Immunogenetics (2004) 55:818–824
DOI 10.1007/s00251-004-0652-7

ORIGINAL PAPER

Michael F. Criscitiello · Niklas E. Wermenstam · Lars Pilström · E. Churchill McKinney

Allelic polymorphism of T-cell receptor constant domains is widespread in fishes

Received: 25 November 2003 / Revised: 18 January 2004 / Published online: 19 February 2004

Abstract T-cell receptor chains contain membrane-proximal constant domains of the immunoglobulin superfamily that are relatively invariant in mammalian species. In contrast, recent studies in the bicolor damselfish have demonstrated surprising allelic polymorphism in the *TCR* alpha (*A*) and *TCR* beta (*B*) "constant" (C) domain genes. This report extends these initial observations beyond Perciformes to two other orders of teleost fishes. Studies in both the Atlantic cod and zebrafish show high levels of polymorphism in the *TCRA* constant genes. Levels of 13% and 15% amino acid nonidentity were found within cod and zebrafish, respectively. Evolutionary analysis of codon usage suggests that positive selection maintains the high number of *TCRAC* alleles in these fish populations. Additionally, investigation of a *TCRB* constant gene from the Beau Gregory, a sister species of the bicolor damselfish, shows no evidence of transpecies maintenance of constant region alleles. These data argue that the T-cell receptor constant domain is being employed by many vertebrates in a manner inconsistent with our current understanding, and may indicate unheralded complexity in signal transduction through the TCR/CD3 complex.

Keywords T-cell receptors · Comparative immunology/evolution · Allelic polymorphism · T lymphocytes

Electronic Supplementary Material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00251-004-0652-7

Introduction

The T-cell receptor (TCR) chains shared by jawed vertebrates are each comprised of two extracellular immunoglobulin superfamily domains, an antigen/MHC reactive variable domain and a membrane-anchored constant domain. A wealth of research has focused on V(D)J somatic cell recombination and the resulting repertoire generating physiology of the variable domains. Meanwhile, a relatively smaller body of work has found that the constant domains function in the maintenance of TCR/CD3 complex quaternary structure and signal transduction (Rubin 1994).

In mammals, the TCR constant domains are relatively invariant, albeit small amino acid differences are known to exist between products of different *TCRBC* or *TCR* gamma (*G*) *C* genes (Hedrick 1984; Lefranc 1989). Comparative studies in non-mammalian organisms have occasionally found more than one sequence for constant domain genes, but they were usually attributed to multiple constant genes (Wilson 1998; Wermenstam 2001; Haire 2002). In contrast, an examination of the TCR genes of the bicolor damselfish, *Stegastes partitus,* indicated significant allelic polymorphism in the constant domains of both the *TCRA* and *TCRB* genes (Kamper and McKinney 2002; Criscitiello 2004). This species maintains a high number of both *TCRA* and *TCRB* alleles in the Caribbean population, and the amino acid differences amongst sampled *TCRAC* and *TCRBC* sequences reaches 20%

M.F. Criscitiello and N.E. Wermenstam contributed equally to this work

Nucleotide sequence data reported here are available in the GenBank database under the accession numbers AJ439464–AJ439499 and AY476721–AY476734.

M. F. Criscitiello
Department of Microbiology and Immunology,
University of Miami School of Medicine,
PO Box 016960, R-138 Miami, FL 33101, USA

N. E. Wermenstam · L. Pilström
Immunology Programme,
Department of Cell and Molecular Biology,
Uppsala University BMC,
Box 596, 751 24 Uppsala, Sweden

E. C. McKinney (✉)
Department of Microbiology,
University of Mississippi Medical Center,
2500 North State Street, Jackson, MS 39216, USA
e-mail: cmckinney@microbio.umsmed.edu
Tel.: +1-601-9841739
Fax: +1-601-9841708

and 23%, respectively. The extraordinary constant domain polymorphism in the damselfish TCR begged the question of whether this was a genetic phenomenon unique to this species or if it was shared by a larger phylogenetic group. We hypothesized the latter, assuming that others had no reason to expect (or look for) more than one "constant" domain allele from other model organisms. The postulate was tested using immunologically characterized species selected to represent two other orders of bony fish, Cypriniformes and Gadiformes. Further, transpecies maintenance of alleles was analyzed for evidence of allelic stability. The data presented here show that divergent teleost species have significant polymorphism in their TCR constant domain genes, which appears unrelated to speciation events.

Materials and methods

Template DNA sources

Zebrafish (*Danio rerio*) cDNA libraries prepared from spleen and kidney were the generous gifts of G. Litman (St. Petersburg, USA). Atlantic cod (*Gadus morhua* L) were raised from eggs and cultivated at the Marine Research Institute of Iceland. Individuals at approximately 20 weeks of age were anesthetized in MS-222 (Sandoz, Basel, Switzerland), and the spleens were rapidly removed, snap frozen in liquid nitrogen and stored at −70 °C until used. Total RNA was prepared from approximately 30 mg of spleen tissue (ten separate individuals) by the method of Chirgwin and co-workers (1979) and poly(A)$^+$ RNA was isolated by chromatography on an oligo-dT column (Pharmacia Biotech, Uppsala, Sweden). Copy DNA was prepared using a cDNA synthesis kit (Pharmacia Biotech). Genomic DNA was isolated from cod erythrocytes as described by Sambrook and co-workers (1989). Beau Gregory (*Stegastes leucostictus*) were netted by Dr. Henry Feddern, marine collector (Taverier, USA), near Conch Reef in the Florida Keys. Liver tissue was homogenized from two *S. leucostictus* samples and genomic DNA was prepared as above.

