Published in final edited form as:

Immunogenetics. 2009 August ; 61(8): 581–596. doi:10.1007/s00251-009-0385-8.

On the genomics of immunoglobulins in the gray, short-tailed opossum *Monodelphis domestica*

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

Center for Evolutionary & Theoretical Immunology, Department of Biology, The University of New Mexico, Albuquerque NM 87131 USA, Ph: 1 505 277 2844, Fax: 1 505 277 0304

Robert D. Miller: rdmiller@unm.edu

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.

Keywords

immunoglobulins; genomics; marsupials; *Monodelphis domestica*

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, TCR μ , 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.

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. musculus* (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. franciscii* (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. musculus* (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. franciscii* (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.

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.

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; Aveskog 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).

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 ($S\mu$) 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 S μ of the pseudogene (Fig. 3a). Although the S regions upstream of both the 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, 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 extracellular 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 Supplementary Fig. 1. Analysis of expressed IgH VDJ recombinations 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 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 (Supplementary Fig. 1). 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 IGHI segments have been identified in the opossum genome, four of which appear to be functional: IGHI1, 2, 4, and 5 (Fig. 1). The other two, IGHI3 and 6, both contain in frame stops and they appear to be pseudogenes. The IGHI segments are organized in two sets of three, with IGHI1, 2, and 3 immediately upstream of the functional copy of the IgM C region exons and IGHI4, 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 IGHI1 and 4 as a pair and IGHI3 and 6 as a pair share >90% nucleotide identity within pairs but less than 40% between pairs. IGHI2 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 IGHI3 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 IGHI1 and 2 have been 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. 6). 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. 6; 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. 6; 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. 6 and 7). 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. 6). The remaining 18 IGKV genes comprise three previously undiscovered subgroups bringing the total to seven IGKV subgroups in the opossum (Fig. 6 and 7). Five of the seven IGKV subgroups contain both functional and pseudogene copies (Fig. 6, 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. 6; Supplementary Table 2).

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. 6). 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. 7). 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).

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. 8). 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. 8; 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. 7 and 8; Supplementary Table 3). In addition to the original three subgroups, a fourth IGLV subgroup was identified (IGLV4.1 in Fig. 7 and 8), 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. 7; Lucero et al. 1998).

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; Aveskog 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 S μ 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 (Figs. 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 cartilaginous fishes (Lee et al. 2000). This is in contrast to the only other known mammalian germ-line joined V gene, the V μ J gene found in TCR μ , a unique TCR also discovered in marsupials (Parra et al. 2007). V μ J 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 V μ J 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 (Supplementary Fig. 1). 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 (Aveskog 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.

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

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

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

Supplementary Figure 1 legend: Nucleotide sequence and amino acid translations of all three reading frames of the nine opossum IGHD segments. * indicates in-frame stop codons.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to acknowledge support from National Science Foundation award IOS-0641382 and a National Institutes of Health Institutional Development Award supporting the Center for Evolutionary and Theoretical Immunology's core molecular biology facility. The authors also wish to thank Joseph Volpe and Thomas Kepler at Duke University for being willing to include the opossum germ-line gene segments at the SoDA website, making analysis of antibody diversity publicly available.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410. [PubMed: 2231712]
- Aveskogh M, Hellman L. 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* 1998;28:2738–2750. [PubMed: 9754561]
- Aveskogh M, Pilström L, Hellman L. Cloning and structural analysis of IgM (μ chain) and the heavy chain V region repertoire in the marsupial *Monodelphis domestica*. *Dev Comp Immunol* 1999;23:597–606. [PubMed: 10579388]
- Baker ML, Belov K, Miller RD. Unusually similar patterns of antibody V segment diversity in distantly related marsupials. *J Immunol* 2005;174:5665–5671. [PubMed: 15843567]
- Bell RG. Marsupial immunoglobulins: the distribution and evolution of macropod IgG2, IgM and light chain antigenic markers within the sub-class metatheria. *Immunol* 1977;33:917–924.
- Bell RG, Lynch NR, Turner KJ. Immunoglobulins of the marsupial *Setonix brachyurus* (quokka). *Immunol* 1974;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. Reconstructing an ancestral mammalian immune supercomplex from a marsupial MHC. *PLoS Biol* 2006;4:e46.10.1371/journal.pbio.0040046 [PubMed: 16435885]
- Belov K, Harrison GA, Rosenberg GH, Miller RD, Cooper DW. Isolation and comparison of the IgM heavy chain constant region from Australian (*Trichosurus vulpecula*) and American (*Monodelphis domestica*) marsupials. *Dev Comp Immunol* 1999;23:649–656. [PubMed: 10579393]
- Belov K, Hellman L. Immunoglobulin genetics of *Ornithorhynchus anatinus* (platypus) and *Tachyglossus aculeatus* (short-beaked echidna). *Comp Biochem Physiol, Part A Mol Integr Physiol* 2003;136:811–819.
- Das, S.; Nozawa, M.; Klein, J.; Nei, M. Evolutionary dynamics of the immunoglobulin heavy chain variable region genes in vertebrates; *Immunogenetics*. 2008. p. 47-55. <http://dx.doi.org/10.1007/s00251-007-0270-2>
- Deakin JE, Olp JJ, Graves JAM, Miller RD. Physical mapping of immunoglobulin loci *IGH@*, *IGK@*, and *IGL@* in the opossum (*Monodelphis domestica*). *Cytogenet Genome Res* 2006;114:94.
- Deane, EM.; Cooper, DW. Immunological development in pouch young marsupials. In: Tyndale-Biscoe, CH.; Janssens, PA., editors. *The Developing Marsupial*. Springer-Verlag; Berlin: 1988. p. 190-199.

