European Journal of Immunology

The dynamic TCRδ: TCRδ chains in the amphibian *Xenopus tropicalis* utilize antibody-like V genes

Zuly E. Parra[1], Yuko Ohta[2], Michael F. Criscitiello[2], Martin F. Flajnik[2] and Robert D. Miller[1]

[1] Center for Evolutionary and Theoretical Immunology, Department of Biology, University of New Mexico, Albuquerque, NM, USA
[2] Department of Microbiology and Immunology, University of Maryland at Baltimore, Baltimore, MD, USA

The content and organization of the *Xenopus tropicalis* TCRα/δ locus was determined. This locus is highly conserved among tetrapods, with the genes encoding the TCRδ chains embedded with those encoding TCRα. However, the frog *TCRα/δ* is unusual in that it contains V genes that appear indistinguishable from those in the *IgH* locus (VH). These V genes, termed VHδ, make up 70% of the V genes at the TCRδ locus and are expressed exclusively in TCRδ chains. Finding TCRδ chains that use antibody-like V domains in frogs is similar to the situation in shark TCRδ variants and TCRμ in marsupials. These results suggest that such unconventional TCR may be more widespread across vertebrate lineages than originally thought and raise the possibility of previously unrealized subsets of T cells. We also revealed close linkage of *TCRα/δ*, *IgH*, and *Igλ* in *Xenopus* which, in combination with linkage analyses in other species, is consistent with the previous models for the emergence of these antigen receptor loci.

Key words: Amphibian · Evolution · Immunoglobulin · TCRδ

Supporting Information available online

Introduction

B and T lymphocytes using somatically diversified antigen-binding receptors mediate the adaptive immune responses in all jawed vertebrates [1]. These receptors contain antigen-binding V domains encoded by exons that normally exist in a segmented form in the germ-line DNA and are assembled by DNA recombination during development. In B cells this recombination takes place at the *Ig* loci-encoding molecules that serve both as cell-surface BCR as well as secreted antibodies. In the case of T cells, recombination occurs at the genes encoding the TCRα, β, γ, and δ chains, which form part of the cell surface αβ or γδ TCR complexes [2]. The genes encoding the Ig and TCR chains share similar organization and structural features consistent with their common ancestry [3]. The genes encoding the IgH, TCRβ, and TCRδ chains use V, D, and J gene segments to assemble and encode the V domain, whereas the IgL, TCRα, and TCRγ chains use just V and J. In all cases, these gene segments are flanked by conserved recombination signal sequences (RSS) that are site-specific targets of the endonuclease activity of RAG [4].

The genes encoding the TCRα and δ chains are unique among the loci undergoing V(D)J recombination in several ways. In all tetrapods examined so far, they are interspersed at a single *TCRα/δ* locus [5–9]. This single locus encodes two chains whose tightly regulated expression is mutually exclusive resulting in distinct T-cell lineages, the αβ and γδT cells [10, 11]. In most

Correspondence: Prof. Robert D. Miller
e-mail: rdmiller@unm.edu

cases, TCRα and δ chains share a common pool of V that, depending on the chain, are recombined to either a DJδ or directly to a Jα segment.

In addition to the complex genetics of the *TCRα/δ* locus, TCRδ appears to have a high degree of evolutionary plasticity. Approximately one quarter of the shark TCRδ chains are expressed in an alternative isoform called New Antigen Receptor (NAR)-TCR that contains a double V structure [12]. Interestingly, each of the two V domains require V(D)J rearrangement, and the *N*-terminal V is more similar to the V region of an antibody discovered in the nurse shark called the IgNAR than it is to TCR V. More recently, a novel TCR locus, *TCR*μ, was discovered in marsupials that shares an evolutionary history with TCRδ [13]. TCRμ also encodes a double V structure and both V genes are also more similar to VH than to TCR V. Marsupials also have a conventional *TCRα/δ* locus with a prototypic mammalian organization and, therefore, TCRμ is not a substitute for TCRδ in these mammals [9]. However, the C regions of TCRμ do share greatest sequence similarity to Cδ and appear to have been derived from TCRδ, perhaps during the early evolution of amniotes [13]. TCRμ is also found in the duckbill platypus, consistent with its ancient origins and presence in the common ancestor of all living mammals, and thus an orthologue could still be present in some eutherian (placental) mammals, although so far none have been found [9].

The presence of atypical TCRδ forms with similar features in distantly related species such as cartilaginous fish and non-eutherian mammals suggests that they may be found in other vertebrate lineages. So far, surveys of the chicken, lizard, and frog genomes failed to uncover any gene sequences bearing homology to TCRμ [9] (ZEP and RDM personal observations). However, when investigating the genome of an amphibian, *Xenopus tropicalis*, for evidence of TCR resembling either TCRμ or NAR-TCR, evidence for a third alternative, a variant of TCRδ, was obtained that may give insight into both the emergence and the evolution of Ig and TCR in general and the origins of TCRμ.

Results

Conserved synteny at the X. *tropicalis TCRα/δ* locus

As in all tetrapod species analyzed so far, the genes encoding the *X. tropicalis* TCRα and TCRδ chains are tightly linked, with some TCRδ genes nested among the TCRα (Fig. 1). This genomic region appears stable in tetrapods since the genes flanking the *X. tropicalis TCRα/δ* locus are the same as in birds and mammals, including the olfactory receptors interspersed among the Vα genes (Fig. 1) [5, 7, 9]. Individual V, D, and J gene segments in the *X. tropicalis TCRα/δ* locus were annotated using the convention established by the International ImMunoGeneTics database (http://www.imgt.org) and the recommendations of Koop *et al.* [14]. A total of 71 V gene segments were identified within the *TCRα/δ* locus, many of which share a high degree of sequence identity to those previously reported in *Xenopus* [15] (Supporting Information Table 1). Fifty-two V gene segments, in the same transcriptional orientation as the most 3′ C region (Cα1) are Vα based on nucleotide identity (Figs. 1 and 2). All Vα appeared to be functional based on open reading frames (ORF), an upstream exon encoding a leader sequence, and a canonical RSS. They segregate into 28 subgroups designated Vα3 through Vα30 based on nucleotide identity and phylogenetic relationships (Fig. 2). Also present is a large number of Jα gene segments upstream of Cα1 similar to all mammalian species investigated (75 in the frog compared with 61 in humans, 60 in mice and 53 in the opossum) (Fig. 1 and Supporting Information Fig. 1) [9, 16].

To be consistent with the published *Xenopus* TCRα sequences, the 3′ most Cα was designated as Cα1 and corresponds to a sequence described previously [15] (Supporting Information Fig. 2). Upstream, within the Vα genes, is a second Cα (Cα2) that shares only 51% nucleotide identity with Cα1. Cα2 appears functional by having an ORF and canonical cysteine residues necessary to form the intra-chain disulfide bond, however it has two, three codon insertions located 5′ of each cysteine and two extra cysteines in the β-strand, all of which could affect protein folding (Supporting Information Fig. 2). Attempts to isolate TCRα transcripts containing Cα2 using either 5′ RACE or targeted RT-PCR were unsuccessful using either the cDNA library or the RT-PCR on small intestine, spleen or thymus RNA (data not shown). Furthermore, neither of the two Jα gene (Jα2.1 and 2.2) upstream of Cα2 was identified in transcripts (data not shown).

