

Rearrangement of Immunoglobulin Genes in Shark Germ Cells

By Susan S. Lee,* David Fitch,† Martin F. Flajnik,§ and Ellen Hsu*

From the *Department of Physiology and Pharmacology, State University of New York Health Science Center at Brooklyn, Brooklyn, New York 11203; the †Department of Biology, New York University, New York, New York 10003; and the §Department of Microbiology and Immunology, University of Maryland, Baltimore, Maryland 21201

Abstract

The variable (V), (diversity [D]), and joining (J) region recombinases (recombination activating genes [RAGs]) can perform like transposases and are thought to have initiated development of the adaptive immune system in early vertebrates by splitting archaic V genes with transposable elements. In cartilaginous fishes, the immunoglobulin (Ig) light chain genes are organized as multiple VJ-constant (C) clusters; some loci are capable of rearrangement while others contain fused VJ. The latter may be key to understanding the evolutionary role of RAG. Are they relics of the archaic genes, or are they results of rearrangement in germ cells? Our data suggest that some fused VJ genes are not only recently rearranged, but also resulted from RAG-like activity involving hairpin intermediates. Expression studies show that these, like some other germline-joined Ig sequences, are expressed at significant levels only early in ontogeny. We suggest that a rejoined Ig gene may not merely be a sequence restricting antibody diversity, but is potentially a novel receptor no longer tied to somatic RAG expression and rearrangement. From the combined data, we arrived at the unexpected conclusion that, in some vertebrates, RAG is still an active force in changing the genome.

Key words: V(D)J recombination • rearrangement • immunoglobulin genes • light chain • germline joining

Introduction

The lymphocyte antigen receptors encoded by Ig and TCR genes bind foreign ligands after clonal selection from a vast repertoire of combining sites. The diversity of the combining sites is generated in part by a mechanism that joins different gene segments in a combinatorial fashion, and in the process somatically deletes and sometimes inverts DNA in the lymphocyte. This mechanism and the antigen receptors have so far been found only in jawed vertebrates, e.g., cartilaginous fishes such as sharks and skates being representatives of the oldest of them (1). The unique rearrangement mechanism stirs interest in the origin and evolution of the Ig and TCR gene systems, which have not been identified in hagfish or lamprey, the living representatives of the oldest vertebrates. The apparent absence, too, of the

recombination activating genes (RAG)¹-1 and RAG-2, which encode the enzymes mediating gene rearrangement, suggests that these, and the genes for the antigen receptors, coevolved in an ancestral jawed vertebrate. Recently, it has been found that the RAG enzymes not only cause DNA excision, but can also mediate reintegration of the excised DNA (2, 3), both processes through a biochemical mechanism akin to integrases and transposases. These findings support a hypothesis that the Ig and TCR gene segments, which are all separated by interspersed noncoding DNA with specific recombination signal sequences (RSS) adjacent to the exon joining ends, were generated after the RAG-mediated insertion of a mobile sequence into a prototype Ig gene, rending it into two gene segments, V and J (see Fig. 1, and references 2, 4, 5).

In the course of evolution, the ancestral RAG transposase acquired specialized function in vertebrates, excising mobile sequences in developing lymphocytes. The gene segments thus targeted underwent extensive duplication and reorganization, generating the TCR chains and the H and L chains of antibodies. By and large, in tetrapods Ig and TCR genes are organized as tandem arrays of V, followed

Address correspondence to Ellen Hsu, Department of Physiology and Pharmacology, State University of New York Health Science Center at Brooklyn, 450 Clarkson Ave., Brooklyn, NY 11203. Phone: 718-270-1154; Fax: 718-270-3103; E-mail: hsu@hscbklyn.edu

¹Abbreviations used in this paper: FR, framework region; Myr BP, million years before present; RAG, recombination activating gene; RSS, recombination signal sequence(s); RT, reverse transcription; TdT, terminal deoxynucleotidyl transferase.

by (D and) J gene segments, and by a few C region genes (Fig. 1). In bony fishes, the Ig H chain genes are similarly organized (6, 7). However, the Ig L chain genes in bony fishes, as well as the Ig H and L chain genes in cartilaginous fishes, are organized as multiple clusters each containing VH, D, JH, CH, and VL, JL, CL genes, respectively (8–11). This divergence of gene arrangement among vertebrates suggests that sharks and skates carry one of the organizational alternatives—perhaps the most ancient one—developed early in the evolution of the Ig genes.

In sharks and skates, many clusters contain Ig genes that are rearranged in the germline, sometimes as they can be in lymphocytes (10, 12, 13). L chain genes of any one type of the three so far reported have been found either all in split germline configuration or all as joined VJ (see Fig. 2, bottom). In contrast, the H chain gene clusters contain rearranged, partially rearranged, or nonrearranged V, D, and J genes (12–14). In comparing the H chain joined and split sequences, it was difficult to dissect the nature of the joining events because of the highly varied sequences at the junctions. The process by which these joined genes were created, whether RAG, terminal deoxynucleotidyl transferase (TdT), or the RSS were actually involved, is unclear.

In this study, genes of one L chain type, NS4, have been isolated from the nurse shark *Ginglymostoma cirratum*. They are homologues of the type III L chains in the horned shark (15, 16). Whereas the type III genes are reported to be all split (13), we have identified clones of NS4 V and J genes existing as both germline-joined and nonjoined (see Fig. 2). This mixed combination in one L chain type is a novel finding; furthermore, the NS4 genes are so similar that it is possible to explore not only the evolution among the various joined genes, but also their relationship to the split genes. The probable steps of the germline joining events

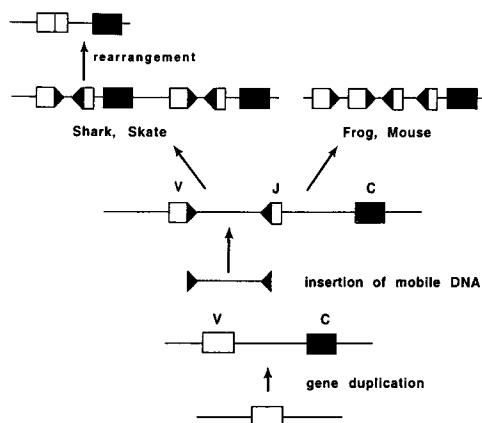


Figure 1. Hypothetical scheme for evolution of Ig L chain genes. The precursor Ig gene underwent gene duplication, generating the V and C exons. Sometime in or before the early Devonian, >400 Myr BP, a mobile DNA sequence was inserted in the V exon, splitting it into V and J gene segments. Dark arrows on the DNA insert indicate the RSS. Gene duplication of the individual genes in tandem generated the overall L chain Ig gene arrangement in tetrapods, whereas duplication of V, J, and C genes produced the multiple cluster organization common in cartilaginous fishes. Based on diagram from reference 4.

were reconstructed, and these were compared with the joints in cDNA rearrangements obtained from lymphocytes of the same individual. Our data strongly suggest that the joined genes are products of a germline V(D)J RAG-mediated recombination event; this genomic evidence supports RAG-like activity occurring in germ cells.

Materials and Methods

Genomic Libraries. Whole blood was obtained from nurse shark Y, and the cells were centrifuged through a Ficoll gradient to separate erythrocytes from PBLs. Genomic DNA was isolated from both populations of cells, and RNA was extracted from the PBLs. DNA from the RBCs was digested with EcoRI and fractionated on a 1% low melt agarose gel. DNA fragments corresponding to 6–4- and 3–2-kb ranges were isolated and separately ligated into λ ZapII phage arms (Stratagene). $2\text{--}2.5 \times 10^5$ bacteriophages from each library were screened with probes from the V region of NS4 and NS3. 90 phages from the 3–2-kb library hybridized with NS4 probe, and among them R4, R6, R18, RE18, R19, and all the “S” clones were identified. R7 was isolated from shark Y PBL genomic DNA library; we believe that R7 is a germline-joined gene because three independent clones were isolated, and two more sequences identical to R7 were isolated from RBC DNA from an unrelated animal, shark A. Each phage isolate was analyzed by PCR using primers to the leader (NSL) and J segment (JL5) of NS4 to determine whether the cloned gene was germline-rearranged as VJ. DNA was extracted from phage lysates and digested with EcoRI to determine the insert size. The selected phages were purified, and the plasmid component of λ ZapII was excised and subjected to the dideoxy sequencing technique (17).

