Structure and Genomic Organization of a Second Cluster of Immunoglobulin Heavy Chain Gene Segments in the Channel Catfish

Seyed H. Ghaffari and Craig J. Lobb

*J Immunol* 1999; 162:1519-1529; ;

http://www.jimmunol.org/content/162/3/1519

**References**
This article cites 37 articles, 12 of which you can access for free at:
http://www.jimmunol.org/content/162/3/1519.full#ref-list-1

**Subscription**
Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

**Permissions**
Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

**Email Alerts**
Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts
Structure and Genomic Organization of a Second Cluster of Immunoglobulin Heavy Chain Gene Segments in the Channel Catfish1,2

Seyed H. Ghaifari and Craig J. Lobb3

The structure, organization, and partial sequence of a 25-kb genomic region containing a second cluster of H chain gene segments in the channel catfish (Ictalurus punctatus) has been determined. Multiple V\textsubscript{H} gene segments, representing different V\textsubscript{H} families, are located upstream of a germ-line-joined VDJ. The VDJ segment has a split leader sequence and a single open reading consistent with that expressed in members of the VH1 family. Downstream of the germline-joined VDJ is a single J\textsubscript{H} segment and two pseudogene exons structurally similar to the C\textsubscript{\mu}1 and C\textsubscript{\mu}2 exons of the functional gene. Both pseudogene exons are multiply crippled with RNA splice sites destroyed, and open reading frames are interrupted by termination codons, insertions, and/or deletions. Sequence alignment of a 10.8-kb region within the second H chain cluster with the genomic sequence of the nine J\textsubscript{H} crippled with RNA splice sites destroyed, and open reading frames are interrupted by termination codons, insertions, and/or deletions. Sequence alignment of a 10.8-kb region within the second H chain cluster with the genomic sequence of the nine J\textsubscript{H} segments and the functional C\textsubscript{\mu} within the first H chain gene cluster indicates that the second H chain gene cluster probably arose by a massive duplication event. The J\textsubscript{H} region of the VDJ, the coding and flanking regions of the single J\textsubscript{H} segment, and the pseudogene C\textsubscript{\mu} exons were readily aligned with homologous segments in the first gene cluster. This duplication event may have extended to include the upstream V\textsubscript{H} segments. A member of the Tc1 mariner family of transposable elements is located downstream of the pseudogene C\textsubscript{\mu}2, which suggests that the transposition may have contributed to the evolution of the duplicated C\textsubscript{\mu}. The Journal of Immunology, 1999, 162: 1519–1529.

The evolution of Ig H chain loci during vertebrate phylogeny has been a dynamic process. Although the general structure of the V\textsubscript{H}, D\textsubscript{H}, and J\textsubscript{H} variable region gene segments has been conserved, the genomic organization and utilization patterns of these gene segments have been subject to considerable “phylogenetic experimentation.” Three general patterns of genomic organization have been characterized. The most familiar is that found in humans and mice, in which various functional V\textsubscript{H} segments representing different members of multiple V\textsubscript{H} gene families are located upstream from D\textsubscript{H} segments that are, in turn, located upstream of J\textsubscript{H} segments. These V region segments undergo gene rearrangement processes to form a V-D-J that is subsequently spliced to an H chain C region gene (1–5). A second organizational pattern is represented in chickens. The locus is comprised by multiple V\textsubscript{H} segments upstream of D\textsubscript{H}, J\textsubscript{H}, and C\textsubscript{\mu} segments; however, in this case only a single V\textsubscript{H} segment is functional. The other V\textsubscript{H} segments are pseudogenes, and their contribution to the diversity of the Ig repertoire is manifested only through gene conversion mechanisms (6, 7). A third genomic organization pattern is that represented by early vertebrates such as sharks, where there are multiple clusters of Ig gene segments. Within each cluster, there are V\textsubscript{H}, D\textsubscript{H}, J\textsubscript{H}, and C\textsubscript{\mu} segments, and gene rearrangement appears restricted to the segments within individual clusters (8). Shark V\textsubscript{H} segments can also be grouped into different V\textsubscript{H} families (9).

