Divergent Evolution and Evolution by the Birth-and-Death Process in the Immunoglobulin V_H Gene Family

Tatsuya Ota and Masatoshi Nei
Department of Biology and Institute of Molecular Evolutionary Genetics, The Pennsylvania State University

Immunoglobulin diversity is generated primarily by the heavy- and light-chain variable-region gene families. To understand the pattern of long-term evolution of the heavy-chain variable-region (V_H) gene family, which is composed of a large number of member genes, the evolutionary relationships of representative V_H genes from diverse organisms of vertebrates were studied by constructing a phylogenetic tree. This tree indicates that the vertebrate V_H genes can be classified into group A, B, C, D, and E genes. All V_H genes from cartilaginous fishes such as sharks and skates form a monophyletic group and belong to group E, whereas group D consists of bony-fish V_H genes. By contrast, group C includes not only some fish genes but also amphibian, reptile, bird, and mammalian genes. Group A and B genes were composed of the genes from mammals and amphibians. The phylogenetic analysis also suggests that mammalian V_H genes are classified into three clusters—i.e., mammalian clans I, II, and III—and that these clans have coexisted in the genome for >400 Myr. To study the short-term evolution of V_H genes, the phylogenetic analysis of human group A (clan I) and C (clan III) genes was also conducted. The results obtained show that V_H pseudogenes have evolved much faster than functional genes and that they have branched off from various functional V_H genes. There is little indication that the V_H gene family has been subject to concerted evolution that homogenizes member genes. These observations indicate that the V_H genes are subject to divergent evolution due to diversifying selection and evolution by the birth-and-death process caused by gene duplication and dysfunctioning mutation. Thus, the evolutionary pattern of this monofunctional multigene family is quite different from that of such gene families as the ribosomal RNA and histone gene families.

Introduction

Monofunctional multigene families, where all member genes have the same function, are generally believed to undergo coincidental or concerted evolution, which homogenizes the DNA sequences of the member genes (Smith et al. 1971; Smith 1974; Zimmer et al. 1980). Good examples of concerted evolution are seen with ribosomal RNA (rRNA) genes (e.g., see Arnheim 1983) and histone genes (e.g., see Matsuo and Yamazaki 1989), where all the member genes have virtually identical coding sequences within species. This is understandable, because these gene families produce a large quantity of gene products of the same function that are essential for the survival of an organism. Concerted evolution is caused either by unequal crossover or intergenic gene conversion.

Concerted evolution was also invoked to explain the evolution of the immunoglobulin heavy-chain variable-region (V_H) genes (Hood et al. 1975; Ohta 1980, 1983). Gojobori and Nei (1984), however, showed that V_H gene (often called “gene segment”) family in the mouse and human undergoes a very slow rate of concerted evolution, if any; that the extent of sequence divergence between member genes in the mouse or the human is substantial; and that some of these genes diverged much earlier than the mammalian radiation (~80 Mya) (also see Tutter and Riblet 1989a).

During the past 10 years, the DNA sequences of V_H genes have been determined for many lower vertebrate organisms (see Litman et al. 1993). We can therefore study the pattern of evolution of the V_H gene family on a much longer evolutionary time scale. This type of study is particularly interesting, because the gene organization of V_H genes is now known to vary extensively among different classes of vertebrates (Du Pasquier 1993). In this paper, we show that the pattern of V_H gene evolution is very different from the typical concerted evolution and that two evolutionary processes, which may be called “divergent evolution” and “evolution by the birth-and-death process,” predominate in
FIG. 1.—Genomic organizations of immunoglobulin heavy-chain genes. Cartilaginous fishes have the \((V_H-D_H-J_H-C_H)_n\) type organization including germ-line-joined genes, whereas most higher vertebrates, including bony fishes and tetrapods, have the \((V_n-J_n-C_n)_n\) type gene organization. Chicken has a single functional variable gene. \(V = \) variable-segment gene; \(D = \) diversity-segment gene; \(J = \) joining-segment gene; \(C = \) constant-segment gene; and \(\psi = \) pseudogene. For more detailed information, see the work of Litman et al. (1993).

the evolution of \(V_H\) genes in vertebrates, despite the fact that these genes have the same function. We first study the long-term evolution of \(V_H\) genes, examining the evolutionary relationships of the genes obtained from various classes of vertebrates, and then a short-term evolution of human \(V_H\) genes.

