# Generation of Immunoglobulin Light Chain Gene Diversity in *Raja erinacea* Is Not Associated with Somatic Rearrangement, an Exception to a Central Paradigm of B Cell Immunity

By Michele K. Anderson,* Michael J. Shamblott,§ Ronda T. Litman,‡ and Gary W. Litman*‡

*From the Departments of *Biochemistry and ‡Pediatrics, University of South Florida, All Children's Hospital, St. Petersburg, Florida 33701; and the §Center of Marine Biotechnology, University of Maryland, Baltimore, Maryland 21202*

## Summary

In all vertebrate species examined to date, rearrangement and somatic modification of gene segmental elements that encode portions of the antigen-combining sites of immunoglobulins are integral components of the generation of antibody diversity. In the phylogenetically primitive cartilaginous fishes, gene segments encoding immunoglobulin heavy and light chain loci are arranged in multiple clusters, in which segmental elements are separated by only 300–400 bp. In some cases, segmental elements are joined in the germline of nonlymphoid cells (joined genes). Both genomic library screening and direct amplification of genomic DNA have been used to characterize at least 89 different type I light chain gene clusters in the skate, *Raja*. Analyses of predicted nucleotide sequences and predicted peptide structures are consistent with the distribution of genes into different sequence groups. Predicted amino acid sequence differences are preferentially distributed in complementarity-determining versus framework regions, and replacement-type substitutions exceed neutral substitutions. When specific germline sequences are related to the sequences of individual cDNAs, it is apparent that the joined genes are expressed and are potentially somatically mutated. No evidence was found for the presence of any type I light chain gene in *Raja* that is not germline joined. The type I light chain gene clusters in *Raja* appear to represent a novel gene system in which combinatorial and junctional diversity are absent.

A central paradigm of contemporary molecular immunology is that the generation of antibody and T cell antigen receptor diversity is associated with the rearrangement of segmental elements in individual somatic cells committed to the B or T lineages (1, 2). In mammals, this rearrangement process is associated with combinatorial as well as junctional diversity, which is achieved by deletions and nontemplated additions at the segmental junctions. These somatic changes, along with hypermutation of rearranged variable regions, result in the diversification of the antibody repertoire (3). The relative contributions of these mechanisms to the final immunoglobulin specificity vary among species, as does the total number of individual gene loci. The primary avian model, the chicken, possesses one immunoglobulin heavy chain locus (4) and one immunoglobulin light chain locus (5), whereas cartilaginous fish, such as sharks, have >100 chromosomally dispersed immunoglobulin heavy chain loci and possibly an equal number of immunoglobulin light chain loci (6, 7). Previously, our laboratory demonstrated that the horned shark, *Heterodontus francisci* (6), and the little skate, *Raja erinacea* (8), possess two major types of immunoglobulin heavy chain gene organization in which the segmental elements ($V_H$, $D_H$, and $J_H$) are either unjoined or joined in the germline of nonlymphoid cells. On the basis of identities between the predicted coding and flanking region sequences of unjoined and joined genes, it appears that none of the joined heavy chain genes identified to date are pseudogenes. However, the transcription of these genes has not been demonstrated, and their role in adaptive immunity is not understood (9).

Sequence analyses have revealed that there are at least three types of light chain clusters in the cartilaginous fish, some of which are also $V_L$–$J_L$ joined in the germline. Type I light chain genes, which were originally isolated using an antibody screening approach (10), are unjoined in *Heterodontus* (7); however, several joined light chain genes have been identified in *Raja* (9). Type II light chain genes, which are related to type I light chain genes by ~38% amino acid identity, have been characterized in *Heterodontus*, *Hydrolagus colliei* (spotted ratfish), *Raja* (11), and *Carcharhinus plumbeus* (sandbar shark) (12), and appear to be joined in each species. Type III light chain genes, which are ~60% related in amino acid sequence to mammalian κ light chain genes in the V

 J. Exp. Med. © The Rockefeller University Press • 0022-1007/95/07/0109/11 $2.00
Volume 182 July 1995 109–119

region, have been characterized at the genomic level in *Heterodontus*, in which they are unjoined (11), as well as at the cDNA level in *Ginglymostoma cirratum* (nurse shark) (13). However, we have not been able to detect a type III homologue in either *Raja* or *Hydrolagus* despite extensive attempts using a variety of strategies, including PCR amplification and direct probing (Anderson, M. K., and J. P. Rast, unpublished observations). To establish that germline joining is of functional significance, the distributions of nucleotide sequence differences within a significant portion of the gene family must be characterized and the transcriptional status of the gene must be established. Using several genomic selection strategies combined with analyses of cDNAs, we provide evidence that all type I light chain genes in *R. erinacea* are joined in the germline, diversified, and transcribed, and appear to be somatically mutated.

## Materials and Methods

*Animals.* Adult specimens of *R. erinacea* were obtained from the Marine Biological Laboratory (Woods Hole, MA). After the animals were sacrificed, tissues were processed immediately.

*DNA Libraries.* A genomic library was constructed from *R. erinacea* nucleated RBCs. RBCs were carefully separated from other cells in whole blood by low speed centrifugation. Cytological staining of such preparations shows these to be devoid of leukocytes. High molecular weight DNA was extracted as previously described (14), treated with RNase, partially digested with Sau3A, and ligated to Lambda DASH® (Stratagene, La Jolla, CA) arms. The unamplified library consisted of ~7.5 × $10^6$ PFU, which corresponds to ~2.7 genomic equivalents, assuming 7 pg per haploid genome (15). The library was amplified on the bacterial host P2392, a P2 lysogen of LE392.

A cDNA library was constructed essentially as previously described (16), using *R. erinacea* spleen RNA from the same animal that was used in genomic library construction. RNA was extracted using a commercially available method (RNAzol; Cinna/Biotecx Laboratories, Houston, TX), and mRNA was purified using Dynabeads (Dynal, Inc., Oslo, Norway). The mRNA was converted to cDNA using a commercially available cDNA synthesis method (Pharmacia LKB Biotechnology Inc., Piscataway, NJ). The cDNA was ligated into λgt11 arms, packaged using Gigapack® Gold (Stratagene) packaging mixes, and plated on *Escherichia coli* Y1088. The library was amplified to 4 × $10^9$ PFU/ml.

*Probes and PCR Primers.* DNA probes were generated by PCR using primers directed against specific nucleotide sequences of corresponding cDNA or genomic clones. HFL is a 681-bp *H. francisci* type I light chain probe (10) that complements the variable ($V_L$) and constant ($C_L$) regions and cross-hybridizes with *Raja* type I light chain genes. $HFV_L$ (288 nucleotides) and $HFC_L$ (303 nucleotides) complement the *Heterodontus* type I $V_L$ and $C_L$ regions, respectively. $REV_LC_L$ (642 nucleotides) complements the *Raja* type I light chain, and $REV_L$ (260 nucleotides) complements the $V_L$ region. PCR-generated probes were labeled by a modification (8) of the random hexanucleotide priming method (17) to a specific activity of $10^8$–$10^9$ cpm/μg.

The primer pair SKVN-X (CCC**TCTAGA**ATTCCAGTCCTGAATCAAA; XbaI linkered) and SKCC-S (CCC**GAGCTC**GTTCGCGCAAGATGATGAG; SstI linkered) was used to amplify by PCR a 1.6-kb fragment, which included the 972-bp J–C intervening sequence (IVS)[1] of genomic light chain clone λsk102 (the restriction site is indicated in boldface in these and subsequent primer structures). The 3′ cDNA primer SKJC (AC**AAGCTT**GGAATGAGAACGGCTGC; HindIII linkered) was used with the SKVN-X primer to generate $V_L$–$J_L$ region products from genomic isolates. The primer pair RELCFR1 (**AAGCTT**CGTCTCCGCGGCACAGA; HindIII linkered) and RELCFR3 (**TCTAGA**CACGGTGCCGATGGTCA; XbaI linkered) was used to generate the GPLC (genomic PCR-derived light chain; framework region 1 [FR1]–FR3 amplified from genomic DNA) clones (see Fig. 1). The primer pair RELCFR1 and SKLCJ (**GAGCTC**TTCAGGATGGTCCCAGG; SstI linkered) was used to generate the FJ (FR1–$J_L$ amplified from genomic DNA) series, and the primer pair RELCFR2 (**AAGCTT**TATCGACAGCGTCCCGG; HindIII linkered) and SKLCJ was used to generate the FR2–$J_L$ series (FR2–$J_L$ amplified from genomic DNA). The rationale for the use of these various primer pairs is described below.