Studies with the channel catfish have provided insight into the early evolutionary patterns of Ig gene organization and genetic diversity. The genomic organization of H chain gene segments in the catfish, a teleost fish, is different from that represented in sharks. The C\textsubscript{\mu} gene, which encodes the four-domain C region of the predominant serum Ig and Ab of catfish (10, 11), exists as a single genomic copy, as shown in quantitative gene titration experiments using probes derived from the C\textsubscript{\mu}3 and C\textsubscript{\mu}4 exons (12, 13). This general conclusion has been extended to C\textsubscript{\mu} genes in other teleost fish (14–17). In addition, there is now conclusive evidence that V\textsubscript{H} gene families diverged within the phylogeny of bony fish. There are at least seven catfish V\textsubscript{H} gene families that together probably represent >120 different V\textsubscript{H} gene segments (18). The V\textsubscript{H} gene segments representing these different V\textsubscript{H} families are interspersed and closely linked (19). Additional diversity within the catfish H chain V region is attributed to different D\textsubscript{H} as well as J\textsubscript{H} segments. The J\textsubscript{H} locus contains nine functional J\textsubscript{H} segments that are tightly clustered within a 2.2-kb region located immediately upstream from C\textsubscript{\mu} (20, 21). Thus, these combined studies indicate that the genomic organization of H chain genes typically associated with higher vertebrates appears to have had its phylogenetic foundation at the level of the bony fishes.

Southern blot studies using genomic DNA obtained from the nucleated erythrocytes of all individual channel catfish examined to date have indicated the presence of an additional H chain C region gene. This additional gene hybridizes with probes derived from the C\textsubscript{\mu}1 and C\textsubscript{\mu}2, but not with probes derived from either the C\textsubscript{\mu}3 or C\textsubscript{\mu}4 exons. This gene shares higher similarity with C\textsubscript{\mu}1 rather than C\textsubscript{\mu}2, as determined by the relative signal intensities of the genomic fragments observed under stringent or relaxed hybridization conditions (13, 20). In the present study, the locus containing this additional gene has been characterized. The results provide
new insights into the early evolutionary patterns of Ig gene organization.

Materials and Methods

Screening of genomic libraries

The genomic libraries used in the study were constructed in λ DASH II (Stratagene, La Jolla, CA) using high m.w. DNA obtained from the erythrocytes of an individual channel catfish (Ictalurus punctatus) as previously described (20). The libraries were screened with a BstEII-PstI fragment from cDNA clone NG13, which encodes the channel catfish Cμ1 and Cμ2 domains (13). Positive clones identified with the above approach were screened under high or low stringency hybridization conditions with the following channel catfish H chain cDNA probes: Cμ1, a 335-bp BstEII-Stxl fragment from clone NG13; Cμ2, a 248-bp Stxl-EcoRI fragment from clone NG13; Cμ3, a 271-bp EcoRI fragment from cDNA clone HG103 (12); and Cμ4, a 578-bp EcoRI fragment from clone HG103 that also encodes the C terminus and 3' untranslated region. In addition, the positive clones were screened with the OJ23 oligonucleotide that corresponds to the noncoding strand of Jγ segments (21). From these analyses, two clones, designated C2 and C16, were chosen for additional study.

Genomic analysis and sequencing of genomic channel catfish clones

Ordered restriction maps of clones C2 and C16 were generated with a gene-mapping kit using alkaline phosphatase-conjugated T3 and T7 oligonucleotide probes (Stratagene). Southern blots of the restricted genomic inserts from these clones were hybridized with the following radiolabeled probes: Cμ1, Cμ2, Cμ3, Cμ4, OJ23, and VH probes, representing seven different catfish VH families (VH1 to VH7) (19, 20, 22). The blots were hybridized under both high or low stringency conditions with the Cμ probes (13) and under high stringency conditions with the VH probes (22). A 10.8-kb region, represented by a S4b-EcoRI fragment, an overlapping 3.1-kb PstI fragment, and a 3.8-kb PstI fragment, was derived from clone C2, and each was subcloned in M13 mp18. Overlapping unidirectional deletion subclones of one strand were obtained using exonuclease III (Promega, Madison, WI) and sequenced using vector primers. The other strand was sequenced with various oligonucleotide primers. The complete 10,767-bp sequence has been given the database accession number AF068137. The derived 10.8-kb sequence was aligned with the genomic sequence of the region containing the nine JH segments and the functional VH probes. Clones C15 hybridized with VH1, VH2, VH3, VH4, and VH6 probes, whereas C2 hybridized only with the VH1 and VH3 probes.

The inserts from the C2 and C15 clones were mapped with seven different restriction enzymes, and the locations of the regions that hybridized with the VH, JH, and CMu probes were identified (Fig. 1, top). These results indicated that clones C2 and C15 overlapped; the total genomic distanced spanned was about 25 kb. Seven VH segments representing five different VH families were interspersed and closely linked within a region spanning about 14 kb. Each of the VH segments mapped to regions that were upstream of the Cμ2 hybridizing regions.