Material and Methods

Background Information

Immunoglobulin or antibody molecules consist of two heavy chains and two light chains. Both the heavy and light chains are composed of the variable region and the constant region (C). The variable region is responsible for antigen-binding, whereas the constant region is responsible for effector function. The heavy-chain variable region is encoded by the variable-segment \((V_H)\), diversity-segment \((D_H)\), and joining-segment \((J_H)\) genes, whereas the light-chain variable region is encoded by the variable-segment \((V_L)\) and joining-segment \((J_L)\) genes. However, the major portion of the variable region is determined by either \(V_H\) or \(V_L\). The \(V_H\) and \(V_L\) genes are evolutionarily homologous with each other, but their sequence similarity is quite low.

The \(V_H\) (as well as \(V_L\)) genes can be subdivided further into two hypervariable or complementarity-determining regions (CDR1 and CDR2) and three framework regions (FW1, FW2, and FW3). (CDR3 is determined primarily by the \(D_H\) and \(J_H\) genes, so it was excluded in this paper). The human and mouse genomes contain \(>100\) \(V_H\) genes and a few to a dozen \(D_H\) and \(J_H\) genes (see fig. 1). Each of the \(V_H\)'s is recombined with one of the \(D_H\)'s, one of the \(J_H\)'s, and one of the \(C_H\)'s (constant region genes), to form a heavy-chain gene (Tonegawa 1983). This process therefore generates a large number of different immunoglobulins that are necessary for defending an individual from many different foreign antigens. Essentially the same humoral immune system exists for most of the bony fishes and tetrapods.

In lower vertebrates such as sharks and skates, however, the genomic organization of the immunoglobulin genes is different from that of higher vertebrates, which may be described as \((V_n-D_n-J_n-C_n)\). In lower vertebrates, the basic unit of repeat of genes is \((V-D-D-J-C)\) (fig. 1); that is, the linked gene group \(V_H-D_H-J_H-C_H\) is repeated several hundred times in the genome,
and the repeat seems to be scattered on different chromosomes (Litman et al. 1993). Therefore, the immunoglobulin genes in these organisms may be represented by (V-D-D-J-C)ₙ.

Of course, there are many exceptions to this rule. Some units of shark genes have the Vₜ and Dₜ genes that are already united (VDD-J-C) in the germ-line genome, whereas in some other units all Vₜ, Dₜ's, and Jₜ are united (VDDJ-C) (Litman et al. 1993). In the little skate (Raja erinacea), (V-DJ-C) are also present. Furthermore, in the coelacanth (Latimeria chalumnae), which belongs to the lobe-finned fishes, (V-D) is apparently repeated many times in the 5' side of the J and C genes (Amemiya et al. 1993). Actually, the number of lower vertebrates studied so far is very small, so it is likely that some new types of genomic organization of Vₜ genes will be discovered in the near future.

Another group of vertebrates that has a deviant genomic organization of the Vₜ genes is that of avian species represented by chicken (for review, see McCormack et al. 1991). In these species, there are only one functional Vₜ gene and ~80 Vₜ pseudogenes in the genome (Reynaud et al. 1989), and the antibody diversity is generated by gene conversion of the functional gene by the pseudogenes (fig. 1). (The same process operates for the light-chain genes as well; Reynaud et al. 1987; Thompson and Neiman 1987.) However, this system may not apply for all avian species, because in the Muscovy duck there seem to be at least two functional Vₜ genes (McCormack et al. 1989). It is interesting that a somewhat similar pattern has been observed in the rabbit. In this organism, the D-proximal Vₜ gene is preferentially used, though there are ~100 Vₜ genes and many of them seem to be functional (Knight 1992). The immunoglobulin diversity in this organism seems to be generated by gene conversion and somatic mutation.

In the primitive jawless vertebrates such as lampreys and hagfishes, an extensive search for immunoglobulin genes has been conducted, but no genes that are clearly homologous to the immunoglobulin genes in cartilaginous fishes or higher vertebrates have been identified, though antibodies are clearly present in these organisms (Litman et al. 1993). Therefore, a completely different immune system may exist in these species.

