A novel multigene family encodes diversified variable regions

Scott J. Strong*, M. Gail Mueller†, Ronda T. Litman*, Noel A. Hawke‡, Robert N. Haire*, Ann L. Miracle§, Jonathan P. Rast¶, Chris T. Amemiya‖, and Gary W. Litman*†‡§**

Departments of *Pediatrics, and ‡Medical Microbiology and Immunology, University of South Florida College of Medicine, St. Petersburg, FL 33701; †All Children's Hospital, and §Institute for Biomolecular Sciences, St. Petersburg, FL 33701; ¶Division of Biology, California Institute of Technology, Pasadena, CA 91125; and ‖Boston University School of Medicine, Center for Human Genetics, Boston, MA 02118

Edited by Martin G. Weigert, Princeton University, Princeton, NJ, and approved October 15, 1999 (received for review July 23, 1999)

Antigen recognition in the adaptive immune response by Ig and T-cell antigen receptors (TCRs) is effected through patterned differences in the peptide sequence in the V regions. V-region specificity forms through genetically programmed rearrangement of individual, diversified segmental elements in single somatic cells. Other Ig superfamily members, including natural killer receptors that mediate cell-surface recognition, do not undergo segmental reorganization, and contain type-2 C (C2) domains, which are structurally distinct from the C1 domains found in Ig and TCR. Immunoreceptor tyrosine-based inhibitory motifs that transduce negative regulatory signals through the cell membrane are found in certain natural killer and other cell surface inhibitory receptors, but not in Ig and TCR. In this study, we employ a genomic approach by using the pufferfish (*Spheroides nephelus*) to characterize a nonrearranging novel immune-type receptor gene family. Twenty-six different nonrearranging genes, which each encode highly diversified V as well as a V-like C2 extracellular domain, a transmembrane region, and in most instances, an immunoreceptor tyrosine-based inhibitory motif-containing cytoplasmic tail, are identified in an ≈113 kb P1 artificial chromosome insert. The presence in novel immune-type receptor genes of V regions that are related closely to those found in Ig and TCR as well as regulatory motifs that are characteristic of inhibitory receptors implies a heretofore unrecognized link between known receptors that mediate adaptive and innate immune functions.

V gene diversity | immunoreceptor tyrosine-based inhibitory motif | evolution | adaptive immunity | innate immunity

Ig and T-cell antigen receptor (TCR) genes are the primary mediators of highly specific adaptive immune responses. Recognition of antigens by these two structurally related but functionally distinct types of antigen-binding receptors is achieved through specific polypeptide folding patterns in N-terminal V regions that are created by both somatic rearrangement of individual segmental elements encoding V, diversity (D), and J regions as well as through nontemplated mechanisms that introduce additional sequence variation (1, 2). Patterns of shared sequence identity, organizational similarities, and a common rearrangement mechanism in Ig and TCR found in jawed vertebrate species are consistent with their origin from a common ancestral form in the distant evolutionary past and diversification in structure and organization throughout vertebrate phylogeny (3). Although V regions have undergone diversification during vertebrate evolution, comparisons of both Ig and TCR, as well as CD8β, a nonrearranging V region-containing gene expressed on the surface of T cells (4, 5), indicate general conservation of short-sequence motifs in the second and third framework regions (FR2 and FR3) of the V region (6). Sharing of such short-sequence regions forms the basis for a strategy that has been used to identify TCR gene homologs as well as Ig genes and an Ig-like gene in species representing diverse forms of jawed vertebrates (7–9). Because this technology depends on only the sharing of sufficient nucleotide identity between the short primer(s) and template as well as a defined length separating the priming sites, it also is capable of amplifying segments of related, non-Ig/TCR genes sharing these features (8). Although there is no functional predicate for this approach, here we ask a basic question as to whether or not any other immune-type receptors possess diversified families of V regions equivalent to those found in Ig and TCR.

To address this question, we used the pufferfish (*Spheroides nephelus*), which possesses a compact genome (≈1/8 that of man) and shares synteny with higher vertebrate genomes, as a model system. The smaller genome size facilitates characterization of extended genomic regions and the identification of regulatory features (10, 11). In addition, the lower proportion of noncoding regions in pufferfish minimize artifactual primings, which complicate the use of low-stringency PCR strategies in larger genomes. Employing the short primer PCR approach in conjunction with the pufferfish model, we report the presence of an extended, diversified multigene family of novel V region-containing genes that share properties of Ig and TCR as well as possess structural features that are common to mediators of innate immune function.

Materials and Methods

PCR Amplification of Novel Genes and Identification of Corresponding cDNAs. The methods for the isolation of DNA and mRNA from Southern pufferfish tissues, PCR amplification of cDNA and genomic DNA templates with short primers, as well as screening and isolation of cDNAs from a Southern pufferfish λgt11 spleen cDNA library by using short primer amplicons have been described (7–9). pGFPfs, an in-frame selection vector, was used to maximize selection of in-frame amplification products (12).