The genomic sequence of a 10.8-kb region contained in clone C2 was determined to provide understanding into the unexpected organizational pattern of the VH, JH, and Cμ gene segments within this region. The sequence initiated at the 5' end of clone C2 (upstream of the VH1 segment) and extended to the first PstI site located downstream of the Cμ2 hybridizing region (Fig. 1, top). The sequence of this region is shown in Fig. 2.

Structure of VH segments located in the second H chain gene cluster

Sequencing studies confirmed that a VH1 segment was located near the upstream end of the C2 clone. This gene segment, designated VH1.1, encodes a leader sequence that is split by a 105-bp intron, an open reading frame that encodes readily identified FR4 and CDR regions, and an RSS consisting of a heptamer (CA-CACGTG), a 22-bp spacer, and a nonamer (ACAAAAACT). In comparison with VH1.1, a previously characterized and probably functional germline segment of the VH1 family, VH1.1 also appears functional (Fig. 3A). The total nucleotide similarity between VH1.1 and VH1.1 is 91% (Table I). The FR regions are generally conserved and exhibit a similar FR nucleotide similarity of 96%. Sequence diversity, however, is present in the CDR regions; the nucleotide similarity of the combined CDR regions is 76%, with a predicted amino acid similarity of 68%.

Downstream of the VH1.1 segment is a member of the VH3 family (designated VH3.3a). This segment is a pseudogene because it lacks a leader exon. In comparison with VH3.1, a previously characterized and probably functional VH3 member, VH3.3a sequence begins within the intron of the split leader sequence (Fig. 3B). There are no other obvious sequence defects; the open reading frame extends downstream to the RSS, and the predicted FR regions are similar to those representing a VH3 segment. The overall nucleotide sequence similarity with the VH3.3a segment extending from FR1 through the end of the FR3-encoded region is 80% (Table I). Sequence diversity between the CDR-encoded regions is also evident.

A germline-joined VDJ segment is located downstream of the VH3 segments

Downstream of the VH3.3a segment is a VH1 hybridizing segment that is distinct in its structure. This segment is a germline-joined VDJ segment, and in comparison with cDNA as well as VH1.1 and VH1.1a, the VDJ sequence appears functional (Fig. 3A). The split leader sequence contains appropriate RNA donor and acceptor splice sites. The 115-bp leader intron is longer and more diverse in sequence than the 105-bp leader introns found in either VH1.1 or VH1.1a. A single open reading frame initiates within the split leader at the appropriate ATG start codon and extends through the body.

4 Abbreviations used in this paper: FR, framework region; CDR, complementarity-determining region; RSS, recombination signal sequence; H, heavy chain of immunoglobin; L, light chain of immunoglobin; SINE, short interspersed repetitive elements.
of the coding region. In comparison with V_H 1.1 and V_H 1.a (extending from FR1- through the end of the FR3-encoded regions), the VDJ shares overall nucleotide similarities of about 94 and 90%, respectively, with the highest diversity present within the CDR-encoded regions (Table I). The VDJ sequence does not contain a V_H RSS, but, rather, continues downstream through a sequence encoded by D_H and J_H segments. In comparison with the nine J_H segments located upstream of the characterized C_M, the region extending from the middle of the encoded CDR3 through the end of the FR4 region (sequence positions 4039–4087) is most similar to the coding region of J_H 7 (Fig. 4).

The presence of a germline-joined VDJ has not been previously described in bony fish. The DNA used to construct the phage library was derived from erythrocytes. It was important to verify that the VDJ was joined in the germline rather than potentially representing a rearrangement event that had occurred in erythrocytes. Inspection of the VDJ sequence indicated that a PstI site was located within the CDR2-encoded region and a HindIII site was located within the CDR3-encoded region. Genomic DNA from liver, testes, and erythrocytes of three different catfish was restricted with either PstI or HindIII and Southern blots hybridized with two different probes. Southern blots of PstI restricted DNA were hybridized with the C_M1 probe. These results showed that the predicted 3.1-kb PstI fragment was present in the DNA obtained from each of these tissues (Fig. 5A). The restriction fragment containing the C_M1 exon of the known gene, which was previously shown to exhibit PstI restriction fragment length polymorphisms (20), was represented by the characteristic 16- or 10-kb fragments. Southern blots of HindIII-restricted DNA were hybridized with a 1.6-kb PstI-SstI probe. This probe was derived from the region located between the C_M1 and C_M2 regions of the C2 clone (see Fig. 1). These results showed that a HindIII fragment of 7.9 kb was detected in the restricted DNA from each of these tissues (Fig. 5B). This sized fragment also agrees with that predicted from the derived sequence of the C2 clone; therefore, it is concluded that the VDJ is joined in germline.