Polymerase chain reaction cloning

TCR constant gene alleles were amplified using *Pfu*Turbo polymerase (Stratagene, La Jolla, Calif.) or *Taq* DNA polymerase (Promega, Madison, Wis.) with low cycle number. Zebrafish *TCRAC* sequences were amplified with primers C5 (5′-AACTGAAGTGAAGCCGAATAT-3′, forward) and C3 (5′-GCTCATCCACGCTTTGAAAAGTCA-3′, reverse) from kidney and spleen cDNA libraries. The resulting smear, migrating around 300 bp, was ligated into pGEMT (Promega) after addition of deoxyadenosine with *Taq* DNA polymerase (Promega) for TA cloning. Nested primers C5.2 (5′-TTACCAAGTGGGAAACTCATGCCTG-3′, forward) and C3.2 (5′-GTCACCAGCACATTGAAAACGACTG-3′, reverse) were used to PCR screen transformed *Escherichia coli* DH5α cells before Perfectprep plasmid preparation (Eppendorf, Hamburg, Germany). A sense primer (5′-GACTCTGCTCTGTACTACTGT-3′) and an antisense primer (5′-GTGGATCCACAGCCGTAGG-3′) hybridizing to the Atlantic cod *TCRA* V and C regions, respectively, were used to amplify corresponding DNA in spleen cDNA from ten different individuals (Wermenstam and Pilstrom 2001). DNA of desired size was extracted from 1% agarose gels supplemented with ethidium bromide, using a StrataPrep PCR Purification Kit (Stratagene, La Jolla, USA). Purified fragments were ligated into pPCR-Script Amp SK(+) vector using the PCR-Script Amp Cloning Kit (Stratagene). Beau Gregory *TCRBC* were amplified with primers SK48 (5′-CCTTCAGCAAAGGAGTGCAG-3′, forward) and SK61 (5′-GAGGAAACGGTTTTAACGA-3′, reverse) from liver genomic DNA. Products were cloned into pGEMT and prepared as for the zebrafish clones.

DNA sequencing

Cod sequences were obtained using Thermo Sequenase Radiolabeled Terminator Cycle Sequencing Kit (USB, Cleveland, USA), employing the method of Sanger and co-workers (1977). Zebrafish and Beau Gregory clones were analyzed by cycle sequencing resolving the products on a Global IR dual laser acrylamide-gel system (LICOR, Lincoln, N.E.). All coding regions were sequenced in both directions (both strands) using fluorescently labeled common SP6 and T7 primers.

Sequence data analysis

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

Mutational selection (K_a/K_s) analysis was computed for *TCRAC* nucleotide sequences pairwise using the synonymous/non-synonymous analysis program (SNAP) at http://www.hiv-web.lanl.gov. SNAP calculates K_a (nonsynonymous substitutions per 100 nonsynonymous sites) and K_s (synonymous substitutions per 100 synonymous sites) based on an algorithm described by Li and co-workers (1985) and then modified by others (Li 1993; Pamilo and Bianchi 1993; Ota and Nei 1994). A nucleotide substitution at a designated position was considered polymorphic only when an independent second clone was found showing mutation to the same nucleotide at that position. This definition is meant to discount the consideration of mutations introduced during reverse transcription or clone replication. Therefore, a nonsynonymous 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 and the nonpolymorphic positions were averaged for each of the data sets. The tree in Fig. 4B was drawn using MEGA 2.1 by unweighted pair-group with arithmetic means (UPGMA) (Kumar 2001). The confidence of the nodes was assessed with 500 bootstrap replications.

Results

Zebrafish and cod *TCRAC* sequences

High fidelity polymerase was used to amplify nine novel *TCRAC* sequences from zebrafish (*Danio rerio*) cDNA libraries. With the addition of four published sequences (Haire 2000) and one EST from the databases, 14 different TCRα constant domain amino acid sequences were analyzed. Eighteen of 109 constant domain amino acid residues show polymorphisms that are corroborated in another independent clone, resulting in 17% polymorphism as shown in Fig. 1.

