

The Evolution of Vertebrate Antigen Receptors: A Phylogenetic Approach

M. H. Richards* and J. L. Nelson†

*Department of Biological Sciences, Brock University, St. Catharines, Ontario, Canada; and †Program in Human Immunogenetics, Fred Hutchinson Cancer Research Center, Seattle, Washington

Classical T cells, those with $\alpha\beta$ T-cell receptors (TCRs), are an important component of the dominant paradigm for self-nonself immune recognition in vertebrates. $\alpha\beta$ T cells recognize foreign peptide antigens when they are bound to MHC molecules on the surfaces of antigen-presenting cells. $\gamma\delta$ T cells bear a similar receptor, and it is often assumed that these T cells also require specialized antigen-presenting molecules for immune recognition, which we term “indirect antigen recognition.” B-cell receptors, or immunoglobulins, bind directly to antigens without the help of a specialized antigen-presenting molecule. Phylogenetically, it has been assumed that T-cell receptors and the genes that encode them are a monophyletic group, and that “indirect” antigen recognition evolved before the split into two types of TCR. Recently, however, it has been proposed that $\gamma\delta$ -TCRs bind directly to antigens, as do immunoglobulins (Ig’s). This calls into question the null hypothesis that indirect antigen recognition is a common characteristic of TCRs and, by extension, the hypothesis that all TCR gene sequences form a monophyletic group. To determine whether alternative explanations for antigen recognition and other historical relationships among TCR genes might be possible, we performed phylogenetic analyses on amino acid sequences of the constant and variable regions which encode the basic subunits of TCR and Ig molecules. We used both maximum-parsimony and genetic distance-based methods and could find no strong support for the hypothesis of TCR monophyly. Analyses of the constant region suggest that TCR γ or δ sequences are the most ancient, implying that the ancestral immune cell was like a modern $\gamma\delta$ T cell. From this $\gamma\delta$ -like ancestor arose $\alpha\beta$ T cells and B cells, implying that indirect antigen recognition is indeed a derived property of $\alpha\beta$ -TCRs. Analyses of the variable regions are complicated by strong selection on antigen-binding sequences, but imply that direct antigen binding is the ancestral condition.

Introduction

In the classical paradigm for T-cell-mediated immune recognition in mammals, an $\alpha\beta$ T-cell receptor (TCR) on the surface of a T cell recognizes and binds to a peptide antigen that is bound to a major histocompatibility complex (MHC) molecule on the surface of a specialized antigen-presenting cell (APC). When a T cell encounters a peptide-MHC complex in which either component is foreign, it can respond appropriately, for instance, by lysing the target or by providing help for B-cell antibody responses. However, not all T cells carry the same type of TCR molecule, and only $\alpha\beta$ T cells react to the MHC+peptide complex. A second class of T cells, called $\gamma\delta$ T cells, is less well understood, perhaps because in humans, these T cells comprise only 1%–5% of peripheral blood lymphocytes (Haas, Pereira, and Tonegawa 1993). However, in other human tissues, especially gut and reproductive epithelia, or in the peripheral blood of other mammals such as ungulates, they may comprise almost half of the T cells (Haas, Pereira, and Tonegawa 1993). The role of $\gamma\delta$ T cells in the vertebrate immune system is only just beginning to be understood (Brenner, Strominger, and Krangel 1988; Raullet 1989; Allison and Havran 1991; Kaufmann 1996), but comparative studies in immunology point to an important role in primary defense against infection (Flajnik 1998).

Despite this growing body of evidence, it is still sometimes assumed that $\gamma\delta$ -TCRs function similarly to

$\alpha\beta$ -TCRs (Bartl, Baltimore, and Weissman 1994; Flajnik 1998), recognizing antigens only when they are presented by some sort of specialized antigen-presenting molecule analogous to an MHC molecule (we refer to this as “indirect antigen binding”). However, recent studies (Schild et al. 1994; Rock et al. 1994; Chien and Jores 1995) indicate that the process of antigen recognition by $\gamma\delta$ T cells may be fundamentally different from that in $\alpha\beta$ T cells. In fact, $\gamma\delta$ -TCRs appear to recognize antigens in a manner similar to the antigen recognition processes of immunoglobulins, the receptors of B cells. Immunoglobulins bind directly to antigens and do not require specialized antigen processing and presentation as do $\alpha\beta$ T cells.

TCRs and immunoglobulins (Ig’s) are the most closely related members of a large protein family called the immunoglobulin superfamily (IGSF), many of whose members are involved in immune recognition or cellular adhesion (Hunkapiller and Hood 1989). TCR and Ig molecules are clearly more similar to each other than to other IGSF molecules (Marchalonis, Schluter, and Edmundson 1997), and the genes that encode them display highly similar patterns of organization and rearrangement during transcription and translation (Hunkapiller and Hood 1989) that are not observed in any other genes. In phylogenetic terms, $\alpha\beta$ - and $\gamma\delta$ -TCR molecules and genes are often thought of as a monophyletic group whose sister group is the Ig’s, and this assumption is one reason that $\gamma\delta$ - and $\alpha\beta$ -TCRs have been assumed to function similarly. As illustrated in figure 1A, since $\alpha\beta$ -TCRs exhibit indirect antigen binding, and since they are closely related to the $\gamma\delta$ -TCRs, it has been assumed that this mode of antigen binding evolved before the split between the two. However, it would be equally parsimonious to assume that indirect antigen

Key words: T-cell receptor, immunoglobulin, immune system evolution.

Address for correspondence and reprints: M. H. Richards, Department of Biological Sciences, Brock University, St. Catharines, Ontario L2S 3A1, Canada. E-mail: mrichard@spartan.ac.brocku.ca.

Mol. Biol. Evol. 17(1):146–155, 2000

© 2000 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

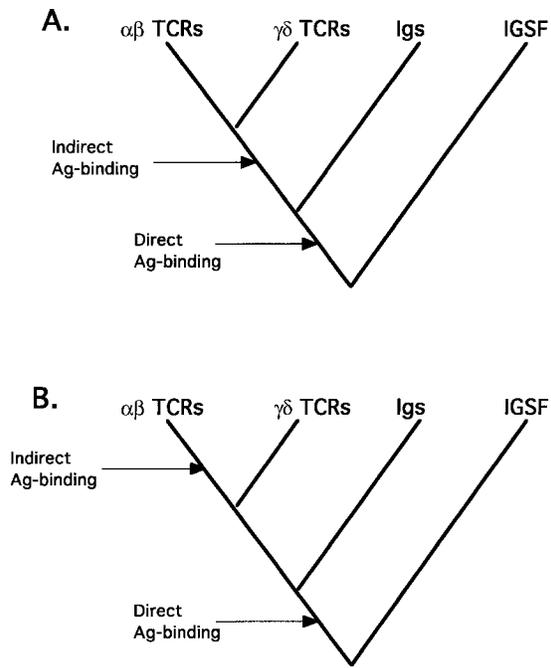


FIG. 1.—Null hypotheses for phylogenetic relationships among members of the immunoglobulin superfamily (IGSF). It is generally assumed that the T-cell receptors (TCRs) and immunoglobulins (Ig's) each comprise a monophyletic group and that these two groups are sister groups within the IGSF. Other IGSF molecules bind directly to their ligands, so direct binding must be an ancestral characteristic of the entire group. "Antigen binding" per se arose when the molecules evolved to be part of an immune system. A, Based on this hypothesis of relationship, immunologists often assume that the antigen-binding processes of $\alpha\beta$ - and $\gamma\delta$ -TCRS are similar: since $\alpha\beta$ -TCRs recognize antigen indirectly, so too must $\gamma\delta$ -TCRs. B, A second possibility, even with this pattern of relationship, is that indirect antigen binding is a derived characteristic of $\alpha\beta$ -TCRs.

binding is a derived characteristic of $\alpha\beta$ T cells (fig. 1B). Moreover, it is possible that the phylogenetic relationships in figure 1 are incorrect, specifically, that the TCR genes are not monophyletic. This would lead to different conclusions about the evolution of immune cells and antigen recognition and binding by $\alpha\beta$ -TCRs, $\gamma\delta$ -TCRs, and also Ig's.

