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. Ecological 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