PCR. Oligonucleotide primers to the NS4 leader sequence (NSL) and the JL5 were synthesized by GIBCO BRL (NSL, 5'-GATTTTCICAYATICARCT-3'; JL5, 5'-CTTGGTTCCTT-TACCGAA-3', where I is inosine, Y is T/C, and R is G/A), and used in PCR under the following cycling conditions: 96°C for 1 min, 46°C for 2 min, 72°C for 2 min for 39 cycles, and a final cycle with the 72°C step extended to 15 min. DNA templates consisted of 100–500 ng genomic DNA per 0.25-ml reaction, or of denatured phage lysate not exceeding 17% of the final volume.

Reverse transcription (RT)-PCR assays were performed as described previously (18). Oligo dT was used for priming first strand synthesis in total RNA from PBLs of shark Y. RT-PCR products were cloned into pGEM-T-Easy (Promega).

Probes. A joined VJ NS4 sequence of 314 bp was initially obtained from shark Y genomic DNA by amplification with primers NS4A (5'-ATCACIATGACICARTC-3') in framework region (FR)1 and JL5 in the J segment (16); the sequence was later identified as part of R4 in the genomic library. The germline-joined NS3 sequence is published (16) and was also amplified using primers in FR1 (19) and J. The I probe containing intervening sequence between V and J was obtained from split NS4 gene S5 (similar in structure to S1 or S10), which contained a HincII site in the CDR3 in the 3' region of the V and a SacI site in the RSS of the J gene end. S5 was subjected to restriction endonuclease digestion (New England Biolabs), and the ~500-bp DNA fragment was isolated from a low melt gel, using GeneClean II (Bio101), and radiolabeled (Prime-It II; Stratagene). Hybridization and blotting procedures have been described elsewhere (20).

Animals. Nurse sharks were captured off the coast of the Florida Keys, and blood was obtained from the caudal sinus into

heparinized syringes. Blood cells from one individual, shark Y, were separated into erythrocyte and PBL fractions, which were used in the cloning and expression studies described under Genomic Libraries and PCR above. DNA extracted from either erythrocytes or noncollected blood cells was obtained from other individuals.

Phylogenetic Analyses. Nucleotide alignments were constructed manually. In rare instances of difficulty in alignment, homology assignments were informed by comparing the derived amino acid sequences. PAUP*4.0beta4 (Phylogenetic Analysis Using Parsimony [and Other Methods] v4.0b2a; D.L. Swofford, Sinauer Associates) was used with five character and taxon partitions (data sets) of nucleotide or amino acid characters and taxa: (A) the coding regions only minus the first three codons not present in h6 (all taxa); (B) the coding regions and intron (NS4 family genes only); (C) the entire nucleotide alignment (NS4 genes, except S31 and S32); (D) only the first and second positions of the coding regions (all taxa); and (E) the derived amino acid alignment, except the first three residues (all taxa). Methods used to evaluate trees in searches were maximum parsimony (MP), minimum evolution (ME), and minimum likelihood (ML) as implemented in PAUP*. The neighbor-joining algorithm (NJ) was also used to estimate phylogenies. In MP analyses, characters were weighted equally; transformations were either weighted equally or transversions were weighted twice that of transitions. Gaps in the alignment were either ignored (coded as missing) or each gap was counted as a single insertion/deletion (indel) event, regardless of the number of sites covered by the gap. Before ME, NJ, and ML analyses, Modeltest 3.0 (21) was used with PAUP* to test a variety of models in a hierarchical fashion to find the simplest model that fit the data not significantly worse than more complex (generalized) models. The model identified and used for tree searches by the distance and likelihood methods was K80+ Γ (the two-parameter model of Kimura [22] with an assumed transition/transversion rate ratio of 1.2134, corrected for rate variation across sites with a gamma distribution shape parameter of 0.3301, estimated as four discrete rate categories). In analyses with ≤ 11 taxa, exhaustive searches were made, except when the ML method was used. In all other analyses, heuristic branch-swapping (tree-bisection-reconnection, TBR) searches were performed, with starting trees obtained in 1,000 replicates of random sequence additions. Additionally, in parsimony searches, the 3,000 best trees were kept and used for additional branch-swapping searches. For evolutionary reconstructions based on the inferred phylogenetic networks, rooting of the nurse shark NS4 phylogeny employed the outgroup method with all sequenced genes of the homologous type III family of horned shark and representatives from the more divergent NS3/type II and NS5/type I families. This rooting could only be used for data sets A, D, and E. However, analyses with these data sets revealed that SE7 is the most basally diverged member of the NS4 family; thus, SE7 was used for rooting with data sets B and C. For a final estimation of branch lengths, the ML method was used with the HKY85 model (23) with base frequencies and transition/transversion ratio estimated from the data, corrected for invariance and rate variation across sites by a gamma distribution approximated as four discrete rate categories (with parameters estimated from the data by PAUP*).

Statistical Analyses. Support for different branches in the estimated trees was evaluated by bootstrapping (with replacement), jackknifing (50% deletion), decay index ("Bremer support index"; 24). Except with ML (where 100 replications were used), 2,000 bootstrap and jackknife replications were used for each analysis. Sets of $\sim 2,900$ – $3,200$ trees collected during tree searches

were compared statistically with the KH (25, 26) test, as implemented in PAUP*. For each such analysis, a strict consensus of unrejected trees was calculated to identify relationships that remained fully resolved.

Dating. We assumed that the point of divergence among the NS4 and type III genes corresponded to the point of divergence between the two shark species. The earliest point at which rearrangements were likely to have occurred was then calculated as the fraction of the total length from the species divergence to the most closely related functional gene where each rearranged gene diverged. Only the nonrearranged, functional genes S31 and S10 were used so as to eliminate artefactual inflation of rate estimates due to increased rates in pseudogenes. Using only the most closely related functional genes for calibrating the molecular clock also reduces the error due to rate differences among genes. This proportion was then applied to the species divergence date between nurse shark and horned shark, assumed to be 180 million years before present (Myr BP), as estimated from fossil data (27), to derive the earliest date at which germline rearrangements could have occurred.

Results

Isolation of NS4 L Chain Genes. A probe to the NS4 V region revealed a series of hybridizing EcoRI-digested DNA fragments by genomic Southern blotting: ~ 6.0 , 5.1, 4.6, and 4.3 kb, many bands at 2.8–2.65, 2.4, and 2.1 kb (Fig. 2, panel V). To isolate these sequences, two genomic libraries were constructed from size-selected erythrocyte DNA fragments of 6–4 and 3–2 kb. In the latter library, 90 phages were selected and PCR analysis was performed with primers to the leader and J gene segment. Nine phages gave NS4-hybridizing PCR fragments of ~ 550 bp, four of $\sim 1,000$ bp, and the rest $\sim 1,078$ bp. This last, largest group consisted of nonrearranged ("split") NS4 genes with a leader intron of 186 bp and an intervening sequence of ~ 500 bp (502–524 bp) separating the V and J gene segments. All EcoRI inserts were in the 2.6–2.8-kb range. The second group of four phages also consisted of split genes, with a shorter V-J intervening sequence, and the insert size was 2.5 kb; the only one analyzed was S32.

The nine phages that produced PCR fragments of 550 bp all contained germline-rearranged NS4 genes. The EcoRI insert sizes varied, six being ~ 2.1 kb, two 2.4 kb, and one 4.6 kb. A probe (I) consisting of the intervening sequence between V and J hybridized to all but these EcoRI fragments (Fig. 2, panel I; arrows in lane 1). A side-by-side comparison of the cloned joined gene of 4.6 kb (RE18) showed that it comigrated with the same band on a genomic Southern blot (not shown); both hybridized with V and not I probe.

