGCP1.

3.3.3. TRGV segments

Combining the TRGV segments identified in the genomic scaffolds and expressed transcripts, 16 TRGV segments were identified: 14 functional and 2 pseudogenes. The CDR region lengths of the TRGV segments are summarized in Table 1. The functional TRGV segments could be separated into three subgroups based on 75% nucleotide identity (Fig. S9). The inclusion of the pseudogenes created four subgroups. This is comparable to human (4) and mouse (5), but fewer than cattle (10) and sheep (11). Only one of the four manatee subgroups had at least 75% identity to any human TRGV subgroup, and three of the four human subgroups were not represented in any of the species compared (75%; Fig. 3). Due to the lack of sequence similarity to human and the lack of the genomic organization for *T. m. latirostris*, locus synteny was not analyzed for the TRG locus.

Even with the incomplete genomic data, manatees have an above average number of TRGV segments compared to other species (Fig. 5). Considering the species with known circulating γδ T cell proportion, the γδ "low" species (human and mouse) had approximately half the number of TRGV segment of the γδ "high" species (cattle and sheep). The *T. m. latirostris* TRGV numbers are intermediate to the two γδ "high" species, but no work has yet been done to determine their proportion of circulating γδ T cells.

3.3.4. TRGJ segments

There were seven TRGJ segments identified in sequence, two of which were found on genomic scaffolds (Fig. S10). The TRGJ segments can be separated into two identity groups: TRGJ1, TRGJ2, TRGJ3, and TRGJ4 share 90.7–98.1% identity, whileTRGJ4, TRGJ5, and TRGJ6 share 85.4–97.8% identity. Between the two groups, the highest identity is 53.3%. Similar to the TRGV segments, there are more TRGJ segments in the γδ "high" species; however, *T. m. latirostris* fell between the "high" and "low" groups in TRGJ segments (Fig. 5). There may be more TRGJ segments that are not yet sequenced by either expression or genomic scaffolds.

3.3.5. Expression profile

From 2579 total sequences, we obtained 338 unique functionally rearranged TRG transcripts between the four individuals' PBLs using the TRG transmembrane reverse primer. The expressed transcripts contained 14 functional TRGV segments and seven TRGJ segments. The TRGV segment with the highest expression level was

	J1-1	J1-2	J1-3	J1-4	J1-5	J1-6	J2-1	J2-2	J2-3	J2-4	
V1-1		1				1	2			1	1.02
V2-1		2		1			2		1	1	1.43
V4-1	3	1	2	1			2	1		1	2.24
V5-1	4	6		4			2	3		1	4.08
V6-2	2	1		3	2		3	1	1		2.65
V4-4		3		1							0.82
V4-5	1			2			1				0.82
V8-1		4	1	1	3	3	1		2		3.06
V9-1	6	17	1	2			9	5	2		8.57
V12-1		2		1			3	3		1	2.04
V16-1	4	3	3	1	1	2	2	4	4	1	5.10
V5-8		1			1	1					0.61
V15-2		2									0.41
V8-4	4	11		4	3		4		1		5.51
V12-3	2	1						1		1	1.02
V13-2	4	13	5	4	10	1	12	10	2	3	13.06
V15-4	2						1		1		0.82
V8-5	2	12	1	3	1		1			4	4.90
V13-3	1						2	1			0.82
V15-5	3	1	3	2	1	1	5	1	4	1	4.49
V16-4	2	5	1	3	1	1	4	3	4		4.90
V16-5	2	6		2	2		4	2	1	1	4.08
V16-8	1	3		1		1		1	1	6	2.86
V18-1		1									0.20
V20-1	5	20	3	11	4		20	6	3	1	14.90
V21-1		1		1			1			1	0.82
V23-1	1	4	2		1		7	1	3		3.88
V23-2		1					2				0.61
V24-1		3	2	4	2	1	3	1	2	3	4.29
	10.00	25.51	4.90	10.61	6.53	2.45	18.98	8.98	6.53	5.51	

Fig. 8. TRB combinatorial diversity is skewed to a few V-J Combinations. The expressed TRBV segments are indicated in the rows and the expressed TRBJ segments are indicated in the columns. The percent of the total number of expressed transcripts that each segment was used is indicated at the end of each row or column. The number of expressed transcripts each V/J combination was observed is indicated in the cells. The cells are colored on a scale based on their values: the lowest values are in red, the midpoint values are in yellow, and the highest values are in green. Empty cells indicate combinations that were not observed. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

TRGV3-3, which was expressed in 49.0% of the transcripts (Fig. 12). The remaining TRGV segments were expressed in 7.2–0.3% of the transcripts. The TRGJ segment that dominated expression was TRGJ5, which was expressed in 53.9% of the transcripts. The remaining TRGJ segments were expressed in 10.1–1.3% of the transcripts. Of the transcripts expressing TRGV3-3, 97.4% of the were rearranged with TRGJ5. The CDR3 length ranged from 5 to 21 amino acids with an average of 10.56 (Table 1).

Since the TRGV and TRGJ combinations did not reveal any pattern to suggest a hypothetical genomic organization, we looked at the TRGJ and TRGC combinations (Fig. 12). While there were some exceptions, some general patterns arose: TRGJ2 spliced to TRGC1, TRGJ1 and TRGJ6 spliced to TRGC3.1, TRGJ3 spliced to TRGC3.2, TRGJ4 and TRGJ7 spliced to TRGC2, and TRGJ5 spliced to TRGC4.

D1: GGGACAGGGGAGC D2: GGGACTTGGGGGTAG

D1-D2-J Cluster 1

BV8-1 J1-3 T1 1413: CGACCTCGGACAGGGAGGACT
BV8-1 J1-2 T1 1415: GCTACCCTCCGACAGGCCGATATGGAC
BV9-1 J1-1 T1 1413: ATTTGACAGGTTTGGAC
BV13-2 J1-2 T1 1411: GGCGTGAGCAGGGAAGTGGGGGAT
BV13-2 J1-2 T3 1413: TCGGACCTCAGGGGAGCGGGAAT
BV13-2 J1-2 T4 1414: ATGTGGCAGGATCGACAAAAGGGGATTAT
BV8-5 J1-2 T7 1414: GGACAGCCGGGTATCGTCGGGCAGGGT
BV13-3 J1-1 T1 1414: AGGACACGGGAG
BV16-4 J1-4 T2 1414: TAGAACAGCCGGGA
BV16-4 J1-2 T2 1414: CCGATTCAAAGAGCTGGGCGTGGGAGAGA
BV16-5 J1-1 T1 1411: AGGGACGGAGAATTGGGGTATTGCAAACACCCTACCG
BV20-1 J1-3 T2 1414: TTGGATTATTGGCATCT
BV24-1 J1-3 T1 1413: GGAGCGATTGGAGCTCT
BV23-1 J1-2 T1 1413: AATGACAGGTGGGGT

D1-D2-J Cluster 2

BV4-1 J2-2 T1 1414: GTGATAGGGGAGCGGGGGTATACCGG
BV13-2 J2-2 T1 1415: ATTCTAGGGGCTTGACTTGGGGGCAGGGC
BV15-5 J2-1 T3 1415: GACAGAGCTGGT
BV16-8 J2-1 T3 1414: CTCAGCATCCGGTTCAGGGGAGCTTGGGAG
BV20-1 J2-1 T2 1413: GGAGACAGGTTTGGAGATGAGCAG
BV20-1 J2-1 T2 1414: AGGGGATTGGGCTAC
BV23-1 J2-1 T2 1414: ATCTCAAGGACAGGAGGGATATAC

D2-D1-J Cluster 1

BV8-4 J1-1 T1 1413: ACAATTGGGCGACCGGAG
BV8-5 J1-4 T1 1414: AATGGGGGATGACAGGG

D2-D1-J Cluster 2

BV2-1 J2-4 T1 1414: AGGGAATACAGGGGG
BV9-1 J2-1 T1 1411: GGTCGGGGGAATGGGAGCTACAAT
BV13-2 J2-2 T1 1411: CTCGGACTTACAGGGATAACG
BV13-2 J2-3 T1 1414: GCTGGGGGTCCGACAGTCCT
BV16-4 J2-1 T1 1411: GGGGGGTACAGGGGAGCAGCTACAAT
BV16-4 J2-2 T1 1413: CCGGACCCAGGGCAGAGAC
BV20-1 J2-1 T1 1411: CGGGTACAGGGAGGGAAT
BV12-3 J2-4 T1 1413: GGACCCACAGTTG

Fig. 9. TRB can form D-D fusions in both 5′ and 3′ directions. A list of CDR3 sequences from expressed TRB transcripts that exhibit D-D fusions with J segments from DJC Cluster 1 and 2. The sequence highlighted in blue is derived from TRBD1 and the sequence highlighted in green is derived from TRBD2. The sequences are divided into four groups: D1-D2-J Cluster 1; D1-D2-J Cluster2; D2-D1-J Cluster 1; D2-D1-J Cluster 2. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

4. Discussion

4.1. Chain-specific features that determine functional differences

Although these four antigen receptor genes are created by the same genomic rearrangement mechanism and have the same basic function, they use different strategies to create diversity that gives each chain a different niche. Due to the evolutionary position of *T. m. latirostris* in the basal clade of Eutheria, it is possible that these features that are similar to human are shared by most other eutherians (Foley et al., 2016).

TRA has the highest number of V and J segments and the most diverse V and J segments in sequence identity. This combinatorial diversity was also seen in the expressed transcripts, with the most frequent V and J combination only comprising 4.3% of the sequenced repertoire. The TRA CDR3 lengths are comparable to TRB despite not using any D segments. This is achieved by having the longest J segments of any chain (Rock et al., 1994). This increases binding diversity in the CDR3α that interacts with the peptide within the MHC binding cleft (Jorgensen et al., 1992). However, it is limited to genomically encoded CDR3 diversity because there is very little N and P addition. TRBV and TRBJ segments are less numerous and sequence diverse than TRA, but TRB has the ability to

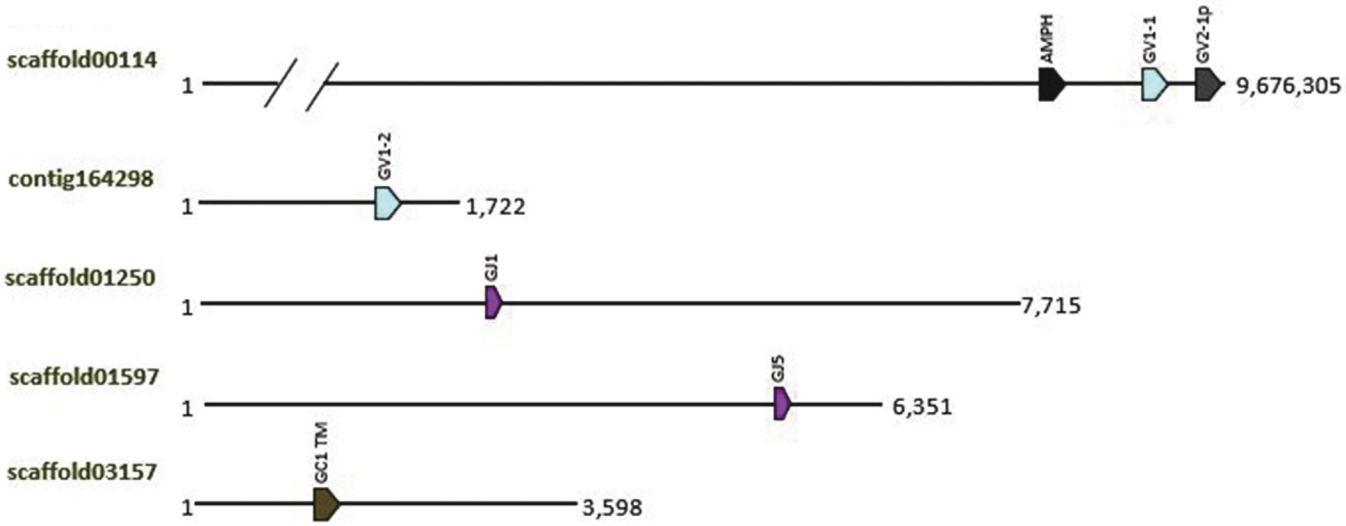


Fig. 10. The TRG locus is poorly assembled. Small pieces of the *T. m. latirostris* TRG locus is located on five genomic scaffolds. The numbers along each line indicate nucleotide position in the scaffold. The color coding is as follows: medium blue blocks are functional TRGV segments, grey are pseudogene TRGV segments, pink are TRGV J segments, and brown are TRG transmembrane regions. V segments are numbered by subgroup. The points on the blocks indicate transcriptional orientation. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

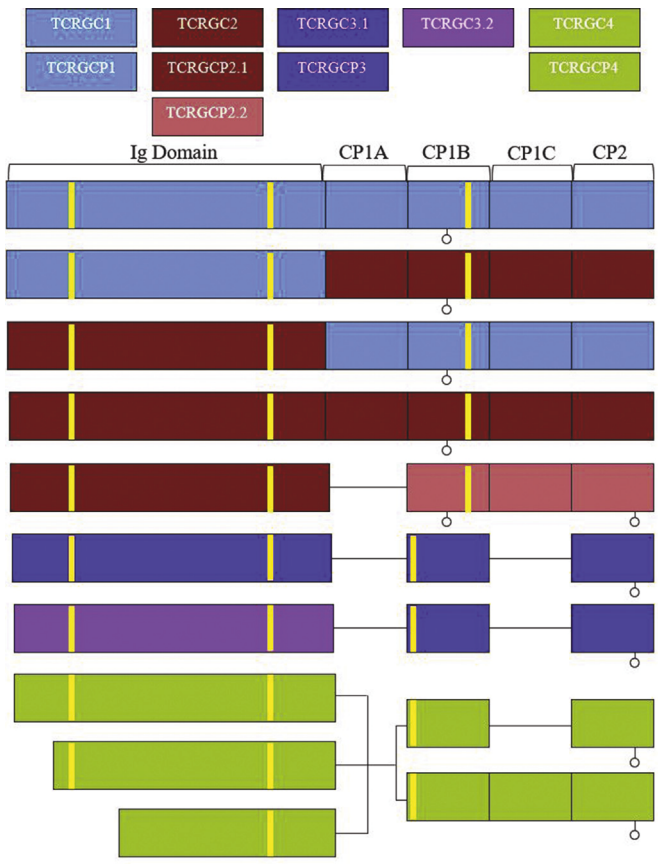


Fig. 11. TRG constant region and connecting peptide regions have variable expression patterns. Cartoon representations of the TRG constant region and connecting peptide configurations seen in expressed transcripts. The Ig domain constant region is Exon 1 (E1), the connecting peptide predicted exons are E2A, E2B, and E2C, and the connecting peptide and transmembrane exon is E3. Cysteine residues are indicated by yellow bars and possible N-linked glycosylation sites are indicated by lollipops. Regions of the same color indicate constant region-connecting peptide combinations that are seen most frequently. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

create diverse CDR3 in a non-genomically encoded way. While one TRBV and J combination did not dominate expression, there were ten combinations that were more frequent than the others that comprised 27.8% of the repertoire, seven of which used the same two TRBV segments. Therefore, there seems to be selection for certain TRBV segments that may bind the *T. m. latirostris* MHC alleles with higher affinity since TRB is more responsible for binding MHC than TRA (Luz et al., 2002). The addition of a D segment increases the amount of random N and P addition, and the TRB chain can create D-D fusions in both upstream and downstream order with J segments from either cluster.

While the TRD chain may rearrange with more V segments in the TRAD locus than we identified, TRD used considerably fewer V segments than the TRA constant region in expressed transcripts, half of which were from the same TRADV subgroup. The number of V segments that can rearrange to TRD may be underestimated due to the small number of sequences obtained. While the same concentration of cDNA amplicon was included in the pooled sample for the PacBio SMRT library preparation, it did not receive the same coverage as the other chains. However, similar expression patterns were observed across all four individuals, so other TRDV segments would likely have low expression rates. There are usually some TRD-specific V segments between the pool of TRADV and the TRDD segments, but this subset in *T. m. latirostris* were all pseudogenes (Glusman et al., 2001; Parra et al., 2008; Parra and Miller, 2012). The segmental diversity was also limited to only one TRDD and TRDJ, which is the most limited TRD segmental diversity we know of to date in mammals. Two V and J combinations dominated expression with 63.1% of the repertoire. However, the TRD-only V segments had the longest average CDR1 of the four chains, and the average CDR2 length was significantly longer than human, mouse, and sheep TRDV. The TRD CDR3 region was also longer on average than the other four chains due to extensive N and P addition rather than D-D fusions as seen in other species (Piccinni et al., 2015; Uenishi et al., 2009). Therefore, the TRD CDR3 has even more non-genomically encoded CDR3 diversity than TRB and can extend further beyond the planar structure of the paratope than the other chains. The TRGV have the longest average CDR2 of the four chains, but also the shortest average CDR3 length since it does not have a

	J1	J2	J3	J4	J5	J6	J7	
V1-1				2				0.52
V1-2		2						0.52
V1-3	1	1		4			3	2.32
V1-4	3			5		1	2	2.84
V1-5	1	2		4			7	3.61
V1-6	2	8		4		2	8	6.19
V1-7	3	13		4		3	5	7.22
V1-8	1			1		1		0.77
V1-9		4			1		2	1.80
V3-1				1				0.26
V3-2		2		1	20			5.93
V3-3			5		185			48.97
V4-1	5	4		3	1			3.35
V4-2		3		1	2		1	1.80
	4.12	10.05	1.29	7.73	53.87	1.80	7.22	
	J1	J2	J3	J4	J5	J6	J7	
C1		37		1	2	7	1	12.37
C2		1		31	1		20	13.66
C3.1	15				3		8	6.70
C3.2			5					1.29
C4					198			51.03

Fig. 12. TRG combinatorial diversity is dominated by one TRGV/TRGJ combination. The expressed TRGV segments are indicated in the rows and the expressed TRGJ segments are indicated in the columns. The percent of the total number of expressed transcripts that each segment was used is indicated at the end of each row or column. The number of expressed transcripts each V/J combination was observed is indicated in the cells. The cells are colored on a scale based on their values: the lowest values are in red, the midpoint values are in yellow, and the highest values are in green. Empty cells indicate combinations that were not observed. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

compensatory mechanism for the lack of D segments. The TRG repertoire had one V and J combination that comprised 47.7% of the repertoire, which is the highest frequency of a single V-J pair in any of the chains. The lower combinatorial diversity is consistent with the temporal/spatial segregation of γδ T cells seen in other species (Asarnow et al., 1988; Triebel et al., 1988). However, there is still enough expressed diversity from the *T. m. latirostris* circulating T cells from PBLs to identify above average V segment numbers in lieu of genomic locus assembly.

