An evolutionarily mobile antigen receptor variable region gene: Doubly rearranging NAR-TcR genes in sharks

Michael F. Criscitiello*, Mark Saltis*, and Martin F. Flajnik*†‡

*Department of Microbiology and Immunology, University of Maryland, Howard Hall, Suite 324, 660 West Redwood Street, Baltimore, MD 21201-1559; and †National Aquarium in Baltimore, 501 East Pratt Street, Baltimore, MD 21202

Edited by Max D. Cooper, University of Alabama at Birmingham, Birmingham, AL, and approved January 24, 2006 (received for review August 15, 2005)

Distinctive Ig and T cell receptor (TcR) chains define the two major lineages of vertebrate lymphocyte yet similarly recognize antigen with a single, membrane-distal variable (V) domain. Here we describe the first antigen receptor chain that employs two V domains, which are generated by separate VDJ gene rearrangement events. These molecules have specialized "supportive" TcRδV domains membrane-proximal to domains with most similarity to IgNAR V. The ancestral NAR V gene encoding this domain is hypothesized to have recombined with the *TRD* locus in a cartilaginous fish ancestor >200 million years ago and encodes the first V domain shown to be used in both Igs and TcRs. Furthermore, these data support the view that γ/δ TcRs have for long used structural conformations recognizing free antigen.

cartilaginous fish | evolution | γ/δ T cells | T cell receptor | V(D)J rearrangement

The hallmark characteristics of the adaptive immune system are specificity and memory, both conferred by the V domains of Ig and T cell receptors (TcR). All genes encoding antigen receptor V domains are in the same superfamily and are generated by the same gene rearrangement mechanism, yet all jawed vertebrates studied have discrete Ig and TcR loci. Cartilaginous fish are the oldest animals having an adaptive immune system centered on rearranging antigen receptors. They have all four types of TcR (α, β, γ, δ) (1), three Ig isotypes [IgM (2, 3), IgW (4), and IgNAR (new antigen receptor) (5)], the recombination-activating gene recombinase (6), and polymorphic MHC genes (7, 8). Studies of modern sharks may shed light on the origins of adaptive immunity (9).

Although sharks have much of the basic molecular hardware required for adaptive immunity, some differences have been noted between their antigen receptors and those of other vertebrates. Shark Ig loci are found in many "clusters" as opposed to the single translocon organization common to mammals. Each of the hundreds of Ig loci in the shark genome contains V, D (diversity), J (joining), and C (constant) genes and may be prejoined in the germ line (10), begging questions of the regulation of rearrangement (11). It was originally suggested that horned shark *TRB* was multicluster (12), but this species seems to be an exception, because all four TcR genes are single translocon loci in skate (1). In addition to monomeric and pentameric IgM, cartilaginous fish have at least two other IgH isotypes. The poorly understood IgW isotype occurs in multiple forms (13) [as does IgM from elasmobranchs (14)]. IgNAR, which is apparently found only in cartilaginous fish, binds antigen by means of a single V domain (15), which is no more similar to IgV than to TcR V (16, 17). The *IgNARV* gene undergoes extensive hypermutation resulting in affinity maturation (18). Only the conventional α, β, γ, and δ TcR chains with single C and V domains have been described from shark or other vertebrates (19).

Here we describe a unique antigen receptor chain that blends characteristics of (previously incompatible) Ig and TcR into a TcR chain with two V domains, each encoded by separate rearranging VDJ segments, on a membrane-anchored TcRδ C domain. The membrane-distal domain, christened NAR-TcRV, is most related to the IgNARV domain and was recombined into the *TRD* locus >200 million years ago. Considering the reported interaction with antigen by IgNARV domains, this TcR chain provides evidence for direct antigen binding by γ/δ TcRs.

Results

RACE PCR Reveals a NAR Domain. We discovered an Ig superfamily V domain while studying the genetics and expression of TcRδV families in the nurse shark *Ginglymostoma cirratum*. Using a reverse primer to the TcRδC domain, 5′ RACE PCR from RNA of several tissues amplified products of two sizes (Fig. 1*A*). Cloning and sequencing of the low-molecular-weight band yielded typical TcRδ leader-V-D-J-C-encoding transcripts, but the higher band unexpectedly contained transcripts encoding an additional V domain, N-terminal to TcRδV. This domain is most similar to V domains of IgNAR, the Ig class of cartilaginous fish that does not associate with Ig light (IgL) chains (Figs. 1*B* and 2). We named this V domain NAR-TcRV, which reflects its relatedness to IgNARV and its expression as a TcR, and we call the entire chain NAR-TcR (genes encoding the NAR-TcRV domain are denoted *NT*) (Fig. 3*A*). *NT* gene transcripts encode a typical leader peptide 5′ to the *NTV* segment and have a *NTJ* segment spliced directly to a *TRDV* segment. A parsimonious model of a γ/δ TcR complex including a NAR-TcR δ chain is depicted in Fig. 1*B*.

Several forms of NAR-TcRV were identified, all most related to IgNARV (≈30–40% identity; Fig. 1*C*) (20). β-Strands c and d deviate most from the Ig superfamily prototype V domain, and the canonical tryptophan of the WYRK motif is not present (21). NAR-TcRV lacks the cysteines in the c and d strands and complementarity-determining region (CDR)3 that make the additional disulfide bonds in some IgNARV domains (17). However, a conserved cysteine in the a–b loop is free of intradomain cysteine partners in all NAR-TcRV (see below).

Dedicated TcRVδ Domains Support NAR-TcRV Domains. Genes encoding the same subfamilies of NAR-TcRV and TcRδV domains are always found in the same RNA transcripts (e.g., *NTV1* is only found in transcripts 5′ of *TRDV1*). Alignment of four TcRδV subfamilies that support NAR-TcRV (families 1–4) with typical δV subfamilies (families 5–8) shows the supporting δV domains to lack leader peptides and share a cysteine residue in CDR1 (Fig. 1*D*). Threading these sequences onto crystal structures for IgNARV and human TcRδ predicts that the cysteine in the a–b loop of the NAR-TcRV domains and in CDR1 of supporting TcRδV domains are in close

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: TcR, T cell receptor; RSS, recombination signal sequence; CDR, complementarity-determining region; PBL, peripheral blood leukocyte.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. DQ022688–DQ022718).

‡To whom correspondence should be addressed. E-mail: mflajnik@som.umaryland.edu.

