The Journal of Immunology
RESEARCH ARTICLE | FEBRUARY 01 1991

Heavy chain variable region gene families evolved early in phylogeny. Ig complexity in fish. FREE

S H Ghaffari; ... et. al

J Immunol (1991) 146 (3): 1037–1046.
https://doi.org/10.4049/jimmunol.146.3.1037

0022-1767/91/1463-1037$02.00/0
THE JOURNAL OF IMMUNOLOGY

Vol. 146, 1037–1046, No. 3, February 1, 1991
Printed in U.S.A.

HEAVY CHAIN VARIABLE REGION GENE FAMILIES EVOLVED EARLY IN PHYLOGENY

Ig Complexity in Fish[1]

SEYED H. GHAFFARI AND CRAIG J. LOBB[2]

From the Department of Microbiology, University of Mississippi Medical Center, Jackson, MS 39216-4505

The V regions of channel catfish H chain cDNA clones have been analyzed. Based upon sequence relationships and hybridization analyses, five different groups of VH genes are identified whose definition is consistent with that of five different VH families. Genomic Southern blots indicate that as many as 100 different germ-line VH genes are likely represented by these families. The sequence diversity between identified members of these different families is similar in magnitude to the divergence represented between members of different human or mouse VH families. The FR regions are the most conserved regions when members of different catfish VH families are compared; specific amino acid positions appear to be highly conserved in phylogeny. Equally important is that diversity is represented in complementarity-determining regions CDR1 and CDR2 in members of the different families as well as in members of the same VH family. These results suggest that an extensive repertoire of VH genes can contribute to antibody diversity in this lower vertebrate. Sequence comparisons indicate that one of the catfish VH families shares considerable structural similarity to several higher vertebrate VH gene families—a relationship which suggests that this VH family may be ancestral to some VH gene families of higher vertebrates. Characteristic of the genomic organization of higher vertebrate H chains, catfish appear to have different VH families wherein a VH gene likely undergoes functional recombination with putative DH gene segments and one of apparently several different JH segments. The recombined V region is expressed with the same C region gene. These combined results suggest that bony fishes are the earliest known phylogenetic representatives to have evolved extensive V region gene families.

The ability of an antibody to recognize a specific Ag is dependent upon the V regions in the component H and L chains. The structural diversity represented in the V region results from the recombination of gene segments known as V and J. In addition, known H chains have an additional V region coding segment referred to as the diversity or DH segment. Recombination of V, (D) and J segments is mediated by specific recombination signal sequences and is accompanied by the processes of somatic mutation, junctional diversity, and N region additions. Through these processes the inherent structural diversity of the V regions in H and L chains arises (reviewed in Reference 1).

The V region in defined H chains is about 110 amino acids in length and is composed of four FR[3] regions and three hypervariable or CDR regions; the CDR regions are the primary areas that provide antibody specificity. The first three FR regions and the first two CDR regions are encoded by VH genes, the CDR3 is encoded by contributions from VH, DH, and JH gene segments, and FR4 is encoded by JH gene segments. Comparisons of the amino acid sequence of the V regions in different H chains have shown that these sequences can be arranged into related groups (2). Subsequent approaches comparing nucleic acid similarities of VH genes further extended these observations to define families of VH genes wherein members of the same VH family show greater than 80% nucleic acid similarity, whereas the similarity between members of different VH families is generally less than 70%. Genomic hybridization analyses have been used to estimate the number of likely VH members in each VH family. Through these approaches the number of potentially different VH genes in the genomic repertoire has been estimated (3, 4).

Although the general system of recombination between individual VH, DH, and JH gene segments has been demonstrated in human and mouse systems, phylogenetic studies have provided additional insights into mechanisms that can create antibody diversity. In the chicken, antibody diversity arises by a hyperconversion mechanism wherein a single functional copy of a V gene undergoes segmental gene conversion between an extensive V gene pseudogene pool (5, 6). A strikingly different mechanism has been shown to function in sharks in which there are at least 100 individual clusters of H chain genes. In each cluster there are VH, DH, JH, and CH gene segments. Sequence analysis has shown that the gene

Received for publication June 26, 1990.
Accepted for publication October 30, 1990.
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
[1] This work was supported by Grant AI 23052 from the National Institutes of Health, Bethesda, MD. The nucleotide sequence data reported in this paper have been submitted to the EMBL/GenBank database under the following accession numbers: M58668 (NG10); M58669 (NG21); M58670 (NG22); M58671 (NG41); M58672 (NG54); M58673 (NG64); M58674 (NG66); and M58675 (NG77).
[2] Please address correspondence and reprint requests to Craig J. Lobb, Ph.D., Department of Microbiology, University of Mississippi Medical Center, 2500 North State Street, Jackson, MS 39216-4505.

[3] Abbreviations used in this paper: FR, framework; CDR complementarity-determining regions; PAM, accepted point mutations per 100 amino acid residues.

segments in each of these clusters are highly related. For example, pairwise comparisons of nine shark VH segments indicate a ~86% overall nucleotide similarity (7). Clearly neither of these lower vertebrates displays the genomic arrangement of Ig V region genes known in higher vertebrates.

Our recent analyses, however, have shown that the genomic arrangement of the H chain genes in catfish appears quite distinct from that found in sharks. Quantitative gene titration experiments have shown that there is only a single genomic copy of the catfish CH gene expressed in characterized full length H chain cDNA (8, 9). The phylogeny of single copy CH genes seems to have been established at the level of bony fishes early in vertebrate phylogeny. Phylogenetic lineages reflecting single copy CH genes have also been identified in the amphibian toad *Xenopus* (10), as well as in another teleost fish (11).

With these findings a major phylogenetic question arises: has structural diversity of V region genes coevolved with single copy Ig C region genes? In earlier studies we had shown that there was considerable FR1 diversity in the amino acid sequences of purified catfish H chains (12). This variation suggested there was likely to be extensive genomic VH diversity. The results presented here show that different VH genes, putative DH segments as well as potentially different JH gene segments, are expressed with the same C region gene. These VH genes can be placed into related groups whose definition is consistent with different VH gene families.

MATERIALS AND METHODS

Construction of catfish cDNA library and analysis of recombinant clones. A channel catfish (*Ictalurus punctatus*) H chain cDNA library was constructed from the poly(A^+) RNA obtained from in vitro stimulated peripheral blood leukocytes (8, 9). Briefly, first strand cDNA synthesis of this library was initiated by priming with a complementary oligomer located near the 3′ end of the CH2 domain of the H chain mRNA. The cDNA library was screened by hybridization with the use of a radiolabeled oligomer that was located eight nucleotides upstream from the oligomer used in priming first strand cDNA synthesis. The reported cDNA clones were obtained from this library and the inserts were sequenced in both directions by the chain termination elongation reaction by using both plasmid primers and synthetic oligonucleotides as described earlier (8, 9).

