

Research article

Open Access

Diversity and repertoire of IgW and IgM VH families in the newborn nurse shark

Lynn L Rumfelt*^{1,2}, Rebecca L Lohr³, Helen Dooley³ and Martin F Flajnik^{1,3}

Address: ¹Department of Microbiology and Immunology, University of Miami School of Medicine, Miami, FL 33101 USA, ²Department of Immunology, University of Toronto, Sunnybrook & Women's College Health Sciences Centre, 2075 Bayview Avenue, Toronto, Ontario M4N 3M5 Canada and ³Department of Microbiology and Immunology, University of Maryland School of Medicine, 655 W. Baltimore St., Rm. BRB-13009, Baltimore, MD 21201 USA

Email: Lynn L Rumfelt* - Lynn.Rumfelt@sw.ca; Rebecca L Lohr - Rebecca_Lohr@yahoo.com; Helen Dooley - HDooley@som.umaryland.edu; Martin F Flajnik - MFlajnik@som.umaryland.edu

* Corresponding author

Published: 06 May 2004

Received: 27 January 2004

BMC Immunology 2004, 5:8

Accepted: 06 May 2004

This article is available from: <http://www.biomedcentral.com/1471-2172/5/8>

© 2004 Rumfelt et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Adult cartilaginous fish express three immunoglobulin (Ig) isotypes, IgM, IgNAR and IgW. Newborn nurse sharks, *Ginglymostoma cirratum*, produce 19S (multimeric) IgM and monomeric/dimeric IgM_{I_g}, a germline-joined, IgM-related VH, and very low amounts of 7S (monomeric) IgM and IgNAR proteins. Newborn IgNAR VH mRNAs are diverse in the complementarity-determining region 3 (CDR3) with non-templated nucleotide (N-region) addition, which suggests that, unlike in many other vertebrates, terminal deoxynucleotidyl transferase (*TdT*) expressed at birth is functional. IgW is present in the lungfish, a bony fish sharing a common ancestor with sharks 460 million years ago, implying that the IgW VH family is as old as the IgM VH family. This nurse shark study examined the IgM and IgW VH repertoire from birth through adult life, and analyzed the phylogenetic relationships of these gene families.

Results: IgM and IgW VH cDNA clones isolated from newborn nurse shark primary and secondary lymphoid tissues had highly diverse and unique CDR3 with N-region addition and VDJ gene rearrangement, implicating functional *TdT* and *RAG* gene activity. Despite the clear presence of N-region additions, newborn CDR3 were significantly shorter than those of adults. The IgM clones are all included in a conventional VH family that can be classified into five discrete groups, none of which is orthologous to IgM VH genes in other elasmobranchs. In addition, a novel divergent VH family was orthologous to a published monotypic VH horn shark family. IgW VH genes have diverged sufficiently to form three families. IgM and IgW VH serine codons using the potential somatic hypermutation hotspot sequence occur mainly in VH framework I (FR1) and CDR1. Phylogenetic analysis of cartilaginous fish and lungfish IgM and IgW demonstrated they form two major ancient gene groups; furthermore, these VH genes generally diversify (duplicate and diverge) within a species.

Conclusion: As in ratfish, sandbar and horn sharks, most nurse shark IgM VH genes are from one family with multiple, heterogeneous loci. Their IgW VH genes have diversified, forming at least three families. The neonatal shark Ig VH CDR3 repertoire, diversified via N-region addition, is shorter than the adult VDJ junction, suggesting one means of postnatal repertoire diversification is expression of longer CDR3 junctions.

Background

The major components of adaptive immunity, including Ig, T cell receptors, and MHC class I and class II, have been identified in the ancient jawed vertebrate taxon cartilaginous fish (Chondrichthyes), which includes sharks, skates, rays (elasmobranchs), and ratfish (holocephali) (Figure 1) [1,2]. Additionally, genes required for lymphopoiesis and Ig gene rearrangement/diversification, such as *PU.1*, *Ikaros*, *RAG1*, *RAG2* and *TdT*, as well as primary and secondary lymphoid tissues, are also present within this vertebrate class [3-8]. Cartilaginous fish genomic Ig gene segments are organized in ~10 kb clusters, each containing one variable segment (V), two to three diversity segments (D) depending on the *VH* class, a joining segment (J), and constant gene exons (C); in all expression studies to date these gene segments undergo rearrangement only within a cluster [1,9,10]. Thus far three major Ig classes, IgM, IgNAR, and IgW, have been identified in adult elasmobranchs [1,2]. However, newborn shark pups require postnatal development to express adult levels of 7S IgM and IgNAR [11-13]. In lieu of 7S IgM and IgNAR, neonates preferentially secrete 19S IgM and IgM_{1gj} (so-named because it is related to conventional IgM; 1 refers to the 1st IgM gene expressed during ontogeny and expressed in primary lymphoid tissue; and gj refers to the germline-join of this *VH* gene [5]), a monotypic *VH* expressed from a non-somatically rearranged (germline-joined) *VDJ* locus [5,11,12]. Whereas IgNAR protein is present in low amounts in the sera of newborn sharks, study of IgNAR mRNA at birth revealed a novel isotype (type 3) of reduced diversity, in part due to germline-joining of two of the three *DH* genes in the specific cluster [5,14]. Interestingly, this IgNAR type 3 is most highly expressed at birth and may undergo positive selection since its CDR3 length is highly constrained [14]. As with IgM_{1gj} expression, when shark pups mature IgNAR type 3 gene transcription declines and expression of conventional IgNAR types 1 and 2 comes to predominate. This switch in expression correlates with increased diversity in CDR3 due to rearrangement of the three *DH* genes and N-region additions [5,14]. A molecular study in the embryonic and newborn clearnose skate, *Raja eglanteria*, demonstrated that IgM and IgW *VH* CDR3 increase in diversity by 2–3 fold as they mature to adulthood [15]. Together these results suggest cartilaginous fish may have a developmental program for expression of their Ig genes, progressing from an innate-like, restricted repertoire to a complex, diverse repertoire. These findings are similar to previous analyses of the ontogeny of mouse, human, and frog Ig gene expression [16-18]. However, unlike those other species, there is clear evidence for N-region additions in CDR3 during embryonic and neonatal life at every rearranging H chain locus so far analyzed.

In all elasmobranchs examined, secretory IgW transcripts are expressed in two forms, a full-length, long form and a truncated, short form that is probably derived by alternative splicing [19,20]. Recent molecular characterization of Ig H chains also identified both IgW forms in the African lungfish, *Protopterus aethiopicus*, a lobe-finned bony fish, which phylogenetic studies suggest are closely related to land vertebrates (tetrapods) (refer to Figure 1) [1,21,22]. Since molecular phylogenetic analyses determined that the cartilaginous fish (Chondrichthyes) separated prior to the divergence of bony fishes (Osteichthyes – lobe-finned and ray-finned) from the other jawed vertebrates [23] the discovery of IgW in the lungfish suggests that it was present in the common ancestor of bony and cartilaginous fishes 460 million years ago (MYA). As well this discovery provides an impetus for further work to determine if this Ig class is present in one or more groups of the ray-finned bony fish, i.e. chondrosteian, and holostean, and teleost, and the cartilaginous fish group of holocephali [21,24] since to date only IgM and IgD have been identified in the ray-finned bony fishes and only IgM in the holocephali [25].

To further our understanding of Ig class expression and regulation during ontogeny, we herein describe IgM and IgW *VH* family expression in the newborn nurse shark, *Ginglymostoma cirratum*.

Results

Newborn IgM is expressed from multiple loci of one *VH* family

Developmental regulation of IgM expression was examined by screening two cDNA libraries constructed from newborn nurse shark epigonal organ (primary lymphoid tissue) and spleen (secondary lymphoid tissue). Libraries were probed with a canonical IgM *VH* and partial *CH1* probe under high and low stringency hybridization conditions [5]. Deduced amino acid (AA) sequences of the leader, *VH*, and partial *CH1* for 25 positive clones were compared to the predicted AA sequence of the single previously published adult nurse shark IgM (Genbank accession #M92851 [26]) and aligned in Clustal W (Figure 2) [27]. These clones formed six groups (I-VI) that differ in *VH* leader, FR and CDR. Clones in Groups I-V are all clustered within one *VH* family, and as has been reported for other elasmobranchs; nevertheless CDR1 and CDR2 are diverse among the different groups. Group VI, which consists only of clone 36E, is distinct from the other groups and is discussed separately below.

Excluding the positions of *VDJ* rearrangement (CDR3), *VH* members within a group are quite similar in FR and CDR1-2, suggesting that they are expressed from one locus or several highly related loci. Indeed, microheterogeneity of 5'UT sequences of the 24 clones identified