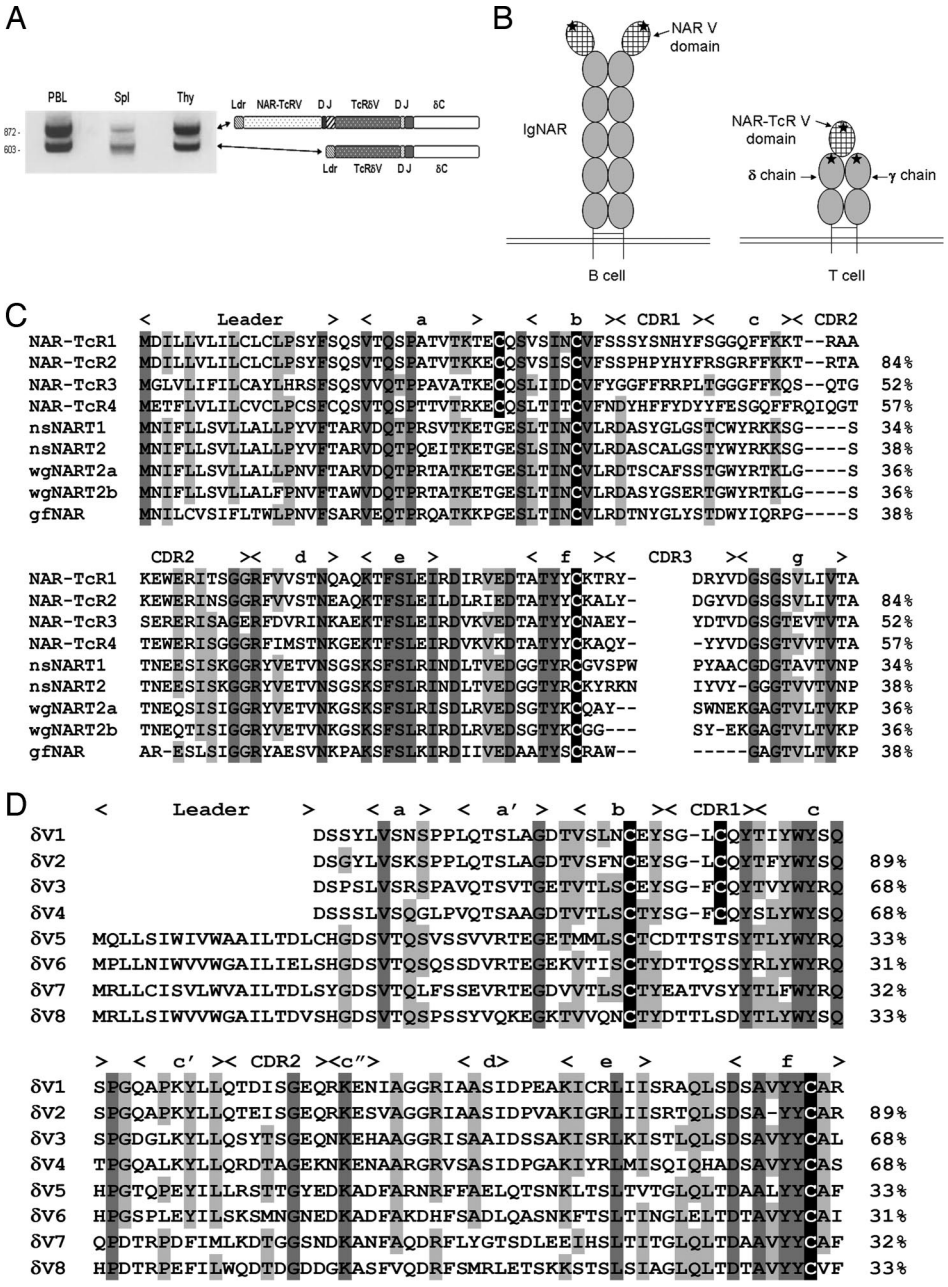

Fig. 1. NAR-TcRV is a V domain supported on TcRδ. (*A*) Products of 5′ RACE PCR with *TRDC* primers using RNA from shark PBL, spleen, and thymus. (*B*) Cartoon depiction of IgNAR and the hypothetical γ/δ TcR with NAR-TcR antigen receptors. Stars mark CDR3s generated from somatic rearrangement. (*C*) Predicted amino acid sequences of four families of NAR-TcRV aligned with IgNARV from nurse shark (nsNART1, U18721; nsNART2, U18680), wobbeygong (wgNART2a, AF336092; wgNART2b, AF336091), and guitarfish (gfNAR, AY524298). Residues conserved are highlighted in gray, and Ig superfamily canonical intradomain and putative interdomain cysteines are highlighted in black. β-Strands and CDRs are noted above the alignment. Percent amino acid identity to the top sequence is shown on the right. (*D*) Predicted amino acid sequences of eight families of TcRδV, four of which support NAR-TcRV. Highlighting and notation are the same as in *C*.

proximity and could form a disulfide bond (R. L. Stanfield, personal communication) (22, 23). Thus, the NAR-TcRV domain likely has two covalent anchors to the δV domain.

Genomic Organization of a Doubly Rearranging Receptor. High variability in the sequence and length of CDR3 in NAR-TcRV sequences strongly suggested that these genes undergo V(D)J rearrangement events (Fig. 4*A*), prompting an investigation of *NTV* and supporting *TRDV* genomic organization. RT-PCR resulted in some incompletely rearranged transcripts that revealed the intergenic sequence between *V* and *D*, and *D* and *J* segments encoding the NAR-TcRV domain (Fig. 4*B*). In two transcripts, the associated *TRDV* gene is completely rearranged to the typical TcR *DD* and *DJ* gene segments. Genomic PCR using *NTV* forward primers and reverse primers to the corresponding *TRDV* revealed that a cluster comprising one *NTV*, one *D*, and one *J* segment is 5′ of each *TRDV* exon that supports NAR-TcRV. These *TRDV*s lost the exon encoding the leader peptide, and thus they can be expressed as protein only when attached to the N-terminal NAR-TcRV domain (Figs. 1*A* and 4*B*). The *NTJ* exon splices to a canonical splice site in the *TRDV* exon. The *NTV*, *D*, and *J* segments are bordered by recombination signal sequences (RSS) known to regulate recombination-activating gene-mediated rearrangement events (24, 25). Like in IgH loci heptamers and nonamers separated by 23-nucleotide spacers lie 3′ and 5′ of the *NTV* and *J* segments, respectively, and *NTD* segments are flanked by RSSs with 12-nucleotide spacers. Like in all other vertebrates examined, *TRDD* segments, which are flanked by RSS with 5′ 12-nucleotide and 3′ 23-nucleotide spacers, rearrange to a *J* segment RSS having a 12-nucleotide spacer. There is neither rearrangement between different *NTV* clusters (as is true of the *IgH* and *IgL* clusters of the cartilaginous fish) (10, 26) nor joining of *NTV* segments with *DD* or *DJ* segments. The incompletely rearranged transcripts provide preliminary evidence that the *TRDV*, *D*, and *J* segments rearrange

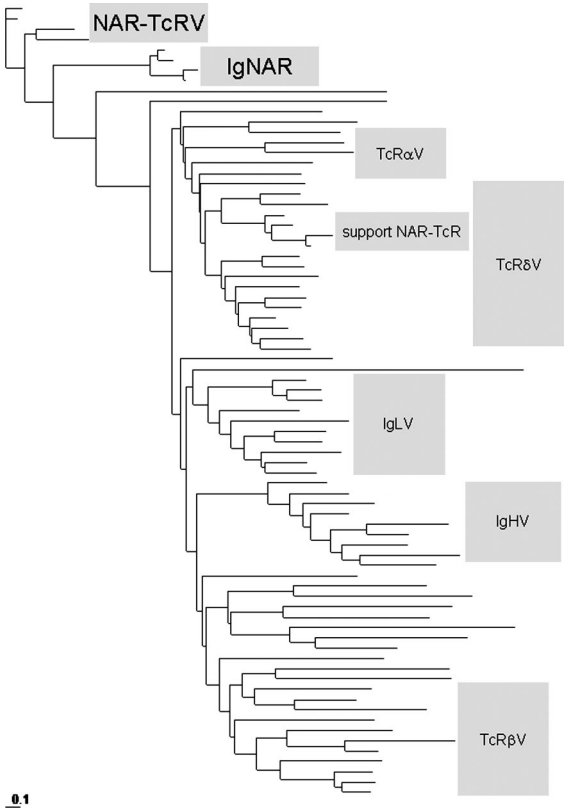

Fig. 2. Phylogenetic analysis of Ig superfamily V domains. See Table 1 and Fig. 6, which are published as supporting information on the PNAS web site, for accession numbers of sequences used as well as the same data presented in a radial phylogram.