We decided to examine the immunological null hypothesis that the antigen recognition functions of $\alpha\beta$ and $\gamma\delta$ T cells must be similar by constructing phylogenies of TCR and Ig genes in order to reconstruct the phylogenetic relationships among these molecules. To do so requires phylogenetic studies based on outgroup analysis, because we need to investigate the branching order of the TCR- α , - β , - γ , and - δ sequences, as well as how the TCRs as a whole relate to the immunoglobulin sequences as a whole. Previous studies (e.g., Bernstein et al. 1996; Greenberg et al. 1995, 1996; Schluter, Bernstein, and Marchalonis 1997) have investigated relationships of TCR and Ig genes, but they included few TCR sequences and usually have not included outgroup analysis, precluding inferences about branching order among all six TCR and Ig types.

Materials and Methods

Amino acid sequences for TCR- α , - β , - γ , and - δ chains, Ig heavy (IgH) and light (IgL) chains (only λ

chains were included), and a recently discovered novel antigen receptor in sharks were retrieved from the GenBank, GenPept, and PIR databases (table 1). Although TCRs and Ig's are known to exist in all jawed vertebrates, GenBank's representation of the various types of sequences among various animal taxa is uneven. The current data set (compiled in January 1998) represents an effort to include as many animal species as possible without undue focus on overrepresented groups such as humans, apes, and rodents. Chondrichthyan sequences are included among all four types of TCR and both types of immunoglobulin (table 1).

TCRs and Ig's are composed of subunits which are themselves composed of homologous regions known as the constant (C) and variable (V) regions. Generally, the C regions comprise the backbone of the receptor molecule, while the V regions include the sites of actual antigen binding. The C and V regions may be under very different selective pressures and could have very different evolutionary histories and so are analyzed separately. IgH's contain differing numbers of constant regions, depending on the type of Ig molecule they form (e.g., IgM, IgG, etc.). To minimize the effects of such structural variation, the constant regions of IgH sequences are represented by the most membrane-proximal C region, usually the C-4 region of IgM, which is thought to be homologous across taxa (Fellah et al. 1992). The constant and variable regions are joined together by shorter sequences called joining (J) and diversity (D) regions; thus, TCRs and Ig's are heterodimers composed of one VJC and one VDJC molecule. The J and D regions are highly variable even among TCR lineages of a single individual (Schatz, Oettinger, and Schliessel 1992) and so are unlikely to be phylogenetically useful.

Sequences were aligned using the program PILEUP (Feng and Doolittle 1987) in the GCG package, version 8.1-UNIX (Devereaux, Haeberli, and Smithies 1984). The program CLUSTAL W (Higgins and Sharp 1988) was also used and gave similar results, but results presented here are based primarily on the PILEUP alignments. Various gap penalties were tried. A lower gap penalty results in the insertion of more gaps in the alignments and greater similarity among sequences. An extremely low gap penalty tends to eliminate any phylogenetic signal, since aligned sequences are either highly similar or full of missing characters. An excessively high gap penalty is likely to lead to misalignment in cases in which insertions and deletions have occurred. A gap penalty of 2.0 (using PILEUP) guaranteed perfect alignment of two highly conserved motifs common to all TCR and Ig sequences, so this gap penalty was adopted for all alignments. The two motifs are the highly conserved WYRK/Q and YFCA motifs in the second and third framework regions of the V region genes. Overall, then, the alignments were conservative. The V region alignment, including the complementarity determining regions (CDRs), was truncated at a consensus length of 134 amino acids, ending at the highly conserved G-G hinge region between the V and D/J regions in the fourth framework region. Excluding the CDRs resulted in a consensus length of only 64. The C region

alignment was truncated at a consensus length of 150 amino acids, the approximate length of the IgL C regions.

Two different tree-building methods were used. The parsimony criterion was used to find the set of trees that minimizes the number of changes in character states across the whole tree, ignoring invariant or uninformative characters. Using the program PAUP, version 3.1 (Swofford 1993), parsimony trees were found by heuristic search with the branch-and-bound search option and 1,000 iterations. Each iteration starts the search with a different tree (i.e., set of branching patterns) and finds the shortest set of trees based on that starting point. Through dozens of trials and hundreds of hours of tree building with this method, we found that the set of shortest trees was almost always found in the first 10 iterations. Successive approximations character weighting (SACW; Carpenter 1988) was used to further improve tree resolution. Each character was reweighted by its goodness of fit to the most parsimonious trees, and the trees were recalculated. The process was repeated until a constant tree topology was obtained. Thus, each parsimony consensus tree should represent the best fit between tree topology and the current data set.

The second tree-building method was based on the use of genetic distance matrices and the neighbor-joining method of Saitou and Nei (1987). These trees were calculated using the PHYLIP (version 3.57c) programs PROTDIST and NEIGHBOR (Felsenstein 1993). One thousand neighbor-joining trees were calculated based on a bootstrap of 1,000 separate genetic distance matrices; the final trees are consensus of these 1,000 trees.

Trees based on outgroup comparison can be used to infer both patterns of relationship among taxa and branching order over time. Several members of the IGSF have been proposed to be representative of a primordial IGSF molecule which probably had only one Ig-like region that would have been ancestral to both the C and the V regions of TCRs and Ig's (Hunkapiller and Hood 1989). One of these is the TCR-associated cell surface molecule CD3. CD3 exists in three forms, two of which, CD3- γ and CD3- δ , were used as outgroup sequences. Several other molecules were also tested as outgroups, including the cell surface molecules CD3- ϵ , Thy-1, CD4, and CD8 (Williams 1987), but this did not result in improved resolution of TCR and Ig topologies.