Hybridization of these filters with the probe of another L chain type, NS3, identified three hybridizing phages that all contained DNA with similar restriction enzyme sites. In previous genomic Southern analyses, there appeared to be at least two copies of NS3 in the 3–2-kb region. Based on these results, and on comparisons of the NS4 signal from dilutions of total EcoRI-digested DNA with that of the selected 3–2-kb fraction, we estimate that 200,000 phages

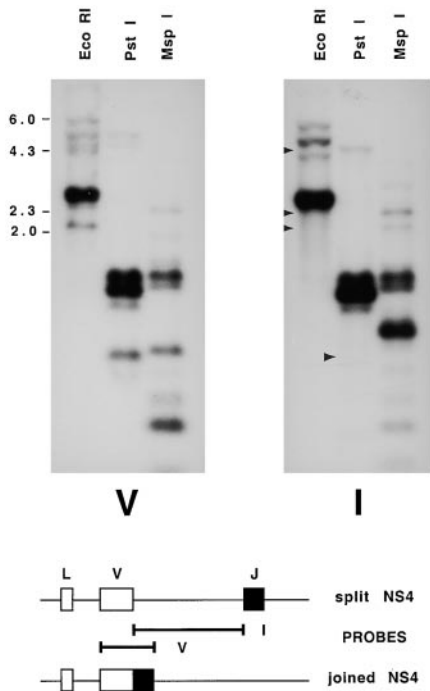


Figure 2. Genomic Southern blot hybridized with probes to the coding and intervening regions of NS4. Top, erythrocyte DNA from shark Y was digested with EcoRI, electrophoresed on a 1.2% agarose gel, and transferred onto a nylon filter. The filters were hybridized with radiolabeled DNA sequences from the V region of NS4 (V) and the intervening sequence between V and J (I). Arrows in the EcoRI lane point to the location of V-hybridizing bands at 2.1, 2.4, and 4.6 kb that were not detected by the I probe. These were cloned and contained germline-rearranged NS4 genes. Arrow in the PstI lane points to a fragment of 727–744 bp containing joined genes R4, R7, R18, R6, and R19, predicted from DNA sequence in Fig. 3 (dots over positions 79 and ~816); accordingly, this fragment does not hybridize to I probe. The slower migrating bands in the same lanes at ~1,150, 1,250, and 1,400 bp correspond to sizes expected with the presence of an intervening sequence of ~500 bp; these bands accordingly hybridized with both V and I probes. Molecular size markers from λ phage DNA digested with HindIII are shown. The PstI fragment sizes were calculated from another gel; not shown. Bottom, location of V and I probes from cloned split (S5) and joined (R4) genes.

with 3–2-kb inserts represented ~2 genomes of nurse shark DNA, assuming a genome size of $6\text{--}8 \times 10^9$ bp. Thus, in this region there are at least 45 NS4 genes, 3–5 of which are germline rearranged. In the 6–4-kb range there are five to six NS4 genes, at least one of which is germline-joined (Fig. 2); the latter (RE18) was cloned in the 3–2-kb library. Thus, our estimates suggest a total of 50 NS4 L chain genes in shark Y, 10% of which is not in the split germline configuration.

Six unique germline-joined NS4 genes were identified; some may be alleles, two (R6, R19) are pseudogenes. Except for RE18, all were obtained in duplicate from the screening of three independent genomic libraries prepared from shark Y, and thus assuredly originate from germline sequence. No split gene was identified in any cloned EcoRI fragment of 2.1 kb, consistent with the absence of I probe hybridization to the genomic DNA (Fig. 2).

The nucleotide sequences are shown in Fig. 3 and compared with nonjoined genes S1, S10, S31, and S32. The J gene segments of the latter sequences are also shown separated by slashes that indicate exclusion of the RSS and the intervening sequence. The predicted amino acid sequences are shown in Fig. 4.

Germline Origin of Joined NS4 Genes. There are two PstI sites in both R and S sequences, at position 79 between octamer and leader, and at positions 806–833 after J gene segment (Fig. 3, bold dots). Thus, the sequences R4, R7, R18, R6, and R19, but not RE18, are expected to produce fragments of 727–754 bp when digested with PstI. In genomic Southern blots, such a band was observed with V probe and not with I probe (Fig. 2, panel I; arrow in second lane). Split sequences such as S1 and S10 would contain in addition an intervening sequence of ~500 bp, and the predicted PstI fragments would be 1,200 bp. Such corresponding genomic DNA PstI fragments hybridized with both V and I probes (Fig. 2, both panels, second lanes; see legend).

Additional experiments demonstrating the germline origin of the R genes, in particular R4/R7, are shown in Fig. 5. DNA samples from seven wild nurse sharks were analyzed by genomic Southern blotting. The signal from the EcoRI 2.1-kb fragment, from which only joined genes have been identified, was present in all individuals and identical in intensity in the erythrocyte and PBL lanes of shark Y (Fig. 5 A, V, lanes 6 and 7), showing that contribution of somatically rearranged NS4 sequence, even in the PBL DNA, was not visible compared with the germline-rearranged genes.

There is a ClaI site in CDR2 of the R4 and R7 sequences (Fig. 3); restriction enzyme analyses confirmed that digestion with ClaI yielded fragments of ~1,150 and 980 bp. Bands of this size hybridizing with V probe were observed in genomic DNA digested with EcoRI and ClaI (Fig. 5 B, lanes 2 and 4, arrows). R6 and R19 are on 2.1-kb EcoRI fragments but do not contain ClaI sites (Fig. 5 B, diagram on right). Accordingly, the signal from 2.1-kb fragments is reduced but not abolished in the EcoRI-ClaI-digested DNA.

Phylogenetic Analyses. The NS4 coding sequences, together with the type III L chain homologues in horned shark and representatives from other L chain types, were subjected to several different methods of phylogenetic estimation. Only one tree appeared as a best tree in all of the analyses (Fig. 6). The various L chain types grouped together as NS4/type III, NS3/type II, and NS5/type I, suggesting that these are the respective orthologs in the two species. The nurse shark NS4 gene family is significantly supported as a monophyletic clade separate from the homologous type III gene family of the horned shark, indicating that gene expansion and diversification occurred independently in the two species lineages.

Three groups of NS4 genes emerged, clade I (S1, S10, S31, R4, R7, and R18), clade II (S32, R6, R19, RE18), and clade III (SE7), and the joined genes grouped with split genes. Assuming that once L chain genes were joined in the germline they cannot be reverted to split genes, it is es-

```

10      20      30      40      50      60      70      80      90     100
R4  GTGTGACGCGCTGACGTGCTCTGTCTCTGCTGATGCAATTACATACTGGGGAACAGCCAGTGTGAGACTCTCACAGGCTGCAGCTTTATAAAGCAGGGC 100
R7  .....G..G.....C.....A..... 100
R18 .....G..G.....C.....A..... 100
RE18 .....G..G.....CC.....N..A..T.....A..A..... 100
R6  .....G..G...T...C..T.....T.....A..... 100
R19 .....G..G.....C.....C.....A..... 100
S1  .....G..G.....A..... 100
S10 .....G..G.....A..... 100
R4  CCGGTGAAGGGAGAGTTTATCAAGAACACACAGGACAGGACAGCACACCACTGATTTCCACACATCCAGCTGATTTGGCCTCTGGCATCTCTGCCTTGC 200
R7  .....G..G.....G.....C..... 200
R18 .....G..G.....GG.....GG.....C..... 200
RE18 .....A...T...G..G...GG...GG.....C.....A..... 200
R6  .....G..G.....G.....C.....T..CA.. 200
R19 .....A...G..T...GG.....C..... 200
S32 .....G..G.....GG.....C.....A..C.. 50
S1  .....A..... 200
S10 ..... 200
S31 ..... 50
R4  ACGTACAATCTCTCTCTCCACTTGAATCTTTAGCTGATCTCTTTTCACTCCAATCCACTGACTTGTGATTTAAGGGAAAATGCTGAAACCAGAG 300
R7  ..... 300
R18 ..... 300
RE18 .....C.....A..... 300
R6  .....C.....A..C..A.....A..... 300
R19 .....C.....C.....A..... 300
S32 .....T.....C.....A..C.....C.....T.....G.....G.....G..... 150
S1  ..... 300
S10 ..... 300
S31 .....C..... 150
R4  GAATGCTCAGTGTCTCCGGCAATCTCTCAATTGACAGGCTCTTTCTGTGACAGTTCGACTTCACCTGTGTTCTCTCTGACCCCTCAGGTATCAGTGGGG 400
R7  ..... 400
R18 .....AA.....C..... 400
RE18 .....CA...A...GAAG..G.....T..... 400
R6  .....T.....AA.....G..... 400
R19 .....CA..T..C...GAA..G..G.....T..... 400
S32 .....C..G..AA..G.....T..... 249
S1  .....C.....T..... 400
S10 .....C..... 400
S31 .....A.....T..... 250
R4  FR1→ ACATACCATGACCCAGT-----CTCCCC-----GGTGCCTGTCAGTGGGACAGCCGCAACCATCACTGATACGGCCAGT 477
R7  .....A.....T..... 477
R18 ..... 477
RE18 .....T...A.....GTACCCCGACCCCCC..A...CCAACACC..A.....TG.....TC 477
R6  .....G..T...A...T.....C.....C.....G.....A..T.....TC 480
R19 .....T...A...T.....C..T.....A..... 324
S32 .....T.....C.....T..... 477
S1  .....T..... 477
S10 .....T..... 477
S31 ..... 327
R4  AAAAGCATTTCACAGCTACCTTGCCTGGTACCAGCAGCAGGAGAGGACAGAAACCCCTCTCTCTGATCTATTATGCAACAATTCGATACACAGGAGCTCCCG 577
R7  C..TCT...AG..A..G.....G..... 577
R18 C..TCT...AG.....C..... 577
RE18 C..TATTC..GGA...TG...AA.....A.....G..A...G...T...C..... 577
R6  C..TATTC..GGA...TG...AA.....A.....G..A...GT..T..... 600
R19 C..TATTC..GGA...TG...AA.....T.....G..A...GG..A..T..... 580
S32 C..TATTC..GGA...GTG...AA...A.....G.....CG...C...T..... 424
S1 C..TCTG..AG.....GTG...AA...A.....G.....GG...C...G..... 577
S10 C..TCT...AG..A.....GC..... 577
S31 C..TCT...TA.....GC..... 427
R4  AGCGATTCAGCGCAGTGGATCAGGSAACAGTTTCACCCCTGACAAATCAGCAACCTTCAGAAATGAGGATGTGCGCTGACTATTACTGCTGGGTTCAGATGG 677
R7  .....T.....A.....A...T..TA 677
R18 .....T.....A.....A..C---AGT..AA 674
RE18 .....T.....AA..CAGTCC..A..AA..A.....CA.....A..C---A..G..CA 697
R6  .....T.....AA..CAGTCC..A..AA..A.....CA.....A..C---A..G..CA 677
R19 .....T.....A.....AA..CAGTCC..A..AA..A.....CA.....A..C---A..G..CA 677
S32 .....T.....A.....AA..CAGTCC..A..AA..A.....CA.....A..C---A..G..TC 521
S1 .....T.....A.....T..... 677
S10 .....T.....A.....T..TA 677
S31 .....T.....T..A.. 527
R4  CAGCTACCCTCGA---TCTTGGCTTCGGTAAAGGAACCAAGCTCAGACTGAGTAAGTAACCTCAATCCTCCGTTATTAACCCCTTCAATACTGAAACTTT 774
R7  .....C..T..... 774
R18 .....C...T.....T..... 768
RE18 .....T...T.....A.....T..... 768
R6  .....T..C..T.....A.....T..... 791
R19 .....T..T.....A...T.....GA---C..... 765
S32 .....T...T...T...//A..A..A.....T.....T.....G..NN.....T..... 765
S1 .....T...T...//GT..A.....A.....G..... 765
S10 .....C..T...//GT.....A.....T..... 765
S31 .....T...//GT..... 765
R4  CTCAAACACAAATGCTGAATTTGCAAGGGGTTACTGGGAGCTGCAGTGAATGAAATGGTTGGTTGAGGAGCGCTGTGCTAAAAGGGGGCC 867
R7  ..... 867
R18 .....A.....T.....A.....C..... 861
RE18 .....A.....A.....A.....N..C..... 861
R6  .....A.....T.....A.....A.....C..... 884
R19 .....A.....T.....A.....A.....C..... 857
S1 .....G.....T.....C..... 857
S10 .....G.....A.....C..A.....CC..... 857

