Coevolution of Immunoglobulin Heavy- and Light-Chain Variable-Region Gene Families

Tatyana Sitnikova and Chen Su
Institute of Molecular Evolutionary Genetics, and Department of Biology, The Pennsylvania State University

The gene families encoding the immunoglobulin variable regions of heavy (V_H) and light (V_L) chains in vertebrates are composed of many genes. However, the gene number and the extent of diversity among V_H and V_L gene copies vary with species. To examine the causes of this variation and the evolutionary forces for these multigene families, we conducted a phylogenetic analysis of V_H and V_L genes from the species of amniotes. The results of our analysis showed that for each species, V_H and V_L genes have the same pattern of clustering in the trees, and, according to this clustering pattern, the species can be divided into two groups. In the first group of species (humans and mice), V_H and V_L genes were extensively intermingled with genes from other organisms; in the second group of species (chickens, rabbits, cattle, sheep, swine, and horses), the genes tended to form clusters within the same group of organisms. These results suggest that the V_H and V_L multigene families have evolved in the same fashion: they have undergone coordinated contraction and expansion of gene repertoires such that each group of organisms is characterized by a certain level of diversity of V_H and V_L genes. The extent of diversity among copies of V_H and V_L genes in each species is related to the mechanism of generation of antibody variety. In humans and mice, DNA rearrangement of immunoglobulin variable, diversity, and joining-segment genes is a main source of antibody diversity, whereas in chickens, rabbits, cattle, sheep, swine, and horses, somatic hypermutation and somatic gene conversion play important roles. The evolutionary pattern of V_H and V_L multigene families is consistent with the birth-and-death model of evolution, yet different levels of diversifying selection seem to operate in the V_H and V_L genes of these two groups of species.

Introduction

A typical immunoglobulin (Ig) molecule is a Y-shaped tetramer composed of two identical heavy (H) chains and two identical light (L) chains. The H and L chains consist of variable (V) and constant (C) domains. Both V_H and V_L domains contain the framework region (FR) and the complementarity-determining region (CDR), where CDRs are sites of interaction with antigens (Kabat et al. 1991). The V_H domains are encoded by variable (V_H), diversity (D_H), and joining-segment (J_H) genes, and the V_L domains are encoded by variable (V_L) and joining-segment (J_L) genes (Tonegawa 1983).

The V_H genes from various vertebrate species have been classified into five groups (A–E), of which three (A, B, and C) are shared by mammals, amphibians, and bony fishes (Tutter and Riblet 1989; Schroeder, Hillson, and Perlmutter 1990; Ota and Nei 1994). The evolution of V_L genes is more complex than that of V_H genes. First, in mammalian species there are two types of light chains: κ and λ. The genes encoding κ and λ chains are located on different chromosomes, and the λ-chain genes are polyphyletic with respect to the κ-chain genes (Rast et al. 1994; Haire et al. 1996). Second, the divergence of mammalian V_L genes seems to have started earlier than that of V_H genes: the presence of V_L-like genes in the genomes of sharks as well as in mammals suggests that the mammalian V_L groups originated more than 470 MYA (Rast et al. 1994).

The level of divergence among copies of V_H and V_L genes varies with species. However, the cause of this variation is not clear. One way to determine the forces affecting the evolution of V_H and V_L multigene families is to examine whether or not there is a correlation in their evolution. If the variation in the number and the diversity of these genes occurs due to random genetic drift, all processes of duplication, gene loss, and sequence divergence of V_H and V_L genes are expected to be independent, given that the V_H, V_L, and V_L multigene families are located on different chromosomes. On the other hand, a correlated pattern of multigene family expansion or contraction (that is, coevolution) of V_H and V_L genes indicates that there are common factors affecting the evolution of these multigene families, such as population dynamics or selection, given that these genes have essentially the same function and are expected to be exposed to similar selective pressure.