Low cycle number PCR was used to amplify *TCRAC* sequences from spleen cDNA of ten different individual Atlantic cod, *Gadus morhua* (Fig. 2). Seventeen different cod TCRα constant domain amino acid sequences were revealed by this analysis, not inclusive of polymorphisms appearing in only one clone. Eighteen of 112 residues show corroborated polymorphisms, giving a percent polymorphism of 16%. Several of the fish analyzed yielded three or four distinct sequences, consistent with

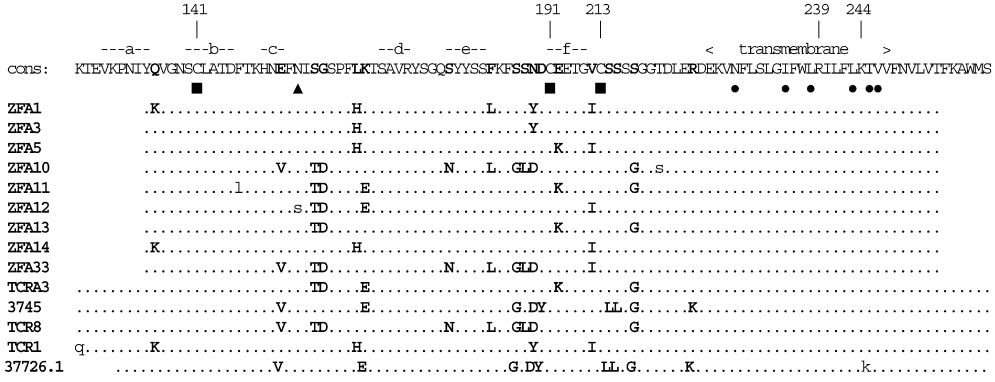
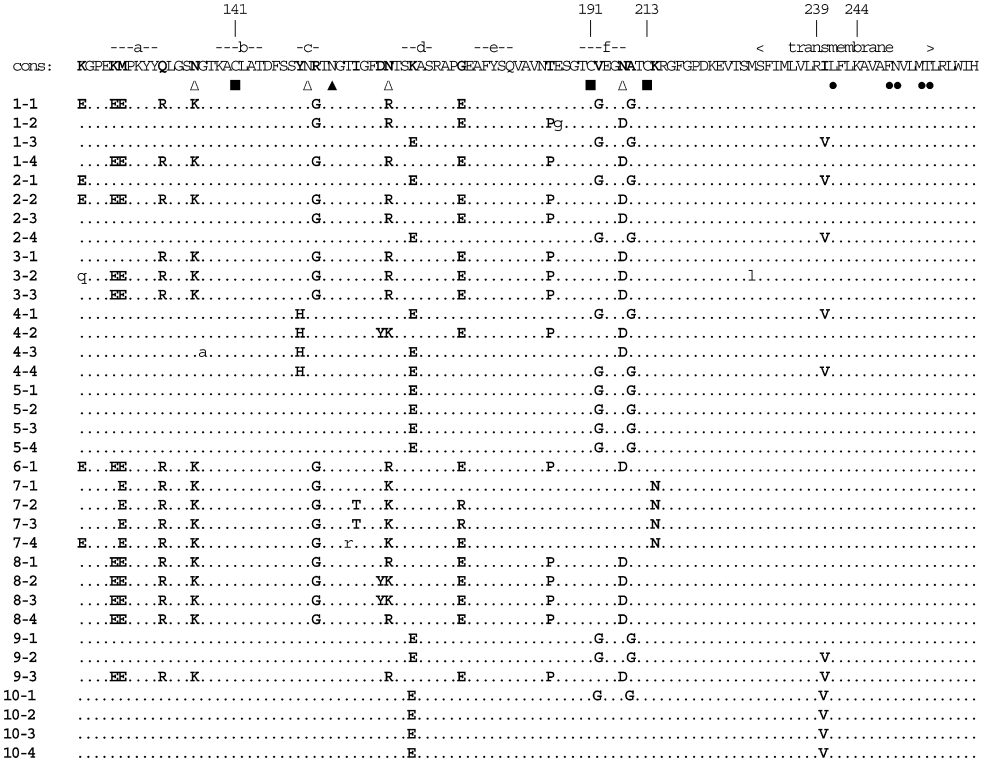


Fig. 1 Deduced amino acid sequences of zebrafish *TCRA* constant genes. Clone numbers are shown to the *left* of sequences. *Dots* represent identity with the consensus shown at the *top*. *Triangles* denote asparagine residues that could potentially be sites for N-linked glycosylation. *Squares* mark conserved cysteine residues. *Circles* show conserved CART motif residues (Campbell 1994). *Above* the consensus sequence some characteristic residues are numbered according to Kabat and co-workers (Schiffer 1992), and the predicted β-strands of the immunoglobulin domain and the transmembrane region are shown. *Boldface type* marks polymorphic residues that are observed in more than one clone. *Lowercase type* marks potential polymorphisms that are not yet supported by a second clone. The last five sequences are previously published or submitted sequences for zebrafish *TCRAC* (Haire et al. 2000)

Fig. 2 Deduced amino acid sequences of cod *TCRA* constant genes. The same designations are used as in Fig. 1. In addition, *open triangles* denote those glycosylation sites that are affected by polymorphisms. Clone numbers contain the individual fish number before the hyphen followed by the clone number from that individual



allelic polymorphism at two *TCRAC* gene loci, as has been described in the bicolor damselfish (Criscitiello et al. 2004).