Within the Vα gene segments is a Cδ (Cδ2) in the same orientation as both Cα (Figs. 1 and 3, and Supporting Information Fig. 2). Transcripts containing Cδ2 were not detectable by Northern blot analysis of various tissues including thymus RNA, but could be amplified by 5′ RACE using thymus cDNA (data not shown). All Cδ2 transcripts amplified contained either Vα4 or Vα5, which are the two of the four V genes located immediately upstream of Cδ2 and in the same orientation (Fig. 1, GenBank accession numbers GQ262036–GQ262039). A comparison of the CDR3 sequences from these clones to the germ-line sequence failed to identify sequences corresponding to D segments (data not shown). In addition, all clones amplified contained the same J segment (Jδ2.1). However, searching the genomic sequence for conserved FR4 (*e.g.* FGXG) and RSS motifs identified the second Jδ (Jδ2.2) (Fig. 1 and Supporting Information Fig. 1). Although Cδ2 and Cα2 were found expressed at low levels or not at all they constitute a mini-locus (from Vα3 to Cα2) resembling the organization found in conventional mammalian *TCRα/δ* loci with V segments upstream of a Jδ-Cδ-Jα-Cα cluster. Whether this mini-locus represents an ancient or derived *TCRα/δ* organization remains to be determined.

An unconventional TCRδ

At the 5′ end of the *Xenopus TCRα/δ* locus is a cluster of gene segments in reverse transcriptional orientation, encoding a TCRδ

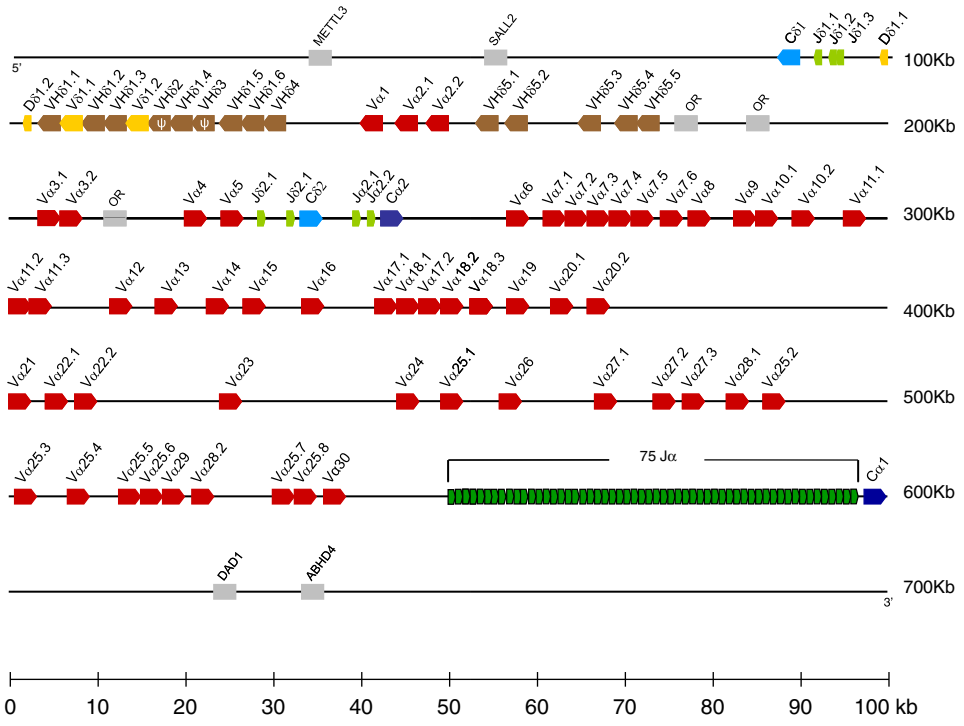

Figure 1. Representation of the X. *tropicalis TCRα/δ* locus. TCR Vα (red), Vδ (yellow), VHδ (brown), Jδ (light green), Jα (dark green), and Dδ (orange) gene segments and two Cα (dark blue) and two Cδ (light blue) were numbered according to their position in the locus or to prior description. Transcriptional orientation is demonstrated by the direction of the arrow in each segment. Syntenic genes shown in gray are those conserved with mammalian *TCRα/δ* locus: *methyl-transferase like 3* (*METTL3*), zinc finger protein (*SALL2*), olfactory receptors (*OR*), *defender against cell death gene 1* (*DAD1*), and *abhydrolase domain-containing protein 4gene* (*ABHD4*).

chain (Fig. 1). A single Cδ (Cδ1), the most 5′ TCR gene in the locus, shares 61% nucleotide identity with Cδ2 (Figs. 1 and 3 and Supporting Information Fig. 2). Present are 19 V genes of which three are Vα (Vα1, Vα2.1, and 2.2) and two are Vδ (Vδ1.1 and 1.2) (Figs. 1 and 2). The three Vα genes belong to two subgroups (Vα1 and Vα2) sharing 63% nucleotide identity and fall into the same phylogenetic clade as *X. tropicalis* Vα subgroups 10 and 11 (group I, Fig. 2). Phylogenetic analyses revealed that the Vδ fall within a well-supported clade with opossum Vδ2 and Vδ3, human Vδ2 and mouse Vδ4 (group F, Fig. 2). *X. tropicalis* Vδ gene segments share the highest similarity with opossum Vδ3 (45.9%) when compared with mammalian and avian Vα/δ sequences. *X. tropicalis* Vδ genes also have an unpaired cysteine residue located in the e-strand of FR3 (Fig. 4). Some of the Vδ gene segments encoding supporting V domains in shark NAR-TCR also contain a cysteine in the e-strand (GenBank accession no. DQ022688 and DQ022691) [12].

The 14 remaining V genes were compared with Ig and TCR V genes from other vertebrates and found to be most similar to VH based on nucleotide identity and phylogenetic analyses. These genes have been designated VHδ due to their similarity to VH but their location in the *TCRα/δ* locus and exclusive expression with Cδ1 (Figs. 1, 5, 6). VHδ genes segregated into five sub-groups based on 80% or greater nucleotide identity (Fig. 1 and Supporting Information Table 2). Two VHδ (VHδ2 and VHδ3) are pseudo-genes due to in-frame stop codons; however, the rest appear functional. In phylogenetic analyses, the *Xenopus* VHδ all clustered with clan II VH gene segments (Fig. 5). The VHδ genes form two distinct clades, one containing VHδ1 through 4 and the other VHδ5 only (Fig. 5). VHδ1 through 4 share less than 60% nucleotide identity with the *Xenopus* clan II VH subgroups, VH5, 9 and 11 [17, 18]. In contrast, the clade that includes VHδ5 also contains VH genes from *X. laevis* and *X. tropicalis* (*i.e.* V that have been identified in IgH transcripts). Indeed, VHδ5 genes in the *TCRα/δ* locus share a surprisingly high degree of identity (91–96%) with VH5 genes in the *IgH* locus, consistent with the VHδ5 genes being paralogous to VH5 genes.

Upstream of Cδ1 are three Jδ genes identified by first searching the *TCRα/δ* genomic region for conserved RSS and investigating flanking regions for sequences encoding conserved TCR Jδ motifs (Fig. 1). Both the Jδ were confirmed in transcripts containing Cδ1 (Fig. 6).

Expression of Cδ1 with Vα, Vδ, and VHδ

To determine which V genes are used in Cδ1 transcripts, 5′ RACE was performed. As expected, the Vα and Vδ genes oriented upstream of Cδ1 were recombined and transcribed with Cδ1 (Fig. 6, Table 1). Transcripts containing Vδ1 genes could be amplified from the cDNA library, as well as small intestine, spleen, and thymus RNA individually (Table 1, Fig. 6). Cδ1 transcripts containing Vα2 were amplified from the cDNA library, small