Construction and Screening of a Pufferfish P1 Artificial Chromosome (PAC). A genomic PAC library was constructed in pCYPAC6 (accession no. AF133437; C.T.A., unpublished work) and encompasses six to seven genome equivalents (nonrecombinant clones are <10%) (13). High-density nylon colony filter arrays were initially screened by using standard procedures and subsequently screened with probes complementing pufferfish Ig heavy chain (IgH; T. Ota and C.T.A., unpublished observation), Ig light chain (IgL) (N.A.H., unpublished observations), TCRα

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

Abbreviations: TCR, T-cell antigen receptor; CDR, complementarity determining region; FR, framework region; ITIM, immunoreceptor tyrosine-based inhibitory motif; NITR, novel immune-type receptor; PAC, P1 artificial chromosome; TM, transmembrane; C2, type-2 C domain; V/C2, V-like C2; Cyt, cytoplasmic region; FGXG, glycine bulge; GXG, glycine bulge; IgL, Ig light chain.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF110144, AF110145, and AF094698).

See commentary on page 14672.

**To whom reprint requests should be addressed. E-mail: litmang@allkids.org.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

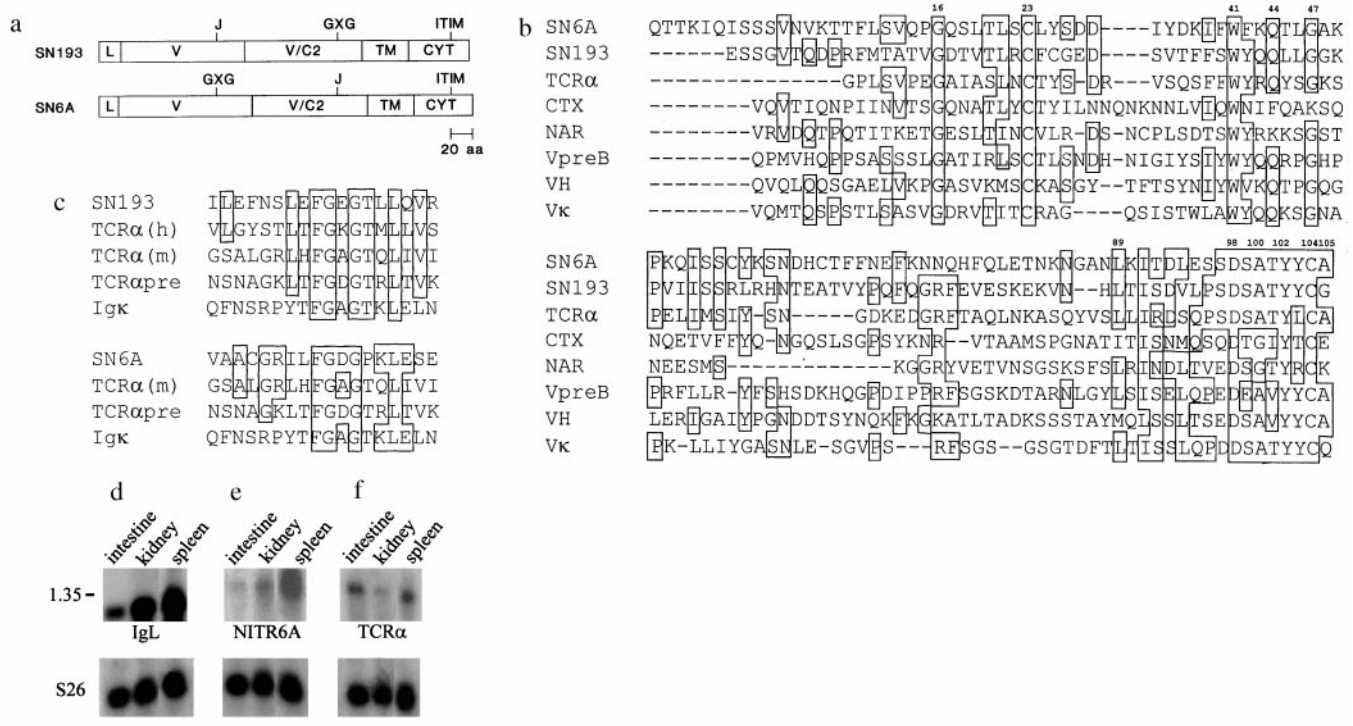

Fig. 1. (*a*) cDNA structures of SN193 and SN6A indicating functional boundaries. The relative scale is indicated. (*b*) Alignments of the putative V regions of NITR genes SN6A and SN193 with the V regions of: human TCRα (M13725), *Xenopus* CTX (XLU43330), nurse shark new antigen receptor (ECU18713), mouse VpreB (MMVPRE83), mouse V_H (AF045029), and human V_κ (PIR S46374). Residues shared by half or more of the sequences and/or by both NITR genes and one other sequence are enclosed. Alignments are based on CLUSTAL W (14) and visual optimization. Numbers over conserved amino acid residues correspond to International Immunogenetics Database (IMGT) designations (http:/imgt.cnusc.fr:8104/). Sequence identity relationships to nucleotide and protein databases were based on BLASTN and BLASTX analyses, respectively. (*c Upper*) Alignment of the putative J region-located C-terminal of the V Ig domain of SN193 with the J regions of: human TCRα (PIR S57872); mouse TCRα (M64239); rat TCRα precursor (PIR F27579); and rat Igκ (M62828). Mammalian J regions were selected on the basis of the highest alignment scores (BLASTX) with SN193. (*c Lower*) Alignment of the putative J region-located C-terminal to the V/C2 Ig domain of SN6A with the same three mammalian prototypes, selected in the same manner as above. Identical residues shared between the NITR gene segments and the mammalian prototypes are enclosed. (*d–f Upper*) Northern blot analyses of poly(A)$^+$ mRNA (1.5 μg per track) recovered from intestine, cranial kidney, and spleen hybridized with probes complementing *S. nephelus*: (*d*) IgL; (*e*) NITR cDNA SN6A V and V/C2 domains (NITR6A); and (*f*) TCRα. (*d–f Lower*) Following removal of bound probe, blots were rehybridized with a probe complementing *Spheroides* S26. Probes were labeled to equivalent specific activity. Exposure times were 7 days for IgL and TCRα, 18 days for NITR6A, and 2 days for S26. Reduced exposure and overexposure are entirely consistent with the patterns presented. An RNA size standard is indicated (1.35 kb). Equivalent expression patterns were obtained when referenced to the expression of β actin and when total RNA was used and referenced to 18S RNA.