Sequence comparisons. The sequence comparisons, alignments, and randomization values reported were done as described earlier (8) by using programs of the Protein Identification Resource. Multiple alignments of VH sequences were done using the PROPHET System ALIGN program, based on a described algorithm (13). PAM values for each possible pair were determined by using the unitary matrix and the percentage differences converted to PAM as described (14). The phylogenetic trees obtained from the above approach were verified by an alternative approach wherein measurements of difference with the log odds matrix were used (15). The construction of the phylogenetic trees was assisted by computer programs developed in the PROPHET by G. Gentry of this department (16). These programs implement the algorithms of Fitch and Margoliash (17). The phylogenetic tree presented in Figure 5 was defined by using the unitary matrix and has a percent SD of 7.59, which compares favorably with those reported for other trees (15–17). The best fit phylogenetic tree determined by the log odds matrix approach gave essentially the same branching topology as shown in Figure 5 with only slight differences in the limb lengths.

The representative sequences for the VH families compared herein were taken from the following references: X24, 7183, and J606 (sequences 55.1, 37.1, and 22.1, respectively) (18); J558 (sequence B1-8) (19); S107 (sequence S107) (20); 36-60 (sequence 36-60) (21); Q52 (sequence M141) (22); V31 and VGAM3.8 (sequences V31, and VGAM 3.8, respectively) (4); V10 (sequence MRL-DNA4) (23); *Elops* (sequence 14501) (11) *Caiman* (sequence C3) (24); horned shark (sequence HXIA) (25); *Xenopus* VH1 (sequence VH1) (26); and *Xenopus* VHII (sequence XIg8) (10).

Genomic analyses. Channel catfish blood was obtained from animals collected from a nearby commercial processing plant (Delta Pride, Belzoni, MS). The isolation of high m.w. genomic DNA from catfish E was previously described (9). The designated cDNA fragments used as probes for the catfish VH family hybridization analyses were: VH1, a 300-bp *Pst*I fragment from NG70; VH2, a 378 bp *Pvu*II fragment from NG41; VH3, a 266-bp *Pst*I fragment from NG54, and a 244-bp *Pst*I fragment from NG21; VH4, a 260-bp *Pvu*II-*Bst*EII fragment from NG10; and VH5, a 364-bp *Pvu*II-*Bam*HI fragment from NG66. Each of these fragments was randomly primed to a specific activity of approximately 1×10^9 cpm/μg, and hybridization reactions were performed under high stringency conditions as previously described (9).

RESULTS AND DISCUSSION

VH gene families in the channel catfish. A catfish H chain cDNA library was screened to define full length V region recombinant clones. The cDNA inserts from positive recombinants were sized, mapped and hybridized with a catfish VH gene probe (obtained from clone NG70), which was defined earlier in our report on the full length sequence of the catfish H chain (9). Seven clones, predicted to contain cDNA inserts representing genes from undefined catfish VH families, as well as one clone that cross-hybridized with the NG70 probe (NG64), were chosen for sequence analysis.

The nucleotide and predicted amino acid sequences of these cDNA inserts is shown in Figures 1 and 2. The defined V region in these clones is divided into the likely 5′-untranslated region, the leader sequence, and the FR and CDR regions based upon prior sequence comparisons of NG70 with higher vertebrate V regions (9) as well as amino-terminal sequence analyses of purified catfish H chains (12). The catfish V regions extending through the end of the FR3 region exhibited key structural coding features of known VH genes (DH and JH gene segments are discussed later in the paper). These features included the conserved V region intradomain cysteines, the conserved tryptophan located at the beginning of FR2, and the conserved terminal VH coding sequence Y-Y-C-A-R located at the end of FR3. When the nucleotide sequence similarities in the VH regions of these nine clones were compared, the VH sequences could be placed into five different groups consistent with different VH families (Fig. 3). These families were designated VH1 through VH5 with the prototype sequence for each family designated by the inserts defined in clones NG70, NG41, NG54, NG10, and NG66, respectively.

Genomic VH complexity in the channel catfish. Before estimating the number of VH genes by genomic Southern blot analyses, it was essential to determine if potential VH family probes cross-hybridized. Restriction fragments from each of the designated VH family prototypes were analyzed as potential probes; each of the fragments extended from the leader sequence and ended within the FR3 of the V region (see *Materials and Methods*). The plasmids containing the sequenced cDNA inserts were transfered to nitro-cellulose in dot-blot experiments and duplicate blots were probed with the labeled fragments. These experiments showed that positive hybridization occurred only between members of the same VH family (data not shown). Because these probes did not cross-hybridize, these probes were used in Southern blot genomic analyses.

Genomic DNA was isolated from the nucleated E of eight individual catfish and restricted with *Eco*RI. South-