*DNA and cDNA Library Screening.* To identify several representative type I *Raja* light chain clones, a portion of the *Raja* genomic library (~280,000 PFU) was plated, lifted onto nitrocellulose, hybridized with the radiolabeled probe HFL (*Heterodontus* type I) in SET (0.6 M NaCl, 0.2 M Tris, 0.02 M EDTA, 0.5% SDS, 0.1% sodium pyrophosphate) at 65°C for 16–24 h, and washed in 1× SSC (0.15 M NaCl, 0.015 M sodium citrate), 0.1% SDS, 0.05% sodium pyrophosphate at 52°C. The cDNA library (~160,000 PFU) was screened with HFL and $REV_LC_L$ as previously described.

*Genomic PCR.* Genomic DNA was used as a template to potentially amplify portions of the $V_L$–$J_L$ regions of all type I light chains in the germline using the primer pairs previously described. 1 μl of genomic DNA (3 μg/μl) from the same animal used in cDNA and genomic library construction was mixed gently with 9 μl of $H_2O$, heated to 95°C for 7 min, and held at 80°C. PCR (10×) buffer (500 mM KCl, 100 mM Tris-Cl, pH 8.3, 15 mM $MgCl_2$, 0.1% gelatin), dNTPs, primers, and 2.5 U of Taq polymerase (0.5 μl of 5 U/μl; Applied Biosystems, Inc., Foster City, CA) were added, followed by 35 cycles at 94°C, 1 min; 55°C, 2 min; 72°C, 4 min. The same conditions were used in parallel analyses in which the template was omitted. This served as a control to rule out potential reagent-based contamination. PCR products were extracted with phenol–chloroform, precipitated with ethanol, digested with appropriate restriction enzymes, ligated into M13 mp18, plated, and selected by absence of LacZ expression and by dot blot hybridization under relaxed stringency to the $REV_L$ probe. To obtain the opposite direction strand for sequencing, PCRs were performed using M13-specific primers. The resulting products were digested with appropriate restriction endonucleases and subcloned into M13 mp19.

*DNA Sequencing and Analysis.* DNA sequences were determined by the dideoxynucleotide chain termination method (18) using [$^{35}$S]dATP and Sequenase® (United States Biochemical Corp., Cleveland, OH). Oligodeoxynucleotide 18–20-mer primers were used to extend sequences. Analyses of sequences were performed using commercially available software (IntelliGenetics Suite™; IntelliGenetics, Palo Alto, CA), including GEL for managing the sequencing projects and IFIND and GENALIGN for sequence comparisons. All sequences contained herein have been submitted to

[1] *Abbreviations used in this paper:* CDR, complementarity-determining region; FR, framework region; IVS, intervening sequence; RAG, recombination-activating gene.

GenBank and assigned accession numbers U19001–U19025 for the FJ series, U19045–U19097 and U19203 for the GPLC series, U19187–U19202 for the 27***/28*** series, U19204–19208 for the RRLC series, and U19209 for SK102.

The neighbor-joining tree was constructed using the MEGA (Molecular Evolutionary Genetic Analysis) program kindly provided by Tatsuya Ota and Masatoshi Nei (Pennsylvania State University, University Park, PA; 19). The evolutionary distances between gene pairs over the region examined (codons 19–72, according to Kabat et al. [20]) were calculated as the proportion of nucleotide sites at which the two sequences differed (*p* distances). The *p* distances were used to construct the tree using the neighbor-joining method (21). Four *Heterodontus* light chain sequences described previously (7, 10) were used as an outgroup to root the tree.

## Results

*Complete Nucleotide Sequence of a Raja Type I Light Chain Gene Cluster.* Initially, a *Raja* genomic library was screened with the heterologous (*Heterodontus*) type I light chain probe HFL and seven clones that were shown by partial restriction mapping to be unique (λsk102, λ27101, λ27102, λ27103, λ27104, λ27105, and λ27106) were selected. A 2.0-kb EcoRI–PstI $HFV_L^+$ fragment and a 0.6-kb EcoRI $HFC_L^+$ fragment from one of the isolates (λsk102) were subcloned and sequenced. In addition, PCR was performed using λsk102 DNA as a template and the SKVN-X/SKCC-S primer pair (Fig. 1). A 1.6-kb fragment that overlapped the other fragments was identified and subcloned into M13, and the DNA sequence was determined.

The complete genomic sequences of the $V_L$ and $C_L$ segments of λsk102 are shown in Fig. 2. A typical split leader is noted, with the last three codons contiguous with FR1. The $V_L$, $J_L$, and $C_L$ segments of λsk102 are in close linkage, exhibiting the cluster-type pattern of gene organization described by us earlier (22). The *Raja* type I light chain genomic locus as represented by λsk102 is $V_L$–$J_L$ joined in the germline, unlike the *Heterodontus* type I light chain gene loci, which contain an ~350-bp $V_L$–$J_L$ IVS (7). Furthermore, λsk102 does not contain the typical light chain regulatory octamer (ATTTGCAT) present 134 bp 5′ of the putative start codon in the *Heterodontus* type I light chain gene locus (7). An octamer-like sequence (ATTTGTAT; 7/8 match with consensus) is located 297 bp upstream of the putative start codon. The consensus enhancer elements (TCATGTG and CAGATG), which are found within the *Heterodontus* $J_L$–$C_L$ IVS (7), are not present in the $J_L$–$C_L$ IVS of λsk102. However, an octamer-like sequence (ATTTGCAG; 7/8 match with consensus) and its reverse complement (TTGCAAAT; 7/8 match with consensus) are located in the $J_L$–$C_L$ IVS, in positions similar to those of the putative enhancer elements found in *Heterodontus*. The functional significance of these motifs is uncertain.

*Raja Type I Light Chain Gene Clusters Are $V_L$–$J_L$ Joined in the Germline.* Several different approaches were used to determine the nature, extent, and degree of sequence diversity of the type I light chain gene family. PCR analyses were performed on the other six genomic light chain clones that were isolated initially (λ27101, λ27102, λ27103, λ27104, λ27105, and λ27106) using the primer pair SKVN-X/SKCC-S (Fig. 1) to assess whether they are germline joined. An ~1.6-kb product was detected in each case, consistent with the $V_L$–$C_L$ linkage distance determined by direct cloning and sequencing for λsk102 and thereby with germline joining of $V_L$ and $J_L$. An additional 32 clones were identified and isolated from a *Raja* genomic library based on hybridization to HFL under low stringency conditions (14). Similar PCR analyses of each using the SKVN-X/SKCC-S primer pair yielded only ~1.6-kb products. An ~400-bp product was amplified from each of the second group of 32 clones (the SKLCGEN series) using the SKVN-X and SKJC primers. Sequence analyses of the products con- firmed that each of these genes is $V_L$–$J_L$ joined and that 17 of the isolates are unique (i.e., they can be distinguished by at least a one nucleotide sequence difference), as shown in Fig. 3. All of the sequences are 413 bp in length, with the exception of λ28402, λ28423, λ28417,

