Published in final edited form as:
Immunogenetics. 2014 June ; 66(6): 411–426. doi:10.1007/s00251-014-0774-5.

Unusual evolutionary conservation and further species-specific adaptations of a large family of Nonclassical MHC class Ib genes across different degrees of genome ploidy in the amphibian subfamily *Xenopodinae*

Eva-Stina Edholm[*], **Ana Goyos**[†], **Joseph Taran**[*], **Francisco De Jesús Andino**[*], **Yuko Ohta**[#], and **Jacques Robert**[*]

[*]Department of Microbiology and Immunology, University of Rochester Medical Center, Rochester, NY 14642

[†]Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305-5126

[#]Department of Microbiology and Immunology, University of Maryland at Baltimore

Abstract

Nonclassical MHC class Ib (class Ib) genes are a family of highly diverse and rapidly evolving genes wherein gene numbers, organization and expression markedly differ even among closely related species rendering class Ib phylogeny difficult to establish. Whereas among mammals there are few unambiguous class Ib gene orthologs, different amphibian species belonging to the anuran subfamily *Xenopodinae* exhibit an unusually high degree of conservation among multiple class Ib gene lineages. Comparative genomic analysis of class Ib gene loci of two divergent (~65 million years) *Xenopodinae* subfamily members *X. laevis* (allotetraploid) and *X. tropicalis* (diploid) shows that both species possess a large cluster of class Ib genes denoted as *Xenopus/Silurana* nonclassical (*XNC/SNC*). Our study reveals two distinct phylogenetic patterns among these genes: some gene lineages display a high degree of flexibility, as demonstrated by species-specific expansion and contractions, whereas other class Ib gene lineages have been maintained as monogenic subfamilies with very few changes in their nucleotide sequence across divergent species. In this second category, we further investigated the *XNC/SNC10* gene lineage that in *X. laevis* is required for the development of a distinct semi-invariant T cell population. We report compelling evidence of the remarkable high degree of conservation of this gene lineage that is present in all 12 species of the *Xenopodinae* examined, including species with different degrees of ploidy ranging from 2, 4, 8 to 12N. This suggests that the critical role of *XNC10* during early T cell development is conserved in amphibians.

Keywords

Amphibians; *Xenopus*; XNC/SNC; class I-like; MHC evolution; XNC10

Communicating Author: Dr. Jacques Robert, Department of Microbiology and Immunology, University of Rochester Medical Center, Rochester, NY 14642; Phone (585) 275-1722; FAX (585) 473-9573; Jacques_Robert@urmc.rochester.edu.

Introduction

Major histocompatibility complex (MHC) class I genes constitute key components of the vertebrate immune system, encoding molecules that play integral roles in both adaptive and innate immune responses. MHC class I genes are further subdivided into classical MHC class Ia (class Ia) and non-classical MHC class Ib (class Ib) genes, which are structurally similar but functionally disparate. Class Ia genes have been described in all groups of jawed vertebrates (Gnathostomes) and encode highly polymorphic molecules essential for $\alpha\beta CD8^+$ T cell differentiation and function (Kasahara et al. 1995; Flajnik and Kasahara 2001). In addition, Gnathostomes have variable numbers of heterogeneous, typically non-polymorphic class Ib genes involved in a variety of functions including both immune and non-immune related roles (Adams and Luoma 2013; Rodgers and Cook 2005). While class Ia genes play a central role in the adaptive immune response, class Ib molecules are generally considered as regulators of innate immune responses, particularly as NK and innate-like T cell receptor ligands (Adams and Luoma 2013). In particular, recent studies in mammals have elucidated the roles of the class Ib molecules CD1d and MR1 in the differentiation of distinct semi-invariant T cell subsets, which are involved in regulating the immune responses to a variety of different microbes (Le Bourhis et al. 2010; Matsuda and Gapin 2005; Skold and Behar 2003; Treiner et al. 2003) and reviewed in (Kronenberg and Gapin 2002; Le Bourhis et al. 2011). However, despite these recent advances the functional roles of many class Ib molecules are still not fully elucidated.

Among challenges in studying the functional roles of class Ib molecules is the rapid evolution undergone by these genes, culminating in extensive variation in their gene numbers and genomic organization even among closely related species (Nei et al. 1997; Nei and Rooney 2005; Piontkivska and Nei 2003)(Adams and Parham 2001). Furthermore, distinct class Ib genes appear to have been subjected to variable selective pressures resulting in multiple species-specific specializations including expansions and/or contractions of individual gene families (Adams and Luoma 2013; Adams and Parham 2001; Flajnik and Kasahara 2001). Combined these factors contribute to the difficulty in establishing evolutionary relationships and class Ib phylogeny across different species. As such, although comparative analyses have revealed functional similarities between certain class Ib molecules of evolutionary divergent species, few unambiguous class Ib orthologs or even homologs have been described.

Homologs of CD1, one of the most conserved class Ib gene families described to date, are present in both mammals and birds (Baker and Miller 2007; Miller et al. 2005) albeit, depending on the species, both the presence of different isoforms and the total number of CD1 genes in the genome varies (Dascher 2007). CD1 proteins bind and present lipid antigens to both conventional and unconventional T cells providing a way for the immune system to recognize and respond to both self and non-self (exogenous as well as endogenous) lipid and glycolipid antigens. Similarly, homologs of the MHC-1-like related protein 1 (MR1) gene have been described in a number of different mammalian species including placentals and marsupials (Tsukamoto et al. 2013). Monomorphic MHC class I-like molecule structurally similar to MR1 termed YF1*7.1 has also been described in chickens (Hee et al. 2010; Kjer-Nielsen et al. 2012). In mouse, MR1 binds and presents

microbial vitamin B metabolites (Kjer-Nielsen et al. 2012), thus representing an additional nonclassical dependent recognition for a specific set of antigens. Despite the presence of multiple class Ib genes in ectothermic vertebrates, to date no CD1 or MR1 homolog have been identified in these species. Moreover, although amphibians as well as cartilaginous and bony fish posses a number of distinct and divergent class Ib genes very little is known about the biological relevance and phylogenetic relationships of these ectothermic vertebrate class Ib genes.

Amphibians represent a key phylogenetic position connecting mammals with vertebrates of more ancient origin such as bony and cartilaginous fish, (reviewed in (Robert and Ohta 2009)). In addition, amphibians belonging to the anuran subfamily *Xenopodinae* is one of few extant vertebrate groups containing natural polyploid species ranging from diploid (2N) to dodecaploid (12N) (Evans 2008). Representative species from two separate genera belonging to the *Xenopodinae* subfamily *X. laevis* and *X. (Silurana) tropicalis* (Bewick et al. 2012) both posses a large number of class Ib genes (Flajnik et al. 1993; Goyos et al. 2011). These class Ib genes, like their mammalian counterparts, are heterogeneous, less polymorphic, and less ubiquitously expressed compared to class Ia. However, in contrast to most mammalian class Ib genes, we have observed an unusually high degree of conservation of distinct class Ib gene subfamilies between *X. laevis* and *X. tropicalis* (Goyos et al. 2011). Notably, this high degree of conservation is apparent despite the fact that these species are estimated to have diverged from the last common ancestor more then 65 million years ago, roughly corresponding to the split between rodent and primate ancestors (Evans 2008; Evans et al. 2004). Class Ib genes isolated from *X. laevis* and *X. tropicalis* have been grouped into multiple subfamilies based on sequence similarities in the α1 and α2 domains (Flajnik et al. 1993; Goyos et al. 2011). The nomenclature of these amphibian class Ib genes was based on a three-letter code whereby the first letter refers to the species genus and thus XNC for *Xenopus* non-classical and SNC for those genes coming from X. *tropicalis* (originally named S*ilurana tropicalis)* non-classical, and were numbered based on the order of discovery and subsequently according to homology and/or orthology with previously described class Ib gene subfamilies. Furthermore, we have recently demonstrated that the *X. laevis* class Ib gene, *XNC10* is required for the development and function of a population of semi-invariant T (iT) cells (Edholm et al. 2013). Notably, XNC10 and its unequivocal *X. tropicalis* ortholog represent a unique class Ib lineage, divergent from other XNC/SNC genes, and it has been suggested that this gene is conserved in other *Xenopus* species (Goyos et al. 2009; Goyos et al. 2011). To further delineate the evolutionary relationships among all XNC/SNC class Ib genes and XNC/SNC10 in particular, we have identified and characterized all individual subfamilies of *X. laevis* class Ib genes and conducted a comprehensive genomic and phylogenetic study of class Ib genes between *X. laevis* and *X. tropicalis*. Furthermore, we demonstrate that the *XNC/SNC10* gene lineage is highly conserved among divergent amphibian species of the *Xenopodinae* subfamily regardless of genome ploidy.

Material and Methods

Experimental animals

Outbred strain of *X. laevis, X. borealis* and *X.* (*Silurana) tropicalis* were from the *Xenopus laevis* Research Resource for Immunology at the University of Rochester (http://www.urmc.rochester.edu/smd/mbi/xenopus/index.htm. All animals were handled under strict laboratory and UCAR regulations (100577/2003-151), and discomfort was minimized at all times.

Identification of XNC genes

Putative *XNC* genes where identified and annotated using the BLAST-like alignment tool (BLAT) algorithm with nucleotide sequences of all known *XNC* and *SNC* α1, α2 and α3 domains as queries against the *X. laevis* genome using both the USSC genome bioinformatics *Xenopus* genomes hosted at NIMR (http://genomes.nimr.mrc.ac.uk, Nov, 2012 assembly and the *Xenopus laevis* Genome project http://xenopus.lab.nig.ac.jp, XenVis 2.0 assembly. The nucleotide and amino acid sequences of known genes were retrieved from GenBank using ENTREZ at http://www.ncbi.nlm.nih.gov.

Identification of XNC/SNC10 homologs

XNC/SNC10 homologs were isolated from genomic DNA using either degenerate primers designed to recognize consensus XNC/SNC10 motifs in the α1 and α2 encoding exons or sub-lineage specific primer (sTable 1). Genomic DNA from *X. laevis, X. borealis* and *X. tropicalis* was purified from liver using Trizol reagent (Invitrogen, CA) following the manufacturers protocol. Genomic DNA samples from *X. gilli, X. mulleri, X. clivii, X. andreii, X. amietti, X. vestitus, X. ruwenzoriensis, X. longipes, X. tropicalis and X. paratropicalis* were kindly provided by Ben Evans (McMaster University, Hamilton, ON, Canada). A total of 50 ng of genomic DNA was used for PCR. Typical parameters were 5 min at 94°C, followed by 30 cycles of 94°C for 30 sec, 58°C to 61°C (depending on the primer pair) for 30 sec, and 72°C for 1 min, with a final extension cycle of 72°C for 10 min. All RT-PCR products where cloned into the pGEM-T easy vector (Promega, WI) and verified by sequencing.