A J_H segment is located downstream of the V_H and VDJ segments

A single J_H segment is located about 400 bp downstream of the germline-joined VDJ. This J_H segment (designated J_H b) has a 5’ RSS consisting of a nonamer, 23-bp spacer, and a consensus heptamer. In a multiple sequence alignment with the nine known J_H segments, J_H b is most similar to J_H 8 (Fig. 4). The nonamers are identical, there are only two base pair differences in the 23-bp spacer, the heptamers are identical, and there are four base pair differences in the coding region that result in two changes in the predicted amino acid sequence. The J_H b sequence encodes the hallmark tryptophan, and the sequence terminates with an appropriate RNA splice site. Thus, J_H b appears structurally functional and homologous to J_H 8.

Top. The restriction enzymes used to map genomic λ clones C15 and C2 were HindIII (H), BamHI (B), PstI (P), SstI (S), EcoRI (E), BsrEI (Bs), and SalI (Sa). The location and V_H family designation of the V_H segments are indicated. The location of V_H segments solely contained in clone C15 are shown in the middle of the respective restriction fragment that hybridized with the V_H family-specific probe. Bottom. Schematic sequence alignment of the gene segments in λ clones C2 and C7. Clone C7 contains the nine functional J_H segments, and the functional C_M gene characterized in earlier cDNA and genomic studies (12, 13, 20, 21, 24). The areas of sequence homology are shown by cross-hatched regions placed between these loci. The gaps shown within the loci are not sequencing gaps, but are introduced to show regions of sequence deletions.
FIGURE 2. Sequence and genomic alignment of the gene segments within the two channel catfish H chain gene clusters. The complete 10,767-bp genomic sequence of the second heavy chain gene cluster characterized in this report is shown. The aligned genomic sequence of the first gene cluster containing the nine J H segments and the functional Cm gene is shaded. The predicted amino acid sequence encoded by the gene segments is shown. Alignment gaps, which are introduced to maximize sequence identities, are indicated by dashes.
FIGURE 2. continued.
FIGURE 3. Sequence comparison of selected V_H segments located within the second channel catfish H chain gene cluster. A, The genomic sequence of V_H1.1, a previously characterized member of the channel catfish V_H1 family (accession no. U09719), is aligned with the V_H1.a and the VDJ segments. Arrows are shown to demarcate regions corresponding to the encoded split leader sequence, FR, and CDR of the H chain V region (25). Nucleotide identity is indicated by dots, whereas gaps, which are introduced to maximize similarity, are indicated by dashes. The RNA donor and acceptor splice sites within the leader intron are underlined. The heptamer (7 mer), spacer, and nonamer (9 mer) of the RSS are indicated.

B

Downloaded from http://www.jimmunol.org/ by guest on June 9, 2017
Genomic alignment and structure of Cµ regions within the second H chain gene cluster

The above results indicated that the J region sequence within the VDJ was most similar to J H 7 and that J H b was most similar to J H 8. Genomic sequence alignments were performed to understand this relationship. For clarity, the JH segments (J H 1 to J H 9) and the domain exons and introns of the functional Cµ gene will be termed the first gene cluster. The V H segments, VDJ, and J H b segment as well as the downstream regions that hybridized with Cµ probes will be termed the second gene cluster.

Sequence homologies between the clusters began within the J-coding region of the VDJ of the second gene cluster and the J H 7 segment of the first cluster (Fig. 1, bottom, and Fig. 2). The region between VDJ and J H b readily aligned with the intron between J H 7 and J H 8. The 140-bp region upstream of the J H b RSS shared 99% identity with the 140-bp region upstream of the J H 8 RSS. Sequence homology between the gene clusters continues downstream until position 4878 of the second gene cluster. At this point, an alignment gap (representing the J H 9 segment and its immediate upstream and downstream flanking regions) had to be introduced. Sequence positions 4879–5895 of the second cluster and positions 4325–5429 of the first cluster could be subsequently aligned, although gaps (ranging in size from 10 to 115 bp) were required to accommodate insertions or deletions that were principally located in regions of di- and trinucleotide repeats. A larger gap of 197 bp was necessary to align the region of tetranucleotide (TGTA) repeats upstream of the Cµ1. The sequence of the Cµ1 exon is highly conserved within the second gene cluster. The total similarity, as determined by the percentage of nucleotide identities, is 92% (second gene cluster positions 5896–6183 aligned with first gene cluster positions 5430–5735). The alignment also shows that two deletions have occurred. The first is a single base pair deletion that corresponds to a location near the beginning of the Cµ1-coding region. The second deletion, represented as a 17-bp gap, corresponds to a region that aligns near the middle of the Cµ1 exon. Both these deletions result in frameshift mutations as well as multiple downstream stop codons. Thus, the Cµ1 in the second gene cluster must be considered a pseudogene.