Mutation analysis

The K_a/K_s ratios of zebrafish and cod *TCRAC* suggest that selection is acting on the polymorphic residues to favor amino acid replacements and maintain high allelic polymorphism (Fig. 3). The nonpolymorphic positions of zebrafish and cod *TCRAC* show typical K_a/K_s ratios of less than one (zebrafish, 0.21; cod, 0.09), indicative of purifying selection to conserve amino acid sequence. The K_a/K_s ratios of the polymorphic positions, however, elevate to 3.72 for the zebrafish and 2.24 for the cod. Such a favoring for amino acid replacement substitutions provides strong evidence that natural selection is actively maintaining allelic polymorphism at *TCRAC* by balancing selection. It also demonstrates that the selection main-

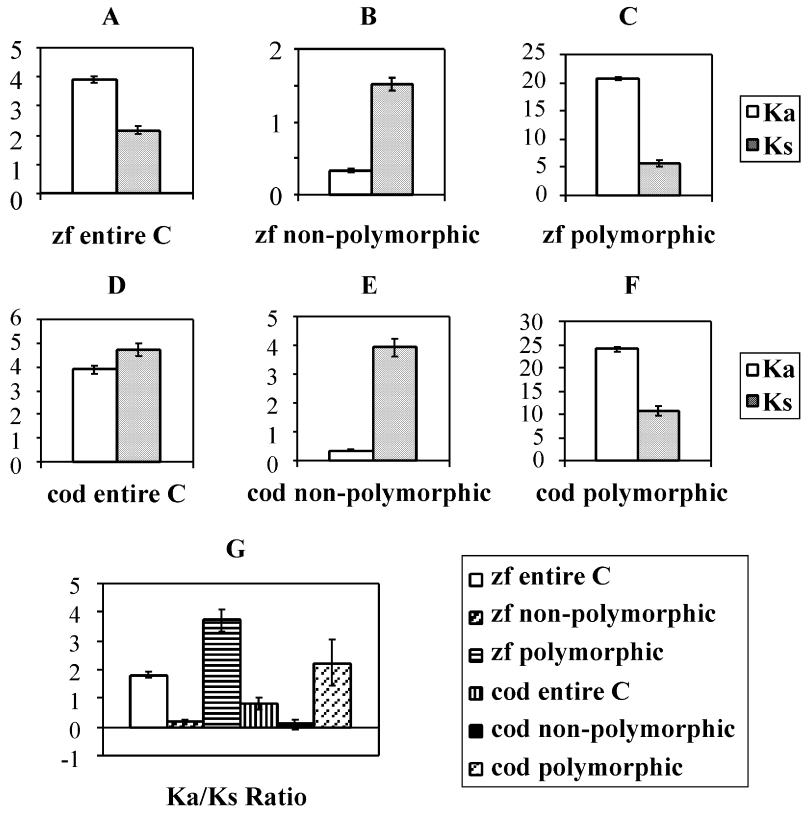


Fig. 3A–F K_a and K_s values at polymorphic and nonpolymorphic regions of *TCRAC*. Charts **A–F** show average numbers of nonsynonymous substitutions per 100 nonsynonymous sites (K_a) and average number of synonymous substitutions per 100 synonymous sites (K_s) amongst pairwise comparisons of the 14 zebrafish (*zf*) and 35 cod *TCRAC* sequences. These are shown for the entire *TCRAC* [**A, D**], the nonpolymorphic residues [**B, E**], and the polymorphic residues [**C, F**]. Chart **G** plots the ratio of K_a to K_s for each. *Error bars* denote standard error of the mean

taining this diversity is focused at the polymorphic residues.

Transpecies maintenance of alleles

The finding of high K_a/K_s ratios in the polymorphic residues of fish *TCRAC* prompted an investigation for another characteristic of MHC alleles, transpecies polymorphism. MHC alleles have been maintained for long periods of time, often pre-dating speciation events. This is demonstrated by the clustering of similar alleles without regard to species in trees derived from MHC alignments (Lawlor 1988).

A recent phylogeny of the genus *Stegastes,* based upon mitochondrial genes, identified two sister species of *S. partitus* that could be used in a transpeciation study: *S. leucostictus*, the Beau Gregory, and *S. otophorus* (personal communication from H.A. Lessios of the Smithsonian Tropical Research Institute). Although selection for high levels of allelic polymorphism have now been described for both alpha and beta TCR constant domain genes (Fig. 1; Criscitiello et al. 2004), TCR beta was chosen for this analysis due to a greater number of available sequences for comparison. *TCRBC1* alleles from the readily available Beau Gregory were cloned, sequenced and compared with alleles from the bicolor damselfish (Fig. 4A; Kamper and McKinney 2002). A dendrogram of this alignment shows that alleles from each fish segregate by species (Fig. 4B), indicating that these alleles appeared after the speciation event that gave rise to *S. leucostictus* and *S. otophorus*. Hence, these data imply that TCR constant alleles are not conserved for long stretches of evolutionary time, as are MHC alleles.