Sequence analysis

Nucleotide and amino acid pairwise alignments were made using CLUSTALW (Thompson et al. 1997) and neighbor-joining trees were drawn using MEGA v5.2.2 (Tamura et al. 2011). Genetic distances were calculated by estimating the number of amino acid substitutions using the p-distance method employing pairwise deletion of gaps. Numbers on nodes represent percentages above 50 of 10,000 bootstrap replicates supporting each partition. Putative ligand binding residues were determined based on alignment with human HLA-A. Accession numbers of species used are as follows: Hs MR1: AAC72900; Oa MR1: ACJ60632, Mm MR1: NP_032235, Ss MR1: XP_003130402, Me MR1: BAM66417, Md MR1: BAM66416, Hs HLA-E: AAA52655, Hs HLA-G: NP_002118, Mm Qa-1: XP_003945787, Mm Qa-2: NP_001188389, Pp MHC-R9: AAF03410, MHC-R6: AAF03409, Amcr-AU: AC668408, Xt MHC: NP_001106387, Xl MHC: AAA16064, Hs

HLA-A2: EAX03243, Gg CD1d.1: BAE19763; Gg Cd1d.2: AAQ75347, Hs CD1a: CAA28049, Hs CD1b: BAJ20663, Hs CD1c: CAG33361, Hs CD1d: AAA51935, CD1e: CAB93156, Mm CD1d-1: NP_031666, Mm CD1d-2: ACM45456, Dn CD1: XP_004447409, Ss CD1: XP_005663293, Cs CD1: XP_004443514, Ec CD1: AEO72063.

Quantitative PCR

RNA from developmental stage 53 *X. laevis* tadpole tissues was prepared using Trizol reagent (Invitrogen) and treated with DNase (Ambion, Life technologies), according to the manufacturer's protocol. 500 ng total RNA was transcribed into cDNA using oligo-dT and iScript reverse transcriptase (Bio-Rad). qPCR was performed with specific primers for XNC10.1 and XNC10.2 and GAPDH was used as a template control (sTable I). Quantitative PCR gene expression was performed using the delta^delta CT method using the ABI 7300 real-time PCR system and PerfeCta® SYBR Green FastMix ROX (Quanta Bioscience Inc). Thermocycling parameters were 2 min at 95°C, followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min. Expression analysis of XNC10.1 and XNC10.2 were normalized to the GAPDH endodgenous control and against the lowest observed tissue expression. Expression analysis was performed using the ABI sequence detection system software (SDS) and all primers were validated prior to use.

Results

3.1 Genomic organization of the XNC/SNC loci

In silico analysis of the *X. tropicalis* genome have revealed the existence of at least 29 *SNC* genes (including pseudogenes) that based on sequence similarities in the α1/α2 domains where grouped into 14 subfamilies (Goyos et al. 2011). Based on Southern blot analysis, X. *laevis* was estimated to have approximately 20 *XNC* genes per haplotype and to date *XNC* gene transcripts have been grouped into 11 different subfamilies (Flajnik et al. 1993; Goyos et al. 2009). However, in the absence of genomic sequence information, the exact number of *XNC* genes and subfamilies was unknown. Recently, in addition to the genome of the only diploid frog in the *Xenopodinae* subfamily, *X. tropicalis* (Hellsten et al. 2010), the genome of the allotetraploid frog, *X. laevis* has been fully sequenced and annotated using a male of the MHC homozygous J inbred strain (http://genomes.nimr.mrc.ac.uk and http://xenopus.lab.nig.ac.jp). To define the genomic organization of *XNC* genes and determine the level of gene synteny between *X. laevis* and *X. tropicalis XNC/SNC* loci, we annotated the XNC gene models and performed a comparative analysis with the *X. tropicalis* class Ib loci. Our previous findings based on the partially annotated *X. tropicalis* genome (version 4.1) indicated the presence of two separate scaffolds each containing multiple *SNC* genes (Goyos et al. 2011). By mapping these two scaffolds to the now fully assembled *X. tropicalis* genome we determined that they are in fact linked and located at the end of chromosome 8, approximately 65 Mbp distant from the *X. tropicalis* classical MHC class Ia gene (Fig. 1).

To characterize the genomic organization of *XNC* genes, we searched the recently available X. *laevis* genomic databases. Because the *X. laevis* genome assembly is still in progress, we found a few orphan scaffolds with XNC sequences including *XNC1* and *XNC2* that mapped on scaffolds 214755 and 54732 of 110 and 36 kbp, respectively. In addition to *XNC2,*

scaffold 54732 contained another putative XNC gene (*XNC6.4*) and a partial XNC gene (*XNC14*). Only *XNC1* was found on scaffold 214755. All the remaining *XNC* genes mapped to one of two scaffolds, 3760 and 326 (http://xenopus.lab.nig.ac.jp, version XenVis 2.0). Combined these two scaffolds encompass 0.83 Mbp that contain a total of 21 predicted *XNC* genes and a number of class Ib-like gene fragments (Fig. 1 and Table 1). Our search resulted in the identification of all the previously defined *XNC* genes (Flajnik et al. 1993; Goyos et al. 2009) plus 12 additional *XNC* genes. Out of these 12 new genes, seven displayed clear homology to previously defined *XNC* subfamilies, including four genes classified as members of the *XNC8* subfamily (*XNC8.2*, *XNC8.3*, *XNC8.4* and *XNC8.5*) and three as belonging to the *XNC6* (*XNC6.2*, *XNC6.3* and *XNC6.4*) subfamily. This suggests that some of the identified XNC genes are members of distinct subfamilies including more than one gene.

To infer orthologous relationships among *X. laevis* and *X. tropicalis* class Ib genes, multiple amino acid alignments of XNC and SNC α1/α2 domains were used to generate a neighbor-joining phylogenetic tree. As previously described, orthologous subfamilies were defined as any clade with a bootstrap value of 80 or higher (Goyos et al. 2011) (Fig. 2). Of the 12 new unassigned XNC sequences three (*XNC14*, *XNC13.1* and *XNC13.5)* clearly cluster with distinct SNC gene subfamilies. One *XNC* gene designated as a new XNC subfamily, *XNC17*, showed similar levels of homology to *XNC1*, *2* and *3* genes with 62–66% amino acid identity in the α1/α2 domains. This *XNC* gene, while clustering together with the *XNC/SNC* subfamily lineages 1, 2, 3, 4 and 16, did not have a clear subfamily relationship among any of the described *XNC/SNC* genes.

Additionally, using the *SNC10* α1 deduced amino acid sequence as search criteria; we identified a new *XNC* gene adjacent to the previously described *XNC10* gene. This gene showed the highest degree of similarity to the *XNC/SNC10* lineage with a α1 amino acid identity of 66% and 50% to *SNC10* and *XNC10* respectively. Moreover, as suggested by phylogenetic analysis supported by high bootstrap values, this *XNC* gene clustered within the *XNC/SNC10* lineage that formed a branch separate from both the classical MHC class Ia genes and the other class Ib genes (Fig 2). Hence, we designated this gene as *XNC10.2* to indicate that it is a distinct member of the *XNC/SNC10* lineage.

Comparative analysis of the deduced *XNC* and *SNC* loci shows that although the number of genes in a given subfamily and the organization of individual genes vary between the two *Xenopus* species, there is an overall high degree of genetic synteny (Fig 1). The majority of class Ib genes of both *X. laevis* and *X. tropicalis* are clustered together in the same transcriptional orientation, with only a few other genes interspersed. In addition to the expressed class Ib genes (indicated by a * in Fig. 1), and typical to MHC gene loci, both *X. laevis* and *X. tropicalis* have multiple class Ib pseudogenes and/or gene fragments at varying stages of disintegration, ranging from nearly complete genes to a single exon. Finally, both *XNC* and *SNC* genes are flanked by a number of SLAM-family member-like genes as well as a T-lymphocyte surface antigen Ly-9-like gene that shows clear locus conservation between the two *Xenopus* species.

3.2 Evolution of distinct XNC/SNC Gene lineages: examples of both conservation and divergence

Despite the relatively long evolutionary time (~ 65 million years) separating *Xenopus* and *Silurana* genera, multiple class Ib genes within the representative species *X. tropicalis* (*Silurana*) and *X. laevis* (*Xenopus*) exhibit an unusual degree of conservation (Goyos et al. 2011), which is further supported by the conserved gene synteny. Aside from this overall high degree of evolutionary preservation, we also found examples of species-specific adaptations, such as expansions and contractions of specific XNC/SNC subfamilies.

Examples of evolutionary conservation include the three monogenetic subfamilies *XNC/SNC3*, 4 and 5. Among these, the XNC/SNC5 genes show the highest degree of conservation (83% and 84% amino acid identity in the α1 and the α2 domain, respectively). In contrast, examples of species-specific divergences include genes that are present and diversified or expanded in one species but not in the other species. In this category, the *XNC8* subfamily is expanded in *X. laevis*, while no corresponding gene was found in *X. tropicalis*. The *XNC8* subfamily is comprised of five genes that (with the exception of *XNC8.2*) are organized in consecutive order (Fig. 1). Using RT-PCR, all *XNC8* gene-specific transcripts, with the exception of *XNC8.4*, were isolated and confirmed by sequencing. *XNC8.3* was predominantly expressed in the thymus, whereas *XNC8.1*, *8.2* and *8.5* were ubiquitously expressed (data not shown). Conversely, in *X. tropicalis* the *SNC16* subfamily was expanded while no corresponding ortholog was found in X. *laevis*. A second set of species-specific specializations is constituted by gene families present in both *X. laevis* and *X. tropicalis* but with varying degrees of expansion such as *XNC/SNC6*, *XNC/SNC7* and *XNC/SNC13*. Finally, putative homologs of functional class Ib genes such as *XNC9* are present in one species but are only detected as pseudogenes and/or gene fragments in another species suggesting that this gene is being silenced and lost in a species-specific manner. In summary, while the majority of class Ib gene lineages are preserved in these two divergent amphibian species, the extent of gene duplications and mutations within individual lineages vary.