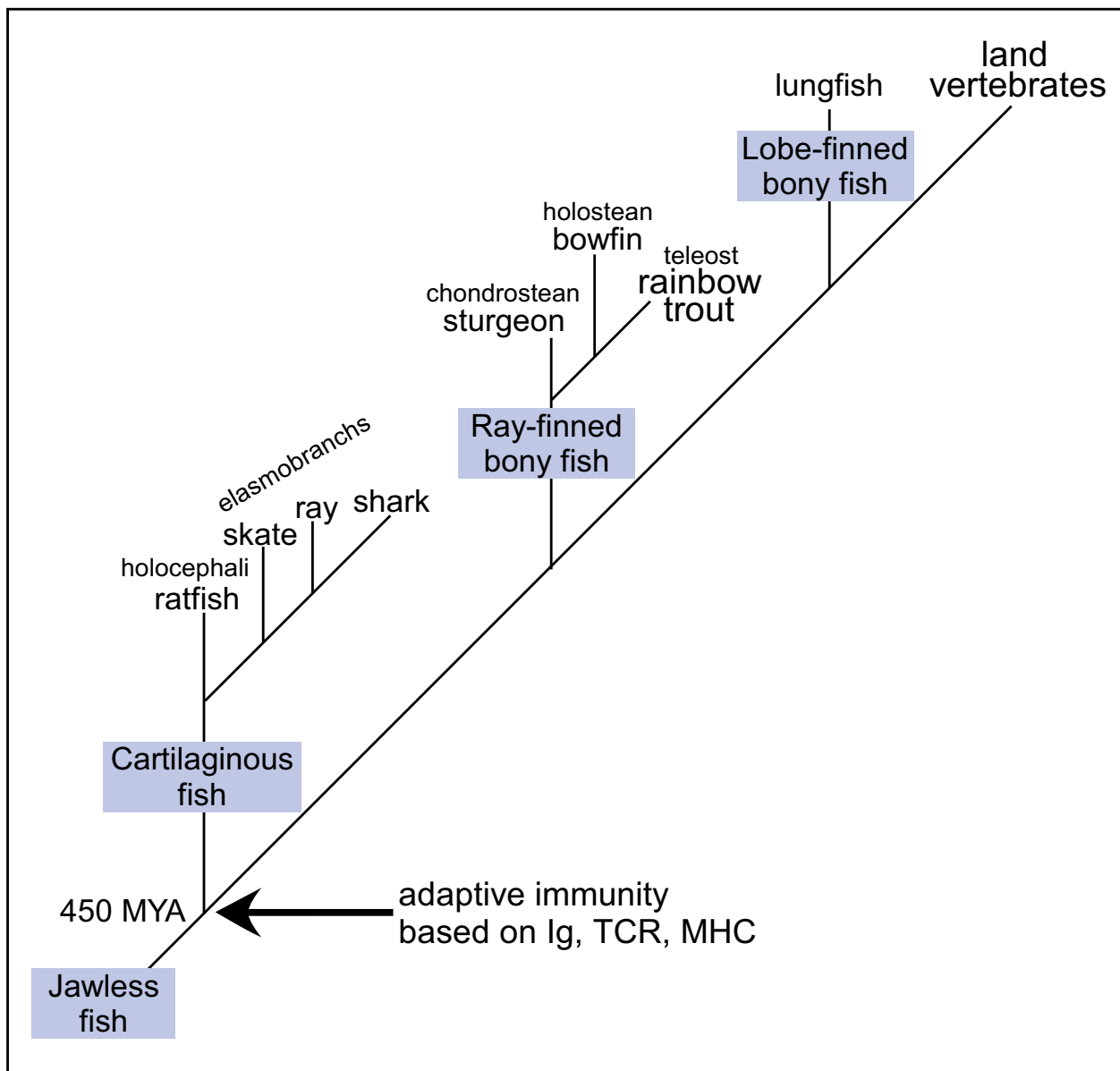


Figure 1
Phylogenetic tree. Cartilaginous fish, which includes the Holocephali ratfish and Elasmobranch sharks, skates and rays, shared a common ancestry with the ray-finned and lobe-finned bony fishes 460 million years ago (MYA). Cartilaginous fish have been identified as the earliest extant vertebrate group with major components of the adaptive immune system, i.e., Ig, TCR and MHC.

expression of 15 highly related loci from the five groups (data not shown). We estimated from previous Southern blotting analysis that there were ~15–25 IgM loci in the nurse shark [5]. Thus, our library screening results would suggest that most IgM loci are expressed at birth, which is consistent with our previous immunohistochemistry

study demonstrating that the great majority of splenic B cells in newborn nurse sharks are IgM⁺ [12]. If this estimate of gene number is confirmed in future studies of nurse shark *VH* genes, then this elasmobranch species has far fewer IgM genes as compared to the published esti-

	Leader	FR1	CDR1	FR2	CDR2	FR3
IgM _{Igj}	AF327520	MIRMQSAISLSLLTSISCVSRSDIMLTQPEAETSIPGGSLKLTCKISGFDL	EYHGMM	/WVRQIQGQGLEWLI	/YYYIVSDTHYNPFGINRFTAS	
I	30E	///MSTIFFSLLLTLFLSCVQS/QIILTQKVAETGRPGGTLTLTCKTSGFNL	GNDWMO	/WIRQVPGQGLEWLL	EYKSSSSNNYAPGVKARFTAS	
	22E	///				
	54S	///				
II	1E	//MTTMI--L---N---EEVT-I-PE--N-H---SMR-----D-	DSYA-S	//-V-----IV///	Y-YG-Y--D---AI-D	
	72S	//MTTMI--L---N---EEVT-I-PE--N-H---SMR-----D-	DSYA-S	//-V-----IV///	YSYG-Y--D---AI-D	
	35S	//MTTMI--L---N---EEVT-I-PE--N-H---SMR-----D-	DSYA-S	//-V-----IV///	YSYG-Y--D---AI-D	
	14E	///MTTMI--L---N---EEVT-I-PE--N-H---SMR-----D-	DSYT-S	//-V-----IV///	Y-YA-Y--D---AI-D	
III	17S	///////HE-	-T-G	//-V-----C-V///	T-Y-----Y---EI-G	
	6E	METTXLX--L-----/EVT-I-PE--GNSQ---S-K-----D-	-T-ETG	-V-----C-V///	T-Y-----Y---EI-G	
	57S	//MTTL--L-----/EVT-I-PE--GNSQ---S-K-----D-	-T-G	//-V-----C-V///	T-Y-----Y---EI-G	
IV	55E	///////S-----D-	SSYY-G	//-V-----V///	S-Y-----Y---AI-G	
	26E	//MTX---L--F---R---/EVT---PE--NSQ---S-T-----D-	SSYY-G	//-V-----V///	S-Y-----Y---AI-G	
	15S	//LP---L--F---R---/EVT---PE--NSQ---S-T-----D-	SSYY-G	//-V-----V///	S-Y-----Y---AI-G	
	24S	//MTT---L--F---R---/EVT---PE--NSQ---S-T-----D-	SSYY-G	//-V-----V///	S-Y-----Y---AI-G	
	46S	//MTT---L--F---R---/EVT---PE--NSQ---S-T-----D-	SSYY-G	//-V-----V///	S-Y-----Y---AI-G	
	29S	//MTT---L--F---R---/EVT---PE--NSQ---S-T-----D-	SSYY-G	//-V-----V///	T-Y-----N-Y---AI-D	
	73S	//MTT---L--F---R---/EVT---PE--NSQ---S-T-----D-	SSYY-G	//-V-----V///	T-Y-----N-Y---AI-D	
V	2S	//MMT---L---AL-P---/E-T-I-PE--T-H---S-S-----D-	SSS-Y	//-----IV///	Y-Y-----M---AI-D	
	49S	//MMT---L---AL-P---/E-T-I-PE--T-H---S-S-----D-	SSS-Y	//-----IV///	Y-Y-----M---AI-D	
	92851	//MMT---L---AL-P---/E-T-I-PE--T-H---S-S-----D-	SSS-Y	//-----IV///	Y-Y-----M---AI-D	
	2E	//MMT---L---AL-P---/E-T-I-PE--T-H---S-S-----D-	SSS-Y	//-----IVGVD	SY-Y-----M---AI-D	
	50S	//MMT---L---AL-P---/E-T-I-PE--T-H---S-S-----D-	SSS-Y	//-----IV///	Y-Y-----M---AI-D	
	47S	//MMT---L---AL-P---/E-T-I-PE--T-H---S-S-----D-	SSS-Y	//-----IV///	Y-Y-----M---AI-D	
	21S	//MMT---L---AL-P---/E-T-I-PE--T-H---S-S-----D-	SSS-Y	//-----IV///	Y-Y-----M---AI-D	
	27S	//MMT---L---AL-P---/E-T-I-PE--T-H---S-S-----D-	SSS-Y	//-----IV///	Y-Y-----M---AI-D	
VI	36E	//MQLIPCLCF--SCFTH-R-/N-V-N-SPPVSKR-ESHK-S-SV--S-	DSHVH	//-LK-F--KR-----A//	YRNP-DT-----EG-IIPS	
	FR3	CDR3	FR4	CHI		
IgM _{Igj}	AF327520	KDISNNIALDIV//NLKTEDSAIYYCARGST	YWGQGT//	RVTVTEEKPFPPPTLYGLISFNQHQHTGSSVTVYVCLATDYS	PDVIRVT	
I	30E	KDTSNNI FALEMK//NLKI EDTAIYYCAKSHGEKHC	////G YLD	HWGQGT//MVT	VTAVTTS SPILYGLVSSCQQHNTDGSVIFGCLAMGYSPDVTKVT	
	22E	///				
	54S	///				
II	1E	I-----//S-----	////SQ LQDWDDPF	-Y-----//	-L-P-T-----QTN	-D-----T-S
	72S	I-----//S-----	RDDSGGSPILGP	-Y-----//	-L-P-T-----QTN	-D-----T-S
	35S	I-----//S-----	R//////DGVILG	-Y-----//	-L-P-T-----QTN	-D-----T-S
	14E	I-----//S-----	XDY//////SAWDTIG	-Y-----//	-L-P-T-----A-Q-NE	-D-----T-S
III	17S	---T/R---	TN//////S LCN	-----//	-L-P-T-----G-S	-D-----A-S
	6E	---ETTR---	TI//////F GN	-MET-	-L-P-T-----G-S	-D-----A-S
	57S	---T/R---	TI//////F GN	-----//	-L-P-T-----G-S	-D-----A-S
IV	55E	---/R---	RRG//////GDW-RFAN	-----//	-L-P-T-----V-S	-D-----A-S
	26E	---/R---	TH//////LQSILGSFAN	-----//	-L-P-T-----V-S	-D-----A-S
	15S	---/R---	R//////ASVEREFAN	-----//	-L-P-T-----V-S	-D-----A-S
	24S	---/R---	R//////DGGVGEFAN	-----//	-L-P-T-----V-S	-D-----A-S
	46S	---/R---	S//////SGY-EFAN	-----//	-L-P-T-----V-S	-D-----A-S
	29S	---/S---	R//////VRGEAV-N	-----//	-L-P-T-----E-S	-D-----A-S
	73S	---/S---	R//////DTVGGAN	-----//	-L-P-T-----E-S	-D-----A-S
V	2S	---R//SV-D---	TESG//////GGGAGY	-----//	-TA-P-T-----G-I	-Y-----D-----AS
	49S	---R//SV-D---	TDIS//////GRVAGY	-----//	-TA-P-T-----G-I	-Y-----D-----AS
	92851	---R//SV-D---	TRRMSGY//EYLGGHSGY	-----//	-TA-P-T-----G-I	-Y-----D-----AS
	2E	---R//SV-D---	TRSTA//////PLCY	-----//	-TA-P-T-----G-I	-Y-----D-----AS
	50S	---R//SV-D---	T//////LG Y	-----//	-TA-P-T-----G-I	-Y-----D-----AS
	47S	---R//SV-D---	TVG//////G Y	-----//	-TA-P-T-----G-I	-Y-----D-----AS
	21S	---R//SV-D---	TTVG//////G Y	-----//	-TA-P-T-----G-I	-Y-----D-----AS
	27S	---R//SV-D---	TDSG//////G Y	-----//	-TA-P-T-----G-I	-Y-----D-----AS
VI	36E	RSS-TAYIEIK//SFRLE	K//////			

Figure 2
Alignment of VH and partial CHI of IgM cDNA clones. The deduced amino acid sequences of 25 IgM cDNA clones (Genbank accession # AY609247-AY609274) were aligned with the conventional adult nurse shark IgM (Genbank accession #M92851 [32]) in ClustalV. Dashes (-) indicate identity with the first sequence and gaps introduced to correct alignment are indicated by slashes (/). Identification of framework and CDR are based on previous work [26]. Differences in the framework and CDR separated the clones into six groups (I-VI). The CDR3 varies in length and sequence thus most clones are unique, demonstrating diversity is present at birth in the nurse shark.

mates of 100 IgM loci for the horn shark and 50 loci for the skate [1,28].

