5-1-2012

Lymphocyte evolution and ontogeny in non-eutherian mammals

Xinxin Wang

Xinxin Wang
Candidate

Biology
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Robert D. Miller, Chairperson

Eric S. Loker

Coenraad Adema

Michael Criscitiello

LYMPHOCYTE EVOLUTION AND ONTOGENY IN NON-EUTHERIAN MAMMALS

by

XINXIN WANG

M.S. Biology, Institute of Hydrobiology,
Chinese Academy of Sciences, 2006

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Biology

The University of New Mexico
Albuquerque, New Mexico

May, 2012

ACKNOWLEDGMENTS

I would like to thank the four members of my committee: the chair of my committee, Rob Miller for his direction and guidance, who has taught me techniques of researching and writing. Thanks are also due to the other members of my committee, Sam Loker, Coen Adema and Michael Criscitiello for their invaluable advice and support.

Thanks to Michelle Baker for making available to us a colony of opossums. Thanks to Joseph Volpe from Duke University for facilitating the analyses of the data. Thanks to George Rosenberg and Jennifer Hathaway of the Molecular Biology Facility. Thanks to Jennifer Kavka for arranging my trips to several conferences.

Thanks to Jon Olp and Alana Sharp for helping with the collection of data.

Thanks to Zuly Parra for being a good friend who has been helping me in many ways.

Finally, I want to thank my parents Yuncheng Wang and Lijun Liu who have always supported and trusted me. I cannot become the person who I am today without their love.

My research was financially supported by: the National Science Foundation grant (IOS 0641382) awarded to Robert D Miller.

Lymphocyte Evolution and Ontogeny In Non-Eutherian Mammals

By

Xinxin Wang

M.S., Biology, Institute of Hydrobiology,

Chinese Academy of Sciences, 2006

Ph.D., Biology, University of New Mexico, 2012

ABSTRACT

Based on their different reproductive modes and evolutionary history, extant mammals are divided into three lineages: eutherians, marsupials and monotremes. Marsupials are the closest relatives to eutherians. They are born immune-incompetent. The characteristic that they develop their immune system postnatally makes them unique models to study newborn immune development and maternal immunology. As monotremes comprises the basal lineage of mammals, the study of monotreme immunology will provide unique insights regarding the origin of the mammalian immune system. Unfortunately, compared to our in-depth knowledge of eutherian immunology, especially humans and mice, knowledge of marsupial and monotreme immunology is scarce.

The goal of the first part of this research was to investigate the development of immune-competence in a model marsupial, the gray short-tailed opossum, *Monodelphis domestica.* This research specifically looked at the ontogeny of B cells with emphasis on diversity of immunoglobulin genes during postnatal development. It takes advantage of genomic resources available for the opossum and a captive colony at the UNM Biology Department. To achieve these goals, the content and genomic organization of Ig heavy chain and light chain loci was determined. Opossum Ig heavy chain are of low germ-line diversity while light chains have high germ-line diversity. This suggests that opossums rely more on light chains than heavy chains for repertoire diversity (Wang et al. 2009).

Using the detailed genomic information available of Ig loci, the timing of B cell ontogeny and Ig repertoire diversity was then determined. Opossum newborns start heavy chain VDJ recombination within the first 24 hours postpartum. The expression of the surrogate L chains occurs at day 6 postnatally. The subsequent rearrangement of the Igλ and Igκ L chain genes occur at days 7 and 8 postnatal, respectively. The diversity of early B cell H chains is limited and reduced in N region additions, as has been seen in fetal humans and mice, but lacks bias in the V, D and J segments used. Different from H chains, L chains develop much more diverse VJ recombinations and high IgL repertoire diversity when first expressed. Collectively the results demonstrate that B cell development is entirely postnatal in the opossum. The earliest time-point that an opossum has mature B cells is at the starting the second week of life (Wang et al. in preparation). These results are consistent with earlier work demonstrating that most marsupial species, including opossums are unable to generate an antibody response until the second week.

A second goal of my research was the characterization of a novel T cell receptor in the duckbill platypus *Ornithorhynchus anatinus.* Although different from the previous goal, it nonetheless uses a non-eutherian model to address broader questions regarding immunoglobulin and T cell receptor gene evolution. TCRμ is a new T cell receptor that was first identified in marsupials and does not exist in eutherians (Parra et al. 2007). Homology searches of the platypus genome with opossum TCRμ sequence have identified a homologue of this unconventional TCRμ in platypus. Platypus TCRμ is expressed in a double V domains structure and these resemble Ig V more than conventional TCR V domains. Different from opossum TCRμ, platypus TCRμ requires two rounds of somatic recombination to assemble both V domains. The identification of TCRμ in platypus indicates that TCRμ as an ancient T cell receptor has been lost in eutherians.

Outline of dissertation

Chapter one is published as Baker, M.L., Wang, X., and Miller, R.D. Marsupial immunoglobulin and T cell receptor genomics. In: Deakin J, Waters P & Graves J, editor. *Marsupial Genetics and Genomics*. New York: Springer; 2010. pp. 357-380. This chapter is an overview of marsupial immunology and summarizes the organization, content, and evolutionary history of what is known about marsupial immunoglobulin and T cell receptor loci from a genomics perspective. It serves as background for what is presented in chapters two and three. This chapter also introduces the discovery of the unconventional T cell receptor, TCRμ in opossums. This provides a background overview for the study of TCRμ in platypus in chapter four.

Chapter two is published as Wang, X., Olp, J.J., Miller, R.D.2009. On the genomics of immunoglobulins in the gray, short-tailed opossum *Monodelphis domestica*. Immunogenetics 61: 581-596. This chapter presents a detailed genomic analysis and annotation of the Ig loci in the short-tailed opossum *Monodelphis domestica* generated from analyses of the available whole genome sequence for this species. This is the first non-eutherian mammal, and one of the few vertebrate species of any lineage, which such a detailed organization of the Ig loci has been determined and fully annotated. This annotation serves as a resource for establishing the genomic contribution to the Ig repertoire diversity during B cell ontogeny. This genomic study also identified a germ-line joined VH gene, VH3.1, which is the only one thus far known from in mammals. This chapter serves as additional background for chapter 3.

Chapter three is formatted for a planned submission to the Journal of Immunology as Wang, X., Sharp, A.R., Miller, R.D. Postnatal ontogeny of the Ig repertoire and B cell maturation in the opossum, *Monodelphis domestica*. Marsupials are unique among mammals because they achieve their immune-competence after birth. This study is the first to investigate marsupial B cell lineage development at a molecular level starting prior to birth. The results support B cell ontogeny in opossums being entirely postnatal. Using the Ig genomic information presented in chapter two, this chapter analyzes the genomic contribution of Ig repertoire diversity during B cell ontogeny. Also in this study the only germ-line joined VH gene thus far known from mammals, VH3.1, has been analyzed for its contribution to Ig repertoire diversity.

Chapter four is published as Wang, X., Parra, Z.E., Miller, R.D. 2011. Platypus TCRμ provides insight into the origins and evolution of a uniquely mammalian TCR locus. Journal of Immunology 187: 5246-5254. This chapter is very different from the other chapters in that it focuses on monotreme immunology. This chapter presents the characteristics of TCRμ in the platypus. TCRμ is an unconventional T cell receptor that has been identified in marsupials (one lineage of non-eutherian mammals) but is absent in eutherian species thus far studied. The identification of TCRμ in the platypus, which presents another lineage of non-eutherian mammals, provides insight that TCRμ as an ancient T cell receptor that has been lost in eutherians thus far examined. TCRμ V domains resemble Ig V domain more than conventional TCR V domains. The origin of TCRμ in platypus provides unique insight into understanding the evolutionary history of the genes encoding immunoglobulins and T cell receptors.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 4

LIST OF TABLES

CHAPTER 1

MARSUPIAL IMMUNOGLOBULIN AND T CELL RECEPTOR GENOMICS

By Michelle L. Baker[#], Xinxin Wang, and Robert D. Miller*

*Center for Evolutionary and Theoretical Immunology and Department of Biology, University of New Mexico, Albuquerque NM 87131 USA. [#]Current address: CSIRO Livestock Industries, Australian Animal Health Laboratory, PO Bag 24, Geelong, Victoria 3220, Australia.

MLB and RDM drafted the manuscript. XW collected data, data analyses and generated all figures.

Manuscript published in In: Marsupial Genetics and Genomics (2010) Ed Deakin J, Waters P & Graves J; Springer New York, pp 357-380.

Abstract

The development of marsupial genomic resources has provided a wealth of information for the study of the development and function of the adaptive immune system in these remarkable mammals. Here is reviewed the organization, content, and evolutionary history of the genes encoding the antigen specific receptors of the adaptive immune system, the immunoglobulins and T cell receptors. Included is one of the more unexpected outcomes of marsupial genomics, the discovery of a previously unknown TCR chain, not found in eutherian mammals. In addition, how the genomic complexity of the Ig and TCR genes influences their expression and its relationship to what is known about marsupial immune responses is also presented. Although this review focuses primarily on work done for *Monodelphis domestica*, research on other marsupial species is also included.

Introduction

The evolution of the gnathostomata, or jawed vertebrates, was accompanied by a number of innovations in the adaptive immune system, the hallmarks of which are antigen specific responses resulting not only in the control and elimination of pathogens but also the establishment of immune memory. Unique to the jawed vertebrates are two antigen receptor systems that mediate specific recognition of antigens: the immunoglobulins and the T cell receptors. Immunoglobulins (Ig) are synthesized by B lymphocytes and are expressed both as cell surface molecules (the B cell receptor [BCR]) where they participate in antigen recognition and cell activation, and also as secreted forms (antibodies) where they directly mediate pathogen clearance. T cell receptors (TCR) on the other hand exist solely as recognition and activating receptors on the surface of T lymphocytes. The mode of antigen recognition also differs between Ig and TCR. Ig generally bind intact, native antigen structures, whereas TCR bind antigens that have been processed into typically small peptide fragments that are presented on the cell surface bound by molecules of the Major Histocompatibility Complex (MHC).

The protein domains of both Ig and TCR that directly contact antigens are called the variable (V) domains. A unique feature of these domains is that they are encoded by exons that exist as incomplete or fragmented gene segments in the germ-line genome (Tonegawa 1983). The gene segments, named V, D (diversity), and J (joining) genes, are assembled into a complete exon through somatic DNA rearrangements in developing lymphocytes, the so-called V(D)J recombination. This process is dependent on the

recombination activating gene (RAG) recombinase system, which is an endonuclease complex that recognizes specific nucleotide sequences, called the recombination signal sequences (RSS), that flank each gene segment (reviewed in Sekiguchi et al., 2004). The possible combinations of gene segments, along with variation in nucleotides at the junctions between recombined segments, results in the typical mammalian Ig and TCR loci being able to encode on the order of 10^{15} to 10^{17} different receptor specificities. It is this vast potential receptor repertoire that is the basis for having clonally diverse B and T cells, providing a broad recognition coverage of rapidly evolving pathogens.

The V, D, and J gene segments, along with the exon encoding the constant (C) regions of the Ig and TCR, are generally found organized in two possible ways: translocon or cluster. The translocon organization is the most common and the V, D, and J gene segments exist in distinct arrays as $(V)_n$ - $(D)_n$ - $(J)_n$ - C. This organization is found in all conventional TCR genes analyzed so far, and the Ig genes of boney fish and tetrapods (Flajnik 2002). Based on recent analysis of the genomic organization TCRγ chain in a cartilaginous fish it appears that the translocon is the ancestral form for TCR (Chen et al. 2009). For clarification, the term "conventional" TCR is meant to denote the four TCR chains that are known to be present in all jawed-vertebrates, α, β, γ and δ, and to distinguish them from the atypical TCRμ, found only so far in marsupials and monotremes (Parra et al. 2007, 2008). Genes with the cluster organization have the gene segments organized in cassettes of $(V – [D] – J – C)_n$ with multiple, sometimes hundreds of such cassettes or clusters being found throughout the genome. This organization has

mostly been found in the Ig genes of cartilaginous fishes, and the number of D segments per cluster can vary (Flajnik 2002).

Although clearly sharing a common origin, Ig and TCR have followed distinct evolutionary paths, resulting in contrasting degrees of conservation and plasticity between the two receptor systems. For example, two isotypes of TCR have been identified in all jawed vertebrates: the αβTCR, composed of a heterodimer of an alpha and beta chain, and the γδTCR composed of gamma and delta chains. The conservation of these two isotypes across the whole of jawed vertebrates may be a reflection of their use solely as receptors for antigen recognition. Only recently have TCR isotypes other than the conventional α, β, γ, and δ chains been discovered. These include the TCRμ chain found in marsupials and monotremes, but apparently lacking in eutherian mammals, as well as the NAR-TCR chain in cartilaginous fishes (Baker et al 2005a, Criscitiello et al 2006, Parra et al. 2007, 2008)

In contrast to the relatively conserved nature of the TCR, the Ig have evolved a wider array of isotypes, some of which appear to be specific to particular vertebrate lineages. The typical Ig is composed of four proteins, two identical heavy chains and two identical light chains, with the isotype of the secreted antibody being determined by its heavy chain. Differences in the C regions of the heavy chains impart different functional roles on the isotype of the secreted antibody. Some isotypes, such as IgM, have been conserved throughout gnathostome evolution, whereas others such as IgG and IgE are lineage specific (mammalian in this case). IgD, on the other hand, is an apparently

ancient Ig isotype but appears to have a spotty distribution across vertebrate lineages (Ohta and Flajnik 2006).

In addition to there being a greater diversity of effector functions associated with Ig heavy chain C regions, there is also greater variation in the genetic contributions to diversity of binding specificities in antibodies among species. Notably there is an association between genomic organization and complexity of the Ig genes and patterns of B cell ontogeny (reviewed in Butler 1997). In rabbits and chickens for example, V(D)J recombination alone is not sufficient to produce a diverse primary antibody repertoire and in both these species developing B cells migrate to a secondary site (the appendix in rabbits and the Bursa of Fabricius in chickens) where the rearranged V(D)J segments are further mutated before the B cells enter the primary pool. In contrast mice and humans are able to generate diverse antibody specificities solely through V(D)J recombination due to a diverse and complex array of gene segments (reviewed in Butler 1997). Therefore, even lineages as closely related as lagomorphs (rabbits) and rodents (mice) can be distinctly different in how their primary antibody repertoires are constructed and these are a direct reflection of the genomic content and organization of their Ig genes.

The altricial nature of the newborn marsupial makes them ideal models to study early development in the immune system and the evolution of maternal immunity. Some cross-reactive antibodies directed against evolutionarily conserved peptide sequences present in human T and B cells and a limited number of marsupial specific reagents have been used successfully for immunohistochemistry on marsupial tissues and cells from adults

(Wilkinson et al, 1995; Jones et al., 1993; Hemsley et al., 1995; Baker et al., 1999) and developing marsupials (Coutinho et al., 1993, 1995; Canfield et al., 1996; Baker et al., 1999; Cisternas and Armati, 2000). Overall, these studies have demonstrated a pattern of T and B cell distribution in lymphoid tissues similar to that of eutherian mammals. However, overall such research has been limited by a scarcity of marsupial specific reagents.

The completion of the first marsupial genome project along with extensive analysis of expressed Ig and TCR chains has provided a near complete picture of the genomic contributions to Ig and TCR diversity and expression in the gray short-tailed opossum, *Monodelphis domestica* (Mikkelsen et al., 2007; Parra et al., 2008, 2009; Wang et al., 2009). The level of detail available helps put this species on par with more developed eutherian model organisms such as mice and humans. In addition, the Ig and TCR genes of several other marsupial species have been analyzed at least at the level of cDNA clones providing the opportunity to make some generalizations regarding the evolution and function of these receptors in metatherians.

Marsupial B cell development, antibody responses, and Ig genetics

Some of the earliest experiments performed on marsupial immune systems were measuring antibody responses (Kalmutz 1962). This is not surprising given that the methods for detection and quantification of serum Ig were developed by the early 20th century. Early studies of antibody responses in marsupials were consistent with differences in both the kinetics and magnitude of primary and secondary antibody

responses compared with eutherian mammals. Prolonged primary antibody responses have been reported in *M. domestica*, *Setonix brachyurus* and *Trichosurus vulpecula*, with elevated IgG responses lasting 37, 26 and 15 weeks respectively (Croix et al., 1989, Yadav 1971; Deakin et al., 2005). Secondary responses have been described as being delayed and lower in magnitude compared with those of eutherian mammals. In addition, the presence of a substantial amount of IgM in the secondary response supports poor isotype switching in marsupials (Shearer et al., 1995; Croix et al., 1989; Stanley et al., 1972; Wilkinson et al., 1992a). More recently, studies using variations on the traditional immunization protocols used for eutherian mammals have reported more comparable antibody responses to those of eutherian mammals (Kitchener et al., 2002, Asquith et al., 2006; Deakin et al., 2005; McLelland et al., 2005).

The immunoglobulin heavy chain locus

From the analysis of cDNA sequences it was determined that marsupials have a typically mammalian repertoire of Ig isotypes. Four Ig heavy chain (IgH) isotypes (M, G, E, and A) have been described in marsupials (Miller et al., 1998; Aveskogh and Hellman 1998, Aveskogh et al., 1999; Belov et al. 1999a, 1999b, 1999c; Daly et al., 2007). The only isotype that appears to be missing is IgD, which is found in the other two mammalian lineages, the eutherians and the monotremes, and some non-mammalian lineages (Ohta and Flajnik 2006; Wang et al. 2009).

Prior to the completion of the *M. domestica* genome, the locations of the genes encoding the IgH chains (*IGH*) were localized to chromosome (chr) 1 using *in situ* hybridization on

metaphase chromosomes (Deakin et al 2006a). The availability of the opossum whole genome sequence has provided a more detailed view of the genomic landscape surrounding the *IGH* locus and its overall organization. The loci on the centromeric side of *IGH* share a high degree of conserved synteny between opossum chr 1 and human and mouse chr 14 and 12, respectively (Figure 1). Genes on the telomeric side of opossum *IGH* are also found on human chr 14 and mouse chr 12. This is not contiguous, conserved synteny since, in contrast to opossum, human and mouse *IGH* is the most telomeric locus on their respective chromosomes. The overall architecture of the opossum *IGH* locus is fairly typical for mammalian orthologues with a translocon organization (Figure 2). This pattern is only broken by the presence of two D segments interspersed amongst the V segments and a duplication of a region containing three J segments and partial, pseudogene copy of the IgM constant regions (Figure 2; Wang et al 2009). This latter duplication contains an insertion of a long interspersed (LINE) element. Insertion of retro-elements, LINE and endogenous retroviral (ERV) elements, has clearly contributed to the evolution of the *IGH* locus in the opossum, particularly in the region encoding the constant genes (Wang et al 2009). They are not only associated with both an unusual duplication that resulted in a pseudogene copy of IgM but also may have contributed to the loss of IgD in *M. domestica* (see below). The exons encoding the constant regions are also in the conserved order of 5'-M > G > E > A-3' (Figure 2).

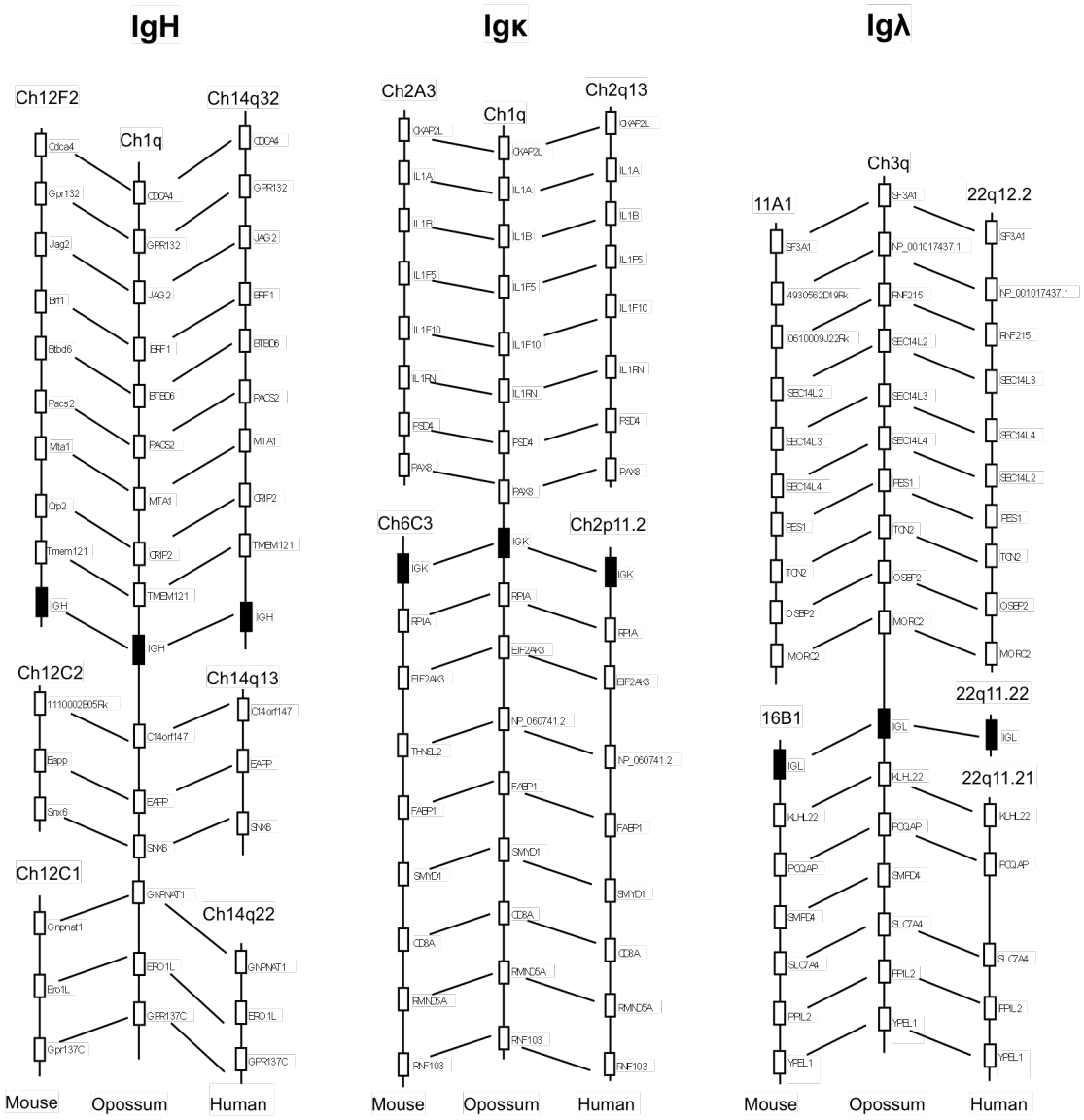


Figure 1. Chromosome maps of the regions containing the *IGH* (right), *IGK* (center), and *IGL* (left) loci, comparing the opossum (*M. domestica*) with the regions in human and mouse bearing conserved synteny. The maps are oriented with the opossum chromosome oriented as centromeric at the top and telomeric at the bottom. Gene lists and locations used are based on the current Ensembl annotations for opossum, human, and mouse genome assemblies (www.ensembl.org).

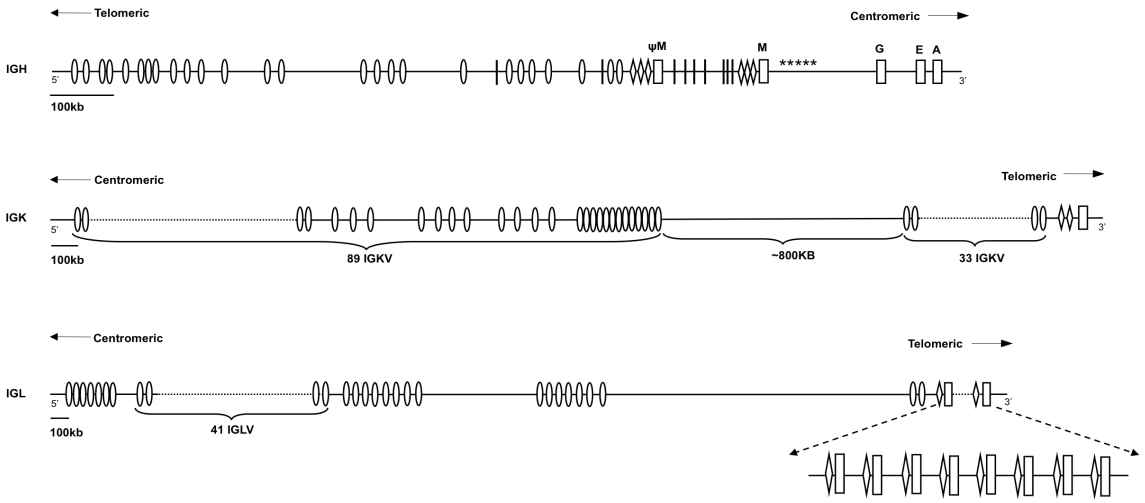


Figure 2. Gene maps of the *IGH, IGK,* and *IGL* loci. Vertical ovals indicate V gene segments, vertical lines indicate D segments, vertical diamonds indicate J segments, and vertical rectangles indicate C regions. For the *IGH* locus, C region genes are distinguished based on isotype (M, G, E, and A) by letters above the rectangle. ψ indicates pseudo-gene. Asterisks indicate region rich in L1 and ERV elements immediately 3' of the functional IgM constant region exons. Brackets below the IGK locus indicate the two large clusters of V gene segments separated by a region devoid of these coding sequences. An expanded view of the eight J-C clusters in the *IGL* locus are shown below the main line.

From the analysis of Ig heavy chain transcripts in the opossum and other marsupial species there appeared to be only a single IgG. This is in contrast to predictions made from analysis of serum Ig in *M. domestica* and other marsupials (Bell et al., 1974; Bell 1977; Shearer et al., 1995). Analysis of the whole genome assembly revealed only a single set of exons encoding conserved IgG constant regions. The genomic organization and content are consistent with what was predicted from the cDNA analysis of there being three extra-cellular constant region domains and a single hinge region encoding exon in marsupial IgG (Aveskogh and Hellman 1998; Belov et al., 1999a). The discrepancy between the serum protein data and the genomic data may have a number of possible explanations. First Ig heavy chains are often glycosylated and it is possible that there are distinct patterns of glycosylation that made it appear that there was more than one IgG subclass based on molecular weight. Alternatively, the methods used to purify the serum IgG may have led to false conclusions. The majority of these experiments used Staphylococcal Protein A (SpA) based purification that is expected to preferentially enrich for IgG through binding to its C domain. However, marsupial IgM has also been shown to bind SpA, so it is possible that this purification method inadvertently enriches for multiple heavy chain isotypes. SpA has also been shown to act as a super-antigen for human Ig containing clan III V domains (Silverman and Goodyear 2002). Since all marsupial VH are also clan III it is possible that SpA is binding the V domains of some marsupial Ig rather than the C domains (Baker et al., 2005b; Wang et al., 2009).

The exons encoding IgD are typically immediately downstream of those encoding IgM and IgD and IgM can be co-expressed in some species through alternative mRNA

splicing. A thorough search of the *M. domestica* genome sequence failed to identify any coding sequences with homology to IgD (Wang et al., 2009). Furthermore, the region of the opossum genome expected to contain IgD contains two large repetitive insertions, each over 15 kb in length composed of a mixture of LINE and ERV elements (Wang et al 2009). Whether a similar insertion is present in other marsupials isn't known. How old these insertions are and if they contributed to the loss of IgD in the opossum lineage are also not known. Nonetheless, it is pretty clear that the absence of IgD in the opossum is due to a gene loss given the ancient nature of this isotype and its presence in the other mammalian lineages (Ohta and Flajnik 2006; Wang et al. 2009).

The diversity of expressed IgH V domains (VH) has been published for two marsupial species, *M. domestica* and the brushtail possum *T. vulpecula* (Miller et al 1998, Aveskogh et al., 1999; Baker et al., 2005b). In addition the VH repertoire of the tammar wallaby *Macropus eugenii* has also been extensively sampled at the cDNA level (Figure 3). Several general conclusions could be drawn from the analysis of the diversity of marsupial Ig cDNA clones. The first is all expressed marsupial VH belong to a single evolutionary clan, called clan III. This is an ancient and ubiquitous clan and is the most conserved of all the VH clans so it is not surprising that they are present in marsupials (Tutter and Riblet 1989). What is surprising is that all marsupials species studied so far have only clan III VH genes and they are all highly similar, forming a "marsupial-only" monophyletic clade (Baker et al., 2005b). This is not due to the analyses being performed on species that are closely related. The divergence times for *M. domestica, T. vulpecula,* and *M. eugenii* are on par with that of eutherian lineages such as humans and

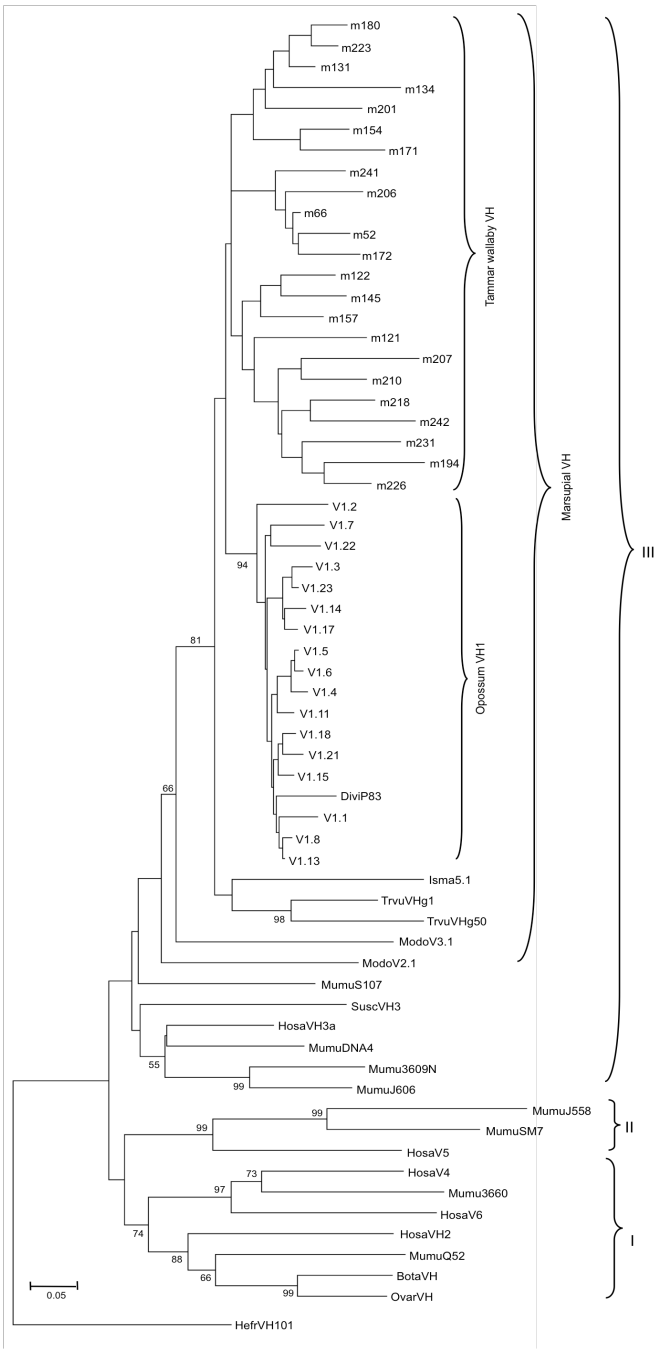


Figure 3. Phylogenetic analysis of VH gene sequences from marsupials and a representative other mammalian and non-mammalian species. The Neighbor Joining method of Saitou and Nei contained within the MEGA package was used for this analysis (Kumar et al., 2004). A VH sequence from the horn shark, *Heterodontus francisci* (Hefr) is used as the outgroup. On the right, the three major clans of VH genes are indicated as I, II and III. The single large marsupial specific VH clade is also indicated and has high bootstrap value. The two large clades containing tammar wallaby and opossum VH1 are also bracketed. Other species included are: the Northern brown bandicoot *Isoodon macrourus* (Isma), brushtail possum *Trichosurus vulpecula* (Trvu), *M. domestica* (Modo) VH that are not VH1, Virginia opossum *Didelphis virginiana* (Divi), mouse *Mus musculus* (Mumu), human *Homo sapien* (Hosa), cow *Bos taurus* (Bota), and sheep *Ovis aries* (Ovar).

mice that have more diverse VH repertoires (Baker et al., 2005b; Das et al., 2008).

One of the more unusual features in the opossum *IGH* locus is the presence of a partially germ-line joined V gene, designated VH3.1 (Wang et al., 2009). This gene segment is longer than a typical V gene by 14 codons and appears to be a V segment already fused to a D segment in the germ-line DNA. The sequence and genomic organization of this VH gene segment has been confirmed by direct sequencing of *M. domestica* genomic DNA to ensure that it is not an assembly artifact (see GenBank accession no. EU592040). Typical Ig and TCR V gene segments are composed of two exons separated by a short intron. Exon 1 encodes most of the leader (L) sequence and exon 2 encodes most of the extracellular V domain and ends at the RSS. The partially germ-line joined VH3.1 gene segment retains both the intron and RSS and the latter contains a 12 base pair spacer typical of DH segments, rather than the 23 bp spacer of a VH segment RSS. VH3.1, therefore, appears to be the result of RAG-mediated recombination (Wang et al., 2009). Partial or fully germ-line joined gene segments have been found in the Ig loci of cartilaginous and boney fishes and birds, however this is the first such Ig gene segment to be described in a mammal (Kokubu et al., 1988; Reynaud et al., 1989; Ventura-Holman and Lobb 2002). Their origin is thought to be due to ectopic activation of V(D)J recombination in germ cells (Lee et al 2000). Full or partially germ-line joined V genes have been shown to contribute to the antibody repertoires in non-mammalian species. In sharks they are used in the early neonatal antibody repertoire (Rumfelt et al., 2001). In chicken they contribute by being a source of sequence variation by modifying the primary antibody repertoire through gene conversion mechanisms (Reynaud et al., 1989).

Interestingly the structure of VH3.1 resembles that of V pseudogenes found in the chicken *IGH* locus, which are also partially germline joined V to D gene segments (Reynaud et al. 1989). VH3.1 appears to be completely functional, with an open reading frame (ORF) and canonical RSS (Wang et al 2009). Whether it contributes to antibody diversity in the opossum is not known, however, and transcripts containing VH3.1 in adult opossum splenic RNA have not been found (unpublished observation). So far there also does not appear to evidence of gene conversion influencing the opossum repertoire either.

The Ig light chain loci

Marsupials, like all mammals investigated so far, have two Ig light chain isotypes: kappa (Igκ) and lambda (Igλ). Light chain gene sequences have been reported for *M. domestica, T. vulpecula,* and *M. eugenii* (Miller et al 1999, Lucero et al 1998, Belov et al., 2001, 2002a, Baker et al., 2005b; Daly et al., 2007). The *IGK* and *IGL* loci have been localized to the long arms of chromosomes 1 and 3, respectively, in *M. domestica* (Deakin et al., 2006a). The region of opossum chr 3 containing *IGL* has conserved synteny with human chr 22 (Figure 1). This conservation is not as well maintained in mice, however, where chr 16 has conserved synteny with human chr 22 and opossum chr 3 on the telomeric side of *IGL*, and 11 on the centromeric side. The *IGK* locus in opossum is also at a break point in synteny with eutherian mammals. This region of opossum chr 1 is conserved with regions on both short and long arms of human chr 2 and with mouse chr 2 and 6 (Figure 1).

The overall genomic organization of the opossum *IGK* and *IGL* loci are typical of mammalian light chain genes and match what was predicted from analysis of cDNA clones (Miller et al 1999, Lucero et al 1999, Wang et al 2009). This includes a single C region and only two J gene segments in *IGK*, and eight J and C pairs in *IGL* (Figure 2). The V gene segments in *IGK* are organized in two large clusters, with inverted reading orientation, separated by an approximately 800 kb region. Inverted V genes are not a problem since V(D)J recombination can occur through inversion rather than deletion of intervening DNA (Sekiguchi et al., 2004). This organization is similar to that seen in human *IGK* where it is clear that a large inverted duplication occurred during the evolution of this locus (Zachau 2004). This does not appear to be what has occurred in *M. domestica*, however. Phylogenetic analysis of the V gene segments in each cluster reveal that they belong to distantly related subgroups (Miller et al., 1999; Wang et al., 2009).

In contrast to the limited diversity of marsupial VH genes, the light chain V region genes, Vκ and Vλ, are highly diverse and the different marsupial Vκ and Vλ clades intersperse amongst that of other mammalian lineages suggesting the retention of a more ancient diversity than in VH (Miller et al 1999; Lucero et al 1999; Baker et al 2005b, Wang et al 2009). This has resulted in speculation that Ig light chains contribute more to the overall antibody diversity in marsupials than does heavy chains (Baker et al., 2005b).

Marsupial B cell ontogeny and immuno-competence

The availability of fully annotated *IGH, IGK* and *IGL* has facilitated the analysis of both B cell ontogeny and diversity. Transcripts containing IgM heavy chains with completely recombined V genes can be detected within the first 24 hours of postnatal life in *M. domestica* (unpublished observation). This is significantly earlier than what has been reported previously for the brushtail possum *T. vulpecula*, where the earliest time-point investigated was day 10 (Belov et al., 2002b). The diversity of the IgH chains at these early time-points is limited and the V-D-J junctions are devoid of additional nucleotides added by the enzyme terminal deoxynucleotidyl transferase, which adds N nucleotides without the use of a template. N nucleotides are found in the adult repertoire (Miller et al., 1998; unpublished observations). Whether this limited diversity is due to limited numbers of B cells in the newborn, or represents a programmed bias as has been seen in eutherian fetal antibodies is not known. The absence of N nucleotides in early opossum IgH chains is similar to what has been seen for mouse neonatal antibodies and may be an analogous ontogenic pattern (Feeney 1992; Benedict et al., 2000).

In contrast to IgH, light chain gene transcription is not detectable until day 8 of postnatal life in *M. domestica* (unpublished observation). This suggests that, although B cell ontogeny is initiated peri-natally, mature functional B cells are not present until the second week. These results support that mature B cells also appear later in development than mature conventional T cells in the opossum (Parra et al 2009). These results are also consistent with early studies of antibody responses in a related marsupial species, the Virginia opossum *Didelphis virginiana*, which found antibody responses to a T

dependent antigen not being detectable until the newborn opossums were at least two weeks of age (Kalmutz 1962; Rowlands et al., 1964). Therefore, the newborn marsupial is fully dependent on maternal antibodies for protection during the first week, which, in *M. domestica*, are entirely transferred via the milk (Samples et al. 1986).

T cell receptor genomics

Critical to the function of T cells are their TCR, which establish not only the antigen specificity of individual clones but also their phenotype and function. T cells that utilize αβ TCR differentiate into the well-known helper, cytotoxic (killer), and regulatory (suppressor) phenotypes. This differentiation is determined by selection for combinations of binding affinity for self-peptides and self-MHC molecules in the thymus. The random combinations of V, D and J segments encoded at the TCR loci establish these binding affinities. Hence, understanding the structure and organization of the TCR genes in any given species is critical to understanding T cell function and fate.

Marsupial T cell mediated responses have attracted attention, in part, because of a general sense that they are deficient relative to that of eutherian mammals. For example, rejection of primary skin grafts occurs with similar kinetics in marsupials and eutherians, however marsupial second set rejection is relatively slower (Infante et al., 1991; Stone et al 1997). Delayed type hypersensitivity (DTH) responses are also slower and weaker in magnitude than those of eutherians (Turner et al., 1972; Taylor and Burrell, 1968). *In vitro* mitogenic responses in all marsupials tested to date are similar to eutherian mammals, however mixed lymphocyte reactions (MLR) are negligible or absent (Stone et al., 1997;

Fox et al., 1976; Wilkinson et al., 1992a; 1992b; Stone et al., 1998; Baker et al., 1998; Baker and Gemmell, 1999; Baker et al., 1999; Turner et al., 1972; Woods et al., 2007). There is no clear explanation for the weak MLR or slow DTH responses in marsupials. Low MHC polymorphism could account for the absence of MLR observed in Tasmanian devils, which have undergone a considerable population bottleneck (Woods et al, 2007). However, in at least one other marsupial, *M. domestica*, polymorphism at classical MHC genes is comparable to human HLA, so absence of MHC diversity in this species is not a likely explanation for poor MLR (Gouin et al., 2007). In addition, antibody isotype switch in secondary immune responses in marsupials are not as robust as they are in eutherian mammals (Stone et al., 1996; Croix et al., 1989; Stanley et al., 1972; Wilkinson et al., 1992a). Since isotype switching is also T cell dependent, a number of investigators have questioned the sufficiency of T cells in humoral immune responses in marsupials as well. The availability of marsupial genomic resources has provided the means to investigate T cell development and diversity to gain some insight into marsupial T cell biology.

T cell maturation in all gnathostomes is dependent on a functional thymus, the gland-like lymphoid organ typically located in the upper thoracic region. It is in the thymus where T cells commit to a lineage, αβ or γδ, and where they undergo V(D)J recombination and selection to establish their antigen specificity and MHC restriction. A unique, and so far not fully understood, aspect of thymus biology in marsupials is the presence of an extra set of thymuses in the cervical vertebrae region in some family groups. The diprotodont species, including Burramydae, Phalangeridae, Tarsipedidae and Macropodidae, have

both cervical and thoracic thymuses with the exception of the koala that possesses only a cervical thymus. The polyprotodont species, including Caenolestidae, Dasyuridae, Thylacinidae, Notoryctidae and Peramelidae, possess only a thoracic thymus, and Didelphidae marsupials have paired thoracic thymuses (Deane and Cooper, 1988; Yadav, 1973; Hubbard et al., 1991; Haynes, 2001).

Although the thymus is present in newborn marsupials it is generally undifferentiated and undergoes much of its development during postnatal life. In marsupials with both cervical and thoracic thymuses both appear to be histologically identical and to have the same function, but the development of the thoracic thymus generally lags behind that of the cervical thymus (Kathiresan, 1969; Stanley et al., 1972). T lymphocytes differentiate in the thymus before populating other sites in the body. Peripheral lymphoid organs such as the spleen, lymph nodes and gut associated lymphoid tissue typically develop into functional lymphoid tissue only after the thymus has reached maturity. T cell responses are correspondingly low at birth, and neonates are dependent on maternal antibody for protection until their immune systems are functional (Hubbard et al., 1991; Samples et al., 1986).

Immunohistochemistry using cross-reactive antibodies on developing lymphoid tissue from marsupials has provided evidence for the appearance of T cells soon after birth. $CD3^+$ T cells have been identified in the thoracic thymus as early as day 2 postpartum in *T. vulpecula* and *D. albiventris* (Baker et al., 1999; Coutinho et al., 1995). In contrast, the thymus of *M. eugenii* does not have CD3 positive T cells earlier than day 12

postpartum (Old and Deane, 2003). Subsequent appearance of $CD3^+$ T cells appear in the spleen by day 21 and 25 postpartum in *M. eugenii* and *T. vulpecula* respectively (Baker et al., 1999; Old and Deane, 2003). These immunohistochemistry results are somewhat consistent with recent analysis of T cell ontogeny in the opossum *M. domestica* based on expression of mature, rearranged TCR transcripts described below (Parra et al. 2009).

As with the Ig genes, the annotation of the TCR loci was greatly assisted by the published analysis of marsupial cDNAs encoding TCR chains, analysis which has already revealed considerable homology between marsupial TCRα, β, γ and δ chains to those of eutherian mammals. TCR genes encoding the α and β chains of the αβ TCR have been characterized in *M. domestica*, *T. vulpecula* and *M. eugenii*, and for the γδ TCR, the δ chain has been characterized in *M. eugenii*, *M. domestica* and *Isoodon macrourus* and the γ chain has been identified among expressed sequence tags (EST) from *I. macrourus* (Baker et al., 2001; Zuccolotto et al., 2000; Harrison et al., 2003; Baker et al., 2005a; Baker et al., 2007). The diversity of expressed V domains in cDNAs has been described only for the TCRδ chain from *M. eugenii* and the TCRα and β chains of developing *M. domestica*. At least three different subfamilies of *M. eugenii* Vδ genes were expressed in cDNA from mammary gland associated lymphoid tissue, consistent with a high level of V segment diversity in *M. eugenii* (Harrison et al., 2003). In *M. domestica*, a highly diverse Vα and Vβ repertoire was observed, with preferential use of V segments during development as discussed below (Parra et al., 2009).

Physical mapping of the conventional TCR loci in *M. domestica* has localized TCRα/δ to chromosome 1p, TCRβ to chr 8q, and TCRγ to chr 6q by *in situ* hybridization of metaphase chromosomes (Deakin et al., 2006b). The availability of the whole genome sequence of *M. domestica* has provided a more detailed view of the genomic organization of TCR loci in this marsupial, demonstrating a high degree of synteny with TCR loci from other mammals and a translocon arrangement (Figures 4 and 5; Parra et al., 2008).

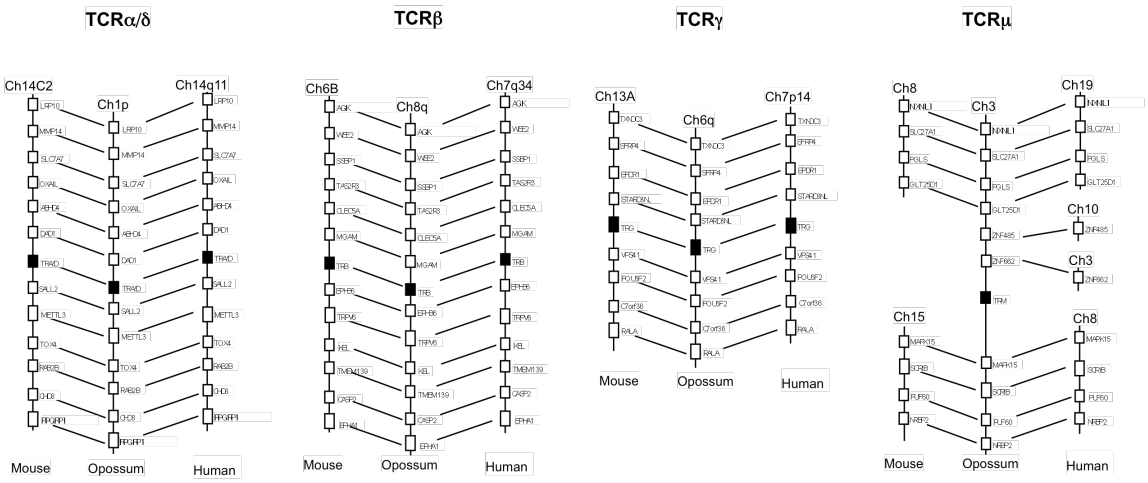


Figure 4. Chromosome maps of the regions containing, left to right, the TRA/D, TRB, TRG, and TRM loci, comparing the opossum (*M. domestica*) with the regions in human and mouse bearing conserved synteny. The nomenclature is by the convention where the locus encoding the TCRα/δ chains is TRA/D, TCRβ is TRB, and TCRγ is TRG, TCRμ is TRM. The maps are oriented with the opossum chromosome oriented as centromeric at the top and telomeric at the bottom. Gene lists and locations used are based on the current Ensembl annotations for opossum, human, and mouse genome assemblies (www.ensembl.org).

