Veterinary Immunology and Immunopathology 167 (2015) 156–165

Contents lists available at ScienceDirect

Veterinary Immunology and Immunopathology

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

Characterization of antibody V segment diversity in the Tasmanian devil (*Sarcophilus harrisii*)

Beata Ujvari[a,b,*], Katherine Belov[b]

[a] Centre for Integrative Ecology, Deakin University, Victoria, Australia
[b] Faculty of Veterinary Science, University of Sydney, New South Wales, Australia

ARTICLE INFO

Article history:
Received 29 April 2015
Received in revised form 25 July 2015
Accepted 6 August 2015

Keywords:
Immunoglobulin
Heavy and light chains
Variable region
Marsupial
Tasmanian devil

ABSTRACT

The Tasmanian devil (*Sarcophilus harrisii*) immune system has recently been under scrutiny because of the emergence of a contagious cancer, which has decimated devil numbers. Here we provide a comprehensive description of the Tasmanian devil immunoglobulin variable regions. We show that heavy chain variable (VH) and light chain variable (VL) repertoires are similar to those described in other marsupial taxa: VL diversity is high, but VH diversity is restricted and belongs only to clan III. As in other mammals, one VH and one Vλ germline family and multiple incomplete Vκ germline sequences were identified in the genome. High Vκ variation was observed in transcripts and we predict that it may have arisen by gene conversion and/or somatic mutations, as it does not appear to have originated from germline variation. Phylogenetic analyses revealed that devil VL gene segments are highly complex and ancient, with some lineages predating the separation of marsupials and eutherians. These results indicate that although the evolutionary history of immune genes lead to the expansions and contractions of immune gene families between different mammalian lineages, some of the ancestral immune gene variants are still maintained in extant species. A high degree of similarity was found between devil and other marsupial VH segments, demonstrating that they originated from a common clade of closely related sequences. The VL families had a higher variation than VH both between and within species. We suggest that, similar to other studied marsupial species, the complex VL segment repertoire compensates for the limited VH diversity in Tasmanian devils.

1. Introduction

The adaptive/combinatorial immune system arose about 400–450 million years ago, and from its onset had the potential to recognize and distinguish all possible antigens, whether of 'self', 'altered self' or 'non-self' origin (Marchalonis et al., 2006). The basis for this ability lies in the structural organization and assembly of immunoglobulins: while gene segments specifying the constant regions and frameworks are genetically inherited, diversity in the recognition units are generated by somatic mutations and by somewhat random rearrangements. Immunoglobulins are composed of two identical light chains and two identical heavy chains. The heavy chain determines the subclass of each antibody, while based on small polypeptide sequence differences the light chains can be classified as either kappa (κ) or lambda (λ). The chains are held together by inter- and intra-chain disulfide bonds. Based on the variability in the amino acid sequences both the light and heavy chains can be divided into two segments, a variable and a constant region. Antibodies bind antigens primary through their variable region. Unique immunoglobulin variable regions are produced by somatic recombination of tandemly repeated subgenes, called variable (V), diversity (D) and joining (J) segments to generate an exon encoding the V domain. The heavy chain variable region is formed by random assembly of one V, one D and one J gene segment, while the light chain variable region is generated by combining one V and one J segment. The VDJ subgenes evolve via gene duplication, divergence and deletion generating multiple copies of these gene segments and providing a broad array of genes available for recombination. Mature Ig V segment diversity is created by three distinct mechanisms: (1) germline diversity, (2) gene conversion, and (3) somatic mutation (Sun et al., 2013). The combination of these processes produces large numbers of antibodies with different antigen binding specificity. The diversity of VH sequences originates from the presence of divergent germline VH families, where family clustering is determined by VH genes sharing ≥75% nucleotide identity.

* Corresponding author at: Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Victoria, Australia.
E-mail address: beata.ujvari@deakin.edu.au (B. Ujvari).

http://dx.doi.org/10.1016/j.vetimm.2015.08.001
0165-2427/

Based on their degrees of identity VH genes from various vertebrate species have been classified into five groups (A–E) (Sitnikova and Su, 1998), of which mammals possess three (A–C) (or clans I–III) (Kirkham et al., 1992).

