Characterization of the opossum immune genome provides insights into the evolution of the mammalian immune system

Katherine Belov, Claire E. Sanderson, Janine E. Deakin, et al.

Genome Res. 2007 17: 982-991 originally published online May 10, 2007
Access the most recent version at doi:10.1101/gr.6121807

Letter

Characterization of the opossum immune genome provides insights into the evolution of the mammalian immune system

Katherine Belov,[1,7] Claire E. Sanderson,[1] Janine E. Deakin,[2] Emily S.W. Wong,[1] Daniel Assange,[3] Kaighin A. McColl,[3] Alex Gout,[3,4] Bernard de Bono,[5] Alexander D. Barrow,[6] Terence P. Speed,[3] John Trowsdale,[6] and Anthony T. Papenfuss[3]

[1]Faculty of Veterinary Science, University of Sydney, Sydney NSW 2006, Australia; [2]ARC Centre for Kangaroo Genomics, Research School of Biological Sciences, The Australian National University, Canberra ACT 0200, Australia; [3]Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, Parkville 3050, Australia; [4]Department of Medical Biology, The University of Melbourne, Victoria 3010, Australia; [5]European Bioinformatics Institute, Cambridge CB10 1SD, United Kingdom; [6]Immunology Division, University of Cambridge, Cambridge CB2 5DS, United Kingdom

The availability of the first marsupial genome sequence has allowed us to characterize the immunome of the gray short-tailed opossum (*Monodelphis domestica*). Here we report the identification of key immune genes, including the highly divergent chemokines, defensins, cathelicidins, and Natural Killer cell receptors. It appears that the increase in complexity of the mammalian immune system occurred prior to the divergence of the marsupial and eutherian lineages ~180 million years ago. Genomes of ancestral mammals most likely contained all of the key mammalian immune gene families, with evolution on different continents, in the presence of different pathogens leading to lineage specific expansions and contractions, resulting in some minor differences in gene number and composition between different mammalian lineages. Gene expansion and extensive heterogeneity in opossum antimicrobial peptide genes may have evolved as a consequence of the newborn young needing to survive without an adaptive immune system in a pathogen laden environment. Given the similarities in the genomic architecture of the marsupial and eutherian immune systems, we propose that marsupials are ideal model organisms for the study of developmental immunology.

[Supplemental material is available online at www.genome.org.]

A defining characteristic of marsupials is that they give birth to underdeveloped young (Tyndale-Biscoe and Renfree 1987). The young emerges from the uterus without immunological tissues or organs, and their immune system develops outside the sterile confines of the uterus (Old and Deane 2000). These features make them ideal model organisms for developmental immunology studies. However, research in this area has been limited, due to a general lack of understanding of marsupial immunology. Early studies suggested that marsupials have limited capacity for antibody class switching, a lack of allogeneic recognition, and a deficient memory response (Stone et al. 1996). The inability to isolate many of the more divergent immune molecules in the laboratory fuelled debates about the level of complexity of the marsupial immune system (Jurd 1994; Young and Deane 2003; Wong et al. 2006). However, recent molecular studies are starting to uncover a genomic architecture of the marsupial immune system that is very similar to that of eutherian mammals. Highly conserved immune genes, including the immunoglobulins M, A, G, and E (Miller and Belov 2000), T cell receptors α, β, γ, δ (Baker et al. 2001), and major histocompatibility complex (MHC) genes (Miska and Miller 1999; Gouin et al. 2006), were isolated in the laboratory, their gene clusters characterized (Belov et al. 2006; R.D. Miller, in prep.), and divergent immune molecules, including Th1 and Th2 cytokines, identified (Wong et al. 2006).

The key difference that separates the immune systems of marsupials and eutherians is that marsupial neonates have no adaptive immune system, while eutherian young are born with a relatively well developed adaptive immune system (Old and Deane 2000). The gray short-tailed opossum (*Monodelphis domestica*) is born after a gestation period of only 15 d. The young are born without immunological tissues or organs, and their immune system develops in the presence of pathogens. Unlike most other marsupials, the opossum does not have a pouch, and spends its first 16 d attached to one of the mother's teats. The mechanism by which the immunologically naïve young survives pathogen challenge during this period is not fully understood. Maternal antibodies are not passed across the placenta in the opossum (Samples et al. 1986) and the young cannot produce their own antibodies until day 7 postpartum. Therefore, during the first 7 d of life, the young are heavily reliant on passive immunity through lactation. The young permanently suckle for 16 d, followed by a 20-d period of intermittent suckling. Intermittent suckling corresponds with exposure to new pathogens and an increase in immunocompetence. The young are weaned at 60 d of age (Tyndale-Biscoe 2005), corresponding to the termination of absorption of antibodies across the gut epithelium (Wild et al. 1994). It is possible that the marsupial young are

[7]Corresponding author.
E-mail kbelov@vetsci.usyd.edu.au; fax 61-2-9351-3957.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.6121807. Freely available online through the *Genome Research* Open Access option.

17:982–991 ©2007 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/07; www.genome.org

reliant on innate immunity during their first few weeks of life. Marsupial-specific expansions of innate immune genes would support this notion.

The availability of the first marsupial genome (Mikkelsen et al. 2007) has allowed the in silico identification of immune genes in the opossum genome. We have not included immune genes that are encoded in gene segments, such as immunoglobulins and T cell receptors and specifically focus on several key immune multigene families, which are potentially involved in the protection of the immunologically naïve young: the defensins, cathelicidins, chemokines, and Natural Killer (NK) receptor gene clusters.

Results and Discussion

The immunome

In this study, we used the human Immunogenetic Related Information Source (IRIS) (Kelley et al. 2005a) to search the opossum genome for homologous genes, which are likely to have an immune function. We aligned 1528 human immune-related proteins from the IRIS database with the opossum genome; 1405 aligned successfully ($E < 1 \times 10^{-5}$), a further 48 aligned with lower stringency ($0.01 < E < 1 \times 10^{-5}$), and 75 proteins did not align at all ($E > 0.01$). We constructed chains of high-scoring segment pairs (HSPs) and aligned these against both the IRIS and RefSeq databases. By taking the reciprocal best hit (with E-value $< 1 \times 10^{-5}$) we were able to find 1111 putative orthologs. Gene predictions are available via keyword search on our genome browser at http://bioinf.wehi.edu.au/opossum/. Overall, it appears that the genetic content of the marsupial immune system is as extensive as that of their eutherian counterparts.

Genes not identified using this method included the rapidly evolving gene families encoding the Natural Killer cell receptors, interleukins, interferons, chemokines, defensins, and cathelicidins. Not surprisingly, many of these genes are also unrepresented or poorly represented in the opossum genebuilds (data not shown). The identification and characterization of these genes are the focus for this paper. We were able to identify members of all of these gene families by taking into account gene features including genome location (conserved synteny) and presence of key domains (e.g., immunoglobulin and C-type lectin domains). Our gene predictions can be found at http://bioinf.wehi.edu.au/opossum/, with additional information on identification of chemokines described in the Supplemental Material.

