Published in final edited form as:

Immunogenetics. 2006 June ; 58(0): 362–373. doi:10.1007/s00251-006-0112-7.

Ancient divergence of a complex family of immune-type receptor genes

John P. Cannon,

Department of Pediatrics, University of South Florida College of Medicine, and USF/ACH Children's Research Institute, 830 First Street South, St. Petersburg, FL 33701, USA

Immunology Program, H. Lee Moffitt Cancer Center and Research Institute, 12902 Magnolia Avenue, Tampa, FL 33612, USA

Robert N. Haire,

Department of Pediatrics, University of South Florida College of Medicine, and USF/ACH Children's Research Institute, 830 First Street South, St. Petersburg, FL 33701, USA

Immunology Program, H. Lee Moffitt Cancer Center and Research Institute, 12902 Magnolia Avenue, Tampa, FL 33612, USA

M. Gail Mueller,

Department of Molecular Genetics, All Children's Hospital, 801 Sixth Street South, St. Petersburg, FL 33701, USA

Ronda T. Litman,

Department of Pediatrics, University of South Florida College of Medicine, and USF/ACH Children's Research Institute, 830 First Street South, St. Petersburg, FL 33701, USA

Donna D. Eason,

Department of Pediatrics, University of South Florida College of Medicine, and USF/ACH Children's Research Institute, 830 First Street South, St. Petersburg, FL 33701, USA

Immunology Program, H. Lee Moffitt Cancer Center and Research Institute, 12902 Magnolia Avenue, Tampa, FL 33612, USA

Deborah Tinnemore,

Molecular Genetics Program, Benaroya Research Institute at Virginia Mason, 1201 Ninth Avenue, Seattle, WA 98101, USA

Chris T. Amemiya,

Molecular Genetics Program, Benaroya Research Institute at Virginia Mason, 1201 Ninth Avenue, Seattle, WA 98101, USA

Tatsuya Ota, and

Department of Biosystems Science and Hayama Center for Advanced Studies, The Graduate University for Advanced Studies (Sokendai), Hayama, Kanagawa 240-0193, Japan

Gary W. Litman

Department of Pediatrics, University of South Florida College of Medicine, and USF/ACH Children's Research Institute, 830 First Street South, St. Petersburg, FL 33701, USA

Immunology Program, H. Lee Moffitt Cancer Center and Research Institute, 12902 Magnolia Avenue, Tampa, FL 33612, USA

Department of Molecular Genetics, All Children's Hospital, 801 Sixth Street South, St. Petersburg, FL 33701, USA, litmang@allkids.org, Tel.: +1-727-5533602, Fax: +1-727-5533610

Abstract

Multigene families of activating/inhibitory receptors belonging to the immunoglobulin superfamily (IgSF) regulate immunological and other cell–cell interactions. A new family of such genes, termed modular domain immune-type receptors (MDIRs), has been identified in the clearnose skate (*Raja eglanteria*), a phylogenetically ancient vertebrate. At least five different major forms of predicted MDIR proteins are comprised of four different subfamilies of IgSF ectodomains of the intermediate (I)- or C2-set. The predicted number of individual IgSF ectodomains in MDIRs varies from one to six. MDIR1 contains a positively charged transmembrane residue and MDIR2 and MDIR3 each possesses at least one immunoreceptor tyrosine-based inhibitory motif in their cytoplasmic regions. MDIR4 and MDIR5 lack characteristic activating/inhibitory signalling motifs. MDIRs are encoded in a particularly large and complex multigene family. MDIR domains exhibit distant sequence similarity to mammalian CMRF-35-like molecules, polymeric immunoglobulin receptors, triggering receptors expressed on myeloid cells (TREMs), TREM-like transcripts, NKp44 and FcR homologs, as well as to sequences identified in several different vertebrate genomes. Phylogenetic analyses suggest that MDIRs are representative members of an extended family of IgSF genes that diverged before or very early in evolution of the vertebrates and subsequently came to occupy multiple, fully independent distributions in the present day.

Keywords

Ig superfamily; Evolution; Activating/inhibitory signalling

Introduction

The diversification of the immunoglobulin gene superfamily (IgSF) during metazoan evolution has produced an extremely complex set of proteins with varying roles including neuronal development, muscle cell function, and specific immune recognition. Furthermore, somatic changes in the rearranging immunoglobulin (Ig) and T-cell antigen receptor gene families give rise to an excess of 10^{14} receptor specificities and constitute the basis for adaptive immunity in jawed vertebrates. The mammalian IgSF activating/inhibitory leukocyte regulatory receptors also serve many roles in immune cells, including mediation of NK cell activity (Lanier 2001; McQueen and Parham 2002; Parham 2005; Vivier and Malissen 2005), recognition of Ig Fc regions (Davis et al. 2002a; Ravetch and Lanier 2000) and other functions that likely remain to be identified. Activating/inhibitory immune-type proteins constitute the single most diverse group of the mammalian IgSF receptors; however, their genetic regulation, developmental control, functional interactions, and evolutionary origins are understood only in part. One line of investigation that eventually could shed light on several of these issues is to characterize counterparts of these genes in species that occupy representative positions in vertebrate phylogeny (Nikolaidis et al. 2005; Vierterlboeck et al. 2005).

The variable (V) region-containing novel immune-type receptor (NITR) genes have been described in several species of bony fish, including two prominent genome model systems, pufferfish (Strong et al. 1999) and zebrafish (Yoder et al. 2004). The NITRs are the largest single family of activating/inhibitory leukocyte regulatory receptor genes that has been described to date (Litman et al. 2001, 2005). A short primer PCR strategy was used to search for counterparts of bony fish NITRs in clearnose skate (*Raja eglanteria*), a cartilaginous fish

that diverged in evolution before the bony fish. Although it was not possible to identify NITR transcripts, another broadly divergent, heretofore-unrecognized multigene family encoding activating/inhibitory IgSF members, termed modular domain immune-type receptors (MDIRs), was identified. No gene or protein bearing significant similarity across the entire length of any MDIR has been described outside of cartilaginous fish; however, individual MDIR domains exhibit varying degrees of sequence similarity with a variety of mammalian proteins, as well as with other Ig domains that are encoded in bony fish, amphibian, and avian genomes. Phylogenetic analyses of the newly discovered genes suggest that over the course of evolution of jawed vertebrates, extensive domain duplication, exchange, and divergence likely occurred among a major supra-set of IgSF members involved in cell–cell interactions and immune-type functions.