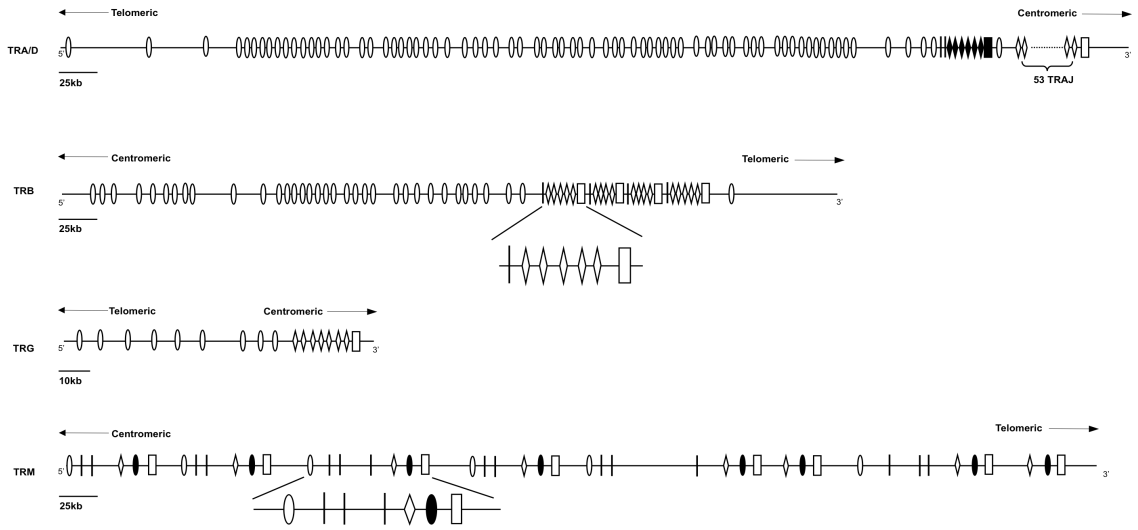


Figure 5. Gene maps of the loci encoding the TCRα, β, γ, δ, and μ chains. Vertical ovals indicate V gene segments, vertical lines indicate D segments, vertical diamonds indicate J segments, and vertical rectangles indicate C regions. The nomenclature is by the convention where the locus encoding the TCRα/δ chains is TRA/D, TCRβ is TRB, TCRγ is TRG, and TCRμ is TRM. In the case of the TRA/D locus the Jα and Cα genes are indicated with open symbols and the Jδ and Cδ as filled symbols. In the TRM locus the filled oval indicates the germ-line joined gene segment Vμj. Below the TRB locus is an expaned view of a single TCRβ D, J, and C cluster. Below the TRM locus is an expanded view of a single TCRμ cluster.

TCRα/δ *locus*

Consistent with physical mapping of the TCR loci, the TCRδ genes in the genome assembly are clustered within the TCRα locus, resembling the organization found in other amniotes (Figure 5, Chein et al 1987; Deakin et al 2006b; Kubota et al 1999). The TCRα/δ locus is the largest and most complex of the conventional TCR loci. The V segments within the TCRα/δ locus are either utilized exclusively by TCRα or by both TCRα and TCRδ chains (Parra et al. 2008, 2009).

The number and complexity of D and J segments in the opossum's TCR loci also appears to be comparable to eutherian mammals and non-mammalian vertebrates (Figure 5, Partula et al., 1996; Parra et al., 2008). For example, there are 53 TCRα J segments located between the TCRδ C and TCRα C all of which appear to be functional. This high number of Jα segments is thought to confer the ability of developing T cells to undergo secondary V to J rearrangements, increasing the possibility of successfully generating a functional TCRα chain (Marolleau et al. 1988).

TCRβ *locus*

In addition to the high level of conserved synteny surrounding the TCRβ locus, the structure of the locus is also highly conserved (Figures 4 and 5). As in humans and mice, the opossum TCRβ D, J and C segments are arranged in tandem cassettes each containing a single D, four or five J, and a single C. There are four such cassettes in the *M. domestica* TCRβ locus, all of which appear to be functional (Figure 5; Parra et al., 2008). Noteworthy is the presence of a single V gene segment on the 3' side of the terminal C

region exon in the opossum (Figure 5). This Vβ gene segment is in the opposite transcriptional orientation to the D, J, and C segments and can and is used by V(D)J recombination creating an inversion rather than a deletion (Parra et al., 2008). This gene segment is conserved in eutherians as well, consistent with this arrangement being ancient in therian mammals (Parra et al. 2008).

TCRγ *locus*

The locus encoding TCRγ chains is the smallest of the conventional TCR loci in *M. domestica*, with the fewest number of V gene segments and only a single C region gene. It is also highly generic in its organization, a model for the definition of the translocon organization (Figure 5).

Genomic organization influences early diversity of αβT cells

Once the content and organization of the conventional TCR has been determined, both the germ-line and somatic contributions to T cell diversity can be determined during ontogeny. Such studies have been conducted using *M. domestica* as the model species. mRNA transcribed from functionally recombined TCRα and TCRβ genes are both detectable within the first postnatal day in newborn opossums, indicating that they are born with T cells that have at least reached the point of development when they can undergo selection for binding MHC molecules and self-peptides (Parra et al., 2009). This development is taking place prior to histological evidence of a clear thymus in this species (Hubbard et al., 1991). TCRδ transcripts were also detectable within the first postnatal 24 hours. In contrast to the other three conventional TCR chains, TCRγ is the

last to be expressed during postnatal development in *M. domestica*, not detected until the second postnatal week (Parra et al., 2009). This was a surprising finding since it suggests that αβT cells appear in ontogeny prior to γδT cells, which is the reverse of eutherian mammals where γδT cells develop first (Allison and Havran 1991; Parra et al., 2009). What biological significance there may be to this difference is not known.

Developing αβT cells rearrange their TCRβ chain genes prior to TCRα and analysis of the V gene segments being used in these early rearrangements indicates that their genomic position in the TCRβ locus does not influence their use. However, there is a preferential use of TCRβ gene segments that share micro-homology between the V, D and J (Parra et al., 2009). This micro-homology is evident in the rearrangements as nucleotides at the VD and/or DJ junctions that could be contributed by either germ-line gene segment. This is not unlike what has been seen in fetal Ig and TCR repertoires in some eutherian species (Feeney 1992; Zhang et al. 1995). In contrast, micro-homology does not appear to bias the rearranged gene segments in TCRα chains. Rather, genomic position of V segments appears to have an influence in early TCRα rearrangments, with there being a bias for the most 3' or C proximal V genes. By the second week of life both TCRα and TCRβ chains are random and diverse, and similar to adult (Parra et al. 2009).

In summary, having spatial information as well as content has revealed differences in the way V(D)J rearrangements are being biased in early developing marsupial T cells, in this case even between different TCR loci within the same cell. How the bias in gene segment

use early in development clearly limits the diversity of T cells, but how this influences function during the first week of postnatal life remains to be determined. The presence of completely rearranged TCR loci does not necessarily indicate that there are functionally mature T cells present. Early studies of early immuno-competence in marsupials have found T dependent responses, both humoral and cell mediated, lacking during the first week of postnatal life (Kalmutz 1962; La Via et al., 1963; Rowlands et al., 1964; Ashman et al., 1975). This was thought to be due to lack of functional T cells at the earlier time-points, but may also be explained by limited diversity in the pool of functional T cells available in the neonate (Parra et al., 2009).

A novel TCR in marsupials

One of the more remarkable discoveries made while examining marsupial genomes, at least from an immunological perspective, was presence of a fifth TCR chain, TCRμ. (Baker et al., 2005a; Parra et al., 2007). TCRμ has all the canonical amino acid residues necessary to form a heterodimer with another chain, however what partner it pairs with remains unknown. TCRμ was originally described as a divergent TCRδ chain due to the homology of its C region (Baker et al., 2005a). However the genes encoding this chain are a separate locus from the conventional TCR loci (Parra et al., 2007). In addition to sequence homology to other TCR, it is clear that TCRμ is a *bona fide* TCR in that it undergoes V(D)J recombination in developing thymocytes (Parra et al., 2007). Based on limited phylogenetic surveys TCRμ also appears to be common to all marsupials but is absent from all eutherian species checked so far (Parra et al., 2007, 2008). Not surprisingly then, the region that flanks the opossum's TCRμ locus on chromosome 3

shares no conserved synteny with the genomes of eutherian mammals (Figure 4; Parra et al., 2008). This lack of conservation between the chromosomal regions in the opossum that contains TCRμ genes with that of non-marsupial species has confounded uncovering the evolutionary history of this locus. This is particularly unfortunate given the apparent hybrid nature of this unusual TCR chain; TCRμ appears, based on sequence similarity, to be a hybrid between a TCR and an Ig (Parra et al., 2007). The constant region of TCRμ is most closely related to TCRδ, while the variable domain is homologous to the variable domains of immunoglobulin heavy chains. It is therefore likely that TCRμ may have been the result of a recombination between ancient TCRδ and Ig loci (Parra et al., 2008). Any evidence of such a recombination however appears to be lost in that the regions flanking the TCRα/δ locus in amniotes is highly conserved, showing no sign of a break in synteny (Figure 4, Parra et al., 2008).

In addition to its apparent hybrid features, *M. domestica* TCRμ is also the first, and so far only, TCR to have a clear cluster style organization, rather than translocon (Figure 5). The number of clusters may vary between species with the opossum having eight and the Northern brown bandicoot, *I. macrourus*, likely having only two for example (Baker et al. 2005a; Parra et al 2008). In the opossum six of the eight clusters are complete with each containing two classes of V segments, as well as D, J and C genes. The two classes of V segments are a single non-rearranged V gene segment (Vμ) that can be somatically recombined with the D and J segments, and pre-joined V segment (Vμj) that is a complete exon that appears to be derived from V, D, and J gene segments already recombined in the germ-line DNA. The two partial clusters lack Vμ and Dμ gene

segments. Phylogenetic analysis of the gene segments in each cluster reveals that the clusters are due to whole cluster duplications. Unlike the Ig genes of cartilaginous fishes, which are also in clusters and the clusters are scattered throughout the genome, the opossum TCRμ clusters are in a tandem array (Parra et al. 2008).

TCRμ is also unusual for a TCR in that it is expressed in at least two mRNA isoforms, designated TCRμ1.0 and TCRμ2.0. Both isoforms are transcribed in the thymus and spleen but TCRμ2.0 appears to predominate in peripheral lymphoid tissue (Parra et al., 2007). The chain predicted to be encoded by the TCRμ2.0 isoform contains two V domains; one that is somatically recombined and another that is Vμj. TCRμ1.0 contains only the pre-joined Vμj domain and is predicted to be structurally more similar to conventional TCR chains (Parra et al., 2007). The TCRμ2.0 isoform is generated when the unrearranged Vμ gene segment is recombined to the D and J segments and this exon is spliced to the Vμj exon during mRNA processing (Parra et al., 2007). The TCRμ1.0 isoform is generated by initiated transcription upstream of the Vμj exon and does not require V(D)J recombination. Given the paucity of TCRμ1.0 transcripts in peripheral lymphoid tissues it is not clear if this is a functionally translated form. Furthermore, only two of the eight opossum TCRμ clusters are capable of producing the TCRμ1.0 isoform due to in frame stops in the leader sequence of Vμj in the other six clusters. All six complete clusters are capable of generating the TCRμ2.0 isoform, and one of the partial clusters is capable of producing the TCRμ2.0 isoform by using Vμ gene segments from upstream clusters (Parra et al., 2007). This latter observation may explain the evolutionary advantage to maintaining the clusters as a tandem array even though they

can function independently. It is possible that inter-cluster V(D)J recombination provides the opportunity to generate additional diversity that would not be possible if the clusters were not closely physically linked.

Vμj is the first, and so far only germ-line joined V segment to be described in a TCR. It and the VH3.1 gene segment in the *M. domestica IGH* locus are the only germ-line joined V genes to be described in mammals for either TCR or Ig, although VH3.1 is a partially joined gene. In contrast to VH3.1, which clearly appears to be due to the action of RAG based recombination on the DNA in a germ cell, Vμj appears to be the result of retro-transposition, involving an RNA intermediate (Parra et al., 2007; Wang et al., 2009). How or where this took place is not clear, although it seems likely to also have been in a germ cell for it to be in the germ-line. The recent discovery of a TCRμ homologue in the duckbill platypus genome, lacking the germ-line joined Vμj gene segment will hopefully give some insight into its origins (Parra et al., 2008, unpublished observations).

The discovery of TCRμ in the marsupial genome illustrates not only the dynamic nature of the evolution of the antigen receptor loci but also the value of investigating the genomes of more, distantly related species. Not long ago, the TCR appeared to be fairly static with all jawed vertebrates having a similar, if not identical, complement of receptor chains. This view is more likely due to the fact that much of comparative immunology was proceeded by discoveries being made first in humans and mice followed by identifying homologues in other species. It is harder to ask the question "what does a non-model species have that is lacking in humans and mice?" Comparative genomics

makes asking such questions feasible. What does TCRμ do? Now that's a tough question to answer.

Conclusions

Comparative genomics provides not only the fodder for understanding the evolution of organisms and their genomes, but also the resources for physiological studies of non-model species that do not always have as sufficient a tool-kit as would be desired. The genomic analysis of at least one marsupial species, *M. domestica*, has put this species on par with humans, mice and a very short list of other gnathostomate species in regards to being able to study their adaptive immune responses. In 1977 Ashman lamented the scarcity of research on marsupial immune systems and asked if this field would have a brighter future (Ashman 1977). The past decade has seen amazing growth and discoveries in the studies of marsupial immunity. Given the emerging problems such as the Tasmanian devil facial tumor disease, it is hopefully not too late (McCallum and Jones 2006).

References

Allison JP, Havran WL (1991) The immunobiology of T cells with invariant gamma delta antigen receptors. Ann. Rev Immunol 9: 679-705.

Ashman R, Keast D, Stanley NF, Waring H (1975) The immunological responses of marsupials. Am Nat 15:155-166

Ashman RB (1977) Marsupial immunology: A brighter future? Dev and Comp Immunol 1: 283-284.

Asquith KL, Kitchener AL, Kay DJ (2006) Immunisation of the male tammar wallaby (*Macropus eugenii*) with spermatozoa elicits epididymal antigen-specific antibody secretion and compromised fertilisation rate. J Reprod Immunol 69:127-147

Aveskogh M, Hellman L (1998) Evidence for an early appearance of modern post-switch isotypes in mammalian evolution; cloning of IgE, IgG and IgA from the marsupial *Monodelphis domestica*. Eur J Immunol 28: 2738-2750

Aveskogh M, Pilström L, Hellman L (1999) Cloning and structural analysis of IgM (μ chain) and the heavy chain V region repertoire in the marsupial *Monodelphis domestica.* Dev Comp Immunol 23: 597-606

Baker ML, Gemmell E, Gemmell RT (1998) Physiological changes in brushtail possums, *Trichosurus vulpecula*, transferred from the wild to captivity. J. Exper Zool 280: 203-212.

Baker ML, Gemmell RT (1999) Physiological changes in brushtail possums relocated from Armidale into captivity in Brisbane, Australia. J. Exper Zool 284: 42-49.

Baker, M.L., Gemmell, E. and Gemmell, R.T. (1999). Ontogeny of the immune system of the brushtail possum, *Trichosurus vulpecula*. The Anatomical Record. 256: 354-365.

Baker ML, Rosenberg GH, Zuccolotto P, Harrison G, Deane EM, Miller, RD (2001). Further characterization of T cell receptor chains of marsupials. Dev Comp Immunol 25: 495-507.

Baker ML, Osterman AK, Brumburgh S (2005a) Divergent T cell receptor delta chains from marsupials. Immunogenetics. 57: 665-673.

Baker ML, Belov K, Miller RD (2005b) Unusually similar patterns of antibody V segment diversity in distantly related marsupials. J Immunol 174:5665-5671

Baker ML, Indiviglio S, Nyberg AM, Rosenberg GH, Lindblad-Toh K, Miller RD, Papenfuss AT (2007) Analysis of a set of Australian northern brown bandicoot expressed

sequence tags with comparison to the genome sequence of the South American grey short tailed opossum. BMC Genomics. 8: 50.

Bell RG (1977) Marsupial immunglobulins: the distribution and evolution of macropod IgG2, IgM and light chain antigenic markers within the sub-class metatheria. Immunol 33:917-924

Bell RG, Lynch NR, Turner KJ (1974) Immunoglobulins of the marsupial *Setonix brachyurus* (quokka). Immunol 27:1103-1115

Belov K, Harrison GA, Miller RD, Cooper DW (1999a) Isolation and sequence of a cDNA coding for the heavy chain constant region of IgG from the Australian brushtail possum, *Trichosurus vulpecula*. Mol Immunol 36:535-541.

Belov K, Harrison GA, Miller RD, Cooper DW (1999b) Molecular cloning of the brushtail possum (*Trichosurus vulpecula*) immunglobulin E heavy chain constant region Mol Immunol 36:1255-1261

Belov K, Harrison GA, Rosenberg GH, Miller RD, Cooper DW (1999c) Isolation and comparison of the IgM heavy chain constant region from Australian (*Trichosurus vulpecula*) and American (*Monodelphis domestica*) marsupials. Dev Comp Immunol 23: 649-656

Belov K, Harrison GA, Miller RD, Cooper DW (2001) Characterisation of the kappa light chain of the brushtail possum (*Trichosurus vulpecula*). Vet Immunol Immunopathol 78:317-324

Belov K, Harrison GA, Miller RD, Cooper DW (2002a) Molecular cloning of four lambda light chain cDNAs from the Australian brushtail possum (*Trichosurus vulpecula*). Eur J Immunogenet 29:95-99.

Belov K, Nguyen M-AT, Zenger KR, Cooper DW (2002b) Ontogeny of immunoglobulin expression in the brushtail possum (*Trichosurus vulpecula*). Dev Comp Immunol 26: 599-602.

Benedict CL, Gillian, S. Thai T, Kearney JF (2000). Terminal deoxynucleotidyl transferase and repertoire development. Immunol Rev 175:150-157.

Butler JE (1997) Immunoglobulin gene organization and the mechanism of repertoire development. Scand J Immunol 45: 455-462.

Canfield P, Hemsley S, Connolly J (1996) Histological and immunohistological study of the developing and involuting superficial cervical thymus in the koala (*Phascolarctos cinereus*). J Anat 189:159-169.

Chen H, Kshirsagar S, Jensen I, Lau K, Covarrubias R, Schluter S, Marchalonis J (2009) Characterization of arrangement and expression of the T cell receptor γ locus in the sandbar shark. Proc Natl Acad Sci USA 106: 8591-8596.

Chien YH, Iwashima M, Kaplan KB, Elliot JF, Davis MM: (1987) A new T-cell receptor gene located within the alpha locus and expressed early in T-cell differentiation. Nature 327: 677-682.

Cisternas PA, Armati PJ (2000) Immune system cell markers in the northern brown bandicoot, *Isoodon macrourus*. Dev Comp Immunol 24(8): 771-82.

Coutinho HB, King G, Sewell HF, Tighe P, Coutinho VB, Robalinho TI, Carvalho AB (1993) Immunocytochemical study of Peyer's patches follicular-associated epithelium in the marsupial, Didelphis albiventris. Dev Comp Immunol 17:537–548.

Coutinho HB, Sewell HF, Tighe P, King G, Nogueira JC, Robalinho TI, Coutinho VB, Cavalcanti VMS (1995) Immunocytochemical study of the ontogeny of the marsupial *Didelphis albiventris* immune system. J Anat 187: 37–46.

Criscitiello MF, Saltis M, Flajnik MF: (2006) An evolutionarily mobile antigen receptor variable region gene: Doubly rearranging NAR-TcR genes in sharks. Proc Natl Acad Sci USA 2006, 103: 5036-5041.

Croix DA, Samples NK, VandeBerg JL, Stone WH (1989) Immune response of a marsupial (*Monodelphis domestica*) to sheep red blood cells. Dev Comp Immunol 13: 73-78.

Daly KA, Digby M, Lefèvre C, Mailer S, Thomson P, Nicholas K, Williamson P (2007) Analysis of the expression of immunoglobulins throughout lactation suggests two periods of immune transfer in the tammar wallaby (*Macropus eugenii*). Vet Immunol Immunopathol 120: 187-200.

Das S, Nozawa M, Klein J, Nei M (2008) Evolutionary dynamics of the immunoglobulin heavy chain variable region genes in vertebrates. Immunogenetics 60: 47-55

Deakin JE, Belov K, Curach NC, Green P, Cooper DW (2005) High levels of variability in immune response using antigens from two reproductive proteins in brushtail possums. Wild Res 32: 1-6.

Deakin JE, Olp JJ, Graves JAM, Miller RD (2006a) Physical mapping of immunoglobulin loci *IGH@*, *IGK@*, and *IGL@* in the opossum (*Monodelphis domestica*). Cytogenet Genome Res 114: 94

Deakin JE, Parra ZE, Graves JAM, Miller RD: (2006b) Physical Mapping of T cell receptor loci (TRA@, TRB@, TRD@ and TRG@) in the opossum (*Monodelphis domestica*). Cytogenet Genome Res 112: 342K.

Deane EM, Cooper DW (1988) Immunological development in pouch young marsupials. In: Tyndale-Biscoe CH, Janssens PA, eds. The Developing Marsupial. Springer-Verlag, Berlin, pp 190-199

Feeney AJ (1992) Predominance of VH-D-JH junctions occurring at sites of short sequence homology results in limited junctional diversity in neonatal antibodies. J Immunol 149: 222-229.

Fox DH, Rowlands Jr DT, Wilson DB (1976) Proliferative reactivity of opossum peripheral blood leukocytes allogeneic cells, mitogens, and specific antigens. Transplantation 21:164-168.

Flajnik MF (2002) Comparative analyses of immunoglobulin genes: surprises and portents. Nat Rev Immunol 2: 688-698.

Gouin N, Wright AM, Miska KB, Parra ZE, Samollow PB, Baker ML, Miller RD (2006). Modo-UG, a marsupial non-classical MHC class I locus. Immunogenetics. 58: 396-406.

Harrison GA, Taylor CL, Miller RD, Deane EM (2003) Primary Structure and Variation of the T-Cell Receptor δ Chain from a Marsupial, *Macropus eugenii.* Immunology Letters 88:117-125

Haynes JI (2001) The marsupial and monotreme thymus, revisited. J Zool, Lond 253: 167-173.

Hemsley SW, Canfield PJ, Husband AJ (1995) Immunohistological staining of lymphoid tissue in four Australian marsupial species using species cross-reactive antibodies. Immunol Cell Biol 73:321–325.

Hubbard GB, DG Saphire, SM Hackleman, MV Silva, JL VandBerg, Stone WH (1991) Ontogeny of the thymus gland of a marsupial (*Monodelphis domestica*). Lab Anim Sci 41:227

Infante AJ, Samples NK, Croix DA, Redding TS, VandeBerg JL, Stone WH (1991) Cellular immune response of a marsupial, *Monodelphis domestica*. Dev Comp Immunol 15:189-99.

Jones M, Cordell JL, Beyers A, Tse AGD, Mason DY (1993) Detection of T and B cells in many animal species using cross-reactive antipeptide antibodies. J Immunol 150: 5429–5435.

Kalmutz SE (1962) Antibody production in the opossum embryo. Nature 193: 851-853

Kathiresan S (1969) Morphological studies on the thymus. Part I. Thymus of the marsupial *Trichosurus vulpecula*. Indian J Med Res 57:939–945.

Kitchener AL, Edds LM, Molinia FC, Kay DJ (2002) Porcine zonae pellucidae immunization of tammar wallabies (*Macropus eugenii*): fertility and immune responses. Reprod Fertil Dev 14:215-223.

Kokubu F, Litman R, Shamblott MJ, Hinds K, Litman GW (1988) Diverse organization of immunoglobulin VH gene loci in a primitive vertebrate. EMBO J 7:3413-3422.

Kubota T, Wang JY, Gobel TWF, Hockett RD, Cooper MD, Chen CH (1999) Characterization of an avian (*Gallus gallus domesticus*) TCR α/δ locus. J Immunol 163: 3858-3866.

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 5:150-163

La Via MF, Rowlands DT, Block M (1963) Antibody formation in embryos. Science 140: 1219-1220

Lee SS, Fitch D, Flajnik MF, Hsu E (2000) Rearrangement of immunoglobulin genes in shark germ cells. J Exp Med 191: 1637-1648

Lucero JE, Rosenberg GH, Miller RD (1998) Marsupial light chains: Complexity and conservation of Igl in the opossum *Monodelphis domestica*. J Immunol 161: 6724-6732

Marolleau JP, Fondell JD, Malissen M, Trucy J, Barbier E, Marcu K, Cazenave PA, Primi D (1988) The joining of germ-line Vα to Jα genes replaces the preexisting Vα-Jα complex in a T cell receptor α positive T cell line. Cell 55:291.

McCallum H, Jones M (2006) To lose both would look like carelessness: Tasmanian devil facial tumour disease. PLoS Biology 4: e342.

McLelland DJ, Kirkland PD, Rose KA, Dixon RJ, Smith N (2005) Serologic responses of Barbary sheep (*Ammotragus lervia*), Indian antelope (*Antilope cervicapra*), wallaroos (*Macropus robustus*), and chimpanzees (*Pan troglodytes*) to an inactivated encephalomyocarditis virus vaccine. J Zoo Wildl Med 36:69-73

Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A, Jurka J, Kamal M, Mauceli E, Searle SMJ, Sharpe T, Baker ML, Batzer MA, Benos PV, Belov K, Clamp M, Cook A, Cuff J, Das R, Davidow L, Deakin JE, Fazzari MJ, Glass JL, Grabherr M, Greally JM, Gu W, Hore TA, Huttley GA, Jirtle RL, Koina E, Lee JT, Mahony S, Marra MA, Miller RD, Nicholls RD, Oda M, Papenfuss AT, Parra ZE, Pollock DD, Ray DA, Schein JE, Speed TP, Thompson K, VandeBerg JL, Wade CM, Walker JA, Waters PD, Webber C, Weidman JR, Xie X, Zody MC, Broad Institute Genome Sequencing Platform, Broad Institute Whole Genome Assembly Team, Marshall Graves JA, Ponting CP, Breen M, Samollow PB, Lander ES,

Lindblad-Toh KS (2007) Genome of the marsupial *Monodelphis domestica* reveals lineage-specific innovation in coding sequences. Nature 447: 167-178

Miller RD, Grabe H, Rosenberg GH (1998) The V_H Repertoire of a Marsupial: *Monodelphis domestica*. J Immunol 160: 259-265

Miller RD, Bergemann ER, Rosenberg GH (1999) Marsupial light chains: Igk with four V families in the opossum *Monodelphis domestica***.** Immunogenetics 50: 329-335

Ohta Y, Flajnik M (2006) IgD, like IgM, is a primordial immunoglobulin class perpetuated in most jawed vertebrates. Proc Natl Acad Sci USA 103: 10723-10728

Old JM, Deane EM (2003) The detection of mature T- and B- cells during development of the lymphoid tissues of the tammar wallaby (*Macropus eugenii*). J Anat 203: 123-131.

Parra ZE, Baker ML, Schwarz R, Deakin JE, Lindblad-Toh K, Miller RD (2007). Discovery of a new T cell receptor in marsupials. Proc Natl Acad Sci USA 104:9776-9781.

Parra ZE, Baker ML, Hathaway J, Lopez AM, Trujillo J, Sharp A, Miller RD (2008) Comparative genomic analysis and evolution of the T cell receptor loci in the opossum *Monodelphis domestica.* BMC Genomics 9:111.

Parra ZE, Baker ML, Lopez AM, Trujillo J, Volpe JM, Miller RD (2009) TCRm recombination and transcription relative to the conventional TCR during postnatal development in opossums. J Immunol 182:154-163.

Partula S, de Guerra A, Fellah JS, Charlemagne J (1996) Structure and diversity of the TCR alpha-chain in a teleost fish. J Immunol 157: 207-212

Reynaud CA, Dahan A, Anquez V, Weill JC (1989) Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. Cell 59: 171-183

Rowlands, DT, Lavia, MF, Block, MH (1964) The blood forming tissues and blood of the newborn opossum (*Didelphys virginiana*). II. Ontogenesis of antibody formation to flagella of *Salmonella typhi*. J Immunol 93: 157-164.

Rumfelt LL, Avila D, Diaz M, Bartl S, McKinney EC, Flajnik MF (2001) A shark antibody heavy chain encoded by a nonsomatically rearranged VDJ is preferentially expressed in early development and is convergent with mammalian IgG. Proc Natl Acad Sci USA 98: 1775–1780

Samples NK, VandeBerg JL, Stone WH (1986) Passively acquired immunity in the newborn of a marsupial (*Monodelphis domestica*). Am J Reprod Immunol Microb 11:94-97

Sekiguchi J, Alt FW, and Oettinger M. (2004) The mechanisms of V(D)J recombination. In: Honjo T, Alt FW, Neuberger M, eds. Molecular Biology of B Cells. Academic Press, London, pp 62-82.

Shearer MH, Robinson ES, VandeBerg JL, Kennedy RC (1995) Humoral immune response in a marsupial Monodelphis domestica: anti-isotypic and anti-idiotypic responses detected by species-specific monoclonal anti-immunoglobulin reagents. Dev Comp Immunol. 19(3):237-46.

Silverman GJ, Goodyear CS (2002) A model B-cell superantigen and the immunobiology of B lymphocytes. Clin Immunol 102:117-134.

Stanley NF, Yadav M, Waring H, Eadie M. (1972) The effect of thymectomy on response to various antigens of a marsupial *Setonix brachyrus* (Quokka). Aust J Exp Biol Med Sci 50:689-702.

Stone WH, Manis GH, Hoffman ES, Saphire DG, Hubbard GB, VandeBerg JL. (1997) Fate of allogeneic skin transplantations in a marsupial (*Monodelphis domestica*). Lab Anim Sci 47:283-7.

Stone WH, Bruun DA, Manis GS (1996) The immunobiology of the marsupial *Monodelphis domestica*. In: Stolen JS, Fletcher T.C. Bayne CJ et al. eds. Modulators of immune responses, the evolutionary trail, Breckenridge series 2 vol. 11 SOS Publications, Fairhaven, NJ, pp. 149–165.

Stone WH, Bruun DA, Foster EB, Manis GS, Hoffman ES, Saphire DG, VandeBerg JL, Infante AJ. (1998) Absence of a significant mixed lymphocyte reaction in a marsupial (*Monodelphis domestica).* Lab Anim Sci 48:184.

Taylor DL, Burrell R (1968). The immunological responses of the North American opossum (*Didelphis virginiana*). J Immunol 101: 1207-1216.
Tonegawa S (1983) Somatic generation of antibody diversity. Nature 302:575–581.

Turner KJ, Alpers MP, Wight AM (1972). Delayed hypersensitivity in the marsupial *Setonix brachyurus* (Quokka). J Immunol 108: 1675-1680.

Tutter A, Riblet R (1989) Conservation of an immunoglobulin variable-region gene family indicates a specific, noncoding function. Proc Natl Acad Sci USA 86: 7460-7464

Ventura-Holman T, Lobb CJ (2002) Structural organization of the immunoglobulin heavy chain locus in the channel catfish: the IgH locus represents a composite of two gene clusters. Mol Immunol 38: 557-564.

Wang X, Olp, JJ, Miller RD (2009). On the genomics of immunoglobulins in the gray, short-tailed opossum *Monodelphis domestica.* In review

Wilkinson R, Kotlarski I, Barton M (1992a) Koala lymphoid cells: analysis of antigen-specific responses. Vet Immunol Immunopath 33:237-47.

Wilkinson R, Kotlarski I, Barton M, Phillips P (1992b) Isolation of koala lymphoid cells and their in vitro responses to mitogens. Vet Immunol Immunopathol 31:21-33.

Wilkinson R, Barton M, Kotlarski I (1995) Identification of koala lymphocytes using an anti-human CD3 antibody. Dev Comp Immunol 19:537–545.

Woods GM, Kreiss A, Belov K, Siddle HV, Obendorf DL, Muller HK (2007) The Immune Response of the Tasmanian Devil (*Sarcophilus harrisii*) and Devil Facial Tumour Disease. EcoHealth. 4: 338-345.

Yadav M (1971) The transmissions of antibodies across the gut of pouch-young marsupials. Immunology 21:839-851.
Yadav M (1973) The presence of the cervical and thoracic thymus lobes in marsupials. Aust J Zool 21:285-301.

Zachau HG (2004) Immunoglobulin kappa genes of human and mouse. In: Honjo T, Alt FW, Neuberger M ed. Molecular Biology of B Cells. Academic Press, London, pp 27-36

Zhang Y, Cado D, Asarnow DM, Komori T, Alt FW, Raulet DH, Allison JP (1995) The role of short homology repeats and TdT in generation of the invariant γδ antigen receptor repertoire in the fetal thymus. Immunity 3:439.

Zuccolotto PD, Harrison GA, Deane EM (2000) Cloning of marsupial T cell receptor α and β constant region cDNAs. Immunol Cell Biol 78:103-109.

CHAPTER 2

ON THE GENOMICS OF IMMUNOGLOBULINS IN THE GRAY, SHORT-TAILED OPOSSUM *MONODELPHIS DOMESTICA*

By Xinxin Wang, Jonathan J. Olp, and Robert D. Miller*.

*Center for Evolutionary and Theoretical Immunology and Department of Biology, University of New Mexico, Albuquerque NM 87131 USA

XW and RDM conceived of the study, participated in its design, and drafted the manuscript. XW and JJO generated the data. XW performed the sequencing reaction and data analyses.

Manuscript published in Immunogenetics (2009) 61: 581-596.

Abstract

Annotated maps of the *IGH, IGK,* and *IGL* loci in the gray, short-tailed opossum *Monodelphis domestica* were generated from analyses of the available whole genome sequence for this species. Analyses of their content and organization confirmed a number of previous conclusions based on characterization of cDNAs encoding opossum immunoglobulin heavy and light chains and limited genomic analysis, including: i) the predominance of a single IGHV subgroup and clan; ii) the presence of a single IgG subclass; iii) the apparent absence of an IgD; and iv) the general organization and V gene complexity of the *IGK* and *IGL* light chain loci. In addition several unexpected discoveries were made including the presence of a partial germ-line joined IGHV segment, the first germline joined Ig V gene to be found in a mammal. In addition was the presence of a larger number of IGKV subgroups than had been previously identified. With this report, annotated maps of the Major Histocompatibility Complex, T cell receptor, and immunoglobulin loci have been completed for *M. domestica*, the only non-eutherian mammalian species for which this has been accomplished, strengthening the utility of this species as a model organism.

Introduction

The gray, short-tailed opossum *Monodelphis domestica* is among the better-developed metatherian (marsupial) model species for biomedical research. With the recent completion of its whole genome sequence, it is arguably the premier model marsupial (Samollow 2006). As with all marsupials, the newborn opossum is highly altricial making it ideal for the study of early development in the immune system and the evolution of maternal transfer of immunity (Deane and Cooper 1988). In addition, *M. domestica* has its uses as a model organism for a variety of human diseases including skin cancer, hypercholesterolemia, and neurological development and regeneration to name a few (VandeBerg and Robinson 1997; Ley et al. 2000; Fry and Saunders 2000). The utility of *M. domestica* as a model organism for cancer, infectious diseases, and early development can only be further enhanced by continued characterization of the genes encoding the components of the immune system. Many of the components of both the innate and adaptive immune system have been identified in the opossum genome (Wong et al. 2006). In addition, detailed genomic analyses of the Major Histocompatibility Complex and T cell receptor (TCR) loci have already been published, including a newly discovered TCR locus, TCRμ, which is not found in eutherian ("placental") mammals (Belov et al. 2006; Parra et al. 2007, 2008). Here we complete the analysis of genes encoding antigen receptors of the adaptive immune system in *M. domestica* by presenting a detailed, annotated description of the immunoglobulin (Ig) heavy and light chain loci.

Previously, we had physically mapped the loci encoding the opossum Ig chains, the heavy chain (*IGH*), kappa (*IGK*) and lambda (*IGL*) light chains to chromosomes 1 (*IGH* and *IGK*) and 3 (*IGL)* (Deakin et al. 2006). The content and diversity of expressed opossum Ig heavy and light chains have also been inferred from analysis of transcribed Ig mRNA (or cDNA) (Aveskogh et al. 1998, 1999; Lucero et al. 1998; Miller et al. 1998, 1999). A number of observations emerged from these studies, including an apparent greater diversity of variable (V) gene segment subgroups in the light chains than in heavy, a pattern that appears to be common to other marsupial species as well (Baker et al 2005). Furthermore, the majority of expressed IGHV gene segments appeared to belong to a single V subgroup (Baker et al 2005; Miller et al. 1998; Aveskogh et al. 1999). A second IGHV subgroup, IGHV2 was also known but appeared to contain only a single gene segment. Both subgroups belong to clan III of mammalian IGHV, as do all marsupial IGHV isolated so far (Aveskogh et al 1999; Miller et al. 1998; Baker et al. 2005).

The IGH constant (C) regions identified by analyses of opossum cDNAs included what appeared to be a single IgM, IgG, IgE and IgA (Aveskogh et al. 1998, 1999; Belov et al. 1999; Miller et al. 1998). This is in contrast to most eutherian ("placental") and prototherian (monotreme) mammals studied which have multiple IgG, IgA, and/or IgE subclasses encoded by separate sets of exons (Belov and Hellman 2003). The presence of only a single IgG isotype based on cDNA sequence contradicted previous serum Ig analyses that supported the presence of at least two IgG subclasses in *M. domestica* as well as other marsupial species (Bell 1977; Bell et al. 1974; Shearer et al. 1995), and

remained an unresolved question. Furthermore, no cDNAs encoding a heavy chain with homology to IgD had been reported for any marsupial species (Miller and Belov 2000 and unpublished observations).

The recent completion of the *M. domestica* whole genome sequence has facilitated finer scale analyses of the organization and content of the Ig loci (Mikkelsen et al. 2007). In addition to providing detailed genomic maps of the three Ig loci, the results of these analyses presented here both confirm previous predictions made based on the cDNA analyses and a limited amount of genomic DNA sequence available, and also reveal some surprises not uncovered in the transcriptome.

Materials and Methods

Whole genome sequence analysis

The analyses presented here were made using MonDom5, the current complete *M. domestica* genome assembly, available at GenBank under the accession number AAFR03000000 (Mikkelsen et al. 2007).

IGH, IGK, and IGL cDNA sequences from *M. domestica* and *Trichosurus vulpecula* were used in a homology search against the M. domestica genome project with the aid of the BLAST algorithm (Baker et al. 2005; Belov et al. 1999; Aveskogh et al. 1999; Miller et al. 1998, 1999; Lucero et al, 1998). Scaffolds identified from the *M. domestica* genome project as containing Ig sequences were compared with these cDNA to identify genomic V, D, J, and C gene segments. The beginning and end of each coding exon were identified by the presence of mRNA splice sites or flanking recombination signal sequence (RSS) sites.

To scan MonDom5 specifically for sequences corresponding to exons encoding the constant domains of IgD, sequences from both the extracellular and transmembrane domains from human (GenBank accession number AAH21276), mouse (AAB59654), horse (AAU09793), and catfish (AAC60133) IgD were used to perform both nucleotide (BLASTN) and translated (TBLASTN) alignments of both the entire opossum genome and an isolated region only containing the opossum *IGH* locus (Altschul et al. 1990). Using the same method exons homologous to IgD exons were identified in the recently

completed platypus genome assembly Ornithorhynchus_anatinus-5.0 available at Ensembl (www.ensembl.org). This is a species for which no cDNA sequence for IgD was previously available, much like the opossum.

Sequences that correspond to switch (S) regions were identified upstream of both the functional and pseudogene copies of the IgM C regions. They were identifiable as containing repeat sequence composed of pentameric repeat sequences of GAGCT and GGGCT conserved in other mammalian species (Nikaido et al. 1982, Mills et al. 1990)

RNA extraction

Opossum tissues were collected and stored in RNAlater (Ambion, Austin, TX) at 4°C for 24 hours and stored long term at -80°C. RNA extraction was performed using the Trizol RNA extraction protocol (Invitrogen, Carlsbad, CA). All procedure involving the use of live animals were approved under institutional protocol 07UNM005.

Reverse transcription, PCR and Sequencing

Reverse transcription-polymerase chain reaction (RT-PCR) was performed using the GeneAmp RNA PCR Core Kit (Applied Biosystems, Foster City, CA). PCR amplification was performed using Advantage TM-HF 2 PCR (BD Biosciences, CLONTECH Laboratories, Palo Alto, California) with the conditions for all primer combinations: a long denaturation at 94 °C for 1 minute for one cycle, followed by 34 cycles of denaturation at 94 °C for 30 seconds, annealing at 62 °C for 4 minutes, and a final single extension period of 68 °C for 5 minutes.

All oligonucleotide sequences used for PCR primers are presented in Table 1. 3' cDNA ends were generated by the rapid amplification of cDNA ends (RACE) approach using the Gene Racer Kit (Invitrogen, Carlsbad, CA) following manufacturer's recommended protocol. Primers complementary to the 3' most CH exon based on cDNA sequence of each of the IgH isotypes, were used to amplify transcripts containing the complete TM form of each heavy chain, including the 3' untranslated region (UTR). Additional primers based on the TM1 exon were used in nested PCR to confirm the 3' ends for each isotype. These sequences have been deposited in GenBank, accession numbers pending.

To confirm that the IGHV3.1 gene segment, which contains a germ-line joined D and RSS, was not an assembly artifact, an oligonucleotide based on sequence 5' of the L exon of this gene segment in the assembly was paired with another based on sequence 3' of the RSS were used to amplify the entire gene segment by PCR (Table 1). These primers amplified a 855 bp fragment from *M. domestica* genomic DNA, which was cloned and sequenced. The sequence has been deposited in GenBank under accession number EU592040.

In the MonDom5 assembly, the exon encoding the CH1 domain of the pseudogene copy of the IgM C regions contained a gap at the start of the exon. This gap was filled by using PCR to amplify this region from *M. domestica* genomic DNA using primers based on sequences that flank the gap. The sequence has been deposited in GenBank, accession number pending.

Table 1. Sequences and description of oligonucleotide primers used.

Name	Sequence (5'-3')	Description
IgM-TMF1	CTTCTCCCCACCAGATGTGT	Forward primer on IGM CH4 to acquire TM and 3'UTR
IgM-TMF2	GGGGAAGTGAATGCTGAAGA	Forward primer on IGM TM1 to confirm 3' end of TM1, TM2 and 3'UTR
IgM-TMR	TGAGGCTGTTGTCCACAGAT	Reverse primer on IGM TM1 to confirm 5' end of TM1
IgA-TMF1	ACTTCACCCAGCAAACCATC	Forward primer on IGA CH4 to acquire TM and 3'UTR
IgA-TMF2	CCCATGACAGTGACCTTCCT	Forward primer on IGA TM to confirm 3' end of TM and 3'UTR
IgA-TMR	GAACATGCTGAGCAGGAACA	Reverse primer on IGA TM to confirm 5' end of TM
IgG-TMF1	CCAAATTTCCCAGAGGACAA	Forward primer on IGG CH3 to acquire TM and 3'UTR
IgG-TMF2	CTGAGACCAGGGATGGAGAG	Forward primer on IGG TM1 to confirm 3' end of TM1, TM2 and 3'UTR
IgG-TMR	AGGGATCATTGGCTGCTTTA	Reverse primer on IGG TM1 to confirm 5' end of TM1
IgE-TMF1	AGAGGACCACACTGGACACC	Forward primer on IGE CH4 to acquire TM and 3'UTR
IgE-TMF2	ACACGTGCACAGAAATGGAG	Forward primer on IGE TM1 to confirm 3' end of TM1, TM2 and 3'UTR
IgE-TMR	GCCTGCTGGATCATGTTCTT	Reverse primer on IGE TM1 to confirm 5' end of TM1
pCmuF	ATGCTGTGGTGCCTATCTCC	Forward primer to fill the gap from pseudo-M CH1 exon
pCmuR	GGTGTCTTGTTCCCCTCAAA	Reverse primer to fill the gap from pseudo-M CH1 exon
VH3F	GGAGACAGGTGGTTTGCATT	Forward primer 5' of IGHV3.1 genomic sequence
VH3R	CTGGTCCACTTGTTTTTCTGG	Reverse primer 3' of IGHV3.1 RSS genomic sequence

PCR products were cloned using TOPO TA cloning Kit (Invitrogen, Carsbad, CA) and sequenced using BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). All sequences reported are based on sequencing both strands of each clone. Sequences were analyzed using Sequencher 3.0 (Gene Codes, Ann Arbor, MI) and compared with the GenBank database and the MonDom5 assembly using the BLAST algorithm (Altschul et al. 1990).

Phylogenetic analysis

All phylogenetic tree reconstruction was based on analyses done using nucleotide alignments. Gaps in the nucleotide sequences were determined by first aligning the amino acid translations to establish gap position and then converting the sequence back to nucleotide using the BioEdit program (Hall 1999). In this way, nucleotide gaps were established based on codon position. Based on the nucleotide alignments, phylogenetic trees were constructed by the neighbor joining (NJ) method of Saitou and Nei (1987) using the MEGA software package (Kumar et al. 2004).

IGHV sequences from other species used in phylogenetic analyses presented were: Virginia opossum, *Didelphis virginiana* (Divi) IGHV is unpublished and was provided by Dr. R. Riblet. Possum, *T. vulpecula* (Trvu), IGHV were AAL87470, AAL87474; bandicoot, *I. macrourus* (Isma) AY586158; Mouse, *M. musulus* (Mumu), IGHV clan representatives were as follows: 3360, K01569; 3609N, X55935; DNA4, M20829; J558, Z37145; J606, X03398; Q52, M27021; S107, J00538; SM7, M31285; VH11, Y00743. Human, *H. sapiens* (Hosa), IGHV sequences were obtained from the VBASE database.

Pig, *S. scrofa* (Susu) VH was U15194. Cow, *Bos taurus* (Bota) IGHV was AF015505. Sheep, *O. aries* (Ovar) IGHV was Z49180. Horned shark, *H. fransciscii* (Hefr) IGHV was X13449.

IGKV and IGLV used in the analyses were: Human, *H. sapiens* (Hosa) IGKV and IGLV sequences were obtained from the VBASE database. Mouse, *M. musulus* (Mumu), IGKV family representatives were as follows: IGKVR1, X13938; IGKCLM, Z72384; IGKCAM2, M24937; IGVKID, M63611; mouse IGLV were as follows: VL1, X82687; VLX, D38129. Possum, *T. vulpecula* (Trvu), IGKV was AAL87498; IGLV were VL126, AAM09977; VL12, AAM09961; Rat, *R. norvegicus* (Rano) IGKV was U39609. Hamster, *C. migratorius* (Crmi) IGKV was U17165. Horse, *E. caballus* (Eqca), IGKV was X75611. Sheep, *O. aries* (Ovar) IGKV was X54110. Rabbit, *O. cuniculus* (Crcu) were VL2, M27840, VL3, M27841. Chicken, *G. gallus* (Gaga) IGLV was M96972. Horned shark, *H. fransciscii* (Hefr) IGLV, X15316 was used as an outgroup.

Dot matrix analysis of duplications

Comparisons of the *IGH* genomic sequence were done using Jdotter (http://pgrc.ipk-gatersleben.de/jdotter/).

Analysis of V, D, and J gene segment functionality

The identified germ-line gene segments from the opossum *IGH, IGK,* and *IGL* loci have been uploaded to the Somatic Diversification Analysis (SoDA) site, which is a web based software tool for analyzing germ-line and somatic contributions to expressed V(D)J

diversity (http://dulci.org/soda/) (Volpe et al. 2006). Both heavy and light chain V(D)J recombinations isolated either by RT-PCR or from a cDNA library were analyzed using SoDA to determine which gene segments were used and the contribution of P and N nucleotides to the mature V(D)J sequence.

Annotation of the MonDom5 assembly of the **M. domestica** *genome*

Supplementary Tables 1, 2 and 3 include the location of each identified coding segment in the *IGH, IGK,* and *IGL* loci, respectively. Provided is the beginning and end positions of each exon and the where in exon 2 of the V genes sequence predicted to encode the leader peptide ends and the extra-cellular V domain starts. Also indicated are those gene segments that are pseudogenes, the reason why they were labeled pseudogenes, and their transcriptional orientation relative to the constant region genes.