Results

Constant Region Trees

The C region alignment was 150 characters long. Of these, 143 characters were informative. Using the parsimony criterion, 73 C region trees of length 2,344 were found. These were reduced to 2 trees of length 2,345 using SACW (fig. 2). In these trees, the earliest branches are represented by TCR sequences, specifically TCR- γ . Of the four TCR groups, only TCR- α is monophyletic. The three chondrichthyan TCR- δ sequences appear to be more similar to each other than to their tetrapod homologs, which are associated with IgL. IgH sequences are nested well within the IgL sequences, im-

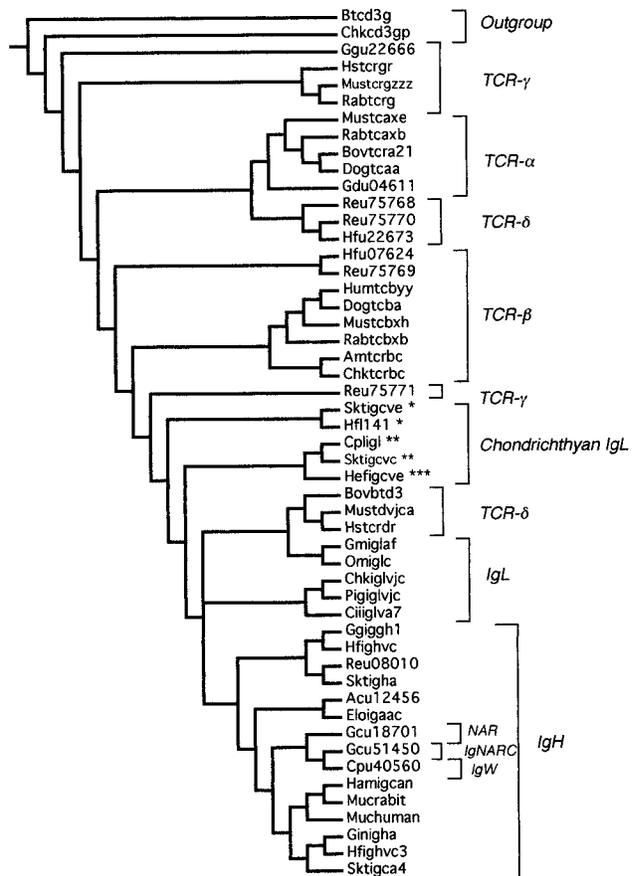


FIG. 2.—Nelson parsimony consensus tree for the constant region genes. The original set of 73 equally parsimonious trees was reduced using successive approximations character weighting to 2 trees of length 2,345 with a consistency index of 0.566. Asterisks indicate the chondrichthyan group I (*), II (**), and III (***) immunoglobulin light-chain genes.

plying that the light chains are older and gave rise to the heavy chains. The IgL sequences as a whole are polyphyletic, with the chondrichthyan group I (Hfl141, Sktigcve), II (Cpligl, Sktigcvc), and III (Hefigcvc) IgL sequences forming the earliest branches. The IgH sequences comprise a monophyletic group that includes shark Novel Antigen Receptor (NAR, Gcu18701), IgNARC (Gcu51450), and IgW (Cpu40560).

A second set of trees was calculated based on a bootstrap analysis of genetic distances (fig. 3). These trees are probably superior to the parsimony trees, since they include the following monophyletic groups: TCR- α/δ , TCR- β , TCR- γ , and all immunoglobulins. As in the previous trees, the earliest branches are TCR sequences, although in this analysis the oldest type is TCR- α/δ . The shark NAR, IgNARC, and IgW clearly are closely associated with the IgH sequences. The IgL sequences are polyphyletic, with osteichthyan and tetrapod IgL representing branches completely separate from the chondrichthyan group I, II, and III light chains.

Variable Region

The variable regions of T-cell receptors and immunoglobulins are divided into framework (FR) and

complementarity-determining (CDR) regions. The latter are highly variable, conferring epitope-specific sensitivity on the receptor, and a priori are expected not to be phylogenetically informative. Therefore, the CDRs were excluded, leaving a V region alignment with a consensus length of 64 characters, of which 55 were informative.

Using the parsimony criterion, an initial set of 119 trees of length 894 steps was found. These were reduced to 2 trees of length 896 using SACW (fig. 4). The IgH sequences, including IgW (Cpu40560), comprise the basal branch, giving rise to the TCRs as a whole, which in turn eventually give rise to the IgL group. IgL and TCR sequences are more closely related to each other than either is to the IgH sequences. None of the TCR groups is monophyletic. The position of axolotl TCR- β (Amttcrb) is problematic, rendering the IgL sequences paraphyletic. We note that if a second amphibian sequence, *Xenopus* TCR- β (U60436; Chretien et al. 1997), is added to the analysis, the axolotl sequence remains associated with IgL, although the *Xenopus* sequence groups with all other TCR- β sequences (data not shown). If Amttcrb is removed, TCR- β and IgL become monophyletic (data not shown). However, the tetrapod IgL sequences (those not including the chondrichthyan group I, II, and III sequences; table 1) do comprise a monophyletic group. The position of IgNARC (Gcu51450) is unclear: while it is probably associated with the IgH group, it appears as a separate branch intermediate between the IgH and TCR sequences. However, the shark NAR is clearly associated with TCR- γ and embedded in the TCR+IgL group.

Based on genetic distances, the neighbor-joining consensus tree (fig. 5) also indicates that IgH make up the basal branch and that TCR+IgL compose a monophyletic group. However, this tree differs from the parsimony tree in important respects. As was the case for the C region trees, the distance-based trees are probably superior to the parsimony-based trees, since they include the following monophyletic groups: IgH (including IgW and IgNARC), all IgL, TCR- α/δ , and TCR- γ . Perhaps oddly, the TCR- β sequences are polyphyletic, splitting into two groups, one comprising rabbit (Rabtcxbx), human (Humtcbxy), and axolotl (Amttcrb) sequences, and the other comprising shark (Hfu07624 and Reu75769), mouse (Mustcbxh), and chicken (Chktrbc) sequences. This may be an artificial consequence of attraction of the Rabtcxbx and Humtcbxy sequences to the problematical sequence Amttcrb (see above), which we found often appeared in odd places on different trees (we did not exclude it because we have no objective grounds for believing this sequence to be odd, and because the addition of the *Xenopus* TCR- β sequence, as noted above, does not solve the problem [data not shown]). Polyphyly of the TCR- β sequences was also found when the CDR regions were not excluded, using both the maximum-parsimony and the genetic distance-based methods (trees not shown). The positions of IgW and IgNARC are clearly embedded within the IgH groups. However, as in the parsimony tree, NAR is associated with the TCR(+IgL) sequences.

Discussion

$\gamma\delta$ T Cells as Primordial Immune Cells

Neither the V nor the C region trees support the hypothesis of monophyly of the TCR genes, but they disagree as to the overall branching patterns of the four TCR and two Ig sequence sets. Trees based on C region sequences suggest that TCR sequences evolved earliest, that these gave rise to IgL sequences, and that the last to appear were the IgH sequences. In marked contrast, trees based on V region sequences suggest that IgH sequences are the oldest sequences, followed by the TCRs, and then the IgL. These two hypotheses of relationship among TCR and Ig sequences might seem to be mutually exclusive were it not that the C and V region genes are probably subject to quite different selective regimes, due to the different roles of the C and V regions in the molecule. Topological relationships among the V region sequences could be more reflective of functional similarity than of historical descent, due to convergence of the types of sequences required for direct (as in the Ig's) versus indirect (as in the $\alpha\beta$ -TCRs) antigen binding. By the same token, the C region sequences, which are not involved in antigen recognition and binding, might be less likely to be convergent due to such constraints, and their relationships might be more reflective of phylogenetic history. Furthermore, the evolutionary history of the gene sequences does not necessarily reflect their current roles as members of modern TCR and Ig receptors, since their functions in modern molecules could have arisen after the genes themselves began to diverge.