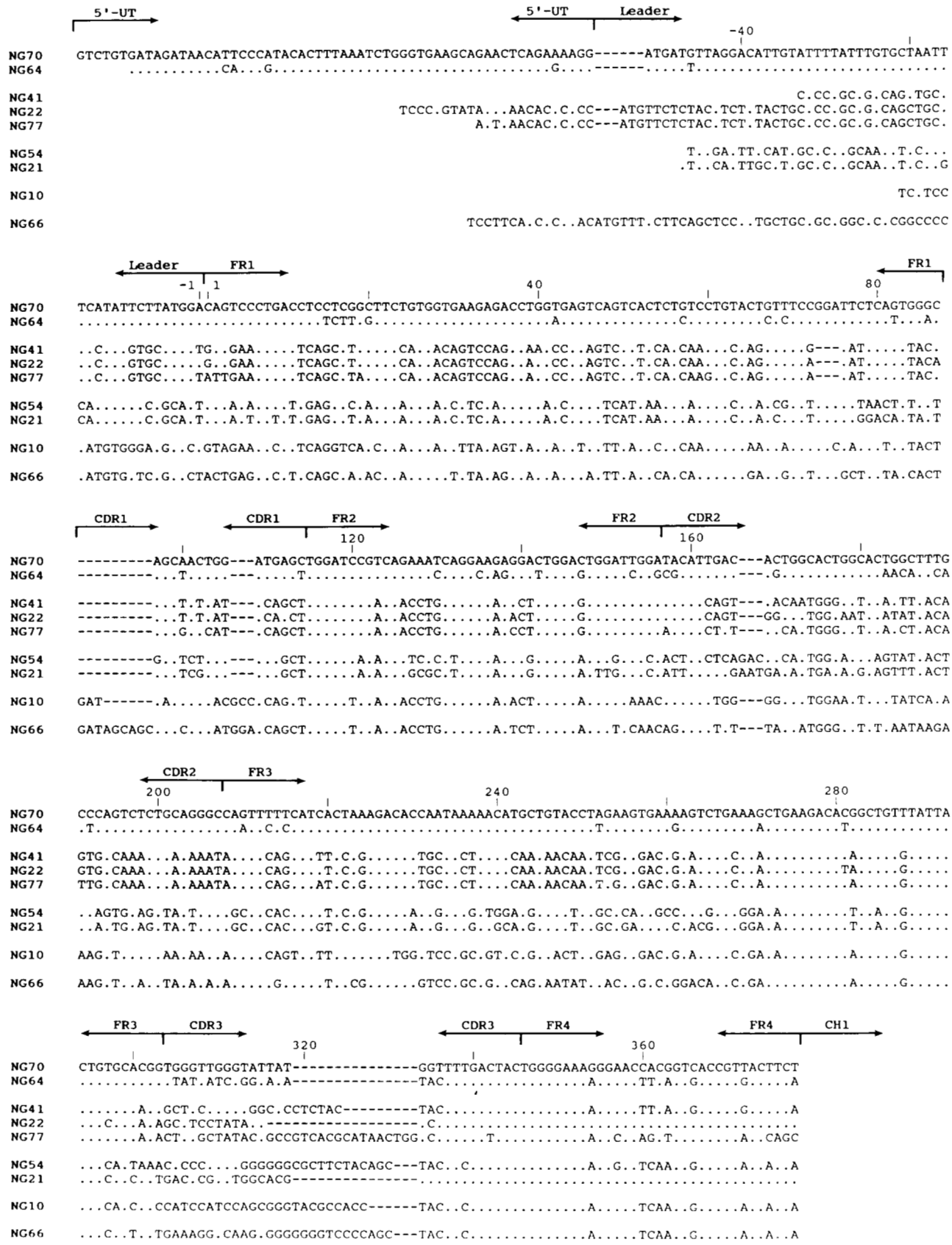

Figure 1. Nucleotide sequences of channel catfish Ig H chain V regions defined in nine cDNA clones. The nucleotide sequence of the C region is not shown because it was identical to that previously reported (9). Demarcation of the 5′ untranslated region (5′-UT) from the leader sequence (*Leader*) was at the location of the first translation start codon ATG; the demarcations of FR and CDR are based upon Kabat et al., (2) with the aid of the ALIGN program of the Protein Identification Resource computer facility. The sequences are compared with the NG70 sequence, and nucleotide identities are indicated by *dots* whereas gaps introduced to maximize homology are indicated by *dashes*.

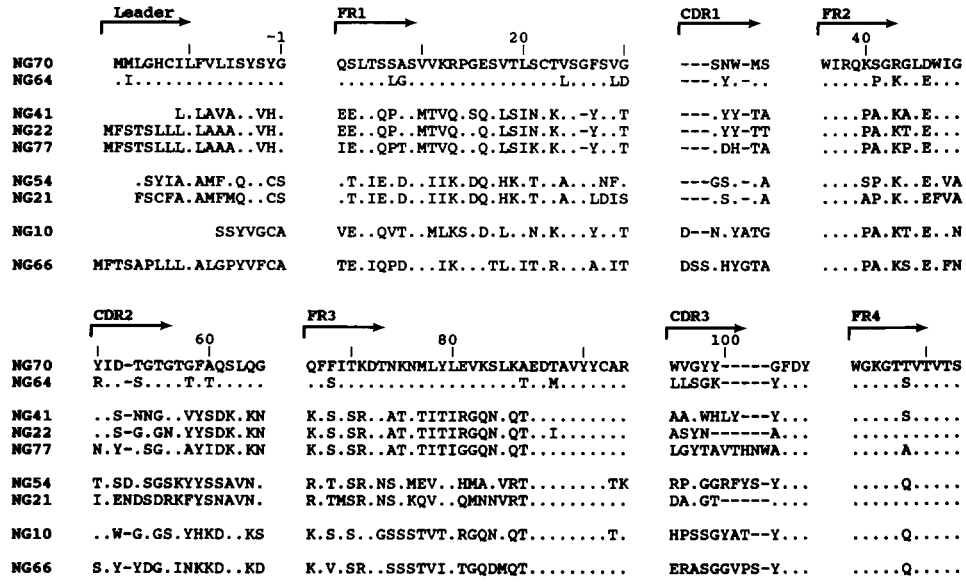

Figure 2. Predicted amino acid sequences of channel catfish Ig H chain V regions defined in nine cDNA clones. The names of the sequences and the demarcations of the leader, FR, and CDR regions are as in Figure 1.

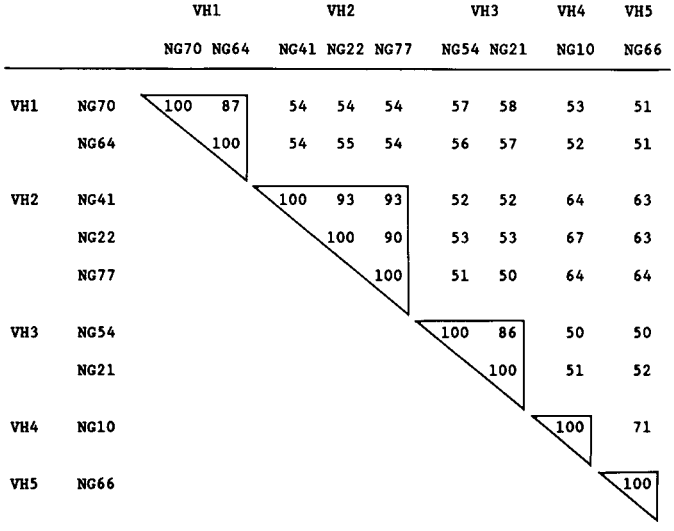

Figure 3. The nucleotide similarity between the VH-coding regions defined in the catfish cDNA clones. The sequences were compared from the beginning of the FR1 region through the end of the FR3 region as aligned in Figure 1. The boxed values indicate the clones that are likely members of the same VH family; these catfish VH families are designated as VH1 to VH5.

ern blots of the restricted DNA were hybridized with probes for each of the catfish VH family prototypes under the identical hybridization conditions used in the dot-blot analyses. These results showed that each of the five probes gave a different Southern blot pattern of restriction fragments (Fig. 4). The range in the number of different restriction fragments represented in these different fish for each VH probe was: NG70, 22–28; NG41, 20–24; NG54, 17–21; NG10, 11–15; and NG66, 28–32. By this approach, these results suggest that there are about 100 genomic VH gene represented by these five families. This estimate assumes that a given restriction fragment contains only one VH gene and that comigration of similar sized restriction fragments does not occur. Thus, it appears that each of these catfish VH families is relatively complex. In this regard, similar Southern blotting approaches when used to analyze human or mouse VH gene families have shown that many VH families appear to be considerably more simple (3, 4, 27, 28). In the mouse, for example, the VH families 36-60, S107, V10, J606, X24, and VGAM 3.8 are represented by as few as 2 and perhaps no more than 10 restriction fragments. In fact, by Southern blotting approaches only the murine VH J558 family would appear to be significantly more complex than these catfish families.

In these analyses, the hybridization intensity of the genomic restriction fragments that were detected with the same probe varied in an individual fish. This could result from comigration of similar sized fragments, potential differences in the number of VH genes on a given restriction fragment as well as potential differences between the probe and VH genes located on the fragment. In an effort to gain some insight into hybridization intensity variation of the fragments seen in an individual fish, as well as to determine if different probes from the same VH family gave the same general VH family restriction pattern, two VH probes obtained from the VH3 family (derived from NG-54 and NG-21) were labeled to the same specific activity and hybridized to duplicate Southern blots. These results showed that the same general pattern of VH3 restriction fragments was detected. There were differences, however, in the intensity of some of the fragments that hybridized with these two probes. Some of the fragments that exhibited relatively high intensity with one probe were less intense with the alternative probe and vice versa. Although the basis for this variation will not be understood until sufficient genomic VH elements are isolated and characterized, these results seem to suggest that there is a partial spectrum of closely related genes within a VH family.

Therefore, bony fishes seem to be the earliest known phylogenetic representatives to have evolved extensive genomic VH gene diversity. This finding distinguishes the VH genes in bony fishes from those defined in sharks, where only a single VH family is known (additional phylogenetic relationships with other vertebrate VH families are discussed later in the paper).