Nomenclature used was according to the method used the IMGT database (http://imgt.cines.fr/). Opossum V gene segments were numbered according to their order on the genome, from the 5' to 3' end of the locus, whereas D and J gene segments were numbered from 3' to 5'. V genes were designated with the subgroup number followed by a period and the individual number. V genes that belonged to subgroups that had previously been identified retained the original subgroup numbering followed by their order along the chromosome (Miller et al. 1998, 1999; Lucero et al. 1998).

Results

The opossum IGH locus

Previously the opossum *IGH* locus was localized to the centromeric end of the long arm of chromosome 1 (Deakin et al. 2006). Analysis of the genomic sequence from this region revealed that the *IGH* locus is oriented with its 3' end containing the *IGHC* genes being centromeric and its 5' end containing *IGHV* genes being telomeric (Fig. 1). The opossum *IGH* locus spans 1,418 kb from the exon encoding the leader (L) of the most 5' *IGHV* (VH2.1) to the 3' end of the terminal exon of the IgA C region (Fig. 1). This total length is an estimate at this point since the assembly contains gaps (Fig. 1), in the region containing the *IGHV* segments. An overall view of the *IGH* locus reveals that the gene segments are distributed fairly evenly and are in the same reading orientation (Fig. 1). The location of each of the identified coding regions in this assembly is provided in Supplementary Table 1.

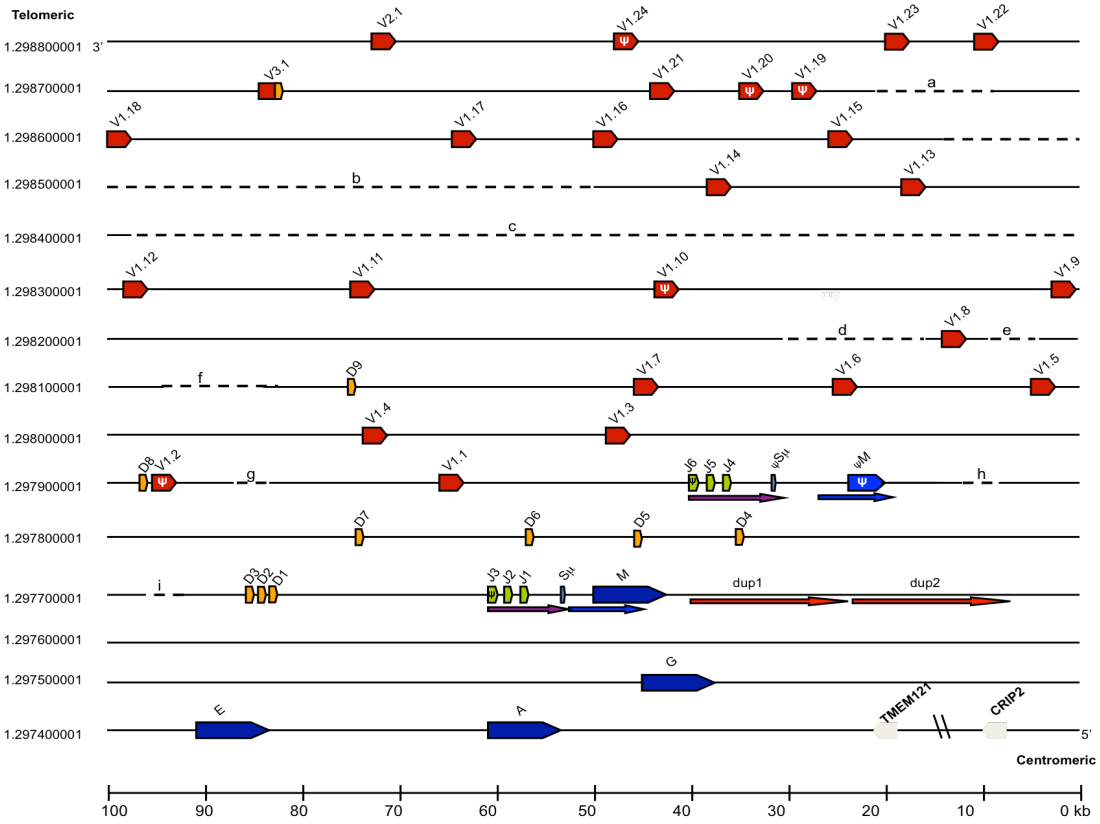


Figure 1. Diagram of the opossum *IGH* locus with the V, D, J, and C gene exons shown in red, yellow, green and blue, respectively. V segments are designated with the subgroup number followed by a period and a designated number. Direction of transcription is indicated on each segment by the arrow built into the symbol. Presumptive pseudogenes are indicated ψ. The arrows beneath the line indicate the location of gene duplication. The dashed lines designated from a through g indicated the gaps in the MonDom5 genome sequence, which are 12kb, 51kb, 98kb, 16kb, 3.5kb 10kb and 400bp, respectively. The scale across the bottom in is kilobases. The position in the assembly listed as a column (1.XXXXXX) on the left designates the location as being chromosome 1 and the nucleotide position along the chromosome. Syntenic genes shown are transmembrane protein 121 (*TMEM121*) and cysteine-rich protein 2 (*CRP2*).

The IGH *constant regions:* Homologues corresponding to C regions of the single, functional IgM, IgG, IgE, and IgA were identified using previously isolated cDNA sequence for each of these heavy chain isotypes (Miller et al. 1998; Belov et al. 1999; Aveskogh et al. 1998, 1999). For each of the IgH chain isotypes present in the opossum there is a single functional set of constant region exons (Fig. 1 and 2; Supplementary Table 1). All cDNAs reported so far for each of the Ig heavy chain isotypes encoded secretory forms of antibody, therefore sequences corresponding to the TM regions were not available. To identify the 3' ends of the membrane forms of each of the heavy chain isotypes, 3' RACE was performed on adult splenic RNA using primers specific for the exon encoding the CH4 for M and E, and CH3 for G and A. This strategy successful lead to the identification of the exons encoding the TM and 3' UTR sequences. For M, G and E there are two TM exons, TM1 and TM2, for IgA there is only a single TM exon (Fig. 2). Therefore all coding exons reported here have been confirmed as being present in mRNA transcripts (or cDNA).

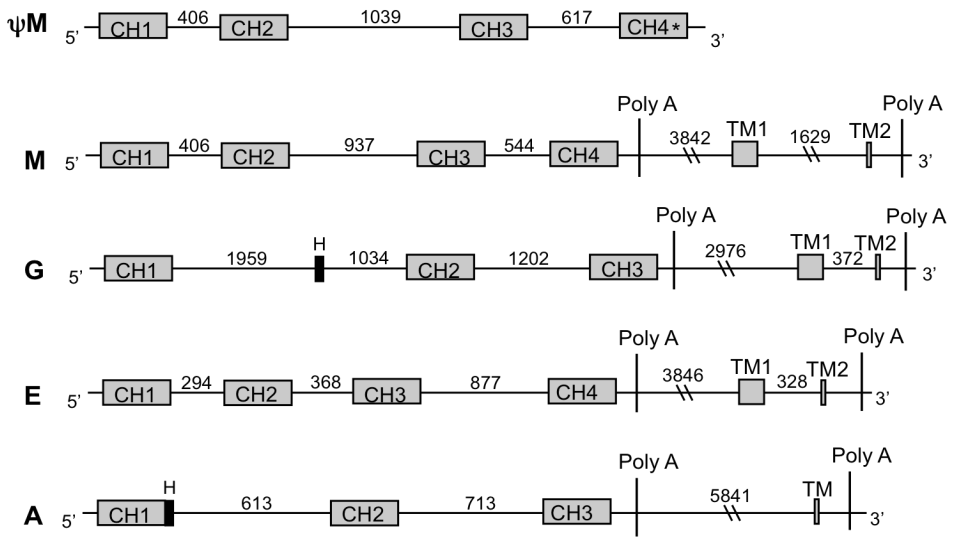


Figure 2. Exon and intron organization of the opossum IgH constant region genes. Exons are showed with red rectangles and 3'UTR are showed with yellow rectangles. Numbers on the lines indicate the length of the introns in base pairs. H designates sequences encoding identified hinge regions. The presumptive M pseudogene is indicated with ψ. The asterisk indicates the exon with an in-frame stop codon. Polyadenylation sites are identified as having a conserved AATAAA signal sequence.

As predicted from analysis of opossum H chain cDNAs, the extra-cellular domain structure for the four isotypes is fairly typical of mammalian Ig (Fig. 2). IgM and IgE each have four CH domain encoding exons and no evidence of a hinge region. IgG and IgA each have three CH exons. IgG also has a single hinge encoding exon similar to human IgG1, 2 and 4. The sequence encoding the IgA hinge region is an extension of the CH1 exon rather than the CH2 as in humans, other primates and rodents (Kawamura et al. 1992; Osborne et al. 1988). In this regard the opossum IgA hinge region genetics is more like that of the platypus where the hinge sequence is part of the exon encoding the CH1 domain (Belov and Hellman 2003; and unpublished analysis of the platypus genome)

In addition to the single functional IgM in the opossum there is also a second partial set of exons with near identity to the CH1 through 4 domains of IgM, located 169 kb upstream of the functional exons (Fig. 1 and 2). This appears to be due to an apparent duplication of a region containing three *IGHJ* gene segments, the switch region for IgM (Sμ) and IgM exons CH1 through CH4, resulting in a partial duplication of IgM (Fig. 2 and 3a). The absence of exons encoding the TM regions, along with the presence of an in-frame stop codon in the CH4 exon and our inability to detect any transcripts using these exons (not shown), resulted in this second upstream IgM being designated a pseudogene. The distance between the duplicated IGHJ segments and the partial partial IgM pseudogene is greater than it is for the functional copy (Fig. 3a). This length difference is due to the insertion of a LINE element just 3' of the region corresponding to the Sμ of the pseudogene (Fig. 3a). Although the S regions upstream of both the

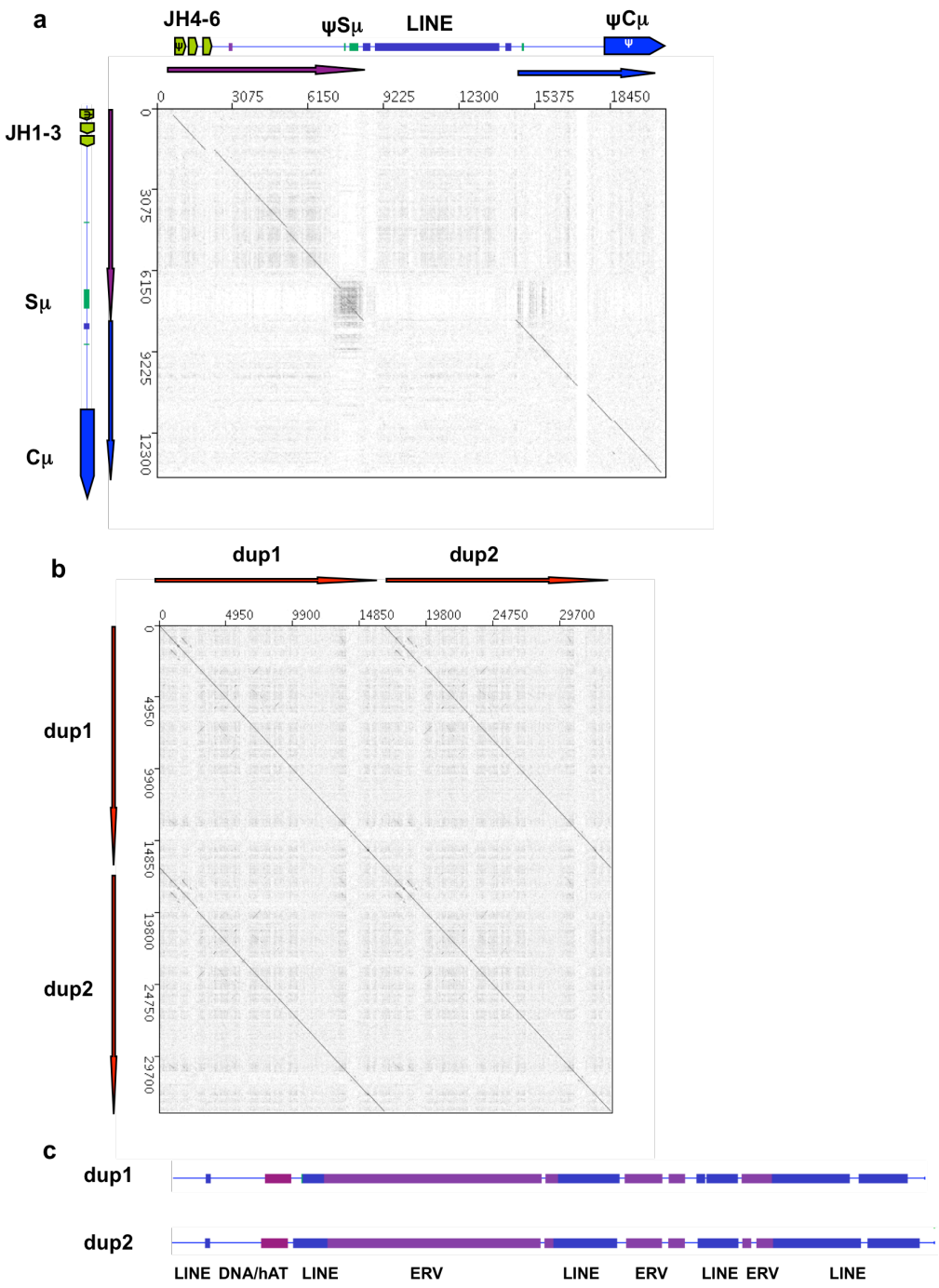


Figure 3. Dot plot analysis of duplicated regions within the *IGH* locus. **a** the region from IGHJ6 to the ψM constant region exons compared with the region spanning IGHJ3 to the functional M constant region exons. The region corresponding to the LINE element insertion is indicated. **b** the duplicated regions 3' of the functional M constant regions shown in Fig. 1 compared with itself. **c** diagram of the two duplicated regions located 3' of the functional M constant regions shown in Fig. 1 indicating the presence of LINE and ERV related sequences.

functional and non-functional copy of IgM were identified, corresponding S regions could not be identified in the regions upstream of G, E and A constant regions.

No marsupial IgD has been reported to date. The area predicted to contain IgD C region, in particular the region 3' of the IgM exons, between IgM and IgG, as well as the whole MonDom5 assembly were thoroughly searched for coding sequences that might correspond to a putative IgD (see Materials and Methods). This search included the trace sequences that remain unaligned to the whole genome sequence which are not necessarily part of the final assembly. These searches all failed to detect sequences with homology with mammalian IgD. As a positive control for the search criteria, the same methods were applied to the recently completed platypus whole genome sequence and revealed an IgD present in this species (not shown). In addition, the region in the sequence assembly that would be expected to contain IgD C exons, *i.e.* immediately 3' of the IgM C exons, is a relatively complete sequence and does not contain any large or suspicious gaps (Fig. 1). However there is a large duplicated region rich in repetitive DNA including LINE and endogenous retroviral (ERV) elements starting 1.5 kb 3' of the exons encoding IgM (Fig. 1, 3b, and 3c).

The IGH V *gene segments:* A total of 25 IGHV segments were identified on the assembly scaffold the IGH locus assembly present on chromosome 1. Twenty three of these gene segments belong to the previously identified IGHV1 subgroup (Fig. 1, Supplementary Table 1). Of the 23 IGHV1 gene segments, 18 appear to be fully functional based on containing leader sequences, open reading frames (ORF) and what appear to be

functional recombination signal sequences (RSS). The remaining five all contain in frame stop codons or are partial sequences and have been designated pseudogenes (Fig. 1, Supplementary Table 1). The largest gaps in the assembly in this region of the opossum genome are found amongst the IGHV gene segments (Fig. 1), making it possible that there are as yet unidentified IGHV in the opossum. It seems unlikely, however, that it will be a large number of additional gene segments given the consistency of current state of the assembly with previous analyses of IgH cDNAs and Southern blots (Miller et al. 1998).

Phylogenetic analysis including all the the germline IGHV gene sequences reveals that the IGHV1 family form a monophyletic clade interspersed by only the single available IGHV gene sequence from the North American opossum *Didelphis virginiana* (Fig. 4). Sister to, or immediately outside, this clade are the available IGHV genes from Australian marsupial species, the tammar wallaby, bandicoot and brushtail possum. These results confirm previous speculation that the diversity of the majority of IGHV segments in the opossum is fairly limited (Miller et al. 1998; Aveskogh et al. 1999; Baker et al. 2005).

In agreement with previous Southern analysis, there is only a single gene segment belonging to the IGHV2 subgroup in the opossum genomic sequence (Fig. 1, Miller et al. 1998). This gene segment is physically the most distal or 5' gene segment in the locus (Fig. 1) and phylogenetically the most divergent marsupial IGHV identified so far, being outside of the clade containing all other available marsupial IGHV sequences (Fig. 4,

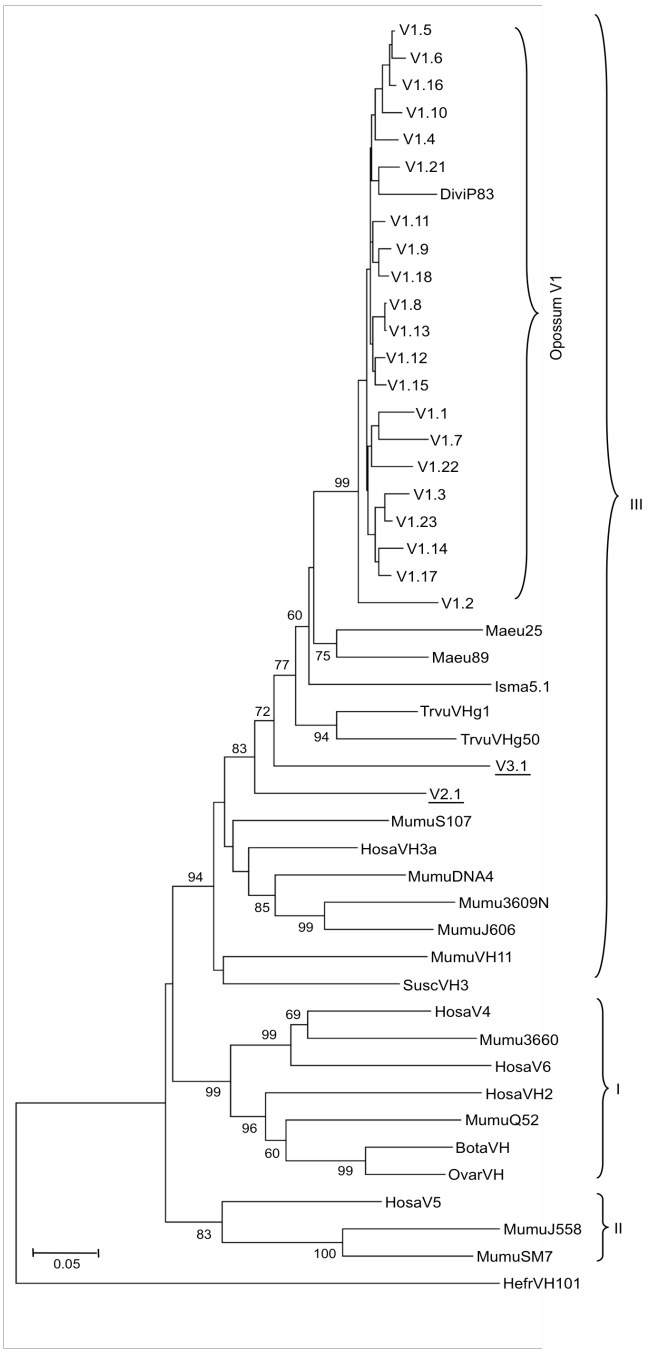


Figure 4. Phylogenetic tree based on alignments of IGHV sequences from opossum and other representative mammals and non-mammals. This tree was constructed using the NJ method. The numbers at the branch nodes indicate percent boostrap values on 1000 replicates. Only those values greater than 50% are shown. Also no values even greater than 50% are shown for within the opossum IGHV1 clade. Roman numerals on the right indicate the three IGHV clans. Species names are abbreviated to the first two letters of their scientific names: *B. taurus* (Bota), *D. virginiana* (Divi), *H. sapiens* (Hosa), *H. fransciscii* (Hefr), *I. macrourus* (Isma), *M. eugenii* (Maeu), *M. musculus* (Mumu), *O. aries* (Ovar), *S. scrofa* (Susc), *T. vulpecula* (Trvu). The *M. domestica* IGHV are those that do not contain a species designation. The *M. domestica* IGHV1 subgroup is bracketed and IGHV2.1 and 3.1 are underlined. Scale bar is branch length based on substitution rates (Kumar et al. 2004).

Baker et al. 2005).

One unexpected result from the analysis of the opossum genomic sequence was the discovery of a IGHV gene segment that represents a third, previously unrecognized subgroup. One unusual feature of this IGHV segment, designated IGHV3.1, is the length of the exon encoding the extra-cellular V domain which has an ORF that is 43 bp longer than typical opossum IGHV gene segments (Fig. 5). Further scrutiny of this gene segment revealed a RSS with canonical heptamer and nonamer sequences. They are separated, however, by a 12 bp spacer typical of D segments rather than the 23 bp spacer in IGHV RSS (Fig. 5). Given the unusual nature of the IGHV3.1 sequence, and the possibility of assembly artifacts, we cloned and sequenced IGHV3.1 directly from *M. domestica* genomic DNA and confirmed the sequence in the assembly as being accurate (not shown; sequence deposited in GenBank under accession no. EU592040). In summary, IGHV3.1 appears to be a partially joined gene segment where a D segment has been recombined to the end of the V in the germ-line.

The IGH D *gene segments:* Nine potential IGHD gene segments were identified in the opossum genome by scanning the IGH genomic region for conserved RSS sequences and examining the local sequence for a nearby, second RSS in the opposite orientation (Fig. 1, Table 2). The translations of the alternative reading frames for all nine are presented in Fig. 6. Analysis of expressed IgH VDJ recombinations using using the SoDA software package revealed that all but IGHD7 are being used and are therefore functional (not shown). This includes IGHD9, which is located amongst the IGHV gene segments and, in

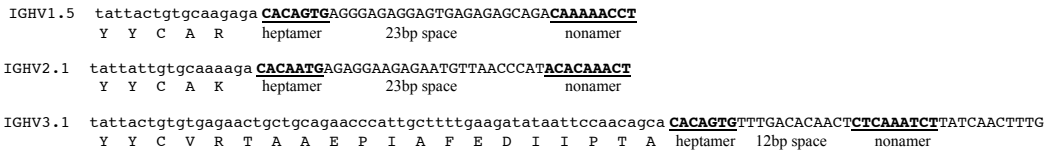


Figure 5. Comparison of the germ-line 3' coding sequences including the last five codons and RSS of a single representative functional germline IGHV1 gene and the IGHV2.1 gene with the corresponding region in IGHV3.1. The RSS regions corresponding to the heptamer, nonamer, and spacer are indicted.

Table 2. Sequences of IGHD exons and RSS flanking IGHD

IGHD	Nonamer	12 bp Spacer	Heptamer	Coding Sequence	Nonamer	23 bp Spacer	Heptamer
D1	TGCTTTTGA	CTTAGTCTGTGT	CACTGTG	GATACAGTAACTAC	CACAGTG	GCAAACACTAAT	TCTAAAAGT
D2	TGTTTATGA	CTTGCTCTGTGT	CATTGTG	GTTTATAGCTGGGGTA	CACAGTG	ATATATTTCAAA	TCTAAAACT
D3	GGATTCCAT	ACAGTAATGTAA	CACTGTG	TAACTACGGGTATAGCTA	CACAGTG	TTAATCCCTATT	ACAAAAATT
D4	AGATTCTGT	GTTGGACTCTGT	CACTGTG	CTATATTATTGTAGTAGTGGTATTTGCTACGAC	CACAGTG	ATAGATCTCCCA	TCAAAAACC
D5	AGGATGTTG	CCAAGGATGTAT	CACTGTG	CTATTAACTGGTGGTTGGAACTAC	CACAGTG	ACAAAGAACTTA	TCCAAAACC
D6	AGATTTTGA	ATGAGAATATAT	CATAGTG	TACTACTACAGTGGTAAC	CACATTG	AAACAACCTTTG	ACAAAAACC
D7	AGATTATGT	GCTAGACTGTGT	CACTGTG	GTATTATTGTTATAGTGATTATTGGGGCT	CACAGTG	ATCTCCAATCAA	AAACCCATC
D8	TAGATATAT	ACCTCCTGGTTC	CACTGTG	ATAAGCATTGAGTGGTACCTAGTCTAATCAACCTTTGA ATTC	CACAGAT	CATCCCTTATTC	CCCCAACCA
D9	TATAACAAT	CCGTCCAGCAAA	CACTGTG	TCCTTGTATCTCCAACCTTAGGCCATCT	CACATGG	AATATACATCTG	TCTGTGGCT

```
D1    gatacagtaactac
       D  T  V  T
        I  Q  *  L
         Y  S  N  Y


D2    gtttatagctggggta
       V  Y  S  W  G
        F  I  A  G  V
         L  *  L  G


D3    taactacgggtatagcta
       *  L  R  V  *  L
        N  Y  G  Y  S
         T  T  G  I  A


D4    ctatattattgtagtagtggtatttgctacgac
       L  Y  Y  C  S  S  G  I  C  Y  D
        Y  I  I  V  V  V  V  F  A  T
         I  L  L  *  *  W  Y  L  L  R


D5    ctattaactggtggttggaactac
       L  L  T  G  G  W  N  Y
        Y  *  L  V  V  G  T
         I  N  W  W  L  E  L


D6    tactactacagtggtaac
       Y  Y  Y  S  G  N
        T  T  T  V  V
         L  L  Q  W  *


D7    gtattattgttatagtgattattggggct
       V  L  L  L  *  *  L  L  G
        Y  Y  C  Y  S  D  Y  W  G
         I  I  V  I  V  I  I  G  A


D8    ataagcattgagtggtacctagtctaatcaacctttgaattc
       I  S  I  E  W  Y  L  V  *  S  T  F  E  F
        *  A  L  S  G  T  *  S  N  Q  P  L  N
         K  H  *  V  V  P  S  L  I  N  L  *  I


D9    tccttgtatctccaaccttaggccatct
       S  L  Y  L  Q  P  *  A  I
        P  C  I  S  N  L  R  P  S
         L  V  S  P  T  L  G  H
```

Figure 6. Nucleotide sequence and amino acid translations of all three reading frames of the opossum IGHD segments. * indicates in-frame stop codons.

recombination where it is used it is only in combination with IGHV segments that were upstream or 5' to the D segment (*e.g.* IGHV1.8 and 1.15 in Fig. 1). Furthermore, IGHD4, which is among the longer the D segments, encodes a pair of cysteines in one of its first reading frame (Fig. 6). Analysis of a large set of splenic IgH cDNAs revealed that this reading frame is used in V(D)J recombinations using IGHD4 (not shown).

The IGH J *gene segments:* Six IGHJ segments have been identified in the opossum genome, four of which appear to be functional: IGHJ1, 2, 4, and 5 (Fig. 1). The other two, IGHJ3 and 6, both contain in frame stops and they appear to be pseudogenes. The IGHJ segments are organized in two sets of three, with IGHJ1, 2, and 3 immediately upstream of the functional copy of the IgM C region exons and IGHJ4, 5, and 6 upstream of the IgM pseudogene. Several lines of evidence point to the two sets of J gene segments being created by the same duplication that gave rise to the second partial copy of IgM. The first is their genomic organization (Fig. 1 and 3a). Secondly, sequence analysis reveals IGHJ1 and 4 as a pair and IGHJ3 and 6 as a pair share >90% nucleotide identity within pairs but less than 40% between pairs. IGHJ2 and 4 also share greater identity with each other (79%) than with any other segments (77% or less) but the difference is less extreme. Lastly the mutations in IGHJ3 and 6 that render them pseudogenes are identical and it is likely they were already non-functional prior to the duplication event. From analysis of a large set of IgH cDNAs from adult opossum spleen, so far only IGHJ1 and 2 have were found to be used in transcribed V(D)J recombinations (Aveskogh et al. 1999; and data not shown)

The opossum IGK *locus*

The *IGK* locus was previously mapped to the distal end of the long arm of chromosome 1 in *M. domestica* (Deakin et al. 2006). The region of opossum genome assembly MonDom5 containing the *IGK* genes was analyzed and found to be 3,196 kb in length and appears to be well assembled, containing only three small gaps (Fig. 7). As predicted previously from Southern blot and cDNA sequence analyses, there is only a single IGK C region gene and two IGKJ gene segments in the opossum *IGK* locus (Fig. 7; Miller et al. 1999). Of the three Ig loci however, *IGK* is the most complex with respect to number and diversity of V segments with a total of 122 IGKV gene segments identified (Fig. 7; Supplementary Table 2). Previous analysis of IGK cDNAs in opossum revealed four IGKV subgroups (Miller et al. 1999). These four subgroups make up the majority (104) of the total V gene segments present in the opossum *IGK* locus (Fig. 7 and 8). The original subgroup designations of IGKV1 through 4 based on cDNA analysis were retained for consistency. As a result the ordering of gene segments along the *IGK* locus is not in numerical order, where IGKV1 gene segments are the most C proximal and IGKV2 are the most distal (Fig. 7). The remaining 18 IGKV genes comprise three previously undiscovered subgroups bringing the total to seven IGKV subgroups in the opossum (Fig. 7 and 8). Five of the seven IGKV subgroups contain both functional and pseudogene copies (Fig. 7, Supplementary Table 2). The two exceptions are IGKV5 and IGKV6 whose five and one gene segments, respectively, appear to be fully functional. In contrast only four of the twelve IGKV7 subgroup members appear to be functional (Fig. 7; Supplementary Table 2).

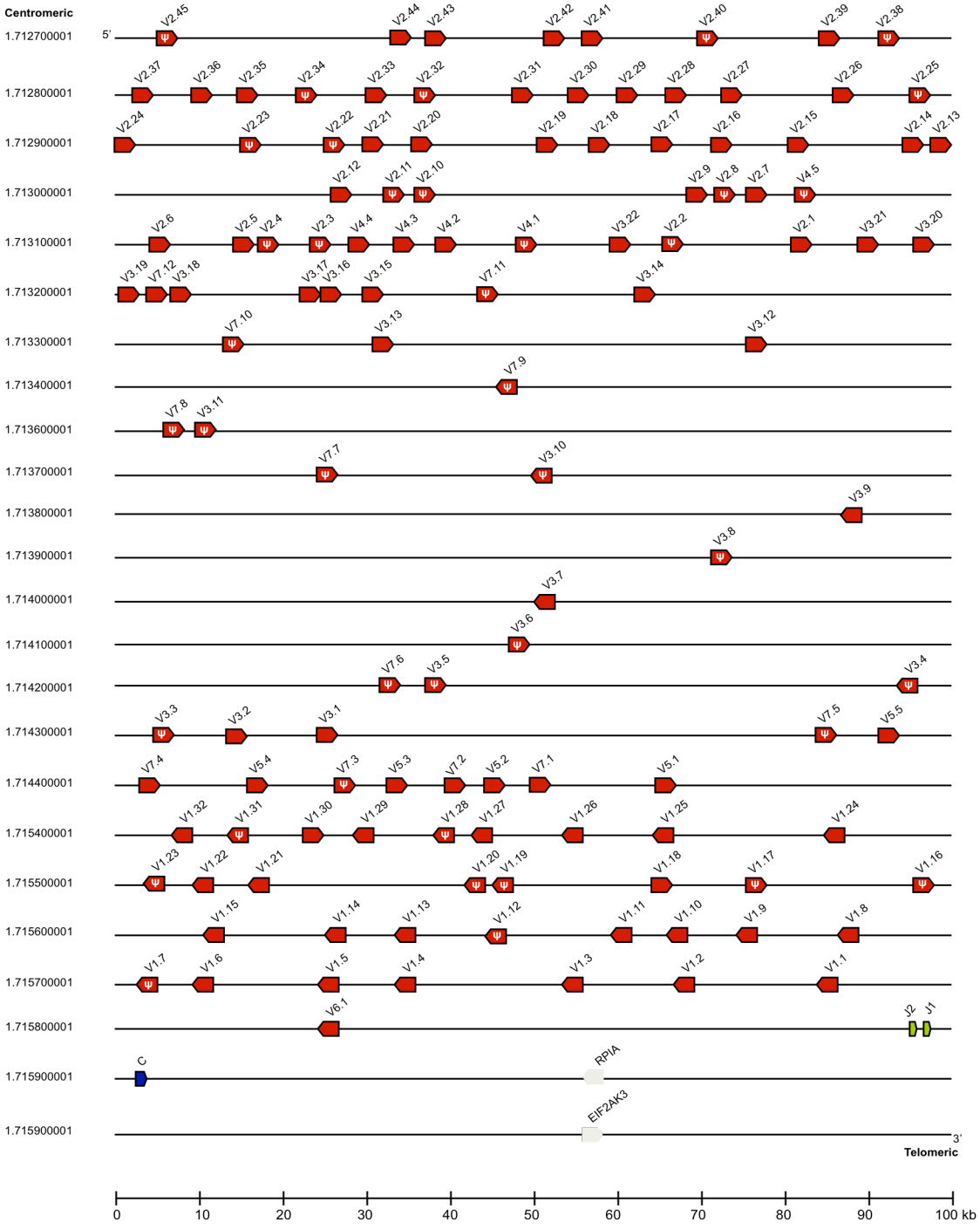


Figure 7. Diagram of the opossum *IGK* locus. The V, J and C genes are showed in red, green and blue, respectively. IGKV segments are designated with the subgroup number followed by a period and a designated number according to their order. IGKJ segments are designated according to their order. Direction of transcription is indicated by the shape of each segment. Presumed pseudogenes are indicated with the respective sign (ψ). Syntenic genes shown are ribose-5-phosphate isomerase (*RPIA*) and eukaryotic translation initiation factor 2-alpha kinase 3 precursor (*EIF2AK3*).

Apparent from the genomic organization of the *IGK* locus is that the IGKV gene segments exist in two large clusters separated by an approximately 800 kb region that is sparse with V segments (Fig. 7). In addition, the two dominant IGKV subgroups in each cluster, IGKV1 in the C proximal cluster and IGKV2 in the distal cluster, are for the most part in opposite transcriptional orientation. Most IGKV1 genes are in reverse reading frame relative to the J and C genes whereas IGKV2 subgroup members are in the same orientation. This organization is somewhat reminiscent of the structure of human *IGK* locus that also contains two large clusters of V genes in inverted orientation relative to each other (Kawasaki et al. 2001; Zachau 2004). Dot matrix analysis of the *IGK* region did not reveal any large genomic duplications that might explain this inverted organization (not shown). Furthermore, the phylogenetic relationship between IGKV1 and V2 segments does not support a recent duplication within the opossum *IGK* locus either (Fig. 8). Rather, IGKV1 and V2 appear to be the result of a more ancient duplication predating at least the divergence of marsupial and eutherian mammals. In other words it does not look as if the two clusters are the product of a large inverted duplication similar to what has been seen in the human *IGK* locus (Zachau 2004; Kawasaki et al. 2001).

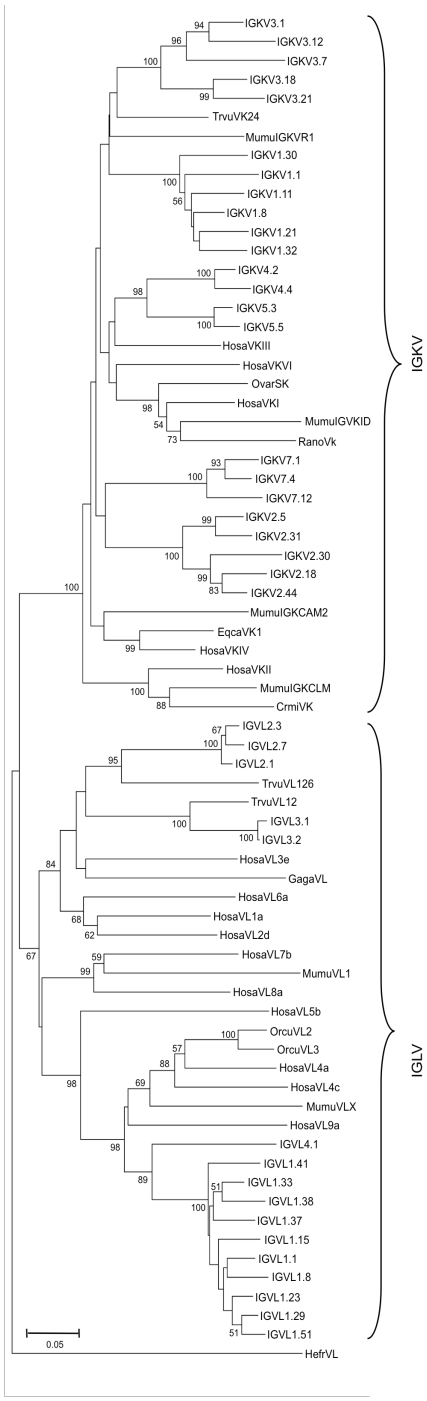


Figure 8. Phylogenetic tree based on alignments of IGKV and IGLV sequences representing each subgroup from opossum and other representative mammals and non-mammals constructed using the NJ method. Not all *M. domestica* V sequences were included due to the large number. The numbers at the branch nodes indicate percent bootstrap values on 1000 replicates. Only those values greater than 50% are shown. Species designations are abbreviated to the first two letters of their scientific names as in Fig. 1 legend with the inclusion of: *C. migratorus* (Crmi), *E. caballus* (Eqca), *G. gallus* (Gaga), *O. anatinus* (Oran), and *R. norvegicus* (Rano). The *M. domestica* IGKV and IGLV sequences are those without a species designation. The IGKV and IGLV clades are indicated by brackets. Scale bar is branch length based on substitution rates (Kumar et al. 2004).

The opossum IGL *locus*

The *IGL* locus was previously located to the distal end of the long arm of opossum chromosome 3 (Deakin et al. 2006). This region of opossum genome assembly MonDom5 also appears to be well assembled, although there are more sequence gaps than were found in *IGK* (Fig. 9). The *IGL* locus is also the longest of the three Ig loci, spanning 3,797 kb in length. The location of each coding segment within the *IGL* locus is provided in Supplementary Table 3. As predicted earlier, the IGLJ and C gene genes are organized in J-C pairs, much like has been described in other mammals (Lucero et al. 1998; Lefranc and Lefranc 2004). Based on Southern blot and cDNA sequence analyses it was estimated that there were at least six J-C pairs in *M. domestica* (Lucero et al. 1998). This number is fairly close to the actual eight J-C pairs found in the MonDom5 assembly (Fig. 9; Supplementary Table 3).

From the analysis of a large set of IGL cDNA clones, three subgroups of IGLV gene segments were identified, of which IGLV1 was clearly the most abundant based on Southern blot analysis (Lucero et al. 1998). Of the 64 total IGLV gene segments identified within the *IGL* locus, 54 belong to the IGLV1 subgroup, all but six of which appear functional by having an ORF and conserved RSS (Fig. 8 and 9; Supplementary Table 3). In addition to the original three subgroups, a fourth IGLV subgroup was identified (IGLV4.1 in Fig. 8 and 9), which is a single, apparently functional V gene segment. Similar to IGKV and in contrast to marsupial IGHV, the IGLV subgroups intersperse amongst V genes of other species, as described previously (Fig. 8; Lucero et al. 1998).

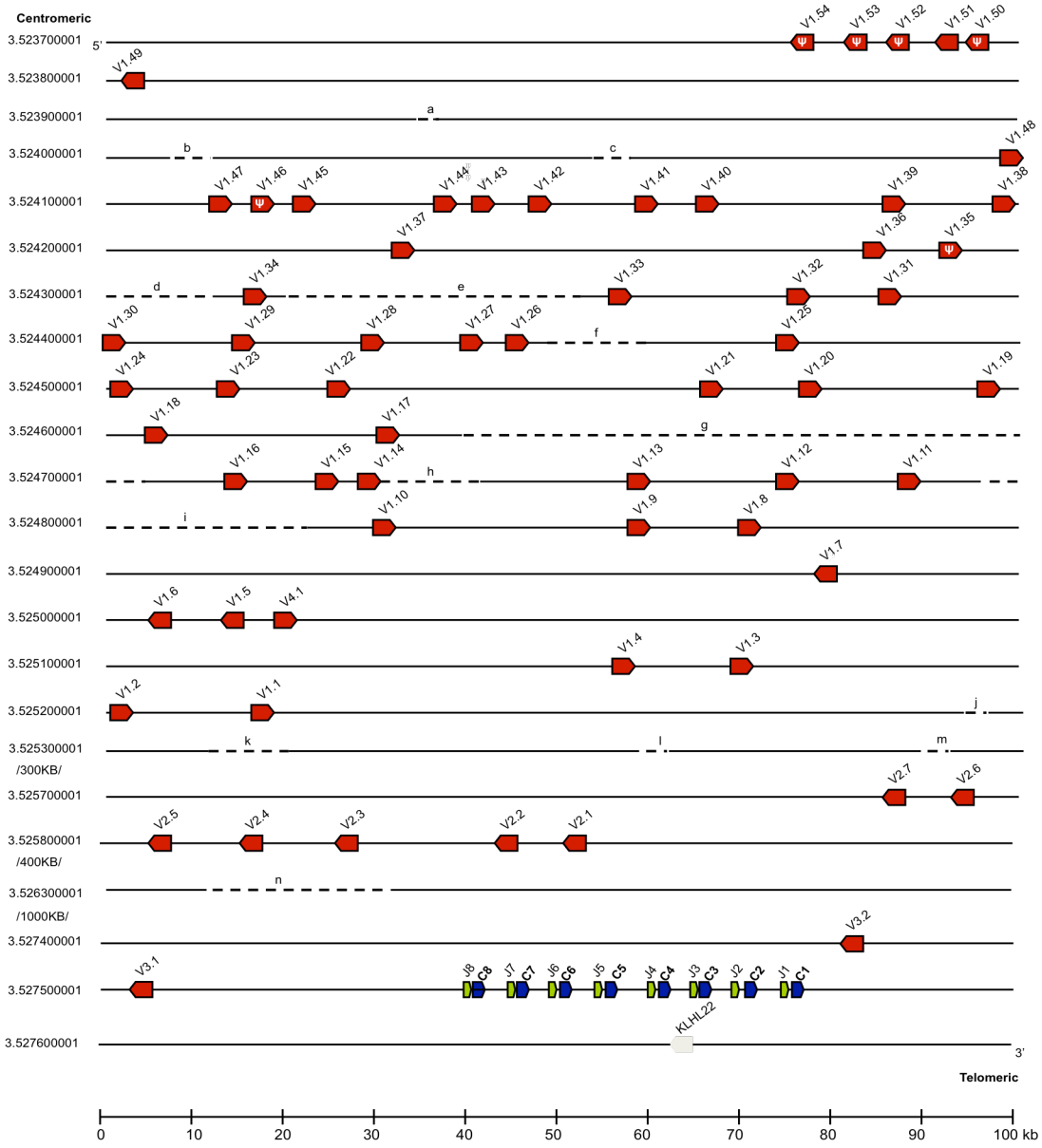


Figure 9. Diagram of the opossum *IGL* locus. V, J and C genes are showed in red, green and blue, respectively. IGLV segments are designated with the subgroup number followed by a period and a designated number according to their order. IGLJ gens are designated according to their order. Direction of transcription is indicated by the shape of each segment. Presumed pseudogenes are indicated with the respective sign (ψ). Syntenic gene shown is Kelch-like protein 22 (*KLHL22*).

Unaligned Genomic Sequence

The *M. domestica* genome assembly contains traces that assembled to short scaffolds but which did not assemble to the longer chromosomes. These are provided as the unassigned or unassembled (Un) chromosome associated with MonDom5. Some of these sequences are clearly allelic to loci assigned to the assembled chromosomes, a problem created by the fact that the individual animal sequenced was not fully inbred (Mikkelsen et al. 2007). Searching the unassigned scaffolds for sequences corresponding to C regions of Ig loci revealed only a single scaffold (Un 60100001) that contained three IGHJ segments and a complete set of IgM C region exons, appearing to be an allele of the region containing the functional IGHJ through IgM C regions. There were also eight IGHV gene segments identified among the Unassigned scaffolds, all belonging to the IGHV1 subfamily (Supplementary Table 1). The other two are partial sequences and were excluded from this analysis. It is difficult to say if these IGHV are alleles of V gene segments assembled in the IGH locus or represent missing sequences, perhaps located in the gaps present in the current assembly (Fig. 3). It is worth noting that the total number of IGHV present in the chromosome 1 of the MonDom5 assembly is not substantially different from that predicted earlier by Southern blot analysis (Miller et al 1998). Therefore we suspect that many of the unassigned IGHV sequences represent allelic variants excluded from the assembly. There is a single IGHJ in the unassigned sequences that is highly similar IGHJ2 and, based on nearly identical flanking sequence appears to be a second allele of IGHJ2 (not shown). There are no recognizable sequences resembling IGHD segments in the unaligned sequences. There is one IGKV present in the unassigned sequences which is identical to VK7.7, including flanking sequences

covering a 4 kb region. It is not apparent why this sequence was excluded from the assembly since it is not an allelic variant but identical to the assembled sequence. There are also nine IGLV gene segments all having identity to the IGLV1 family.

Discussion

M. domestica is the first, and so far only, marsupial species for which an assembled whole genome sequence has been produced (Mikkelsen et al. 2007, Renfree 2007). And with this report the opossum becomes the first non-eutherian mammal, and one of the few vertebrate species of any lineage, for which detailed organization of the Ig loci has been determined and fully annotated.

As with all marsupial species, the opossum presents a number of immunological problems, particularly with respect to immunity. Marsupial young are born highly altricial, being developmentally equivalent at birth to eight-week human embryos in many respects (Deane and Cooper 1988). Much of the development of the immune system appears to occur entirely postnatally (Deane and Cooper 1988; Parra et al 2009). *M. domestica*, like most marsupial species that have been studied do not transfer Ig from mother to fetus trans-placentally, but depend entirely on transfer of milk antibodies for maternal immunity (Samples et al. 1986). The one known exception is the tammar wallaby *Macropus eugenii*, for which there is clear evidence of prenatal transfer of Ig during pregnancy (Renfree 1973; Deane et al 1990). Fortunately, the tammar wallaby is one of the other marsupial species for which there is an active genome project, which will facilitate comparative studies of maternal immunity in marsupials in the future (Wakefield and Graves 2003). A long-term goal is to determine when during development the marsupial young become immuno-competent and what are the contributions of maternal immunity to protection during postnatal development. Determining the germ-line gene segments that contribute to antibody diversity in the

opossum provides, in part, the information necessary to evaluate the state of B cell development and the level of diversity being generated at different ontogenic time-points. Therefore, one of the immediate goals of this research was to develop detailed annotation of the Ig heavy and light chain loci in the genome of *M. domestica.*

The *IGH* locus in the MonDom5 assembly appears fairly complete, at least in that the organization and complexity of the locus is consistent with previous sequence analyses of IgH chain cDNAs from opossum. For example, and in spite of sequence gaps, the total number of functional *IGHV* gene segments present in the assembly (18 IGHV1 and one IGHV2) is not very different from the approximately 15 functional V segments predicted from cDNA sequences and Southern blot analyses (Miller et al. 1998). Furthermore, the presence of only single copies of each of the heavy chain isotypes (M, G, E and A) and only two IGHJ segments being used were all predicted from cDNA analyses (Belov et al 1999; Aveskogh et al 1998, 1999; Miller et al. 1998).