In this paper, we conduct a phylogenetic analysis of V_L genes from various species of amniotes to study the evolutionary patterns of these two multigene families and the levels of diversity of V_H and V_L genes in various species. This study will allow us to understand the mechanisms of the evolution of V_H and V_L multigene families and the peculiarities of immune systems in various organisms.

Materials and Methods

In this phylogenetic analysis we used representative functional V_H and V_L sequences from species for which a sufficient number of both V_H and V_L sequences were available as of September 1997. These species are humans, mice, sheep, cattle, swine, horses, rabbits, and chickens. Several cDNA V_H sequences from horses have recently become available (Schrenzel et al. 1997), yet we did not include them in our analysis because they...
contain many deletions in FRs. We also used functional $V_H$ and $V_L$ sequences from Muscovy duck and the domestic duck to study the relationship of these sequences with their counterparts from chickens (no $V_L$ genes have been found in birds). The chicken $V_H$ and $V_L$ pseudogenes were not used, since they are closely related to a single functional chicken $V_H$ or $V_L$ gene, respectively, and they form a tight cluster in the trees (e.g., see Ota and Nei [1995] for the tree for chicken $V_H$ genes).

The DNA sequences of human $V_H$ and $V_L$ genes were obtained from the VBASE database [Tomlinson et al. 1996], whereas most other sequences were retrieved from GenBank. For humans, we used representative sequences from each of the 7 $V_H$ and 10 $V_L$ families with the exception of the single-gene family $V_{10}$, which did not consistently cluster with other sequences and was probably originated by recombination. From the seven human $V_L$ families, we used representative sequences of three major families. (In gene notations for $V_L$ genes [Tomlinson et al. 1996] and for mouse $V_H$ and $V_L$ genes [Strohal et al. 1989; Kofler et al. 1992], the first number designates the family that the gene represents.)

In the analysis, we used germ line sequences of $V_H$ and $V_L$ genes whenever they were available; otherwise, we used cDNAs. (The cDNA sequences are marked by asterisks.) Since somatic mutations preferentially occur in CDRs rather than in FRs (Gojobori and Nei 1986; Reynaud et al. 1995; Ignatovich et al. 1997), and we used FRs for our analysis, the occasional use of cDNA sequences should not affect the results of our analysis. Below are given the names of the sequences used (except human sequences), together with the GenBank accession numbers.

$V_H$ Sequences Used in the Analysis—(1) mouse ($M. musculus$): $V_H^{1a}$ (X02459), $V_H^{1b}$ (Z66551), $V_H^{1c}$ (X02462), $V_H^{1d}$ (146384), $V_H^{2a}$ (M24271), $V_H^{2b}$ (X14511), $V_H^{2c}$ (X14515), $V_H^{3a}$ (Z37144), $V_H^{3b}$ (X03091), $V_H^{4a}$ (X55984), $V_H^{5a}$ (L14737), $V_H^{5b}$ (K02152), $V_H^{6a}$ (M03520), $V_H^{6b}$ (L22748), $V_H^{9a}$ (X70091), $V_H^{9b}$ (X70098), $V_H^{10}$ (M21470), $V_H^{11}$ (M27841), $V_H^{12}$ (X02239), $V_H^{13}$ (X55935), $V_H^{14}$ (X02562); (2) mouse d. ($M. domesticus$): $V_H^{7a}$ (C18973); (3) rabbit ($Oryctolagus cuniculus$): $V_H^{1a}$ (M03702), $V_H^{1b}$ (L14737), $V_H^{1c}$ (M24271), $V_H^{2a}$ (M03702), $V_H^{2b}$ (X03091), $V_H^{3a}$ (Z37144), $V_H^{3b}$ (X02462), $V_H^{4a}$ (X55984), $V_H^{4b}$ (L14737), $V_H^{5a}$ (L14737), $V_H^{5b}$ (K02152), $V_H^{6a}$ (M03520), $V_H^{6b}$ (L22748), $V_H^{9a}$ (X70091), $V_H^{9b}$ (X70098), $V_H^{10}$ (M21470), $V_H^{11}$ (M27841), $V_H^{12}$ (X02239), $V_H^{13}$ (X55935), $V_H^{14}$ (X02562); (3) cattle ($B. taurus$): 1a (U31106), c2/1a (U32249), c8/1a (U32254), c10/1a (U32255), c7/1b (U32253), c9/1b (U32251), c11/1b (U33226), c14/1b (U33229); (4) horse ($Equus caballus$): $pHL2^*$ (L07563), $pHL4^*$ (L07564), $pHL10^*$ (L07570), $pHL11^*$ (L07571), V4A (Home, Ford, and Gibson 1992); (5) rabbit: (O. cuniculus): $V_L^{2}$ (M27840), $V_L^{3}$ (M27841), $pDH7^*$ (D00071), $pDH8^*$ (M25617), $c7^*$ (X57729); (6) chicken ($G. gallus$): $V_L^{1a}$ (M15095); (7) Muscovy duck ($Cairina moschata$): V1 (M25724), V5 (M25720); (8) domestic duck ($A. platyrhynchos$): $\lambda^*$ (X820690); (9) VpreB genes: human: VpreB (M34927), mouse: VpreB1 (X05557).