The IgM clones were analyzed for the conserved residues necessary to maintain the Ig structural fold using the ImmunoGeneTics (IMGT) system [29-31]. Representa-

tives of the most divergent groups I and VI were compared to the originally isolated nurse shark IgM M92851 VH (Figure 3) [32]. Residues essential for the Ig fold are the hydrophobic GlyLeuGluTrp in FR2 positions 49-52, Trp41, and Cys in strands B and F at positions 23 and 104 required for the canonical disulfide bridge [31]. These

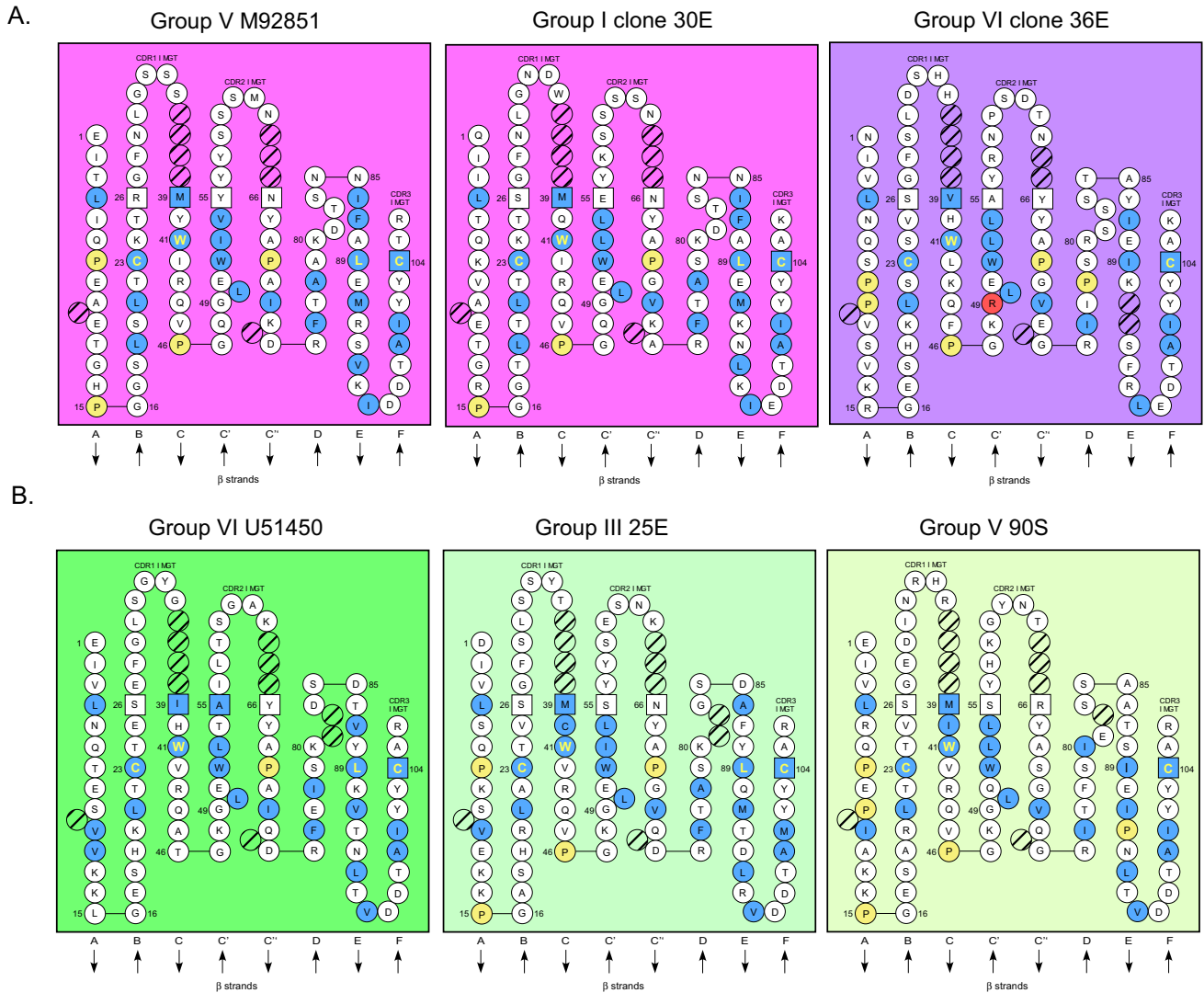


Figure 3
Newborn IgM and IgW VH families conserve residues important for the Ig fold motif. The VH structures are formatted per the IMGT numbering system [29]. (A) IgM Ig fold is maintained in IgM VH family members. The nurse shark canonical IgM Genbank accession # M92851 [32] was compared to the more divergent members from group I clone 30E and group VI clone 36E. Hydrophobic amino acids (hydropathy index with positive value: I, V, L, F, C, M, A) and tryptophan (W) at positions at which more than 50 % of analysed IG and TR sequences were positive are highlighted in blue [32]. Prolines (P) are marked in yellow and conserved Ig fold residues Cys at positions 23 and 104, and Trp (W) at position 41 are highlighted in yellow lettering. This modeling demonstrates that Ig fold motifs strand B and F Cys and hydrophobic core residues are conserved in these VH domains. Hatched circles represent residues of the canonical VH IMGT structure that are not utilized in the shark VH. (B) The IMGT numbering system was used to compare the VH domains of the conventional IgW U51450 [46] to representatives of the IgW VH divergent groups III and V. These divergent VH families maintain the Ig fold motif and utilize Pro in the FR similar to IgM (refer to part A).

residues and their positions are conserved in members of the five IgM groups, indicating preservation of the Ig fold structure. The FR and CDR for the clones were analyzed for their similarity to the original VH (Table 1). All groups

were least similar to the original VH in the FR1 (69–83% identity) and CDR1 (33%), and most similar in FR2 and FR3 (78–93%). CDR2 was more conserved than CDR1, with groups II-IV most similar to the original sequence. FR

Table 1: Deduced amino acid sequence percent similarity of canonical IgM (group V) VH regions to IgM_{Igj} and newborn IgM groups I-IV, and VI.

	FR1	CDR1	FR2	CDR2	FR3	VH
IgM_{Igj}	69	17	64	44	59	58
I	69	33	86	69	84	74
II	83	33	93	75	84	80
III	79	33	78	75	84	77
IV	76	33	86	81	84	78
VI	38	0	50	44	38	38

conservation and variation in the CDR are consistent with findings in the horn shark, sandbar shark, and higher vertebrates [33]. Membership within a specific *VH* family has been defined as greater than or equal to 75% identical AA (greater than or equal to 70% identical nucleotides); thus *VH* in nurse shark groups I-V are within a single family (refer to Table 1) [31,35]. Additionally an ATG codon (Methionine) in CDR1 is present in all IgM clones, consistent with a previous study showing this codon at this position is preserved throughout vertebrate phylogeny (refer to Fig 2) [35]. Any nucleotide substitution in this codon leads to an AA replacement, suggesting a role for its evolutionary conservation. Generally, CDR1 and CDR2 contain codons that either target the hypermutation machinery (see below) or are minimally degenerate (e.g. ATG) to maximize the potential for AA modifications in these regions of the V domain [36].

The CDR3 sequences of each clone were unique in length and sequence, suggesting gene rearrangement and N-region addition via *RAG1/2* and *TdT* action, respectively. This was further investigated by comparison of CDR3 nucleotide sequences to a representative genomic horn shark IgM locus (X13447, [33]) in Table 2. The *DH* gene sequence that exactly matched the horn shark sequence (and thus could be encoded by a *DH* segment) is underlined (at least 4 bases corresponding to the *D* segment), and putative N-region additions are indicated by italicized G and C nucleotides [37]. At least four of 22 newborn clones (18%) putatively utilize both *DH1* and *DH2* genes, as do most adult *VH* clones. The majority of newborn *VH* may rearrange only the *DH1* gene, or more likely, based upon the RSS sequences of the rearranging segments in most shark clusters (V-23, 12-D1-22, 12-D2-12, 23-) there may instead be extensive trimming of the *DH2* segment during rearrangement so that it cannot be recognized in the cDNA clones. There is a GC bias in the non-D-encoding CDR3 sequences for ~75% of the clones implicating *TdT* activity, similar to previous findings for neonatal IgNAR *VH* [14]. In summary, newborn IgM is expressed from multiple heterogeneous loci of one *VH* family that have conserved the Ig fold structure, and these

expressed *VH* genes are diverse in CDR3, indicative of *TdT* and *RAG* activity at birth.

cDNA clone 36E in group VI is an unrearranged *VH* containing intronic sequence. The deduced sequence of 36E has only 38–50% AA identity with the conventional *VH*, and thus forms a separate *VH* family (Table 1) [31]. This *VH* family contains the residues important for the Ig fold, with the exception of an Arg replacing the typical Gly at position 49 (Figure 3). Clone 36E *VH* has more charged residues, especially in CDR1 and FR2, than the canonical *VH*, and the CDR are not enriched in Ser, unlike the other IgM *VH*. National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) [38–40] identified this *VH* family is most similar to a monotypic horn shark IgM *VH* (79%) (accession #Z11776 [41]). Clone 36E *VH* has 56% and 48% identical AA residues, respectively, to the sandbar shark IgW (accession #U50606 [42]) and nurse shark IgW (accession #U51450 [43]), and only 36% identical AA to a representative sandbar shark IgM *VH* [44] (data not shown). This novel *VH* family, therefore, is most related to the unusual horn shark *VH* and it was present in their common ancestral gene pool ~120 MYA [24]. In addition, this *VH* has similarity to both IgM and IgW *VH* genes, suggesting a more ancient origin (see below). Identification of a productively rearranged transcript containing the constant gene exons for this novel IgM *VH* locus is needed to permit its placement in the appropriate Ig class and enable further study.