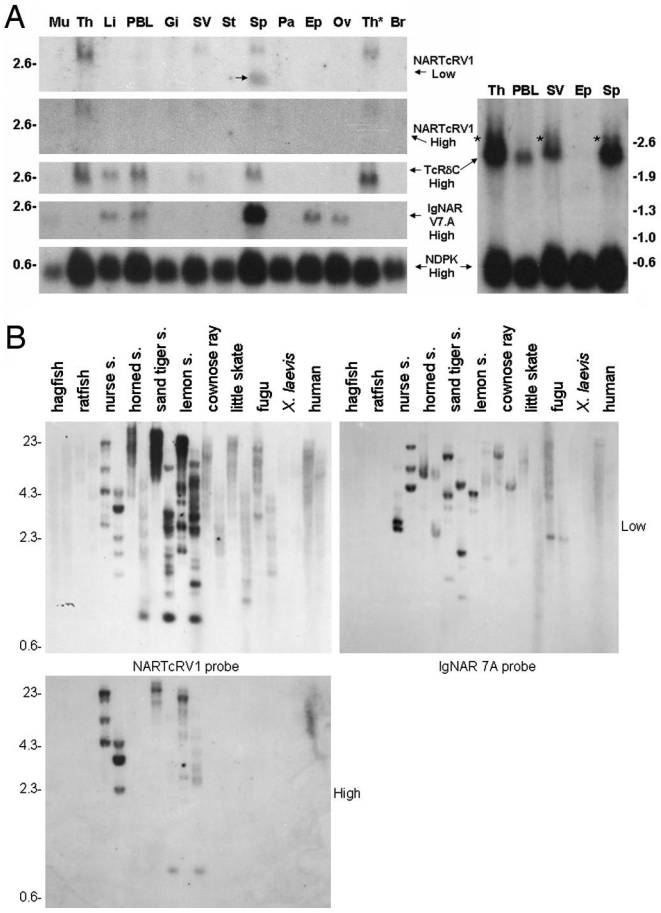

Fig. 3. NAR-TcR is expressed in shark lymphoid tissues and is found in modern sharks. (*A*) Tissues used in the Northern blot are muscle, thymus, liver, PBLs, gill, spiral valve (intestine), stomach, spleen, pancreas, epigonal (a primary lymphoid organ of cartilaginous fishes), ovary, thymocytes, and brain. Marker (in kb) is shown flanking the blots. Probes were generated from V domain encoding genes of *NTV* and *IgNAR*, *TRDC*, and the loading control *NDPK*. An arrow marks the cross-hybridizing IgNAR band, and asterisks highlight the NAR-TcR encoding transcripts recognized by *TRDC* probing a blot allowed to migrate further. (*B*) Southern blot of HindIII- and PstI-digested genomic DNA probed with *NTV1* and *IgNAR 7A*. Marker (in kb) is shown on left, and higher or lower wash stringency is indicated.

before *NTV*, *D*, *J* joining. As mentioned, the CDR3 sequences at both junctions are diverse in length and sequence (Fig. 4*A*). Notably, the TcRδV that support NAR-TcRV are no less diverse in CDR3 than the typical TcRδVs (data not shown) despite the fact that the antigen-binding site of this domain might be occluded by NAR-TcRV. Both identified *TRDD* segments are used in the supporting *TRDV* rearrangements, and no preference for particular *TRDJ* segments has been observed.

Tissue Expression. Highest expression of *NTV* was detected in the thymus followed by spleen, spiral valve (shark intestine), and peripheral blood (Fig. 3*A*). The *NTV1* probe cross-hybridized with the faster migrating *IgNAR* transcript under low-stringency conditions in the spleen. This band is not surprising considering the sequence homology between the two gene families and the high levels of *IgNAR* expression in spleen (27). Increasing the wash stringency resulted in loss of the cross-hybridizing *IgNAR* band. The *NTV1* signal was also decreased because of the range of different NAR-TcRV family members (Fig. 1*C*). Hybridization with a typical *TRDV* probe that does not support NAR-TcRV revealed the same expression pattern (data not shown) as does a *TRDC* probe that recognizes all *TRD* transcripts regardless of *V* usage, with the exception of higher relative liver and peripheral blood leukocyte (PBL) expression for the *TRDC* probe and a higher relative expression of *NTV1* in the spiral valve. Signal strength from the membrane-expressed, individual *NTV* family probes is, as expected, much less than that of secreted *Ig* or housekeeping probes and only a fraction of that of pan-V recognizing *TRDC*. Unlike what has been hypothesized for some unusual γδ TcR in mammals (28), the NAR-TcR tissue distribution is consistent with the great junctional diversity in CDR3 in suggesting that the chains have not been selected to recognize tissue-specific ligand. A second blot was prepared with immune tissue RNA, electrophoresed longer, and probed with only *TRDC* and *NDPK* to confirm that the higher 2.7-kb *TRDC*-positive transcript is indeed NAR-TcR and not old activity from previous *NTV1* probings. This second blot shows significant enrichment of NAR-TcR in the spiral valve.

Phylogenetic Distribution of NAR-TcR and Number of Families. The existence of NAR-TcR in other organisms was examined by genomic Southern blotting (Fig. 3*B*). All sharks tested (separated in phylogeny by up to 200 million years) were positive and had multiple copies of *NTV*, even under higher-stringency conditions. Weakly hybridizing bands were seen with the batoid species (ray and skate), fugu, and even human, which were washed away under high-stringency conditions; thus, it is not clear whether NAR-TcR existed in the ancestor of all cartilaginous fish and even perhaps all extant vertebrates. However, *in silico* investigations of fugu, *Xenopus*, and mammalian databases have failed in finding *NTV* homologues or telltale leaderless *TRDV* genes. An IgNARV probe hybridized to different bands on the same blot, dismissing cross-