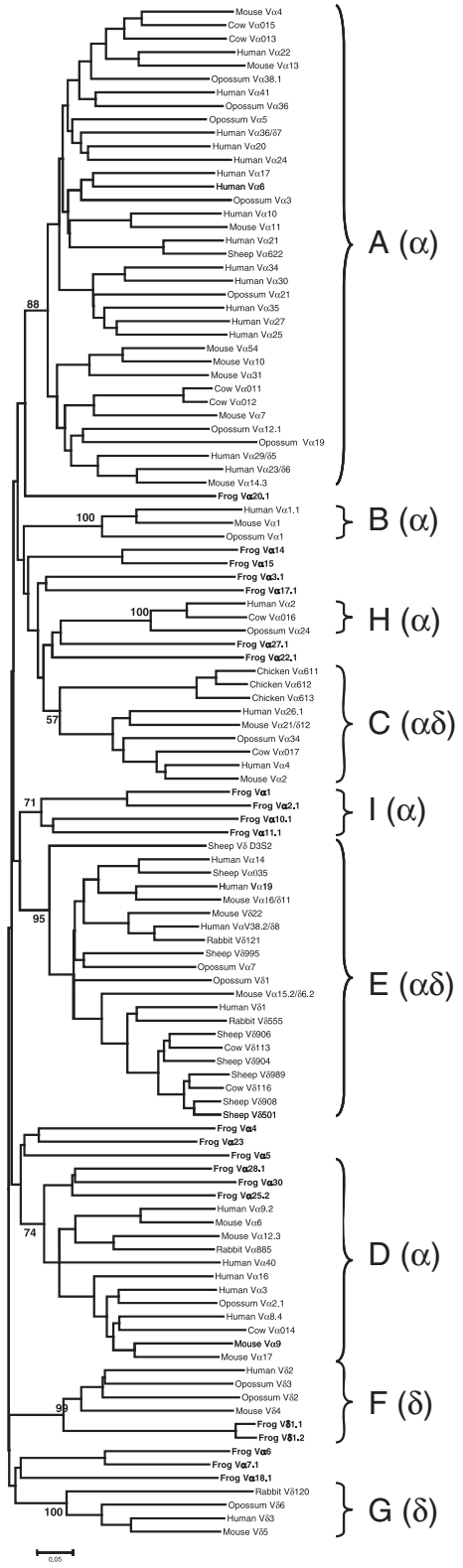

Figure 2. Phylogenetic tree of *X. tropicalis* Vα and Vδ compared with mammalian and avian genes using the neighbor joining method. Similar results were found using the minimum evolution method. *X. tropicalis* sequences are shown in bold. Vα and Vδ genes fall into nine groups that are indicated by brackets. Bootstrap values shown are based on 1000 replicates. A distance bar is shown below the tree.

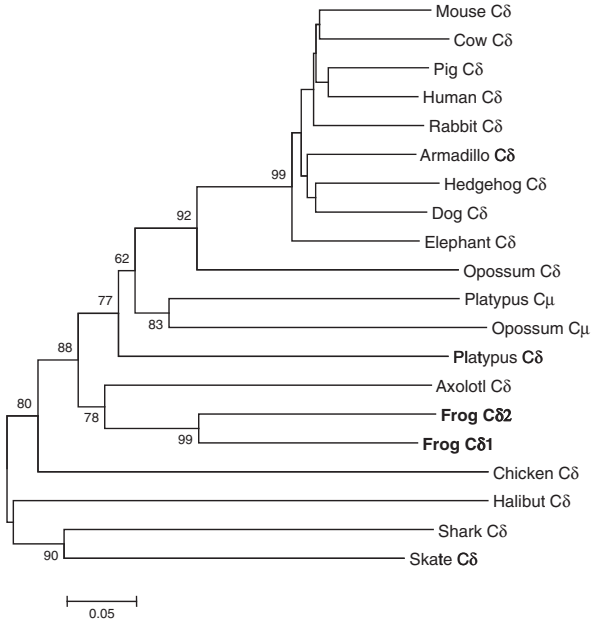

Figure 3. Phylogenetic analyses of *X. tropicalis* TCRδ constant regions. The tree shown includes TCRμ sequences from platypus and opossum. Alignments were analyzed using the minimum evolution distance method. Similar results were found using the neighbor joining method. Bootstrap values shown are based on 1000 replicate samples. The distance bar is shown under the tree.

intestine, spleen, and thymus, whereas Vα1 transcripts were found only in the cDNA library and small intestine (Table 1).

Significantly, Cδ1 transcripts containing recombined VHδ genes were also amplified by 5′ RACE from all tissues (Fig. 6, Table 1). Only VHδ1, 4 and 5 were amplified, consistent with these three being functional (Fig. 1, Table 1). These results confirm that *Xenopus* TCRδ chains can be expressed with VH-like V genes, as well as Vα and Vδ.

Using nucleotide sequences corresponding to CDR3 from Cδ1 transcripts, two Dδ (Dδ1.1 and Dδ1.2) genes were identified (Fig. 1). Transcripts containing Vα, Vδ, and VHδ all used these same two Dδ, which are similar in length (12 and 11 bp, respectively) and share a region of micro-homology, three nucleotides in length (Fig. 6 and 7). This micro-homology appears to facilitate recombination in clones lacking P or N nucleotides in the junction between D segments (Fig. 6). Both Dδ have ORF in all three reading frames and all three are used (Fig. 7 and data not shown). However, there was a slight bias toward the first reading frame of Dδ1.1 (12 out of 26 clones) and third reading frame Dδ1.2 (10 out of 25 clones) both of which encode a tyrosine and two hydrophobic amino acids (Fig. 7). The Dδ are flanked by RSS with 12 bp spacers on the 5′ side and 23 bp on the 3′ side, as is typical of TCR D (Fig. 7) [19]. However, the RSS are not well conserved, having four different heptamer sequences (Fig. 7). This asymmetrical RSS organization facilitates the use of multiple D in individual recombinants, as has been described previously for TCRδ chains [13, 19]. The majority of Cδ1 transcripts (24 out of 32) encoded CDR3 using both Dδ genes irrespective of the V or J genes being used (Fig. 6). The remaining eight clones contained only a single Dδ. Four were Dδ1.1, three were Dδ1.2, and one was

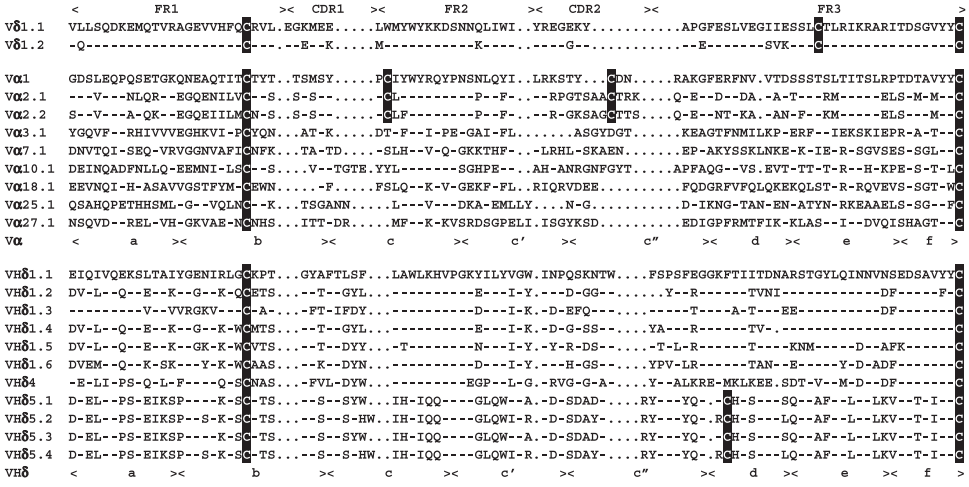

Figure 4. Alignment of predicted amino acid sequences of *X. tropicalis* Vδ, representative Vα, and VHδ regions. Canonical cysteine residues involved in intrachain disulfide bonds and extra cysteine residues potentially used for interchain interactions are highlighted in black. Frameworks and complementarity determining regions are indicated above the alignment. β-Strands are indicated under the alignment.

indeterminate due to excessive trimming. Therefore, when single Dδ are used there does not seem to be a preference.

It is noteworthy that all of the *X. tropicalis* TCRδ clones characterized encoded single V and C domains, irrespective of the type of V being used. There is no evidence for *X. tropicalis* TCRδ chains having three extracellular Ig domains, two V and one C, as has been found for shark NAR-TCR and mammalian TCRμ [12, 13].