IgW VHs are diverse and consist of multiple families expressed at birth

The ontogeny of IgW expression in neonate sharks was investigated by screening nurse shark pup spleen and epigonal organ cDNA libraries with an IgW *VH* and partial *CH1* probe [5]. Clones were negatively selected for IgM *VH*⁺ cross-hybridizing clones and positively selected for strong, intermediate, and weak hybridization intensity signals. Their deduced AA sequences were aligned to the published nurse shark IgW *VH* (accession #U51450 [43]) in Clustal W and shown to form six groups (Figure 4). The

Table 2: Newborn IgM VH regions predominantly use only the DH1 gene. The aligned IgM cDNA clones from Figure 2 were analyzed for DH gene usage by comparison to the horn shark genomic DH genes (Accession #X13447 [33]). The DH gene portion that exactly matches the horn shark is underlined and italicized nucleotides indicate non-templated GC nucleotide additions. Clones that could not be conclusively identified as either DH gene are indicated by (?). CDR3 identical in sequence to both DH genes are indicated by (D1/D2). Clones 22E and 36E excluded due their non-productive rearrangement.

	Horn shark X13447	DH1 GGTACAGCAGTGGGT	DH2 ATATCTGGGTG	D1 GTA VG	D2 ISGW
I	54S	GTGGGTACTGGT		VGTG	
	30E	AGCCACGGGGAGAGGAGCCAGAGCCACTGTGGCTAT CTT (D1/D2)		SHGERSQSHCGYL	
II	1E	CAACTACAGGACTGGGATGATCCA		QLQDWDDP	
	72S	GACGACAGTGGGGGTCTCCC	ATACTGGGTCCGACCTACTAT	DDSGGSP	ILGPYY
	35S		GACGGAGTGATACTGGGA		DGVILG
III	14E	GACTACAGTG	CCTGGGATACCATTTGGT	DYS	AWDTIG
	17S		ACAACAACTCTCTGGGT		TTNSLG
	6E		ACAATATTCGGT		TIFG
IV	57S	ACTACTTTCGGT (?)		TTFG	
	55E		AGGGGGGGGACTGGGTAGGTTCCGCT		RGGDWGRFA
	26E	AAGACCCACCTACAGTCTATA	CTGGGGTCTTCGCT	KTHLQSI	LGSFA
	15S	GCCTCAGTGGAGCGGAATTCGCT		ASVEREFA	
	24S	GATGGTGGGGTAGGGGAATTCGCT (D1/D2)		DGGVGEFA	
	46S	AGCAGTGGGTACGAATTCGCT		SSGYEFA	
	29S	GTCAGGGGGGAGGCGGT (?)		VRGEAV	
V	73S	GACACAGTAGGGGGAGCG		DTVGGA	
	2S	ACAGAAAGTGGG	GGGGTGGGCGCGGT	TESG	GGGAG
	49S	ACAGATATCAGTGGCGCGGTAGCCGT		TDISGRVAG	
	M9285I	ACAAGCCATGAGTGGGTATGAATACT	TGGGCGGCCACTCCGCT	TRRMSGY EY	LGGHSG
	2E	ACCCGATCTACGGCGCCCTTTGT (?)		TRSTAPLC	
	50S	ACAGGGGGTACTTGGGT (D1/D2)		TLG	
	47S	ACAGTGGGCGGT		TVGG	
	21S	ACTACAGTGGGT		TTVG	
	27S	ACAGACAGTGGG		TDSG	

percent similarity of the FR and CDR AA sequences to the original VH region is summarized in Table 3. Groups I and II are members of the original VH family, with 73–78% identity. Groups III-V form three new IgW VH families, with 46–56% AA identity to the original VH [31]. Interestingly, similar to our observations for the novel IgM VH family clone 36E, CDR1 in groups II and V contain more charged residues than are found in the other groups. CDR3s of each IgW clone in groups I and V are diverse and contain unique AA residues, again suggesting that TdT is active on rearranging IgW loci in neonates. A genomic clone for nurse shark IgW has not been isolated and therefore identification of DH gene usage in the junctions is not possible. 5'UT microheterogeneity shows that IgW VH groups are expressed from 11 loci, suggesting that several highly related loci are represented by each group (data not shown). High stringency Southern blotting analysis of the conventional IgW VH family indicated that 4–6 loci exist in the nurse shark, and low stringency conditions revealed more VH+ bands, which may have been divergent IgW (or IgM) VH families [43]. Multiple IgW VH families were previously indicated by *in situ* hybridization results in the skate, *Raja eglanteria* [45]. Thus these data together suggest that the nurse shark and skate have an equivalent number of IgW VH loci.

In the first constant domain (CH1) of conventional Ig H chains a Cys in the A strand is typically present that forms a disulfide bond with light (L) chains; this position is Gly in group II clones and Arg in group I clone 8S (Figure 4) [31]. A Cys located in a more C-terminal position may be available for L chain association (data not shown). Mutation of the canonical Cys used for L chain association also is seen in IgM_{1gi} where the Ig heavy chain protein does associate with Ig light chain protein [5], and modeling of other Cys more C-terminal in the CH1 domain has shown they likely are available for disulfide bonding to L chains [5].

IgW VH (#U51450 [43]) structure from IMGT [46] was compared to representative members of the divergent groups III and V (refer to Figure 3). Both of these VH families maintain the conserved residues and positions required for the Ig fold. In addition, they contain more Pro residues in strands A and C than conventional IgW VH, which are also present in these same strands in cartilaginous fish IgM VH (ratfish, nurse and horn sharks, Figure 3) [26,33,47]. In summary, our results, along with previous data, demonstrate the existence and expression of several IgW VH families in newborns. These VH gene families conserve residues necessary for maintenance of

	Leader	FR1	CDR1	FR2	CDR2
I	14S	MGIAPNLCVLLVCLTGVWSEITLTQPESVVKKPGESHRLTCTVSGFSL	SSNSIHW	VVRQAPGKGLEWLA	AIASSGSKYYAPAVRD
	27S	-----	-----	-----	-----
	105S	-----	-----M-----	-----	-----
	1E	-----	-----YGM-----	-----	-----G-----
	8S	-----	-----M-----	-----/-----	-----S-----
	5E	-----	-----YGM-----	-----	-----G-R-----
II	130S	-----L-----K-----L--Y--A-----	DDYWM	-----	-----G-----G
	37E	-----L-----K-----L--Y--A-----	DDYWM	-----	-----G-----G
III	25E	-----F-L-----R-D-V-S--K--E--A--A-----	---YTMC---	---V-----	---ILSYE-ESN-N---G-Q-
IV	20E	-----G--L-----DQ--V-I--PIA--V--AI-----	RRYSG/////	---V-----Q--	LSYHKGYNTR--CG-HG
	90S	-----G--L-----Q--V-R--PIA-----A-----	EDINRHRM	---V-----Q--	LSYHKGYNTR--SG-QG
	9E	-----G--L-----Q--V-R--PIA-----A-----	EDINRHRM	---V-----Q--	LSYHKGYNTR--SG-QG
V	92S	-----G--L-----Q--V-R--PIA-----A-----	EDINRHRM	---V-----Q--	LSYHKGYNTR--SG-QG
	99S	-----G--L-----Q--V-R--PIA-----A-----	EDINRHRM	---V-----Q--	LSYHKGYNTR--SG-QG
	40S	-----G--L-----Q--V-R--PIA-----A-----	EDINRHRM	---V-----Q--	LSYHKGYNTR--SG-QG
	114S	-----G--L-----Q--V-R--PIA-----HEA-----	EDINRHRM	---V-----Q--	LSYHKGYNTR--SG-QG
	76S	-----G--L-----Q--V-R--PIA-----A-----	EDINRHRM	---V-----Q--	LSYHKGYNTR--SG-QG
VI	51450	---H-----L-----V-N-T-----L---K---E-E-G---	---GYG---	---T-----T---	---LT--A-----IQ-

	FR3	CDR3	FR4	CH1	*
I	14S	RFEISKDS/DTVYLQVTSLRVDDTAIYYCAS	//////////	RMGGSGTFLEVTSVAQSAPSVYISNP	SCDMNSNQD
	27S	-----/-----	//////////	-----	-----
	105S	-----/-----	//////////	GAYW	-----
	1E	-----/GA-----X-----	RDVYSWVNMRDYL	DYW	-----
	8S	-----/-----K---DF---	RG//////////	SYW	-----LF-----R-K---E
	5E	-----/GA-----	RD//////////	TVRG//////////	//////////
II	130S	-----/N-----N-----	RD//////////	WVDYW	-----DI-IK---L-STYG-TS---A
	37E	-----/N-----N-----	RD//////////	WVCVW	-----DI-IK---L-STYG-TS---A
III	25E	--TA--G-/DAF--M-D-----M---	RDN//////////	LMGY-EW	-----DV-IK---L-STF--TX---
IV	20E	-ITF-IE-SAATSIEIPN-T-----	RDAVN/////	GVVYW	-----DV-IK---L-ST---TSF-H-
	90S	-ITF-IE-SAATSIEIPN-T-----	REAGN/GVWL///	AHW	-----DV-IK---L-ST---TS---
	9E	-ITF-IE-SAATSIEIPN-T-----	REAGN/GVWT/////	//////////	//////////
	92S	-ITF-IE-SAATSIEIPN-T-----	REAGN/GVWGRYLAHW	-----	-----DV-IK---L-ST---TS---
V	99S	-ITF-IE-SAATSIEIPN-T-----	REAGN/GVWGRYLAHW	-----	-----DV-IK---L-ST---TS---
	40S	-ITF-IE-SAATSIEIPN-T-----	REAGN/GVWGRYLAHW	-----	-----DV-IK---L-ST---TS---
	4S	//////////	REAGN/////	NGG/////	//////////
	114S	-ITF-IE-SAATSIEIPN-T-----	REAGNG/////	VWIAHW	-----DV-IK---L-ST---TS---
	76S	-ITF-IE-SAATSIEIPN-T-----	REAGN/GVWGRYLAHW	-----	-----DV-IK---L-ST---TS---
VI	51450	-----/-----K--N-T-----	RGYHSGHATPPYLDY	W-D-----	-----DV-VK---V-ST---TS---

Figure 4
IgW VH form multiple families. The deduced amino acid sequences of the IgW cDNA clones (Genbank accession #AY609225-AY609246, AY531553-AY531554, AY524297) were compared to the predicted amino acid sequence of conventional IgW U51450 [43] and aligned in ClustalW with differences in FR and CDR separating the clones into six groups. Dashes (-) indicate identity with the first sequence and gaps introduced to correct alignment are indicated by slashes (/). The CDR3 are unique for most clones demonstrating the neonatal repertoire is diverse in the IgW isotype. The Cys (*) typically important for light chain association is mutated in several members of groups I and II.

the Ig fold, and they undergo rearrangements containing N-region additions in neonates.