```

Figure 3. Nucleotide sequences of NS4 genes containing the leader, V, and J exons are compared with the first sequence, R4. Dots indicate identity, dashes indicate gaps. R4, R7, R18, RE18, R6, and R19 are germline rearranged as VJ. S1, S10, S31, and S32 are not joined, and their V gene segments are shown up to but not including the RSS. Slashes indicate that the intervening sequence and the RSS on either side are left out. These sequences recommence with the coding region of the J gene segments and the 3' noncoding sequence. Regulatory signals as the octamer and splice sites are bold-faced, and their locations are based on homology with other vertebrate L chain genes. The location of the 3' splice signal of J is derived from NS4 full-length cDNA sequence, where the amino acid preceding the first residue of the constant region is arginine (reference 15), so that the last nucleotide in J is adenylyl. The coding region is underlined; the leader sequence is split by an intron. The beginning of the V region is indicated by an asterisk over the first base. The division of the V region into FR and CDR is indicated (reference 19). The two PstI sites (CTGCAG) referred to in the Fig. 2 legend and text are indicated by bold dots.

estimated that there were at least five independent rearrangement events (Fig. 6, arrows) in the evolutionary lineage giving rise to the extant nurse shark genes. Accordingly, because S1, S31, SE7, and S10 are not joined, it is most parsimonious to infer that the ancestor at node A and indeed at all internal nodes of the tree, with the possible ex-

ception of the ancestor of R6 and R19, must have been nonrearranged. For instance, if R7 and R4 were closest relatives, we would not have been able to exclude the possibility that they shared a rearranged ancestral gene, but as it is clear that S31, a split gene, is the closest relative of R4, R7 and R4 arose from independent joining events.

	Leader	FR1	CDR1	FR2
R4	MISHIQLIWLAFVCVAGISGDITMTQ	-SPPVLSVGLGQTATITCTASKSIYSYLAWYQQREGQKPSLLIY		
R7	Q..SND	
R18	Q..S	
RE18	I.....I.N.....	ITYSG.V.....Q	
R19	ST.N.FL.....	A..YM.ITYSG.VN.....*	
S32	I.N.F.....R.....	TTYSD.CVN.....E	
S1	Q.VS	
S10	Q..SN	
S31	Q..N	

	CDR2	FR3	CDR3	FR4
R4	YATNRYTGVSERFSGSGSGTFTLTISNVQNEVDVADYCLGADGSGYPLAFGKGTGLRL			
R7	QS..YY..SS	
R18	IQ--YNSSPY	
RE18	E..K.F.....Y.....	QQ--Y.SFPY	
R19	E..R.F.....KQSNKY.....	QQ--HSF/T.....R	
S32	D..T.F.....*R..KQSNKY.....	QQ--..SFP	
S1	G.....D.....	Q..Y..S	
S10	V.....	QS..YY..SS	
S31	V.....	YS..S	

Figure 4. Derived amino acid sequences of NS4 genes. The amino acid sequences of most of the cloned NS4 genes shown in Fig. 2 are compared with R4. Dots indicate identity, dashes indicate gaps. R19 and S32 contain stop codons (asterisks), and one nucleotide was left out of CDR3 in R19 in order to obtain a reading frame that is comparable to the other sequences (indicated by a hyphen). R6 carried too many disruptions and is not shown.

When Did the Germline Recombination Occur? The most closely related pairs of joined and split genes are R7/S10 and R4/S31 (Fig. 3). The similarity of germline-rearranged genes to the split ones suggests that R4, for instance, may have arisen from a gene cluster similar to S31. Dating the divergence of R4 from S31 would give the earliest time at which the R4 germline rearrangement could have occurred. The rooting of the NS4 phylogeny is described in Materials and Methods, and the R4/S31 divergence point was calculated as a proportion of total branch length (0.129/0.134) to S31 from the species divergence point (node B) at 180 Myr BP (27); that is, 180 Myr \times [180 Myr (0.129/0.134)], 7 Myr BP. Thus, R4 is estimated to have diverged from S31 at least 7 Myr BP, and the recombination event creating the R4 joined gene occurred some time between 0 and 7 Myr BP. The estimated dates of the other rearranged genes are 11 Myr BP for R7/S10 and 49 Myr BP for R18/S10. The earliest date of origin of RE18 was measured as a fraction of the length between RE18 terminus and node B, which was found to be 38 Myr BP.

Molecular Events in Germline Joining. Fig. 7 shows two possibilities (a and b) for the recombination events generating the R7 joint. In both cases, the 3' flank of the R7 gene segment coding region is assumed to have been the same length as in S10, and the 5' flank of the R7 J gene segment is assumed to have been the same length as other clade I NS4 J gene segments so far analyzed (Fig. 7, bottom).

One possibility (a) is that the R7 J 5' flank was different from the S10 J flank in the first two bases. The recombination proceeded by cleavage of DNA adjacent to the RSS, and joining of the blunt coding ends was without any nucleotide addition or loss. The second possibility (b) is that after removal of the intervening sequence, hairpin coding ends were formed. Single-strand nicking on both the flanks occurred at least two nucleotides from the ends, and subsequent processing resulted in a two-base P region at the V end and a two-base deletion of the J gene end.