A comparison of immune gene numbers between the opossum, human, mouse, and chicken is shown in Table 1. The immune gene content of the opossum is on par with that of humans and mice and it is likely that the diversification of immune genes that resulted in the highly complex immune system of eutherian mammals occurred prior to the divergence of the marsupial and eutherian lineages. The immune system of chickens appears to be much simpler than that of mammals. This is supported by the fact that the chicken MHC is "minimal essential" (Kaufman et al. 1999) and is much more compact and simple than the MHC of mammals. Genomic and EST sequencing has led to the recent discovery of many chicken immune genes including interleukins, chemokines, and genes involved in the Toll-like pathway (Smith et al. 2004; Burt 2005). The marsupial data will allow the prediction of ancestral genes phylogenetically through the alignment of mammalian sequences. Searching the chicken genome with ancestral sequences will reduce the evolutionary distance between bird and human immune genes and aid the identification of highly divergent genes.

Table 1. Comparison of gene family numbers in human, mouse, opossum, and chicken genomes

	Human	Mouse	Opossum	Chicken
Cathelicidin	1	1	12	3
Beta-defensin	39	52	32	13
Alpha-defensin	10	6	1	0
Theta-defensin	1ps	0	0	0
Chemokine	47	45	31	24
KLRA1 (Ly49)	1ps	16	0	0
NKG2D/KLRK1	1	1	1	0
CD69	1	1	1	0
KLRC	4	3	0	0
Ig-like receptor	30	10	~45	103

The large scale organization of identified immune genes in the opossum genome is shown in Figure 1. It should be noted that our search strategy was largely based on cross-species searches. As a consequence, it is possible that some novel opossum immune genes may have been overlooked. In general, immune genes are randomly distributed across the genome, with the exception of several immune gene clusters, including the leukocyte receptor complex (LRC), the Natural Killer complex (NKC), the major histocompatibility complex (MHC) and paralogous regions, immunoglobulin heavy and light chain genes, the T cell receptors genes, the toll-like receptor genes (TLR), and the antimicrobial peptide genes.

A large number of immune genes are also seen on the unordered chromosome, which consists of genomic segments that have not yet been assigned to chromosomes. The high number of immune genes on the unordered chromosome is probably the result of difficulties in genome assembly around genes belonging to large gene families that evolve by duplication and diversification, and around alleles and between haplotypes with variable copy number. This is supported by the fact that we identified related genes for all of the key gene families, studied in this paper, on the unordered chromosome.

The fact that genes that evolve by tandem duplication (including defensins, C-type lectin, and immunoglobulin-like receptor genes) are found in clusters in the opossum genome is not surprising. However, clustering of unrelated immune genes is more interesting. For instance, unrelated genes within the MHC region have remained clustered for extended periods of time. The MHC is one of the most intensively studied immune gene clusters, yet the evolution of this region is still subject to debate. The presence of four MHC paralogous regions in higher vertebrates has been used to suggest that the vertebrate genome underwent two rounds of duplication in its early evolution (Kasahara 1997). Evidence for this has mainly been provided by comparing the presence and arrangement of these genes in human with the arrangement in the proto-MHC in amphioxus (Abi-Rached et al. 2002). The paralogous region on human chr9q32–q34 is thought to correspond to the ancestral region from which the MHC evolved (Vienne et al. 2003). However, the recent elucidation of the gene organization of the opossum MHC demonstrated that the organization of the human MHC is derived, as the deduced ancestral mammalian MHC organization did not have the Class I and II genes separated by the Class III region (Belov et al. 2006). The ancestral organization of the opossum MHC, in conjunction with the fact that the gene content is similar to that in humans,

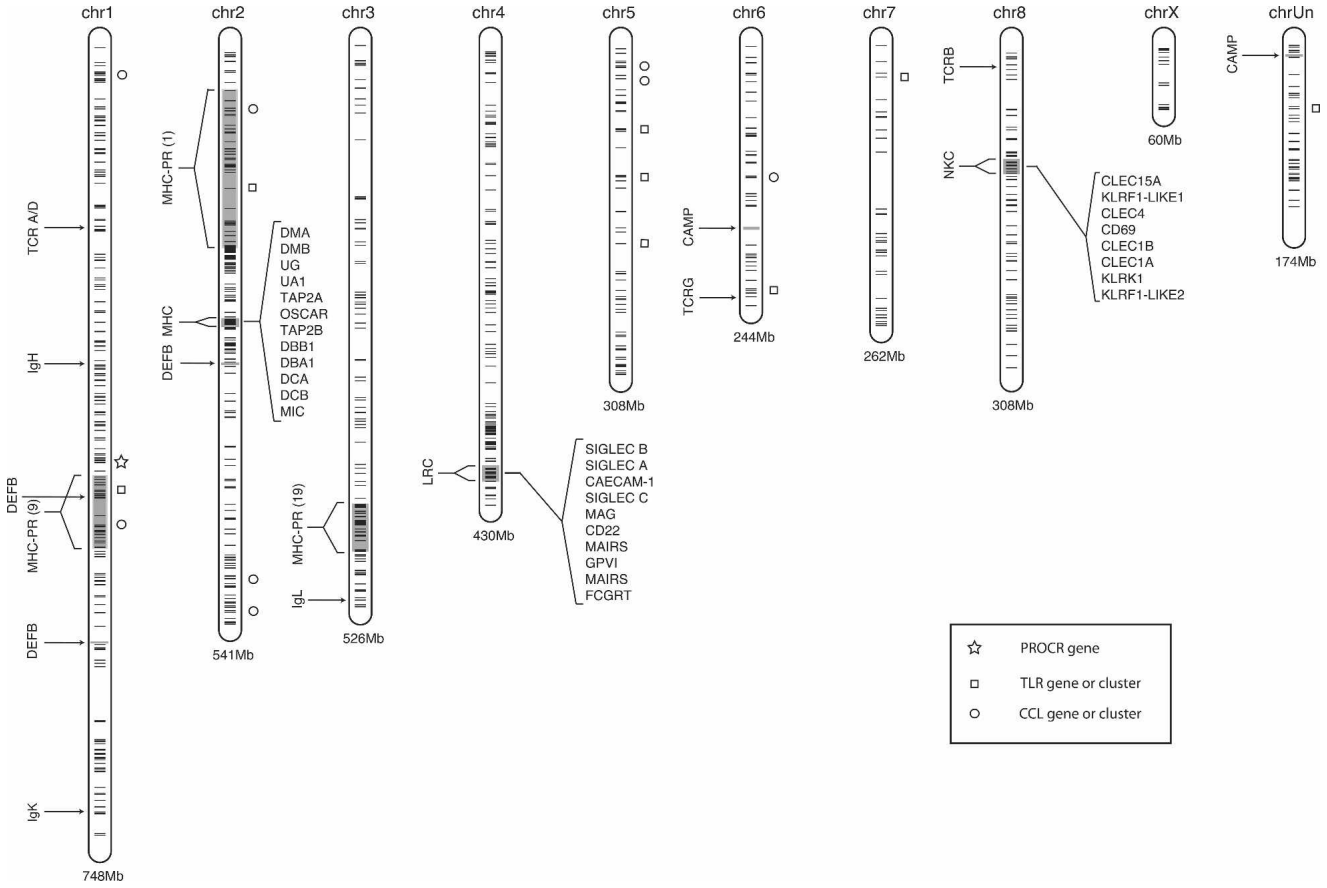

Figure 1. Chromosome map showing locations of immune gene clusters in the opossum. Locations of key immune gene clusters are shown.

makes the opossum an ideal model for studying the evolution of MHC paralogous regions. The opossum MHC paralogous regions were identified on chromosomes 1, 2, and 3 and correspond to human paralogous regions on chromosomes 9, 1, and 19, respectively (details in Supplemental Material). The order of genes is highly conserved between the two MHC regions. The paralogous region on opossum chromosome 1 and its counterpart on human chromosome 9 are well conserved with only a few rearrangements. The other two regions have undergone extensive rearrangement.

