



Published in final edited form as:

Immunogenetics. 2012 August ; 64(8): 571–590. doi:10.1007/s00251-012-0616-2.

Evolution of the B7 family: Co-evolution of B7H6 and NKp30, identification of a new B7 family member, B7H7, and of B7's historical relationship with the MHC

Martin F. Flajnik,

Dept of Microbiology and Immunology University of Maryland 685 West Baltimore Street, HSF1 Suite 380, Baltimore MD 21201 USA

Tereza Tlapakova,

Faculty of Science, Department of Cell Biology Charles University in Prague Vinicna 7, Prague 2, Czech Republic, 12844

Michael F. Criscitiello,

Department of Veterinary Pathobiology, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX 77843 USA

Vladimir Krylov, and

Faculty of Science, Department of Cell Biology Charles University in Prague Vinicna 7, Prague 2, Czech Republic, 12844

Yuko Ohta

Dept of Microbiology and Immunology University of Maryland 685 West Baltimore Street, HSF1 Suite 380, Baltimore MD 21201 USA yota@som.umaryland.edu

Abstract

The B7 family of genes is essential in the regulation of the adaptive immune system. Most B7 family members contain both variable (V)- and constant (C)-type domains of the immunoglobulin superfamily (IgSF). Through *in silico* screening of the *Xenopus* genome and subsequent phylogenetic analysis, we found novel genes belonging to the B7 family, one of which is the recently discovered *B7H6*. Humans and rats have a single *B7H6* gene, however many *B7H6* genes were detected in a single large cluster in the *Xenopus* genome. The *B7H6* expression patterns also varied in a species-specific manner. Human *B7H6* binds to the activating natural killer receptor, NKp30. While the *NKp30* gene is single-copy and maps to the MHC in most vertebrates, many *Xenopus NKp30* genes were found in a cluster on a separate chromosome that does not harbor the MHC. Indeed, in all species so far analyzed from sharks to mammals, the number of *NKp30* and *B7H6* genes correlates well, suggestive of receptor-ligand co-evolution. Furthermore, we identified a *Xenopus*-specific B7 homolog (*B7HXen*) and revealed its close linkage to *B2M*, which we have demonstrated previously to have been originally encoded in the MHC. Thus, our study provides further proof that the B7 precursor was included in the proto MHC. Additionally, the comparative analysis revealed a new B7 family member, B7H7, which was previously designated in the literature as an unknown gene, *HHLA2*.

Keywords

B7 family; MHC; evolution; natural killer receptors; genetic linkage; *Xenopus*

Introduction

B7 ligands are expressed on the cell surface of many different cell types including APCs and their interaction with receptor molecules on T cells provide activating and/or inhibitory signals that regulate T cell activation and tolerance (Collins et al. 2005). Some inhibitory B7 ligands are also expressed on tumor cells, resulting in suppression of immune responses (Keir et al. 2008; Zou and Chen 2008). Therefore, stimulating or attenuating the interactions of B7 ligands and their receptors holds therapeutic potential for autoimmune diseases and cancer (Callahan et al. 2010; Dinesh et al. 2010; Loos et al. 2010). Most known B7 members contain so-called Ig superfamily (IgSF) domains (Williams and Barclay 1988), with an N-terminal variable (V)-type domain, followed by one constant (C)-type domain and a transmembrane (TM) region. V-type IgSF domains are usually responsible for interacting with their ligands. C-type IgSF domains generally fall into two major categories, C1 and C2, based on their beta strand and loop compositions (Williams and Barclay 1988). The C2-type is more primitive because it is found in certain immune molecules such as NK receptors and FcRs in vertebrates, but it is most commonly found in adhesion molecules in both invertebrates and vertebrates (Du Pasquier et al. 2004). In contrast, the C1-type is found primarily in molecules involved in adaptive immunity, and the C domains found in B7 molecules are mostly of this unusual type. Many B7 family members have been identified from jawed vertebrate (gnathostome) classes (Hansen et al. 2009; Zhang et al. 2009) and phylogenetic analyses have shown their V domains to be related to tapasin (TAPBP), poliovirus receptor (PVR), and other virus receptor V domains (e.g. Cocksackievirus and Adenovirus receptors), and distantly related to butyrophilins (BTN) and myelin oligodendrocyte glycoproteins (MOG) (Fahrer et al. 2001; Henry et al. 1999; Linsley et al. 1994).

Xenopus species shared a common ancestor with humans ~350 million years ago, and these amphibians are “connecting” organisms linking fish to higher vertebrate taxa (Hellsten et al. 2010; Vogel 1999). Compared to the bony fish model, in which the genome has been disrupted presumably due to rapid expansion and contraction of species and a class-specific genome-wide duplication (Postlethwait et al. 2004), the *Xenopus* genome is much more stable, and some genes display unexpected syntenic relationships that are primordial (Ohta et al. 2006). Such genetic stability distinguishes *Xenopus* as a model taxon for comparative analysis. Thus, by examining syntenic relationships in *Xenopus*, in concert with overall comparative analyses in all chordates, we hope to shed light on the natural history of the vertebrate genome, especially regarding immunity.

While examining the *Xenopus* genome for MHC-related genes, we found two polygenic clusters of highly divergent genes that contain both V- and C1-type IgSF domains. Phylogenetic analysis showed that the V-IgSF domain clustered with the known B7 family members and subsequently we found orthologs from other vertebrate species, including human. Indeed, while we were preparing this paper, a human gene was reported as the newest B7 family member, *B7H6* (Brandt et al. 2009). *B7H6* binds to the activating natural killer receptor (NKR), NKp30, through its V domain (Brandt et al. 2009). *B7H6* is upregulated under tumor transformation or stress conditions and NK cells eliminate such cells either directly via cytotoxicity or indirectly by cytokine secretion (Baratin and Vivier 2010). *NKp30* is encoded in the MHC in mammals and possesses a single V-IgSF domain followed by a TM region (Pende et al. 1999). The cytoplasmic region of NKp30 interacts with the ITAM-containing adaptor molecule, CD3 ζ , resulting in activation signals to NK cells (Pende et al. 1999). Interestingly, the NKp30 V domain is a unique type, the so-called “VJ”-domain, which resembles the precursor of antigen receptors that has been found in other genes involved in immunity and cellular interactions (Bartl et al. 1994; Du Pasquier L. 2002; Thompson 1995). Unlike the V domains of the antigen receptors in which diversity is

generated by RAG-mediated somatic rearrangement of 'V' and 'J' segments, the "VJ"-type IgSF is germline-encoded in a single exon. It is believed that a single "VJ"-IgSF exon was split into separate 'V' and 'J' exons by a transposon containing signals recognized by the RAG proteins which gave rise to the somatically rearranging lymphocyte antigen receptor genes (Du Pasquier et al. 2004). Thus, the "VJ"-IgSF is assumed to have predated the emergence of antigen receptors.

In this paper we conducted a genomic analysis of *Xenopus B7H6* and *NKp30* loci. We found a positive correlation between the number of loci for these receptors and ligands throughout vertebrate evolution. We also uncovered another (and probably the last) B7 family member, which we designate B7H7. We further performed a phylogenetic analysis of all known B7 family members in all gnathostomes. Lastly, our analysis provides unequivocal evidence that B7 family members were originally encoded in the MHC.

Materials and Methods

Data-mining the EST and genomic databases

Initially, the new *Xenopus* B7 family genes were found serendipitously with tBLASTn searches using IgSF domains of chicken MHC class II DM. Both the *X. tropicalis* and *X. laevis* EST databases and the *X. tropicalis* genomic assembly version 4.1 at the Department of Energy Joint Genome Institute (JGI: www.genome.psi/jgi.doe.gov/Xenopus) were used. All 'hits' were further BLASTx searched against the eukaryotes protein databases to confirm their novelty. We then searched for all *Xenopus* B7 families by using B7 IgSF domains from other species and confirmed the orthology by BLAST and later phylogenetic trees. When there were *Xenopus* EST sequences available, genomic scaffolds were searched with BLASTn (using *X. tropicalis*) or tBLASTx (using *X. laevis*). Analysis of genomic synteny was obtained from either Gene Page at the NCBI web site or Map Viewer, with further manual annotation when necessary (as well as classical segregation analyses by Southern blotting, see below). *Xenopus* NKp30 genes were identified by using human NKp30 as bait and confirmed as detailed above. To search for *NKp30* in shark databases, we used primarily *Xenopus NKp30* as bait and again confirmed as described above. Sequences were searched in the NCBI databases and 'Elephant Shark Genome Project' web site (<http://esharkgenome.imcb.a-star.edu.sg/>). Transmembrane prediction was made using TMHMM server v2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>).

Phylogenetic tree analysis

The V domains of the deduced amino acid sequences of B7 family members were aligned using ClustalX, and neighbor-joining bootstrapping trees (500 trials) (Saitou and Nei 1987) were made and viewed in the TreeView 1.6.6. program (Page 1996). For phylogenetic tree analysis of NKp30, amino acid alignments of variable domains were initially made in Bioedit with ClustalW employing gap opening penalties of 10 and gap extension penalties of 0.1 for pairwise alignments then 0.2 for multiple alignments and the protein weighting matrix of Gonnet or Blossum (Hall 1999; Tamura et al. 2007). These alignments were then modified by hand and converted back to nucleotides for computation of distances and trees. MEGA was then used to infer the phylogenetic relationships. A neighbor-joining tree was built using distances computed with the maximum composite likelihood method. All three positions of codons were included in the analysis and 525 nucleotide columns were included. The maximum likelihood tree was computed with the method of Tamura and Nei (Tamura et al. 2007) and had a log likelihood of -5921.4319. The maximum likelihood tree employed 203 positions as all columns with gaps or missing data were excluded. The minimum evolution tree resulted in a sum branch length of 9.74464 and used all 525 positions. These NKp30 trees were bootstrapped 1000 times.

Southern blotting

10 µg of genomic DNA isolated from the erythrocytes of an MHC-defined *X. laevis* family (Ohta et al. 1999) and *X. tropicalis* was digested with 80 units of restriction endonucleases overnight, followed by further digestion with 40 units of enzymes overnight. DNA was then separated on 0.8% agarose gels and blotted onto nitrocellulose membranes by capillary transfer (Whatman). Membranes were then hybridized with a radiolabeled cDNA probe under high stringency conditions (50% formamide) and washed under stringent conditions (Bartl et al. 1997). The same membranes were often re-used for different probes after the radioactivity decayed.

Fluorescence *In situ* hybridization on chromosomes

Pre-Sertoli cells as a source of metaphase spreads were obtained from testes of *X. tropicalis* juveniles. Hypotonic shock was done by 40 mM KCl followed by three washes with fixative (methanol : acetic acid, 3:1). Cell suspension was then dropped onto glass slide with subsequent immersion in 50% acetic acid for two seconds. The ~1kb cDNA probes were labeled by Dig-11-dUTP using the Decalabel Labelling Kit (Fermentas). Simultaneous hybridization of 100-150 ng of probe with chromosomes was performed using simple hybridization mixture of 50% formamide in $2 \times$ SSC, in a humidifying chamber overnight at 37°C. After several high stringency washes, slides were blocked and incubated with antidigoxigenin-POD antibody. Signal visualization was done using tyramide coupled with tetramethylrhodamine (TSA-TMR kit-Perkin Elmer) (Krylov et al. 2007). Nuclei and chromosomes were counterstained with DAPI (shown in blue).

Expression

Gene expression was examined by northern blotting using 10-15 µg total RNA from various tissues. We extracted total RNA from various tissues of a Lewis-strain rat, and *X. laevis*. Human total RNA from 18 different tissues was purchased from Invitrogen (Carlsbad, CA). The human and rat probes were made by subcloning the RT-PCR fragment amplified using primers encompassing the two IgSF domains (i.e. inter-exons) based on the GenBank sequence entries from liver, spleen, and testis RNA. The DNA fragments of expected size (~550 bp) were amplified from all the tissues in both human and rat.