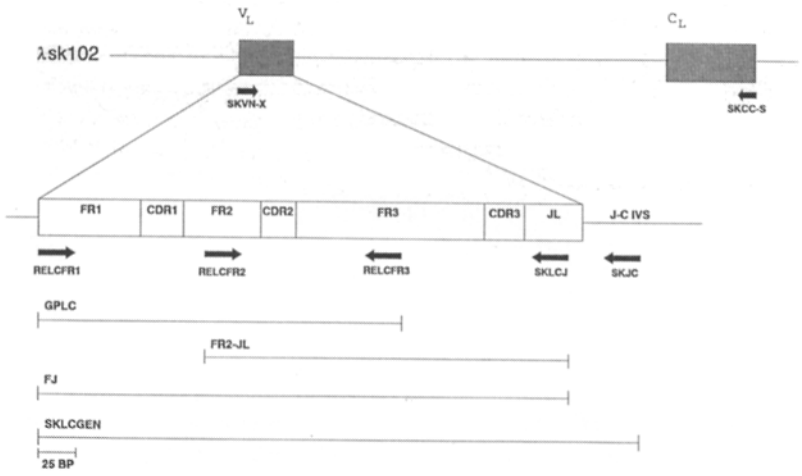


**Figure 1.** Approximate locations of PCR priming sites used to obtain the full nucleotide sequence of λsk102 and to generate the SKLCGEN, GPLC, FR2–$J_L$, and FJ series clones. Schematic representation of PCR primers and products in relation to the genomic type I light chain locus of *Raja*. The top portion of the figure represents an entire genomic locus. SKVN-X complements the $V_L$ FR1 region, and SKCC-S complements the end of the constant region, based on the nucleotide sequence of clone λsk102. The $V_L$ region is expanded in the lower portion of the figure to depict the regions complemented by the primers RELCFR1, RELCFR2, RELCFR3, SKLCJ, and SKJC. The GPLC series clones (235 bp) are PCR products from genomic DNA primed with RELCFR1 and RELCFR3. The FR2–$J_L$ series clones (242 bp) are PCR products from genomic DNA primed with RELCFR2 and SKLCJ. The FJ series clones (318 bp) are PCR products from genomic DNA primed with RELCFR1 and SKLCJ. The SKLCGEN series clones (413 bp) are PCR products from genomic library isolates primed with SKVN-X and SKJC.

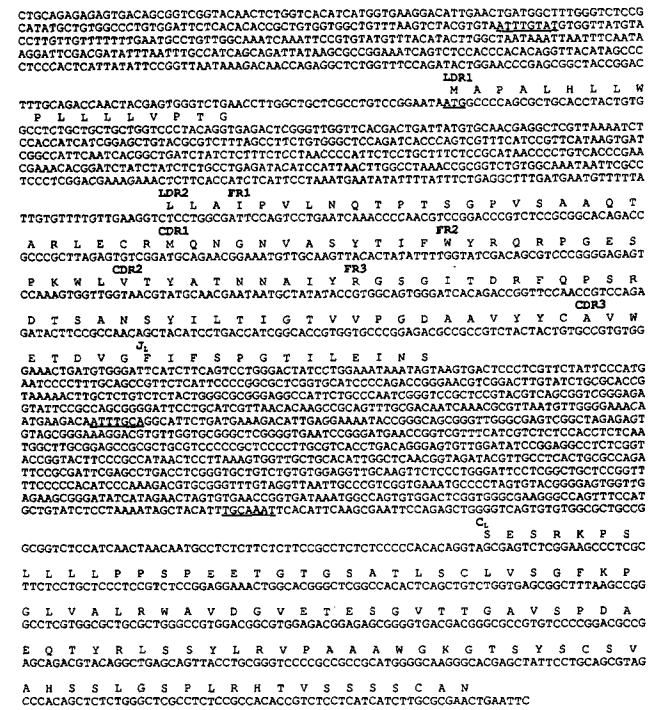

**Figure 2.** Nucleotide sequence of λSK102. The predicted amino acid sequence is shown above the second nucleotide of each codon in the coding regions. The start codon (ATG), octamer-like sequence, and enhancer-like sequences are underlined. The leader (*LDR*) regions are separated by intervening DNA sequence. In the variable region, FRs and CDRs are designated at the beginning of each region. The stop codon at the 3′ end of the constant ($C_L$) region is denoted by a period.

and λ28406, which possess one fewer codon in the $J_L$ segment. In addition, the λ28402 gene is truncated in the $J_L$ region. Five of these genes (λ28425, λ28403, λ27104, λ28401, and λ28409) appear to be pseudogenes, owing to the presence of two shared mutations–one in FR2 and one in FR3, that result in stop codons. A sixth putative pseudogene, λ28406, has a stop codon at the 3′ end of the $J_L$ segment.

To identify additional type I light chain genes that possibly would not be incorporated into the library and to assess whether or not these genes are germline joined, the 17 unique $V_L$–$J_L$ gene sequences recovered from the *Raja* λ genomic library were compared and the regions exhibiting the highest degree of nucleotide identity were identified. PCR primers complementing these regions (Fig. 3) were used to amplify products from *Raja* genomic DNA (Fig. 1). The primers RELCFR1 and SKLCJ were used to amplify the regions between and partially including FR1 and $J_L$ (the FJ series), which were separated by subcloning and sequenced. Each of the 29 FJ sequences obtained is 318 bp long, and 25 of these are distinct within this series. The sequences indicate that all 25 genes are $V_L$–$J_L$ joined. In the region covered by these primers, there are a total of 34 unique sequences among the 25 FJ (25 of 34) and the 17 SKLCGEN (9 of 34) sequences. The 8 additional SKLCGEN sequences differ from the 34 other sequences outside the region covered by these primers (see Fig. 3), except for λ27101, which matches FJ36 in this region. The predicted amino acid sequences of 23 of the 25 FJ sequences are unique. An alignment of the FR1–$J_L$ region sequences of the 34 unique sequences between the RELCFR1 and SKLCJ primers is shown in Fig. 4. The majority of these substitutions are nonsynonymous. There are 303 substitutions (as compared with the consensus sequence)/1,648 nucleotides (18%) in the complementarity-determining regions (CDRs) and 96 substitutions/5,508 nucleotides (1.7%) in the FRs, with the highest occurrence of substitutions in CDR3 (13.6%). Furthermore, 94% of the substitutions in the CDRs are replacement substitutions versus 75% replacement substitutions in the FRs. A distribution of the variability of amino acid residues at each position, shown in Fig. 5, illustrates the highest incidence of amino acid substitutions in CDR2, followed by CDR3. Multiple substituted positions are evident in the FRs as well, but many of these fall below the isolated substitution line, indicating that only one of the genes is substituted at this position. It should also be noted that certain substitutions occur as "blocks" that are shared between multiple sequences. However, the distribution of these blocks is not consistent with regard to the specific regions shared by the different genes, as might be expected among genes that diverged from a common ancestral gene. For example, FJ3 shares the block (as seen in Fig. 4 compared with consensus) "C-CG-" (CACGA) at the 3′ end of CDR1 with FJ4, FJ16, FJ9, FJ25, FJ20, FJ26, and FJ15. However, at the 3′ end of CDR3, FJ3 shares the block "C-G-AC" (CGGTAC) with FJ28, FJ29, FJ23, and FJ2, which share a different block in CDR1. Other examples of differential block distributions are also evident in this alignment.

Previous studies have shown that PCR has a strong bias toward amplification of smaller products; i.e., germline-joined templates are amplified preferentially over unjoined templates even when the overall length of the product varies by <350 bp (reference 23 Anderson, M. K., and G. W. Litman, unpublished observations). In an attempt to circumvent or at least minimize PCR length bias and possibly detect unjoined type I light chain genes, the primers RELCFR2 and SKLCJ were used to amplify the CDR3–$J_L$ junctions and the surrounding regions from genomic DNA. If a typical (*Heterodontus*-like) type I light chain gene IVS had been present, the product would have been ~600 bp; however, a single 242-bp fragment was recovered, consistent with uniform germline joining, which subsequently was established by the subcloning and sequencing of four different clones. Notably, each of these was identical in overlapping regions to type I genes selected by different methods, i.e., PCR priming of genomic DNA and λ clones (data not shown).