The potential somatic hypermutation motif AGY is biased to FR1 and CDR1

The V region of secretory IgNAR in the adult nurse shark has a high frequency of mutation in adult sharks, yet this region is seldom mutated in young pups suggesting that the newborn immune system requires further maturation to enable an effective immune responsive environment

[14,48]. Previously the Ser codon motif AGY (Y=C/T) had been identified as a heavily targeted site of IgNAR somatic hypermutation [48]. Therefore, we analyzed the newborn IgM and IgW VH FR and CDR for the AGY potential hypermutation motif and Ser codon TCN (N=C/G/T/A), which tends not to be somatically mutated (Figure 5). The AGY motif was utilized significantly more frequently in IgM VH FR1 (p < 0.01) and CDR1 (p < 0.01) than the TCN codons and IgW also preferred the AGY codon motif for CDR1 (p < 0.01). Both IgM and IgW VH genes encode the

Table 3: Deduced amino acid sequence percent similarity of IgW (group VI) VH regions to newborn IgW groups I-V.

	FR1	CDR1	FR2	CDR2	FR3	VH
I	72	50	86	69	90	78
II	66	33	86	63	90	73
III	59	33	71	38	62	56
VI	52	–	71	13	45	50
V	55	17	71	19	45	46

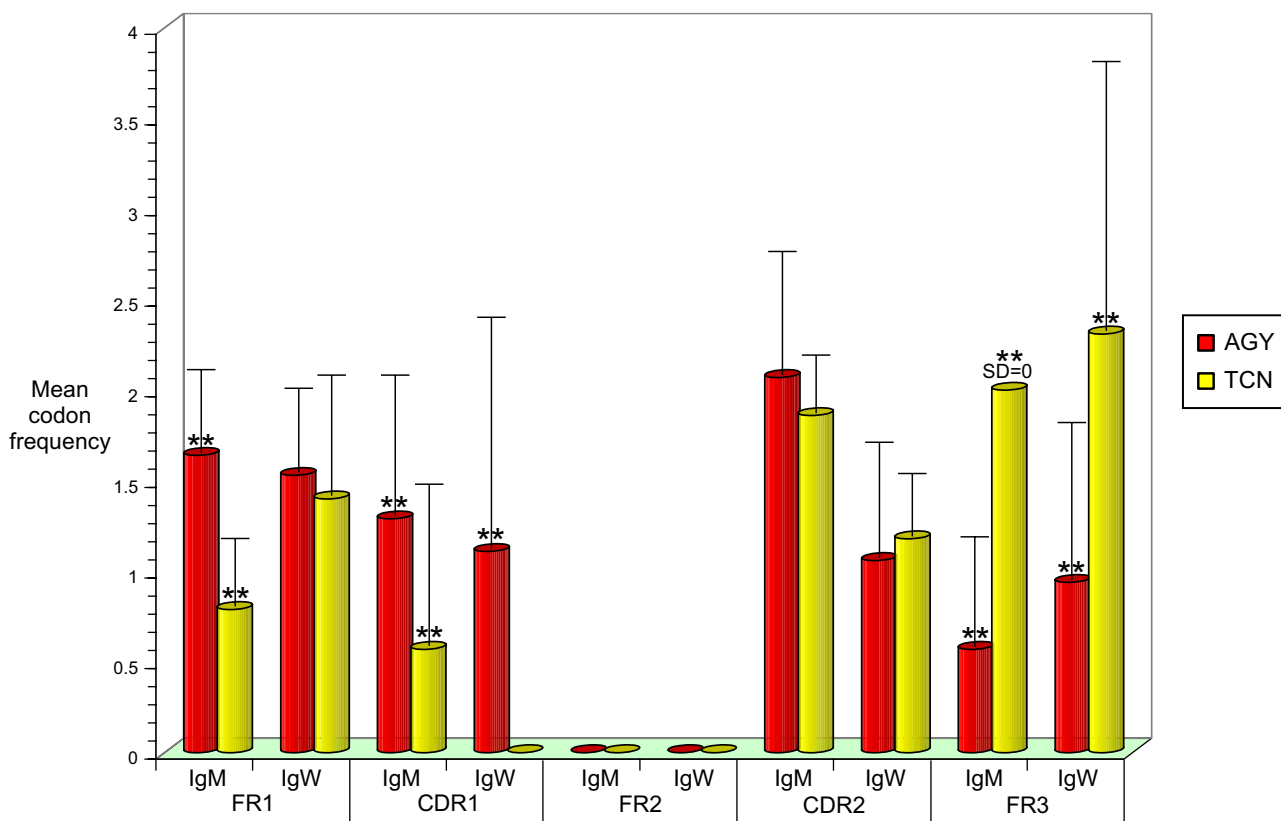


Figure 5
Potential somatic hypermutation motifs are restricted to the FR1 and CDR1 for IgM and IgW VH genes. The FR and CDR nucleotide sequences for all IgM and IgW cDNA clones were analyzed for the Ser codon potential somatic hypermutation motif AGC/T (AGY red square) and the non-hypermutated Ser codon nucleotide sequence TCC/T/A/G (TCN yellow square). IgM significantly uses the AGY sequence in the FR1 and CDR1 ($p < 0.01$, one-way factorial analysis of variance (ANOVA), $n = 23$) while IgW uses this motif in the CDR1 ($p < 0.01$, ANOVA, $n = 16$). The non-hypermutated motif was significantly enriched in the FR3 region for both Ig VH families ($p < 0.01$, ANOVA) and neither motif was present in all FR2 analyzed.

TCN sequence more frequently than AGY in FR3 ($p < 0.01$), but encode neither AGY nor TCN in FR2 implying a resistance to hypermutation in these regions, as seen in previous studies of tetrapod Ig genes [34]. Thus, the nurse shark VH genes have potential hypermutation sites

focused in FR1 and CDR1, while FR2 and FR3 lack motifs conducive to targeting somatic hypermutation.

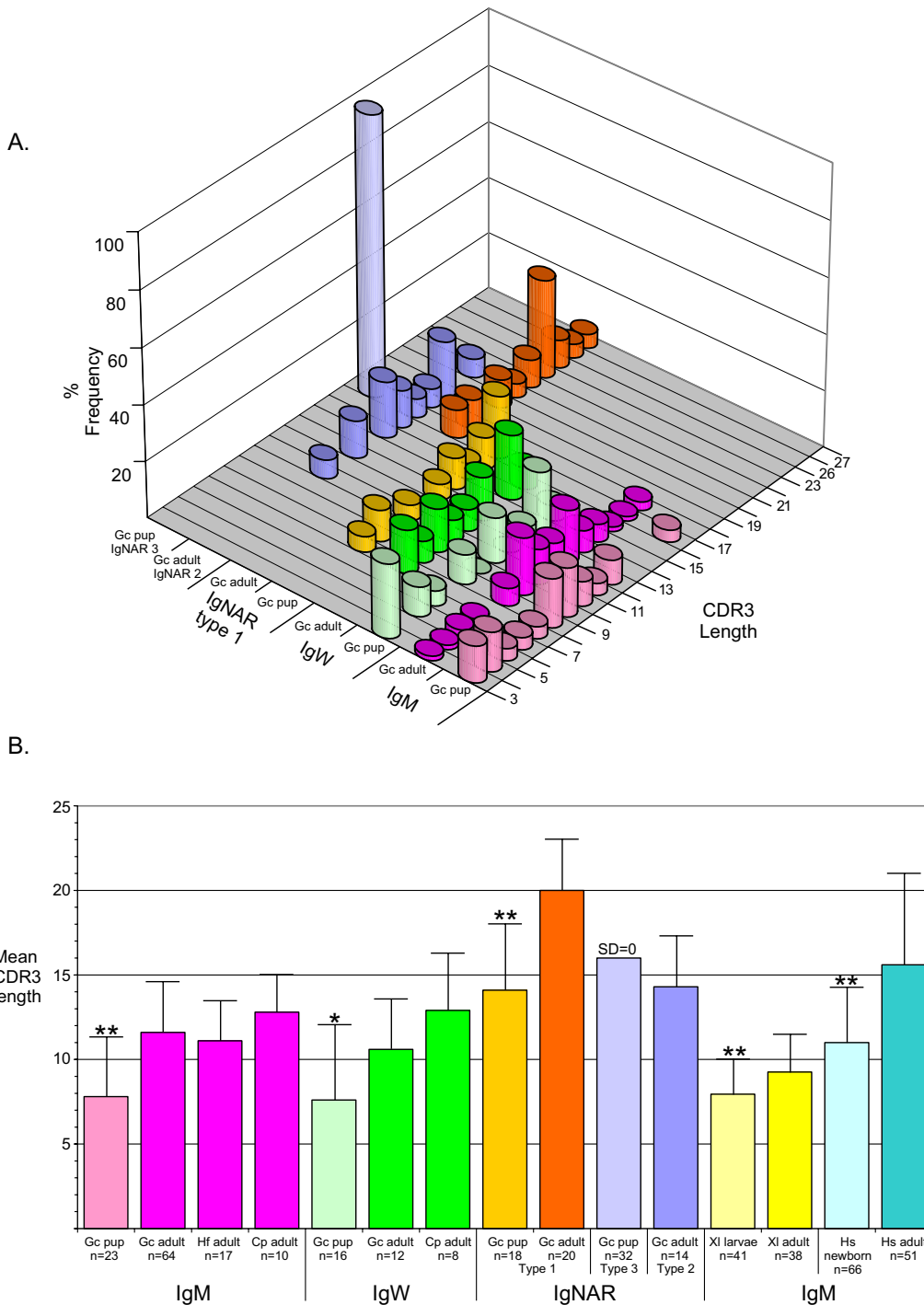


Figure 6
Newborn IgW and IgM CDR3 are significantly shorter than the adult correlate. A). Newborn and adult nurse sharks (Gc) frequency of specific CDR3 region lengths for IgM, IgW and IgNAR were compared. Newborn VH CDR3 is shorter in length more frequently than the adult for each Ig isotype except IgNAR type 3 (which is not expressed in detectable amounts in the adult). Sample size for each Ig isotype in newborns and adults are listed in part B. B). The mean CDR3 lengths for *Xenopus laevis* (Xl) [49,51] and human (Hs) [64] newborn and adult IgM VH sequences were compared to the newborn and adult shark IgM, IgW and IgNAR VH CDR3 mean lengths. Significantly the newborn shark CDR3 lengths for each Ig class follow the same pattern of preference for shorter CDR3 lengths as seen in the newborn frog (larvae) and human. Significance determined by ANOVA test with ** = p value <0.01 and * = p value <0.05.