3.3 Identification of amphibian class Ib gene lineage subgroups

Phylogenetic analysis based on nucleotide and deduced amino acid multiple alignments of the α1/α2 domains shows that *XNC/SNC* genes form a monophyletic clade with respect to other vertebrate class Ia and class Ib genes. This supports the hypothesis that these genes evolved from a common ancestor. However, despite the overall high level of evolutionary conservation between class Ib gene in the two species, it is also apparent from the phylogenetic analysis that individual class Ib gene subfamilies have undergone further species-specific lineage expansions and contractions, likely reflecting differences in selective pressures thought to influence class Ib evolution. Indeed, the different class Ib gene subfamilies are divergent, sharing within a species between 30–68% amino acid identity in the α1/α2 domains. To gain further insight into individual relationships of different class Ib genes, we performed a detailed comparisons of all class Ib genes thus far identified in *X. laevis* and *X. tropicalis*. Focusing on either individual α1, α2 (data not shown) or combined α1/α2 domains that together form the putative peptide binding region, the *Xenopus/Silurana* class Ib genes segregate into seven clusters, supported by high bootstrap values (Fig. 2 and

sFig1). With the exception of *SNC15*, which cluster within the class Ib gene clade but forms a separate branch distinct from other class Ib genes, all the other clusters contain both *XNC* and *SNC* genes.

In contrast to this loci-based clustering of the α1/α2 domains, the α3 domains of *XNC* and *SNC* genes cluster according to species and show a relatively high degree of conservation (67–94% amino acid identity) (Fig. 3A, B). One of the evolutionary conserved MHC class I signature features is the putative CD8α-binding site (**QD**TE-$(x)_{16}$-AxxV), in particular residues Q226 and D227. We examined the presence of this motif in the *XNC* and *SNC* α3 domains. As previously reported for *X. laevis* class I sequences, this motif is poorly conserved in *XNCs* and class Ia sequences (Flajnik et al. 1993). Instead, the majority of XNC sequences are characterized by a **Y**SEE-$(x)_{16}$-VxxV motif. Comparably, we found that all SNC sequences, with the exception of *SNC6* and *SNC14*, are characterized by a **H**SEE-$(x)_{16}$-VxxV motif (Fig 3A). Similarly, specific residues in CD8α, which in mammals have been shown to be important for class I interaction (i.e., Arg4, Lys21 and Leu25), are not conserved in either *X. laevis* or *X. tropicalis* CD8α sequences (Chida et al. 2011) which suggests a species-specific co-evolution of MHC class Ia and class Ib molecules with their putative cognate T cell co-receptor. Within the *XNC/SNC* α3 cluster the *XNC/SNC14* subfamily is the most divergent, and, as supported by high bootstrap values, forms a separate branch away from the main clade. This suggests that the *XNC/SNC14* gene lineage was one of the first to diverge from the *Xenopodinae* class Ia.

To obtain further evidence of class Ib gene lineage relationships, we analyzed the molecular features and potential conservation of the putative putative ligand binding regions of aligned sequences within each cluster (Table 2 **and** sFig1). In mammals, classical MHC class Ia peptide binding involves side chain-independent recognition of the peptide main chain via nine invariant, highly evolutionary conserved amino acid residues clustered in two shallow pockets, A and F, located at either ends of the antigen binding groove. As previously described (Flajnik et al. 1993; Robert and Ohta 2009) these specific residues can be identified based on deduced amino acid alignments with human class Ia HLA-A invariant residues and are highly conserved in the *X. laevis* class Ia sequences (Bjorkman et al. 1987; Saper et al. 1991). In contrast, these residues are not conserved in the mammalian non-peptide binding nonclassical MHC molecules CD1d (Beckman et al. 1994; Brigl and Brenner 2004) and MR1 (Kjer-Nielsen et al. 2012; Patel et al. 2013). Similarly, for each XNC/SNC cluster, with the exception of a conserved Trp at position 147 in the F pocket, and two conserved Tyr at positions 7 and 171 respectively in the A pocket, the canonical class Ia peptide binding residues within each cluster were distinct from each other as well as from those of human HLA-A and *X. laevis*/*X. tropicalis* class Ia sequences. Thus the lack of full conservation of peptide anchoring tyrosine residues in the A and F pocket of XNC/SNC sequences suggests that these molecules bind non-peptide based antigens. However, a notable feature is that among gene members within each XNC/SNC cluster, these putative ligand binding residues are conserved suggesting that genes within a specific cluster might have a similar specialization, possibly through interaction with a conserved antigenic ligand motif (Table 2). For example, the *XNC/SNC10* gene lineage (cluster 1), exhibits five residues distinct from class Ia and other XNC/SNC proteins. In the A pocket, a notable

substitution (Try167 to Glu/His167) results in the loss of Trp 167, which in class Ia blocks the amino-terminal extension of peptides (Hansen et al. 2007). Although it is difficult to decipher the nature of specific class Ib ligands by sequence analysis alone, it is likely that these different gene lineages encode proteins that bind distinct antigens. Collectively these data suggest that the large family of XNC/SNC class Ib genes can be subdivided into different groups with presumably functionally divergent roles.

3.4 The XNC/SNC10 gene lineage is highly conserved in multiple polyploid species of the Xenopodinae subfamily

Recently, we demonstrated that the expression of XNC10.1 is essential for the development of a distinct subset of semi-invariant T cells (Edholm et al. 2013). Although the equivalent to these class Ib-restricted iT cells have not yet been identified in *X. tropicalis*, the high degree of conservation of primary sequence and differential gene expression profiles between XNC10.1 and SNC10 suggest that SNC10 performs a similar function in *X. tropicalis*. Based on Southern blot analysis it has been suggested that the XNC10.1 and SNC10 genes have been differentially retained across *Xenopodinae* species. To further characterize the evolutionary relationships of these genes, we investigated whether *XNC/SNC10.1* and 10.2 gene orthologs were present among *Xenopodinae* species. We focused our investigation on exon 2 (encoding the α1 domain) and exon 3 (encoding the α2 domain) as these have the highest degree of divergence. As a template for PCR we used genomic DNA isolated from 10 species belonging to the *Xenopus* genus with varying degrees of ploidy ranging from 2N to 12N, as well as 2 species belonging to the *Silurana* genus (Table 3). Using degenerate primers targeting conserved *XNC/SNC10* sequence motifs we obtained XNC/SNC10-related sequences of exon 2 and 3 from all species, thus confirming the preservation of *XNC/SNC* gene in all *Xenopodinae* species tested. We then designed gene-specific primers to exon 2 and 3 of *XNC10.1*, *XNC10.2* and *SNC10*. We obtained XNC10.1-related sequences from all species sampled from *Xenopus* genus, but none from the *Silurana* genus. Similarly, we isolated SNC10-related sequences from the two *Silurana* species (*X. tropicalis* and *X. paratropicalis*). In addition, using primers specific for the SNC10 exon 2 we obtained a gene fragment from *X. gilli* that upon sequencing was confirmed as SNC10-like. However, despite numerous efforts, we were unable to isolate an SNC10-like exon 3 from *X. gilli*. Comparably, XNC10.2-related sequences were only isolated from four species belonging to the *Xenopus* genus (*X. laevis*, *X. borealis*, *X. andreii* and *X. ruwenzorienzis*) suggesting that the *XNC10.2* gene has, in the majority of species tested, either been lost or diverged beyond the typing system employed here. Notably, one of the XNC10.2 variants isolated from *X. ruwenzorienzis* has a four-nucleotide insertion in exon 2, resulting in a frameshift and premature termination of the α1 sequence.

From the 10 *Xenopus* species tested, we identified three predominant XNC10.1 exon 2 sequences (Fig 4A) and six XNC10.1 exon 3 sequences (Fig 4B), which overall showed a remarkably high degree of sequence conservation. Notably, for each species regardless of ploidy, we isolated a maximum of three different sequences. Similarly, for XNC10.2 three different sequences were isolated from *X. borealis* (4N), while for the remaining species (*X. laevis*/2N, *X. andreii*/8N and *X. ruwenzorienzis*/12N) two different sequences with either one or two nucleotide differences were isolated. From the two representative species of the

Silurana genus, two different SNC10 exon 2 and a single exon 3 sequence were identified. Overall, the three genes within the XNC/SNC10 gene lineage display a high degree of sequence conservation and with the exception of positions 146 and 147 in the α2 domain of XNC10.2, the putative ligand binding residues are absolutely conserved in all sequences isolated for a specific gene (Table 4). Notably, all the SNC/XNC10 sequences display an overall conservation of putative invariant anchoring residues in the F pocket, while displaying more variability in the putative A pocket residues consistent with a potentially conserved C-terminus Ag docking moiety.

To elucidate the evolutionary history of the XNC10/SNC10 lineage, we performed phylogenetic analysis (Fig. 5) revealing that both exon 2 and 3 of the different SNC10 and XNC10.1/2 sequences represent three clearly distinctive groups. Furthermore, when comparing exon 2 sequences, the SNC10 sequences cluster with the XNC10.2 sequences with a high bootstrap value. Conversely, for exon 3, the SNC10 clusters with the XNC10.1 sequences, while the XNC10.2 sequences supported by high bootstrap values form a distinct cluster. Collectively these data suggests that the three *XNC/SNC10* genes share a common ancestor. It is also interesting to note that within each main branch the different sequence variants cluster together in a non-species distinctive pattern.

3.5 Differential expression pattern of the XNC/SNC gene lineage

Both *X. laevis* XNC10.1 and *S. tropicalis* SNC10 are primarily expressed in the adult and larval thymi and are detected on thymocytes from early onset of thymic organogenesis (3 days post fertilization; developmental stage 39; (Goyos et al. 2009; Goyos et al. 2011). We were therefore interested in determining the expression patterns of the XNC10 sub-lineage i.e XNC10.1 and XNC10.2 in *X. laevis*. Accordingly, we performed qPCR gene expression using XNC10.1 and XNC10.2 specific primers. In contrast to the predominant thymic expression of the XNC10.1 gene, the XNC10.2 gene was predominantly expressed in the gill, kidney and liver (Fig 6).

Discussion

Although class Ib genes have been identified across jawed vertebrates, from elasmobranch to mammals, (reviewed in (Flajnik and Kasahara 2001)), they are very heterogeneous in sequences, gene number, organization and expression profiles. Moreover, the dynamic evolution of class Ib genes has resulted in extensive diversification of their molecular structure and function. This evolutionary diversification has resulted in multiple species-specific adaptations and, even between closely related species, there are few unambiguous class Ib orthologs (Adams and Luoma 2013; Kulski et al. 2002). The work presented here, markedly extending previous findings (Goyos et al. 2011), highlights the unusual high degree of class Ib conservation across divergent amphibian species and provides new insight into the evolution of class Ib genes.