The *TCRα/δ* and *IgH* loci are linked in *Xenopus*

The evidence for *Xenopus* TCRδ using VH-like genes raises questions regarding linkage between *TCRα/δ* and *IgH* in frogs. This was investigated by RFLP analysis for Cδ2, VH, and Cλ genes using pedigreed *X. laevis*. All loci analyzed in *X. tropicalis* share the same linkage patterns as *X. laevis*, making the latter a suitable proxy for this analysis [20]. Complete co-segregation was found for Cδ and VH in all offspring, suggesting a very tight linkage between *TCRα/δ* and *IgH* (Fig. 8). This result was confirmed using a housekeeping gene, *SLC7A7*, adjacent to IgF on scaffold 972 from the *X. tropicalis* genome, which also co-segregated with TCRδ (Fig. 8). Although *IgF* is not present on the same scaffold as the other four IgH isotypes in version 4.1 of the *X. tropicalis*, it is located directly downstream of *IgY* (YO and MFF personal observation). Co-segregation with *Igλ* was also detected with the exception of two (10%) recombinants, consistent with *Igλ* also being linked to *IgH/TCRα/δ*. These results are consistent with the linkage map at http://tropmap.biology.uh.edu where *Igλ*, *TCRδ*, and *IgH*, in that order, are on linkage group 1, corresponding to *X. tropicalis* chromosome 1.

No evidence for trans-locus somatic recombination between *IgH* and *TCRδ* in *X. tropicalis*

Linkage between *Xenopus IgH* and *TCRα/δ* raised the possibility of the use of a common pool of V genes. To investigate this, VHδ were used to search *Xenopus* expressed sequence tag (EST) databases for evidence of their expression with different C genes. A single Cδ1 *X. laevis* EST (EB475751) containing a VHδ1 was identified (VHδ751 in Fig. 5). All other EST from both *X. tropicalis* and *X. laevis* containing V sequence with greater than 90% nucleotide identity to VHδ5 contained FR4 with the WGXG motif conserved in JH genes, and several contained sufficient C region sequence to confirm that they were frog IgM or IgX transcripts, not TCRδ (Supporting Information Table 2). These EST also shared greater nucleotide identity with VH genes located on scaffold 1168 in the *X. tropicalis* genome assembly. For example, EST CF592362 is 100% identical to a V on scaffold 1168, but only 95.6% identical to VHδ5.1. A detailed analysis of scaffold 1168 revealed that it contains at least 20 VH genes and, along with scaffold 928, represents part of the *IgH* locus [21, 22]. Therefore, it appeared that these EST are not TCRδ, but rather are IgH transcripts.

To directly test whether VHδ are used in IgH transcripts, primers specific for VHδ1 and five genes were paired with a primer specific for *X. tropicalis* IgM C region and PCR was performed using the spleen cDNA library as a target. Products were obtained only when primers specific for VHδ5 subgroup were used and were found to encode V sequences more related to the VH5 genes on scaffold 1168 (96–99% nucleotide identity) than to VHδ5 genes (91–94% nucleotide identity). The products also contained FR4 sequence more similar to JH (encoding the conserved WGXG motif) than to J genes in the *TCRα/δ* locus. In conclusion, although the *Xenopus IgH* and *TCRα/δ* loci are linked and contain related V genes, there is no evidence for somatic trans-rearrangements between the loci.

Discussion

For much of the past two decades, the story of TCR structure and genetics has been one of conservation. This was recently

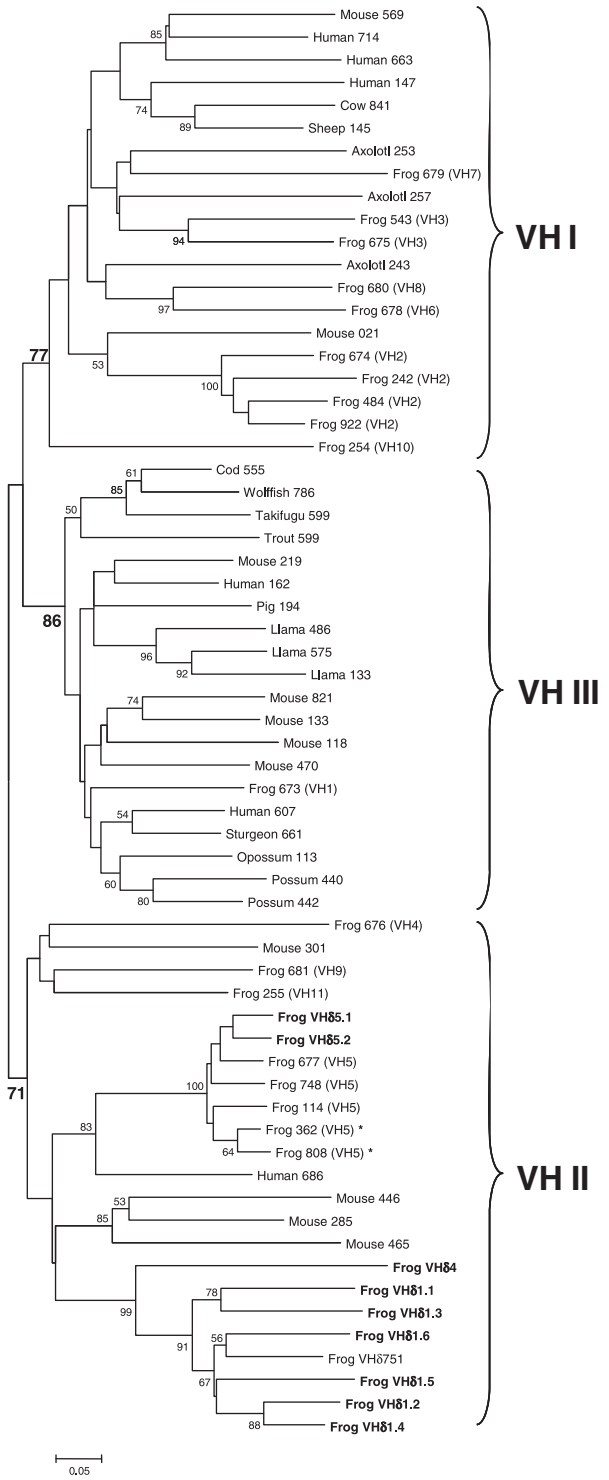

Figure 5. Phylogenetic tree showing the *X. tropicalis* VHδ gene segments compared with Ig VH gene segments from different species. *X. tropicalis* VHδ gene segments are indicated in bold. The VHδ pseudogenes, VHδ2 and VHδ3, are not included in this tree but cluster with VHδ1 and VHδ4 subgroups when included. Frog VH sequences indicated with an asterisk are those from *X. tropicalis*, with the remaining being from *X. laevis*. Three distinct VH clans are indicated by brackets. The tree was generated using the minimum evolution method. Similar results were found using the neighbor joining method. Numbers at the nodes are bootstrap values based on 1000 replicates. A distance bar is shown at the bottom of the tree.

illustrated when the genomic organization of the sandbar shark *TCRγ* locus was found to have the translocon organization similar to that of mammals [23]. Furthermore, the genomic region containing *TCRα/δ* has been relatively stable over the 340 million years of evolution separating amniotes from amphibians [7, 9, 24]. This made all the more surprising the discovery of VH-related genes within the *Xenopus TCRα/δ* locus, a characteristic not found previously in any tetrapod. It should be stressed that the VHδ genes are not orthologues of VH genes in the *IgH* locus [17, 18]. Rather, the results support that are they are paralogues of VH genes expressed exclusively in TCRδ chains.

