Available online at www.sciencedirect.com
SCIENCE DIRECT®
Developmental and Comparative Immunology 28 (2004) 781–792
Developmental & Comparative Immunology
www.elsevier.com/locate/devcompimm

Allelic polymorphism of TCRα chain constant domain genes in the bicolor damselfish

Michael F. Criscitiello[a,b], Sondra M. Kamper[a], E. Churchill McKinney[c,*]

[a]*Department of Microbiology and Immunology, University of Miami School of Medicine, Miami, FL 33101, USA*
[b]*Department of Microbiology and Immunology, Univ. Maryland School of Medicine, 655 W. Baltimore St, 13-009, Baltimore, MD 21201, USA*
[c]*Department of Microbiology, University of Mississippi Medical Center, 2500 N. State Street, Jackson, MS 39216, USA*

Received 1 October 2003; revised 19 December 2003; accepted 23 December 2003

Abstract

T cell receptor (TCR) chains are composed of two extracellular domains, the membrane-distal variable domain and the membrane-proximal constant domain. Data presented here show that the *TCRA* 'constant' (C) domain of damselfish exhibits considerable allelic polymorphisms that appear to be positively selected. Each of 32 damselfish *TCRAC* clones showed different patterns of atypical polymorphism in the constant region. Twenty-three of the 121 TCRα constant region amino acid residues show substitutions, clustered mainly in the loops between the beta strands. Coding regions of the *TCRAC* genes differ by up to 8% at the nucleotide level and 20% at the amino acid level. Southern hybridization, polymorphism segregation, and genomic cloning data suggest allelic polymorphism at two *TCRAC* genes, distinguished by a single amino acid. K_A/K_S ratios suggest that balancing selection is acting to maintain polymorphisms at the variable sites of one of these genes, but not the other, in a manner comparable to the peptide binding regions of MHC. Nonetheless, each *TCRAC* gene is spliced to variable and joining segments similar to those described in other species. These data suggest that our understanding of the function of the TCR constant domains of these vertebrates is incomplete.

Keywords: T cell receptors; Comparative immunology/evolution; T lymphocytes

1. Introduction

T lymphocytes play pivotal roles as both regulators and executioners of adaptive immunity in mammals and probably in all gnathostomes [1]. The mammalian heterodimeric, membrane anchored T cell receptor (TCR) endows the cell with antigen specific reactivity. The TCRα and TCRβ polypeptide chains each contain a membrane-distal variable domain that is generated from V, (D), and J gene segments by combinatorial processes mediated by the recombination activating gene (RAG) products. These rearrangements produce a repertoire of TCRα and TCRβ variable domains within the organism capable of recognizing diverse MHC/peptide

Abbreviations: CART, motif, conserved antigen receptor transmembrane motif; CDR, complementarity determining region; FR, framework region; RAG, recombination activating genes.

* Corresponding author. Tel.: +1-601-984-1739; fax: +1-601-984-1708.
E-mail address: cmckinney@microbio.umsmed.edu (E.C. McKinney).

0145-305X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.dci.2003.12.004

antigen combinations. Each polypeptide also contains a membrane-proximal constant domain that functions to translate receptor binding via the variable domain into a cytoplasmic response. The constant domains also coordinate the many members of the TCR/CD3 complex to form a multi-chain glycoprotein raft in the plasma membrane that is capable of antigen binding, receptor aggregation, and the initiation of cytoplasmic signal transduction cascades. As the nomenclature indicates, the genes encoding the variable and constant domains of these TCR chains have been found to be just that—variable and constant, respectively. It has been thought that the distinct functions of the domains required diversity in the variable domain but static conservation of the constant domain [2].

Homologues of the genes necessary for the basic antigen receptor hardware used by the mammalian immune system have been found in most major groups of jawed vertebrates. Immunoglobulin, TCR, MHC and RAG genes have all been described from diverse classes, including cartilaginous fishes, bony fishes and amphibians [3]. The constant domains of TCR genes from ectothermic vertebrates are reported to contain little if any variability [4]. However, it has been recently shown that the bicolor damselfish, *Stegastes partitus*, has high allelic polymorphism at each of two TCRβ constant region loci [5]. This observation led to the testable hypothesis that the damselfish TCRβ chain's heterodimeric partner may also exhibit polymorphism.

Here is described the results of testing that hypothesis, i.e. sequence and genomic Southern blot analyses of the *TCRA* genes of this damselfish, including mutation selection analysis of novel constant domain polymorphisms. Damselfish *TCRAC* was found to be not only polymorphic, but polygenic as well, with extraordinary differential selection upon alleles of the two genes. In addition to questioning our current understanding of TCR function, this observation will complement existing TCRα sequence data from other ectotherms in efforts to distinguish the phylogenetically fundamental features of this critical receptor from those that may be accessory [6–14].

2. Materials and methods

2.1. cDNA library screening

Messenger RNA isolated from thymus, spleen and pronephros of 50 damselfish was used in cDNA library construction in the Uni-Zap XR lambda vector (Stratagene, La Jolla, CA). This library was amplified once and screened with a ^{32}P-labeled catfish TCRα probe (gift of M. Wilson, Jackson, MS) hybridizing at 42 °C in 30% formamide. This yielded one clone with significant inferred amino acid homology to TCRα V and J genes from other species, although there was no C gene segment in this clone. This partial clone was subsequently used to re-screen the library twice under the same low stringency conditions, yielding nine full-length TCRα homologues. To find more diverse clones, a heterogeneic damselfish Cα probe was made using sequence information from the initial nine clones to avoid preliminary polymorphic regions as primer sites. Primers C2 (5′-ACCTTTTCAACGATTCTGAGG-3′) and C4 (5′-TCACTGACTGATCCACAGCCGCATC-3′) were designed to amplify ~400 bp of the Cα gene segment from the library for probe DNA. Four more rounds of library screening using this heterogeneic damselfish TCRAC probe produced 32 distinct TCRα clones. Positive phage clones were excised to plasmid form, amplified in SOLR cells, and purified with PerfectPrep spin columns (Brinkman/Eppendorf, Westbury, NY).

2.2. Genomic PCR cloning

TCRα constant gene fragments were amplified using *Pfu*Turbo polymerase (Stratagene, La Jolla, CA). After addition of deoxyadenosine tails, these PCR products were ligated into the PGEMT-easy vector (Promega, Madison, WI). Ligations were transformed into competent DH5α and purified with Perfectprep plasmid columns (Brinkman/Eppendorf). Short segments (21–22 amino acids) from genomic clones originally amplified to compare intron sequences are included as additional polymorphism segregation support for the two TCRAC genes.