The number of IgHV genes in these three groups varies among different mammals (Das et al., 2008; Sitnikova and Su, 1998): primates, rodents and bats have multiple VH families, but other mammals (including both eutherians, monotremes and marsupials) show limited VH diversity (Baker et al., 2010; Mainville et al., 1996; van Dijk et al., 1993) with only a single VH family or a set of recently derived VH families present in the germline (Baker et al., 2005; Butler, 1997; Johansson et al., 2002). The number of functional and pseudo-VH genes also varies between species: humans and mice have over 40 functional VH genes (Das et al., 2008), while rabbits only use one VH segment to generate 70–90% of the expressed antibody repertoire (Knight, 1992).

Compared to the VH genes, the mammalian VL genes most likely originated more than 470 million years ago (Rast et al., 1994) and their evolution has followed a more complex path: the genes coding the two light chains (κ and λ) are located on different chromosomes, and the λ-chain genes appear to be polyphyletic to the κ-chain genes (Rast et al., 1994; Sitnikova and Su, 1998). Mammalian VL genes are grouped into six clans (five Vλ sequences and 1 Vκ) – which are further divided into 10 subgroups (Sitnikova and Nei, 1998; Sitnikova and Su, 1998; Sun et al., 2013). Tetrapod species utilize κ and λ chains to different extents, but the overall complexity of the germline Vλ and Vκ repertoire tends to favor the preferential use of one isotype over another (Almagro et al., 1998; Sun et al., 2013). Mammals and reptiles utilize both types of light chains: κ and λ, birds possess only λ chains; and both κ and λ orthologues have been identified in amphibians and fish (reviewed in Sun et al., 2013; Wu et al., 2010).

Species that have a restricted VH repertoire tend to have limited VL repertoires as well (Sun et al., 2013; Wu et al., 2010). Divergent taxa, such as rabbits, cattle, platypus and chicken both have limited germline VH and VL diversity. In contrast, other organisms, such as humans, sheep and mice, show both high germline VH and VL diversity. The two marsupial species investigated so far, the gray short-tailed opossum (*Monodelphis domestica*) and the common brushtail possum (*Trichosurus vulpecula*) provide an exception to this rule. They show limited VH diversity but high VL diversity (Baker et al., 2005; Miller, 2010; Miller et al., 1998). In this study we characterize immunoglobulin VH and VL diversity in the Tasmanian devil, a species under threat of extinction due to a contagious cancer (Hawkins et al., 2006).

2. Materials and methods

2.1. Analysis of the Tasmanian devil whole genome and transcriptome sequences

VH segments were identified in the whole genome sequence of the Tasmanian devil available in the Ensembl (Murchison et al., 2012) database (DEVIL7.0, GCA_000189315.1) by using the BLAT, BLASTN and BLASTX algorithms. The reference genome of Tasmanian devils was generated from fibroblast cell lines originated from a 5-year old female devil. Sequencing libraries were sequenced from both ends, yielding 2.87 × 109 pairs of 100 bp sequence reads. Additional "mate pair" libraries were generated and 50 bp was sequenced from both ends in order to assist with genome assembly (Murchison et al., 2012). The N50 length (50% of the assembled genome lies in blocks of the N50 size or longer) for supercontigs is 1847.19 kb and is 20.13 kb for contigs. The total number of bases in supercontigs is 3.17 Gb and in contigs is 2.93 Gb (Murchison et al., 2012).

Additionally, BLASTN searches were performed on as yet unpublished Tasmanian devil spleen and lymph node transcriptomes to identify expressed V sequences (sample accession numbers: ERS624952, ERS624953, secondary accession: SAMEA3170515Spleen, SAMEA3170516 Lymph (Papenfuss, A. T. WEHI, pers. commun.))