Results and Discussion

Identification of *Xenopus B7H6* genes

While searching for *Xenopus DM* genes (an MHC class II-related gene) by BLAST using the immunoglobulin superfamily (IgSF) domain of DM genes from other species (e.g. chicken), we found many sequences containing C1-type IgSF domains in the databases. Subsequently, we performed BLASTp and BLASTx searches for each sequence and discovered that one partial EST (BX732750; *X. tropicalis*) sequence encodes a novel gene. Using this partial sequence, we found two closely related full-length *X. laevis* sequences (AAH84632 and AAH82955) and an AAH84632 ortholog (AAI66200) from the *Xenopus* databases (Figure 1a). The deduced amino acid (aa) sequences of all clones had two extracellular domains, an N-terminal variable (V) and C-terminal C1-type, based on characteristic aa residues in each type of domain (shaded in grey in Figure 1a), followed by a predicted type I transmembrane (TM) region (underlined in Figure 1a). Additionally, there were often long cytoplasmic tails containing 4 to 5 IgSF domains (e.g. AAH84632, AAI66200), (Online Resource 1a), whereas there were no obvious cytoplasmic IgSF domains in AAH82955. BLASTp searches using V-IgSF domains showed that the *Xenopus* clones are most similar to the V domains of some genes of unknown function from various species (see below for further details) followed by the TAP-binding protein-like (TAPBP-R)

with an E-value $\sim e^{-12}$, and the extracellular C1-IgSF was most similar to TAPBP and MHC class II C1-IgSF domains with E-values $\sim e^{-10}$. The presence of C1-IgSF domains that are distantly related to the C1-IgSF domains in various other immune genes (e.g. TCR, Igs, MHC I, MHC II, TAP-binding protein (TAPBP)) seems to be unique to these molecules. Interestingly, some of the cytoplasmic IgSF domains were C1-type and matched to C1-IgSF domains of MHC class I and IgH/L with low but significant similarities ($\sim e^{-7}$ and $\sim e^{-4}$, respectively), while other non-C1-type domains seemed to be degenerate (e.g. Domain-5 of AAH84632 “lost” the N-terminal-half (Online Resource 1a)) and did not significantly match other sequences. Although the cytoplasmic domains may contribute to a unique signaling mechanism, and /or interaction with adaptor molecules that contribute to downstream signaling events, in this report we focus only on the extracellular domains where the receptor-ligand interaction occurs in all B7 family members so far examined.

To examine whether such IgSF molecules were *Xenopus*-specific, we examined the genes with unknown function that matched with the “best hits” (E-values $< e^{-8}$) during the BLAST searches. Accession numbers of these genes were CAD97811 (cDNA from human uterus; also previously named AC406 (Du Pasquier 2000)), XP_001076811 (predicted gene; rat), XP_601599 (predicted gene; cow), XP_542518 (predicted gene; dog), XP_001501683 (predicted gene; horse), and CAH89395 (cDNA from orangutan kidney). Although some sequences were only predicted gene models and may not be full-length (marked as * in Figure 1a), all of them contained the V-C1 type domain architecture (Figure 1b). The C1 domains from these genes show low but significant similarity to MHC class II IgSF domains. As mentioned in Introduction, while we were characterizing these molecules, the human ortholog was reported as the newest B7 molecule, B7H6, a ligand for the NKR NKp30 (Brandt et al. 2009).

Since the membrane distal V domains are involved in receptor-ligand interactions and contain conserved characteristic features of each gene family, we constructed phylogenetic trees using V-IgSF domains. Preliminary phylogenetic tree analysis with various molecules, including antigen receptors, NKR, signal-regulatory protein (SIRPs) and poliovirus receptor (PVRs), revealed clustering of B7H6 to the other B7 family members (data now shown). Thus, we used only the B7 family members to identify the orthologous relationships (Figure 2). *Xenopus* IgSF genes clustered tightly with B7H6 from other species, and based on the BLAST results and phylogenetic analysis, we concluded that the *Xenopus* IgSF gene is a B7H6 ortholog. Note that the cytoplasmic IgSF domains seen in *Xenopus* were not detected in other species (Figure 1b). The deduced aa alignment shows conservation among mammalian species (Figure 1a), and the variability is more apparent in the N-terminal halves of both the V- and C1-IgSF domains. There is a surprisingly high conservation of amino acid residues in the D-loop to E-strand region of the C1-IgSF domain (boxed in Figure 1a) which may be used either for homodimerization or interaction with other molecules on the same cell. The TM regions are well predicted in all three *Xenopus* sequences (underlined in Figure 1a). There is no GAG polyprotein homologous regions in the 3' untranslated region as was found in the original human sequence, thus this viral insertion seems to be primate-specific (Kaifu et al. 2011).

B7H6 expression

A recent report showed that the human *B7H6* was not detected in normal human tissues, but was expressed by human tumor cell serving as a damage-associated molecular pattern to trigger innate immunity (Baratin and Vivier 2010; Brandt et al. 2009). We examine whether this expression pattern was evolutionary conserved. High *B7H6* expression in an unimmunized rat was detected in lymph node, muscle, spleen, and testis, and lower expression was found in mucosal tissues (colon, ileum, jejunum and lung) (Figure 3a). Thus, there is a different expression pattern in human and rat.

Xenopus B7H6 expression was also studied in unimmunized animals. There were at least five distinct transcripts showing tissue-specific expression patterns in blood, colon, intestine, and spleen, and lower amounts in skin and other tissues (Figure 3b). Multiple transcripts were expected since we identified many different-sized EST sequences from various tissues and developmental stages (see below; and Online Resource 2). Thus, the diverse bands likely reflect expression of different genes and/or differentially spliced forms. Since we used a probe derived from the BX732750 sequence that only contains two cytoplasmic IgSF domains (Online Resource 1a), we repeated the northern blotting with the V-IgSF domain of AAH84632, which under low stringency conditions should recognize all frog B7H6 members (Online Resource 1b), and we obtained identical results (data not shown). Collectively, the differential expression pattern in various species implies that B7H6 can acquire new functions in different vertebrate taxa.

Genomic analysis of *Xenopus B7H6* gene

To obtain *B7H6* genomic information, we searched the *X. tropicalis* genomic scaffolds (<http://genome.jgi-psf.org/Xentr4/Xentr4.home.html>) by BLASTn and tBLASTn and found 12-13 predicted gene models in two scaffolds (707 and 1420) (Online Resource 2). Most of them contain a domain structure similar to AAH824632 and AAI66200 (Figure 1a, b; Online Resource 1a) with external V-C1, TM, and cytoplasmic C1-IgSF and “Ig-like” domains. Fifty-seven unique IgSF domains were annotated, each of which being encoded by a single exon. We found that some C1 or “Ig-like” domains are more similar to each other and they were often found in the same order like cassettes in multiple gene models (data not shown); thus, these IgSF domains were likely generated by duplications, further suggesting that the expansion of the *Xenopus B7H6* genes was driven by *cis*-duplications and subsequent exon shuffling.

B7H6 genes from all mammalian species are either single-copy or completely absent (Figures 1, 2, and Table I), and therefore the expansion of *B7H6* genes seems to be species-specific. The orthologous genes were lost in mouse (Brandt et al. 2009), chicken and opossum, and are not present in fish (Table I). The species-specific expression pattern, in combination with expansion of genes in *Xenopus* and loss of the gene in some species, suggest that *B7H6* evolves rapidly and might be co-evolving with its receptor genes. The deduced aa sequences for all V domains contain the conserved cysteine (C), tryptophan (W), but no GXG motif in the G-strand (i.e. not a “VJ”-type IgSF domain). The alignment of the *Xenopus* V-IgSF domains showed broad variability (57-93%) throughout the entire domain (Online Resource 1b).

Identification of a *Xenopus*-specific *B7H6* subfamily

During the searches, we found additional full-length sequences (AAH85212 (*X. laevis*) and AAI61104 (*X. tropicalis*)) with low but significant identity to *Xenopus B7H6*. They contain multiple C1-IgSF domains as well as a distinctive V-IgSF domain (Online Resource 1c). The deduced aa sequences revealed three extracellular IgSF domains (V-Ig-C1), followed by a TM region, and either a long cytoplasmic tail with four “Ig-like” domains (AAH85212) or no domains (AAI61104) (Figure 1b). BLASTp searches revealed that the V-IgSF domains were most similar to the V domains of B7H1 and TAPBP-R (Du Pasquier 2000) with an E-value $\sim e^{-8}$. Like the results with B7H6, extracellular Ig-like and C1-IgSF domains of this related family were most similar to IgL chain and MHC class II (E-value $\sim e^{-5}$), and the four cytoplasmic Ig-like domains of AAH85212 matched weakly to IgH, IgL, and MHC (E-values $\sim e^{-7}$). Exhaustive BLAST searches did not uncover orthologs from other species, suggesting that these genes are *Xenopus*-specific B7 homologues, and thus we designate them as *B7HXen*.

Phylogenetic tree analysis did not show strong orthology to other B7 family members with all low bootstrap values (arrow in Figure 2). Interestingly, B7HXen clustered with the membrane distal V-IgSF domain of hagfish IgSF3 (Haruta et al. 2006). The results were consistent between several sets of data and numerous analyses using different parameters (data not shown), and thus it is possible that B7HXen might be evolutionarily older than most B7 family members. Unfortunately, related genes were not found in the existing databases of lower deuterostomes such as amphioxus, sea squirt, or sea urchin so the hypothesis will be difficult to verify.

We found eleven *X. tropicalis* B7HXen domain-containing scaffolds (scaffolds-1119, 867 etc.) containing a total of 31 predicted gene models (Online Resource 2), of which thirteen V-IgSF domains were found at the N-termini of the gene models. Most gene models had a similar domain structure as AAH85212 (V-Ig-C1-TM) with zero to >10 cytoplasmic Ig domains (Figure 1b, OnlineResource1c, and 2). Like B7H6, the deduced aa sequences for all V domains contain the canonical conserved aa but no G-strand GXG motif. Similar to B7H6, alignment of V-IgSF domains from the gene models showed high variability (Online Resource 1d) (some sequences shared only 36% identity) distributed equally throughout the domain.

NKp30

As mentioned above, human B7H6 binds to the activating NKR, NKp30 (Brandt et al. 2009), and thus we examined the phylogenetic representation of NKp30 in the vertebrate lineage. It has been previously reported that NKp30 genes are not present outside of mammals (Kaifu et al. 2011), and furthermore that NKRs evolve rapidly, making it difficult to find orthologues among divergent taxa (Yoder and Litman 2011). Therefore, we were surprised that our BLAST searches using human NKp30 identified NKp30-like genes in the *Xenopus* databases (Online Resource 3). There were a total of seventeen predicted gene models in six scaffolds in the *X. tropicalis* genome (88, 215, 3675, 3771, 15172, 18746), with twelve models clustered on scaffold-215. There are three gene models that contain two V-IgSF domains in scaffold-215; however, since NKp30 in all other vertebrates has only one V-IgSF domain and all *Xenopus* EST sequences only contain a single domain, most likely the models either identify pseudogenes or assembly artifacts owing to the difficulty in identifying the short leader and transmembrane exons. The number of gene models in *X. tropicalis* (17-20) corresponds well with the number of bands detected by Southern blotting (~20) using a single-exon probe (Figure 4a). The *Xenopus NKp30-like* genes are also quite polymorphic, showing different restriction fragment length polymorphism (RFLP) patterns in the sixteen individuals tested (four are shown in Figure 4a).

Like *B7H6*, many NKp30-like EST sequences were found in databases from various tissue sources (Online Resource 3). Based upon sequence similarity, *X. tropicalis NKp30-like* genes were categorized into four types: 1) BX-type (only one gene model 215-12 matches to EST sequence BX773925); 2) New (or 88-2)-type (gene model 88-2 has no matching EST sequence); 3) ES-type (four gene models (215-2, 88-1, 15172-1, 3771-1) matching EST sequence ES685492 and many other ESTs); and 4) DT-type (eleven gene models (215-1, 3~11, 3675) matching the EST sequences DT447366 and EL837335) (see details in Online Resource 3). The DT-type is most abundant in the genome, but based on the relative levels of EST clones the ES-type may be expressed at the highest levels. The aa sequences of individual clones are >90% identical within each type, but only 40-60% identical between different types (Figure 4b). Although, we do not know the function at this time, the phylogenetic analysis using V-IgSF domains from many different molecules supports the true orthology of *Xenopus* NKp30 to NKp30 of mammalian species (Figure 4c, Online Resources 4). All four types of *Xenopus* NKp30 contain the “VJ” IgSF type ‘GXG’ motif. Among the different types, there may be certain aa changes responsible for interaction with

B7H6 based upon the human data (bold letters in Figure 4b (Li et al. 2011)). Since *Xenopus B7H6* and *NKp30* are both polygenic and contain divergent V-IgSF domains, there may be a preference for binding between particular receptors and ligands, *i.e.* certain NKp30 isoforms may be co-evolving with some *Xenopus B7H6*. Human NKp30 has six splice variants (Kaifu et al. 2011) some of which are differentially expressed in a tissue-specific manner (Delahaye et al. 2011). *Xenopus* EST ES685492 contains one ITIM motif (double underlined in Figure 4b) in the cytoplasmic region and it lacks the arginine (R) in the TM that could interact with ITAM-containing adaptors; thus, as is typical for other vertebrate NKR-but not specifically for NKp30 in other species-*Xenopus* NKp30 differing in their cytoplasmic tails might be either stimulatory or inhibitory.

