Immunogenetics (1995) 42: 204–212

ORIGINAL PAPER

Jonathan P. Rast · Robert N. Haire · Ronda T. Litman
Susan Pross · Gary W. Litman

Identification and characterization of T-cell antigen receptor-related genes in phylogenetically diverse vertebrate species

Received: 29 March 1995 / Revised: 19 April 1995

Abstract Characterization of the structure, multiplicity, organization, and cell lineage-specific expression of T-cell receptor (*TCR*) genes of nonmammalian vertebrate species is central to the understanding of the evolutionary origins of rearranging genes of the vertebrate immune system. We recently described a polymerase chain reaction (PCR) strategy that relies on short sequence similarities shared by nearly all vertebrate *TCR* and immunoglobulin (*Ig*) variable (*V*) regions and have used this approach to isolate a *TCR* beta (*TCRB*) homolog from a cartilaginous fish. Using these short PCR products as probes in spleen cDNA and genomic libraries, we were able to isolate a variety of unique *TCR* and *TCR*-like genes. Here we report the identification and characterization of a chicken *TCR* gamma (*TCRG*) homolog, apparent *Xenopus* and pufferfish *TCR* alpha (*TCRA*) homologs, and two horned shark *TCR* delta (*TCRD*)-like genes. In addition, we have identified what could be a novel representative of the *Ig* gene superfamily in the pufferfish. This method of using short, minimally degenerate PCR primers should speed progress in the phylogenetic investigations of the *TCR* and related genes and lend important insights into both the origins and functions of these unique gene systems.

The nucleotide sequence data reported in this paper have been submitted to the EMBL/GenBank nucleotide sequence databases and have been assigned the accession numbers U22666 (*Gd*186cDNA), U22667 (*Gd*187cDNA), U22668 (*Gd*186), U22669 (*Gd*187), U22670 (*Hf*2A), U22671 (*Hf*191Y), U22672 (*Hf*191YcDNA), U22673 (*Hf*2AcDNA), U22674 (*Sn*YYC191), U22675 (*Sn*YYC193), U22678 (*Sn*YYC193cDNA), U22679 (*Xl*11), and U23067 (*Sn*YFC191)

J. P. Rast · R. N. Haire · R. T. Litman · G. W. Litman (✉)
University of South Florida, All Children's Hospital, 801 Sixth Street South, St. Petersburg, FL 33701, USA

S. Pross
University of South Florida, 12901 Bruce B. Downs Boulevard, Tampa, FL 33612, USA

Introduction

Nonmammalian models are invaluable for resolving fundamental issues relating to the control of expression of eukaryotic genes, including those of the immune system. The developmental regulation of *TCR* gene expression is of particular interest in terms of the immune function of T cells and the pathogenesis of immunological disease (Davis and Bjorkman 1988). Peptide sequencing, cross-hybridization, and polymerase chain reaction (PCR) amplification employing degenerate primers have been employed to characterize *TCR* genes in a number of different nonmammalian systems, including: the *A* (Gobel et al. 1994) and *B* (Tjoelker et al. 1990) *TCR*s in avians as well as a *TCRB*-type gene in the axolotol, a neotenic amphibian (Fellah et al. 1993), and a teleost fish (Partula et al. 1994). A number of other species that have proven useful in understanding the developmental control of gene expression are potentially valuable model systems for understanding the ontogenetic control of T-cell function; however, insufficient nucleotide sequence identity for cross-hybridization with known genes, difficulties in developing immunological reagents that are specific for TCRs (or accessory molecules), and complications in obtaining T-cell lines in some of these species have confounded the study of the structure and function of their TCR systems.

Recently, we described a PCR amplification approach employing short, minimally degenerate nucleotide primers that complement variable region amino acid motifs that are found in nearly all TCR isotypes and *Ig* light chain proteins. Using this technique we identified *TCRB* gene homologs in the horned shark (*Heterodontus francisci*, a representative chondrichthyan species; Rast and Litman 1994). We now report the application of this method to the identification of *TCR* and *TCR/Ig*-like genes in: the chicken (*Gallus domesticus*), which is well suited for ontogenetic studies of T- and B-cell lineages; the African clawed frog (*Xenopus laevis*), in which metamorphosis is associated with expression of specific genes; and the pufferfish (*Spheroides nephelus*), in which the small genome size can facilitate studies of gene

structure and organization (Brenner et al. 1993). We also describe the structure of additional *TCR* genes in the horned shark that are not related to the *TCRB* homolog described previously. These studies establish that the PCR amplification of genomic DNA employing short, minimally degenerate primers is a broadly applicable approach for identifying both *TCR* and related novel gene types.

Materials and methods

DNA isolation

Chicken, African clawed frog, pufferfish, and horned shark DNA were extracted using methods previously adapted by our laboratory (Kokubu et al. 1987).

RNA isolation, mRNA isolation, and cDNA library construction

Pufferfish RNA was extracted using the RNAzol (Biotecx, Friendswood, TX) method, and poly A+ RNA was isolated using a magnetic bead approach as described previously. cDNA was synthesized using a kit (Pharmacia LKB, Piscataway, NY) and cloned into λgt11 (Stratagene, La Jolla, CA).

PCR primers

The designations and structures of the various primer pairs utilized in these investigations are shown; cloning site(s) are boxed and preceded by spacer nucleotides. Y = C or T; R = A or G; N = A, C, G or T; and J = C or A.

TCRWYRQ	CC	GAATTC	TGGTAYCRNCA
YYCBAM	C	GGATCC	GCRCARTARTA
YFCBAM	C	GGATCC	GCRCARAARTA

While these primers have been used successfully to amplify product(s) from genomic DNA in a variety of species, they sometimes yield considerable amounts of nontarget, apparent artifactual by-product. This phenomenon appears to be species-specific, and in the case of African clawed frog, an additional primer series containing inosine (I) was synthesized in an attempt to reduce this problem (Bartl and Weissman 1994).

SallWYRQ	CC	GTCGAC	TGGTAYJGICA
AIY(F/L)C	CC	GTCGAC	GCACAGARRTAIIIRGC
AIYYC	CC	GGATCC	GCRCARTARTAIIIRGC

cDNA and genomic DNA libraries

African clawed frog (Haire et al. 1990) and horned shark (Shamblott and Litman 1989) spleen cDNA libraries have been described previously. The pufferfish cDNA library was synthesized from spleen mRNA essentially as described (Shamblott and Litman 1989). The unamplified pufferfish spleen cDNA library consists of ~1 × 10^6 pfu. The chicken (λ Zap cloning vector, Stratagene) spleen cDNA library was supplied by C.-L. Chen and M. Cooper. The amplified African clawed frog genomic DNA library has been described previously (Haire et al. 1991).