(8), and TCRβ (N.A.H., unpublished observations) genes to infer linkage (or lack thereof) of those genes with PACs containing novel immune-type receptor (NITR) genes.

Subcloning from PACs and Determination of DNA Sequence. PACs were digested with various restriction enzymes and subcloned in pBluescript SK(+) (Stratagene). Exonuclease III/S1 nuclease deletion subcloning of fragments up to ≈12 kb in length employed the Double Stranded Nested Deletion Kit (Amersham Pharmacia). Automated DNA sequencing was performed with models 4000L and 4200-IR2 sequencers (Li-Cor, Lincoln, NE) with use of the ThermoSequenase (Amersham International) and SequiTherm Long Read (Epicentre Technologies, Madison, WI) cycle sequencing kits. Sequences were analyzed and assembled by using the INTELLIGENETICS SUITE DNA analysis software (IntelliGenetics). Sequence identity relationships to available nucleotide and protein databases were based on BLASTN and BLASTX analyses, respectively. Phylogenetic trees were constructed from CLUSTAL W (14) amino acid alignments with the METREE computer program (15).

RNA Blot Analysis. Poly(A)$^+$ mRNA was recovered by using oligodT magnetic beads, subjected to electrophoresis in 1% agarose containing 2.2 M formaldehyde, transferred to nitrocellulose and hybridized with radioactively labeled probes employing standard methods. Phosphorimage (STORM) scans were quantified with the IMAGEQUANT software package (Molecular Dynamics).

Results

Novel Immune-Type Receptor Genes Contain V and J Regions as Well as Immunoreceptor Tyrosine-Based Inhibitory Motifs (ITIMs). Two forms of short primer PCR technology, which are based on shared identity patterns in (*i*) FR2 and FR3 of Ig/TCR V regions and (*ii*) FR2 and the J regions, were used to amplify spleen mRNA (cDNA) and genomic DNA, respectively (7–9). cDNA priming yielded ≈200 bp fragments with characteristic features of V region segments that in turn hybridized to two near full-copy length spleen cDNAs, designated SN193 and SN6A. DNA priming yielded an ≈200-bp fragment that hybridized to a 5′ truncated cDNA designated SN5.1, that subsequently was characterized at the genomic level. All three genes encode a leader, V; V-like C2 (V/C2) (6); and transmembrane (TM) domains as well as cytoplasmic regions (Cyt) regions, which contain consensus ITIMs (Fig. 1*a*). J regions with glycine bulges ([FGXG] and [GXG]) are found in opposite order in the two cDNAs shown in Fig. 1*a*.

The V domains of cDNAs SN193 and SN6A and the genomic form of SN5.1 are ≈44–50% related at the nucleotide level and exhibit less than 70% identity at the predicted peptide level, the

operational definition of distinct families. The alignment affinities in BLAST searches of V regions of SN193, SN6A, and the other NITR genes, all of which are in the same transcriptional orientation, are typically highest for TCRα, Ig heavy chain, and Igκ light chain V regions; highest identities with TCRβ and Igλ are less frequent. Eight canonical positions: C^{23}, W^{41}, G^{47}, L^{89}, D^{98}, A^{100}, Y^{102}, and C^{104} (12), are highly (>95%) conserved in Ig heavy chain, Igκ and Igλ light chains, and TCRα and TCRβ. These positions also are conserved in SN193 and SN6A (Fig. 1*b*) and are found in >95% of the individual NITRs encoded in PAC 19B20 (see below) with the exception of G^{47}, which is present in ≈80% of the potentially functional NITR genomic sequences. G^{16}, Q^{44}, and A^{105}, which are shared by ≈80% of Ig and TCR genes, are highly conserved in NITRs.

NITRs also are related to Ig and TCR through significant sequence identity in J regions, which includes characteristic [GXG] glycine bulges (16) at two sites in the extracellular domains (Fig. 1*a*). In SN193, an [FGXG]-anchored J-like sequence is present at a position equivalent to that of J regions in Ig and TCR, between the V region (defined as the N-terminal through one amino acid C-terminal of C^{104}) and the C-terminal V/C2 domain; a second [GXG] motif is encoded between the V/C2 domain and TM region (see below). In SN6A, an [FGXG] region is located between the V/C2 domain and the TM region (Fig. 1*c*); a second [GXG] motif is located between the V sequence region and the V/C2 domain (Fig. 1*a*).