The discovery of a Class I-like gene called *PROCR*, which is an endothelial protein receptor, on opossum chromosome 1 is interesting. Opossum *PROCR* was identified within a block of genes that are located on human chromosome 20, adjacent to an MHC paralogous region. Maruoka et al. (2005) found that *PROCR* and *CD1* are more closely related to each other than they are to other Class I sequences and suggested that they arose by tandem duplication of MHC Class I genes within the MHC, followed by translocation of *PROCR* to a location outside the MHC (Maruoka et al. 2005). We propose that *PROCR* may have evolved as a component of a block duplication during the formation of the MHC paralogous regions.

The discovery of clustering of individual MHC and NK receptor genes is becoming increasingly common. The chicken MHC contains C-type lectin genes and Class I-related paralog *CD1* genes (Rogers et al. 2005), while the opossum MHC contains *Oscar*, a gene found in the LRC of eutherian mammals (Belov et al. 2006). *A2M*, a complement component paralog, is found within the NKC of the opossum and humans, but just outside the NKC in rodents. Class I-related genes, *Mill1* and *Mill2*, are located near the LRC in mice (Kasahara et al. 2002). *TAPBPL*, a tapasin-like gene is located near the NKC region in humans (Teng et al. 2002) but is not part of the NKC region in opossum (instead being found on chromosome 2). The MHC of *Xenopus* contains *XMIV* genes that are believed to be Ig-like receptors containing ITIMs and ITAM motifs (Ohta et al. 2006). Colocalization of MHC Class I ligands and their NK receptors in an ancestral immune supercomplex would have allowed the coevolution of ligands and receptors. However, clustering does not appear to be essential for immune function, and over time the genes within these clusters have dispersed, resulting in different organization in different lineages.

The antimicrobial peptides

Eminent marsupial reproductive biologists have long appreciated that surgical incisions on pouch young returned to a nonsterile pouch do not become infected (Renfree and Tyndale-Biscoe 1978). The antimicrobial peptides (AMP) may play a key role in protecting immunologically vulnerable young during early development, as they provide immediate defense through their antibiotic activity and direct killing of a wide variety of microorganisms. AMPs can be divided into two main groups: cathelicidins and defensins. The opossum genome contains 12 cathelicidins, 32 beta-defensins, and a single alpha-defensin gene.

Cathelicidins

Twelve cathelicidin genes (details in Supplemental Material) were found in a single cluster on opossum chromosome 6. This region displays conserved synteny with the cathelicidin clusters found in eutherian mammals and birds (Uzzell et al. 2003; Chang et al. 2006; Xiao et al. 2006). The opossum cathelicidin antimicrobial mature peptide sequences are highly heterogeneous. The high level of sequence divergence in the opossum cathelicidin genes is highlighted by the long branches separating them on a phylogenetic tree (Fig. 2), and amino acid identities of the precursor peptides ranging from 5% to 95% and the mature peptides from 1% to 94%. The level of heterogeneity in opossum cathelicidin genes is much greater than that seen within other species. This may have resulted from positive Darwinian selection on cathelicidin gene sequences by a highly diverse array of microbial pathogens and could be related to the young's reliance on innate immunity for the first week of life.

Defensins

Defensins are a family of cationic AMPs with three pairs of intramolecular cysteine disulfide bonds. They belong to one of the most rapidly evolving mammalian gene families. On the basis of their size and spacing of disulfide bonds, defensins can be classified into three subfamilies: alpha, beta, and theta. A total of 37 putative defensins genes and pseudogenes were identified in the genome of the opossum. Details are given in Supplemental Table 1.

The opossum beta-defensin genes are arranged in three clusters, which display conserved synteny to the four or five clusters found in humans, rodents, and dogs (summarized in Fig. 3; Patil et al. 2005). The chicken genome contains only 13 beta-defensin genes, which are found in a single cluster of 86 kb on chromosome 3 (Lynn et al. 2004; Xiao et al. 2004). This cluster displays conserved synteny with the two clusters on human chromosome 8 (clusters A and B) and does not contain any alpha- or theta-defensins. The evolutionary relationship of the opossum genes was determined phylogenetically, as described in the Supplemental Material. The phylogenetic tree and synteny map allowed us to assign opossum beta-defensins to known eutherian defensin gene families and to detect some opossum specific lineage expansions. For instance, *MdoDEFB20* is clearly orthologous to dog *DEFB140* and rat and mouse *Defb51*, while *MdoDEFB21* is orthologous to rodent *Defb14* and human and dog *DEFB103* defensin gene families (Supplemental Fig. 3). A large opossum-specific expansion is seen in cluster A. It is possible that these genes have a specific role in marsupial immunity that is not required in eutherians, such as protection of underdeveloped young.

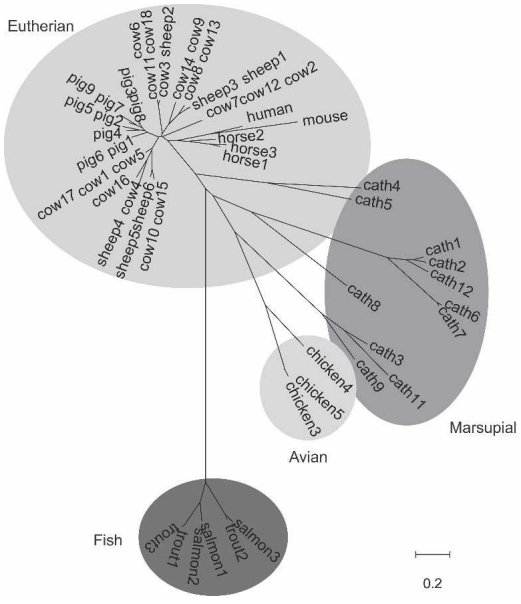

Figure 2. Phylogenetic tree of mammalian cathelicidin genes.

A single alpha-defensin was found in the opossum genome (details in Supplemental Material) on chromosome 1, in conserved syntenic cluster A (Fig. 3). Previously, alpha-defensins had only been identified in eutherian mammals, including humans, rodents, lagomorphs, and horses (Patil et al. 2004). The discovery of a marsupial alpha-defensin indicates that the gene duplication of a beta-defensins that gave rise to the alpha-defensin lineage (Patil et al. 2004) occurred prior to the divergence of the marsupial and eutherian lineages, 180 million years ago. The alpha-defensin loci are conserved in syntenic chromosomal regions (Fig. 3). Identification of a marsupial alpha-defensin genes and beta-defensins orthologs between marsupials and eutherians suggests that the mammalian antimicrobial repertoire diversified rapidly through gene duplication and chromosomal translocation after the divergence of birds and mammals, ~310 million years ago, but prior to the divergence of marsupials and eutherians, ~180 million years ago.

The NK receptors

Natural Killer (NK) cells play a key role in the innate immune system, destroying virally infected cells and cancer cells. NK cells contain activating and inhibitory receptors that respond to counter-receptors on target cells. The detection of suitable targets by NK cells is mediated by two major families of receptor molecules: the immunoglobulin superfamily (IgSF) and the C-type lectin superfamily (CLSF). The ligands for many of these receptors are the MHC Class I molecules. In eutherian mammals, immunoglobulin superfamily NK receptors are found in the leukocyte receptor complex (LRC) while C-type lectin-like domain containing NK receptor genes are found in the Natural Killer complex (NKC).