$V_L$ Sequences Used in the Analysis—(1) mouse ($M. musculus$): $V_L^{4/5}$ (V17 from R. Kofler’s database; see Strohal et al. 1989), $V_L^{23}$ (J00574), $V_L^{28}$ (M24937); (2) rabbit (O. cuniculus): $V_L^{18a}$ (X00977), $V_L^{18b}$ (X02336), $V_L^{19a}$ (X02337), $V_L^{19b}$ (X02338), $V_L^{9b}$ (X14364).

In the tree for $V_H$ genes, horned shark ($Heterodontus francisci$) 101 (X13449) and little skate ($Raja erinacea$) 278 (X15124) $V_H$ sequences were used as outgroups, because they belong to $V_H$ group E, which branched off before the divergence of groups A, B, and C (Ota and Nei 1994). In the tree for $V_L$ genes, horned shark 122 (X15316) and little skate SK102 (U19209) $V_L$ type I sequences were used as outgroups, because elasmobranch type I $V_L$ genes have been shown to be the most ancient group of vertebrate $V_L$ genes (Rast et al. 1994).

Phylogenetic Analysis

The $V_H$ and $V_L$ sequences were aligned by using the CLUSTAL W computer program (Thompson, Higgins, and Gibson 1994) with minor manual adjustments done by using the SEQED computer program (Zarkhikh et al. 1991). In the analysis, we used the framework regions of $V_H$ and $V_L$ sequences, because the CDRs are highly variable, which makes them difficult to align. Furthermore, we used only first and second codon positions of the sequences in distance computation, because third codon positions are saturated due to the ancient divergence of these genes.

Phylogenetic analysis was done using the MEGA computer program ([Kumar, Tamura, and Nei 1993] and the TreePack computer program (I. Belyi, personal communication), which can be obtained from the anonymous ftp account ftp.transarc.com in the directory pub/staff/belyi. First, Kimura’s (1980) two-parameter model was employed to estimate distances between sequences. In computation of distances, we ignored all sites with gaps for each pair of sequences (pairwise deletion option in the MEGA program). This option was appropriate in this case, because the number of gaps was small, and in most cases, the gaps were shared by a group of closely related sequences. We used the neighbor-joining (NJ, Saitou and Nei 1987) method to reconstruct trees. The reliability of the trees was examined by the bootstrap probability (BP) and con-
Results