PCR priming

Approximately 0.5–1 μg of the respective genomic DNA templates was used in the PCR reactions. The standard buffer reaction consisted of 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin, 10 mM Tris, pH 8.3; 0.2 mM of each deoxyribonucleotide triphosphate; primers at 0.1 μM; 2.5 units *Taq* polymerase was added in a 100 μl final volume and was overlaid with mineral oil. The thermal cycling protocol was 1 min, 94 °C → 1 min, 45 °C → 1 min, 72 °C for 30 cycles. The final cycle was extended for 10 min at 72 °C. Ten microliters of the PCR mixture was electrophoresed on a 4% NuSieve 3:1 agarose gel (FMC, Rockland, ME), bands in the 180–210 bp range were cored with Pasteur pipettes, the plugs were placed in 100 μl TE (10 mM Tris, 1 mM ethylenediaminetetraacetate, pH 8.0) and melted at 95 °C. Individual samples (1 μl) were subjected to a second round of 25 PCR cycles, in the same buffer, with the same primers, and using the identical cycling profile indicated above. Mineral oil was extracted by shaking with chloroform, sodium acetate was added to 0.3 M, two volumes of ethanol were added, the samples were precipitated at -20 °C, and the precipitate was washed in 70% ethanol. The product was resuspended in 10 μl of TE, of which 2 μl was digested with *Eco* RI and *Bam* HI and electrophoresed in a 4% NuSieve 3:1 agarose gel. The appropriate size band was then excised and the DNA product separated from the gel matrix in a Spin-X filter (Costar, Cambridge, MA) according to the manufacturer's recommended procedure and cloned into appropriately linearized M13 (mp18 and mp19).

cDNA and genomic library screening

Spleen cDNA libraries were lifted onto charged nylon filters (Hybond N+; Amersham, Arlington Heights, IL) and hybridized with probes derived by PCR amplification of the sequenced M13 clones. The probes were labeled with ^{32}P-dCTP using a random primer method based on Feinberg and Vogelstein (1983). Hybridizations were at 65 °C for 16 h in 0.6 M NaCl, 0.02 M EDTA, 0.5% NaPyrophosphate, 0.2 M Tris, pH 8.0. The filters were washed at 52 °C in 1 × SSC (0.15 M NaCl, 0.015 M sodium citrate), 0.5% NaPyrophosphate, 0.1% sodium dodecyl sulfate and exposed to X-ray film. Genomic library screening was carried out under the same conditions.

DNA sequence analysis

The dideoxynucleotide chain termination method with ^{35}S-dATP was employed for sequencing of the initial PCR amplification products, the cDNAs, and genomic clones. The FASTDB program of IntelliGenetics (Mountain View, CA; Brutlag et al. 1990) was used for initial databank searches of genomic PCR amplification products and cDNA sequences. Parameter settings for analyses of genomic PCR product were: similarity matrix, unitary; K-tuple = 2; translation frame = 3; mismatch penalty = 1; joining penalty = 20; gap penalty = 1; window size = 32; gap size penalty = 0.05; cutoff score = 10. Searches were made against GenBank 79, and initial matches were then sorted by optimized score. For initial database searches and sequence identity comparisons, the amino acid residues included the complementary portion of the primer.

Results

Chicken TCRG

Since several groups had encountered difficulty in identifying *TCRG* genes using a number of different approaches (M. Cooper and W. McCormack, personal communications), detection of the putative chicken *TCRG* locus was used to further examine the capabilities of short primer PCR amplification. Agarose gel electrophoretic analyses of the products of PCR amplification of chicken genomic DNA employing primers (TCRWYRQ and YYCBAM/YFCBAM) yielded distinct bands in the ~190 base pair (bp) range. The ~190 bp fragment was isolated and subcloned; 18 different clones were selected for DNA sequenc-

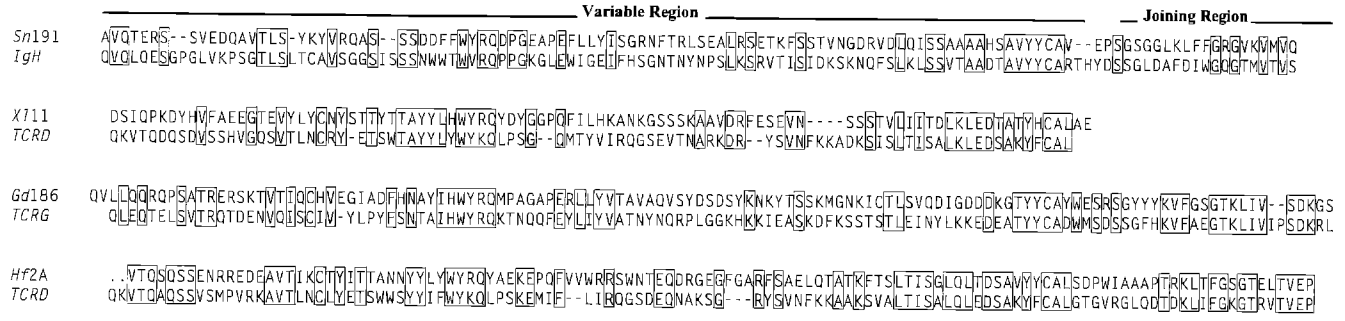
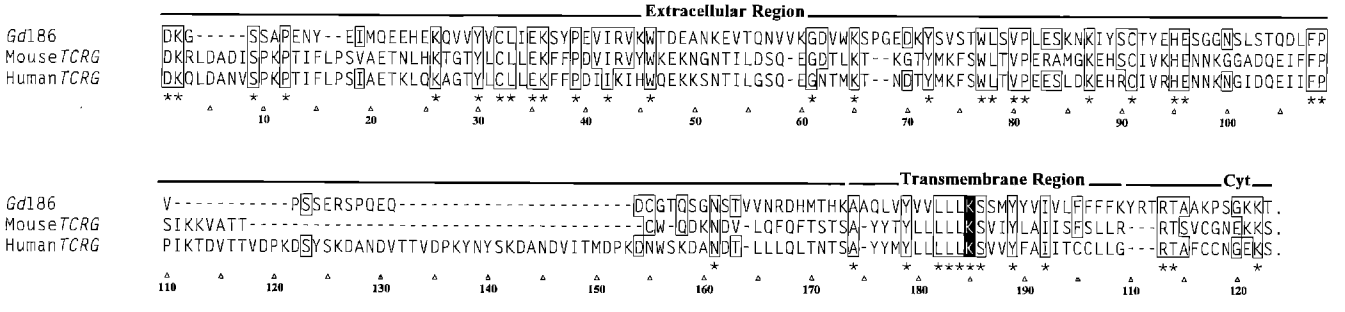