The discovery of NAR-TCR in sharks and TCRμ in marsupials provides precedence for TCR chains using antibody-like V genes [12, 13]. One question is whether the characteristics of mammalian TCRμ, frog TCRδ and shark NAR-TCR are the result of convergent evolution or homology by descent. This can be addressed by considering the likely origins of VH-like genes in the *TCRδ* locus. It is noteworthy that *Xenopus* is one of the genera with a diverse pool of VH genes at their *IgH* locus. *Xenopus* has retained all three ancient VH clans (I, II, and III) and these VH genes segregate into eleven subgroups [17, 18]. This level of diversity is in contrast to many species such as chicken, platypus, opossum, and rabbit that have only single VH subgroup, most often clan III [25–27]. The *X. tropicalis* VHδ are most related to clan II VH. Platypus TCRμ V genes, on the other hand, are related to VH clan III. In contrast, the V genes used in marsupial TCRμ fall outside the clan I, II, or III designations, and are from a sister clade to the VH [13]. Alternatively, shark NAR-TCR create the *N*-terminal V domain using V genes related to IgNAR, a type of antibody unique to cartilaginous fish [12, 13]. Therefore, the V genes being used in each case appear independently derived, consistent with the convergent mode of evolution.

The biological significance of *Xenopus* TCRδ using VH-like genes related to clan II VH (VH5, 9 and 11) is unknown. The VH clans appear to evolve at different rates with clan III being more conserved and widespread [28]. However, it is noteworthy that, in *Xenopus*, B cells expressing clan II VH appear later in ontogeny [29]. A question to be addressed is whether there is a similar late expression of VHδ genes in developing *Xenopus* γδT cells. This would seem unlikely since VH and VHδ are distinct sets of V genes in separate loci and expressed in distinctly different lymphocyte lineages. Furthermore, the majority of V genes available for rearrangement and expression of TCRδ chains are VHδ. Whether they ultimately comprise the majority of γδT cells and the order of their appearance in frog ontogeny both remain to be determined.

If the V genes used in shark NAR-TCR, frog TCRδ, and marsupial TCRμ represent convergent evolution then the origin of VHδ genes in the *X. tropicalis TCRα/δ* locus needs to be considered. The relationship between *X. tropicalis* VHδ and clan II VH is consistent with a block duplication inserting multiple VH into the *TCRα/δ* locus. VHδ5 genes in particular share a high degree of similarity with VH5 genes in the *IgH* locus, consistent with either a recent introgression into the *TCRα/δ* locus or a more ancient duplication that has been subject to

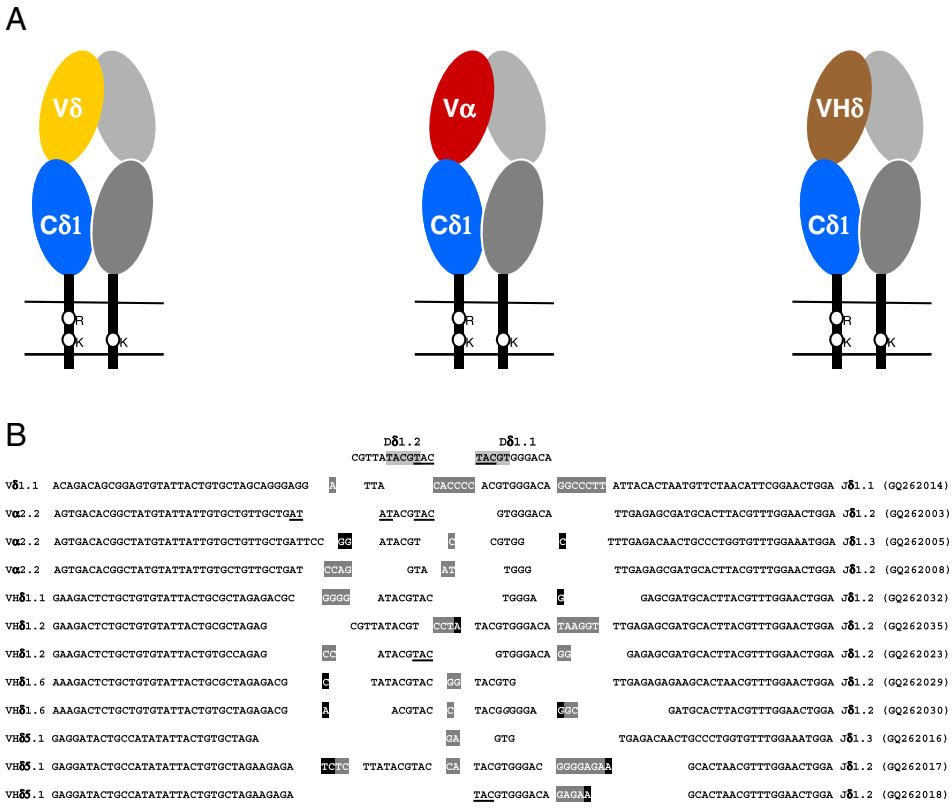

Figure 6. (A) Diagram of expressed isoforms of *X. tropicalis* Cδ1. (B) CDR3 nucleotide alignment of representative sequences. Germline D sequences are labeled on top of the alignment. Vα, Vδ, and VHδ gene segments are indicated on the left–hand side and J gene segments on the right. Highlighted in gray is a 5 bp sequence present in both Dδ segments. Microhomologies in the germline and CDR3 regions are underlined; P and N nucleotides (shown in white letters) are highlighted in black and gray, respectively. GenBank accession numbers are shown in parenthesis on the right.

Table 1. Cδ1 transcripts amplified from different *X. tropicalis* tissues

	Vα		Vδ		VHδ	
Tissue/subgroup	Vα1	Vα2	Vδ1	VHδ1	VHδ4	VHδ5
cDNA library	+	+	+	+	NF[a)]	+
Small intestine	+	+	+	+	NF	+
Spleen	NF	+	+	+	NF	+
Thymus	NF	+	+	+	+	+

[a)]Not found.

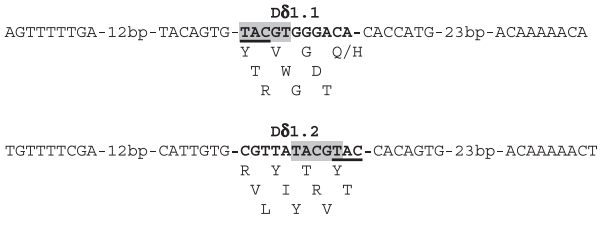

Figure 7. The germline *X. tropicalis* Dδ gene segments. Highlighted in gray is a 5 bp sequence present in both Dδ segments. Reading frames and RSS of the two Dδ gene segments found recombined with Vα, Vδ, and VHδ.

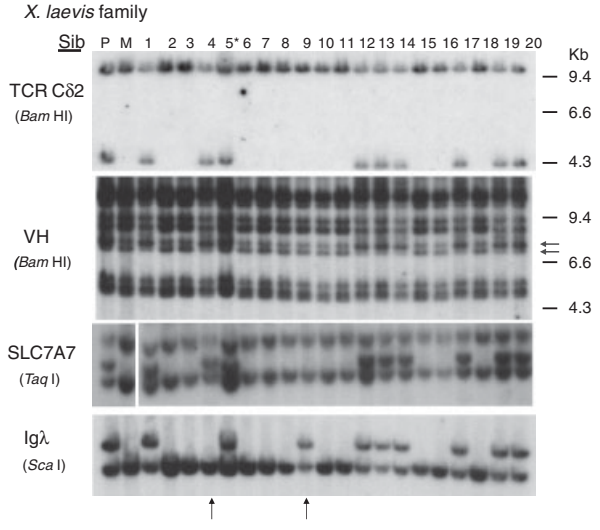

Figure 8. Close linkage of TCRα/δ, IgH, and Igλ loci in the *Xenopus* genome. Probes and restriction enzyme giving the RFLP are indicated on the left. Arrows on the right indicate polymorphic bands with VH that segregate with the bottom band of TCRδ. Arrows below indicate recombinants (siblings ♯4 and 9) between TCRδ and Igλ. Size markers are indicated on the right. The bands for SLC7A7 and Igλ are found between 2.3 and 4.3 kb marker bands. Sibling ♯5 (asterisk) is a natural triploid. P, paternal; M, maternal.

ongoing nonhomologous recombination or gene-conversion that preserved the high-sequence similarity. Given the close linkage between *IgH* and *TCRα/δ* in *Xenopus*, such nonhomologous recombination is highly likely. The similarity between VHδ5 and VH5 genes and their designation as the respective "fifth" subgroup in each locus is purely coincidental. VH5 were identified as the fifth new subgroup among a collection of *Xenopus* IgH transcripts [18]. VHδ5, on the other hand, was annotated as the fifth subgroup in linear order in the *TCRα/δ* genomic sequence using conventional TCR nomenclature (*Materials and Methods*). In contrast to VHδ5, the VHδ1 through four subgroups are less similar to any of the known *Xenopus* VH subgroups, sharing less than 60% nucleotide identity with the latter. This suggests that either these VHδ are older introgressions of VH genes into the *TCRα/δ* locus and were later lost in the *IgH* locus, or are genes that have diverged at a higher rate than VHδ5.