Genomic DNA was also subjected to PCR amplification using the primers RELCFR1 and RELCFR3 (Fig. 1). This strategy was directed at examining the diversity of germline type I $V_L$ genes, irrespective of joining status. The high degree of nucleotide identity in the sequences chosen for primer design (Fig. 3) between independently isolated genomic library clones (the initial selection via library screening was based on as little as 60% nucleotide identity) provided a strong basis for potentially priming all genomic type I light chain loci. DNA sequence analyses of the 235-bp PCR product(s) amplified from genomic DNA (the GPLC series) using these

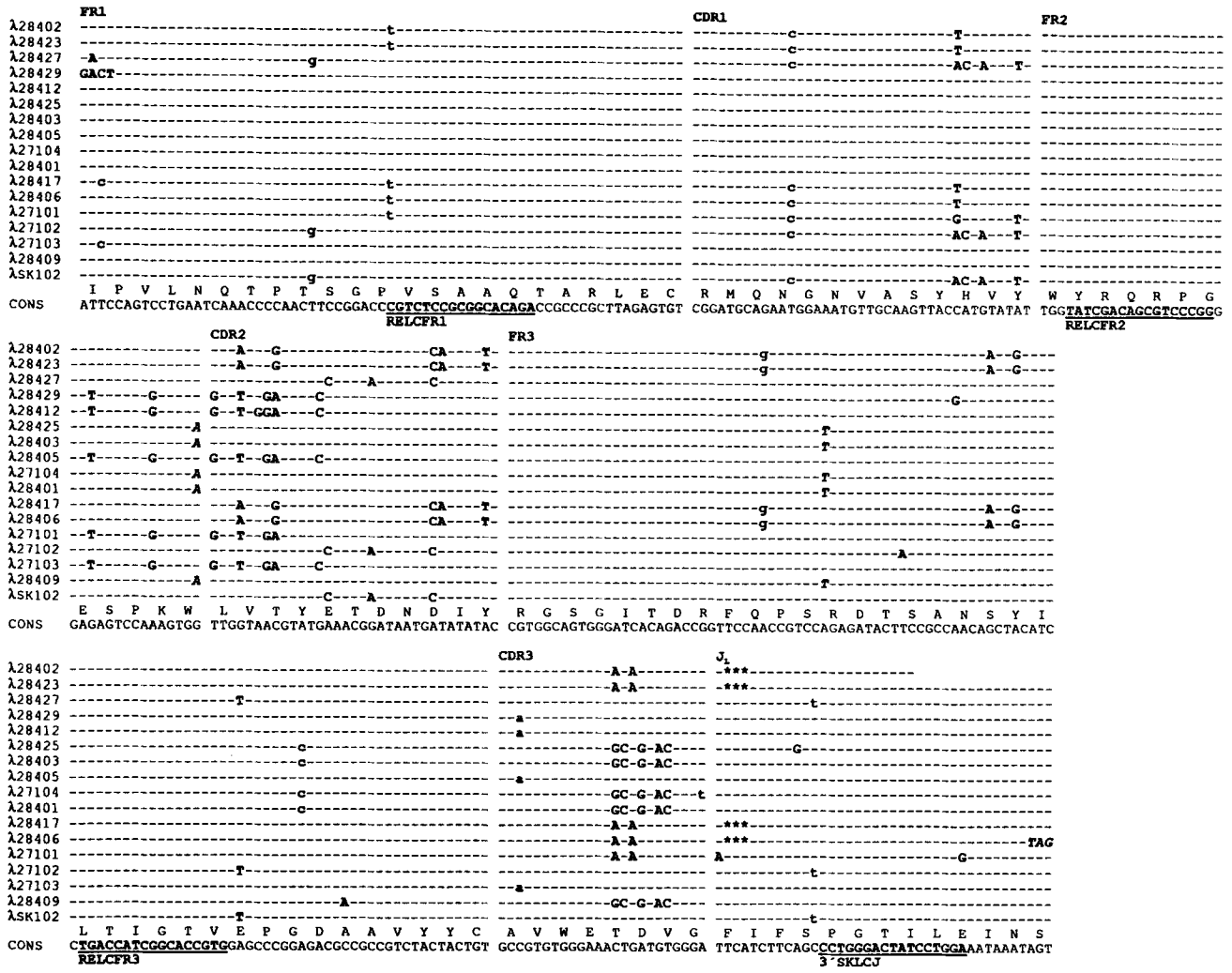


**Figure 3.** Conservation of nucleotide sequence between genomic library isolates in regions used to design primers RELCFR1, RELCFR2, RELCFR3, and SKLCJ, which were used to prime genomic DNA to generate the GPLC, FR2–$J_L$, and FJ clones. Dashes indicate nucleotide identity with the consensus sequence; asterisks, gaps; uppercase letters, nonsynonymous substitutions; and lowercase letters, synonymous substitutions. Predicted amino acids are located above the second nucleotide of each codon.

primers resulted in the identification of 54 unique clones in 127 informative cloning events. 14 of these sequences were represented by more than one clone. The GPLC sequences are at least 90% related from FR1 to FR3 to the other longer clones obtained by direct selection from the genomic library (17 unique sequences) or through priming of genomic DNA with RELCFR1 and SKLCJ (25 unique sequences).

As has been emphasized, there is considerable overlap in the three series of isolates, consistent with a representative selection/amplification of an extended gene family. The nucleotide sequence of GPLC103 matches those of λ27103 and λ28405. GPLC8 matches FJ36 and λ27101. GPLC13 matches FJ26 and FJ20, GPLC55 matches FJ33 and FJ7, GPLC10 matches FJ19 and FJ35, and GPLC82 matches FJ18. Assuming that overlaps between different clones isolated in different ways indicate independent amplification of the same gene, there are 89 type I light chain germline-joined genes that differ by at least 1 bp. Out of these 89 genes, only 7 appear to be pseudogenes (λ28425, λ28409, λ28403, λ27104, λ28401, λ28406, and GPLC73) because of in-frame stop codons resulting from point mutations.

A neighbor-joining tree constructed from an alignment of 93 FR1–FR3 region nucleotide sequences of representative unique genes from each series (Fig. 6) reveals (*a*) that since GPLC, FJ, and SKLCGEN clones are intermixed to different degrees in the different clusters, the relatedness of the genes is not a function solely of the method by which they were isolated; and (*b*) that sequences segregate into four major clusters, the first of which is subdivided into three smaller clusters. One of these consists of only a single representative (GPLC33). The fourth cluster includes five (λ28401, λ28409, λ27104, λ28425, and λ28403) of the seven pseudogenes identified in this study. These putative pseudogenes are identical to FR1–FR3. The other pseudogenes, λ28406 and GPLC73, distribute to the second cluster and one of the subdivisions of the first cluster, respectively.

*Comparison of Genomic and cDNA $V_L$–$J_L$ Sequences.* To determine whether any of the type I genes are transcribed,