Relative Expression of NITRs in Lymphopoietic Tissues Differs from Ig and TCR. The relative expression of *Spheroides* IgL (Fig. 1*d*), NITR gene SN6A (Fig. 1*e*), and TCRα (Fig. 1*f*) are compared in intestine, cranial kidney [which is lymphohematopoietic in bony fish (17)], and spleen. TCRα was expressed predominantly and at equivalent levels in intestine and spleen whereas IgL was expressed equivalently in kidney and spleen. When normalized to S26 ribosomal RNA, the expression of NITR gene SN6A is distinguished from Ig and TCR by at least seven times higher expression in the spleen than in the kidney and was found at much lower levels in the intestine.

NITRs Represent a Diversified Multigene Family. To further characterize the NITR gene family, a pufferfish PAC library encompassing six to seven genomes and consisting of <10% nonrecombinant clones was screened with probes specific for cDNAs: SN193, SN6A, and SN5.1 and a PAC contig was defined (Fig. 2*a*). All three probes hybridized to PAC 19B20, which contained the longest insert, and its complete nucleotide sequence (113,426 bp) was determined. Twenty-six NITR genes, which are in the same transcriptional orientation, were identified and can be grouped into 13 different V and 5 different V/C2 families defined as exhibiting ≥70% predicted peptide sequence identity; two NITRs are pseudogenes. ITIMs are present in the 3′ cytoplasmic exons (Cyt2) of a large number of genes (Figs. 1*a* and 2*b*; see below). Variation occurs in intron lengths (Fig. 2*c*).

The leader peptides of NITR genes are encoded in exon I and vary in overall length from 16 to 20 amino acids, according to criteria set forth for TCRα (18). Exon II encodes both the V and V/C2 regions, which form the extracellular domains of the NITR genes (Figs. 2*c* and 3). This unusual presence of two Ig domains in a single exon in NITRs is likely a derived feature of the pufferfish. Ten of the V region domains are in the same family as SN6A and are designated as V type-1, two additional families of V domains each contain two members, and the remaining 10 V domain sequences are members of distinct V families (Figs. 2*b* and 3). BLASTP searches of V regions only yield convincing identities with V regions of Ig and TCR genes.

The C-terminal extracellular domain is difficult to classify. It possesses a V-like character but is missing a strand relative to typical V domains (a C2 character), thus the V/C2 designation.

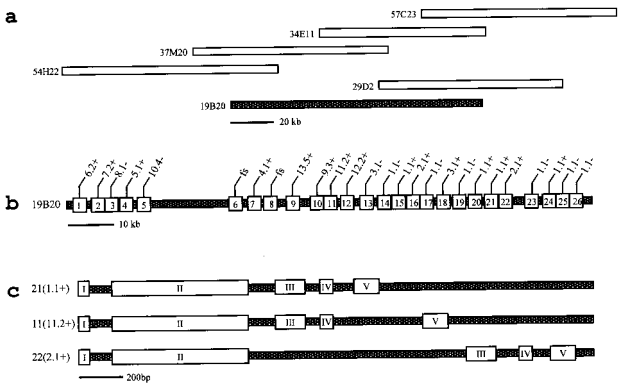

Fig. 2. (*a*) Contig of PACs containing NITR genes based on the overlap of end sequences and the complete nucleotide sequence of PAC 19B20. Clones are oriented 5′ to 3′ starting from PAC 54H22, relative to PAC 19B20. This orientation is based on partial sequences of other clones and is consistent with partial restriction analyses. (*b*) Organization of the NITR genes encoded in the PAC 19B20 insert. Each gene consists of five exons, encoding: (I) a (split) leader peptide; (II) putative extracellular V and V/C2 Ig domains; (III) transmembrane region; (IV) N-terminal cytoplasmic region; and (V) C-terminal cytoplasmic region. Genes are designated on the basis of the sequence families of the V (1–13) and V/C2 (1–5) Ig domains, respectively, as well as the presence (+) or absence (−) of an ITIM in exon V. For example, NITR 1 (6.2^{+}) contains a type-6 V domain and type-2 V/C2 domain, and possesses an ITIM in exon V. fs, frameshift. (*c*) Variation in the organization of exons and introns in three NITR genes; classification is the same as in *b*.

Although the I-set is a V/C2 intermediate, NITR V/C2 regions lack other residues that are characteristic of I-set molecules (19). Notably, when BLAST searches are performed on NITR V/C2 regions, certain NITR genes score significantly with the new antigen receptor, predominantly driven by C-terminal identities (20). However, as shown in Fig. 1*b*, the new antigen receptor V shares only limited identity with Ig/TCR/NITR V. Seventeen V/C2 sequences share >90% overall amino acid identity and are classified as type-1 (Figs. 2*b* and 3). The V/C2 type-2 family consists of four genes. Three other NITR genes (5, 9, and 10) cannot be grouped with each other or other V/C2 families and are treated as outliers. Absolute conservation of peptide structure (as defined in Fig. 3) occurs at only 19 positions in the V domain but occurs at 64 positions in the V/C2 domain.