- Deane EM, Cooper DW, Renfree MB. Immunoglobulin G levels in fetal and newborn tammar wallabies (*Macropus eugenii*) Reprod. Fert Dev 1990;2:369–375.
- Fry EJ, Saunders NR. Spinal repair in immature animals: a novel approach using the South American opossum *Monodelphis domestica*. Clin Exp Pharmacol Physiol 2000;27:542–547. [PubMed: 10874515]
- Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 1994;41:95–98.
- Johansson J, Aveskogh M, Munday G, Hellman L. 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 2002;168:5155–5162. [PubMed: 11994470]
- Kawasaki K, Minoshima S, Nakato E, Shibuya K, Shintani A, Asakawa S, Sasaki T, Klobeck HG, Combriato G, Zachau HG, Shimizu N. Evolutionary dynamics of the human immunoglobulin kappa locus and the germline repertoire of the V kappa genes. Eur J Immunol 2001;31:1017–1028. [PubMed: 11298326]
- Kawamura S, Saitou N, Ueda S. Concerted evolution of the primate immunoglobulin α -gene through gene conversion. J Biol Chem 1992;267:7359–7367. [PubMed: 1559979]
- Kumar S, Tamura K, Nei M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 2004;5:150–163. [PubMed: 15260895]
- Lee SS, Fitch D, Flajnik MF, Hsu E. Rearrangement of immunoglobulin genes in shark germ cells. J Exp Med 2000;191:1637–1648. [PubMed: 10811858]
- Lefranc, MP.; Lefranc, G. Immunoglobulin lambda (IGL) genes of human and mouse. In: Honjo, T.; Alt, FW.; Neuberger, M., editors. Molecular Biology of B Cells. Academic Press; London: 2004. p. 37-59.
- Ley RD, Reeve VE, Kusewitt DF. Photobiology of *Monodelphis domestica*. Dev Comp Immunol 2000;24:503–516. [PubMed: 10785275]
- Lucero JE, Rosenberg GH, Miller RD. Marsupial light chains: Complexity and conservation of Igl in the opossum *Monodelphis domestica*. J Immunol 1998;161:6724–6732. [PubMed: 9862702]
- Melchers, F.; Kincade, P. Early B cell development to a mature, antigen-sensitive cell. In: Honjo, T.; Alt, FW.; Neuberger, M., editors. Molecular Biology of B Cells. Academic Press; London: 2004. p. 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, Grealis 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. Genome of the marsupial *Monodelphis domestica* reveals lineage-specific innovation in coding sequences. Nature 2007;447:167–178. [PubMed: 17495919]
- Miller RD, Grabe H, Rosenberg GH. The V_H Repertoire of a Marsupial: *Monodelphis domestica*. J Immunol 1998;160:259–265. [PubMed: 9551979]
- Miller RD, Bergemann ER, Rosenberg GH. Marsupial light chains: Igk with four V families in the opossum *Monodelphis domestica*. Immunogenetics 1999;50:329–335. [PubMed: 10630297]
- Miller RD, Belov K. Immunoglobulin genetics of marsupials. Dev Comp Immunol 2000;24:485–490. [PubMed: 10785273]
- Mills FC, Brooker JS, Camerini-Otero RD. Sequences of human immunoglobulin switch regions: implications for recombination and transcription. Nucleic Acids Res 1990;18:7305–7316. [PubMed: 2124350]
- Muyldermans S, Atarhouch T, Saldanha J, Barbosa JA, Hamers R. Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. Protein Eng 1994;7:1129–1135. [PubMed: 7831284]
- Nikaido T, Yamawaki-Kataoka Y, Honjo T. Nucleotide sequences of switch regions of immunoglobulin Cepsilon and Cgamma genes and their comparison. J Biol Chem 1982;257:7322–7329. [PubMed: 6282840]