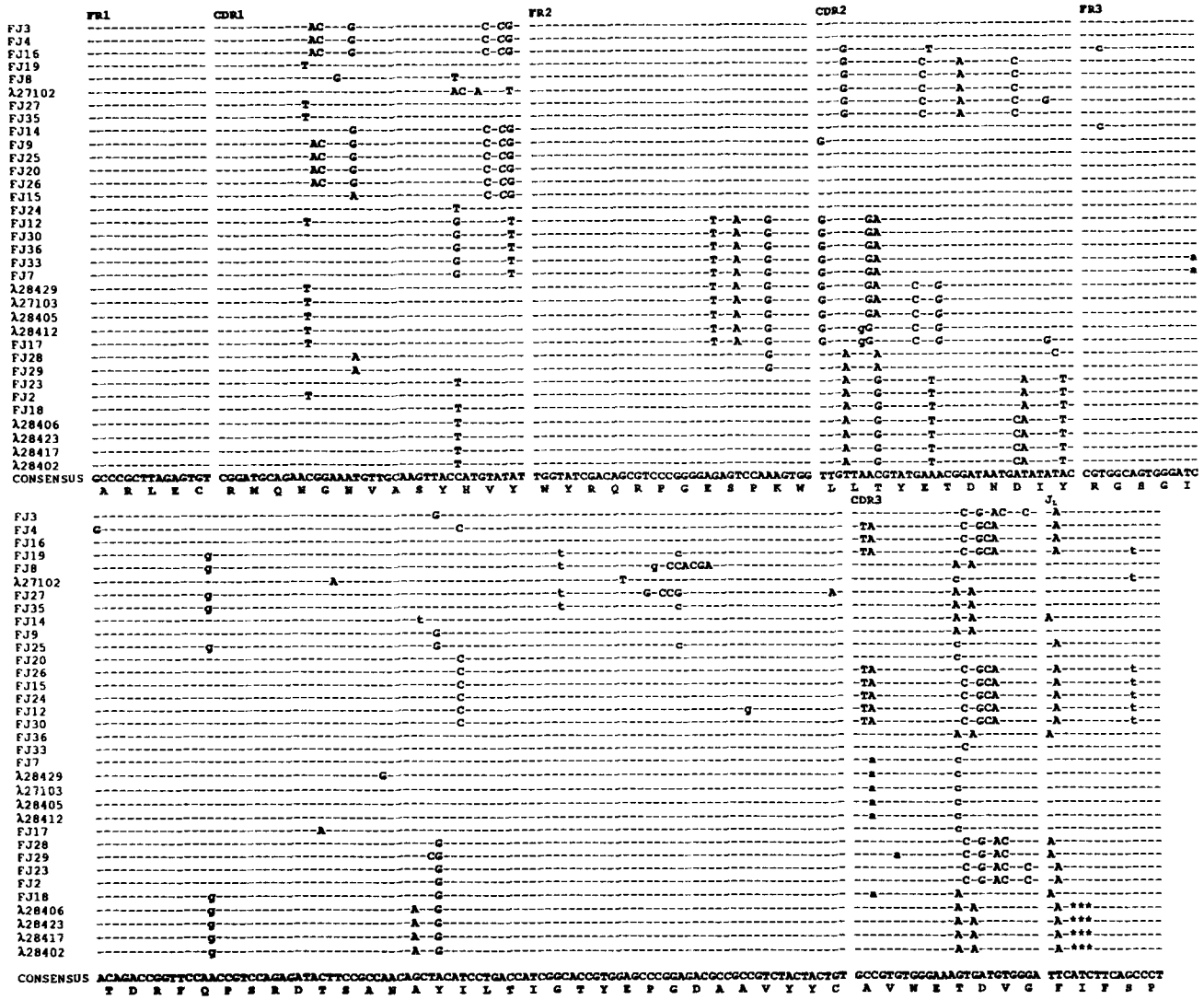


**Figure 4.** Alignment of regions between but not including RELCFR1 and SKLCJ primers of FJ clones, derived from genomic PCR, and SKLCGEN clones, derived from genomic library isolates. *Dashes*, nucleotide identity with the consensus sequence; *uppercase letters*, nonsynonymous substitutions; *lowercase letters*, synonymous substitutions. The FR, CDR, and J regions are denoted at the beginning of each region and are set apart from one another by a space. The predicted amino acid sequence of the consensus nucleotide sequence is shown below the second nucleotide of each codon. Note the abundance of substitutions in the CDRs as compared with the FRs.

a cDNA library was screened with the probe $REV_LC_L$, and the clones RRLC18, RRLC8, RRLC47, RRLC50, and RRLC36 were selected for analysis. The five cDNAs were compared with the consolidated, 89-member data base of GPLC, FJ, and SKLCGEN sequences (Fig. 7). An exact match was identified between the overlapping portions of cDNAs RRLC8, FJ18, and GPLC82. Similarly, an exact match was identified between the overlapping portions of RRLC47 and GPLC55. RRLC47 exhibits a 1-bp mismatch with FJ7 and with FJ33 in the regions that overlap. RRLC36 differs from GPLC82 by 2 bp and from FJ18 by 3 bp. RRLC50 exhibits a 3-bp mismatch with GPLC10 and a 4-bp mismatch with FJ35. RRLC18 differs from GPLC82 by 1 bp and from FJ18 by 11 bp. These data are consistent with the transcription of at least two germline-joined type I light chain genes. The other genes most likely represent somatically mutated forms or could represent allelic or pseudoallelic forms of these genes.

## Discussion

In mammals, combinatorial diversity results from the different recombination possibilities presented by tandemly arrayed segmental elements. Somatic joining of these elements is associated with the generation of diversity through both nucleotide deletions and nontemplated additions at the joining boundaries. Cartilaginous fish possess multiple clusters of immunoglobulin gene loci that recombine within but not between clusters (23), limiting combinatorial diversity. However, extensive junctional diversity and at least some degree

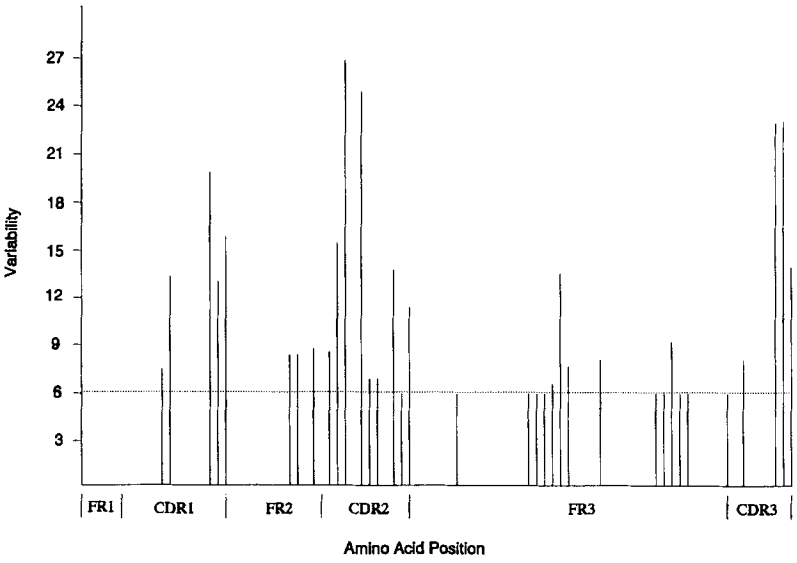


**Figure 5.** Variability at each predicted amino acid position of the *Raja* type I light chain $V_L$ gene regions. Variability was determined by dividing the number of different amino acids at each position by the frequency of the most common amino acid (36). Relative positions of amino acid residues in $V_L$ as well as FRs and CDRs are indicated. The dotted line indicates the variability corresponding to a single amino acid change at that position.

of somatic mutation have been demonstrated in rearranged immunoglobulin heavy chain clusters of the most intensively studied cartilaginous fish, *H. francisci* (23). An apparent second limitation in diversity arises from varying degrees of germline joining of ~50% of the heavy chain loci in this species (6). Similar preliminary findings of germline joining of type I light chain genes in *Heterodontus* (11), *Carcharhinus* (12), and *Raja* (11) suggest that the phenomenon is widespread and functionally significant. We provide unequivocal evidence that at least 41 of the 89 unique type I immunoglobulin light chain genes examined in *Raja* are germline $V_L$–$J_L$ joined, in contrast to the absence of joining in all 63 type I $V_L^+$ gene clusters that have been characterized in *Heterodontus* (7). Other experiments directed at amplifying potential $V_L$–$J_L$ intervening sequences in *Raja* failed to identify unjoined genes. Finally, we provide evidence that is consistent with transcription of at least some germline-joined genes, suggesting some physiological relevance. Immunoglobulin light chain clusters in *Raja* potentially represent an immunoglobulin gene system in which extensive germline diversification compensates for the loss of combinatorial and junctional diversity.

Although these studies did not result in the cloning and sequencing of all type I $V_L$ genes, it is likely that a representative sampling of type I light chain genes has been achieved. Specifically, 17 individual clones were identified and isolated from a genomic library using an approach that is capable of identifying genes that are only 60% related to the probe sequence. The ~90% relatedness between the overall sequences of the 17 clones indicates an absence of family divergence in $V_L$ regions, a finding entirely consistent with previously reported studies of *Heterodontus* $V_H$ genes, which are ~90% related. In addition, PCR primers designed on the basis of conserved regions of these 17 clones were used to amplify portions of $V_L$–$J_L$ genes directly from genomic DNA to identify genes that had not been incorporated in the library. Significantly, all of the genes amplified by the various primer pairs are related closely in nucleotide sequence to each other and to the genomic library isolates. Although all of the PCR products spanning the $V_L$ to $J_L$ regions are germline joined, it is possible that unusually long $V_L$–$J_L$ IVSs exist that would not be identified using these PCR strategies. However, the FR2–$J_L$ primers, which should have been able to amplify a *Heterodontus*-like (unjoined) type I light chain, because of the relatively short IVS characteristic of this type of gene, yielded germline-joined amplification products. If any unjoined type I clusters exist, they either must be present in very low abundance (relative to the joined type), possess extremely long $V_L$–$J_L$ IVSs, or differ appreciably from the joined genes in the primer regions. However, the identity between germline genes and cDNAs described in the next discussion is inconsistent with the last possibility.