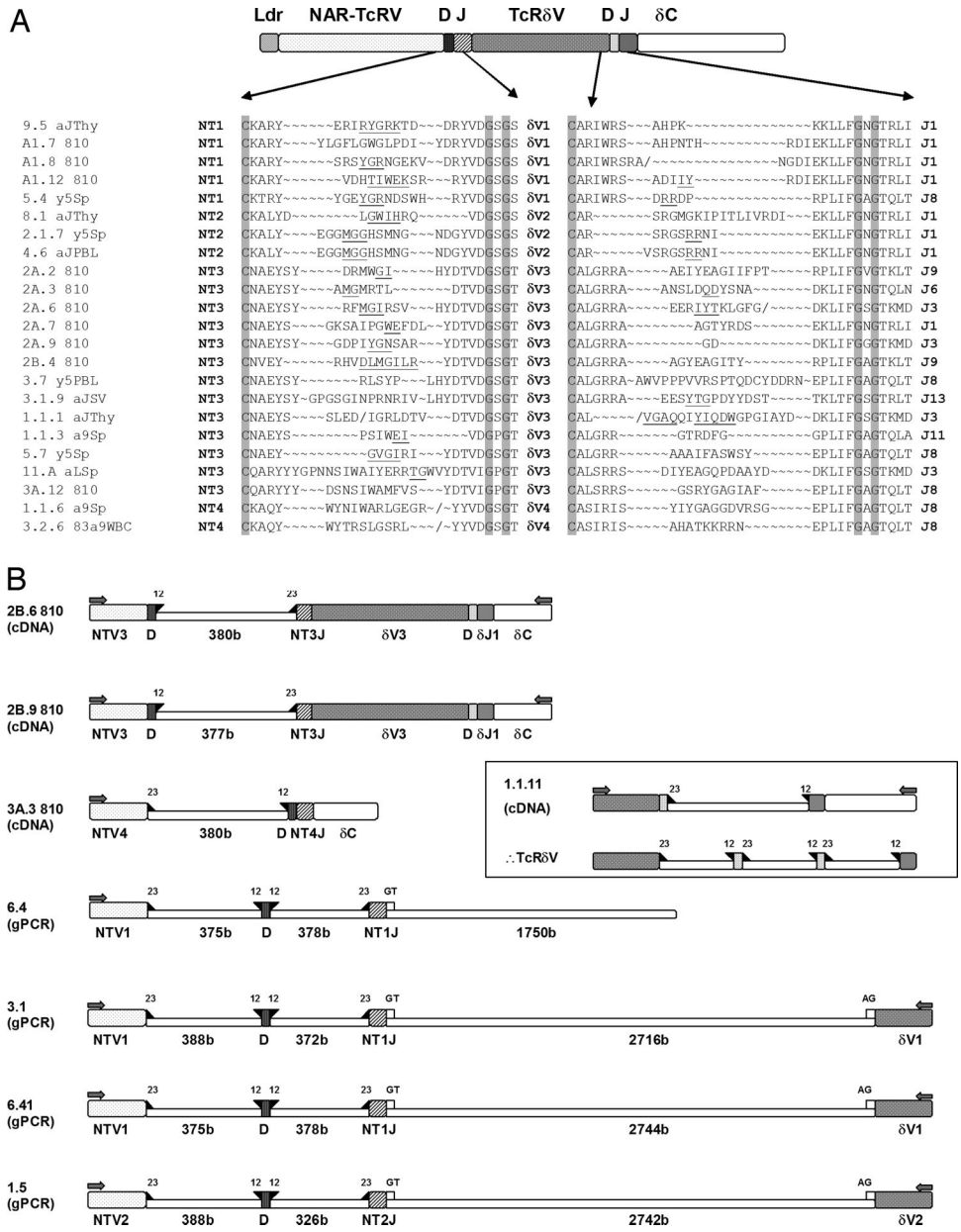

Fig. 4. NAR-TcR is doubly rearranging. (*A*) Focus on amino acid sequences of both NAR-TcRV and TcRδV CDR3s. Bold abbreviations indicate NAR-TcRV, TcRδV, and TcRδJ (unpublished data) families used in clones named at left. Conserved cysteine of the f strand and GXG motif of the g strand are highlighted flanking each CDR3. Known and predicted D-segment-encoded amino acids are underlined, ''~'' are gaps introduced to align Vs and Js, and ''/'' marks frameshifts resulting from nonfunctional rearrangements. (*B*) Diagrams of clones showing *NTV* assembly mechanism. Recombination and splicing signals are marked by filled triangles for RSSs and open squares for GT/AG intronic splice sites. An incompletely rearranged 5′ RACE clone from PBL shows RSS 3′ of *TRDD* and 5′ of *TRDJ*. These data suggest that the mammalian RSS system for δ is also used in shark (*Inset*) (52, 53).

hybridization to the IgNAR loci as an explanation for NAR-TcR's existence in sharks (Fig. 3*B*).

We have identified four nurse shark NAR-TcRV families through cDNA and genomic cloning, but the Southern blot suggests that more members await discovery. Consistent with this proposal, screening of a 10× coverage nurse shark bacterial artificial chromosome (BAC) library with a *NTV1* probe resulted in 208 clones, 168 of which also hybridized to *TRDV* probes. In contrast, no NAR-TcRV clone contained the single-copy *TRDC* gene, suggesting that the *NTV* genes are >150 kb away from *TRDC*. Five of the 11 *TRDC*-positive BAC clones were also positive for *TRDV* genes that do not support NAR-TcRV, which advocates a large, complex *TRD* locus with *NTV*-supportive *TRDV* miniclusters upstream of the traditional *V*, *D*, and *J* arrays (Fig. 5). No BAC clones positive for any TcR probes were also positive for *IgNAR*.

Initial characterization of nurse shark TcRβ and γ revealed no NAR-like mRNA transcripts by 5′ RACE or cDNA screenings (unpublished observations), nor did differential screening of a shark cDNA library with *TRDC* and *NTV1* probes. Published studies of horned shark TcRβ also did not reveal NAR-TcR (12, 29).

Relatedness of NAR-TcRV to IgNARV and Other Antigen Receptors. Amino acid alignment and phylogenetic analysis of many vertebrate V domains confirms the close relationship between NAR-TcRV and IgNARV and substantiates their distinction from TCR and Ig V domains (Fig. 2). It is peculiar that NAR-TcRV and IgNARV group not only outside of Ig and TcR V domains but also basal to diverse chordate V domains (30–35). We hesitate to infer from the tree an ancestral relationship between the NAR-TcR and IgNAR domains and those of conventional Ig and TcR, despite the parsimony of a single binding domain and the position of the IgNARV/NAR-TcRV cluster, because establishing alignments of such diverse V domains is challenging. Whatever the exact history of these gene families, V domains of NAR-TcR and IgNAR are clearly distinct from those of conventional Ig and TcR, as shown by genetic distance (Fig. 2), gross domain structure (Fig. 1 *B* and *C*), and mechanism of antigen binding (proven for IgNAR) (15).

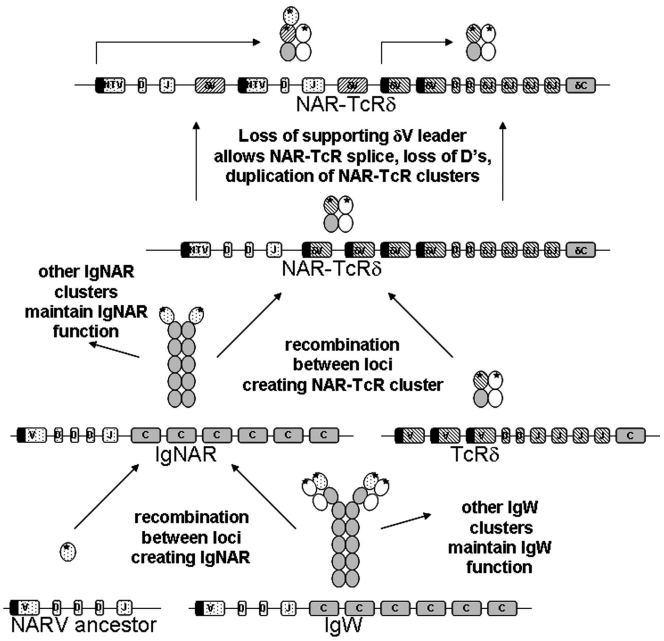

Fig. 5. Hypothetical scheme for history of NAR variable domains. Depictions of genomic loci are simplified; filled bars indicate variable segments with leader exons upstream. Direction of cross-hatching distinguishes *TRDV* segments dedicated to supporting NAR-TcRV.