These results also bear on the link between γδTCR and IgH noted previously, including recent evidence that TCRγ genes in sharks may undergo somatic mutation similar to that of Ig [23]. In mice and humans, Vδ are more similar to VH in CDR3 length [30]. Ig VH and TCR Vδ both have CDR3 heterogeneous in size and associated with partners (IgL and TCRγ, respectively) with generally shorter CDR3. In contrast, the CDR3 of both Vβ and Vα are relatively similar, presumably due to the constraints of MHC restriction. These observations led to proposals that γδT cells recognize antigens differently than αβT cells, such as recognizing free, unprocessed antigens [30]. Indeed, while conventional αβT cells are MHC restricted, some γδT cells have been shown to respond in an MHC un-restricted manner to unprocessed viral antigens [31]. Furthermore, so-called "innate" γδT cells with limited TCR diversity in mice and humans can recognize self-molecules such as the MHC-class I chain-related MICA (in humans) and RAE-1 (in mice) or metabolic by-products such as isoprenyl-phosphate [32]. If TCR with antibody-like V domains are products of convergence, it is likely antigen recognition is driving this evolution. We predict that such T cells, including *Xenopus* γδT cells using VHδ, may bind microbial or other pathogen-associated epitopes in a manner more like B cells. In the case of γδTCR, the δ chain would be performing the role similar to H chain in antibodies, where the H chain often takes a predominant role in antigen recognition [33]. It seems unlikely that shark NAR-TCR$^+$ T cells or mammalian TCRμ$^+$ T cells would be the equivalent of the "innate" γδT cells in eutherians, given the diversity of their receptors. Rather, this model provides some explanation for the consistent finding that, over evolutionary time, the *TCRδ* locus has demonstrated extraordinary plasticity, and we predict that the unusual TCRδ chains in sharks and frogs, and TCRμ in marsupials and monotremes, are involved in adaptive immunity.

The capacity to bind free antigen in solution may have provided a strong pressure for rapid evolution of the VH repertoire, resulting in the clans and families that emerged early in the Ig. Perhaps what is being observed in sharks, frogs, and non-eutherian mammals has been the TCRδ taking advantage of this unique VH diversity and incorporating gene segments into its own repertoire, which in mammals resulted in the evolution of a separate locus, TCRμ. Why this has not been found in any eutherian mammal, such as humans and mice, is not clear [9]. It is possible that γδT cells entirely satisfy the role of direct antigen recognition in eutherians.

An atypical feature shared by NAR-TCR and TCRμ is the expression of an extracellular form containing three Ig-superfamily domains [12, 13]. There was no evidence for *Xenopus* TCRδ chains expressed in a double V form, nor does it appear that the *Xenopus TCRα/δ* locus could encode such a form. Double Vs require tandem array of rearranging V, D, and J segments as has been found in shark NAR-TCR and the platypus TCRμ homologue or, alternatively, the presence of a second germ-line joined V, as is the case for marsupial TCRμ [12, 13]. Although transcripts encoding double V were not found for *X. tropicalis* TCRδ, many of the V segments expressed with Cδ1 encode un-paired cysteine residues in positions similar to that of VNAR and the supporting Vδ that are used in the NAR-TCR (Fig. 4). These cysteines are presumed to form inter-domain disulfide bonds that help to stabilize the double V structure. What role they may play in *Xenopus* TCRδ chains remains to be determined. None of the other V genes in the locus contained unpaired cysteines, although the majority of the Vα expressed with Cα1 and are not used in the TCRδ chain. In addition, such unusual V genes also raise the obvious question of whether all of the *Xenopus* TCRδ chains are associated with TCRγ or are found in other complexes. This is a question that remains to be answered for NAR-TCR and TCRμ as well.

The common origin of Ig and TCR is beyond doubt when considering gene organization, conserved RSS, and common enzymatic machinery mediating somatic recombination [34]. Although many of the characteristics in *Xenopus* TCRδ, NAR-TCR, and TCRμ appear to be the result of convergent evolution, the structure of *Xenopus TCRα/δ* locus, and the linkage analyses are consistent with prevailing models of the origin and evolution of Ig and TCR genes [35, 36]. For example, the inverted genomic region containing VHδ and Cδ1 is reminiscent of the duplication and inversion event predicted to give rise to the current *TCRα/δ* locus configuration in mammals [9, 35]. The tight linkage of *IgH*, TCRα/δ, and *Igλ* in *Xenopus* is consistent with what is found in many species, arguing for a common ancestral arrangement [35]. This synteny is consistent with a model where the genes encoding both chains of an ancestral antigen receptor heterodimer were linked and then duplicated *en bloc* to create two gene complexes [35]. If the original antigen receptor was TCR-like, then the two complexes, βγ and αδ, may have been derived from the second genome-wide duplication believed to have occurred early in vertebrate history [37]. The αδ complex has remained linked in the tetrapods and some fish species, whereas βγ has been separated [35, 38, 39]. The presence of an ancestral Igλ chain in the original complex is parsimonious, since it shares the RSS with a 23 bp spacer type with all other antigen receptor V genes. Different lineages of receptor genes would have been free to independently evolve their own RSS configurations, most likely through inversions, resulting in the symmetrical 12 bp spacer RSS flanking DH genes

and the 12 bp spacer RSS flanking V genes in the *Igκ* and *Igσ* loci [36, 40].

In conclusion, NAR-TCR in sharks, TCRμ in noneutherian mammals, and now TCRδ in *Xenopus* comprise what appears to be a growing list of TCRδ or δ-like chains utilizing antibody-like V domains. How long this list of species will become is unknown; however, these observations may reveal a new mechanism for antigen recognition by T cells that has remained undiscovered because of its absence in commonly studied eutherian mammals.

Materials and methods

Genome screening and identification of V–D–J gene segments

Version 4.1 of the *X. tropicalis* genome assembly was analyzed using the BLAST algorithm at www.ncbi.nlm.nih.gov/BLAST along with the *Xenopus* EST BLAST server at http://www.sanger.ac.uk/ and ENSEMBL at www.ensembl.org. Known Vα and Jα gene segments and the Cα1 genes were located using available sequences [15, 41]. Novel Vα, Jα, and Cα genes and all Vδ, Dδ, Jδ, and Cδ genes were identified by similarity to homologues from other species. Conserved RSS were also used to identify V, D, and J gene segments. The location for all *TCRα/δ* genes in version 4.1 of the *X. tropicalis* assembly is provided in Supporting Information Tables 1 and 2.