Previous investigators had reported the presence of at least two IgG subclasses in several marsupial species including *M. domestica* (Bell 1977; Bell et al. 1974; Shearer et al. 1995) but genomic analysis reveals that there is only a single IgG in *M. domestica*. In many cases marsupial IgG was defined based on binding to Staphylococcal Protein-A (SpA). All marsupial IGHV described so far are clan III segments (Fig. 4; Miller et al. 1998; Baker et al. 2005) similar to that which in humans bind SpA as a super-antigen (Silverman and Goodyear 2002). SpA binding in the marsupial studies likely resulted in a mix of Ig isotypes due to inadvertent binding to the common IGHV family. Indeed it was

noted that serum IgM from *M. domestica* binds SpA and it is possible this result is best explained through binding to the V domain, although this would need to be shown (Shearer et al. 1995).

Internal duplications and insertions appear to have contributed to the evolution of the opossum *IGH* locus, especially in the region containing the C genes. One duplication gave rise to additional IGHJ gene segments and a non-functional copy of the IgM constant region genes. The distance between the duplicated J segments and the partial IgM pseudogene is greater than it is for the functional copy and is due to the insertion of a LINE element just 3' of the region corresponding to the Sμ of the pseudogene (Fig. 3a). It is possible that this mobile element contributed to the local genomic instability that resulted in the duplication event. However, it is also possible that this insertion occurred later since interspersed repeat type retroelements such as LINEs and SINEs are fairly common in the opossum genome. In fact the opossum genome contains the greatest fraction of such repetitive elements amongst animal genomes sequenced so far (Mikkelsen et al 2007).

The insertion of repetitive elements may also have contributed to loss of IgD in this species. So far, no cDNA clones corresponding to an IgD have been reported for any marsupial species (Miller and Belov 2000). Homology based searches of the region expected to contain IgD, and the whole genome sequence in MonDom5 were negative as well. Furthermore, there is a large duplicated region rich in repetitive DNA including LINE and ERV elements downstream of the IgM constant region where IgD would be

expected to be located and it is possible their insertion contributed to a loss of the IgD in this species (Fig.s 1, 3b, and 3c). Whether other marsupials contain this duplicated region is not known. However, recent analyses that support IgD being an ancient isotype, and from its presence in the platypus as well as eutherian mammals, it is clear that the absence of IgD in the opossum represents a gene loss in this marsupial (Ohta et al. 2006; Wilson et al. 1997; and data not shown).

One of the unanticipated results of the analysis of the *IGH* locus in the opossum was the presence of a third IGHV subgroup that appears to be a partially germ-line joined Ig V gene, the first such to be described in a mammal to our knowledge. Based on scrutiny of its sequence IGHV3.1 appears to be fully functional by having a typical leader sequence, intron, and ORF. The presence of both an intron separating the exon encoding the L sequence from the rest of the V domain in IGHV3 and an RSS at the end of the coding sequence is consistent with this gene segment not being generated by retro-transposition. Rather it appears to be the product of direct recombination activation gene (RAG) mediated V to D recombination in the germ-line similar to what is thought to have created the germ-line joined V genes in cartilagenous fishes (Lee et al. 2000). This is in contrast to the only other known mammalian germ-line joined V gene, the Vμj gene found in TCRμ, a unique TCR also discovered in marsupials (Parra et al. 2007). Vμj appears to have involved a retrotransposition step in its creation due to the lack of an intron separating the L and V exons. In other words Vμj has the characteristics of a processed gene that is still functional (Parra et al. 2007, 2009). Whether IGHV3 contributes to antibody diversity in the opossum remains to be determined, however,

preliminary attempts to identify heavy chain cDNA clones containing IGHV3 from opossum adult spleen have been unsuccessful (not shown). This may not be surprising given that rearrangement of IGHV3.1 to a J segment would be an atypical V(D)J recombination since typically D to J rearrangement precede V to D in developing B cells (reviewed in Melchers and Kincade 2004). Whether IGHV3 can serve as a substrate for RAG recombination or perhaps contribute to diversity in other ways such as through gene conversion remains to be determined. The latter is an intriguing possibility given that the IGHV pseudogenes in the chicken that are used in gene conversion to diversify the primary antibody repertoire are themselves partially germline joined (V-D) gene segments similar to IGHV3 (Reynaud et al. 1989).

The discovery of a new IGHV subgroup in opossum, whether functional or not, supports that at one time marsupials may have had greater available V gene diversity than is currently extant in the *IGH* locus. However, opossum IGHV3 like all other marsupial IGHV found so far is still a member of clan III, the most conserved or widespread of the heavy chain V genes (Baker et al 2005; Tutter and Riblet 1989). These results are also consistent with a recent large analysis of IGHV genes by Das and colleagues (2008) that included many of the germline opossum IGHV. Our analysis of the Ig light chain loci also are consistent with earlier conclusions that, in contrast to *IGH*, the *IGK* and *IGL* loci have a great deal of sequence diversity and complexity, supporting the hypothesis that light chains may contribute more to antibody diversity than heavy chains in opossums particularly, and perhaps in marsupials in general (Baker et al 2005).

IGHD4, one of the longest the D segments used, encodes a pair of cysteines in one of its first reading frame (Fig. 6). Analysis of a large set of splenic IgH cDNAs revealed that this reading frame is used (not shown). It is possible that this D segment, when used in the first reading frame, is used when internal cysteine bridges are needed for stability in particularly long CDR3 regions, much like as has been described in the duckbill platypus, camel, cow, and shark (Johansson et al. 2002, Muyldermans et al. 1994; Roux et al. 1998, Saini et al. 1999).

From analysis of a large set of IgH cDNAs from adult opossum spleen, only IGHJ1 and 2 have were found to be used in expressed VDJ recombinations. This result may explain the earlier estimates of only two IGHJ genes in opossum based analysis of heavy chain cDNAs (Aveskogh et al 1999). Based on the organization of the IGH locus this may not be surprising given that IGHJ1 and 2 are the apparently functional J segments immediately upstream of the functional IgM C region exons and downstream of the majority of IGHD segments (Fig. 1).

The organization and complexity of the opossum *IGK* and *IGL* loci, including the estimated number of J and C genes, are also similar to what was predicted previously (Lucero et al. 1998; Miller et al. 1999). There were additional IGKV and IGLV gene subgroups uncovered in the genomic sequence, however there were in all cases relatively smaller families in gene copy number and are likely to be rare in the repertoire, perhaps explaining why they were missed in earlier analyses. The consistency with previous

predictions is not meant to belittle the value in determining the genomic organization, rather it is meant to support confidence in the MonDom5 assembly.

In conclusion: detailed, annotated genomic maps of the Ig loci have now been established for the first time for a marsupial mammal. This annotation serves as a resource for further analysis of B cell diversity and ontogeny in *M. domestica*, helping to establish what are the germ-line versus somatic contributions the expressed antibody repertoire. These results also solidify many of the conclusions regarding Ig locus genomic organization in this species that up until now have been primarily speculation based on cDNA and limited genomic DNA analysis.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410

Aveskogh M, Hellman L (1998) Evidence for an early appearance of modern post-switch isotypes in mammalian evolution; cloning of IgE, IgG and IgA from the marsupial *Monodelphis domestica*. Eur J Immunol 28: 2738-2750

Aveskogh M, Pilström L, Hellman L (1999) Cloning and structural analysis of IgM (μ chain) and the heavy chain V region repertoire in the marsupial Monodelphis domestica. Dev Comp Immunol 23: 597-606

Baker ML, Belov K, Miller RD (2005) Unusually similar patterns of antibody V segment diversity in distantly related marsupials. J Immunol 174:5665-5671

Bell RG (1977) Marsupial immunglobulins: the distribution and evolution of macropod IgG2, IgM and light chain antigenic markers within the sub-class metatheria. Immunol 33:917-924

Bell RG, Lynch NR, Turner KJ (1974) Immunoglobulins of the marsupial *Setonix brachyurus* (quokka). Immunol 27:1103-1115

Belov K, Deakin JE, Papenfuss AT, Baker ML, Melman SD, Siddle HV, Gouin N, Goode DL, Sargeant TJ, Robinson MD, Wakefield MJ, Mahony S, Cross JGR, Benos PV, Samollow PB, Speed TP, Graves JAM, Miller RD (2006) Reconstructing an ancestral mammalian immune supercomplex from a marsupial MHC. PLoS Biol 4: e46 doi:10.1371/journal.pbio.0040046

Belov K, Harrison GA, Rosenberg GH, Miller RD, Cooper DW (1999) Isolation and comparison of the IgM heavy chain constant region from Australian (*Trichosurus vulpecula*) and American (*Monodelphis domestica*) marsupials. Dev Comp Immunol 23: 649-656

Belov K, Hellman L (2003) Immunoglobulin genetics of *Ornithorhynchus anatinus* (platypus) and *Tachyglossus aculeatus* (short-beaked echidna). Comp Biochem Physiol, Part A Mol Integr Physiol 136: 811-819

Das S, Nozawa M, Klein J, Nei M (2008) Evolutionary dynamics of the immunoglobulin heavy chain variable region genes in vertebrates. Immunogenetics 60: 47-55

Deakin JE, Olp JJ, Graves JAM, Miller RD (2006) Physical mapping of immunoglobulin loci *IGH@*, *IGK@*, and *IGL@* in the opossum (*Monodelphis domestica*). Cytogenet Genome Res 114: 94

Deane EM, Cooper DW (1988) Immunological development in pouch young marsupials. In: Tyndale-Biscoe CH, Janssens PA (ed) The Developing Marsupial. Springer-Verlag, Berlin, pp 190-199

Deane EM, Cooper DW, Renfree MB (1990) Immunoglobulin G levels in fetal and newborn tammar wallabies (*Macropus eugenii*) Reprod. Fert Dev 2:369-375

Fry EJ, Saunders NR (2000) Spinal repair in immature animals: a novel approach using the South American opossum *Monodelphis domestica*. Clin Exp Pharmacol Physiol 27:542-547

Hall TA (1994) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41: 95-98

Johansson J, Aveskogh M, Munday G, Hellman L (2002) Heavy chain V region diversity in the duck-billed platypus (*Ornithorhynchus anatinus*): long and highly variable complementarity-determining region 3 compensates for limited germline diversity. J Immunol 168: 5155-5162

Kawasaki K, Minoshima S, Nakato E, Shibuya K, Shintani A, Asakawa S, Sasaki T, Klobeck HG, Combriato G, Zachau HG, Shimizu N (2001) Evolutionary dynamics of the human immunoglobulin kappa locus and the germline repertoire of the Vkappa genes. Eur J Immunol 31: 1017-1028

Kawamura S, Saitou N, Ueda, S (1992) Concerted evolution of the primate immunoglobulin α-gene through gene conversion. J Biol Chem 267: 7359-7367

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 5:150-163

Lee SS, Fitch D, Flajnik MF, Hsu E (2000) Rearrangement of immunoglobulin genes in shark germ cells. J Exp Med 191: 1637-1648

Lefranc MP, Lefranc G (2004) Immunoglobulin lambda (IGL) genes of human and mouse. In: Honjo T, Alt FW, Neuberger M (ed) Molecular Biology of B Cells. Academic Press, London, pp 37-59

Ley RD, Reeve VE, Kusewitt DF (2000) Photobiology of *Monodelphis domestica*. Dev Comp Immunol 24: 503-516

Lucero JE, Rosenberg GH, Miller RD (1998) Marsupial light chains: Complexity and conservation of Igl in the opossum *Monodelphis domestica*. J Immunol 161: 6724-6732

Melchers F, Kincade P (2004) Early B cell development to a mature, antigen-sensitive cell. In: Honjo T, Alt FW, Neuberger M (ed) Molecular Biology of B Cells. Academic Press, London, pp 101-126

Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A, Jurka J, Kamal M, Mauceli E, Searle SMJ, Sharpe T, Baker ML, Batzer MA, Benos PV, Belov K, Clamp M, Cook A, Cuff J, Das R, Davidow L, Deakin JE, Fazzari MJ, Glass JL, Grabherr M, Greally JM, Gu W, Hore TA, Huttley GA, Jirtle RL, Koina E, Lee JT, Mahony S, Marra MA, Miller RD, Nicholls RD, Oda M, Papenfuss AT, Parra ZE, Pollock DD, Ray DA, Schein JE, Speed TP, Thompson K, VandeBerg JL, Wade CM, Walker JA, Waters PD, Webber C, Weidman JR, Xie X, Zody MC, Broad Institute Genome Sequencing Platform, Broad Institute Whole Genome Assembly Team, Marshall Graves JA, Ponting CP, Breen M, Samollow PB, Lander ES, Lindblad-Toh KS (2007) Genome of the marsupial *Monodelphis domestica* reveals lineage-specific innovation in coding sequences. Nature 447: 167-178

Miller RD, Grabe H, Rosenberg GH (1998) The V_H Repertoire of a Marsupial: *Monodelphis domestica*. J Immunol 160: 259-265

Miller RD, Bergemann ER, Rosenberg GH (1999) Marsupial light chains: Igk with four V families in the opossum *Monodelphis domestica***.** Immunogenetics 50: 329-335

Miller RD, Belov K (2000) Immunoglobulin genetics of marsupials. Dev Comp Immunol 24: 485-490

Mills FC, Brooker JS, Camerini-Otero RD (1990) Sequences of human immunoglobulin switch regions: implications for recombination and transcription. Nucleic Acids Res 18: 7305-7316

Muyldermans S, Atarhouch T, Saldanha J, Barbosa JA, Hamers R (1994) Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. Protein Eng 7:1129-1135

Nikaido T, Yamawaki-Kataoka Y, Honjo T (1982) Nucleotide sequences of switch regions of immunoglobulin Cepsilon and Cgamma genes and their comparison. J Biol Chem 257: 7322-7329

Ohta Y, Flajnik M (2006) IgD, like IgM, is a primordial immunoglobulin class perpetuated in most jawed vertebrates. Proc Natl Acad Sci U S A 103: 10723-10728

Osborne BA, Golde TE, Schwartz RL, Rudikoff S (1988) Evolution of the IgA Heavy Chain Gene in the Genus Mus. Genetics 119: 925-931

Parra ZE, Baker ML, Hathaway J, Lopez AM, Trujillo J, Sharp A, Miller RD (2008) Comparative genomic analysis and evolution of the T cell receptor loci in the opossum *Monodelphis domestica*. BMC Genomics 9:111

Parra ZE, Baker ML, Lopez AM, Trujillo J, Volpe JM, Miller RD (2009) TCRμ recombination and transcription relative to the conventional TCR during postnatal development in opossums. J Immunol 182:154-163

Parra ZE, Baker ML, Schwarz R, Deakin JE, Lindblad-Toh K, Miller RD (2007) Discovery of a new T cell receptor in marsupials. Proc Natl Acad Sci U S A 104: 9776-9781

Renfree MB (1973) The composition of fetal fluids of the marsupial *Macropus eugenii*. Dev Biol 33: 62-79

Renfree MB (2007) The genome of a Gondwanan mammal. Bioessays 29: 1073-1076

Reynaud CA, Dahan A, Anquez V, Weill JC (1989) Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. Cell 59: 171-183

Roux KH, Greenberg AS, Greene L, Strelets L, Avila D, McKinney EC, Flajnik MF (1998) Structural analysis of the nurse shark (new) antigen receptor (NAR): molecular convergence of NAR and unusual mammalian immunoglobulins. Proc Natl Acad Sci U S A 95: 11804-11809

Saitou N, Nei M (1987) The neighbor joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425

Saini SS, Allore B, Jacobs RM, Kaushik A (1999) Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM Abs. Eur J Immunol 29:2420-2426

Samollow PB (2006) Status and applications of genomic resources for the gray, short-tailed opossum, *Monodelphis domestica*, and american marsupial model for comparative biology. Australian Journal of Zoology 54: 173-196

Samples NK, VandeBerg JL, Stone WH (1986) Passively acquired immunity in the newborn of a marsupial (*Monodelphis domestica*). Am J Reprod Immunol Microb 11:94-97

Shearer MH, Robinson ES, VandeBerg JL, Kennedy RC (1995) Humoral immune response in a marsupial *Monodelphis domestica*: Anti-isotypic and anti-idiotypic responses detected by species specific monoclonal anti-immunoglobulin responses. Dev Comp Immunol 19: 237-46

Silverman GJ, Goodyear CS (2002) A model B-cell superantigen and the immunobiology of B lymphocytes. Clin Immunol 102:117-134

Tutter A, Riblet R (1989) Conservation of an immunoglobulin variable-region gene family indicates a specific, noncoding function. Proc Natl Acad Sci U S A 86: 7460-7464

VandeBerg JL, Robinson ES (1997) The laboratory opossum (*Monodelphis Domestica*) in biomedical research. In: Saunders NR, Hinds LA (ed) Marsupial Biology: Recent Research, New Perspectives. UNSW, Sydney, pp 238-253

Volpe JM, Cowell LG, Kepler TB (2006) SoDA: implementation of a 3D alignment algorithm for inference of antigen receptor recombinations. Bioinformatics 22: 438-444

Wakefield MJ, Graves JAM (2003) The kangaroo genome. Leaps and bounds in comparative genomics. EMBO Rep 4: 143-147

Wilson M, Bengten E, Miller NW, Clem LW, Du Pasquier L, Warr GW (1997) A novel chimeric Ig heavy chain from a teleost fish shares similarities to IgD. Proc Natl Acad Sci U S A 94: 4593-4597

Wong ES, Young LJ, Papenfuss AT, Belov K (2006) In silico identification of opossum cytokine genes suggests the complexity of the marsupial immune system rivals that of eutherian mammals. Immunome Res 2:4

Zachau HG (2004) Immunoglobulin kappa genes of human and mouse. In: Honjo T, Alt FW, Neuberger M(ed) Molecular Biology of B Cells. Academic Press, London, pp 27-36

CHAPTER 3

POSTNATAL ONTOGENY OF THE IG REPERTOIRE AND B CELL MATURATION IN THE OPOSSUM, *MONODELPHIS DOMESTICA*

By Xinxin Wang, Alana R. Sharp and Robert D. Miller*

*Center for Evolutionary and Theoretical Immunology and Department of Biology, University of New Mexico, Albuquerque NM 87131 USA

XW and RDM conceived of the study, participated in its design, and drafted the manuscript. XW and ARP generated the data. XW performed the sequencing reaction and data analyses.

Manuscript formatted for a planned submission to Journal of Immunology

Abstract

Marsupials are a lineage of mammals noted for giving birth to highly altricial young that complete much of their "fetal" development externally, attached to a teat. Postnatal B cell ontogeny and diversity was investigated in a model marsupial species, the gray short-tailed opossum, *Monodelphis domestica.* The results support the initiation of Ig heavy chain V(D)J recombination is occurring within the first 24 hours postpartum. The expression of the surrogate L chain *VpreB3* was detected on postnatal day 6 with subsequent rearrangement of the Igλ and Igκ L chain genes occurring at postnatal days 7 and 8, respectively. The timing of the likely appearance of B cells based on gene rearrangements is consistent with previous analysis on the timing of antibody responses in newborn marsupials. The diversity of early B cell H chains is limited and reduced in N region additions, as has been seen in fetal humans and mice, but lacks bias in the V, D and J segments used. Rather the lack of N region additions may facilitate rapid expansion of B cell numbers in the newborn opossum. Newborn L chain diversity is, from the start, comparable to that of the adult, consistent with the hypothesis that L chains contribute extensively to antibody diversity in this species.

Introduction

The degree of immune competence of newborn mammals can vary quite a bit between species. A newborn mouse, for example, is much less developed than the more immunologically precocious newborn cow or pig (Butler et al 2006; Ishino et al 1991). Whether a species is considered altricial or precocial at birth, of course, is relative (Derrickson 1992). The marsupials are one of three living lineages of mammals (placentals, marsupials, and monotremes [*e.g.* the egg laying platypus]), which differ substantially in their state of development at birth. Marsupials, such as opossums and kangaroos, when compared with any placental mammal are born in an extremely altricial state. Indeed, the developmental state of the newborn marsupial immune system has been equated to that of a human embryo at the eighth to tenth week of gestation or a mouse or rat at the tenth day of gestation (La Via et al 1963, Deane & Cooper, 1984). Therefore, much of the development occurring in prenatal humans and other placental mammals is postnatal in marsupials, making marsupials unique models of early immune system development.

Indicative of their altricial state, newborn marsupials are unable to initiate endogenous immune responses until they are generally at least a week old. The North American opossum *Didelphis virginiana*, for example was unable to generate a T-dependent antibody response until greater than seven days of age (Kalmutz 1962, La Via et al 1963, Rowlands et al 1972). Similar results have been found with other marsupial species (Stanley et al., 1972). The ability to generate cell-mediated immune responses such as

transplant rejection also develops late in newborn marsupials. Skin allografts are usually tolerated prior to 12 days of age in species where it has been tested (LaPlante et al 1969, Yadav et al 1971). Allografts at later time points, however, are rejected consistent with the eventual appearance of functional T cells. The postnatal development of immune-competence is also consistent with the appearance of cells expressing lymphocyte markers in newborn marsupials. In tammar wallabies, for example, $CD3^+$ lymphocytes were first detected in the thymus at postnatal day 12, however cells expressing the B cell marker CD79b (Ig-b) could be found in gut associated lymphoid tissue as early as day 7 (Old and Deane 2003).

Placental species can differ in the diversity of the antibody repertoire generated during fetal development when compared with that of the adult. One mechanism for this difference has been associated with the lack of N-additions in the VDJ junctions, due to low or absent expression of TdT (Feeney 1990). The absence of N-additions is thought to achieve multiple goals in early B cell development. For one, it is associated with bias in VDJ recombination driven by short sequence homology. Such bias appears to be one mechanism for preferentially generating beneficial idiotypes specific for common pathogens such as the protective anti-phosphorylcholine response in mice (Benedict and Kearney 1999). In addition, B cells without N additions more rapidly populate lymphoid tissues, which may be advantageous early in ontogeny when first seeding peripheral lymphoid sites, even at the expense of resulting in less efficient responses (Schelonka et al., 2011). Limited N-additions in early B cell development is not universal to all placental species, however. In pigs for example there is limited diversity of IgH

complementarity determining region-3 (CDR3) in early fetal development, but this is not due to lack of N-additions. Rather it is likely due to limited B cell numbers early in ontogeny (Butler et al. 2000). Unlike humans and mice, pigs have restricted recombinatorial diversity, using a limited variety of V, D, and J segments to derive their IgH repertoire. This may have compensated for this limitation through increased CDR3 diversity earlier in development than is found in humans and mice (Butler et al 2000).

In species such as marsupials, where B cell development appears to be initiated postpartum, it is not known whether there are changes in the repertoire that are analogous to the fetal-to-adult transition found in humans and mice. Here we investigate that question and establish the timing of critical steps in B cell development in a model marsupial species.

The gray, short-tailed opossum, *Monodelphis domestica* is one of the better-developed model marsupials (Samollow 2006). They are easily bred in captivity, are not seasonal breeders, and are pouchless providing easy access to newborn opossums while they are attached to the teats. A high quality whole genome sequence is available and the content and organization of their germ-line TCR and Ig genes has been established (Mikkelsen et al., 2007, Parra et al 2008, Wang et al., 2009). The opossum has single IgM, IgG, IgE, and IgA isotypes, along with the Igκ and Igλ L chains (Miller et al 1998, Wang et al 2009; Lucero et al 1998, Aveskogh & Hellman, 1998; Miller et al 1999). The IgH locus contains three VH families that are all closely related within VH clan III (Miller et al 1998; Wang et al 2009). Family VH1 is composed of 24 V gene segments of which 19

are functional while family VH2 is a single, functional gene segment. Family VH3 is also only a single V gene, however it is atypical in that it is germ-line joined to a DH segment (Wang et al 2009). VH3 appears functional in that it contains an open reading frame (ORF) and conserved recombination signal sequence (RSS) and is the only known germ-line joined VH gene found in mammals. However, VH3 was not found in the expressed IgH repertoire in previous studies (Miller et al 1998, Aveskogh et al 1999, Wang et al 2009). In contrast to the IgH chains with limited germ-line VH diversity, the opossum Ig L chains have a diverse V gene germ-line diversity (Baker et al 2005, Wang et al 2009). There are 122 V genes divided into seven families in the Igκ locus and 64 V gene segments divided into four families in the Igλ locus. Utilizing the available genomic information for Ig genes and B cell markers the ontogeny of the Ig repertoire and timing of B cell development was investigated in the opossum.

Materials and Methods

Tissue collection, RNA extraction and Complementary DNA (cDNA) synthesis

All procedures using live animals were approved under institutional protocol 07UNM005. *M. domestica* typically give birth in the evening and, for the purposes of this study, the next morning was counted as postnatal day 1 (P1). Due to their small size, opossums P10 or less in age were either extracted whole or using the abdominal region containing the liver, gut, spleen, and bone marrow. For opossums P10 or older the data presented here were derived using individual tissues. For embryonic tissues, pregnancies were timed from the point of ovulation, which occurs on average five days following the pairing of females with a male (Fadem 1985). Embryos collected for the experiments described here were 10 mm in length consistent with being at day 14 of gestation, within 24 hours of birth.

All tissues were either used immediately or stored in RNAlater (Ambion, Austin, TX) at 4˚C for 24 hours and long term at -80˚C. Total RNA was extracted using Trizol RNA extraction protocol (Invitrogen, Carlsbad, CA). Reverse transcription-polymerase chain reaction (RT-PCR) was performed using the Superscript III First-Strand kit (Invitrogen, Carlsbad, CA).

PCR and Sequencing

PCR amplification was performed using Advantage TM-HF 2 PCR (BD Biosciences, CLONTECH Laboratories, Palo Alto, California) with the following conditions for all

primer combinations: denaturation at 94 ˚C for 1 min, followed by 34 cycles of denaturation at 94 ˚C for 30 sec, annealing at 62 ˚C for 4 min, and a final single extension period of 68 ˚C for 5 min. PCR products are cloned using TOPO TA cloning Kit (Invitrogen, Carsbad, CA) and sequenced using BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). All sequences reported are based on sequencing both strands of each clone. Chromatograms are analyzed using Sequencher 4.9 (Gene Codes, Ann Arbor, MI).

Ig transcript collection and junctional diversity analyses

To amplify Ig transcripts, forward primers were designed individually complementary to each V gene family. Forward primers that complementary to framework region (FR) -1 in the VH1, VH2 and VH3 families are: 5'-CCTGCAAAGCTTCTGGATTC, 5'-CATGCATTGGATACGACAGG and 5'-GGACATCTCTGCACCTCTCC, respectively. Forward primers complementary to FR1 of Vκ1, Vκ2, Vκ3, Vκ4, Vκ5, Vκ6 and Vκ7 families are: 5'-TCCCTGGCTGTGTCTC(T,C)AGG, 5'-CCAGCCTCTGTGTCTGTGTC, 5'-TGTGATGACCCAGACTCCAG, 5'-TCCAGCCTCTTTGTCCAGAT, 5'-TCCATCCTCTCTGTCTGCAA, 5'-AATCTCCTGCCTCCCTGTCT and 5'-CAGCCTCAGTGTCTGTGAGC, respectively. Forward primers for FR1 of Vλ1, Vλ2, Vλ3 and Vλ4 families are: 5'-GGTGACTCAGCCTCCCTCT, 5'-TGTCCATGTCTCTGGGAGAA, 5'-GATTCCCTCCATGTCTGTGG and 5'-CCTCCTTGGGAACCACAGTA, respectively. Reverse primers complementary to first exons of Cμ, Cγ, Cα and Cε are: 5'-CAGCACTTTGGTTTTGGTAGG, 5'-TTGCAGGTATATGACTGAGAGGAC, 5'-

TCACCAGTTCTAGAGTCACAGAGG and 5'-CAGATGTGGGATCATAAGTAGCTG, respectively. The reverse primers that complementary to Cκ and Cλ are: 5'-TGGTTGGAAGATGAAGGCAG and 5'-ACCATAGGCCATGACCATGG, respectively.

The germ-line opossum VH, DH, JH, Vκ, Jκ, Vλ and Jλ gene sequences are in the Somatic Diversification Analyses (SoDA) database (https://dulci.org/soda/) (Volpe et al., 2006). IgH and L transcripts amplified and sequenced were analyzed using the SoDA website to determine which germ-line gene segments were contained in each clone and to analyze their junctional diversity and identify P and N nucleotides.

Characterization of the opossum* VpreB *gene

To identify opossum *VpreB*, mouse *VpreB1, VpreB2,* and *VpreB3* sequences (Genbank Accession Numbers: NM_016982, BC141459, and NM_009514) were used to perform an *in silico* search of the whole genome version MonDom5 (GenBank accession number AAFR03000000) with the aid of the BLAST algorithm (Altschul et al 1990; Mikkelsen et al 2007). A 5' partial gene sequence matching VpreB3 was found amongst opossum unassembled sequences, Un.55000001-60000000. To complete the partial *VpreB* gene sequence primers were designed to flank a 329 bp gap in the whole genome sequence at the 5' end of the gene (5'-AGGAGGGCCTTCTCAGGA and 5'-GCTCCTGCTCCTCTTCATTG) and a fragment that successfully covered the gap was cloned and sequenced. To investigate transcription of the opossum *VpreB3* gene, primers were designed for exons 1 and 2 (5'-AGGAGGGCCTTCTCAGGA and 5-GCTCCTGCTCCTCTTCATTG, respectively) and used in RT-PCR.

VpreB phylogenetic analyses were performed on nucleotide alignments using the neighbor-joining method in MEGA4 (Tamura et al 2007). Amino acid translations were first aligned to establish gap position and then converted back to nucleotide using the BioEdit program (Hall 1994). The GenBank accession numbers of the sequences used in the phylogenetic analysis were: opossum VpreB3: JN863116; human VpreB1, CR456609; human VpreB3, NM_013378; mouse VpreB1, NM_016982; mouse VpreB2, BC141459; mouse VpreB3, NM_009514; chicken VpreB3, XM_415223; rat VpreB1, NM_001108845; rat VpreB2, NM_001134788; rabbit VpreB1, AY351269; rabbit VpreB2, AY351268; rabbit VpreB3, XM_002724010. Opossum Vκ and Vλ sequences used in the alignment are from opossum assembly version 5.1 (Mikkelsen et al. 2007; Wang et al., 2009). Human Vλ sequences used in the analyses were obtained from the VBASE database. Mouse Vλ were as follows: Vλ1, X82687; VλX, D38129; Vλ, M34598. Possum Vλ126, AY074464; Possum Vλ12, AY074448; rabbit Vλ2, M27840; rabbit Vλ3, M27841; Chicken Vλ, AB061561; Shark Vλ, X15316.

Results

Initiation of H chain VDJ rearrangement in the opossum

Previous analyses revealed that productively rearranged and expressed αβ TCR could be detected using RT-PCR within the first 24 postnatal hours in opossums (Parra et al 2009). This time window was chosen to begin investigating the initiation of B cell development by attempting to amplify full or partially rearranged IgH transcripts. RT-PCR was performed using primers specific for each of the three opossum VH families paired with a primer for the first exon of each of the opossum H chain C regions. Nine P1 opossums from six different litters were used. Only PCR using primers specific for IgM transcripts utilizing VH1 family members could be successfully amplified from P1 animals (Fig. 1A). Using the VH1 primers two types of transcripts could be amplified from P1 animals; one ranged from 505 to 538 bp and contained transcribed but un-rearranged, germ-line VH genes, the other ranged from 432 to 447 bp in length and contained productive VDJ recombinants (Fig. 1A). When embryonic tissue at gestational day 14 (E14) was used neither transcript were detected (not shown).

The germ-line transcripts detected contained un-recombined VH genes with their RSS intact that, using a cryptic mRNA splice site downstream of the RSS, were spliced to the start of the first C exon of IgM (Fig. 1B). These sterile, germ-line VH transcripts were detected in all nine P1 animals and some P2 and P3 animals and most likely represent the sterile VH transcripts generated during the initiation of V to DJ recombination in pro-B cells (Yancopoulos & Alt, 1985) (Fig. 1A and results not shown). Of 28 independent germ-line transcripts characterized, 26 used the VH1.1 gene segment, which is the most

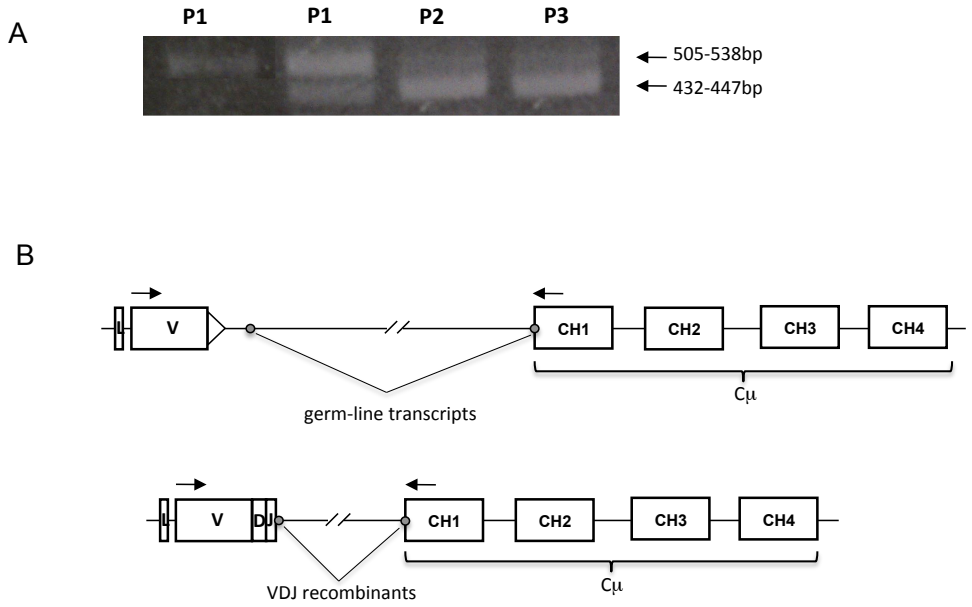


Figure 1. Ig heavy chain gene rearrangement during postnatal opossum development. *A*, Gel images of germ-line, sterile IgM transcripts (505-538 bp range) IgH and VDJ recombinants (432-447 bp range) from representative P1, P2, and P3 individuals. *B*, Diagram showing different mRNA splice variants from the germ-line, sterile IgM transcripts (top) and functional VDJ recombinants (bottom). The locations of the primers used for RT-PCR in *A* are indicated by arrows.

DH proximal of the VH genes (Wang et al 2009). The remaining two used VH1.3 and 1.4, respectively, which are the next two most DH proximal, functional VH genes (not shown) (Wang et al 2009). The majority (26 out of 28) germ-line transcripts were spliced to the CH1 exon of the functional IgM locus. The remaining were spliced to an IgM pseudogene found upstream of the functional copy in the opossum (not shown) (Wang et al 2009).

Transcripts from productively rearranged H chain VDJ genes were detected in only five of the nine P1 animals (Figs. 1A and 2). All were using VH1 family members recombined to different DH and JH genes and spliced to the functional IgM C region. B cells using VH2.1 were not detected until P4 at the earliest and those using VH3.1 were not detected at all in the first six weeks (Fig. 3A). Consistent with VDJ recombination taking place within the first 24 hours postpartum in both B and T cells, Rag-1 transcription could also be detected at this time-point (Fig. 3A) (Parra et al 2009). IgD co-expression was not investigated as opossums lack an IgD (Wang et al 2009). B cells that had switched to IgG, IgE or IgA were not detected in the spleen until much later ages, weeks five, six, and eight, respectively (Fig. 3B). When a mucosal site, the gut, was also tested IgA transcripts were not detected until week eight, as in the spleen (not shown).

	Animal	Clone #	VDJ	VH 3' end	P	N	P	DH	P	N	P	JH 5' end
P1	1	1/1/01	VH1.6_DH6_JH1	CAAGAGA	TC	GG		TACTACAGT		C		TTGATTTCTGGGGAAGGG
	2	1/2/01	VH1.6_DH3_JH2	CAAGAG				CGGGTATAGC				TGATAACTGGGGCAAGGGGA
	2	1/2/02	VH1.7_DH3_JH1	CAAGAGA				ACTACGGGTATAGCT				ACAACTGGATTGATATCTGGGGAACGG
	3	1/3/01	VH1.16_DH2_JH2	CAAGAGA				TTATAGCTG			GT	ACTACTTTGATAACTGGGGCAAGGGGA
	4	1/4/01	VH1.6_DH6_JH2	CAAGAGA				ACTACAGTA				ACTACTTTGATAACTGGGGCAAGGGGA
	5	1/5/01	VH1.9_DH1_JH2	CAGGAGA		G	C	GATACAGTAACT				ACTACTTTGATAACTGGGGCAAGGGGA
P2	1	2/1/01	VH1.6_DH6_JH2	CAAGAGA				ACTACAGT				ACTACTTTGATAACTGGGGCAAGGGG
	1	2/1/03	VH1.23_DH6_JH2	CAAGACA				CTACTACAGTGGTA				ACTTTGATAACTGGGGCAAGGGG
	1	2/1/05	VH1.8_DH3_JH2	CA		AG		AACTACGGGTATAG		C		CTATTTTGAATTCTGGGGCAAG
	1	2/1/06	VH1.21_DH1_JH2	CAAGAGA	TC			GATACAGTAACT				ACTGGGGCAAGGGG
	2	2/2/01	VH1.6_DH3_JH2	CAAGAG				CGGGTATAGC				TGATAACTGGGGCAAGGGG
	3	2/3/01	VH1.15_DH2_JH2	CAGGAGAT	CT			AGCTGGG				TACTTTGATAACTGGGGCAAGGGG
P3	1	3/1/01	VH1.15_DH2_JH2	CAGGAGAT	CT			AGCTGGG				TACTTTGATAACTGGGGCAAGGGG
	1	3/1/02	VH1.6_DH6_JH2	CAAGAGA				ACTACAGT				ACTACTTTGATAACTGGGGCAAGGGG
	1	3/1/03	VH1.15_DH1_JH2	CAA				GATACAGTA				ACTACTTTGATAACTGGGGC
	1	3/1/04	VH1.7_DH3_JH1	CAAGAG				ACTACGGGTATAGCT				ACAACTGGTTTGATTTCTGG
	1	3/1/05	VH1.6_DH3_JH2	CAAGAG				CGGGTATAGC				TGATAACTGGGGCAAGGGG
	1	3/1/06	VH1.6_DH1_JH2	CAAGAGA				TACAGTAAC				TGATAACTGGGGCAAGGGG
	1	3/1/07	VH1.6_DH6_JH2	CAAG				ACTACAGTGG				TACTTTGATAACTGGGGCAAGGGG
	2	3/2/01	VH1.6_DH2_JH2	CAAGA				CTACAGTGG				TACTTTGATAACTGGGGCAAGGGG
	3	3/3/01	VH1.4_DH3_JH2	CAGGAGA	TC			GGGTAT				ACTACTTTGATAACTGGGGCAAGGGG

Figure 2. Representative IgH VDJ junctions showing only the end of the VH through the start of the JH for P1 through P3 animals. The age of the animals is indicated on the left and individual opossums in each age group are numbered in the first column, followed by the clone number in the second column. Nucleotides identified as P, N or DH, as determined using the SoDA website are indicated at the top. The specific V, D, and J germ-line gene segments used in each clone are indicated in the third column. Micro-homology at the VD and DJ junctions is underlined.

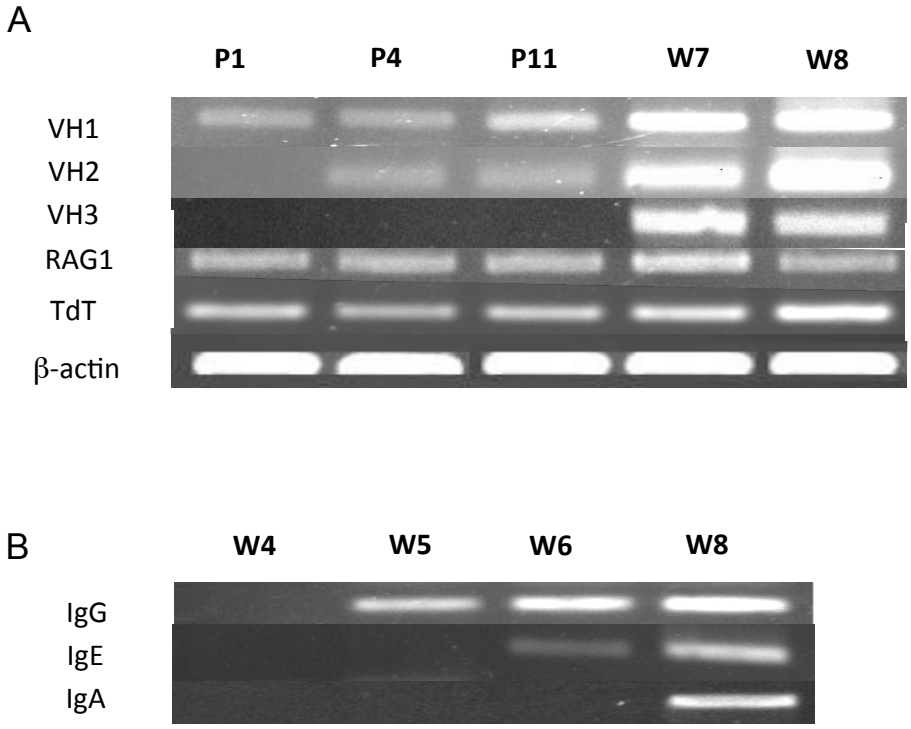


Figure 3. *A*, The expression of opossum Ig heavy chains using the three different VH families along with RAG1 and TdT detected by RT-PCR at the ages indicated. *B*, The expression of different IgH isotypes using VH1 family specific primers detected by RT-PCR in postnatal opossum spleen.

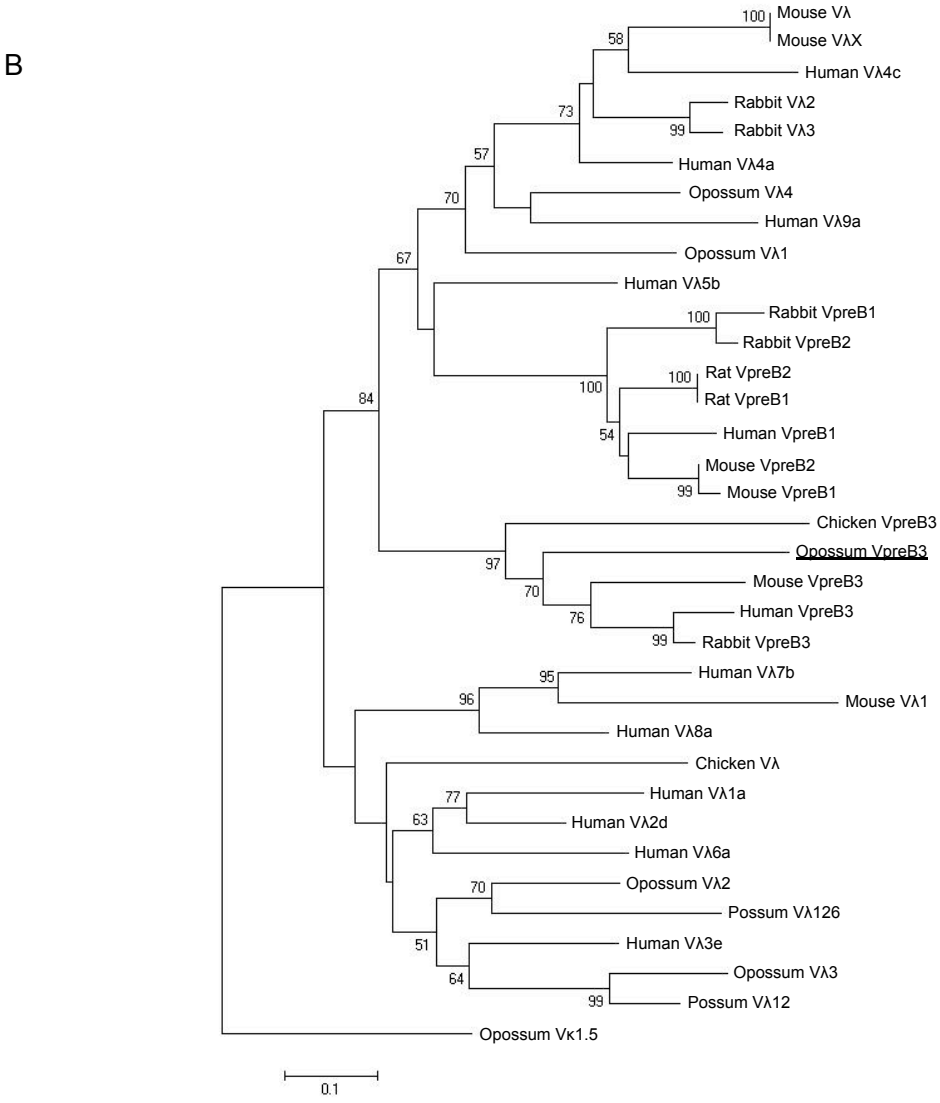
Identification, evolution and expression of the opossum* VpreB3 *gene

Coincident with successful IgH gene recombination, and preceding L chain gene rearrangement during B cell development is the expression of the surrogate L chains, VpreB and λ5, necessary for assembling the pre-BCR (Kudo & Melchers 1987). Marsupial VpreB and λ5 have not been described previously, and attempts to identify an opossum homologue of *VpreB1, VpreB2* and *λ5* by *in silico* screening of the *M. domestica* genome were unsuccessful (not shown). *In silico* screening using mouse *VpreB3,* however, matched a partial gene sequence in the opossum whole genome that appeared to contain a 329 bp gap at the 5' end of the gene. The complete opossum *VpreB3* gene sequence was obtained by filling the gap in the genomic sequence using PCR (Fig. 4A).