It is also notable that the supporting TcRδV domains form a separate cluster within the major TcRδV group. The genetic distances between the supporting δV are similar to those of the NAR-TcRV, consistent with *en bloc* duplications of *NTV*/supporting *TRDV* genes over evolutionary time (Fig. 5).

Discussion

The maintenance of many functional NAR-TcR and their supporting TcRδV with conserved cysteines (presumably for interdomain stability), many *NTV* genes in shark genomes divorced hundreds of millions of years ago, and high expression in lymphoid tissues are all consistent with NAR-TcR playing an important role in the shark immune system. NAR-TcR transcripts comprise 21 of 99 of the TcRδ repertoire by sequenced full-length 5′ RACE clone count, and the higher band accounts for 51%, 42%, and 53% of the amplification products from PBL, spleen, and thymus by semiquantitative densitometry in Fig. 1*A*. Because the IgNARV domain recognizes antigen as a single chain (15, 22), and additional domains have not been found on shark TcRγ, it is likely that the NAR-TcRV domain will recognize antigen in a similar fashion. Crystallography of human γ/δ has shown a smaller angle between the C and V domains than is seen in α/β heterodimers (23); such an orientation in shark could accommodate steric constraints of an additional V domain in an immunological synapse. This angle could lend accessibility to ligand binding at either V domain, so we cannot rule out use of the typical γV/δV binding site. However, we propose that the NAR-TcRV uses the typical γ/δ TcR as a scaffold (Fig. 1*B*), and thus after antigen recognition by the NAR-TcR V domain the T cell signaling machinery is directed to induce cytokine secretion or cell killing. Quaternary structure modeling predicts that the TcRδV CDR3 might be occluded by the NAR-TcRV domain, which would replace the TcR γ/δ Vs in antigen recognition. Like other Igs, IgNAR has both transmembrane and secreted forms, and its gene undergoes extensive somatic mutation after antigenic stimulation. NAR-TcR neither has a secreted form (single band by Northern blotting and absence of alternative splicing from sequence data) nor is somatically mutated (extensive sequence analysis of particular families); thus, although the NAR-TcRV sequence is most related to IgNARV, its other basic characteristics are the same as those of TcRs.

Hallmark transmembrane charged residues of TcRδC, as well as its short cytoplasmic tail void of obvious signaling potential, predict that the NAR-TcR δ chain is part of a γ/δ/CD3 complex for cell surface expression and signal transduction. This notion is consistent with the short cytoplasmic tails found on IgNAR and other antigen receptor chains from cartilaginous fish to man. CD3 and CD79 genes have been found in diverse vertebrate groups, including the amphibian *Xenopus* and fish *Takifugu* (36, 37), and we predict that these signaling chains will be found in elasmobranchs. Thus, it is likely that the evolutionarily mobile NAR V domain signals via CD79 orthologues on B cells (as IgNAR) and CD3 orthologues on T cells (as NAR-TcRδ).

An *IgNAR VDJ* cluster may have recombined with *TRDV* genes sometime early in the evolution of modern sharks, 200 million years ago (Fig. 5). The original *NTV-DV* gene cluster then duplicated many times and is apparently used to a different extent in different shark taxa (as exemplified by fewer bands in the horn shark digests in Fig. 3*B*). We previously suggested that an *IgNAR VDJ* cluster recombined with an *IgW* cluster (an extant elasmobranch Ig isotype), which then gave rise to the *IgNAR* isotype gene, a possibility that is supported by these data (16). The origins of the NAR V domain is a mystery (38), because it does not show high similarity to any of the known antigen receptor V domains; the simplest interpretation is that the NAR V arose as a gene encoding a single domain, which was distributed to different antigen receptor families (Fig. 5).

Lateral gene transfer and endosymbiosis has forced major reconsideration of prokaryotic evolution (39). Horizontal gene transfer mechanisms have occurred in eukaryotes (40) but are often excluded from evolutionary hypotheses in favor of simpler Darwinian vertical-descent models. The concept of lateral gene flow is also a common theme for evolutionary immunologists, in which the demonstrable transposase activity of the recombination-activating gene products suggested an invading transposon in antigen receptor evolution (41, 42). The NAR domain now requires consideration as a nomadic gene shuttled intragenomically by means of conventional recombination. The exon-shuffling theory asserts that exon-encoded domains facilitated evolutionary swapping and experimentation of functional protein domains (43), and the best empirical evidence for this theory has been obtained from analysis of exon–intron structure of complex eukaryotic genes in relation to the proteins they encode (44). The use of the NAR domain in two very different antigen receptors on two very different lymphocytes is the most explicit example of the exon-shuffling theory of which we are aware.

Comparative antigen receptor studies throughout vertebrates have thus far shown contrasting natural histories for Ig and TcR. Although Ig have used various genomic arrangements, isotypes, and somatic diversification mechanisms in different vertebrates to achieve repertoire diversification (reviewed in ref. 45), TcR studies have yielded relatively few surprises (19). This dichotomy has been linked to the different stringencies required for function as secreted binders of free antigen and recognition of peptide presented by MHC molecules (46). However, because most γ/δ T cells are not MHC-restricted (47), they may have been afforded evolutionary freedom more similar to that of the direct binding B cell receptors (Igs). Therefore, perhaps it is not surprising that a mobile genetic element encoding an antigen-binding domain would thrive on TcRδ, giving it an additional tool for antigen recognition by means of a protruding single domain in addition to the planar recognition landscape possible with the TcRδ and TcRγ V domains. The NAR-TcR on a clonally expandable T cell could conceivably recognize cell-bound antigen of fungi, parasites, or virally infected cells and direct cellular cytotoxicity in a manner akin to antibody-dependent cellular cytotoxicity by natural killer cells through the TcR instead of Ab and FcR.

Materials and Methods

RACE PCR. 5′ RACE PCR products were amplified by using TcRδC primer (5′-GCTGGCCAGACAGACTGCAGCTTGGACAGC-3′) and nested TcRδC primer (5′-TTGGTGGATAAAAAGC-CGAC-3′) from adult shark PBLs, spleen, and thymus total RNA and then separated in 1% agarose and visualized with ethidium bromide. The SMART RACE system (BD Biosciences) was used according to the manufacturer's protocol with 2 × 20 cycles annealing at 68°C. All products were cloned into pCR2.1 with TA Cloning kit (Invitrogen) and sequenced by the University of Maryland's Biopolymer Core facility.

RT-PCR. Clones in Fig. 4*A* ending in 810 and incompletely rearranged cDNA clones in Fig. 4*B* were gathered with NAR-TcR to TcRδC RT-PCR by using the nonnested δC primer (Fig. 1*A*) and primers specific for NAR-TcRV1 (5′-ACGCGTG-CAGCGAAGGAGTGG-3′), NAR-TcRV2 (5′-CATCCCTAT-CATTATTTTAGGAGT-3′), NAR-TcRV3 (5′-TTCAAA-CAAAGCCAGACAGGATCAGAG-3′), and NAR-TcRV4 (5′-CTCAGGCAAACCCAGACCAAGACAGAC-3′). Oligo-dT-primed cDNA was made from 5 μg of RNA as described and used as template for PCR amplification (14).

Genomic PCR. PCR from genomic DNA isolated from shark erythrocytes using described NAR-TcRV primers (Fig. 2*A*) and δV1 (5′-TCTCTGCTCTCCTGAAATGTCTGT-3′) and δV2 (5′-TCTCTGCTCTCCTGAAATTTCTGT-3′) reverse primers shows the arrangement of two entire NAR-TcRV clusters. One microgram of genomic DNA was used as template in a PCR of five cycles annealing at 50°C followed by 25 cycles annealing at 54°C.