Relationships between the different catfish VH families. Inspection of the aligned nucleic acid sequences (Fig. 1) shows that sequence diversity between the representives of the different VH families is reflected throughout the VH coding region. There were no examples of

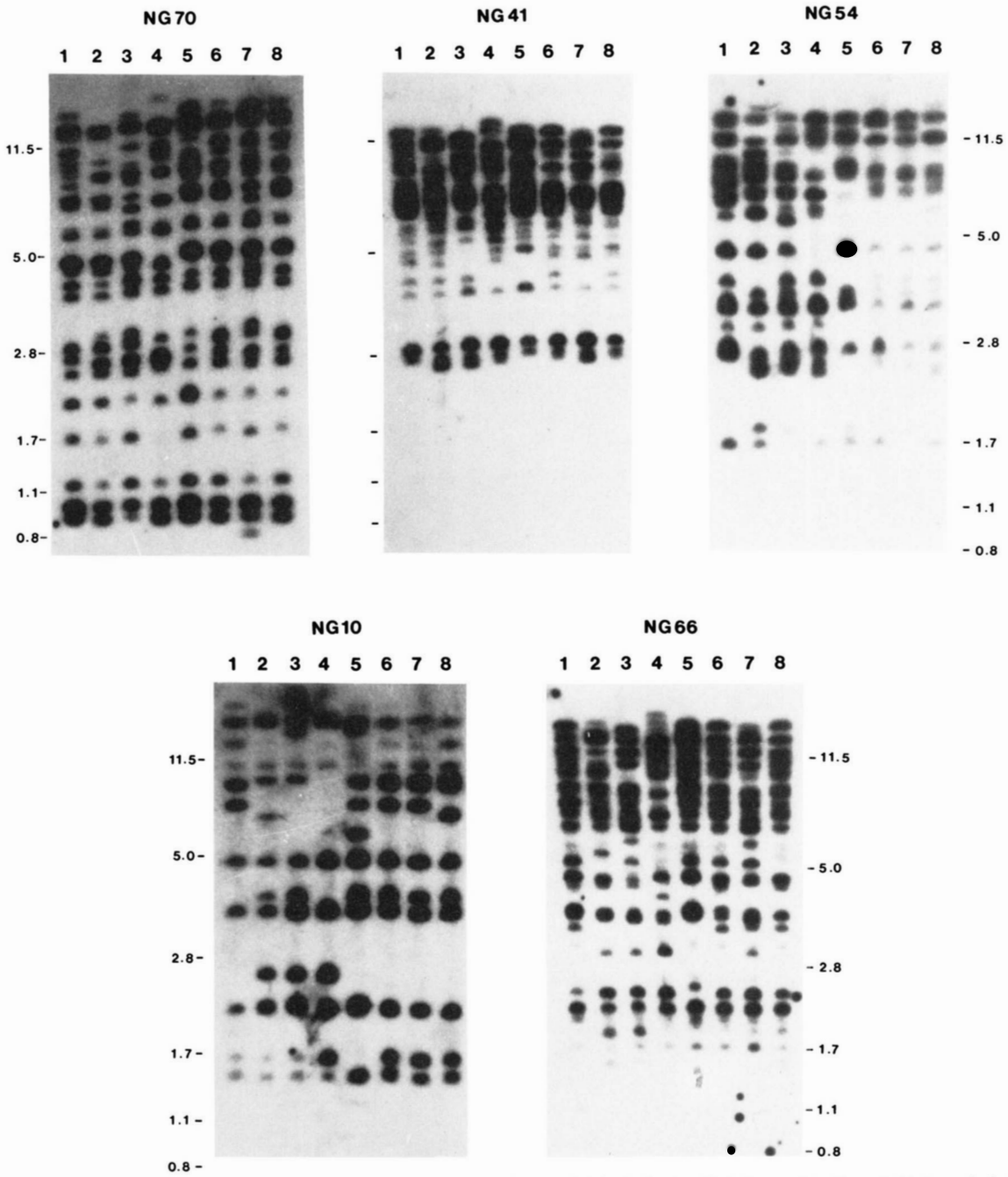

Figure 4. Southern blots of *Eco*RI-digested catfish genomic DNA isolated from eight individual catfish (*lanes 1* to *8*); each blot was hybridized with the indicated catfish VH family probe. The sizes shown in kilobases were measured by using *Pst*I-digested λ DNA.

long, continuous stretches of identical nucleotides (i.e., >10) between members of different VH gene families. To determine the extent of similarities between the VH genes in the different families, the sequence similarities in the FR and CDR regions were compared (Table I). These comparisons showed that the FR2 region was the most conserved; the percent nucleotide similarity ranged from 60% to 93% (Table I). This conservation was also reflected in the predicted amino acid sequence wherein 9 of the 14 amino acids (64%) were generally conserved. FR2 is the smallest of the FR regions and the conserved tryptophan located at the beginning of FR2 forms part of a conserved sequence that is postulated to serve as essential structural role by allowing Ag to have optimal interaction with the flanking CDR regions (reviewed in Reference 29).

The FR3 regions were generally more conserved than the FR1 regions in the different families; the overall nucleotide similarity in the FR3 regions ranged from 51% to 72%, whereas the similarities in the FR1 regions ranged from 46% to 68%. Although these ranges were about equal, the similarities in the FR1 regions were generally dispersed, whereas the similarities in the FR3 regions were generally more localized. The lowest similarities defined in these comparisons were in the CDR regions. Nucleotide comparisons of both CDR regions showed that both exhibited extensive variation; the similarities in the CDR1 regions ranged from 25% to 60% whereas the similarities in the CDR2 regions ranged from 35% to 64%. Thus the selective pressure that was apparently applied early in the phylogeny of vertebrate VH

TABLE I

Percent of nucleotide and amino acid similarities in the FR and CDR regions between representative members of different catfish VH families[a]

Catfish VH Families Compared	FR1	CDR1	FR2	CDR2	FR3	FRT	CDRT	TOT
VH1 vs VH2	46	40	76	35	65	60	37	54
	(38)	(20)	(64)	(31)	(50)	(48)	(29)	(44)
VH1 vs VH3	60	53	74	35	60	63	39	57
	(48)	(40)	(57)	(18)	(47)	(49)	(23)	(43)
VH1 vs VH4	49	38	62	48	58	56	45	53
	(48)	(14)	(57)	(38)	(47)	(49)	(30)	(45)
VH1 vs VH5	49	26	60	40	62	56	35	51
	(45)	(11)	(50)	(25)	(47)	(47)	(20)	(40)
VH2 vs VH3	47	60	74	37	54	55	42	52
	(24)	(20)	(64)	(29)	(44)	(40)	(27)	(37)
VH2 vs VH4	59	48	83	50	72	69	49	64
	(55)	(29)	(86)	(44)	(75)	(69)	(39)	(62)
VH2 vs VH5	54	48	81	54	72	67	52	63
	(45)	(44)	(79)	(38)	(69)	(61)	(40)	(56)
VH3 vs VH4	48	24	64	41	57	55	36	50
	(31)	(0)	(64)	(24)	(53)	(47)	(18)	(39)
VH3 vs VH5	55	26	64	41	51	55	37	50
	(41)	(11)	(64)	(12)	(53)	(51)	(12)	(41)
VH4 vs VH5	68	56	93	65	72	74	61	71
	(52)	(33)	(86)	(44)	(75)	(68)	(40)	(61)