Nucleotide Sequences Used

At the present time, there are >600 Vₜ gene sequences that have been determined for the various groups of organisms mentioned above. This number includes both germ-line and cDNA sequences, and the majority of them come from the human and mouse (Kabat et al. 1991). For studying the long-term evolution of Vₜ genes, it is necessary to use representative genes from each species and to cover a wide range of organisms. We have therefore included (a) almost all sequences from a species where a small number of genes have been studied but (b) a relatively few representative sequences from a species where a large number of genes have been studied. For the study of long-term evolution, no pseudogenes were included. In both human and mouse, >60 functional germ-line genes have been sequenced, but Schroeder et al. (1990) showed that they can be classified into three major clusters, i.e., clans I–III (also see Kirkham et al. 1992). Our preliminary analysis of ~600 mammalian Vₜ genes, which were obtained from GenBank and Kabatnuc databases, has also suggested that these genes can be grouped into the three major clusters (data not shown). We have therefore used 11 and 7 representative sequences from the mouse and the human, respectively, covering the three major clusters, i.e., clans I–III. More than 30 different Vₜ sequences are also available from the African toad Xenopus laevis. We chose 11 representative sequences from them. For other organisms, we used almost all available genes after exclusion of closely related ones—except for chicken, where we used only one gene, which is used exclusively for antibody generation. For the purpose of rooting the Vₜ gene tree, we included a horned shark light-chain Vₜ gene and a human preB gene, which is known to be homologous to vertebrate lambda Vₜ genes (Kudo and Melchers 1987).

The total number of Vₜ and outgroup genes used for our study of long-term evolution was 57, and they are listed in table 1. We used the germ-line sequences whenever possible, to avoid the effect of somatic mutation (Tonegawa 1983). In a few species, however, we used cDNA, because germ-line genes were not available and sequence divergences were so large that the effect of somatic mutation was negligible. The cDNA sequences are denoted by asterisks in table 1. Since we used many genes from many different organisms, we designated each gene by the first letter of both the genus name and the species name plus the gene designation used by the original authors, except in the African toad (X. laevis), where we used “Xe” instead of “Xl” because “1” (cl) is easily confused with “1” (one).

For the study of “short-term” evolution, we included pseudogenes as well as functional genes and used 10 Vₜ genes belonging to the clan I genes and 28 genes belonging to the clan III genes. (Clan II genes were not used, because there was only one pseudogene.) All of these data were taken from recent exhaustive studies of human Vₜ genes by Shin et al. (1991) and Matsuda et al. (1993). These Vₜ genes are scattered throughout a 0.8-Mb region of the human Vₜ gene cluster of chromosome 14 and are intermingled with different Vₜ clan genes.
Table 1

V_{H} Genes Used for Phylogenetic Analysis in Figure 4

<table>
<thead>
<tr>
<th>V Gene(s)* (references)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heavy chains:</td>
</tr>
<tr>
<td>Cartilaginous fishes:</td>
</tr>
<tr>
<td>Sharks:</td>
</tr>
<tr>
<td>Nurse shark (<em>Ginglymostoma cirratum</em>)</td>
</tr>
<tr>
<td>Horned shark (<em>Heterodontus francisci</em>)</td>
</tr>
<tr>
<td>Skate:</td>
</tr>
<tr>
<td>Little skate (<em>Raja erinacea</em>)</td>
</tr>
<tr>
<td>Bony fishes:</td>
</tr>
<tr>
<td>Ray-finned fishes:</td>
</tr>
<tr>
<td>Goldfish (<em>Carassius auratus</em>)</td>
</tr>
<tr>
<td>Ladyfish (<em>Elops saurus</em>)</td>
</tr>
<tr>
<td>Channel catfish (<em>Ictalurus punctatus</em>)</td>
</tr>
<tr>
<td>Rainbow trout (<em>Oncorhynchus mykiss</em>)</td>
</tr>
<tr>
<td>Lobe-finned fishes:</td>
</tr>
<tr>
<td>Coelacanth (<em>Latimeria chalumnae</em>)</td>
</tr>
<tr>
<td>Amphibians:</td>
</tr>
<tr>
<td>African toad (<em>Xenopus laevis</em>)</td>
</tr>
<tr>
<td>Reptiles:</td>
</tr>
<tr>
<td>Caiman (<em>Caiman crocodylus</em>)</td>
</tr>
<tr>
<td>Birds:</td>
</tr>
<tr>
<td>Chicken (<em>Gallus domesticus</em>)</td>
</tr>
<tr>
<td>Mammals:</td>
</tr>
<tr>
<td>Human (<em>Homo sapiens</em>)</td>
</tr>
<tr>
<td>Mouse (<em>Mus musculus</em>)</td>
</tr>
<tr>
<td>Rabbit (<em>Oryctolagus cuniculus</em>)</td>
</tr>
<tr>
<td>Light chains (outgroups):</td>
</tr>
<tr>
<td>Horned shark (<em>Heterodontus francisci</em>)</td>
</tr>
<tr>
<td>Human (<em>Homo sapiens</em>)</td>
</tr>
</tbody>
</table>

* An asterisk (*) indicates a cDNA sequence.

Phylogenetic Analysis

In the study of the evolutionary relationships of V_{H} genes for the entire group of vertebrates, we used amino acid sequences rather than nucleotide sequences, because the V_{H} genes are highly divergent and the nucleotide differences at certain positions apparently have reached the saturation level. All amino acid sequences were aligned by using the program CLUSTAL V (Higgins et al. 1992), with some modification by visual inspection.