2.3. DNA sequencing

Initial sequencing was performed using the dideoxynucleotide chain termination method with

^{35}S-dATP. All clones were eventually analyzed by cycle sequencing resolving the products on a Global IR dual laser acrylamide gel system (LI-COR, Lincoln, NE). All coding regions were completely sequenced in both directions (both strands) using fluorescently labeled common M13F, M13R, T3 or T7 primers.

2.4. Sequence data analysis

Sequence data was managed with the DNAStar software suite (Lasergene, Madison, WI) and the BioNavigator web-based program suite (www.bionavigator.com). Sequences were aligned using Megalign (DNAStar) with the CLUSTALV algorithm and the PAM250 residue weight table then adjusted manually.

Mutational selection (K_A/K_S) analysis was computed for αC domain nucleotide sequences pairwise using the SNAP (Synonymous/Non-synonymous Analysis Program) at www.hiv-web.lanl.gov. SNAP calculates K_A (non-synonymous substitutions per 100 non-synonymous sites) and K_S (synonymous substitutions per 100 synonymous sites) based on an algorithm described by Li [15] and then modified by others [16–18]. Polymorphic codons were defined as those encoding residues that were mutated to the same different residue in at least two of the 32 clones. Therefore, a non-synonymous mutation in only one clone was not considered a polymorphic site in this analysis. The many possible pairwise values for K_A and K_S for the entire domain, the polymorphic positions, the non-polymorphic positions and the polymorphic positions in each gene were averaged for each of the data sets.

2.5. Southern blotting

Genomic DNA (gDNA) was extracted from liver tissue of wild-caught damselfish (gift of M. Schmale, Miami, FL). Ten micrograms DNA samples were digested overnight with various restriction endonucleases, electrophoresed on 1% agarose gels, and transferred to nitrocellulose using the Turboblotter system (Schleicher and Schuell, Keene, NH). Damselfish specific primers were synthesized to produce probes complementing distinct Vα gene families (depending on template clone) or the Cα region: Vα1 sense 5′-ATGCTGTCACTGCATTACTGG-3′, Vα2-3 sense 5′-ATGCTGTCAGTGAACTTGAGT-3′, Vα1-3 anti-sense 5′-AGCACAGTAGTACACAG-CAGA-3′, Cα sense 5′-CCATCCTACTTCAAA-3′, Cα anti-sense GGATTCTTCACATGCA-3′. Hybridizations were performed under high stringency conditions with final washes at 60 °C in 0.1 × SSC/0.1% SDS.

3. Results

Thirty-two damselfish TCRα cDNA clones that contained nucleotide differences in the protein coding regions were identified. Twenty-eight clones were full-length containing V, J and C segments, and only one of these (DFA23) had a frame-shift mutation between the V and J segments. Four clones were incomplete, i.e. two contained constant gene segments with no V or J elements (DFA20, DFA56), one had a truncated 5′V followed by normal J and C segments (DFA126), and one had a truncated 5′V followed by a normal C segment without an intervening J segment (DFA128). These C segment data are summarized in Fig. 1.

3.1. Constant segments

Damselfish *TCRAC* in most aspects resembles those of other vertebrates (Fig. 2). Conserved cysteines distanced approximately 50 amino acids apart are predicted to form the intradomain disulfide bond. β-Strands a through f can readily be assigned although they are shorter and less conserved in the unorthodox immunoglobulin-fold of the TCRαC domain [2]. The hinge region, characteristically short in teleost fishes [4], contains a cysteine that in mammals forms the covalent linkage with a cysteine in the β chain; in the fish β chain this cysteine is absent. The damselfish transmembrane region contains the core of the conserved antigen receptor transmembrane (CART) motif, and maintains charged arginine and lysine residues for association with other TCR/CD3 complex chains [19,20]. As in mammals, the fish cytoplasmic tail is short and lacks appropriate docking motifs to recruit signaling adaptor molecules. Hence, fish CD3 homologues with immunoreceptor tyrosine-based activation motifs appear