V<sub>H</sub> Gene Tree

The phylogenetic tree of V<sub>H</sub> sequences from various vertebrate species presented in figure 1 shows that V<sub>H</sub> genes form three clusters that correspond to three V<sub>H</sub> groups (A, B, and C; Ota and Nei 1994). However, various V<sub>H</sub> groups are not necessarily represented in the genomes of all species: humans and mice possess V<sub>H</sub> genes from all three groups, whereas V<sub>H</sub> genes from birds, rabbits, cattle, sheep, and swine are restricted to one V<sub>H</sub> group, with genes of each species forming their own clusters within group B or C. (When included in the analysis, the V<sub>H</sub> pseudogene sequences from chickens form a tight cluster with a single functional chicken V<sub>H</sub> sequence.) The chicken and duck V<sub>H</sub> sequences appear to be closely related and form a reliable cluster in the tree. Interestingly, the rabbit possesses a more heterogeneous V<sub>H</sub> repertoire than do other species with restricted V<sub>H</sub> repertoires, and the cluster of rabbit V<sub>H</sub> sequences is not statistically supported, even though they have been assigned to a single family based on sequence similarity (Knight and Tunyaplin 1995). This disagreement exists because, in rabbit nucleotide sequences, first and second codon positions, which we used for our analysis, are more variable than third codon positions, possibly due to positive selection operating in these genes (C. Su and M. Nei, personal communication). When all three codon positions are used, rabbit V<sub>H</sub> sequences form a reliable cluster in the tree (data not shown).

Interestingly, three antibody species, sheep, cattle, and swine, all have restricted V<sub>H</sub> repertoires. However, two closely related species, sheep and cattle, have genes belonging to group B, whereas all swine sequences belong to group C. This suggests that the common ancestor of artiodactyls must have possessed the V<sub>H</sub> genes from both groups B and C, with group B having been lost in the swine lineage and group C having been lost in the ruminant lineage. In addition, V<sub>H</sub> genes from the horse, a species that is phylogenetically close to artiodactyls (Novacek 1992), also seem to have restricted diversity and belong to group B (data not shown; Navarro et al. 1995; Schrenzel et al. 1997).

V<sub>L</sub> Gene Tree

In the phylogenetic tree for V<sub>L</sub> sequences (fig. 2), we identified six V<sub>L</sub> groups which are not always statistically supported but remain stable in the tree when all available V<sub>L</sub> sequences are used (data not shown). Five V<sub>L</sub> groups contain V<sub>k</sub> sequences and V<sub>pB</sub> sequences (genes expressed in preB-cells), whereas one group contains V<sub>k</sub> genes exclusively. The V<sub>L</sub> groups in our tree are consistent with two classifications of V<sub>k</sub> genes suggested by the earlier studies (Hayzer 1990; Zezza, Stewart, and Steiner 1992). However, in Hayzer's (1990) classification, V<sub>k</sub> groups A, B, and C have been classified as one group and in Zezza, Stewart, and Steiner's (1992) classification, V<sub>k</sub> groups A and B have been placed into one group, and D and E have been placed into another group.

According to our tree, the V<sub>k</sub> sequences from each species have a level of diversity similar to that of their V<sub>H</sub> genes. The V<sub>k</sub> sequences from humans belong to several groups, whereas the sequences from chickens, rabbits, cattle, and sheep are restricted to one V<sub>k</sub> group. The two remaining species, mice and horses, show a somewhat unusual pattern. In mice, in contrast to numerous V<sub>H</sub> genes (see fig. 1), there are only three V<sub>k</sub> genes, two belonging to group C and one belonging to group D. The high diversity of numerous mouse V<sub>k</sub> genes (~140 genes; Kirschbaum, Jaenichen, and Zachau 1996) compensates for the scarcity of V<sub>k</sub> genes and partly explains a highly skewed k:λ chain ratio of expression (95:5). In horses, all but one of the published V<sub>k</sub> sequences belong to group A. Although the remaining gene, V<sub>k</sub>4, belongs to group B, it seems to have low expression, given that it does not code for any cDNA or protein identified so far in horses (Home, Ford, and Gibson 1992). This suggests that, similar to those of sheep and cattle, the horse V<sub>k</sub> repertoire effectively used for antibody production is limited and restricted to V<sub>k</sub> group A. Interestingly, in contrast to sequences from sheep, cattle, and horses, a single V<sub>k</sub> sequence available from swine belongs to group C (data not shown), which suggests that the swine possesses V<sub>k</sub> genes belonging to the family not represented in other artiodactyls, similar to the case observed for swine V<sub>H</sub> genes. However, more data are necessary to draw firm conclusions about the swine V<sub>k</sub> repertoire.