A

```
        cagccactctgcctccctgacagccatggccccccacctgctggttctgctcttgattcctgtcctcctggcacgtgagtga
                                    M  A  P  H  L  L  V  L  L  L  I  P  V  L  L  A
tcctccatctctttggcgagccaggcgggaggggcaccctgggggccacccgctggcagtggccctttctcctccccaccagcgtgggc
tgtctccgaggcttgggcagagacagtagccaggagggccaccaggtgacctggaagagggatggggagggggaggagggccttctcagg
aggctgagaagaggctccacaccagcgggagtgaccggaggaatgccaaggggcccccatcagggaggaggggctgcccacagaggaatgg
gctggctcgggaagcctggtactgagctcctcatccctggcagtgttcaagcaacgcacagaggcccggtgcctgcttgttggtgtgtg
cttggactcagtggccactgagatccccagctgtggcatgggtaccttgatacaaggacgatggcatcaggctggtgggtgggtaggc
ctttgggcccttgggcgggctgttgggctggtgggggggccagcgggccttgggcccttgggccggtgggcgggtaagtcggtgggcct
ttgggcaggctggtgggggggccgttggaccgttgggcagccgttggcccaggagctccgggaagcgggggggggggaggcccctgtggg
cactcaagcttccccacacctgggctccccagaggggccaggaaggaggtgctgccagcctgagggcattggggtgtgtctcccccag
agtcccagcgggcacccagccagcccgaagctctggtggtgtttccgggccagacggccaggctgctgtgctccctgcagccagaagtg
   S  Q  R  A  P  S  Q  P  E  A  L  V  V  F  P  G  Q  T  A  R  L  L  C  S  L  Q  P  E  V
gccatctccgagcggggcatctcgtggttccagcagttccccggcagcgccccccgcttcctgctgtactactacaatgaagaggagca
 A  I  S  E  R  G  I  S  W  F  Q  Q  F  P  G  S  A  P  R  F  L  L  Y  Y  Y  N  E  E  E  Q
ggagcgccggccggggctcccagagcgctttggggcctccaaggatgccacccacaatgcctgcatcctgaccatctccccggtgcagc
  E  R  R  P  G  L  P  E  R  F  G  A  S  K  D  A  T  H  N  A  C  I  L  T  I  S  P  V  Q
cagaggacgaagctgactactactgctccttggcctactagggccctcctctcttctttcccagcttcccccctccctttcccagccct
P  E  D  E  A  D  Y  Y  C  S  L  A  Y  *
```

B

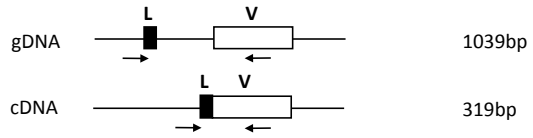


C

```
                <     leader        ><         FR1        ><    CDR1     ><      FR2     >
ModoVpreB3  MAPH-LLVLLLIPVLLAQS--QRAPSQPEALVVFPGQTARLLCSLQP-EVAISERGISWFQQFPGSAPRFLLYY 70
HosaVpreB3  MACR-CLSFLLMGTFLSVS--QTVLAQLDALLVFPGQVAQLSCTLSPQHVTIRDYGVSWYQQRAGSAPRYLLYY 71
OrcuVpreB3  MACR-YLALFLTGAFLAVS--QPVLTQPDALLVFPGQVAQLSCTLSPQHASIWDYGVSWYQQRAGSAPRYLLYY 71
MumuVpreB3  MACPGCLPLLLIGTFVAVF--QPTLTQPDAFSVFPGQDAHLSCTINSQHATAGDIGVSWYQQQPGSAP-HLLYY 72
GagaVpreB3  MVLG-FMLLLLAGTAGSACRAQPVLTQPAAVQVLPGETARLSCVLSP-QYNISDFGITWYQQRPGQALRYLLYY 72
MumuVpreB2  MAWTSVLLMLLAHLTGCGP--QPMVHQPPSASSSLGATIRLSCTLSN-DHNIGIYSIYWYQQRPGHPPRFLLRY 71
MumuVpreB1  MAWTSVLLMLLAYLTGCGP--QPMVHQPPLASSSLGATIRLSCTLSN-DHNIGIYSIYWYQQRPGHPPRFLLRY 71

                <  CDR2  ><               FR3              ><           CDR3         >
ModoVpreB3  YNEEEQERRPGLPERFGASKDATHNACILTISPVQPEDEADYYCSLAY---------------------- 118
HosaVpreB3  RSEEDHHRPADIPDRFSAAKDEAHNACVLTISPVQPEDDADYYCSVGYGFSP------------------- 123
OrcuVpreB3  RSEEDHHRPEDVPDRFSASKDAAHNACILTINPVQPEDDADYYCSVGYTSDL------------------- 123
MumuVpreB3  YAEEEHYRPADIPDRFSATVDAAHNACILTISPVLPEDDADYFCSIAHTFEP------------------- 123
GagaVpreB3  NTERDKHKSARIPDRFSATKDLVHNACILIIAVAQEEDNGRYFCSLPLTINWL----------------- 125
MumuVpreB2  FSHSDKHQGPDIPPRFSGSKDTARNLGYLSISELQPEDEAVYYCAVGLRSHEKKRMEREWEGEKSYTDLGS 142
MumuVpreB1  FSHSDKHQGPDIPPRFSGSKDTTRNLGYLSISELQPEDEAVYYCAVGLRSQEKKRMEREWEGEKSYTDLGS 142
```

D



E

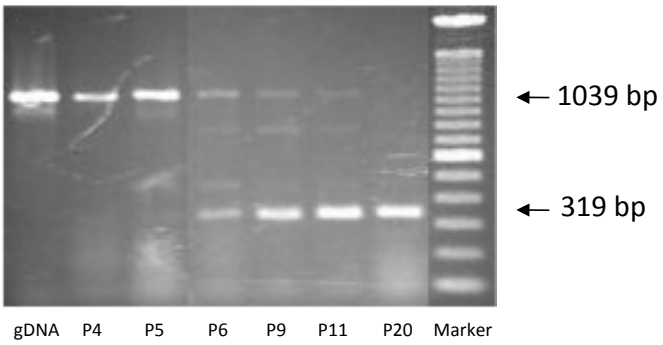


Figure 4. Characterization of the opossum VpreB3 gene. *A*, Nucleotide sequence of opossum VpreB3 genomic sequence with deduced amino acids below. mRNA splice sites are underlined. The stop codon TAG is marked with asterisk. *B*, Phylogenetic tree based on nucleotide alignments of VpreB1, 2, and 3 along with Vl sequences from human, mouse, rabbit, chicken, and opossum (underlined). The opossum Vk1.5 sequence was used as out-group. *C*, Alignment of deduced amino acid sequence of opossum VpreB3 with the sequences of mouse VpreB1, 2, and 3, human VpreB1 and 2, rabbit VpreB3, and chicken VpreB3. Leader peptide, and regions corresponding to the FRs and CDRs are indicated above the alignment. Conserved cysteines are shaded gray, the conserved HXAC motif is highlighted in black. *D*, Diagram showing VpreB3 gene structure for the genomic (top) and cDNA (bottom) organization. Leader region and V domain are represented by solid and open boxes, respectively. Location of the primers used for RT-PCR, complementary to leader exon and V exon are indicated as arrows. *E*, Gel image of VpreB3 PCR products at different ages indicated. The cDNA samples contain contaminating genomic DNA (gDNA) that serves as an internal positive control. Actin controls were included as positive controls for the cDNA (not shown). In P4 and P5 only the 1039 bp genomic fragment amplified. The 319 bp product from the cDNA begins to appear and increases with age at the expense of amplifying the gDNA target.

Known *VpreB* genes consist of two exons encoding the leader and extracellular V domain, respectively. Based on sequence similarity and predicted mRNA splice sites, the presumptive opossum *VpreB* exons were identified (Fig. 4A). These were used to design PCR primers that amplified a cDNA clone from splenic mRNA from an eight-week-old animal. When compared to the genomic sequence, the cDNA confirmed the opossum *VpreB* gene structure (Fig. 4A). When compared with VpreB genes from placental mammals and chickens, the opossum gene, not surprisingly, clustered with other *VpreB3* genes in a phylogenetic analysis (Fig. 4B). When the translated sequence was aligned to VpreB from other species, a conserved HXAC motif was present in all VpreB3, including opossum, but absent from VpreB1 and 2 (Fig. 4C) (Rosnet et al., 2004). These results were consistent with the opossum having a *VpreB3* homologue and this gene was used to investigate the timing of expression during B cell development. Using primers that flank the intron in opossum *VpreB3* a product of 319 bp was amplified by RT-PCR from mRNA from P6 animals as the earliest time-point (Figs. 4D, 4E and data not shown).

Initiation of L chain gene rearrangement in the developing opossum

Marsupial Ig L chain V gene segments are more abundant and diverse than are the H chain V genes (Lucero et al 1998, Miller et al 1999, Baker et al 2005, Wang et al 2009). To amplify Igλ transcripts, primers specific for FR1 of each Vλ family members were used with a primer that could pair with all eight known Cλ genes. The same approach was applied to Igκ transcripts where there is a single known Cκ. The earliest time-points where Igλ and Igκ transcripts containing VJ rearrangements could be detected were in P7 and P8 animals, respectively (Fig. 5).

In summary to this point, it appeared that B cell development, as it relates to Ig gene rearrangement and transcription was absent in the last 24 hrs of gestation but is initiated within the first 24 hrs postpartum. However B cell development is delayed with VpreB3 gene expression and Ig L chain rearrangements first being detected six and seven days after the initiation of H chain rearrangement, respectively (Fig. 6).

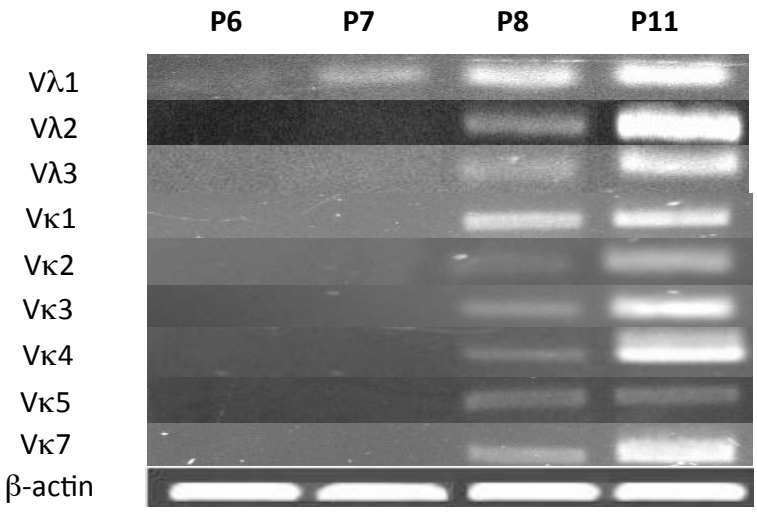


Figure 5. Ig light chain gene rearrangement during postnatal opossum ontogeny. The expression of opossum Ig light chain was determined by RT-PCR at the ages indicated using primers specific for each V family and the relevant C gene. β-actin was used as a positive control.

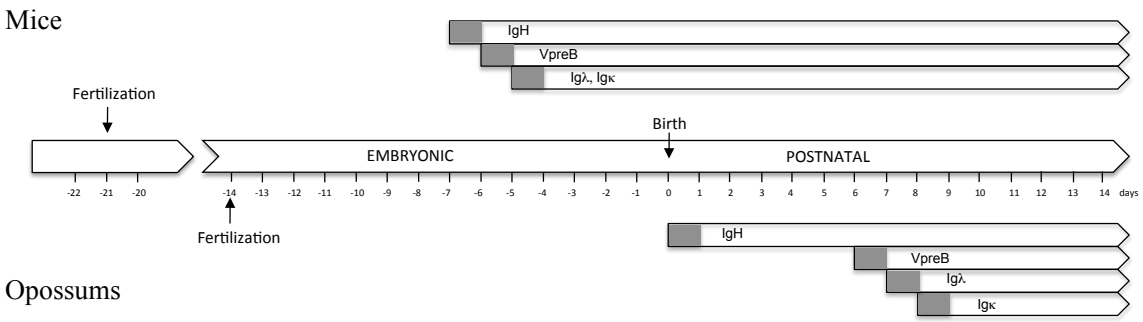


Figure 6. Diagram comparing key stages of B cell development in mice and opossums relative to birth. The periods of gene expression are indicated by the horizontal bars. The gray zone indicates a 24 hour period over which expression is initiated.

IgH repertoire development in the opossum

The diversity of expressed H chain rearrangements during opossum ontogeny was also investigated. Transcribed VDJ rearrangements were amplified by RT-PCR from animals at different ages and products were cloned and sequenced. For each P1 animal, eight to twelve clones from at least two independent PCR reactions were characterized. Productive H chain VDJ rearrangements were successfully amplified from only five of the nine P1 animals (Fig. 2). In three P1 animals only a single, productive rearrangement was detected; the sequences were different and unique to each individual. From the remaining two animals, one yielded two different productive rearrangements, and the other a single productive and a non-productive rearrangement (P1 animals no. 2 and 5 in Fig. 2, and data not shown). In the case of the non-productive rearrangement, it is missing most of CDR3 and FR4 having rearranged the VH1.23 to approximately the last third of JH1 (not shown). Whether this was due to excessive trimming during VDJ recombination or direct V to J recombination is not obvious and there is no evidence of a cryptic RSS heptamer in JH1 (not shown). Of the six different productive VDJ rearrangements, three used the VH1.6 gene. The remaining three used VH1.7, VH1.9, and VH1.16 gene segments, respectively (Fig. 2). Four different DH segments were used in these rearrangements (DH1, 2, 3, and 6) with two (DH3 and DH6) each being used twice. There are only two JH used in the opossum; two of the six P1 clones used JH1 and the four remaining used JH2. Animals at ages P2 and P3 were also examined and found to be similar to P1 individuals. Four out of six of the P2 and P3 age animals, like P1 animals, yielded only a single VDJ recombinant. The other two animals, one P2 and one

P3 in age, yielded much greater diversity of sequences with four unique VDJ from the P2 and seven unique VDJ from the P3 (Fig. 2).

Of the 21 unique VDJ rearrangements isolated from P1, P2, and P3 animals, only three contained N-additions (Fig. 2). Previous analyses of fetal mouse IgH rearrangements revealed the absence of N-additions was associated with the use of micro-homology to direct VDJ recombination (Feeney 1990). Evidence of micro-homology, however, was observed in less than half (eight of 21) of the VDJ recombinants isolated from P1, P2 and P3 animals (Fig. 2). Although N-additions were rare in the first three postnatal days, they were not completely absent and transcription of TdT in P1 and older tissues was detectable by RT-PCR (Fig. 3A).

In the first three postnatal days, nine different VH genes were used, however there was some bias in that nearly half (10 of 21) used VH1.6. This bias, however, did not correlate with evidence of micro-homology at the V(D)J junctions (Fig. 2). There was also no apparent positional bias for the VH being used in the early developing repertoire as the nine VH used were scattered throughout the locus (Wang et al 2009). Furthermore, the RSS are nearly identical between VH1 V genes and were also not likely to influence recombination frequency (Fig. 7).

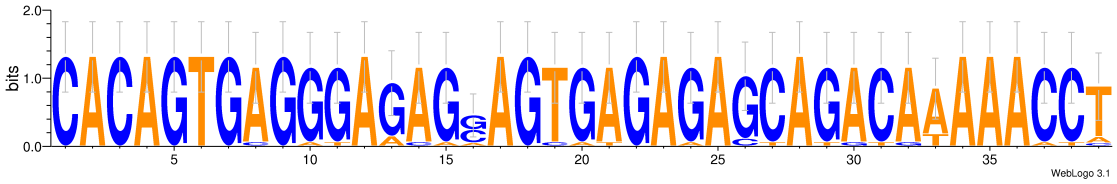
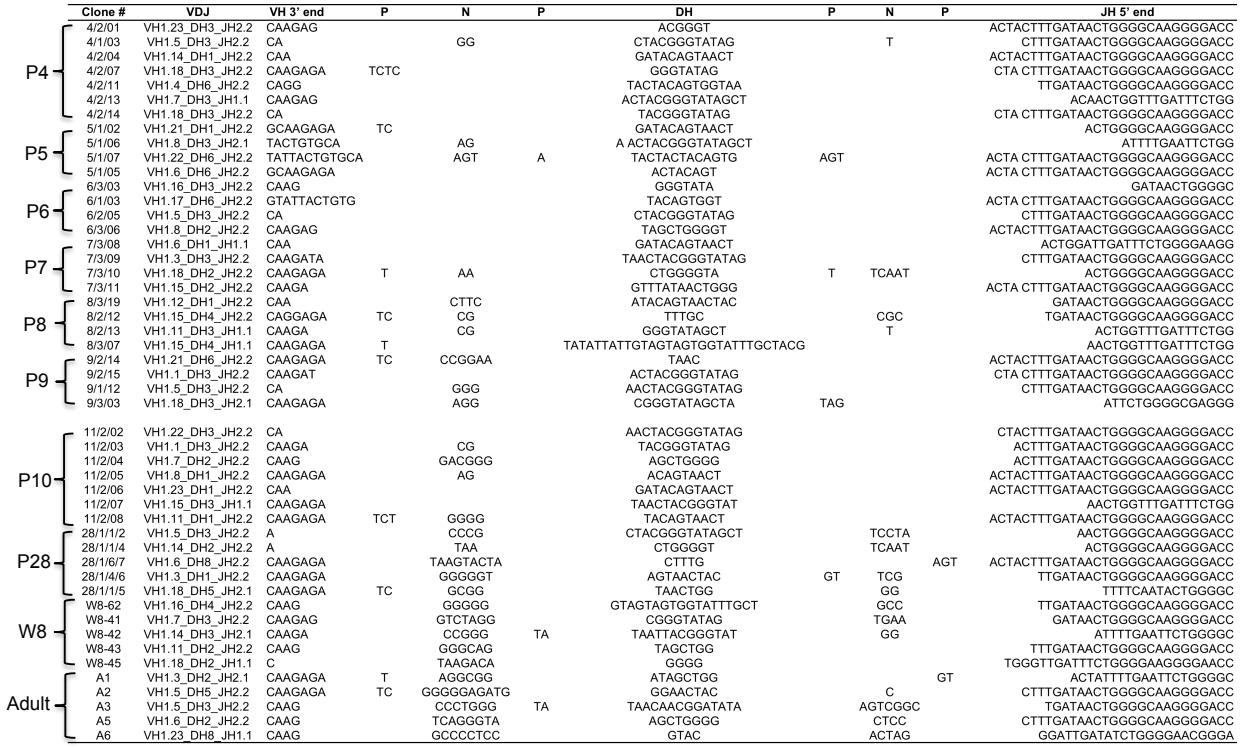


Figure 7. Logo analysis of the combined RSS flanking the germ-line VH1 gene segments in the opossum genome demonstrating little variation in the heptamer, spacer, or nonamer regions. The opossum VH RSS were compared using a WebLogo application available at http://weblogo.berkeley.edu/.

There are 19 functional VH1 family members in the opossum *IgH* locus (Wang et al 2009). By P8, 18 of them had been found to be used in VDJ rearrangements. The exception is VH1.13 that, although appearing functional in the germ-line, was never used at any age (Figs. 2 and 8). The frequency of clones containing N-additions remained a somewhat similar 21 to 37% from P4 through P8. By P11 the frequency of CDR3 containing N-additions jumped to 45%. By week 8 the frequency was 100% where it remained in adults. The frequency of clones containing N-additions paralleled the number of N-nucleotides added to the junctions to some degree. For the first six days the average number of N-additions remained less than one. In P7 to P11 animals the length of the CDR3 contributed by N-additions increased to an average greater than two. By P28 the average number of N-additions continued to increase to 4.1. By week 8 and beyond the average number of N-additions is substantially larger at eight to eleven total nucleotides added per recombinant. Similarly the CDR3 length remained a fairly similar 29 to 33 nucleotides from P1 through P11 and then there is an increase after P11 to greater than 35 that are associated with increased N-additions (Fig. 9 & Table I).



	Clone #	VDJ	VH 3' end	P	N	P	DH	P	N	P	JH 5' end
P4	4/2/01	VH1.23_DH3_JH2.2	CAAGAG				ACGGGT				ACTACTTTGATAACTGGGGCAAGGGGACC
	4/1/03	VH1.5_DH3_JH2.2	CA		GG		CTACGGGTATAG		T		CTTTGATAACTGGGGCAAGGGGACC
	4/2/04	VH1.14_DH1_JH2.2	CAA				GATACAGTAACT				ACTACTTTGATAACTGGGGCAAGGGGACC
	4/2/07	VH1.18_DH3_JH2.2	CAAGAGA	TCTC			GGGTATAG				CTA CTTTGATAACTGGGGCAAGGGGACC
	4/2/11	VH1.4_DH6_JH2.2	CAGG				TACTACAGTGGTAA				TTGATAACTGGGGCAAGGGGACC
	4/2/13	VH1.7_DH3_JH1.1	CAAGAG				ACTACGGGTATAGCT				ACAACTGGTTTGATTTCTGG
	4/2/14	VH1.18_DH3_JH2.2	CA				TACGGGTATAG				CTA CTTTGATAACTGGGGCAAGGGGACC
P5	5/1/02	VH1.21_DH1_JH2.2	GCAAGAGA	TC			GATACAGTAACT				ACTGGGGCAAGGGGACC
	5/1/06	VH1.8_DH3_JH2.1	TACTGTGCA		AG		A ACTACGGGTATAGCT				ATTTTGAATTCTGG
	5/1/07	VH1.22_DH6_JH2.2	TATTACTGTGCA		AGT	A	TACTACTACAGTG	AGT			ACTA CTTTGATAACTGGGGCAAGGGGACC
	5/1/05	VH1.6_DH6_JH2.2	GCAAGAGA				ACTACAGT				ACTA CTTTGATAACTGGGGCAAGGGGACC
P6	6/3/03	VH1.16_DH3_JH2.2	CAAG				GGGTATA				GATAACTGGGGC
	6/1/03	VH1.17_DH6_JH2.2	GTATTACTGTG				TACAGTGGT				ACTA CTTTGATAACTGGGGCAAGGGGACC
	6/2/05	VH1.5_DH3_JH2.2	CA				CTACGGGTATAG				CTTTGATAACTGGGGCAAGGGGACC
	6/3/06	VH1.8_DH2_JH2.2	CAAGAG				TAGCTGGGGT				ACTACTTTGATAACTGGGGCAAGGGGACC
P7	7/3/08	VH1.6_DH1_JH1.1	CAA				GATACAGTAACT				ACTGGATTGATTTCTGGGGAAGG
	7/3/09	VH1.3_DH3_JH2.2	CAAGATA				TAACTACGGGTATAG				CTTTGATAACTGGGGCAAGGGGACC
	7/3/10	VH1.18_DH2_JH2.2	CAAGAGA	T	AA		CTGGGGTA	T	TCAAT		ACTGGGGCAAGGGGACC
	7/3/11	VH1.15_DH2_JH2.2	CAAGA				GTTTATAACTGGG				ACTA CTTTGATAACTGGGGCAAGGGGACC
P8	8/3/19	VH1.12_DH1_JH2.2	CAA		CTTC		ATACAGTAACTAC				GATAACTGGGGCAAGGGGACC
	8/2/12	VH1.15_DH4_JH2.2	CAGGAGA	TC	CG		TTTGC		CGC		TGATAACTGGGGCAAGGGGACC
	8/2/13	VH1.11_DH3_JH1.1	CAAGA		CG		GGGTATAGCT		T		ACTGGTTTGATTTCTGG
	8/3/07	VH1.15_DH4_JH1.1	CAAGAGA	T			TATATTATTGTAGTAGTGGTATTTGCTACG				AACTGGTTTGATTTCTGG
P9	9/2/14	VH1.21_DH6_JH2.2	CAAGAGA	TC	CCGGAA		TAAC				ACTACTTTGATAACTGGGGCAAGGGGACC
	9/2/15	VH1.1_DH3_JH2.2	CAAGAT				ACTACGGGTATAG				CTA CTTTGATAACTGGGGCAAGGGGACC
	9/1/12	VH1.5_DH3_JH2.2	CA		GGG		AACTACGGGTATAG				CTTTGATAACTGGGGCAAGGGGACC
	9/3/03	VH1.18_DH3_JH2.1	CAAGAGA		AGG		CGGGTATAGCTA	TAG			ATTCTGGGGCGAGGG
P10	11/2/02	VH1.22_DH3_JH2.2	CA				AACTACGGGTATAG				CTACTTTGATAACTGGGGCAAGGGGACC
	11/2/03	VH1.1_DH3_JH2.2	CAAGA		CG		TACGGGTATAG				ACTTTGATAACTGGGGCAAGGGGACC
	11/2/04	VH1.7_DH2_JH2.2	CAAG		GACGGG		AGCTGGGG				ACTTTGATAACTGGGGCAAGGGGACC
	11/2/05	VH1.8_DH1_JH2.2	CAAGAGA		AG		ACAGTAACT				ACTACTTTGATAACTGGGGCAAGGGGACC
	11/2/06	VH1.23_DH1_JH2.2	CAA				GATACAGTAACT				ACTACTTTGATAACTGGGGCAAGGGGACC
	11/2/07	VH1.15_DH3_JH1.1	CAAGAGA				TAACTACGGGTAT				AACTGGTTTGATTTCTGG
	11/2/08	VH1.11_DH1_JH2.2	CAAGAGA	TCT	GGGG		TACAGTAACT				ACTACTTTGATAACTGGGGCAAGGGGACC
P28	28/1/1/2	VH1.5_DH3_JH2.2	A		CCCG		CTACGGGTATAGCT		TCCTA		AACTGGGGCAAGGGGACC
	28/1/1/4	VH1.14_DH2_JH2.2	A		TAA		CTGGGGT		TCAAT		ACTGGGGCAAGGGGACC
	28/1/6/7	VH1.6_DH8_JH2.2	CAAGAGA		TAAGTACTA		CTTTG			AGT	ACTACTTTGATAACTGGGGCAAGGGGACC
	28/1/4/6	VH1.3_DH1_JH2.2	CAAGAGA		GGGGGT		AGTAACTAC	GT	TCG		TTGATAACTGGGGCAAGGGGACC
	28/1/1/5	VH1.18_DH5_JH2.1	CAAGAGA	TC	GCGG		TAACTGG		GG		TTTTCAATACTGGGGC
W8	W8-62	VH1.16_DH4_JH2.2	CAAG		GGGGG		GTAGTAGTGGTATTTGCT		GCC		TTGATAACTGGGGCAAGGGGACC
	W8-41	VH1.7_DH3_JH2.2	CAAGAG		GTCTAGG		CGGGTATAG		TGAA		GATAACTGGGGCAAGGGGACC
	W8-42	VH1.14_DH3_JH2.1	CAAGA		CCGGG	TA	TAATTACGGGTAT		GG		ATTTTGAATTCTGGGGC
	W8-43	VH1.11_DH2_JH2.2	CAAG		GGGCAG		TAGCTGG				TTTGATAACTGGGGCAAGGGGACC
	W8-45	VH1.18_DH2_JH1.1	C		TAAGACA		GGGG				TGGGTTGATTTCTGGGGAAGGGGAACC
Adult	A1	VH1.3_DH2_JH2.1	CAAGAGA	T	AGGCGG		ATAGCTGG			GT	ACTATTTTGAATTCTGGGGC
	A2	VH1.5_DH5_JH2.2	CAAGAGA	TC	GGGGGAGATG		GGAACTAC		C		CTTTGATAACTGGGGCAAGGGGACC
	A3	VH1.5_DH3_JH2.2	CAAG		CCCTGGG	TA	TAACAACGGATATA		AGTCGGC		TGATAACTGGGGCAAGGGGACC
	A5	VH1.6_DH2_JH2.2	CAAG		TCAGGGTA		AGCTGGGG		CTCC		CTTTGATAACTGGGGCAAGGGGACC
	A6	VH1.23_DH8_JH1.1	CAAG		GCCCCTCC		GTAC		ACTAG		GGATTGATATCTGGGGAACGGGA

Figure 8. Representative IgH VDJ junctions showing only the end of the VH through the start of the JH for P4 through adult animals. The age of the animals is indicated on the left and individual opossums in each age group are numbered in the first column, followed by the clone number in the second column. Nucleotides identified as P, N or DH, as determined using the SoDA website are indicated at the top. The specific V, D, and J germ-line gene segments used in each clone are indicated in the third column. Micro-homology at the VD and DJ junctions is underlined.

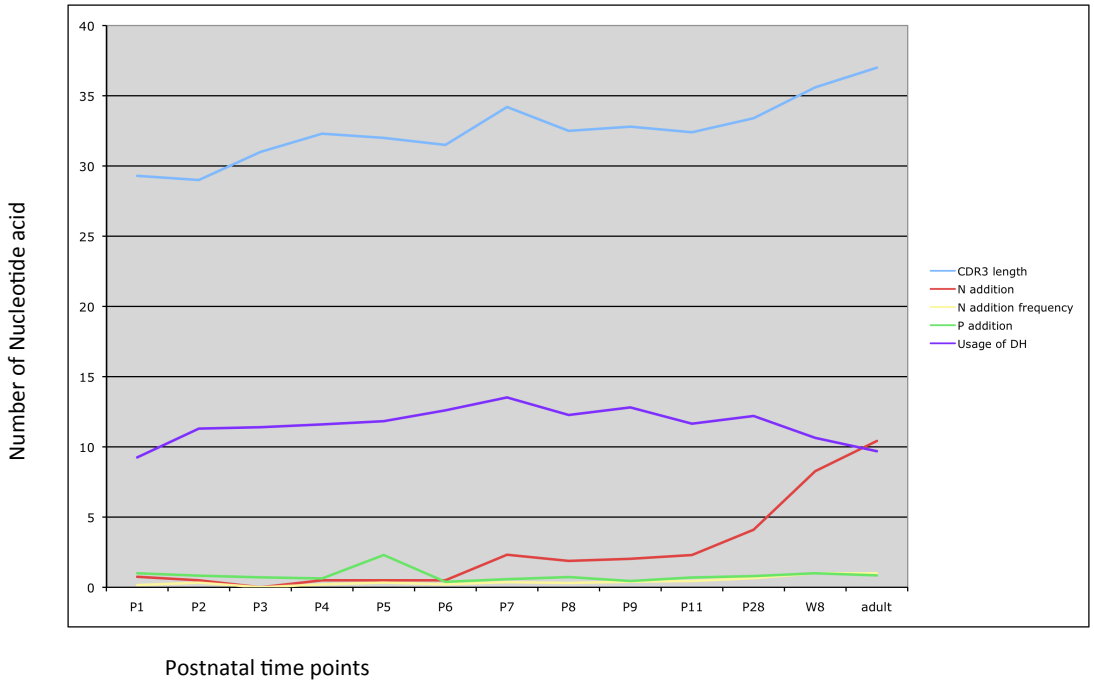


Figure 9. Comparision of CDR3 length, P and N nucleotide addition length and DH length in opossums IgH transcripts at different ages.

Table I. CDR3 length, P and N nucleotide addition length, N nucleotide addition frequency and DH length in opossums IgH, Igκ and Igλ transcripts at different ages.

		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P28	W8	Adult
	CDR3 length[1]	229.3	29	31	32.3	32	31.5	34.2	32.5	32.8		32.4	33.4	35.6	37
	N addition[2]	0.75	0.5	0	0.5	0.5	0.5	2.32	1.88	2.03		2.3	4.1	8.27	10.42
	N addition frequency	16%	33%	0	21%	33%	21.40%	36.80%	30.80%	40%		45%	65.20%	100%	100%
VH1	**P addition**[3]	1	0.83	0.71	0.63	2.3	0.39	0.58	0.73	0.45		0.7	0.8	1	0.85
	DH length[4]	9.25	11.3	11.4	11.6	11.83	12.6	13.52	12.27	12.81		11.65	12.2	10.64	9.69
	CDR3 length							41	40	40.2		41		40.9	40.3
	N addition							0.16	2	1.8		1		3.2	4.17
Vλ1	**N addition frequency**							16%	21%	20%		33%		41%	32.5%
	P addition							0	0.06	0		0		0	0.09
	Microhomology frequency							33%	40%	50%		30%		0	0
	Clone No.							6	38	4		32		6	32
	CDR3 length								29.3	30		30.2		30.5	30
Vλ2	**N addition**								0.28	0.13		0.27		0	0.4
	N addition frequency								14%	23%		9%		0	20%
	P addition								0.08	0		0.6		0	0
	Microhomology frequency								57%	22%		18%		0	0
	Clone No.								7	9		11		5	6
	CDR3 length								31	30.5		29.8		30.2	30
Vλ3	**N addition**								0.66	0.33		0.33		1.2	0
	N addition frequency								33%	17%		11%		40%	0
	P addition								0.17	0		0		0	0
	Microhomology frequency								50%	33%		33%		0	0
	Clone No.								6	6		9		5	2
	CDR3 length								27		27	28.3		29.2	26.7
Vκ1	**N addition**								0		0	0.12		0.36	0
	N addition frequency								0		0	12%		18%	0
	P addition								0		0	0		0	0
	Microhomology frequency								12%		16%	5%		0	0
	Clone No.								16		6	17		11	11
	CDR3 length								27.3					28.5	27
Vκ2	**N addition**								0					0.13	0.6
	N addition frequency								0					13%	42%
	P addition								0					0.07	0.14
	Microhomology frequency								11%					0	0
	Clone No.								18					15	8
	CDR3 length											29.2		28.3	30
Vκ3	**N addition**											0		0.28	0
	N addition frequency											0		14%	0
	P addition											0		0	0
	Microhomology frequency											18%		0	0
	Clone No.											11		7	2
	CDR3 length								30		27				28.3
Vκ4	**N addition**								0		0				0
	N addition frequency								0		0				0
	P addition								0		0				0
	Microhomology frequency								0		0				0
	Clone No.								8		4				1
	CDR3 length													30	
Vκ5	**N addition**													0	
	N addition frequency													0	
	P addition													0	
	Microhomology frequency													0	
	Clone No.													9	
	CDR3 length								27			28.3		29.2	
Vκ7	**N addition**								0			0.37		0	
	N addition frequency								0			12%		0	
	P addition								0			0.12		0	
	Microhomology frequency								20%			12%		0	
	Clone No.								15			8		2	

[1, 2, 3, 4] Calculated in nucleotides.

The opossum *IgH* locus contains nine DH genes (Wang et al 2009) and all nine are used (Fig. 10). Four different DH gene segments (DH1, 2, 3, and 6), however, account for 85% all the VDJ recombinants isolated. These also happen to be the four shortest germ-line DH genes (Wang et al 2009). All four were found in VDJ recombinants in the P1 repertoire and remained the dominant DH genes at early time-points (Fig. 2). Eight of the nine DH, including the four used most frequently have ORF in two of three frames, but this did not appear to influence their use (not shown). Therefore, there does not appear to be a bias for DH gene segments used, except for a preference for short DH that remains through to adulthood. The contribution that the DH genes made to CDR3 length did not vary as the opossums mature (range of 9 to 14 nucleotides) (Table I). Rather the increase in CDR3 length in older animals was due to increased N-additions as described above. The opossum *IgH* locus contains six JH segments, of which four appear functional based on genomic sequence (Wang et al 2009). However only two, JH1 and JH2, are used and these are the two immediately upstream of the functional IgM C region (Fig. 2 and 8) (Wang et al 2009). They were not used equally, however, as JH2 was found in 85% of all VDJ rearrangements and this bias begins at P1 (Fig. 2).

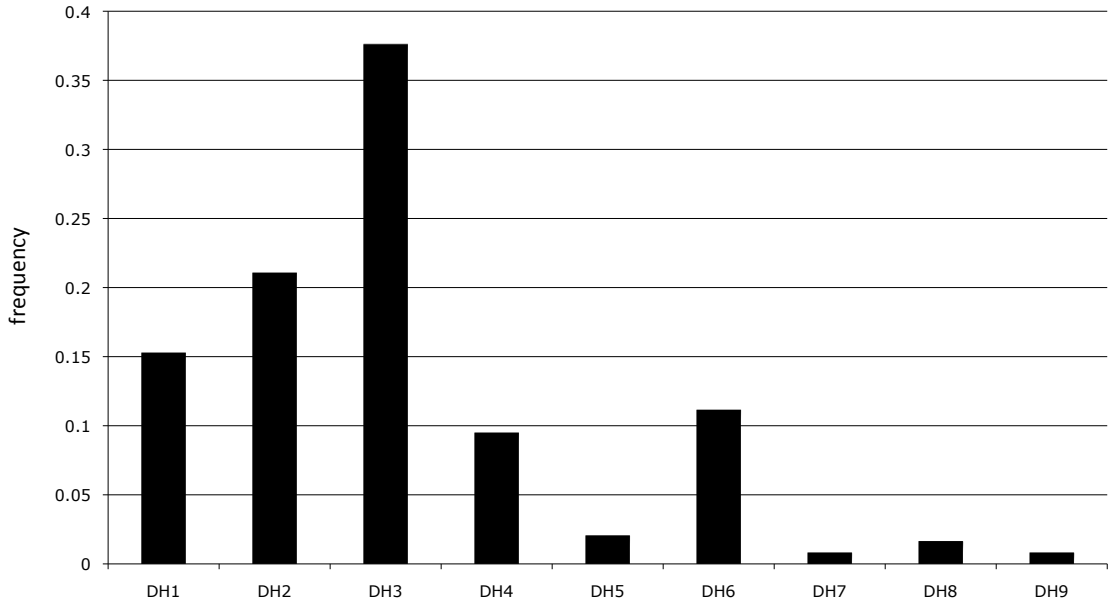


Figure 10. DH usage frequency in opossums IgH transcripts. Analyses of the usage frequency of nine DHs in IgH clones obtained from all ages detected.

Opossum L chain diversity in early ontogeny

There are 64 total V genes in the opossum Igλ locus and based on nucleotide identity they have been grouped into four subfamilies (Lucero et al 1998, Wang et al 2009). Of the 64 V genes, 58 appear functional based on genomic analysis (Wang et al 2009). The Vλ4 family contains only a single gene segment that appears functional, however this gene was never detected in expressed VJ rearrangements at any age (not shown). As shown previously, Igλ transcripts were first detected at P7 but only by RT-PCR using Vλ1 specific primers (Fig. 5). Vλ1 is the largest V family, containing 83% (48 out of 58) of the functional Vλ genes. The diversity of Vλ1-Jλ rearrangements at P7 was limited and only a total of six different recombinants were isolated from two different individuals. These six Vλ1 clones used five different Vλ genes, which were scattered across the 1.5 Mb region of the opossum Igλ locus (Fig. 11) (Wang et al 2009). There was no evidence of bias for Vλ based on position at any age in the opossum (Fig. 11). As in other mammalian species, the opossum Igλ locus contains tandem J-C pairs; there are eight opossum Cλ each with their own upstream Jλ (Lucero et al 1998, Wang et al 2009). Five of the six P7 rearrangements used Jλ8-Cλ8, which is the most V proximal of the J-C pairs (Fig. 12). By P8, Igλ transcripts containing V genes from all three expressed Vλ families could be detected (Fig. 12). Of 38 Igλ transcripts sequenced from P8, almost half (16 out of 38) also used the Jλ8-Cλ8 cluster. This bias for the most V proximal J-C pair, however, appeared to be gone by P11 (Fig. 12).

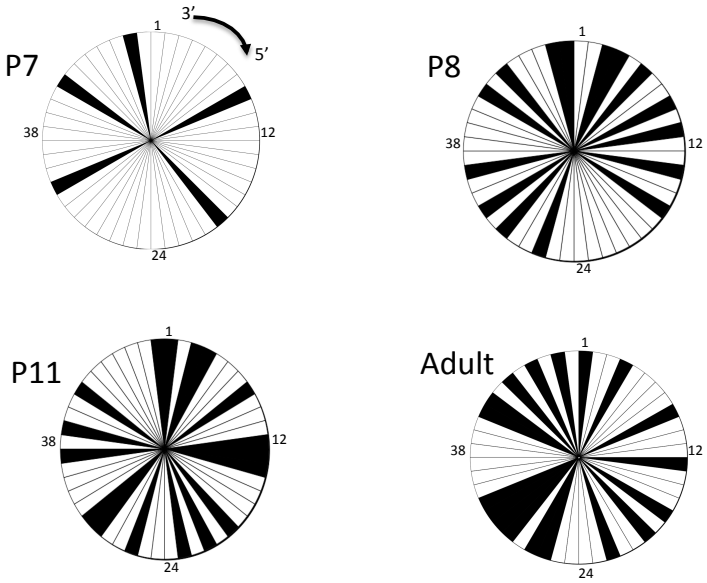


Figure 11. Expression of Vλ1 genes in opossums at different ages. Pie diagrams representing the detection of transcripts containing Vλ1 gene segments in opossums at ages of P7, P8, P11 and Adult. Vλ1 gene segments are ordered clockwise on the pie diagrams from 3' to 5' in the Igλ locus. Due to the large number of Vλ1 gene segments, not all of them are labeled in the figure. For positional reference, four gene segments are numbered on four positions of the clock. Black filled pie wedges indicate positive detection of that specific V gene in a Igλ transcript by RT-PCR at the specific ages.

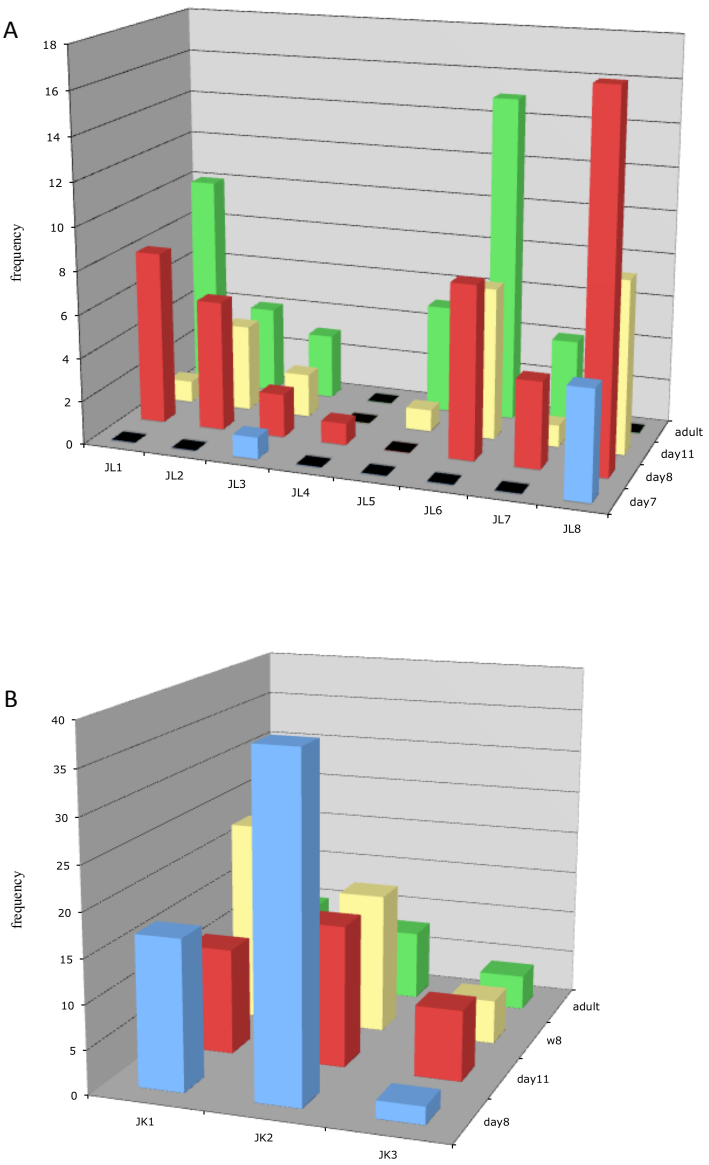


Figure 12. Usage frequency of Jκ and Jλ in opossums at different ages. *A*, column chart representing the frequency of transcripts containing Jλ genes in the opossums at ages P7, P8, P11 and adult. *B*, column chart representing the frequency of transcripts containing Jκ genes in the opossums at ages P8, P11, week 8 and adult.

In addition to P7 and P8, Igλ transcripts were collected from animals at ages P9, P11, and week 8. Furthermore, Igλ sequences from adult opossum were available from a previous study and were included in this analysis (Lucero et al 1998). A comprehensive analysis of the CDR3 length found that it remained rather unchanged over developmental time, however the CDR3 of Vλ1 recombinants were 25% longer than Vλ2 and Vλ3 due to Vλ1 gene segment coding regions being longer (Table I). Approximately a third to a half of all Vλ CDR3 contained evidence of microhomology in the P7 through P11 animals. By week 8 and in the adult there was no evidence of using microhomology (Table I). Less than half of the L chain clones in the opossum contained N-additions at any age, consistent with down-regulation of TdT later in pre-B cells undergoing L chain rearrangement as has been seen in other species (Li et al., 1993).

Functional Igκ V-J rearrangements were first detected at P8 (Fig. 5). There are 122 V genes divided into seven V families in the opossum Igκ locus, however only 82 Vκ appear functional based on genomic sequence (Wang et al 2009). V genes from six of the seven Vκ families could be detected in the expressed Igκ repertoire at P8 (Fig. 5). Missing was the Vκ6 family, which contains only a single V gene and was never used at any age (not shown). There was no bias in the selection of Vκ genes used at P8 based on position in the Igκ locus; Vκ genes from across the locus were found to be used on the first day Igκ transcripts could be detected (not shown) (Wang et al 2009). As with Igλ clones, the Igκ CDR3 lengths did not vary significantly with age, and most clones lacked N region additions (Table I). The opossum Igκ locus contains a single Cκ gene and previously only two Jκ had been identified (Miller et al 1999, Wang et al 2009). While

analyzing expressed Vk-Jk recombinants, a third Jκ designated Jκ3, was found and mapped 3' of Jκ1 and Jκ2 in the Igκ locus (Fig. 13). All three Jκ were used starting on P8 (Fig. 12). Using Jκ3 to search the opossum whole genome sequence a fourth J, Jκ4, was also found. However, although appearing functional, Jκ4 was not used at any age. The expressed Vκ-Jκ rearrangements were analyzed for evidence of micro-homology at the junctions. Micro-homology could be detected in clones at early time-points independent of which Vκ or Jκ were used (Fig. 14).

In summary, the diversity of expressed κ and λ L chains is greater in early B cells than is H chain diversity. This observation is consistent with a previous hypothesis that marsupials may rely more heavily on L chain germ-line diversity than H chain for developing their antibody repertoire (Baker et al 2005). Indeed, developing opossums appear to express a repertoire of Ig λ and κ chains at the earliest time-points that is equivalent to that of the adult.

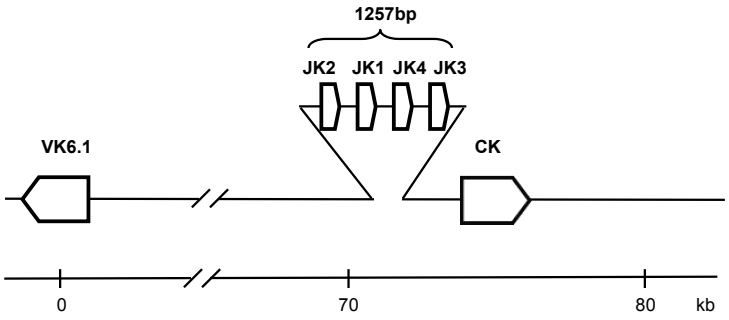


Figure 13. Partial map of the opossum Igκ locus. The location of two new identified Jκ (Jκ3 and Jκ4) gene segments are indicated in the map of opossum Igκ locus.

Clone #	VJ	V 3' end	Microhomology	J 5' end
1132/8/2	VK1.30_JK1	TGCCTACAAGTCAATAGCTGGCCTC	T	ATACATTCGGAGACGGAACCCAAC
1132/8/3	VK1.2_JK1	CTGCCTACAACTCAGTAGTTATCCTC	T	ATACATTCGGAGACGGAACCCAAC
1133/8/1/8	VK2.35_JK2	TCAAGGCACTCGCTGGCCTTCT	A	CGTTCGGCGGTGGGACCAAGGTGG
1133/8/2/1	VK2.9_JK3	TATCAACACTCTCACTCACCGG	TC	ATGTTTGGTGCAGGGACCAAGGTGGAGATAAAAC
1135/8/2/6	VK3.21_JK2	ACTGTCACCAGTATGATGTATCCCC	T	TGGACGTTCGGCGGTGGGACCAAGGTGGAAATTAAAC
1135/8/2/7	VK3.17_JK2	CTGTCAGCAATATCATAATTC	T	TGGACGTTCGGCGGTGGGACCAAGGTGGAAATTAAAC
1139/8/2/2/2	VK7.5_JK3	AAGGATCTTTATATCCT	CTCA	TGTTTGGTGCAGGGACCAAGGTGGAGATAAAAC
1139/8/2/2/3	VK7.5_JK2	AAGGATCTTTATATCC	T	TGGACGTTCGGCGGTGGGACCAAGGTGGAAATTAAAC

Figure 14. Representative partial Igκ transcripts collected from P8 opossums. The clone numbers are indicated in the first column. The specific Vκ and Jκ germ-line gene segments used in each clone are indicated in the third column. Micro-homologies at the VJ junctions are indicated in the fourth column.