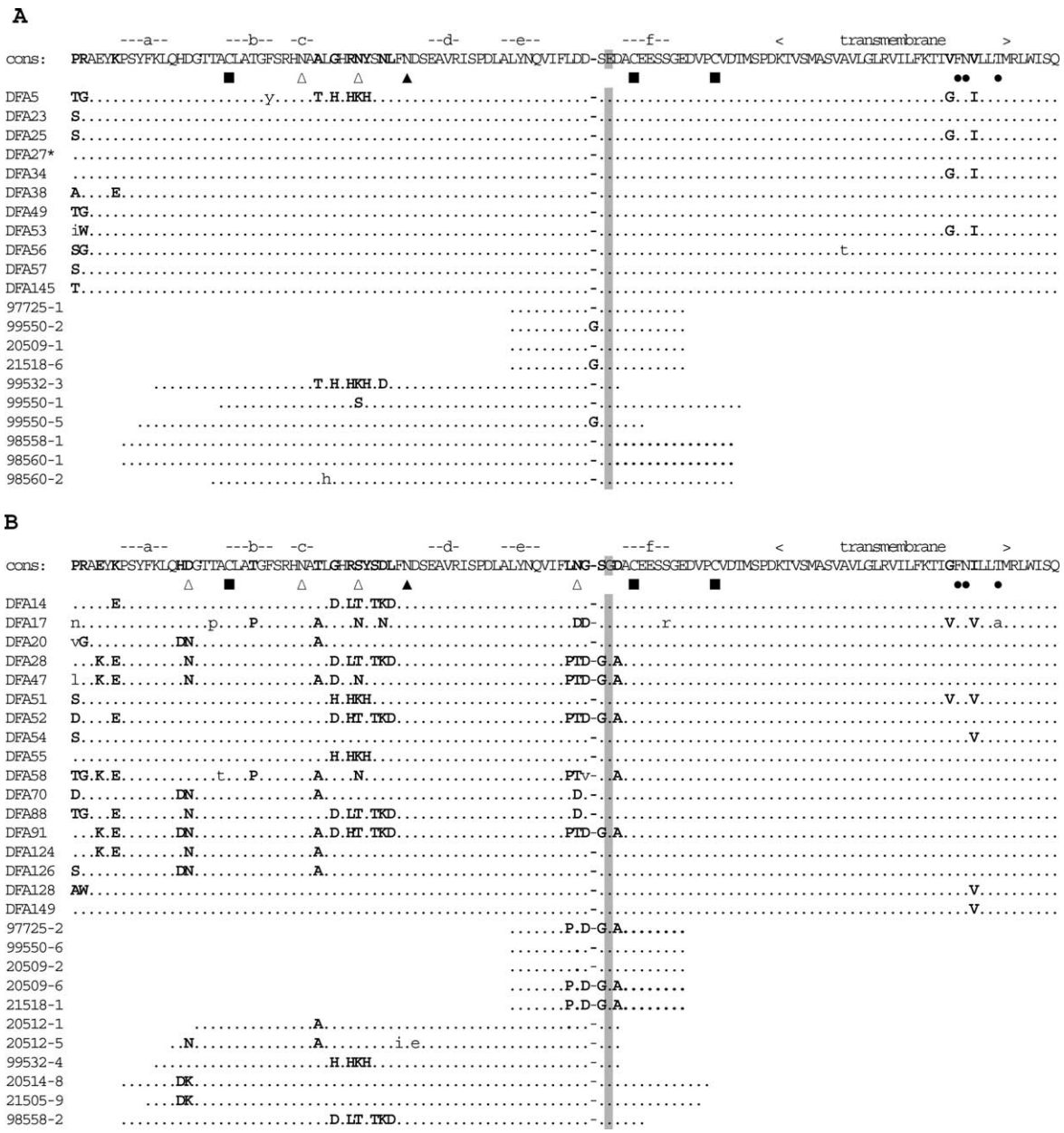

Fig. 1. Deduced amino acid sequences of damselfish *TCRAC* genes. Sequences are divided into *TCRAC1* (A) and *TCRAC2* (B) alleles. Clone numbers are shown to the left of sequences. cDNA clones begin with DFA, and genomic clones contain a five digit fish number with the clone number following the hyphen. The DFA27 amino acid sequence is also representative of clones DFA29, DFA39, DFA50 and DFA137. Dots represent identity with the consensus shown at the top. Triangles denote asparagine residues that could potentially be sites for N-linked glycosylation; open triangles denote those glycosylation sites that are affected by polymorphisms. Squares mark conserved cysteine residues. Circles show conserved CART motif residues [19]. The predicted β-strands of the immunoglobulin domain and the transmembrane region are shown above the consensus sequence. Since the crystal structure of ectothermic TCR has not been analyzed, the b-strands were assigned conservatively based on alignments of the domain with those from other species. Boldface type marks polymorphic residues that are observed in more than one clone. Lower case type marks potential polymorphisms that are not yet supported by a second clone. DNA sequence data have been deposited in Genbank with accession numbers AY198341–AY198372 and AY269843–AY269863.

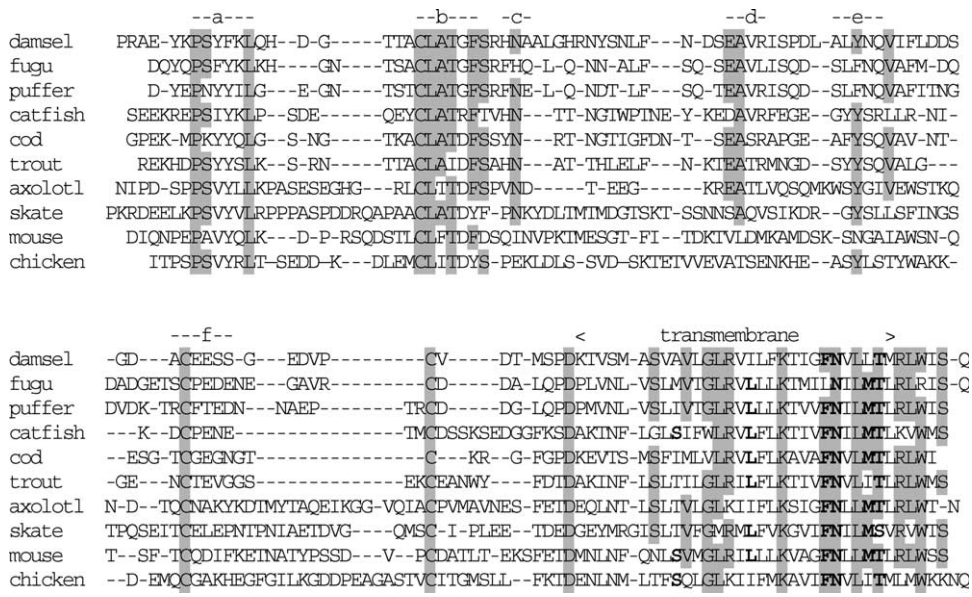

Fig. 2. Alignment of inferred amino acid sequences of vertebrate TCRαC regions. Dashes indicate gaps introduced into sequences for alignment. For the polymorphic damselfish, the consensus sequence from all clones (both genes) is used. Shaded residues highlight conservation between the damselfish consensus sequence and other vertebrates. Conserved residues of the TCRα CART motif are in bold [19]. Predicted positions of β-strands of the Ig-like domain and the transmembrane region are shown above the alignment.

necessary, albeit not yet demonstrated, for tyrosine kinase recruitment and signal transduction.

In fact, the only striking feature of the damselfish TCRα constant domain is that it is clearly not constant (Fig. 1). Of the 32 constant domain cDNAs sequenced, 27 different deduced amino acid sequences were found (84%). Additionally, 23 of the 121 amino acid positions of the TCRα constant domain are polymorphic (19%) as defined by the occurrence of at least two different corroborated amino acids at a single position in different alleles. An amino acid substitution at a designated position is considered polymorphic only when an independent second clone is found showing mutation to the same amino acid at that position. This definition is meant to discount the consideration of mutations introduced during reverse transcription or clone replication. The polymorphisms are concentrated before the a strand, in the c–d loop, and in the carboxyl end of the e–f loop. Fewer polymorphic residues are also found in the a–b loop, carboxyl end of strand b, and the transmembrane region.

Genomic PCR from six individual fish yielded *TCRA* clones containing at least two different sequences from each animal. These sequences contained only one uniform difference: the occurrence of glycine (G) or glutamic acid (E) in the second residue before β-strand f (Fig. 1). When the library-derived clones are sorted according to glycine or glutamic acid at that position, the two groups of sequences differ noticeably in the frequency of polymorphisms. The clones with glycine in this position contain substantially more polymorphisms in the Ig portion of the gene than those employing glutamic acid. Southern hybridization data (Fig. 3A) show two major hybridizing bands in most digests suggesting two distinct *TCRAC* genes. These data led to the assignment of the sequences to gene *TCRAC1* or *TCRAC2* based on the glycine or glutamic acid residue. In addition, two different alleles from one gene, either *TCRAC1* or *TCRAC2*, were found in four of ten animals surveyed. These three independent sets of data suggest the existence of two gene segments for this Cα domain, making allelic polymorphism at two *TCRAC* genes in the damselfish population the most attractive explanation of this sequence diversity. The Southern blot data also exclude the possibility that numerous *TCR* loci are the source of the *TCRAC* variation rather than polymorphism.