Discussion

Zebrafish and cod *TCRAC* sequences

Analysis of the amino acid substitutions found at polymorphic sites shows that the positioning of polymorphic residues in the TCRαC of cod and zebrafish are somewhat similar. Most of the diversity is found in the exposed loops connecting the β-strands of the domain (83% in zebrafish and 56% in the cod), and is focused around the c-d loop, the e-f loop and f strand. The majority of replacements are not conservative. In the zebrafish, 6/8 (75%) substitutions in these external positions change the charge or hydrophobicity of the R group compared with 7/10 (70%) in the remainder of the domain. In the cod TCRαC, amino acid substitutions tended to be more radical throughout but absolutely so in these exposed portions, 6/6 (100%) versus 10/12 (83%) in the rest of the domain. As might be expected, the radical amino acid changes are reflected in widely divergent isoelectric points (5.110–9.705) for the allelic forms of the TCRαC domain of some, but not all, fish. (Tables of substitutions and Pi values are in the Electronic Supplementary Material) One individual cod transcribed genes encoding

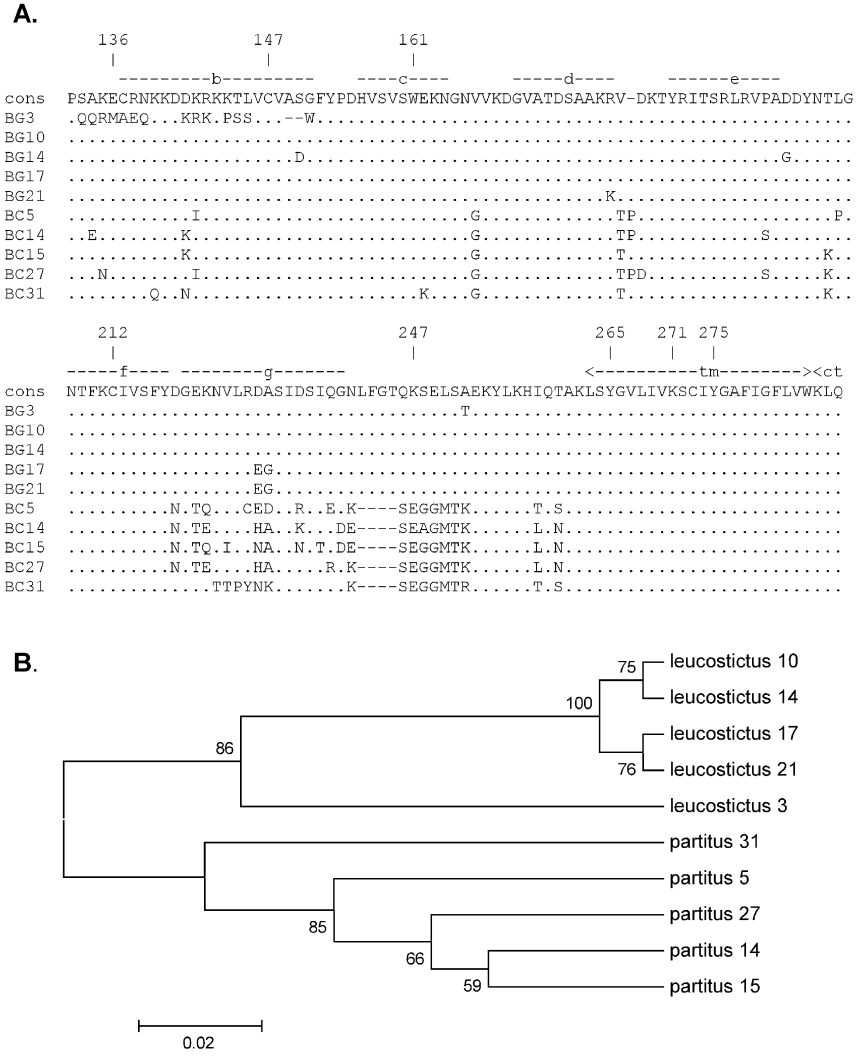


Fig. 4A, B Comparison of *TCRBC1* alleles of damselfish sister species. **A** Amino acid alignment of novel *Stegastes leucostictus* (BG) and published *S. partitus* (BC) *TCRBC1* alleles (Kamper and McKinney 2002). Some characteristic residues are numbered according to convention (Schiffer et al. 1992). Predicted β-strands of the constant domain are shown *above* the consensus sequence. *Dots* represent identity with the consensus, *dashes* mark gaps introduced for alignment. **B** Dendrogram drawn from the alignment in **A**

TCRαC that diverged by nearly two points (cod no. 2, 6.940–8.910), while three other animals had polymorphic sequences but no Pi divergence. Charged groups on the surface of the TCR may interact with the surrounding medium or with other transmembrane proteins of the TCR/CD3 complex. By contrast, nonpolar groups on the surface may lead to instability and thus usually have a distinct purpose when found on the outside of a molecule, such as binding subunits of complex proteins. In mammalian TCRαC these external loops and the exposed f strand, where the frequencies of polymorphisms and of radical amino acid substitutions are concentrated in these fish, are predicted to articulate with CD3ε (Wang 1998). Interestingly, these polymorphic areas of TCRα are adjacent to equally divergent and polymorphic segments of TCRβC in the damselfish (Kamper and McKinney 2002).