A partial germ-line joined VH contributes to the opossum Ig repertoire

Opossums are unique amongst mammals in that they have a partially germ-line joined VH gene. The lone member of its own family, VH3.1 is already fused to a DH and the RSS at the 3' end of the gene segment has a 12 bp spacer similar to DH genes, rather than a 23 bp spacer found in VH genes (Wang et al 2009). Whether VH3.1 contributes to the IgH repertoire was further investigated. First, the large database of expressed H chain VDJ recombinants provided a resource to test the possibility that VH3 may be used to modify other recombined VH genes during B cell development through gene-conversion. This was a reasonable possibility given that the contribution of partially germ-line joined pseudo-VH through gene-conversion is essential to Ig diversity in avian species (Reynaud et al 1989). A total of 240 independent VDJ recombinants from all ages were aligned based on the VH1 gene used. Although single nucleotide variation was found between clones using the same VH gene, there was no evidence of longer (greater than one or two nucleotides) stretches of sequence replacement derived from VH3.1 or any other VH gene that would be indicative of gene-conversion occurring (not shown).

Next, the possibility that VH3.1 undergoes VDJ recombination and is expressed directly in the IgH repertoire was investigated. VH3.1 appears to be functional in that it contains a leader sequence, open reading frame and a conserved RSS. However, VH3.1 was not found in previous analyses of the expressed opossum H chain repertoire (Miller et al 1998, Aveskogh et al 1999). Furthermore, attempts to specifically amplify transcripts containing VH3.1 from P1 through P28 individuals were also unsuccessful (Wang et al 2009). Transcripts containing productive VH3.1 recombinants, however, were detected

in splenic mRNA from week seven and eight individuals but not week 4 (Fig. 3A and result not shown). Since VH3.1 contains its own germ-line joined DH gene it was of interest to investigate junctional diversity in clones using VH3.1. Nineteen unique VH3.1 clones were sequenced from splenic mRNA from an eight-week-old animal. In contrast to splenic clones using VH1 or VH2 genes where 100% are productive, only 52% (10 of 19) using VH3.1 were productive (Fig. 15).

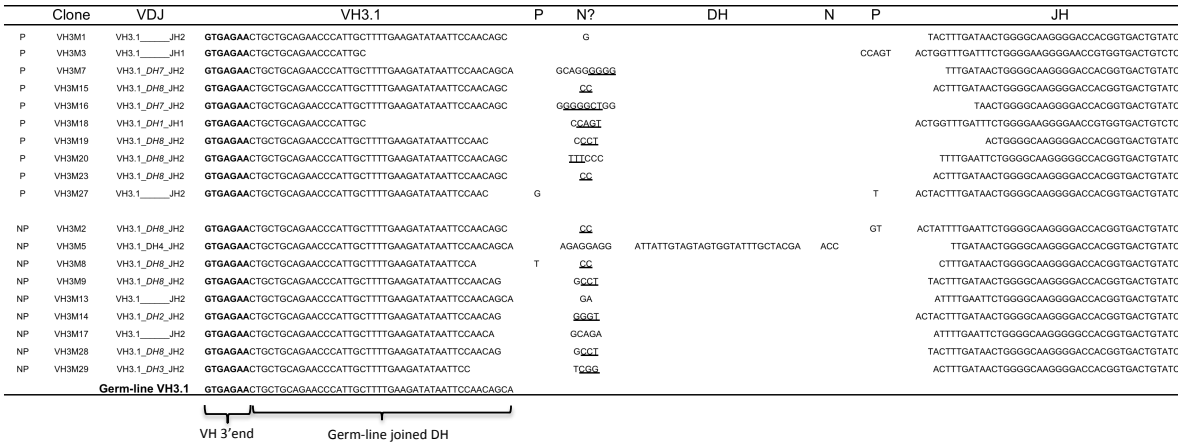


Figure 15. Alignment of Ig rearranged transcripts using the partially germ-line joined VH3.1 gene segments. The first column indicates grouped rearrangements being productive (P) or non-productive (NP). The second column indicates which JH was used and a possible DH contribution in italics. The sequence corresponding to the end of VH3.1 is shown in the third column with the nucleotides from the V region in bold. P and N additions in the junctions are indicated. Those N-nucleotides that could possibly have come from DH are underlined. The bottom sequence shows the 3' end of the germ-line VH3.1.

The majority of VH3.1 clones, both productive and non-productive rearrangements, demonstrated very little trimming of the germ-line joined DH segment (Fig. 15). This DH adds 44 nucleotides added to the end of VH3.1, which is substantially longer than the most commonly used conventional opossum germ-line DH, which range from 14 to 18 nucleotides prior to any trimming (Wang et al 2009). Consequently the CDR3 of VH3.1 clones were long, averaging 20 codons (range 15 to 23 codons), compared with an average of 12.3 codons for CDR3 of VH1 clones in the adult repertoire. Seven of the 10 productive VH3.1 rearrangements contained additional nucleotides in the junction, which ranged from 1 to 9 nucleotides, that could not be accounted for from the germ-line VH3.1 or JH used (Fig. 15). Some of these additional nucleotides could be from the inclusion of an extra DH in addition to the DH already germ-line joined to VH3.1. However, this putative DH contribution was never more than six nucleotides in length and, rather, may be N-additions from TdT activity. This is in contrast to week eight clones using VH1 family members where, as described above, 100% of clones contained N-additions (Fig. 8 and data not shown). Furthermore, recombination that would include an extra DH would violate the 12/23 rule since the RSS flanking VH3.1 contains a 12 bp spacer due to the germ-line joined DH and all the opossum DH genes also contain 12 bp spaced RSS (Wang et al 2009). Indeed, only a single VH3.1 clone contained a clear additional DH segment and this was non-productive rearrangement (see clone VH3M5 in Fig. 15). Therefore it seems more likely that the use of VH3.1 involves a direct VH to JH (VD → J) recombination with the inclusion of P and N nucleotides to the junction.

Discussion

The results presented here support B cell ontogeny in the gray short-tailed opossum being entirely postnatal. It appears that pro-B cell commitment occurs within the first 24 postnatal hours as evident by P1 animals having at least initiated transcription of germ-line, sterile VH transcripts, whereas neither the germ-line sterile VH transcripts nor mature VDJ recombinants were detected in embryonic tissues from the period 24 hours prior to birth. Although not previously characterized in a marsupial, the sterile VH transcripts are likely the equivalent of those described in mice and humans and associated with chromatin remodeling and initiation of V to DJ recombination (Yancopoulos and Alt 1985, del Blanco et al 2011). Only half of the P1 animals tested had progressed to having B cells with H chain VDJ rearrangements further supporting that P1 opossums are on the cusp of pre-B cell development and that some had not progressed very far. Following VpreB expression and L chain gene rearrangement, the earliest time-point opossums would appear to have B cells with functionally rearranged H and L chain genes would be P7. This conclusion is consistent with earlier work on the development of immune-competence in Virginia opossums and other marsupial species, which were able to generate antibody responses only in the second postnatal week (Kalmutz 1962, La Via et al 1963, Rowlands et al 1972, Stanley et al 1972).

This current study is the first to investigate marsupial B cell lineage development at a molecular level starting prior to birth. Another recent study of early marsupial B cell development investigated CD79a/CD79b transcription in the tammar wallaby *Macropus*

eugenii pouch young (Duncan et al 2010). Transcription of both these BCR subunits in bone marrow, cervical thymus, and lung, and CD79a alone on spleen, gut and blood tissues was detected in P10 *M. eugenii*. Unfortunately P10 was the earliest time point tested and no other component of the BCR was investigated. Furthermore, immunohistochemistry had already shown that B cell development in wallabies was initiated earlier than P10, as $CD79b^{+}$ cells could be detected in P7 *M. eugenii* pouch young (Old and Deane 2003). A similar study of B cell development in the brushtail possum (*Trichosurus vulpecula*) reported the detection of IgM transcripts in pouch young spleen and liver by P10 (Belov et al 2002). This study also, unfortunately, did not look at earlier time-points nor did it investigate the diversity of IgH at early developmental stages. It is noteworthy, however, that *T. vulpecula* B cells that had switched to IgG were not detected until over three months of age. This is significant in that it is following the time when this species transitions from being firmly attached to the teats and, late in the second month, begin to suckle intermittently similar to newborn placental mammals (Lyne et al 1959). Similarly an IgG switch was not detected in the opossum until the fifth postnatal week, which is also after opossums have also ceased to be firmly attached to the teats and start suckling intermittently and moving around independently (VandeBerg 1990). *M. domestica* is a pouch-less marsupial and once the young detach from the teats they are maintained in a nest much like mice, rather than being held in a pouch. This transition may be associated with increased exposure to environmental antigens after the young, although not yet weaned, are more actively rooting around and increasing the likelihood of ingesting pathogens and parasites, driving B cell maturation into the memory pool and class switch. If antigen exposure were obligatory for driving early

isotype switch in marsupials this would be in contrast to placental species where fetal B cells initiate class switch independently of antigen exposure (Milili et al 1991, Butler et al 2001).

A significant difference between the results presented here and the *T. vulpecula* study is the appearance of IgA. In *T. vulpecula*, IgA transcripts were detected early at P18, prior to IgG (Belov et al 2002). In the opossum, IgA was detected late, around the eighth postnatal week in the spleen and even at mucosal sites such as the gut. This may represent species-specific differences, but more likely is due to differences in the history of the animals used. In the experiments described here, the opossums are derived from a long-term captive colony kept under pathogen free conditions. In the *T. vulpecula* study the animals were wild-caught and likely at greater risk of prior exposure to mucosal pathogens that may have stimulated an early IgA response (Belov et al 2002). Furthermore, if antigen exposure is driving early isotype-switch in marsupials after they have detached from the teat this speaks to just how tight a barrier the nipple provides to protect the newborn marsupial from antigen exposure.

M. domestica is the only mammal found so far to have a germ-line joined VH gene (Wang et al 2009). VH3.1 is partially germ-line joined to a DH and contains a functional DH-type RSS, with a 12 bp spacer (Wang et al 2009). Early analyses of expressed adult IgH diversity failed to uncover VH3.1, and its discovery only came about from a detailed analysis of the IgH genomic region (Miller et al 1998, Aveskogh et al 1999, Wang et al 2009). Previous attempts to detect VH3.1 expression focused on early developmental

time-points. This was, in part, based on evidence from species where germ-line joined V genes are more common, such as in sharks. In sharks, germ-line joined VH are expressed in the early developing IgH repertoire (Wang et al 2009, Rumfelt et al 2001). The prevailing hypothesis is that shark germ-line joined V genes may have been evolutionarily selected to encode idiotypes specific for common antigens or to generate antibodies that perform housekeeping roles (Rumfelt et al 2001). Such antibodies would be expressed early in ontogeny and not depend on the randomness of V(D)J recombination for their generation. This seemed a reasonable possibility for VH3.1 as well given a VHDH germ-line joined gene would encode a pre-established CDR1, CDR2 and most of CDR3. However, this does not appear to be the case for the opossum. B cells that successfully recombine VH3.1 appear late in the adult repertoire and were difficult to detect compared with other VH gene segments. This does not mean they do not perform some select function, however it does not appear to be analogous to the situation for germ-line joined V genes in sharks. In other words, VH3.1 does not appear to perform a specific function in the early Ig repertoire in juvenile opossums.

An alternative hypothesis for how VH3.1 might have contributed to the opossum IgH repertoire through gene conversion was also tested. This mechanism is critical to the development of antibody diversity in birds where the donor sequences are also partially germ-line joined VDH, although they are pseudogenes (Reynaud et al 1989). There was no evidence in the opossum, however, of mutation of rearranged and expressed VH through gene conversion by sequences donated by VH3.1, or any other VH genes.

It may not be surprising that VH3.1 use is rare and went undetected in previous studies. B cells using VH3.1 have a higher frequency of non-productive rearrangements than do those using VH1 and VH2. In addition VDJ recombination using VH3.1 is essentially an out-of-order gene rearrangement. Pro-B cells typically rearrange first a DH to a JH followed by VH to DJH (Alt et al 1984, Hardy et al 1991). Rearrangements involving VH3.1 are essentially a reverse order VDH to JH. Furthermore, VH protein domains using VH3.1 would have unusually long CDR3, which may have biological consequences due to potential instability. The range of CDR3 length in VH3.1 clones were comparable to what has been found for cattle H chains and longer than found for the platypus, two species noted for have long, diverse IgH CDR3 (Johansson et al 2002, Saini et al 1999). Unlike what is found in cattle and platypus, however, the opossum VH3.1 clones lack additional cysteines that are thought to help stabilize long CDR3 (Johansson et al 2002, Saini et al 1999). Furthermore, long H chain CDR3 constrain the ability of H chains to pair with some L chains properly (Saini et al 2003). Therefore, it is possible that opossum B cells that rearrange VH3.1 are at a selective disadvantage and, for reasons enumerated above, rarely successful. How VH3.1 may contribute to the functional Ig repertoire, if at all, remains to be determined.

The use of somatic mutation, such as gene conversion, to diversify the primary antibody repertoire is associated with absence of germ-line V gene diversity. Birds and some placental species such as rabbits are noted for having limited VH gene diversity and reliance of somatic mutation for generating a diverse primary Ig repertoire, typically taking place in a gut associated lymphoid tissue (GALT) such as the *Bursa of Fabricius*

or the appendix (reviewed in Butler 1997). In such species both the Ig H and L chain loci have limited germ-line V diversity and are somatically mutated. Mice and humans in contrast have diverse germ-line V repertoires and do not use mutation to further diversify the primary repertoire. Opossums and other marsupials have limited VH diversity like birds and rabbits (Miller et al 1998, Wang et al 2009). But as described in the previous paragraph they do not rely on somatic mutation for diversity like birds and rabbits. In contrast, opossums and other marsupials have diverse L chain V genes (Miller et al 1999, Lucero et al 1998, Wang et al 2009, Baker et al 2005). Other investigators have hypothesized that germ-line H and L chain V gene diversity co-evolves (Sitnikova and Su 1998) Marsupials appear to break this rule (Baker et al 2005). Indeed, we previously hypothesized that marsupials depend upon L chain diversity for their overall Ig diversity (Baker et al 2005). The results presented here remain consistent with that hypothesis as there is early use of a diverse Vλ and Vκ repertoire in the developing opossum.

While characterizing B cell development in the opossum, a *M. domestica* VpreB3 homologue was identified, the first surrogate L chain to be characterized in a marsupial. Three *VpreB* genes, *VpreB1, VpreB2* and *VpreB3,* have been described in humans and mice (Shirasawa et al 1993, Rosnet et al 1999). Only a single *VpreB* related gene was found in the opossum and it is a *VpreB3* homologue. Phylogenetic analyses of the three *VpreB* gene lineages support their origin by three gene duplication events. An earlier duplication gave rise to the more ancient VpreB3, which is present in birds, marsupials, and placental mammals (Rosnet et al 1999, Rosnet et al 2004). The second duplication, which appears to be specific to placental mammals, gave rise to a VpreB that later

duplicated again to create VpreB1 and VpreB2. VpreB1 and VpreB2 share 97% amino acid identity, while VpreB3 shares only 37% amino acid identity with VpreB1 and 2 (Mårtensson et al 2007). The chicken *VpreB3* gene is linked to the *Igλ* locus, as are *VpreB1* and *VpreB2* in humans and mice (Rosnet et al 1999, Rosnet et al 2004, Kudo & Melchers 1987). In mice, however, the *VpreB3* gene is on chromosome 10, non-syntenic to *Igλ* (Rosnet et al 1999). Unfortunately the VpreB3 gene was found amongst the unassembled sequences in the opossum genome database and, therefore, it is not known if *VpreB3* is syntenic to *Igλ* in the opossum.

VpreB1 and 2, along with λ5, are known to be components of the extra-cellular pre-BCR, which also includes IgM H chain and the CD79a and CD79b signaling molecules (Karasuyama et al 1990, Tsubata & Reth 1990, Mårtensson et al 2007). The role of VpreB3 is less well defined. VpreB3 has been linked to retention of free L chains in the endoplasmic reticulum (ER) in chickens, and in mice has been found to associate with nascent H chains and λ5 in the ER but does not traffic to the cell surface (Ohnishi & Takemori 1994, Mårtensson et al 2007). Given its conservation in marsupials as well as eutherians, this may be an ancient function that has been retained in these vertebrate lineages (Rosnet et al 2004). We were unable to identify homologues of *VpreB1, VpreB2,* or *λ5* in the opossum genome. To our knowledge they have not been identified in any species other than placental mammals and it is possible that they do exist in the opossum and are too divergent to identify by *in silico* methods. Alternatively they may be absent and B cell development in the opossum uses an alternative pre-BCR. Indeed, genetically altered mice have been shown to transport intact IgM H chain to the cell

surface that lack surrogate L chains but signal progression of B cell development (Schuh et al 2003, Su et al 2003, Geraldes et al 2007). Perhaps something analogous is occurring during opossum B cell ontogeny.

Having identified at least VpreB3, three stages of B cell development could be distinguished in the opossum: H chain gene rearrangement, VpreB3 expression, and L chain gene rearrangement. In fetal mice the transition from H chain rearrangement, surrogate L expression, and L chain rearrangement and expression occurs over a one to two day period (Palacios & Samaridis 1992). This process is surprisingly drawn out in the newborn opossum, occurring over a period of a nearly a week. The delay appears to be primarily due to the lag time between rearranging the H chain genes and expression of the surrogate L chains. The reason for this lag-time is not entirely evident. It is possible that B cells that have successfully rearranged their H chain replicate to take advantage of a successful H chain rearrangement by producing daughter clones that undergo independent L chain rearrangements. This hypothesis is consistent with L chains being more important than H chains for overall Ig diversity in opossums (Baker et al 2005). It is unlikely that the delay is an artifact from being unable to detect low level VpreB and rearranged L chain gene expression at earlier time-points as the detection of H chain transcripts was not difficult; in other words, it does not appear to be due to low cell number.

There is limited or complete absence of addition of N-nucleotides to the VDJ junctions of IgH genes in opossum neonatal B cells. However, there is no evidence that this results in

any particular bias in the selection of VH gene segments early in B cell development, either related to genomic position or micro-homology as has been seen for some placental mammals. One question then is why do early opossum B cell H chains lack N-additions? It is clear that early developing T cells are using TdT to add N-additions to the junctions of TCRα chain genes at the same time-point in development in the opossum (Parra et al 2009). One possible explanation is provided by recent observations by Schelonka and colleagues who found that reconstituting mice with B cells from TdT deficient bone marrow resulted in a more rapid repopulation of the spleen and peritoneal cavity compared to wild-type bone marrow (Schelonka et al 2011). A hypothesis is that the lack of N-additions in early B cells favors clones that more easily pass through some developmental checkpoint and therefore dominate the early seeding of lymphoid tissues. This appears to create a biological trade-off however. B cells lacking H chain N-additions may populate secondary lymphoid sites more readily, but at the expense of their ability to respond to a variety of pathogens (Schelonka et al 2011). It is possible that marsupials are gambling on this trade-off and using early B cells to rapidly populate tissues and drive the development of the secondary lymphoid organs, while relying on maternal Ig absorbed through the milk for protection. Although opossums do not receive any maternal Ig trans-placentally they do rapidly absorb antibodies from the milk soon after the initiation of suckling (Samples et al 1986).

Some aspects of early postnatal B cell development in the opossum recapitulate what has been found for fetal B cell development in humans and mice. A significant increase of the H chain CDR3 length during the development from fetal/neonate to adult has been

previously reported in mice and humans (Feeney 1992, Tonnelle et al 1995). A trend towards increasing H chain CDR3 length is also found in the opossum. There is little evidence for bias in the VH genes used in the earliest opossum H chain rearrangements as has been seen for mice, however (Feeney 1990, 1992). This may not be surprising since there is not much diversity to select from in the germ-line opossum VH repertoire (Wang et al 2009).

References

Adamski, F. M., and J. Demmer. 1999. Two stages of increased IgA transfer during lactation in the marsupial, *trichosurus vulpecula* (Brushtail possum). *J. Immunol.* 162: 6009–6015.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403–410.

Alt, F. W., G. D. Yancopoulos, T. K. Blackwell, C. Wood, E. Thomas, M. Boss and R. Coffman., et al. 1984. Ordered rearrangement of immunoglobulin heavy chain variable region segments. *EMBO J.* 3: 1209–1219.

Aveskogh, M. and L. Hellman. 1998. Evidence for an early appearance of modern post-switch isotypes in mammalian evolution; cloning of IgE, IgG and IgA from the marsupial *Monodelphis domestica*. *Eur. J. Immunol.* 28: 2738–2750.

Aveskogh, M., L. Pilstrom and L. Hellman. 1999. Cloning and structural analysis of IgM (μ chain) and the heavy chain V region repertoire in the marsupial Monodelphis domestica. *Dev. Comp. Immunol.* 23: 597–606.

Baker M. L., K. Belov and R. D. Miller. 2005. Unusually similar patterns of antibody V segment diversity in distantly related marsupials. *J. Immunol.* 174: 5665–5671.

Belov, K., M. A. Nguyen, K. R. Zenger, and D. W. Cooper. 2002. Ontogeny of immunoglobulin expression in the brushtail possum (*Trichosurus vulpecula*). *Dev. Comp. Immunol.* 26: 599–602.

Benedict, C. L. and J. F. Kearney. 1999. Increased junctional diversity in fetal B cells results in a loss of protective anti-phosphorylcholine antibodies in adult mice. *Immunity* 10: 607–617.

Butler, J. E., M. Sinkora, N. Wertz, W. Holtmeier, and C. D. Lemke. 2006. Development of the neonatal B and T cell repertoire in swine: implications for comparative and veterinary immunology. *Vet. Res.* 37: 417–441.

Butler, J. E. 1997. Immunoglobulin gene organization and the mechanism of repertoire development. *Scand. J. Immunol.* 45: 455–462.

Butler, J. E., P. Weber, M. Sinkora, J. Sun, S. J. Ford, and R. K. Christenson. 2000. Antibody repertoire development in fetal and neonatal piglets. II. Characterization of heavy chain complementarity-determining region 3 diversity in the developing fetus. *J Immunol.* 165: 6999–7010.

Butler, J. E., J. Sun, P. Weber, S. P. Ford, Z. Rehakova, J. Sinkora, and K. Lager. 2001. Antibody repertoire development in fetal and neonatal piglets. IV. Switch recombination,

primarily in fetal thymus, occurs independent of environmental antigen and is only weakly associated with repertoire diversification. *J. Immunol.* 167: 3239–3249.

Deane, E. M. and D. W. Cooper. 1984. Immunology of pouch young marsupials. I. Levels of immunoglobulin, transferrin and albumin in the blood and milk in euros and wallaroos (hill kangaroos: *Macropus robustus*, Marsupialia). *Dev. Comp. Immunol.* 8: 863–876.

Decker, D. J., N. E. Boyle, J. A. Koziol and N. R. Klinman. 1991. The expression of the Ig H chain repertoire in developing bone marrow B lineage cells *J. Immunol.* 146: 350–361.

del Blanco, B., V. García, A. García-Mariscal, and C. Hernández-Munain. 2011. Control of V(D)J Recombination through Transcriptional Elongation and Changes in Locus Chromatin Structure and Nuclear Organization. *Genet. Res.* 2011: 1–10.

Derrickson, E M. 1992. Comparative reproductive strategies of altricial and precocial eutherian mammals. *Funct. Ecol.* : 6: 57–65.

Duncan, L., K. Webster, V. Gupta, S. Nair, and E. Deane. 2010. Molecular characterisation of the CD79a and CD79b subunits of the B cell receptor complex in the gray short-tailed opossum (*Monodelphis domestica*) and tammar wallaby (*Macropus eugenii*): Delayed B cell immunocompetence in marsupial neonates. *Vet. Immunol. Immunopathol. 136* : 235–247.

Fadem, B. H. 1985. Evidence for the activation of female reproduction by males in a marsupial, the gray short-tailed opossum (*Monodelphis domestica*). *Biol Reprod* 33*:* 112–116.

Feeney, A. J. 1990. Lack of N Regions in Fetal and Neonatal Mouse Immunoglobulin V D J functional Sequences. *J. Exp. Med.* 172: 1377–1390.

Feeney, A. J. 1992. Predominance of VH-D-JH junctions occurring at sites of short sequence homology results in limited junctional diversity in neonatal antibodies. *J. Immunol.* 149: 222–229.

Geraldes, P., M. Rebrovich, K. Herrmann, J. Wong, H. M. Jäck, M. Wabl, and M. Cascalho. 2007. Ig heavy chain promotes mature B cell survival in the absence of light chain. *J. Immunol.* 179:1659–1668.

Hall, T. A. 1994. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic. Acids. Symp. Ser.* 41: 95–98.

Hardy, R. R., C. E. Carmack, S. A. Shinton, J. D. Kemp, and K. Hayakawa. 1991. Resolution and characterization of pro-B and pre-pro-B cell stages in normal mouse bone marrow. *J. Exp. Med.* 173: 1213–1225.

Hombach, J., T. Tsubata, L. Leclercq, H. Stappert, and M. Reth. 1990. Molecular-Components of the B-Cell Antigen Receptor Complex of the Igm Class. *Nature* 343: 760–762.

Ishino, S., K. Kadota, Y. Matsubara, H. Agawa, and N. Matsui. 1991. Immunohistochemical studies on ontogeny of bovine lymphoid tissues. *J. Vet. Med. Sci.* 53: 877–882.

Johansson, J., M. Aveskogh, B. Munday, and L. Hellman. 2002. Heavy chain V region diversity in the duck-billed platypus (*Ornithorhynchus anatinus*): long and highly variable complementarity-determining region 3 compensates for limited germline diversity. *J. Immunol.* 168: 5155–5162.

Kalmutz, S. E. 1962. Antibody production in the opossum embryo. *Nature* 193: 851–853.

Karasuyama, H., A. Kudo, and F. Melchers. 1990. The proteins encoded by the VpreB and lambda 5 pre-B cell-specific genes can associate with each other and with mu heavy chain. *J. Exp. Med.* 172: 969–972.

Kokubu, F., R. Litman, M. J. Shamblott, K. Hinds and G. W. Litman. 1988. Diverse organization of immunoglobulin VH gene loci in a primitive vertebrate. *EMBO J.* 7: 3413–3422.

Koyama, M., K. Ishihara, H. Karasuyama, J. L. Cordell, A. Iwamoto, T. Nakamura. 1997. CD79a/CD79b heterodimers are expressed on pro-B cell surfaces without associated μ heavy chain. *Int. Immunol.* 9: 1767–1772.

Kudo, A., and F. Melchers. 1987. A second gene, VpreB in the lambda 5 locus of the mouse, which appears to be selectively expressed in pre-B lymphocytes. *EMBO J.* 6: 2267–2272.

LaPlante, E. S., R. Burrell, A. L. Watne, D. L. Taylor, and B. Zimmermann. 1969. Skin allograft studies in the pouch young of the opossum. *Transplantation* 7: 67–72.

La Via, M. R., D. T. Jr Rowlands, and M. Block. 1963. Antibody formation in embryos. *Science* 140: 1219–1220.

Li, Y. S., K. Hayakawa, and R. R. Hardy. 1993. The regulated expression of B lineage associated genes during B cell differentiation in bone marrow and fetal liver. *J. Exp. Med.* 178: 951–960.

Lucero, J. E., G. H. Rosenberg and R. D. Miller. 1998. Marsupial light chains: complexity and conservation of λ in the opossum *Monodelphis domestica*. *J. Immunol.* 161: 6724–6732.

Lyne, A. G., P. E. Pilton, and G. B. Sharman. 1959. Oestrous cycle, gestation period and parturition in the marsupial Trichosurus vulpecula. *Nature* 183: 623–623.

Malecek, K., V. Lee, W. Feng, J. L. Huang, M. F. Flajnik, Y. Ohta, E. Hsu. 2008. Immunoglobulin heavy chain exclusion in the shark. *PLoS Biol.* 6: e157.

Mårtensson, I.-L., R. A. Keenan, and S. Licence. 2007. The pre-B-cell receptor. *Curr. Opin. Immunol.* 19: 137–142.

Mikkelsen, T. J., M. J. Wakefield, B. Aken, C. T. Amemiya, J. L. Chang, S. Duke, M. Garber, A. J. Gentles, L. Goodstadt, A. Heger, et al. 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* 447: 167–177.

Milili, M., M. Fougereau, P. Guglielmi, and C. Schiff. 1991. Early occurrence of immunoglobulin isotype switching in human fetal liver. *Mol. Immunol.* 28: 753–761.

Miller, R. D., H. Grabe, and G. H. Rosenberg. 1998. VH Repertoire of a Marsupial (*Monodelphis domestica*). *J. Immunol.* 160: 259–265.

Miller, R. D., R. Elizabeth, and G. H. Bergemann. 1999. Marsupial light chains: IGK with four V families in the opossum *Monodelphis domestica*. *Immunogenetics* 50: 329–335.

Ohnishi K, T. J. Takemori. 1994. Molecular components and assembly of mu.surrogate light chain complexes in pre-B cell lines. *Biol. Chem.* 269: 28347–28353.

Old, J. M., and E. M. Deane. 2000. Development of the immune system and immunological protection in marsupial pouch young. *Dev. Comp. Immunol.* 24: 445–454.

Old, J. M., and E. M. Deane. 2003. The detection of mature T- and B-cells during development of the lymphoid tissues of the tammar wallaby (*Macropus eugenii*). *J. Anat. 203*: 123–131.

Palacios, R., and J. Samaridis. 1992. Fetal liver pro-B and pre-B lymphocyte clones: expression of lymphoid-specific genes, surface markers, growth requirements, colonization of the bone marrow, and generation of B lymphocytes in vivo and in vitro *Mol. Cell. Biol.* 12: 518–530.

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

Parra, Z. E., M. L. Baker, A. M. Lopez, J. Trujillo, J. M. Volpe, R. D. Miller. 2009. TCR mu recombination and transcription relative to the conventional TCR during postnatal development in opossums. *J Immunol.* 182: 154–163.

Reynaud. C.A., A. Dahan, V. Anquez, and J. C. Weill. 1989. Somatic hyperconverslon diversifies the single VH gene of the chlcken with a hlgh incidence in the D region. *Cell* 59: 171.

Rosnet, O., M. Mattei, O. Delattre, and C. Schiff. 1999. VPREB3: cDNA characterization and expression in human and chromosome mapping in human and mouse. *Cytogenet. Cell. Genet.* 87: 205–208.

Rosnet, O., C. Blanco-Betancourt, K. Grivel, K. Richter and C. Schiff. 2004. Binding of free immunoglobulin light chains to VpreB3 inhibits their maturation and secretion in chicken B cells. *J. Biol. Chem.* 279: 10228–10236.

Rowlands, D. T., D. Blakeslee, and H. H. Lin. 1972. The early immune response and immunoglobulins of opossum embryos. *J Immunol.* 108: 941–946.

Rumfelt, L. L., D. Avila, M. Diaz, S. Bartl, E. C. McKinney, and M. F. Flajnik. 2001. A shark antibody heavy chain encoded by a nonsomatically rearranged VDJ is preferentially expressed in early development and is convergent with mammalian IgG. *Proc. Natl. Acad. Sci. USA* 98: 1775–1780.

Saini, S. S., B. Allore, R. M. Jacobs, and A. Kaushik. 1999. Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM Abs. *Eur. J. Immunol.* 29: 2420.

Saini, S. S. 2003. Bovine IgM antibodies with exceptionally long complementarity-determining region 3 of the heavy chain share unique structural properties conferring restricted VH + Vλ pairings. *Int. Immunol.* 15: 845–853.

Samollow, P. B. 2006. Status and applications of genomic resources for the gray, short-tailed opossum, *Monodelphis domestica*, an American marsupial model for comparative biology. *Aust. J. Zool.* 54: 173–196.

Samples, K., J. L. Vandenberg, and W. H. Stone. 1986. Passively acquired immunity in the newborn of a marsupial (*Monodelphis domestica*). *Am. J. Reprod. Immunol. Microbiol.* 11: 94–97.

Schelonka, R. L., I. I. Ivanov, A. M. Vale, R. A. Dimmitt, M. Khaled, and H. W. Schroeder. 2011. Absence of N addition facilitates B cell development, but impairs immune responses. *Immunogenetics* 63: 599–609.

Schuh, W., S. Meister, E. Roth, and H.-M. Jäck. 2003. Cutting edge: signaling and cell surface expression of a mu H chain in the absence of lambda 5: a paradigm revisited *J Immunol.* 171: 3343–3347.

Schelonka R. L., I. I. Ivanov, A. M. Vale, R. A. Dimmitt, M. Khaled, H. W. Jr Schroeder. 2011. Absence of N addition facilitates B cell development, but impairs immune responses. *Immunogenetics* 63: 599–609.

Shirasawa, T., K. Ohnishi, S. Hagiwara, K. Shigemoto, Y. Takebe, K. Rajewsky, and T. Takemori. 1993. A novel gene product associated with mu chains in immature B cells. *EMBO J.* 12: 1827–1834.

Sitnikova, T. and C. Su. 1998. Coevolution of immunoglobulin heavy-and light-chain variable-region gene families. *Mol. Biol. Evol.* 15: 617–625.

Stanley, N. F., M. Yadav, H. Waring, M. Eadie. 1972. The effect of thymectomy on response to various antigens of a marsupial *Setonix brachyrus* (Quokka). *Aust. J. Exp. Biol. Med. Sci.* 50: 689–702.

Su, Y. W., A. Flemming, T. Wossning, E. Hobeika, M. Reth, and H. Jumaa. 2003. Identification of a Pre-BCR Lacking Surrogate Light Chain. *J. Exp. Med.* 198: 1699–1706.

Tamura, K., J. Dudley, M. Nei and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 *Mol. Biol. Evol.* 24: 1596–1599.

Tonnelle, C., A. M. Cuisinier, L. Gauthier, V. Guelpa-Fonlupt, M. Milili, C. Schiff, and M. Fougereau. 1995. Fetal versus adult pre-B or B cells: the human VH repertoire. *Ann. NY Acad. Sci.* 764: 231–241.

Tsubata, T., and M. Reth. 1990. The products of pre-B cell-specific genes (lambda 5 and VpreB) and the immunoglobulin mu chain form a complex that is transported onto the cell surface. *J. Exp. Med.* 172: 973–976.

VandeBerg, J. L. 1990. The Gray Short-Tailed Opossum (*Monodelphis Domestica*) as a Model Didelphid Species for Genetic Research. *Aust. J. Zool.* 37: 235–247.

Volpe, J. M., L. G. Cowell, T. B. Kepler. 2006. SoDA: implementation of a 3D alignment algorithm for inference of antigen receptor recombinations. *Bioinformatics* 22: 438–444.

Wang, X., J. J. Olp, R. D. Miller. 2009. On the genomics of immunoglobulins in the gray, short-tailed opossum *Monodelphis domestica*. *Immunogenetics* 61: 581–596.

Yadav, M., M. Eadie, N. F. Stanley. 1971. The thymus glands of a marsupial, *Setonix brachyurus* (quokka) and their role in immune responses. V1 Maternal transfer of immunoglobulins. *Aust. J. Exp. Biol. Med. Sci.* 49: 322–332.

Yancopoulos, G. D., and F. W. Alt. 1985. Developmentally controlled and tisue-specific expression of unrearranged VH gene segments. *Cell* 40: 271–281.

CHAPTER 4

PLATYPUS TCRμ PROVIDES INSIGHT INTO THE ORIGINS AND EVOLUTION OF A UNIQUELY MAMMALIAN TCR LOCUS

By Xinxin Wang, Zuly E. Parra, and Robert D. Miller*

*Center for Evolutionary and Theoretical Immunology and Department of Biology, University of New Mexico, Albuquerque NM 87131 USA

XW and RDM conceived of the study, participated in its design, and drafted the manuscript. XW and ZEP generated the data. XW performed the sequencing reaction and data analyses.

Manuscript published in Journal of Immunology (2011) 187: 5246-5254.

Abstract

TCRμ is an unconventional TCR that was first discovered in marsupials and appears to be absent from placental mammals and non-mammals. Here we show that TCRμ is also present in the duckbill platypus, an egg-laying monotreme, consistent with TCRμ being ancient and present in the last common ancestor of all extant mammals. As in marsupials, platypus TCRμ is expressed in a form containing double V domains. These V domains more closely resemble antibody V than that of conventional TCR. Platypus TCRμ differs from its marsupial homologue by requiring two rounds of somatic DNA recombination to assemble both V exons and has a genomic organization resembling the likely ancestral form of the receptor genes. These results demonstrate that the ancestors of placental mammals would have had TCRμ but it has been lost from this lineage.

Introduction

Conventional T cells exist in two distinct lineages based on the composition of their TCR heteroduplex: αβ T cells use a TCR composed of α and β chains while γδ T cells use γ and δ chains. Like Ig, the Ag binding V domains of the TCR chains are encoded by exons that are assembled from gene segments by somatic DNA recombination. All jawed vertebrates have both αβ and γδ T cells and the genes encoding these four TCR chains are highly conserved both in sequence and organization (1-3). Recently, a fifth locus encoding TCR chains, named *TCRμ*, were found in marsupial mammals (4). *TCR*μ contains C regions related to TCRδ but is transcribed in a form that would include double V domains that are more related to IgH V (VH) than to TCR V genes (2, 4, 5). TCRμ does not substitute for TCRδ in marsupials since the genes encoding conventional TCRδ chains are highly conserved and expressed (2, 6).

TCRμ genes are distinct and unlinked to those that encode conventional TCR chains and have atypical gene organization. The N-terminal V of TCRμ (Vμ) is encoded by somatically recombined genes (V, D, and J), with the recombination taking place in thymocytes, resulting in clonal diversity (4). The second, C-proximal V domain (Vμj) is encoded by an exon where the V, D, and J genes are already pre-joined in the germ-line DNA and are relatively invariant (4). This is the only known example of germ-line joined V genes being used in a TCR. The *TCRμ* locus is also organized in tandem clusters, which is also atypical of TCR genes (2, 4).

Searching the available placental mammal, avian, and amphibian genomes failed to uncover TCRμ orthologues (2). However, here we show that TCRμ is present in a monotreme, the duckbill platypus *Ornithorhyncus anatinus*. The monotremes are oviparous mammals that last shared a common ancestor with marsupials and placentals at least 165 million years ago (MYA) (7). The genomic organization of the platypus *TCRμ* locus reveals insight into the evolution of this uniquely mammalian TCR locus and supports its ancient presence in mammals.

Materials and Methods

Whole genome analysis and annotation

Analyses were performed using the platypus genome assembly Version 5.0.1 available at GenBank (http://www.ncbi.nlm.nih.gov/genome/guide/platypus/). Marsupial Cμ sequences were used to search based on homology using the BLAST algorithm (4, 5, 8). Scaffolds containing Cμ sequences were retrieved and exon boundaries were determined by the presence of canonical mRNA splice sites. Platypus cDNA sequences were used to search against the *O. anatinus* genome project to identify the genomic V, D and J gene segments. The beginning and end of each coding exon of V, D and J gene segments were identified by the presence of mRNA splice sites or flanking recombination signal sequences (RSS). Figure 1 shows the location of each TCRμ V, D, J and C segments on the scaffolds. Platypus TCRδ chain C region sequence (GenBank accession number XM_001516959) was used to identify the single copy platypus Cδ on scaffold 588, which is separate from any of the scaffolds containing the putative platypus TCRμ sequences.

PCR and cDNA analyses

A spleen cDNA library constructed from tissue from a Tasmanian platypus was screened by PCR (9). All PCR primer sequences used in this study are presented in Table I. PCR amplification was performed using Advantage^{TM}-HF 2 PCR (BD Biosciences, Clontech Laboratories, Palo Alto, California) with the conditions: denaturation at 94 ºC for 1 min for 1 cycle, followed by 34 cycles of 94ºC for 30 s, annealing/extension at 62 ºC for 4 min, and a final extension period of 68 ºC for 5 min. Forward and reverse primers

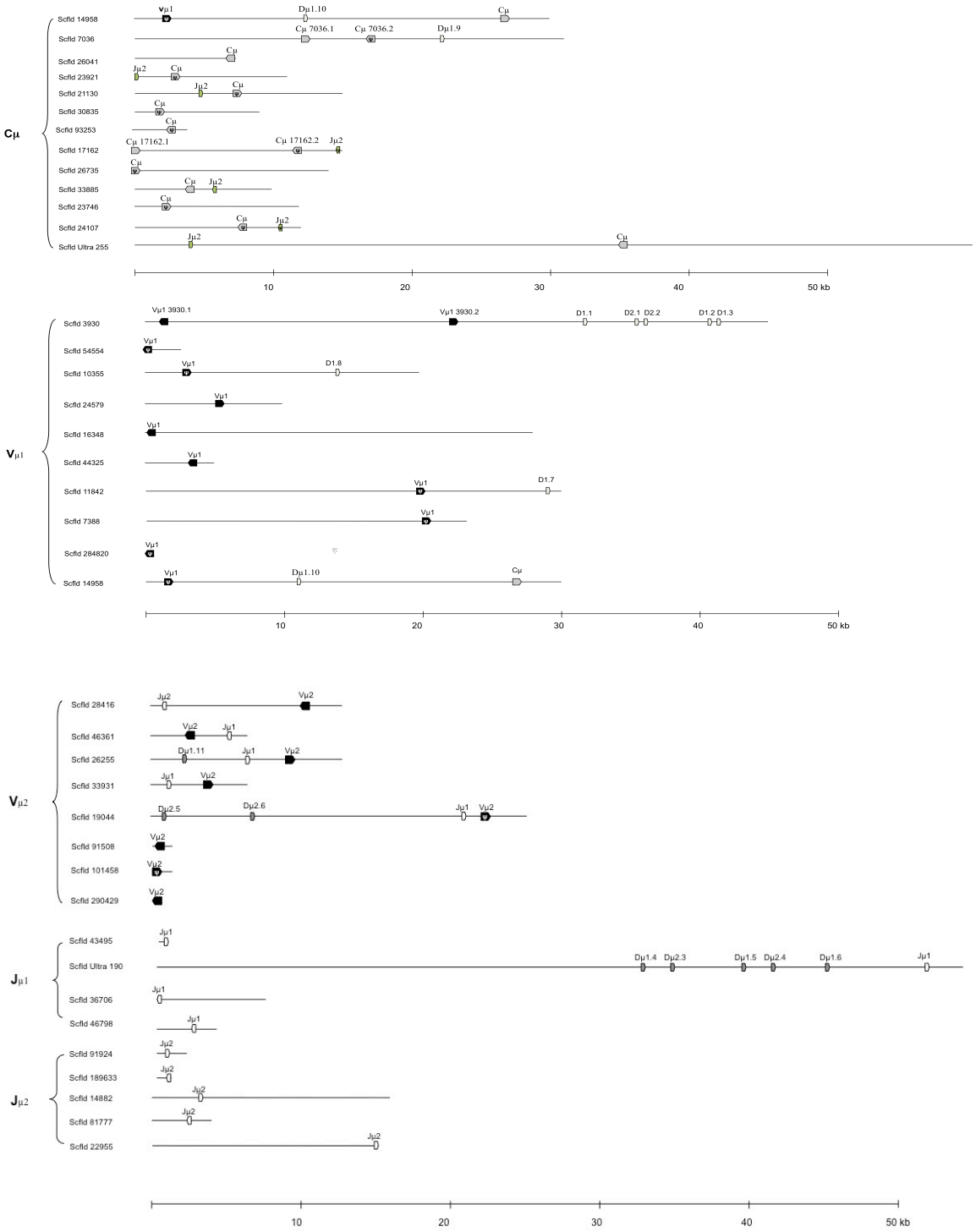


FIGURE 1. Diagram of the scaffolds containing TCRμ genes from the platypus whole genome assembly version 5.0.1. Vμ (black), Dμ (dark grey) Jμ (white) and Cμ (light grey) gene segments are shown in their relative position and transcriptional. Presumptive pseudogenes are indicated with ψ.

Table I. Sequences and description of oligonucleotide primers used.

Sequence (5'-3')	Orientation	Region
CCTGGGCAGTGGGGGCCATGGCCTG	R	Cμ
GGGATAGTAATCTTTCACCAGGCAAG	R	Cμ
AGCAAGTTCAGCCTGGTTAAG	F/R	λ gt10 vector
ATTATGAGTATTTCTTCCAGGGTA	F/R	λ gt10 vector
CCCAACCCATGGTCTTTGTCATG	F	Cμ
GGAACCAGAGCTTCGCTGCTTGCC	F	Cμ
AACCATGCTGGTCCAGGTC	F	5' UTR
CAGGAGGGAAATGATTCAGG	R	3' UTR
CGGAAACAAAAGAAGGCAGA	R	3' UTR
CGTGAAATACTCGGGGGAAT	F	Vμ1
AGGCTCTGCATTGATCTTCG	F	Vμ2

complementary to sequence internal to the platypus Cμ exon were paired with primers in the λgt10 vector used to construct the library to amplify clones containing the 5' and 3' un-translated regions (UTR) (10). This approach generated the partial cDNA sequences analyzed. Full-length platypus TCRμ cDNA sequences were isolated by PCR using primers complementary to 5' and 3' UTR. PCR products were cloned using TOPO TA cloning Kit (Invitrogen, Carsbad, CA) and sequenced using BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). The GenBank accession numbers of the cDNA sequences described here are: clone 21, GU458338; clone 26, GU458339; clone 2.22, GU458341; clone 3815, GU475137; clone 1951, GU475138; clone 1953, GU475139; clone 1954, GU475140; clone 1955, GU475141; clone 4951, GU475142; clone 4942, GU475143; clone 786, GU475144; clone 6, GU458343; clone 17, GU475135; clone 2.34, GU458340; clone 10 GU264000; clone 36, GU475136; clone 4966, GU475145; clone 1.22, GU458342.

Phylogenetic analysis

Phylogenetic analyses were performed on nucleotide alignments using the MEGA4 program (11) with UPGMA (Unweighted Pair Group with Arithmetic mean), MP (Maximum parsimony), Neighbor-Joining (NJ) and ME (Minimum Evolution) methods. Amino acid translations were first aligned to establish gapping and then converted back to nucleotide using the BioEdit program (12).

The GenBank accession numbers of the sequences used in the phylogenetic analyses of TCRμ C and V region sequences were: Cβ sequences are Echidna, AY423735; Platypus,

XM_001509180; Opossum, AY014507; Human, AF043178; Mouse, FJ188408. Cγ sequences are Opossum, DQ499632; Platypus, DQ011295; Human, X15019; Mouse, X03802. Cα sequences are Echidna, DQ011301; Platypus, XM_001507799; Opossum, AY014504; Human, FJ79357; Mouse, DQ186679. Cδ sequences are Platypus, XM_001516959; Human, M21624; Mouse, M37694; Bandicoot, AY955295; Opossum, XP_001379771; Wallaby, AY238447; Frog, GQ262033 and GQ262033; Chicken, XM_423780. Cμ sequences are Wallaby, AY956350; Bandicoot, AY955293; Opossum Cμ sequences are from MonDom5 scaffold 3.430000001-435000000 (13). The sequences of platypus Cμ used in the alignment are from platypus assembly version 5.0.1 and scaffold locations are presented in Fig. 1. Wallaby Vδ48, AY238448; Wallaby Vδ51, AY238451; Bandicoot Vδ46, DQ076246; Cattle Vδ13, D16113; Human Vα96, Z14996; Human Vα34, AB360834; Human Vβ04, M27904; Mouse Vβ16 M15616; Cattle Vβ19, D90129; Rabbit Vβ19, D17419; Sheep Vβ11, AF030011; Human Vγ29, M13429; Mouse Vγ38 , M13338; Cattle Vγ 88, U73188; Sheep Vγ98, Z12998; Platypus Vγ95, DQ011295; Platypus Vγ19, DQ011319; Shark NAR62, AY114762; Shark NAR78, AY114978; Shark NAR82, AY261682; Shark NAR60, EU213060; Shark TNAR05, DQ022705; Shark TNAR88, DQ022688; Shark TNAR10, DQ022710; Opossum Vμ sequences are DQ979402, DQ979398, EF503722, EF5037719, DQ979397, DQ979396, EF503721, EF503718. The sequences of platypus Vμ used in the alignment are from platypus assembly version 5.0.1 and locations are presented in Fig. 1. Frog VHδ sequences are GQ262028, GQ262032, GQ232013; IgVH sequences are: possum VH50, AAL87470; possum VH1, AAL87474; bandicoot VH5.1, AY586158; Opossum VH sequences are from MonDom5 scaffold 1.295000001-300000000 (12). Mouse VH3660,

K01569; mouse VH3609N, X55935; mouse VHDNA4, M20829; mouse VHJ588, Z37145; mouse VHJ606, X03398; mouseVHQ52, M27021; mouse VHS107, J00538; mouse VHSM7 M31285; mouse VH11, Y00743; mouse VH98, AJ851868. Human VH sequences were obtained from the VBASE database. Pig VH3, U15194; cattle VH, AF015505; sheep VH, Z49180; echidna VH7g, AY101438; echidna VH8g, AY101439; echidna VH51g, AY101442; platypus VH29, AF381294; platypus VH26, AF381293; platypus VH3, AF381314; platypus VH53, AF381304.