The genomic alignment between the clusters could be extended about 160 bp downstream of the Cµ1 exon, at which point sequence similarity between the clusters ended. The region extending between positions 6347 and 8629 of the second cluster exhibited no apparent sequence similarity with any region in the first gene cluster. Beginning at position 8630 of the second cluster and extending downstream through about position 9879, sequence similarity with the flanking and coding regions of the Cµ2 exon of the first cluster re-emerged.

### Table I. Percent of nucleotide and predicted amino acid similarities in the FR and CDR regions of V H and VDJ segments compared with representative germline VH segments

<table>
<thead>
<tr>
<th>V H Segments Compared</th>
<th>FR1</th>
<th>CDR1</th>
<th>FR2</th>
<th>CDR2</th>
<th>FR3</th>
<th>CDRT</th>
<th>FRT</th>
<th>TOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>V H 1.a vs V H 1.1</td>
<td>97.7</td>
<td>66.7</td>
<td>88.1</td>
<td>78.4</td>
<td>97.8</td>
<td>75.8</td>
<td>95.9</td>
<td>91.3</td>
</tr>
<tr>
<td>VDJ vs V H 1.1</td>
<td>96.6</td>
<td>73.3</td>
<td>95.2</td>
<td>88.2</td>
<td>96.8</td>
<td>84.8</td>
<td>96.4</td>
<td>93.8</td>
</tr>
<tr>
<td>VDJ vs V H 1.a</td>
<td>94.3</td>
<td>46.7</td>
<td>92.9</td>
<td>86.3</td>
<td>94.6</td>
<td>77.3</td>
<td>94.1</td>
<td>90.3</td>
</tr>
<tr>
<td>V H 3.a vs V H 3.1</td>
<td>86.2</td>
<td>66.7</td>
<td>88.1</td>
<td>56.9</td>
<td>85.4</td>
<td>59.1</td>
<td>86.2</td>
<td>80.1</td>
</tr>
</tbody>
</table>

*The percent of nucleotide and amino acid (shown in parentheses) similarity in the three FR and two CDRs of the indicated segments were determined from the aligned sequences shown in Fig. 3. The number of sequence identities was used to calculate the percent similarity (sequence identities/maximum number of residues compared). The similarities of the three combined FR regions (FRT), the two combined CDR regions (CDRT), and the overall total (TOT) similarities are indicated.*

**FIGURE 4.** Sequence comparison of the Iµ segments located in the first and second heavy chain gene clusters of the channel catfish. The nonamer, spacer, heptamer, and coding region are aligned with the Iµ7 gene segment to maximize nucleotide and deduced amino acid similarity. Nucleotide and amino acid identities are indicated by dots, whereas gaps introduced to maximize homology are indicated by dashes. The demarcation of the CDR3-FR4 region of the H chain is indicated (25).
FIGURE 5. Southern blots of restricted genomic DNA isolated from liver (L), testes (T), or erythrocytes (E) from three different channel catfish. A, The DNA was restricted with PstI and hybridized under stringent conditions with a Cµ1 probe. B, The DNA was restricted with HindIII and hybridized under stringent conditions with a 1.6-kb PstI–SstI probe derived from the region residing between the Cµ1 and Cµ2 pseudogenes of the second H chain gene cluster (see Fig. 2). The indicated sizes of the hybridizing fragments are measured from the sizes of PstI-digested λ DNA.

The alignment between the gene clusters extends about 190 bp downstream of the Cµ2 exon. At this point, sequence similarity ended, although the sequence of the second gene cluster was extended downstream about 0.9 kb. This distance would be expected to span the distance to Cµ3, if such was present within the second gene cluster. The distance from Cµ2 to the Cµ3 exon in the first gene is 447 bp (13, 24). The lack of nucleotide similarity within the sequence regions coupled with the fact that no region in clone C2 hybridized under relaxed stringency with either Cµ3 or Cµ4 probes are consistent with the conclusion that there are no additional regions of extended similarity between the gene clusters.