The second possibility strongly suggests that the germline

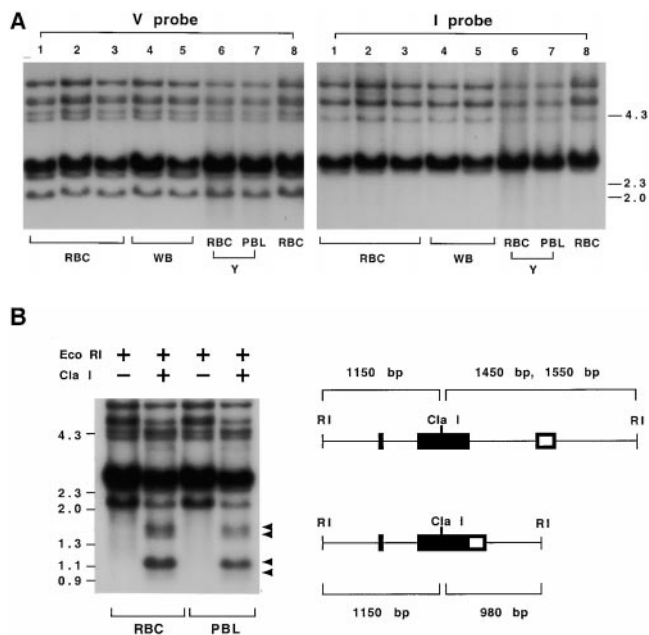


Figure 5. Germline origin of rearranged NS4 genes. (A) EcoRI-digested DNA from seven wild nurse sharks hybridized with V and I probes. Lanes 1–3, DNA from RBC fraction of ficolled blood of three individuals; lanes 4 and 5, DNA from whole blood (WB) of two sharks; lanes 6 and 7, DNA from erythrocyte (RBCs) fraction (lane 6) and PBL fraction (lane 7) of ficolled whole blood from shark Y. The genomic DNA libraries were generated from these samples (see Materials and Methods). Lane 8, DNA from erythrocytes from shark B. The DNA samples were digested with EcoRI, electrophoresed together on one 1% agarose gel, and transferred to a nylon filter; the filter was cut in two and one half was hybridized to probe V, the other to probe I (see legend to Fig. 2). Arrows indicate bands at 4.6, 2.4, and 2.1 kb that hybridize to V but not to I. The band at 5.4 kb is variable in hybridization to I probe among individuals. When the V and I films were superimposed, the bands at 2.4 kb detected by V probe are distinct from the faint bands at \sim 2.5 kb detected by I probe. (B) Identification of R4/R7 sequences in genomic DNA. The DNA from shark Y obtained from erythrocytes (lanes 1 and 2) and from PBLs (lanes 3 and 4) was digested with EcoRI and ClaI (lanes 2 and 4) or EcoRI alone (lanes 1 and 3). Hybridization with V probe showed that some of the EcoRI fragments containing NS4 V genes carried ClaI sites as well. Arrows point to ClaI and EcoRI-digested fragments that hybridize with V probe at 1,550, 1,450, 1,150, and 980 bp. The markers used are λ HindIII and ϕ X174 HaeIII, at 4.36, 2.3, 2.0, 1.35, 1.08, and 0.872 kb. The interpretation of the digestion patterns is shown in the diagram. Top, split NS4 genes at 2.6 and 2.7 kb such as S1 and S31 with ClaI site in CDR2. Bottom, joined NS4 genes at 2.1 kb such as R4 and R7 with ClaI site in CDR2. Joined NS4 genes such as R6 and R19 at 2.1 kb are not expected to be affected by ClaI, so the 2.1-kb signal in lanes 2 and 4 is reduced but not abolished.

recombination occurred by V(D)J rearrangement mechanisms described in lymphocytes, as the coding ends would have undergone an intermediate hairpin stage to achieve the R7 joint. By comparison, P region is also found occasionally in NS4 cDNA sequences cloned from shark Y. In the examples shown, cDNA 6 and cDNA 23, which are similar in CDR3 to S10, both contain P and probably N regions. Even if the cDNAs were rearrangements to a J segment different from the one found closest to S10, some of the nucleotides in CDR3 still appear to be nontemplated.

Analysis of the joints of R4, R18, RE18, R6, and R19

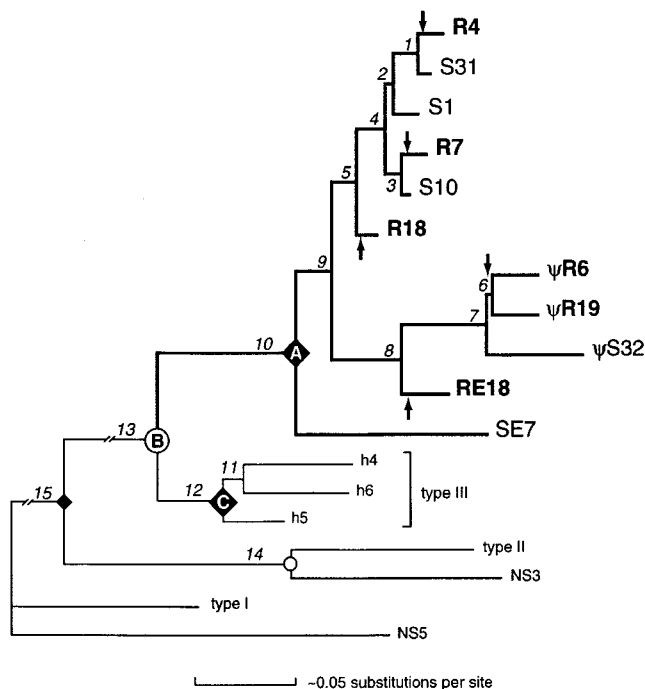


Figure 6. Phylogenetic tree of shark L chain V sequences. The relationships among the nurse shark NS4 family of genes are emphasized by bold branches. Other L chain types from nurse shark (NS3 and NS5) and horned shark (types I–III) were included in the analysis and used to root the tree. Branch lengths were estimated by likelihood using the HKY85+I+ Γ model (reference 23), correcting for invariance and rate variation across sites as detailed in Materials and Methods. Nodes representing presumed speciation events are indicated as circles; e.g., node B corresponds to the divergence between all nurse shark NS4 genes and all horned shark type III genes. Nodes representing the most recent common ancestor of all members within a particular gene family are designated by ◆; e.g., all the NS4 genes derived from a common ancestral gene at node A and all the type III genes derived from a common ancestral gene at node C. Arrows designate branches in which germline rearrangements in NS4 genes most likely occurred. Lengths of internal branches (numbered segments) are approximately proportional to the estimated accumulation of nucleotide substitutions per site (scale at bottom), except for branches 13 and 15, which are ~10 times as long and are truncated in the figure. The most strongly supported branches are 1, 4, 7, 8, 10, 12, 13, 14, and 15. For example, in one of many parsimony jackknife analyses of only the coding regions (assuming transversions were weighted twice that of transitions, ignoring alignment gaps, with 2,000 replications of a random selection of 50% of the nucleotide characters), each branch appeared in the following percentages of replicates (heuristic searches with 1,000 random taxon additions each): 1, 88%; 2, 61%; 3, 71%; 4, 82%; 5, 66%; 6, 51%; 7, 97%; 8, 86%; 9, 61%; 10, 95%; 11, 31%; 12, 98%; 13, 100%; 14, 100%; and 15, 100%. In a similar analysis that included noncoding regions (and excluded taxa for which these regions were not sequenced), jackknife values were (2,000 replications of exhaustive searches): 1+2, 37%; 3, 47%; 4, 75%; 5, 50%; 6, 98%; and 7+8, 94%. Horned shark cDNAs h4, h5, and h6 (available from EMBL/GenBank/DBJ under accession nos. L25561, L25562, and L25563, respectively) were isolated by cross-hybridization with NS4 sequence. Based on separate phylogenetic analyses of V and C regions, type III was identified as a homologue of NS4 (reference 13); this conclusion is also supported by the current phylogram. All available type III sequences grouped strongly together, as did all the representative NS4 sequences (germline-rearranged R7, R4, R18, R6, R19, and RE18, and split S1, S10, S31, S32, and SE7), suggesting gene diversification after the species diverged. Pseudogenes are indicated. Other sequences are from different L chain types in horned shark (type I, X15316; type II, L25560) and nurse shark (NS3 [reference 16], NS5 [reference 39]). SE7 is the most basally divergent member of the NS4 gene family identified to date, and its sequence is not shown in Fig. 3.