Phylogenetic extent of TCRα allelic polymorphism

The initial observation of high allelic polymorphism in the TCR constant domains of the bicolor damselfish raised the question of the evolutionary extent of this phenomenon. The analyses of *TCRAC* of cod and zebrafish described here and the previous study in damselfish (Criscitiello et al. 2004) suggest that polymorphism in these genes is common in bony fishes. Discounting the less parsimonious explanation of three independent evolutionary origins, the presence of polymorphism in representatives of three Superorders (zebrafish in Ostariophysi, cod in Paracanthopterygii, damselfish in Acanthopterygii) advocates its presence in their most recent ancestor and its descendants. The clade descendant of this ancestor includes over 22,000 described species (mostly Ostariophysi and Perciform Acanthopterygii; cladograms in Electronic Supplemetary Material). Thus, a majority of (the ~44,000) vertebrate species (Nelson 1994) are predicted to have misnomered TCR "constant" domains. The relatively stable *TCRAC* allele from inbred strains of mice [with only three

corroborated polymorphic nucleotides, all synonymous (Hayday et al. 1985; Winoto et al. 1985; Kuchroo et al. 1994)], has fostered the idea that TCR constant regions, like those of immunoglobulin, are relatively invariant. However, RFLP studies of the TCR among wild mice indicate that significant polymorphisms are detectable, particularly in the *TCRBC* (Nobuhara et al. 1989). It is likely that sequencing of wild *Mus TCR* constant genes would reveal polymorphisms beyond those detected with RFLP. Nonetheless, few RFLP patterns have been documented for the human *TCR* constant domains, although some polymorphic sites have been identified in the intervening sequences (Robinson and Kindt 1985, 1987; Bragado 1994). Thus, vertebrate TCR constant domains appear to contain significant polymorphisms even among certain mammals, although the functional significance of these variations is currently unknown.

Mutation analysis

The degeneracy of the genetic code allows distinction between mutations that do not change the amino acid (synonymous) and those that cause amino acid replacements (nonsynonymous). In most genes, the number of synonymous substitutions per 100 synonymous sites (K_s) exceeds the number of nonsynonymous substitutions per 100 nonsynonymous sites (K_a). This occurs because most nonsynonymous mutations are deleterious to protein structure and are therefore quickly removed from the gene pool by purifying selection. Thus, most genes have a K_a/K_s ratio less than one. When this analysis was applied to the peptide-binding regions of human MHC class I genes, however, K_a/K_s ratios for the peptide-binding regions were found to be greater than one (Hughes and Nei 1988), as balancing natural selection favors amino acid replacements to maintain high allelic diversity (Hughes 1999). Heterozygosity is advantageous at MHC loci in order to bind a variety of peptides and thus resist a variety of pathogens. The functional significance of the location of polymorphisms within the *TCRAC* is less obvious. They do not appear to interfere with polar or hydrophobic contacts between the TCRα and TCRβ, since they are primarily on the exterior of the molecule. That many polymorphic residues map to portions of the molecule that juxtapose CD3ε in mammalian crystal structures supports the idea that polymorphisms modulation signal transduction. However, the notion that multiple constant region polymorphisms may alter the conformation of the V domain, and generate additional V region diversity cannot be excluded.

Transpecies polymorphism

Although only five Beau Gregory *TCRBC* were sequenced, they were compared with more than 40 damselfish *TCRBC* (unpublished data), and only those most similar were presented here. None of the sequences was shared between the two species. The lack of transpecies polymorphism at damselfish *TCRBC1* is inconsistent with diversification for antigen binding. However, it may be consistent with diversifying the range of signal transduction capability through the TCR/CD3 complex, which may be shielded from pathogen-driven selection. This analysis implies that alteration in TCR/CD3 interaction is a more likely consequence of TCR constant region polymorphism than variation in TCR variable region structure.

Conclusion

Whether or not signaling, receptor diversity or another unidentified function is affected by C-domain polymorphisms, TCR complexity must be beneficial to fish populations or it would not be actively maintained by selection. Since each individual may have a maximum of four TCRα alleles, it is likely that the benefit of these changes is at the population level. Current data show no evidence for transpecies polymorphism suggesting that polymorphisms arise and are maintained separately within each species. Whatever physiological advantage is provided by *TCRC* variability, it appears to be prevalent among fish species, and may exist in other vertebrates. Whether the presumed benefit has been deleted or supplanted by additional complexities of the human immune system, significant TCRC variation appears to have been lost before the emergence of man.

Acknowledgements This work was supported in part by grants from the National Science Foundation to E.C.M. (MCB-0211785), from the National Institutes of Health to M. Schmale (NS36998) and to P. Walsh (P30 ES 05705), and from the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (Formas) (22.3/2001-0793) to L.P. The experiments presented here comply with the current laws of the United States and Sweden.

References

Bragado R, Garcia A, Trevino MA, Vilches C, Lopez de Castro JA (1994) Allelic polymorphism in the coding region of human TCR C alpha gene and characterization of structural variability in the alpha chain constant domain. Int Immunol 6:223–230

Campbell KS, Backstrom BT, Tiefenthaler G, Palmer E (1994) CART: a conserved antigen receptor transmembrane motif. Semin Immunol 6:393–410

Carrol RH (1988) Vertebrate paleontology and evolution. Freeman, New York

Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18: 5294–5299

Criscitiello MF, Kamper SM, McKinney EC (2004) Allelic polymorphism of TCRα chain constant domain genes in the bicolor damselfish. Dev Comp Immunol (in press)