The greater sequence diversification in V relative to V/C2 domains is supported by phylogenetic analyses in which markedly different branch lengths are evident between the corresponding V (Fig. 4*a*) and V/C2 (Fig. 4*b*) trees. Two major gene clusters are evident in these comparisons, excluding the outliers 5, 9, and 10. Specifically, two interrelated V-region subgroups are defined: IA corresponds to the SN6A-type genes, and IB is more divergent; genes 3 and 4 cannot be classified effectively. Although V and V/C2 divergence trees display similar topology, some cases of extensive V divergence and minimal V/C2 divergence are notable. Genes 21 and 22 are relatively divergent and belong to V subgroups IA and IB, respectively, whereas their V/C2 domains vary only by three nucleotides and are identical at the amino acid level. The most likely explanation for this observation is that domain shuffling has occurred within the extended NITR gene locus. This same phenomenon occurs with exons III, IV, and V, which segregate incongruently with respect to exon II (see below). These latter observations are particularly intriguing because V and V/C2 domains are encoded in a single exon.

In addition to the J-like sequences identified in SN193, SN6A, and SN5.1, BLAST searches of the other NITR genes identified a 14-amino acid string which is ≥64% identical at the predicted

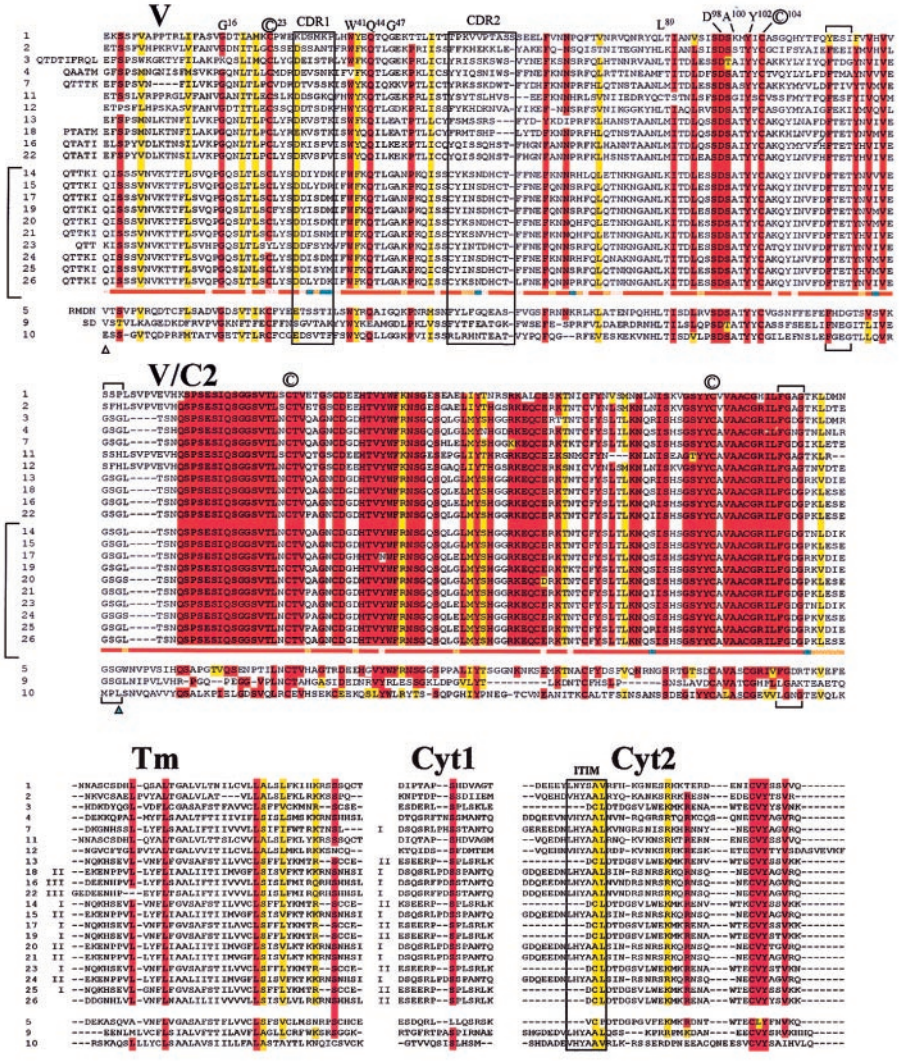

Fig. 3. The predicted peptide structures of the V, V/C2, TM, Cyt1, and Cyt2 of the 24 NITR genes encoded in PAC 19B20; pseudogenes 6 and 8 are not included. Genes are listed in an order that facilitates intergenic comparisons; vertical brackets in V and V/C2 enclose the 10 members of the SN6A type-1 gene family. △, Uncertainties with regard to the mature start site of NITR genes, referenced to gene 10; alternative start sites are predicted, as described (18). The basis for the designation of V and V/C2 domains in exon II is based largely on the homologous relationship of the NITR genes to Ig and TCR V domains (Fig. 1*b*), the shorter lengths of V/C2 domains and the relative locations of V/C2 exon/intron boundaries (▲) in two other species of bony fish, in which an intron separates the (two) extracellular Ig domains. The outlier genes (5, 9, and 10) are separated. Note the interspersed color bands and extensive regions of nonidentity in the three designated outliers, particularly in V/C2. For reference purposes, positions that are (highly) conserved between V_H, V_κ, V_λ, TCRVα, and TCRVβ are designated in one-letter code by using International Immunogenetics Database (IMGT) numbering (Fig. 1*b* legend); conserved cysteine is circled. The location (by reference to Ig/TCR designations) of sequence regions corresponding to CDR1 and CDR2 as well as the boundaries of the glycine bulge (J homology) regions, [(F)GXG], are shown by brackets above and below the alignments. Absolute identity is shown in red (up to one difference is allowed); substitutions that result in changes which retain functional groups, defined