As mentioned earlier, a V_{H} gene consists of CDRs and FWs. However, CDRs are highly variable and contain many insertions/deletions (indels) when the V_{H} and V_{L} sequences from diverse organisms are included (see fig. 2), so they were excluded from the analysis. In the present analysis even the FW regions showed some indels. All sites showing indels were excluded from the analysis. Therefore, the total number of amino acid sites used for phylogenetic analysis was 64. Three sequences, one from each of the horned shark (Hf1113), little skate (Re20*), and African toad (Xe8), had several indels in FW2 and FW3, and it was difficult to align them with other sequences (fig. 2). Therefore, they were not used for the present analysis. In the study of a “short-term” evolution of V_{H} genes, we used nucleotide sequences, since the extent of sequence divergence was relatively small in this case. Because we were interested in the
The dynamics of the evolution of Vu genes in this study, we included both functional genes and pseudogenes. There was no difficulty in the alignment of the sequences, except in the case of a few pseudogenes. These few pseudogenes were excluded from the analysis.

The phylogenetic analysis of both amino acid and nucleotide sequence data was conducted by using the minimum-evolution method (Rzhetsky and Nei 1992). This method has a solid theoretical basis and is known to be efficient in obtaining correct phylogenetic trees (Rzhetsky and Nei 1993). The evolutionary distances used were the Poisson-correction distance (see Nei 1987, p. 41), for amino acid sequence data, and Jukes and Cantor’s (1969) distance, for nucleotide sequence data. The reliability of the tree obtained was tested by computing the standard error of each interior branch and conducting a t-test (Rzhetsky and Nei 1992).

**Results**

**Evolutionary Relationships of Vh Genes from Diverse Organisms of Vertebrates**

The Poisson-correction distances for all pairs of 55 VH genes, and two outgroup sequences are presented in figure 3, whereas the phylogenetic tree obtained is given in figure 4. Figure 4 shows that the VH genes in vertebrates can be classified into five groups, i.e., groups A, B, C, D, and E. Group A includes all mammalian clan I genes and a few Xenopus genes, whereas group B consists of all mammalian clan II genes and some Xenopus genes. Group C includes mammalian clan III genes and VH genes from chicken, cai man (reptile), Xenopus, coelacanth, and some bony fishes. Group D consists of bony-fish VH genes only, whereas all group E genes are from cartilaginous fishes.

Figure 4 shows that the first evolutionary split of VH genes occurred between group E genes and others. This is reasonable, because the genomic organization of the genes from cartilaginous fishes is different from that of other vertebrates and because cartilaginous fishes separated from the other vertebrates ~450 Mya. The second split of VH genes in the rest of vertebrates occurs between bony fishes (group D) and tetrapods, though group C includes some bony (ray-finned)-fish and coelacanth (lobe-finned fish) genes. However, the confidence probability (CP) (1−t) of group D genes is relatively low (63%). This is caused mainly by the fact that the genes CaVH99A, IpNG64* and Es1450* in this group are relatively closely related to the genes XeVH1, Es1450*, and OmRTVH431 of group C (fig. 3). This somewhat anomalous relationship of genes may be due to either stochastic errors of amino acid substitution or some intergenic recombination or gene conversion events that occurred a long time ago. The next level of splitting is between group A genes and group (B+C) genes.

The group A gene cluster also has a relatively low CP value. This low CP value (64%) is mainly due to the fact that the genes CaVH99A, IpNG64* and Es1450* in this group are relatively closely related to the genes XeVH1, Es1450*, and OmRTVH431 of group C (fig. 3). The final clusters of group B and C genes show a relatively high CP value.

Within each of these groups, there are several clusters that are highly significant. In group E, for example, the skate genes are substantially different from the shark genes and form a cluster with a 99% probability. Apparently, this cluster separated from the shark cluster in an early stage of evolution, and, if more data accumulate and the cluster is still supported, we may have to separate it from the shark group. Note that sharks and skates diverged ~200 Mya (Carroll 1988, p. 62–83). Significant clusters (CP>95%) are also observed in all other groups. An interesting one is the mouse VH cluster...
FIG. 3.—Pairwise Poisson-correction distances for 55 VH and two VL sequences. The five major groups of VH genes and two VL outgroup genes are separated by lines. An asterisk (*) indicates a sequence deduced from a cDNA.
FIG. 4.—Phylogenetic tree of 55 $V_H$ and two $V_L$ (outgroup) genes. There are five major groups (A, B, C, D, and E) of $V_H$ genes identified from this tree, though the grouping is not necessarily statistically significant. The number given to each interior branch is the probability at which the branch length is different from 0 (confidence probability). Confidence probability <60% is not given. The branch lengths are measured in terms of the number of amino acid substitutions, with the scales given below the tree.
Table 2
Jukes-Cantor Distances (×100) for 10 Human Group A VH Genes