It was previously recognized that NKp30 is present in rat, cow, and dog (Figure 4b), but that mouse *NKp30* is a pseudogene (Hollyoake et al. 2005;Kaifu et al. 2011). Neither could we identify *NKp30* in opossum, chicken or bony fish. The presence or absence of both *NKp30* and *B7H6* in different species correlate precisely (Table I); this finding, in concert with the expansion of both genes in *Xenopus*, is strongly suggestive of co-evolution of this receptor/ligand pair over hundreds of millions of years.

Mapping of *B7H6*, *B7HXen*, and *NKp30* in the *Xenopus* genome

To further study the evolutionary origin of *B7H6*, we mapped both *B7H6* and *B7HXen* onto *X. tropicalis* chromosomes using the TSA-FISH system (Macha et al. 2003). *B7H6* and *B7HXen* were mapped onto the telomeres of *X. tropicalis* chromosomes 4 and 9, respectively (Figure 5a). Since no other regions hybridized to the probes, all genes were tightly clustered only in these chromosomal regions, suggesting that expansion of the genes was likely the result of the *cis*-duplications. The expansion of genes in the telomeric regions seems to be a common phenomenon since it was observed for other *Xenopus* genes: *e.g.* non-classical *Xenopus* MHC class I genes are found in a large telomeric cluster on the same chromosome as MHC (Courtet et al. 2001;Flajnik et al. 1993).

B7H6—To gain more insight into the genomic synteny of *B7H6* family members, we further analyzed genes neighboring *B7H6* in the genomic scaffolds. Scaffold-707 contains two other genes (*PLEKHA7* and *RPS13*) both of which map in the vicinity of *B7H6* on human chromosome 11p15 (Online Resource 2). The scaffold-1420 contains only one other gene, *SLC6A11* which maps to the telomeric region of human chromosome 3p25.3 (Online Resource 2). We also examined the *X. tropicalis* genetic map (<http://tropmap.biology.uh.edu/>) for synteny, but there were no microsatellites associated with the *B7H6*-containing scaffolds and thus no other associated genes or linkage groups could be established.

B7HXen—Most of the *B7HXen*-containing scaffolds were small and did not contain other genes, and none of the gene models in larger scaffolds (*e.g.* 1119 and 867) revealed clear orthology to any human gene. However, upon examination of the *X. tropicalis* genetic map, we found that the *B7HXen*-containing scaffold-1146 was located near scaffold-673, which contains *beta-2 microglobulin (B2M)* (Stewart et al. 2005). To confirm their close linkage, we performed a segregation analysis using genomic Southern blotting of a *X. laevis* family with 20 siblings. Unlike the true diploid *X. tropicalis*, *X. laevis* is tetraploid (Kobel and L.Du Pasquier 1986); nevertheless, genetic synteny is usually conserved between these species. Consistent with the tetraploidy in *X. laevis*, two independently segregating *B2M* loci are present (groups c and d in Figure 5b) (Stewart et al. 2005), whereas the *B7HXen* seemed to be diploidized (*i.e.* silenced on one chromosome), showing only four allelic segregation patterns derived from one locus (shown as groups 1-4). One set of *B2M* loci (group d) matched perfectly to the *B7HXen* RFLP pattern except for sibling 13 (*B2M RFLP*

'a' to *B7HXen* group 1 and 2; *B2M RFLP* 'b' to *B7HXen* group 3 and 4), confirming the close linkage of *B7HXen* and *B2M* (Figure 5b). *B2M* is linked to the shark MHC, and in conjunction with other phylogenetic evidence the presence of *B2M* in the MHC is primordial (Ohta et al. 2011). Therefore, the linkage of *B7HXen* to *B2M* in *Xenopus* further implies that the precursor of B7 family genes was present in the proto MHC before genomic expansion and extensive reorganization during vertebrate evolution.

NKp30—*NKp30* maps to the MHC in human and other species, but it is not in the *Xenopus* MHC; instead, we found a cluster of genes with a very similar domain structure as *NKp30*, the *Xenopus* MHC-linked Immunoglobulin superfamily Variable genes (*XMIV*), in the same location as human *NKp30* (Ohta et al. 2006). Both *XMIV* and *NKp30* contain “VJ”-type IgSF domain with the ‘GXG’ motif and we predicted that these genes were derived from a common ancestor. However, phylogenetic tree analysis did not reveal a close relationship between *XMIV* and *NKp30* genes (data not shown), suggesting that they may be evolving at different rates. *Xenopus NKp30* maps to linkage group (LG) 7 (LG 4 in the new assignment (Wells et al. 2011)). Scaffold-88 contains three microsatellites, 036C01, 040A12, and 022B03, in the corresponding position 12.56-13.26cM, and is sandwiched between scaffolds-128 and 1. Scaffolds-128 and 1 contain the *adamts18* and *e2f4* genes, respectively (Wells et al. 2011), and map to the long arm of chromosome 4 from centromere to telomere orientation (Khokha et al. 2009; Wells et al. 2011) (Figure 6). Thus, *Xenopus NKp30* is located in the middle of the long arm of chromosome 4. Note that *B7H6* mapped to the telomere of the short arm of chromosome 4 (Figure 6).

In order to further study the evolutionary history of *NKp30* and *XMIV* genes, we examined the synteny around *NKp30* on scaffold-215. *NKp30* genes are clustered at the 5' end of the scaffold, adjacent to beta-1,3-glucuronyltransferase 1 (*B3GAT1*) (human chr 11q25), followed by genes found in the telomeric region of human chromosome 3p21.3 (Online Resource 3). *NKp30* genes are located at the 3'-end of scaffold-88 that contains genes mostly found in human chromosomes 3p21-25 and 22q13 (Online Resource 3) and these regions do not contain MHC paralogous genes and were never recognized as MHC paralogous regions (Flajnik and Kasahara 2010) (Figure 7). All of the other scaffolds were short and only contain a single *NKp30* gene. Inconsistent synteny around *NKp30* genes between *Xenopus* and other species suggests that there has been a genomic modification at and around the locus, perhaps the result of translocation from the other end of the chromosome (Figure 6); this scenario is consistent with the great expansion of genes in this species.

B7 family members contain V-IgSF domains that share a common ancestor with TAPBP (Greenwald et al. 2005), which maps to the MHC in all species so far examined. Based on the similar domain structure and the presence of “VJ”-type IgSF domains in both *XMIV* and *NKp30*, it is reasonable to predict that *XMIV* and *NKp30* share a common ancestor that was linked in the primordial MHC. The close linkage of *B7HXen* and *B2M*, and the mapping of *B7H6* and *NKp30* to the same chromosome, further suggests that the precursor of B7 was also linked to the primordial MHC. Furthermore, the human chromosome region to which *B7H6* maps (11p) has been considered to be an MHC-paralogous region (Du Pasquier 2009; Flajnik and Kasahara 2010) (Figure 7). Thus, we hypothesize that *XMIV* and *NKp30* were generated by a *cis*-duplication from an NKR precursor before the genome-wide duplication in the pre-adaptive immune complex (PIC), after which the MHC paralogous regions experienced differential silencing; this resulted in *XMIV* remaining with the MHC, and the movement of *NKp30/B7H6* to another region. The chromosome containing *NKp30* and *B7H6* was further modified, presumably by translocation, leaving *B7H6* at the telomere and *NKp30* near the center. Since we detected no gene remnants of *NKp30* or *B7H6* near the large clusters, the major gene expansions of both *NKp30* and *B7H6* probably took place

after the separation of these loci (Figure 6). Interestingly, *B3GAT1*, which is immediately adjacent to *NKp30* genes in *Xenopus* scaffold-215, is found on the telomeric end of the human chromosome 11q, whereas *B7H6* is found in the center of the chromosome (Figure 7); this suggests that the genome modification likely took place after the emergence of the amphibian lineage.

The phylogenetic tree was not informative regarding the affiliation of *B7HXen* to other B7 family members, but its linkage to *B2M* might be more revealing. The *B2M* translocation from its original location in the MHC (discussed in (Ohta et al. 2011)), the close linkage of *B2M* and *B7HXen* in *Xenopus*, and the wide distribution of most B7 families in MHC paralogous regions (see below, Figure 7) support the hypothesis that B7 precursor ('B7p' in Figure 6) was encoded in the proto MHC. Our hypothesis is also consistent with the previous observation of a connection of B7 to the MHC (Henry et al. 1999). In addition, *B7H3* is on the same chromosome as *B2M* in many mammalian species, but *Xenopus B7H3* is not closely linked to *B2M* (Figure 6). We suggest that following the co-translocation of *B2M* and a B7 precursor early in vertebrate history, a secondary translocation occurred in *Xenopus* splitting *B7H3* and *B7HXen* onto two chromosomes. *B7HXen* subsequently expanded greatly, and thus it is difficult to ascribe its orthology to other B7 members; in contrast, the translocated *B7H3* remained a single-copy gene and evolved at a slower rate (Figure 6; genomic data not shown).

Identification of Shark *B7H6* and *NKp30* genes

Since some B7 orthologues were found in the shark databases, and based on the presence of *NKp30* genes in the *Xenopus* genome, we further examined the cartilaginous fish databases. We used the human B7H6 V domain to search the elephant shark (E-shark) genomic scaffolds (Venkatesh et al. 2007) and found 5 sequences hitting with E-values under e^{-5} : four sequences contain only V domains (AAVX01166257.1, AAVX01015110.1, AAVX01166967.1, AAVX01076204.1) and one sequence contains a C domain (AAVX01231033.1). Phylogenetic tree analysis showed clustering with other B7H6, confirming the orthology (Figure 2). All V sequences showed diversity distributed throughout the domain (Online Resource 5), and thus shark B7H6 seems to belong to a multigene family. We also identified a cartilaginous fish (*Squalus acanthias*, spiny dogfish) EST sequence (ES788778) that showed high similarity to other species' NKp30 (E-values $<e^{-9}$). Using dogfish NKp30 as bait, we identified five *NKp30* genes from the E-shark genome (Online Resource 5; the V-IgSF domain of one representative sequence is shown in Figure 4b). Phylogenetic tree analysis showed clear clustering of the shark NKp30 to other species' NKp30 (Figure 4c). Like NKp30 in most vertebrates, a "VJ"-IgSF domain is present, but the C-terminal Tm domains were not identified by the gene models, perhaps due to the small exon size or the short scaffolds (Figure 4b). Such conservation of *NKp30* during vertebrate evolution was unexpected since *NKp30* is not detected in any bony fish species and rapid expansion and contraction of NKR families in a species-specific manner is a general rule (Yoder and Litman 2011), especially the activating NKR. The presence of both *B7H6* and *NKp30* genes in cartilaginous fish species demonstrates that this system is evolutionary old, in fact the most ancient among NKR families. Thus, we suggest that the selection pressure on *NKp30* is exceptional and quite different from most other NKR, perhaps a result of its co-evolution with the self ligand, *B7H6*. The expansion of both *B7H6* and *NKp30* in *Xenopus* and shark, and the absence of both genes in other species, correlate well with a co-evolving receptor/ligand relationship, perhaps dating back to when their precursor genes were encoded in the proto MHC.

Our phylogenetic tree analysis further showed an unexpected sister grouping of NKp30 with *NKp44*. *NKp44* is expressed on the surface of NK cells and contains a V-IgSF domain (but is not of the 'VJ'-type) (Cantoni et al. 1999) and non-mammalian orthologs were found only

in some bony fish species including carp (Stet et al. 2005), trout (Kock and Fischer 2008; Ostergaard et al. 2009), and zebrafish (Ostergaard et al. 2009). *NKp44* maps in the TREM cluster region outside the MHC in the human and mouse genomes (Allcock et al. 2003). The clustering of the NKp30 and NKp44 V domains in phylogenetic analysis, their association with the MHC, and their presence in lower vertebrates identify these immune molecules as ancient and unique members of the NKR family.