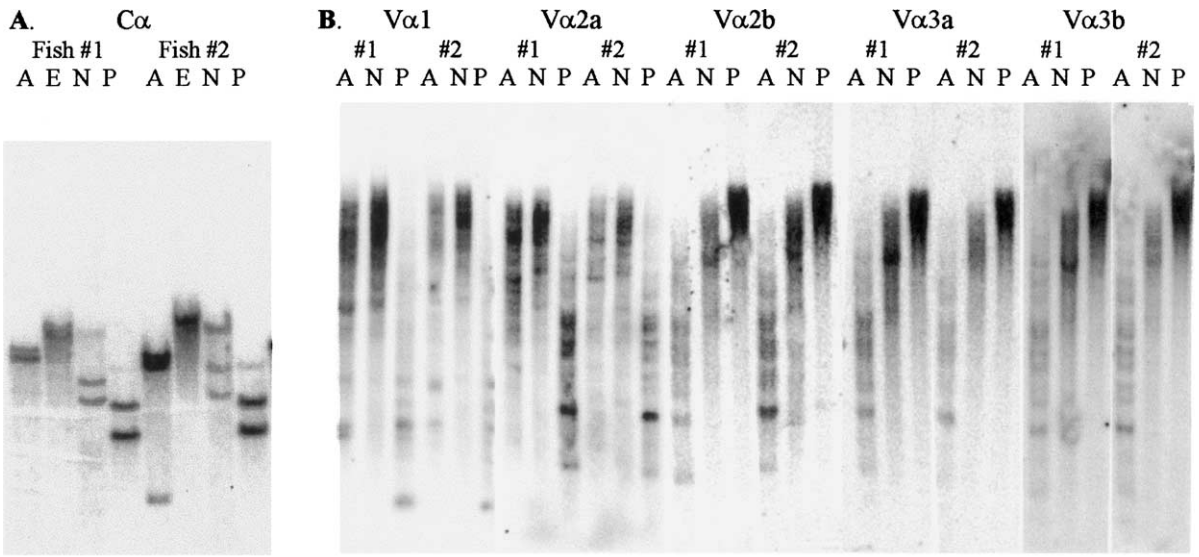

Fig. 3. Southern blot analysis of damselfish TCRA. Damselfish liver DNA from two individual fish (#1 and #2) was digested to completion with a panel of enzymes, electrophoresed, and transferred to nylon. The nylon filter was then hybridized with a ^{32}P TCRα probe. (A) Cα probe, restriction endonucleases are (left to right) A, AseI; E, EcoRI; N, NsiI and P, PstI. (B) Vα family probes, restriction enzymes are A, N, and P.

3.2. Selection of mutations

To determine what types of selective processes are acting on these variant *TCRAC* genes, the rates of substitution at non-synonymous sites (K_A) and at synonymous sites (K_S) were computed for the entire C region, separately for the polymorphic and non-polymorphic positions, and for only the polymorphic residues of *TCRAC1* or *TCRAC2*. In the case of most genes, purifying positive selection keeps K_S much greater than K_A. However, balancing positive selection can act to maintain diversity and is evidenced by a K_A greater than K_S. As shown in Fig. 4, K_A is unusually high in damselfish *TCRAC*. When this analysis is applied to different portions of the molecule, a clearer picture emerges. The polymorphic and non-polymorphic residues (as defined for either damselfish *TCRAC* gene in Fig. 1) were analyzed separately for each gene. The non-polymorphic positions are under strong purifying selection ($K_A/K_S = 0.1167$), maintaining much of the domain's amino acid sequence by suppressing non-synonymous mutations. Conversely, the polymorphic positions show significant positive selection pressure to diversify ($K_A/K_S = 5.6586$). When this polymorphism is further dissected between the two genes, it is clear that *TCRAC1* ($K_A/K_S = 1.0061$) is not experiencing the strong evolutionary pressure to diversify as *TCRAC2* ($K_A/K_S = 9.7757$).

Most of the amino acid polymorphisms in this admittedly small sampling result from single nucleotide substitutions in the most common codon used at a site. Only 30 codons (of >1600 analyzed) require more complex substitution pathways; 19 are accounted for by a single nucleotide substitution from the second most common codon. This pattern of substitution is consistent with the idea that positive selection is acting to maintain naturally occurring mutations that support allelic diversity at polymorphic residues in *TCRAC2*.

3.3. V and J segments

The 28 full-length variable segment sequences were aligned and a matrix of their pairwise nucleotide identities was generated as well as a dendrogram estimating their relationships using parsimony analysis (data not shown). These were used to assign the V sequences into three families based on nucleotide identities of greater than 77% and the topography of

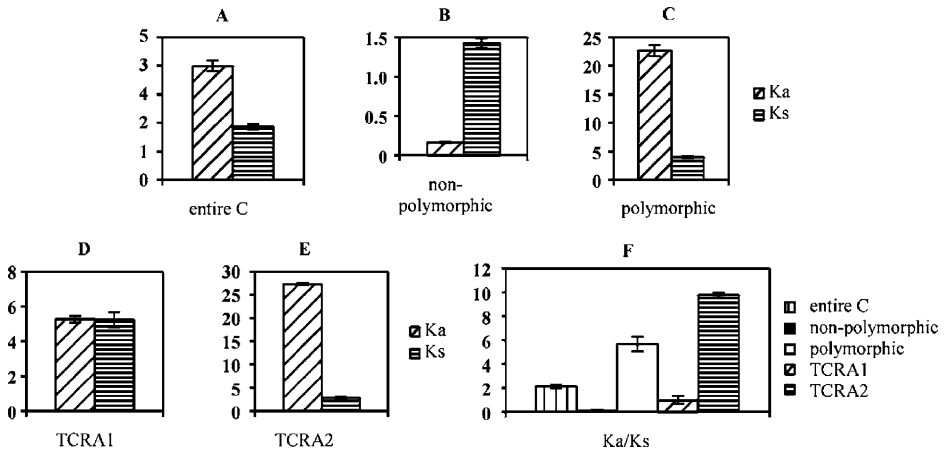

Fig. 4. K_A and K_S values at polymorphic and non-polymorphic residues of damselfish *TCRAC*. Charts A–E show average numbers of non-synonymous substitutions per 100 non-synonymous sites (K_A) and average numbers of synonymous substitutions per 100 synonymous sites (K_S) amongst pair wise comparisons of the 32 damselfish *TCRAC* sequences. These are shown for the entire Cα coding region, the non-polymorphic residues, and the polymorphic residues in A, B and C, respectively. In charts D and E, the cDNA clones from *TCRA1* and *TCRA2* are analyzed separately. Chart F plots the ratio of K_A to K_S for each. Error bars denote standard error of the mean.