- Ohta Y, Flajnik M. IgD, like IgM, is a primordial immunoglobulin class perpetuated in most jawed vertebrates. *Proc Natl Acad Sci U S A* 2006;103:10723–10728. [PubMed: 16818885]
- Osborne BA, Golde TE, Schwartz RL, Rudikoff S. Evolution of the IgA Heavy Chain Gene in the Genus *Mus*. *Genetics* 1988;119:925–931. [PubMed: 2842228]
- Parra ZE, Baker ML, Hathaway J, Lopez AM, Trujillo J, Sharp A, Miller RD. Comparative genomic analysis and evolution of the T cell receptor loci in the opossum *Monodelphis domestica*. *BMC Genomics* 2008;9:111. [PubMed: 18312668]
- Parra ZE, Baker ML, Lopez AM, Trujillo J, Volpe JM, Miller RD. TCR μ recombination and transcription relative to the conventional TCR during postnatal development in opossums. *J Immunol* 2009;182:154–163. [PubMed: 19109146]
- Parra ZE, Baker ML, Schwarz R, Deakin JE, Lindblad-Toh K, Miller RD. Discovery of a new T cell receptor in marsupials. *Proc Natl Acad Sci U S A* 2007;104:9776–9781. [PubMed: 17535902]
- Renfree MB. The composition of fetal fluids of the marsupial *Macropus eugenii*. *Dev Biol* 1973;33:62–79. [PubMed: 4789603]
- Renfree, MB. The genome of a Gondwanan mammal; Bioessays. 2007. p. 1073-1076. <http://dx.doi.org/10.1002/bies.20670>
- Reynaud CA, Dahan A, Anquez V, Weill JC. Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. *Cell* 1989;59:171–183. [PubMed: 2507167]
- Roux KH, Greenberg AS, Greene L, Strelets L, Avila D, McKinney EC, Flajnik MF. 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* 1998;95:11804–11809. [PubMed: 9751746]
- Saitou N, Nei M. The neighbor joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;4:406–425. [PubMed: 3447015]
- Saini SS, Allore B, Jacobs RM, Kaushik A. Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM Abs. *Eur J Immunol* 1999;29:2420–2426. [PubMed: 10458755]
- Samollow PB. 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* 2006;54:173–196.
- Samples NK, VandeBerg JL, Stone WH. Passively acquired immunity in the newborn of a marsupial (*Monodelphis domestica*). *Am J Reprod Immunol Microb* 1986;11:94–97.
- Shearer MH, Robinson ES, VandeBerg JL, Kennedy RC. 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* 1995;19:237–46. [PubMed: 8595822]
- Silverman GJ, Goodyear CS. A model B-cell superantigen and the immunobiology of B lymphocytes. *Clin Immunol* 2002;102:117–134. [PubMed: 11846453]
- Tutter A, Riblet R. Conservation of an immunoglobulin variable-region gene family indicates a specific, noncoding function. *Proc Natl Acad Sci U S A* 1989;86:7460–7464. [PubMed: 2798416]
- VandeBerg, JL.; Robinson, ES. The laboratory opossum (*Monodelphis Domestica*) in biomedical research. In: Saunders, NR.; Hinds, LA., editors. *Marsupial Biology: Recent Research*. New Perspectives; UNSW, Sydney: 1997. p. 238-253.
- Volpe JM, Cowell LG, Kepler TB. SoDA: implementation of a 3D alignment algorithm for inference of antigen receptor recombinations. *Bioinformatics* 2006;22:438–444. [PubMed: 16357034]
- Wakefield MJ, Graves JAM. The kangaroo genome. Leaps and bounds in comparative genomics. *EMBO Rep* 2003;4:143–147. [PubMed: 12612602]
- Wilson M, Bengten E, Miller NW, Clem LW, Du Pasquier L, Warr GW. A novel chimeric Ig heavy chain from a teleost fish shares similarities to IgD. *Proc Natl Acad Sci U S A* 1997;94:4593–4597. [PubMed: 9114035]
- Wong ES, Young LJ, Papenfuss AT, Belov K. In silico identification of opossum cytokine genes suggests the complexity of the marsupial immune system rivals that of eutherian mammals. *Immunome Res* 2006;2:4. [PubMed: 17094811]
- Zachau, HG. Immunoglobulin kappa genes of human and mouse. In: Honjo, T.; Alt, FW.; Neuberger, M., editors. *Molecular Biology of B Cells*. Academic Press; London: 2004. p. 27-36.