Materials and methods

Amptrap PCR cloning

Amptrap RACE PCR was performed with skate spleen mRNA/cDNA as described previously (Cannon et al. 2002, 2005). Briefly, seven degenerate oligonucleotides (VYWFR-SfiIB, VYWF-SfiIB, YWFK-SfiIB, YWFR-SfiIB, WFK-SfiIB, WFR1-SfiIB, WFR2-SfiIB; Table 1A) were designed to include conserved amino acid codons in teleost NITRs, centered around the invariant IgSF tryptophan (Strong et al. 1999; Yoder et al. 2001), as well as an *Sfi*IB restriction recognition site. For PCR, each 3′-primer was paired with the SMART 5′-RACE anchor oligonucleotide, which contains an *Sfi*IA site. After amplification, unfractionated PCR products were digested with *Sfi*I, size separated using Chromaspin-400 gel filtration columns (BD Biosciences Clontech, Palo Alto, CA, USA) and ligated to the *Sfi*I-linearized Amptrap vector G7311, containing *Sfi*IA and *Sfi*IB-compatible ends (Cannon et al. 2002). Ligated plasmids were used to electrotransform *Escherichia coli* strain DH10β. Cloned skate cDNAs encoding start codons and secretion signal peptides in frame with the desired amino acid motifs were selected by growth on LB plates containing kanamycin and ampicillin. Individual cDNA clones were sequenced at random and surveyed for IgSF characteristics.

Full-length cDNA isolation

cDNA libraries representing spleen and thymus mRNA from single-skate specimens, constructed in the λZAPII vector (Stratagene, La Jolla, CA, USA), were screened by standard hybridization procedures using ³²P-labeled partial MDIR cDNA probes. MDIR-positive phagemids were excised in vivo using filamentous helper phage and purified from bacterial cultures using Qiagen Miniprep columns (Qiagen, Valencia, CA, USA).

DNA sequencing and genomic Southern blot analysis

Both methods were performed as described previously by our laboratory (Cannon et al. 2002). Hybridization probes were amplified by PCR using various MDIR cDNAs as templates, purified by agarose gel electrophoresis, and labeled with ³²P-dCTP. Oligonucleotide primer pairs used for generating probes are: MDIR1IgD1: MDIR1-IgD-F + MDIR1-IgD-R; MDIR2IgD1: MDIR2-IgD1-F + MDIR2-IgD1-R; MDIR2IgD2: MDIR2-IgD2-F + MDIR2-IgD2-R; MDIR4-IgD2: MDIR4-IgD2-F + MDIR4-IgD2-R (Table 1B).

Construction, screening, and characterization of skate MDIR bacterial artificial chromosome (BAC) clones

The genome size of *R. eglanteria* is estimated to be 3.8×10⁹ bp/haploid genome. A BAC library (pBAC-VMRC9) representing ~5–6X coverage and consisting of ~120 kb average insert length has been produced from a single male individual by adaptation of standard

procedures (Amemiya et al. 1996; Miyake and Amemiya 2004). Library screening (Amemiya et al. 1996), BAC DNA isolation, end-labeling and domain-type-specific hybridization were carried out as described (Yoder et al. 2001). Automated fingerprint analyses of restriction fragments of BAC DNA (Marra et al. 1997) were carried out by the Genome Sciences Centre, Vancouver, British Columbia. Oligonucleotide probes for refining of contig assignments were synthesized by Operon Biotechnologies (Huntsville, AL, USA), labeled with ^{32}P -dCTP using terminal deoxynucleotidyl transferase and hybridized to Southern blots of BAC DNA restriction fragments (6X SSC, 5X Denhardt's solution, 0.5% sodium dodecyl sulfate, 0.05% sodium pyrophosphate, 50°C).

PCR haplotyping

For PCR "haplotyping" of MDIR-encoding exons, BAC DNA was isolated from 5 ml of an overnight culture using standard alkaline lysis procedures and re-dissolved in 50 μl of TE buffer. One microliter of a 1:10 dilution of each DNA sample was used as a template for a PCR reaction using (a) TypeIIcon-F1 + either TypeIIcon-R1 or TypeIIcon-R2 oligonucleotides for I-type domains or (b) TypeIIIcon-F1 + either TypeIIIcon-R1 or TypeIIIcon-R2 oligonucleotides for C2-type domains (Table 1B). Amplicons were cloned in pGEM-T (Promega, Madison, WI, USA) and sequenced to assess the complexity of exons amplified from each BAC.

Phylogenetic analyses

Amino acid sequences of homologous genes to MDIRs, identified by BLASTP analyses, were retrieved from GenBank. Sequences were progressively aligned by the CLUSTALW program (Thompson et al. 1994) and adjusted manually by consideration of pairwise alignments obtained by BLASTP. Phylogenetic analyses were conducted by the neighbor-joining (NJ) method (Saitou and Nei 1987) implemented in MEGA3 software (Kumar et al. 2004).

Results and discussion

MDIRs are a diverse set of predicted IgSF proteins in a cartilaginous fish

Seven oligonucleotides complementing sequences that include the conserved V-domain tryptophan codon of available NITR gene sequences were used to prime skate spleen cDNA (Table 1A). Subsequent Amptrap PCR and selection yielded two different amplification products, each encoding a secretion signal peptide plus approximately 35 N-terminal amino acids of an IgSF domain. Hybridization probes complementing these partial IgSF coding sequences were used to screen skate spleen and thymus cDNA libraries. Five distinct, full-copy-length cDNA sequences were identified, encoding a diverse set of predicted IgSF transmembrane proteins; at least one transcript appears to be alternatively spliced. Due to sequence similarity between individual IgSF domains from these molecules and domains from multiple different mammalian immune-type receptor genes (described further below), this set of genes has been termed MDIRs (Fig. 1). The general features of these genes suggest that they belong to a single family, which is supported by subsequent hybridization and domain-specific sequence analyses (see below).

MDIR1 contains a basic amino acid in its transmembrane domain (YYVIWNIMRWIFLVILLVWGIVT; predicted using the TMHMM algorithm), potentially allowing interaction with activating membrane-signaling adaptors containing acidic transmembrane residues. The cytosolic domain of MDIR2 contains a consensus immunoreceptor tyrosine-based inhibitory motif (ITIM) (M YAA V), as well as two additional ITIM-like motifs (YAR V, YAS V). The cytosolic domain of MDIR3 contains three consensus ITIMs (IT YAV V, VV YAD V and SI YAS V). MDIR4 and MDIR5 contain

positively charged amino acids near both ends of their predicted transmembrane regions; however, only one of two prediction algorithms (TMHMM) predicts that either basic residue resides in the membrane. Cytoplasmic signaling motifs are not evident in MDIR4 or MDIR5.

Individual MDIR IgSF domains can be classified into four distinct sequence types based on their level of amino acid identity (Fig. 1). The single IgSF domain of MDIR1 (type I), the N-terminal IgSF domains of MDIR2 through MDIR5 (type II) and certain other domains of MDIR3 through MDIR5 (also type II) are best classified as intermediate (I)-set IgSF domains (Harpaz and Chothia 1994). In contrast, the membrane-proximal domains of MDIR2 through MDIR5 (type III) and four other Ig domains seen in MDIR3 through MDIR5 (type IV) are of the C2-set. The I-set domains demonstrate very low (4–13%) amino acid identity with the C2-set domains; however, the type III C2-set domains share 52–63% amino acid identity with the C2-set domains of type IV. In addition, the type I domain of MDIR1 shares 31–40% amino acid identity with the other I-set MDIR domains (of type II). Designating MDIR1 as belonging to the same family as MDIRs 2–5 is based on this similarity, as well as the linkage of types I and II domains in several BACs (see below).