The TRD and TRG CDR3 length inequality is more similar to the Ig heavy and light chain, whereas the CDR3s of TRA and TRB are more equal, which is similar to the patterns seen in human and mouse (Rock et al., 1994). This may indicate that recognition of both self classical-MHC and foreign peptide restrains the TRB CDR3 length to maintain close contact of all CDRs to the MHC-peptide complex. The γδ T cells can recognize various epitopes like haptens and non-classical MHC that do not present antigen at all, such as MICA and T22 (Wingren et al., 2000; Zeng et al., 2014). The γδ T cells are able to recognize ligands in various ways that rely on different CDRs and are not limited to maintain one configuration (Xu et al., 2011). The conservation of these antigen recognition patterns indicates the maintenance of specific immunological roles for these two heterodimer receptors across evolution.

4.2. Influence of genomic organization on evolution

Not only do the gene segments of each chain have specific features that align with their individual niches, but the genomic organization of each chain has also influenced the way the loci have evolved over the mammalian timeline that correlates to their individual functions. The antigen receptor loci create V segment diversity through “birth-and-death” evolution resulting from extensive duplication events (Su and Nei, 2001). Duplications of V segment regions termed homology units have been identified in the TR loci of human, mouse, and cattle (Connelley et al., 2014; Glusman et al., 2001; Rowen et al., 1996). However, homology units between species have not been identified previously. By comparing human and *T. m. latirostris* TR loci, we identified conserved synteny blocks of certain V subgroups between the two species. Due to the *T. m. latirostris* position in eutherian evolution, these synteny blocks can be considered a characteristic of the ancestral eutherian progenitor.

The most conserved TR locus in sequence similarity and gene order is the TRAD locus. Even among species that encounter vastly different pathogens such as human and *T. m. latirostris*, the majority of their V segments have at least 75% identity with the other species. This level of sequence conservation could be maintained because T cells tend to recognize internal epitopes that are more conserved across strains of viruses or species of bacteria than the external epitopes recognized by antibodies (Wahl et al., 2009). Not only does the TRAD locus have more V segments that are within synteny blocks, but there are also more V segments with interspecies subgroup identity outside those blocks than the TRB locus. Therefore, the TRAD locus is more maintained as a whole than the TRB locus. The TRAD locus has a complex structure that is maintained at the 5′ end by the olfactory genes and the 3' end by the nested TRD locus. There are also complex regulatory features such as promotors and chromatin-folding motifs that orchestrate the exclusive and shared V segment rearrangement between the two constant regions (Hawwari and Krangel, 2005). The TRAV segments also have more exclusivity of use for certain T cell subtypes. For example, there is one invariant TRAV-TRAJ combination used for the iNKT T cell, where the TRB chain is not invariant, but limited (Porcelli et al., 1993). There is also exclusive use of the TRAV1 segment in mucosal associated invariant T (MAIT) cells, while the TRB chain is oligoclonal (Reantragoon et al., 2013). The need to maintain this complex organization and specific functions may explain the observation of long-range duplications that contain several synteny blocks in the TRA locus versus the short-range duplications of individual synteny blocks seen in the TRB locus.

The TRB locus is less complex and has less locus synteny than the TRAD locus. There are fewer synteny blocks in the TRB locus, and two of the three blocks only contain two V subgroups. The longest block, SBβ3, is also more fragmented and less consistently maintained than the longest TRA block, SBα5. There is also less locus position synteny than the TRAD locus. The overall TRB locus is less complex than the TRAD locus, but there is some maintenance of locus structure across species concerning the D-J-C clusters. These clusters have been identified in most studied mammals and avians, and there is evidence of multiple TRBC regions in teleosts

(Kamper and McKinney, 2002; Yang et al., 2017). Therefore, it is not surprising that the greatest extent of V segment subgroup conservation is seen in the more 3' end of the locus closest to the D-J-C clusters.

It can be surmised that the structure of the TRADV locus is more important to maintain than the TRB and TRG loci, and therefore has less freedom to diverge. This could be due to the complex regulation of V segment expression between the TRA and TRD constant regions, the more specific use of certain TRAV segments for specific functions, or a lesser contribution of the TRAV segment in MHC recognition, which puts less selective strain on the locus to co-evolve with MHC. The TRB locus is less constrained to maintain a complex locus structure, and can therefore diverge and select for TRBV segments that complement the MHC alleles that are specific to each species. This may be the reason the TRB is more involved in MHC binding than TRA (Luz et al., 2002). This coevolution could also explain the increased proportion of pseudogenes observed in the TRB locus in all mammals (Parra et al., 2008).

The TRG locus is the least conserved in subgroup identity and in locus organization. Even amongst mammals, there are variations in numbers of constant regions, numbers of distinct loci, and the way regions cluster (no clusters, J-C clusters, or V-J-C clusters). The degree of species specificity could indicate limited functional conservation between species. There is still very little that is understood about γδ T cells, even in human and mouse. There is even less understood about the similarities and differences between γδ T cell roles in γδ high (cattle and sheep) vs γδ low species (human and mouse). Additional functional data is needed between species to understand the consequence of minimizing locus organization maintenance across species.

Although the TRG locus is not yet assembled, we can make some assumptions based on the expressed transcripts. While there are no strong expression patterns between The TRGV segments and the TRGC regions, there are patterns between the TRGJ segments and TRGC regions. This hints at a multi-cluster organization that has several copies of the same V segment sequence in different clusters. This is also seen in the high TRGV/J combinatorial diversity in *T. m. latirostris*, which is typically restricted in other species by within-cluster diversity with few inter-cluster rearrangements (Pereira and Boucontet, 2004). Each TRGJ segment has one TRGC region that dominates its expression, but it is not an exact 1:1 ratio. Both TRGC4 and TRGC3.2 have exclusive J-C pairings, but TRGC1, 2, and 3.1 do not. It is unclear whether each J-C pairing is its own cluster or if there are multiple TRGJ segments per cluster. However, multiple clusters using the same Ig constant domain could explain the discrepancies between the constant domain and connecting peptide combinations seen in the expressed transcripts. Having multiple clusters with similar gene segments could explain why this locus did not assemble well by Illumina short read assembly.

We attempted to determine phylogenetic relationships between the V segments of mammalian species, similar to the clan system used in the Ig genes. Previous papers have used this analysis to determine V segment groups shared between species. However, the number of groups changes depending on the number of species included in the analysis (Connelley et al., 2014; Su et al., 1999). The most thorough of these studies defined groups for each chain by including *Monodelphis domesticus* V segments with human, mouse, cow, sheep, chicken, and rabbit (Parra et al., 2008). Our attempts to recreate the phylogenetic groups with the addition of the *T. m. latirostris* V segments were unsuccessful; we observed inconsistent tree topography and low bootstrap values to support these groups (data not shown). We also attempted to use phylogenetic relationships to support our analysis of subgroup conservation by 75% nucleotide identity (Figs. S11–13). While conservation patterns were supported in most cases, some nodes were unresolved or had low bootstrap values. The inconsistency of strong phylogenetic relationships for the TR genes could be due to higher divergence in sequence than we see for the Ig genes. For example, the human IgHV are separated into 10 subgroups, while the human TRADV are separated into 44 subgroups. Therefore, we relied on nucleotide identity and locus synteny to determine evolutionary relationships over the phylogenetic trees.

4.3. VHD in jawed vertebrates

While it is not functional, this is the first time that a VHD remnant has been identified in a eutherian mammal. These IG-like V segments that are expressed with the TRD constant region are seen throughout jawed vertebrates, but are missing in some evolutionary radiations. They are present in shark, they seem to have been lost in teleosts, but are present in amphibians (Criscitiello et al., 2010; Parra et al., 2010; Seelye et al., 2016). In birds, some species (zebra finch) have a VHD in their TRAD locus, while some (chicken and turkey) have a VHD in a TRD locus separate from the TRAD locus (Parra et al., 2012b). They are present in some marsupials (present in platypus, but not in opossum), which also have a separate TR chain (TRM) that rearranges a IG-like V segment and a fused IG-like V-J segment to a TRD-like constant region (Parra et al. 2008, 2012a). A functional VHD or evidence of a TRM locus has yet to be identified in a eutherian mammal. The VHD segment seems vulnerable to translocation and loss, which may be a result of being located in the dynamic region between the pool of TRADV segments and the often-deleted nested TRD locus. The identification of this VHD remnant leads to the possibility that this V segment may be present as a functional component in an afrotherian or a xenarthran that has yet to be studied.

4.4. Conclusions

Comparative analysis of these complex loci is essential for understanding the evolutionary forces that have shaped and maintained them to fulfill specific functions. If we just look at human and mouse, we only get a snapshot of the current state of these loci, but nothing of the evolutionary process to get there. Mouse is not an ideal comparison to human because there is very little subgroup identity between the two species. Cattle have maintained more sequence similarity, but their TR loci have expanded greatly, which makes genomic comparison more difficult to perform. Cattle are also within Boreoeutheria, so they are less informative of evolutionary history. Marsupials and eutherians share almost no subgroup identity, which also makes them difficult to compare. However, *T. m. latirostris* is a part of Afrotheria, one of the more basal clades of eutherian mammals. They also have maintained similar number of functional V segments, similar number of subgroups, and sequence similarity to human, which provided the opportunity to compare subgroup order. This revealed locus synteny blocks that had not been identified before. With this new information, we can predict ancestral patterns for all eutherians that could be informative of features that are important for maintaining individual chain function in earlier vertebrates that do not share the same locus organization.

Acknowledgements

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 https://doi.org/10.1016/j.dci.2018.04.007.

References

Asarnow, D.M., Kuziel, W.A., Bonyhadi, M., Tigelaar, R.E., Tucker, P.W., Allison, J.P., 1988. Limited diversity of γδ antigen receptor genes of thy-1+ dendritic epidermal cells. Cell 55, 837–847.

Connelley, T.K., Degnan, K., Longhi, C.W., Morrison, W.I., 2014. Genomic analysis offers insights into the evolution of the bovine tra/trd locus. BMC Genom. 15 (994), 1471–2164.

Criscitiello, M.F., Ohta, Y., Saltis, M., McKinney, E.C., Flajnik, M.F., 2010. Evolutionarily conserved tcr binding sites, identification of t cells in primary lymphoid tissues, and surprising trans-rearrangements in nurse shark. J. Immunol. 184 (12), 6950–6960.

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

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. Lond. B Biol. Sci. 371 (1699).

Glusman, G., Rowen, L., Inyoul, L., Boysen, C., Roach, J.C., Smit, A.F.A., Wang, K., Koop, B.F., Hood, L., 2001. Comparative genomics of the human and mouse t cell receptor loci. Immunity 15, 337–349.

Hawwari, A., Krangel, M.S., 2005. Regulation of tcr delta and alpha repertoires by local and long-distance control of variable gene segment chromatin structure. J. Exp. Med. 202 (4), 467–472.

Jorgensen, J.L., Esser, U., Fazekas de St. Groth, B., Reay, P.A., Davis, M.M., 1992. Mapping t-cell receptor-peptide contacts by variant peptide immunization of single-chain transgenics. Nature 355, 224–230.

Kamper, S.M., McKinney, C.E., 2002. Polymorphism and evolution in the constant region of the t-cell receptor beta chain in an advanced teleost fish. Immunogenetics 53 (12), 1047–1054.

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.

Luz, J.G., Huang, M., Garcia, K.C., Rudolph, M.G., Apostolopoulos, V., Teyton, L., Wilson, I.A., 2002. Structural comparison of allogeneic and syngeneic t cell receptor–peptide-major histocompatibility complex complexes. J. Exp. Med. 195 (9), 1175–1186.

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.

Parra, Z.E., Baker, M.L., Hathaway, J., Lopez, A.M., Trujillo, J., Sharp, A., Miller, R.D., 2008. Comparative genomic analysis and evolution of the t cell receptor loci in the opossum monodelphis domestica. BMC Genom. 9, 111.

Parra, Z.E., Lillie, M., Miller, R.D., 2012a. A model for the evolution of the mammalian t-cell receptor alpha/delta and mu loci based on evidence from the duckbill platypus. Mol. Biol. Evol. 29 (10), 3205–3214.

Parra, Z.E., Miller, R.D., 2012. Comparative analysis of the chicken tcr alpha/delta locus. Immunogenetics 64, 641–645.

Parra, Z.E., Mitchell, K., Dalloul, R.A., Miller, R.D., 2012b. A second tcrdelta locus in galliformes uses antibody-like v domains: insight into the evolution of tcrdelta and tcrmu genes in tetrapods. J. Immunol. 188 (8), 3912–3919.

Parra, Z.E., Ohta, Y., Criscitiello, M.F., Flajnik, M.F., Miller, R.D., 2010. The dynamic tcrdelta: tcrdelta chains in the amphibian xenopus tropicalis utilize antibody-like v genes. Eur. J. Immunol. 40 (8), 2319–2329.

Pereira, P., Boucontet, L., 2004. Rates of recombination and chain pair biases greatly influence the primary γδ tcr repertoire in the thymus of adult mice. J. Immunol. 173 (5), 3261–3270.

Piccinni, B., Massari, S., Caputi Jambrenghi, A., Giannico, F., Lefranc, M.P., Ciccarese, S., Antonacci, R., 2015. Sheep (ovis aries) t cell receptor alpha (tra) and delta (trd) genes and genomic organization of the tra/trd locus. BMC Genom. 16, 709.

Porcelli, S., Yockey, C.E., Brenner, M.B., Balk, S.P., 1993. Analysis of t cell anitgen receptor (tcr) expression by human peripheral blood cd4-8- a/b t cells demonstrates preferential use of several vb genes and an invariant tcr a chain. J. Exp. Med. 178, 1–16.

Reantragoon, R., Corbett, A.J., Sakala, I.G., Gherardin, N.A., Furness, J.B., Chen, Z., Eckle, S.B., Uldrich, A.P., Birkinshaw, R.W., Patel, O., Kostenko, L., Meehan, B., Kedzierska, K., Liu, L., Fairlie, D.P., Hansen, T.H., Godfrey, D.I., Rossjohn, J., McCluskey, J., Kjer-Nielsen, L., 2013. Antigen-loaded mr1 tetramers define t cell receptor heterogeneity in mucosal-associated invariant t cells. J. Exp. Med. 210 (11), 2305–2320.

Rock, E.P., Sibbald, P.R., Davis, M.M., Chien, Y.-h., 1994. Cdr3 length in antigen-specific immune receptors. J. Exp. Med. 179, 323–328.

Rowen, L., Koop, B.F., Hood, L., 1996. The complete 658-kilobase DNA sequence of the human beta t cell receptor locus. Science 272, 1755–1762.

Seelye, S.L., Chen, P.L., Deiss, T.C., Criscitiello, M.F., 2016. Genomic organization of the zebrafish (danio rerio) t cell receptor alpha/delta locus and analysis of expressed products. Immunogenetics 68 (5), 365–379.

Su, C., Jakobsen, I., Gu, X., Nei, M., 1999. Diversity and evolution of t-cell receptor variable region genes in mammals and birds. Immunogenetics 50, 301–308.

Su, C., Nei, M., 2001. Evolutionary dynamics of the t-cell receptor vb gene family as inferred from the human and mouse genomic sequences. Mol. Biol. Evol. 18 (4), 503–513.

Triebel, F., Faure, F., Graziani, M., Jitsukawa, S., Lefranc, M.P., Hercend, T., 1988. A unique v-j-c-rearranged gene encodes a gamma protein expressed on the majority of cd3+ t cell receptor-a/b- circulating lymphocytes. J. Exp. Med. 167, 694–699.

Tucker, K.P., Hunter, M.E., Bonde, R.K., Austin, J.D., Clark, A.M., Beck, C.A., McGuire, P.M., Oli, M.K., 2012. Low genetic diversity and minimal population substructure in the endangered Florida manatee: implications for conservation. J. Mammal. 93 (6), 1504–1511.

Uenishi, H., Eguchi-Ogawa, T., Toki, D., Morozumi, T., Tanaka-Matsuda, M., Shinkai, H., Yamamoto, R., Takagaki, Y., 2009. Genomic sequence encoding diversity segments of the pig tcr delta chain gene demonstrates productivity of highly diversified repertoire. Mol. Immunol. 46 (6), 1212–1221.

Vaccarelli, G., Miccoli, M.C., Antonacci, R., Pesole, G., Ciccarese, S., 2008. Genomic organization and recombinational unit duplication-driven evolution of ovine and bovine t cell receptor gamma loci. BMC Genom. 9, 81.

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.

Wahl, A., Schafer, F., Bardet, W., Buchli, R., Air, G.M., Hildebrand, W.H., 2009. Hla class i molecules consistently present internal influenza epitopes. Proc. Natl. Acad. Sci. U.S.A. 106 (2), 540–545.

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.

Wingren, C., Crowley, M.P., Degano, M., Chien, Y.-h., Wilson, I.A., 2000. Crystal structure of a yd t cell receptor ligand t22: a truncated mhc-like fold. Science 287, 310–314.

Xu, B., Pizarro, J.C., Holmes, M.A., McBeth, C., Groh, V., Spies, T., Strong, R.K., 2011. Crystal structure of a yd t-cell receptor specific for the human mhc class i homolog mica. Proc. Natl. Acad. Sci. Unit. States Am. 108 (6), 2414–2419.

Yang, Z., Sun, Y., Ma, Y., Li, Z., Zhao, Y., Ren, L., Han, H., Jiang, Y., Zhao, Y., 2017. A comprehensive analysis of the germline and expressed tcr repertoire in white peking duck. Sci. Rep. 7, 41426.

Zeng, X., Meyer, C., Huang, J., Newell, E.W., Kidd, B.A., Wei, Y.L., Chien, Y.H., 2014. Gamma delta t cells recognize haptens and mount a hapten-specific response. Elife 3 e03609.