While having a restricted V<sub>k</sub> repertoire, sheep and cattle show some deviation from the clustering pattern observed for their V<sub>H</sub> genes. Even though all sheep and cattle V<sub>k</sub> sequences belong to V<sub>k</sub> group A, they do not form separate clusters as V<sub>H</sub> genes do; instead, they are intermingled with each other and with the horse sequences. Note, however, that two sheep V<sub>k</sub> genes, 5.1 and 16.1, which are responsible for about 50% of the antibodies produced by the organism (Reynaud et al. 1995), are closely related. On the other hand, of the six sheep V<sub>k</sub> families, even the most divergent sheep V<sub>k</sub> family, V, which contains only pseudogenes and therefore is not represented in our tree, still belongs to the V<sub>k</sub> group A (data not shown; see Reynaud, Dufour, and Weill 1997).

Most V<sub>k</sub> genes sequenced so far are from species with high κ-chain expression: humans, mice, and rabbits. In our tree, the V<sub>k</sub> genes from these species show the same clustering pattern as did their V<sub>H</sub> and V<sub>k</sub> genes: the human and mouse V<sub>k</sub> sequences are intermingled with each other and belong to many V<sub>k</sub> groups, whereas the rabbit V<sub>k</sub> genes form one cluster and are restricted to one group (fig. 2). This is consistent with a detailed phylogenetic analysis of V<sub>k</sub> sequences from various spe-
Coevolution of $V_H$ and $V_L$ Multigene Families

The constitution of the immune system varies with species. In table 1, we summarize the results of our phylogenetic analyses and compile the information on the numbers and the levels of diversity of immunoglobulin $V_H$ and $V_L$ genes in the species discussed in the paper. The data in the table show that even though the number of $V_H$ and $V_L$ genes for a particular organism may differ, the levels of diversity exhibited in these two gene families are similar. On this basis, we can identify two groups of organisms. The species of the first group, including humans and mice, contain the diverse repertoires of $V_H$, $V_L$, and $V_K$ genes. In contrast, the species of the second group, including chickens, rabbits, sheep, cattle, swine, and horses, have $V_H$, $V_L$, and $V_K$ genes with a restricted diversity, and all genes sequenced so far in each species belong to one group of $V_H$, $V_L$, or $V_K$ genes, respectively. Since $V_H$, $V_K$, and $V_L$ genes are not completely sequenced in the species of the latter
group, it is not known whether they contain more distantly related functional \( V_H \), \( V_L \), or \( V_K \) genes; however, the cDNA analysis suggests that the genes belonging to a single group are expressed most (e.g., Dufour, Malinge, and Nau 1996; Sinclair, Gilchrist, and Aitken 1997).

These two groups of species also differ by the sites and the stages of B-cell development. In the species of the first group (including humans and mice), the DNA rearrangement of Ig genes and the development of lymphoid stem cells into B cells occur in the bone marrow and continue during the lifetimes of the organisms (Klein 1990, pp. 15–17). Then, B cells migrate to secondary lymphoid organs, where they encounter antigens and undergo antigen-dependent selection. In contrast, in the species of the second group, other lymphoid organs, such as the spleen (in chickens, cattle, and sheep), as well as the blood and yolk sac (in chickens), serve as sites of DNA rearrangement (Masteller et al. 1997; Meyer et al. 1997). Then, B cells migrate to the gut-associated lymphoid tissue (GALT), such as the bursa (in chickens; McCormack, Tjoelker, and Thompson 1991), ileal Peyer’s patches (in sheep and cattle and probably in swine and horses; Reynaud et al.