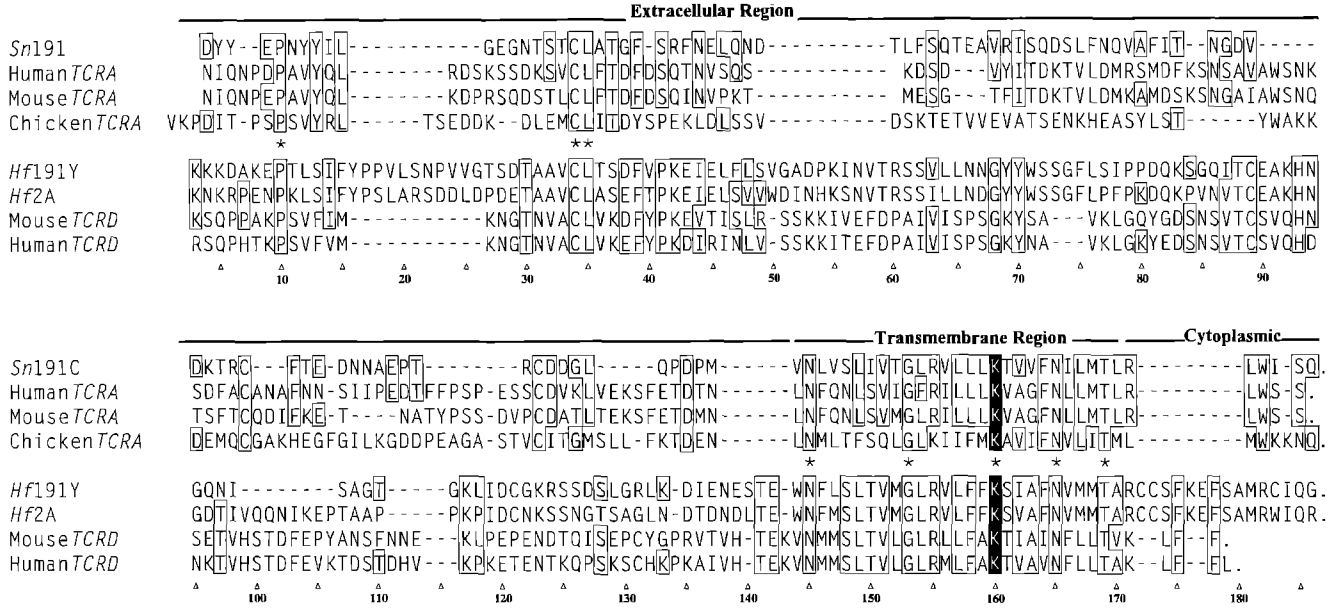
Fig. 1 Alignments of *V* regions selected with probes complementing *V* region segments of genomic DNA with best matches from a search of GenBank using FASTDB. Highest scoring matches are illustrated to establish that sequences are *Ig/TCR*-type *V* regions and are not intended to classify a given sequence as homologous to a particular mammalian gene type (e. g., *IgH*, *TCRD*, etc.). Top matches are as follows: *Sn*191, human *IgH* (HSIGHXX2); *Xl*11, cattle *TCRD* (BOVBTDV2); *Gd*186, mouse *TCRG* (MUSTCRGZZX); *Hf*2A, human *TCRD* (HSTCRDR). *J* regions are included except in the case of *Xl*11, which represents a genomic *V* segment

ing. Two clones (*Gd*186 and *Gd*187) were identified as *TCR* candidates based on identity with *Ig* and *TCRV* regions, including the essentially invariant Leu[75] and Asp[84] (Kabat et al. 1991). The GenBank sequences exhibiting highest amino acid identity with the two candidate segments were 30%–33% identical (excluding the sequence segments complemented by the primer sequences) and include mammalian *Ig* V_H and horned shark light chain (HFL141). A mouse *TCRG* (MUSTCRGZZX) gene produced the tenth highest score and was the only *TCR* in the ten highest optimized scores. The analyses with the chicken amplification products are typical of our experience with this general approach, in which the initial findings with genomic segments do not consistently predict the ultimately determined identity of the individual amplification product(s), which can be determined when constant (*C*) region sequence data are obtained in cDNA analyses. Because the chicken possesses only single V_H and *Ig* V_L families (Reynaud et al. 1985, 1989) and both *TCRBV* (Tjoelker et al. 1990) and *TCRAV* (Gobel et al. 1994) sequences are known, it is likely that gene segments *Gd*186 and *Gd*187 represent new *TCR* segments.

The M13 mp18 clones containing the candidate chicken V_{TCR} segments (*Gd*186 and *Gd*187) were subjected to PCR amplification in order to derive gene-specific probes, which were then each used to screen ~6 × 10^5 chicken spleen cDNA recombinant pfus. Four positively hybridizing plaques were purified and sequenced. The predicted amino acid sequences of the *V* and *C* regions of the *Gd*186cDNA are illustrated in Figures 1 and 2, respectively. The alignment shown of the *V* and joining (*J*) region is with mouse *TCRG* and consists of 46/121 identities. The highest-ranking alignment score is with *TCRG*, when the *J* region is included; however, the basis for the assignment of this clone as a *TCRG* homolog is based on extensive identities of the associated *C* region with the corresponding regions of the mouse and human *TCRG* genes. Specifically, 26 complete identities are shown for chicken, mouse, and human sequences in the putative extracellular Ig domain (defined here as from the beginning of the *C* region to 20 amino acids beyond the second conserved *C* region cysteine; residues 1-110 in Figure 2). In addition, *Gd*186cDNA shares ten additional extracellular Ig domain identities with either mouse or human (Pelicci et al. 1987) *TCRG* that they do not share with each other. By comparison, human and mouse *TCRG* share 78 identities in the extracellular Ig domain. Few identities are evident in the extracellular spacer region (residues 111-173 in Figure 2), which contains extensive deletions and/or insertions. The transmembrane regions of the chicken *Gd*186cDNA, mouse, and human genes are identical at 9 of 25 positions, including a conserved Lys[312] (position 185 in Figure 2) that probably