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.8</td>
<td></td>
<td>3.5</td>
<td>7.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.18</td>
<td></td>
<td>6.1</td>
<td>7.7</td>
<td>6.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.46</td>
<td></td>
<td>4.0</td>
<td>5.5</td>
<td>4.0</td>
<td>6.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>ψ1.17</td>
<td></td>
<td>10.4</td>
<td>11.0</td>
<td>11.3</td>
<td>12.7</td>
<td>11.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ψ1.24</td>
<td></td>
<td>9.8</td>
<td>10.4</td>
<td>8.7</td>
<td>10.4</td>
<td>9.3</td>
<td>16.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>ψ1.27</td>
<td></td>
<td>21.2</td>
<td>23.8</td>
<td>21.8</td>
<td>26.5</td>
<td>22.5</td>
<td>31.5</td>
<td>29.3</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>ψ1.58</td>
<td></td>
<td>10.4</td>
<td>11.0</td>
<td>9.8</td>
<td>12.1</td>
<td>9.8</td>
<td>16.8</td>
<td>13.2</td>
<td>28.6</td>
</tr>
<tr>
<td>10</td>
<td>5.51</td>
<td></td>
<td>27.2</td>
<td>31.5</td>
<td>29.3</td>
<td>29.3</td>
<td>30.0</td>
<td>37.7</td>
<td>33.0</td>
<td>37.7</td>
</tr>
<tr>
<td>11</td>
<td>Xenopus 11.1b*</td>
<td>50.8</td>
<td>53.7</td>
<td>51.7</td>
<td>48.9</td>
<td>49.8</td>
<td>56.7</td>
<td>53.7</td>
<td>61.0</td>
<td>54.7</td>
</tr>
</tbody>
</table>

NOTE.—"ψ" denotes a pseudogene.

* Outgroup gene.

in group A. Although this cluster belongs to the mammalian clan I, it is quite different from the human gene belonging to the same clan.

An important feature that emerges from figure 4 is that the three mammalian V_H clans separated a long time ago before mammals and amphibians diverged (~350 Mya), because group A, B, and C genes are shared by both mammals and *Xenopus*; that is, mammalian V_H gene clans I-III have coexisted in the genome, for >350 Myr, as separate lineages. Figure 4 shows that group C includes V_H genes from both bony fishes and tetrapods (including mammalian clan III genes) and that separation of tetrapod group C genes from bony-fish group C genes occurred after group A, B, and C genes diverged. This suggests that the divergence of group A, B, and C genes occurred even before bony fishes and tetrapods diverged ~400 Mya. It is impressive that different copies of V_H genes have persisted for such a long time in the tetrapod genome, and this raises a serious question about the importance of concerted evolution in V_H genes.

It should be noted that the mammalian V_H clans I-III may not exist in many mammalian species. Indeed, all V_H genes in the rabbit seem to belong to clan III, and they are relatively closely related to each other in terms of sequence divergence (data not shown). In the rabbit, however, the V_H gene, which is located proximally to the D segment gene cluster, is used with a high frequency (70%-90%) for generating antibody diversity, though there are >100 V_H genes in the genome and the rest of the genes seem to be used with a low frequency (Knight 1992).

A more extreme case of preferential use of one V_H gene is known in chicken. In this organism, only one V_H gene (V_H1) is functional, and the other V_H genes are all pseudogenes (Reynaud et al. 1989; McCormack et al. 1991). These genes again belong to group C (see fig. 4) and are relatively closely related to each other (authors' unpublished data). The V region diversity of this organism is also generated by somatic gene conversion.

Evolutionary Relationships of V_H Genes in the Human Genome

In the study mentioned above, we chose distantly related V_H genes from diverse organisms to study the pattern of long-term evolution. To understand the dynamics of evolutionary change of V_H genes, however, it is necessary to examine the evolutionary relationships of genes within either the same species or closely related species. The most extensive study of the structure and physical map of V_H genes has been done in the human. Matsuda et al. (1993) constructed the physical map of the 0.8-Mb DNA fragment that contains 64 V_H genes, and the nucleotide sequences of these genes are available from the GenBank.