Newborn IgW and IgM CDR3 are shorter than in adults

Newborn mouse, human, and frog *VH* CDR3s are less diverse and shorter in length due to non-random *VH* gene usage and lack of *TdT* activity [16,17,49-51]. In newborn elasmobranchs *TdT* is expressed and functional early in shark ontogeny, as shown in previous studies of *TdT* expression and implied by the diversity of the newborn CDR3 repertoire in the IgM and IgW classes (Figures 2, 4, Table 2) and IgNAR types 1 and 2 [5,14,48]. The frequency of IgW and IgM *VH* CDR3 lengths in newborn nurse sharks were compared to adult nurse, horn, and sandbar sharks (Figure 6A). Newborn IgM *VH* utilize significantly shorter CDR3 than the adult with the newborn mean length 7.8 AA ($p < 0.01$, range 3–17 AA, $n = 23$) and the adult mean length 11.6 AA ($p < 0.01$, range 4–18 AA, $n = 64$). This is consistent with our findings that expressed newborn IgM *VH* tend to have only one recognizable *DH* gene in CDR3 (Table 2). As well, newborn IgW *VH* CDR3s are significantly shorter than adult CDR3s with the newborn mean 7.6 AA ($p < 0.05$, range 2–13, $n = 16$) and the adult mean 10.6 AA ($p < 0.05$, range 7–15, $n = 12$). Previously published newborn and adult IgNAR were also included in the *VH* CDR3 length analysis, although IgNAR CDR3s are longer because this *VH* gene undergoes four rearrangement events (*V-D1-D2-D3-JH*). Neonatal IgNAR *VH* Type 1 CDR3 mean length 14.1 AA ($p < 0.01$, range 8–20 AA, $n = 18$, [14]) was also significantly shorter than those of the adult CDR3 mean 20 AA ($p < 0.01$, range 15–26, $n = 20$, [14]). As stated previously, the IgNAR Type 3 *VH* locus has two germline-joined *DH* genes and its length is always 16 AA ($n = 32$) for all clones analyzed, suggesting that this *VH* is positively selected during ontogeny on a self-ligand [14]. As Type 3 *VH* expression wanes during postnatal maturation it was not possible to compare this gene's expression in adult life [14]. Together these results show that newborn IgW, IgM, and IgNAR *VH* CDR3 are significantly shorter than those of adults, indicating that the postnatal Ig *VH* repertoire is not entirely mature. In addition, the results suggest that the rearrangement of Ig genes and the development of B cells from immature precursors persist in primary lymphoid tissues after birth in cartilaginous fish. To confirm these findings the shark Ig *VH* CDR3 lengths were compared to previously published newborn and adult frog and human IgM *VH* CDR3s (Figure 6B). Despite the fact that all newborn sharks *VH* CDR3 contain N-region additions, indicative of functional *TdT* gene activity not apparent in frog and human newborn sequences, the contrast of CDR3 sizes in young and old sharks is similar to that of tetrapod CDR3 during ontogeny. Thus, a developmental program of postnatal maturation, defined by longer CDR3, is demonstrated in sharks, frogs, and humans, which spans 460 million years of evolution.

IgM and IgW classes were present in the ancestors of all jawed vertebrates

¹The identification of IgW in the lungfish, close relatives of land vertebrates, substantially changed the phylogenetic distribution and evolution of this Ig class, since these lobe-finned bony fish shared a common ancestor with cartilaginous fish more than 460 MYA (refer to Figure 1) [21]. ²Accordingly, IgW *VH* family is more ancient than previously believed and may be the primordial Ig rather than IgM, as initially hypothesized [43]. ³The lungfish discovery allowed a more thorough phylogenetic analysis of the IgM and IgW *VH* families (Figure 7). ⁴The neighbor-joining phylogram was constructed using *VH* FR1-3, excluding the CDR to reduce effects caused by positive selection, and was rooted using the outgroup mouse (*M. musculus*) L chain V kappa (accession #29725591) [52-55]. ⁵Representative members from each newborn IgM and IgW *VH* group were included in the tree, in addition to representative IgM and IgW *VH* sequences from the lungfish (*P. aethiopicus*), ratfish (*H. colliei*-holocephali), skate (*R. erinacea*-elasmobranch), guitarfish (*R. productus*-ray, elasmobranch), sandbar shark (*C. plumbeus*-elasmobranch), horn shark (*H. francisci*-elasmobranch), and nurse shark (*G. cirratum*) (refer to Figures 1,2 and 4). Notably this analysis reveals that IgM and IgW cluster as separate *VH* families among the cartilaginous and bony fish lineages, demonstrating that the IgW *VH* family is ancient. The monotypic horn shark IgM (accession #Z11776) and its ortholog, the nurse shark divergent IgM group VI clone 36E, segregate with the IgW *VH* group indicating that they have features of both the IgM and IgW *VH* families. This phylogenetic analysis confirms the findings of the lungfish IgW discovery, namely that IgM and IgW *VH* are ancient gene families that were present in the common ancestor of cartilaginous and bony fish, which most likely was a placoderm, an extinct heavily-armored fish group considered to be the earliest vertebrates with jaws [24]. Finally, most of the shark and skate IgM sequences, excluding clone 36E, cluster within a species, which suggests that the various *VH* groups emerged after the divergence of each species.

Discussion

We show that the newborn nurse shark expresses five IgM *VH* genes from one conventional *VH* family, as shown previously for the adult horn shark, sandbar shark, and ratfish [33]. Additionally, newborn nurse sharks express at least two other *VH* families (36E and IgM_{1gj} [5]). Newborn IgM *VH* CDR3 are diverse with N-nucleotide additions, which shows that *TdT* is active in embryonic life, as previously documented for newborn IgNAR CDR3 [5,14]. The majority of IgM loci are expressed at birth. Among the five *VH* groups these loci differ mainly in the CDR1 and CDR2, implying that diversity-enhancing selection pressures are active in germline CDR as previously reported for

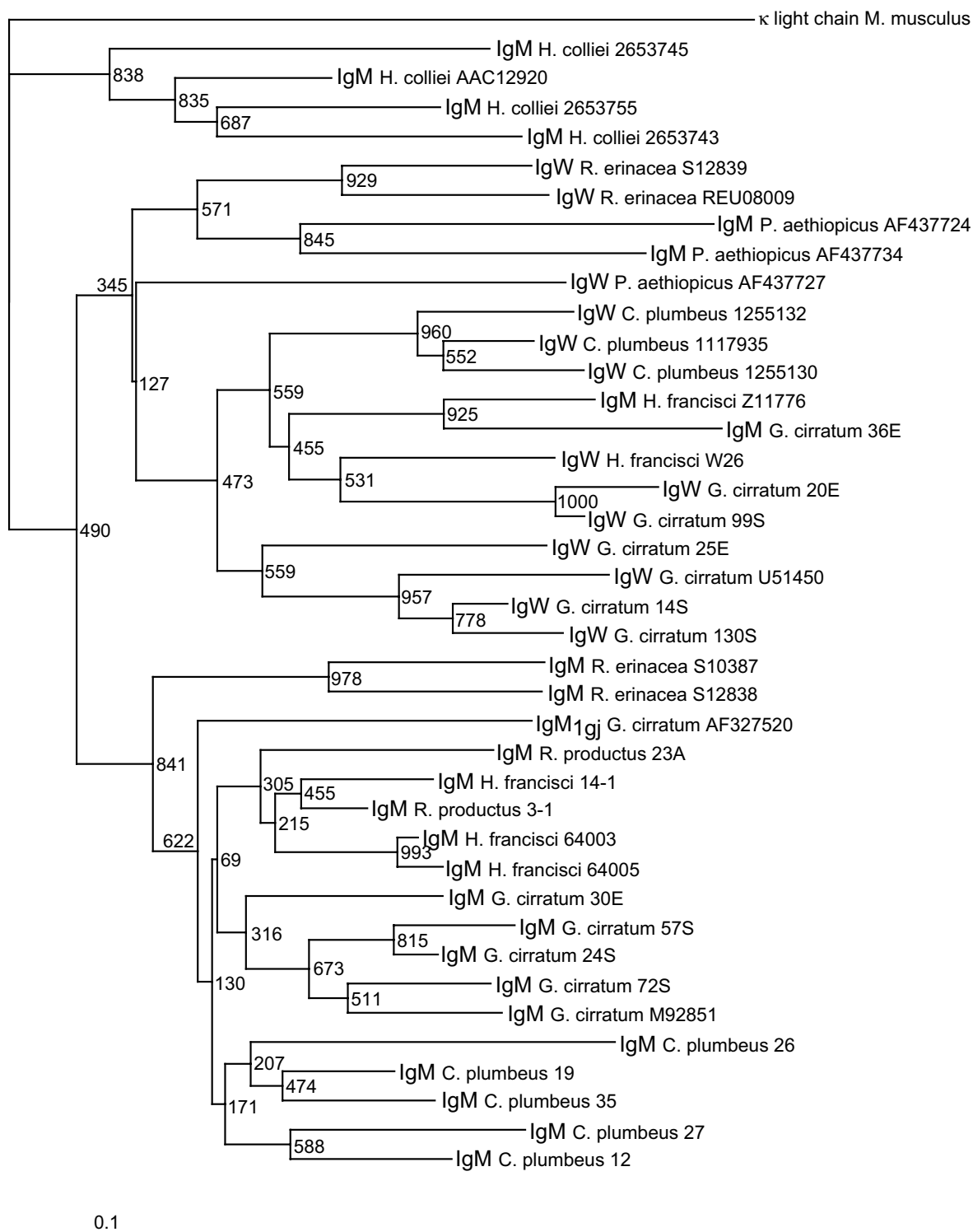


Figure 7

IgW and IgM VH are ancient gene families. A neighbor-joining tree was drawn in PHYLIP using cartilaginous fish and lobe-finned bony fish IgM and IgW VH sequences FR1-FR3 omitting the CDRI-2 from published sources (accession numbers noted next to sample) and this report (identified by clone numbers from Figures 2 and 4). The tree shows the various fish species cluster together by VH family rather than by their evolutionary relatedness indicating the IgW VH gene family is as ancient as the IgM VH gene family.

other vertebrates [53]. The IgM *VH* CDR3 length is significantly shorter in neonates, showing that a developmental program may control the CDR3 length. One explanation for this finding may be that there is expression of different *VH* clusters in adults that have longer *DH* and *JH* gene segments, as is seen in the mouse [56]; this possibility seems unlikely as germline diversity seems similar in neonates and adults (although the adult nurse shark repertoire must be studied in more detail to prove this point). Secondly, *TdT* expression in pro/pre B cells in adults may be higher than in young animals [57]. Finally, *TdT* activity may be modified (e.g. decreased N-region addition or increased exonuclease activity – our data suggest that *DH2* segments may be extensively trimmed in neonatal junctions) by expression of splice variants or developmental regulation of DNA-dependent protein kinase, a known modulator of *TdT* activity [58,59].