Fig. 2 Alignment of putative chicken *TCRGC* region with mouse and human *TCRG* prototypes. *Dashes* indicate gaps introduced to optimize alignment. Residues matching the chicken sequence are enclosed in *boxes*. *Asterisks* indicate residues conserved in all three sequences. *Periods* at the end of sequences indicate stop codons. A lysine, which has been identified in all *TCR*s and postulated to interact with the γ chain of *CD3*, is shown in reverse image, in the transmembrane region. Extracellular and cytoplasmic (Cyt) regions are designated accordingly. The human γ sequence is taken from Pelicci and co-workers (1987). The mouse γ is from Hayday and co-workers (1985)





interacts with CD3γ in mammals (Alcover et al. 1990). The mouse and human sequences share additional identities with each other in this region that are not shared with the chicken sequence. Six of 14 residues in the short cytoplasmic region (position 110–123 in Figure 2) of the *Gd*186cDNA show identity with the mouse and/or human sequence. An additional cDNA representing a second *TCRGV* region family was identified with the probe derived from the *Gd*187 PCR product (data not shown). Finally, the designation of the *Gd*186/187 as *TCRG* genes is supported entirely by various molecular genetic studies in which they have been employed as probes (C.-L. Chen and M. D. Cooper, personal communication).

Fig. 3 Alignment of putative pufferfish clone *Sn*191C and horned shark clones *Hf*191Y and *Hf*2A *C* regions with human, mouse, and chicken *TCRA* (GDU04611), and human and mouse *TCRD*. *Dashes* indicate gaps introduced to optimize alignment. Residues shared between fish and *TCRA* sequences and between shark and *TCRD* sequences are enclosed in *boxes*. *Asterisks* indicate residues conserved among all shown *TCRA*, *TCRD*, and query sequences. *Periods* at the end of the sequences indicate stop codons. A lysine, which has been described in all *TCR*s and postulated to interact with the γ chain of *CD3* is shown in reverse image in the transmembrane region. All sequences are aligned to each other, as it is difficult to establish homology to tetrapod *TCRA* vs *TCRD*. GenBank accession numbers of the sequences used in the alignment are as follows: human $TCRAD_2$, HSTCELA1; mouse *TCRA*, MMU07658; chicken *TCRA*, GDU04611; mouse *TCRD*, MMTRDNU; human *TCRD*, HUMTCRGC

Pufferfish TCRA-like gene and a novel V-related gene

The same approach that had been employed with chicken DNA was adapted successfully with pufferfish DNA. In this case, PCR amplification using primer sets TCRWYRQ/YYCBAM and TCRWYRQ/YFCBAM yielded relatively uncomplicated patterns. Prominent PCR bands in the ~190 bp range were recovered and subcloned; 15 individual clones were subjected to complete sequence analysis. Despite this relatively small sampling, three sequence types (represented by *Sn*YYC191, *Sn*YFC191, and *Sn*YYC193) were found to exhibit significant identity at the predicted peptide level with V_L/V_{TCR} segments. The sequence of the *Sn*YYC191 genomic amplification product exhibits ~38% amino acid identity with horned shark light chain, ~38% identity with mouse *TCRAV8* (S53024), and 42% identity with dog *TCRAV* (DOGTCAA). A full copy length cDNA (SnYYC191cDNA) was recovered when this sequence segment was employed to screen a spleen cDNA library. Identities of *Sn*YYC191cDNA with a human V_H (HUMIGHGU) sequence are evident in both V_H and J_H region segments; however, this gene could be a pseudogene, owing to the absence of the highly conserved Cys^{23} (Kabat et al. 1991; Fig. 1). The *C* region of *Sn*YYC191cDNA exhibits significant sequence identity with mammalian *TCRAC* and *TCRDC* regions. However, this particular clone contains an apparent frameshift (based on identity with mammalian prototypes). A probe complementing the putative *C* region of *Sn*YYC191cDNA was derived and used to screen the same cDNA library. Three additional, unique clones with identical sequences and open reading frames (resulting from a 2 bp insertion relative to *Sn*YYC191cDNA) were identified, and the predicted amino acid sequence of one of these (*Sn*191C) is shown in Figure 3 relative to several tetrapod *TCRAC* regions. When the sequence of the same *TCR* is compared with mammalian *TCRD* sequences, it is apparent that the pufferfish sequence is more *A*-like, sharing 18 identities in the extracellular region (residues 1-120 in Figure 3) with human and mouse *TCRA*; most of these occur in the extracellular Ig domain. Only nine positions in the extracellular region are shared by all four species. Ten additional positions are shared by the pufferfish *TCR* and either the human or mouse sequences, but not both. Chicken *TCRA* (GDU04611) shares five additional identities with pufferfish, which are not shared with the mammalian prototypes. The transmembrane and cytoplasmic regions of all four genes are closely related.