If, as indicated by the C region trees, TCR- γ and $-\alpha/\delta$ sequences represent the earliest immune receptor sequences, and since the TCR- α/δ sequences are highly convergent (Guglielmi et al. 1988), this suggests that $\gamma\delta$ -TCRs represent the oldest of the extant immune receptors, whereas $\alpha\beta$ -TCRs and Ig's are more recent. The V region trees, which may reflect functional relationships among antigen-binding sequences, indicate that the IgH's are the earliest, implying that the earliest immune receptors had Ig-like antigen-binding properties. Despite these important distinctions in V and C region tree topologies, both data sets indicate that direct antigen-recognition is the primitive condition for T and B cells, while indirect antigen recognition is clearly a derived characteristic of $\alpha\beta$ T cells. Indeed, it would be startling if this were not the case, since direct ligand recognition and binding is typical of other cell surface adhesion molecules in the IGSF (Hunkapiller and Hood 1989).

Even before the antigen receptors of T cells were identified and long before it was realized that two classes of T cells existed, it was suggested that the ancestral immune cells were probably T cells (Marchalonis 1977). Since then, several authors (Marchalonis and Schluter 1990; Stewart 1992; Thompson 1995; Flajnik 1998) have suggested that primordial immune cells were $\gamma\delta$ -like. Many characteristics of $\gamma\delta$ T cells fit the role of a first line of immune defense (Brenner, Strominger, and Krangel 1988; Raulet 1989; Allison and Havran 1991). The primordial immune cell would have to have been

Table 1
Sequences Used in Phylogenetic Analyses

Gene	Species	Locus Name	Accession No.	Data Set
CD3.....	<i>Bos taurus</i> (cow)	Btd3g	X53268	C, V
	<i>Gallus gallus</i> (chicken)	Chkcd3gp	M59925	C, V
TCR- α	<i>Bos taurus</i> (cow)	Bovtcr21	D10394	C
	<i>Canis familiaris</i> (dog)	Dogtcaa	M97511	C
	<i>Homo sapiens</i> (human)	Humtcaxy	M13741	V
	<i>Gallus domesticus</i> (chicken)	Gdu04611	U04611	C, V
	<i>Mus musculus</i> (mouse)	Mustcaxe	M16119	C, V
	<i>Oryctolagus cuniculus</i> (rabbit)	Rabtcaxb	M12885	C
	<i>Raja eglantera</i> (skate)	Reu75749	U75749	V
	<i>Raja eglantera</i> (skate)	Reu75768	U75768	C
	<i>Ovis garies</i> (sheep)	Shptcra	M55622	V
TCR- β	<i>Ambystoma mexicanum</i> (axolotl)	Amtcrbc	X70168	C
	<i>Ambystoma mexicanum</i> (axolotl)	Amttcrb	L33400	V
	<i>Gallus gallus</i> (chicken)	Chkcrbc	M81149	C, V
	<i>Canis familiaris</i> (dog)	Dogtcb	M97510	C
	<i>Heterodontus francisci</i> (horn shark)	Hfu07624	U07624	C, V
	<i>Homo sapiens</i> (human)	Humtcbyy	M12886	C, V
	<i>Mus musculus</i> (mouse)	Mustcbxb	M11456	C, V
	<i>Oryctolagus cuniculus</i> (rabbit)	Rabtcxb	M13895	C, V
	<i>Raja eglantera</i> (skate)	Reu75769	U75769	C, V
TCR- γ	<i>Bos taurus</i> (cow)	Bovtcr08	D16119	V
	<i>Gallus gallus</i> (chicken)	Ggu22666	U22666	C, V
	<i>Homo sapiens</i> (human)	Hstcrgr	Y00790	C
	<i>Mus musculus</i> (mouse)	Mustcrgzzz	M54989	C, V
	<i>Ovis aries</i> (sheep)	Oatcvg3	Z13003	V
	<i>Oryctolagus cuniculus</i> (rabbit)	Rabterg	D38134	C, V
	<i>Raja eglantera</i> (skate)	Reu75761	U75761	C
	<i>Raja eglantera</i> (skate)	Reu75771	U75771	V
TCR- δ	<i>Bos taurus</i> (cow)	Bovbtd3	D90419	C
	<i>Heterodontus francisci</i> (horn shark)	Hfu22673	U22673	C, V
	<i>Homo sapiens</i> (human)	Hstcrdr	X06557	C, V
	<i>Mus musculus</i> (mouse)	Mustdvjca	M37694	C
	<i>Raja eglantera</i> (skate)	Reu75768	U75768	C
	<i>Raja eglantera</i> (skate)	Reu75770	U75770	C, V
NAR.....	<i>Ginglymostoma cirratum</i> (nurse shark)	Gcu18701	U18701	C, V
IgH.....	<i>Amia calva</i> (bowfin) IgM	Acu12456	U12456	C
	<i>Ambystoma mexicanum</i> (axolotl)	Amtigmua	L20243	V
	<i>Anas platyrhynchos</i> (duck) IgY	B46529	B46259 ^a	V
	<i>Carcharhinus leucas</i> (bull shark)	Clu50611	U50611	V
	<i>Carcharhinus plumbeus</i> (sandbar shark)	Cpu40560	U40560	C, V
	<i>Elops saurus</i> (ladyfish)	Eloigaac	M26812	C, V
	<i>Ginglymostoma cirratum</i> (nurse shark) IgNARC	Gcu51450	U51450	C, V
	<i>Gallus gallus</i> (chicken) IgD	Ggiggh1	X07174	C
	<i>Ginglymostoma cirratum</i> (nurse shark)	Ginigha	M92851	C, V
	<i>Mesocricetus auratus</i> (Syrian hamster) IgM	Hamigcan	M23867	C
	<i>Heterodontus francisci</i> (horn shark)	Hfighvc	X07781	C
	<i>Heterodontus francisci</i> (horn shark)	Hfighvc3	X07784	C
	<i>Homo sapiens</i> (human) IgM	Muc_human	P01871 ^b	C
	<i>Oryctolagus cuniculus</i> (rabbit) IgM	Muc_rabbit	P03988 ^b	C
	<i>Ovis aries</i> (sheep)	Oaigmu	X59994	V
	<i>Raja eglantera</i> (skate)	Reu08010	U08010	C, V
	<i>Raja erinacea</i> (skate) IgM	Sktigca4	M29677	C
<i>Raja erinacea</i> (skate) IgM	Sktigha	M29672	C, V	
<i>Xenopus laevis</i> (African clawed frog) IgM	Xeligha	M20484	V	
IgL.....	<i>Gallus gallus</i> (chicken)	Chkiglvjc	M24403	C, V
	<i>Cairina moschata</i> (duck)	Ciiiglva7	M25726	C
	<i>Carcharhinus plumbeus</i> (sandbar shark) group II	Cplig1	M81314	C, V
	<i>Gadus morhua</i> (Atlantic cod)	Gmiglaf	X68516	C
	<i>Heterodontus francisci</i> (horn shark) group III	Hefigcve	L25561	C, V
	<i>Heterodontus francisci</i> (horn shark) group I	Hefiglaaa	M64307	V
	<i>Heterodontus francisci</i> (horn shark) group I	Hfl141	X15315	C, V
	<i>Homo sapiens</i> (human)	Hsiglv11	X14615	V
	<i>Mus musculus</i> (mouse)	Musiglcvx	L31404	V
	<i>Oncorhynchus mykiss</i> (rainbow trout)	Omiglc	X65260	C, V
	<i>Sus scrofa</i> (pig)	Pigiglvjc	M59322	C
	<i>Raja erinacea</i> (skate) group II	Sktigcvc	L25566	C, V
	<i>Raja erinacea</i> (skate) group I	Sktigcvc	L25568	C, V