The new B7 family member: HHLA2

During the search for other B7 family members from the *Xenopus* databases, we uncovered an additional new sequence (AAH96014) that was most similar to B7H4. *Xenopus* AAH96014 matched with significant similarities (E-value $\sim e^{-20}$) to human HERV-H LTR-associating 2 (HHLA2) having an unknown function (Mager et al. 1999). *HHLA2* was so-named because a sequence in the 3'-untranslated region has polyadenylated long terminal repeats (LTR), but no further characterization has been done. Our phylogenetic tree analysis clearly shows that *HHLA2* is a new B7 member, and we re-designate it *B7H7*. Our database searches revealed orthologs in hoofed mammals (e.g. horse, pig), chicken, opossum, salmon, and shark, but only pseudogenes in rodents (mouse and rat) (Table 1). The aa sequence alignment of HHLA2 showed a similar domain structures in all species, with conservation of the canonical aa residues for IgSF domains (Figure 8a). HHLA2 contains three extracellular IgSF domains; one V-type, followed by a degenerate Ig-like domain that still was identified as a C1-type, one V/C2-IgSF domain, and a TM region (Figure 8b). BLAST searches revealed that the V-IgSF domains from all species were most similar to either B7H3 or B7H4, depending on the species. The similarity of V-IgSF domain to B7H3 and B7H4 by BLAST was confirmed by the phylogenetic tree analysis, clearly demonstrating HHLA2's close relationship to other B7 family genes (Figure 2). The other two IgSF domains matched to B7H4 ($\sim e^{-4}$), TCR/IgL ($\sim e^{-6}$), BTN ($\sim e^{-7}$), MHC class II ($\sim e^{-5}$), and CD4 ($\sim e^{-7}$), along with other non-immune genes (ROR2 and NRCAM). Human *B7H7* is expressed in gut, kidney, and lung, and it was speculated that *B7H7* is expressed in epithelial cells. However, *B7H7* was also detected in lymphocytes in the original analysis (Mager et al. 1999). *Xenopus HHLA2* is expressed highly in intestine, kidney, liver, and stomach and weakly in pancreas, colon and skin (Figure 8c), suggesting that it may be involved in mucosal immunity. We identified two bands on the northern blot, and upon reexamination of the databases, we found another EST sequence (AAI67398) for *Xenopus HHLA2*. Both EST clones are splice variants, and AAI67398 lacks the TM region (Figure 8a). Both clones are full-length (2,300bp (AAH96014) and 5365bp (AAI67398)), and it seems that the rapidly migrating band with the lower expression in northern blotting correlates with the TM-containing splice variant whereas the higher expression band correlates with the secreted variant. *Xenopus HHLA2* is found in genomic scaffold-23 and genetic synteny between human, chicken, and *Xenopus* is well conserved (Figure 8d). Human *HHLA2* is found on chromosome 3q13.33 near B7.1 and B7.2, which is a region rich in genes containing IgSF domains and one of the regions assigned to be a MHC paralog (Du Pasquier et al. 2004; Zucchetti et al. 2009) (Figure 7).

Phylogenetic tree analysis of B7 family

Building on a previous study that centered on teleost fish (Hansen et al. 2009), we conducted updated phylogenetic tree analysis including B7H6 and B7H7. We also searched for B7 family members from shark databases and added them in the analyses. Our phylogenetic tree analysis showed that B7 family members were categorized into six major groups (Figure 2) and we discuss each in turn.

B7.1/B7.2 group—The CD28 ligands, B7.1 (Selvakumar et al. 1992) and B7.2 (Reeves et al. 1997), were the founding B7 molecules and are the best characterized B7 family

members. Both *B7.1* and *B7.2* are present in mammals and their redundant roles in costimulation and their chromosomal positions had implied that they were recent duplicates. However, the presence of *B7.2* in *Xenopus* and a *B7.1* ortholog in all teleost fish (called B7R; (Hansen et al. 2009; Zhang et al. 2009)), as well as co-stimulation by trout B7R (Zhang et al. 2009), suggest that they are ancient. Interestingly, B7R formed a unique cluster within the B7 group in our phylogenetic tree analysis. Bony fish experienced a species-specific duplication of the genome (Meyer and Van de 2005) that generated separate clades for some genes and a general disruption of their genome (Postlethwait 2007). We could not detect opossum or *Xenopus B7.1*, or both shark *B7.1* and *B7.2* (Table I).

B7H2 group—Phylogenetic analysis showed that B7H2 is most similar to B7.1 and B7.2. *B7H2* was not found in teleost fish and genomic synteny is not well conserved between *Xenopus* and other species. However, B7H2 maps to human chromosomal region 21q22.3, which is likely to be paralogous to chromosomal region 3q13-22 where B7.1 and B7.2 are found (Du Pasquier et al. 2004) (Figure 7). Thus, B7H2 and B7.1/B7.2 may be derived from an immediate common ancestor. Another area that is paralogous to these regions is chromosome 1q23-25, an MHC-paralog (Figure 7). Therefore the origin of these two B7 groups also has an MHC connection.

B7H1/B7DC group—*B7H1* and *B7DC* are adjacent to each other in all species that have both genes (e.g. human and *Xenopus*), and thus they were predicted to have arisen by a *cis*-duplication. B7H1 and B7DC genes were reported previously in the chicken genome (Hansen et al. 2009), however our analysis showed that the both clustered with mammalian B7H1, and thus are clearly duplicated B7H1 genes (Figure 2). Two *cis*-duplicated genes were identified from the *Xenopus* genome and each picked up B7H1 and B7DC by Blast. Teleost fish have only one gene, which matches equally to B7H1 and B7DC in the Blast analysis (e.g. fugu CAF93166 in Figure 2), suggesting the pre-duplicated state. There are two E-shark sequences (AAVX01047137.1 and AAVX01029406.1: similar to each other ~76% at the aa level) that are most similar to B7H1. Phylogenetic analysis including all vertebrate species revealed that both *Xenopus* B7H1 and B7DC, the two shark B7H1, and *fugu* B7H1/DC clustered together, but did not group with mammalian B7H1 and B7DC. Thus, *Xenopus* B7DC may not be orthologous to mammalian B7DC. Taken together, the timing of the separation of B7H1 and B7DC may be relatively recent.

B7H3/B7H4/B7H7 group—Here we demonstrate that *HHLA2* (Mager et al. 1999) is a new B7 family member, *B7H7*. *B7H7* is found in an IgSF-rich genomic region on human chromosome 3q (Du Pasquier et al. 2004) along with *B7.1* and *B7.2*, and it was likely generated by an ancient duplication from the common ancestor of *B7.1* and *B7.2* (Figure 7). A previous report noticed a weak similarity of B7H7 (HHLA2) to the B7 family (Fahrer et al. 2001), but we have found that B7H7 clearly is related to B7H3 and B7H4 in the phylogenetic analysis (figure 2). A *B7H3* ortholog was identified from skate (Hansen et al. 2009) (with E-values $\sim e^{-76}$ and $\sim e^{-45}$ to human entire molecule and V-IgSF domains, respectively) and E-shark (AAVX01064573.1). Similarly, we found B7H4 (AAVX01079670.1) and B7H7 (AAVX01019552.1) from the E-shark genome, suggesting that this group of B7 family member is evolutionarily old (Figure 2). The functions of these newer B7 family members are not fully known, but their phylogenetic conservation suggests important roles in immune regulation. There are several other B7-related genes in the shark and bony fish databases that clustered separately but belong to this group. However, we could not detect clear orthology to specific B7 family members (data not shown) and thus we did not include them in our analyses.

B7H5 group—B7H5, also known as VISTA or Gi24, was recently discovered as an inhibitory B7 molecule expressed by APCs and T cells (Wang et al. 2011) and it is conserved among different species. We found dogfish shark EST sequence (e.g. DV500095) and E-shark sequence (AAVX01038049.1) with low E-values $\sim e^{-27}$ and $\sim e^{-18}$ to *Xenopus* and human, respectively.

V domain of B7H5 matches to the V domains of T cell receptors in Blastp analysis and B7H5 consistently clustered distantly from other B7 members with long branch-lengths when the tree was made un-rooted, and thus it was used to root the trees. B7H5 maps to human chromosomal region 10q22.1, which has not been assigned to any MHC paralogous region (Figure 7). Furthermore, B7H5 has a domain structure V-TM and lacks the C domain, which is distinctly different from other B7 family members (Table I). All of the data to date suggest that B7H5 may be derived from a different precursor than all other B7 family members.

B7H6 group—The presence of shark B7H6, the clustering of B7H6 with the TAPBP/TAPBPR group, and the presence of the unique C1 domain in B7H6 suggest that B7H6 arose early in vertebrate evolution and was lost in bony fish. In addition, the discovery of *Xenopus* and shark B7H6 has further highlighted the plastic features of B7 family genes: 1) *The number of IgSF domains*: Most B7 family members have two-domain structures (Table I), but human B7H3 has duplicated its V- and C-IgSF domains, generating a longer form with 4 IgSF domains (Steinberger et al. 2004; Sun et al. 2002). Other members have three-domain structures (e.g. B7HXen and B7H7). Thus, it is clear that B7 family members need not conform to the two-domain paradigm. The cytoplasmic IgSF domains in *Xenopus B7H6* and *B7HXen* are so far unique in vertebrates, and duplication of these domains has also been detected. These *Xenopus* cytoplasmic IgSF domains, which presumably play some role in signaling events, are puzzling and await biochemical analysis. Since receptor/ligand interactions occur via the extracellular domains in the B7 family, our assignment of orthology has been established with these domains. 2) *Gene duplication*: As we found for the *Xenopus B7H6* and *B7HXen* genes, B7 genes are often duplicated. B7H1 and B7DC were generated by tandem duplications, and *Xenopus B7H4* was duplicated in tandem, resulting in two genes with $\sim 48\%$ similarities at the aa level (Table I). Other species-specific B7 genes may have emerged by duplication as well. For example, one B7 gene model (Xt1303) from *X. tropicalis* scaffold-1303 is adjacent to the orthologous gene next to *B7.1* (Figure 2). Xt1303 has a predicted three-domain structure and BLAST searches matched closest to B7H1 with $\sim e^{-10}$ for all three domains, but it does not cluster in the tree with any particular family member. 3) *B7 precursor candidates*: BLAST searches and phylogenetic tree analysis further revealed a connection between *Xenopus*-specific *B7HXen* and hagfish *IGSF3* genes (Haruta et al. 2006) (Figure 2), suggesting that hagfish *IGSF3* may be related to the common precursor of all B7 family members; if true, this would be the first B7 family member detected in taxa more primitive than gnathostomes. In the sea urchin, there are more than 50 genes encoding at least three types TM proteins with V- and C1-like Ig domains having low but intriguing sequence similarity to Ig/TCR/MHC, as well as a number of Ig-domain-containing proteins with similarity to the B7 co-stimulatory molecules (Sodergren et al. 2006). Our database searches did not detect any of these genes, but clearly they are worth examining in future expression/functional studies.

Genomic history of the B7 family

The chromosomal distribution of the B7 family members into paralogous regions in the human genome is consistent with the ancient genome-wide duplications (WGD) early in the history of vertebrates (Flajnik and Kasahara 2010; Ohno S 1970) (right and middle columns in Figure 7). All B7 members are distributed into the MHC paralogous regions;

chromosomes 1 (*B7H4*) and 9 (*B7H1*, *B7DC*) as “major” paralogs, and 11 (*B7H6*) and 15 (*B7H3*) as “minor” paralogs (Flajnik and Kasahara 2010). Additionally, from the study on IgSF genes in the *Ciona* genome, chromosomes 3 (*B7.1*, *B7.2*, *B7H7*) and 21 (*B7H2*) have been identified for their connection to the MHC paralogous regions on chromosomes 1 (Du Pasquier et al. 2004) and 19q (Zucchetti et al. 2009) (left column in Figure 7), suggesting that these chromosomal regions were associated with the proto-MHC and translocated out before WGD. Since B7 family members are found in MHC paralogous regions, the B7 precursor must have arisen before the emergence of the vertebrates (Proto-MHC in Figure 7). *B7H4* genes may have been differentially silenced in human and *Xenopus*, resulting in two paralogous regions; the *Xenopus B7H4* gene is linked to genes found on human chromosome 3q13 in the vicinity of *B7.1*, while human *B7H4* is on chromosome 1 (Figure 7). Alternatively, *Xenopus B7H4* could have been translocated from the equivalent location to human chromosome 1. Furthermore, the similarity of the B7 V domains to V domains found in MHC-linked BTN and MOG is also consistent with the B7 family’s original presence in the proto MHC (Henry et al. 1999).