The composition of the LRC and NKC differ markedly between species, as these gene families evolve quickly, with rapid lineage specific expansions and contractions (Kelley et al. 2005b). Different eutherian lineages make use of genes from the two different superfamilies to carry out analogous functions. Humans preferentially use LRC-encoded *KIR* (killer cell immunoglobulin-like receptors) genes to engage nonclassical MHC Class I molecules, while rodents use NKC-encoded *KLRA* (*Ly49*) genes for the same role. The functional homology of human *KIR*s and mouse *KLRA* genes is believed to have arisen by convergent evolution. The study of the genetic composition of these clusters in different species provides us with clues about their evolution. Here we report the identification of the NKC and LRC regions in the opossum and provide some preliminary descriptions of their gene content and organization. Specifically, we report the absence of *KLRA* genes in the NKC (or elsewhere) and an expansion of unique Ig-like domain containing genes in the LRC.

The Natural Killer complex (NKC)

The opossum NKC is located on chromosome 8 and contains nine C-type lectin receptor genes. Gene prediction was con-

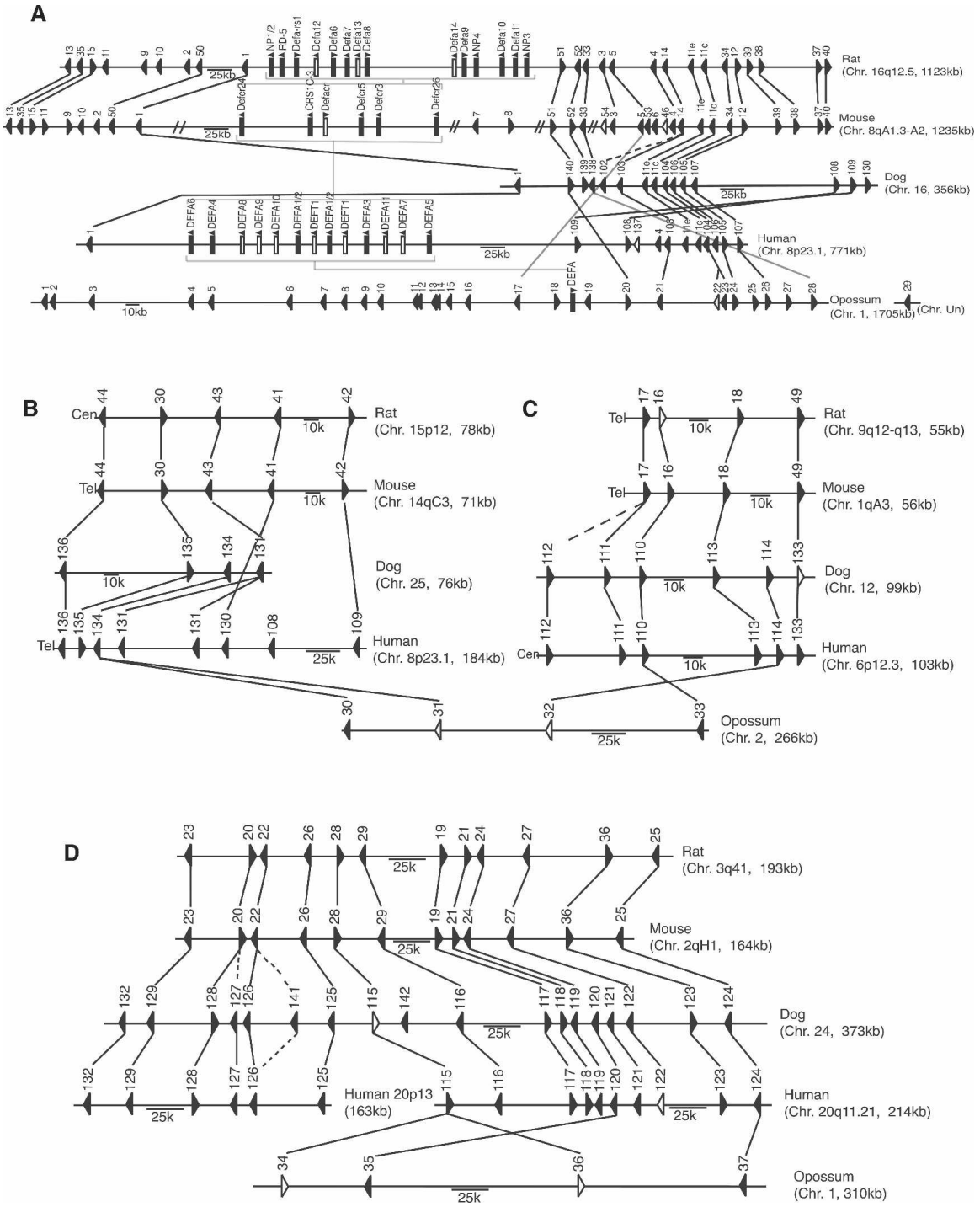

Figure 3. Synteny map of mammalian alpha- and beta-defensin gene clusters (modified from Patil et al. 2005). Four different beta-defensin clusters (*A, B, C, D*) are found in rat, mouse, and dog. Five clusters are found in humans due to a split in cluster *D*. The opossum contains three clusters suggesting that eutherian clusters *B* and *C* evolved from a single cluster in ancestral mammals. The alpha-defensins (cluster A) are denoted by rectangles. Arrows show transcriptional orientation. Pseudogenes are shown in white, and putative functional genes in black. Orthologs, based on phylogenetic analyses, are connected.

ducted around the C-type lectin domains, and putative genes were assigned to gene families based on phylogenetic analysis (Supplemental Fig. 6). Orthologs for *CLEC15A*, *CLEC4E*, *KLRF1-like1*, *CLEC4*, *CD69*, *CLEC1B*, *CLEC1A*, *KLRK1*, *KLRF1-like2* were identified. Eight of these genes were previously identified by Hao et al. (2006) from an earlier genome assembly. The level of amino acid identity between the opossum sequences ranged from 4% to 26%. The organization of the opossum NKC is similar to that seen in eutherians (Fig. 4), with at least seven NKC genes predating the divergence of marsupials and eutherians. The identifica-