[a] The percent of nucleotide and amino acid (shown in parentheses) similarities in the three framework (FR) and two CDR regions of the VH coding regions were determined from the aligned sequences shown in Figure 1 and 2. The similarities of the three combined FR regions (FRT), and two combined CDR regions (CDRT), and the overall total (TOT) similarities are indicated. The prototype sequences for each of the five catfish VH families (VH1, VH2, VH3, VH4, and VH5) were used for these comparisons (NG70, NG41, NG54, NG10, and NG66, respectively). Aligned randomized nucleotide versions of NG54 with NG66 gave an average total percent similarity of 48.3 ± 1.8.

genes was not equally distributed over the entire VH coding region. Lastly, there were also structural features that were only represented in certain catfish VH families. The FR1 region of the VH2 family was one codon shorter than the others; the CDR1 regions of the VH4 and VH5 families were seven and nine codons in length respectively, whereas in the others the CDR1 regions were five codons in length. Lastly, the CDR2 regions in the VH3 family were one codon longer than those found in the others. Clearly additional family members will need to be analyzed to determine if these codon length distinctions are characteristic of particular VH families. These features, however, suggest that VH families may have arisen from different lineages of VH genes (see also below).

Relationships between the individual members of a catfish VH family. Clones likely representing different members of the same VH family were also sequenced to provide some insight into the similarity of VH genes within a family. Without experimental evidence there was no a priori reason to assume that a primitive multigene VH family was composed of related genes rather than multiple copies of nearly identical genes. The VH1 family was represented by clones NG70 and NG64, the VH2 family was represented by clones NG41, NG22, and NG77, and the VH3 family was represented by clones NG54 and NG21. The comparisons of the nucleotide and amino acid sequence similarities in the FR and CDR regions in these clones are shown in Table II. The FR regions were the most conserved; the total FR nucleotide similarities ranged from 88% to 96% and this similarity was reflected in the comparisons of each of the three FR regions. The total nucleotide similarities in the CDR regions, however, were considerably less and ranged from 73% to 84%. Of particular importance was the predicted diversity at the amino acid level. The total amino acid similarity of the CDR regions ranged from 52% to 84%, and diversity was exhibited in both CDR1 and CDR2 regions. Comparisons of the two most divergent members of the VH2 family, NG22 and NG77, are particularly noteworthy. Although the overall total nucleotide similarity between these two was 90%, the total nucleotide similarity between the CDR regions was 73%. The amino acid similarity was only 40% and 56% for the CDR1 and CDR2 regions, respectively. The diversity in the CDR regions was not reflected within the intermediate FR2 region; the FR2 sequence was essentially identical in these two genes (98% similarity). These comparisons suggest that, in addition to the evolutionary pressures that lead to the divergence of different VH families, there may have been accompanying selective pressure toward diversity of the CDR regions in closely related genes. This selective pressure would significantly expand the potential antibody repertoire when closely related VH genes were expressed.

The relationships of catfish VH families to representative vertebrate VH families. The data bases were searched to define genes that had the highest similarity to the representative catfish VH prototype sequences. The highest optimized alignment scores were with VH genes. Aligned amino acid sequence comparisons showed strong conservation in specific residues; these residues (identified by using the numbering system of Kabat et al. (2) included: Leu(−10) in the leader sequence; Leu-4, Cys-22, and Gly-26 in FR1; Trp-36, Gln-39, Gly-42, Leu-45, and Glu-46 in FR2; and Asp-86, Ala-88, Tyr-90, Tyr-91, and Cys-92 in FR3. These residues are highly conserved residues in higher vertebrate VH sequences (2).

To define possible phylogenetic relationships, the catfish prototype VH sequences were compared with nucleotide sequences representing 10 of the murine VH families as well as representative lower vertebrate sequences (Table III). The lower vertebrate sequences included VH families from the toad (*Xenopus*), the crocodile (*Caiman*), the ladyfish (*Elops*), and the horned shark (*Heterodontus*). If it is assumed that the similarities in the CDR regions can vary extensively depending upon which VH genes

TABLE II

Percent of nucleotide and amino acid similarities in the FR and CDR regions between representative members of the same catfish VH family[a]

VH Family	Clones Compared	FR1	CDR1	FR2	CDR2	FR3	CDRT	FRT	TOT
VH1	NG64 vs NG70	87	87	81	79	93	81	88	87
		(83)	(80)	(79)	(75)	(91)	(76)	(85)	(83)
VH2	NG41 vs NG22	96	93	98	79	96	83	96	93
		(96)	(80)	(93)	(75)	(97)	(76)	(96)	(92)
	NG41 vs NG77	92	80	98	85	97	84	95	93
		(86)	(60)	(93)	(63)	(97)	(62)	(92)	(85)
	NG22 vs NG77	93	73	98	73	96	73	95	90
		(89)	(40)	(93)	(56)	(94)	(52)	(92)	(83)
VH3	NG54 vs NG21	91	87	91	73	87	76	89	86
		(86)	(80)	(86)	(53)	(75)	(59)	(81)	(76)

[a] The nucleotide and amino acid (shown in parentheses) similarities were determined as described in Table I.

are compared, then the comparison of the total FR similarities may provide a more consistent indicator of phylogenetic relationships. In this regard the sequence comparisons with the catfish VH3 family were of particular interest. In 11 of these 15 vertebrate VH families, the catfish VH3 family had the highest total FR similarity. The highest FR similarity was with the *Elops* VH sequence (76%); a sequence relationship which suggests that closely related VH genes have been identified in two different teleost fish. A strong relationship also appears to extend to other lower vertebrate VH families. Of the five catfish VH families, the VH3 family also shared the highest similarity with the VH families in *Xenopus*, *Caiman*, and shark.

The comparisons with the murine VH families were also of interest. The 7183, S107, X24, V10, and J606 each shared about 60% or higher total FR similarity with the catfish VH3 sequences. The highest FR similarity was with the 7183 family (66%). Murine families 7183, X24, S107, and J606 are known to share a close structural relationship and have been placed in the same major subgroup of murine VH families (30). Murine families 7183, S107, and X24 are also among the VH families that are located most proximal to the D region (30).