A Tc1-like transposon is located within the second H chain gene cluster

In earlier studies, Henikoff (26) had conducted database searches to determine whether transposons known from bacteria, invertebrates, and plants were present in other classes of organisms. New members of the Tc1/mariner family of invertebrate transposons were detected by searching databanks with conserved amino acid sequence blocks derived from sequences of Tc1 family members. In his study a Tc1 homologue was identified within the first H chain gene cluster of the channel catfish; this homologue (designated IpTc1) resided within the intron between the transmembrane 1 and transmembrane 2 exons (26). With the present data indicating that similarities between the gene clusters ended abruptly, analyses were conducted to determine whether a transposon was also present within the second gene cluster. Using the search methodology of Henikoff and colleagues (27), Tc1 conserved amino acid sequence blocks A–F were used to analyze the second gene cluster sequence. These comparisons showed that a member of the Tc1/mariner family was located in inverted orientation downstream of the Cµ2 pseudogene segment. A sequence alignment of IpTc1 and the Tc1 homologue within the second gene cluster (designated IpTc2) is shown in Fig. 6. IpTc1 is flanked by an imperfect 85-bp inverted repeat that bears features common to Tc1-like transposons, including a match to the CAGTGC consensus preceded by an AT-rich region (26). IpTc1 is defective in that the coding sequence is interrupted by stop codons and frame shifts. The element conserves coding sequence blocks A–E; sequence block F is absent. In comparison, IpTc2 lacks terminal repeats, the defective coding sequence begins within conserved sequence block D, and coding sequence similarities end after sequence block E. There are also stop codons and frame shifts that interrupt the coding sequence.

The conservation of amino acid sequence blocks in members of Tc1-like transposable elements has also been examined by Radice and colleagues (28). In their study they used PCR approaches to amplify Tc1-like sequences from Atlantic salmon, rainbow trout, and zebrafish. Sequence alignments of amplified sequences supported Henikoff’s hypothesis that a conserved conceptual protein representing a single long open reading frame could be reconstructed. The conceptual amino acid sequence of the ancestral Tc1-like transposon was compared with the IpTc1 and IpTc2 sequences. These analyses also indicate that IpTc2 could not be aligned outside the indicated area (Fig. 6). IpTc2 appears to have suffered extensive deletions that removed both the amino- and carboxyl-terminal regions of the open reading frame. Thus, sequence similarities between the first and second H chain gene clusters appear to terminate with a Tc1-like element downstream of the Cµ2 pseudogene segment.

Discussion

The second H chain gene cluster analyzed in this report is a complex Ig locus characterized by multiple, closely linked V_H segments, a germline-joined VDJ, a single J_H segment, and two Cµ pseudogene segments. Previously we characterized the genomic organization of channel catfish V_H gene segments (19). Hybridization studies indicated that the average V_H-positive λ clone (insert size of 16–18 kb) contained members of four different families. Four clones studied in detail represented 65 kb and contained 21 V_H segments; the average distance between segments was about 3 kb. None of these clones had the identical V_H organiza-

nizational pattern, and none had V_H segments in the same gene order as that found in the present study. The V_H segments located upstream of the Cµ pseudogene segments are closely linked (average distance between segments of about 2 kb). The V_H segments represent five different V_H families, and the members of these different families are interspersed. The structure and organizational pattern of the V_H segments located upstream of the pseudogene Cµ segments are consistent with these earlier studies.
A characteristic feature of the second gene locus is the germline-joined VDJ. Germline-joined VDJ segments have been identified in some gene clusters found in sharks. It appears that about half these clusters exhibit VH or VDJ joining in germline (8). The structure of the catfish VDJ and its associated flanking regions is consistent with landmark features predicted to occur as the result of an earlier gene rearrangement event. The single open reading frame encodes an amino acid sequence similar to that predicted by both cDNA studies as well as limited genomic studies with other VH 1 members. The V_{H1} reading frame extends through the end of FR3 but does not terminate with an RSS. Rather, the open reading frame extends to encode a CDR3 region similar in length to that observed in cDNA and presumably partially encoded by DH segments (22). The sequence of the encoded FR4 region is also consistent with that coded by known JH segments. It is not clear, however, whether the germline-joined VDJ is expressed. Primers suitable for the selective detection of the VDJ have been used in PCR studies to amplify first-strand cDNA derived from PBL, and no amplification product has been observed. Whether the absence of the product represents low utilization patterns or perhaps a developmental or tissue-specific pattern of gene expression has not been determined. It is nonetheless interesting that germline-joined VDJ segments are only known to occur in lower vertebrates, but the origin, significance, and the impact of their presence are not known.