Table S1. PCR primers and conditions

	Forward Primer	Reverse Primer	Polymerase	Initial	Denature	Annealing	Elongation	# Cycles	Final
TRA 5' RACE	GGACACTGACATGGACTGAAGGAGTA	ATCGGTGAACAGGCAGACAG	Hot Start	95°C for 15:00	95°C for 0:30	62.5°C for 0:30	72°C for 1:00	35	72°C for 5:00
TRB 5' RACE	GGACACTGACATGGACTGAAGGAGTA	TCGGGAGATCTCGACCTCTG	Hot Start	95°C for 15:00	95°C for 0:30	55°C for 0:30	72°C for 1:00	35	72°C for 5:00
TRD 5' RACE	GGACACTGACATGGACTGAAGGAGTA	ACCAGACAAGCAACCTTCGT	Hot Start	95°C for 15:00	95°C for 0:30	64°C for 0:30	72°C for 1:00	35	72°C for 5:00
TRG 5' RACE	GGACACTGACATGGACTGAAGGAGTA	AGAATCCCACGATGGCCAAG	Hot Start	95°C for 15:00	95°C for 0:30	68°C for 0:30	72°C for 1:00	35	72°C for 5:00

Table S2. Gene segment genomic scaffold positions

Scaffold	Segment	Start	End	Orientation	Functional?
scaffold00273	TRAV1-1	436,064	435,783	-	Y
	TRAV2-1	319,645	319,358	-	Y
	TRAV3-1	267,548	267,255	-	Y
	TRAV4-1	237,637	237,344	-	Y
	TRAV5-1	197,933	197,675	-	N
	TRAV6-1	162,770	162,480	-	Y
	TRAV7-1	144,995	144,718	-	N
	TRAV8-1	133,719	133,423	-	N
	TRAV9-1	127,407	127,129	-	Y
	TRAV10-1	114,231	113,950	-	Y
	TRAV11-1	85,387	85,113	-	N
	TRAV12-1	66,686	66,405	-	Y
	TRAV13-1	49,002	48,774	-	N
	TRAV14-1	23,396	23,125	-	N
	TRAV15-1	9,767	9,474	-	Y
scaffold00745	TRAV16-1	15,149	14,865	-	Y
	TRAV17-1	10,884	10,608	-	N
	TRAV18-1	3,347	3,060	-	N
scaffold01496	TRAV15-2	1,343	1,050	-	Y
scaffold03036	TRAV19-1	342	608	+	Y
contig164553	TRAV20-1	414	118	-	Y
scaffol00268	TRAV21-1	2,317,770	2,317,536	-	N
	TRAV22-1	2,299,921	2,299,640	-	Y
	TRAV23-1	2,299,380	2,299,111	-	N
	TRAV24-1	2,288,291	2,287,986	-	Y
	TRAV25-1	2,277,360	2,277,082	-	N
	TRAV26-1	2,271,605	2,271,312	-	Y
	TRAV20-2	2,263,326	2,263,051	-	N
	TRAV27-1	2,254,688	2,254,404	-	Y
	TRAV28-1	2,219,757	2,219,524	-	N
	TRAV29-1	2,211,584	2,211,337	-	N
	TRDV1-1	2,203,033	2,202,737	-	N
	TRAV30-1	2,172,873	2,172,583	-	Y
	TRAV16-2	2,151,173	2,150,948	-	N
	TRAV17-2	2,145,252	2,145,054	-	N
	TRAV6-2	2,130,161	2,129,871	-	Y
	TRAV7-2	2,113,982	2,113,701	-	N
	TRAV8-2	2,103,568	2,103,275	-	Y
	TRAV9-2	2,097,211	2,096,995	-	N
	TRAV10-2	2,084,258	2,083,977	-	Y
	TRAV3-2	2,052,225	2,051,932	-	Y
	TRAV11-2	2,037,004	2,036,735	-	N
	TRAV31-1	2,024,658	2,024,368	-	Y
	TRAV13-2	2,016,102	2,015,825	-	N
	TRAV32-1	2,002,562	2,002,318	-	N
	TRAV14-2	1,991,731	1,991,459	-	N

	TRAV15-3	1,981,268	1,980,975	-	Y
	TRAV33-1	1,951,265	1,950,984	-	Y
	TRAV4-2	1,942,104	1,941,805	-	N
	TRAV34-1	1,931,186	1,930,898	-	Y
	TRAV33-2	1,921,285	1,921,004	-	Y
	TRAV24-2	1,902,316	1,902,011	-	Y
	TRAV25-2	1,889,480	1,889,181	-	Y
	TRAV20-3	1,873,415	1,873,122	-	Y
	TRAV27-2	1,865,941	1,865,654	-	Y
	TRAV29-2	1,827,835	1,827,588	-	N
	TRDV1-2	1,817,254	1,816,946	-	Y
	TRAV35-1	1,806,956	1,806,691	-	N
	TRAV7-3	1,773,438	1,773,153	-	N
	TRAV8-3	1,762,743	1,762,450	-	N
	TRAV9-3	1,745,587	1,745,306	-	Y
	TRAV10-3	1,731,740	1,731,459	-	Y
	TRAV36-1	1,714,489	1,714,190	-	Y
	TRAV32-2	1,702,289	1,702,011	-	Y
	TRAV14-3	1,684,091	1,683,783	-	Y
	TRAV15-4	1,674,302	1,674,012	-	Y
	TRAV34-2	1,662,024	1,661,737	-	N
	TRAV33-3	1,653,670	1,653,392	-	N
	TRAV4-3	1,641,359	1,641,066	-	N
	TRAV19-2	1,624,815	1,624,540	-	N
	TRAV22-2	1,607,618	1,607,360	-	N
	TRAV24-3	1,589,365	1,589,092	-	N
	TRAV25-3	1,578,518	1,578,225	-	N
	TRAV26-2	1,572,756	1,572,473	-	N
	TRAV20-4	1,548,773	1,584,477	-	Y
	TRAV27-3	1,539,873	1,539,586	-	Y
	TRAV37-1	1,520,072	1,519,823	-	N
	TRAV28-2	1,510,869	1,510,669	-	N
	TRAV29-3	1,505,720	1,505,487	-	N
	TRDV1-3	1,501,882	1,501,634	-	N
	TRAV35-2	1,494,174	1,493,936	-	N
	TRAV30-2	1,417,524	1,417,291	-	N
	TRAV5-2	1,413,103	1,412,825	-	Y
	TRAV7-4	1,402,111	1,401,821	-	Y
	TRAV8-4	1,383,303	1,383,010	-	N
	TRAV9-4	1,376,539	1,376,338	-	N
	TRAV10-4	1,344,831	1,344,550	-	Y
	TRAV36-2	1,336,845	1,336,546	-	N
	TRAV34-3	1,329,959	1,329,714	-	N
	TRAV33-4	1,315,303	1,315,022	-	N
	TRAV4-4	1,304,404	1,304,145	-	N
	TRAV19-3	1,295,933	1,295,653	-	N
	TRAV24-4	1,275,935	1,275,645	-	Y

	TRAV38-1	1,260,493	1,260,161	-	N
	TRAV25-4	1,248,895	1,248,710	-	N
	TRAV26-3	1,243,192	1,242,959	-	N
	TRAV20-5	1,235,379	1,235,083	-	N
	TRAV27-4	1,227,905	1,227,621	-	N
	TRAV37-2	1,185,891	1,185,672	-	N
	TRAV28-3	1,176,214	1,176,026	-	N
	TRAV29-4	1,152,475	1,152,216	-	N
	TRAV39-1	1,136,282	1,136,058	-	N
	TRDV1-4	1,130,666	1,130,358	-	Y
	TRAV40-1	1,110,141	1,109,851	-	Y
	TRAV41-1	1,106,634	1,106,335	-	Y
	TRAV41-2	1,084,726	1,064,427	-	Y
	TRAV42-1	1,039,321	1,039,040	-	Y
	TRDV2-1	956,587	956,294	-	N
	TRDV3-1	937,894	937,617	-	N
	TRDV4-1	932,785	932,496	-	N
	TRD-VH	922,199	922,065	-	N
	TRDV5-1	876,808	877,113	+	Y
	TRDD1	899,942	899,929	-	Y
	TRDJ1	898,365	898,321	-	Y
	TRAJ50	869,443	869,390	-	Y
	TRAJ49	866,678	866,619	-	Y
	TRAJ48	866,027	865,968	-	Y
	TRAJ47	863,213	863,151	-	Y
	TRAJ46	862,555	862,493	-	Y
	TRAJ45	859,331	859,266	-	Y
	TRAJ44	856,368	856,312	-	Y
	TRAJ43	855,475	855,419	-	Y
	TRAJ42	852,704	852,651	-	Y
	TRAJ41	851,629	851,567	-	Y
	TRAJ40	850,759	850,700	-	Y
	TRAJ39	849,659	849,606	-	Y
	TRAJ38	848,588	848,526	-	Y
	TRAJ37	846,351	846,292	-	Y
	TRAJ36	844,422	844,363	-	Y
	TRAJ35	843,769	843,710	-	Y
	TRAJ34	842,855	842,814	-	N
	TRAJ33	842,076	842,020	-	Y
	TRAJ32	841,430	841,371	-	Y
	TRAJ31	839,920	839,861	-	Y
	TRAJ30	838,544	838,488	-	Y
	TRAJ29	837,852	837,799	-	Y
	TRAJ28	836,817	836,758	-	Y

	TRAJ27	835,213	835,166	-	Y
	TRAJ26	833,375	833,331	-	Y
	TRAJ25	832,295	832,239	-	Y
	TRAJ24	831,265	831,200	-	Y
	TRAJ23	830,614	830,558	-	Y
	TRAJ22	827,931	827,875	-	Y
	TRAJ21	826,633	826,574	-	Y
	TRAJ20	825,315	825,256	-	Y
	TRAJ19	823,494	823,438	-	Y
	TRAJ18	822,774	822,721	-	Y
	TRAJ17	821,451	821,389	-	Y
	TRAJ16	820,241	820,182	-	Y
	TRAJ15	818,872	818,813	-	Y
	TRAJ14	817,787	817,731	-	Y
	TRAJ13	817,097	817,047	-	Y
	TRAJ12	815,377	815,321	-	Y
	TRAJ11	814,804	814,748	-	Y
	TRAJ10	813,804	813,742	-	Y
	TRAJ9	811,913	811,854	-	Y
	TRAJ8	811,332	811,276	-	Y
	TRAJ7	809,845	809,789	-	Y
	TRAJ6	807,475	807,416	-	Y
	TRAJ5	805,749	805,693	-	N
	TRAJ4	803,411	803,352	-	Y
	TRAJ3	802,174	802,115	-	Y
	TRAJ2	801,614	801,552	-	Y
	TRAJ1	800,650	800,591	-	Y
	TRDC	885,022	884,744	-	Y
	TRAC	798,765	798,508	-	Y
scaffold00093	TRBV1-1	6,212,037	6,211,744	-	Y
	TRBV2-1	6,208,417	6,208,127	-	Y
	TRBV3-1	6,200,615	6,200,346	-	N
	TRBV4-1	6,191,227	6,190,934	-	Y
	TRBV5-1	6,189,465	6,189,172	-	Y
	TRBV6-1	6,187,568	6,187,305	-	N
	TRBV4-2	6,180,328	6,180,073	-	N
	TRBV5-2	6,178,550	6,178,326	-	N
	TRBV6-2	6,177,191	6,176,880	-	Y
	TRBV7-1	6,175,175	6,174,930	-	N
	TRBV4-3	6,170,638	6,170,345	-	Y
	TRBV5-3	6,168,892	6,168,599	-	Y
	TRBV6-3	6,167,312	6,167,026	-	N
	TRBV4-4	6,142,721	6,142,428	-	Y
	TRBV5-4	6,140,978	6,140,686	-	Y
	TRBV6-4	6,139,509	6,139,259	-	N

	TRBV4-5	6,133,285	6,132,992	-	Y
	TRBV5-5	6,129,197	6,128,949	-	N
	TRBV8-1	6,124,233	6,123,940	-	Y
	TRBV5-6	6,119,969	6,119,697	-	N
	TRBV9-1	6,118,602	6,118,306	-	Y
	TRBV5-7	6,114,835	6,114,545	-	Y
	TRBV10-1	6,106,330	6,106,056	-	N
	TRBV11-1	6,089,649	6,09,401	-	N
	TRBV12-1	6,086,734	6,086,441	-	Y
	TRBV13-1	6,082,140	6,081,844	-	N
	TRBV14-1	6,070,168	6,069,875	-	Y
	TRBV15-1	6,068,011	6,067,713	-	N
	TRBV16-1	6,050,272	6,049,976	-	Y
	TRBV16-2	6,046,863	6,046,585	-	N
	TRBV8-2	6,042,373	6,042,083	-	Y
	TRBV5-8	6,038,890	6,038,615	-	Y
	TRBV9-2	6,037,610	6,037,317	-	N
	TRBV5-9	6,024,752	6,024,460	-	N
	TRBV10-2	6,017,119	6,016,823	-	Y
	TRBV14-2	6,010,692	6,010,465	-	N
	TRBV15-2	6,008,763	6,008,461	-	Y
	TRBV8-3	5,999,947	5,999,628	-	N
	TRBV5-10	5,996,425	5,996,147	-	N
	TRBV12-2	5,994,271	5,993,994	-	N
	TRBV15-3	5,990,409	5,990,107	-	N
	TRBV16-3	5,982,742	5,982,466	-	N
	TRBV8-4	5,971,664	5,971,371	-	Y
	TRBV5-11	5,967,400	5,967,107	-	Y
	TRBV11-2	5,964,590	5,964,339	-	N
	TRBV12-3	5,961,594	5,961,302	-	Y
	TRBV13-2	5,956,985	5,956,689	-	Y
	TRBV17-1	5,953,394	5,953,157	-	N
	TRBV14-3	5,952,330	5,952,037	-	Y
	TRBV15-4	5,950,185	5,949,883	-	Y
	TRBV8-5	5,934,379	5,934,086	-	Y
	TRBV5-12	5,930,850	5,930,576	-	N
	TRBV12-4	5,928,702	5,928,419	-	N
	TRBV13-3	5,924,090	5,923,794	-	Y
	TRBV17-2	5,920,496	5,920,271	-	N
	TRBV14-4	5,919,441	5,919,148	-	Y
	TRBV15-5	5,909,120	5,908,818	-	Y
	TRBV16-4	5,888,840	5,888,544	-	Y
	TRBV16-5	5,884,884	5,884,588	-	Y
	TRBV16-6	5,875,470	5,875,173	-	N
	TRBV16-7	5,868,457	5,868,160	-	N
	TRBV16-8	5,848,415	5,848,119	-	Y
	TRBV18-1	5,842,961	5,842,668	-	Y

	TRBV19-1	5,840,234	5,839,941	-	Y
	TRBV20-1	5,833,787	5,833,494	-	Y
	TRBV21-1	5,830,281	5,829,988	-	Y
	TRBV22-1	5,828,841	5,828,526	-	N
	TRBV23-1	5,823,451	5,823,152	-	Y
	TRBV23-2	5,819,710	5,819,415	-	Y
	TRBV24-1	5,778,830	5,779,120	+	Y
	TRBD1	5,804,484	5,804,472	-	Y
	TRBD2	5,795,392	5,795,378	-	Y
	TRBJ1-1	5,803,865	5,803,821	-	Y
	TRBJ1-2	5,803,751	5,803,707	-	Y
	TRBJ1-3	5,803,245	5,803,198	-	Y
	TRBJ1-4	5,802,702	5,802,655	-	Y
	TRBJ1-5	5,802,029	5,801,985	-	Y
	TRBJ1-6	5,801,585	5,801,535	-	Y
	TRBJ2-1	5,794,499	5,794,458	-	Y
	TRBJ2-2	5,794,055	5,794,008	-	Y
	TRBJ2-3	5,793,871	5,793,827	-	Y
	TRBJ2-4	5,793,554	5,793,510	-	Y
	TRBC1	5,799,118	5,798,735	-	Y
	TRBC2	5,790,179	5,789,796	-	Y
scaffold00114	TRGV1-1	36,786	36,484	-	Y
	TRGV2-1	4,343	4,057	-	N
contig164298	TRGV1-2	359	670	+	Y
scaffold01250	TRGJ1	5,038	4,985	-	Y
scaffold01597	TRGJ5	5,570	5,620	+	Y
scaffold03157	TRGTM1	1,007	1,183	+	Y