Our linkage analysis in *Xenopus* showed close linkage of *B7HXen* to *B2M*, *XMIV* mapping to MHC (Ohta et al. 2006), and *NKp30* mapping to the same chromosome as *B7H6*. In conjunction with the mapping of *NKp30* to the MHC in other species, all of these data suggest that the B7 precursor was likely encoded in the proto MHC along with *B2M* (shown from shark linkage study (Ohta et al. 2011)), NKRs (shown from linkage analysis in chicken (Kaufman et al. 1999) and opossum (Belov et al. 2007)), and perhaps antigen receptors (Flajnik and Kasahara 2010). Antigen receptors play the central role in adaptive immunity and their evolutionary origin has been speculated to be associated with the MHC. One theory for the origin of antigen receptors is that the C1-type IgSF domains became associated with “VJ” domains encoded within the same genetic region (Bartl et al. 1994; Du Pasquier L. 2002; Du Pasquier et al. 2004). The presence of NKRs of the germline-joined “VJ”-type in the MHC (i.e. *NKp30* and *XMIV*), and the fragments of synteny with C1-IgSF-containing B7 homologs (i.e. *B7H6* and *B7HXen*) that are associated with the MHC certainly are consistent with this scenario (Figure 6).

Conclusion—In this paper we have shown: 1) the general scheme for evolution of the B7 family; 2) a new member of the B7 family, perhaps the last one to be discovered that is evolutionarily conserved; and 3) the evolutionary history of the *B7H6* and *NKp30* families, and we speculate that they persist over evolutionary time in a co-evolving receptor/ligand unit. Furthermore, we perpetuate the theme that B7 family members and NKR (genes mostly in LRC and NKC) were encoded in the proto MHC. Various B7 family members have been identified from gnathostomes (Hansen et al. 2009; Zhang et al. 2009), but so far not from jawless vertebrates or invertebrates, strongly suggesting that the immune regulation associated with most B7 family members arose with the emergence of the Ig/TCR/MHC-based adaptive immune system. Indeed, the emergence of a large number of both B7 and TNF superfamily members (Collette et al. 2003), all involved in regulating adaptive immune responses, occurred rapidly in concert with the advent of adaptive immunity. When observing this system from an evolutionary perspective it is fascinating how complicated networks of regulation must have fallen into place rapidly to harness the nascent Ig/TCR/MHC-based immune system. This also suggests that while the jawless fish have adaptive immunity complete with T and B cells, the gnathostome system must be regulated in a much more complicated fashion since the majority of costimulatory molecules, as well as most members of the cytokine and chemokine families arose after the split between agnathans and gnathostomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Du Pasquier for critical reading and his insightful suggestions. We also thank Caitlin Doremus for critical reading. This project is supported by National Institutes of Health Grant RR006603 and AI02877 to YO, KO, AK and MFF, and Ministry of Education, Youth and Sports of Czech Republic MSM0021620858 and Grant Agency of Czech Republic P502/11/P522 for VK and TT.

Reference List

- Allcock RJ, Barrow AD, Forbes S, Beck S, Trowsdale J. The human TREM gene cluster at 6p21.1 encodes both activating and inhibitory single IgV domain receptors and includes NKp44. *Eur. J. Immunol.* 2003; 33:567–577. [PubMed: 12645956]
- Baratin M, Vivier E. B7-H6 : a novel alert signal for NK cells. *Med. Sci. (Paris)*. 2010; 26:119–120. [PubMed: 20188033]
- 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. [PubMed: 9550410]
- Bartl S, Baltimore D, Weissman IL. Molecular evolution of the vertebrate immune system. *Proc. Natl. Acad. Sci. U. S. A.* 1994; 91:10769–10770. [PubMed: 7971960]
- Belov K, Sanderson CE, Deakin JE, Wong ES, Assange D, McColl KA, Gout A, de BB, Barrow AD, Speed TP, Trowsdale J, Papenfuss AT. Characterization of the opossum immune genome provides insights into the evolution of the mammalian immune system. *Genome Res.* 2007; 17:982–991. [PubMed: 17495011]
- Brandt CS, Baratin M, Yi EC, Kennedy J, Gao Z, Fox B, Haldeman B, Ostrander CD, Kaifu T, Chabannon C, Moretta A, West R, Xu W, Vivier E, Levin SD. The B7 family member B7-H6 is a tumor cell ligand for the activating natural killer cell receptor NKp30 in humans. *J. Exp. Med.* 2009; 206:1495–1503. [PubMed: 19528259]
- Callahan MK, Wolchok JD, Allison JP. Anti-CTLA-4 antibody therapy: immune monitoring during clinical development of a novel immunotherapy. *Semin. Oncol.* 2010; 37:473–484. [PubMed: 21074063]
- Cantoni C, Bottino C, Vitale M, 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; 189:787–796. [PubMed: 10049942]
- Collette Y, Gilles A, Pontarotti P, Olive D. A co-evolution perspective of the TNFSF and TNFRSF families in the immune system. *Trends Immunol.* 2003; 24:387–394. [PubMed: 12860530]
- Collins M, Ling V, Carreno BM. The B7 family of immune-regulatory ligands. *Genome Biol.* 2005; 6:223. [PubMed: 15960813]
- Courtet M, Flajnik M, Du Pasquier L. Major histocompatibility complex and immunoglobulin loci visualized by in situ hybridization on *Xenopus* chromosomes. *Dev. Comp Immunol.* 2001; 25:149–157. [PubMed: 11113284]
- Delahaye NF, Rusakiewicz S, Martins I, Menard C, Roux S, Lyonnet L, Paul P, Sarabi M, Chaput N, Semeraro M, Minard-Colin V, Poirier-Colame V, Chaba K, Flament C, Baud V, Authier H, Kerdine-Romer S, Pallardy M, Cremer I, Peaudecerf L, Rocha B, Valteau-Couanet D, Gutierrez JC, Nunes JA, Commo F, Bonvalot S, Ibrahim N, Terrier P, Opolon P, Bottino C, Moretta A, Tavernier J, Rihet P, Coindre JM, Blay JY, Isambert N, Emile JF, Vivier E, Lecesne A, Kroemer G, Zitvogel L. Alternatively spliced NKp30 isoforms affect the prognosis of gastrointestinal stromal tumors. *Nat. Med.* 2011; 17:700–707. [PubMed: 21552268]
- Dinesh RK, Hahn BH, Singh RP. PD-1, gender, and autoimmunity. *Autoimmun. Rev.* 2010; 9:583–587. [PubMed: 20433954]

- Du Pasquier L. Several MHC-linked Ig superfamily genes have features of ancestral antigen-specific receptor genes. *Curr Top Microbiol Immunol.* 2002; 266:57–71. [PubMed: 12014203]
- Du Pasquier L. MHC Evolution, Structure and Function. Springer-Verlag; Tokyo: 2000. Relationships among the genes encoding MHC molecules and the specific antigen receptors; p. 53–65.
- Du Pasquier L. Fish ‘n’ TRIMs. *J. Biol.* 2009; 8:50. [PubMed: 19519941]
- Du Pasquier L, Zucchetti I, De Santis R. Immunoglobulin superfamily receptors in protochordates: before RAG time. *Immunol. Rev.* 2004; 198:233–248. [PubMed: 15199966]
- Fahrer AM, Bazan JF, Papathanasiou P, Nelms KA, Goodnow CC. A genomic view of immunology. *Nature.* 2001; 409:836–838. [PubMed: 11237003]
- Flajnik MF, Kasahara M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat. Rev. Genet.* 2010; 11:47–59. [PubMed: 19997068]
- Flajnik MF, Kasahara M, Shum BP, Salter-Cid L, Taylor E, Du Pasquier L. A novel type of class I gene organization in vertebrates: a large family of non-MHC-linked class I genes is expressed at the RNA level in the amphibian *Xenopus*. *EMBO J.* 1993; 12:4385–4396. [PubMed: 8223448]
- Greenwald RJ, Freeman GJ, Sharpe AH. The B7 family revisited. *Annu. Rev. Immunol.* 2005; 23:515–548. [PubMed: 15771580]
- Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series.* 1999; 41:95–98. Ref Type: Journal (Full).
- Hansen JD, Du Pasquier L, Lefranc MP, Lopez V, Benmansour A, Boudinot P. The B7 family of immunoregulatory receptors: a comparative and evolutionary perspective. *Mol. Immunol.* 2009; 46:457–472. [PubMed: 19081138]
- Haruta C, Suzuki T, Kasahara M. Variable domains in hagfish: NICIR is a polymorphic multigene family expressed preferentially in leukocytes and is related to lamprey TCR-like. *Immunogenetics.* 2006; 58:216–225. [PubMed: 16541254]
- Hellsten U, Harland RM, Gilchrist MJ, Hendrix D, Jurka J, Kapitonov V, Ovcharenko I, Putnam NH, Shu S, Taher L, Blitz IL, Blumberg B, Dichmann DS, Dubchak I, Amaya E, Detter JC, Fletcher R, Gerhard DS, Goodstein D, Graves T, Grigoriev IV, Grimwood J, Kawashima T, Lindquist E, Lucas SM, Mead PE, Mitros T, Ogino H, Ohta Y, Poliakov AV, Pollet N, Robert J, Salamov A, Sater AK, Schmutz J, Terry A, Vize PD, Warren WC, Wells D, Wills A, Wilson RK, Zimmerman LB, Zorn AM, Grainger R, Grammer T, Khokha MK, Richardson PM, Rokhsar DS. The genome of the Western clawed frog *Xenopus tropicalis*. *Science.* 2010; 328:633–636. [PubMed: 20431018]
- Henry J, Miller MM, Pontarotti P. Structure and evolution of the extended B7 family. *Immunol. Today.* 1999; 20:285–288. [PubMed: 10354554]
- Hollyoake M, Campbell RD, Aguado B. NKp30 (NCR3) is a pseudogene in 12 inbred and wild mouse strains, but an expressed gene in *Mus caroli*. *Mol. Biol. Evol.* 2005; 22:1661–1672. [PubMed: 15872155]
- Joyce MG, Tran P, Zhuravleva MA, Jaw J, Colonna M, Sun PD. Crystal structure of human natural cytotoxicity receptor NKp30 and identification of its ligand binding site. *Proc. Natl. Acad. Sci. U. S. A.* 2011; 108:6223–6228. [PubMed: 21444796]
- Kaifu T, Escaliere B, Gastinel LN, Vivier E, Baratin M. B7-H6/NKp30 interaction: a mechanism of alerting NK cells against tumors. *Cell Mol. Life Sci.* 2011; 68:3531–3539. [PubMed: 21877119]
- Kaufman J, Milne S, Gobel TW, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S. The chicken B locus is a minimal essential major histocompatibility complex. *Nature.* 1999; 401:923–925. [PubMed: 10553909]
- Keir ME, Butte MJ, Freeman GJ, Sharpe AH. PD-1 and its ligands in tolerance and immunity. *Annu. Rev. Immunol.* 2008; 26:677–704. [PubMed: 18173375]
- Khokha MK, Krylov V, Reilly MJ, Gall JG, Bhattacharya D, Cheung CY, Kaufman S, Lam DK, Macha J, Ngo C, Prakash N, Schmidt P, Tlapakova T, Trivedi T, Tumova L, bu-Day A, Geach T, Vendrell E, Ironfield H, Sinzelle L, Sater AK, Wells DE, Harland RM, Zimmerman LB. Rapid gynogenetic mapping of *Xenopus tropicalis* mutations to chromosomes. *Dev. Dyn.* 2009; 238:1398–1406. [PubMed: 19441086]
- Kobel HR, Du Pasquier L. Genetics of polyploid *Xenopus*. *Trends Genet.* 1986; 12:310–314. Ref Type: Journal (Full).