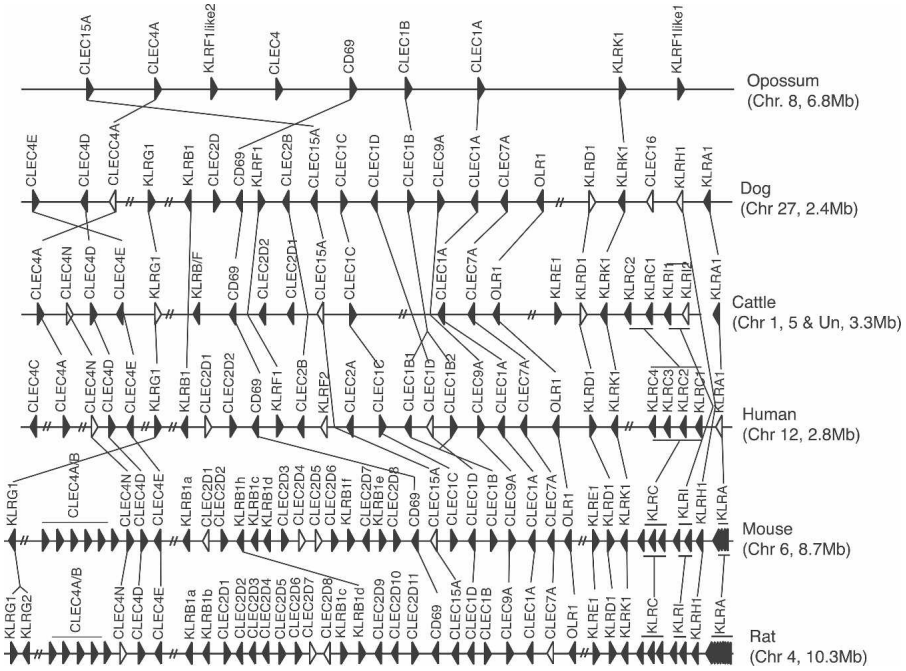

Figure 4. Comparative map showing opossum and eutherian NKC gene organization (modified from Hao et al. 2006). Arrows show transcriptional orientation. Pseudogenes are shown in white and putative functional genes in black.

tion of this gene cluster indicates that the NKC appeared prior to the divergence of the marsupial and eutherian lineages, but the paucity of C-type lectin genes within the opossum NKC when compared with eutherian mammals, and especially the rodent lineages, suggests that rapid gene duplication occurred within the NKC in certain eutherian mammals.

The presence of *KLRK1* in the opossum is interesting. *KLRK1* (*NKG2D*) is a homodimer that is expressed on NK cells, macrophages, and T cells (Kelley et al. 2005b). The ligands for *KLRK1* include the MHC Class I-associated molecules MICA and MICB (Sentman et al. 2006). Classical MHC Class I genes evolve rapidly and tend to form species-specific clusters on phylogenetic trees. Therefore, orthologous relationships are not seen between opossum and human MHC Class I genes (Miska and Miller 1999). *MIC* is an exception. A *MIC* ortholog is present in the MHC of the opossum (Belov et al. 2006), so it is perhaps not surprising that its receptor is also highly conserved. The opossum *KLRK1* sequence shares 60% amino acid identity with its human and 65% identity with its cattle, dog, and rat counterparts.

The leukocyte receptor complex (LRC)

The opossum LRC is located on chromosome 4. It contains a large expansion of Ig domains, which are found in five separate clusters (shown in Fig. 5). It appears that the automated gene builds were generally unable to correctly identify these genes. This is not surprising given the complex and variable nature of the eutherian and avian LRC genes. The human LRC (Chr 19q13.4) contains ~30 IgSF receptor genes (Martin et al. 2002). These genes can be divided into several multigene families based on gene organization, phylogeny, and structure. They include the killer cell immunoglobulin-like receptors (*KIR*) and leukocyte Ig-like receptors (*LILR*) and leukocyte-associated Ig-like receptors (*LAIR*), platelet glycoprotein VI (*GP6*), the natural cytotoxicity receptor 1 (*NCR1*), and the receptor for the IgA fragment (*FCAR*). The number of Ig-like domains in these molecules varies, as can be seen in Supplemental Figure 7. All of these genes are believed to have evolved from a common ancestral gene containing an Ig domain. The human LRC contains ~15 *KIR* genes. Different haplotypes contain different numbers of genes that are believed to have arisen as a result of nonreciprocal crossovers. The *KIR* genes evolve rapidly and share high levels of sequence similarity. Primate, cattle, and pig genomes also contain *KIR* genes within their LRC. The human LRC also contains 13 *LILR* genes.

The mouse LRC does not contain any *KIR* or *LILR* genes. They do, however, have an expansion of paired immunoglobulin-like receptor (*PIR*) genes. The PIR molecules are expressed on a wide range of immune cells and, like *KIR* and *LILR*, also interact with MHC Class I ligands. There are at least seven *PIR* genes in the mouse genome, and they contain six Ig domains (for review, see Takai and Ono 2001).

Similarly, the chicken does not contain any *KIR* or *LILR* homologs, but the chicken LRC contains 103 immunoglobulin-like activating and inhibitory receptors, which are tightly clustered on a microchromosome (Laun et al. 2006). These chicken immunoglobulin-like receptors (*CHIR*s) contain one or two Ig domains (Dennis et al. 2000).

The opossum LRC contains 154 Ig-like domains. Two Ig-like domains belong to *CEACAM1* (also known as *CD66*). Five *SIGLEC* genes were recognized (details in Supplemental Material), including *MAG* (*SIGLEC3*) and *CD22* (*SIGLEC2*). Adjacent to *CD22* are 124 Ig-like domains with similarity to *KIR* and *LILR* Ig-like domains. A further 33 similar domains are found on the unordered chromosome. It is interesting to note that in eutherian mammals, *KIR*s, *LILR*s, and *PIR*s have the same transcriptional orientation, while chicken *CHIR*s and marsupial Ig-like receptors do not. Database searches with these domains suggest that these genes have a common ancestry with avian *CHIR* and eutherian *KIR*, *LILR*, *GP6*, and *PIR* genes. However, BLAST searches cannot assign domains to gene families, as opossum LRC domains hit multiple eutherian LRC genes with significant *E*-values. Reciprocal blasts did not identify the original sequence. An inability to assign opossum genes to the level of gene family was consistently seen with all opossum Ig-like domains. Therefore, we used phylogenetic analyses to examine the evolutionary relationship of marsupial immunoglobulin-like receptors with their likely counterparts in eutherian mammals and birds.

Ig-like domains within the LRC belong to the C2-set and evolved from an ancestral element. Nikolaidis et al. (2005) have separated the mammalian LRC Ig-like domains into two groups: MI and MII types, and avian Ig-like domains into CI and CII types. They found that all mammalian *KIR* genes are composed of MII domains, *LILR* domains 2 through 4 are of the MII type, whereas *LILR* domain 1s are of the MI type. Chicken *CHIR* domain 1 sequences were CI type, whereas the second domain of CHIR molecules belonged to CII (summarized in Supplemental