We examined the evolutionary relationships of V_H genes belonging to groups A and C (groups a and c in the notation of Matsuda et al.), as mentioned earlier. Both of these groups include many pseudogenes, but some of the pseudogenes were truncated or unalignable with the functional genes. We used the *Xenopus* gene 11.1b as an outgroup for group A genes and used the caiman gene C3 for group C genes. (Closely related genes are preferable as outgroups.)

The Jukes-Cantor distances for group A genes are presented in table 2, and the phylogenetic tree for these genes is given in figure 5. This figure clearly shows that pseudogenes evolve much faster than functional genes. This is expected because pseudogenes do not have functional constraint and thus accumulate mutations more rapidly than do functional genes (Li et al. 1981; Miyata and Yasunaga 1981; Kimura 1983). This result is again
FIG. 5.—Phylogenetic tree of 10 group A human V<sub>H</sub> genes. All sequences except for Xenopus 11.1b were taken from Shin et al. (1991) and Matsuda et al. (1993). The Xenopus gene used here is the one of the closest outgroup gene (see fig. 4). ψ = Pseudogene. The branch lengths are measured in terms of the number of nucleotide substitutions, with the scales given below the tree. The confidence probabilities of interior branches are indicated by asterisks: * , >90%; and ** , >95%.

inconsistent with the view that the V<sub>H</sub> genes are subject to frequent (germ-line) gene conversion or unequal crossover events. If intergenic gene conversion or unequal crossover occurs very often, we would expect that both functional and pseudogenes would be mixed and that most genes in the genome would have similar nucleotide sequences. The phylogenetic tree in figure 5 does not show this pattern. It is interesting to see that all functional genes have evolved nearly at the same rate and that some functional genes are quite close to each other. Gojobori and Nei (1984) estimated that the rate of nucleotide substitution for FWs is $1.01 \times 10^{-9}$, $0.65 \times 10^{-9}$, and $2.63 \times 10^{-9}$ per site per year per lineage for the first, second, and third codon positions, respectively. Therefore, the average rate is $1.43 \times 10^{-9}$. If this rate is reliable, even the closest pair of functional genes (Hs1.2 and Hs1.8) seem to have diverged $\sim 12$ Mya. The most divergent gene, Hs5.51, seems to have diverged from the other five functional genes $\sim 100$ Mya. This suggests that many functional V<sub>H</sub> genes have persisted for a long time in the human genome. When all group A, B, and C genes are considered together, they have survived for $>400$ Myr, as mentioned earlier.

Figure 6 shows the phylogenetic tree for group C genes (distance estimates are not shown). The pseudogenes in this group also generally evolve much faster than functional genes. There are a few pseudogenes whose branch lengths are nearly the same as those of the functional genes. These pseudogenes probably lost their function very recently. At any rate, this tree also does not support the view that intergenic gene conversion or unequal crossover plays an important role to homogenize the member genes.

The mouse genome is also known to contain many pseudogenes as well as functional V<sub>H</sub> genes. Unfortunately, there seems to be no systematic study of V<sub>H</sub> gene sequences in this organism. However, using GenBank sequences available, we again constructed a phylogenetic trees for 39 group A mouse V<sub>H</sub> genes (data not shown). This tree also showed a tendency for pseudogenes to evolve faster, and there was no clear-cut evidence of sequence homogenization.

Discussion

Concerted Evolution

As mentioned earlier, multigene families are believed to be subject to unequal crossover, which generates new duplicate genes or deletes some extant duplicate genes (Smith et al. 1971; Smith 1974). If this process continues, duplicate genes in a multigene family tend

FIG. 6.—Phylogenetic tree of 28 group C human V<sub>H</sub> genes. All sequences except for Caiman C3 were taken from Matsuda et al. (1993). The Caiman gene used here is the one of the closest outgroup gene (see fig. 4). The confidence probabilities of interior branches are indicated by asterisks: * , >90%; ** , >95%, and *** , >99%.
to have similar nucleotide sequences even in the presence of mutation. More recently, intergenic gene conversion was considered as an additional mechanism acting to homogenize the member genes of a multigene family (Slighlom et al. 1980). These two mechanisms are thought to be the major causes of concerted evolution that homogenizes the member genes of such gene families as RNA genes and histone genes (fig. 7A) (Smith 1974; Ohta 1980; Arnheim 1983).