The germline-joined state of the type I light chain genes in *Raja* results in a loss of junctional diversity, in addition to the lack of combinatorial diversity, because of the cluster-type organization of the gene loci. However, there are several lines of evidence which suggest that junctional diversity is not as important in the generation of light chain diversity as it is in generating heavy chain diversity, both in mammals and in the more phylogenetically distant vertebrates. For example, significant limitations in both nontemplated (N) additions and deletions at the $V_L$–$J_L$ coding joints have been demonstrated in mammalian $V_\kappa$ (24) and $V_\lambda$ (25) genes. In addition, CDR3 length and variability are more limited in both murine and human light chains as compared with heavy chains (26). Other systems, including the avian (4) and ovine (27) $V_\lambda$ genes, are diversified primarily by gene conversion and somatic point mutation, respectively. Perhaps the high level of somatic diversification associated with the presence of the D segment(s) in heavy chains provides a strong selective advantage for recombination of these genes. In the absence of D segments, the adaptive advantage of gene rearrangement in the generation of light chain diversity would be markedly reduced. There are no described cases of an entirely germline-joined heavy chain isotype, whereas this and

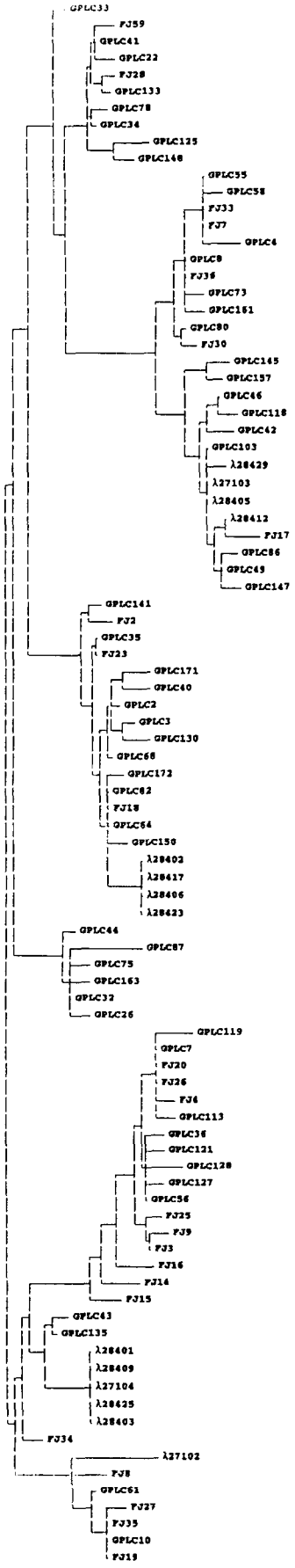


**Figure 6.** Neighbor-joining tree of the proportion of nucleotide differences between the FR1 and FR3 regions of the *R. erinacea* type I light chain $V_L$ regions. The tree was constructed by the neighbor-joining method in the MEGA program (19) based on the proportions of nucleotide differences (*p* distance) between pairwise comparisons of the regions between but not including primers RELCFR1 and RELCFR3. The *p* distance is obtained by dividing the number of nucleotide differences by the total number of nucleotides compared. The tree was rooted using four *Heterodontus* type I light chain sequences trimmed to the same length as the *Raja* sequences. Scale: Each – is approximately equal to the *p* distance of 0.001132.

previous studies (11, 12) provide strong evidence that two types of light chain genes are entirely germline joined, at least within a species. It is therefore conceivable that the level of germline heterogeneity in the *Raja* type I $V_L$–$J_L$ segments, especially in combination with somatic mutation mechanisms, compensates for the lack of both combinatorial and junctional diversifying mechanisms in this system.

The degree of germline diversity of the type I $V_L$–$J_L$ segments is a function of the number of unique genes, the amount of nucleotide identity between these genes, and the regional distribution of the differences. Although the 89 unique type I light chain $V_L$ (FR1–FR3) genes described in this study indicate a high level of germline-encoded heterogeneity (i.e., there are only 32 potentially functional human $V_\kappa$ genes [28]), many of these genes differ from each other by only 1–2 bp. Although these 1–2-bp mismatches may indicate Taq polymerase or Sequenase misincorporation, previous studies (28), including a large study from our laboratory (29), have indicated that such errors are negligible ($\sim$1/1,800) over relatively short nucleotide sequences and are inconsistent with repeated instances of recovery of identical sequences using different isolation and detection methods. These minor differences also could potentially represent allelic or pseudoallelic forms, although there is no a priori reason to assume that such differences would be associated with small versus large differences. Furthermore, these differences predominate in the CDRs and favor replacement substitutions, both of which are functionally significant.

By analogy to higher vertebrate immunoglobulin, the clustering of nucleotide sequence differences in CDRs would maximize variability in antigen-binding regions (30). Such nonsomatic variation would be subject to direct evolutionary selection, unlike the rearranged immunoglobulin genes of higher vertebrates, which are selected indirectly at a somatic level. The higher levels of variability at all three CDRs indicates that positive Darwinian selection may be acting specifically at these regions to maintain heterogeneity, which has been shown to occur also in mammalian CDR1 and CDR2 regions (31). Many of the substitutions in both the FRs and the CDRs occur in adjacent positions as blocks, similar to those observed in human $\kappa$ genes (28). Such blocks are typically indicative of genes that were derived from duplication and modification of an ancestral gene. However, the distribution of these blocks among the different regions of different genes indicates that they have not arisen solely by the process of evolutionary duplication and diversification. It also is unlikely that this distribution is a result of recombination of multiple templates during the PCR amplification since the sequences do not represent simple hybrids of known genes and some of these genes are represented independently by closely related cDNA sequences. Furthermore, the extension time of 4 min should be long enough to limit recombinants to <2% (32). In addition, one of the clones identified in the direct genomic library screening (λ27101) exhibits this type of distribution; i.e., in the FR1, CDR1, and CDR3 of λ27101, blocks of nucleotide sequence are shared with λ28402, λ28423, λ28417, and λ28406, whereas in CDR2, λ27101 does not share blocks with these clones, but instead shares blocks with λ28429, λ28412, λ28405, and λ27103. It seems more likely