Genomic complexity of the MDIRs

Genomic Southern blotting and hybridization with probes complementing the types I–IV MDIR IgSF domains indicate that the type I domain is encoded at a single locus with distinguishable alleles or at relatively few loci, whereas the other domain types (II–IV) likely are encoded at multiple loci (Fig. 2). The less complex hybridization patterns evident with the probe complementing the type III domain relative to types II and IV could be due to either a lower number or more divergent forms of type III exons in the skate genome. Although Southern blotting patterns in large, diverse multigene families, particularly those in which individual genes can consist of divergent domains of the same types, are complicated to interpret, the extensive genomic complexity implied in Fig. 2 cannot easily be attributed to cross-hybridization between the different domains, as only the types III and IV probe sequences exhibit 70% or more identity at the nucleotide level (78% across a region of 116 nucleotides). The overall complexity seen in the hybridization patterns is consistent with the complexity of the MDIR locus/loci inferred from BAC analyses (see below).

Screening of a *R. eglanteria* BAC library with MDIR domain type I- to type IV-specific probes provides an estimate of the size and complexity of the MDIR-encoding locus/loci (Table 2). From a total of 68 MDIR-positive BACS, six hybridized to the MDIR1-specific probe, consistent with MDIR1 being encoded at a single locus (or very few loci). In addition, a total of 58 BACs hybridized to a type III domain-specific probe. Given that all known multi-domain MDIRs contain only a single type III domain, this number suggests that roughly 10–12 MDIR2- to MDIR5-related genes are present in the genome. However, the potential complexity of MDIR-hybridizing fragments within single BACs coupled with the close sequence identity between certain individual MDIR domain families makes definitive enumeration of MDIR genes or projection of the structure of the MDIR gene locus/loci difficult. Such assignments will require whole genome sequencing and detailed sequence annotation (Yoder et al. 2004).

To survey potential MDIR-encoding exons (domain types II and III) represented in isolated BACs, DNA preparations from 28 individual BAC clones were subjected to PCR “haplotyping” using consensus oligo-nucleotides. Although this analysis is not at saturation, CLUSTAL W (Thompson et al. 1994) alignment of the translated amplicons demonstrates the presence of at least 13 distinct type II, nine type III, and two type IV MDIR exons, as

well as three likely type III pseudoexons, in the individual genome from which the BAC library was constructed (Fig. 3).

MDIR-positive BACs were subjected to DNA end-labeling and automated fingerprinting after restriction enzyme digestion (data not shown), resulting in the definition of 10 contigs, comprised of two to ten constituent BACs each. Preliminary assignment of short overlaps among individual BACs within separate contigs implies that several of the ten current contigs may reside in a larger “supercontig”, although definitive conclusions will require larger-scale BAC sequencing. Two BACs represent single-tons within the limits of interpretation of currently available restriction enzyme digestion-hybridization patterns. Of note, the BACs hybridizing with both types I and II domain probes physically link the MDIR1-type gene(s) to a type II exon.

MDIRs are a large and diverse set of IgSF molecules that likely diverged early in vertebrate evolution

BLASTP database searches using full-length predicted MDIR proteins failed to produce high-confidence matches across the entire length of any MDIR. Subsequent BLASTP homology searches were conducted independently for representative members of each MDIR IgSF domain subclass to examine the relationships of MDIRs to other proteins at the level of individual domains. Depending on the query sequence used, the domain-specific analyses produced two distinct groups of matches. Types I and II MDIR domains consistently identified various mammalian CMRF-35-related molecules [CMRF-35, myeloid-associated Ig-like receptors [MAIRs] I/II (Yotsumoto et al. 2003), CMRF-35-like molecules [CLMs] (Chung et al. 2003; Jackson et al. 1992) and immune receptor expressed on myeloid cells 1 (Alvarez-Errico et al. 2004) and 2 (Aguilar et al. 2004)] and several mammalian polymeric immunoglobulin receptors (pIgRs) (Fig. 4).

The types I and II MDIR domains also bear resemblance to the extracellular domains of NKp44 (Cantoni et al. 1999), triggering receptors expressed on myeloid cells (TREMs) and TREM-like transcripts (TLTs) (Colonna 2003), Fc α / μ receptors (Shibuya et al. 2000), novel Ig-like transcripts from the carp (Stet et al. 2005), hepatitis A virus cellular receptor (Shakhov et al. 2004), and adenosine A3 receptor (Salvatore et al. 1993). Notably, a large number of the Ig domains identified using types I and II domain query sequences appear to possess distinct second disulfide bonds at the corresponding C–C' beta strand region (Cantoni et al. 2003; Hamburger et al. 2004) as well as Ig V-set signature residues and lengths. In addition, the amino acid motif KYWC, which is found at the conserved IgSF tryptophan residue in the type I and most of the type II domains of skate MDIRs, is also found in the IgSF domain of the Fc α / μ receptor (Shibuya et al. 2000) and in various domains of mammalian pIgRs and CMRF-35/MAIR/CLM (Fig. 4). The KYW segment appears essential for the interaction of pIgR with its immunoglobulin ligands (Bakos et al. 1991). NKp44 and certain TREMs/TLTs share the broader KXWC motif at this position.

BLASTP searches using isolated types III and IV MDIR domains identified regions of distant sequence relationships to certain mammalian Fc receptor homologs (FcRHs) (Davis et al. 2001). In one representative case, BLASTP analysis using the sixth IgSF domain of MDIR4 as a query identified similarity to the fifth IgSF domain of human FcRH5 (BLAST E value=4 $\times$ 10⁻⁷). In addition to FcRHs, a protein with unknown function (chicken XP_415669, a possible homolog of human LOC284021 on chromosome 17q23.3 or rat mast cell antigen 32), PECAM1, titin, CTH/CTM, CD22, and CD169 also were identified, although their E values are higher than those of FcRHs. Thus, while it appears certain from BLAST and phylogenetic analyses (see below) that CMRF-35/MAIR/CLM, pIgR, Fc α / μ R, and TREM/TLT/NKp44 of higher vertebrates diverged before or at an early stage of vertebrate evolution, the evolutionary relationships of all of these proteins to the skate

MDIRs cannot be resolved at this time. MDIRs may represent an independent, entirely distinct lineage within this very large family of IgSF proteins.

MDIR-related sequences in other species

Peptide sequences of individual skate MDIR domains identified MDIR types I and II domain-related sequences in TBLASTN queries of currently available genome sequence assemblies from human, mouse, chicken, Western clawed frog, and zebrafish. Delineation of distinct MDIR-related ESTs is difficult in genomic sequences in chicken, Western clawed frog, and zebrafish, not unlike the low representation seen in our efforts to recover MDIR cDNAs in skate. Searches using MDIR types III and IV domains generally did not produce high-confidence matches; no genomic loci with significant similarity across the entire length of any skate MDIR were identified, even permitting relatively loose peptide identity.