conservatively as: G or A; I, L, M, or V; K or R; S or T; F or Y; D or E; and N or Q, are shown in yellow. Recognized TM and Cyt1 families are designated by Roman numerals to the left of the alignments. ITIMs in Cyt2 are enclosed. Variation among the members of the SN6A type-1 gene family (enclosed in a vertical bracket) are shown in the continuous horizontal color bars below gene 26. Positions at which no variation occurs are shown in red; a single amino acid difference in one family member is shown in pink; one amino acid substitution in two or more family members is shown in orange; two or more different substitutions in multiple family members are shown in blue. The comparison is referenced to the predicted start site of gene 10.

peptide level with J regions of Ig and TCR, between the hyperconserved [YYC(A/V)] in V/C2 and the 3′ conserved gt dinucleotide, which serves as the exon II/III splice donor (Fig. 3). If conservative substitutions are considered, identities are as high as ≈91% for certain NITR J regions. Furthermore, six amino acids (18 nucleotides) separate the canonical J motif [FGXG] in V/C2 and the gt dinucleotide splice donor located immediately 3′ in all but gene 11. This spacing is a conserved structural feature of Ig/TCR J regions and is an inferred feature in the N-terminal J region of gene 10 (corresponding to SN193; Fig. 3), supporting evolutionary relationships between NITR, Ig, and TCR genes.

Sequence Variation in NITRs Is Regionalized. As indicated above, the sequence variation in the V domain greatly exceeds that in the V/C2 domain, a pattern that is reminiscent of the relationship between V and C1 domains in both Ig and TCR genes found throughout jawed vertebrate phylogeny (3). Furthermore, the patterns of sequence variation are reminiscent of complementarity determining regions (CDR1 and CDR2) in Ig and TCR. CDR3 contributions are somewhat more difficult to define because this region is formed through germ-line and somatic contributions in Ig and TCR, and in this sense are not analogous to NITR genes. Nevertheless, two types of CDR3 can be considered based on established CDR3 boundary definitions

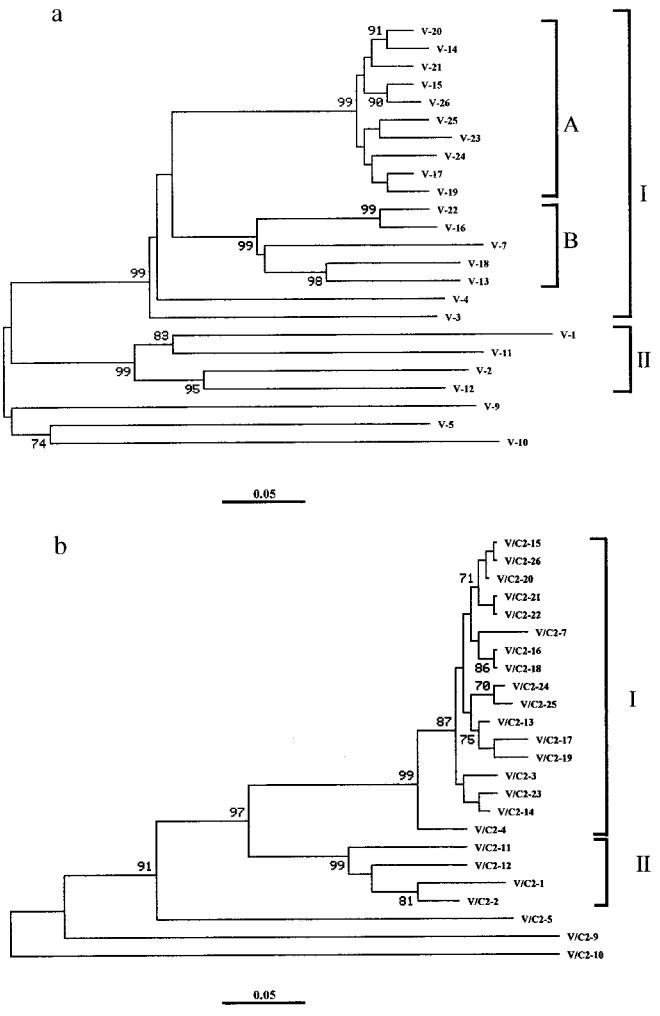

Fig. 4. Unrooted minimum phylogenetic trees comparing (*a*) V and (*b*) V/C2 domains of NITR genes. Trees are based on amino acid sequences by using the proportion of difference as an estimate of distance (15). The bar at the bottom indicates the number of amino acid changes per site. Positions containing gaps were excluded for pairwise distance estimates. Groups and subgroups are indicated by I or II and A or B, respectively. Numbers to the left of nodes indicate estimates of percentage probability that the length of the branch separating the adjacent cluster is greater than zero (i.e., that the cluster is statistically significant). Only values ≥70% are shown. Six statistically similar trees were obtained for the V domain and four for the V/C2 domain comparisons; however, the trees differed only in minor aspects of topology within the groups.