the tree [6,7]. These data were also used to further subdivide the Vα2 and Vα3 families into two groups, a and b, that share insertion/deletion patterns and have greater than 88% identity. Southern hybridization experiments with distinct Vα family probes show many hybridizing bands of differing patterns (Fig. 3B). This suggests the presence of large Vα families, similar to those described in other vertebrates.

Intra-familial substitutions in the variable segments are found within both the complementarity determining regions (CDR) and the other parts of the molecule (Fig. 5). The most conserved regions of the V segments are the leader peptide and the framework regions, although family Vα1, as well as clone DFA50 and clone DFA57 use different leaders than the other sequences. The most interesting feature of these V1 family sequences is the absence of a cysteine before CDR1 in Vα1 family members that would be expected to complete the disulfide bond to the cysteine in framework region 3. This covalent bond is usually necessary to hold the two β-sheets of an immunoglobulin domain together in a β-barrel configuration. Curiously, many flounder Vαs submitted to databases also lack this cysteine. The presence of this cysteine in clones DFA50 and DFA57 argues against their inclusion in the Vα1 family. There is considerable diversity at the amino acid level between the V family or subgroup consensus sequences and the four clones not assigned to families (as high as 63% dissimilar). Further, there is less than 54% identity between these four sequences suggesting that each belongs to a different V family. This level of heterogeneity in the variable gene segment is consistent with that seen in mammalian as well as other ectothermic species [6,21].

Of the 29 Jα regions analyzed, only five appear to involve repeated usage of the same J gene segment, suggesting a much larger Jα repertoire than that sampled here (Fig. 6). Only one clone with a V–J frame-shift was identified (DFA23). As previously shown in other fish, damselfish TCRα V–J junctions seem to have a higher proportion of functional rearrangements than β V–D–J junctions [13]. The FGXG pattern is highly conserved in these Jα segments as it is in most TCR and immunoglobulin light chain genes in other species. Only DFA47 deviated from this consensus motif, i.e. it substituted valine for glycine128. Although the 5′ ends of the J segments are difficult to determine without genomic data, it is evident from these sequences that varying J lengths and/or N diversity can alter the CDR3 length by as many as three amino acids (Fig. 6). Maximal diversity is expected in fish CDR3 if, as in mammals, it is implicated in interactions with peptide/MHC.