Multiple sequence alignments have been shown to be useful to define relationships among members of sequence families, to reveal conserved regions, and to predict possible structural relationships among distantly related proteins (reviewed in Reference 16). Although numerous approaches have been used to demonstrate sequence relationships, one extensively used method is based upon the model of Dayhoff et al., wherein evolutionary distance is measured in PAM units (14). With this approach phylogenetic trees were constructed to compare the different catfish VH families with representative vertebrate VH families. The phylogenetic tree having the lowest S.D. is shown in Figure 5. Several conclusions can be made from these analyses. Consistent with the percent similarities obtained by the pairwise comparisons presented in Tables I to III, these analyses suggest that there is a close phylogenetic relationship between the VH2, VH4, and VH5 catfish families. These analyses also suggest that catfish families VH1 and VH3 share a somewhat closer phylogenetic relationship to each other than they do to the other three catfish families. These results also suggest that these major groups of catfish VH families diverged in an ancestral animal which phylogenetically preceded the radiation of bony fishes (catfish) and sharks

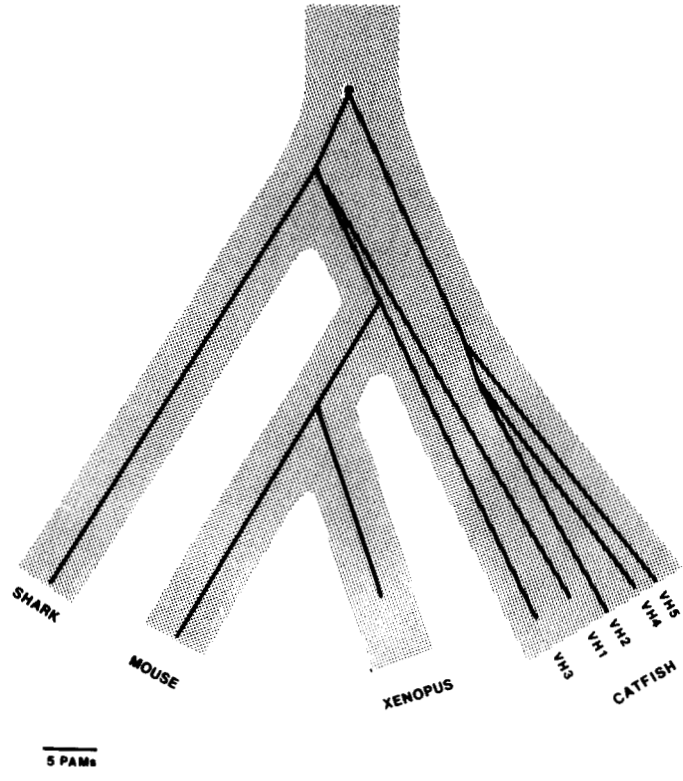

Figure 5. Phylogenetic tree for catfish VH family sequences. The reference sequences for each of the five catfish VH families (VH1 to VH5) are compared with reference sequences representing some other vertebrate VH families (mouse VH 7183, *Xenopus* VH1, horned shark HXIA) as described in the text. The lengths of each branch of the tree are in PAM. The *dot* at the top of the tree represents a VH gene in an unknown ancestral animal which likely underwent a duplication event to give rise to the descending branches of the tree.

(horned shark). In this regard, the fossil record is incomplete. It is known, however, that the first bony fishes appear in the fossil record at a stage that preceeds the first sharks. This fossil evidence has led one author to comment "even though (sharks) are generally considered to be "primitive" fishes, it is doubtful whether they are truly more primitive than the bony fishes" (31).

The percent similarity values as well as the PAM analyses suggest that the catfish VH3 family, as represented by NG54, shares a close phylogenetic relationship with other vertebrate VH families. This sequence relationship is also observed in a multiple nucleotide sequence alignment (Fig. 6). Areas of identical nucleotide sequences are apparent within the FR regions. The similarities are most

TABLE III

Percent nucleotide sequence similarity of five catfish VH families compared to representative vertebrate VH families[a]

VH Families Compared	Catfish VH Families														
	VH1			VH2			VH3			VH4			VH5		
	FRT	CDRT	TOT	FRT	CDRT	TOT	FRT	CDRT	TOT	FRT	CDRT	TOT	FRT	CDRT	TOT
Q52	56	40	53	59	46	56	54	47	53	54	49	53	57	39	53
7183	58	50	57	53	49	52	66	49	62	51	54	52	54	41	52
VGAM 3.8	57	50	54	57	44	53	54	38	49	51	33	45	55	36	48
V36-60	54	48	53	51	54	52	53	44	51	47	54	49	51	36	47
X24	61	55	60	55	42	52	63	40	58	54	44	52	54	36	49
S107	59	36	54	54	43	51	59	39	54	50	36	47	54	32	48
J606	54	44	52	48	43	47	61	42	56	54	36	50	52	35	47
V31	53	39	50	54	41	51	54	42	51	50	44	49	53	43	50
J558	54	49	53	53	52	52	53	46	51	48	42	47	48	44	47
V10	55	44	53	54	44	51	62	43	57	54	41	51	54	38	50
Xenopus VH1	60	59	60	56	49	54	69	56	66	57	51	55	58	42	53
Xenopus VHII	61	51	58	55	51	54	65	49	61	48	48	48	45	47	45
Elops	65	48	61	57	38	53	76	61	72	54	32	49	56	36	51
Caiman	58	52	57	50	30	45	65	50	61	53	35	48	52	36	48
Shark	54	43	52	52	37	49	58	47	55	52	26	46	51	32	46

[a] The nucleotide similarities were determined as described in Table I. Aligned randomized nucleotide versions of catfish VH3 (NG54) with mouse 7183 gave an average total percent similarity of 49.1 ± 1.7.

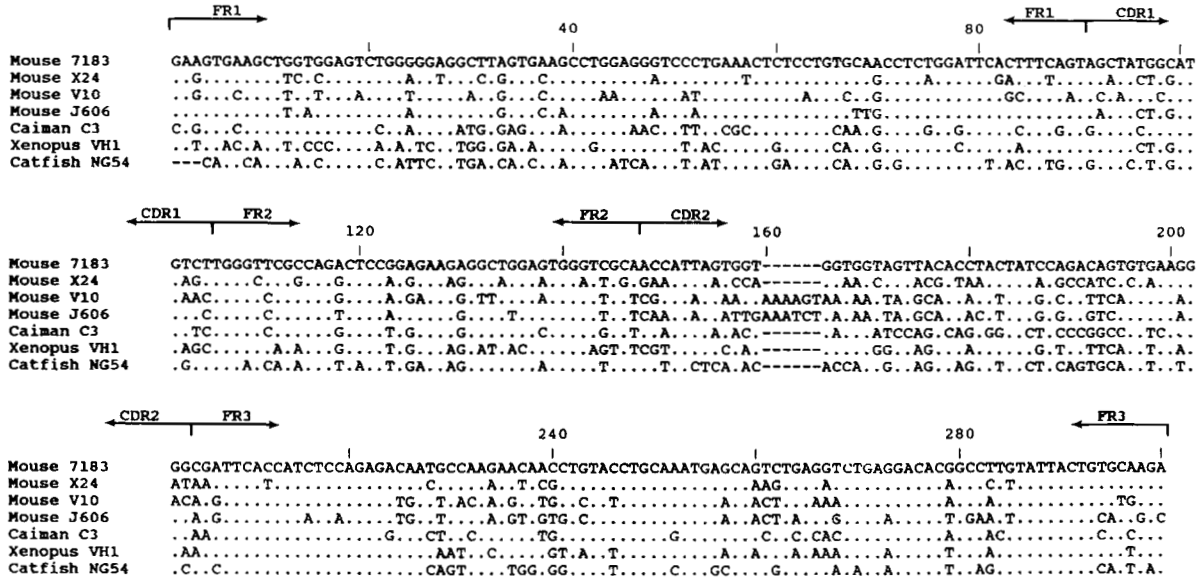

Figure 6. Multiple alignment of representative members of vertebrate VH families compared with the catfish NG54 sequence. The sequences are aligned with the mouse VH7183 sequence and nucleotide identities are indicated by *dots* whereas gaps introduced to maximize homology are indicated by *dashes*.

evident as continuous stretches of common nucleotides that lie within the approximate first and last thirds of the FR3 region. Tutter and Riblet have recently postulated that the mouse 7183 family shares a close relationship with an ancestoral VH gene (32). They also postulated that conserved coding region sequences, such as chi octomer found within FR1 as well as an embedded heptamer within FR3, may provide important noncoding functions during recombination processes. Earlier studies had suggested that this embedded FR3 heptamer may mediate VH gene replacement in a downstream VDJ rearrangement (33, 34). Although a chi octomer is not present within the FR1 of the catfish sequences, the embedded FR3 heptamer TACTGTG (see VH7183, Fig. 6 nucleotide positions 289–295) is represented in some catfish families (see Fig. 1, families VH1 and VH2, nucleotide positions 282–298). This heptamer is contained in a 10-base sequence TATTACTGTG; the first 8 bases in this sequence are conserved in the catfish sequences as well as in other vertebrate sequences. Hence, these analyses support the hypothesis that the mouse VH7183 family likely descended from a phylogenetically conserved VH gene family.