- Kock H, Fischer U. A novel immunoglobulin-like transcript from rainbow trout with two Ig-like domains and two isoforms. *Mol. Immunol.* 2008; 45:1612–1622. [PubMed: 18035417]
- Krylov V, Tlapakova T, Macha J. Localization of the single copy gene *Mdh2* on *Xenopus tropicalis* chromosomes by FISH-TSA. *Cytogenet. Genome Res.* 2007; 116:110–112. [PubMed: 17268187]
- Li Y, Wang Q, Mariuzza RA. Structure of the human activating natural cytotoxicity receptor NKp30 bound to its tumor cell ligand B7-H6. *J. Exp. Med.* 2011; 208:703–714. [PubMed: 21422170]
- Linsley PS, Peach R, Gladstone P, Bajorath J. Extending the B7 (CD80) gene family. *Protein Sci.* 1994; 3:1341–1343. [PubMed: 7527261]
- Loos M, Hedderich DM, Friess H, Kleeff J. B7-h3 and its role in antitumor immunity. *Clin. Dev. Immunol.* 2010; 2010:683875. [PubMed: 21127709]
- Macha J, Tlapakova T, Krylov V, Kopsky V. Xstir polymorphism and absence of sex linkage in *Xenopus laevis* ME2 gene. *Folia Biol. (Praha).* 2003; 49:115–117. [PubMed: 12859020]
- Mager DL, Hunter DG, Schertzer M, Freeman JD. Endogenous retroviruses provide the primary polyadenylation signal for two new human genes (HHLA2 and HHLA3). *Genomics.* 1999; 59:255–263. [PubMed: 10444326]
- Meyer A, Van de PY. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays.* 2005; 27:937–945. [PubMed: 16108068]
- Ohno, S. *Evolution by gene duplication.* Springer-Verlag; New York: 1970.
- Ohta Y, Goetz W, Hossain MZ, Nonaka M, Flajnik MF. Ancestral organization of the MHC revealed in the amphibian *Xenopus*. *J. Immunol.* 2006; 176:3674–3685. [PubMed: 16517736]
- Ohta Y, Powis SJ, Coadwell WJ, Haliniewski DE, Liu Y, Li H, Flajnik MF. Identification and genetic mapping of *Xenopus* TAP2 genes. *Immunogenetics.* 1999; 49:171–182. [PubMed: 9914331]
- Ohta Y, Shiina T, Lohr RL, Hosomichi K, Pollin TI, Heist EJ, Suzuki S, Inoko H, Flajnik MF. Primordial Linkage of {beta}2-Microglobulin to the MHC. *J. Immunol.* 2011; 186:3563–3571. [PubMed: 21321107]
- Ostergaard AE, Martin SA, Wang T, Stet RJ, Secombes CJ. Rainbow trout (*Oncorhynchus mykiss*) possess multiple novel immunoglobulin-like transcripts containing either an ITAM or ITIMs. *Dev. Comp Immunol.* 2009; 33:525–532. [PubMed: 19013192]
- Page RD. TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 1996; 12:357–358. [PubMed: 8902363]
- Pende D, Parolini S, Pessino A, Sivori S, Augugliaro R, Morelli L, Marcenaro E, Accame L, Malaspina A, Biassoni R, Bottino C, Moretta L, Moretta A. Identification and molecular characterization of NKp30, a novel triggering receptor involved in natural cytotoxicity mediated by human natural killer cells. *J. Exp. Med.* 1999; 190:1505–1516. [PubMed: 10562324]
- Postlethwait J, Amores A, Cresko W, Singer A, Yan YL. Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet.* 2004; 20:481–490. [PubMed: 15363902]
- Postlethwait JH. The zebrafish genome in context: ohnologs gone missing. *J. Exp. Zool. B Mol. Dev. Evol.* 2007; 308:563–577.
- Reeves RH, Patch D, Sharpe AH, Borriello F, Freeman GJ, Edelhoff S, Disteche C. The costimulatory genes *Cd80* and *Cd86* are linked on mouse chromosome 16 and human chromosome 3. *Mamm. Genome.* 1997; 8:581–582. [PubMed: 9250865]
- Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 1987; 4:406–425. [PubMed: 3447015]
- Selvakumar A, Mohanraj BK, Eddy RL, Shows TB, White PC, Dupont B. Genomic organization and chromosomal location of the human gene encoding the B lymphocyte activation antigen B7. *Immunogenetics.* 1992; 36:175–181. [PubMed: 1377173]
- Sodergren E, Weinstock GM, Davidson EH, Cameron RA, Gibbs RA, Angerer RC, Angerer LM, Arnone MI, Burgess DR, Burke RD, Coffman JA, Dean M, Elphick MR, Etensohn CA, Foltz KR, Hamdoun A, Hynes RO, Klein WH, Marzluff W, McClay DR, Morris RL, Mushegian A, Rast JP, Smith LC, Thorndyke MC, Vacquier VD, Wessel GM, Wray G, Zhang L, Elsiek CG, Ermolaeva O, Hlavina W, Hofmann G, Kitts P, Landrum MJ, Mackey AJ, Maglott D, Panopoulou G, Poustka AJ, Pruitt K, Sapojnikov V, Song X, Souvorov A, Solovyev V, Wei Z, Whittaker CA, Worley K, Durbin KJ, Shen Y, Fedrigo O, Garfield D, Haygood R, Primus A, Satija R, Severson T, Gonzalez-Garay ML, Jackson AR, Milosavljevic A, Tong M, Killian CE, Livingston BT, Wilt FH,

Adams N, Belle R, Carbonneau S, Cheung R, Cormier P, Cosson B, Croce J, Fernandez-Guerra A, Geneviere AM, Goel M, Kelkar H, Morales J, Mulner-Lorillon O, Robertson AJ, Goldstone JV, Cole B, Epel D, Gold B, Hahn ME, Howard-Ashby M, Scally M, Stegeman JJ, Allgood EL, Cool J, Judkins KM, McCafferty SS, Musante AM, Obar RA, Rawson AP, Rossetti BJ, Gibbons IR, Hoffman MP, Leone A, Istrail S, Materna SC, Samanta MP, Stolc V, Tongprasit W, Tu Q, Bergeron KF, Brandhorst BP, Whittle J, Berney K, Bottjer DJ, Calestani C, Peterson K, Chow E, Yuan QA, Elhaik E, Graur D, Reese JT, Bosdet I, Heesun S, Marra MA, Schein J, Anderson MK, Brockton V, Buckley KM, Cohen AH, Fugmann SD, Hibino T, Loza-Coll M, Majeske AJ, Messier C, Nair SV, Pancer Z, Terwilliger DP, Agca C, Arboleda E, Chen N, Churcher AM, Hallbook F, Humphrey GW, Idris MM, Kiyama T, Liang S, Mellott D, Mu X, Murray G, Olinski RP, Raible F, Rowe M, Taylor JS, Tessmar-Raible K, Wang D, Wilson KH, Yaguchi S, Gaasterland T, Galindo BE, Gunaratne HJ, Juliano C, Kinukawa M, Moy GW, Neill AT, Nomura M, Raisch M, Reade A, Roux MM, Song JL, Su YH, Townley IK, Voronina E, Wong JL, Amore G, Branno M, Brown ER, Cavalieri V, Duboc V, Duloquin L, Flytzanis C, Gache C, Lapraz F, Lepage T, Locascio A, Martinez P, Matassi G, Matranga V, Range R, Rizzo F, Rottinger E, Beane W, Bradham C, Byrum C, Glenn T, Hussain S, Manning G, Miranda E, Thomason R, Walton K, Wikramanayke A, Wu SY, Xu R, Brown CT, Chen L, Gray RF, Lee PY, Nam J, Oliveri P, Smith J, Muzny D, Bell S, Chacko J, Cree A, Curry S, Davis C, Dinh H, Dugan-Rocha S, Fowler J, Gill R, Hamilton C, Hernandez J, Hines S, Hume J, Jackson L, Jolivet A, Kovar C, Lee S, Lewis L, Miner G, Morgan M, Nazareth LV, Okwuonu G, Parker D, Pu LL, Thorn R, Wright R. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science*. 2006; 314:941–952. [PubMed: 17095691]

- Steinberger P, Majdic O, Derdak SV, Pfistershammer K, Kirchberger S, Klauser C, Zlabinger G, Pickl WF, Stockl J, Knapp W. Molecular characterization of human 4Ig-B7-H3, a member of the B7 family with four Ig-like domains. *J. Immunol.* 2004; 172:2352–2359. [PubMed: 14764704]
- Stet RJ, Hermsen T, Westphal AH, Jukes J, Engelsma M, Lidy Verburg-van Kemenade BM, Dortmans J, Aveiro J, Savelkoul HF. Novel immunoglobulin-like transcripts in teleost fish encode polymorphic receptors with cytoplasmic ITAM or ITIM and a new structural Ig domain similar to the natural cytotoxicity receptor NKp44. *Immunogenetics.* 2005; 57:77–89. [PubMed: 15702329]
- Stewart R, Ohta Y, Minter RR, Gibbons T, Horton TL, Ritchie P, Horton JD, Flajnik MF, Watson MD. Cloning and characterization of *Xenopus* beta2-microglobulin. *Dev. Comp Immunol.* 2005; 29:723–732. [PubMed: 15854684]
- Sun M, Richards S, Prasad DV, Mai XM, Rudensky A, Dong C. Characterization of mouse and human B7-H3 genes. *J. Immunol.* 2002; 168:6294–6297. [PubMed: 12055244]
- Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 2007; 24:1596–1599. [PubMed: 17488738]
- Thompson CB. New insights into V(D)J recombination and its role in the evolution of the immune system. *Immunity.* 1995; 3:531–539. [PubMed: 7584143]
- Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP, Johnson J, Dandona N, Viswanathan LD, Tay A, Venter JC, Strausberg RL, Brenner S. Survey sequencing and comparative analysis of the elephant shark (*Callorhynchus milii*) genome. *PLoS. Biol.* 2007; 5:e101. [PubMed: 17407382]
- Vogel G. Frog is a prince of a new model organism. *Science.* 1999; 285:25. [PubMed: 10428694]
- Wang L, Rubinstein R, Lines JL, Wasiuk A, Ahonen C, Guo Y, Lu LF, Gondek D, Wang Y, Fava RA, Fiser A, Almo S, Noelle RJ. VISTA, a novel mouse Ig superfamily ligand that negatively regulates T cell responses. *J. Exp. Med.* 2011; 208:577–592. [PubMed: 21383057]
- Wells DE, Gutierrez L, Xu Z, Krylov V, Macha J, Blankenburg KP, Hitchens M, Bellot LJ, Spivey M, Stemple DL, Kowis A, Ye Y, Pasternak S, Owen J, Tran T, Slavikova R, Tumova L, Tlapakova T, Seifertova E, Scherer SE, Sater AK. A genetic map of *Xenopus tropicalis*. *Dev. Biol.* 2011; 354:1–8. [PubMed: 21458440]
- Williams AF, Barclay AN. The immunoglobulin superfamily--domains for cell surface recognition. *Annu. Rev. Immunol.* 1988; 6:381–405. [PubMed: 3289571]
- Yoder JA, Litman GW. The phylogenetic origins of natural killer receptors and recognition: relationships, possibilities, and realities. *Immunogenetics.* 2011; 63:123–141. [PubMed: 21191578]

- Zhang YA, Hikima J, Li J, LaPatra SE, Luo YP, Sunyer JO. Conservation of structural and functional features in a primordial CD80/86 molecule from rainbow trout (*Oncorhynchus mykiss*), a primitive teleost fish. *J. Immunol.* 2009; 183:83–96. [PubMed: 19535623]
- Zou W, Chen L. Inhibitory B7-family molecules in the tumour microenvironment. *Nat. Rev. Immunol.* 2008; 8:467–477. [PubMed: 18500231]
- Zucchetti I, De SR, Grusea S, Pontarotti P, Du Pasquier L. Origin and evolution of the vertebrate leukocyte receptors: the lesson from tunicates. *Immunogenetics.* 2009; 61:463–481. [PubMed: 19404636]



Figure 1.

a) Alignment of B7H6 from various vertebrate species. Gaps and identical aa residues are indicated with dashes and dots, respectively. Canonical aa residues characteristic of IgSF domains are shaded in grey, potential glycosylation sites are marked with dotted underlines, and predicted TM regions are underlined. IgSF strand designation is shown above sequences and potential glycosylation sites are marked with dotted underlines. Only external IgSF and predicted TM (when available) domains are shown; the cytoplasmic regions including the internal Ig domains in *Xenopus* sequences are shown in Online Resource 1. Bold letters indicate the interaction sites with NKp30 based on the crystal structure (Li et al. 2011). * indicates predicted genes. Conserved aa residues in the D-loop to E-strand region of the C1 domain are boxed. Accession numbers are listed in Figure 2 legend. This figure includes sequences from representative vertebrates in different classes. b) Domain structures of B7H6 and B7HXen. Human B7H6 represents all species except *Xenopus*. Number of cytoplasmic Ig domains varies depending on the sequences: n=5 for AAH84632 and AAI66200; n=0 for AAH82955 and AAI61104; n=4 for AAH85212; various numbers were predicted in gene models from the *X. tropicalis* genome (details in Online resources 1a, 1b, and 2).