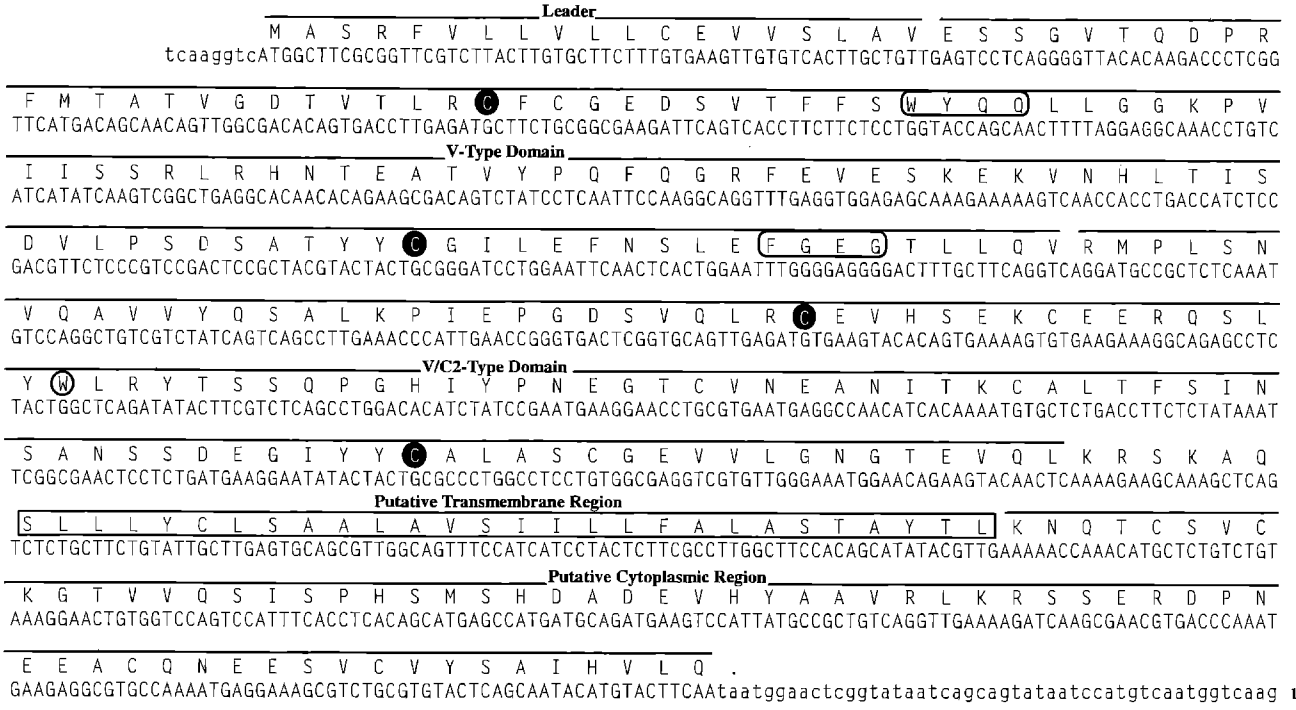


Fig. 4 Nucleotide sequence and predicted peptide sequence of pufferfish sequence *Sn*193cDNA. Putative leader, extracellular Ig domains, transmembrane, and cytoplasmic regions are designated. *TCR/Ig/CD8*-like V (WYQQ) and *J* (FGXG) motifs are *circled*. The conserved cysteine and tryptophan residues in the Ig domain are *circled*. The stop codon is indicated by a *period*

By contrast, *Sn*YFC191, which was amplified using a different 3′ primer, YFCBAM, exhibits significant identity only with *TCR*s, including ~38% maximal nucleotide identity with *TCRB* (OSU18125_1, rainbow trout). *Sn*YFC191 appears to represent a *TCRB* homolog, by comparison with various vertebrate prototypes. When this amplified segment was used as a probe, no hybridization was evident with a spleen cDNA library consisting of ~6×10^5 recombinants, thus precluding further assignment. This lack of hybridization of candidate V_{TCR} genomic sequences is to be expected, since *V* segments of *TCR*s within a single species are highly diverse (Concannon et al. 1986; Klein et al. 1987) and expression of specific *V* segments is known to be both tissue- and developmental stage-specific.

*Sn*YYC193 exhibits ~37% identity with human V_H and V_K (HUMIGKAZB) but lacks identity with *TCR*s within the top 17 comparison sequences. When used to screen the spleen cDNA library, *Sn*YYC193 hybridized to one cDNA clone (*Sn*YYC193cDNA) and its nucleotide and predicted amino acid sequences are shown in Figure 4. Although *Sn*YYC193cDNA contains an *Ig/TCR/CD8B*-like *V* region followed by an Ig domain, it is most probably not a *TCR*. By comparison with *Ig*, *TCR*, and various cell surface antigens that are recognized as members of the *Ig* gene superfamily, several protein domains in *Sn*YYC193 can be assigned tentatively: leader, *V*-type domain, and a second *V*- or *C2*-type domain (Williams and Barclay 1988). Putative transmembrane and cytoplasmic regions were assigned on the basis of hydropathicity analyses (Kyte and Doolittle 1982). At the predicted peptide level, the second Ig domain most closely matches a V_H region but at a low percentage of identity. The transmembrane and cytoplasmic regions lack significant identity with sequences contained in GenBank. Although the genomic sequence of *Sn*YYC193cDNA has not yet been determined, a PCR from the *WYRQ* region to the putative *J* region using genomic DNA as a template produced a band of a size equal to that produced with the cDNA, inconsistent with the existence of a rearranging mechanism. It is unlikely that this gene is of the germline-joined types that we have described previously for immunoglobulins of cartilaginous fish (Kokubu et al. 1988; Rast et al. 1994). These findings emphasize the ability of the short primer PCR approach to identify genetic systems, both rearranging and nonrearranging, which relate to *Ig* and *TCRs* through shared Ig domain sequences.

Genomic sequence of a TCR-like gene from the African clawed frog

African clawed frog genomic DNA was PCR-amplified employing primers that had proven successful with both chicken and pufferfish DNA. Amplification products in the appropriate size range were recovered, along with an unusually large number of products that were not apparent *Ig* or *TCR* homologs. Although two *TCRA*-like sequences were identified, neither hybridized with spleen cDNAs (data not shown). Based on previous experience with African clawed frog *Ig V* genes, which are highly derived (Haire et al. 1990), it was not unreasonable to assume that minor sequence differences between African clawed frog DNA and the primers could preclude amplification. In

Fig. 5 Complete nucleotide sequence of African clawed frog genomic clone *Xl*11. Noncoding sequence is in *lowercase* and putative coding sequence in *uppercase*. Conserved leader donor and acceptor splice site nucleotides are in *reverse image squares*. Conserved cysteines are in *reverse image circles*. Conserved WYRQ is enclosed, as are recombination signal sequences, which are shown in *uppercase* and separated by 23 nucleotides