Subgroup	A/D	V	-------A------> (1-15)	-----B----> (16-26)	CDR1 (27-38)	---C---> (39-46)	---C'---> (47-55)	CDR2 (56-65)	---C"---> (66-74)	----D----> (75-84)	-----E-----> (85-96)	---F---> (97-104)	CDR3
TRAV1	+/-	1-1	..GTV	GQDIQQ.	PTELTAME	GNFVQVNCTYQ	TSG.....FNG	LFWYQQSD	GGAPTFLSY	NVL....DGL	EKR.....G	HFSSFLSLSS	ASSYLLLTELQM NDSARYLC AVR.
TRAV2	+/-	2-1	..MGR	GEKVEQSPPLLSIQD	GESSFSNCTYT	DSA.....STY	FLWYKQEP	GRGPQLLIQ	ILGN...ADK	KQD.....Q	RFTVLLNKKE	KHFSLHIADPHP	GDSATYFC AVR.
TRAV3	+/-	3-1	..GTN	GDSVIQTEGLVTLSE	GAAMILNCTYQ	ISYTT....TPF	LFWYVQYL	NKAPQLLLK	SSTE...SQG	TDS.....R	GFQANHVKSN	SSFHLKKASVQT	SDSAVYYC ALS.
	+/-	3-2	..GTN	GDSVTQTEGLVTVSE	GASMNLNCNYR	TNYTT....SPL	LFWYVQYL	NKAPHLLLK	SSTE...SQG	TED.....Q	GFQANLVKSE	NSFHLQKRSLQI	SDSALYYC AMR.
TRAV4	+/-	4-1	.GEIR	AQSVTQPDVPVSVSE	GDSLELRCIYS	YSG.....TVY	LYWYVQYP	HKVPQLLLK	YFPG..DTLV	KGI.....K	GFEAEFRKNE	SSFNLRKHSAHW	NDSADYFC AVS.
TRAV5	+/-	5-2	..GVR	GVKVEQSPSALSLQE	GTNSTLRCNYS	TTT.....NN	VQWFRQTP	GGGLISLFF	IAS.....GT	KQN.....G	RLKSTINSKE	LYSTLHIATSQL	EDSATYLC AVE.
TRAV6	-/-	6-1	.GVSG	NNKVEQSSGSLIILK	GENCTLQCNYL	VSP.....FNN	LRRYKEDT	GRGPISLII	MKYS...ESK	KSN.....G	RYTVTLDATA	KCSFLHLPASQL	SDSASYIC AVG.
	+/-	6-2	.GASG	NNQVEQSPGSLIVLE	GENCTLQCNYS	VSS.....FNS	LRWYKEDT	GRGPVSLII	MTYS...ESK	KSN.....G	RYAVTLDATA	KFSFLHLTASQL	SDSASYIC AVG.
TRAV7	+/-	7-4	GTPSQ	QKKVEQSPESLSIPE	GTIASLNCTYE	DRA.....SQY	FMWYRQYS	GKGPELLMS	IYS....KDD	KKE.....G	RFTAQLNKGS	QYVSLLITDSQP	SDSATYLC VVS.
TRAV8	-/-	8-2	.GGTR	AQLVTQLDIHITLSE	RVLLELRCNYS	SFI.....LPN	FFWYVQYP	NRGFQILLK	YIVG..NTMI	KGI.....K	HFEAELKKSE	TPFHLRKPSVHW	SNSAESFC ATS.
TRAV9	+/-	9-1	..GTR	AQRVTQPEDHIFVFE	GAPVKIKCNYS	YSG.....SPV	LSWYVQYP	KKGLQLLLG	HI.....SG	DSI.....K	GFSADLDKGK	MTFHLKKPSAQE	EDSAMYYC AVS.
	-/-	9-3	.GGTR	AQRVTQPEDHIFVFE	GAPEQIKCNYS	YSG.....SPV	LSWFVQYP	KKGLHLLLG	HI.....SG	DSI.....K	GFSADLDKGK	MTFHLKKPSAQE	EDSAMYYC AVS.
TRAV10	+/-	10-1	...GV	NSQVEQNPEALTIQE	GENATMNCSYK	ASI.....NN	LQWYRQDS	GRGLVQLIL	IHSN...EKE	KTN.....G	RLRATLDTST	KSSFLFITASQA	ADTAAYFC AID.
	-/-	10-2	...GV	NSQVEQNPEALTIQE	GENATMNCSYK	TSI.....NY	LQWYRQDS	GRGLVLLIL	IRSN...EKK	KPN.....G	RLTATLDTST	TSSFLFITASQA	TDTASYFC AID.
	-/-	10-3	...GV	NSQVEQNPEALTILE	NEHATMNCSCN	TSI.....NN	LQWYRQDS	GRGVVLLIL	IHSN...EKE	TPN.....E	RLKATLDTST	KSSFLFITASQA	ADTAAYFC AID.
	-/-	10-4	...GV	NSQVEQNPETLPIQE	GENATKSCSYK	TSI.....DS	LQWYRQDS	GRGFVLLIL	IRSN...EKE	TSA.....G	RLRATLDINS	KRSFLFITTSQA	ADTAAYFC AID.
TRAV12	+/-	12-1	.GSTG	KDQVEQSPRALILQE	GDNVSLNCSYS	VSK.....FGG	LQWYKQDP	GNGPDLLFL	LYSV...GDE	EQK.....E	RLRGTLSK..	KESSLHITASKP	EDSATYLC AV..
TRAV14	+/-	14-3	KEKSD	REQVKQNHQSLTVQE	GGTSVLNCAYD	NSA.....FDY	FPWYRQYP	GKGLVLLIA	IRSV...ESK	KVE.....G	RFTVFFNKTG	KHSSVHIVASQP	GDSATYFC AAS.
TRAV15DV3	+/+	15-1	.GSIV	AQKVTQVNSATSAQE	GEAVTLDCLYE	TSET.....AYS	ILWYKQLP	SGKIIFLIR	QSSS..SQNA	KED.....	RYSINFQKED	KSISLIISTLQM	EDSAKYFC ALW.
	+/+	15-2	.GSIV	AQKVTQVKSATSAQE	GAAVTLDCLFE	TSAT.....VYS	IFWYKQLP	SGKIIFLIY	QSSS..GQNA	NKG.....	RYSIKFQAEN	KSISLIISTLQL	EDSAKYFC ALW.
	+/+	15-3	.GSIV	AQKVTQVKSATSTQE	GEAVTLDCLYE	TSAT.....VYY	IFWYKQLP	SGKIIFLIH	QSSS..SQNA	EEG.....	RYSINFQKED	KFISLIISTLQM	EDSAKYFC ALW.
	-/+	15-4	.GSIV	AQKVTQVNSATSAQE	GETVTLDCLYE	TSET.....VYV	ILWYKQLP	SGKIIFLIH	QSSS..GENA	KKG.....	RYSIKFQKEK	KSIGLIISTLQL	EDSAKYFC AL..
TRAV16	+/-	16-1	..VCL	GENVEQYPSTLSVLE	GSSSVINCTYS	DSG.....STY	FPWYKQEP	GKGPQLIID	IRSN...VDR	NYQ.....Q	RFTVLLDKKA	KHLSLYISAAQP	GDSAVYLC AA..
TRAV19	+/-	19-1	..GKS	TQQLEQSPQFLSIQE	GDNFTTHCNSS	SVF.....PT	FQWYRQRP	GEGPVLLLM	LMKG...GEV	KKQ.....K	RLTAHFREAK	KDSSLHITAAQP	EDAGLYLC AGA.
TRAV20DV4	+/+	20-1	.GCSM	AEKVNQAQITITMQE	GEAVTLGCMYE	TKWS.....SYG	LYWYKQPP	SGEMVFLIQ	QEY...SKPN	AKQ.....D	RYSVNFQKAN	KSINLTISSLQL	ADSAKYFC ALWD
	-/-	20-3	.GCSM	AEKVNQTQTTITMQK	GEAVTLGCMYE	TRWS.....SYA	LYWYKQPP	CGEMVFLIQ	QEYS..KLNA	KHD.....	HYSVNFQKTN	KSINRTISSLQL	PDSAKSFC ALW
	-/-	20-4	.GRSM	TEKVNQAQITITMQE	GEAVTLGCTYE	TRWS.....SHA	LYWYKQPP	SGEMVFLIQ	QEYS..KPNA	KQD.....	RYSVNFQKAN	KSIILTISSLQL	ADSAKYFC AVWD
TRAV22	-/-	22-1	.VIRV	EMNVEQSPRVLILQE	GRNSSMTCNVS	VSI.....TS	VQWFQQSP	GGHLIPLFY	IAS.....GM	QQK.....G	RLESTVNTNE	LYSHLHIRDAQP	GDSATYFC IVE.
TRAV24	+/-	24-1	PKKKD	DQQVKQSPPSRKTQE	GEISIFNCDYN	NDL.....FDY	FVWYKKSP	GKGPALLIS	IHSG...ADG	NEE.....G	RFTVFLNKSA	KHLSLHIAASQP	GDSATYFC AAS.
	+/-	24-2	PKEKD	DQQVKQNPPSLKTQE	GEISILNCDYN	NDL.....FAY	FVWYKKSP	AKGPAPLIS	VSPG...VDG	NEE.....G	RFTVFLNKSA	KHLSLRIAASQP	GDSATYFC AAS.
	-/-	24-4	.RKTM	TSRFKQSPPSLKTQE	VEISILNCDYN	NNL.....FDY	LVWHKKSP	AKGLALLIP	IRSG...VNR	NQV.....G	RLTVFLNKNT	KHLSLHIAGSQP	GDSATYFC AAS
TRAV25	-/-	25-2	.RSSM	SQRVIQSQMAISMQE	GKTVELDCSYE	TNFA.....YYV	LFWYKSLP	SGEMIFLIL	QDTD..TELI	TTQ.....D	RYFLSFQKAP	KTIKLIISTAQL	EESATYFC GLRQ
TRAV26	+/-	26-1	VVSGQ	KEDIVQSPSILNVWE	REMAVINCSYK	EEA.....LNY	FPWYKKEA	GKGLNLLIE	IRSN...VDR	NQD.....G	RFTVLLNKKA	KHVSLHISATQP	GDSALYLC AAS.
TRAV27	+/-	27-1	.GITG	DAKTTQ.PNSMESTE	GKPVNLPCNHS	TIGG.....SEY	IYWYRQTP	HQGPEYVIH	GMN.....NN	VTN.....E	MASLTIPTGR	KSSTLILPHVTL	RDTAVYYC MVR.
	+/-	27-2	.GVTG	DAKTTQ.PNSMESTE	GEPVNLPCNHS	AISG.....NEY	IYWYRQIP	HQGPEYLIH	GLN.....NN	QTN.....E	MASLTIPTGR	KSSTLILPHVTL	RDTAVYYC MVRD
	+/-	27-3	.GITG	DAKTTQ.PNSMESTE	GKRVNLLCNHS	TISG.....NEY	IYWYRQIP	HQGPQYVIH	GIN.....NN	VTN.....E	MASLTIPTGR	KTSTLILPHVTL	RDTAVYYC MVRD
TRAV30	+/-	30-1	.GGNN	GDTVTQTEGQVTLSE	GAFLTVNCSYE	TTG.....YPV	LFWYVQYS	HEGLQLLLK	ATKA...KET	GSN.....K	SFEATYNKET	SSFHLKKASVQE	SDSAVYYC VLG.
TRAV31	-/-	31-1	.GVSS	KQEVQQSPAALNVQE	GNSFILNCSYT	DSA.....VYF	IQWYRQDL	GKGLTPLLL	IQSN...QRE	QMS.....G	RLKVSLDKSA	RHSALYLAASQT	GDSATYLC AVR.
TRAV32	-/-	32-2	..GMR	GVQVEQSPLVLSLQE	GASCTLRCNFS	TSV.....SI	LQWFRQDP	GERLTTLFY	IPS.....GT	KQN.....G	RLNVTTGTKE	RHSSLYISSSQT	TDSAIYFC AVS.
TRAV33	-/-	33-1	..EGN	GQQVKQIPQFLLVQE	GENFTTFCNST	STF.....SN	LQWYKQKP	GGCPVLLMM	LVKA...GEV	KEQ.....K	RLTSQFGKTR	KDSSLHITATQA	ADVSTYFC AM..
	-/-	33-2	..EGN	GQQVKQIPQFLLVQE	GENFTTYCNST	STF.....NN	LRWYKQKP	GGCPVLLMM	LVKT...GEV	KEQ.....K	RLTAQFGKTR	KDSFLHITAAQA	ADVGTYFC AI..
TRAV34	+/-	34-1	.GVSS	KFNVEQSPQSLHIQE	GESTNFTCSFP	SSS.....FYA	LHWYSWEL	AKGPKVLFV	LLSN...GDE	QKE.....R	RIRATLNTTE	GYSYLYIKGSQP	EDSATYLC A...
TRAV36	+/-	36-1	.GSTM	VQEVTQDQPATSVVE	KEAVTLDCVYK	TQDT.....SYA	LFWYKQPP	SGEMVFLIH	QESY..KEQN	TRQ.....G	RYSLNFQKEA	KSISLTITASQP	GDSAVYFC ALRD
TRAV40	-/-	40-1	.GSSG	QLQVEQSLAALQVKE	GEGFTLNCSYK	DSV.....SDF	FQWFGQNP	GQGLTSLIQ	MSSN...VCN	KVS.....G	RVTARLNKGD	QHFSLHMKDSPL	LDSVIFLC AAT.
TRAV41DV5	+/-	41-1	.ASSA	AQMVTQSKQQKSVRE	TESVTLDCTYD	TKDS.....NYY	LFWYKQLP	RGKMIFIIL	QEAF..KQQN	ATE.....N	RFSVNFQKAA	KSFSLKIADSQL	EDAAMYFC ALSG
	+/+	41-2	.GYST	AQTVTQSQRQMSIRE	TESVTLDCTYD	TKDS.....NYY	LFWYKQFP	RGGMIFIIR	QDGF..KQQS	ATK.....N	RFSVNFRKAA	KSFSLRISDSQL	EDTAMYFC VYSS
TRAV42	+/-	42-1	.GISG	KNEVEQSPPYLSVQE	GELVTINCSYS	TGM.....TT	LQWLQKNP	GGGIVSLFI	LSL.....EM	KKK.....E	RLSATINTEE	RHSSLHITASQP	RDSATYLC AVE.

Figure S1. TRAV IMGT Numbering. IMGT protein display of the T. m. latirostris TRADV segments. The strand and loops are defined by the IMGT numbering system for a V-DOMAIN and are indicated above the sequences. Conserved cysteines are in pink, the conserved tryptophan (W) and hydrophobic amino acid motifs are in blue, and the N-linked glycosylation sites are in green. Framework regions are in black and the CDR regions are in red. The first column indicates the T. m. latirostris TRADV subgroup each V segment belongs to, the second column indicates expression with the TRA/TRD constant region, and the third column indicates the individual V segment number.

Subgroup	EXP	V	-------A------>	-----B---->	CDR1	---C--->	---C'--->	CDR2	---C"--->	----D---->	-----E----->	---F--->	CDR3
			(1-15)	(16-26)	(27-38)	(39-55)	(47-55)	(56-65)	(66-74)	(75-84)	(85-96)	(97-104)	
TRDV1	-	1-2	.GSRR AGWIRPQETEVLGTE	GKSVNLLCYYG	ASTA....PIVD	LYWYQQYL	TRAPKYTLY	KRR....SYT	GKGDASF.A	IEWFYSKSAA	NMTNLCTRNLVP	ADTAKYYC	ALRN
	+	1-4	.GSRR AGWIRPQETEMSVTE	GKSVNLLCYYD	ASTG....GAVD	LYWYRQYP	NQAPEYILY	RGW....GYT	GKGDASF.A	VERFYSKSAA	DTTNLYIRNLLL	ADTATYYC	ALRD
TRDV5	+	5-1	DEGML CNEVIQSPTEQTVGL	GGEAMLLCTYK	LEIS.....NPD	LFWYRMRP	DHSFQFVLY	RDN.....SR	SKDADFTQG	RFSVKHSQTQ	KTFHLVISLVRA	EDSATYYC	AYAP

Figure S2. TRDV IMGT Numbering. IMGT protein display of the T. m. latirostris TRDV segments. The strand and loops are defined by the IMGT numbering system for a V-DOMAIN and are indicated above the sequences. Conserved cysteines are in pink, the conserved tryptophan (W) and hydrophobic amino acid motifs are in blue, and the N-linked glycosylation sites are in green. Framework regions are in black and the CDR regions are in red. The first column indicates the T. m. latirostris TRDV subgroup each V segment belongs to, the second column indicates expression TRD constant region, and the third column indicates the individual V segment number.

```
TRAJ1:   --YAGVASKLQFGKGTRVTI--
TRAJ2:   -NTGAMTG..T.....Q..V--
TRAJ3:   --GYSSDT.VI..A...LRVQP
TRAJ4:   --SS.DY...I..A...LAVHP
TRAJ5:   ---DTGS..T..S...LQVQP
TRAJ6:   --AS.GNY..T..R..NLVVHP
TRAJ7:   ---DYGNNRFT.....Q.LVTP
TRAJ8:   ---NTGFQ..V..T..QLLVIP
TRAJ9:   --GNTGGFRVV..T...LFVEA
TRAJ10: -GLG.G.D.IT..M..WLKVEP
TRAJ11: ---NSGYN..T.....MLLVSP
TRAJ12: ---KDGGY.WI..S...LLVRP
TRAJ13: -----LYNTFI..S...LSVKP
TRAJ14: ---YQAGNA.I.....TI.VSS
TRAJ15: --GLLS.H.MV..R..MLKVEL
TRAJ16: --INTARN..T..E..M.LVKP
TRAJ17: -DRGSSLGR.Y.....QL.VWP
TRAJ18: ----SNNY..T..A..P.IVRA
TRAJ19: ---VYD.NRFF..S..KLSVKP
TRAJ20: --SS.SGW..A..S..QL.VVP
TRAJ21: --NYNQGG..I..Q..ELSVKP
TRAJ22: ---EEQGFSFI.....LLVKP
TRAJ23: ---NTS.G..T..D..TL.VKP
TRAJ24: AYTGTGSYG.T.....KLSVTP
TRAJ25: ---NTGDRQ.V.....LSVIS
TRAJ26: -------N.II.....QLHVLP
TRAJ27: ------NNIRV..S..QLVVKP
TRAJ28: --...AGN..I..T..LLSVKP
TRAJ29: ----DSNYQ.IW.S..KLF.KP
TRAJ30: ---SSSTN.FI..A...LQVFP
TRAJ31: --QT..GGV.H..S..Q.IVKP
TRAJ32: --PTSG.N..I..T...L.VFS
TRAJ33: ---SSSNTR.I..Q..TLQVKP
TRAJ34: ----PNSLYCG..R.WLP----
TRAJ35: --NV.SSRE.IW.L..SLAVNP
TRAJ36: --NSNAG.M.T..G...LMVKP
TRAJ37: --TTSGSY.YV..A...LKVLT
TRAJ38: -N.G.SQRN.I.....KLSVKP
TRAJ39: ----NNNNAPH..A...LRVIP
TRAJ40: --TTDS.R..V..T...LQVTL
TRAJ41: -NTG.SSYQ.T.....QLI.QP
TRAJ42: ----NYGN.FI..S..ILRVIP
TRAJ43: ---NTGNE..Y..TA.SL.VIP
TRAJ44: ---VVSYN..I..Q..SLS.IP
TRAJ45: NTGGAGYG..T..Q..IL.VYP
TRAJ46: -NTGSIIY..TL.....L..VNP
TRAJ47: -IRESGL..V..Q...L.VNP
TRAJ48: --DT.SNN..K.....TLSVRP
TRAJ49: --TQ.RLE..F.....KL.VNP
TRAJ50: ----KSTKT.T.....QLIVSL
```

Figure S3. TRAJ Alignment. Amino acid alignment of the T. m. latirostris TRAJ segments. The FGXG motif is highlighted in grey. TRAJ segment identifiers that are bolded represent TRAJ segments found in expressed transcripts, identifiers that are red represent pseudogene TRAJ segments. Dots represent identity to the first sequence and dashes represent gaps in the alignment.