In contrast to the other vertebrate genomes examined, which generally yield between 10 and 30 MDIR type-I-and type-II-related exons in TBLASTN searches, >100 candidate exons encoding MDIR type-I- and type-II-like domains were identified in the partial zebrafish genome (ZV5 Ensemble v34, http://www.ensembl.org/Danio_rerio/index.html). The most extensive set of these sequences is located in two distinct clusters in the two megabases (Mb) spanning the 45- to 46-Mb region of chromosome 2. Zebrafish relatives of the carp novel Ig-like transcript genes (Stet et al. 2005) also are located within this region, implying not only a potential relationship of MDIRs to novel Ig-like transcripts but also a relationship of novel Ig-like transcripts to a much larger family of zebrafish genes. Subsequent TBLASTN searches of nearby genomic sequence identified a third cluster, in the 47-Mb region, with distant similarity to the C2-set MDIR types III and IV domains.

It is important to recognize that allelic complexity confounds the assembly process in the current Zebrafish Genome Project, potentially resulting in artificial increases in the numbers of genes assigned to individual loci, as described previously for NITR genes (Yoder et al. 2004). TBLASTN searches for MDIR-like exons in two pufferfish genome models, *Takifugu rubripes* and *Tetraodon nigroviridis*, thus far have yielded approximately 10–20 sequences related to types I and II domains and very few sequences related to types III and IV domains in both species. However, the assemblies of both *Takifugu* and *Tetraodon* remain incomplete and considerably fragmented, a particular disadvantage in the analysis of large, closely interrelated gene families within a species. The question of whether the numerous zebrafish MDIR-like sequences represent a very large gene complex in teleost fish or, rather, reflect a very high degree of allelic polymorphism in a somewhat smaller complex must await the complete resolution of the zebrafish, pufferfish, and possibly other teleost genomes.

MDIRs exhibit particularly complex phylogenetic relationships

Reduced homology in pairwise BLAST comparisons and correspondingly higher *E* values can be due to rapid rates of sequence change, rather than to extended passage of time since divergence. In addition, the lengths of sequences available for comparison have some effect on the *E* values obtained. Thus, lower BLAST *E* values do not necessarily imply closer evolutionary or functional relationships. To confer overall statistical inferences to the projected patterns of MDIR evolution relative to other vertebrate Ig domains, phylogenetic analyses were performed. To improve confidence in the conclusions: 1) only sequences with *E* values less than 10^{-1} by BLASTP analyses were used; 2) those sequences that were difficult to align due to insertions/ deletions or absence of critical residues were eliminated to avoid ambiguities in sequence alignment; and 3) divergent sequences encoding peptides with extensive deletions or lacking the classical Ig-fold disulfide bond cysteine residues were eliminated.

The N-terminal (types I and II) IgSF domains of MDIR1 and MDIR2-5 associate weakly with vertebrate TREM4/ CLM9 and the other vertebrate CMRF-35/MAIR/CLM family members, respectively (Fig. 5, left side). Because this tree is unrooted and the results are not supported strongly by bootstrap analysis, the evolutionary origin of sequences is not implied. Thirty-eight zebrafish MDIR-like domains, representing the V/I-set exon clusters from the 45- and 46-Mb regions of chromosome 2, and one MDIR-like domain encoded on zebrafish chromosome 9, were also included. Each of the two clusters from zebrafish chromosome 2 occupies a distinct branch of the tree, although all of the zebrafish domains included in the tree exhibit distant relationships to skate MDIR type I/II domains (Fig. 5, right side). Within the branch of the tree that contains the skate MDIRs, MDIRs 2-5 form a single family that exhibits modest intergenic sequence divergence. Analysis of nucleotide substitution frequencies (Ps/Pn) among individual domains implies that MDIR1 diverged very early from the other MDIRs, and that MDIR2 likely separated subsequently from MDIR3/ MDIR4/MDIR5 (Table 3).

MDIR types III and IV domains did not generate trees of sufficient significance to allow firm conclusions (Fig. 6), although (as indicated above) the mammalian FcRH proteins were found to associate weakly with these C2-set MDIR domains. Furthermore, a set of 24 representative C2-set domains encoded by the 47-Mb region of zebrafish chromosome 2 (linked tightly to the zebrafish type I/II MDIR-like exons) also distantly associates with the skate MDIR domains. Nevertheless, given the relatively low significance of this tree, it remains possible that the C2-set domains of MDIRs represent a lineage unique to cartilaginous fish.

Overall organizational and sequence motifs among MDIR-related proteins

The relatively high degree of amino acid sequence variation in the constituent domains of MDIRs also is observed in CMRF-35/MAIR/CLM, TREM/TLT, and FcRH proteins. CMRF-35-like Ig domains generally bear the closest sequence identity of any mammalian proteins to skate MDIRs. Although the CMRF-35 locus has been linked to psoriasis susceptibility (Speckman et al. 2003), molecular ligands of the CMRF-35/MAIR/CLM family are currently unknown. The products of MDIR and MDIR-like genes also bear similarity to pIgR and Fc α / μ R proteins; however, it is unclear whether or not their functions are related to interactions with Ig constant regions. The extraordinary heavy chain “isotype” complexity in cartilaginous fish (Litman et al. 1999), nevertheless, raises the possibility that Fc receptors in these species may be more extensive and diverse.

The MDIR domains exhibit similarity to many pIgRs, and there is very distant similarity of the MDIR types III and IV domains to FcRHs, which likely function in B-cell regulation (Davis et al. 2001, 2002b, 2005; Ehrhardt et al. 2003, 2005; Leu et al. 2005); however, the membrane and cytosolic features of MDIRs are generally more similar to those seen in CMRF-35/MAIR/CLM and FcRH molecules than to those of pIgR. Although similarity in expression profiles could potentially reveal relationships to the expression patterns of complex activating/inhibitory IgSF members found in other vertebrates, the only significant MDIR gene expression that we have observed to date is in the epigonal organ, which contains B and T lymphocytes, as well as appreciable numbers of myeloid cells, and lacks an equivalent tissue in other vertebrates (unpublished observation). Notably, the expression of MDIRs is not prominent in either the Leydig organ (a rich source of B lymphocytes) or the thymus (which contains B and T lymphocytes).

Evolutionary implications of MDIRs in activating/inhibitory signaling

Overall patterns of sequence similarity plus conserved organizational features of MDIRs and other IgSF molecules are consistent with a certain level of evolutionary relatedness;

however, interpreting such relationships is confounded by extensive evolutionary distances, as well as the apparent specialization within members of the particular IgSF families to which comparisons can be made (Fig. 7). The most parsimonious interpretation of the sequence similarities of the I-set MDIR domains to CMRF-35/MAIR/CLM/pIgR and the C2-set MDIR domains to the FcRH proteins is that there has been at least one significant domain duplication/exonic interchange event between two large, distinct families of IgSF genes during the evolution of vertebrates (Posada and Crandall 2001). While selective pressures shaping this particular pattern of evolution are not clear, it is conceivable that there is selective advantage to maintaining a large and diverse set of activating/ inhibitory IgSF surface receptors that are organized in such a way as to promote the interchange(s).