We conducted an extensive comparative analysis of *Xenopodinae* class Ib genes, including class Ib sequences from closely related species (within either the *Xenopus* or *Silurana* genus) and more divergent species (between representatives of the *Xenopus* and *Silurana* genus). This study reveals two major phylogenetic patterns. On the one hand, some gene

lineages have been maintained as monogenetic subfamilies with remarkably few changes in their nucleotide sequence across divergent species, consistent with a strict functional conservation. On the other hand, other class Ib gene lineages display a higher degree of flexibility, as demonstrated by species-specific adaptations including expansion and contractions of specific gene family, which attests to the distinct selective pressures thought to influence class Ib evolution resulting in functional divergence. In addition to the variable selective pressures exerted on the different *XNC/SNC* gene families, phylogenetic analyses also confirm our initial report (Goyos et al. 2011) suggesting that in *Xenopodinae spp* the α3 domain has undergone strong evolutionary constraints different from those exerted on the α1/α2 domains. This phylogenetic relationship, whereby the α3 domain remains more conserved and clustered according to species, while the α1/α2 part of the same molecule are more divergent is reminiscent of the family of receptors recognizing CD94, which includes the human HLA-E, mouse H2-Qa1 and rat RT-BM (Joly and Rouillon 2006). This supports the idea that concerted evolution takes place among class Ib genes within a species. However, juxtaposed to this more divergent α1/α2 domain, there is also a strong conservation of the α1/α2 putative ligand binding residues within each phylogenetically defined *XNC/SNC* cluster suggesting that these specific residues are under negative selection, possibly to ensure that a distinct function (such as antigen presentation) is maintained.

Both *X. laevis* and *X. tropicalis* have a large number of class Ib genes, their genomes contain at minimum of 25 *XNC* and 29 *SNC* genes, respectively, and the majority if not all of these genes are linked and located distally from the class Ia gene. Out of the 17 distinct class Ib gene lineages identified to date, 11 display orthologous relationships between *X. laevis* and *X. tropicalis.* This degree of conservation of class Ib genes, with the exception of CD1 and MR1, is in contrast to that observed across mammals and indicative of a different pattern of class Ib evolution in amphibians. This implies that various *XNC/SNC* gene lineages have evolved to perform distinct functions. In mammals both CD1 and MR1 are located outside the MHC region (Adams and Luoma 2013) and have been shown to function as restricting element for distinct populations of semi-invariant T cells. These include the CD1d restricted iNKT cells (Bendelac et al. 1995; Bendelac et al. 2007) and the MR1 restricted mucosal associated invariant T (MAIT) cells (Treiner et al. 2003). CD1 and MR1 are less heterogeneous than class Ia and most class Ib genes, and both CD1 and MR1 orthologous genes have been described in most mammals. CD1 orthologs have also been identified in chickens suggesting that this gene was present prior to the mammalian/avian split corresponding to ~310 million years of evolution (Miller et al. 2005). Notably, none of the class Ib genes of *X. laevis* or *X. tropicalis* identified to date share significant sequence identity with either CD1 or MR1. Additionally, by close range synteny analysis (Y. Ohta, personal communication), neither CD1 nor MR1 like genes were found in the class Ib genomic loci of either *X. tropicalis* or *X. laevis*.

The fact that class Ib genes are present in all taxa of jawed vertebrates attest to their evolutionary primordial origins However, the evolutionary history of class Ib genes in lower vertebrates remains largely unknown and the phylogenetic relationships of these genes are difficult to outline. For example, a large number of divergent class Ib genes have been

identified in the urodele amphibian, *Ambystoma mexicanium*, although none of these genes are orthologous to any of the *XNC/SNC* genes identified to date (Sammut et al. 1999). Furthermore, in contrast to *Xenopus* where all class Ib genes identified to date form a single linkage group outside the MHC, Southern blot analysis suggest that in *A. mexicanium* class Ib, class Ia as well as MHC class II genes are linked (Sammut et al. 1999) Elasmobranchs also have variable numbers of class Ib genes including highly divergent species-specific genes (Bartl et al. 1997; Wang et al. 2003). Similarly, teleosts encode multiple class I genes that based on evolutionary relationships have been grouped into distinct lineages (U-, Z/ZE-, L- and S-lineage) (Lukacs et al. 2010). These teleost class I lineages are differentially distributed in divergent species, with some lineages like the U-lineage, containing both class Ia and class Ib like genes being broadly represented. In contrast, genes belonging to the L-lineage are highly divergent class Ib genes and have so far only been identified in salmonids and cyprinids (Dijkstra et al. 2007). Notably, Atlantic cod (*Gadhus Morhua*) has lost MHC class II genes coinciding with a large expansion (more than 100 genes) of their U-lineage class I genes. Based on phylogenetic analysis, these genes segregate into two distinct clades that display distinct ratios of non-synonymous to synonymous mutation and differences in the average nucleotide diversity per site, suggesting possible neofunctionalization for this MHC class I lineage (Star et al. 2011). Interestingly, in contrast to other vertebrate groups, the MHC class I and II regions of teleosts are located on different linkage groups, which might have facilitated independent evolution of these two systems (Lukacs et al. 2010). However, despite the presence of class Ib genes in all ectothermic vertebrates, to date no conserved class Ib gene ortholog have been identified across distant species.

A unique feature of the *Xenopus* immune system not found in mammals is the drastic physiological transition between larval and adult stages that occurs during metamorphosis. In particular, while class Ia transcripts are detected in tadpoles (Goyos and Robert, unpublished), there is no consistent class Ia surface protein expression until the onset of metamorphosis (Flajnik et al. 1986; Rollins-Smith et al. 1997; Salter-Cid et al. 1998). Irrespective of this, tadpoles are immunocompetent and have circulating $CD8^+$ T cells (Barlow and Cohen 1983). This could either be due to low levels (below antibody detection) of class Ia protein expression capable of supporting T cell development and function. Alternatively, it has been suggested that during this life stage, class Ib molecules may compensates for the low class Ia expression. In support of this we have shown that XNC10.1, in a manner similar to CD1d and MR1, is required for the development and function of a distinct semi-invariant T cell population (Edholm et al. 2013). Also, we recently showed that in the spleen of early developmental tadpoles (before optimal class Ia surface expression), there is a preponderance of six unique invariant TCRα rearrangements in the $CD8^-$ and $CD8^{dim}$ T cell populations (Edholm et al. 2013). In light of this, it is tempting to speculate that there are additional class Ib-restricted iT cell populations in the tadpole and that this pool of innate-like T cells are capable of facilitating immune effector functions. Considering the wide expansion of class Ib genes in many ectothermic vertebrate species (*Xenopus, A. mexicanium, G. Morhua*), it is possible that ectothermic vertebrates relies heavily on class Ib-dependent innate-like invariant T cells. This type of immunity may provide selective advantage for organisms that undergo rapid external development by

maximizing the use of a small number of lymphocytes and minimizing negative selection in the thymus.

Our present findings indicate that despite the otherwise high variability across the genomes of these species including varying levels of polyploidy the *XNC/SNC10* gene lineage is conserved in all *Xenopodinae* species examined to date. Notably, *XNC10.1* and *SNC10* share the same expression pattern with a predominant and developmentally early thymocyte expression, from the onset of thymic organogenesis suggesting that these molecules have similar conserved functions. Phylogenetic analysis confirms that *XNC10.1*, *XNC10.2* and *SNC10* represent three distinct genes and that each species tested has at least one and sometimes two of these genes in their genome. Whereas *XNC10.1* is represented in all species tested of the *Xenopus* genus, *SNC10* sequences, with the exception of an SNC10-like α1 sequence isolated from *X. clivii*, are exclusively found in the *Silurana* genus. Comparably, XNC10.2 is found sporadically throughout the *Xenopus* genus suggesting that this gene has either been lost or undergone extensive modifications in the majority of *Xenopus* species. Therefore, although the XNC10.2 gene is found in *X. laevis, X. borealis, X. andreii* and *X. ruwenzoriensis*, there may be a combinatorial effect of lower XNC10 gene dosage in polyploidy species (X. *borealis* and X. *laevis* are tetraploid, *X. andreii* is octotetraploid and *X. ruwenzoriensis* is dodecaploid) and XNC10.2 locus degradation, which can preclude their identification by PCR. Interestingly, in *X. ruwenzoriensis* (12N) close to half of the α1 sequences isolated were nonfunctional, suggesting that this gene is approaching non-functionality. Nevertheless, gene expression studies in both *X. laevis* and X. *borealis* show that at least in these two species both XNC10.1 and 10.2 genes are expressed at the mRNA level. Moreover, in X. *laevis* XNC10.1 and 10.2 display distinct non-overlapping expression patterns suggesting that these molecules have distinct functional roles.

Collectively, these results provide strong supportive evidence that class Ib genes represent an evolutionary ancient gene family that perform functional roles distinct from those of class Ia. Indeed, the high degree of genetic conservation of multiple nonclassical MHC class Ib lineages observed in the anuran subfamily *Xenopodinae* imply preservation for biologically important functions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Tina Martin for the expert animal husbandry. We also gratefully thank Drs. Martin Flajnik and Leon Grayfer, as well as Nikesha Haynes-Gilmore for helpful discussions and critical reading of the manuscript. This research was supported by grants R24-AI-059830 from National Institute of Allergy and Infectious Diseases (NIH/NIAID).

Bibliography

Adams EJ, Luoma AM. The adaptable major histocompatibility complex (MHC) fold: structure and function of nonclassical and MHC class I-like molecules. Annu Rev Immunol. 2013; 31:529–61. [PubMed: 23298204]

Adams EJ, Parham P. Species-specific evolution of MHC class I genes in the higher primates. Immunol Rev. 2001; 183:41–64. [PubMed: 11782246]

Baker ML, Miller RD. Evolution of mammalian CD1: marsupial CD1 is not orthologous to the eutherian isoforms and is a pseudogene in the opossum Monodelphis domestica. Immunology. 2007; 121:113–21. [PubMed: 17244156]

Barlow EH, Cohen N. The thymus dependency of transplantation allotolerance in the metamorphosing frog Xenopus laevis. Transplantation. 1983; 35:612–9. [PubMed: 6346599]

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; 159:6097–104. [PubMed: 9550410]

Beckman EM, Porcelli SA, Morita CT, Behar SM, Furlong ST, Brenner MB. Recognition of a lipid antigen by CD1-restricted alpha beta+ T cells. Nature. 1994; 372:691–4. [PubMed: 7527500]

Bendelac A, Lantz O, Quimby ME, Yewdell JW, Bennink JR, Brutkiewicz RR. CD1 recognition by mouse NK1+ T lymphocytes. Science. 1995; 268:863–5. [PubMed: 7538697]

Bendelac A, Savage PB, Teyton L. The biology of NKT cells. Annu Rev Immunol. 2007; 25:297–336. [PubMed: 17150027]

Bewick AJ, Chain FJ, Heled J, Evans BJ. The pipid root. Syst Biol. 2012; 61:913–26. [PubMed: 22438331]