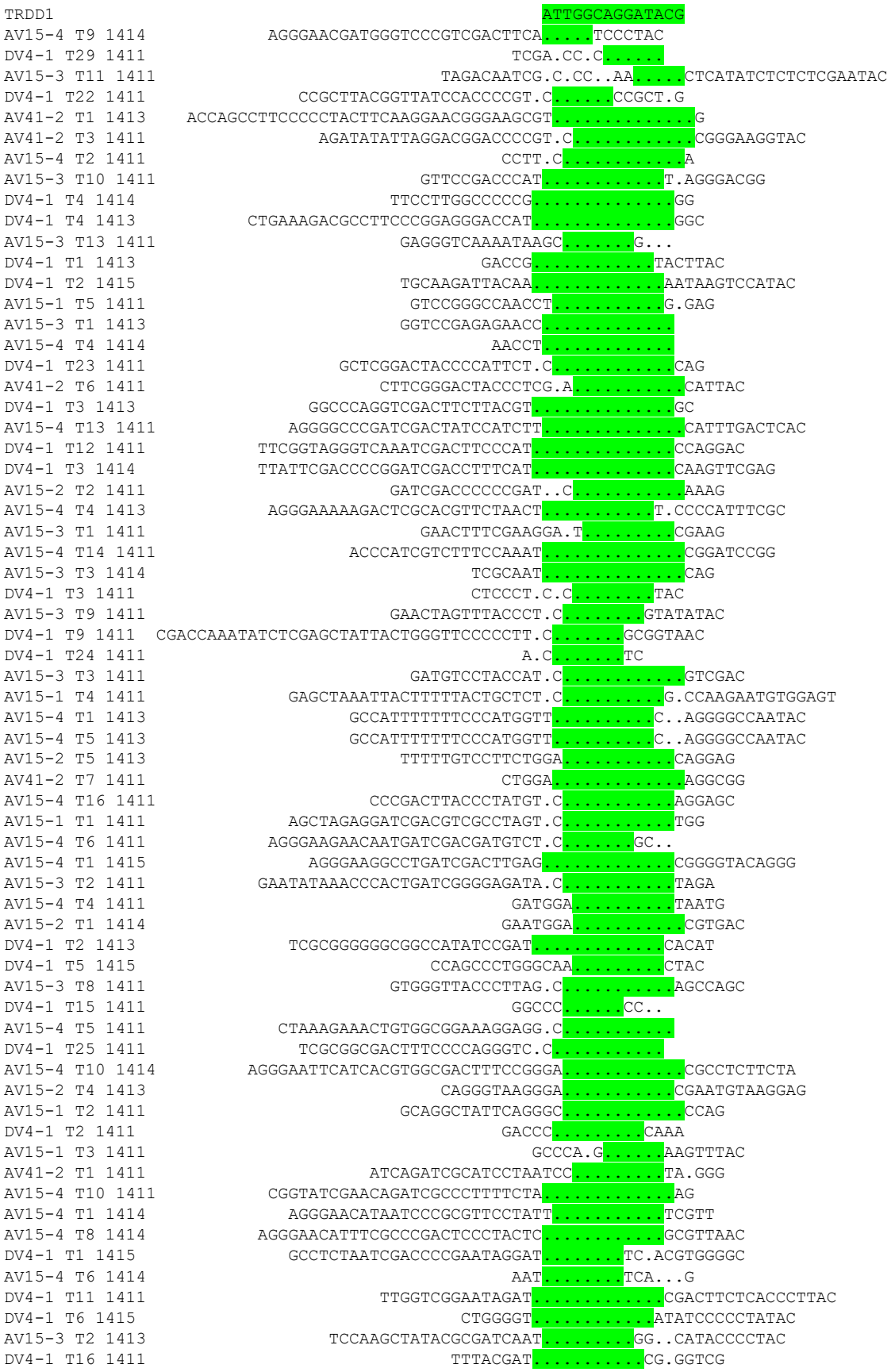

TRDD1 ATTGGCAGGATACG
AV15-4 T9 1414 AGGGAACGATGGGTCCCGTCGACTTCA.....TCCCTAC
DV4-1 T29 1411 TCGA.CC.C......
AV15-3 T11 1411 TAGACAATCG.C.CC..AA.....CTCATATCTCTCTCGAATAC
DV4-1 T22 1411 CCGCTTACGGTTATCCACCCCGT.C......CCGCT.G
AV41-2 T1 1413 ACCAGCCTTCCCCCTACTTCAAGGAACGGGAAGCGT.............G
AV41-2 T3 1411 AGATATATTAGGACGGACCCCGT.C.............CGGGAAGGTAC
AV15-4 T2 1411 CCTT.C.............A
AV15-3 T10 1411 GTTCCGACCCAT.............T.AGGGACGG
DV4-1 T4 1414 TTCCTTGGCCCCCG.............GG
DV4-1 T4 1413 CTGAAAGACGCCTTCCCGGAGGGACCAT.............GGC
AV15-3 T13 1411 GAGGGTCAAAATAAGC.......G...
DV4-1 T1 1413 GACCG.............TACTTAC
DV4-1 T2 1415 TGCAAGATTACAA.............AATAAGTCCATAC
AV15-1 T5 1411 GTCCGGGCCAACCT.............G.GAG
AV15-3 T1 1413 GGTCCGAGAGAACC.............
AV15-4 T4 1414 AACCT.............
DV4-1 T23 1411 GCTCGGACTACCCCATTCT.C.............CAG
AV41-2 T6 1411 CTTCGGGACTACCCTCG.A.............CATTAC
DV4-1 T3 1413 GGCCCAGGTCGACTTCTTACGT.............GC
AV15-4 T13 1411 AGGGGCCCGATCGACTATCCATCTT.............CATTTGACTCAC
DV4-1 T12 1411 TTCGGTAGGGTCAAATCGACTTCCCAT.............CCAGGAC
DV4-1 T3 1414 TTATTCGACCCCGGATCGACCTTTCAT.............CAAGTTCGAG
AV15-2 T2 1411 GATCGACCCCCCGAT..C.............AAAG
AV15-4 T4 1413 AGGGAAAAAGACTCGCACGTTCTAACT.............T.CCCCATTTCGC
AV15-3 T1 1411 GAACTTTCGAAGGA.T.........CGAAG
AV15-4 T14 1411 ACCCATCGTCTTTCCAAAT.............CGGATCCGG
AV15-3 T3 1414 TCGCAAT.............CAG
DV4-1 T3 1411 CTCCCT.C.C.........TAC
AV15-3 T9 1411 GAACTAGTTTACCCT.C.........GTATATAC
DV4-1 T9 1411 CGACCAAATATCTCGAGCTATTACTGGGTTCCCCCTT.C.......GCGGTAAC
DV4-1 T24 1411 A.C.......TC
AV15-3 T3 1411 GATGTCCTACCAT.C.............GTCGAC
AV15-1 T4 1411 GAGCTAAATTACTTTTTACTGCTCT.C.............G.CCAAGAATGTGGAGT
AV15-4 T1 1413 GCCATTTTTTTCCCATGGTT.............C..AGGGGCCAATAC
AV15-4 T5 1413 GCCATTTTTTTCCCATGGTT.............C..AGGGGCCAATAC
AV15-2 T5 1413 TTTTTGTCCTTCTGGA.............CAGGAG
AV41-2 T7 1411 CTGGA.............AGGCGG
AV15-4 T16 1411 CCCGACTTACCCTATGT.C.............AGGAGC
AV15-1 T1 1411 AGCTAGAGGATCGACGTCGCCTAGT.C.............TGG
AV15-4 T6 1411 AGGGAAGAACAATGATCGACGATGTCT.C.......GC..
AV15-4 T1 1415 AGGGAAGGCCTGATCGACTTGAG.............CGGGGTACAGGG
AV15-3 T2 1411 GAATATAAACCCACTGATCGGGGAGATA.C.............TAGA
AV15-4 T4 1411 GATGGA.............TAATG
AV15-2 T1 1414 GAATGGA.............CGTGAC
DV4-1 T2 1413 TCGCGGGGGGCGGCCATATCCGAT.............CACAT
DV4-1 T5 1415 CCAGCCCTGGGCAA.........CTAC
AV15-3 T8 1411 GTGGGTTACCCTTAG.C.............AGCCAGC
DV4-1 T15 1411 GGCCC.......CC..
AV15-4 T5 1411 CTAAAGAAACTGTGGCGGAAAGGAGG.C.............
DV4-1 T25 1411 TCGCGGCGACTTTCCCCAGGGTC.C.............
AV15-4 T10 1414 AGGGAATTCATCACGTGGCGACTTTCCGGGA.............CGCCTCTTCTA
AV15-2 T4 1413 CAGGGTAAGGGA.............CGAATGTAAGGAG
AV15-1 T2 1411 GCAGGCTATTCAGGGC.............CCAG
DV4-1 T2 1411 GACCC.........CAAA
AV15-1 T3 1411 GCCCA.G.......AAGTTTAC
AV41-2 T1 1411 ATCAGATCGCATCCTAATCC.........TA.GGG
AV15-4 T10 1411 CGGTATCGAACAGATCGCCCTTTTCTA.............AG
AV15-4 T1 1414 AGGGAACATAATCCCGCGTTCCTATT.............TCGTT
AV15-4 T8 1414 AGGGAACATTTCGCCCGACTCCCTACTC.............GCGTTAAC
DV4-1 T1 1415 GCCTCTAATCGACCCCGAATAGGAT.........TC.ACGTGGGGC
AV15-4 T6 1414 AAT.........TCA...G
DV4-1 T11 1411 TTGGTCGGAATAGAT.............CGACTTCTCACCCTTAC
DV4-1 T6 1415 CTGGGGT.............ATATCCCCCTATAC
AV15-3 T2 1413 TCCAAGCTATACGCGATCAAT.........GG..CATACCCCTAC
DV4-1 T16 1411 TTTACGAT.............CG.GGTCG

Figure S4. TRDD CDR3 Alignment. Nucleotide alignment of the expressed CDR3 regions (excluding the V and J germline sequence) to the germline TRDD1 sequence. Each sequence is identified by the TRADV segment used in that transcript, a number to differentiate transcripts within an individual that use the same TRADV (T), and the individual the transcript was sequenced from (141X). Dots represent identity to the germline TRDD1, and the dots highlighted in green are the CDR3 sequence most likely contributed by the TRDD1.

Subgroup	EXP	V	-------A------> (1-15)	-----B----> (16-26)	CDR1 (27-38)	---C---> (39-46)	---C'---> (47-55)	CDR2 (56-65)	---C"---> (66-74)	----D----> (75-84)	-----E-----> (85-96)	---F---> (97-104)	CDR3
TRBV1	+	1-1	GSA	DSGITQRPKYLVVGM	GIKKSLICEQH	LGH.....NS	MYWYKQVP	EKPPELMFQ	FNY....KEL	VTNETTS.S	RFSPKCR.DS	SHLYLHVDALEP	NDSAVYLC	ASSK
TRBV2	+	2-1	GPL	DSAVSQTPKNLVTQM	GSQGSLYCKQD	LGH.....QT	MYWYRQDA	AGLLRVMFG	YSN....REP	ILNESVP.S	RFSPECR.DK	ALLSLHINALEP	SDSALYLC	AGS.
TRBV4	+	4-1	GAV	GSGVTQTPKHLIKAR	GEHVMMSCFPL	SEH.....LS	VSWYQKAQ	GQGPKFLVE	YYN....GEE	RDKGDIP.D	RFSGKQF.SD	YRSELNLSAVEL	GDSAVFLC	ASSL
	-	4-3	GPL	DSGVTQTPKHLIKAP	GQQVMLSCSPL	SGH.....LS	VSWYQQAQ	GQGPKFLLE	YFS....GKE	RDKGHIP.G	RFSGRQF.SD	YHSELNVSALEL	ADSAVFLC	ASSL
	+	4-4	GPV	DSGVTQTLRHLIKTP	GQQVMLSCSPH	SGH.....RY	VSWYQQAQ	GQGPKFLLQ	YYD....GKE	IDKGDIP.D	RLSGRQF.SD	NHSELNLTILEK	GDSAVFLC	ASSL
	+	4-5	GPV	DSGVTQTPRHLIKAR	GQQVTLSCSPL	SGH.....LS	VYWYQQAQ	GQGPKFLLQ	YYY....GKE	KDKGDIP.D	RFSGQQF.SD	YRSELTLSALEP	GDSAVFLC	ASSL
TRBV5	+	5-1	GHT	DAGVTQTPRHKVTRT	GGNVTLQCAQD	MNH.....EC	MYWYRQDP	GLGLRLIHY	SVL....IGD	TQKGDIP.D	GYSVSRS.NK	EHFPLTLGSAAP	SQTAVYFC	ASSV
	-	5-3	GHM	DAGVTQTQRQKVTRT	GGNVTLQCTQD	MKH.....NY	MYWYRQDP	RLELLLMHY	SPG....VGD	IVKGDIL.Y	GYCVSRS.NT	EHFPLTLGSAAP	SQTAVYCC	ASSV
	-	5-4	GHT	DAGITQTPRHQITRT	GRYVTLQCTQD	MNH.....EN	MYWYRQDP	GLELWLIHY	APD....VGD	TYKGEIP.S	GYSVSRS.NA	EHFPLTLESAIP	SKTAVYFC	TSGA
	-	5-7	GNT	DADIIQSPRHKITKT	GGYVTLQCTPD	INN.....DY	MYWYLQDP	ALGMQLMHY	SAG....VGE	TYKGDIP.S	GYNVSRS.NI	QHFLLLLGSAAP	SQTVVYFC	TSNP
	+	5-8	GHT	DAGVTQTPRHKVTRT	GGIVTLQCTQD	MNH.....NY	MYWYRQDP	GLGLRLIHY	SSG....VRV	TDKGDLP.D	GYNVSRS.NT	RNFPLMLESAAP	SQTAVYFC	ASSL
	-	5-11	GHK	DAGITQTPRHKVTRT	GGNVTLQCSQD	LNH.....DN	VYWYQQDP	GLGLLLIHY	SVD....VCV	TDKGDIP.D	GYSVSIS.NT	ERFPLTLGSAVR	LQTAVYFC	TSSV
TRBV6	+	6-2	GHT	GAGVSQVPRHRVTKK	GQAVALSCDPI	SGH.....VG	LYWYRQNL	GQGLELLTY	FQN....KFP	GDTSGMPEA	RFSADRP.DG	SSSTLRIQPVEP	GDSAVYLC	ASSL
TRBV8	+	8-1	GPV	DTGVVQSPRHLIAEK	GGQATLQCHPA	SGH.....RS	VYWYQQTR	GQGPKFLFE	LYK....GEQ	QEKGDIQ.G	RFSAQQF.SN	DSSELNVRALEQ	GDSAVFLC	ASSP
	-	8-2	GPV	DAGIVQSPRHIITEN	GGQATPQCHPV	SEH.....RS	ESRYQQAQ	GQGPKILVE	YYE....KAD	RNKGDIL.D	RFSVQQF.PN	YSSELKLSALEQ	GDSAMFLC	ASSS
	+	8-4	GPV	DPEVIQSPRHFITRK	GGQAILQCHPA	SGH.....TN	VYWYQQTQ	GQGPQFLFR	YYN....GKQ	QEKGDIP.D	RFSGQQF.SE	DHSELNLSALEL	GDSAVFLC	ASSL
	+	8-5	GPV	DTGTVQSPRHIITEK	GGQATLKCHPV	SGH.....RS	VYWYQQAQ	GQGPKFLIQ	YYD....KTE	RNKGDIP.D	RFSAQQF.SN	YSSELKLSTLEK	GDSAMFLC	ASSI
TRBV9	+	9-1	EHS	EAGVVQTPRHKITEV	RQSVALWCDPV	SSH.....NT	LYWYRQMP	GQGPELLVN	FQN....EAV	LDDAQLPKN	RFSAERL.KG	ADSTLRIQAAEP	GDSAVYLC	ASSL
TRBV10	-	10-2	EHT	DARVTQTPRHKVAEM	GHRVTLRCQLI	SGD.....NT	LFWYRQTS	AWGLELLIF	FQN....QAV	IEQVEMLKD	WSSAEMP.DR	SFSTLKIQHAEL	GDSAIYLC	ASSS
TRBV12	+	12-1	GSM	GAMVIQSPKYRVTSV	GKLVSLSCSQN	LNH.....DT	MYWYQQKQ	SQAPKLLFY	YYE....TNF	NKEADTS.D	NFQPQRP.NT	SFCGLDIGSPGL	ADSAVYLC	ASSR
	+	12-3	GPM	GAMVTQSPKYQVTKV	GKLVCLRCSQS	LNH.....NT	MYWYQQKQ	SQAPKLLFY	YYN....RDF	NKETDTS.D	NFQPTRP.NT	SFCGLDIRSPVM	GDLAVYLC	ASSR
TRBV13	+	13-2	GSF	DAEVTQTPRHLLKGK	GQKVKMDCVPI	KGH.....SY	VYWYQQIP	AKEFKFLIS	FQN....EKI	FDGNEMPKE	RFSAECP.QD	SACSLEIQPAAL	QDSAMYFC	ASRV
	+	13-3	GSR	DAEVTQTPRHLLKGK	GQKVKMDCVPI	KGH.....SY	VYWYQQIP	AKEFKFLIS	FQN....EKI	FDGNEMPKE	RFSAECP.QD	SACSLEIQPAGL	QDSAMYFC	ASRV
TRBV14	-	14-1	GTT	DSGITQTPKYLVQEE	GQDVTLKCEQT	FNH.....DS	MYWYRQDL	GKEAGLVYY	SPI....TND	VQGGDAA.A	GYSTSQG.KK	ECFPFVVTWAQR	NQTALYLC	ASSL
	-	14-3	GTT	DSGITQTPKYLVREE	GQNVSLKCEQT	FNH.....DS	MYWYRQDP	GHGLRLVYY	STI....TNY	VQEGDAA.A	GYRASRK.KM	EHFPLVVTRAQR	NQTALYLC	ASSL
	-	14-4	GTT	DSGITQTPKYLVREE	GQNVSLKCEQT	FNH.....DS	MYWYLQDP	GHGLRLIYY	SLV....SNN	VQEGDAA.A	GYRASRK.KK	EHFPLMVTRAQR	NQTALYLC	ASSL
TRBV15	+	15-2	GPV	SALISQHPSRAICKS	GASVKIQCHSM	GIQ.....AMT	MLWYRQLP	GQSFTLIAT	SNQG...SSV	TYEQGFTEA	KFPISHP.DL	TFSTLMVTSAQV	EDSSLYFC	GVI.
	+	15-4	GPL	SALVSQQPSRTICKS	GASVKIQCRSM	DIQ.....AAV	MLWYHQLP	GQSFTLIAT	SNQG...SRV	TYEQGFTEA	KFPISHP.DL	TFSTLTVTSAQA	KDSSLYFC	GAS.
	+	15-5	GPV	GALVSQHPSRAICKS	GASVKIQCRST	GIQ.....AWT	ILWYHQPP	RQSFTLIAT	SNQG...SSV	TYEQGFTKA	KFPISHP.NL	TFSTLTVTSGQA	EDSSLYFC	GAS.
TRBV16	+	16-1	GSM	DTEVTQTPRQLVKAR	EQTARMECVPI	KGH.....EY	VYWYRQKL	GEELKFLIY	FNN....EEV	FDESGMSDK	RFSAQCP.KN	STCSMEIQSTEP	RDSAVYFC	ASSL
	+	16-4	GSM	ETEVTQTPRQLVKAR	QQTAKMDCVPM	EGH.....AY	VFWYRQKL	GEELKFLIY	FQN....EDA	MDKSGMPND	RFSAKCP.EN	SPCSMEIQSMEP	GDSAVYFC	ASSQ
	+	16-5	GSM	DTEVTQTPRQLVKAR	QQIARMDCVPI	KQH.....VH	VYWYRQKL	REELKFLAY	LKN....EET	IDGSGLPDK	RFSAHCS.KN	SSCSMEIQSTEL	GDSAVYFC	ASSL
	+	16-8	GSM	DTEVTQTPRQLVITT	KQRATMDCVPI	RQH.....NR	VYWYRQKL	GEELKFLTY	LQN....KEA	LDKSGLPDG	RFSAQCT.QN	SHCSMKIQSTEL	KDSGMYFC	ASSP
TRBV18	+	18-1	GSM	DTGVTQIPRNRIAQT	GKKVTLQCSQT	KGH.....KY	MYWYRQDP	GLGLQLIYY	SLD....VND	INKGEAS.K	GYSVFRK.EQ	DKFSLSLEAATP	NQIALYLC	ATSY
TRBV19	-	19-1	GPV	KADVSQTPRHYVTGT	GKKITLECSQT	LGH.....DY	MYWYRQDP	GRELQLMHY	FYS....VNS	TEKGELS.S	RWTVSRR.RK	EHFSLVLESASP	AESSRYLC	ASSS
TRBV20	+	20-1	GAE	DAVVTQFPRHRVLGK	GKELTLECSQK	MNH.....LV	MYWYRQVP	GHGLQLIYY	STG....TGS	IEDGDVT.E	GYSVSRD.KT	QNFPLTLASTSP	SQTAVYLC	ASSE
TRBV21	+	21-1	GPA	EPGVSQTPRHRIAKS	GDSLTVACSQD	LDY.....EV	MYWYRQDP	GRGLQLLHY	SIN....VKI	VQRGDLP.D	GYRVSRE.EK	GRFLLTLESAVP	NQTALYLC	ASSF
TRBV23	+	23-1	GSAF	GVLVSQKPSRHICQH	GTPVTIQCQVD	TQV.....NR	MFWYHQPP	GQSLILIAT	ANQG...SEA	TYESGFTKD	KFAINHP.DF	TFSTLTVKNSSP	EDSSVYLC	SAG.
	+	23-2	GSVF	GVLVSQKPSRDICQH	GTAVKIQCQVD	TQV.....TS	MFWYRQPP	GQGLTLIAT	ANQG...SEA	TYKSGFTKD	KFPINYP.NL	TFSTLTVTKSSP	EDSGIYLC	STG.
TRBV24	+	24-1	DVG	AQTILQWPVVEVRQV	GSSLSLQCTVT	GTS.....NPT	LYWYWQAT	GGAPQLLFY	SLD.....MD	NIKGEAP.K	HFSAFRP.RD	GQFTLNADKLQP	GDSGVYLC	AWD.