Evidence for multiple DH and JH gene segments. The CDR3/FR4 region of mammalian V regions is encoded by contributions from VH, DH, and JH gene segments. Our earlier sequence comparisons had identified two likely catfish JH coding sequences, each about 45 nucleotides in length, which encoded the last four amino acids of the CDR3 and the entire FR4 region. These coding sequences shared about 70% nucleotide sequence similarity with known mouse, human, and shark JH segments. Present sequence comparisons suggest that there may be as many as five different JH segments that can recombine with VH genes from different VH families (Fig. 7). For example, the identical JH coding region sequence was defined in clones NG70, NG22, and NG21; each of these clones represent members of different VH families (VH1, VH2, and VH3, respectively). A different JH coding region sequence identified in NG1 (VH4 family) was identical to

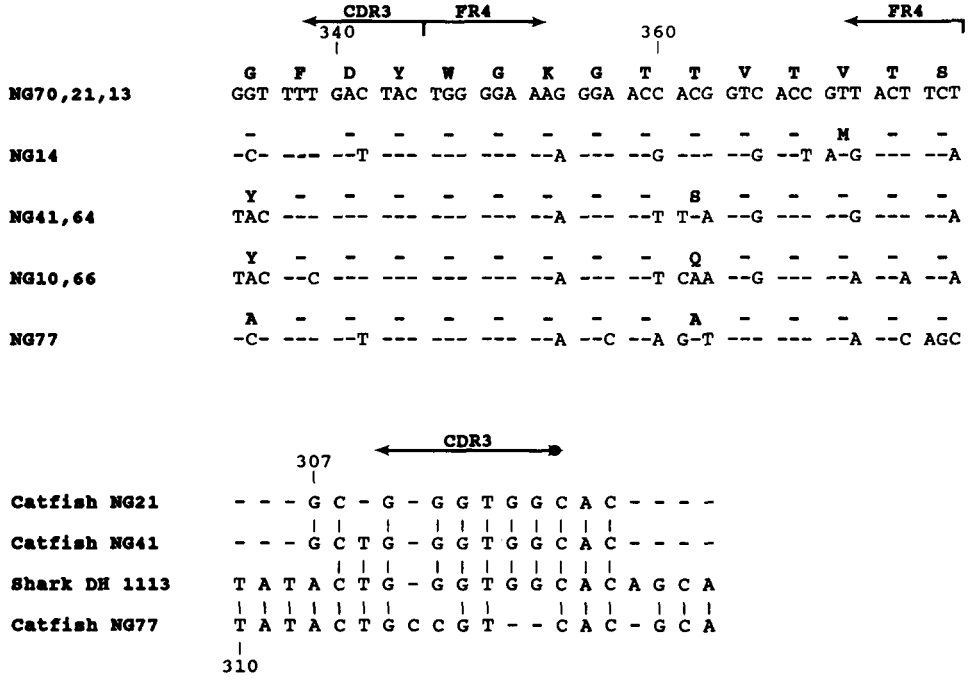

Figure 7. JH and DH coding segments in the CDR3/FR4 region of various catfish cDNA sequences. *Upper,* the nucleotide and predicted amino acid sequence of five likely JH coding regions is shown; clones having identical sequences are indicated. *Lower,* likely catfish DH-encoded sequences compared with a known shark DH gene.

that identified in clones representing the VH3 and VH5 families (NG54, and NG66, respectively). The JH coding sequences in clones NG54 and NG22 are not shown in Figure 7; each of these differed from a concensus sequence by one nucleotide. In NG54 nucleotide position 335 was C, otherwise it was identical to the JH coding sequences in NG10 and NG66. In NG22 nucleotide position 357 was G, otherwise it was identical to the JH coding sequences in NG70, NG21, and NG13. The conservation of structure is evident in the JH coding regions. Although the nucleotide similarities ranged from about 67% to 89%, only two or three of the 15 encoded amino acids varied. Amino acid position 108 (see Fig. 2) was the most variant in the catfish JH sequences; this position is also highly variant in both mouse and human JH sequences (2). Ongoing analyses indicate that JH gene segments are probably not linked to individual VH elements because single as well as mixed sequence oligonucleotide probes complementary to the defined JH coding regions define distinct genomic restriction fragments (data not shown).

The CDR3 comparisons reveal potential information concerning DH gene segments in catfish. First, the length of the CDR3 region in these clones varied extensively ranging from 24 to 42 nucleotides. Second, there was little sequence similarity in the CDR3 between members of the same VH family. Third, and perhaps most important, there were specific sequences within the CDR3 of certain clones that shared sequence similarity with known DH gene segments (Fig. 7). The nine base sequence GGGTGGCAC was identical in the CDR3 of both the NG21 and NG41 clones, and this sequence is identical to a sequence contained in the 1113 DH gene segment of the shark (7).

In conclusion, these analyses show that the channel catfish as a representative of bony fishes has likely evolved an extensive VH repertoire. The catfish VH sequences identified in this study can be placed into different related groups whose definition is consistent with that of five different VH gene families. Comparisons show that the sequence diversity defined between different family members is similar in magnitude to the divergence represented between members of different human or mouse VH families. These studies also suggest that evolutionary pressure seems to have been applied early in vertebrate phylogeny to derive different V region CDR sequences; sequence differences are apparent in both the CDR1 and CDR2 regions and are reflected in members of different VH families as well as in members of the same VH family. Characteristics of the genomic organization of higher vertebrate H chains, the catfish seems to have different VH families wherein a VH gene likely undergoes functional recombination with putative DH gene segments and one of apparently several different JH segments; this recombined V region is joined to a four domain C region gene. This C region gene is represented in the genome as a single copy. The ongoing efforts to characterize genomic V region elements of catfish H chains should continue to provide important insight into the phylogeny of Ig structure and function.

Acknowledgments. We would like to thank James R. Hayman, Julia Cay Jones, and Tereza Ventura-Holman for their valuable assistance in these studies. We also extend our appreciation to Dr. Glenn Gentry for his help and thoughtful discussions regarding the construction of phylogenetic trees.