Haire RN, Rast JP, Litman RT, Litman GW (2000) Characterization of three isotypes of immunoglobulin light chains and T-cell antigen receptor alpha in zebrafish. Immunogenetics 51:915–923

Haire RN, Kitzan Haindfield MK, Turpen JB, Litman GW (2002) Structure and diversity of T-lymphocyte antigen receptors alpha and gamma in *Xenopus*. Immunogenetics 54:431–438

Hayday AC, Diamond DJ, Tanigawa G, Heilig JS, Folsom V, Saito H, Tonegawa S (1985) Unusual organization and diversity of T-cell receptor alpha-chain genes. Nature 316:828–832

Hedrick SM, Nielsen EA, Kavaler J, Cohen DI, Davis MM (1984) Sequence relationships between putative T-cell receptor polypeptides and immunoglobulins. Nature 308:153–158

Hughes A (1999) Adaptive evolution of genes and genomes. Oxford University Press, New York

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–170

Kamper SM, McKinney EC (2002) Polymorphism and evolution in the constant region of the T-cell receptor beta chain in an advanced teleost fish. Immunogenetics 53:1047

Kuchroo VK, Collins M, al-Sabbagh A, Sobel RA, Whitters MJ, Zamvil SS, Dorf ME, Hafler DA, Seidman JG, Weiner HL, et al (1994) T cell receptor (TCR) usage determines disease susceptibility in experimental autoimmune encephalomyelitis: studies with TCR V beta 8.2 transgenic mice. J Exp Med 179:1659–1664

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17: 1244–1245

Lawlor DA, Ward FE, Ennis PD, Jackson AP, Parham P (1988) HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. Nature 335:268–271

Lefranc MP, Rabbitts TH (1989) The human T-cell receptor gamma (TRG) genes. Trends Biochem Sci 14:214–218

Li WH (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol 36:96–99

Li WH, Wu CI, Luo CC (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174

Nelson JS (1994) Fishes of the world. Wiley, New York

Nobuhara H, Kuida K, Furutani M, Shiroishi T, Moriwaki K, Yanagi Y, Tada T (1989) Polymorphism of T-cell receptor genes among laboratory and wild mice: diverse origins of laboratory mice. Immunogenetics 30:405–413

Ota T, Nei M (1994) Variance and covariances of the numbers of synonymous and nonsynonymous substitutions per site. Mol Biol Evol 11:613–619

Pamilo P, Bianchi NO (1993) Evolution of *Zfx* and *Zfy* genes: Rates and interdependence between genes. Mol Biol Evol 10:271

Robinson MA, Kindt TJ (1985) Molecular determination of T-cell receptor alpha and beta chain genotypes in human families. Hum Immunol 14:195–205

Robinson MA, Kindt TJ (1987) Genetic recombination within the human T-cell receptor alpha-chain gene complex. Proc Natl Acad Sci USA 84:9089–9093

Rubin B, Arnaud J, Caspar-Bauguil S, Conte F, Huchenq A (1994) Biological function of the extracellular domain of the T-cell receptor constant region. Scand J Immunol 39:517–525

Sambrook J, Fritsch E, Maniatis T (1989) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schiffer M, Kabat EA, Wu TT (1992) Subgroups of Tcr alpha chains and correlation with T-cell function. Immunogenetics 35:224–234

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 (1998) 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 17:10–26

Wermenstam NE, Pilstrom L (2001) T-cell antigen receptors in Atlantic cod (*Gadus morhua* l.): structure, organisation and expression of TCR alpha and beta genes. Dev Comp Immunol 25:117–135

Wilson MR, Zhou H, Bengten E, Clem LW, Stuge TB, Warr GW, Miller NW (1998) T-cell receptors in channel catfish: structure and expression of TCR alpha and beta genes [published erratum appears in Mol Immunol 1998 35(18):1219]. Mol Immunol 35:545–557

Winoto A, Mjolsness S, Hood L (1985) Genomic organization of the genes encoding mouse T-cell receptor alpha-chain. Nature 316:832–836

Supplementary Material

Phylogenetic position and relationships amongst fish studied. A) Cladistic representation of the phylogeny of damselfishes of the genus *Stegastes*, with the Sergeant Major, *Abudefduf saxatilis,* as outgroup. This tree is based on mitochondrial sequences and was a personal communication from H.A. Lessios of the Smithsonian Tropical Research Institute. B) Order to Superorder level cladistic view of the phylogenetic relationships between species in this study. Number of currently described species follows each taxonomic group, followed by a representative common name in parentheses. This tree was adapted from that of John G. Lundberg on the Teleostei page of the Tree of Life web project (Carrol RH 1988, Nelson JS 1994).

Carrol RH (1988) Vertebrate paleontology and evolution. New York, W. H. Freeman & Co.
Nelson JS (1994) Fishes of the World. New York, NY, John Wiley & Sons, Inc.