Figure S5. TRBV IMGT Numbering. IMGT protein display of the T. m. latirostris TRBV segments. The strand and loops are defined by the IMGT numbering system for a V-DOMAIN and are indicated above the sequences. Conserved cysteines are in pink, the conserved tryptophan (W) and hydrophobic amino acid motifs are in blue, and the N-linked glycosylation sites are in green. Framework regions are in black and the CDR regions are in red. The first column indicates the T. m. latirostris TRBV subgroup each V segment belongs to, the second column indicates expression TRB constant region, and the third column indicates the individual V segment number.

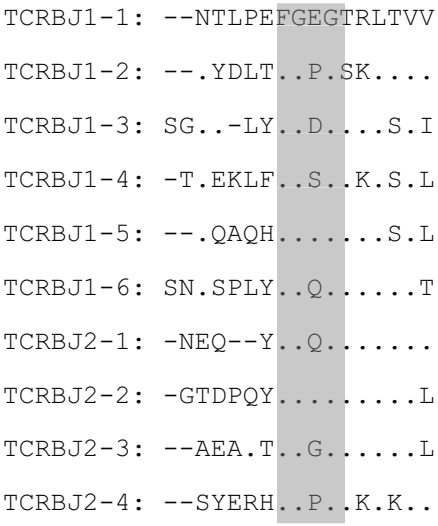
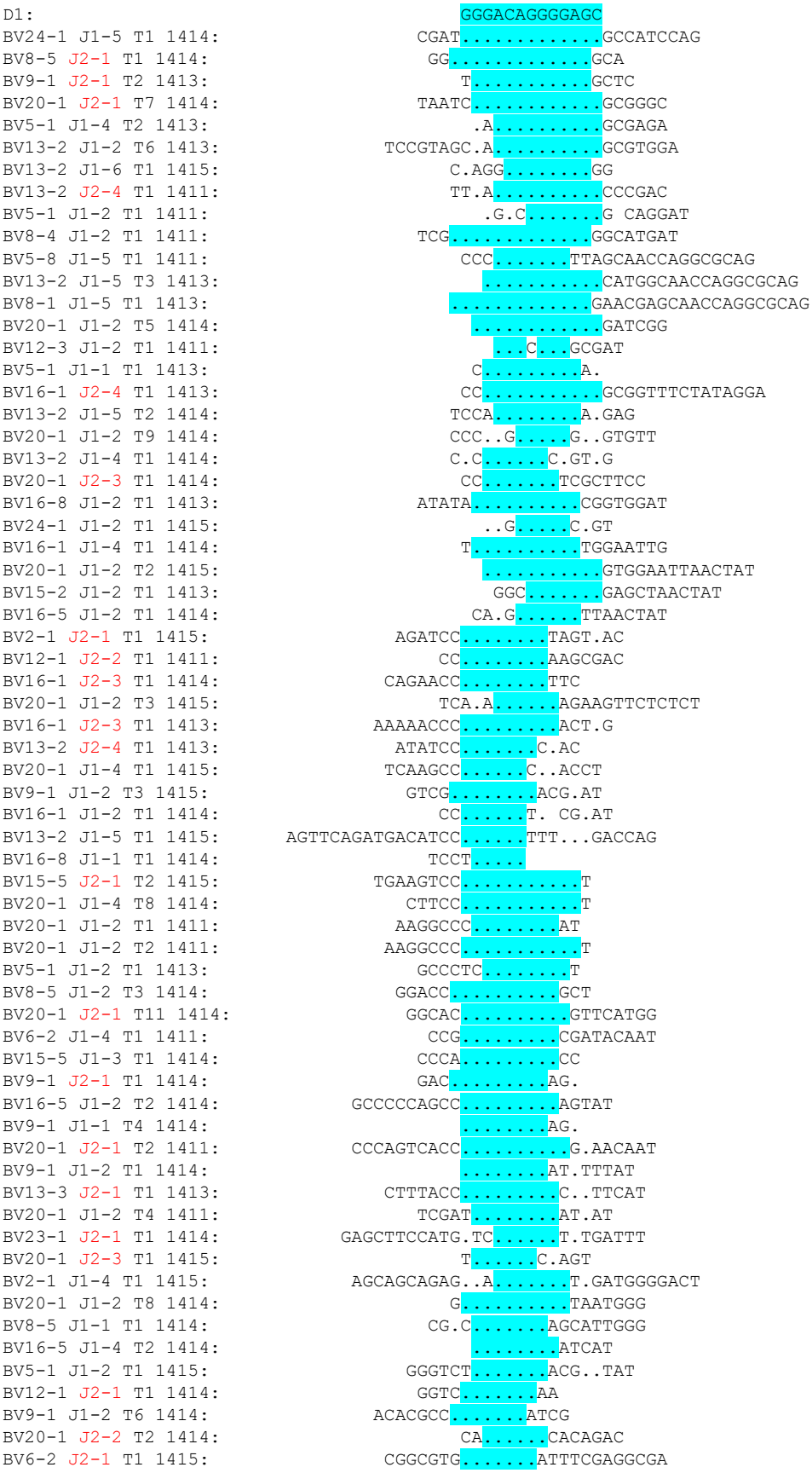
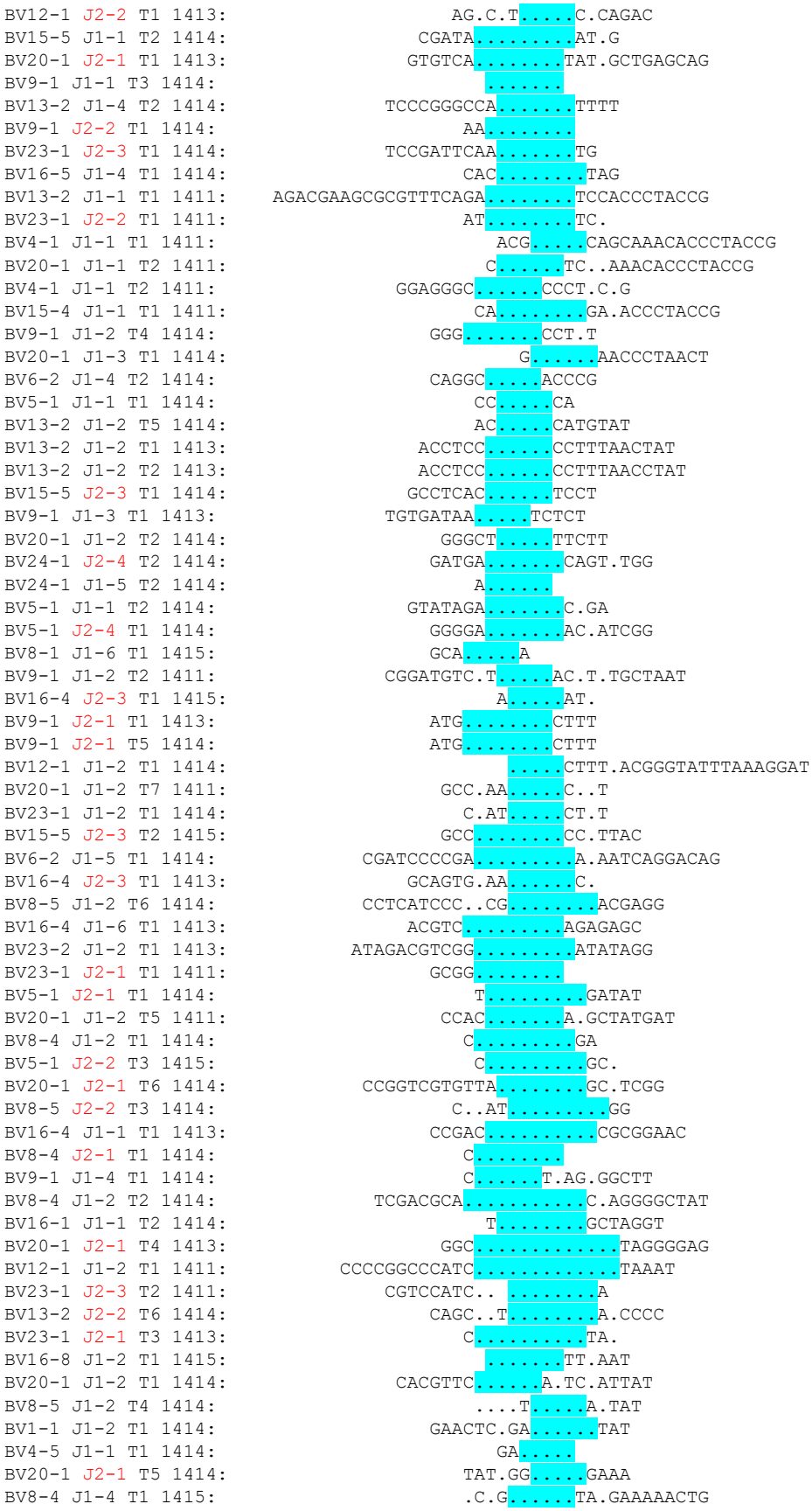
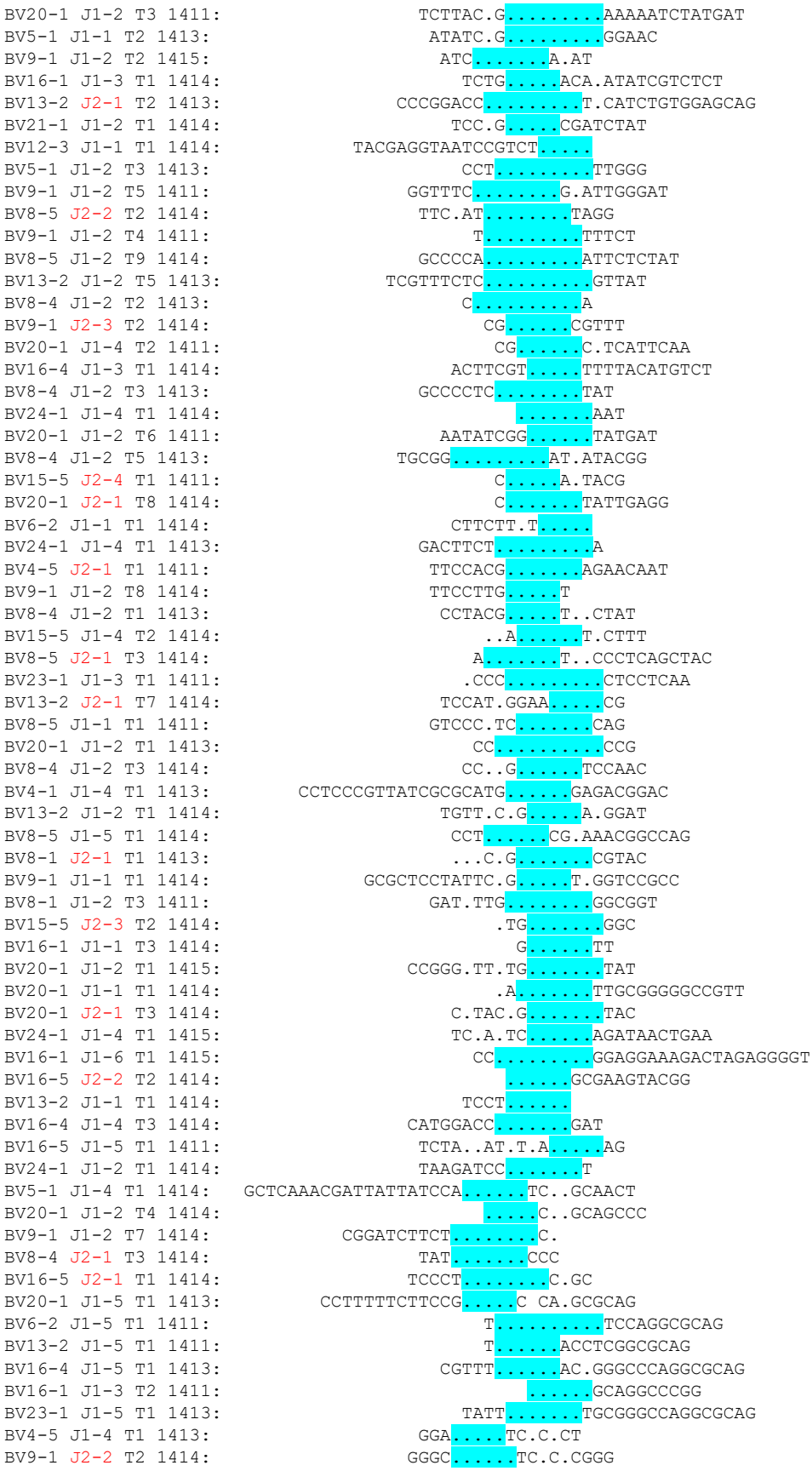
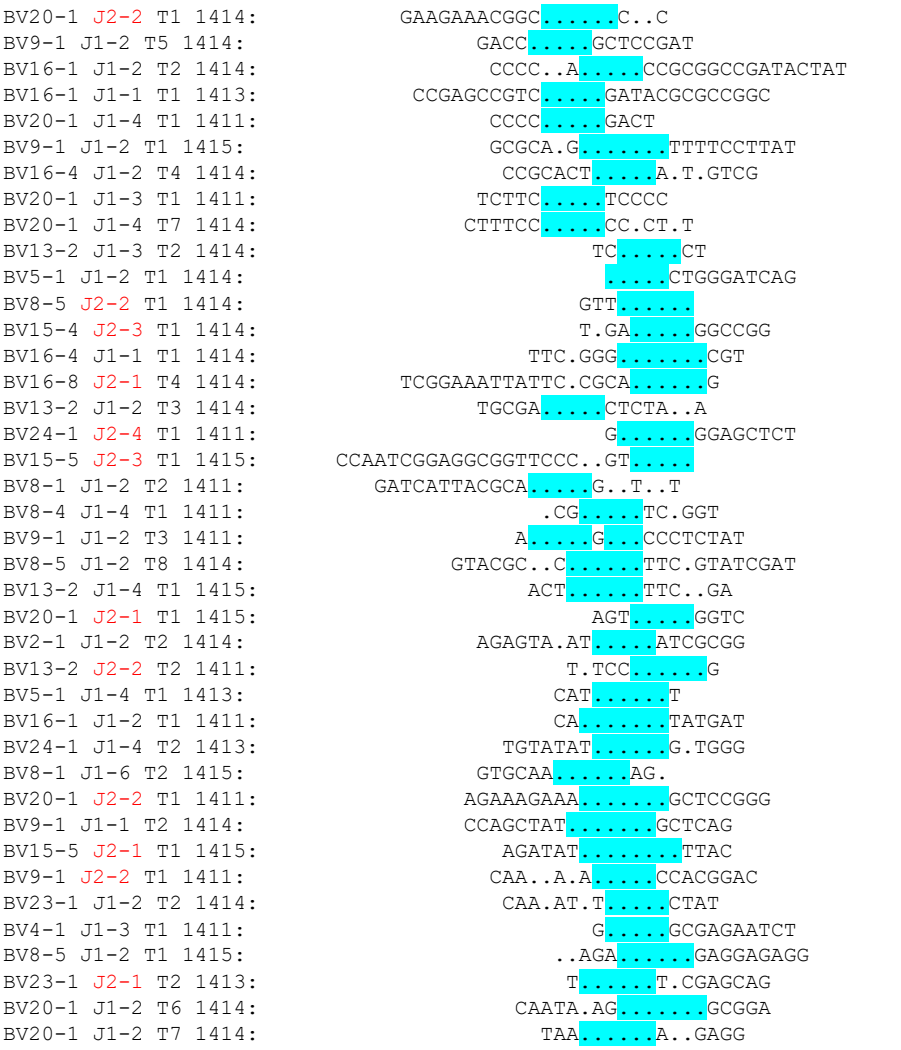


Figure S6. TRBJ Alignment. Amino acid alignment of the T. m. latirostris TRBJ segments. The FGXG motif is highlighted in grey. Dots represent identity to the first sequence and dashes represent gaps in the alignment.