REFERENCES

1. **Max, E. E.** 1989. Immunoglobulins: molecular genetics. In *Fundamental Immunology,* W. E. Paul, ed. Ravon Press, New York, p. 235.
2. **Kabat, E. A., T. T. Wu, M. Reid-Miller, H. M. Perry, and K. S. Goltesman.** 1987. *Sequences of Proteins of Immunological Interest.* National Institutes of Health, Bethesda, MD.
3. **Brodeur, P. H., and R. Riblet.** 1984. The immunoglobulin heavy chain variable region (Igh-V) locus in the mouse. I. One hundred Igh-V genes comprise seven families of homologous genes. *Eur. J. Immunol. 14:922.*
4. **Winter, E., A. Rodbruch, and U. Krawinkel.** 1985. Members of novel VH gene families are found in VDJ regions of polyclonally activated B-lymphocytes. *EMBO J. 4:2861.*

5. **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.*
6. **Reynaud, C.-A., A. Dahan, V. Anquez, and J.-C. Weill.** 1989. Somatic hyperconversion diversifies the single VH gene of the chicken with a high incidence in the D region. *Cell 59:171.*
7. **Kokubu, F., R. Litman, M. J. Shamblott, R. Hinds, and G. W. Litman.** 1988. Diverse organization of immunoglobulin VH gene loci in a primitive vertebrate. *EMBO J. 7:3413.*
8. **Ghaffari, S. H., and C. J. Lobb.** 1989. Cloning and sequence analysis of channel catfish heavy chain cDNA indicate phylogenetic diversity within the IgM immunoglobulin family. *J. Immunol. 142:1356.*
9. **Ghaffari, S. H., and C. J. Lobb.** 1989. Nucleotide sequence of channel catfish heavy chain cDNA and genomic blot analyses: implications for the phylogeny of Ig heavy chains. *J. Immunol. 143:2730.*
10. **Schwager, J., C. A. Mikoryak, and L. A. Steiner.** 1988. Amino acid sequence of heavy chain from *Xenopus laevis* IgM deduced from cDNA sequence: implications for evolution of immunoglobulin domains. *Proc. Natl. Acad. Sci. USA 85:2245.*
11. **Amemiya, 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.*
12. **Lobb, C. J., and M. O. J. Olson.** 1988. Immunoglobulin heavy chain isotypes in a teleost fish. *J. Immunol. 141:1236.*
13. **Needleman, S. B., and C. D. Wunsch.** 1970. A general method applicable to the search for similarities in amino acid sequence of two proteins. *J. Mol. Biol. 48:443.*
14. **Dayhoff, M. O., R. M. Schwartz, and B. C. Orcutt.** 1978. A model of evolutionary change in proteins. *Atlas of Protein Sequence and Structure.* Vol. 5, Suppl. 3. M. O. Dayhoff, ed. National Biomedical Research Foundation, Washington, D.C., p. 345.
15. **Feng, D. F., M. S. Johnson, and R. F. Doolittle.** 1985. Aligning amino acid sequences: comparison of commonly used methods. *J. Mol. Evol. 21:112.*
16. **Gentry, G. A., M. Lowe, G. Alford, and R. Nevins.** 1988. Sequence analyses of herpesviral enzymes suggest an ancient origin for human sexual behavior. *Proc. Natl. Acad. Sci. USA 85:2658.*
17. **Fitch, W. M., and E. Margoliash.** 1967. Construction of phylogenetic trees. A method based on mutation distances as estimated from cytochrome c sequence is of general applicability. *Science 155:279.*
18. **Hartman, A. B., and S. Rudikoff.** 1984. VH genes encoding the immune response to beta-(1,6)-galactan: somatic mutation in IgM molecules. *EMBO J. 3:3023.*
19. **Bothwell, A. L. M., M. Paskind, M. Reth, T. Imanishi-Kari, K. Rajewsky, and D. Baltimore.** 1981. Heavy chain variable region contribution to the NP family of antibodies: somatic mutation evident in a gamma 2a variable region. *Cell 24:625.*
20. **Early, P., H. Huang, M. Davis, K. Calame, and L. Hood.** 1980. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D, and JH. *Cell 19:981.*
21. **Near, R. I., E. Juszczak, S. Y. Hung, S. A. Sicari, M. N. Margolies, and M. L. Gefter.** 1984. Expression and rearrangement of homologous immunoglobulin VH genes in two mouse strains. *Proc. Natl. Acad. Sci. USA 81:2167.*
22. **Sakano, H., R. Maki, Y. Kurosawa, W. Roeder, and S. Tonegawa.** 1980. Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. *Nature 286:676.*
23. **Kofler, R.** 1988. A new murine Ig VH gene family. *J. Immunol. 140:4031.*
24. **Litman, G. W., K. Murphy, L. Berger, R. Litman, K. Hinds, and B. W. Erickson.** 1985. Complete nucleotide sequences of three VH genes in *Caiman*, a phylogenetically ancient reptile: evolutionary diversification in coding segments and variation in the structure and organization of recombination elements. *Proc. Natl. Acad. Sci. USA 82:844.*
25. **Litman, G. W., L. Berger, K. Murphy, R. Litman, K. Hinds, and B. W. Erickson.** 1985. Immunoglobulin VH gene structure and diversity in *Heterodontus*, a phylogenetically primitive shark. *Proc. Natl. Acad. Sci. USA 82:2082.*
26. **Yamawaki-Kataoka, Y., and T. Honjo.** 1987. Nucleotide sequence of variable region segments of the immunoglobulin heavy chain of *Xenopus laevis*. *Nucleic Acids Res. 15:5888.*
27. **Kodaira, M., T. Kinashi, I. Umemura, F. Matsuda, T. Noma, Y. Ono, and T. Honjo.** 1986. Organization and evolution of variable region genes of the human immunoglobulin heavy chain. *J. Mol. Biol. 190:529.*
28. **Berman, J. E., S. J. Mellis, R. Pollock, C. L. Smith, H. Suh, B. Heinke, C. Kowal, U. Surti, L. Chess, C. R. Cantor, and F. W. Alt.** 1988. Content and organization of the human Ig VH locus definition of three new VH families and linkage to the Ig CH locus. *EMBO J. 7:727.*
29. **Hasemann, C. A., and J. D. Capra.** 1989. Immunoglobulins: structure and function. In *Fundamental Immunology*, W. E. Paul, ed. Raven Press, New York, p. 209.
30. **Brodeur, P. H., G. E. Osman, J. J. Mackle, and T. M. Lalor.** 1988. The organization of the mouse IgH-V locus. *J. Exp. Med. 168:2261.*
31. **Colbert, E. H.** 1969. *Evolution of the Vertebrates: A History of the Backboned Animals through Time.* John Wiley, New York, p. 44.
32. **Tutter, A., and R. Riblet.** 1989. Conservation of an immunoglobulin variable region gene family indicates a specific, noncoding function. *Proc. Natl. Acad. Sci. USA 86:7460.*
33. **Kleinfield, R., R. R. Hardy, D. Tarlinton, J. Dangl, L. A. Herzenberg, and M. Weigert.** 1986. Recombination between an expressed immunoglobulin heavy-chain gene segment in a Lyl^+ B-cell lymphoma. *Nature 322:843.*
34. **Reth, M. G., S. Jackson, and F. W. Alt.** 1986. V_HDJ_H formation and DJ_H replacement during pre-B differentiation: non-random usage of gene segments. *EMBO J. 5:2131.*