A growing body of evidence obtained from mouse and human studies underscores essential roles for activating/ inhibitory IgSF members in early phases of immune recognition as well as in the differentiation and regulation of other cell lineages that are involved indirectly in immune function (Cerwenka and Lanier 2001; Dietrich et al. 2000; Trowsdale 2001). Although such molecules exhibit significant differences in their extracellular domains, members of a given family often mediate similar ligand-binding functions. Most families of activating/inhibitory IgSF members share a common mechanism of intracellular regulation, which involves activating and inhibitory forms of the receptors that act in opposing directions and participate in fine regulation of signaling cascades and cellular responses.

The mammalian leukocyte receptor complex encodes several multigene families of activating/inhibitory IgSF receptors, including products of the killer cell immunoglobulin-like receptor, immunoglobulin-like transcript/ leukocyte, immunoglobulin-like receptor/ paired immunoglobulin-like receptor and leukocyte-associated immunoglobulin-like receptor loci (Borges and Cosman 2000; Dietrich et al. 2000; Kubagawa et al. 1999; Trowsdale 2001; Wilson et al. 2000). Other multigene families of activating/ inhibitory receptors in mammals include the classical Fc receptor (Daeron 1997; Davis et al. 2002b; Ravetch and Bolland 2001; Ravetch and Lanier 2000), signal regulatory protein (Colonna 2003) and Ly-49 protein families (Nakamura and Seaman 2001). An emerging picture of the roles of these receptors, which are often co-expressed by many different subsets of cells, involves a large number of ligand interactions, often of relatively low affinity, the integration of which is used to make signaling decisions regarding the fate or function of the expressing cells.

Evolutionary inferences regarding the large families of activating/inhibitory leukocyte regulatory receptors have been limited mainly to humans and mice; however, recent work has extended to address relationships of the chicken *CHIR* gene complex to genes of the mammalian leukocyte receptor complex (Nikolaidis et al. 2005; Viertlboeck et al. 2005). A family of activating/inhibitory IgSF members also has been reported in a jawless vertebrate; however, similarity of these proteins to potential orthologs in jawed vertebrates is remote (Suzuki et al. 2005).

This study demonstrates the existence of a group of complex families of surface receptors that appeared before or at a very early stage of vertebrate evolution. Evidence is presented for the general presence of large families of IgSF receptor gene families, which are equivalently organized but may not share high degrees of sequence similarity, throughout at least the major radiations of jawed vertebrates. The basis for this interspecific “rederivation” is unknown, given the remarkable similarity that other exclusive membrane-bound immunoreceptors, such as T-cell antigen receptors, exhibit within similar phylogenetic boundaries (Litman et al. 2005).

Further insight into the evolution of such systems will be realized through the recent prioritization of skate for whole-genome sequencing, as well as the availability of the chicken genome (International Chicken Genome Sequencing Consortium 2004) and nearly complete resolution of the zebrafish genome. Given the size of the skate MDIR multigene family and its relationship to other extended families of activating/inhibitory immune-type receptor genes in higher vertebrates, future investigations will likely be particularly informative in terms of broad issues relating to the evolution of complex receptor families.

Acknowledgments

We thank Barbara Pryor for editorial assistance. GWL is supported by grants from the National Institutes of Health and the All Children's Hospital Foundation. The experiments described herein comply with the current laws of the United States of America.

References

- Aguilar H, Alvarez-Errico D, Garcia-Montero AC, Orfao A, Sayos J, Lopez-Botet M. Molecular characterization of a novel immune receptor restricted to the monocytic lineage. *J Immunol.* 2004; 173:6703–6711. [PubMed: 15557162]
- Alvarez-Errico D, Aguilar H, Kitzig F, Brckalo T, Sayos J, Lopez-Botet M. IREM-1 is a novel inhibitory receptor expressed by myeloid cells. *Eur J Immunol.* 2004; 34:3690–3701. [PubMed: 15549731]
- Amemiya, CT.; Ota, T.; Litman, GW. Construction of P1 artificial chromosome (PAC) libraries from lower vertebrates. Lai, E.; Birren, B., editors. San Diego, CA: Academic; 1996. p. 223-256.
- Bakos MA, Kurosky A, Goldblum RM. Characterization of a critical binding site for human polymeric Ig on secretory component. *J Immunol.* 1991; 147:3419–3426. [PubMed: 1940346]
- Borges L, Cosman D. LIRs/ILTs/MIRs, inhibitory and stimulatory Ig-superfamily receptors expressed in myeloid and lymphoid cells. *Cytokine Growth Factor Rev.* 2000; 11:209–217. [PubMed: 10817964]
- Cannon JP, Haire RN, Litman GW. Identification of diversified immunoglobulin-like variable region-containing genes in a protochordate. *Nat Immunol.* 2002; 3:1200–1207. [PubMed: 12415263]
- Cannon JP, Haire RN, Pancer Z, Mueller MG, Skapura D, Cooper MD, Litman GW. Variable domains and a VpreB-like molecule are present in a jawless vertebrate. *Immunogenetics.* 2005; 56:924–929. [PubMed: 15650874]
- Cantoni C, Bottino C, Vitale A, Pessino A, Augugliaro R, Malaspina A, Parolini S, Moretta L, Moretta A, Biassoni R. NKp44, a triggering receptor involved in tumor cell lysis by activated human natural killer cells, is a novel member of the immunoglobulin superfamily. *J Exp Med.* 1999; 189:787. [PubMed: 10049942]
- Cantoni C, Ponassi M, Biassoni R, Conte R, Spallarossa A, Moretta A, Moretta L, Bolognesi M, Bordo D. The three-dimensional structure of the human NK cell receptor NKp44, a triggering partner in natural cytotoxicity. *Structure.* 2003; 11:725–734. [PubMed: 12791260]
- Cerwenka A, Lanier LL. Natural killer cells, viruses and cancer. *Nat Rev Immunol.* 2001; 1:41–49. [PubMed: 11905813]
- Chung DH, Humphrey MB, Nakamura MC, Ginzinger DG, Seaman WE, Daws MR. CMRF-35-like molecule-1, a novel mouse myeloid receptor, can inhibit osteoclast formation. *J Immunol.* 2003; 171:6541–6548. [PubMed: 14662855]
- Colonna M. TREMs in the immune system and beyond. *Nat Rev Immunol.* 2003; 3:445–453. [PubMed: 12776204]
- Daeron M. Fc receptor biology. *Annu Rev Immunol.* 1997; 15:203–234. [PubMed: 9143687]
- Davis RS, Wang Y-H, Kubagawa H, Cooper MD. Identification of a family of Fc receptor homologs with preferential B cell expression. *Proc Natl Acad Sci USA.* 2001; 98:9772–9777. [PubMed: 11493702]