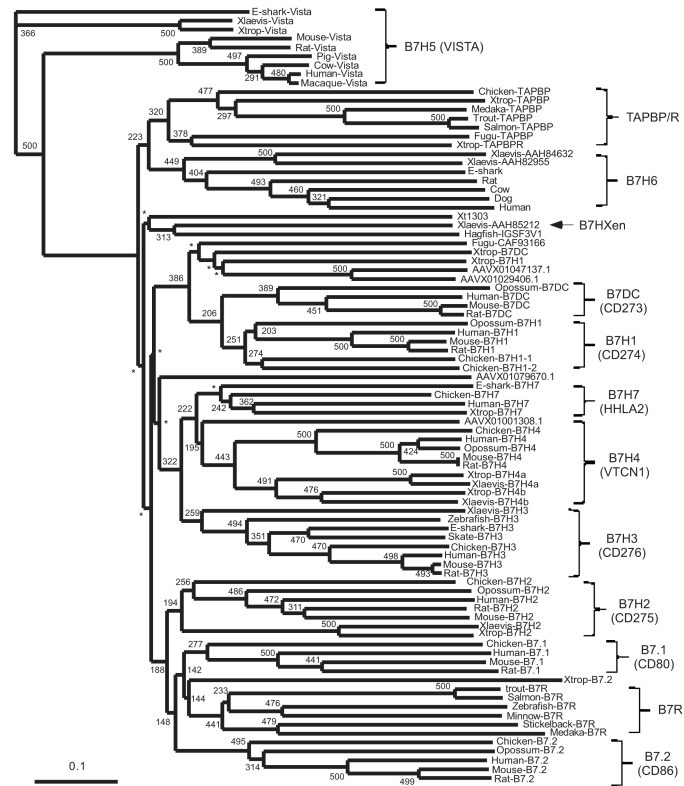


Figure 2. Phylogenetic tree analysis of all members in the B7 family. Deduced aa sequences of V domains are used for this analysis. The bootstrapping tree was constructed using the Neighbor-Joining method with 500 runs (Saitou and Nei 1987). Bootstrapping values are shown at the branch, * indicates less than 100 values. Our analysis consistently showed VISTA (gi24; B7H5) (Wang et al. 2011) to be the most divergent member of B7 family, thus we used it as outgroup. The scale correlates with the divergence time. Accession numbers used for this analysis are: **B7.1**- human (NP_005182), mouse (CAJ18316). Rat (AAC02262), chicken (CAJ18316); **B7.2**- human (CAG46642), mouse (P42082), rat (NP_064466), opossum (XP_001371519), chicken (XP_001371519), *X. tropicalis* (protein ID: 510857 at JGI genome site); **B7R**- trout (ACH58053), zebrafish (CN017839), minnow (DT293786), stickleback (DN656970), salmon (DW580717), medaka (AM306315); **B7H1**- human (NP_054862), mouse (EDL41681), rat (XP_001079572), opossum (XP_001371389), chicken (XP_424811, CF251191), fugu (CAF93166), *X. tropicalis* (protein ID: 16544 at JGI genome site); B7DC- human (AAP13471), mouse (Q9WUL5), rat (EDM13096), opossum (XP_001371408), fugu (CAF93166), *X. tropicalis* (Protein ID:162957 at JGI genome site), E-shark (AAVX14147467.1, AAVX01029406.1); **B7H2**- human (BAA31628), mouse (Q9JHJ8), rat (XP_574731), opossum (XP_001377700), chicken (CAA70058), *X. laevis* (translated from BX848980), *X. tropicalis* (Protein ID: 170158 at JGI genome site); **B7H3**- human (AAK15438), mouse (EDL25954), rat (Q7TPB4), opossum (XP_413702), chicken (Q68EV1), *X. laevis* (XP_700973), zebrafish (XP_001341518), skate (translated from DT378804), E-shark (AAVX01064573.1); **B7H4**- human (AAP37283), mouse (AAP37284), rat (AAH95842), opossum (XP_001365506), chicken (XP_416546), *X. tropicalis* (a: CAJ83577; b: protein ID:396923 at JGI genome site), *X. laevis* (a: AAH41253; b: AAH44000), E-shark (AAVX01001308.1); **B7H5**- human (AAH20568), Rhesus Monkey

(NP_001177807), mouse (Q9D659), rat (AAH98723), pig (XP_001928572), cow (AAI23905), *X. tropicalis* (AAH88869), *X. laevis* (AAH73319), E-shark (AAVX01038049.1); **B7H6**- human (CAH18335), cow (NP_001193721), rat (XP_001076811), dog (XP_542518), *X. tropicalis* (AAI66200), *X. laevis* (AAH84632), E-shark (AAVX01166257.1); **B7HXen**- *X. tropicalis* (AAI61104), *X. laevis* (AAH85212), **B7H7**- human (Q9UM44), *X. tropicalis* (AAH96014), *X. laevis* (ACH85300), E-shark (AAVX01019552.1); **TAPBP**- chicken (CAA06327), *X. tropicalis* (XP_002938728), medaka (BAB83851), trout (AAZ66041), salmon (ABO13869), fugu (CAF97838); **TAPBPR**- *X. tropicalis* (XP_002941962); Hagfish IGSF3 (BAE93396); *X. tropicalis* unknown B7 family in the scaffold-1303 (Protein ID: 461476). Opossum B7.1 and chicken B7H7 were retrieved from the genomic sequences, NW_001581960 and NW_001471429, respectively. BLAST searches with the V-IgSF of B7H5 (aliases; Gi24, VISTA, C10orf54) (Wang et al. 2011) matched to B7H4 with an E-value $\sim e^{-10}$, but consistently separated from other members of the B7 family with long branch lengths and therefore was used to root the tree. When orthology was not strong, E-shark sequences are shown with accession numbers starting with AAVX. All E-shark sequences are deduced from the genomic scaffolds.

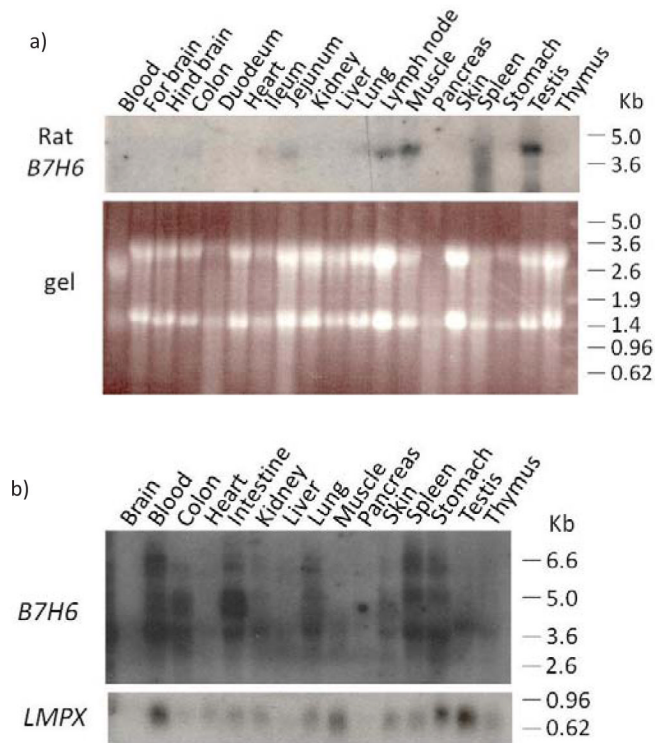


Figure 3.

Tissue expression pattern of rat (a) and *Xenopus* (b) *B7H6*. Northern blotting was done using total RNA from various tissues from a *Lewis* rat and adult *X. laevis*. Ubiquitously expressed *LMPX* was used as a control for *Xenopus*, and the gel image is shown for rat. Size markers are indicated at right. Extracellular IgSF domains were used as probes.

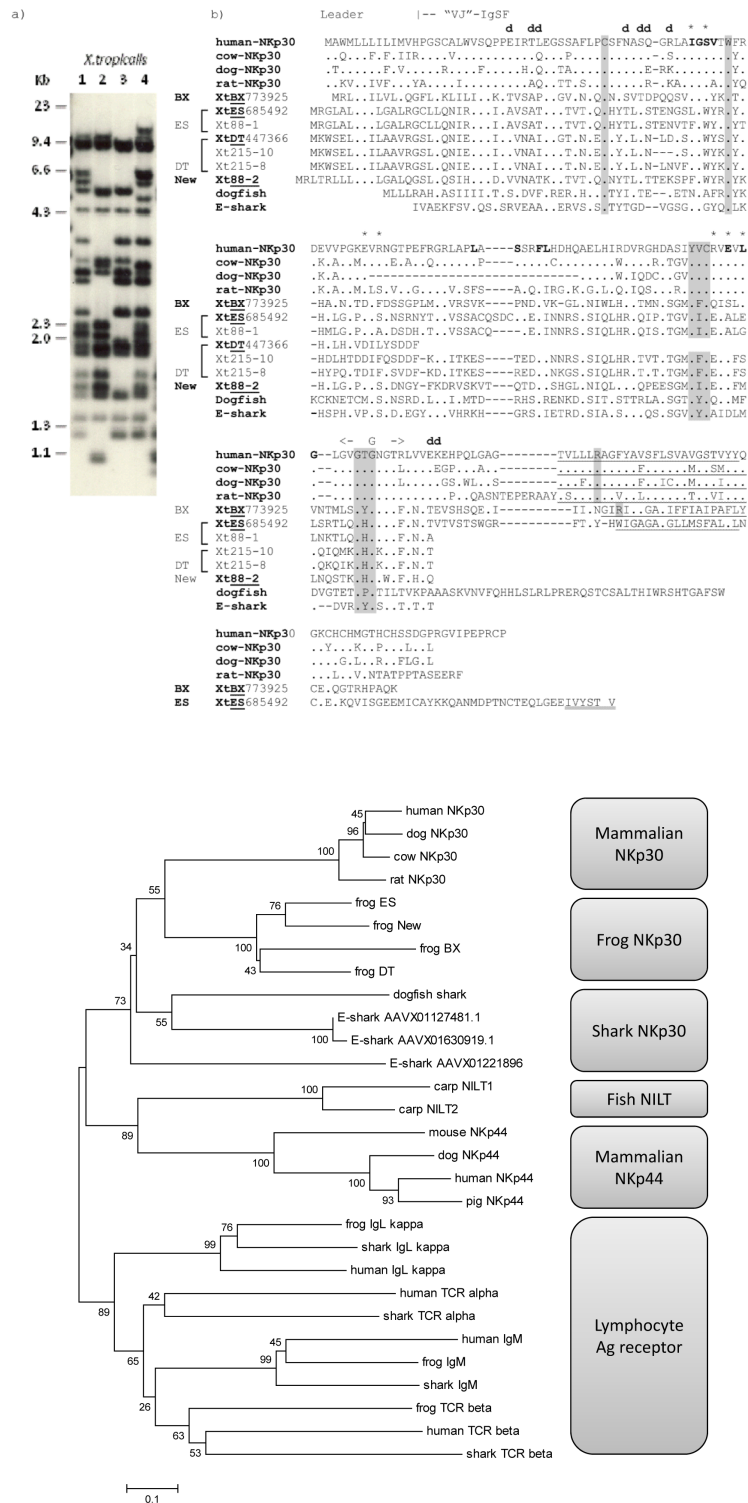


Figure 4. Presence of NKp30 in lower vertebrates. a) *Xenopus NKp30* is polygenic and polymorphic. Southern blotting with a single V-IgSF exon probe shows many genes in the *X. tropicalis* genome. Four different unrelated individual *X. tropicalis* are shown, revealing RFLP-based