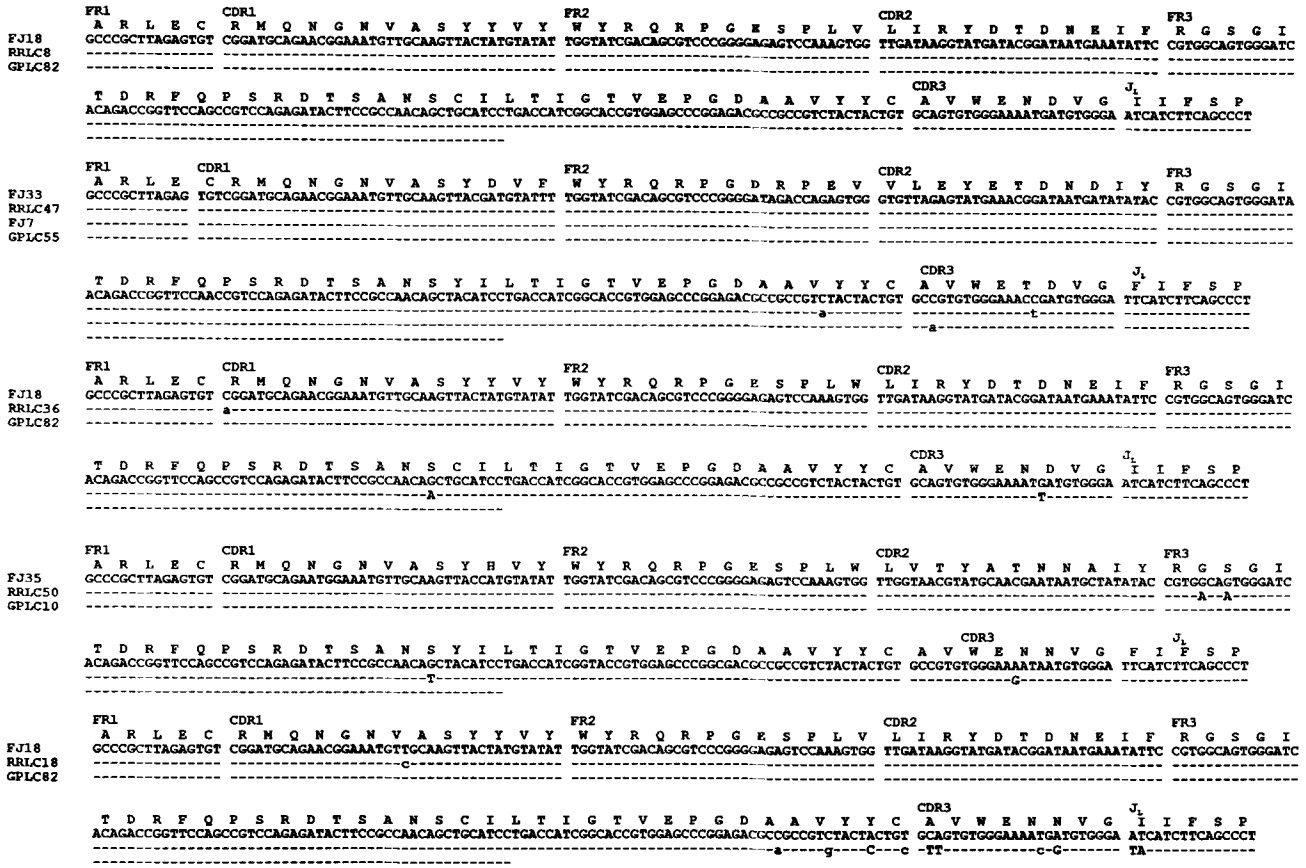

**Figure 7.** Alignment of the *Raja* type I light chain cDNAs and the type I $V_L$ germline-joined genes with which they share the greatest nucleotide sequence identity. Each cDNA is aligned with the genomic sequences from each group (i.e., SKLCGEN or FJ, and GPLC) with which it has the least number of mismatches. The top lines of each set are genomic sequences, the second lines represent the cDNA sequences, and the bottom lines are other closely matching genomic sequences. Dashes indicate nucleotide identity; FR, CDR, and $J_L$ regions are indicated at the beginning of each region, and each region is separated by a space. The predicted amino acid sequence of each cDNA is shown above the second nucleotide of each codon. Nonsynonymous differences in nucleotide sequence are indicated with uppercase letters; synonymous differences are indicated with lowercase letters.

that these genes are templates for multiple germline gene conversion events.

Although both sharks and skates exhibit ~50% germline joining in the IgM-type heavy chain gene clusters, it has not been possible to demonstrate that these genes are expressed. The identical sequences shared by two cDNAs (out of five) and germline isolates, including one sequence that extends through CDR3 to $J_L$, indicate that at least some of the type I light chain germline-joined genes in *Raja* are transcribed. In addition, findings with several other cDNAs are consistent with minor somatic change, or with expression of nearly identical allelic or pseudoallelic forms, as indicated previously (data not shown). Further distinguishing apparent somatic changes as arising by somatic mutation versus gene conversion in this system is particularly difficult owing to (*a*) the existence of multiple closely related gene loci; (*b*) the absence of approaches for obtaining purified lymphoid cell preparations or cell lines; and (*c*) the absence of a method for distinguishing joined genes from lymphoid versus nonlymphoid sources. Currently, the most reliable way to distinguish these events in lymphoid cells involves comparison of genes from a nonlymphoid source with cDNAs, which should be representative of genes from lymphoid sources. Comparisons between the cDNAs and the germline-joined genes in and around the mismatched nucleotides do not provide evidence for somatic gene conversion events in the *Raja* type I light chain gene loci, but rather are consistent with point mutation, although the lack of exact matches between some cDNAs and the germline-joined genes defined in this study also could be explained by the presence of additional germline-joined genes that were not amplified by PCR or integrated into the genomic library. The absence of an octamer in the region 100–150 bp 5′ of the putative start codon distinguishes the *Raja* gene from both the *Heterodontus* type I light chain genes (7), which are not joined, and the *Heterodontus* and *Raja* type II light chain genes (11), which are germline joined. This difference suggests that these genes are under a type of control other than that normally associated with vertebrate light chains. Examination of additional 5′ and 3′ sequences will help clarify this issue.

The available evidence suggests that germline-joined genes represent a derived rather than a primordial arrangement of segmental elements. The existence of unjoined TCR and immunoglobulin heavy chain genes in cartilaginous fish (6, 33)

suggests that the hypothetical common ancestor gene of TCR, heavy chain, and light chain genes was unjoined. Furthermore, the presence of unjoined type I light chain genes in *Heterodontus* indicates that the type I light chain gene in the hypothetical common cartilaginous fish ancestor was unjoined, whereas all type II light chain genes appear to be joined. Therefore, it seems likely that two independent events occurred during evolution, possibly involving in part the activation of a recombination-activating gene (RAG) homologue in germline cells, resulting in the joining of type II light chain genes at one early time point (affecting all cartilaginous fish), and the joining of type I light chain genes at another time after the divergence of the sharks and skates. A RAG homologue has recently been identified in a cartilaginous fish (34). The hypothesis that the joining of type I and II light chain genes represents independent events is supported by the finding of variability in length in CDR3 of type II, but not type I, light chain genes. Specifically, the type II light chain genes of *Carcharhinus* (12), *Raja*, and *Hydrolagus* (ratfish) (11) exhibit CDR3 segments of varying lengths within each species (5–12 predicted amino acids); all of the *Raja* type I light chain genes examined in this study possess CDR3 segments of 8 predicted amino acids. Moreover, this difference may reflect variation in the lengths of the original recombining segmental elements, or it may reflect specific types of joining processes, one of which (type II) resembles somatic rearrangement in the addition and/or deletion of nucleotides at the junctional boundaries. It has been suggested that the restrictions in CDR3 length observed in $\alpha/\beta$ TCR genes are caused by functional constraints in their recognition of MHC–antigen complexes, whereas the immunoglobulin genes and the $\gamma/\delta$ TCR genes are less restricted, since their specificities are not MHC restricted (26, 35). Therefore, it is possible that the difference in CDR3 length variability between type I and II light chain genes reflects varying affinities and specificities for distinct epitope types. Precommitted specificities presumably conferred a selective advantage to the germline-joined genes, resulting in the eventual loss of unjoined light chain genes.

It would be very informative to relate the *Raja* sequences described here to serum light chain proteins to determine the usage of type I versus type II light chains, particularly since both type I light chains described in this study and type II light chains described previously (11) may exist only in the germline-joined form. These two light chain types are ~36% related at the predicted peptide level and appear to be transcribed at similar levels in the spleen (11); i.e., equivalent numbers of type I and II genes are identified in library screening. Although unequivocal distinction of type I and II light chain genes is difficult, owing to intrinsic peptide heterogeneity, preliminary peptide analyses of serum immunoglobulin (pooled) light chains suggest the expression of both type I and type II light chain genes (Anderson, M. K., unpublished observations).

Weighing the relative contributions of various diversifying mechanisms in the context of overall immune function in lower vertebrate species is complicated. Although at first consideration the phenomenon described here appears to limit diversity, it needs to be recognized that (*a*) out-of-frame joinings, which could eliminate two out of three recombinations, do not occur; (*b*) potentially deleterious self-reactive antibodies would be selected against; (*c*) these genes would be ideally suited for a situation in which selection mechanisms, particularly complex cellular interactions associated with antigen recognition and expansion of antibody-producing clones, may not be present or may be inefficient; and (*d*) there are far larger numbers of gene loci compared with mammalian systems. In addition, the amount of somatic mutation operating in this system is still unknown and may provide a significant amount of additional diversity. In this regard, further examination of how antigen receptors are expressed on immunocytes of cartilaginous fish is critical to understanding the net functional consequences and selective advantages of the unique form of germline joining described here and the independent multicluster form of gene organization that encodes other types of antigen receptors in these species.