Newborn IgW *VH* expression is from at least three rather divergent *VH* families [43]. CDR3 lengths of IgW are also significantly shorter than that of adults, as seen for IgM, implying that a developmental program selects for increased CDR3 length as the animals mature.

Potential somatic hypermutation Ser AGY motifs are preferentially encoded in the FR1 and CDR1, whilst the untargeted TCN motif is present in FR3. No Ser residues at all are encoded in FR2. This tendency to target mutations predominantly in the CDR is expected, but is paradoxical in FR1 of IgM *VH*. There are generally fewer mutations in FR1 in Ig from all vertebrates so far examined. This suggests that despite the presence of hypermutation targeting motifs in this region, there is a higher order of control over the mutational mechanism that inhibits the targeting to FR1 AGY hotspots.

The IgM and IgW *VH* phylogenetic analysis suggest that both *VH* families were present in the cartilaginous and bony fish ancestral pool. A more thorough understanding of the evolution of these *VH* families will require identification of IgW in the cartilaginous fish class holocephali and in bony fish groups such as the chondrosteans, holosteans and teleosts. Furthermore, isolation of all of the germline genes is required to examine relative expression levels of the various genes, and the level of diversity gained by somatic hypermutation after antigenic stimulation of B cells.

Conclusion

Most IgM loci expressed at birth in nurse shark are from one heterogeneous highly-related *VH* family which differs in the CDR1-2. This repertoire is increased in its diversity via N-region addition in CDR3. In spite of the N-region addition the newborn *VDJ* junction is significantly shorter than the adult, most likely due to extensive trimming of

the *DH2* segment during rearrangement. In contrast, nurse shark IgW *VH* genes have diverged to form at least three *VH* families expressed at birth.

Methods

Animals

Nurse shark pups were delivered by Caesarian section from a gestating female shark near term as described [60].

cDNA library construction and screening

Tissues were dissected from nurse shark pups and total RNA was isolated as described [5]. The newborn pup spleen and epigonal cDNA libraries were constructed as described previously [5]. Libraries were plated and screened with canonical nurse shark IgM *VH* probe under both high and low stringency conditions as described [5,61]. Canonical nurse shark IgW *VH* and partial *CH1* probe was amplified by PCR from plasmid DNA containing cloned cDNA insert using specific primers and labeled as described [5,43]. Clones were selected and isolated based on hybridization signal intensity of strong, intermediate and weak with more than 60 clones analyzed for each Ig class.

Alignment and phylogenetic analysis

IgW and IgM cDNA clones were translated into amino acid sequences using the EXPASY translate tool [62] and aligned in ClustalW v1.8 [63] for alignment analysis or ClustalX v 1.8 for phylogenetic tree analysis [27,55]. Phylogenetic tree analysis was performed using the *VH* regions from FR1-FR3 excluding CDR1-2. Amino acid sequences were aligned in ClustalX v1.8 using the multiple alignment parameter which does pairwise alignments in the Gonnet series protein weight matrix under default conditions of 10.00 gap opening, 0.20 gap extension, and 30% delay divergent sequences. A neighbor-joining (NJ) tree in PHYLIP output was drawn using a dendrogram as the guide and the reliability of branching order was determined by 1000 replications (bootstrap analysis) [52,55]. The NJ phylogram tree was drawn in Treeview v1.6.6 and rooted using mouse VL kappa as outgroup [54]. The NJ tree was labeled in Canvas v9.0 (ACD Deneba Software, Miami, FL, USA). Genbank and Swiss-Prot and TrEMBL accession numbers used for phylogenetic analysis are as follows: *M. musculus* kappa LC 29725591; *P. aethiopicus* IgW AF437727 clone 28; *P. aethiopicus* IgM AF437734 clone 76; *P. aethiopicus* IgM AF437724 clone 27; *R. erinacea* IgM S10387; *R. erinacea* IgM S12838; *R. erinacea* IgW S12839; *R. erinacea* IgW REU08009; *C. plumbeus* IgW 1117935; *C. plumbeus* IgW1255130; *C. plumbeus* IgW 1255132; *H. colliei* IgM AAC12920; *H. colliei* IgM 2653745; *H. colliei* IgM 2653755; *H. colliei* IgM 2653743; *H. francisci* IgM 64003; *H. francisci* IgM 64005; *H. francisci* IgM Z11776; *H. francisci* IgW C6-26m13f (clone W26) P83907; *H. francisci* IgM clone 14-1 AY612427; *C.*

plumbeus IgM clones 12, 19, 26, 27, 35 [44]; *G. cirratum* IgW U51450; *G. cirratum* IgM_{1gJ} AF327520; *G. cirratum* IgM M92851; *G. cirratum* IgM 57S AY609270; *G. cirratum* IgM 30E AY609260; *G. cirratum* IgM 36E AY609263; *G. cirratum* IgM 72S AY609272; *G. cirratum* IgM 24S AY609256; *G. cirratum* IgW14S AY609229-AY609230; *G. cirratum* IgW 20E AY609231-AY609232; *G. cirratum* IgW 25E AY531553-AY531554; *G. cirratum* IgW 99S AY609242-AY609243; *G. cirratum* IgW 130S AY609246; *R. productus* IgM clone 23A AY612424-AY612425; *R. productus* IgM 3-1 AY612426.

Adult and newborn CDR3 length comparisons

Adult nurse shark IgM VH cDNA Genbank accession numbers AY608337-AY608404 (clones Mary M2-Mary M34, Jesus M3-M47, and Joseph M1-M27). Newborn nurse shark IgM cDNA clones 1E, 2E, 2S, 6E, 14E, 15S, 17S, 21S, 22E, 24S, 26E, 27S, 29S, 30E, 35S, 36E, 46S, 47S, 49S, 50S, 54S, 55E, 57S, 72S, 73S (Genbank accession numbers AY609247-AY609274). Newborn nurse shark IgW cDNA clones 1E, 5E, 8S, 9E, 14S, 20E, 25E, 27S, 37E, 40S, 76S, 90S, 92S, 99S, 105S, 114S, 130S (Genbank accession numbers AY609225-AY609246, AY531553-AY531554, AY524297). Sandbar shark and horn shark IgM accession numbers listed in the phylogenetic tree analysis were used for determination of CDR3 length. Human (Hs) preterm neonate and adult IgM VH cDNA sequences were obtained from Zemlin et al [64] and *Xenopus laevis* (Xl) 5-48 day larvae and adult IgM VH cDNA sequences published in Schwager et al [51] and Du Pasquier et al [49].

Authors' contributions

LR dissected shark tissue, isolated RNA, constructed and screened cDNA libraries, isolated cDNA clones, performed phylogenetic and statistical analyses, wrote the manuscript draft and made the figures. BL screened nurse shark PBL (Y) cDNA library with IgW VH probe and isolated, analyzed and aligned positive clones. Those results have been reported in a paper currently in submission. HD provided 64 IgM VH cDNA sequences from three immunized adult nurse sharks for the statistical study of adult IgM CDR3 lengths. MF delivered the shark pups by Caesarian section, and participated in the bleeding and dissection of the sharks; he also coordinated the study, participated in the analysis of results, and the manuscript draft. All authors read and approved the final manuscript.

Acknowledgements

This study was supported by NIH grant RR06603 (MFF).

References

- Litman GW, Anderson MK, Rast JP: **Evolution of antigen binding receptors.** *Annu Rev Immunol* 1999, **17**:109-147.
- Flajnik MF, Rumpf LL: **The immune system of cartilaginous fish.** *Curr Top Microbiol Immunol* 2000, **248**:249-270.
- Haire RN, Miracle AL, Rast JP, Litman GW: **Members of the Ikaros gene family are present in early representative vertebrates.** *J Immunol* 2000, **165**:306-312.
- Anderson MK, Sun X, Miracle AL, Litman GW, Rothenberg EV: **Evolution of hematopoiesis: Three members of the PU.1 transcription factor family in a cartilaginous fish, *Raja erinacea*.** *Proc Natl Acad Sci U S A* 2001, **98**:553-558.
- Rumpf LL, Avila D, Diaz M, Bartl S, McKinney EC, Flajnik MF: **A shark antibody heavy chain encoded by a nonsomatically rearranged VDJ is preferentially expressed in early development and is convergent with mammalian IgG.** *Proc Natl Acad Sci U S A* 2001, **98**:1775-1780.
- Bernstein RM, Schluter SF, Bernstein H, Marchalonis JJ: **Primordial emergence of the recombination activating gene 1 (RAG1): sequence of the complete shark gene indicates homology to microbial integrases.** *Proc Natl Acad Sci U S A* 1996, **93**:9454-9459.
- Schluter SF, Marchalonis JJ: **Cloning of shark RAG2 and characterization of the RAG1/RAG2 gene locus.** *FASEB J* 2003, **17**:470-472.
- Zapata AG, Torroba M, Sacedón R, Varas A, Vicente A: **Structure of the lymphoid organs of elasmobranchs.** *J Exp Zool* 1996, **275**:125-143.
- Hinds KR, Litman GW: **Major reorganization of immunoglobulin VH segmental elements during vertebrate evolution.** *Nature* 1986, **320**:546-549.
- Du Pasquier L, Flajnik MF: **Origin and Evolution of the Vertebrate Immune System.** *Fundamental Immunology* 4th edition. Edited by: Paul W E. Philadelphia, Lippencott-Raven; 1999:621.
- Fidler JE, Clem LW, Small PA: **Immunoglobulin synthesis in neonatal nurse sharks (*Ginglymostoma cirratum*).** *Comp Biochem Physiol* 1969, **31**:365-371.
- Rumpf L, McKinney E, Taylor E, Flajnik M: **The Development of Primary and Secondary Lymphoid Tissues in the Nurse Shark *Ginglymostoma cirratum*: B-Cell Zones Precede Dendritic Cell Immigration and T-Cell Zone Formation During Ontogeny of the Spleen.** *Scand J Immunol* 2002, **56**:130-148.
- Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC, Flajnik MF: **A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks.** *Nature* 1995, **374**:168-173.
- Diaz M, Stanfield RL, Greenberg AS, Flajnik MF: **Structural analysis, selection, and ontogeny of the shark new antigen receptor (IgNAR): identification of a new locus preferentially expressed in early development.** *Immunogenetics* 2002, **54**:501-512.
- Miracle AL, Anderson MK, Litman RT, Walsh CJ, Luer CA, Rothenberg EV, Litman GW: **Complex expression patterns of lymphocyte-specific genes during the development of cartilaginous fish implicate unique lymphoid tissues in generating an immune repertoire.** *Int Immunol* 2001, **13**:567-580.
- Kearney JF, Won WJ, Benedict C, Moratz C, Zimmer P, Oliver A, Martin F, Shu F: **B cell development in mice.** *Int Rev Immunol* 1997, **15**:207-241.
- Zemlin M, Schelonka RL, Bauer K, Schroeder HW: **Regulation and chance in the ontogeny of B and T cell antigen receptor repertoires.** *Immunol Res* 2002, **26**:265-278.
- Herzenberg LA: **Toward a layered immune system.** *Cell* 1989, **59**:953-954.
- Harding FA, Amemiya CT, Litman RT, Cohen N, Litman GW: **Two distinct immunoglobulin heavy chain isotypes in a primitive, cartilaginous fish, *Raja erinacea*.** *Nucleic Acids Res* 1990, **18**:6369-6376.
- Anderson MK, Strong SJ, Litman RT, Luer CA, Amemiya CT, Rast JP, Litman GW: **A long form of the skate IgX gene exhibits a striking resemblance to the new shark IgW and IgNARC genes.** *Immunogenetics* 1999, **49**:56-67.
- Ota T, Rast JP, Litman GW, Amemiya CT: **Lineage-restricted retention of a primitive immunoglobulin heavy chain isotype within the Dipnoi reveals an evolutionary paradox.** *Proc Natl Acad Sci U S A* 2003, **100**:2501-2506.
- Zhu M, Yu X: **A primitive fish close to the common ancestor of tetrapods and lungfish.** *Nature* 2002, **418**:767-770.
- Takezaki N, Figueroa F, Zaleska-Rutczynska Z, Klein J: **Molecular phylogeny of early vertebrates: monophyly of the agnathans as revealed by sequences of 35 genes.** *Mol Biol Evol* 2003, **20**:287-292.