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC. Structure of the human class I histocompatibility antigen, HLA-A2. Nature. 1987; 329:506–12. [PubMed: 3309677]

Brigl M, Brenner MB. CD1: antigen presentation and T cell function. Annu Rev Immunol. 2004; 22:817–90. [PubMed: 15032598]

Chida AS, Goyos A, Robert J. Phylogenetic and developmental study of CD4, CD8 alpha and beta T cell co-receptor homologs in two amphibian species, Xenopus tropicalis and Xenopus laevis. Dev Comp Immunol. 2011; 35:366–77. [PubMed: 21075137]

Dascher CC. Evolutionary biology of CD1. Curr Top Microbiol Immunol. 2007; 314:3–26. [PubMed: 17593655]

Dijkstra JM, Katagiri T, Hosomichi K, Yanagiya K, Inoko H, Ototake M, Aoki T, Hashimoto K, Shiina T. A third broad lineage of major histocompatibility complex (MHC) class I in teleost fish; MHC class II linkage and processed genes. Immunogenetics. 2007; 59:305–21. [PubMed: 17318646]

Edholm ES, Albertorio Saez LM, Gill AL, Gill SR, Grayfer L, Haynes N, Myers JR, Robert J. Nonclassical MHC class I-dependent invariant T cells are evolutionarily conserved and prominent from early development in amphibians. Proc Natl Acad Sci U S A. 2013; 110:14342–7. [PubMed: 23940320]

Evans BJ. Genome evolution and speciation genetics of clawed frogs (Xenopus and Silurana). Front Biosci. 2008; 13:4687–706. [PubMed: 18508539]

Evans BJ, Kelley DB, Tinsley RC, Melnick DJ, Cannatella DC. A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. Mol Phylogenet Evol. 2004; 33:197–213. [PubMed: 15324848]

Flajnik MF, Kasahara M. Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. Immunity. 2001; 15:351–62. [PubMed: 11567626]

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; 12:4385–96. [PubMed: 8223448]

Flajnik MF, Kaufman JF, Hsu E, Manes M, Parisot R, Du Pasquier L. Major histocompatibility complex-encoded class I molecules are absent in immunologically competent Xenopus before metamorphosis. J Immunol. 1986; 137:3891–9. [PubMed: 3537126]

Goyos A, Ohta Y, Guselnikov S, Robert J. Novel nonclassical MHC class Ib genes associated with CD8 T cell development and thymic tumors. Mol Immunol. 2009; 46:1775–86. [PubMed: 19237199]

Goyos A, Sowa J, Ohta Y, Robert J. Remarkable conservation of distinct nonclassical MHC class I lineages in divergent amphibian species. J Immunol. 2011; 186:372–81. [PubMed: 21115732]

Hansen TH, Huang S, Arnold PL, Fremont DH. Patterns of nonclassical MHC antigen presentation. Nat Immunol. 2007; 8:563–8. [PubMed: 17514210]

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; 328:633–6. [PubMed: 20431018]

Joly E, Rouillon V. The orthology of HLA-E and H2-Qa1 is hidden by their concerted evolution with other MHC class I molecules. Biol Direct. 2006; 1:2. [PubMed: 16542007]

Kasahara M, Flajnik MF, Ishibashi T, Natori T. Evolution of the major histocompatibility complex: a current overview. Transpl Immunol. 1995; 3:1–20. [PubMed: 7551974]

Kronenberg M, Gapin L. The unconventional lifestyle of NKT cells. Nat Rev Immunol. 2002; 2:557–68. [PubMed: 12154375]

Kulski JK, Shiina T, Anzai T, Kohara S, Inoko H. Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. Immunol Rev. 2002; 190:95–122. [PubMed: 12493009]

Le Bourhis L, Guerri L, Dusseaux M, Martin E, Soudais C, Lantz O. Mucosal-associated invariant T cells: unconventional development and function. Trends Immunol. 2011; 32:212–8. [PubMed: 21459674]

Le Bourhis L, Martin E, Peguillet I, Guihot A, Froux N, Core M, Levy E, Dusseaux M, Meyssonnier V, Premel V, Ngo C, Riteau B, Duban L, Robert D, Huang S, Rottman M, Soudais C, Lantz O. Antimicrobial activity of mucosal-associated invariant T cells. Nat Immunol. 2010; 11:701–8. [PubMed: 20581831]

Lukacs MF, Harstad H, Bakke HG, Beetz-Sargent M, McKinnel L, Lubieniecki KP, Koop BF, Grimholt U. Comprehensive analysis of MHC class I genes from the U-, S-, and Z-lineages in Atlantic salmon. BMC Genomics. 2010; 11:154. [PubMed: 20205726]

Matsuda JL, Gapin L. Developmental program of mouse Valpha14i NKT cells. Curr Opin Immunol. 2005; 17:122–30. [PubMed: 15766670]

Miller MM, Wang C, Parisini E, Coletta RD, Goto RM, Lee SY, Barral DC, Townes M, Roura-Mir C, Ford HL, Brenner MB, Dascher CC. Characterization of two avian MHC-like genes reveals an ancient origin of the CD1 family. Proc Natl Acad Sci U S A. 2005; 102:8674–9. [PubMed: 15939884]

Nei M, Gu X, Sitnikova T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci U S A. 1997; 94:7799–806. [PubMed: 9223266]

Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. Annu Rev Genet. 2005; 39:121–52. [PubMed: 16285855]

Patel O, Kjer-Nielsen L, Le Nours J, Eckle SB, Birkinshaw R, Beddoe T, Corbett AJ, Liu L, Miles JJ, Meehan B, Reantragoon R, Sandoval-Romero ML, Sullivan LC, Brooks AG, Chen Z, Fairlie DP, McCluskey J, Rossjohn J. Recognition of vitamin B metabolites by mucosal-associated invariant T cells. Nat Commun. 2013; 4:2142. [PubMed: 23846752]

Piontkivska H, Nei M. Birth-and-death evolution in primate MHC class I genes: divergence time estimates. Mol Biol Evol. 2003; 20:601–9. [PubMed: 12679545]

Robert J, Ohta Y. Comparative and developmental study of the immune system in Xenopus. Dev Dyn. 2009; 238:1249–70. [PubMed: 19253402]

Rodgers JR, Cook RG. MHC class Ib molecules bridge innate and acquired immunity. Nat Rev Immunol. 2005; 5:459–71. [PubMed: 15928678]

Rollins-Smith LA, Barker KS, Davis AT. Involvement of glucocorticoids in the reorganization of the amphibian immune system at metamorphosis. Dev Immunol. 1997; 5:145–52. [PubMed: 9587715]

Salter-Cid L, Nonaka M, Flajnik MF. Expression of MHC class Ia and class Ib during ontogeny: high expression in epithelia and coregulation of class Ia and lmp7 genes. J Immunol. 1998; 160:2853–61. [PubMed: 9510188]

Sammut B, Du Pasquier L, Ducoroy P, Laurens V, Marcuz A, Tournefier A. Axolotl MHC architecture and polymorphism. Eur J Immunol. 1999; 29:2897–907. [PubMed: 10508264]

Saper MA, Bjorkman PJ, Wiley DC. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 A resolution. J Mol Biol. 1991; 219:277–319. [PubMed: 2038058]

Skold M, Behar SM. Role of CD1d-restricted NKT cells in microbial immunity. Infect Immun. 2003; 71:5447–55. [PubMed: 14500461]

Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrom M, Gregers TF, Rounge TB, Paulsen J, Solbakken MH, Sharma A, Wetten OF, Lanzen A, Winer R, Knight J, Vogel JH, Aken B, Andersen O, Lagesen K, Tooming-Klunderud A, Edvardsen RB, Tina KG, Espelund M, Nepal C, Previti C, Karlsen BO, Moum T, Skage M, Berg PR, Gjoen T, Kuhl H, Thorsen J, Malde K, Reinhardt R, Du L, Johansen SD, Searle S, Lien S, Nilsen F, Jonassen I, Omholt SW, Stenseth NC, Jakobsen KS. The genome sequence of Atlantic cod reveals a unique immune system. Nature. 2011; 477:207–10. [PubMed: 21832995]

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011; 28:2731–9. [PubMed: 21546353]

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 1997; 25:4876–82. [PubMed: 9396791]

Treiner E, Duban L, Bahram S, Radosavljevic M, Wanner V, Tilloy F, Affaticati P, Gilfillan S, Lantz O. Selection of evolutionarily conserved mucosal-associated invariant T cells by MR1. Nature. 2003; 422:164–9. [PubMed: 12634786]

Tsukamoto K, Deakin JE, Graves JA, Hashimoto K. Exceptionally high conservation of the MHC class I-related gene, MR1, among mammals. Immunogenetics. 2013; 65:115–24. [PubMed: 23229473]

Wang C, Perera TV, Ford HL, Dascher CC. Characterization of a divergent non-classical MHC class I gene in sharks. Immunogenetics. 2003; 55:57–61. [PubMed: 12715247]