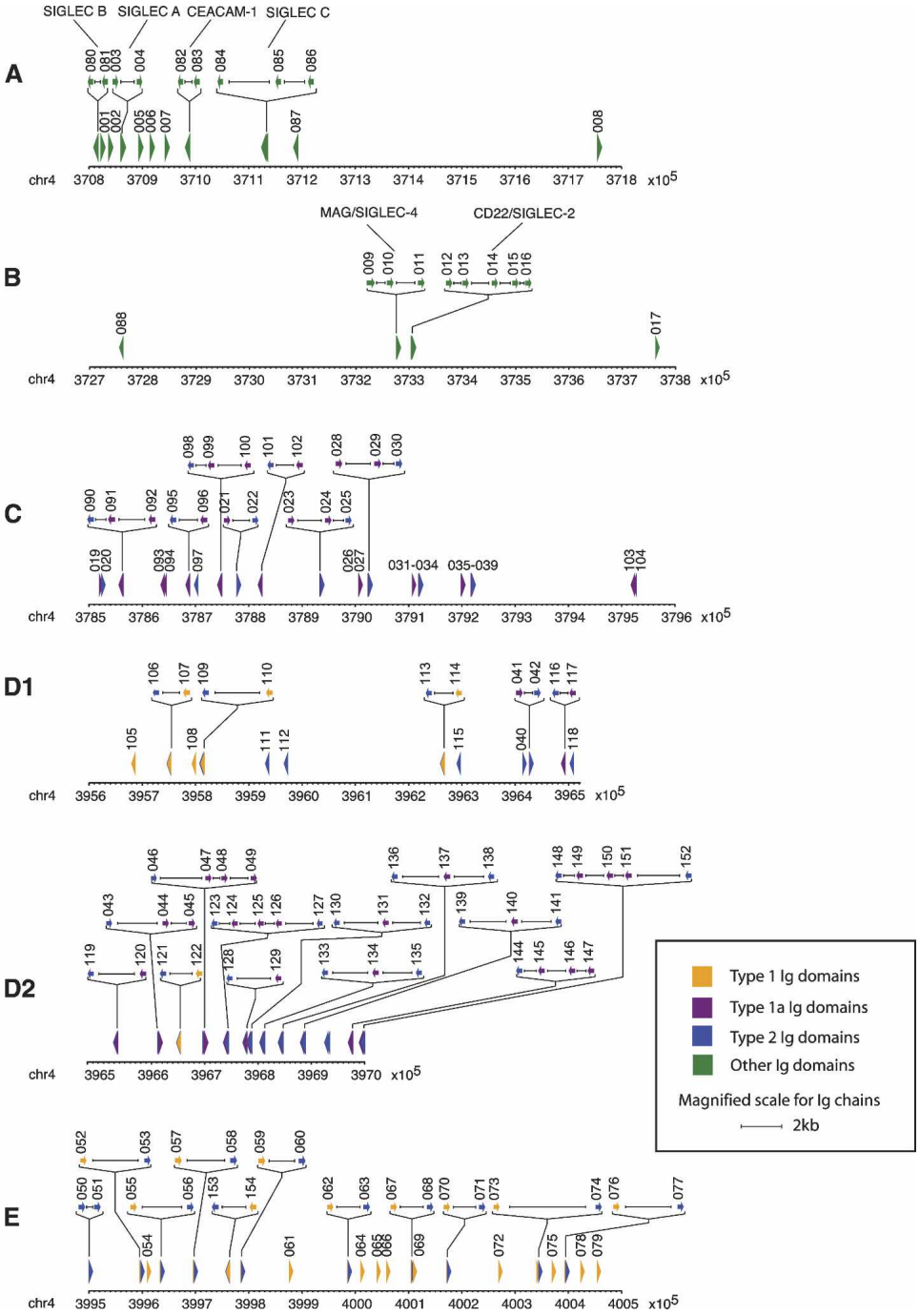

Figure 5. Manual curation of opossum LRC Ig-like domains to predict *MAIR* Ig domain structure. Domains are colored based on their location on the phylogenetic tree in Supplemental Figure 9.

Fig. 7). They proposed that a common ancestor for MI and CI genes existed before the separation of birds and mammals.

Phylogenetic analyses with 157 opossum Ig-like domains from chromosome 4 and the unordered chromosome were conducted. Our analyses confirmed that the opossum Ig-like domains are related to eutherian and avian LRC genes (Supplemental Fig. 9). The marsupial Ig-like receptors fell into three distinct clades on the tree. These clades have been designated I, Ia, and II. Clade I contains the eutherian MI and the chicken CI sequences. Clade II contains the eutherian MII and the chicken CII sequences. Clade Ia is a sister group to the MI and CI clades, and contains only opossum sequences. Caution should be applied when interpreting the results of this tree, as bootstrap support was low due to the short length of the sequences.

Due to the rapid nature of the evolution of the LRC genes, only three opossum Ig domains could be assigned to known avian or eutherian gene families. Ig-like domains 112, 113, and 114 intersperse with eutherian MII and MI domains. Gene pre-

diction around domains 112, 113, and 114 resulted in the identification of a gene that was most similar to *GP6*, based on BLAST searches. Phylogenetic analysis suggests that the predicted gene may be a *GP6* homolog (Supplemental Fig. 10). In eutherian mammals, *GP6* plays an important role in collagen-induced activation and aggregation of platelets. *GP6* has not been identified in chicken. Its presence in marsupials and eutherians indicates that it is not under the same selection pressure as other LRC-encoded genes.

The other 154 marsupial Ig-like domains form lineage specific clusters on the tree and do not intersperse within the eutherian *KIR* and *LILR* or avian *CHIR* Ig domains. Since these Ig-like domains could not be assigned to eutherian or avian gene families based on phylogeny, we have named them *MAIR*s (Marsupial immunoglobulin-like receptors).

We adopted a strategy for identifying chains of Ig-like domains rather than gene prediction for *MAIR*s, as the structure of related eutherian genes and gene transcripts is variable. We identified 45 putative *MAIR* open reading frames (ORFs) (38 on chromosome 4 and seven on the unordered chromosome), which may be composed of two to five domains. *MAIR*s containing only one Ig domain are also possible. The phylogenetic designation of the different domains is shown in Figure 5. It appears that two domain molecules can either contain domains of type I and II, resembling *CHIR*s, *NCR1*, *GP6*, and *FCAR*. Predicted ORFs containing domains of type Ia and II could feasibly resemble two domain *KIR*s, while three domain *MAIR*s may resemble three domain *KIR*s. Four domain ORFs were also predicted and may be counterparts of *LILR*s. cDNA data are not available for marsupial Ig-like receptors. Confirmation of these predictions will be conducted in the laboratory in the near future. This is particularly important given that the region appears rearranged when compared to humans, and there may be issues with the assembly in this region (M. Grabherr, pers. comm.). Moreover, this region is highly polymorphic and contains many SNPs and indels. Assembly of the LRC has also been problematic in other species (Laun et al. 2006), and we will not be able to make any firm conclusions on the structure of the region until the entire region has been sequenced from BACs.

Over the past 310 million years, Ig-like receptor evolution in the LRC has been dynamic, with no discernible orthology between domains in the birds, marsupials, and eutherians. The opossum LRC contains an expansion of Ig-like receptors that is larger than that in the human genome. This diversification is likely to have been driven by exposure to different pathogens or pathogen pressures on newborns, as NK cell lysis does not require prior exposure to pathogens.

Conclusions

The elucidation of specific differences in immune function between different mammalian lineages helps increase our understanding of the evolution and function of our complex immune system. Marsupials last shared a common ancestor with birds 310 million years ago and with eutherians 180 million years ago. These large evolutionary distances have meant that identification of homologous immune genes in the laboratory has been difficult. The availability of the opossum genome has allowed us to identify highly divergent immune genes in these distantly related mammals. We conclude that the genetic architecture of the marsupial immune system is very similar to that of eutherian mammals. It is interesting to note that the number of innate immune genes in the opossum genome is not significantly different from that in eutherians, indicating that gene expansion alone cannot account for protection of immunologically naïve young. Instead, it appears that different pathogen pressures in different environments have resulted in gene expansions and contractions of immune gene families across all the mammalian lineages, accounting for the amazing diversity and complexity of immune responses we see today.

Availability of this genomic sequence opens the door to a new era of research into immune responses of marsupials and, perhaps more importantly, the emergence of a new model organism for studies on mammalian developmental immunology. We predict that the immune system of the opossum is similar enough to that of humans, that comparative studies will allow us to gain important information about the development of the immune system and mechanisms of protection of immunologically naïve young. The opossum is an ideal model organism because they are small, housed in mouse cages, and are highly prolific (average litter size is eight). Moreover, experimental manipulation and tissue collection do not require invasive procedures on the mother.