![Phylogenetic tree of 63 \( V_L \) sequences and 11 \( V_K \) sequences from the species of amniotes. The six \( V_L \) groups and four clusters for sequences from sheep, cattle, horses, birds, and rabbits are indicated by brackets. Two type I \( V_L \) sequences from cartilaginous fishes were used as outgroups. For this analysis, we used first and second codon positions only (150 nt).](https://academic.oup.com/mbe/article-abstract/15/6/617/1026269)
evolutionary scenario of V genes

groups.

V L and J L genes for light chains to generate an antibody repertoire (Mage 1993). In these organs, Ig genes undergo antigen-independent somatic diversification by hypermutation or gene conversion. Interestingly, in these species, a small number of rearrangement events seems to be involved in generation of an antibody repertoire. This is because the time allowed for the colonization of GALT by B cells is very limited, such that only a small number of B-cell progenitors invades GALT and gives rise to the entire B-cell pool of the organism (Masteller et al. 1997; Reynaud, Dufour, and Weill 1997).

In connection with the differences in B-cell development, different strategies are employed to generate diverse antibodies in the two groups of organisms (for reviews see Knight and Tunyaplin 1995; Weill and Reynaud 1996; Butler 1997). The species of the former group rely mostly on the combinatorial rearrangement of various V H , D H , and J H genes for heavy chains and V L and J L genes for light chains to generate an antibody variety. In contrast, in the species of the latter group, the diversity of antibodies tends to be generated mostly by extensive somatic hypermutation or gene conversion. (However, the role of somatic diversification in the generation of antibodies in horses has not been elucidated yet.)

To understand the causal relationships of the characteristics of these two groups of species, let us consider a possible scenario of evolution of V H and V L gene groups.

Evolutionary Scenario of V Genes

Contrary to intuition, the two groups of species characterized by the levels of homogeneity of V H and V L repertoires are not monophyletic. To demonstrate this, let us draw a scheme of evolution of V H gene groups under the assumption that the evolutionary tree of the species studied in this paper is known (fig. 3). (The evolutionary relationships and divergence times for these species were obtained from Novacek [1992], Hedges et al. [1996], and Kumar and Hedges [1998].) This scheme applies to both V H and V L gene families, because the species-specific pattern of the V L gene family is essentially the same as that of V H genes except for the exact number of groups represented in each species.

To account for the presence of many V H groups in both amphibians and mammals (see Ota and Nei 1994), we have to assume that V H groups emerged by gene duplication in the common ancestor of tetrapods. As mentioned earlier, the level of diversity of V H genes is associated with specific sites and stages of B-cell development and specific mechanisms of generation of an antibody variety. Thus, the combinatorial rearrangement of various genes could have been important for antibody production in the ancestral species of tetrapods. However, the site of DNA rearrangement in amphibians is not known, and DNA rearrangement seems to occur in two waves, one at the larval stage and the other after metamorphosis, rather than throughout the life of the organism (Du Pasquier 1993).

In the common ancestor of amniotes, many V H groups have been represented, because at least two mammalian species, humans and mice, retained three V H groups in their genomes. In contrast, in the bird lineage, the V H groups A and B have been deleted from the genome so that a single V H functional gene and all V H pseudogenes from chickens are closely related and belong to V H group C. Further, the restricted V H repertoire is observed in the rabbit and artiodactyl–horse lineages. Therefore, there were three independent events of contraction of V H repertoires (in the avian, artiodactyl–horse, and rabbit lineages), and each gene repertoire contraction is associated with the recruitment of a specific organ for extensive somatic diversification of Ig genes (the bursa, appendix, or ileal Peyer’s patches) (fig. 3).