V gene segment	N J gene segment
R7	
S10	TGTCAGAGTGCATATTACAGCTCCTCCGA GTTTGCGTTC
R7 a.	TGTCAGAGTGCATATTACAGCTCCTCCGA TCTTGCCTTC
R7 b.	TGTCAGAGTGCATATTACAGCTCCTCCGATC TTGCGTTC
cDNA	
S10	TGTCAGAGTGCATATTACAGCTCCTCCGA GTTTGCGTTC
cDNA23	TGTCAGAGTGCATATTACAGCTCCTCCGATC A TAGGTTC
CDNA6	TGTCAGAGTGCATATTACAGCTCCTCCGATCCG CC TAGGTTC
R4	
S1	TGTCAGGGTGCATATGGCAGCTACTCCGA GTTTGCGTTC
S31	TGTCAGGGTGCATATAGCAGCTACTCCGA GTTTGCGTTC
R4 a.	TGTCAGGGTGCAGATGGCAGCTACCCCGA TCTTGCCTTC
R4 b.	TGTCAGGGTGCAGATGGCAGCTACCCCGATC TTGCGTTC
RE18	
S3	TGTCAGCAAAGTAGGACCTACCCCTGGG GTTTACGTTTC
RE18 a.	TGTCAGCAATAGCAGCTTCCC GTATGCGTTC
RE18 b.	TGTCAGCAATAGCAGCTTCCCGTA TGCCTTC
R6, R19, R18	
S5*	TGTCACAAGGTTACAGCTGGTCTAA GCATACGTTTC
S3	TGTCACAAGGTTAGGACCTACCCCTGGG GTTTACGTTTC
S32*	TGTCACAAGGTTCCAGCTTCCCCTA ACATACGTTTC
R6 a.	TGTCACAAGGTTACAGCTTCCCCTTTA CGTTC
R6 b.	TGTCACAAGGTTACAGCTTCCCCTTT ACCTTC
R6 c.	TGTCACAAGGTTACAGCTTCCCCTT TTTACGTTTC
R19 a.	TGTCACAAGGTTACAGCTTCTCTTA CGTTC
R19 b.	TGTCACAAGGTTACAGCTTCTCTT TTAGGTTTC
R18 a.	TGTCACAAGGTTAAGCAGCTCCCCTA TGGCTTC
R18 b.	TGTCACAAGGTTAAGCAGCTCCCCT ATGCGTTC
Coding flanks of various split NS4 gene segments	
Clade I	
S1	TGTCAGGGTGCATATGGCAGCTACTCCGA GTTAGCGTTC
S10	TGTCAGAGTGCATATTACAGCTCCTCCGA GTTTGCGTTC
S37	TGTCAGGGTGCATATAGCAGCTACTCCGA GTTTGCGTTC
S31	TGTCAGGGTGCATATAGCAGCTACTCCGA GTTTGCGTTC
SE32	TGTCAGAGTGCATATGGCAGCTACTCCGA GTTTGCGTTC
S5*	TGTCACAAGGTTACAGCTGGTCTAA GCATACGTTTC
Clade II	
S32*	TGTCACAAGGTTCCAGCTTCCCCTA ACATACGTTTC
Clade III	
S3	TGTCAGCAAAGTAGGACCTACCCCTGGG GTTTACGTTTC
S11	TGTCAGCAAAGTAGGACCTACCCCTGG GTTTACGTTTC
S15*	TGTCAGCAAAGTGAGAGCTTCCCCTGA GTTTACGTTTC
SE7	TGTCAGCAAAGTGAGAGCTTCCCTTGG CCTTACGTTTC

Figure 7. Derivation of the joints in NS4 germline-rearranged genes. The V and J flanks of split genes are shown without RSS. The possible processing events producing the joints of germline-joined R7, R4, RE18, R6, R19, and R18 are shown aligned with the most similar non-joined NS4 coding ends isolated in these studies. Assuming that the split ancestor of R7 resembled S10, the R7 joint could have been produced by (a) cleavage at the RSS and relegation of blunt ends, or (b) cleavage at the RSS, formation of hairpin at V and J ends, single-stranded nick occurring two or more nucleotides from end, processing resulting on P region (underlined), and deletion of two bases from the J flank. Pathway R7 (a) does not preclude hairpin formation, but there is no processing evidence for it. Under cDNA, germline split S10 is compared with cDNA sequences obtained by RT-PCR from shark Y that share similar CDR3. The analysis of the cDNA joints suggest P region in both sequences. The sequences in the cDNA joints that do not appear to belong to either V or J may be N region. R4 is similar to R7. RE18 joint could have resulted from deletion in the V coding end (a), or J coding end (b), or both (c). The flanks of nonjoined ancestor of R6, R19, and R18 could have resembled S32, in which case deletion only from J end occurred during recombination (a). In all other possibilities, nucleotides were removed from either end. Examples of coding flanks from NS4 split genes are included for comparison; the RSS resemble those from other vertebrates. Clade I refers to the cluster of genes including R4, R7, and R18, clade II to that including R6, R19, and RE18 (see Fig. 6); clade III genes are characterized by divergent FR3 and CDR3 of 24 bp; complete sequences are not shown. Asterisk indicates split gene is pseudogene. S32 is shown in Figs. 3 and 4; S15 has TGA in FR2; S5 sequence is not shown but carries TGA in FR3, a GT–AT splice signal combination in leader intron, and noncanonical RSS heptamer sequences.

also suggest that the recombination events may have entailed modification or removal of nucleotides from either flank similar to R7 (Fig. 7). The most obvious difference from cDNA is that in somatically rearranged sequences, there are stretches of several bases that may be N region. Therefore, germline rearrangement of the NS4 L chain gene segments appears to occur much as it does in lymphocytes, and in the R sequences there is little evidence of TdT activity.

Although our phylogenetic analyses pointed to the NS4 joined genes being products of ancestral split genes rather than the other way around, additional supporting evidence comes from examination of the coding sequence on either side of the RSS in the split genes, showing no indication of the characteristic integration site duplication of 3–5 bp that occurs in V(D)J-mediated RSS insertions (2, 3, 9).

Germline-rearranged NS4 in Shark Population. The germline-rearranged NS4 genes R4 and R7 carry a ClaI site in CDR2, and in the CDR3 of R4 there is also a BsgI site. Erythrocyte DNA from seven outbred nurse sharks was subjected to PCR using the NSL and JL5 primers. As the 550-bp fragments amplify far more efficiently than the 1-kb non-rearranged genes (16), after 40 cycles almost the only PCR product strongly hybridizing to V probe was that from the rearranged genes and this was found in all 7 samples (not shown). When digested with ClaI (402 and 156 bp), or ClaI and BsgI (402, 131, and 25 bp), all samples contained a subpopulation of DNA with ClaI sites, and at least four of the samples contained PCR fragments with restriction endonuclease sites similar to R4. R18, another joined gene defined by a unique EcoRV site in FR3 in combination with the absence of a universal AclI site in FR3, was also detected in all shark samples. Moreover, PCR-generated sequences identical to genes R4 and R7 were cloned from shark A and genes R4 and R18 cloned from shark B.

All the animals in Fig. 5 have also been analyzed for polymorphism at the MHC class I genes, and no two samples show identical RFLP (data not shown). The levels of MHC class I and class II polymorphism in this population of sharks are as high as those found in mouse and human populations (28; and Flajnik, M.F., unpublished data). Thus, the presence of germline-joined NS4 genes in each animal tested is not the result of a recent bottleneck. In combination with results shown in Fig. 5 A, we conclude that most of the germline-rearranged genes are widespread in the nurse shark population in the Florida Keys, and that the recombination events that generated them do not occur frequently in the germ cells.

Expression of Germline-joined Genes. R4 and R18 are distinguishable from other NS4 sequences by the restriction enzyme analyses described above. RT-PCR was performed on PBL RNA from shark Y, using primers NSL and JL5, and no significant amounts of either R4 or R18 sequence could be detected in PCR products of the expected cDNA size of ~350 bp. 40 cDNA joints were examined, and none matched those of the R genes. However, analyses of spleen RNA from neonatal shark pups showed the presence of both; this was confirmed by isolation of these se-

quences from pup spleen and pup epigonal cDNA libraries (data not shown).

Discussion

The Origin of Germline-joined Ig Genes. When germline-rearranged Ig genes were discovered in the horned shark (12), it was initially not clear whether these sequences were the products of recombination in germ cells, or an archaic form of Ig genes from whence was derived gene segments capable of recombination. Litman and coworkers subsequently demonstrated that while the horned shark L chain type I genes were in germline-split configuration, the homologue in skates was entirely germline joined (13, 29). It appeared probable, but not conclusive, that joining or splitting event(s) occurred after divergence of *Batoidea*, at least 200 Myr BP (27). If the joined genes originated from split ones, certain questions arise. Is the recombination mechanism in germ cells similar to that described in lymphocytes? Does Ig gene recombination occur frequently, and do these joined genes have a selected function? If the process is really mediated by RAG enzymes present in germ cells, is the excised DNA reinserted, allowing the chance of creating another split gene?