We would like to thank Barbara Pryor and Victoria Haire for editorial assistance as well as Annette Perry, Anna Zilch, and Eric Diner for technical assistance. The authors are indebted to Dr. Martin Flajnik for valuable suggestions regarding this work.

This work was supported by a grant to G. W. Litman from the National Institutes of Health (RO1AI23338).

Address correspondence to Dr. Gary W. Litman, College of Medicine, Department of Pediatrics, All Children's Hospital, 801 Sixth Street South, St. Petersburg, FL 33701-4899.

*Received for publication 10 January 1995 and in revised form 2 February 1995.*

## References

1. Tonegawa, S. 1983. Somatic generation of antibody diversity. *Nature (Lond.).* 302:575–581.
2. Chou, H.S., C.A. Nelson, S.A. Godambe, D.D. Chaplin, and D.Y. Loh. 1987. Germline organization of the murine T cell receptor $\beta$-chain genes. *Science (Wash. DC).* 238:545–548.
3. Desiderio, S.V. 1993. Organization and assembly of immunoglobulin genes. *In* Developmental Immunology. E. Cooper and E. Nisbet-Brown, editors. Oxford University Press, New York.

pp. 129–152.
4. Reynaud, C.-A., A. Dahan, V. Anquez, and J.-C. Weill. 1989. Somatic hyperconversion diversifies the single $V_H$ gene of the chicken with a high incidence in the D region. *Cell.* 59:171–183.
5. Reynaud, C.-A., V. Anquez, A. Dahan, and J.-C. Weill. 1985. A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. *Cell.* 40:283–291.
6. Kokubu, F., R. Litman, M.J. Shamblott, K. Hinds, and G.W. Litman. 1988. Diverse organization of immunoglobulin $V_H$ gene loci in a primitive vertebrate. *EMBO (Eur. Mol. Biol. Org.) J.* 7:3413–3422.
7. Shamblott, M.J., and G.W. Litman. 1989. Genomic organization and sequences of immunoglobulin light chain genes in a primitive vertebrate suggest coevolution of immunoglobulin gene organization. *EMBO (Eur. Mol. Biol. Org.) J.* 8:3733–3739.
8. Harding, F.A., N. Cohen, and G.W. Litman. 1990. Immunoglobulin heavy chain gene organization and complexity in the skate, *Raja erinacea. Nucleic Acids Res.* 18:1015–1020.
9. Litman, G.W., R.N. Haire, K.R. Hinds, C.T. Amemiya, J.P. Rast, and M.A. Hulst. 1992. Evolutionary development of the B cell repertoire. *Ann. N.Y. Acad. Sci.* 651:360–368.
10. Shamblott, M.J., and G.W. Litman. 1989. Complete nucleotide sequence of primitive vertebrate immunoglobulin light chain genes. *Proc. Natl. Acad. Sci. USA.* 86:4684–4688.
11. Rast, J.P., M.K. Anderson, T. Ota, R.T. Litman, M. Margittai, M.J. Shamblott, and G.W. Litman. 1994. Immunoglobulin light chain class multiplicity and alternative organizational forms in early vertebrate phylogeny. *Immunogenetics.* 40:83–99.
12. Hohman, V.S., D.B. Schuchman, S.F. Schluter, and J.J. Marchalonis. 1993. Genomic clone for the sandbar shark lambda light chain: generation of diversity in the absence of gene rearrangement. *Proc. Natl. Acad. Sci. USA.* 90:9882–9886.
13. Greenberg, A.S., L. Steiner, M. Kasahara, and M.F. Flajnik. 1993. Isolation of a shark immunoglobulin light chain cDNA clone encoding a protein resembling mammalian type kappa light chains: implications for the evolution of light chains. *Proc. Natl. Acad. Sci. USA.* 90:10603–10607.
14. Litman, G.W., L. Berger, K. Murphy, R. Litman, K.R. Hinds, and B.W. Erickson. 1985. Immunoglobulin $V_H$ gene structure and diversity in *Heterodontus*, a phylogenetically primitive shark. *Proc. Natl. Acad. Sci. USA.* 82:2082–2086.
15. Stingo, V., and L. Rocco. 1991. Chondrichthyan cytogenetics: a comparison with teleosteans. *J. Mol. Evol.* 32:76–82
16. Kokubu, F., K. Hinds, R. Litman, M.J. Shamblott, and G.W. Litman. 1988. Complete structure and organization of immunoglobulin heavy chain constant region genes in a phylogenetically primitive vertebrate. *EMBO (Eur. Mol. Biol. Org.) J.* 7:1979–1988.
17. Feinberg, A.P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6–13.
18. Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463–5467.
19. Kumar, S., K. Tamura, and M. Nei. 1993. MEGA: Molecular Evolutionary Genetics Analysis, Version 1.0. The Pennsylvania State University, University Park, PA.
20. Kabat, E.A., T.T. Wu, C. Foeller, H.M. Perry, and K. Gottesman. 1991. Sequences of Proteins of Immunological Interest. U.S. Dept. Health and Human Services, Washington, DC.
21. Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406–425.
22. Hinds, K.R., and G.W. Litman. 1986. Major reorganization of immunoglobulin $V_H$ segmental elements during vertebrate evolution. *Nature (Lond.).* 320:546–549.
23. Hinds-Frey, K.R., H. Nishikata, R.T. Litman, and G.W. Litman. 1993. Somatic variation precedes extensive diversification of germline sequences and combinatorial joining in the evolution of immunoglobulin heavy chain diversity. *J. Exp. Med.* 178:825–834.
24. Victor, K.D., K. Vu, and A.J. Feeney. 1994. Limited junctional diversity in $\kappa$ light chains: junctional sequences from $CD43^+$ $B220^+$ early B cell progenitors resemble those from peripheral B cells. *J. Immunol.* 152:3467–3475.
25. Boudinot, P., A.-M. Drapier, P.A. Cazenave, and P. Sanchez. 1994. Mechanistic and selective constraints act on the establishment of V$\lambda$J$\lambda$ junction in the B cell repertoire. *J. Immunol.* 152:2248–2255.
26. 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.
27. Reynaud, C.-A., C.R. Mackay, R.G. Muller, and J.-C. Weill. 1991. Somatic generation of diversity in a mammalian primary lymphoid organ: the sheep ileal Peyer's patches. *Cell.* 64:995–1005.
28. Klein, R., R. Jaenichen, and H.G. Zachau. 1993. Expressed human immunoglobulin $\kappa$ genes and their hypermutation. *Eur. J. Immunol.* 23:3248–3271.
29. Haire, R.N., R.D. Buell, R.T. Litman, Y. Ohta, S.M. Fu, T. Honjo, F. Matsuda, M. de la Morena, J. Carro, R.A. Good, and G.W. Litman. 1993. Diversification, not utilization, of the immunoglobulin $V_H$ gene repertoire is restricted in DiGeorge syndrome. *J. Exp. Med.* 178:825–834.
30. Kirkham, P.M., and H.W. Schroeder. 1995. Antibody structure and the evolution of immunoglobulin V gene segments. *Sem. Immunol.* In press.
31. Tanaka, T., and M. Nei. 1989. Positive Darwinian selection observed at the variable-region genes of immunoglobulins. *Mol. Biol. Evol.* 6:447–459.
32. Meyerhans, A., J.-P. Vartanian, and S. Wain-Hobson. 1990. DNA recombination during PCR. *Nucleic Acids Res.* 18:1687–1691.
33. Rast, J.P., and G.W. Litman. 1994. T cell receptor gene homologs are present in the most primitive jawed vertebrates. *Proc. Natl. Acad. Sci. USA.* 91:9248–9252.
34. Greehalgh, P., and L.A. Steiner. 1995. Recombination activating gene 1 (Rag1) in zebrafish and shark. *Immunogenetics.* 41:54–55.
35. Schild, H., N. Mavaddat, C. Litzenberger, E.W. Ehrich, M.M. Davis, J.A. Bluestone, L. Matis, R.K. Draper, and Y.-H. Chien. 1994. The nature of major histocompatibility complex recognition by $\gamma\delta$ T cells. *Cell.* 76:29–37.
36. Wu, T.T., and E.A. Kabat. 1970. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J. Exp. Med.* 132:211–250.