Methods

Immunome

A total of 1528 known human immune proteins obtained from the IRIS database (Kelley et al. 2005a) were aligned against the opossum genome (MonDom4) using TBLASTN (Altschul et al. 1997). Blast HSPs were chained. This immunome annotation was visualized using Gbrowse (http://www.gmod.org/) with all chains and the best scoring chain made available in separate tracks. To provide context for these immune genes, 27,962 human RefSeq proteins were also aligned (RefSeq v17) against the opossum genome using TBLASTN, and the chain containing the best hit was extracted.

Highly diverged or missing immune genes that did not align (E-value > 0.01) or aligned poorly ($1e \times 10^{-5} < E$-value < 0.01) were sorted into gene families and, in some cases, more sensitive searches using synteny or profile hidden Markov models (HMMs) were undertaken. For example, the chemokines (CCL 1, 2, 7, 8, 11, 13, 15, 16, 18, 20, and 23–26) aligned poorly or not at all. HMMer (http://hmmer.janelia.org) was used to search for members of this gene family using interleukin 8-like chemokine local alignment profile HMM (PF00048) from Pfam (Finn et al. 2006). Hits to exons 2 and 3 in putative family members were chained and extracted for phylogenetic analysis. Gene prediction was not carried out.

Antimicrobials

To identify the location of putative beta-defensin genes in the opossum genome, the six frame translation was searched with HMMer using the beta-defensin profile HMM from Pfam (PF00711), which models the active peptide encoded in the second coding exon. Gene prediction was performed around high quality HMMer hits (E-value ≤ 0.1) using GenomeScan (Yeh et al. 2001). Extracted peptide sequence from the hit itself was supplied to GenomeScan in place of homologous sequence. Gene predictions were then curated, and only predictions with two coding exons that contained the six cysteine defensin motif in the second exon were retained. The beta-defensin profile HMM was updated to include the curated opossum sequences, and this process was repeated with the new profile.

In a few cases GenomeScan predicted long, multi-exon

genes, which included the defensin motif in an internal exon. To improve the predictions of GenomeScan in these cases we constructed a profile HMM of the peptide sequence from the first exons of our curated gene predictions and searched the region upstream of the putative second exon. The extracted peptide sequences from the putative first exon and the mature peptide fragment were provided to GenomeScan and gene prediction was repeated.

In addition to HMMer searches, 850 defensin protein sequences from GenBank and RefSeq (v17) were aligned with the opossum genome assembly (MonDom4) using TBLASTN. All significant blast matches (*E*-value $\leq 1 \times 10^{-5}$) that did not overlap a previously identified HMMer hit were extracted. Gene prediction was performed around these features using GenomeScan supplied with the homologous protein sequence.

A profile HMM was constructed using known alpha-defensin proteins from primates and rodents. This was used to search the six frame translation of the opossum genome. Gene prediction was performed around high quality hits using GenomeScan with human homolog, NP_004075.

To identify possible cathelicidin genes in the opossum genome, the six frame translation was searched with HMMer using the cathelicidin local alignment profile HMM from Pfam (PF00666). This models the conserved domains of the second and third exons. Hits to putative second and third exons were chained to reveal the location of possible cathelicidins. Gene prediction was performed around these features using GenomeScan supplied with NP_004336.2 as a homologous sequence.

Natural Killer cell receptors

A gene map of the NKC was constructed using RefSeq and IRIS protein blast features (see Immunome above). Additionally, we searched for C-type lectin (PF00059) and *KLRA* (PF08391) domains using HMMer. GenomeScan was used to predict NKC sequences.

The location of the LRC was identified using BLAST hits to *KIR*s, *LILR*s, and *SIGLEC*s from the immunome search and from the human–opossum synteny map. This region was then searched for immunoglobulin domains using HMMer and the Pfam immunoglobulin profile HMM (PF00047). High-quality HMMer hits (*E*-value ≤ 0.1) with open reading frames were extracted, and chains of Ig-like domains with gaps <2.5 kb were identified.

Acknowledgments

We thank Yixuan Liu (University of Sydney, Australia) for searching early genome assemblies for AMP genes. We also thank Mike Kamal for access to his synteny map, Manfred Grabherr for information about genome assembly, Nikolas Nikolaidis for sharing his LRC sequence alignments, Hugh Tyndale-Biscoe for helpful comments on the manuscript, and Kerstin Lindblad-Toh and the Broad Institute team for giving us an opportunity to work on the genome project and for their enthusiasm and advice. This work was supported by grants from the Australian Research Council and the University of Sydney to K.B. C.S. and E.W. are supported by scholarships provided by the ARC Centre for Kangaroo Genomics and the Jean Walker Trust.

References

Abi-Rached, L., Gilles, A., Shiina, T., Pontarotti, P., and Inoko, H. 2002. Evidence of en bloc duplication in vertebrate genomes. *Nat. Genet.* **31:** 100–105.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.

Baker, M.L., Rosenberg, G.H., Zuccolotto, P., Harrison, G.A., Deane, E.M., and Miller, R.D. 2001. Further characterization of T cell receptor chains of marsupials. *Dev. Comp. Immunol.* **25:** 495–507.

Belov, K., Deakin, J.E., Papenfuss, A.T., Baker, M.L., Melman, S.D., Siddle, H.V., Gouin, N., Goode, D.L., Sargeant, T.J., Robinson, M.D., et al. 2006. Reconstructing an ancestral mammalian immune supercomplex from a marsupial major histocompatibility complex. *PLoS Biol.* **4:** e46.

Burt, D.W. 2005. Chicken genome: Current status and future opportunities. *Genome Res.* **15:** 1692–1698.

Chang, C.I., Zhang, Y.A., Zou, J., Nie, P., and Secombes, C.J. 2006. Two cathelicidin genes are present in both rainbow trout (*Oncorhynchus mykiss*) and atlantic salmon (*Salmo salar*). *Antimicrob. Agents Chemother.* **50:** 185–195.

Dennis Jr., G., Kubagawa, H., and Cooper, M.D. 2000. Paired Ig-like receptor homologs in birds and mammals share a common ancestor with mammalian Fc receptors. *Proc. Natl. Acad. Sci.* **97:** 13245–13250.

Finn, R.D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., et al. 2006. Pfam: Clans, web tools and services. *Nucleic Acids Res.* **34:** D247–D251.

Gouin, N., Wright, A.M., Miska, K.B., Parra, Z.E., Samollow, P.B., Baker, M.L., and Miller, R.D. 2006. Modo-UG, a marsupial nonclassical MHC class I locus. *Immunogenetics* **58:** 396–406.

Hao, L., Klein, J., and Nei, M. 2006. Heterogeneous but conserved natural killer receptor gene complexes in four major orders of mammals. *Proc. Natl. Acad. Sci.* **103:** 3192–3197.

Jurd, R.D. 1994. "Not proper mammals": Immunity in monotremes and marsupials. *Comp. Immunol. Microbiol. Infect. Dis.* **17:** 41–52.

Kasahara, M. 1997. New insights into the genomic organization and origin of the major histocompatibility complex: Role of chromosomal (genome) duplication in the emergence of the adaptive immune system. *Hereditas* **127:** 59–65.

Kasahara, M., Watanabe, Y., Sumasu, M., and Nagata, T. 2002. A family of MHC class I-like genes located in the vicinity of the mouse leukocyte receptor complex. *Proc. Natl. Acad. Sci.* **99:** 13687–13692.