capable of self- versus nonself-recognition, as well as of protective effector functions such as cytolysis, and its receptor would likely have been membrane-bound rather than secreted. Both of these capabilities are present in $\gamma\delta$ T cells. It might also be expected that the primordial immune cells would have been capable of recognizing a wide array of antigens, many of which could be ingested through the gut. Despite their limited V-gene repertoire in comparison with $\alpha\beta$ V genes, $\gamma\delta$ TCRs display far greater N region (a region of random nucleotide addition between the V and D regions) variability and are potentially capable of recognizing far more antigens than are $\alpha\beta$ T cells. $\gamma\delta$ T cells can respond to antigen challenge very quickly, without a requirement for professional APCs, and have been suggested as a first line of immune defense in mammalian immune systems. On the other hand, $\alpha\beta$ T cells require professional APC and antigen presentation by MHC molecules and exhibit highly specialized interactions with B cells in the maturation of antibody responses. A characteristic of most receptors in the IGSF is that they are membrane-bound, an ancestral characteristic that is retained in $\gamma\delta$ and $\alpha\beta$ T cells. However, B cells can secrete immunoglobulins, a trait that is likely derived (Flajnik 1998). In fact, while many characteristics of $\gamma\delta$ T cells are suggestive of an ancient and primary role in immune defense, the roles of $\alpha\beta$ T cells and B cells imply highly specialized, and probably derived, roles in the immune system, supporting the conclusion that $\gamma\delta$ -like T cells gave rise to $\alpha\beta$ T cells and B cells.

Several lines of structural evidence also support the suggestion that $\gamma\delta$ T cells exhibit ancestral antigen-binding characteristics. $\gamma\delta$ -TCRs recognize a wider variety of antigens than do $\alpha\beta$ -TCRs, including both peptides and phospholipids, and they are especially reactive to conserved molecules such as mycobacterial and heat shock proteins (Constant et al. 1994; Burk, Mori, and DeLibero 1995). An analysis of the length distributions of V region chains, those that actually interact with antigen, in T and B cells indicates greater similarity of $\gamma\delta$ -TCRs to Ig's than to $\alpha\beta$ -TCRs (Rock et al. 1994), and this also strongly suggests that $\gamma\delta$ -TCRs bind directly to antigens, as do immunoglobulins. $\alpha\beta$ T cells display two cell surface markers, CD4 and CD8, that are required for binding of the $\alpha\beta$ -TCR to MHC molecules, but most $\gamma\delta$ T cells carry neither of these; thus, they are not MHC-restricted, and no alternative presentation molecules have been discovered. The differences in antigen binding are probably reflected in the three-dimensional structures of the two types of TCR. Although the structure of $\gamma\delta$ -TCR molecules has not been elucidated, the structure of an $\alpha\beta$ -TCR bound to a peptide+MHC complex has been studied (Garcia et al. 1996). The antigen-binding surface of the $\alpha\beta$ -TCR is fairly flat and similar to the undulating surfaces where Ig's bind to

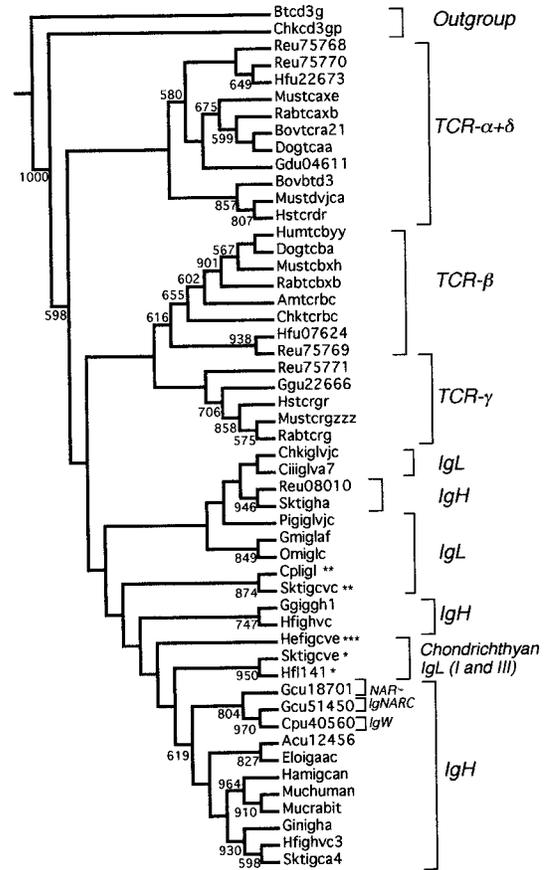


FIG. 3.—Bootstrap neighbor-joining tree for the constant region genes. Numbers indicate bootstrap values over 1,000 trees. Asterisks indicate the chondrichthyan group I (*), II (**), and III (***) immunoglobulin light-chain genes.

proteins, but it is also much smaller than the Ig-binding areas, perhaps bestowing greater complementarity of fit with antigen (Garcia et al. 1996). It may therefore be predicted that the binding surface of the $\gamma\delta$ -TCR is also relatively flat but more like an Ig than an $\alpha\beta$ -TCR in size, as would be expected for recognition of a wider variety of antigens.

Origin of VJ-VDJ Heterodimeric Immune Receptors

All modern TCRs and Ig's are heterodimers composed of one VJC molecule and one VDJC molecule. The only exceptions appear to occur in animals such as camelids that have functioning and abundant IgG molecules composed of heavy-chain dimers without light chains (Hamers-Casterman et al. 1993), and possibly in NAR-bearing cells (Marchalonis et al. 1998). However there is no evidence in our study or in previous studies (e.g., Greenberg et al. 1995; Schluter, Bernstein, and Marchalonis 1997) that the VJC (TCR- α , TCR- γ , and IgL) and the VDJC (TCR- β , TCR- δ , and IgH) sequences

←

NOTE.—TCR = T-cell receptor; NAR = novel antigen receptor; IgH = immunoglobulin heavy chain; IgL = immunoglobulin light chain. All sequences were retrieved from GenBank except where otherwise noted (references may be found in the full GenBank entry).

^a From the PIR database.

^b From the GenPept database.