The concerted (coincidental) evolution was also invoked to explain the evolution of immunoglobulin diversity (Ohta 1980, 1983), as mentioned earlier. In this case, unequal crossover or gene conversion was regarded as a mechanism to increase the genetic diversity (polymorphism) at a locus by introducing new variants from different loci. However, when Gojobori and Nei (1984) conducted a phylogenetic analysis of human and mouse V\textsubscript{H} genes, they noticed that the extent of sequence divergence among different loci is very high and that, if these genes were subject to concerted evolution, the rate of unequal crossover or gene conversion must be very low—two orders of magnitudes lower than that of ribosomal RNA genes. To explain the difference between the two gene families, they proposed that, whereas rRNA genes are subject to purifying selection, V\textsubscript{H} genes are subject to diversifying selection.

Conducting a study on the number of synonymous (\(d_s\)) and nonsynonymous (\(d_N\)) substitutions in CDRs and FWs, Ohta (1992) recently concluded that the amino acid variability in CDRs is primarily caused by gene conversion. This conclusion is based on her observations that \(d_s\) is generally higher in CDRs than in FWs and that \(d_N\) was not necessarily higher than \(d_s\), even in CDRs. She argued that these observations do not support the hypothesis of diversifying selection but are consistent with the gene conversion hypothesis. However, there are three major problems in her study. First, in the study of evolution of V\textsubscript{H} genes, only germ-line sequences should be used; but she used many cDNA sequences. Second, CDRs evolve rapidly and include many indels. Therefore, for the comparison of \(d_s\) and \(d_N\), only closely related sequences should be used. When this is done, \(d_s\) is nearly the same for both CDRs and FWs, and \(d_N > d_s\) in CDRs but \(d_N < d_s\) in FWs (Tanaka and Nei 1989). (In the human, \(d_s\) was higher in CDRs than in FWs in Tanaka and Nei's study; but the difference was not significant.) Third, there are no data showing that gene conversion occurs exclusively in CDRs. Actually, somatic gene conversion occurs in both CDRs and FWs in chicken (McCormack and Thompson 1990). For the above reasons, Ohta's conclusion does not seem to be supported by available data.

Divergent Evolution

In figure 4, we have seen that the major groups of V\textsubscript{H} genes in higher vertebates have been maintained in the genome for \(\geq 400\) Myr. This observation is in rough agreement with the estimate of divergence (300–400 Myr) obtained by Gojobori and Nei (1984) in a statistical analysis of human and mouse V\textsubscript{H} genes (see Addendum in their paper). However, it is in sharp contrast with the evolution of rRNA genes, where even the human and the chimpanzee, which apparently diverged \(~5\) Mya (see Stoneking 1993), do not share the same group of genes (Arnheim 1983). In the present paper, this type of long-term persistence of member genes of a multigene family will be called "divergent evolution," and it is schematically represented in figure 7B.

There are two possible hypotheses to explain this long-term persistence of V\textsubscript{H} genes in the vertebrate genome (Gojobori and Nei 1984); one is to assume that the rate of occurrence of unequal crossover is low in V\textsubscript{H} genes, and the other is the hypothesis that diversifying selection operates among V\textsubscript{H} genes. Gojobori and Nei (1984) estimated the rate of occurrence of unequal crossover by fitting a mathematical theory to the phylogenetic tree for V\textsubscript{H} genes. The rate obtained was \(~100\) times lower than the estimate for the rRNA multigene family. However, since there is no reason to believe that the rate must be lower in V\textsubscript{H} genes than in rRNA genes, they supported the second hypothesis. The recent study of the physical map of V\textsubscript{H} genes by Matsuda et al. (1993) supports Gojobori and Nei's (1984) view, because group A, B, and C V\textsubscript{H} genes are scattered throughout the 0.8-Mb V\textsubscript{H} gene region, suggesting that unequal crossover has occurred relatively frequently in the past (also see Tutter and Riblet 1989b).

The theoretical basis of the second hypothesis is that vertebrate individuals are exposed to a large number of foreign antigens, and thus, if an individual has many

![Figure 7](#)

**Fig. 7.**—Three different modes of evolution of multigene families. 

O = Functional genes; and ● = pseudogenes.
different types of $V_H$ genes, it will enhance the fitness of the individual. This means that a new mutant $V_H$ gene that encodes an antibody molecule matching for a new type of antigens has selective advantage and that old genes that have proved to be useful will be retained in the genome. Indeed, Tanaka and Nei (1989) have shown that the rate of amino acid–altering (nonsynonymous) nucleotide substitutions is higher than the rate of synonymous nucleotide substitution in CDRs of the human and mouse $V_H$ genes. In other words, there is positive Darwinian selection operating for the diversification of $V_H$ genes.