- Davis RS, Dennis GJ, Odom MR, Gibson AW, Kimberly RP, Burrows PD, Cooper MD. Fc receptor homologs: newest members of a remarkably diverse Fc receptor gene family. *Immunol Rev.* 2002a; 190:123–136. [PubMed: 12493010]
- Davis RS, Li H, Chen CC, Wang YH, Cooper MD, Burrows PD. Definition of an Fc receptor-related gene (FcRX) expressed in human and mouse B cells. *Int Immunol.* 2002b; 14:1075–1083. [PubMed: 12202404]
- Davis RS, Ehrhardt GR, Leu CM, Hirano M, Cooper MD. An extended family of Fc receptor relatives. *Eur J Immunol.* 2005; 35:674–680. [PubMed: 15688344]
- Dietrich J, Nakajima H, Colonna M. Human inhibitory and activating Ig-like receptors which modulate the function of myeloid cells. *Microbes Infect.* 2000; 2:323–329. [PubMed: 10758410]
- Ehrhardt GR, Davis RS, Hsu JT, Leu CM, Ehrhardt A, Cooper MD. The inhibitory potential of Fc receptor homolog 4 on memory B cells. *Proc Natl Acad Sci USA.* 2003; 100:13489–13494. [PubMed: 14597715]
- Ehrhardt GR, Hsu JT, Gartland L, Leu CM, Zhang S, Davis RS, Cooper MD. Expression of the immunoregulatory molecule FcRH4 defines a distinctive tissue-based population of memory B cells. *J Exp Med.* 2005; 202:783–791. [PubMed: 16157685]
- Hamburger AE, West AP Jr, Bjorkman PJ. Crystal structure of a polymeric immunoglobulin binding fragment of the human polymeric immunoglobulin receptor. *Structure.* 2004; 12:1925–1935. [PubMed: 15530357]
- Harpaz Y, Chothia C. Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J Mol Biol.* 1994; 238:528–539. [PubMed: 8176743]
- International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature.* 2004; 432:695–716. [PubMed: 15592404]
- Jackson DG, Hart DN, Starling G, Bell JI. Molecular cloning of a novel member of the immunoglobulin gene superfamily homologous to the polymeric immunoglobulin receptor. *Eur J Immunol.* 1992; 22:1157. [PubMed: 1349532]
- Kubagawa H, Chen CC, Ho LH, Shimada T, Gartland L, Mashburn C, Uehara T, Ravetch JV, Cooper MD. Biochemical nature and cellular distribution of the paired immunoglobulin-like receptors, PIR-A and PIR-B. *J Exp Med.* 1999; 189:309–317. [PubMed: 9892613]
- 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]
- Lanier LL. On guard—activating NK cell receptors. *Nat Immunol.* 2001; 2:23–27. [PubMed: 11135574]
- Leu CM, Davis RS, Gartland LA, Fine WD, Cooper MD. FcRH1: an activation coreceptor on human B cells. *Blood.* 2005; 105:1121–1126. [PubMed: 15479727]
- Litman GW, Anderson MK, Rast JP. Evolution of antigen binding receptors. *Annu Rev Immunol.* 1999; 17:109–147. [PubMed: 10358755]
- Litman GW, Hawke NA, Yoder JA. Novel immune-type receptor genes. *Immunol Rev.* 2001; 181:250–259. [PubMed: 11513146]
- Litman GW, Cannon JP, Dishaw LJ. Reconstructing immune phylogeny: new perspectives. *Nat Rev Immunol.* 2005; 5:866–879. [PubMed: 16261174]
- Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B, Wilson RK, McDonald KM, Hillier LW, McPherson JD, Waterston RH. High throughput fingerprint analysis of large-insert clones. *Genome Res.* 1997; 7:1072–1084. [PubMed: 9371743]
- McQueen KL, Parham P. Variable receptors controlling activation and inhibition of NK cells. *Curr Opin Immunol.* 2002; 14:615–621. [PubMed: 12183162]
- Miyake T, Amemiya CT. BAC libraries and comparative genomics of aquatic chordate species. *Comp Biochem Physiol.* 2004; 138:233–244.
- Nakamura MC, Seaman WE. Ligand interactions by activating and inhibitory Ly-49 receptors. *Immunol Rev.* 2001; 181:138–148. [PubMed: 11513135]
- Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* 1986; 3:418–426. [PubMed: 3444411]

- Nikolaidis N, Makalowska I, Chalkia D, Makalowski W, Klein J. Origin and evolution of the chicken leukocyte receptor complex. *Proc Natl Acad Sci USA*. 2005; 102:4057–4062. [PubMed: 15753291]
- Parham P. MHC class I molecules and KIRs in human history, health and survival. *Nat Rev Immunol*. 2005; 5:201–214. [PubMed: 15719024]
- Posada D, Crandall KA. Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol Evol*. 2001; 16:37–45. [PubMed: 11146143]
- Ravetch JV, Bolland S. IgG Fc receptors. *Annu Rev Immunol*. 2001; 19:275–290. [PubMed: 11244038]
- Ravetch JV, Lanier LL. Immune inhibitory receptors. *Science*. 2000; 290:84–89. [PubMed: 11021804]
- Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987; 4:406–425. [PubMed: 3447015]
- Salvatore CA, Jacobson MA, Taylor HE, Linden J, Johnson RG. Molecular cloning and characterization of the human A3 adenosine receptor. *Proc Natl Acad Sci USA*. 1993; 90:10365–10369. [PubMed: 8234299]
- Shakhov AN, Rybtsov S, Tumanov AV, Shulenin S, Dean M, Kuprash DV, Nedospasov SA. SMUCKLER/TIM4 is a distinct member of TIM family expressed by stromal cells of secondary lymphoid tissues and associated with lymphotoxin signaling. *Eur J Immunol*. 2004; 34:494–503. [PubMed: 14768054]
- Shibuya A, Sakamoto N, Shimizu Y, Shibuya K, Osawa M, Hiroyama T, Eyre HJ, Sutherland GR, Endo Y, Fujita T, Miyabayashi T, Sakano S, Tsuji T, Nakayama E, Phillips JH, Lanier LL, Nakauchi H. Fc α / μ receptor mediates endocytosis of IgM-coated microbes. *Nat Immunol*. 2000; 1:441–446. [PubMed: 11062505]
- Speckman RA, Wright Daw JA, Helms C, Duan S, Cao L, Taillon-Miller P, Kwok PY, Menter A, Bowcock AM. Novel immunoglobulin superfamily gene cluster, mapping to a region of human chromosome 17q25, linked to psoriasis susceptibility. *Hum Genet*. 2003; 112:34–41. [PubMed: 12483297]
- Stet RJM, Hermesen T, Westphal AH, Jukes J, Engelsma M, Verburg-van Kemenade BML, Dortmans J, Aveiro J, Savelkoul HFJ. Novel immunoglobulin-like transcripts in teleost fish encode polymorphic receptors with cytoplasmic ITAM and ITIM and a new structural Ig domain similar to the natural cytotoxicity receptor NKp44. *Immunogenetics*. 2005; 57:77–89. [PubMed: 15702329]
- Strong SJ, Mueller MG, Litman RT, Hawke NA, Haire RN, Miracle AL, Rast JP, Amemiya CT, Litman GW. A novel multigene family encodes diversified variable regions. *Proc Natl Acad Sci U S A*. 1999; 96:15080–15085. [PubMed: 10611341]
- Suzuki T, Shin I, Fujiyama A, Kohara Y, Kasahara M. Hagfish leukocytes express a paired receptor family with a variable domain resembling those of antigen receptors. *J Immunol*. 2005; 174:2885–2891. [PubMed: 15728499]
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994; 22:4673–4680. [PubMed: 7984417]
- Trowsdale J. Genetic and functional relationships between MHC and NK receptor genes. *Immunity*. 2001; 15:363–374. [PubMed: 11567627]
- Viertlboeck BC, Habermann FA, Schmitt R, Groenen MA, Du PL, Gobel TW. The chicken leukocyte receptor complex: a highly diverse multigene family encoding at least six structurally distinct receptor types. *J Immunol*. 2005; 175:385–393. [PubMed: 15972672]
- Vivier E, Malissen B. Innate and adaptive immunity: specificities and signaling hierarchies revisited. *Nat Immunol*. 2005; 6:17–21. [PubMed: 15611777]
- Wilson MJ, Torkar M, Haude A, Milne S, Jones T, Sheer D, Beck S, Trowsdale J. Plasticity in the organization and sequences of human KIR/ILT gene families. *Proc Natl Acad Sci USA*. 2000; 97:4778–4783. [PubMed: 10781084]
- Yoder JA, Mueller MG, Wei S, Corliss BC, Prather DM, Willis T, Litman RT, Djeu JY, Litman GW. Immune-type receptor genes in zebrafish share genetic and functional properties with genes