As reported initially in studies with the channel catfish and now confirmed with several different species, bony fish are the first phylogenetic representatives to have evolved single copy H chain C region genes. In terms of hypothesizing the likely origin for the second gene cluster in catfish, several possibilities exist. The cluster might represent a relict of an earlier, multiple H chain gene cluster organizational pattern. This possibility is potentially supported by the recent discoveries that the genes of both L chain classes of the catfish are organized in multiple gene clusters (29, 30). Another possibility is that the second gene cluster evolved via an intermediary RNA duplication process. In the general model of the evolution of processed genes, a transcript is spliced and polyadenylated, and the processed transcript or its cDNA copy is integrated into staggered breaks at a new chromosomal location (31). The structure of the catfish VDJ and its flanking regions, however, bears none of the characteristic features of an RNA intermediate. The leader is split by an intron, the RNA splice site at the end of JH encoded region is intact, the C_{\gamma} pseudogene domains are not spliced to the VDJ, and no extended poly(A) tract is present flanking either the VDJ or the C_{\gamma} pseudogenes. Therefore, hypotheses attributing the evolution of the catfish second H chain gene cluster to a RNA intermediary pathway seem less likely.

As reported initially in studies with the channel catfish and now confirmed with several different species, bony fish are the first phylogenetic representatives to have evolved single copy H chain C region genes. In terms of hypothesizing the likely origin for the second gene cluster in catfish, several possibilities exist. The cluster might represent a relict of an earlier, multiple H chain gene cluster organizational pattern. This possibility is potentially supported by the recent discoveries that the genes of both L chain classes of the catfish are organized in multiple gene clusters (29, 30). Another possibility is that the second gene cluster evolved via an intermediary RNA duplication process. In the general model of the evolution of processed genes, a transcript is spliced and polyadenylated, and the processed transcript or its cDNA copy is integrated into staggered breaks at a new chromosomal location (31). This process has been shown to occur in Ig loci of higher vertebrates as, for example, with a processed human \( e \) gene that has moved to a different chromosome (32). The structure of the catfish VDJ and its flanking regions, however, bears none of the characteristic features of an RNA intermediate. The leader is split by an intron, the RNA splice site at the end of J_{H} encoded region is intact, the C_{\gamma} pseudogene domains are not spliced to the VDJ, and no extended poly(A) tract is present flanking either the VDJ or the C_{\gamma} pseudogenes. Therefore, hypotheses attributing the evolution of the catfish second H chain gene cluster to a RNA intermediary pathway seem less likely.

Another possible origin of the second gene cluster would be through a massive DNA duplication event. In this regard, sequence alignments with the nine J_{H} segments located upstream of the functional C_{\gamma} (designated IpTc1) (26) is underlined. The CAGTGC consensus sequence located within the inverted repeat of IpTc1 transposable elements is indicated by dots in the IpTc1 sequence. Demarcation of the conceptual open reading frame of Tc1 sequence blocks A–E is indicated by arrows (27). The Tc1-like transposon located downstream of the C_{\gamma}2 pseudogene of the second H chain gene cluster (designated IpTc2) is aligned with the IpTc1 sequence. Amino acids corresponding in sequence to the conceptual protein of an ancestral Tc1 transposon from bony fish are underlined (28).

![Figure 6](http://www.jimmunol.org/) Sequence comparison of Tc1-like transposons located within the first and second heavy chain gene clusters of the channel catfish. The imperfect 85-bp inverted repeat of the Tc1-like transposon located between the C_{\gamma} transmembrane exons of the functional C_{\gamma} (designated IpTc1) (26) is underlined. The CAGTGC consensus sequence located within the inverted repeat of Tc1 transposable elements is indicated by dots in the IpTc1 sequence. Demarcation of the conceptual open reading frame of Tc1 sequence blocks A–E is indicated by arrows (27). The Tc1-like transposon located downstream of the C_{\gamma}2 pseudogene of the second H chain gene cluster (designated IpTc2) is aligned with the IpTc1 sequence. Amino acids corresponding in sequence to the conceptual protein of an ancestral Tc1 transposon from bony fish are underlined (28).
within the second cluster contributes to a functional C region seems remote. Both pseudogene exons are multiply crippled, with RNA splice sites destroyed, and open reading frames interrupted by termination codons, insertions, and/or deletions. Thus, the duplication event that gave rise to the second gene cluster appears to have extended downstream to minimally encompass the first two exons of the Cμ. Hybridization studies indicated that no additional Cμ exons were present. Because Cμ3 and Cμ4 are generally conserved even between species, these exons should have been detected by the approaches used. If additional downstream regions were duplicated, then these sequences have probably extensively diverged from that known in the functional gene cluster.