D1: GGGACAGGGGAGC
BV24-1 J1-5 T1 1414: CGAT.....GCCATCCAG
BV8-5 J2-1 T1 1414: GG.....GCA
BV9-1 J2-1 T2 1413: T.....GCTC
BV20-1 J2-1 T7 1414: TAATC.....GCGGGC
BV5-1 J1-4 T2 1413: .A.....GCGAGA
BV13-2 J1-2 T6 1413: TCCGTAGC.A.....GCGTGGA
BV13-2 J1-6 T1 1415: C.AGG.....GG
BV13-2 J2-4 T1 1411: TT.A.....CCCGAC
BV5-1 J1-2 T1 1411: .G.C.....G CAGGAT
BV8-4 J1-2 T1 1411: TCG.....GGCATGAT
BV5-8 J1-5 T1 1411: CCC.....TTAGCAACCAGGCGCAG
BV13-2 J1-5 T3 1413: .....CATGGCAACCAGGCGCAG
BV8-1 J1-5 T1 1413: .....GAACGAGCAACCAGGCGCAG
BV20-1 J1-2 T5 1414: .....GATCGG
BV12-3 J1-2 T1 1411: ...C...GCGAT
BV5-1 J1-1 T1 1413: C.....A.
BV16-1 J2-4 T1 1413: CC.....GCGGTTTCTATAGGA
BV13-2 J1-5 T2 1414: TCCA.....A.GAG
BV20-1 J1-2 T9 1414: CCC..G.....G..GTGTT
BV13-2 J1-4 T1 1414: C.C.....C.GT.G
BV20-1 J2-3 T1 1414: CC.....TCGCTTCC
BV16-8 J1-2 T1 1413: ATATA.....CGGTGGAT
BV24-1 J1-2 T1 1415: ..G.....C.GT
BV16-1 J1-4 T1 1414: T.....TGGAATTG
BV20-1 J1-2 T2 1415: .....GTGGAATTAACTAT
BV15-2 J1-2 T1 1413: GGC.....GAGCTAACTAT
BV16-5 J1-2 T1 1414: CA.G.....TTAACTAT
BV2-1 J2-1 T1 1415: AGATCC.....TAGT.AC
BV12-1 J2-2 T1 1411: CC.....AAGCGAC
BV16-1 J2-3 T1 1414: CAGAACC.....TTC
BV20-1 J1-2 T3 1415: TCA.A.....AGAAGTTCTCTCT
BV16-1 J2-3 T1 1413: AAAAACCC.....ACT.G
BV13-2 J2-4 T1 1413: ATATCC.....C.AC
BV20-1 J1-4 T1 1415: TCAAGCC.....C..ACCT
BV9-1 J1-2 T3 1415: GTCG.....ACG.AT
BV16-1 J1-2 T1 1414: CC.....T. CG.AT
BV13-2 J1-5 T1 1415: AGTTCAGATGACATCC.....TTT...GACCAG
BV16-8 J1-1 T1 1414: TCCT.....
BV15-5 J2-1 T2 1415: TGAAGTCC.....T
BV20-1 J1-4 T8 1414: CTTCC.....T
BV20-1 J1-2 T1 1411: AAGGCCC.....AT
BV20-1 J1-2 T2 1411: AAGGCCC.....T
BV5-1 J1-2 T1 1413: GCCCTC.....T
BV8-5 J1-2 T3 1414: GGACC.....GCT
BV20-1 J2-1 T11 1414: GGCAC.....GTTCATGG
BV6-2 J1-4 T1 1411: CCG.....CGATACAAT
BV15-5 J1-3 T1 1414: CCCA.....CC
BV9-1 J2-1 T1 1414: GAC.....AG.
BV16-5 J1-2 T2 1414: GCCCCCAGCC.....AGTAT
BV9-1 J1-1 T4 1414: .....AG.
BV20-1 J2-1 T2 1411: CCCAGTCACC.....G.AACAAT
BV9-1 J1-2 T1 1414: .....AT.TTTAT
BV13-3 J2-1 T1 1413: CTTTACC.....C..TTCAT
BV20-1 J1-2 T4 1411: TCGAT.....AT.AT
BV23-1 J2-1 T1 1414: GAGCTTCCATG.TC.....T.TGATTT
BV20-1 J2-3 T1 1415: T.....C.AGT
BV2-1 J1-4 T1 1415: AGCAGCAGAG..A.....T.GATGGGGACT
BV20-1 J1-2 T8 1414: G.....TAATGGG
BV8-5 J1-1 T1 1414: CG.C.....AGCATTGGG
BV16-5 J1-4 T2 1414: .....ATCAT
BV5-1 J1-2 T1 1415: GGGTCT.....ACG..TAT
BV12-1 J2-1 T1 1414: GGTC.....AA
BV9-1 J1-2 T6 1414: ACACGCC.....ATCG
BV20-1 J2-2 T2 1414: CA.....CACAGAC
BV6-2 J2-1 T1 1415: CGGCGTG.....ATTTCGAGGCGA


BV12-1 J2-2 T1 1413:
BV15-5 J1-1 T2 1414:
BV20-1 J2-1 T1 1413:
BV9-1 J1-1 T3 1414:
BV13-2 J1-4 T2 1414:
BV9-1 J2-2 T1 1414:
BV23-1 J2-3 T1 1414:
BV16-5 J1-4 T1 1414:
BV13-2 J1-1 T1 1411:
BV23-1 J2-2 T1 1411:
BV4-1 J1-1 T1 1411:
BV20-1 J1-1 T2 1411:
BV4-1 J1-1 T2 1411:
BV15-4 J1-1 T1 1411:
BV9-1 J1-2 T4 1414:
BV20-1 J1-3 T1 1414:
BV6-2 J1-4 T2 1414:
BV5-1 J1-1 T1 1414:
BV13-2 J1-2 T5 1414:
BV13-2 J1-2 T1 1413:
BV13-2 J1-2 T2 1413:
BV15-5 J2-3 T1 1414:
BV9-1 J1-3 T1 1413:
BV20-1 J1-2 T2 1414:
BV24-1 J2-4 T2 1414:
BV24-1 J1-5 T2 1414:
BV5-1 J1-1 T2 1414:
BV5-1 J2-4 T1 1414:
BV8-1 J1-6 T1 1415:
BV9-1 J1-2 T2 1411:
BV16-4 J2-3 T1 1415:
BV9-1 J2-1 T1 1413:
BV9-1 J2-1 T5 1414:
BV12-1 J1-2 T1 1414:
BV20-1 J1-2 T7 1411:
BV23-1 J1-2 T1 1414:
BV15-5 J2-3 T2 1415:
BV6-2 J1-5 T1 1414:
BV16-4 J2-3 T1 1413:
BV8-5 J1-2 T6 1414:
BV16-4 J1-6 T1 1413:
BV23-2 J1-2 T1 1413:
BV23-1 J2-1 T1 1411:
BV5-1 J2-1 T1 1414:
BV20-1 J1-2 T5 1411:
BV8-4 J1-2 T1 1414:
BV5-1 J2-2 T3 1415:
BV20-1 J2-1 T6 1414:
BV8-5 J2-2 T3 1414:
BV16-4 J1-1 T1 1413:
BV8-4 J2-1 T1 1414:
BV9-1 J1-4 T1 1414:
BV8-4 J1-2 T2 1414:
BV16-1 J1-1 T2 1414:
BV20-1 J2-1 T4 1413:
BV12-1 J1-2 T1 1411:
BV23-1 J2-3 T2 1411:
BV13-2 J2-2 T6 1414:
BV23-1 J2-1 T3 1413:
BV16-8 J1-2 T1 1415:
BV20-1 J1-2 T1 1414:
BV8-5 J1-2 T4 1414:
BV1-1 J1-2 T1 1414:
BV4-5 J1-1 T1 1414:
BV20-1 J2-1 T5 1414:
BV8-4 J1-4 T1 1415:


BV20-1 J1-2 T3 1411:
BV5-1 J1-1 T2 1413:
BV9-1 J1-2 T2 1415:
BV16-1 J1-3 T1 1414:
BV13-2 J2-1 T2 1413:
BV21-1 J1-2 T1 1414:
BV12-3 J1-1 T1 1414:
BV5-1 J1-2 T3 1413:
BV9-1 J1-2 T5 1411:
BV8-5 J2-2 T2 1414:
BV9-1 J1-2 T4 1411:
BV8-5 J1-2 T9 1414:
BV13-2 J1-2 T5 1413:
BV8-4 J1-2 T2 1413:
BV9-1 J2-3 T2 1414:
BV20-1 J1-4 T2 1411:
BV16-4 J1-3 T1 1414:
BV8-4 J1-2 T3 1413:
BV24-1 J1-4 T1 1414:
BV20-1 J1-2 T6 1411:
BV8-4 J1-2 T5 1413:
BV15-5 J2-4 T1 1411:
BV20-1 J2-1 T8 1414:
BV6-2 J1-1 T1 1414:
BV24-1 J1-4 T1 1413:
BV4-5 J2-1 T1 1411:
BV9-1 J1-2 T8 1414:
BV8-4 J1-2 T1 1413:
BV15-5 J1-4 T2 1414:
BV8-5 J2-1 T3 1414:
BV23-1 J1-3 T1 1411:
BV13-2 J2-1 T7 1414:
BV8-5 J1-1 T1 1411:
BV20-1 J1-2 T1 1413:
BV8-4 J1-2 T3 1414:
BV4-1 J1-4 T1 1413:
BV13-2 J1-2 T1 1414:
BV8-5 J1-5 T1 1414:
BV8-1 J2-1 T1 1413:
BV9-1 J1-1 T1 1414:
BV8-1 J1-2 T3 1411:
BV15-5 J2-3 T2 1414:
BV16-1 J1-1 T3 1414:
BV20-1 J1-2 T1 1415:
BV20-1 J1-1 T1 1414:
BV20-1 J2-1 T3 1414:
BV24-1 J1-4 T1 1415:
BV16-1 J1-6 T1 1415:
BV16-5 J2-2 T2 1414:
BV13-2 J1-1 T1 1414:
BV16-4 J1-4 T3 1414:
BV16-5 J1-5 T1 1411:
BV24-1 J1-2 T1 1414:
BV5-1 J1-4 T1 1414:
BV20-1 J1-2 T4 1414:
BV9-1 J1-2 T7 1414:
BV8-4 J2-1 T3 1414:
BV16-5 J2-1 T1 1414:
BV20-1 J1-5 T1 1413:
BV6-2 J1-5 T1 1411:
BV13-2 J1-5 T1 1411:
BV16-4 J1-5 T1 1413:
BV16-1 J1-3 T2 1411:
BV23-1 J1-5 T1 1413:
BV4-5 J1-4 T1 1413:
BV9-1 J2-2 T2 1414:


BV20-1 J2-2 T1 1414: GAAGAAACGGC.....C..C
BV9-1 J1-2 T5 1414: GACC.....GCTCCGAT
BV16-1 J1-2 T2 1414: CCCC..A.....CCGCGGCCGATACTAT
BV16-1 J1-1 T1 1413: CCGAGCCGTC.....GATACGCGCCGGC
BV20-1 J1-4 T1 1411: CCCC.....GACT
BV9-1 J1-2 T1 1415: GCGCA.G.....TTTTCCTTAT
BV16-4 J1-2 T4 1414: CCGCACT.....A.T.GTCG
BV20-1 J1-3 T1 1411: TCTTC.....TCCCC
BV20-1 J1-4 T7 1414: CTTTCC.....CC.CT.T
BV13-2 J1-3 T2 1414: TC.....CT
BV5-1 J1-2 T1 1414: .....CTGGGATCAG
BV8-5 J2-2 T1 1414: GTT.....
BV15-4 J2-3 T1 1414: T.GA.....GGCCGG
BV16-4 J1-1 T1 1414: TTC.GGG.....CGT
BV16-8 J2-1 T4 1414: TCGGAAATTATTC.CGCA.....G
BV13-2 J1-2 T3 1414: TGCGA.....CTCTA..A
BV24-1 J2-4 T1 1411: G.....GGAGCTCT
BV15-5 J2-3 T1 1415: CCAATCGGAGGCGGTTCCC..GT.....
BV8-1 J1-2 T2 1411: GATCATTACGCA.....G..T..T
BV8-4 J1-4 T1 1411: .CG.....TC.GGT
BV9-1 J1-2 T3 1411: A.....G...CCCTCTAT
BV8-5 J1-2 T8 1414: GTACGC..C.....TTC.GTATCGAT
BV13-2 J1-4 T1 1415: ACT.....TTC..GA
BV20-1 J2-1 T1 1415: AGT.....GGTC
BV2-1 J1-2 T2 1414: AGAGTA.AT.....ATCGCGG
BV13-2 J2-2 T2 1411: T.TCC.....G
BV5-1 J1-4 T1 1413: CAT.....T
BV16-1 J1-2 T1 1411: CA.....TATGAT
BV24-1 J1-4 T2 1413: TGTATAT.....G.TGGG
BV8-1 J1-6 T2 1415: GTGCAA.....AG.
BV20-1 J2-2 T1 1411: AGAAAGAAA.....GCTCCGGG
BV9-1 J1-1 T2 1414: CCAGCTAT.....GCTCAG
BV15-5 J2-1 T1 1415: AGATAT.....TTAC
BV9-1 J2-2 T1 1411: CAA..A.A.....CCACGGAC
BV23-1 J1-2 T2 1414: CAA.AT.T.....CTAT
BV4-1 J1-3 T1 1411: G.....GCGAGAATCT
BV8-5 J1-2 T1 1415: ..AGA.....GAGGAGAGG
BV23-1 J2-1 T2 1413: T.....T.CGAGCAG
BV20-1 J1-2 T6 1414: CAATA.AG.....GCGGA
BV20-1 J1-2 T7 1414: TAA.....A..GAGG
D2: GGGACTTGGGGGTAG
TCRB V1-1 J1-6 T1 1411: GA.CCC.....TAAC
TCRB V13-2 J1-1 T3 1414: TCTCG.....
TCRB V20-1 J2-2 T4 1414: T..C.....CTAGATTTAGGGG
TCRB V8-1 J1-2 T1 1411: .CGTC.C.....TTACATAT
TCRB V8-1 J2-3 T1 1415: .CTGCGC.....ATCG
TCRB V20-1 J2-1 T10 1414: A.....CCAAC
TCRB V16-5 J2-1 T1 1413: GAA.....ATAG
TCRB V16-5 J2-1 T2 1413: GAA.....ATAG
TCRB V24-1 J2-1 T1 1414: .A.....TCA
TCRB V13-2 J2-4 T2 1413: T.....G.....T..C
TCRB V15-5 J1-2 T1 1413: .....G...TC..T
TCRB V24-1 J2-1 T1 1413: .....A.CCGGGTCTAGGGAC
TCRB V13-2 J2-1 T1 1415: GAAT.....CTAC
TCRB V13-3 J2-1 T1 1414: ATCC..A.....ACCTACGG
TCRB V15-5 J2-1 T2 1414: .....CAGCAGCTAC
TCRB V24-1 J2-4 T1 1414: CTAGAAGAAAATTCT.....GA.C
TCRB V13-2 J1-5 T1 1413: CCTTGAACACATC...AA.....C.ACCAGGCGCAG
TCRB V23-1 J2-1 T1 1413: ......C....CAGCTAC
TCRB V16-5 J2-1 T1 1411: CCAAAACT..A.....ACTCTTGG
TCRB V20-1 J2-2 T3 1414: CAACA..ACC.....GGC
TCRB V8-4 J2-1 T2 1414: AGCGGTC.....TAGCTTCTCC
TCRB V16-4 J2-2 T1 1411: ACGTC.ACCA.....
TCRB V13-2 J2-1 T6 1414: TC.....CGTTCCATATC

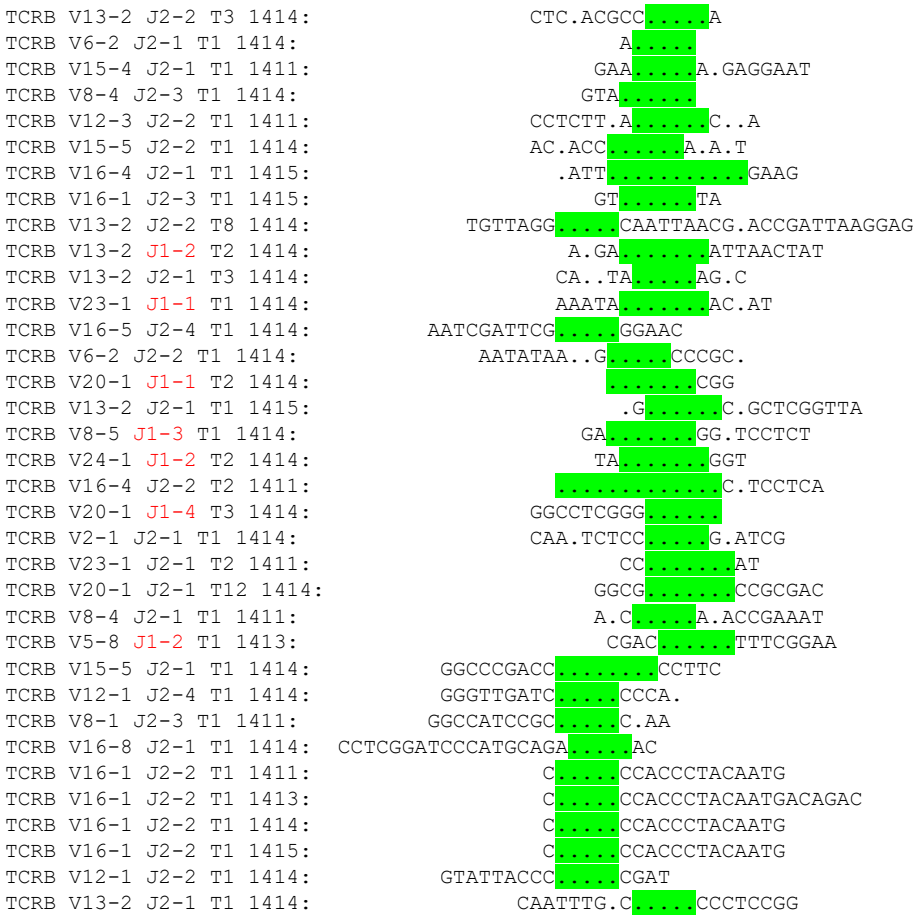


Figure S7. TRBD CDR3 Alignment. Nucleotide alignment of the expressed CDR3 regions (excluding the V and J germline sequence) to the germline TRBD sequences. Each sequence is identified by the TRBV and TRBJ segment used in that transcript, a number to differentiate transcripts within an individual that use the same TRBV/TRBJ (T), and the individual the transcript was sequenced from (141X). Sequence names in red indicate sequences that were inter-cluster rearrangements (TRBD1 to cluster 2 TRBJ or TRBD2 to cluster 1 TRBJ). Dots represent identity to the germline TRBD, the dots highlighted in blue are CDR3 sequence most likely contributed by the TRBD1, and the dots highlighted in green are CDR3 sequence most likely contributed by the TRBD2.