addition, as indicated above, our experience with other systems suggests that interference/competition from various genomic DNA sequences can complicate amplification and identification. Consequently, a second set of PCR primers, complementing an additional position and containing an inosine triplet, integrated to bridge a nonconserved codon, was synthesized. A single sequence of interest was identified as a *V* region of the novel ρ light chain isotype (Stewart et al. 1993), using the AIYYC primer together with the TCRWYRQ or Sal1WYRQ primer. Equal intensity products corresponding to ~180, 190, 200, and 220 bp were amplified using the third framework region (*FR3*) primer AIY(F/L)C. These products were electrophoretically resolved and subcloned in M13 mp18/19. Approximately 10 subclones corresponding to each band were subjected to DNA sequence analysis. Of these isolates, the 180, 200, and 220 bp components lack sequence identity with *Ig* and *TCRs*. Similarly, most clones derived from the ~190 bp product lack either an open reading frame or significant sequence identity with *Ig*-type genes. However, one sequence was identified that exhibits similarity to the rat *Ig* light chain and to *TCRAV*. A probe complementing the *TCRAV*-like sequence was employed to screen a spleen cDNA library (~5 × 10^5 recombinant pfu), but no positively hybridizing plaques were identified. Similarly, hybridization was not detected in high representation screenings of either a phorbol myristate acetate-stimulated African clawed frog thymus cDNA library or a cDNA library derived from thymus, spleen, and liver of this species (kindly provided by M. Flajnik). Screening of an amplified genomic library (~2.4 × 10^5 pfu), representing >0.5 equivalent of the African clawed frog genome, yielded six hybridizing plaques. Partial restriction mapping of the purified phage DNA showed these clones to be identical; hybridization was localized to ~1.5 kilobases (kb) and ~0.3 kb *Eco* RI segments, which were subcloned and sequenced. The genomic fragments contained sequence that is identical with the amplified *V* segment (product between primers) interrupted by an internal *Eco* RI restriction site. A recombination signal sequence (*RSS*) is present 3′ to the putative coding region *Xl*11 (Fig. 5). The sequence of this clone was extended to include ~1.3 kb upstream of the *V* segment and 200 bp 3′ of the *RSS*. Notably, the 5′ sequence of this genomic clone did not contain an *Ig*-type octamer such as has been identified ~200–300 bp upstream of the initial codon in 10 of the 11 different V_H gene families in African clawed frog (Haire et al. 1990) as well as in all vertebrate *Ig* light chain genes.

A second class of TCR in a cartilaginous fish

The (TCRWYRQ, YYCBAM, and YFCBAM) primer pairs described above also were used to amplify additional segments of horned shark genomic DNA. In this case, 35 segments were subcloned and sequenced and five separate classes of *V* region-like sequences were identified. Probes complementing these *V* segments were used to screen a spleen cDNA library. One of the probes detected a *TCRB* homolog and has been reported previously (Rast and Litman 1994). Two of these probes failed to hybridize with spleen cDNA, and two other probes (*Hf*2A and *Hf*191Y) hybridized with two cDNAs, both of which were sequenced. The *V* and putative *J* regions of the *Hf*2AcDNA are shown in Figure 1, and the *C* regions of *Hf*2AcDNA and *Hf*191YcDNA are shown in Figure 3. The sequence of the corresponding *V* and *J* regions of *Hf*191FY is truncated but nevertheless is dissimilar to that of *Hf*2AcDNA over a significantly informative segment. Presumably this represents a sequence family difference as has been observed for β segments in this species (Rast and Litman 1994; N. Hawke and J. Rast, unpublished data). Forty-three identities are evident between the *V* regions of *Hf*2AcDNA and human *TCRD* (HSTCRDR), and additional identities are found in the *J* region. This level of sequence identity between the shark gene and mammalian prototypes is

similar to that described previously for the shark *TCRB* gene (Rast and Litman 1994). The relationships of the putative *C* region segments of *Hf*2AcDNA and *Hf*191YcDNA to tetrapod *TCRD* and *TCRA* are illustrated in Figure 3. Twenty-four identities are present between both horned shark sequences and at least one of the mammalian *TCRD* prototype sequences in the Ig domain region (residues 1-110 in Figure 3). There is less identity in the 3′ extracellular region (residues 111-144 in Figure 3), but extensive identity is found in the respective transmembrane regions, as is the case for pufferfish and chicken *TCRA* sequences. Few identities are present in the cytoplasmic region. Although the nucleotide sequences of the transmembrane, cytoplasmic, and 3′ untranslated regions (not shown) of *Hf*191YcDNA and *Hf*2AcDNA are nearly identical, they exhibit only ~75% amino acid relatedness in extracellular sequence, with large indels in the 3′ Ig domain region. Although it is possible that these genes are of the α-type or could be representative of another type, unique to the cartilaginous fishes, the identities of both *V* and *C* equivalent regions to *TCRD* are the tentative basis for their assignment to this *TCR* class. Nevertheless, the distinct possibility exists that the α/δ divergence took place after the divergence of the cartilaginous fishes and the lineage leading to the mammals, making the α/δ distinction in a shark inappropriate.

Discussion

By analogy to earlier observations in human and mouse, which showed extensive diversification of *TCRV* regions as well as both tissue-restricted and developmental stage-specific expression of these genes, it was not unexpected that major difficulties would be encountered in identifying *TCR* genes in more divergent vertebrate species using both cross-hybridization and conventional PCR priming approaches. Although the avian *TCRB* locus has been identified using a pool of DNA probes complementing different mammalian *TCRV* regions (Tjoelker et al. 1990), the avian α locus could not be detected using this approach (M. D. Cooper and C.-L. Chen, personal communication) but was identified using a *CD3*-specific antisera to isolate cell surface TCR molecules, which in turn were subjected to peptide sequencing. The peptide sequence served as the basis for synthesis of gene-specific oligodeoxynucleotide probes. PCR amplification employing degenerate primers proved successful for identifying the *TCRB* locus in one amphibian species, the Axolotl (Fellah et al. 1993); however, our efforts to utilize this technology for this purpose with African clawed frog and shark have not been successful (unpublished observation). Since the *TCRBC* region shows the greatest identity with *Ig* light chains, approaches targeting these identities could be less successful in the identification of other *TCR* types. The amplification of a *TCRB* homolog from shark genomic DNA using minimal degeneracy primers (8- to 16-fold), which complement only 10–11 nucleotides, suggested that this approach could have broader applications for identifying *TCR* genes in other nonmammalian vertebrate species in which neither nucleotide nor peptide sequence information is available for a *TCR* glycoprotein(s), gene(s), or accessory molecule(s). Furthermore, it is reasonable to predict that other (nonclassical *Ig/TCR*) gene families sharing only short sequence identities in *FR2* and *FR3* could be amplified from these species since other known genes (e.g., *CD8B*) also share a variant of this motif.

The studies reported here demonstrate that a variety of *TCR*-like sequences can be primed directly from the genomic DNA of several different phylogenetically primitive species of jawed vertebrates, which are also significant developmental and/or phylogenetic models. The use of genomic DNA as the PCR template for identification of *TCR* homologs is advantageous as there need be no assumptions made regarding tissue- and developmental stage-specific expression. However, a sizable number of unwanted products are produced, the degree of which appears to be species-specific. Thus, although the amplification and cloning steps are relatively straightforward, the approach is dependent on the analysis of sizable numbers of sequences and extensive database reference analyses. Parallel genomic Southern blotting analyses can be used to confirm and partially characterize products and as a means for excluding cross-species PCR contamination. The primers employed in the studies reported here have been designed to eliminate amplification of *Ig* V_H genes; however, it is not possible to restrict amplification of V_L genes, as both V_{TCR} and V_L have the same primer-targeted sequence. The less extensively diversified nature of *Ig* light chain genes in some species minimizes difficulties in resolving their identity in relation to other gene families.