It is probable that the duplication of the Cμ1 and the Cμ2 resulted from the same molecular event, and thus divergence from the functional sequences might be expected to be similar for both exons. This, however, was not the case, and this conclusion suggests that gene correction mechanisms, such as those attributed to molecular drive (33), may favor Cμ1 rather than Cμ2. This hypothesis also has phylogenetic support from our earlier studies. In the characterization of the channel catfish Cμs, phylogenetic comparisons with the Cμ s of various mammalian, shark, and Xenopus sequences showed that catfish Cμ1, rather than Cμ2, was phylogenetically conserved. For example, the murine Cμ1 domain was 33.0% similar when the nucleotide identities in the aligned sequences were compared; Cμ2 was only 20.7% similar. The statistical interpretation of these alignments was judged to be significant when the alignment score was >3 SD above the mean score of the aligned randomized sequences. Aligned randomized versions of catfish Cμ1 with the murine Cμ1 gave an average similarity of 16.07 ± 2.48%; aligned randomized versions of catfish and murine Cμ2 versions gave an average similarity of 16.70 ± 2.46% (13). It is also of interest that in a recent cDNA description of a chimeric catfish H chain, perhaps related to δ, the identified transcripts used the Cμ1 exon, but not the Cμ2-Cμ4 exons (34). This would also support the possibility that gene correction mechanisms may favor Cμ1.

A major consideration of these results relates to the VH segments. The presence of VH segments upstream of the VDJ suggests that these segments may have also arisen by duplication of extended genomic regions. Although Southern blot studies have predicted extensive VH diversity in the catfish, this assumption is based strictly upon the number of different sized restriction fragments that hybridize with family-specific probes. At this point the number of VH segments located upstream of the VDJ is not known. Although there are at least six such segments, the analyses here did not attempt to localize additional upstream overlapping clones. Therefore, the number of VH segments that may have potentially arisen by duplication of a large genomic region is unknown. This question significantly impacts both repertoire analyses as well as estimations of potential immune diversity. We have shown by pulsed field gel electrophoresis that the first and second gene clusters are linked on the same large genomic fragments (T. Ventura-Holman and C. J. Lobb, unpublished). This finding strengthens the hypothesis that the second gene cluster arose by duplication. It is known in Ig loci of other vertebrates, for example the κ locus of man, that extended regions have undergone large duplication events (35, 36). The evolution of the catfish H chain locus may parallel this organizational pattern.

It has now been shown that transposons related to the Tcl/mariner family are present in the genomes in a variety of vertebrates as well as diverse families of bony fishes (26–28, 37). This family of elements transposes directly from DNA to DNA. The location of transposon-like elements in relationship to proximal host genes is generally not known in other species of bony fish, although a Tcl1-like transposon has been located within an intron of the hagfish vasotocin gene (38). Both Tcl1-like transposons presently identified in the channel catfish genome are located within Cμ regions. While this may be because these regions are one of only a few large areas sequenced in detail, it may also be that Cμ-associated sequences are favorable target sites for transposition.

There is an additional feature within the region between the two pseudogene Cμ exons that is noted. Database comparisons indicate that a retroposon of the SINE (short interspersed repetitive elements) family is also present in this region. These elements use reverse transcriptase to transpose by means of an RNA intermediate. They are a common feature in eukaryotic DNA and are believed to be derived from transfer RNAs (39). Okada and colleagues (40) have identified the Smad SINE family in salmonids that shares common 3’ end sequences. They also identified common 3’ end homologous sequences in other teleosts such as the zebrafish, ladyfish, eel, and whitefish. Sequence alignments indicated that the conserved Smad 3’ end sequence is present in reverse orientation at nucleotide positions 7267–7211 of the second H chain cluster. This 55-bp region is 89% identical with the 3’ end of the zebrafish SINE element (accession no. AB004656, nucleotide positions 182–236). SINE and long interspersed repetitive elements are known to share common 3’ ends (40). Whether the 3’ end of the element observed here is common to a long interspersed repetitive elements element will have to await further description of these elements as they become known in catfish. Transposition of these as well as Tcl1-like elements could cause disruption of genes by mutation or could lead to duplication or deletion (or possibly inversion) of chromosomal regions. Transposition events may have contributed to the disruption of primitive multiple H chain gene clusters and the subsequent emergence of single copy H chain gene organization patterns during early vertebrate phylogeny.

In conclusion, this study has shown that there is a second Ig H chain gene cluster in the channel catfish. This gene cluster shares significant structural features in common with the first gene cluster and appears to have arisen as the result of a massive duplication event. The presence of a germline-joined VDJ segment suggests that the second cluster underwent gene rearrangement after its duplication. Multiple VH gene segments representing members of different VH families are located upstream of the germline-joined VDJ segment. Their presence suggests that the duplication event may have contributed to their origin and offers interesting hypotheses of the potential impact this organizational pattern may have on the repertoire of expressed VH segments. These continuing studies should lead to further insight as the relationship of genomic organization and repertoire function is examined.

References