2.2. RNA extraction and RACE-PCR

Blood samples (500 μl/sample) collected from the jugular vein by a veterinarian following the standard operating procedures of the Department of Primary Industries, Parks, Water and Environment (DPIPWE) were stored in RNAprotect Animal Blood Tubes (Qiagen, Germantown, MD) according to the manufacturer's recommendations. All procedures performed in studies involving animals were in accordance with the ethical standards of the institution or practice at which the studies were conducted. RNA was extracted using the RNeasy Protect Animal Blood Kit (Qiagen, Germantown, MD). RNA quality and quantity were quantified on an Agilent 2100 BioAnalyzer (Agilent, Santa Clara, CA).

In order to obtain variable region transcripts RACE PCR was performed using the GeneRacer kit with Superscript III (Invitrogen, CA, USA). 5′ RACE was performed with reverse primers located in the conserved domains of immunoglobulin constant regions. RACE-ready cDNA was generated following the manufacturer's instructions. In brief, the RNA was dephosphorylated and decapped. A 5′ oligo was then ligated to the 5′ end of the RNA. RNA was transcribed into cDNA using Superscript III, with the reaction primed using gene specific 3′ oligos to produce RACE ready cDNA. Gene specific primers, reverse Kappa-Race1 (5′-ACC TGA TTA TCG ACC TTC CAC GTG-3′), Lambda-Race1 (5′-GTA GAA GCC ATT TAC CAG GCA CAC-3′), IgHM-Race2 (5′-TTG GCA GGG AGG ATGA CTT GTG AAG TG-3′), Kappa-Race2 (5′-TGG AAG ATG AAG GCA GAT GGC TGA-3′), Lambda-Race2 (5′-TGT GTC TAG CTC AGG TTT GGA TGG-3′) and IgHM-Race4 (5′-GCA CGT TCA CTT GCC GGG TAG ATCA-3′) were designed using Primer3Plus website (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) from the Tasmanian devil genome (Murchison et al., 2012). The 50 μl RACE PCR contained 50 ng RACE-ready cDNA, 1× High Fidelity PCR Buffer (Invitrogen), 200 μM dNTP (Sigma–Aldrich, NSW, Australia), 10 pm of region specific primers: Kappa-Race1, Lambda-Race1 or IgHM-Race2 (Sigma–Aldrich, NSW, Australia), 5 pm of 5′ RACE primer (Invitrogen) 1 U of Platinum Taq DNA Polymerase High Fidelity (Invitrogen) and 2 mM $MgSO_4$ (Invitrogen). PCR cycling parameters were as follows; samples were heated at 94 °C for 2 min followed by 5 cycles of 94 °C for 30 s and 72 °C for 1 min, followed by 5 cycles at 94 °C for 30 s and 70 °C for 1 min, followed by a further 15 cycles of 94 °C for 30 s, 70–64 °C/−0.4 C/cycle for 30 s, followed by a further 25 cycles of 94 °C for 30 s, 64 °C for 30 s, and 68 °C for 1 min, with a final extension step at 68 °C for 10 min. To obtain gene specific fragments nested PCRs were performed using the 5′ RACE nested primer (Invitrogen) in combination with nested 3′ reverse primers, Kappa Race2, Lambda Race2 and IgHM-Race4. Nested PCR cycling parameters were as follows: samples were heated at 94 °C for 2 min followed by 35 cycles of 94 °C for 30 s, 65 °C for 30 s, and 72 °C for 2 min, with a final extension step at 72 °C for 10 min. The 30 μl nested RACE PCR contained 50 ng RACE-PCR product, 1× High Fidelity PCR Buffer (Invitrogen), 200 μM dNTP (Sigma–Aldrich, NSW, Australia), 20 pm of region specific primers: Kappa Race2, Lambda Race2 and IgHM-Race4 (Sigma–Aldrich, NSW, Australia), 20 pm of 5′ nested RACE primer (Invitrogen), 1 U of Platinum Taq DNA Polymerase High Fidelity (Invitrogen) and 2 mM $MgSO_4$ (Invitrogen). Samples were run on 2% agarose gel and bands were excised. Gel bands were purified using S.N.A.P.™ Columns (Invitrogen). The purified fragments were cloned into plasmids using the TOPO TA Cloning® Kit for