Results

Identification of a platypus TCRμ homologue.

Fifteen gene sequences with similarity to opossum Cμ were identified in the platypus whole genome assembly (14). Searching the unassembled, raw trace sequences from the platypus whole genome shotgun sequence did not uncover any additional genes with homology to opossum Cμ. Six of these contained complete open reading frames (ORF) and were used in all subsequent analyses (Figs. 1 and 2). When compared to opossum Cμ and conventional TCR C genes from a variety of mammals, the platypus sequences had greatest nucleotide identity to opossum Cμ (Table II, Fig. 2). Included in these analyses was the single copy conventional platypus TCRδ C gene which is located on scaffold 588 in the genome assembly, separate from any of the TCRμ related genes (Table II, Fig. 3 and data not shown). Phylogenetic analyses using several models for tree reconstruction result in the platypus and marsupial Cμ together forming a well-supported monophyletic clade consistent with having identified the platypus TCRμ homologue (Fig. 3).

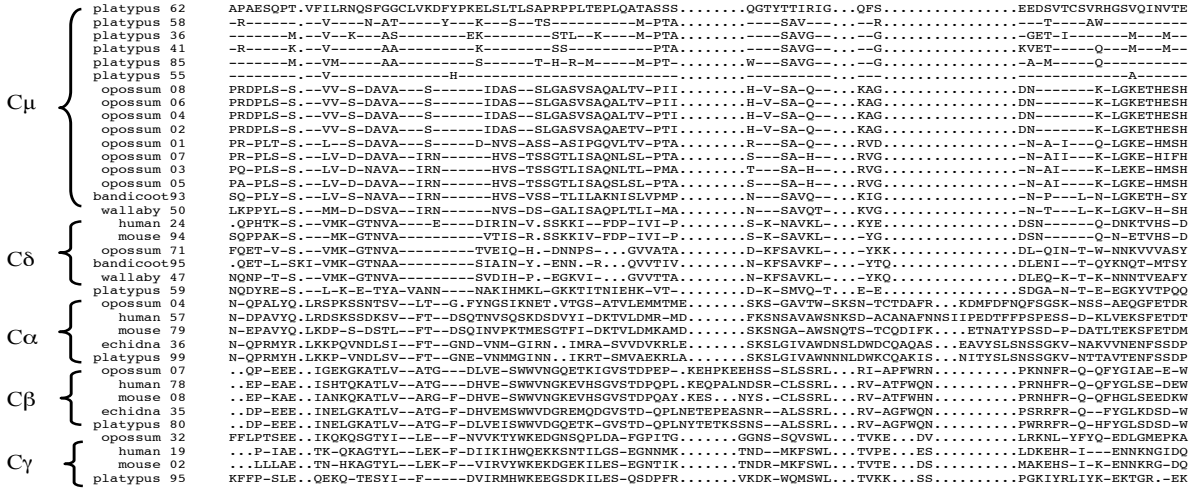


FIGURE 2. Alignment of the predicted amino acid sequences of the six functional platypus Cμ identified in the whole genome assembly with their homologues from marsupials and conventional TCRs from a variety of mammalian species. Dashes indicate identity with the first sequence and dots indicate gaps inserted to generate the alignment. Conserved cysteines involved in intra-domain disulfide bonds are indicated in shade.

Table II. Comparison of platypus Cm with opossum Cm and conventional mammalian TCR C regions

	Platypus Cμ (N=6)	Platypus Cδ (N=1)	Opossum Cμ (N=8)	Cδ[4] (N=5)	Cα[5] (N=5)	Cβ[6] (N=5)	Cγ[7] (N=4)
Platypus Cm (N=6[1])	80-98[2](84[3])	43-47(44)	50-56(52)	41-54(50)	21-26(24)	25-32(29)	27-33(31)
Platypus Cδ (N=1)	43-47(44)	100	43-47(45)	46-53(50)	26-30(28)	29-32(30)	29-33(32)
Opossum Cμ (N=8)	50-56(52)	43-47(45)	75-96(83)	41-54(48)	21-30(25)	26-34(30)	26-33(29)
Cδ (N=5)	41-54(50)	46-53(50)	41-54(48)	55-83(67)	21-30(25)	24-31(27)	28-34(31)
Cα (N=5)	21-26(24)	26-30(28)	21-30(25)	21-30(25)	45-87(54)	22-33(27)	24-33(27)
Cβ (N=5)	25-32(29)	29-32(30)	26-34(30)	24-31(27)	22-33(27)	63-93(72)	28-36(31)
Cγ (N=4)	27-33(31)	29-33(32)	26-33(29)	28-34(31)	24-33(27)	28-36(31)	48-76(54)

[1] Number of sequences included in the comparison
[2] The range of % nucleotide identity
[3] The mean % nucleotide identity
[4] Cδ sequences of human, mouse, opossum, bandicoot, wallaby
[5] Cα sequences of human, mouse, opossum, echidna, platypus
[6] Cβ sequences of human, mouse, opossum, echidna, platypus
[7] Cγ sequences of human, mouse, opossum, platypus

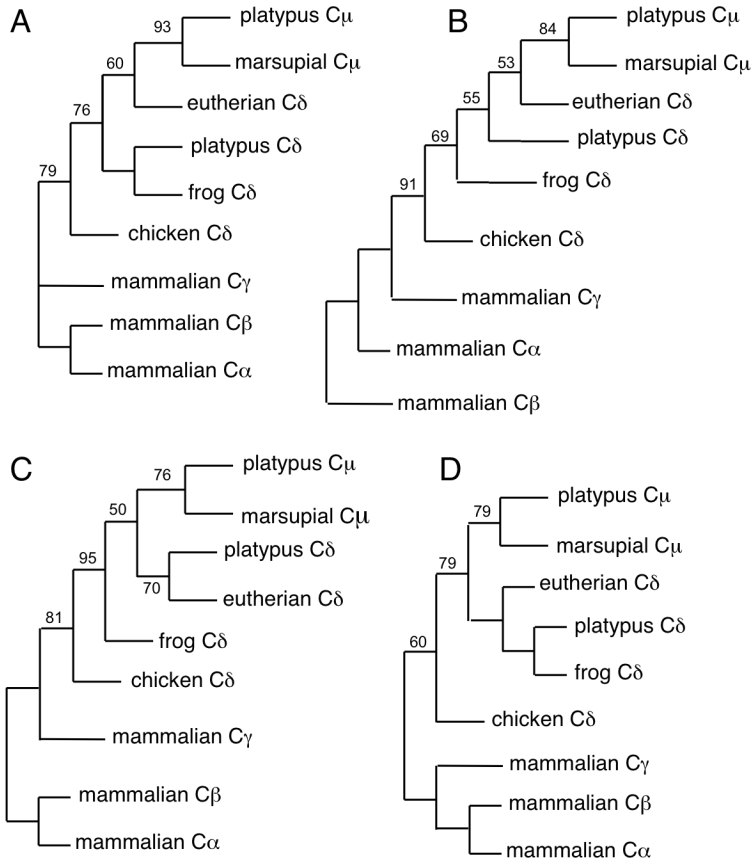


FIGURE 3. Phylogenetic analyses of platypus and marsupial Cμ and C regions from conventional TCR chains. Phylogenetic relationship between Cμ and other conventional TCRs are simplified according to the phylogenetic trees constructed using different methods: *A*, Neighbor-Joining (NJ); *B*, Maximum Parsimony (MP); *C*, Unweighted Pair Group Method with Arithmetic mean (UPGMA); *D*, Minimum Evolution (ME). All phylogenetic analyses are based on nucleotide alignments and branch support is indicated as the percentage of out of 1000 bootstrap replicates.

Platypus TCRμ is transcribed in a double V form.

To investigate the structure of expressed platypus TCRμ, full-length transcripts were isolated from a spleen cDNA library. Transcripts averaged 1300 bp in length, which is longer than a conventional TCR transcript and more similar to the double V encoding opossum TCRμ (Fig. 4). Each encoded a leader (L) peptide followed by two complete V domains, designated V1 and V2 for the 5' (N-terminal) and 3' (C-proximal) domains, respectively. They also contained one C domain along with sequences corresponding to the connecting peptide (CP), transmembrane (TM) and cytoplasmic (CT) regions typical of trans-membrane TCR chains (Fig. 4). The clones encoded conserved residues found in conventional TCR including cysteines forming intra-chain disulfide bonds in the V and C domains as well as inter-chain disulfide bond in the CP (Fig. 4). The framework region (FR) 4 of V1 and V2 contain the sequence YGXG and FXXG, respectively, similar to the conserved FGXG motif in conventional TCR and marsupial TCRμ (4,15,16) (Fig. 4). Also present are two positively charged amino acids (arginine and lysine) in the TM region that, in conventional TCR chains, participate in association with the CD3 signaling complex (17). Comparison to the genomic sequence revealed that the CP is unusual in platypus TCRμ in that it is encoded on two exons, designated CP1 and CP2 with the conserved cysteine in CP2 (Fig. 4). This is unlike the opossum TCRμ and most conventional TCR where the CP is encoded by a single exon (4).

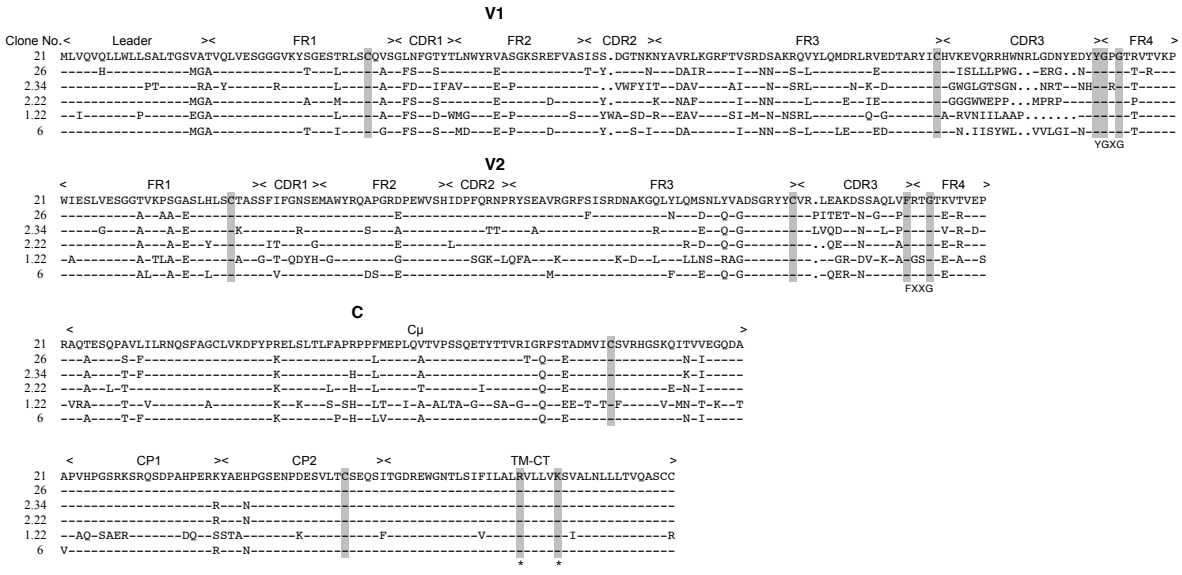


FIGURE 4. Predicted amino acid alignment of full-length platypus TCRμ cDNA clones. Dashes indicate identity and gaps introduced to the alignment are shown as dots. The sequences were divided into the Leader, V1, V2 and C domains. The FR and CDR of the V domains along with the Cm, CP, and TM-CT of C domain were shown above the sequence alignment. Conserved cysteines are shaded. Conserved lysines and arginines are shaded and indicated by *. Conserved residues YGXG and FXXG in FR4 of the V1 and V2 domains, respectively are noted. The borders of CDR and FR were indicated above the sequences.

Both V1 and V2 are encoded by somatically recombined genes.

The germ-line genes encoding the V1 and V2 domains were identified by comparing 18 unique V1 and 16 V2 sequences from both partial and full-length platypus splenic cDNA clones to the genome assembly. V1 and V2 domains share less than 65% nucleotide identity to each other and, by convention, are encoded by different V gene sub-groups designated Vμ1 and Vμ2, respectively. Nine Vμ1 and six Vμ2 genes were identified in the germ-line sequence (Fig. 1). All nine of the Vμ1 genes contained upstream exons encoding a conserved L sequence; however none of the Vμ2 germ-line genes had a L exon (not shown). The sequences corresponding to FR4 in V1 and V2 were also used to identify eight Jμ1 and twelve Jμ2 genes, respectively. Jμ1 and Jμ2 are easily distinguished by length and sequence with Jμ1 being shorter and sharing less than 50% nucleotide identity with Jμ2 genes (Fig. 5). All Vμ and Jμ genes were flanked by conserved RSS, the recognition substrates for the Recombination Activating Gene product (18). The RSS flanking the Vμ and Jμ genes contained 23 and 12 bp spacers, respectively, typical of TCR genes (Fig. 5). In all cDNA sequences analyzed, Vμ1 were recombined to Jμ1 and Vμ2 to Jμ2. These results support both the V1 and V2 domains in platypus TCRμ are encoded by exons that are fragmented in the germ-line DNA and undergo RAG mediated V(D)J recombination.

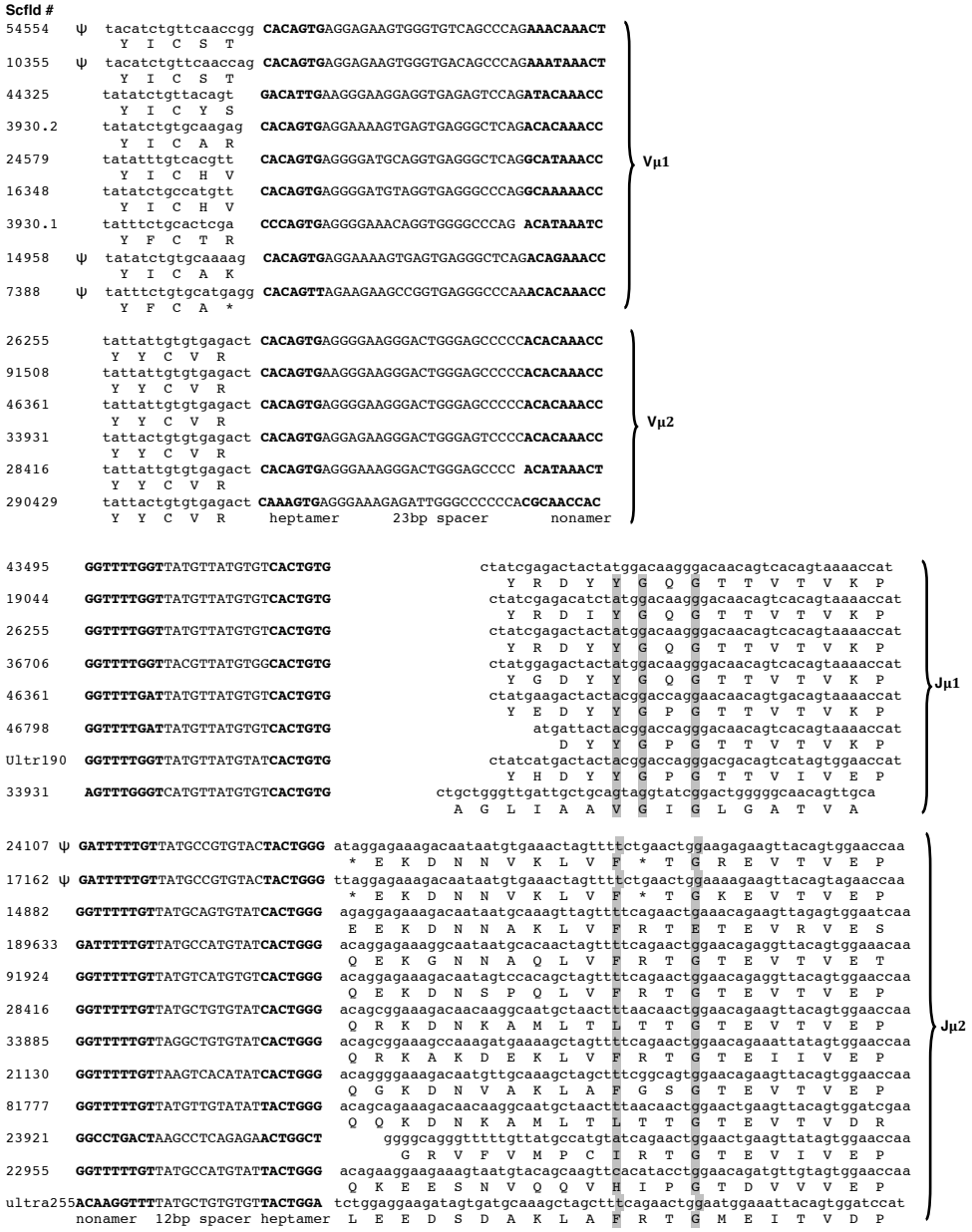


FIGURE 5. Nucleotide sequence and translations of the 3' end of Vm1 and Vm2 gene segments and complete Jm1 and Jm2 gene segments. RSS flanking V and J gene segments in platypus genome are indicated. The scaffolds on which V and J sequences were identified are shown on the left. Pseudogenes are indicated by ψ. Stop codons are indicated as *. Nucleotide sequences of V and J genes are shown in lowercase with amino acid sequences underneath, whereas the RSS sequences are shown in uppercase. Heptamers and nonamers are in bold, and 12 bp or 23 bp spacers are indicated. The YGXG and FXXG conserved motif corresponding to FR4 are shaded.

The sequences corresponding to complementarity determining region 3 (CDR3) differed both in length and diversity between V1 and V2 domains (Fig. 4). The V1 CDR3 are longer and up to 22 codons in length whereas none of the V2 CDR3 exceeded 12 codons. Using the V1 CDR3 sequences identified 35 putative Dμ genes in the platypus genome assembly, all of which were asymmetrically flanked by RSS containing a 12 bp spacer on the 5' side and 23 bp spacer on the 3' side, as is typical of TCR D genes (Fig. 6). Based on length and nucleotide identity the D genes fell into two groups designated Dμ1 and 2. Dμ1 (n = 20) contained coding regions 10 to 13 nucleotides in length while Dμ2 (n = 15) were 18 to 19 nucleotides (Fig. 6). There was greater than 75% nucleotide identity within each group but less than 40% nucleotide identity between Dμ1 and Dμ2 genes. Although Dμ genes could be distinguished in the genomic sequence, individual contributions to the V1 junctions were difficult to establish due to their similarity and short length. Nonetheless it was possible to determine that the Vμ1-Jμ1 junctions contained two, three or four Dμ genes, in roughly a 1:2:1 ratio, similar to the multiple D genes found in opossum TCRμ rearrangements (Fig. 7, Table III). Typical of D gene segments, the Dμ present in V1 junctions were used in multiple reading frames (Fig. 6). The gene segments encoding the V1 domains demonstrated extensive trimming and no evidence of P nucleotide additions, although N nucleotide additions were common (Fig. 7).

A

```
Dμ1.1   GTTTTAGTGTTGGGTACGCATCACTGTG gtaaatatagaac CACAGCGAAACATCCTCATCAGAGCCCTGTGCAAAAACC
                                     V  N  I  E     2
                                      -  I  -  N    1
                                         K  Y  R
Dμ1.2   AGTTTAGTGTTGGCTGTGTATCACTGTG  atgactggagc  CACAGTGAAACTTCCTCGGCAAGACACCGTACAAAAACC
                                      M  T  G       4
                                         -  L  E
                                            D  W  S    3
Dμ1.3   AGTTTAGTGTTGGCTGTGTATCACTGTG  atgactggagc  CACAGTGAAACTTCCTCGGCAAGGCACCGTACAAAAACC
                                      M  T  G       4
                                         -  L  E
                                            D  W  S    3
Dμ1.4   GGTTTAGTGTTTACTGTGTATCACTGTG  acgactacgac  CACAGTGAAACCTCCTTGGCAGGGCCCTGTGCAAAAACC
                                      T  T  T
                                         R  L  R     1
                                            D  Y  D
Dμ1.5   GGTTTAGTGTTGGCTGTGAATCACTGTG atacttacacgag CACGATGAAACTTTCTCGGCAGGACCCTGTGCAAAAACC
                                     I  L  T  R     1
                                        Y  L  H  E
                                           T  Y  T     1
Dμ1.6   GGTTTAGTGTTGGATGTGTATCGCTGTG   atgatggaaa  CACAATGAAACTTCCTTGGCAGAGTCCTGAGCAAAAACC
                                      M  M  E       1
                                         -  W  K
                                            D  G
Dμ1.7   AGTTTAGTGTTCATTGTGTATCACTGTG ctggctgtagaac CACAGTGAAACTTCCTCTGCAAGGCCGTGTTCAAAAACC
                                     L  A  V  E
                                        W  L  -  N   1
                                           G  C  R
Dμ1.8   AATTTAGTGTTGGTCGTGTATCACTGTG gtgggagtggaac CACAGTGAAATTTCCTTTGCTGGGCCCTGGGCAAAACCC
                                     V  G  V  E     7
                                        W  E  W  N   4
                                           G  S  G     1
Dμ1.9   GGTTTAGGGTTGGCCATGCACCACTGTG gtggacataggac CACAGTGAAACATCCTCTGCAAGGCCCTGTGCAAAAATC
                                     V  D  I  G
                                      W  T  -  D
                                        G  H  R      2
Dμ1.10  GTTTTAGTGTTGGCCATACATCACTGTG gtagatagaggac CACAATGAAACGTCCTCGGCAGGGCCTTGTGCAAAAACC
                                     V  D  R  G     1
                                        -  I  E  D   4
                                           R  -  R
Dμ1.11  GGTTTAGTGTTGATGGTGTATCACTGTG attcctggaac   CACAGTGAATCTTCCTCAGCAAGGCTCTGTGCAAAAACC
                                     I  P  G        1
                                        F  L  E      4
                                           S  W  N     1
Dμ1.12  GGTTTAGGGTTGGCCATGCATCACTGTG gtggatataggac CACAGTGAAACGTCCTCTGCAAGGCCCTGTGCAAAAACC
                                     V  D  I  G     1
                                        W  I  -  D   1
                                           G  Y  R     1
Dμ1.13  GGTTTAGTGTTAGCTGTGTATCACTGTG aaggatggagc   CACAGTGAAAGCTCCTCGGCAGGGCCCTGTGTTAAAAAC
                                     K  D  G
                                        R  M  E
                                           G  W  S     2
Dμ1.14  AGTTTAGTGTTCATTGTGTATCACTGTG atgcatgctgtac CACAGTGAAACTTCCTCTGCAAGGAACTGTGCAAAACCC
                                     M  H  A  V
                                     C  M  L  Y
                                      A  C  C
Dμ1.15  AGTTTAGTGTTGGCTGTGTATCACTGTG atgcctggagc   CACAGAGAAACTTCCTCGGCAAGGCACCATGCAAAAACC
                                     M  P  G        2
                                        C  L  E
                                           A  W  S
Dμ1.16  GACTTAGTGTTAGCTGTGTATCACTGTG aaatatggagc   CACGGTGAAACCTCCTTGGCAGGGCCCTGTGTAACAACA
                                     K  Y  G
                                        N  M  E
                                           I  W  S
Dμ1.17  AGTTTAGTGTTCATTGTGTATCACTGTG ctggctgtagtac CACAGTGAAACTTCCTCTGCAAGGCCGTGTTCAAAAACC
                                     L  A  V  V
                                        W  L  -  Y
                                           G  C  S
Dμ1.18  AGTTTAGTGTTCATTGTGTATCACTGTG atgcatggagtac CACAATGAAACTTCCTCTGCAAGGCACTGTGCAAAAACC
                                     M  H  G  V
                                        C  M  E  Y
                                           A  W  S
Dμ1.19  AGTTTAGTGTTCATTGTGTATCACTGTG atgcatgcagtac CACAATGAAACTTCCTCTGCAAGGCACTGTGCAAAAACC
                                     M  H  A  V
                                     C  M  Q  Y
                                      A  C  S
Dμ1.20  GGTTTAGTGTTGGCTATGTATCACTGTG cagactggtat   CACAGTGAAATTTCCTCTGCAGGGCCCCATGCAAAAACC
        nonamer  12bp spacer heptamer Q  T  G      heptamer    23bp spacer        nonamer
                                        R  L  V      1
                                           D  W  Y     1
```

B

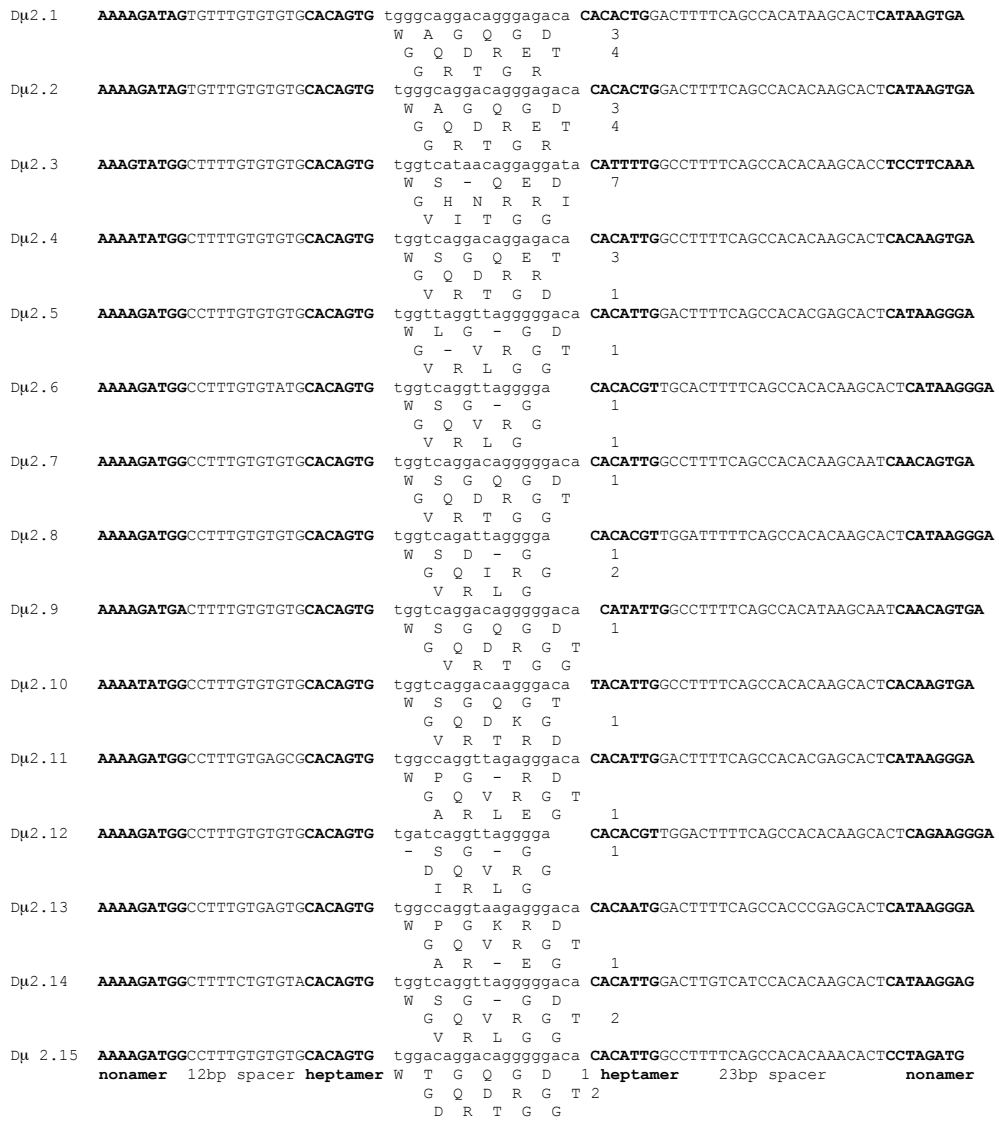


FIGURE 6. Sequences of Dμ (*A*, Dμ1 group and *B*, Dμ2 group) including the flanking RSS and three alternative translations. Dμ coding sequences are shown lowercase and RSS are uppercase. Heptamers and nonamers are in bold. Stop codons are indicated by dashes. For those Dμ that were found used in cDNA sequences, the number of clones found is indicated to the right of the reading frame being used.

A

Clones No.	V1 3' end	N	D	N	D	N	D	N	D	N	J1 5' end
21	CGGTACATCTGCCACGTT	A	AGGAGG	TCCAACGTCGGC	ACTGGA		ATAGG	TTGGG	GGACAA	C	TATGAAGACTACTATGGACCAGGAACAAGA
26	CGGTATATCTGCCACGTT	ATATCGCT	ACTAC	TCCCTTGGGGTT	TGGAAC	G	TGGG		GGAAC		GTGGGAACTACTACGGACCAGGAACAACA
2.22	CGGTATATCTGCCAC	GGGGG	AGGATGG		TGGG		GGAAC	CTCCT	ATGCCT	CGCCCC	TATGAAGACTACTACGGACCAGGAACACCA
3815	CGGTATATCTGCCACGTT		AGACTG	AAGTTGATTTTC	CATAG	TTATGA	GAAC				ACTACTACGGACCAGGAACACCA
1915	CGTTATATCTGCCAC	GTGTT	TGGAGC	ATATGTG	GGAAC	GGGT	TGGG	GCCAT			TATGAAAACTACTACGGACCAGGAACAACG
1953	CGGTATATCTGCCACGTT	AGAGTTCTTT	TGGA		AGACTGGT		TGGG	ATCCTATTAT	GAGGA	C	TACTACGGACCAGGAGCAACA
1954	CGGTATGTCTGTCACGTT	CG	TGGAGC	TTGG	CTTA	GTTAT	GAGGA	C			TATGAGGACTACTACGGACCAGGAACAACA
1955	CGGTATATCTGT	CCTCGA	AGAGGAC	CT	ACTGGA		AGGGGG	CTGGT	TGGG	TCTTGGAGACTC	TATGAAAACCACTACGGACCAGGAACAACA
4951	CGGTATATCTGCCACGTT	AACGT	AATATA	TGTCA	ACTGGT		TGGGC	TCTCT	TGGAT	CT	GAAAACTACTACGGACCAGGAACAACA
4942	CGTTATATGTGCCAC	ACTCCCCACCGTATATTTAAAC	TGGCC	GTCTTTCGCTGGT	ATGAC						AGGAACAACA
786	GTATATCTGCCACGTC	GTCATTATA	TATAGGAGAGGAC	AGGGCT	ATGAC	ACC					CTACTACGGACCAGGAACAACA
6	CGGTACATCTGCCACGTT	AACATTATATCATA	TTGG	CT	GGTC	GTTCTAGGTATA					ACTACTACGGACCAGGAACAACA
17	GATATATCTGCCACG	CCACGGTCTTCGGT	TGGG	GCGCCGT	GTCAGGTT		TGGG	G			ACTACTACGGACGAGGAACAACA
2.34	CGCTACATTTGCCAC	GGT	TGGG	GGT	TGGGC	ACCTCTG	GGAAC	AATCGGACC			CTATGAAAACCACTACGGACGAGGAACAACA
10	CGGTATATCTGTCACGTC	GGTGTCTACAGGA	ACTGGA	ATGGCGTGGGAG	AGGAGA	ACGAGGTTCAATCACAT	GGATGGAG	ACTAACACAT			GACCAGGGACGACAGTCAGAGTGCGACCAT
36	CGGTATATCTGT	GCAAAA	TGGG	TTAAGAAGACGGT	GAAC	TGGCAATACCTGG	GGTC	CTATATTGACGTC			TACGGACCAGGGACAACC
4966	CGATACATCTGTGCAAGA	CAGG	ATGACTGGAG	TC	CTTA	TGCT					ACTACTACGGACCGGGACTACAGTCACAGT
1.22	CGGTACATCTGTGC		GGTCAGG	GTCAATATTATTTTGGCGGC	GCCT						GACTACTACGGACCAGGGACAACA

B

Clones No.	V2 3' end	P	N	D	P	J2 5' end
21	GGCCGCTATTATTGTGTGAGACT		TG			AGGCGAAAGACAGTAGTGCACAGTTAGTTTTCAGAACT
26	GGCCGGTATTATTGTGTG	C	GCCC	GATTA	C	GGAGACAGACAATAGTGGACAGTTACCTTTCAGAACT
2.22	GGCCGGTATTATTGTGTGAGAC					ACAGGAGAAAGACAATAGTGCACAACTGGCTTTCAGAACT
3815	GGCCGGTATTACTGTGTG	C	GCAC	GATTAG		GGAGACAGACAACAGTGCACAGCTACCCTTCAGAACT
1951	GGCCGCTATTATTGTGTGAGACT		GA		C	GAAAGACAATAGTGCGCAGGTGGCTTTCAGAAGT
1953	GGCCGGTATTATTGTGTGAGACT	AG				AGGAGAAAGACAATAGTGCACAGCTAGCTTTCAGAACT
1954	GGCCGGTATTATTGTGTGCGACT					ACAGGAGAAAGACAATAGTGCACGGCTATCTTTCAGAAGT
1955	GGCCGGTATTATTGTGTGAG		GGA			ACAGGAGGAAGACAATACTGCACAGCTACCTTTCAGAACT
4942	GGCCGGTATTATTGTGTGCGACT		GGG			ACAGGAGAAAGACAATAGTGCACGGCTATCTTTCAGAAGT
4951	GGCCGGTATTATTGTGTGAGACT					ACAGGAGAAAGACAGTACTGCACAGGTAGTTTTCAGAACT
6	GGCCGGTATTATTGTGTGAGACT					ACAGGAGAGAGACAATAGTGCACAGCTAGTTTTCAGAACT
17	GGCCGGTATTATTGTGTG	C	GA		GT	ACAGGATAAAGACAATAGTGCGCAGTTAGTTTTCAGAACT
4966	GGCCGGTACTACTGTGTGAGAC					ACAGGGGGCGGACAATGTTGCAGAGGTACCTTTCGGCAGT
1.22	GGCCGGTATTACTGTGTGAGACT	AG				AGGGGAGAGACGATGTTGCAAAGCTAGCTTTCGGCAGT
2.34	GGCCGGTATTATTGTGTGAGACT		TG		TG	CAGGACAAAGACAACAGCGCACTGCTACCTTTCAGAACT
10	GGCCAGTATTATTGTGTGAGAC					ACAGCGGAAAGACACCAAGGCAATGCTGACTTTAACAACT

FIGURE 7. Sequences corresponding to the CDR3 of V (*A*, V1 and *B*, V2) domains from full-length and partial platypus splenic TCRμ cDNAs.

Table III. *The V, D, J and C usage in platypus TCRμ cDNA sequences*

Clone name	Vμ1[#]	Dμ	Dμ	Dμ	Dμ	Jμ1	Vμ2	Dμ?	Jμ2	Cμ
21F	16348	2.3	1.2/1.3	1.9/1.12	1.20	46361	46361		91924	17162.1
26R	16348	1.4	1.8/1.11	1.8/2.1/2.2	1.8/1.11	46361	46361	2.8	91924	17162.1
2.22	16348	1.13	1.8/2.1/2.2	1.8/1.11	1.15	46361	46361		91924	17162.1
3815	16348	1.20	1.9	1.1/1.7/1.8/1.11		46361	46361	2.8	91924	17162.1
1915	16348	1.2/1.3	1.8/1.11	1.8/2.1/2.2		46361	46361		91924	17162.1
1953	16348	1.6/1.8/1.11	1.20	1.8/2.1/2.2	1.10/2.3	46361	46361		91924	17162.1
1954	16348	1.2/1.3	1.5	1.10/2.3		46361	46361		91924	17162.1
1955	16348	1.10	1.2/1.3	2.5	1.8/2.1/2.2	46361	46361		91924	17162.1
4951	16348	1.1	1.2	2.1/2.2	1.12	46361	46361		91924	21130
4942	16348	2.11/2.13	1.2/1.3			46361	46361		91924	17162.1
786*	16348	1.12	1.1			46361				
6	16348	2.12	2.2/2.3			46361	46361		91924	17162.1
17	16348	1.8/2.1/2.2	2.6/2.14	1.8/2.1/2.2		46361	46361		91924	17162.1
2.34	24579	1.8/2.1/2.2	1.8/2.1/2.2	1.8/1.11		46361	46361		91924	17162.1
10	24579	1.2/1.3	2.1/2.2	1.13		36706	28416		28416	26041
36F*	3930.2	1.8/2.1/2.2	1.1/1.7/1.8/1.11	2.3/2.4		46798				
4966	3930	1.2/1.3	1.5			46798	33931		21130	21130
1.22	3930	2.4/2.6	1.15			46798	33931		21130	21130

* 3' partial sequence lacking V2 domain.

[#] The number in the column designates the scaffold on which germ-line gene segment is identified.

In contrast to V1, the CDR3 of 14 of the 16 V2 cDNA sequences could be accounted for entirely by recombination between germ-line Vμ2 and Jμ2 genes, with evidence for P and N nucleotide additions but no Dμ genes being incorporated (Fig. 7, Table III). The remaining two clones contained a short stretch of four or five nucleotides that matches Dμ2.8, and cannot be ruled out as being from a D segment. Whether coincidence or evidence of a D segment is not clear, and is not evident from the genomics where no Dμ has been found between Vμ2 and Jμ2 gene segments (see below). These results are consistent with the longer CDR3 in V1 domains being due to incorporation of multiple D segments and the shorter V2 CDR3 being the result of direct V to J recombination in most if not all junctions.

Platypus TCRμ V genes are related to clan III VH genes.

The relationship Vμ genes have to each other and with V genes from Ig and conventional TCR was investigated by phylogenetic analyses. These analyses included VH from the platypus *IgH* locus (19). The results of these analyses support Vμ1 and Vμ2 each forming their own distinct clades with strong bootstrap support (99-100%) consistent with their designation as separate subgroups (Fig. 8). Furthermore, the platypus Vμ subgroups together form a single clade nested within mammalian clan III VH genes. This is in contrast to the marsupial Vμ (Vμ and Vμj), which are not monophyletic but are closely related to VH (Fig. 8) (4).

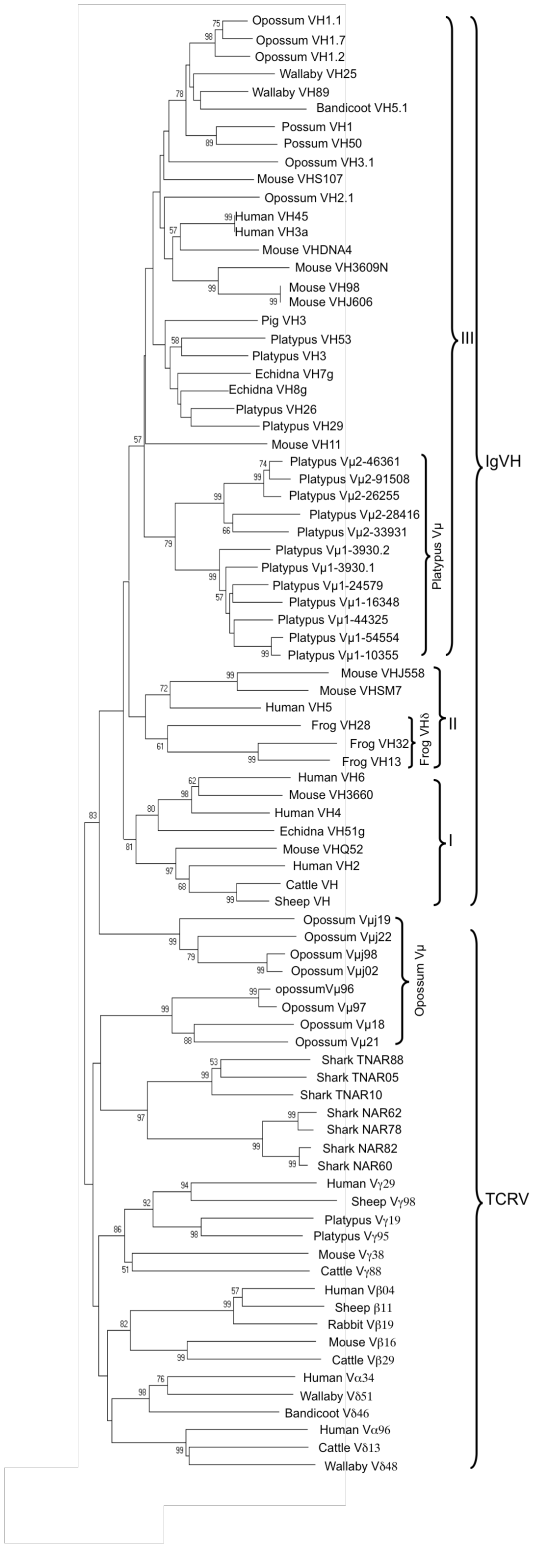


FIGURE 8. Phylogenetic analysis of platypus and marsupial Vμ including V genes from conventional TCR, shark NAR and NAR-TCR and Ig VH. This neighbor-joining tree is based on nucleotide alignments and branch support is indicated as the percentage of out of 1000 bootstrap replicates. Only those nodes with greater than 50% support are indicated. The three major clans of vertebrate VH are indicated by Roman numerals.

Platypus TCRμ genomic organization

The *TCRμ* locus is not fully assembled in the current version of the platypus genome, but rather was scattered on 55 separate scaffolds ranging in length from less than 1 kb up to 64.8 kb (Fig. 1). Seventeen of the 35 Dμ segments were on scaffolds also containing Vμ, Jμ and/or Cμ sequences, supporting their being part of a larger *TCR*μ locus (Fig. 1). Combining the scaffold analyses with the cDNA sequences reveals a minimal model for the organization of the platypus *TCRμ* locus. Three scaffolds contain multiple Dμ either transcriptionally downstream of Vμ1 genes (scaffold 3930) or upstream of a Jμ1 gene (scaffolds Ultra190 and 19044) consistent with the evidence from cDNA sequences having multiple Dμ in the junctions between Vμ1 and Jμ1 genes (Figs. 1 and 9*A*). One scaffold (28416) contains single Vμ2 and Jμ2 genes that correspond to those used in expressed recombinations (Figs. 1 and 9*A*, Table III)**.** However, no Dμ genes were found on this scaffold consistent with the lack of D segments in the majority of Vμ2-Jμ2 junctions (Figs. 1, 7 and 9, Table III).

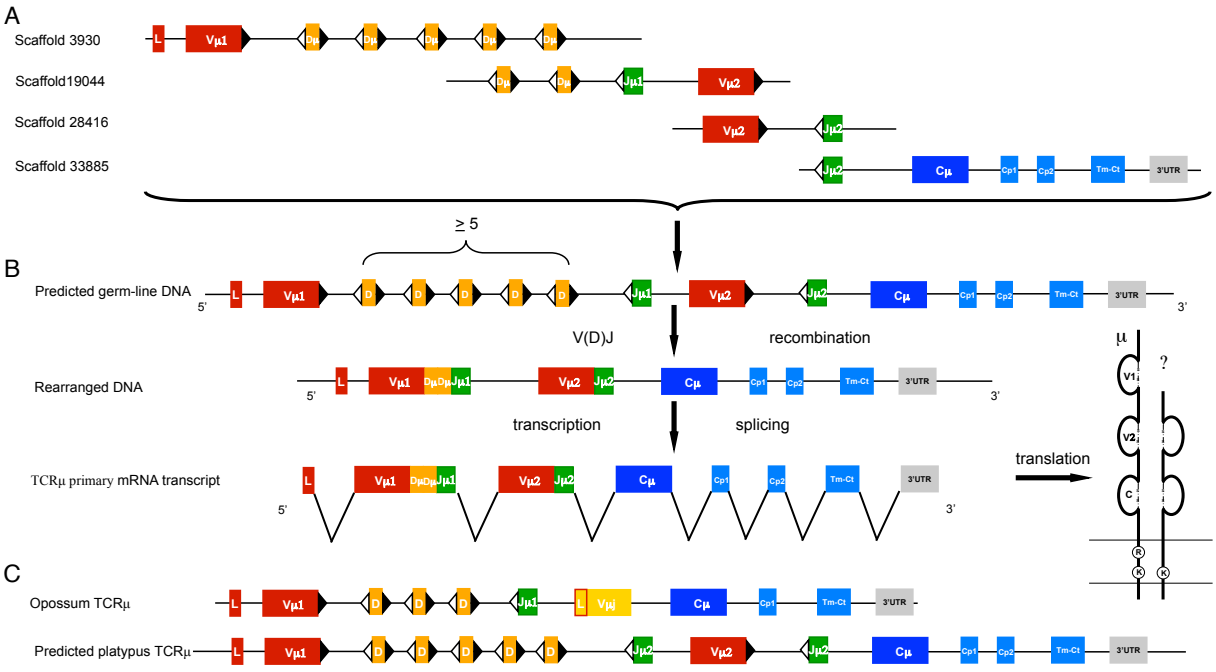


FIGURE 9. Diagrams of the predicted platypus TCRμ gene organization, transcripts, and protein structure. *A*, Representative TCRμ scaffolds containing TCRμ coding sequences. Closed or open triangles flanking the Vμ, Dμ, and Jμ gene segments indicate the presence of 23- or 12-bp spacer RSS, re- spectively. The L sequence, CP, TM-CT, and 39 UTR exons are indicated. *B*, Predicted TCRm germline DNA and rearranged DNA structure and primary TCRm mRNA transcript structure. Conserved R and K residues in the TM region are indicated in the predicted cell surface TCR protein structure. *C*, Comparison of a representative opossum TCRμ cluster with the predicted platypus homolog.

Full-length cDNA clones containing similar or identical Vμ1 sequence also had similar or identical Jμ1, Vμ2, Jμ2 and Cμ (Table III). The most parsimonious explanation for these observations is a cluster organization of platypus TCRμ genes, similar to that found in marsupials (4). In other words, the V, D and J genes encoding V1 domains are upstream of the V and J gene segments encoding V2, followed by Cμ (Fig. 9*B*). Consistent with this prediction, three scaffolds (19044, 26255, and 33931) contain Jμ1 genes upstream of Vμ2 genes and many of the scaffolds containing Cμ genes also contained an upstream Jμ2 (Figs. 1 and 9*A*). A conservative model for the organization of the platypus TCRμ genes is presented in Fig. 9B. The model may be overly conservative since two cDNA clones appeared to use different Vμ1 but the same Jμ1 while two others appeared to use the same Vμ1 recombined to two different Jμ1 (compare clones 2.34, 10 and 17 in Table III). These results imply there may be multiple Vμ1 and Jμ1 in some clusters, or alternatively may be due to trans-cluster recombination as has been found for both opossum TCRμ and shark TCRδ genes (4, 20).