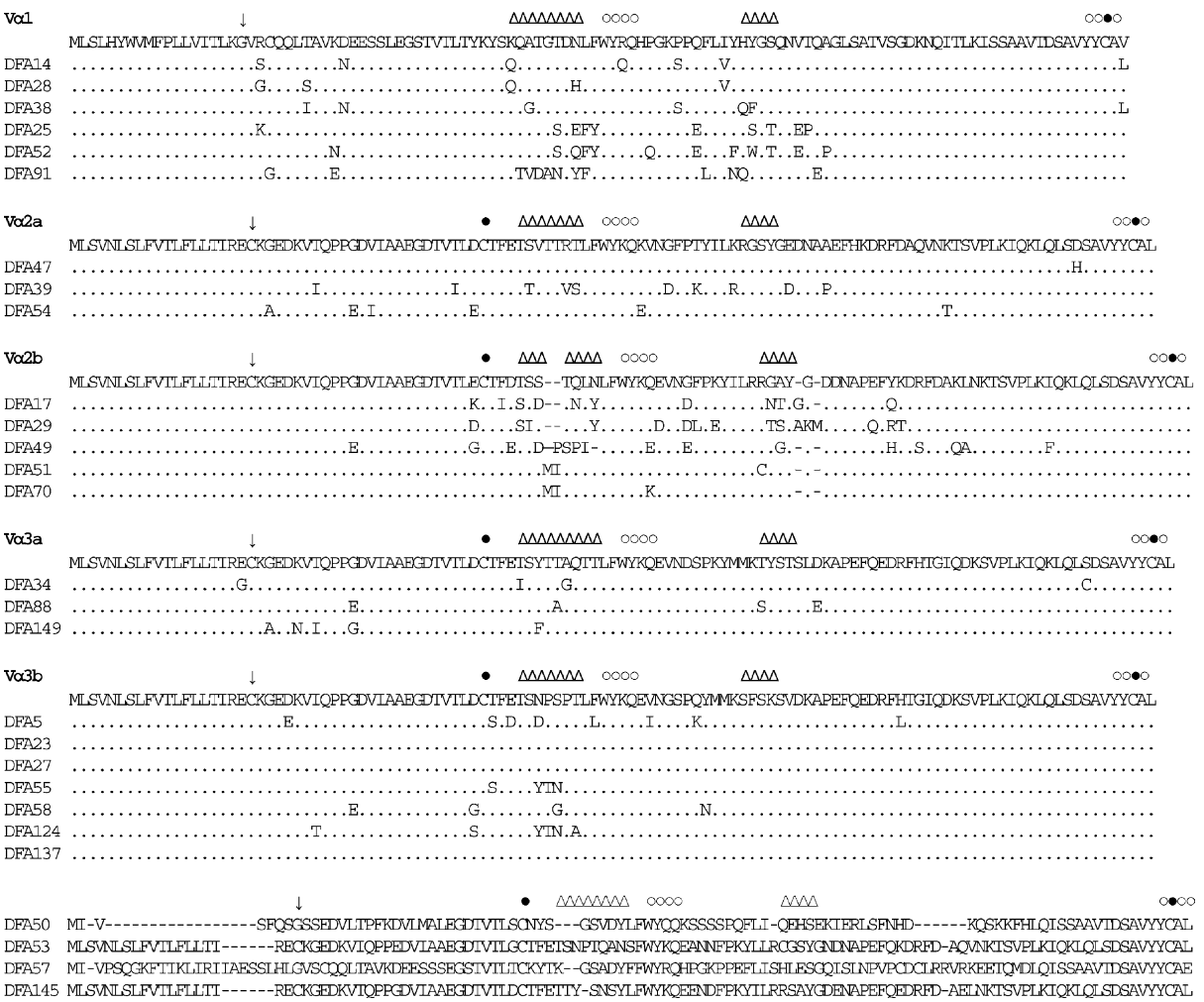

Fig. 5. Amino acid alignments of three V gene families of damselfish. Vα2 and Vα3 families are subdivided into two groups, a and b, that share insertion/deletion patterns and have greater than 88% identity. Clone numbers are shown to the left of each sequence. Dots are used to indicate identity with the consensus sequence shown above each family. Dashes have been introduced to maximize the similarity between family members. The last amino acid of the predicted leader peptide is labeled with an arrow, CDR1 and CDR2 are marked with triangles, and framework region 2 (FR2) and FR3 motifs are denoted with circles. Cysteines thought to form intradomain disulfide bonds are marked with a filled circle. Four clones that did not cluster with the other V sequences on a dendrogram were not assigned to V families and are shown aligned at the bottom of the figure.

4. Discussion

In this report, the first characterization of the TCRα chain-encoding genes from the largest order of vertebrates, Perciformes, is described. In many respects, the damselfish *TCRA* appears similar to that of other vertebrates. Broad diversity in the damselfish V and J gene segments is apparent and the majority of the cDNAs described here are predicted to be recombined, functional transcripts. However, members of the fish Vα1 family are lacking Cys[23] needed to form the disulfide bridge that stabilizes the β-barrel configuration of the V domain. This is rare among vertebrate antigen receptor variable domains, but not unique, since similar sequences have been reported for four other teleost species [8,22; ABO53367.1; AYO62259.1]. Although the pufferfish sequence was isolated from genomic

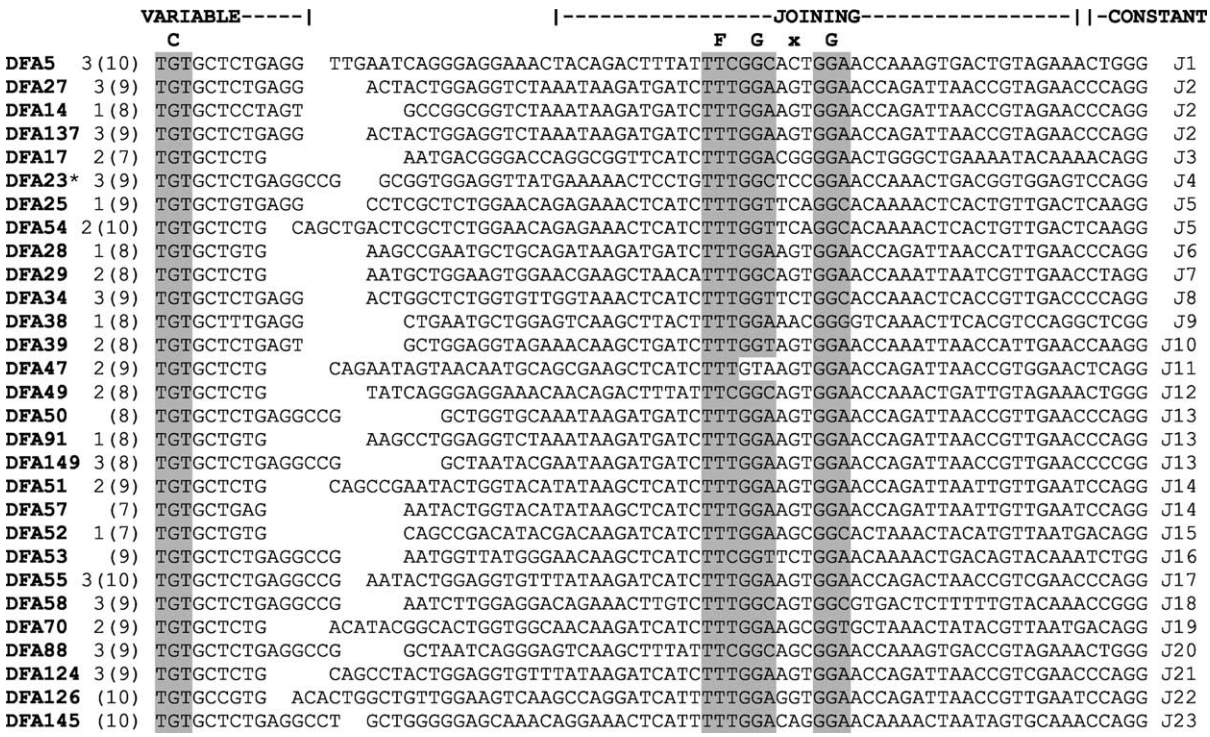

Fig. 6. Alignment of the nucleotide sequences of the joining regions of damselfish *TCRA*. Clone names are listed to the left in boldface. Vα family (if assigned) is listed after the clone name with CDR3 amino acid length in parentheses. Asterisk indicates out-of-frame transcript. The conserved J motif, FGXG, and the conserved 3′ cysteine of the V segment are denoted above the alignment in boldface and shaded. Assigned J families are shown to the right.

DNA, the flounder sequence was identified in a cDNA library, and is assumed to be functional. Both of those sequences have an alternate cysteine at position 3 similar to five of the six damselfish Vα1 family members. The Cys^3 has been postulated to substitute for Cys^{23} in disulfide bond formation [22]. Nonetheless, one Vα1 family member, DFA 91, is lacking even the Cys^3. Perhaps other intramolecular forces allow this molecule to maintain its quaternary structure despite the lack of a covalent bond. Alternatively, this cDNA may represent a non-functional thymic transcript not yet eliminated by selection.

The polymorphic nature of the damselfish TCRα constant domain is remarkable, and contradicts the generally accepted invariant nature of the domain. The positioning of 17 of the 23 polymorphic residues is predicted to be in the loops connecting the β strands. Polymorphism at these positions may be more tolerable in maintaining the tertiary structure of the domain, allowing variation in loops exposed to solvent or other members of the multi-chain complex. Four polymorphic positions, however, fall in the predicted β-strands and two others fall in the transmembrane region. Although the transmembrane region is highly conserved among vertebrates, the conservative substitution of valine for isoleucine is consistent with a transmembrane domain. Previous results from this laboratory described allelic polymorphism at each of two TCRβ constant region genes in damselfish [5]. It seems arguable that some of the exposed polymorphic residues in the TCRα and TCRβ constant domains articulate either with each other, or with members of the CD3 complex.