Assignment of genomic amplification products as *Ig/TCR*- like *V* segments is based only in part on the alignment with family-specific residues. While no positions within the PCR product are absolutely conserved throughout all *V* region proteins, genuine products typically can be identified both by the presence of consensus residues and through database searches, which tend to show matches exceeding 30% overall identity for authentic products. The FASTDB program (Brutlag et al. 1990) is helpful for further characterizing the genomic amplification products; however, several findings in this study demonstrate that the rank ordering of minimally homologous sequence segments can be misleading, as established by subsequent resolution of corresponding full copy length cDNA sequences. Unequivocal assignment of a putative V_T segment as a *TCR* component remains largely dependent on the resolution of a full copy length cDNA sequence, assuming that it is sufficiently homologous to existing database sequences. However, isolation of a corresponding cDNA can be problematic. For example, we identified a large number of PCR fragments from *Danio rerio* (zebrafish) that most likely are representative of several different *TCR* classes but thus far have failed to recover corresponding *TCR*s from several cDNA libraries. This is not unexpected as levels of representation of less than one in 5×10^5 are typical for the gene segments for which corresponding

cDNAs have been identified. Since *TCR*s typically consist of a number of highly diversified *V* region families but relatively few, related *C* region genes, once a *C* region probe has been generated, there is no difficulty in identifying greater numbers of transcripts (cDNAs). In instances where screening with amplified *V* region products has failed, this property of the *TCR* gene family may be exploited in an alternative selection strategy, in which *C* region segments derived from closely related species can be used as probes.

In many cases, homologous cDNAs cannot be found, although their genomic equivalent can be identified. As a general principle, identification of the tissue (mRNA) source for *TCR*-like products that cannot be identified in conventional libraries could employ libraries that are obtained from other tissues and which are representative of different developmental stages and/or are normalized. This latter approach can achieve increases of several order of magnitude in the relative abundance of rare mRNAs through kinetic enrichment (Morgan et al. 1992; Patanjali et al. 1991).

Although it is desirable to have cDNA probes in most studies, if a corresponding cDNA sequence cannot be identified, presumptive identification of a *V* element can be substantiated further by isolating corresponding genomic segments from high representation genomic libraries. Identification of typically conserved sequence features – for instance, recombination signal sequences, leader peptide, and regulatory (promoter) elements – or lack thereof can provide evidence for the identity of these products. If the primed product is not of the *Ig/TCR* rearranging-type, predicting a peptide structure in the absence of mRNA (cDNA) information is considerably more difficult and would require extensive genomic sequencing and/or could utilize exon trap vector approaches (Auch and Reth 1990). Different anchored PCR approaches (see below) also could facilitate identification.

Although the primers used in the present investigation potentially would complement the most common *TCR/IgL* sequences, by no means do they encompass all of the known variants of these motifs. For example, *Ig* heavy chains and the *CD8B* chain have a hydrophobic amino acid in place of the tyrosine in the WYRQ motif (e.g., WVRQ, WLRQ, WIRQ, etc.). Because these motifs are short, it is feasible to synthesize primer sets that exhaust the most common variant sequence possibilities. To avoid exclusion of *V* region PCR products resulting from internal restriction digestion at the cloning step, primers containing a different set of restriction sites can be used in either the primary or secondary PCR. It also is important to note that the analyses of these species with the primer sets described here are far from exhaustive, as relatively few sequences (less than 50) have been analyzed in each case, and more important, in no case has any negative selection technique been implemented to increase the efficiency of screening large numbers of cloned products. These various issues are being addressed in ongoing studies directed at further characterizing the nature of molecules related through the *Ig/TCR V* region domain.

The existence of *Ig* or *TCR* homologs in the extant agnathan vertebrates has not been established. Thus far, although we have sequenced numerous PCR fragments of the predicted, appropriate size, we have been unable to recover *Ig/TCR V* region-related genes from either hagfish (*Eptatretus stouti*) or lamprey (*Petromyzon marinus*). The fact that this technique works readily in all of the broadly representative jawed vertebrate species in which it has been applied, but not in these agnathans, reflects a significant difference between the groups and is consistent with previous functional studies and the failure of numerous earlier attempts to identify true *Ig/TCR* gene homologs in these species. However, these negative findings do not preclude the existence of these and/or related genes in the contemporary agnathans. Homologous genes could be present at low copy number in the agnathan genome and thus be below the kinetic threshold of the technique as applied presently. Alternatively, the *Ig/TCR* light chains/*CD8* divergence may have taken place after that of the jawed vertebrates, and the related genes in the Agnatha may have sequences that differ from the motifs used to construct primers. Finally, if the agnathan genes have diversified following the divergence of the jawed vertebrates, amplification could be extremely difficult to achieve. These studies are being expanded using different primer series and priming parameters.

In addition to the utility of the minimal degeneracy approach for identifying *TCR* homologs in the jawed vertebrates, the method as detailed here, either in its present form or through modification of primer sequences, is potentially capable of amplifying homologous genes, should these exist, in the extant, more phylogenetically primitive species. While rearranging *Ig* and/or *TCR* genes could be restricted to the jawed vertebrates, the possibility exists that progenitor gene systems could be present in more phylogenetically distant forms. Likewise, owing to the low levels of sequence assumption inherent in this approach, there is good reason to believe that the method has considerable potential for uncovering heretofore unknown, homologous genes in mammals and other species that are either rearranging or nonrearranging but nevertheless share a phylogenetic origin with the *Ig*s and *TCR*s.

Acknowledgments We thank Barbara Pryor and Victoria Haire for editorial assistance, Dr. Chen-Lo Chen and Dr. Max D. Cooper for chicken DNA and spleen cDNA library as well as for valuable suggestions regarding the investigation of the *TCRG* gene, and Tatsuya Ota for his constructive criticism of the manuscript. Our thanks also to the Florida Department of Environmental Protection, Marine Research Institute for supplying the *Spheroides nephelus* specimens. This work was supported by a grant to G. W. L. by the National Institutes of Health RO1AI23338. All clones described in this manuscript will be made available for distribution to interested parties.