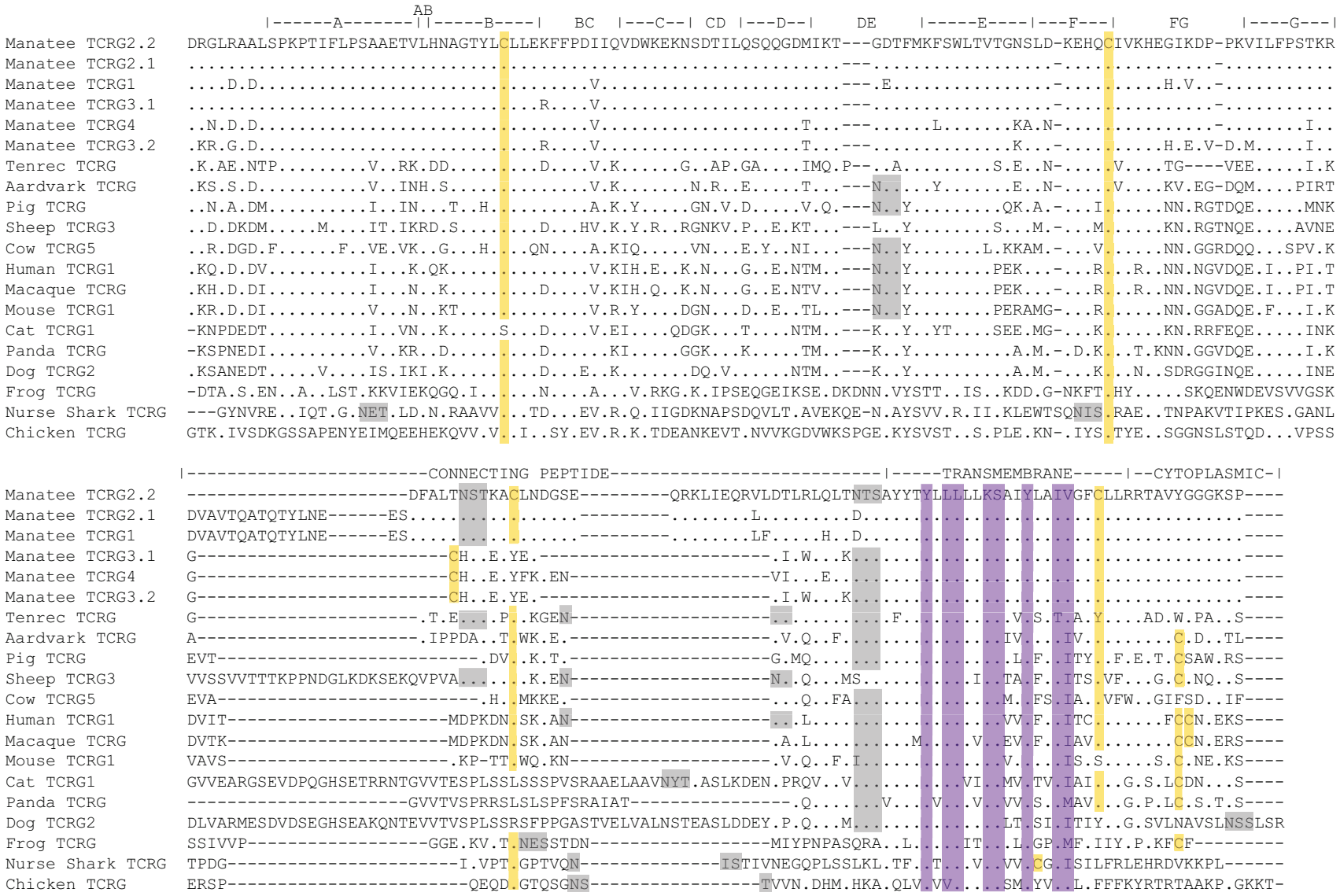


Figure S8. TRGC Comparative Alignment. Amino acid alignment of the TRG constant region from different species. Cysteine residues are highlighted in yellow, N-linked glycosylation sites are highlighted in grey, and the CART motif residues are highlighted in purple. Dots represent similarity to the T. m. latirostris sequence and dashes represent gaps introduced to make the alignment.

Subgroup	EXP	V	-------A------>	-----B---->	CDR1	---C--->	---C'--->	CDR2	---C"--->	----D---->	-----E----->	---F--->	CDR3
			(1-15)	(16-26)	(27-38)	(39-46)	(47-55)	(56-65)	(66-74)	(75-84)	(85-96)	(97-104)	
TRGV1	+	1-1	ASLK	ASILEMKSVSITMQS	GESVEIICDLK	QD.......IYY	IHWYRYQE	GKTLQRILH	FYFP..SSKV	VLDSGFSSE	KYYAYKN.TQ	RICKLVLENLQE	SDSGVYYC	AA
	+	1-2	ASLK	ASILEMKSMSITKQS	GESVEIICDLK	QD.......IDY	IHWYRYQE	GKTLQRILY	FSFS..SSQV	LKDSGFSSE	KYHAYKD.TQ	RICKLVLKNLQE	SDSGVYYC	AA
	+	1-3	ASLK	ASILEMKSMSITKQS	GESVEIICDLK	QD.......IDY	IHWYRYQE	GKILQRILH	FYFP..SSKV	VLDSGFSSE	KYHANKN.TQ	RSCKLVLKNLQE	SDSGVYYC	AA
	+	1-4	ASLK	ASILKMKSMSITKQS	GESVEIICDLK	QN.......IDY	IHWYRYQE	GKTLQRILY	FSFY..SSQV	VKDSGFSSE	KYHAYKN.TQ	SSCKLVLKNLQE	SDSGVYYC	AA
	+	1-5	ASLK	ASILEMKSMSITKQS	GESVEIICDLK	QD.......IDY	IHWYLYQE	GKTLQRILY	FSFS..SSRV	LKDSGFSSE	KYHAYKN.TQ	RICKLVLKNLQE	SDSGVYYC	AT
	+	1-6	ASLK	ASILEMKSMSITKQS	GESVEIICDLK	QF.......TEY	IHWYQYQE	GKTLQRILY	FYFS..SSRV	LKDSGFNSR	KYNAYKN.TQ	SSCKLVLEDLQE	SDSGVYYC	AA
	+	1-7	ASLK	ASILEMKSMSITKQS	GESVEIICDLK	QD.......MNY	IHWYRYQE	GKILQRILH	FYFP..SSKV	VLDSGFSSE	KYHAYKN.TQ	RSCKLFLKNLQE	SDSGVYYC	AA
	+	1-8	ASLK	ASIPEMKSMSITKQS	GESVEIICDLK	QV.......TEY	IHWYRYQE	GKTLQRILN	FFFS..SSQV	VLDSGFSSE	KYHAYKN.TQ	RICKLVLKNLQE	SDSGVYYC	AA
	+	1-9	ASLK	ASILEMKSMSITKQS	GESVEIICDLK	QN.......IDY	IHWYRYQE	GKILQRILY	FYFP..RSQV	VKDSGFNTR	KYNAYKN.TQ	RSCKFVLGSLEE	IDSGVYYC	AV
TRGV3	+	3-1	VGLG	QWKLEQPDMSISKAA	SKSAHITCKVF	VEDF....ENAN	IGWYRLKP	NQGVEYLIY	IVST..TTPN	YASLGGKSK	KLEASRNSHT	STSILKINSVEK	EDEAVYYC	AW
	+	3-2	VGLG	QWKLEQPDMSISKAA	SKSAHITCKVF	VEDF....QNAN	IGWYRLKP	NQGVEYLIY	IIST..TTPT	YASLGGKSK	KLEASKNSHT	STAILKINSVEK	EDEAMYYC	AW
	+	3-3	VGLG	QWKLEQPDMSISKAA	SKSAHITCNVA	VKDF....ENAY	IGWFRLKP	NQGIEYLIY	IHST..TTPN	YASLGGKSK	KLEASKNSRS	STSILKINSVEK	EDEAMYYC	AW
TRGV4	+	4-1	YGLG	LINVEQSQISVSTVV	KKSVDISCHIL	STSF....ENDV	IHWYRLKP	NQALEHLLY	VISS..ENPT	QKRLGE.NN	KFEARKVSYA	STSVFTIHFVEE	EDTAIYYC	AG
	+	4-2	DGLG	LVKVEQSQISVSTVV	KKSVDISCRIS	STSF....ENDV	IRWYRLKP	NQALEHLLY	VIST..ENPT	QKGLGEKNN	KFEARKVSYA	STSVFTIQFVEE	EDTAIYYC	AG

Figure S9. TRGV IMGT Numbering. IMGT protein display of the T. m. latirostris TRGV segments. The strand and loops are defined by the IMGT numbering system for a V-DOMAIN and are indicated above the sequences. Conserved cysteines are in pink, the conserved tryptophan (W) and hydrophobic amino acid motifs are in blue, and the N-linked glycosylation sites are in green. Framework regions are in black and the CDR regions are in red. The first column indicates the T. m. latirostris TRGV subgroup each V segment belongs to, the second column indicates expression TRG constant region, and the third column indicates the individual V segment number.

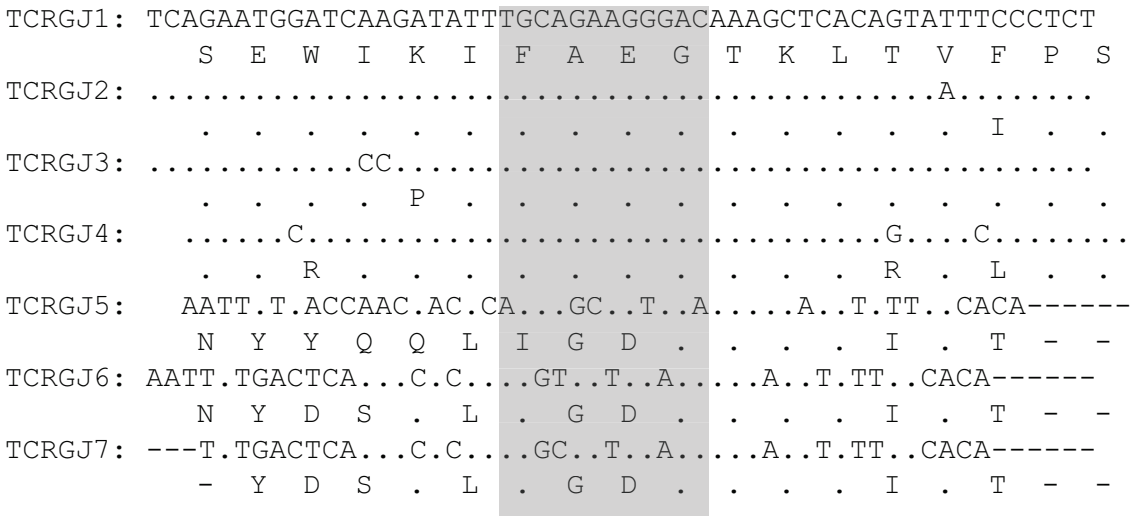


Figure S10. TRGJ Alignment. Nucleotide and amino acid alignment of the T. m. latirostris TRGJ segments. The FGXG motif is highlighted in grey. Dots represent identity to the first sequence and dashes represent gaps in the alignment.

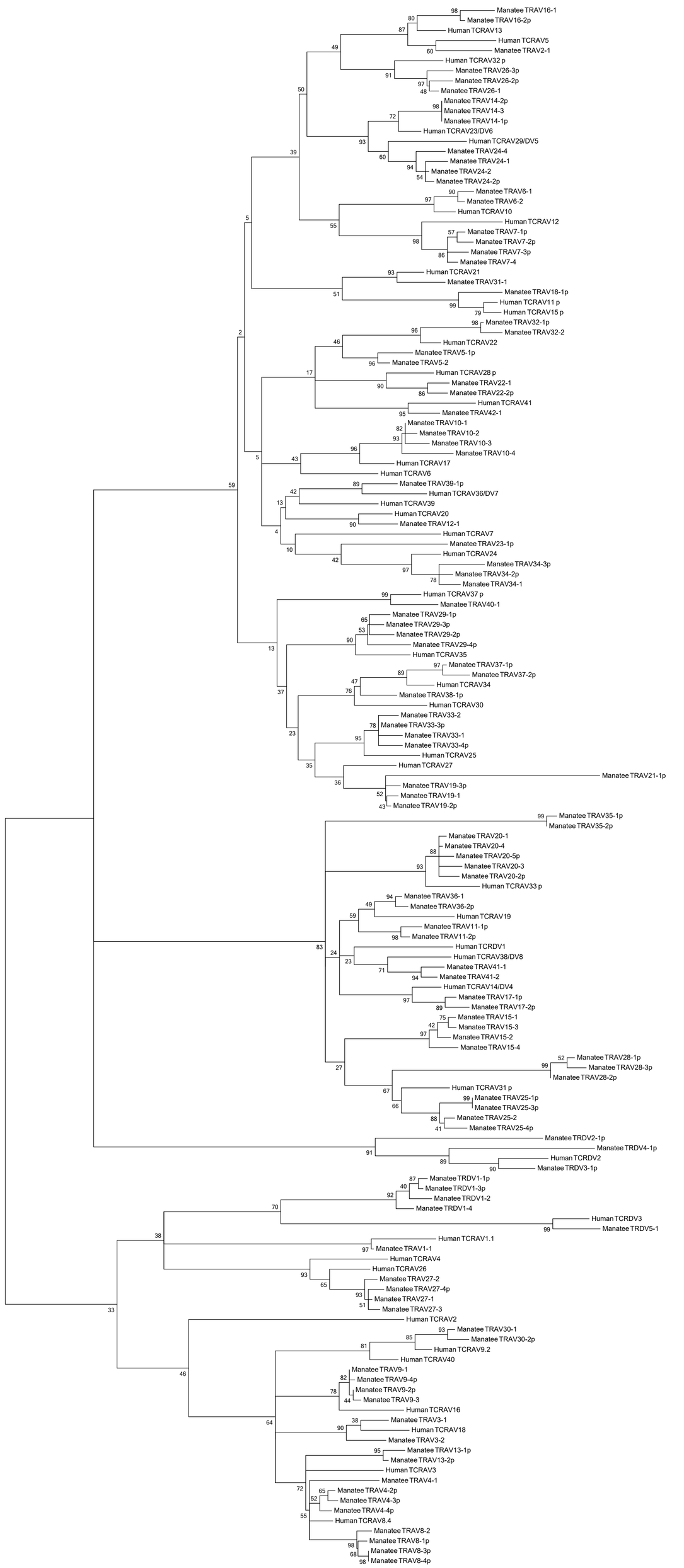


Figure S11. TRADV Phylogenetic Relationships. A maximum-likelihood tree based on the General Time Reversible model including human and manatee TRADV segments. Branch lengths are measured by number of substitutions per site, as indicated by the scale. Topology support from bootstrap replications are indicated at each node.

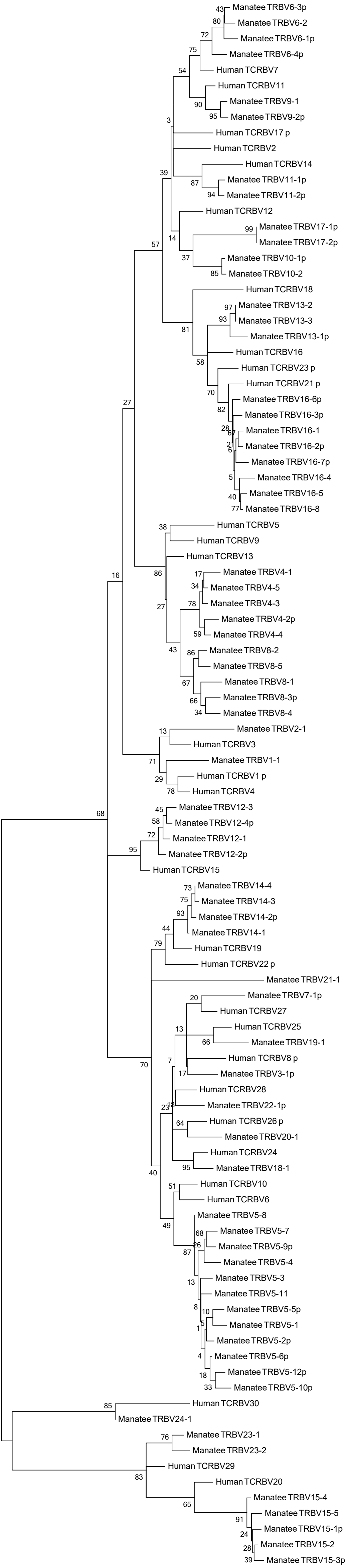


Figure S12. TRBV Phylogenetic Relationships. A maximum-likelihood tree including human and manatee TRBV segments. Branch lengths are measured by number of substitutions per site, as indicated by the scale.

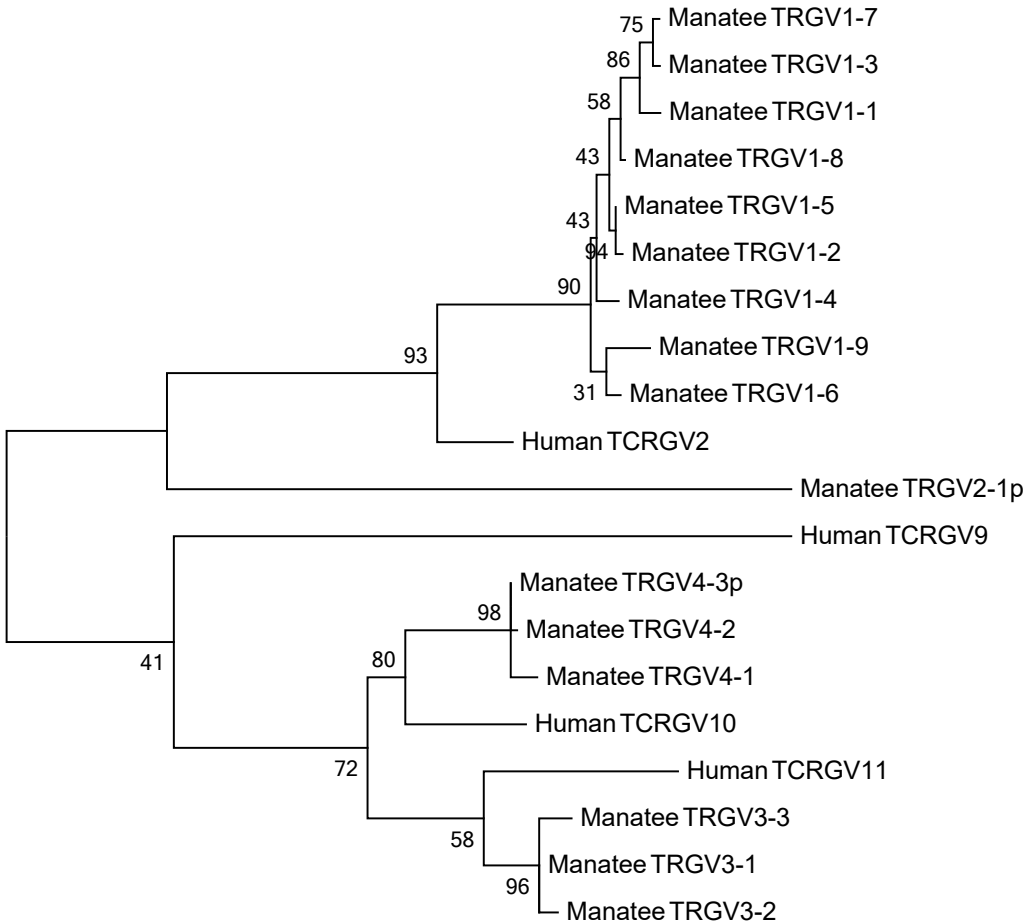


Figure S12. TRBV Phylogenetic Relationships. A maximum-likelihood tree including human and manatee TRBV segments. Branch lengths are measured by number of substitutions per site, as indicated by the scale.