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 is in 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.

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.

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.

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 bootstrap 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. franciscii* (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).

IGHV1.5 tattactgtgcaagaga CACAGTGAGGGAGAGGAGTGAGAGAGCAGACCAAAAACCT
 Y Y C A R heptamer 23bp space nonamer

IGHV2.1 tattattgtgcaaaaga CACAATGAGAGGAAGAGAATGTTAACCCTACACAAACT
 Y Y C A K heptamer 23bp space nonamer

IGHV3.1 tattactgtgtgagaactgctgcagaaccattgcttttgaagatataattccaacagca CACAGTGTTTGACACAACCTCAAATCTTATCAACTTTG
 Y Y C V R T A A E P I A F E D I I P T A heptamer 12bp space nonamer

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 indicated.

Figure 6. 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).

Figure 7.

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. migratorius* (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).

Figure 8.

Diagram of the opossum *IGL* locus. V, J and C genes are shown 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).

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	GGTGTCTTGTTCCTCAAA	Reverse primer to fill the gap from pseudo-M CH1 exon
VH3F	GGAGACAGGTGGTTTGCATT	Forward primer 5' of IGHV3.1 genomic sequence
VH3R	CTGGTCCACTTGTCTTTCTGG	Reverse primer 3' of IGHV3.1 RSS genomic sequence

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	CTTAGTCTCTGT	CACTGTG	GATACAGTAACCTAC	CACAGTG	GCAAAACACTAAT	TCTAAAAAGT
D2	TGTTTATGA	CTTGCTCTGTGT	CATTGTG	GTTTATAGCTGGGGTA	CACAGTG	ATATATTTCAAA	TCTAAAAACT
D3	GGATTCCAT	ACAGTAAATGTAA	CACTGTG	TAACTACGGGTATAGCTA	CACAGTG	TTAATCCCTATT	ACAAAAATT
D4	AGATTCTGT	GTTGGACTCTGT	CACTGTG	CTATATTATTGTAGTAGTGGTATTTGCTACGAC	CACAGTG	ATAGATCTCCCA	TCAAAAAACC
D5	AGGATGTTG	CCAAGGATGTAT	CACTGTG	CTATTAACTGGTGGTTGGAACTAC	CACAGTG	ACAAAAGAACTTA	TCAAAAAACC
D6	AGATTTTGA	ATGAGAAATATAT	CATAGTG	TACTACTACAGTGGTAAC	CACATTG	AAACAACCTTTG	ACAAAAACC
D7	AGATTATGT	GCTAGACTGTGT	CACTGTG	GTATTATTGTATAGTGAATTGGGGCT	CACAGTG	ATCTCCAATCAA	AAACCCATC
D8	TAGATATAT	ACCTCCTGGTTC	CACTGTG	ATAAGCATTGAGTGGTACCTAGTCTAATCAACCTTTGAATTTC	CACAGAT	CATCCCTTATTC	CCCCAACCA
D9	TATAACAAT	CCGTCCAGCAAA	CACTGTG	TCCTTGTATCTCCAACCTTAGGCCATCT	CACATGG	AATATACATCTG	TCTGTGGCT