(21). The V region of gene 10 (corresponding to SN193) encodes 10 amino acids between the hyperconserved C^{104} (Fig. 3) and the J region [GXG], whereas all but four of the other NITRs (which lack [GXG] between the V sequence regions and V/C2 domains), encode 19 amino acids. The predicted CDR3 lengths for the two types are 6 and 15 amino acids, respectively, as defined for TCRs (21). Interfamily variation also occurs in other regions of the NITRs that would be expected to be exposed for binding, including the positions that are consistent with TCR hypervariable region 4 (HV4) (22).

Intrafamily variation is evident in comparisons of V domain sequences of the 10 type-1 NITR genes (Fig. 3). Specifically, one different amino acid is seen at 16 positions, two different amino acids are found at eight positions, and three different amino acids are found at six positions. In contrast, different amino acids occur at only 14 positions in the V/C2 domains, of which six are C-terminal to the highly conserved [FGXG] (Fig. 3). As in the interfamily comparisons, intrafamily variation in NITR genes is greater in the V than in the V/C2 domains and occurs in CDR1 and CDR2, consistent with the classical distinction of V and C regions.

The Characteristics of NITR TM and Cytoplasmic Regions Are Unique. Exon III consists of 43 amino acids and encodes the Ile-, Val-, Leu-rich TM region; three distinct exon III families can be recognized. The C-terminal 8–11 amino acids representing the N-terminal portions of the cytoplasmic tail typically contain from two to four K and/or R residues in proximity to S and/or T residues which appear to be capable of mediating protein kinase C phosphorylation and inositol triphosphate signaling (Fig. 3) (23). Several type-1 genes possess cysteine residues in the C-terminal portion of exon III that potentially could be involved in crosslinking/dimerization. CART motifs, which are characteristic of Ig and TCR TM regions, are not apparent (24), and it is notable that the TM regions are devoid of charged residues, which distinguish them from Ig and T-cell antigen receptors. Exon IV encodes the N-terminal 13 or 15 amino acids of the putative cytoplasmic tail (Cyt1) and like exon III, reveals distinctive families. Absolute association patterns of specific exon III and IV families and exon V (encoding Cyt2) types are not apparent.

Exon V encodes 30–44 amino acids corresponding to the C-terminal portion of the putative cytoplasmic region. A mammalian ITIM consensus (L/V)XYXX(L/V) (25–29) in four different variations is encoded at a single position in exon V in 15 of the PAC 19B20 NITR genes. In addition, there is near absolute conservation of a further downstream ITIM-related sequence (C)YVXXV in exon 5 (Fig. 3). The identification of ITIMs is of considerable significance because these motifs are not present in Ig and TCR, the only other highly diversified V-region-containing member of the Ig superfamily. However, ITIMs are found in a number of other molecules, including: Fc γ receptor IIB (Fcγ RIIB), Ig-like transcripts (ILTs), killer cell Ig-like receptors (KIRs), leukocyte Ig-like receptors (LIRs), paired Ig receptor B cell (PIRB), and signal regulatory protein α (SIRPα) (25, 26), that are known to transduce negative regulatory signals through a mechanism in which tyrosine phosphorylation of ITIMs results in activation of SHP1, SHP2, or SHIP tyrosine phosphatases that can extinguish cell activation by dephosphorylation of receptors (26).

Discussion

Despite its enormous value as a model system for genomics investigations, pufferfish is of limited value for immunological functional studies. The higher relative level of expression of NITR genes in spleen distinguishes these genes from Ig and TCR. Spleen cDNA transcripts (>96% identity) representing four of the 24 potentially functional NITR genes have been isolated (data not shown); seven other NITR transcripts cannot be attributed directly to a gene encoded in 19B20 and presumably are encoded elsewhere in the PAC contig. Screening of the PAC filter arrays with NITR, IgH, IgL, TCRα, and TCRβ revealed no overlaps, consistent with a lack of physical linkage between these loci.

On the basis of the distinctive variation patterns in V, the presence of J, the lack of segmental rearrangement, the presence of a V/C2 domain, and the distribution of ITIMs, NITR genes are unique among all Ig superfamily genes that have been shown to be or possibly are involved in specific immune recognition (26–29). NITR V regions are diversified in a manner resembling that associated with Ig and TCR that mediate adaptive immunity; however, the absence of rearrangement/somatic diversification mechanisms in NITRs categorizes these molecules with innate receptors (30, 31) such as are involved in certain natural

killer recognition processes, although in mammals this function is not mediated through diversified V regions (26). Alternatively, the V diversification patterns of the NITR genes could effect innate recognition capacity for pathogen(s) in some as of yet unspecified cell type. These issues can be addressed in part by identifying and characterizing homologs of NITR genes in higher vertebrates as well as homologs of natural killer-like receptors in lower vertebrates.

It is reasonable to consider NITR genes in the context of ancestral immune recognition genes; such genes possessing VJ continuity in the germ line could have been targets of the transposition event that presumably gave rise to the rearranging form of antigen-binding receptor genes (32, 33). More specifically, if ancestral genes were to function in diverse immune-type innate recognition, they would need to be present in multiple diversified copies, require intracellular signaling capabilities, and if functional at the cell surface, would require some form of association with the surface membrane. NITR genes possess all of these characteristics, and it is likely that their further study will provide insight into the relationships between adaptive and innate immune function and the mechanisms whereby recognition processes became structurally diversified and functionally specialized during evolution.