encoded by the mammalian lymphocyte receptor cluster. *Proc Natl Acad Sci USA*. 2001; 98:6771–6776. [PubMed: 11381126]

Yoder JA, Litman RT, Mueller MG, Desai S, Dobrinski KP, Montgomery JS, Buzzeo MP, Amemiya CT, Trede NS, Wei S, Djeu JY, Humphray S, Jekosch K, Hernandez Prada JA, Ostrov DA, Litman GW. Resolution of the NITR gene cluster in zebrafish. *Proc Natl Acad Sci USA*. 2004; 101:15706–15711. [PubMed: 15496470]

Yotsumoto K, Okoshi Y, Shibuya K, Yamazaki S, Tahara-Hanaoka S, Honda S, Osawa M, Kuroiwa A, Matsuda Y, Tenen DG, Iwama A, Nakauchi H, Shibuya A. Paired activating and inhibitory immunoglobulin-like receptors, MAIR-I and MAIR-II, regulate mast cell and macrophage activation. *J Exp Med*. 2003; 198:223–233. [PubMed: 12874256]

Fig. 1. Predicted structures of proteins MDIR1 through MDIR5; a splice variant of MDIR3 is included. Domain subtypes are indicated by contrasting colors (type I blue, type II black, type III green, type IV red). Amino acid identities among members of the same subtype are as follows: type II, 56–88%; type III, 68–90%; type IV, 72–88%. *ITIM*, immunoreceptor tyrosine-based inhibitory motif. *ITIM*-like motifs (YA[R/S]V), designated “itim”, are present in MDIR2. The dashed arrow designates a partial transcript

Fig. 2.
Genomic complexity of MDIRs. Genomic Southern blot hybridization using probes representing each of the four types of MDIR IgSF domain (I, MDIR1IgD1; II, MDIR2IgD1; III, MDIR2IgD2; IV, MDIR4IgD2). Samples were digested with either *EcoRI* (E) or *HindIII* (H)

Fig. 3. CLUSTALW (Thompson et al. 1994) amino acid sequence alignment of translated PCR amplicons from BAC “haplo-typing” using **a** type II or **b** type III domain consensus primers (alignment provided by <http://www.ebi.ac.uk.clustalw/>). Amino acids encoded by PCR primers have been removed. The stop codon in two of the type III pseudogenes (#1 and #2) is indicated by asterisk; a type III sequence (#3), which contains a gap corresponding to 28 residues, also is potentially a pseudo-gene. Two type IV domains amplified by type III domain primers are represented by sequences 13 and 14 in (**b**)

Fig. 4. Amino acid alignment of MDIR domain types I and II to various vertebrate IgSF domains. The position of the common amino acid motif KXWC is indicated by black shading. Conserved amino acids are indicated by asterisk. Amino acid sequences used in phylogenetic analyses are indicated by (+). Bold-face type indicates regions of β sheet structures as defined by X-ray crystallographic analysis. *Arrows* indicate likely intrachain disulfide bonds. Predicted structural characteristics (strands A through G) are indicated at the bottom

Fig. 5. Neighbor-joining-tree analyses depicting phylogenetic relationships of MDIR domain types I and II to vertebrate proteins listed in GenBank and to selected zebrafish genomic sequences from the 45- to 46-Mb region of chromosome 2 and the 19-Mb region of chromosome 9. Trees are unrooted

Fig. 6. Neighbor-joining-tree analyses depicting phylogenetic relationships of MDIR domain types III and IV to vertebrate proteins listed in GenBank and to selected zebrafish genomic sequences from the 47-Mb region of chromosome 2. Trees are unrooted

Fig. 7. General relatedness of MDIR2 to mouse FcRH, TREM/TLT, pIgR, and CLM/MAIR family molecules. Percentages indicate the range of approximate amino acid identity among constituent IgSF domains, which are distinguished (by family) using different colors. Evolutionary origins of various proteins are not implied. MDIR types I and IV domains (not shown) exhibit equivalent relatedness patterns to types II and III, respectively. *ITIM*, immunoreceptor tyrosine-based inhibitory motif