The specific physiological consequences of damselfish TCR constant region polymorphisms await functional studies of the proteins encoded by these genes. However, statistical methods can indicate what types of natural selection (if any) are likely acting upon particular protein coding nucleotides. The most

striking differences are between sites at which a mutation will change the amino acid (non-synonymous sites) and those at which a mutation will sustain the amino acid (synonymous sites). Analysis of K_A and K_S ratios among pair wise comparisons of divergent sequences can expose those mutations that natural selection is preserving for subsequent generations. A K_A/K_S ratio of <1 is expected if purifying natural selection limits non-synonymous substitutions and conserves the protein sequence. A K_A/K_S ratio of >1 is expected if positive selection is acting to promote incorporation of change at a position. Importantly, a K_A/K_S ratio of 1 is predicted if any substitution at that site is selectively neutral. This analysis was used to show that allelic polymorphism in the peptide binding regions of MHC molecules was maintained by positive selection, while other regions were conserved by purifying selection [23]. The K_A/K_S ratios of both the damselfish *TCRAC* and *TCRBC* are much higher than would be predicted for domains that are conserved to perform quaternary interactions in initiating an adaptive immune response. The elevated K_A/K_S ratios imply that diversification is maintained and favored in certain exposed portions of the TCRα and TCRβ C domains. This selective pressure is most striking in the polymorphic codons of *TCRAC2*, where maintenance of high allelic polymorphism at the population level appears most advantageous.

Two hypotheses have been considered that may account for maintenance of polymorphic TCR constant domains in fish, i.e. TCR variable region diversification or differential signaling through the CD3 peptides. Structural studies of the mammalian TCRαβ heterodimer have shown that the C domains are rigid, and probably incapable of conformational change [24]. The configuration is fixed to such an extent that the orientation between the two domains is the same in different TCRs. Further, the extended f–g loop in mammalian βC constrains lateral movements of the Vβ domain. Thus, quaternary differences between mammalian αβ TCRs are due to variations in the orientation of the Vβ and Vα domains. In fish, however, none of the reported TCRβ peptides contains an extended f–g loop, and in fact, the g strand is quite short, and difficult to assign [4]. This implies that the damselfish receptor may be permitted greater flexibility than its mammalian counterparts. Although the Cα and Cβ cysteines responsible for intradomain stability are conserved throughout phylogeny, the interchain Cα–Cβ disulfide bond is not. All reported teleost sequences lack the required cysteine in the Cβ domain [4], again suggesting flexibility not apparent in mammalian molecules. Further, it has been shown that loss of the intradomain disulfide bond in the constant region of mammalian molecules causes conformational changes in Vβ structure [25]. It is, therefore, plausible that variant, polymorphic residues of the damselfish Cα and Cβ domains may affect the conformation of the V region, and thus provide an additional measure of receptor diversity.

Alternatively, the polymorphic residues in the constant region of the damselfish TCR may affect interactions with CD3 peptides, assuming fish have CD3 homologues (see below). In mammals, the TCR complex is capable of delivering a variety of signals through the CD3 chains resulting in distinct cytokine profiles and effector functions [26]. It is possible that polymorphisms in the damselfish Cα and Cβ domains influence the complexity of antigen-driven TCR/CD3 aggregates, and thus the quality of the signal transmitted. Structural analyses of the mammalian αβTCR suggest that the ectodomain of CD3ε fits into a pocket formed by the constant domains of both chains [24]. If the basic orientation described for mammalian β strands is preserved in damselfish molecules, the most polymorphic regions of the damselfish TCRα and TCRβ constant domains would be concentrated in the pocket formed for CD3ε suggesting a relationship between form and function presumably involving CD3.

To date, only one type of CD3 peptide has been described in ectothermic vertebrates, a CD3ε homolog [27–30]. In the sturgeon, this CD3ε homolog contains polymorphisms in the amino terminus, and the b–c and c–d loops that could juxtapose the TCR constant domains. While it is difficult to envision a system of polymorphic CD3 chains complexing with polymorphic TCR constant domains in a functionally advantageous manner, this hypothesis cannot yet be dismissed. Although intracellular signaling cascades in teleost lymphocytes appear to be analogous to those of mammals [31], it is possible that various combinations of polymorphic domains may alter or even add

specificity to distinct signaling through the CD3 peptide.

Southern hybridization, polymorphism patterns and genomic cloning from individual fish collectively suggest that *TCRAC* is bigenic in the damselfish. This is in contrast to the monogenic situation previously observed in most other vertebrates, including zebrafish [10], skate [12], catfish [7], cod [14], and pufferfish [6], but is concordant with the two different Cα genes recently described for *Xenopus* [9]. The greater number of polymorphisms in *TCRAC2* compared to *TCRAC1* may suggest different functional capabilities for these allelic products. It is possible that TCRα chains encoded by these different constant genes pair with distinct subsets of polymorphic Cβ and CD3ε molecules, with divergent signaling capabilities in the resulting TCR/CD3 complexes. The divergent evolution of the two Cα genes, retaining at least one functional allele at *TCRAC1*, may allow some control over the development of a very complicated molecular juxtaposition. Despite its probable fundamental roles in T cell repertoire generation and the initiation of an adaptive immune response, TCR signal transduction in the damselfish may be determined in part by population level allelic polymorphism. The selection data (Fig. 4) dismiss the idea of neutral drift accounting for all of these polymorphisms and compel investigation of possible functions for this novel diversity. Perhaps this domain has been over-looked as a source of structural diversity for antigen binding or even unheralded complexity in interactions mediating signal transduction.

Comparative studies of the genes encoding these polypeptides in diverse vertebrates are necessary to assess the phylogenetic extent of this phenomenon. It is suspected that the fortuitous small size of the damselfish, which necessitated using many individuals for an immune tissue cDNA library, may have aided the discovery of this polymorphism. Therefore, one can predict that such polymorphism may not be unique to this species and may have been overlooked in studies of other, larger model organisms.

Acknowledgements

We are most thankful to Michael Schmale and Jeffrey VanWye for damselfish, Jack Fell and Gloria Scorzetti for use of and training on their LI-COR sequencing apparatus, and L. William Clem and Melanie Wilson for critical review of this manuscript. This work was supported in part by grants from the National Science Foundation (MCB-0211785) to ECM, and from the National Institutes of Health (NS36998) to M. Schmale and (P30 ES05705) to P. Walsh.