Northern and Southern Blotting. Total RNA was prepared for Northern blotting as described (48), and 15 μg was loaded for each lane. All probes were labeled with ^{32}P dCTP PCR as described (49), and free nucleotides were removed with Quick-Spin Sephedex G-50 columns (Roche). Probes were amplified by using primers to include most of the V segment (NAR-TcRV1, 5′-GACACA-GAGCCCGGCAACGGTG-3′ and 5′-GTGCTTGATTCGTG-GACAC-3′; IgNARV7A, 5′-GCTCGAGTGGACCAAACA-3′ and 5′-ACCGCAACGATACGTGCCAC-3′; TcRδC, 5′-TTA-AGCCGAAATTGTCGGCT-3′ and 5′-AGAAATCCAGA-CTCGGGCAG-3′) or the loading control nucleotide diphosphate kinase (5′-AACAAGGAACGAACCTTC-3′ and 5′-TCACT-CATAGATCCAGTC-3′). Probes routinely labeled to 3×10^7 cpm. Southern blot was performed on HindIII-digested and PstI-digested (Roche) genomic DNA as described (50). Organisms included are Pacific hagfish, spotted ratfish, nurse shark, horned shark, sand tiger shark, lemon shark, cownose ray, little skate, Japanese pufferfish, African clawed frog, and human. Final low-stringency wash conditions were 20-min agitation in 2× SSC/0.1% SDS at 55°C, and final high-stringency wash conditions were 20-min agitation in 0.2× SSC/0.1% SDS at 65°C.

Phylogenetic Analysis. Multiple alignment of 78 chordate V domains, with emphasis on IgH, IgL, and the four TcR from diverse groups of vertebrates was performed in CLUSTALX (51) and then trimmed to 114 informative columns (see Fig. 7, which is published as supporting information on the PNAS web site). The PHYLIP suite was used for subsequent analysis. A distance matrix was created with PROTDIST and used to draw a phylogram with NEIGHBOR. Bootstrap values after 1,000 iterations generated by SEQBOOT are shown.

We thank Yuko Ota for RNA preparation; Helen Dooley and Robin Stanfield for helpful discussions of NAR-TcR structure; Becky Lohr for the identification of the nurse shark TcRδC domain gene; and J. Cerny, M. Diaz, L. Du Pasquier, and D. Nemazee for critique of the manuscript. The work was supported by grants from the National Institutes of Health (to M.F.F. and M.F.C.).

1. Rast, J. P., Anderson, M. K., Strong, S. J., Luer, C., Litman, R. T. & Litman, G. W. (1997) *Immunity* **6,** 1–11.
2. Marchalonis, J. & Edelman, G. M. (1965) *J. Exp. Med.* **122,** 601–618.
3. Marchalonis, J. & Edelman, G. M. (1966) *Science* **154,** 1567–1568.
4. Harding, F. A., Amemiya, C. T., Litman, R. T., Cohen, N. & Litman, G. W. (1990) *Nucleic Acids Res.* **18,** 6369–6376.
5. Greenberg, A. S., Avila, D., Hughes, M., Hughes, A., McKinney, E. C. & Flajnik, M. F. (1995) *Nature* **374,** 168–173.
6. Greenhalgh, P. & Steiner, L. A. (1995) *Immunogenetics* **41,** 54–55.
7. Kasahara, M., McKinney, E. C., Flajnik, M. F. & Ishibashi, T. (1993) *Eur. J. Immunol.* **23,** 2160–2165.
8. Hashimoto, K., Nakanishi, T. & Kurosawa, Y. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 2209–2212.
9. Flajnik, M. F. & Rumfelt, L. L. (2000) *Curr. Top. Microbiol. Immunol.* **248,** 249–270.
10. Hinds, K. R. & Litman, G. W. (1986) *Nature* **320,** 546–549.
11. Eason, D. D., Litman, R. T., Luer, C. A., Kerr, W. & Litman, G. W. (2004) *Eur. J. Immunol.* **34,** 2551–2558.
12. Rast, J. P. & Litman, G. W. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 9248–9252.
13. Kobayashi, K. & Tomonaga, S. (1988) *Mol. Immunol.* **25,** 115–120.
14. Rumfelt, L. L., Avila, D., Diaz, M., Bartl, S., McKinney, E. C. & Flajnik, M. F. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 1775–1780.
15. Stanfield, R. L., Dooley, H., Flajnik, M. F. & Wilson, I. A. (2004) *Science* **305,** 1770–1773.
16. Greenberg, A. S., Hughes, A. L., Guo, J., Avila, D., McKinney, E. C. & Flajnik, M. F. (1996) *Eur. J. Immunol.* **26,** 1123–1129.
17. Roux, K. H., Greenberg, A. S., Greene, L., Strelets, L., Avila, D., McKinney, E. C. & Flajnik, M. F. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 11804–11809.
18. Dooley, H. & Flajnik, M. F. (2005) *Eur. J. Immunol.* **35,** 936–945.
19. Charlemagne, J., Fellah, J. S., De Guerra, A., Kerfourn, F. & Partula, S. (1998) *Immunol. Rev.* **166,** 87–102.
20. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
21. Lesk, A. M. & Chothia, C. (1982) *J. Mol. Biol.* **160,** 325–342.
22. Streltsov, V. A., Varghese, J. N., Carmichael, J. A., Irving, R. A., Hudson, P. J. & Nuttall, S. D. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 12444–12449.
23. Allison, T. J., Winter, C. C., Fournie, J. J., Bonneville, M. & Garboczi, D. N. (2001) *Nature* **411,** 820–824.
24. Max, E. E., Seidman, J. G. & Leder, P. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 3450–3454.
25. Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. (1979) *Nature* **280,** 288–294.
26. Litman, G. W., Anderson, M. K. & Rast, J. P. (1999) *Annu. Rev. Immunol.* **17,** 109–147.
27. Rumfelt, L. L., Diaz, M., Lohr, R. L., Mochon, E. & Flajnik, M. F. (2004) *J. Immunol.* **173,** 1129–1139.
28. Garboczi, D. N. (2005) *Science* **308,** 209–210.
29. Hawke, N. A., Rast, J. P. & Litman, G. W. (1996) *J. Immunol.* **156,** 2458–2464.
30. Pancer, Z., Mayer, W. E., Klein, J. & Cooper, M. D. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 13273–13278.
31. Suzuki, T., Shin, I., Fujiyama, A., Kohara, Y. & Kasahara, M. (2005) *J. Immunol.* **174,** 2885–2891.
32. Cannon, J. P., Haire, R. N. & Litman, G. W. (2002) *Nat. Immunol.* **3,** 1200–1207.
33. Dietrich, J., Cella, M., Seiffert, M., Buhring, H. J. & Colonna, M. (2000) *J. Immunol.* **164,** 9–12.
34. Yoder, J. A., Mueller, M. G., Wei, S., Corliss, B. C., Prather, D. M., Willis, T., Litman, R. T., Djeu, J. Y. & Litman, G. W. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 6771–6776.
35. Chretien, I., Robert, J., Marcuz, A., Garcia-Sanz, J. A., Courtet, M. & Du, P. L. (1996) *Eur. J. Immunol.* **26,** 780–791.
36. Guselnikov, S. V., Najakshin, A. M. & Taranin, A. V. (2003) *Immunogenetics* **55,** 472–479.
37. Guselnikov, S. V., Bell, A., Najakshin, A. M., Robert, J. & Taranin, A. V. (2003) *Dev. Comp. Immunol.* **27,** 727–733.
38. Richards, M. H. & Nelson, J. L. (2000) *Mol. Biol. Evol.* **17,** 146–155.
39. Rivera, M. C. & Lake, J. A. (2004) *Nature* **431,** 152–155.
40. Benveniste, R. E. & Todaro, G. J. (1974) *Nature* **252,** 456–459.
41. Agrawal, A., Eastman, Q. M. & Schatz, D. G. (1998) *Nature* **394,** 744–751.
42. Hiom, K., Melek, M. & Gellert, M. (1998) *Cell* **94,** 463–470.
43. Gilbert, W. (1978) *Nature* **271,** 501.
44. Liu, M. & Grigoriev, A. (2004) *Trends Genet.* **20,** 399–403.
45. Flajnik, M. F. (2002) *Nat. Rev. Immunol.* **2,** 688–698.
46. Eason, D. D., Cannon, J. P., Haire, R. N., Rast, J. P., Ostrov, D. A. & Litman, G. W. (2004) *Semin. Immunol.* **16,** 215–226.
47. Jameson, J., Witherden, D. & Havran, W. L. (2003) *Curr. Opin. Immunol.* **15,** 349–353.
48. Bartl, S., Baish, M. A., Flajnik, M. F. & Ohta, Y. (1997) *J. Immunol.* **159,** 6097–6104.
49. Mertz, L. M. & Rashtchian, A. (1994) *Anal. Biochem.* **221,** 160–165.
50. Greenberg, A. S., Steiner, L., Kasahara, M. & Flajnik, M. F. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 10603–10607.
51. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997) *Nucleic Acids Res.* **25,** 4876–4882.
52. Hata, S., Satyanarayana, K., Devlin, P., Band, H., McLean, J., Strominger, J. L., Brenner, M. B. & Krangel, M. S. (1988) *Science* **240,** 1541–1544.
53. Elliott, J. F., Rock, E. P., Patten, P. A., Davis, M. M. & Chien, Y. H. (1988) *Nature* **331,** 627–631.