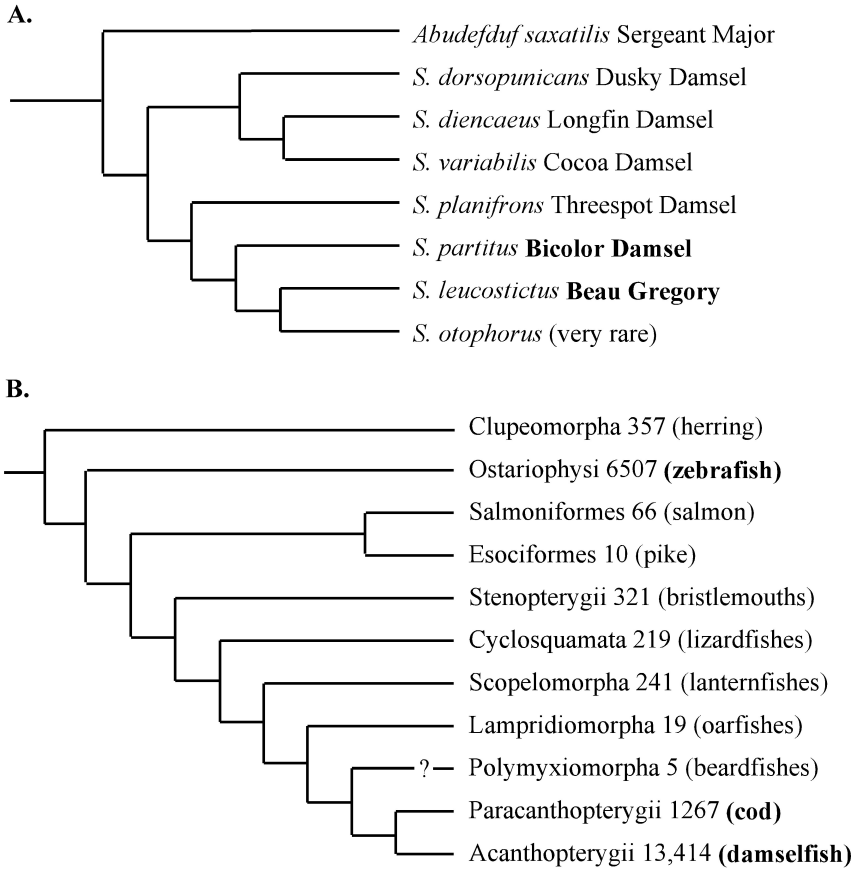


Table 1. Substituted amino acids at polymorphic positions: Zebrafish

Position[a)]	A.A.	Class shift[b)]
10	Q/K	P/+
25	E/V	-/H
29*	S/T	P/P
30*	G/D	G/-
33*	L/H	H/P
34*	K/E	+/-
45	S/N	P/P
50*	F/L	H/H
53*	S/G	P/G
54*	S/L	P/H
55*	N/Y/D	P/P/-
56	D/Y	-/P
58	E/K	-/+
62	V/I	H/H
67	S/G	P/G
68	S/L	P/H
69	S/L	P/H
74	R/K	+/+

[a)] Amino acids in Fig. 1 numbered sequentially, asterisks denote residues predicted to be in the c-d loop, e-f loop and f strand.

[b)] positive charge, +; negative charge, -; hydrophobic, H; polar, P; glycine, G.

Table 2. Substituted amino acids at polymorphic positions: Atlantic cod

Position[a)]	A.A.	Class shift[b]
1	K/E	+/-
5	K/E	+/-
6	M/E	H/-
11	Q/R	P/+
15	N/K	P/+
28	Y/H	P/P
30	R/G	+/G
35*	I/T	H/P
38*	D/Y	-/P
39*	N/R/K	P/+/+
42	K/E	+/-
48	G/E	G/-
59*	T/P	P/H
65*	V/G	H/G
68*	N/D	P/-
69	A/G	H/G
72	K/N	+/P
93	I/V	H/H

[a)] Amino acids in Fig. 2 numbered sequentially, asterisks denote residues predicted to be in the c-d loop, e-f loop and f strand.

[b)] positive charge, +; negative charge, -; hydrophobic, H; polar, P; glycine, G.

Table 3. Isoelectric points of deduced amino acid sequences. The EMBL web server was used to calculate isoelectric points based on C, D, E, H, K, R and Y residues and the two termini. Clone designations are as shown in Figures 1 and 2.

Zebrafish	Isoelectric Point	Cod	Isoelectric Point	Cod	Isoelectric Point
ZFA1	8.465	1-1	7.985	5-4	8.910
ZFA3	7.985	1-2	9.255	6-1	6.940
ZFA5	8.895	1-3	8.910	7-1	9.505
ZFA10	6.925	1-4	8.480	7-2	9.240
ZFA11	6.930	2-1	7.985	7-3	9.705
ZFA12	5.110	2-2	6.940	7-4	8.900
ZFA13	8.475	2-3	8.900	8-1	8.480
ZFA14	8.475	2-4	8.910	8-2	8.865
ZFA33	6.925	3-1	9.505	8-3	8.865
TCRA3	7.985	3-2	7.985	8-4	8.480
3745	7.985	3-3	8.480	9-1	8.910
TCR8	6.925	4-1	8.645	9-2	8.910
TCR1	7.985	4-2	9.505	9-3	8.900
37726.1	7.975	4-3	8.495	10-1	8.910
		4-4	8.945	10-2	8.910
		5-1	8.910	10-3	8.910
		5-2	8.910	10-4	8.910
		5-3	8.910		