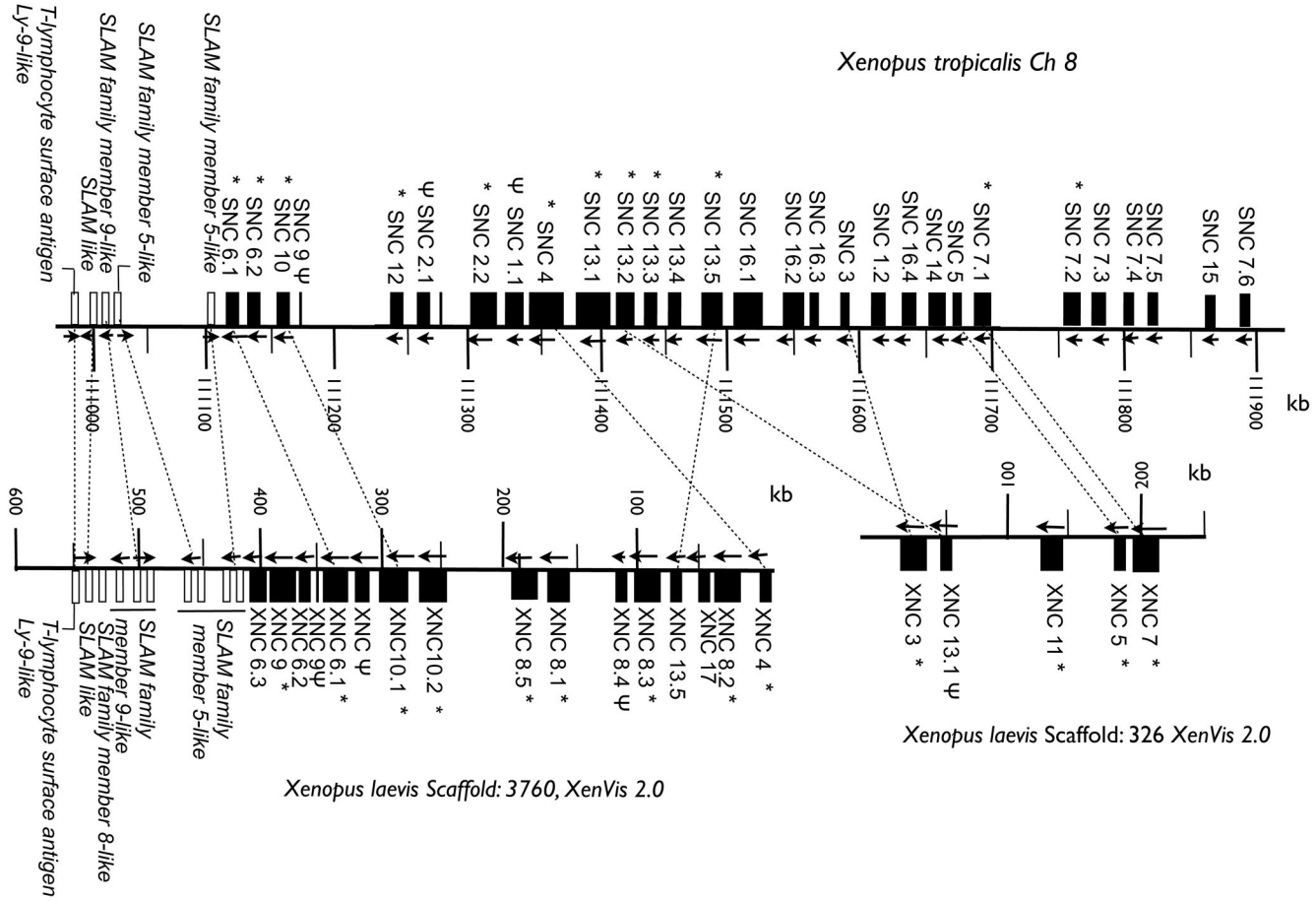


Fig 1. Genomic organization of *X. tropicalis* and *X. laevis* nonclassical MHC genes (SNC and XNC)

Organization of the 29 *SNC* and 21 *XNC* genes including pseudogenes, indicated by Ψ. *SNC* genes are located on chromosome 8 and *XNC* genes are located on two different scaffolds encompassing a total of 0.8 Mbp (drawn to scale). Scaffold numbers for *X. laevis* are based on the *Xenopus laevis* Genome project http://xenopus.lab.nig.ac.jp, XenVis 2.0 assembly and chromosome number and gene location for *X. tropicalis* are based on the USSC genome bioinformatics *Xenopus* genomes hosted at NIMR (http://genomes.nimr.mrc.ac.uk, July, 2013 assembly. Expressed genes, as supported by EST identifications and/or RT-PCR, are indicated by * and arrows indicate transcriptional orientation. Flanking non-class Ib genes are shown as white boxes. Orthologous relationships defined by multiple sequence alignments and phylogenetic analysis between *XNC* and *SNC* genes are indicated by dashed lines.

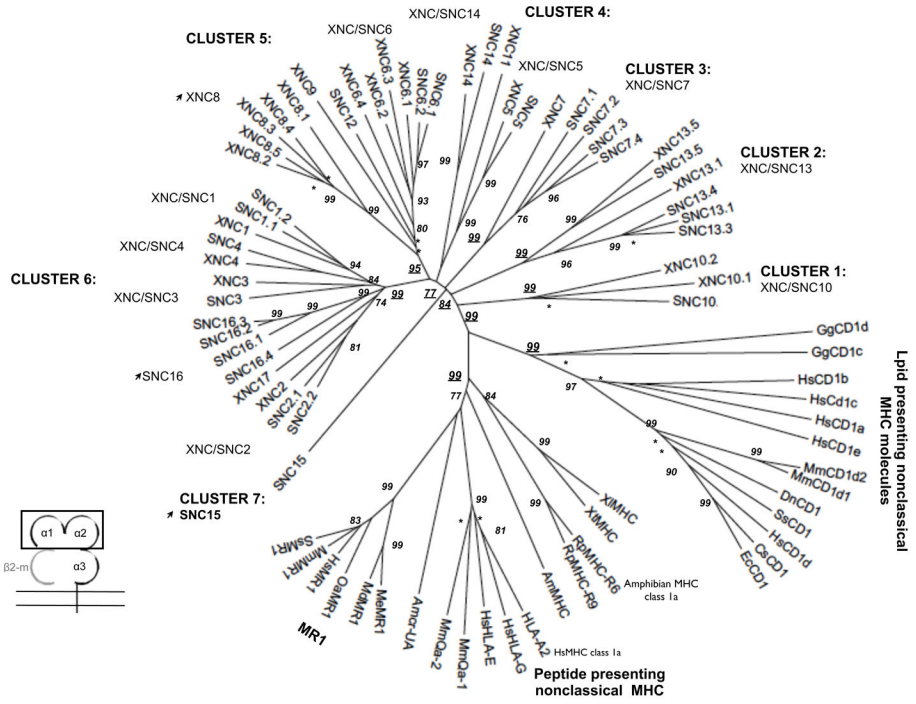


Fig 2. The phylogenetic relationships of XNC and SNC α1/α2 domains compared to selected vertebrate classical and nonclassical MHC genes

The neighbor-joining tree was constructed from amino acid alignments of the alpha 1 and alpha 2 domains using pairwise gap deletions and the p-distance method to estimate evolutionary distance. The tree was drawn using MEGA 5.2. and confidence values were measured using 10,000 bootstrap replications with the values indicated at key nodes and * indicating values < 50. Arrows indicate non-classical subfamilies within the *Xenopodinae* without clear orthologous relationships between *X. laevis* and *X.tropicalis.* Species abbreviations are: Xl, *X. laevis*, Xt, *X. tropicalis*; Hs, human; Mm, mouse; Ss, pig; Gg, chicken; Me, Tammar wallaby; Wd, short-tailed opossum, Oa, sheep; Ec, horse; Cs, rhinoceros; Dn, nine-banded armadillo; Rp, northern leopard frog and Amcr, Galapagos marine iguana.

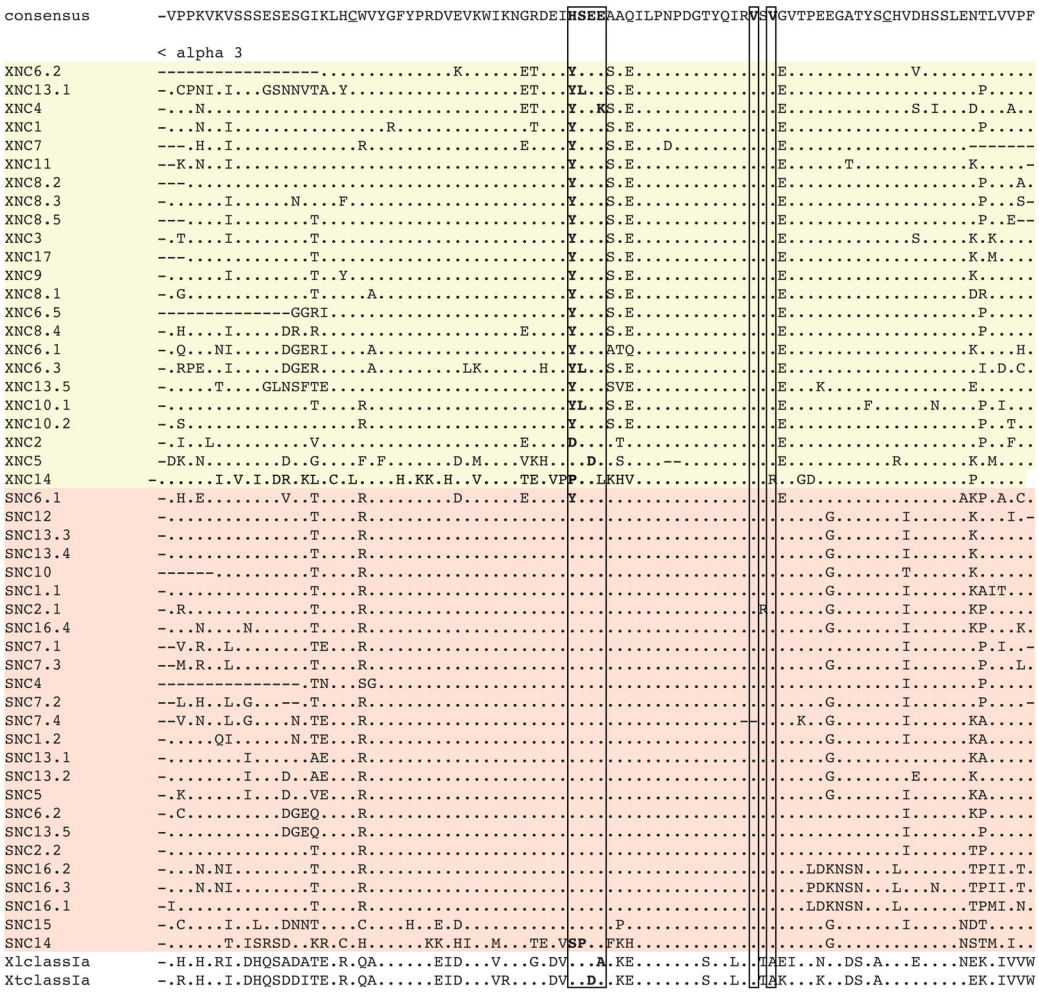


Figure 3A

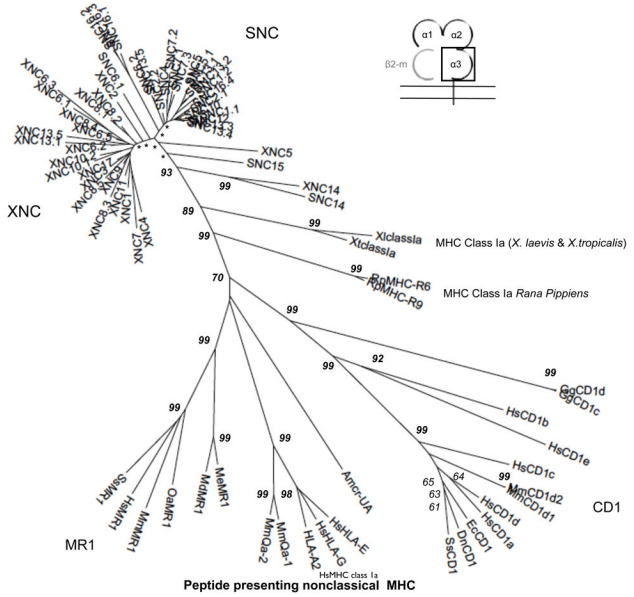


Figure 3B

Fig 3. Multiple deduced amino acid sequence alignment and phylogenetic relationships of the α3 domains of *X. tropicalis* and *X. laevis* nonclassical MHC genes (SNC and XNC)

(A) Deduced amino acid alignment of XNC and SNC α3 domains with *X. laevis* and *X. tropicalis* MHC class Ia. A consensus sequence is shown at the top and dots indicate amino acids identical to this sequence; (-) represent gaps in the alignment and conserved cysteines are in bold and underlined. The MHC class I CD8 binding site is boxed and indicated in bold. **(B)** The neighbor-joining tree was constructed from amino acid alignments of the alpha 3 domains using pairwise gap deletions and the p-distance method to estimate evolutionary distance. The tree was drawn using MEGA 5.2. and confidence values were measured using 10,000 bootstrap replications with the values indicated at key nodes with * indicating values <50. Species abbreviations are: Xl, *X. laevis*, St, *S. tropicalis*; Hs, human; Mm, mouse; Ss, pig; Gg, chicken; Me, tammar wallaby; Wd, short-tailed opossum, Oa, sheep; Ec, horse; Cs, rhinoceros; Dn, nine-banded armadillo; Rp,northern leopard frog and Amcr, Galapagos marine iguana.