Criscitiello *et al*. 10.1073/pnas.0507074103.

Supporting Information

Files in this Data Supplement:

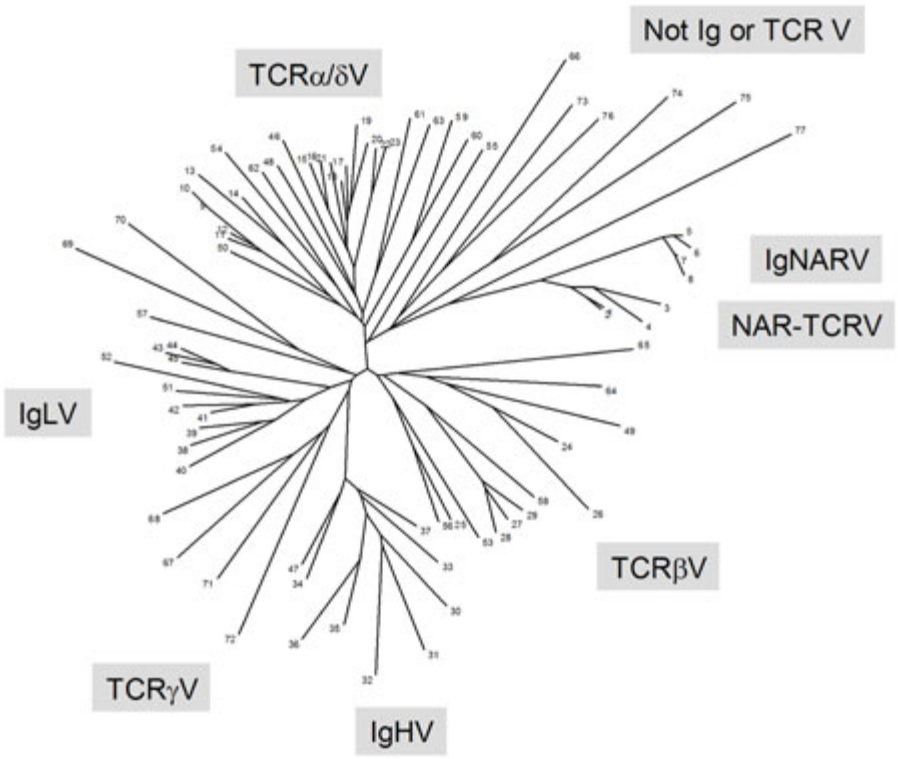

Fig. 6. Radial phylogram of rectangular phylogram in Fig. 2. See Table 1 for individual sequence descriptions and accession numbers.

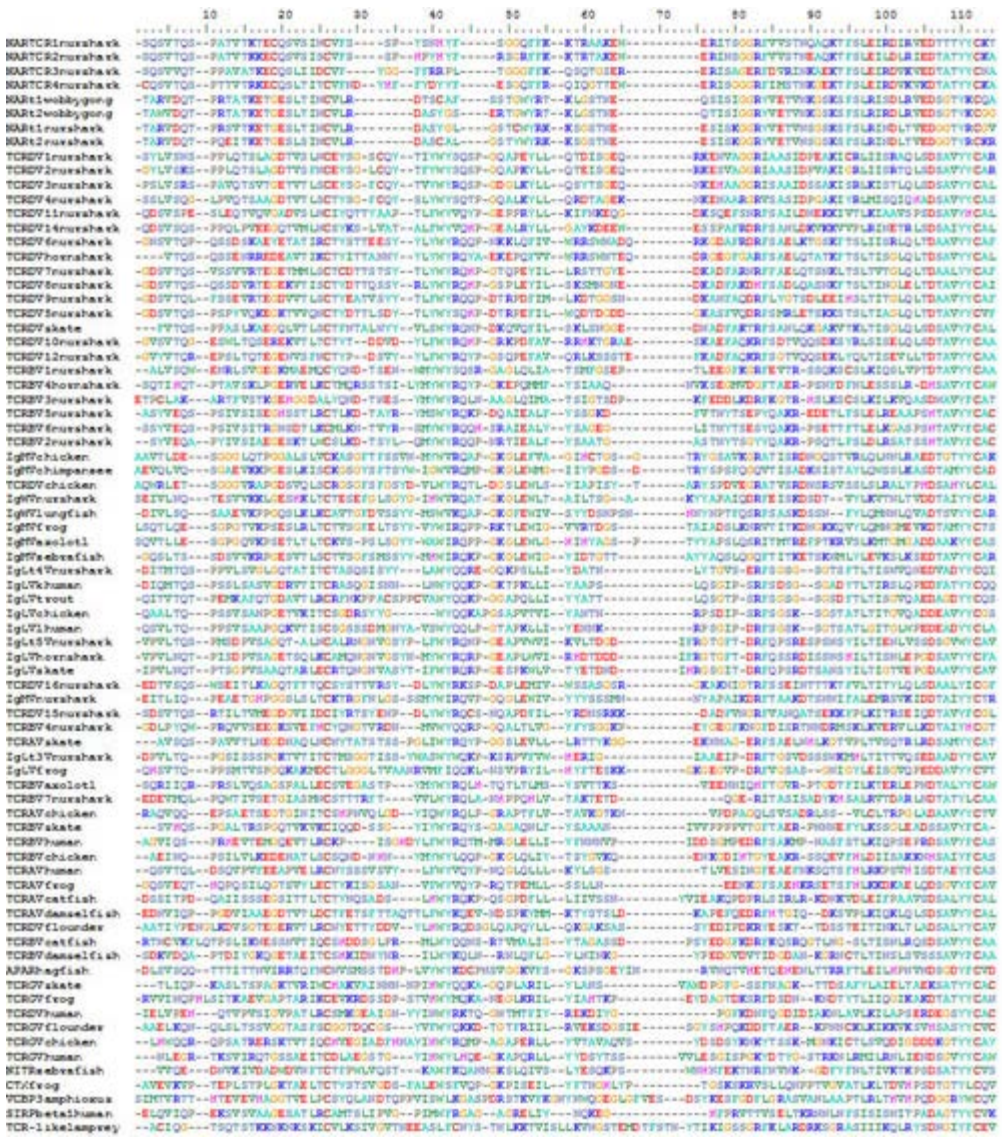

Fig. 7. Amino acid alignment of chordate variable domains used for phylogenetic analysis in Figs. 2 and 6.