To estimate the possible number of TCRμ clusters, the number of unique Cμ sequences that could be isolated from an individual platypus was determined. PCR was performed on genomic DNA from a single platypus using primers designed to amplify all 15 Cμ identified in the genome assembly. Twenty individual clones were sequenced and yielded nine distinct Cμ sequences consistent with at least five Cμ exons per haploid platypus genome (not shown). This number is slightly lower but not significantly different from what would be predicted from the platypus whole genome sequence where 15 different

Cμ were identified or a minimum of eight per haploid genome. Whether this is an artifact of the assembly or normal platypus variation remains to be determined.

Discussion

The discovery of a platypus TCRμ homologue confirms that this unconventional TCR locus is not unique to marsupials but rather it is ancient in the mammalian lineage and appeared prior to the divergence of the prototherian (monotremes) and therian (marsupial and placental) mammals more than 165 MYA (7). TCRμ was clearly retained in the marsupial lineage and, therefore, would have been present in the last common ancestor of marsupials and placental mammals. However, no TCRμ homologue has been identified in placental mammals, consistent with gene loss in this lineage (2). Furthermore, a TCRμ homologue has yet to be found in the available avian, reptilian, and amphibian genomes, consistent with its appearance in the synapsids (mammals and their extinct relatives) after their divergence from the diapsids (birds and reptiles) 310 MYA (2, 21). This conclusion is also consistent with phylogenetic analyses of TCRμ C region genes published previously, where marsupial Cμ appears to diverge from Cδ after the split between mammals and birds (4).

The most distinctive feature common to both marsupial and platypus TCRμ is their transcription in a form predicted to encode three extra-cellular Ig domains (V-V-C) instead of the conventional two domains (V-C). TCR with this characteristic have only been described in one other vertebrate lineage, the cartilaginous fish. Both the elasmobranchs (sharks, rays, and skates) and the holocephalins (ratfish) use an isoform of TCRδ, called NAR-TCR, that also has a double V expressed with a conventional Cδ (22).

There are a number of common characteristics shared between mammalian TCRμ and shark NAR-TCR, as well as distinctive differences (Table IV). In both platypus TCRμ and NAR-TCR the exons encoding both V domains require somatic DNA recombination to be assembled (22). The supporting or V2 domains in NAR-TCR are encoded by a dedicated subset of Vδ gene segments that, like the platypus Vμ2, lack L sequences and would be unable to encode the N-terminus of an extra-cellular protein (22). This is different, however, in marsupials where the exon encoding the V2 domain, called Vμj, is pre-assembled as a germ-line joined gene and contains a L sequence that is contiguous with the exon encoding the extra-cellular V domain (Fig. 9*C*) (4). In the case of marsupial TCRμ this L sequence is left out of the Vμj exon in the mature mRNA due to a canonical RNA splice site at the junctions between the L and V sequences (2, 4). This arrangement makes it possible to transcribe a two-domain form of marsupial TCRμ that contains only Vμj and C region. Indeed, such transcripts are found in the opossum thymus, however, are rare in peripheral lymphoid tissues, leading to the current working hypothesis that it is the double-V form that is the mature, functional chain (4). Furthermore, in the opossum, *Monodelphis domestica,* there are eight tandem clusters of TCRμ genes and in six of these the Vμj L sequences contain mutations rendering them non-functional (2, 4). Therefore, while the shark and platypus have fully deleted the L sequence of the supporting V, the L sequences in marsupials are apparently degenerating due to lack of use.

Table IV. Comparison of the features of TCRμ, shark NAR TCR, and mammalian conventional TCRα/δ.

Locus	Model Species	C gene	Double V	Somatic recombination		Leader sequence		No. of D segments used		CDR3 length[$] (mean)		V nature		Ref
				N-terminal	C-proximal	N-terminal	C-proximal	N-terminal	C-proximal	N-terminal	C-proximal	N-terminal	C-proximal	
TCRμ	Platypus	Cμ	Yes	Yes	Yes	Yes	No	2-4	0?	9-22(14)	9-12(11)	VH Clan III	VH Clan III	herein
	Opossum	Cμ	Yes	Yes	No[§]	Yes	Yes[#]	1-3	NA*	8-29(17)	11	VH-related	VH-related	4
NAR-TCR	Nurse shark	Cδ	Yes	Yes	Yes	Yes	No	1	1-2	9-25(16)	9-27(16)	V-NAR	Vδ	22
TCRα/δ	Xenopus	Cδ1	No	Yes	NA	Yes	NA	1-2	NA	7-20(13)	NA	Vα, Vδ, VHδ (VH Clan II)	NA	23
TCRα/δ	Human	Cδ	No	Yes	NA	Yes	NA	2-3	NA	8-12(15)	NA	Vα, Vδ	NA	24
TCRα/δ	Mouse	Cδ	No	Yes	NA	Yes	NA	2	NA	6-19(13)	NA	Vα, Vδ	NA	24

[$] Range in codons

[§] The C proximal V in marsupial TCRμ is a germ-line joined V

[#] Fused to the V domain exon as the result of retrotransposition

* Not applicable

Both TCRμ and NAR-TCR utilize V domains more similar to antibody V genes than conventional TCR V genes. The N-terminal V domains in NAR-TCR are related to V used in IgNAR, which are light-chainless antibodies unique to cartilaginous fishes (22, 25). As already described, the second V in NAR-TCR is a Vδ gene, making the NAR-TCR appear to be a hybrid between IgNAR and TCRδ (22). In contrast, the genes used to encode both V1 and V2 domains in platypus TCRμ are indistinguishable from mammalian clan III Ig VH genes and unrelated to NAR V genes. Marsupial Vμ and Vμj on the other hand are somewhat intermediary. Vμj are more similar to Ig VH, but do not fall within the three traditional mammalian VH clans, and Vμ appears to be more related to NAR V genes, although this latter relationship is only weakly supported in phylogenetic analyses (Fig. 8).

The current model for the structure of NAR-TCR is an unpaired N-terminal domain, much like the V-NAR domain in IgNAR, binding antigens as a single domain (22, 25). This Ag binding is similar to that which has been described for single V domain IgNAR antibodies in sharks and light-chainless IgG in camels (26, 27). It seems likely that TCRμ is structured similarly to NAR-TCR, with a single, unpaired N-terminal V domain capable of binding antigen directly. Based on conserved residues, including cysteines TCRμ is predicted to form a heterodimer with another TCR chain (4). However, since no other TCR related genes encoding a three-domain chain have been found in the marsupial genome it is predicted that the partner is a conventional two domain TCR chain, likely TCRγ, leaving the N-terminal domain unpaired (2).

The common characteristics found in mammalian TCRμ and shark NAR-TCR raise the question of whether these features are due to homology by descent or convergent evolution. An argument could be made that the evolutionary distance between sharks and mammals is sufficiently vast, and the differences between TCRμ and NAR-TCR extensive enough that each evolved independently and appear analogous due to convergence on a common structure and function. This could imply a common evolutionary pressure shared between cartilaginous fish and early mammals to have T cells capable of binding Ag directly using single domain binding sites.

Phylogenetic analyses of platypus and marsupial TCRμ C region support that they are orthologous genes that would have been found in a last common ancestor of the three living mammalian lineages. However, following the divergence of the oviparous monotremes from the viviparous marsupials and placental mammals, TCRμ appears to have followed different evolutionary paths. In the placental mammals it was lost altogether (2). As discussed earlier, in the marsupials the genes encoding the V2 domain appear to have been replaced in the germ-line by a pre-joined V gene, most likely via retro-transposition (4). This novel marsupial adaptation is consistent with the V2 domains serving strictly supporting roles rather than being Ag-binding and, therefore, requiring little or no clonal variation. In the platypus the TCRμ V2 domain is encoded by somatically recombined genes, but variation remains restricted through limited junctional diversity, with no D segments and few N or P additions in the V-J junctions. Comparisons of the length of the CDR3 region in the platypus and marsupial V2 domains, where they are both relatively short, suggests that D segments, if they were ever

present, were deleted early in the evolution of TCRμ prior to the divergence of prototherians and therians (4). The mean codon length of the platypus V2 CDR3 is the same (n = 11) as that found in the germ-line joined marsupial Vμj genes (Table IV). In contrast the V1 domains of both platypus and opossum TCRμ have comparatively longer and more diverse CDR3 due to the incorporation of multiple D segments during V(D)J recombination in both species (4).

The lack of an intron separating the L from the V in the Vμj exon is evidence of retro-transposition in the evolution of TCRμ in marsupials (4). In other words, Vμj is a functional, partially processed gene. The insertion of joined V genes into the germ-line by retro-transposition would require co-existing retro-elements in the genome and one noteworthy distinction between the opossum and the platypus genomes is the abundance of retro-elements. The opossum has among the highest percentage of retro-elements of any vertebrate genome sequenced (28). In contrast, monotremes are relatively devoid of retro-elements (14, 29). Whether this extreme difference contributed to the evolution of opossum and platypus TCRμ is not known. Furthermore, this explanation is not fully satisfying since processed pseudogenes have been found in the platypus and echidna genomes, consistent with retro-transposition having occurred sometime in the past for some monotreme genes (10).

Phylogenetic analyses support TCRμ being related to and likely derived from a TCRδ ancestor (4, 5). As stated earlier, if TCRμ evolved from a duplication of TCRδ genes it likely occurred after the separation of mammals from birds and reptiles (4). However,

some insight into the origins of TCRμ may come from recent work on the genetics of amphibian TCRδ chains (23). The *TCRα/δ* locus in the frog *Xenopus tropicalis* contains two Cδ genes, one of which, Cδ1, is expressed with V genes called VHδ. These frog VHδ that are indistinguishable from clan II Ig VH genes and, although the *X. tropicalis TCRα/δ* and *Igh* loci are closely linked, the VHδ genes appear to be dedicated for use in TCRδ chains and are not used in IgH chains (23). This close linkage, however, may have facilitated insertion of VH genes among the TCRδ genes in amphibians. The region of the frog *TCRα/δ* locus containing Cδ1 and multiple VHδ genes is distinct and in an inverted transcriptional orientation from the rest of the *TCRα/δ* genes, functioning almost as a separate mini-cluster (23). Amphibians, therefore, appear to be another vertebrate lineage that uses TCRδ chains containing antibody-like V genes. Unlike TCRμ and NAR-TCR, frog TCRδ chains are not expressed with two V domains, however. Rather, *X. tropicalis* TCRδ chains using VHδ are structured like conventional two-domain TCR chains.

It is possible, and seems likely, that the TCRμ locus evolved from genome duplication and translocation of an ancestral region of the *TCRα/δ* locus similar to the Cδ1 region in frogs. Indeed, the discovery of VH genes in the *X. tropicalis TCRα/δ* locus is consistent with their presence in the *TCRδ* locus prior to the evolution of TCRμ. Internal duplications of clusters of V, D, and J segments within the *TCRμ* locus, as hypothesized previously, would then give rise to the double V organization in mammals (2). What remains puzzling is the variation in the source of VH genes used in each lineage. The VHδ in *X. tropicalis* are apparently derived from clan II VH, the platypus Vμ genes are

clan III VH, and although the marsupial Vμ genes are more closely related to VH than TCR V genes but fall outside the clan I, II, and III designation. These observations suggest that the V genes used in TCRδ or TCRμ chains have been replaced over time with different VH lineages, even within the mammals. If the platypus *TCRμ* locus is indeed organized as tandem clusters similar to what has been shown in opossum (4), such gene clusters may facilitate gene replacement and duplication that is not easily achieved by the translocon organization of the conventional TCR genes.

The lack of TCRμ in commonly studied mammals such as humans and mice no doubt contributed to it remaining undiscovered for nearly a quarter century following that of the conventional TCRα, β, γ, and δ (4, 28-31). Determining why placental mammals may have lost this TCR chain will require first determining what function(s) TCRμ$^+$ T cells perform in those species where they are found.

References

1. Rast, J.P., M. K. Anderson, S. J. Strong, C. Luer, R. T. Litman, G. W. Litman. 1997. Alpha, beta, gamma and delta T cell antigen receptor genes arose early in vertebrate phylogeny. *Immunity* 6: 1–11.

2. Parra, Z. E., M. L. Baker, J. Hathaway, A. M. Lopez, J. Trujillo, A. Sharp, R. D. Miller. 2008. Comparative genomic analysis and evolution of the T cell receptor loci in the opossum *Monodelphis domestica. BMC Genomics* 9: 111.

3. Chen, H., S. Kshirsagar, I. Jensen, K. Lau, R. Covarrubias, S. F. Schluter, J. J. Marchalonis. 2009. Characterization of arrangement and expression of the T cell receptor gamma locus in the sandbar shark. *Proc. Natl. Acad. Sci. USA* 106: 8591–8596.

4. Parra Z. E., M. L. Baker, R. Schwarz, J. E. Deakin, K. Lindblad-Toh, and R. D. Miller. 2007. Discovery of a new T cell receptor in marsupials. *Proc. Natl. Acad. Sci. USA* 104: 9776–9781.

5. Baker, M. L., A. K. Osterman and S. Brumburgh. 2005. Divergent T cell receptor delta chains from marsupials. *Immunogenetics* 57: 665–673.

6. Parra, Z. E., M. L. Baker, A. M. Lopez, J. Trujillo, J. M. Volpe and R. D. Miller. 2009. TCRμ recombination and transcription relative to the conventional TCR during postnatal development in opossums. *J. Immunol.* 182: 154–163.

7. Bininda-Emonds, O. R., M. Cardillo, K. E. Jones, R. D. MacPhee, R. M. Beck, R. Grenyer, S. A. Price, R. A. Vos, J. L. Gittleman and A. Purvis. 2007. The delayed rise of present-day mammals. *Nature* 446: 507–512.

8. Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403–410.

9. Vernersson, M., M. Aveskogh, B. Munday and L. Hellman. 2002. Evidence for an early appearance of modern post-switch immunoglobulin isotypes in mammalian evolution (II); cloning of IgE, IgG1 and IgG2 from a monotreme, the duck-billed platypus, *Ornithorhynchus anatinus*. *Eur. J. Immunol*. 32: 2145–2155.

10. Miska, K. B., G. A. Harrison, L. Hellman and R. D. Miller. 2002. The Major Histocompatibility Complex in monotremes: An analysis of the evolution of Mhc class I genes across all three mammalian subclasses. *Immunogenetics* 54: 381–393.

11. Tamura, K., J. Dudley, M. Nei and S. Kumar. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596–1599.

12. Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acid Symp. Ser*. 41: 95–98 .

13. Wang, X., J. J. Olp and R. D. Miller. 2009. On the genomics of immunoglobulins in the gray, short-tailed opossum *Monodelphis domestica*. *Immunogenetics* 61: 581–596.

14. Warren, W. C., L. W. Hillier, J. A. Marshall Graves, E. Birney, C. P. Ponting, F. Grutzner, K. Belov, W. Miller, L. Clarke, A. T. Chinwalla, et al 2008. Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453: 175–183 (2008).

15. Belov, K., R. D. Miller, A. Ilijeski, L. Hellman and G. A. Harrison. 2004. Isolation of monotreme T cell receptor alpha and beta chains. *Immunogenetics* 56: 164–169.

16. Parra, Z. E., T. Arnold, M. A. Nowak, L. Hellman and R. D. Miller. 2006. TCR gamma chain diversity in the spleen of the duckbill platypus (*Ornithorhynchus anatinus*). *Dev. Comp. Immunol.* 30: 699–710.

17. Campbell, K. S., B. T. Backstrom, G. Tiefenthaler and E. Palmer. 1994. CART: A conserved antigen receptor transmembrane motif. *Semin. Immunol*. 6: 393–410.

18. Schatz, D. G., M. A. Oettinger and D. Baltimore. 1989. The V(D)J recombination activating gene, RAG-1. *Cell* 59: 1035–1048.

19. Gambon-Deza, F., C. S. Espinel and S. Magadan-Mompo. 2009. The immunoglobulin heavy chain locus in the platypus (*Ornithorhynchus anatinus*). *Mol. Immunol.* 46: 2515–2523.

20. Criscitiello, M. F., Y. Ohta, M. Saltis, E. C. McKinney and M. F. Flajnik. 2010. Evolutionarily Conserved TCR Binding Sites, Identification of T Cells in Primary Lymphoid Tissues, and Surprising Trans-Rearrangements in Nurse Shark. *J. Immunol.* 184: 6950–6960.

21. Kumar, S. and S. B. Hedges. 1998. A molecular timescale for vertebrate evolution. *Nature* 392: 917–920.

22. Criscitiello, M. F., M. Saltis and M. F. Flajnik. 2006. An evolutionarily mobile antigen receptor variable region gene: doubly rearranging NAR-TcR genes in sharks. *Proc. Natl. Acad. Sci. USA* 103: 5036–5041.

23. Parra, Z. E., K. Ohta, M. F. Criscitiello, M. F. Flajnik and R. D. Miller. 2010. The dynamic TCRδ: TCRδ chains in the amphibian *Xenopus tropicalis* utilize antibody-like V genes. *Eur. J. Immunol.* 40: 2319–2329.

24. Rock, E. P., P. R. Sibbald, M. M. Davis and Y. H. Chien. 1994. CDR3 length in antigen-specific immune receptors. *J. Exp. Med.* 179: 323–328.

25. Greenberg, A. S., D. Avila, M. Hughes, A. Hughes, E. C. McKinney, M. F. Flajnik. 1995. A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature* 374: 168–173.

26. Stanfield, R. L., H. Dooley, M. F. Flajnik and I. A. Wilson. 2004. Crystal structure of a shark single-domain antibody V region in complex with lysozyme. *Science* 305: 1770–1773.

27. Desmyter, A., T. R. Transue, M. A. Ghahroudi, M. H. Thi, F. Poortmans, R. Hamers, S. Muyldermans, and L. Wyns. 1996. Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. *Nature Struct. Biol.* 3: 803-811.

28. Mikkelsen, T. J., M. J. Wakefield, B. Aken, C. T. Amemiya, J. L. Chang, S. Duke, M. Garber, A. J. Gentles, L. Goodstadt, A. Heger, et al 2007. Genome of the marsupial *Monodelphis domestica* reveals adaptive turnover of coding and non-coding sequences. *Nature* 447: 167–177.

29. Kordis, D., N. Lovsin and F. Gubensek. 2006. Phylogenomic analysis of the L1 retrotransposons in deuterostomia. *Syst. Biol.* 55: 886–901.

30. Hedrick, S. M., D. I. Cohen, E. A. Nielsen and M. M. Davis. 1984. Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature* 308: 149–153.

31. Chien, Y. H., N. R. J. Gascoigne, J. Kavaler, N. E. Lee and M. M. Davis. 1984. Somatic recombination in a murine T-cell receptor gene. *Nature* 309: 322–326.

32. Yanagi, Y., Y. Yoshikai, K. Leggett, S. P. Clark, I. Aleksander, T. W. Mak. 1984. A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. *Nature* 308: 145–149.

33. Brenner, M. B., J. Mclean, D. P Dialynas, J. L. Strominger, J. A. Smith, F. L. Owen, J. G. Seidman, S. Ip, F. Rosen, and M. S. Krangel. 1986. Identification of a putative second T-cell receptor. *Nature* 322: 145–149.

APPENDIX I

Supplemenary Tables in Chapter 2

Supplementary Table 1. Locations of IGH gene segments in the MonDom5 assembly

Gene Name[1]	L1 start[2]	L1 end[3]	L2 start[4]	L2 end[5]	V start[6]	V end[7]	Notes[8]
V1.1	1.297866290	1.297866245	1.297866153	1.297866143	1.297866142	1.297865847	
V1.2ψ	1.297895425	1.297895380	1.297895279	1.297895269	1.297895268	1.297894973	*
V1.3	1.297947225	1.297947180	1.297947081	1.297947071	1.297947070	1.297946775	
V1.4	1.297972350	1.297972305	1.297972205	1.297972195	1.297972194	1.297971902	
V1.5	1.298004750	1.298004705	1.298004606	1.298004596	1.298004595	1.298004297	
V1.6	1.298023399	1.298023354	1.298023257	1.298023247	1.298023246	1.298022951	
V1.7	1.298047416	1.298047371	1.298047267	1.298047257	1.298047256	1.298046961	
V1.8	1.298112543	1.298112498	1.298112406	1.298112396	1.29811.2395	1.298112103	
V1.9	1.298202106	1.298202061	1.298201961	1.298201951	1.298201950	1.298201658	
V1.10ψ	1.298243820	1.298243775	1.298243676	1.298243666	1.298243665	1.298243370	*
V1.11	1.298274315	1.298274270	1.298274170	1.298274160	1.298274159	1.298273864	
V1.12	1.298297609	1.298297564	1.298297472	1.298297462	1.298297461	1.298297166	
V1.13	1.298417720	1.298417675	1.298417583	1.298417573	1.298417572	1.298417280	
V1.14	1.298439107	1.298439062	1.298438963	1.298438953	1.298438952	1.298438657	
V1.15	1.298526433	1.298526388	1.298526296	1.298526286	1.298526285	1.298525990	
V1.16	1.298550969	1.298550924	1.298550834	1.298550824	1.298550823	1.298550528	
V1.17	1.298563476	1.298563431	1.298563330	1.298563320	1.298563319	1.298563024	
V1.18	1.298599469	1.298599424	1.298599332	1.298599322	1.298599321	1.298599026	
V1.19ψ					1.298630243	1.298630183	partial
V1.20ψ	1.298633386	1.298633341	1.298633249	1.298633239	1.298633238	1.298633183	partial
V1.21	1.298644672	1.298644627	1.298644527	1.298644517	1.298644516	1.298644224	
V1.22	1.298710837	1.298710792	1.298710705	1.298710695	1.298710694	1.298710399	
V1.23	1.298719939	1.298719894	1.298719810	1.298719800	1.298719799	1.298719504	
V1.24ψ					1.298747248	1.298747050	partial
V2.1	1.298771689	1.298771644	1.298771562	1.298771552	1.298771551	1.298771250	
V3.1	1.298684128	1.298684083	1.298683877	1.298683867	1.298683866	1.298683528	
V1.Ua	Un43298924	Un43298969	Un43299059	Un43299069	Un43299070	Un43299365	
V1.Ubψ	Un43340245	Un43340290	Un43340391	Un43340401	Un43340402	Un43340697	*
V1.Ucψ	Un43347998	Un43348043	Un43348147	Un43348157	Un43348158	Un43348237	partial
V1.Udψ	Un43357692	Un43357692	Un43357692	Un43357702	Un43357703	Un43357882	partial
V1.Ue	Un43369948	Un43369993	Un43370083	Un43370093	Un43370094	Un43370389	
V1.Uf	Un43404139	Un43404184	Un43404276	Un43404286	Un43404287	Un43404582	
V1.Ug	Un43430016	Un43430061	Un43430161	Un43430171	Un43430172	Un43430467	
V1.Uh	Un79397698	Un79397653	Un79397561	Un79397551	Un79397550	Un79397257	
Exon Name	**Exon start**	**Exon end**					
D1	1.297683938	1.297683925					
D2	1.297684466	1.297684451					
D3	1.297686470	1.297686453					
D4	1.297733979	1.297733947					
D5	1.297745145	1.297745122					
D6	1.297757296	1.297757279					
D7	1.297775770	1.297775742					
D8	1.297896360	1.297896319					
D9	1.298074023	1.298073996					
J1	1.297659653	1.297659603					
J2	1.297660023	1.297659976					
J3ψ	1.297661200	1.297661152					*
J4	1.297837888	1.297837838					

J5	1.297838259	1.297838212	
J6ψ	1.297839797	1.297839750	*
ψMCH1	1.297823382	1.297823304	
ψMCH2	1.297822899	1.297822573	
ψMCH3	1.297821534	1.297821223	
ψMCH4	1.297820300	1.297820674	*
MCH1	1.297651274	1.297650963	
MCH2	1.297650557	1.297650231	
MCH3	1.297649294	1.297648977	
MCH4	1.297648433	1.297648039	
MTM1	1.297644197	1.297644082	
MTM2	1.297642453	1.297642445	
GCH1	1.297446317	1.297446024	
GHinge	1.297444065	1.297444024	
GCH2	1.297442990	1.297442673	
GCH3	1.297441471	1.297441143	
GTM1	1.297438167	1.297438037	
GTM2	1.297437665	1.297437582	
ECH1	1.297391034	1.297390735	
ECH2	1.297390441	1.297390112	
ECH3	1.297389744	1.297389424	
ECH4	1.297388547	1.297388213	
ETM1	1.297384367	1.297384251	
ETM2	1.297383923	1.297383840	
ACH1	1.297362168	1.297361948	
Ahinge	1.297361946	1.297361895	
ACH2	1.297361282	1.297360944	
ACH3	1.297360231	1.297359840	
ATM	1.297353999	1.297353815	

Footnotes:

[1]Name of gene segment

[2]Start of Leader exon

[3]End of Leader Exon

[4]Start of Exon 2, the continuation of the Leader sequence

[5] End of Leader sequence in Exon 2

[6]Start of V gene sequence in Exon 2

[7]End of Exon 2

[8]Description of features including reason for being designated a pseudogene: * means stop codon in reading frame.

Supplementary Table 2. Locations of IGK gene segments in the MonDom5 assembly

Gene Name[1]	L1 start[2]	L1 end[3]	L2 start[4]	L2 end[5]	V start[6]	V end[7]	Notes[8]	
V1.1	1.715785240	1.715785189	1.715784961	1.715784951	1.715784950	1.715784664		R
V1.2	1.715768616	1.715768565	1.715768310	1.715768300	1.715768299	1.715768013		R
V1.3	1.715754635	1.715754584	1.715754325	1.715754315	1.715754314	1.715754028		R
V1.4	1.715734848	1.715734797	1.715734547	1.715734537	1.715734536	1.715734250		R
V1.5	1.715726178	1.715726127	1.715725868	1.715725858	1.715725857	1.715725571		R
V1.6	1.715711070	1.715711019	1.715710804	1.715710794	1.715710793	1.715710507		R
V1.7ψ	1.715703291	1.715703242	1.715702982	1.715702972	1.715702971	1.715702876	partial	R
V1.8	1.715688312	1.715688261	1.715688005	1.715687995	1.715687994	1.715687708		R
V1.9	1.715676406	1.715676355	1.715676096	1.715676086	1.715676085	1.715675799		R
V1.10	1.715669168	1.715669117	1.715668861	1.715668851	1.715668850	1.715668565		R
V1.11	1.715660658	1.715660607	1.715660349	1.715660339	1.715660338	1.715660052		R
V1.12ψ	1.715645655	1.715645605	1.715645391	1.715645381	1.715645380	1.715645306	partial	R
V1.13	1.715635448	1.715635397	1.715635138	1.715635128	1.715635127	1.715634841		R
V1.14	1.715628693	1.715628642	1.715628429	1.715628419	1.715628418	1.715628132		R
V1.15	1.715612241	1.715612190	1.715611977	1.715611967	1.715611966	1.715611680		R
V1.16ψ					1.715596278	1.715596390	partial	
V1.17ψ	1.715577704	1.715577755	1.715578014	1.715578024	1.715578025	1.715578311	*	
V1.18	1.715565426	1.715565477	1.715565690	1.715565700	1.715565701	1.715565987		
V1.19ψ	1.715547378	1.715547327	1.715547080	1.715547070	1.715547069	1.715546983	partial	R
V1.20ψ					1.715546959	1.715546679	partial	R
V1.21	1.715518247	1.715518196	1.715517938	1.715517928	1.715517927	1.715517641		R
V1.22	1.715511127	1.715511076	1.715510863	1.715510853	1.715510852	1.715510566		R
V1.23ψ	1.715504319	1.715504269	1.715504061	1.715504051	1.715504050	1.715503764	*	R
V1.24	1.715487443	1.715487395	1.715487136	1.715487126	1.715487125	1.715486839		R
V1.25	1.715468687	1.715468636	1.715468378	1.715468368	1.715468367	1.715468167	partial	R
V1.26	1.715455437	1.715455386	1.715455127	1.715455108	1.715455107	1.715454872	partial	R
V1.27	1.715445141	1.715445090	1.715444831	1.715444821	1.715444820	1.715444534		R
V1.28ψ			1.715440321	1.715440311	1.715440310	1.715440024	no L1	R
V1.29	1.715430093	1.715430045	1.715429786	1.715429776	1.715429775	1.715429489		R
V1.30	1.715424482	1.715424533	1.715424789	1.715424799	1.715424800	1.715425086		
V1.31ψ	1.715413937	1.715413886	1.715413635	1.715413625	1.715413624	1.715413338	*	R
V1.32	1.715409540	1.715409587	1.715409847	1.715409857	1.715409858	1.715410144		
V2.1	1.713181425	1.713181473	1.713181722	1.713181732	1.713181733	1.713182034	*	
V2.2ψ	1.713164212	1.713164260	1.713164491	1.713164501	1.713164502	1.713164797	*	
V2.3ψ	1.713123199	1.713123251	1.713123494	1.713123504	1.713123505	1.713123805	*	
V2.4ψ			1.713118061	1.713118071	1.713118072	1.713118233	partial	
V2.5	1.713115830	1.713115878	1.713116091	1.713116101	1.713116102	1.713116401		
V2.6	1.713106743	1.713106791	1.713107024	1.713107034	1.713107035	1.713107337		
V2.7	1.713075652	1.713075704	1.713075947	1.713075957	1.713075958	1.713076258		
V2.8ψ			1.713071275	1.713071285	1.713071286	1.713071582	*	
V2.9	1.713068993	1.713069041	1.713069290	1.713069300	1.713069301	1.713069604		
V2.10ψ			1.713037449	1.713037459	1.713037460	1.713037722	partial	
V2.11ψ	1.713033751	1.713033791	1.713034048	1.713034058	1.713034059	1.713034357	*	
V2.12	1.713026911	1.713026959	1.713027206	1.713027216	1.713027217	1.713027520		
V2.13	1.712999060	1.712999108	1.712999358	1.712999368	1.712999369	1.712999672		
V2.14	1.712994875	1.712994923	1.712995299	1.712995309	1.712995310	1.712995613		
V2.15	1.712980627	1.712980675	1.712980829	1.712980839	1.712980840	1.712981140		
V2.16	1.712971637	1.712971685	1.712971915	1.712971925	1.712971926	1.712972219		
V2.17	1.712964028	1.712964076	1.712964325	1.712964335	1.713964336	1.712964639		

V2.18	1.712958639	1.712958687	1.712958843	1.712958853	1.712958854	1.712959157		
V2.19	1.712951688	1.712951736	1.712951892	1.712951902	1.712951903	1.712952209		
V2.20	1.712936975	1.712937023	1.712937252	1.712937262	1.712937263	1.712937558		
V2.21	1.712930672	1.712930720	1.712930969	1.712930979	1.712930980	1.712931283		
V2.22ψ	1.712925366	1.712925414	1.712925570	1.712925580	1.712925581	1.712925879	*	
V2.23ψ			1.712917040	1.712917050	1.712917051	1.712917346	*	
V2.24	1.712900827	1.712900875	1.712901125	1.712901135	1.712901136	1.712901439		
V2.25ψ	1.712894859	1.712894907	1.712895063	1.712895073	1.712895074	1.712895376	*	
V2.26	1.712886151	1.712886199	1.712886448	1.712886458	1.712886459	1.712886762		
V2.27	1.712872114	1.712872162	1.712872411	1.712872421	1.712872422	1.712872725		
V2.28	1.712867483	1.712867531	1.712867687	1.712867697	1.712867698	1.712867999		
V2.29	1.712860631	1.712860679	1.712860928	1.712860938	1.712860939	1.712861242		
V2.30	1.712855702	1.712855750	1.712855980	1.712855990	1.712855991	1.712856286		
V2.31	1.712849885	1.712849933	1.712850182	1.712850192	1.712850193	1.712850493		
V2.32ψ	1.712836950	1.712836998	1.712837226	1.712837236	1.712837237	1.712837533	*	
V2.33	1.712831781	1.712831829	1.712832075	1.712832085	1.712832086	1.712832389		
V2.34ψ	1.712821915	1.712821963	1.712822190	1.712822200	1.712822201	1.712822489	*	
V2.35	1.712815389	1.712815437	1.712815686	1.712815696	1.712815697	1.712861000		
V2.36	1.712810511	1.712810559	1.712810715	1.712810725	1.712810726	1.712811029		
V2.37	1.712803629	1.712803677	1.712803926	1.712803936	1.712803937	1.712804240		
V2.38ψ			1.712791481	1.712791491	1.712791492	1.712791785	partial	
V2.39	1.712786038	1.712786086	1.712786336	1.712786346	1.712786347	1.712786642		
V2.40ψ	1.712770533	1.712770581	1.712770816	1.712770826	1.712770827	1.712771121	*	
V2.41	1.712756308	1.712756356	1.712756605	1.712756615	1.712756616	1.712756919		
V2.42	1.712751220	1.712751268	1.712751424	1.712751434	1.712751435	1.712751738		
V2.43	1.712738996	1.712739044	1.712739293	1.712739303	1.712739304	1.712739604		
V2.44	1.712733807	1.712733855	1.712734011	1.712734021	1.712734022	1.712734325		
V2.45ψ	1.712706742	1.712706790	1.712707011	1.712707021	1.712707022	1.712707313	*	
V3.1	1.714325037	1.714325085	1.714325315	1.714325325	1.714325326	1.714325609		
V3.2	1.714315410	1.714315458	1.714315689	1.714315699	1.714315700	1.714315983		
V3.3ψ	1.714306235	1.714306283	1.714306514	1.714306524	1.714306525	1.714306797	partial	
V3.4ψ	1.714293416	1.714293368	1.714293188	1.714093178	1.714293177	1.714292902	partial	R
V3.5ψ	1.714238395	1.714238443	1.714238675	1.714238685	1.714238686	1.714238865	partial	
V3.6ψ			1.714149484	1.714149494	1.714149495	1.714149628	partial	
V3.7	1.714051608	1.714051560	1.714051327	1.714051317	1.714051316	1.714051033		R
V3.8ψ	1.713971424	1.713971472	1.713971700	1.713971710	1.713971711	1.713971998	partial	
V3.9	1.713885094	1.713885046	1.713884813	1.713884803	1.713884802	1.713884515		R
V3.10ψ			1.713750927	1.713750917	1.713750916	1.713750674	*	R
V3.11ψ			1.713611542	1.713611552	1.713611553	1.713611841	no L1	
V3.12	1.713375275	1.713375323	1.713375554	1.713375564	1.713375565	1.713375848		
V3.13	1.713331353	1.713331401	1.713331632	1.713331642	1.713331643	1.713331920		
V3.14	1.713262895	1.713262943	1.713263176	1.713263186	1.713263187	1.713263470		
V3.15	1.713230088	1.713230136	1.713230368	1.713230378	1.713230379	1.713230662		
V3.16	1.713224857	1.713224904	1.713225137	1.713225147	1.713225148	1.713225428		
V3.17	1.713222963	1.713223011	1.713223243	1.713223253	1.713223254	1.713223537		
V3.18	1.713208895	1.713208943	1.713209173	1.713209183	1.713209184	1.713209472		
V3.19	1.713201932	1.713201980	1.713202206	1.713202216	1.713202217	1.713202505		
V3.20	1.713195897	1.713195945	1.713196172	1.713196182	1.713196183	1.713196471		
V3.21	1.713189822	1.713189870	1.713190100	1.713190110	1.713190111	1.713190399		
V3.22	1.713159292	1.713159340	1.713159569	1.713159579	1.713159580	1.713159868		
V4.1ψ			1.713148261	1.713148269	1.713148270	1.713148572	*	
V4.2	1.713140437	1.713140485	1.713140795	1.713140805	1.713140806	1.713141118		
V4.3	1.713134176	1.713134224	1.713134531	1.713134541	1.713134542	1.713134854		

V4.4	1.713127950	1.713127998	1.713128300	1.713128310	1.713128311	1.713128623		
V4.5ψ	1.713080445	1.713080493	1.713080803	1.713080813	1.713080814	1.713081126	*	
V5.1	1.714466772	1.714466820	1.714467132	1.714467142	1.714467143	1.714467444		
V5.2	1.714447543	1.714447591	1.714447904	1.714447914	1.714447915	1.714448216		
V5.3	1.714434216	1.714434264	1.714434576	1.714434586	1.714434587	1.714434888		
V5.4	1.714419163	1.714419211	1.714419524	1.714419534	1.714419535	1.714419836		
V5.5	1.714392198	1.714392246	1.714392559	1.714392569	1.714392570	1.714392871		
V6.1	1.715825259	1.715825205	1.715824955	1.715824945	1.715824944	1.715824658		R
V7.1	1.714451393	1.714451444	1.714451670	1.714451680	1.714451681	1.714451984		
V7.2	1.714441236	1.714441287	1.714441516	1.714441526	1.714441527	1.714441828		
V7.3ψ	1.714428089	1.714428140	1.714428369	1.714428379	1.714428380	1.714428680	*	
V7.4	1.714404736	1.714404787	1.714405026	1.714405036	1.714405037	1.714405338		
V7.5ψ	1.714385904	1.714385955	1.714386185	1.714386195	1.714386196	1.714386495	*	
V7.6ψ	1.714232276	1.714232327	1.714232567	1.714232577	1.714232578	1.714232870	*	
V7.7ψ	1.713723645	1.713723696	1.713723934	1.713723944	1.713723945	1.713724244	*	
V7.8ψ	1.713608008	1.713608059	1.713608298	1.713608308	1.713608309	1.713608608	*	
V7.9ψ			1.713441391	1.713441381	1.713441380	1.713441081	*	R
V7.10ψ	1.713312249	1.713312300	1.713312539	1.713312549	1.713312550	1.713312858	*	
V7.11ψ	1.713243977	1.713244028	1.713244265	1.713244275	1.713244276	1.713244577	*	
V7.12	1.713205428	1.713205479	1.713205716	1.713205726	1.713205727	1.713206025		

Exon Name	**Exon start**	**Exon end**
J1	1.715897140	1.715897178
J2	1.715896774	1.715896811
C	1.715902135	1.715902463

Footnotes:

[1]Name of gene segment

[2]Start of Leader exon

[3]End of Leader Exon

[4]Start of Exon 2, the continuation of the Leader sequence

[5] End of Leader sequence in Exon 2

[6]Start of V gene sequence in Exon 2

[7]End of Exon 2

[8]Description of features including reason for being designated a pseudogene: no L1 means leader exon 1 could not be locaded on the genome sequence; * means stop codon in reading frame; partial means exon sequence was partial on the genome; R means transcription direction is reverse.

Supplementary Table 3. Locations of IGL gene segments in the MonDom5 assembly

Gene Name[1]	L1 start[2]	L1 end[3]	L2 start[4]	L2 end[5]	V start[6]	V end[7]	Notes[8]	
V1.1	3.525216419	3.525216464	3.525216593	3.525216603	3.525216604	3.525216916		
V1.2	3.525201891	3.525201936	3.525202065	3.525202075	3.525202076	3.525202388		
V1.3	3.525169981	3.525170026	3.525170155	3.525170165	3.525170166	3.525170478		
V1.4	3.525156499	3.525156544	3.525156673	3.525156683	3.525156684	3.525156996		
V1.5	3.525013750	3.525013705	3.525013576	3.525013566	3.525013565	3.525013253		R
V1.6	3.525007449	3.525007404	3.525007275	3.525007265	3.525007264	3.525006952		R
V1.7	3.524977328	3.524977283	3.524977154	3.524977144	3.524977143	3.524976831		R
V1.8	3.524868804	3.524868849	3.524868978	3.524868988	3.524868989	3.524869301		
V1.9	3.524858856	3.524858901	3.524859030	3.524859040	3.524859041	3.524859353		
V1.10	3.524829659	3.524829704	3.524829833	3.524829843	3.524829844	3.524830156		
V1.11	3.524786753	3.524786798	3.524786927	3.524786937	3.524786938	3.524787250		
V1.12	3.524772806	3.524772851	3.524772980	3.524772990	3.524772991	3.524773303		
V1.13	3.524758045	3.52475809	3.524758219	3.524758229	3.524758230	3.524758542		
V1.14	3.524728827	3.524728866	3.524729030	3.524729040	3.524729041	3.524729353		
V1.15	3.524724064	3.524724109	3.524724238	3.524724248	3.524724249	3.524724561		
V1.16	3.524712631	3.524712676	3.524712798	3.524712808	3.524712809	3.524713121		
V1.17	3.524629235	3.52462928	3.524629409	3.524629419	3.524629420	3.524629732		
V1.18	3.524605754	3.524605804	3.524605928	3.524605938	3.524605939	3.524606251		
V1.19	3.524596136	3.524596181	3.524596310	3.524596320	3.524596321	3.524596633		
V1.20	3.524575208	3.524575253	3.524575382	3.524575392	3.524575393	3.524575705		
V1.21	3.524564011	3.524564056	3.524564185	3.524564195	3.524564196	3.524564508		
V1.22	3.524525315	3.52452536	3.524525489	3.524525499	3.524525500	3.524525812		
V1.23	3.524512502	3.524512547	3.524512676	3.524512686	3.524512687	3.524512999		
V1.24	3.524502829	3.524502874	3.524503003	3.524503013	3.524503014	3.524503329		
V1.25	3.524473978	3.524474023	3.524474152	3.524474162	3.524474163	3.524474475		
V1.26	3.524445688	3.524445727	3.524445891	3.524445901	3.524445902	3.524446214		
V1.27	3.524440965	3.52444101	3.524441139	3.524441149	3.524441150	3.524441462		
V1.28	3.524429515	3.52442956	3.524429682	3.524429692	3.524429693	3.524430005		
V1.29	3.524415197	3.524415242	3.524415371	3.524415381	3.524415382	3.524415694		
V1.30	3.524400555	3.5244006	3.524400729	3.524400739	3.524400740	3.524401052		
V1.31	3.524385468	3.524385513	3.524385642	3.524385652	3.524385653	3.524385965		
V1.32	3.524375892	3.524375937	3.524376066	3.524376076	3.524376077	3.524376389		
V1.33	3.524357909	3.524357954	3.524358083	3.524358093	3.524358094	3.524358406		
V1.34	3.524283328	3.524283373	3.524283502	3.524283512	3.524283513	3.524283825		
V1.35ψ	3.524231047	3.524231092	3.524231221	3.524231231	3.524231232	3.524231544	partial	
V1.36	3.524199817	3.524199862	3.524199991	3.524200001	3.524200002	3.524200314		
V1.37	3.524186356	3.524186401	3.524186530	3.524186540	3.524186541	3.524186853		
V1.38	3.524164618	3.524164663	3.524164784	3.524164794	3.524164795	3.524165107		
V1.39	3.524159201	3.524159246	3.524159374	3.524159384	3.524159385	3.524159697		
V1.40	3.524148346	3.524148391	3.524148511	3.524148521	3.524148522	3.524148834		
V1.41	3.524138393	3.524138438	3.524138559	3.524138569	3.524138570	3.524138882		
V1.42	3.524120588	3.524120633	3.524120762	3.524120772	3.524120773	3.524121085		
V1.43	3.524440965	3.52444101	3.524441139	3.524441149	3.524441150	3.524441462		
V1.44	3.524429515	3.52442956	3.524429682	3.524429692	3.524429693	3.524430005		
V1.45	3.524415197	3.524415242	3.524415371	3.524415381	3.524415382	3.524415694		
V1.46ψ					3.524117922	3.524118173	partial	
V1.47	3.524112433	3.524112478	3.524112608	3.524112618	3.524112619	3.524112907		
V1.48	3.524099305	3.524099350	3.524099479	3.524099489	3.524099490	3.524099805		
V1.49	3.523803268	3.523803223	3.523803094	3.523803084	3.523803083	3.523802771		R
V1.50ψ	3.523795519	3.523795474	3.523795345	3.523795335	3.523795334	3.523795221	partial	R

V1.51	3.523792695	3.523792650	3.523792521	3.523792511	3.523792510	3.523792198		R
V1.52ψ	3.523789654	3.523789609	3.523789480	3.523789470	3.523789469	3.523789426	*	R
V1.53ψ	3.523783775	3.523783730	3.523783601	3.523783591	3.523783590	3.523783278	*	R
V1.54ψ			3.523778975	3.523778965	3.523778964	3.523778682	partial	R
V2.1	3.525852076	3.525852031	3.525851892	3.525851882	3.525851881	3.525851597		R
V2.2	3.525844800	3.525844755	3.525844615	3.525844605	3.525844604	3.525844320		R
V2.3	3.525828382	3.525828337	3.525828197	3.525828187	3.525828186	3.525827899		R
V2.4	3.525817204	3.525817159	3.525817019	3.525817009	3.525817008	3.525816746		R
V2.5	3.525807617	3.525807572	3.525807432	3.525807422	3.525807421	3.525807137		R
V2.6	3.525792812	3.525792767	3.525792626	3.525792616	3.525792615	3.525792331		R
V2.7	3.525787539	3.525787494	3.525787354	3.525787344	3.525787343	3.525787056		R
V3.1	3.527503467	3.527503422	3.527503271	3.527503261	3.527503260	3.527502970		R
V3.2	3.527481114	3.527481069	3.527480918	3.527480908	3.527480907	3.527480617		R
V4.1	3.525020514	3.525020559	3.525020690	3.525020700	3.525020701	3.525021012		
V1.Ua	Un52402806	Un52402761	Un52402632	Un52402622	Un52402621	Un52402309		R
V1.Ub	Un52413970	Un52413925	Un52413803	Un52413793	Un52413792	Un52413480		R
V1.Uc	Un52428514	Un52428469	Un52428340	Un52428330	Un52428329	Un52428017		R
V1.Ud	Un53293844	Un53293799	Un53293670	Un53293660	Un53293659	Un53293347		R
V1.Ue	Un53298877	Un53298832	Un53298703	Un53298693	Un53298692	Un53298380		R
V1.Uf	Un53317243	Un53317198	Un53317069	Un53317059	Un53317058	Un53316743		R
V1.Ug	Un55477173	Un55477128	Un55476999	Un55476989	Un55476988	Un55476676		R
V1.Uh	Un80111746	Un80111791	Un80111920	Un80111930	Un80111931	Un80112246		R
V1.Ui	Un83558934	Un83558979	Un83559101	Un83559111	Un83559112	Un83559424		R

Exon Name	**Exon start**	**Exon end**
J1	3.527575294	3.527575331
J2	3.527569801	3.527569838
J3	3.527564618	3.527564655
J4	3.527560482	3.527560519
J5	3.527555190	3.527555227
J6	3.527549908	3.527549945
J7	3.527544616	3.527544653
J8	3.527540387	3.527540424
C1	3.527576804	3.527577123
C2	3.527571527	3.527571846
C3	3.527566128	3.527566447
C4	3.527561992	3.527562161
C5	3.527556701	3.527557020
C6	3.527551404	3.527551723
C7	3.527546124	3.527546443
C8	3.527541994	3.527542313

Footnotes:

[1]Name of gene segment

[2]Start of Leader exon

[3]End of Leader Exon

[4]Start of Exon 2, the continuation of the Leader sequence

[5] End of Leader sequence in Exon 2

[6]Start of V gene sequence in Exon 2

[7]End of Exon 2

[8]Description of features including reason for being designated a pseudogene: * means stop codon in reading frame; partial means exon sequence was partial on the genome; R means transcription direction is reverse.

APPENDIX II

Abbreviations:

Ab, antibody;

BCR, B cell receptor;

C, constant;

cDNA, complementary DNA;

CDR, complementarity determining region;

chr, chromosome;

CP, connecting peptide;

CT, cytoplasmic;

D, diversity;

ERV, endogenous retroviral element;

FR, framework region;

Ig H, immunoglobulin heavy chain;

Ig L, immunoglobulin light chain;

J, joining;

L, leader;

LINE, long interspersed element;

MHC, Major Histocompatibility Complex;

MYA, million years ago;

NAR, New Antigen Receptor;

ORF, open reading frame;

RAG, recombination activating gene;

RSS, recombination signal sequences;

TCR, T cell receptor;

TM, transmembrane;

V, variable;

VH, Ig heavy chain V region;