Kaufman, J., Milne, S., Gobel, T.W., Walker, B.A., Jacob, J.P., Auffray, C., Zoorob, R., and Beck, S. 1999. The chicken B locus is a minimal essential major histocompatibility complex. *Nature* **401:** 923–925.

Kelley, J., de Bono, B., and Trowsdale, J. 2005a. IRIS: A database surveying known human immune system genes. *Genomics* **85:** 503–511.

Kelley, J., Walter, L., and Trowsdale, J. 2005b. Comparative genomics of natural killer cell receptor gene clusters. *PLoS Genet.* **1:** 129–139.

Laun, K., Coggill, P., Palmer, S., Sims, S., Ning, Z., Ragoussis, J., Volpi, E., Wilson, N., Beck, S., Ziegler, A., et al. 2006. The leukocyte receptor complex in chicken is characterized by massive expansion and diversification of immunoglobulin-like loci. *PLoS Genet.* **2:** e73.

Lynn, D.J., Higgs, R., Gaines, S., Tierney, J., James, T., Lloyd, A.T., Fares, M.A., Mulcahy, G., and O'Farrelly, C. 2004. Bioinformatic discovery and initial characterisation of nine novel antimicrobial peptide genes in the chicken. *Immunogenetics* **56:** 170–177.

Martin, A.M., Kulski, J.K., Witt, C., Pontarotti, P., and Christiansen, F.T. 2002. Leukocyte Ig-like receptor complex (LRC) in mice and men. *Trends Immunol.* **23:** 81–88.

Maruoka, T., Tanabe, H., Chiba, M., and Kasahara, M. 2005. Chicken CD1 genes are located in the MHC: CD1 and endothelial protein C receptor genes constitute a distinct subfamily of class-I-like genes that predates the emergence of mammals. *Immunogenetics* **57:** 590–600.

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

Miller, R.D. and Belov, K. 2000. Immunoglobulin genetics of marsupials. *Dev. Comp. Immunol.* **24:** 485–490.

Miska, K.B. and Miller, R.D. 1999. Marsupial Mhc class I: Classical sequences from the opossum, *Monodelphis domestica*. *Immunogenetics* **50:** 89–93.

Nikolaidis, N., Klein, J., and Nei, M. 2005. Origin and evolution of the Ig-like domains present in mammalian leukocyte receptors: Insights from chicken, frog, and fish homologues. *Immunogenetics* **57:** 151–157.

Ohta, Y., Goetz, W., Hossain, M.Z., Nonaka, M., and Flajnik, M.F. 2006. Ancestral organization of the MHC revealed in the amphibian *Xenopus*. *J. Immunol.* **176:** 3674–3685.

Old, J.M. and Deane, E.M. 2000. Development of the immune system

and immunological protection in marsupial pouch young. *Dev. Comp. Immunol.* **24:** 445–454.
Patil, A., Hughes, A.L., and Zhang, G. 2004. Rapid evolution and diversification of mammalian alpha-defensins as revealed by comparative analysis of rodent and primate genes. *Physiol. Genomics* **20:** 1–11.
Patil, A.A., Cai, Y., Sang, Y., Blecha, F., and Zhang, G. 2005. Cross-species analysis of the mammalian beta-defensin gene family: Presence of syntenic gene clusters and preferential expression in the male reproductive tract. *Physiol. Genomics* **23:** 5–17.
Renfree, M.B. and Tyndale-Biscoe, C. H. 1978. Manipulation of marsupial embryos and pouch young. In *Methods in mammalian reproduction* (ed. J.C. Daniel), pp. 307–331. Academic Press, New York.
Rogers, S.L., Gobel, T.W., Viertlboeck, B.C., Milne, S., Beck, S., and Kaufman, J. 2005. Characterization of the chicken C-type lectin-like receptors B-NK and B-lec suggests that the NK complex and the MHC share a common ancestral region. *J. Immunol.* **174:** 3475–3483.
Samples, N.K., Vandeberg, J.L., and Stone, W.H. 1986. Passively acquired immunity in the newborn of a marsupial (*Monodelphis domestica*). *Am. J. Reprod. Immunol. Microbiol.* **11:** 94–97.
Sentman, C.L., Barber, M.A., Barber, A., and Zhang, T. 2006. NK cell receptors as tools in cancer immunotherapy. *Adv. Cancer Res.* **95:** 249–292.
Smith, J., Speed, D., Law, A.S., Glass, E.J., and Burt, D.W. 2004. In-silico identification of chicken immune-related genes. *Immunogenetics* **56:** 122–133.
Stone, W.H., Bruun, D.A., and Manis, G.S. 1996. The immunobiology of the marsupial *Monodelphis domestica*. In *Modulators of immune responses, the evolutionary trail* (eds. J.S. Stolen et al.), pp. 149–165. SOS Publications, Fairhaven, NJ.
Takai, T. and Ono, M. 2001. Activating and inhibitory nature of the murine paired immunoglobulin-like receptor family. *Immunol. Rev.* **181:** 215–222.
Teng, M.S., Stephens, R., Du Pasquier, L., Freeman, T., Lindquist, J.A., and Trowsdale, J. 2002. A human TAPBP (TAPASIN)-related gene, TAPBP-R. *Eur. J. Immunol.* **32:** 1059–1068.
Tyndale-Biscoe, H. 2005. *Life of Marsupials*. CSIRO Publishing, Collingwood, Australia.
Tyndale-Biscoe, C.H. and Renfree, M. 1987. *Reproductive physiology of marsupials*. Cambridge University Press, Cambridge.
Uzzell, T., Stolzenberg, E.D., Shinnar, A.E., and Zasloff, M. 2003. Hagfish intestinal antimicrobial peptides are ancient cathelicidins. *Peptides* **24:** 1655–1667.
Vienne, A., Shiina, T., Abi-Rached, L., Danchin, E., Vitiello, V., Cartault, F., Inoko, H., and Pontarotti, P. 2003. Evolution of the proto-MHC ancestral region: More evidence for the plesiomorphic organisation of human chromosome 9q34 region. *Immunogenetics* **55:** 429–436.
Wild, A.E., Burrows, T.D., and Brand, J. 1994. IgG transport across the gut of the suckling opossum (*Monodelphis domestica*). *Dev. Comp. Immunol.* **18:** 75–88.
Wong, E., Young, L.J., Papenfuss, A.T., and 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.
Xiao, Y., Hughes, A.L., Ando, J., Matsuda, Y., Cheng, J.F., Skinner-Noble, D., and Zhang, G. 2004. A genome-wide screen identifies a single beta-defensin gene cluster in the chicken: Implications for the origin and evolution of mammalian defensins. *BMC Genomics* **5:** 56.
Xiao, Y., Cai, Y., Bommineni, Y.R., Fernando, S.C., Prakash, O., Gilliland, S.E., and Zhang, G. 2006. Identification and functional characterization of three chicken cathelicidins with potent antimicrobial activity. *J. Biol. Chem.* **281:** 2858–2867.
Yeh, R.F., Lim, L.P., and Burge, C.B. 2001. Computational inference of homologous gene structures in the human genome. *Genome Res.* **11:** 803–816.
Young, L.J. and Deane, E.M. 2003. Studying marsupial immune function. *Today's Life Science* **1:** 34–36.

Received November 14, 2006; accepted in revised form February 5, 2007.