Table 1. Key to numbered branch termini in Fig. 6 and database accession numbers

No.	Gene	Species	Accession no.
1	*NTV1*	Nurse shark *Ginglymostoma cirratum*	DQ022688
2	*NTV2*	Nurse shark *Ginglymostoma cirratum*	DQ022693
3	*NTV3*	Nurse shark *Ginglymostoma cirratum*	DQ022696
4	*NTV4*	Nurse shark *Ginglymostoma cirratum*	DQ022710
5	*NART1*	Spotted wobbygong *Orectolobus maculatus*	AF336092
6	*NART2*	Spotted wobbygong *Orectolobus maculatus*	AF336091

7	*NART1*	Nurse shark *Ginglymostoma cirratum*	AY114926
8	*NART2*	Nurse shark *Ginglymostoma cirratum*	U18721
9	*TRDV1*	Nurse shark *Ginglymostoma cirratum*	*
10	*TRDV2*	Nurse shark *Ginglymostoma cirratum*	*
11	*TRDV3*	Nurse shark *Ginglymostoma cirratum*	*
12	*TRDV4*	Nurse shark *Ginglymostoma cirratum*	*
13	*TRDV11*	Nurse shark *Ginglymostoma cirratum*	*
14	*TRDV14*	Nurse shark *Ginglymostoma cirratum*	*
15	*TRDV6*	Nurse shark *Ginglymostoma cirratum*	*
16	*TRDV*	Horned shark *Heterodontus francisci*	U22673
17	*TRDV7*	Nurse shark *Ginglymostoma cirratum*	*
18	*TRDV8*	Nurse shark *Ginglymostoma cirratum*	*
19	*TRDV9*	Nurse shark *Ginglymostoma cirratum*	*
20	*TRDV5*	Nurse shark *Ginglymostoma cirratum*	*
21	*TRDV*	Clearnose skate *Raja eglanteria*	U75763
22	*TRDV10*	Nurse shark *Ginglymostoma cirratum*	*
23	*TRDV12*	Nurse shark *Ginglymostoma cirratum*	*
24	*TRBV1*	Nurse shark *Ginglymostoma cirratum*	*
25	*TRBV4*	Horned shark *Heterodontus francisci*	L47447
26	*TRBV3*	Nurse shark *Ginglymostoma cirratum*	*
27	*TRBV5*	Nurse shark *Ginglymostoma cirratum*	*
28	*TRBV6*	Nurse shark *Ginglymostoma cirratum*	*
29	*TRBV2*	Nurse shark *Ginglymostoma cirratum*	*
30	*IGMV*	Chicken *Gallus gallus*	M30348
31	*IGMV*	Chimpanzee *Pan troglodytes*	XM510219
32	*TRDV*	Chicken *Gallus gallus*	AF175433
33	*IGWV*	Nurse shark *Ginglymostoma cirratum*	U51450
34	*IGWV*	Marbled lungfish *Protopterus aethiopicus*	AF437733
35	*IGMV*	African clawed frog *Xenopus laevis*	M20484

36	*IGMV*	Mexican axolotl *Ambystoma mexicanum*	L20243
37	*IGMV*	Zebrafish *Danio rerio*	AF281480
38	*IGLV4*	Nurse shark *Ginglymostoma cirratum*	AY720853
39	*IGLVK*	Human *Homo sapiens*	Y08392
40	*IGLV*	Rainbow trout *Oncorhynchus mykiss*	X65260
41	*IGLV*	Chicken *Gallus gallus*	M96973
42	*IGLVL*	Human *Homo sapiens*	AF216777
43	*IGLV5*	Nurse shark *Ginglymostoma cirratum*	U15144
44	*IGLV*	Horned shark *Heterodontus francisci*	M64307
45	*IGLV*	Clearnose skate *Raja eglanteria*	L25568
46	*TRDV16*	Nurse shark *Ginglymostoma cirratum*	*
47	*IGMV*	Nurse shark *Ginglymostoma cirratum*	M92851
48	*TRDV15*	Nurse shark *Ginglymostoma cirratum*	*
49	*TRBV4*	Nurse shark *Ginglymostoma cirratum*	*
50	*TRAV*	Clearnose skate *Raja eglanteria*	U75749
51	*IGLV*	Nurse shark *Ginglymostoma cirratum*	*
52	*IGLV*	African clawed frog *Xenopus laevis*	L76578
53	*TRBV*	Mexican axolotl *Ambystoma mexicanum*	L33400
54	*TRBV7*	Nurse shark *Ginglymostoma cirratum*	*
55	*TRAV*	Chicken *Gallus gallus*	U04611
56	*TRBV*	Clearnose skate *Raja eglanteria*	U75769
57	*TRBV*	Human *Homo sapiens*	NG001333
58	*TRBV*	Chicken *Gallus gallus*	M81149
59	*TRAV*	Human *Homo sapiens*	U76489
60	*TRAV*	African clawed frog *Xenopus laevis*	AF440816
61	*TRAV*	Channel catfish *Ictalurus punctatus*	AAB02652
62	*TRAV*	Bicolor damselfish *Stegastes partitus*	AY198372
63	*TRDV*	Bastard halibut *Paralichthys olivaceus*	AB177610
64	*TRBV*	Channel catfish *Ictalurus punctatus*	AF038162

65	*TRBV*	Bicolor damselfish *Stegastes partitus*	AF324823
66	*APAR*	Inshore hagfish *Eptatretus burgeri*	AB177610
67	*TRGV*	Clearnose skate *Raja eglanteria*	U75759
68	*TRGV*	African clawed frog *Xenopus laevis*	AF440825
69	*TRDV*	Human *Homo sapiens*	BC007313
70	*TRGV*	Bastard halibut *Paralichthys olivaceus*	AB076073
71	*TRGV*	Chicken *Gallus gallus*	U22666
72	*TRGV*	Human *Homo sapiens*	NG001336
73	*NITR*	Zebrafish *Danio rerio*	AL591482
74	*CTX*	African clawed frog *Xenopus laevis*	U43330
75	*VCBP3*	Lancelet *Branchiostoma floridae*	AF520474
76	*SIRPB1*	Human *Homo sapiens*	BC025286
77	*TCR-like*	Sea lamprey *Petromyzon marinus*	AY686861

Asterisks indicate unpublished Nurse shark *TRDV*, *TRBV*, and *IGLV* sequence data (unpublished observations).