There are two hypotheses to explain the process of V H and V L gene repertoire contraction in several lineages independently (see also Ota and Nei 1995). The first hypothesis is that in the ancestral species of tetrapods, the mechanisms of antibody generation included both combinatorial rearrangement and somatic diversification of immunoglobulin genes upon antigenic challenge of B cells. (In fact, somatic mutations do occur in immunoglobulin genes of all studied tetrapod groups:

<table>
<thead>
<tr>
<th>Species</th>
<th>V H Genes</th>
<th>V H Families</th>
<th>V H Groups</th>
<th>V L Genes</th>
<th>V L Families</th>
<th>V L Groups</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>95 (41)</td>
<td>7</td>
<td>A, B, C</td>
<td>76 (40)</td>
<td>7</td>
<td>7</td>
<td>51 (30)</td>
</tr>
<tr>
<td>Mouse</td>
<td>&gt;126</td>
<td>15</td>
<td>A, B, C</td>
<td>140</td>
<td>19</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Rabbit</td>
<td>100 (4)</td>
<td>1</td>
<td>C</td>
<td>30</td>
<td>1</td>
<td>A</td>
<td>&gt;5</td>
</tr>
<tr>
<td>Sheep</td>
<td>10</td>
<td>1</td>
<td>B</td>
<td>5</td>
<td>1</td>
<td>A</td>
<td>100 (50)</td>
</tr>
<tr>
<td>Cow</td>
<td>10</td>
<td>1</td>
<td>B</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>&gt;20 (7)</td>
</tr>
<tr>
<td>Horse</td>
<td>Limited?</td>
<td>1</td>
<td>B</td>
<td>&gt;20</td>
<td>1</td>
<td>Da</td>
<td>30</td>
</tr>
<tr>
<td>Chicken</td>
<td>100 (1)</td>
<td>1</td>
<td>C</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>26 (1)</td>
</tr>
</tbody>
</table>

NOTE.—Total numbers of genes are shown (in all species but humans, this number is estimated rather than exact), with the numbers of functional genes in parentheses. Numbers of families were identified based on 80% nucleotide similarity among genes. Groups were identified by phylogenetic analyses (Ota and Nei 1994; Nei, Gu, and Sitnikova 1997; Sitnikova and Nei 1998; this paper). Question marks indicate that numbers of V genes for the light λ and κ chains in pigs and for the light κ chain in cattle have not been estimated. The groups indicated for pig sequences are those to which the single available V L genes belong. Dashes indicate that light κ chains for chicken have not been identified.
amphibians, birds, and mammals [Du Pasquier 1993].) Then, in some species, somatic diversification might also become important for generation of a primary antibody repertoire generated before an antigen encounter. This transition could be associated with restrictions on the time allowed for DNA rearrangement or preferential rearrangement of one gene, say, the closest one to the C<sub>H</sub> (or C<sub>L</sub>) region, such that a small number of rearrangement events become involved in the generation of antibodies. Since the maintenance of a diverse repertoire of V<sub>H</sub> and V<sub>L</sub> genes does not confer a selective advantage to the organism, a loss of diverse functional V genes from the genome may occur. The problem with this hypothesis is that the primary reason for such a transition is unknown, and it is not clear why it happened in some species but not in others.

The second hypothesis is that the gene repertoire contraction in some ancestral organisms could occur due to an extended period of population size reduction (Ota and Nei 1995). Such a population size bottleneck could cause accumulation of deleterious mutations by genetic drift, giving rise to many nonfunctional genes and eventual loss of genes, such that many gene groups become unrepresented. During this period, the organisms may live in a stable environment, in which they do not encounter many antigens and can survive without producing diverse antibodies. Then, the reduction of gene diversity could be compensated by employment of somatic diversification of genes.