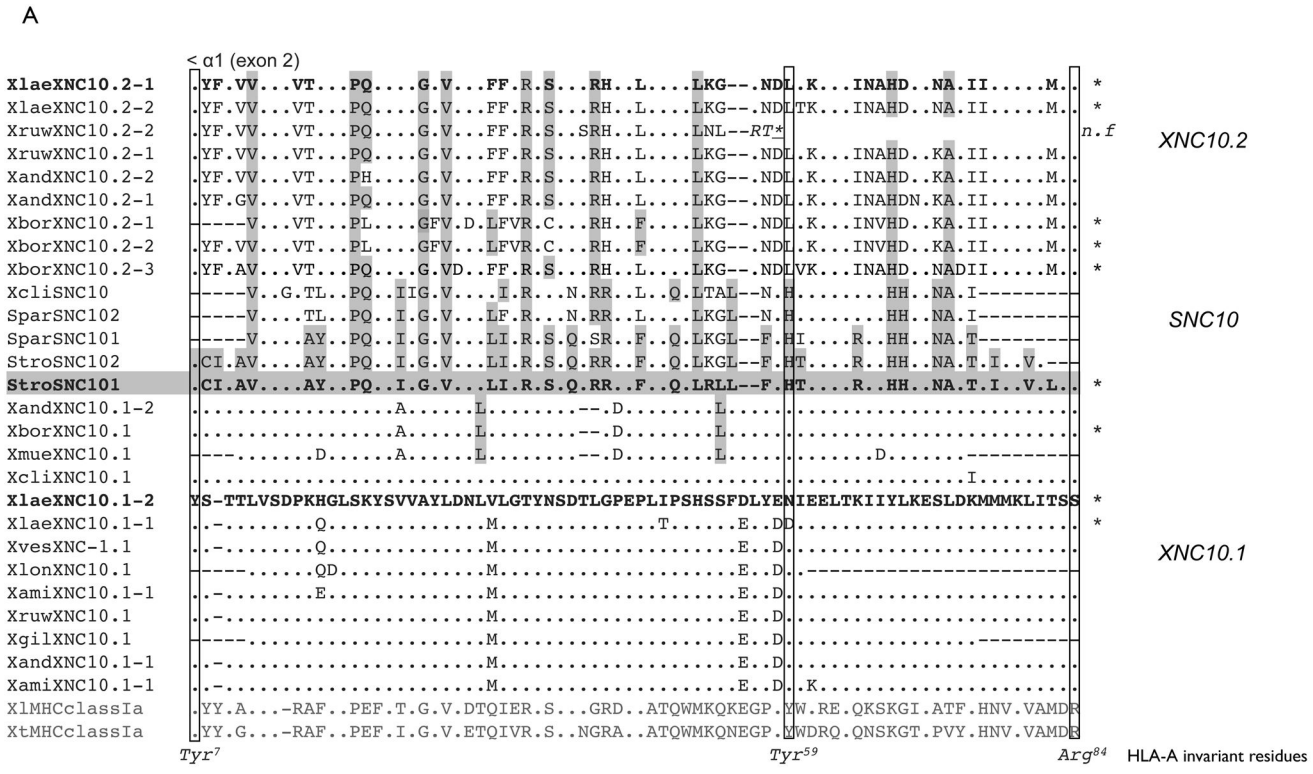


Figure 4A

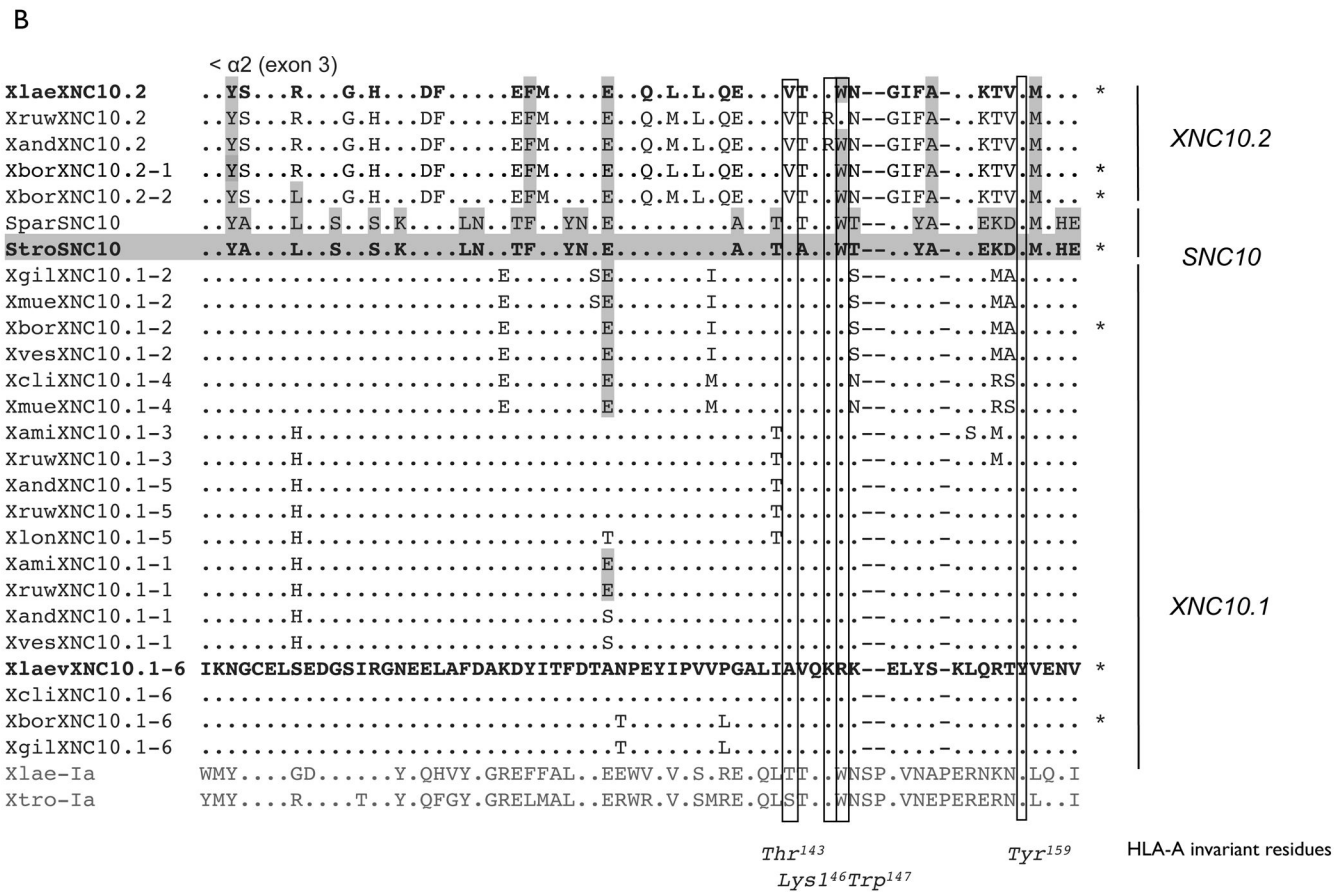


Figure 4B

Fig 4. XNC/SNC10 sequence analysis

Deduced amino acid alignment of **(A)** alpha1/exon 2 and **(B)** alpha 2/exon 3 of XNC/SNC10 genes from different *Xenopodinae* species. Dots indicates amino acid identical to *X. laevis XNC10.1* and grey shading indicate amino acids identical to *X. tropicalis* SNC10 while (-) represents gaps in the alignment. Sequences from the different species grouping as either XNC10.1, XNC10.2 or SNC10 are indicated on the right and * indicate that the gene is expressed. Putative peptide anchoring residues, based on alignment with human HLA-A, are boxed and indicated at the bottom.

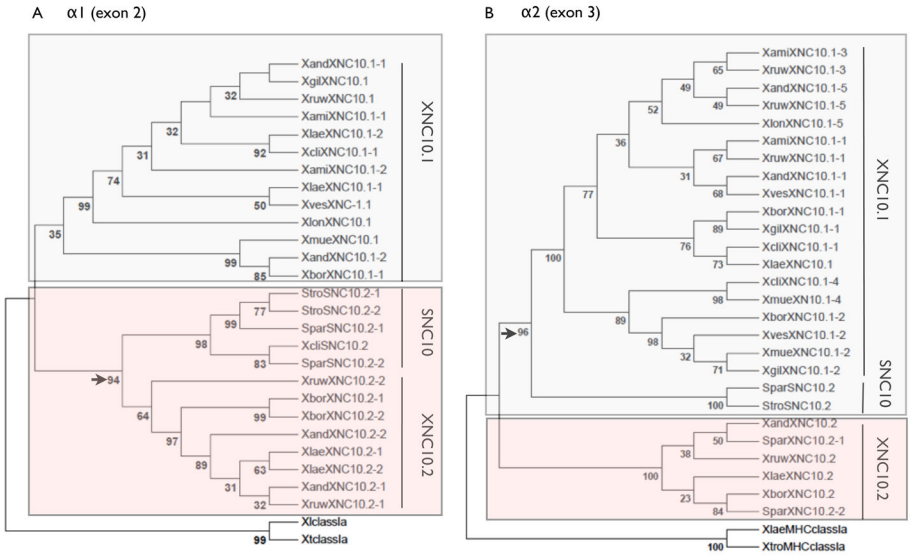


Fig 5. Phylogenetic relationships between XNC/SNC10 gene lineage within multiple species of the *Xenopodinae* subfamily

Neighbor-joining trees were constructed based on amino acid alignments of **(A)** alpha1/exon 2 and **(B)** alpha 2/exon 3 of XNC/SNC10 sequences from different *Xenopodinae* species. Trees were rooted with *X. laevis* and *X. tropicalis* MHC class Ia genes, accession numbers: ABA43373.1 and AAP36728.1 respectively. The trees were drawn using MEGA 5.2. using pairwise gap deletions and the p-distance method to estimate evolutionary distance and confidence values were measured using 10,000 bootstrap replications with the values indicated at key nodes. Trees generated based on nucleotide alignments resulted in similar topology (data not shown).