**Diversifying selection would certainly increase the persistence time of different $V_H$ genes in the genome, but can it explain the existence of three major groups of $V_H$ genes that have survived for >400 Myr?** Theoretically, it can happen by chance when random occurrence of unequal crossover creates new genes or deletes existing ones. However, it is more likely that these three groups are specialized to cope with different groups of antigens, though there is no evidence at the present time. Kirkham et al. (1992) showed that each of the mammalian clans I–III has a unique FR structure that influences the conformation of the antigen-binding site, and they suggested that there might be a differential use of $V_H$ clans in the immune response.

**Evolution by the Birth-and-Death Process**

In the theory of concerted evolution, it is commonly assumed that the total number of duplicate genes of a gene family remains more or less constant (Ohta 1983). Hughes and Nei (1989) and Nei and Hughes (1991, 1992) showed that, in the major-histocompatibility-complex (MHC) gene family, gene duplication often occurs but that many duplicate genes die out because of deleterious mutation. However, the number of functional genes remains more or less constant, since the effects of the birth and death of duplicate genes are balanced out. The dead genes either stay as pseudogenes in the genome or are eliminated by unequal crossover. They called this the "evolution by the birth-and-death process" (fig. 7C).

Figures 5 and 6 show that evolution by the birth-and-death process also occurs in the $V_H$ gene family. Matsuda et al. (1993) estimated that ~50% of $V_H$ genes that exist in the human genome are pseudogenes. Similarly, the mouse genome harbors many pseudogenes (Rathbun et al. 1989). Since the number of functional $V_H$ genes in the genome is similar for many mammalian species, these results indicate that nearly the same number of functional genes as that of pseudogenes is produced by unequal crossover for a given period of evolutionary time. We can therefore conclude that the $V_H$ gene family is subject to evolution by the birth-and-death process, when a relatively short period of evolutionary time is considered.

From the phylogenetic trees given in figures 5 and 6, we previously concluded that concerted evolution does not play an important role in the $V_H$ gene family. This does not mean that there is no homogenization process in the evolution of $V_H$ genes. As long as unequal crossover occurs, there must be some chance of homogenization. We also do not completely rule out the role of intergenic gene conversion. Because gene conversion is an important mechanism of somatic generation of antibody diversity in some organisms (e.g., chicken), it might have played some role in the evolution of genomic structure of $V_H$ genes as well. Nevertheless, as far as currently available data are concerned, the $V_H$ gene family appears to have evolved mainly following the scheme of divergent evolution and evolution by the birth-and-death process.

**Some Remarks**

It should be emphasized that, although we now have a rough picture of the evolution of $V_H$ genes in vertebrates, this picture is based on studies of a limited number of species. If more species are studied, we may find new types of genomic organization of $V_H$ genes. It is rather surprising that the genomic structure of the coelacanth is somewhat similar to that of cartilaginous fishes. It is also interesting that the rabbit system is similar to the chicken system. Although these systems should be studied in more detail, they suggest that antibody diversity is generated in many different ways and that some systems may have evolved independently in different groups of vertebrates. Obviously, more detailed studies of the genomic structure of $V_H$ genes are necessary in many different organisms before we understand the general pattern of evolution of this multigene family.

Previously we suggested that the long persistence of $V_H$ genes in vertebrate genomes is aided by diversifying selection. However, the extent of selective advantage of a new mutant $V_H$ gene is probably small (possibly 1%–2%). The reason for this is that there are organisms in which only one or a few $V_H$ genes are functional and in which others are either pseudogenes or rarely used. If every gene had a unique function that is essential to the survival of the organism, this type of system would never have evolved; somatic mutation and intergenic gene conversion would be too risky and too uneconomical in this case. The fact that many $V_H$ genes die out in the course of evolution also suggests that the adaptive differences between different $V_H$ genes is rather small.

Nevertheless, this small magnitude of adaptive difference ($s$) would be sufficient to explain the long persistence of $V_H$ genes in the vertebrate genome if the population size ($N$) is sufficiently large, because the
maintenance of V_H genes would be determined mainly by Ns. If the theoretical study on frequency-dependent selection with respect to MHC genes (Takahata and Nei 1990) gives any guidance, it suggests that the persistence of different V_H genes would be enhanced substantially by diversifying selection when Ns > 100.

Acknowledgment

This work was supported by research grants from the National Institute of Health (GM-20293) and the National Science Foundation (DEB-9119802) to M.N.

LITERATURE CITED


SCHIFF, C., M. MILILI, and M. FOUGEREAU. 1985. Functional and pseudogenes are similarly organized and may equally contribute to the extensive antibody diversity of the IgV_HII family. EMBO J. 4:1225–1230.


SCHROEDER, H. W. JR., M. A. WALTER, M. H. HOFKER, A.


JAN KLEIN, reviewing editor

Received November 5, 1993

Accepted January 19, 1994