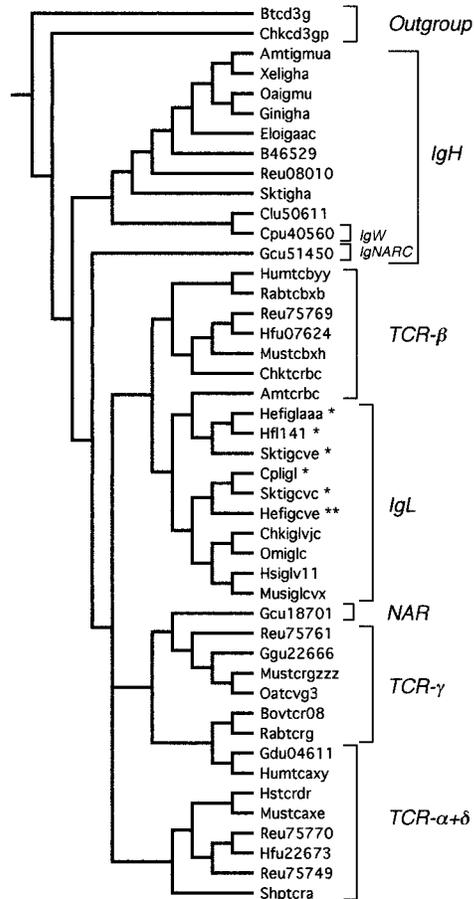


FIG. 4.—Nelson parsimony consensus tree for the variable region genes. The original set of 119 trees was reduced using successive approximations character weighting to 2 trees of length 896 with a consistency index of 0.535. Asterisks indicate the chondrichthyan group I (*), II (**), and III (***) immunoglobulin light-chain genes.

constitute monophyletic groups, as might be predicted if they evolved separately from proto VJ and VDJ molecules (Thompson 1995; Rast et al. 1997). In fact, our analyses suggest that the TCR- α and TCR- δ sequences are equally ancient and could even represent the two most ancient TCR-V sequences (fig. 5). The branching order in figure 5 could be incorrect if the true relationship between TCR- α and TCR- δ sequences is obscured by convergence: mammalian TCR- α and TCR- δ variable genes are so highly convergent that they even display overlapping usage by $\alpha\beta$ - and $\gamma\delta$ -TCRs (Guglielmi et al. 1988; Takihara et al. 1989). Alternatively, if the branching order is correct, then the oldest V region genes are IgH, which gave rise to TCR- $\alpha+\delta$ sequences, from which TCR- γ and TCR- β subsequently arose through duplication (Rast et al. 1997), and from which the IgL sequences also evolved. Thus, the original dimeric molecule could have been a VDJC/VDJC molecule, implying that as TCR and Ig molecules evolved, the D segment has been either retained or lost in different lineages. Thus the origin of the TCR sequences could have coincided with the origin of the VJC/VDJC heterodimeric immune receptor, perhaps replacing some earlier, homodimeric receptor composed of two IgH-like

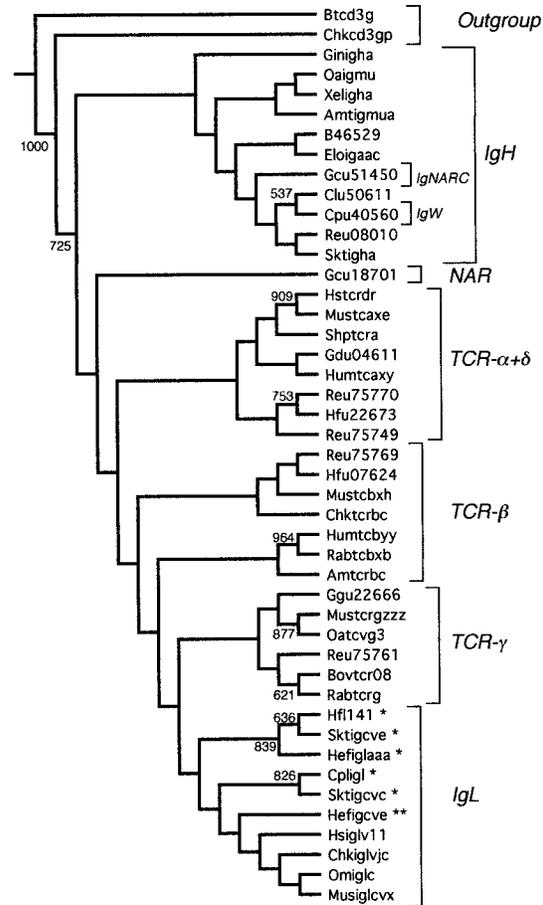


FIG. 5.—Bootstrap neighbor-joining tree for the variable region genes. Numbers indicate bootstrap values over 1,000 trees. Asterisks indicate the chondrichthyan group I (*), II (**), and III (***) immunoglobulin light-chain genes.

chains. Rast and Litman (1998) have also noted the evolutionary lability of D regions, as is evident in vertebrate IgH genes where receptor molecules may be VJ, VDJ, VDDJ, or even VDDDJ (Marchalonis, Schluter, and Edmundson 1997). Although presence or absence of a D segment does not appear to be a phylogenetically useful characteristic, it must be of functional significance at the level of the entire receptor molecule.

Close Relationship of TCR and IgL

Previous studies utilizing fewer TCR sequences have found that the TCR and IgL variable region genes formed one cluster and the IgH genes formed a second cluster (Greenberg et al. 1995; Bernstein et al. 1996; Schluter, Bernstein, and Marchalonis 1997). We also found that the IgL and TCR variable sequences were more closely related to each other than to the IgH sequences, with the IgH variable genes giving rise to the TCR and IgL genes. Interestingly, when the conformations of the binding surfaces (encoded by V region genes) of $\alpha\beta$ -TCR, IgL, and IgH molecules were compared, there seemed to be a closer match of the TCR with IgL than with IgH (Garcia et al. 1996; Marchalonis, Schluter, and Edmundson 1997), as might be expected

if the similarity in molecular conformation reflects an orthologous relationship between TCRs and IgL's.

IgL genes make up two basic groups. The first group comprises the chondrichthyan IgL genes (groups I, II, and III; see table 1), and the second group comprises the λ and κ light-chain genes. In our analyses, the IgL constant genes as a whole do not form a monophyletic group, because the group I, II, and III genes form separate branches that actually appear to be younger than the remaining, "modern," light-chain sequences (figs. 2 and 3). However, the variable region sequences of all light chains appear to be monophyletic (if *Amttcrb* is removed; figs. 4 and 5), with the group I, II, and III sequences being the most ancient, as found by Rast et al. (1994).