While the mixture of rearranged and nonrearranged genes has been reported for horned shark H chain genes (12), such a combination in one L chain type is the first thus described. Unlike horned shark H chain genes, the joined and split NS4 V genes are very closely related in sequence (up to 99% identity), and phylogenetic analysis was possible. An ancient gene duplication giving rise to three NS4 clades occurred within the nurse shark lineage soon after species divergence. Joined genes appeared in two clades, and except for R6/R19, a closer relationship in each case to split genes suggests five independent recombination events.

In this study, the six nurse shark joined sequences identified are found to be the results of rearrangement in the germline occurring at several different points in time. The most closely related joined and split genes, S31 and R4, diverged 7 Myr BP and the recombination events creating the joined R4 gene therefore occurred between 0 and 7 Myr BP. The RAG transposon is thought to have entered the genome of jawed vertebrates >400 Myr BP (2, 5); our calculations suggest that RAG-like activity, as evidenced by recent germline-joining, may be ongoing.

Rearrangement in Germ Cells by RAG-like Enzymes. The close identity of S10 and R7, especially at CDR3, also permitted an analysis of the R7 joint, under the assumption that it arose from a gene similar to S10. Our analyses of the R7 joint and others suggest that the joining process may have involved the RAG recombinase enzymes. RAG studies performed in the mammalian systems have shown that RAG-1 and RAG-2 recognize the RSS and mediated the cleavage of DNA adjacent to the RSS (30–32); the resulting double-stranded breaks on the coding sides of the RSS formed hairpin structures (33). RAG proteins participated in nicking the hairpins (34), and the opened coding ends lost or acquired additional sequence before joining, gener-

ating N and P regions (35, 36). Whereas N region is generated by TdT, P region is acquired as a result of these RAG activities. The possibility that some nucleotides in the R4 and R7 joints are P region suggests a hairpin intermediate like that found in the RAG-mediated pathway. The deletion of nucleotides, which occurred in the shark 5' flank of J segment, is also a characteristic of V(D)J rearrangement. In the joint of each joined gene, no part of the sequence is attributable to the shark RSS (classical heptamers and nonamers); they have been completely excised as if recognized by the recombinase. These observations, and similar processing of the coding ends in the shark cDNA, support the idea that the genes were joined and the intervening DNA excised by recombinase enzymes, as in lymphocytes (37).

Genomic Southern analyses of a second shark B (Fig. 5 A, lane 8) suggest that it too carried NS4 joined genes on EcoRI-digested fragments of 4.6 kb (RE18), 2.4 kb (R6/R19), and 2.1 kb (R4/R7/R18). This observation and the cloning of identical R4, R7, and R18 sequences from other sharks suggest that Ig germline gene rearrangement is not a frequent event, but that once established in the population, the presence of the joined gene is stable.

Antigen Receptors without Rearrangement. In cases where all the identified genes of an isotype are joined, as are L chain genes in sandbar shark (38), in skate (13, 29), and in nurse shark NS3 (16, 39), cDNA sequences have been isolated. However, in the case of mixed split and joined genes of one isotype, as in H chain genes in the horned shark, expression of the germline-joined H chain genes was not detected (40), and this cast doubt on the biological significance of the joined and partially joined H chain genes.

In the NS4 L chain type, four of the six joined genes are rearranged in frame and are potentially functional; all of the joined genes have the same octamer motif and other regulatory sequences as the split genes. Like Litman and co-workers, we were not able to detect significant amounts of germline-joined Ig message in adults, although the major L chain mRNA in nurse shark is the NS4 type (39). However, we also tested nurse shark pups and have detected expression of the two germline-joined NS4 genes that we were screening for, R4 and R18.

Do the joined genes have a special function? The pups not only express germline-joined NS4 genes but also a germline-joined H chain that constitutes more than half of the H chains in serum Ig. Adult nurse sharks, in contrast, express little such message (Flajnik, M.F., unpublished data). The coupling of the germline-joined NS4 L chain and the novel H chain isotype in shark pup lymphocytes presents the possibility that some B cells could differentiate and express antigen receptors independently of RAG activity.

The germline joining of particular Ig gene segments may have occurred fortuitously, but once established in the population, such genes may have evolved specialized function, as suggested by the similarly restricted expression of the newly discovered joined H chain. Seen from this viewpoint, a rejoined Ig gene is not merely a sequence restricting antibody diversity, but can evolve potentially as a novel receptor no longer tied to the rearrangement process.

Our hypothesis does not yet have a direct example, but the case of L chain surrogate V-pre-B is suggestive. From sequence homology and its ability to combine with H chain, V-pre-B is a homologue of L chains (41). It could have been an ancient rejoined gene, although there is little sequence homology with J (or RSS) in its extended 3' end, but it is clearly a V region sequence that no longer has the function in common with other antibody V regions, and because rearrangement is not needed for its role, it is independent of RAG. In other words, V-pre-B is an Ig sequence that no longer requires rearrangement, and its function has likewise evolved away conventional L chain.

A different example of specialized function may be the role evolved by the H chain pseudogenes in the chicken; these are germline-fused, VD sequences that serve as templates for H chain gene conversion. This is discussed in the following section.

Role of RAG in Evolution of Antigen Receptors. We propose that germ cell recombination of Ig genes are infrequent events, but that they may still occur, altering the germline content in nurse sharks. We suggest that the RAG proteins are responsible for generating germline-joined Ig genes. Are there RAG enzymes in the oocyte or sperm? RAG-1 was detected by RT-PCR in zebrafish ovary (42). RAG-2 was found in *Xenopus* oocytes by Northern blotting (43). However, the problem in the fish studies is that the origin of the RAG signal cannot be conclusively identified as being not from blood-borne cells; only in situ hybridization will establish this reliably.

Does germline rearrangement occur whenever the Ig genes are organized in clusters? In bony fishes, the H chain gene organization is like mammals but the L chain genes are in clusters. In catfish, where there are an estimated 60 L chain clusters, no rearranged genes have been amplified by PCR (Hsu, E., unpublished results), but they may just have escaped detection. It is also possible that the inverted orientation of the catfish L chain V genes (44, 45) hinders the success of the rare joining event, as such a rearrangement would require reintegration of the excised DNA by ligation at two sites. However, a second duplicated H chain locus with nonfunctional C region exons was found in catfish, and the H chain gene segments, which are in one transcriptional orientation, are rearranged in the germline as VDJ (46).

In the chicken, antibody diversification is largely created by rearrangement of a single functional V gene at the H or L chain locus, followed by gene conversion modifications, using sequence from multiple upstream pseudogenes as templates (47–50). The H chain pseudogenes are discernibly germline-joined VD sequences. Our interpretation would be that an original VD sequence could have been generated by RAG-mediated recombination in the germ cells; the fused VD underwent gene duplication, coevolving with the dominance of the gene conversion mechanism. A supporting observation is that while rearrangements of VD alone are not normally found in lymphocytes, they are present as such in the germline in shark (12). The existence of VD and VDD joined shark genes suggest that

recombination of Ig gene segments in germ cells are distinguishable from the differentiation signals of lymphocytes.

The current notion is that the RAG proteins can perform like transposases and may have initiated the development of the adaptive immune system in early vertebrates by splitting V genes with transposable elements (2, 5). We suggest that with the discovery of recently rejoined Ig L chain genes and the restricted expression of these and a germline-joined H chain gene, RAG in some vertebrates could be a force in evolution still.

In memory of Charles M. Steinberg. We wish to thank Drs. Chris Roman and Louis Du Pasquier for their comments and criticism, and Dr. John Maisey of the American Museum of Natural History for clarifying elasmobranch phylogeny. We also thank Dr. Churchill McKinney for her help with blood sampling.

This work was supported in part by National Science Foundation grant MCB 9723203 and National Institutes of Health grant RR06603.