References

Alcover, A., Mariuzza, R. A., Ermonval, M., and Acuto, O. Lysine 271 in the transmembrane domain of the T-cell antigen receptor beta chain is necessary for its assembly with the CD3 complex but not for alpha/beta dimerization. *J Biol Chem 265:* 4131–4135, 1990

Auch, D. and Reth, M. Exon trap cloning: Using PCR to rapidly detect and clone exons from genomic DNA fragments. *Nucleic Acids Res 18:* 6743–6744, 1990

Bartl, S. and Weissman, I. L. PCR primers containing an inosine triplet to complement a variable codon within a conserved protein-coding region. *Biotechniques 16:* 246–248, 1994

Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B., and Aparicio, S. Characterization of the pufferfish (Fugu) genome as a compact model vertebrate genome. *Nature 366:* 265–268, 1993

Brutlag, D. L., Dautricourt, J. P., Maulik, S., and Relph, J. Improved sensitivity of biological sequence database searches. *Comp App Bio Sci 6:* 237–245, 1990

Concannon, P., Pickering, L. A., Kung, P., and Hood, L. Diversity and structure of human T-cell receptor β-chain variable region genes. *Proc Natl Acad Sci USA 83:* 6598–6602, 1986

Davis, M. M. and Bjorkman, P. J. T-cell antigen receptor genes and T-cell recognition. *Nature 334:* 395–402, 1988

Feinberg, A. P. and Vogelstein, B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem 132:* 6–13, 1983

Fellah, J. S., Kerfourn, F., Guillet, F., and Charlemagne, J. Conserved structure of amphibian T-cell antigen receptor beta chain. *Proc Natl Acad Sci USA 90:* 6811–6814, 1993

Gobel, T. W. F., Chen, C.-L. H., Lahti, J., Kobota, T., Kuo, C.-L., Aebersold, R., Hood, L., and Cooper, M. D. Identification of T cell receptor alpha-chain genes in the chicken. *Proc Natl Acad Sci USA 91:* 1094–1098, 1994

Haire, R. N., Amemiya, C. T., Suzuki, D., and Litman, G. W. Eleven distinct V_H 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 171:* 1721–1737, 1990

Haire, R. N., Ohta, Y., Litman, R. T., Amemiya, C. T., and Litman, G. W. The genomic organization of immunoglobulin V_H genes in *Xenopus laevis* shows evidence for interspersion of families. *Nucleic Acids Res 19:* 3061–3066, 1991

Hayday, A. C., Saito, H., Gillies, S. D., Kranz, D. M., Tanigawa, G., Eisen, H. N., and Tonegawa, S. Structure, organization, and somatic rearrangement of T cell gamma genes. *Cell 40:* 259–269, 1985

Kabat, E. A., Wu, T. T., Foeller, C., Perry, H. M., and Gottesman, K. *Sequences of Proteins of Immunological Interest, edn. 5,* Washington, D. C. U. S. Dept. Health and Human Services, 1991

Klein, M. H., Concannon, P., Everett, M., Kim, L. D. H., Hunkapiller, T., and Hood, L. Diversity and structure of human T-cell receptor α-chain variable region genes. *Proc Natl Acad Sci USA 84:* 6884–6888, 1987

Kokubu, F., Hinds, K., Litman, R., Shamblott, M. J., and Litman, G. W. Extensive families of constant region genes in a phylogenetically primitive vertebrate indicate an additional level of immunoglobulin complexity. *Proc Natl Acad Sci USA 84:* 5868–5872, 1987

Kokubu, F., Litman, R., Shamblott, M. J., Hinds, K., and Litman, G. W. Diverse organization of immunoglobulin V_H gene loci in a primitive vertebrate. *EMBO J 7:* 3413–3422, 1988

Kyte, J. and Doolittle, R. F. A simple method for displaying the hydopathic character of a protein. *J Mol Biol 157:* 105–132, 1982

Morgan, J. G., Dolganov, G. M., Robbins, S. E., Hinton, L. M., and Lovett, M. The selective isolation of novel cDNAs encoded by the regions surrounding the human interleukin 4 and 5 genes. *Nucleic Acids Res 20:* 5173–5179, 1992

Partula, S., Fellah, J. S., de Guerra, A., and Charlemagne, J. Caractérisation d'ADNc de la chaine β de récepteur des lymphocytes T chez la truite arc-en-ciel. *C R Acad Sci Paris, Sciences de la vie 317:* 765–770, 1994

Patanjali, S. R., Parimoo, S., and Weissman, S. M. Construction of a uniform-abundance (normalized) cDNA library. *Proc Natl Acad Sci USA 88:* 1943–1947, 1991

Pelicci, P. G., Subar, M., Weiss, A., Dalla-Favera, R., and Littman, D. R. Molecular diversity of the human T-gamma constant region genes. *Science 237:* 1051–1055, 1987

Rast, J. P., Anderson, M. K., Ota, T., Litman, R. T., Margittai, M., Shamblott, M. J., and Litman, G. W. Immunoglobulin light chain class multiplicity and alternative organizational forms in early vertebrate phylogeny. *Immunogenetics 40:* 83–99, 1994

Rast, J. P. and Litman, G. W. T cell receptor gene homologs are present in the most primitive jawed vertebrates. *Proc Natl Acad Sci USA 91:* 9248–9252, 1994

Reynaud, C.-A., Anquez, V., Dahan, A., and Weill, J.-C. A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. *Cell 40:* 283–291, 1985

Reynaud, C.-A., Dahan, A., Anquez, V., and Weill, J.-C. Somatic hyperconversion diversifies the single V_H gene of the chicken with a high incidence in the D region. *Cell 59:* 171–183, 1989

Shamblott, M. J. and Litman, G. W. Complete nucleotide sequence of primitive vertebrate immunoglobulin light chain genes. *Proc Natl Acad Sci USA 86:* 4684–4688, 1989

Stewart, S. E., Du Pasquier, L., and Steiner, L. A. Diversity of expressed V and J regions of immunoglobulin light chains in *Xenopus laevis. Eur J Immunol 23:* 1980–1986, 1993

Tjoelker, L. W., Carlson, L. M., Lee, K., Lahti, J., McCormack, W. T., Leiden, J. M., Chen, C.-L. H., Cooper, M. D., and Thompson, C. B. Evolutionary conservation of antigen recognition: the chicken T-cell receptor beta chain. *Proc Natl Acad Sci USA 87:* 7856–7860, 1990

Williams, A. F. and Barclay, A. N. The immunoglobulin superfamily-domains for cell surface recognition. *Annu Rev Immunol 6:* 381–405, 1988