polymorphism. Genomic DNA was digested with *Hind*III. Size markers are shown at left. b) Deduced aa alignment of NKp30 from different vertebrate species. Canonical aa residues characteristics for IgSF domains and the GXG motif are shaded in grey and predicted TM regions are underlined. The potential dimerization interface, and residues that disrupt interaction with B7H6 when mutated, are noted as 'd' and * above the sequence, respectively (Joyce et al. 2011). *Xenopus* sequences retrieved from the genomic scaffolds are terminated after the V-IgSF exon, thus the TM and cytoplasmic regions are not known. Sequences belonging to the four types shown as BX, ES, DT, and New (88-2) on the left side of the alignment and representative sequences from gene models are shown under corresponding EST sequences and bundled with parentheses. *Xenopus* EST sequences are indicated as accession numbers and gene models are shown in Online Resource 3. A potential ITIM motif in XtES685492 is indicated by double underline and positively charged aa residues in TM regions are shaded in grey. Bold letters indicate the interaction sites with B7H6 based on the crystal structure (Li et al. 2011). Accession numbers used for this analysis are: NP_001138939 (human), XP_848928.1 (dog), NP_001035614.1 (cow), NP_861543.1 (rat), ES788778 (spiny dogshark; *Squalus acanthias*), and AAVX01127481 (elephant (E-) shark; *Callorhynchus milii*). c) Neighbor joining phylogenetic tree analysis of NKp30. Nucleotide sequences of V domains are used for this analysis. Alternative methods yielding similar results are shown in Online Resource 4. Additional sequences encoding variable domains of lymphocyte antigen receptors were included as an outgroup: IgL κ from shark (L16765.1), frog (DT059572.1), and human (HE584554.1), TCR α from shark (FJ513688.1), and human (AY232282.1), TCR β from shark (FJ513753.1), frog (NM_001091843.1), and human (L34734.1), and IgM from shark (M92851.1), frog (X14919.1), and human (M87789.1).

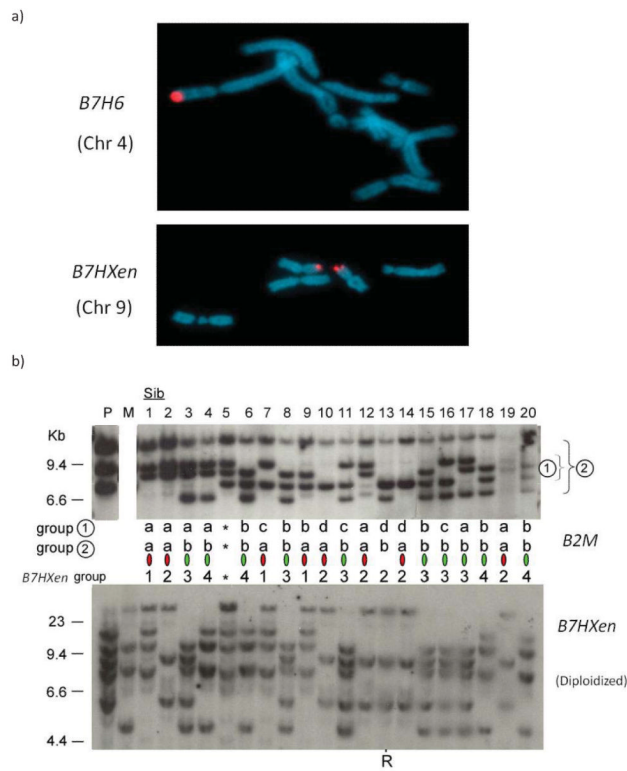


Figure 5. Genetic mapping of *Xenopus B7H6* family members. a) Chromosomal mapping of *B7H6* and *B7HXen* in *X. tropicalis* using the TSA-FISH system. The *Xenopus* chromosome number was assigned following Wells et al (Wells et al. 2011). The coding region probes are shown in red; DAPI counterstaining is shown in blue. b) Linkage of *B7HXen* to the *B2M*. Southern blotting was done using a previously typed *X. laevis* family with 20 siblings along with parents (Paternal and Maternal) (Ohta et al. 1999). Two *B2M* loci are indicated as groups ① and ②, while 4 segregating patterns with *B7HXen* are indicated as numbers 1-4. Matched groups are connected with colored ovals (red or green). Sibling 5 is a natural triploid. Size marker positions are marked at left.

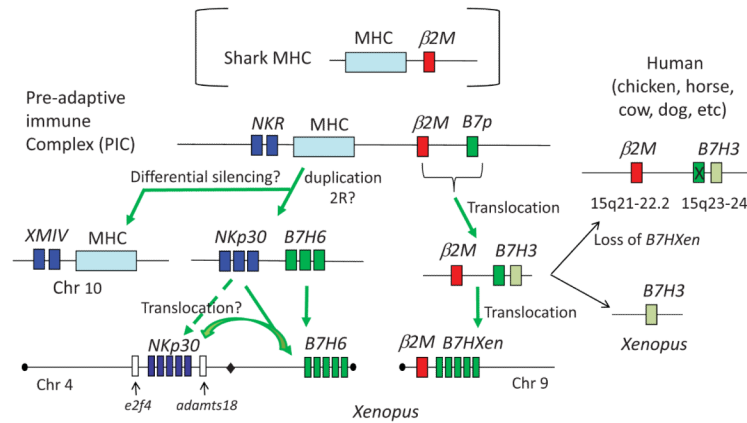


Figure 6.

Model for emergence of the two sets of NKR in *Xenopus* and B7 evolution from the MHC precursor. We hypothesize that the MHC precursor contained precursors of NKR, MHC class I and II, including *beta-2 microglobulin* (*B2M*), and *TAPBP/B7* (denoted as *B7p*) genes. Via two rounds of duplication (2R), four MHC paralogous regions were formed followed by further genetic modification resulting in more paralogous fragments (also see Figure 7). In *Xenopus*, two sets of MHC/NK regions can be followed by differential silencing of the loci, resulting with *XMIV* and *MHC* on chr 10 and *NKp30* and *B7H6* on chr 4. Chromosome 4 may have been further modified by a translocation of *NKp30*, followed by the expansion of *NKp30* and *B7H6* genes. The B7 precursor might have also been translocated along with *B2M*, generating *B7HXen* next to *B2M* on chromosome 9. The synteny of NKR was observed in many vertebrate genomes ((Belov et al. 2007; Kaufman et al. 1999)

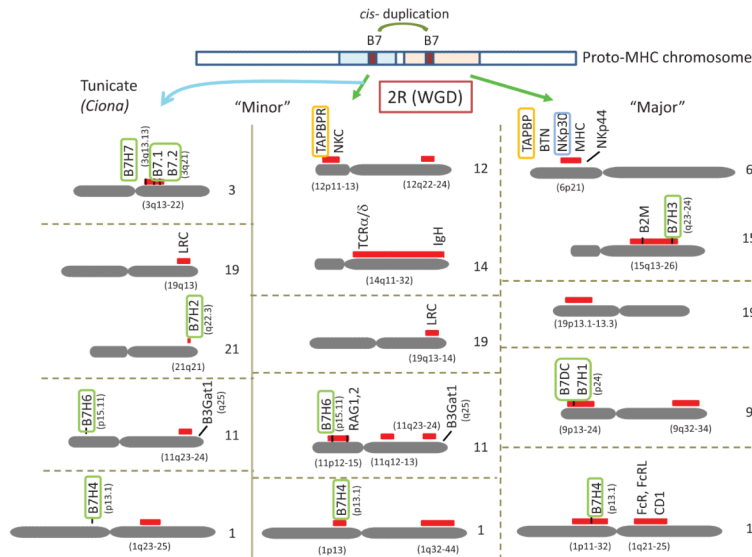


Figure 7. Most B7 family members are distributed on the MHC paralogous regions inferred from the human and *Ciona* genomes. The four sets of “Major” and “Minor” paralogous regions were identified based on the distribution of homologous, duplicated genes related to the MHC and neurotrophins (Flajnik and Kasahara 2010) (right and middle columns). The additional “Minor” paralogues were identified from comparative genomic analysis of IgSF and CTX/JAM family members in the tunicate *Ciona* (a lower deuterostome whose ancestor arose before the emergence of the vertebrates and the 2R whole genome duplication (WGD)) and human genomes (Zucchetti et al. 2009) (far left column). There are other chromosomes that contain paralogous regions, but we only show chromosomes relevant to the B7 family. Paralogous human chromosome numbers are indicated on the right side of the chromosomes. Red bars indicate the paralogous regions among the four sets of paralogous chromosomes. All B7 family members except B7H5 are found in the canonical paralogous regions and they are boxed in green. Since B7 family members are found in both “Major” and “Minor” paralogous sets, we predict that the B7 precursor was duplicated in the proto-MHC and then distributed to both paralogous regions as a result of WGD. Since V domains of TAPBP and B7 are similar, they may share the same common ancestor. TAPBP maps to MHC. TAPBP and related genes are boxed in orange.

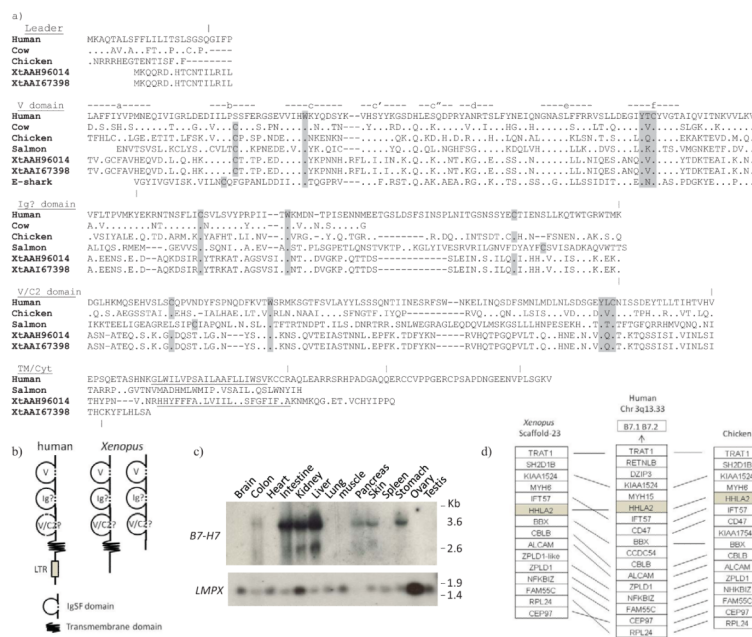


Figure 8.
 a) Deduced amino acid alignment of B7H7 from different species. There is no conservation in the third Ig domain among different species. Canonical aa residues characteristic of IgSF domains are shaded in grey, predicted TM regions are underlined, and IgSF strand designation is shown above sequences. Cow and chicken sequences were only partially retrievable. Exon boundaries are indicated above and below with vertical bars for human and *Xenopus*, respectively. The E-shark sequence was deduced from genomic sequence of AAVX01019552.1 which contains the first two IgSF domains; only V domain is included in this figure. b) Domain structures of B7H7 in human and *Xenopus*. The middle Ig domains have degenerated from the canonical IgSF structure. c) Expression pattern of B7H7 in *Xenopus*. Northern blotting was performed to examine the expression patterns. Ubiquitously expressed LMPX was used as a control and size markers are indicated at right. d) Genomic synteny of B7H7 among *Xenopus*, chicken (chromosome 1), and human. Gene symbols were assigned following the HUGO Gene Nomenclature Committee (HGNC).

Table I

Presence or absence of B7 family genes in different vertebrates.

	Human	Mouse	Rat	Opossum	Chicken	<i>Xenopus</i>	Bony Fish	Shark	Domain architecture
B7.1				---		---	B7R	---?	VC-TM
B7.2							---	---?	VC-TM
B7H1									VC-TM
B7DC					---			---?	VC-TM
B7H2							---	?	VC-TM
B7H3									VC-TM VCVC-TM
B7H4									VjC-GPI
B7H5									V-TM
B7H6		---		---?	---?		---?		VC-TM
B7H7 (HHLA2)		Pseudo	Pseudo						VlIg-TM
NKp30		pseudo		---?	---?		---?		Vj-TM

: single copy; : 2 copies; , : multigene family; --- absence of gene.

Vj: V-IgSF domain with presence of GXG motif in G strand. Ig: IgSF domain.