Submitted: 13 December 1999

Revised: 22 February 2000

Accepted: 23 February 2000

References

1. Du Pasquier, L., and M.F. Flajnik. 1999. Origin and evolution of the vertebrate immune system. *In* Fundamental Immunology. 4th ed. W.E. Paul, editor. Lippincott-Raven, Philadelphia. 605–650.
2. Agrawal, A., Q.M. Eastman, and D.G. Schatz. 1998. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature*. 394:744–751.
3. Hiom, K., M. Melek, and M. Gellert. 1998. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell*. 94:463–470.
4. Sakano, H., K. Hüppi, G. Heinrich, and S. Tonegawa. 1979. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature*. 280:288–294.
5. 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.
6. Amemyia, C.T., and G.W. Litman. 1990. Complete nucleotide sequence of an immunoglobulin heavy chain gene and analysis of immunoglobulin gene organization in a primitive teleost species. *Proc. Natl. Acad. Sci. USA*. 87:811–815.
7. Ghaffari, S.H., and C.J. Lobb. 1992. Organization of immunoglobulin heavy chain constant and joining region genes in the channel catfish. *Mol. Immunol.* 29:151–159.
8. Daggfeldt, A., E. Bengtén, and L. Pilström. 1993. A cluster type organization of the loci of the immunoglobulin light chain in Atlantic cod (*Gadus morhua* L.) and rainbow trout (*Oncorhynchus mykiss* Walbaum) indicated by nucleotide sequences of cDNAs and hybridization analysis. *Immunogenetics*. 38:199–209.
9. Hinds, K.R., and G.W. Litman. 1986. Major reorganization of immunoglobulin V_H segmental elements during vertebrate evolution. *Nature*. 320:546–549.
10. Hohman, V.S., D.B. Schuchman, S.F. Schluter, and J.J. Marchalonis. 1989. Genomic clone for the sandbar shark I light chain: generation of diversity in the absence of gene rearrangement. *Proc. Natl. Acad. Sci. USA*. 90:9882–9886.
11. Shablott, M.J., and G.W. Litman. 1989. Genomic organization and sequences of immunoglobulin light chains in a primitive vertebrate suggest coevolution of immunoglobulin gene organization. *EMBO (Eur. Mol. Biol. Organ.) J.* 8:3733–3739.
12. Kokubu, F., R. Litman, M.J. Shablott, K. Hinds, and G.W. Litman. 1988. Diverse organization of immunoglobulin V_H gene loci in a primitive vertebrate. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:3413–3422.
13. Rast, J.P., M. Anderson, T. Ota, R.T. Litman, M. Margittai, M.J. Shablott, and G.W. Litman. 1994. Immunoglobulin light chain class multiplicity and alternative organizational forms in early vertebrate phylogeny. *Immunogenetics*. 40:83–99.
14. Yoder, J.A., and G.W. Litman. 2000. Immune-type diversity in the absence of somatic rearrangement. *Curr. Top. Microbiol. Immunol.* 248:271–282.
15. 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 κ light chains: implications for the evolution of light chains. *Proc. Natl. Acad. Sci. USA*. 90:10603–10607.
16. Lee, S.S., A. Greenberg, and E. Hsu. 2000. Evolution and somatic diversification of immunoglobulin light chains. *Curr. Top. Microbiol. Immunol.* 248:285–300.
17. Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*. 74:5463–5467.
18. Wilson, M., E. Hsu, A. Marcuz, M. Courtet, L. Du Pasquier, and C. Steinberg. 1992. What limits affinity maturation of antibodies in *Xenopus*—the rate of somatic mutation or the ability to select mutants? *EMBO (Eur. Mol. Biol. Organ.) J.* 11:4337–4347.
19. Kabat, E.A., T.T. Wu, H.M. Perry, K.S. Gottesman, and C. Foeller. 1991. Sequences of Proteins of Immunological Interest. 5th ed. US Department of Health and Human Services, Washington, D.C. xiii–xvii.
20. Lee, A., S. Desravines, and E. Hsu. 1993. IgH diversity in an individual with only one million B lymphocytes. *Dev. Immunol.* 3:211–222.
21. Posada, D., and K.A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics*. 14:817–818.
22. Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16:111–120.
23. Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 21:160–174.
24. Bremer, K. 1994. Branch support and tree stability. *Cladistics*. 10:295–304.
25. Kishino, H., and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in *Hominoidae*. *J. Mol. Evol.* 29:170–179.
26. Templeton, A.R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the humans and apes. *Evolution*. 37:221–244.
27. Cappetta, H., C. Duffin, and J. Zidek. 1993. Chondrichthyes. *In* The Fossil Record. 2nd ed. M.J. Benton, editor. Chapman and Hall, London. 593–609.
28. Kasahara, M., E.C. McKinney, M.F. Flajnik, and T. Ishibashi. 1993. The evolutionary origin of the MHC: polymorphism of class IIA genes in the cartilaginous fish. *Eur. J. Im-*

- munol.* 23:2160–2165.
29. Anderson, M.K., M.J. Shablott, R.T. Litman, and G.W. Litman. 1995. 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. *J. Exp. Med.* 182:109–119.
 30. Schatz, D.G., M.A. Oettinger, and D. Baltimore. 1989. The V(D)J recombination activating gene, RAG-1. *Cell.* 59: 1035–1048.
 31. Oettinger, M.A., D.G. Schatz, C. Gorka, and D. Baltimore. 1990. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science.* 248:1517–1523.
 32. McBlane, J.F., D.C. van Gent, D.A. Ramsden, C. Romeo, C.A. Cuomo, M. Gellert, and M.A. Oettinger. 1995. Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell.* 83:387–395.
 33. Roth, D.B., J.P. Menetski, P.B. Nakajima, M.J. Bosma, and M. Gellert. 1992. V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in scid mouse thymocytes. *Cell.* 70:1–9.
 34. Shockett, P.E., and D.G. Schatz. 1999. DNA hairpin opening mediated by the RAG1 and RAG2 proteins. *Mol. Cell. Biol.* 9:4159–4166.
 35. Alt, F.W., and D. Baltimore. 1982. Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions. *Proc. Natl. Acad. Sci. USA.* 79:4118–4122.
 36. Lafaille, J.J., A. DeCloux, M. Bonneville, Y. Takagaki, and S. Tonegawa. 1989. Junctional sequences of T cell receptor genes: implications for $\alpha\beta$ T cell lineages and for a novel intermediate of V-(D)-J joining. *Cell.* 59:859–870.
 37. Lewis, S.M. 1994. The mechanism of V(D)J joining: lessons from molecular, immunological, and comparative studies. *Adv. Immunol.* 56:27–150.
 38. Schluter, S.F., V.S. Hohman, A.B. Edmundson, and J.J. Marchalonis. 1989. Evolution of immunoglobulin light chains: cDNA clones specifying sandbar shark constant regions. *Proc. Natl. Acad. Sci. USA.* 86:9961–9965.
 39. Greenberg, A.S. 1994. Evolution of the antigen receptor family. Ph.D. thesis. The University of Miami, Miami. 313 pp.
 40. Litman, G.W., J.P. Rast, M.J. Shablott, R.N. Haire, M. Hulst, W. Roess, and R.T. Litman. 1993. Phylogenetic diversification of immunoglobulin genes and the antibody repertoire. *Mol. Biol. Evol.* 10:60–72.
 41. Kudo, A., and F. Melchers. 1987. A second gene, Vpre-B, in the lambda 5 locus of the mouse, which appears to be selectively expressed in pre-B lymphocytes. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:2267–2272.
 42. Willett, C.E., J.J. Cherry, and L.A. Steiner. 1997. Characterization and expression of the recombination activating genes (rag1 and rag2) of zebrafish. *Immunogenetics.* 45:394–404.
 43. Greenhalgh, P., C.E. Oleson, and L.A. Steiner. 1993. Characterization and expression of recombination activating genes (RAG-1 and RAG-2) in *Xenopus laevis*. *J. Immunol.* 151: 3100–3110.
 44. Ghaffari, S.H., and C.J. Lobb. 1993. Structure and genomic organization of immunoglobulin light chain in the channel catfish. *J. Immunol.* 151:6900–6912.
 45. Ghaffari, S.H., and C.J. Lobb. 1997. Structure and genomic organization of a second class of immunoglobulin light chain genes in the channel catfish. *J. Immunol.* 159:250–258.
 46. Ghaffari, S.H., and C.J. Lobb. 1999. Structure and genomic organization of a second cluster of immunoglobulin heavy chain gene segments in the channel catfish. *J. Immunol.* 162: 1519–1529.
 47. 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.
 48. Reynaud, C.A., V. Anquez, H. Grimal, and J.C. Weill. 1987. A hyperconversion mechanism generates the chicken light chain preimmune repertoire. *Cell.* 48:379–388.
 49. 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 on the D region. *Cell.* 40:283–291.
 50. Thompson, C.B., and P.E. Neiman. 1987. Somatic diversification of the chicken immunoglobulin light chain gene is limited to the rearranged variable gene segment. *Cell.* 48: 369–378.