Annotation of the *X. tropicalis* TCRα/δ and VH regions

The *TCRα/δ* locus is located on scaffold number 539 (length 791 252 bp). V and D gene subgroups and individual gene segments were annotated in 5′–3′ order in the genomic sequence in accordance with the International ImMunoGeneTics nomenclature (http://imgt.cines.fr/). For example, Vα2.1 refers to the most 5′ Vα in subgroup 2. J genes were numbered 3′–5′ also according to the established nomenclature [14]. Cα1 was identified as being identical to Cα genes identified previously in *X. laevis* [15]. Cα2 and both Cδ were identified in the genomic sequence as being a C region gene with similarity to homologues in other species. The two Cδ were numbered based on Cδ1 being more frequently isolated among TCRδ transcripts (data not shown). Scaffold 1168 containing VH genes most closely resembling the VHδ found on scaffold 539 was also characterized. The V genes found on this scaffold are utilized in IgH transcripts and this genomic sequence is likely part of the *IgH* locus in *X. tropicalis* (see *Results* section).

RNA extraction, cDNA synthesis, and RT-PCR

The combination intestine, liver, spleen, and thymus cDNA library used was constructed using the Orient Express Oligo(dT) cDNA library kit (Novagen, Madison, WI) [42]. Total RNA was extracted from small intestine, spleen, and thymus from *X. tropicalis* as described previously and under the approval of animal protocol number 1009003 to MFF [42]. Total RNA was treated to remove DNA contamination using the TURBO DNA-free kit (Ambion, Austin, TX) and then used for RT-PCR. RT-PCR were performed using GeneAmp RNA PCR core kit (Applied Biosystmens, Foster City, CA). PCR were performed using the AdvantageTM-HF-2 kit using manufacturer-recommended conditions (BD Biosciences, CLONTECH Laboratories, Palo Alto, CA). PCR products were cloned using the TopoTA cloning® kit (Invitrogen, Carlsbad, CA) and sequenced using the BigDye Terminator Cycle Sequencing kit v3 (Applied Biosystems, Foster City, CA). Sequencing reactions were analyzed on the ABI Prism 3100 DNA automated sequences (PerkinElmer Life and Analytical Sciences, Wellesley, MA). Chromatograms were analyzed using the SequencherTM 4.3 program (Gene Codes, Ann Arbor, MI).

Primers for Cδ1 and Cδ2 semi-nested PCR and 5′ RACE were Cδ1 outside primer, 5′-GCGCTGTACGTTCCATTCAGGGAG-3″; Cδ1-nested primer: 5′-CCAGACAGGCAGTGGGAGAGTCACC-3′; Cδ2 outside primer: 5′-GAAAGTACTGGGGTAACATCTCCC-3″; Cδ1-nested primer: 5′-CAGGCAGCAGGAGACAGACCCTTC-3″.

To investigate whether VHδ genes are utilized in IgH chains, VHδ1 and VHδ5 primers were paired in PCR with a primer specific for *Xenopus* IgM C region using the spleen cDNA as target. IgM C region primer: 5′-AATGGCCAGAATGCTGATTTGAG-3′; VHδ1.2 and VHδ1.3 primer: 5′-GTTCCTGGAAAAGAGATTCT GTAC-3′; VHδ1.5 and VHδ1.6 primer: ′-CTTGGCTGAAA-CATGTTCCTGG-′; VHδ5.1, VHδ5.2, VHδ5.3, VHδ5.4, and VHδ5.5 primer: 5′-GTCCTGCAAGACATCAGGGTATTC-3′.

Phylogenetic analyses

The GenBank accession numbers of all sequences used are in Supporting Information Table 3. Nucleotides corresponding to FR1 through FR3 of all V genes and exon 1 of all C genes were aligned using BioEdit [43, 44]. Phylogenetic trees were constructed using the neighbor joining and minimum evolution methods with 1000 bootstrap replicates using MEGA version 3.0 [45].

Linkage analyses

Erythrocyte genomic DNA from 20 *X. laevis* siblings was digested with various restriction enzymes and blotted onto nitrocellulose membranes (Whatman, ME). Blots were hybridized under high stringency conditions with ^{32}P-dCTP radiolabeled probes and exposed to the X-ray film [46]. Probes used for the experiments include the exons encoding the Cδ2 domain and 3′ UTR, the C domain of Igλ, and a V domain from the VH1 subroup in the *IgH* locus (GenBank NM_001086404) [17, 18]. VH1 was chosen to type the *IgH* locus since it shares less than 60% nucleotide identity to any VHδ and, therefore would not

cross-hybridize. To further confirm Ig-TCR linkage, the SLC7A7 gene near IgF in scaffold-972 was used as probe, specifically nucleotide position 140–1537 of *X. laevis* SLC7A7 (BC072040) [21, 22]. Tetraploid *X. laevis* was used for this analysis due to the availability of pedigreed families for this species [47].

Acknowledgements: Support was provided by National Institutes of Health (IP20RR18754) and the National Science Foundation (IOS-0641382) to R. D. M., and the National Institutes of Health (RR00603) to M. F. F.

Conflict of interest: The authors declare no financial or commercial conflict of interest.

References

1 **Davis, M. M. and Chein, Y. H.,** T cell antigen receptors. In Paul W. E., (Ed.), *Fundamental Immunology*, 6th Edn, Lippincott, Philadelphia 2008, pp. 313–345.

2 **Saito, T.,** T cell antigen receptor, structure, expression and function. *Hum. Cell* 1990. **3**: 183–192.

3 **Litman, G. W., Anderson, M. K. and Rast, J. P.,** Evolution of antigen binding receptors. *Annu. Rev. Immunol.* 1999. **17**: 109–147.

4 **Chien, Y. H., Gascoigne, N. R., Kavaler, J., Lee, N. E. and Davis, M. M.,** Somatic recombination in a murine T-cell receptor gene. *Nature* 1984. **309**: 322–326.

5 **Satyanarayana, K., Hata, S., Devlin, P., Roncarolo, M. G., De Vries, J. E., Spits, H., Strominger, J. L. *et al.*,** Genomic organization of the human T-cell antigen-receptor α/δ locus. *Proc. Natl. Acad. Sci. USA* 1988. **85**: 8166–8170.

6 **Chien, Y. H., Iwashima, M., Kaplan, K. B., Elliot, J. F. and Davis, M. M.,** A new T-cell receptor gene located within the alpha locus and expressed early in T-cell differentiation. *Nature* 1987. **327**: 677–682.

7 **Kubota, T., Wang, J. Y., Gobel, T. W. F., Hockett, R. D., Cooper, M. D. and Chen, C. H.,** Characterization of an avian (*Gallus gallus domesticus*) TCR α/δ locus. *J. Immunol.* 1999. **163**: 3858–3866.

8 **Fisher, C., Bouneau, L., Ozouf-Costaz, C., Crnogorac-Jurcevic, T., Weissenbach, J. and Bernot, A.,** Conservation of the T-cell receptor α/δ linkage in the teleost fish *Tetraodon nigroviridis*. *Genomics* 2002. **79**: 241–248.

9 **Parra, Z. E., Baker, M. L., Hathaway, J., Lopez, A. M., Trujillo, J., Sharp, A. and Miller, R. D.,** Comparative genomic analysis and evolution of the T cell receptor loci in the opossum *Monodelphis domestica*. *Biomed. Chromatogr. Genomics* 2008. **9**: 111.

10 **Clevers, H., Alarcon, B., Wileman, T. and Terhorst, C.,** The T cell receptor/CD3 complex: a dynamic protein ensemble. *Ann. Rev. Immunol.* 1988. **6**: 629–662.

11 **Oltz, E. M.,** Regulation of antigen receptor gene assembly in lymphocytes. *Immunol. Res.* 2001. **23**: 121–133.

12 **Criscitiello, M. F., Saltis, M. and Flajnik, M. F.,** An evolutionarily mobile antigen receptor variable region gene: doubly rearranging NAR-TcR genes in sharks. *Proc. Natl. Acad. Sci. USA* 2006. **103**: 5036–5041.

13 **Parra, Z. E., Baker, M. L., Schwarz, R. S., Deakin, J. E., Lindblad-Toh, K. and Miller, R. D.,** A unique T cell receptor discovered in marsupials. *Proc. Natl. Acad. Sci. USA* 2007. **104**: 9776–9781.