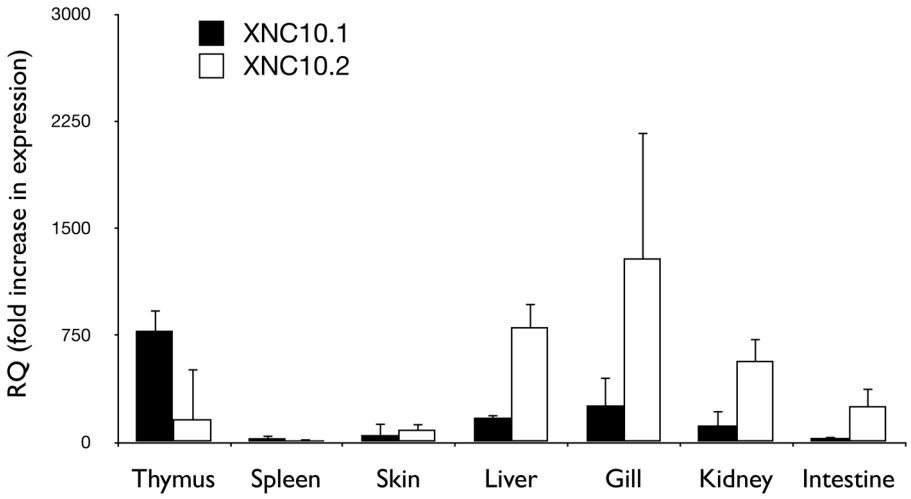


Fig 6. Differential expression of XNC10.1 and XNC10.2

Expression of XNC10.1 and XNC10.2 in various tissues of *X. laevis* and developmental stage 53/54 tadpoles. Results are normalized to an endogenous control (GAPDH) and expressed as fold change compared to expression of XNC10.2 in spleen. All results are presented as mean ± s.e.m, n = 5 for each group.

Table 1

Class Ib residues at conserved positions involved in MHC class Ia terminal peptide anchoring.

Class Ib: XNC/SNC											
	HLA-A2	**CD1d**	**Xl Class Ia**	**MR1**	**CLUSTER 1[a]**	**CLUSTER 2[a]**	**CLUSTER 3[a]**	**CLUSTER 4[a]**	**CLUSTER 5[a]**	**CLUSTER 6[a]**	**CLUSTER 7[a]**
A pocket	Tyr7	Cys	-	-	-	-/Phe (Leu/Val)	- (Phe)	-	-/Leu	- (Phe)	-
	Tyr59	Gln	-	His	Leu/His/Asp/Asn	Met (Ile)	His(Glu)	His	His (Leu)	His (-)	Ala
	Tyr159	Leu	-	Trp	-	-/Phe	Phe (Leu/Ser)	-	- (Asn)	Phe/-	-
	Trp167	Phe	Gly	-	His (Gly)	Arg	His	His (Tyr)	His/Gln (Asn/Tyr)	His (Asn)	Asp
	Tyr171	Leu	-	Phe	- (His)	- (Phe)	- (Phe)	-	-	-	-
F pocket	Arg/Tyr84	Met	-	His	Ser	Ile/Leu	Leu/(Ser)	Phe	Phe/Tyr	Val/Phe	Met
	Thr143	Ala, Pro	-	Thr/Ile	Ala (Val)	Val (Ile/Ala)	Leu	Leu/Met	Met (Val/Ile)	Val (Leu/Met)	Val
	Lys146	Val	-	Ala	Arg/-	Ile (Val/Leu)	GAP	Ser/Arg/Glu/Gln	Gln (leu)	Leu	Gln
	Trp147	Leu	-	Trp	- (Arg)	-	-	-	-	-	Cys
Ligand	**peptides**	**lipids**		**vitamin B metabolites**							

Putative ligand binding residues of Xenopus/Siluriana class Ia and class Ib where predicted based on alignment with human HLA-A.

[a]Cluster refers to phylogenetically related families of XNC/SNC sequences.

- conservation of residue with MHC class Ia; GAP, gap in the sequence such that no equivalent residue is determined.

Bold, unique residue among the differnet XNC/SNC molecules. x/x, equal distribution of two different amino acids (x) amino acid only detected in 1–2 of sequences in the group.

Grey shade, hydrophobic amino acids

Table 2

Putative class Ib residues at conserved positions involved in terminal peptide anchoring.

Class Ib: XNC/SNC											
	HLA-A2	**CD1d**	**Xl Class Ia**	**MR1**	**CLUSTER 1[a]**	**CLUSTER 2[a]**	**CLUSTER 3[a]**	**CLUSTER 4[a]**	**CLUSTER 5[a]**	**CLUSTER 6[a]**	**CLUSTER 7[a]**
A pocket	Tyr7	Cys	-	-	-	-/Phe (Leu/Val)	- (Phe)	-	-/Leu	- (Phe)	-
	Tyr59	Gln	-	His	Leu/His/**Asp**/**Asn**	**Met** (Ile)	His(Glu)	His	His (Leu)	His (-)	**Ala**
	Tyr159	Leu	-	Trp	-	-/Phe	Phe (Leu/Ser)	-	- (Asn)	Phe/-	-
	Trp167	Phe	Gly	-	His (**Gly**)	**Arg**	His	His (Tyr)	His/Gln (Asn/Tyr)	His (Asn)	**Asp**
	Tyr171	Leu	-	Phe	- (His)	- (Phe)	- (Phe)	-	-	-	-
F pocket	Arg/Tyr84	Met	-	His	**Ser**	Ile/Leu	Leu/(Ser)	Phe	Phe/Tyr	**Val**/Phe	**Met**
	Thr143	Ala, Pro	-	Thr/Ile	Ala (Val)	Val (Ile/Ala)	Leu	Leu/Met	Met (Val/Ile)	Val (Leu/Met)	Val
	Lys146	Val	-	Ala	**Arg/-**	Ile (Val/Leu)	**GAP**	Ser/Arg/Glu/Gln	Gln (leu)	Leu	**Gln**
	Trp147	Leu	-	Trp	- (**Arg**)	-	-	-	-	-	**Cys**
Ligand	**peptides**	**lipids**		**vitamin metabolites**							

Putative peptide binding residues of Xenopus/Siluriana class Ia and class Ib where predicted based on alignment with human HLA-A.

[a] Cluster refers to phylogenetically related families of XNC/SNC sequences.

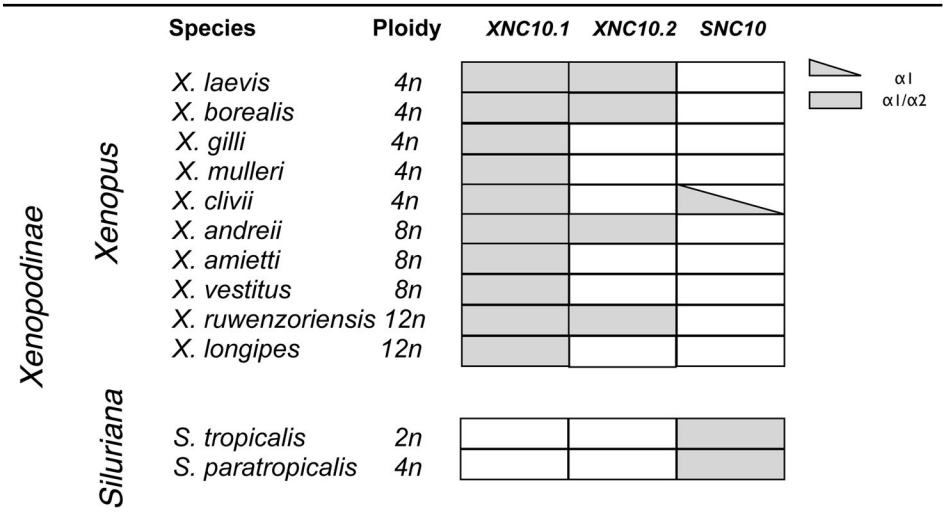
conservation of residue with MHC class Ia; GAP, gap in the sequence such that no equivalent residue is determined.

Bold, unique residue among the differnet XNC/SNC molecules. x/x, equal distribution of two different amino acids (x) amino acid only detected in 1–2 of sequences in the group.

Grey shade, hydrophobic amino acids

Table 3

Species distribution of XNC/SNC10-related exons from polyploid species of the *Xenopodinae* subfamily.



Grey shading indicate presence of XNC10.1 / XNC10.2 and/or SNC10 α1 and α2 encoding exons

Table 4

Putative class Ib residues at conserved positions involved in terminal peptide anchoring.

residue[a]	domain	*XNC10.1*	*XNC10.2*	*SNC10*
Tyr7	α1	Tyr 7/7[b]	Tyr 4/4	Tyr 1/3 nd 2/3
Tyr59	α1	Asn 7/7	Leu 4/4	His 3/3
Tyr159	α2	Tyr 7/7	Tyr 4/4	Tyr 2/2
Trp167	α2	Glu 7/7	His 5/5	His 1/2 nd 1/2
Tyr171	α2	His 7/7	Tyr 5/5	Tyr 1/2 nd 1/2
Tyr/Arg84	α1	Ser 7/7	Ser 4/4	Ser 1/3 nd 2/3
Thr143	α2	Ala 7/7	Val 5/5	Ala 2/2
Lys146	α2	Lys 7/7	Lys 2/5 ; Arg 3/5	Lys 2/2
Trp147	α2	Arg 7/7	Arg 2/5 ; Trp 3/5	Trp 2/2

putative PBR of *Xenopodinae* XNC/SNC10 gene lineage were predicted by alignment with human HLA-A

[a]Numbering based on HLA-A2

[b]Indicate the number of species that have conserved amino acids at the indicated position/ total number of species analyzed

nd= not determined