Relationships of NAR to IgW and TCR

Strongly supported in our analyses of the C region genes (figs. 2 and 3) is a close relationship between the recently discovered shark NAR (Gcu18701) and the IgH's, including IgW (Cpu40560) and IgNARC (Gcu51450). When NAR was first discovered (Greenberg et al. 1995), it was speculated that a NAR-bearing cell might bind to antigen in a manner resembling that of immunoglobulins (Greenberg et al. 1995), and this seems very likely, since comparisons of IgW, NAR, and IgNARC C region genes confirm that they are all genuine immunoglobulins (Schluter, Bernstein, and Marchalonis 1997). Although IgM was long thought to represent the primordial immunoglobulin type (Fellah et al. 1992), the fact that the IgW class is found only in sharks and their allies may indicate that IgW is primordial (Bernstein et al. 1996; Shen et al. 1996; Schluter, Bernstein, and Marchalonis 1997). Our analyses of the C region genes offer little support for this, since they indicate that the IgH sequences are in fact younger than the IgL sequences. On the other hand, the V region sequences do support the hypothesis that IgW sequences evolved earlier than IgM (figs. 4 and 5). There is little support in our analyses for the hypothesis that among the non-IgW heavy-chain sequences, the chondrichthyan IgM sequences are archaic (Shen et al. 1996), but we may not have included sufficient diversity of sequences to address this issue.

Although the C region trees unequivocally group IgW, IgNARC, and NAR molecules with other immunoglobulin heavy chains, the same cannot be said of the V region trees, in which NAR is more closely allied with TCR than with Ig sequences (figs. 4 and 5). This divergent evidence from C and V region trees with respect to the relationship between NAR and other receptor sequences has also been noted in several previous studies (Greenberg et al. 1995, 1996; Schluter, Bernstein, and Marchalonis 1997) and suggests that NAR is not simply a rather odd member of the IgW class, but a molecule with its own distinct role. Klein (1998) has interpreted the "chimeric" nature of the NAR molecule in two ways. First, the NAR V region gene sequence is either highly convergent with or actually orthologous to a fully evolved (*sic*) TCR V region sequence. This is plausible if NAR represents an early branch in TCR phylogeny,

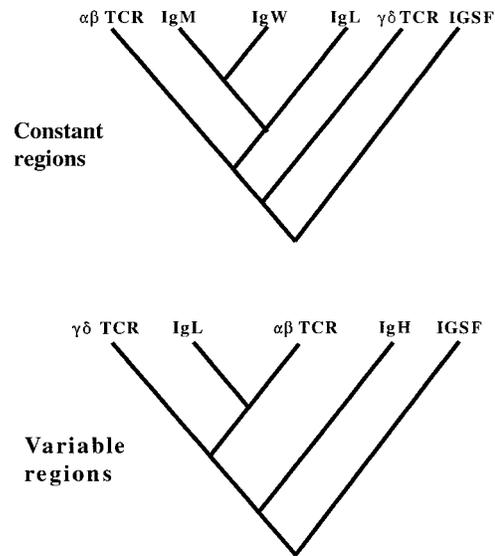


FIG. 6.—The evolution of vertebrate immune receptors: evolutionary relationships based on the results of this and other studies (see text for references).

as is indicated in figure 5 (and in fig. 4 to a lesser extent). Klein's (1998) second suggestion (and the one he prefers) is that the similarity between TCR and NAR V regions is an artifact of phylogenetic analysis, because over the last 400 Myr, any phylogenetic signal would have been obliterated by mutations in the NAR protein sequence. We do not accept this argument, because there is no reason to believe that this should be more true for NAR than for any of the other TCR or Ig sequences which we and others have analyzed, including the IgH V region sequences, which appear to be the most ancient. In general, we believe that our own and other analyses support Rast and Litman's (1998) suggestion that NAR is a divergent IgH gene type. NAR might be divergent not in effector function (since it is a bona fide immunoglobulin), but in the way it binds antigens or in the type of antigens bound.

Conclusions

From a phylogenetic point of view, it is a little surprising that so much effort has been devoted to finding the antigen-presenting molecule for $\gamma\delta$ T cells (perusal of immunological literature generated in the last 15 years indicates the huge amount of effort and resources put into this search). It seems that this search has been based on an assumption that current functions of immune receptors, especially in humans (and perhaps mice; Flajnik 1998) must reflect the evolutionary histories of the genes that encode them, as summarized in figure 1. And yet this study and others, based on comparisons of TCR, IgL, and IgH genes across vertebrate taxa, clearly support an alternative evolutionary history of these sequences, as summarized in figure 6. The major stumbling block to this scheme is the apparent antiquity of the IgH V genes, since it is difficult to imagine what their early function was if the primordial antigen receptor resembled a $\gamma\delta$ T cell and did not use the IgH

V genes for its receptors. On the other hand, the relatively “young age” of the IgH C genes supports the hypothesis that secretion of immune receptors (i.e., modern Ig’s) is a derived characteristic. It is now clear that indirect, MHC-restricted antigen recognition must be a derived characteristic of $\alpha\beta$ T cells, while $\gamma\delta$ T cells exhibit the primitive condition of direct antigen binding. This removes a major stumbling block in several schemes for the evolution of vertebrate immune systems (e.g., Stewart 1992), because it shows that the origins of diverse T-cell characteristics, such as antigen recognition and effector functions, may be separate, independent evolutionary events. Continued phylogenetic analysis of TCR and Ig relationships, especially as more suitable outgroup sequences are obtained from organisms such as cyclostomes or sharks, should continue to improve our understanding of immune system evolution.

Acknowledgments

We thank Laurence Packer and several anonymous reviewers for their very helpful comments and Dan Graur for his patience. This work was supported by NIH operating grants AR-39282 and AI-38583 to J.L.N. We also thank Michael Parker of the FHCRC Biocomputing Shared Resource Center and acknowledge its support from National Cancer Institute award NCI P30 CA15704.