14 **Koop, B. F., Rowen, L., Wang, K., Kuo, C. L., Seto, D., Lenstra, J. A., Howard, S. *et al.*,** The human T-cell receptor TCRAC/TCRDC (Cα/Cδ) region: organization, sequence, and evolution of 976 kb of DNA. *Genomics* 1994. **19**: 478–493.

15 **Haire, R. N., Haindfield, M. K., Turpen, J. B. and Litman, G. W.,** Structure and diversity of T-lymphocyte antigen receptors alpha and gamma in *Xenopus*. *Immunogenetics* 2002. **54**: 431–438.

16 **Giudicelli, V., Chaume, D. and Lefranc, M. P.,** IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res.* 2005. **33**: D256–D261.

17 **Schwager, J., Burckert, N., Courtet, M. and Du Pasquier, L.,** Genetic basis of the antibody repertoire in Xenopus: analysis of the VH diversity. *EMBO J.* 1989. **8**: 2989–3001.

18 **Haire, R. N., Amemiya, C. T., Suzuki, D. and Litman, G. W.,** Eleven distinct VH gene families and additional patterns of sequence variation suggest a high degree of immunoglobulin gene complexity in a lower vertebrate, *Xenopus laevis*. *J. Exp. Med.* 1990. **171**: 1721–1737.

19 **Carroll, A. M., Slack, J. K. and Mu, X.,** V(D)J recombination generates a high frequency of nonstandard TCR delta-associated rearrangements in thymocytes. *J. Immunol.* 1993. **150**: 2222–2230.

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

21 **Ohta, Y. and Flajnik, M. F.,** IgD, like IgM, is a primordial immunoglobulin class perpetuated in most jawed vertebrates. *Proc. Natl. Acad. Sci. USA* 2006. **103**: 10723–10728.

22 **Zhao, Y., Pan-Hammarström, Q., Yu, S., Wertz, N., Zhang, X., Li, N., Butler, J. E. and Hammarström, L.,** Identification of IgF, a hinge-region-containing Ig class, and IgD in *Xenopus tropicalis*. *Proc. Natl. Acad. Sci. USA* 2006. **103**: 12087–12092.

23 **Chen, H., Kshirsagar, S., Jensen, I., Lau, K., Covarrubias, R., Schluter, S. F. and Marchalonis, J. J.,** Characterization of arrangement and expression of the T cell receptor gamma locus in the sandbar shark. *Proc. Natl. Acad. Sci. USA* 2009. **106**: 8591–8596.

24 **Benton, M. J. and Donoghue, P. C.,** Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* 2007. **24**: 26–53.

25 **Ota, T., and Nei, M.,** Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Mol. Biol. Evol.* 1994. **11**: 469–482.

26 **Baker, M. L., Belov, K. and Miller, R. D.,** Unusually similar patterns of antibody V segment diversity in distantly related marsupials. *J. Immunol.* 2005. **174**: 5665–5671.

27 **Johansson, J., Aveskogh, M., Munday, B. and Hellman, L.,** Heavy chain V region diversity in the duck-billed platypus (*Ornithorhynchus anatinus*): long and highly variable complementarity-determining region 3 compensates for limited germline diversity. *J. Immunol.* 2002. **168**: 5155–5162.

28 **Tutter, A. and Riblet, R.,** Conservation of an immunoglobulin variable region gene family indicates a specific, noncoding function. *Proc. Natl. Acad. Sci. USA* 1989. **86**: 7460–7464.

29 **Mussman, R., Courtet, M. and Du Pasquier, L.,** Development of the early B cell population in Xenopus. *Eur. J. Immunol.* 1998. **28**: 2947–2959.

30 **Rock, E. P., Sibbald, P. R., Davis, M. M. and Chien, Y. H.,** CDR3 length in antigen-specific immune receptors. *J. Exp. Med.* 1994. **179**: 323–328.

31 **Sciammas, R., Johnson, R. M., Sperling, A. I., Brady, W., Linsley, P. S., Spear, P. G., Fitch, F. W. and Bluestone, J. A.,** Unique antigen recognition by a Herpesvirus-specific TCR-γδ cell. *J. Immunol.* 1994. **152**: 5392–5397.

32 **Hayday, A. C.,** γδ T cells and the lymphoid stress-surveillance response. *Immunity* 2009. **31**: 184–196.

33 **Wilson, I. A. and Stanfield, R. L.,** Antibody-antigen interactions: new structures and new conformational changes. *Curr. Opin. Struct. Biol.* 1994. **3**: 857–867.

34 **Marchalonis, J. J., Schluter, S. F. and Edmundson, A. B.,** The T-cell receptor as immunoglobulin: paradigm regained. *Proc. Soc. Exp. Biol. Med.* 1997. **216**: 303–318.

35 **Glusman, G., Rowen, L., Lee, I., Boysen, C., Roach, J. C., Smit, A. F., Wang, K. *et al.*,** Comparative genomics of the human and mouse T cell receptor loci. *Immunity* 2001. **15**: 337–349.

36 **Lewis, S. M., Wu, G. E.,** The old and the restless. *J. Exp. Med.* 2000. **191**: 1631–1636.

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

38 **Wang, K., Gan, L., Kunisada, T., Lee, I., Yamagishi, H. and Hood, L.,** Characterization of the Japanese pufferfish (*Takifugu rubripes*) T-cell receptor alpha locus reveals a unique genomic organization. *Immunogenetics* 2001. **53**: 31–42.

39 **Fischer, C., Bouneau, L., Ozouf-Costaz, C., Crnogorac-Jurcevic, T., Weissenbach, J. and Bernot, A.,** Conservation of the T-cell receptor alpha/delta linkage in the teleost fish *Tetraodon nigroviridis*. *Genomics* 2002. **79**: 241–248.

40 **Pilström, L.,** The mysterious immunoglobulin light chain. *Dev. Comp. Immunol.* 2002. **26**: 207–215.

41 **Rast, J. P., Haire, R. N., Litman, R. T., Pross, S. and Litman, G. W.,** Identification and characterization of T-cell antigen receptor-related genes in phylogenetically diverse vertebrate species. *Immunogenetics* 1995. **42**: 204–212.

42 **Ohta, Y., Powis, S. J., Lohr, R. L., Nonaka, M., Du Pasquier, L. and Flajnik, M. F.,** Two highly divergent ancient allelic lineages of the transporter associated with antigen processing (TAP) gene in *Xenopus*: further evidence for co-evolution among MHC class I region genes. *Eur. J. Immunol.* 2003. **33**: 3017–3027.

43 **Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G.,** The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 1997. **24**: 4876–4882.

44 **Hall, T. A.,** BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 1999. **41**: 95–98.

45 **Kumar, S., Tamura, K. and Nei, M.,** MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform.* 2004. **5**: 150–163.

46 **Bartl, S., Baish, M. A., Flajnik, M. F. and 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.

47 **Ohta, Y., Powis, S. J., Coadwell, W. J., Haliniewski, D. E., Liu, Y., Li, H. and Flajnik, M. F.,** Identification and genetic mapping of *Xenopus* TAP2 genes. *Immunogenetics* 1999. **49**: 171–182.

Abbreviations: **EST:** expressed sequence tag · **NAR:** New Antigen Receptor · **ORF:** open reading frame · **RSS:** recombination signal sequences · **VH:** V genes that are indistinguishable from VH but found in the TCR locus

Full correspondence: Prof. Robert D. Miller, MSC03 2020, 1 University of New Mexico, Albuquerque, NM 87131-0001, USA
Fax: +1-505-277-0304
e-mail: rdmiller@unm.edu

Current address: Professor Michael F. Criscitiello, Department of Veterinary Pathobiology, Texas A&M University, College Station, TX 77843, USA

Received: 22/3/2010
Revised: 3/5/2010
Accepted: 12/5/2010
Accepted article online: 18/5/2010