Unfortunately, there is not enough evidence either for or against this hypothesis, because no information concerning the history of population size in different phylogenetic lineages is available. However, to test this hypothesis, one might want to study whether or not there is reduced gene diversity in other multigene families, because a force resulting in a loss of Ig gene diversity must have affected other multigene families as well. Still, it is possible that the reduction of Ig gene diversity is not due to a single mechanism; rather, one explanation may apply to one phylogenetic lineage, while another works for other lineages.

Interestingly, all mammalian species with restricted V<sub>H</sub> and V<sub>L</sub> gene repertoires are domesticated animals. However, domestication is unlikely to be the main reason for the reduction in V<sub>H</sub> and V<sub>L</sub> gene diversity, because domestication started relatively recently (e.g., 9,000–11,000 years ago for domestication of sheep; Maijala 1997). On the other hand, the ecology of the
organism may play a role in the establishment of Ig gene repertoire. In this respect, it would be interesting to study Ig genes in cetaceans (whales and dolphins), which are believed to be close relatives of artiodactyls (e.g., Gatesy 1997; Shimamura et al. 1997), yet occupy a distinctly different ecological niche.

Of the mammalian species with restricted V_H and V_L repertoires, sheep, cattle, and horses retained V_H genes belonging to group B, whereas swine V_H genes belong to group C. (A similar observation has been made for V_L genes; see Results.) Surprisingly, this is in contrast to the phylogenetic relationships of these species: sheep and cattle are closer to swine than to horses. The most reasonable explanation for this is that the common ancestor of artiodactyls and horses had V_H genes representing both groups B and C, and the conservation and expansion of a particular group of V_H genes in the genome reflects the adaptation of the immune system to cope with a certain type of parasites or antigens. In fact, the foods and lifestyles of horses, sheep, and cattle are closer to one another than to those of swine, because sheep, cattle, and horses are herbivorous, whereas swine are omnivorous. Therefore, sheep, cattle, and horses are likely to encounter similar types of antigens, which are different from the antigens to which swine are exposed.

Models of Evolution of V_H and V_L Gene Families

The model that is most consistent with the observed evolutionary pattern of V_H and V_L gene families is the birth-and-death model of evolution (Nei and Hughes 1992; Ota and Nei 1994; Nei, Gu, and Sitnikova 1997; Sitnikova and Nei 1998). It is also similar to the “accordion model,” proposed to explain evolution of MHC genes (Klein et al. 1993). According to the birth-and-death model, some genes are maintained in the genome for a long time, while other genes of a multigene family are deleted or become nonfunctional by deleterious mutations. The persistence of diverse V_H and V_L genes, as well as the presence of numerous nonfunctional genes in the genomes of vertebrate species (e.g., about 40% of V_H, V_s, and V_k genes in the human genome are pseudogenes), is in agreement with the predictions of this model.

However, the high diversity of V_H and V_L genes cannot be accounted for simply by chance; rather, the action of diversifying or directional selection on these genes has to be assumed (Tanaka and Nei 1989; Nei, Gu, and Sitnikova 1997). Thus, in the species of the first group (humans and mice) the selection operates to maintain numerous and diverse copies of V_H and V_L genes. In contrast, the species of the second group (chickens, rabbits, sheep, cattle, horses, and swine) did not retain diverse repertoires of V_H and V_L genes in their genomes due to relaxation of diversifying selection, caused either by enabling somatic diversification of their genes or by a population size bottleneck at some point in their evolution.

Acknowledgment

This study was inspired by discussions with Masatoshi Nei and supported by NIH and NSF grants to M. Nei. We thank Anatoly Ruvinsky for discussion of genetic diversity and lifestyles of domesticated animals. We are grateful to Masatoshi Nei, Andrey Rzhetsky, Andrey Zharkikh, Andy Clark, George Zhang, and Kei Takahashi for their comments on earlier versions of the manuscript. We also thank two anonymous reviewers and Naoyuki Takahata for their comments.

LITERATURE CITED