References

[1] Flajnik MF, Rumfelt LL. The immune system of cartilaginous fish. Curr Top Microbiol Immunol 2000;248:249–70.
[2] Garcia KC, Degano M, Stanfield RL, Brunmark A, Jackson MR, Peterson PA, Teyton L, Wilson IA. An alphabeta T cell receptor structure at 2.5 A and its orientation in the TCR–MHC complex [see comments]. Science 1996;274:209–19.
[3] Du Pasquier L, Flajnik M. Origin and evolution of the vertebrate immune system. In: Paul W, editor. Fundamental immunology, 4th ed. Philadelphia: Lippincott/Raven Publishers; 1999. p. 605–50.
[4] Charlemagne J, Fellah JS, De Guerra A, Kerfourn F, Partula S. T-cell receptors in ectothermic vertebrates. Immunol Rev 1998;166:87–102.
[5] Kamper SM, McKinney CE. Polymorphism and evolution in the constant region of the T-cell receptor beta chain in an advanced teleost fish. Immunogenetics 2002;53:1047–54.
[6] Wang K, Gan L, Kunisada T, Lee I, Yamagishi H, Hood L. Characterization of the Japanese pufferfish (*Takifugu rubripes*) T-cell receptor α locus reveals a unique genomic organization. Immunogenetics 2001;52:174–85.
[7] Wilson MR, Zhou H, Bengten E, Clem LW, Stuge TB, Warr EW, Miller NW. T cell receptors in the channel catfish: structure and expression of TCR α and β genes. Mol Immunol 1998;35:545–7.
[8] Rast JP, Haire RN, Litman RT, Pross S, Litman GW. Identification and characterization of T-cell antigen receptor-related genes in phylogenetically diverse vertebrate species. Immunogenetics 1995;42:204–12.
[9] Haire RN, Kitzan Haindfield MK, Turpen JB, Litman GW. Structure and diversity of T-lymphocyte antigen receptors alpha and gamma in Xenopus. Immunogenetics 2002;54:431–8.
[10] Haire RN, Rast JP, Litman RT, Litman GW. Characterization of three isotypes of immunoglobulin light chains and T-cell antigen receptor alpha in zebrafish. Immunogenetics 2000;51:915–23.
[11] Fellah JS, Kerfourn F, Dumay AM, Aubet G, Charlemagne J. Structure and diversity of the T-cell receptor alpha chain in the Mexican axolotl. Immunogenetics 1997;45:235–41.
[12] Rast JP, Anderson MK, Strong SJ, Luer C, Litman RT, Litman GW. Alpha, beta, gamma, and delta T cell antigen receptor genes arose early in vertebrate phylogeny. Immunity 1997;6:1–11.
[13] Partula S, de Guerra A, Fellah JS, Charlemagne J. Structure and diversity of the TCR alpha-chain in a teleost fish. J Immunol 1996;157:207–12.

[14] Wermenstam NE, Pilstrom L. T-cell antigen receptors in Atlantic cod (*Gadus morhua* l.): structure, organisation and expression of TCR alpha and beta genes. Dev Comp Immunol 2001;25:117–35.
[15] Li W-H, Wu C-I, Luo C-I. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon change. Mol Biol Evol 1985;2:150–74.
[16] Li W-H. Unbiased estimates of the rates of synonymous and nonsynonymous substitution. J Mol Evol 1993;36:96–9.
[17] Ota T, Nei M. Variance and covariances of the numbers of synonymous and nonsynonymous substitutions per site. Mol Biol Evol 1994;11:613–9.
[18] Pamilo P, Bianchi NO. Evolution of the *Zfx* and *Zfy* genes: rates and interdependence between the genes. Mol Biol Evol 1993;10:271–81.
[19] Campbell KS, Backstrom BT, Tiefenthaler G, Palmer E. CART: a conserved antigen receptor transmembrane motif. Semin Immunol 1994;6:393–410.
[20] Exley M, Wileman T, Mueller B, Terhorst C. Evidence for multivalent structure of T-cell antigen receptor complex. Mol Immunol 1995;32:829–39.
[21] Chothia C, Boswell DR, Lesk AM. The outline structure of the T-cell alpha beta receptor. Embo J 1988;7:3745–55.
[22] Nam BH, Hirono I, Aoki T. The four TCR genes of teleost fish: the cDNA and genomic DNA analysis of Japanese flounder (*Paralichthys olivaceus*) TCR alpha-, beta-, gamma-, and delta-chains. J Immunol 2003;170:3081–90.
[23] Hughes AL, Yeager M. Natural selection and the evolutionary history of major histocompatibility complex loci. Front Biosci 1998;3:d509–16.
[24] Wang J, Lim K, Smolyar A, Teng M, Liu J, Tse AG, Hussey RE, Chishti Y, Thomson CT, Sweet RM, Nathenson SG, Chang HC, Sacchettini JC, Reinherz EL. Atomic structure of an alphabeta T cell receptor (TCR) heterodimer in complex with an anti-TCR fab fragment derived from a mitogenic antibody. Embo J 1998;17:10–26.
[25] Caspar-Bauguil S, Arnaud J, Huchenq A, Hein WR, Geisler C, Rubin B. A highly conserved phenylalanine in the alpha, beta-T cell receptor (TCR) constant region determines the integrity of TCR/CD3 complexes. Scand J Immunol 1994;40:323–36.
[26] Backstrom BT, Hausmann BT, Palmer E. Signaling efficiency of the T cell receptor controlled by a single amino acid in the beta chain constant region. J Exp Med 1997;186:1933–8.
[27] Ropars A, Bautz AM, Dournon C. Sequencing and expression of the CD3 gamma/delta mRNA in *Pleurodeles waltl* (urodele amphibian). Immunogenetics 2002;54:130–8.
[28] Park CI, Hirono I, Enomoto J, Nam BH, Aoki T. Cloning of Japanese flounder *Paralichthys olivaceus* CD3 cDNA and gene, and analysis of its expression. Immunogenetics 2001;53:130–5.
[29] Alabyev BY, Guselnikov SV, Najakshin AM, Mechetina LV, Taranin AV. CD3epsilon homologues in the chondrostean fish *Acipenser ruthenus*. Immunogenetics 2000;51:1012–20.
[30] Dzialo RC, Cooper MD. An amphibian CD3 homologue of the mammalian CD3 gamma and delta genes. Eur J Immunol 1997;27:1640–7.
[31] Park H, Zhou H, Bengten E, Wilson M, Chinchar VG, Clem LW, Miller NW. Activation of channel catfish (*Ictalurus punctatus*) T cells involves NFAT-like transcription factors. Dev Comp Immunol 2002;26:775–84.