24. Carroll RL: **Vertebrate Paleontology and Evolution**. New York, W. H. Freeman and Co.; 1988:698.
25. Flajnik MF: **Comparative analyses of immunoglobulin genes: surprises and portents**. *Nat Rev Immunol* 2002, **2**:688-698.
26. Vazquez M, Mizuki N, Flajnik MF, McKinney EC, Kasahara M: **Nucleotide sequence of a nurse shark immunoglobulin heavy chain cDNA clone**. *Mol Immunol* 1992, **29**:1157-1158.
27. 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.
28. Harding FA, Cohen N, Litman GW: **Immunoglobulin heavy chain gene organization and complexity in the skate, Raja erinacea**. *Nucleic Acids Res* 1990, **18**:1015-1020.
29. Lefranc M-P: **Unique database numbering system for immunogenetic analysis**. *Immunology Today* 1997, **18**:509.
30. Williams AF, Barclay AN: **The immunoglobulin superfamily-domains for cell surface recognition**. *Annu Rev Immunol* 1988, **6**:381-405.
31. Frazer JK, Capra JD: **Immunoglobulin: Structure and function**. *Fundamental Immunology* 4th edition. Edited by: Paul W E. Philadelphia, Lippencott-Raven; 1999:37-74.
32. Scaviner D: **Collier de Perles: Nurse shark (Ginglymostoma cirratum) IGHV Rearranged IGHV1S1 (M92851)**. 2003 [http://imgt.cines.fr/textes/IMGTr repertoire/2D-3Dstruct/colliers/Nshark/IGH/IGHV/Ns_IGHV1S1.html].
33. Kokubu F, Litman R, Shablott MJ, Hinds K, Litman GW: **Diverse organization of immunoglobulin VH gene loci in a primitive vertebrate**. *Embo J* 1988, **7**:3413-3422.
34. Golub R, Charlemagne J: **Structure, Diversity, and Repertoire of VH Families in the Mexican Axolotl**. *J Immunol* 1998, **160**:1233-1239.
35. Hsu E: **Canonical VH CDR1 nucleotide sequences are conserved in all jawed vertebrates**. *Int Immunol* 1996, **8**:847-854.
36. Chang B, Casali P: **The CDR1 sequences of a major proportion of human germline Ig VH genes are inherently susceptible to amino acid replacement**. *Immunology Today* 1994, **15**:367-373.
37. Feeney A: **Lack of N regions in fetal and neonatal mouse immunoglobulin V-D-J junctional sequences**. *J Exp Med* 1990, **172**:1377-1390.
38. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST**. 1997, 1997: [<http://www.ncbi.nlm.nih.gov/BLAST/>].
39. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic Acids Res* 1997, **25**:3389-3402.
40. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *J Mol Biol* 1990, **215**:403-410.
41. Hinds-Frey KR, Nishikata H, Litman RT, Litman GW: **Somatic variation precedes extensive diversification of germline sequences and combinatorial joining in the evolution of immunoglobulin heavy chain diversity**. *J Exp Med* 1993, **178**:815-824.
42. Berstein RM, Schluter SF, Shen S, Marchalonis JJ: **A new high molecular weight immunoglobulin class from the carcharhine shark: implications for the properties of the primordial immunoglobulin**. *Proc Natl Acad Sci U S A* 1996, **93**:3289-3293.
43. Greenberg AS, Hughes AL, Guo J, Avila D, McKinney EC, Flajnik MF: **A novel "chimeric" antibody class in cartilaginous fish: IgM may not be the primordial immunoglobulin**. *Eur J Immunol* 1996, **26**:1123-1129.
44. Shen SX, Bernstein RM, Schluter SF, Marchalonis JJ: **Heavy-chain variable regions in carcharhine sharks: development of a comprehensive model for the evolution of VH domains among the gnathanstomes**. *Immunol Cell Biol* 1996, **74**:357-364.
45. Anderson M, Amemiya C, Luer C, Litman R, Rast J, Niiimura Y, Litman G: **Complete genomic sequence and patterns of transcription of a member of an unusual family of closely related, chromosomally dispersed Ig gene clusters in Raja**. *Int Immunol* 1994, **6**:1661-1670.
46. Scaviner D: **Collier de Perles: Nurse shark (Ginglymostoma cirratum) IGHV Rearranged IGHV1S3 (U51450)**. 2003 [http://imgt.cines.fr/textes/IMGTr repertoire/2D-3Dstruct/colliers/Nshark/IGH/IGHV/Ns_IGHV1S3.html].
47. Rast JP, Amemiya CT, Litman RT, Strong SJ, Litman GW: **Distinct patterns of IgH structure and organization in a divergent lineage of chondrichthyan fishes**. *Immunogenetics* 1998, **47**:234-245.
48. Diaz M, Greenberg AS, Flajnik MF: **Somatic hypermutation of the new antigen receptor gene (NAR) in the nurse shark does not generate the repertoire: possible role in antigen-driven reactions in the absence of germinal centers**. *Proc Natl Acad Sci U S A* 1998, **95**:14343-14348.
49. Du Pasquier L, Robert J, Courtet M, Musmann R: **B-cell development in the amphibian Xenopus**. *Immunol Rev* 2000, **175**:201-213.
50. Feeney AJ: **Comparison of junctional diversity in the neonatal and adult immunoglobulin repertoires**. *Int Rev Immunol* 1992, **8**:113-122.
51. Schwager J, Mikoryak CA, Steiner LA: **Amino acid sequence of heavy chain from Xenopus laevis IgM deduced from cDNA sequence: implications for evolution of immunoglobulin domains**. *Proc Natl Acad Sci U S A* 1988, **85**:2245-2249.
52. Saitou N, Nei M: **The neighbor-joining method: a new method for constructing phylogenetic trees**. *Mol Biol Evol* 1987, **4**:406.
53. Tanaka T, Nei M: **Positive darwinian selection observed at the variable-region genes of immunoglobulins**. *Mol Biol Evol* 1989, **6**:447-459.
54. Page RD: **Treeview: an application to display phylogenetic trees on personal computers**. *Comput Appl Biosci* 1996, **12**:357-358.
55. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools**. *Nucleic Acids Res* 1997, **24**:4876-4882.
56. Bangs LA, Sanz IE, Teale JM: **Comparison of D, JH, and junctional diversity in the fetal, adult, and aged B cell repertoires**. *J Immunol* 1991, **146**:1996-2004.
57. Benedict CL, Gilfillan S, Thai TH, Kearney JF: **Terminal deoxynucleotidyl transferase and repertoire development**. *Immunol Rev* 2000, **175**:150-157.
58. Thai TH, Purugganan MM, Roth DB, Kearney JF: **Distinct and opposite diversifying activities of terminal transferase splice variants**. *Nat Immunol* 2002, **3**:457-462.
59. Mickelsen S, Snyder C, Trujillo K, Bogue M, Roth DB, Meek K: **Modulation of terminal deoxynucleotidyltransferase activity by the DNA-dependent protein kinase**. *J Immunol* 1999, **163**:834-843.
60. Ohta Y, Okamura K, McKinney EC, Bartl S, Hashimoto K, Flajnik MF: **Primitive syntenic of vertebrate major histocompatibility complex class I and class II genes**. *Proc Natl Acad Sci U S A* 2000, **97**:4712-4717.
61. Bartl S, Baish MA, Flajnik MF, Ohta Y: **Identification of class I genes in cartilaginous fish, the most ancient group of vertebrates displaying an adaptive immune response**. *J Immunol* 1997, **159**:6097-6104.
62. Swiss-Prot: **ExpPASy (Expert Protein Analysis System) Swiss Institute of Bioinformatics Translate tool**. [<http://ca.expasy.org/tools/dna.html>].
63. EMBL-EBI: **Clustal W**. [<http://www.ebi.ac.uk/clustalw/>].
64. Zemlin M, Bauer K, Hummel M, Pfeiffer S, Devers S, Zemlin C, Stein H, Versmold HT: **The diversity of rearranged immunoglobulin heavy chain variable region genes in peripheral blood B cells of preterm infants is restricted by short third complementarity-determining regions but not by limited gene segment usage**. *Blood* 2001, **97**:1511-1513.