We thank Barbara Pryor for editorial assistance and Drs. David Schatz, Jeffrey Yoder, Louis Du Pasquier, Jim Kaufman, and Donald Wiley for valuable comments on this work. We also thank Dr. Eric Davidson for providing access to the Q-Bot robotics workstation. This work was supported by grants to G.W.L. from the National Institutes of Health, the Valerie Fund, and the Pediatric Cancer Foundation, and grants to C.T.A. from the National Institutes of Health and the National Science Foundation.

1. Tonegawa, S. (1983) *Nature (London)* **302,** 575–581.
2. Blackwell, T. K. & Alt, F. W. (1988) in *Ig Genes*, eds. Hames, B. A. & Glover, D. M. (IRL, Oxford), Vol. 1, pp. 1–60.
3. Litman, G. W., Anderson, M. K. & Rast, J. P. (1999) *Annu. Rev. Immunol.* **17,** 109–147.
4. Parnes, J. R. (1989) *Adv. Immunol.* **44,** 265–311.
5. Norment, A. M., Salter, R. D., Parham, P., Engelhard, V. H. & Littman, D. R. (1988) *Nature (London)* **336,** 79–81.
6. Barclay, A. N., Brown, M. H., Law, S. K. A., McKnight, A. J., Tomlinson, M. G. & van der Merwe, P. A. (1997) *The Leucocyte Antigen FactsBook* (Academic, London), pp. 1–613.
7. Rast, J. P. & Litman, G. W. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 9248–9252.
8. Rast, J. P., Haire, R. N., Litman, R. T., Pross, S. & Litman, G. W. (1995) *Immunogenetics* **42,** 204–212.
9. Rast, J. P., Anderson, M. K., Strong, S. J., Luer, C., Litman, R. T. & Litman, G. W. (1997) *Immunity* **6,** 1–11.
10. Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B. & Aparicio, S. (1993) *Nature (London)* **366,** 265–268.
11. Elgar, G. (1996) *Hum. Mol. Genet.* **5,** 1437–1442.
12. Hawke, N. A., Strong, S. J., Haire, R. N. & Litman, G. W. (1997) *BioTechniques* **23,** 619–621.
13. Amemiya, C. T., Ota, T. & Litman, G. W. (1996) in *Construction of P1 Artificial Chromosome (PAC) Libraries from Lower Vertebrates,* eds. Lai, E. & Birren, B. (Academic, New York), pp. 223–256.
14. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680.
15. Rzhetsky, A. & Nei, M. (1992) *Mol. Biol. Evol.* **9,** 945–967.
16. Chothia, C., Novotny, J., Bruccoleri, R. & Karplus, M. (1995) *J. Mol. Biol.* **186,** 651–663.
17. Zapata, A. G. & Cooper, E. L. (1990) *The Immune System: Comparative Histophysiology* (Wiley, Chichester, U.K.).
18. Wülfing, C. & Plückthun, A. (1995) *Immunol. Today* **16,** 405–406.
19. Harpaz, Y. & Chothia, C. (1994) *J. Mol. Biol.* **238,** 528–539.
20. Greenberg, A. S., Avila, D., Hughes, M., Hughes, A., McKinney, E. C. & Flajnik, M. F. (1995) *Nature (London)* **374,** 168–173.
21. Rock, E. P., Sibbald, P. R., Davis, M. M. & Chien, Y.-H. (1994) *J. Exp. Med.* **179,** 323–328.
22. Wuilmart, C. & Urbain, J. (1991) *Mol. Immunol.* **28,** 931–941.
23. Newton, A. C. (1997) *Curr. Opin. Cell Biol.* **9,** 161–167.
24. Campbell, K. S., Backstrom, B. T., Tiefenthaler, G. & Palmer, E. (1994) *Semin. Immunol.* **6,** 393–410.
25. Vély, F. & Vivier, E. (1997) *J. Immunol.* **159,** 2075–2077.
26. Lanier, L. L. (1998) *Annu. Rev. Immunol.* **16,** 359–393.
27. Kubagawa, H., Burrows, P. D. & Cooper, M. D. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 5261–5266.
28. Borges, L., Hsu, M.-L., Fanger, N., Kubin, M. & Cosman, D. (1997) *J. Immunol.* **159,** 5192–5196.
29. Torkar, M., Norgate, Z., Colonna, M., Trowsdale, J. & Wilson, M. J. (1998) *Eur. J. Immunol.* **28,** 3959–3967.
30. Medzhitov, R. & Janeway, C. A. (1997) *Curr. Opin. Immunol.* **9,** 4–9.
31. Hoffmann, J. A., Kaftos, F. C., Janeway, C. A. & Ezekowitz, R. A. B. (1999) *Science* **284,** 1313–1318.
32. Agrawal, A., Eastman, Q. M. & Schatz, D. G. (1998) *Nature (London)* **394,** 744–751.
33. Hiom, K., Melek, M. & Gellert, M. (1998) *Cell* **94,** 463–470.