LITERATURE CITED

- ALLISON, J. P., and W. L. HAVRAN. 1991. The immunobiology of T cells with invariant $\gamma\delta$ antigen receptors. *Annu. Rev. Immunol.* **9**:679–705.
- BARTL, S., D. BALTIMORE, and I. L. WEISSMAN. 1994. Molecular evolution of the vertebrate immune system. *Proc. Natl. Acad. Sci. USA* **91**:10769–10770.
- BERNSTEIN, R. M., S. F. SCHLUTER, S. SHEN, and J. J. MARCHALONIS. 1996. A new high molecular weight immunoglobulin class from the carcharhine shark: implications for the properties of the primordial immunoglobulin. *Proc. Natl. Acad. Sci. USA* **93**:3289–3293.
- BRENNER, M. B., J. L. STROMINGER, and M. S. KRANGEL. 1988. The $\gamma\delta$ T cell receptor. *Adv. Immunol.* **43**:133–192.
- BURK, M. R., L. MORI, and G. DELIBERO. 1995. Human V γ 9-V δ 2 cells are stimulated in a cross-reactive fashion by a variety of phosphorylated metabolites. *Eur. J. Immunol.* **25**:2052–2058.
- CARPENTER, J. M. 1988. Choosing among multiple equally parsimonious cladograms. *Cladistics* **4**:291–296.
- CHIEN, Y.-H., and R. JORES. 1995. T cells with B-cell-like properties. *Curr. Biol.* **5**:1116–1118.
- CHRETIEN, I., A. MARCUZ, J. FELLAH, J. CHARLEMAGNE, and L. DU PASQUIER. 1997. The T cell receptor beta genes of *Xenopus*. *Eur. J. Immunol.* **27**:763–771.
- CONSTANT, P., F. DAVODEAU, M.-A. PEYRAT, Y. POQUET, G. PUZO, M. BONNEVILLE, and J.-J. FOURNIE. 1994. Stimulation of human $\gamma\delta$ T cells by nonpeptidic mycobacterial ligands. *Science* **264**:267–270.
- DEVEREAUX, J., P. HAEBERLI, and O. SMITHIES. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**:387–395.
- FELLAH, J. S., M. V. WILES, J. CHARLEMAGNE, and J. SCHWAGER. 1992. Evolution of vertebrate IgM: complete amino acid sequences of the constant region of *Ambystoma mexicanum* μ chain deduced from cDNA sequences. *Eur. J. Immunol.* **22**:2595–2601.
- FELSENSTEIN, J. 1993. PHYLIP (phylogeny inference package). Version 3.57c. Distributed by the author, Department of Genetics, University of Washington, Seattle.
- FENG, D.-F., and R. F. DOOLITTLE. 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* **25**:351–360.
- FLAJNIK, M. F. 1998. Churchill and the immune system of ectothermic vertebrates. *Immunol. Rev.* **166**:5–14.
- GARCIA, K. C., M. DEGAN, R. L. STANFIELD, A. BRUNMARK, M. R. JACKSON, P. A. PETERSON, L. TEYTON, and I. A. WILSON. 1996. An $\alpha\beta$ T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. *Science* **274**:209–219.
- GREENBERG, A. S., D. AVILA, M. HUGHES, A. L. HUGHES, E. MCKINNEY, and M. F. FLAJNIK. 1995. A novel antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. *Nature* **374**:168–173.
- GREENBERG, A. S., A. L. HUGHES, J. GUO, D. AVILA, E. C. MCKINNEY, and M. F. FLAJNIK. 1996. A novel “chimeric” antibody class in cartilaginous fish: IgM may not be the primordial immunoglobulin. *Eur. J. Immunol.* **26**:1123–1129.
- GUGLIELMI, P. E., F. DAVI, L. D’AURIOL, J. BORIES, J. DAUSET, and A. BENSUSSAN. 1988. Use of a variable α region to create a functional T-cell receptor δ chain. *Proc. Natl. Acad. Sci. USA* **85**:5634–5638.
- HAAS, W., P. PEREIRA, and S. TONEGAWA. 1993. Gamma/delta cells. *Annu. Rev. Immunol.* **11**:637–685.
- HAMERS-CASTERMAN, C., T. ATARHOUCHE, S. MUYLDERMANS, G. ROBINSON, C. HAMERS, E. BAJYANA SONGA, N. BENDAHMAN, and R. HAMERS. 1993. Naturally occurring antibodies devoid of light chains. *Nature* **363**:446–448.
- HIGGINS, D. G., and P. M. SHARP. 1988. CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. *Gene* **73**:237–244.
- HUNKAPILLER, T., and L. HOOD. 1989. Diversity of the immunoglobulin gene superfamily. *Adv. Immunol.* **44**:1–63.
- KAUFMANN, S. H. E. 1996. $\gamma\delta$ and other unconventional T lymphocytes: what do they see and what do they do? *Proc. Natl. Acad. Sci. USA* **93**:2272–2279.
- KLEIN, J. 1998. In an immunological twilight zone. *Proc. Natl. Acad. Sci. USA* **95**:11504–11505.
- MARCHALONIS, J. J. 1977. *Immunity in evolution*. Harvard University Press, Cambridge, Mass.
- MARCHALONIS, J. J., and S. F. SCHLUTER. 1990. Origins of immunoglobulins and immune recognition molecules. *BioScience* **40**:758–768.
- MARCHALONIS, J. J., S. F. SCHLUTER, R. M. BERNSTEIN, and V. S. HOHMAN. 1998. Antibodies of sharks: revolution and evolution. *Immunol. Rev.* **166**:103–122.
- MARCHALONIS, J. J., S. F. SCHLUTER, and A. B. EDMUNDSON. 1997. The T-cell receptor as immunoglobulin: paradigm regained. *Proc. Soc. Exp. Biol. Med.* **216**:303–318.
- RAST, J. P., M. K. ANDERSON, T. OTA, R. T. LITMAN, M. MARGITTAI, J. J. SHAMBLOTT, and G. W. LITMAN. 1994. Immunoglobulin light chain class multiplicity and alternative organizational forms in early vertebrate phylogeny. *Immunogenetics* **40**:83–99.
- RAST, J. P., M. K. ANDERSON, S. J. STRONG, C. LUER, R. T. LITMAN, and G. W. LITMAN. 1997. α , β , γ , and δ T cell antigen receptor genes arose early in vertebrate phylogeny. *Immunity* **6**:1–11.

- RAST, J. P., and G. W. LITMAN. 1998. Towards understanding the evolutionary origins and early diversification of rearranging antigen receptors. *Immunol. Rev.* **166**:79–86.
- RAULET, D. H. 1989. The structure, function, and molecular genetics of the γ/δ T cell receptor. *Annu. Rev. Immunol.* **7**:175–207.
- ROCK, E. P., P. R. SIBBALD, M. M. DAVIS, and Y.-H. CHIEN. 1994. CDR3 length in antigen-specific immune receptors. *J. Exp. Med.* **179**:323–328.
- SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- SCHATZ, D. G., M. A. OETTINGER, and M. S. SCHLISSEL. 1992. V(D)J recombination: molecular biology and regulation. *Annu. Rev. Immunol.* **10**:359–383.
- SCHILD, H., N. MACADDAT, C. LITZENBERGER, E. W. ERICH, M. M. DAVIS, J. A. BLUESTONE, L. MATIS, R. K. DRAPER, and Y. CHIEN. 1994. The nature of major histocompatibility complex recognition by $\gamma\delta$ T-cells. *Cell* **76**:29–37.
- SCHLUTER, S. F., R. M. BERNSTEIN, and J. J. MARCHALONIS. 1997. Molecular origins and evolution of immunoglobulin heavy-chain genes of jawed vertebrates. *Immunol. Today* **18**:543–549.
- SHEN, S. X., R. M. BERNSTEIN, S. F. SCHLUTER, and J. J. MARCHALONIS. 1996. Heavy-chain variable regions in carcharine sharks: development of a comprehensive model for the evolution of V_H domains among the gnathostomes. *Immunol. Cell Biol.* **74**:357–364.
- STEWART, J. 1992. Immunoglobulins did not arise in evolution to fight infection. *Immunol. Today* **13**:396–399.
- SWOFFORD, D. L. 1993. PAUP: phylogenetic analysis using parsimony. Version 3.1. Illinois Natural History Survey, Champaign.
- TAKIHARA, Y., J. REIMANN, E. MICHALOPOULOS, E. CICCONE, M. LORENZO, and T. W. MAK. 1989. Diversity and structure of human T cell receptor δ chain genes in peripheral blood γ/δ -bearing T lymphocytes. *J. Exp. Med.* **169**:393–405.
- THOMPSON, C. B. 1995. New insights into V(D)J recombination and its role in the evolution of the immune system. *Immunity* **3**:531–539.
- WILLIAMS, A. F. 1987. A year in the life of the immunoglobulin superfamily. *Immunol. Today* **8**:298–303.
- DAN GRAUR, reviewing editor
- Accepted September 28, 1999