Table 1

Oligonucleotide sequences of PCR primers

Name	Sequence
A	
VYWFR-SfiIB	5'-TGGCCGAGGCGGCCNCGRAACCARTANAC
VYWF-SfiIB	5'-GACTGGCCGAGGCGGCCRAACCARTANAC
YWFK-SfiIB	5'-GACTGGCCGAGGCGGCCCYTTTRAACCARTA
YWFR-SfiIB	5'-GACTGGCCGAGGCGGCCNCGRAACCARTA
WFK-SfiIB	5'-GACTGGCCGAGGCGGCCCYTTTRAACCA
WFR1-SfiIB	5'-GACTGGCCGAGGCGGCCNCGRAACCA
WFR2-SfiIB	5'-GACTGGCCGAGGCGGCCCYCTRAACCA
B	
MDIR1-IgD-F*	5'-GACTGGCCATTATGGCCTAAGTGGACCTCGTGCAG
MDIR1-IgD-R*	5'-GACTGGCCACCGCGCCGTTAGTATTATGTTTCCA
MDIR2-IgD1-F*	5'-GACTGGCCATTATGGCCTGTGGGGAAGGACCATG
MDIR2-IgD1-R	5'-AGACGCAGCTTCATCAGAGACTTG
MDIR2-IgD2-F	5'-GTTCTGTGCTTGGATTACTGTCA
MDIR2-IgD2-R*	5'-GACTGGCCACCGCGCCAAGTGTCTTGGTGTGAA
MDIR4-IgD2-F	5'-CTACTCCTGTGCTCAGATATCTGTC
MDIR4-IgD2-R	5'-ACGTTCTGCTCCTTTCAAAAC
TypeIIcon-F1	5'-GTTTCAGGTGCANTGTGGGSRAG
TypeIIcon-R1	5'-TCCACAKCYGTACCAYCCTGNATC
TypeIIcon-R2	5'-ATCCWCCAYAGTCACAGTWA
TypeIIIcon-F1	5'-CCAACCAATGTCTCCTGTCTCGG
TypeIIIcon-R1	5'-CACGTTGACCAATTTACTGGATTT
TypeIIIcon-R2	5'-CACGTTGACCATTTCCTGATTT

A Amptrap, B MDIRs. MDIR-specific primers marked with an asterisk contain SfiI restriction sites at their 5' ends

Table 2

MDIR domain-specific hybridization patterns

# of BACs	Domain-type
2	I
1	II
3	III
4	I, II
1	II, III
3	II, IV
6	III, IV
48	II, III, IV
68	Total

MDIR domain-specific hybridization patterns for MDIR BACs (isolated by iterative screening of a 5–6X representation library with MDIR domain type I- to IV-specific probes)

Table 3

Analysis of nucleotide differences for individual MDIR domains

A									
Type I and Type II	1	2	3	4	5	6	7	8	9
1. MDIR1-IgD									
2. MDIR2-IgD1	0.715								
3. MDIR3-IgD1	0.674	0.177							
4. MDIR3-IgD2	0.683	0.172	0.027						
5. MDIR5-IgD1	0.703	0.170	0.027	0.027					
6. MDIR4-IgD1	0.745	0.217	0.159	0.167	0.146				
7. MDIR4-IgD3	0.723	0.222	0.185	0.186	0.191	0.247			
8. MDIR3-IgD4	0.685	0.191	0.141	0.139	0.171	0.198	0.115		
9. MDIR4-IgD5	0.672	0.189	0.104	0.115	0.134	0.176	0.127	0.067	
10. MDIR5-IgD3	0.676	0.185	0.142	0.153	0.158	0.206	0.122	0.082	0.067
Type III and Type IV									
1	1	2	3	4	5	6	7		
1. MDIR2-IgD2									
2. MDIR3-IgD5	0.245								
3. MDIR4-IgD6	0.264	0.131							
4. MDIR5-IgD4	0.327	0.201	0.139						
5. MDIR3-IgD3	0.318	0.399	0.405	0.418					
6. MDIR4-IgD2	0.327	0.417	0.420	0.421	0.051				
7. MDIR4-IgD4	0.275	0.389	0.385	0.368	0.078	0.077			
8. MDIR5-IgD2	0.283	0.367	0.396	0.406	0.051	0.063	0.091		
B									
Type I and Type II	1	2	3	4	5	6	7	8	9
1. MDIR1-IgD									
2. MDIR2-IgD1	0.423								
3. MDIR3-IgD1	0.441	0.207							
4. MDIR3-IgD2	0.443	0.187	0.060						
5. MDIR5-IgD1	0.454	0.200	0.104	0.060					

6. MDIR4-IgD1	0.453	0.158	0.173	0.154	0.178		
7. MDIR4-IgD3	0.444	0.214	0.195	0.182	0.176	0.180	
8. MDIR3-IgD4	0.443	0.185	0.171	0.154	0.152	0.161	0.101
9. MDIR4-IgD5	0.441	0.198	0.196	0.179	0.182	0.185	0.128
10. MDIR5-IgD3	0.458	0.217	0.209	0.192	0.191	0.193	0.133
						0.060	0.086

Type III and Type IV

1. MDIR2-IgD2	1	2	3	4	5	6	7
2. MDIR3-IgD5							
3. MDIR4-IgD6	0.218						
4. MDIR5-IgD4	0.186	0.071					
5. MDIR3-IgD3	0.218	0.097	0.061				
6. MDIR4-IgD2	0.290	0.298	0.266	0.279			
7. MDIR4-IgD4	0.260	0.278	0.247	0.273	0.071		
8. MDIR5-IgD2	0.256	0.293	0.268	0.292	0.133	0.089	
	0.249	0.260	0.229	0.261	0.080	0.058	0.093

A The proportion of synonymous nucleotide differences per synonymous site (Ps) estimated by Nei and Gojobori's method (Nei and Gojobori 1986). B The proportion of non-synonymous nucleotide differences per non-synonymous site (Pn)

From Table 3, the following points can be made: The theoretical maximum value of Ps is 0.75, if the frequencies of the four different types of nucleotide are assumed to be equal. Therefore, the MDIR1 and MDIR2-MDIR5 sequences appear to have been diverged for a long period of time, as the proportion of synonymous nucleotide differences per synonymous sites are close to saturation

MDIR3-IgD1 and MDIR3-IgD2 are results of recent duplications, as the proportion of synonymous nucleotide differences per synonymous site is small

Two distinct types of MDIR intermediate (I)-type IgSF domains emerge: (a) Type I: MDIR1-IgD1, MDIR2-IgD1, MDIR3-IgD1, MDIR3-IgD2, MDIR5-IgD1, MDIR4-IgD1; (b2) Type II, subtype 2: MDIR3-IgD4, MDIR4-IgD3, MDIR4-IgD5, MDIR5-IgD3

Two distinct C2-types emerge: Type III: MDIR2-IgD2, MDIR3-IgD5, MDIR4-IgD6, MDIR5-IgD4; Type IV: MDIR3-IgD3, MDIR4-IgD2, MDIR4-IgD4, MDIR5-IgD2

Within the MDIR domain type II, subtype 1 or type III categories, the respective domains of MDIR2 show higher differences from MDIR3-5, indicating the MDIR2 may have diverged first from MDIR3-5. The proportion of synonymous nucleotide differences per synonymous site for MDIR3-IgD1/IgD2 vs MDIR5-IgD1, MDIR3-IgD3 vs MDIR5-IgD2, MDIR3-IgD4 vs MDIR5-IgD3, MDIR3-IgD5 vs MDIR5-IgD4 are 0.027, 0.051, 0.082, and 0.201, respectively. Simple duplication of Ig domains of MDIRs does not explain the observation, implying instead an exon shuffling or gene conversion-like event between ancestors of MDIR3 and MDIR5