Origin and Evolution of the Vertebrate Immune System

Edited by L. Du Pasquier and G.W. Litman

With 81 Figures and 17 Tables

Springer 2000
Evolution of Vertebrate Immunoglobulin Variable Gene Segments
T. Ota1, T. Stnikova2, and M. Nei3

1 Introduction .................................................. 221
2 Human Ig Genes ............................................... 222
  2.1 Gene Organization and Germine Diversity of V Gene Segments ...... 222
  2.2 Polymorphism of V Gene Segments ................................ 225
  2.3 Somatic Generation of Ig Diversity during B-Cell Development ..... 226
  2.4 Ig Repertoire and Antigen Recognition ............................ 227
  2.5 Evolution of Human V Gene Segments ............................. 228
3 Cartilaginous Fish Ig Genes .................................... 229
  3.1 Gene Organization and Generation of Diversity .................... 229
  3.2 Evolution of Cartilaginous Fish V Gene Segments ................. 230
4 Chicken Ig Genes ............................................... 232
  4.1 Gene Organization and Generation of Diversity .................... 232
  4.2 Evolution of Chicken V Gene Segments ............................ 233
5 Evolution of V Gene Families .................................... 234
  5.1 Evolution of V\textsubscript{H} Gene Families ....................... 236
  5.2 Evolution of V\textsubscript{L} Gene Families ....................... 238
  5.3 Coevolution of V\textsubscript{H} and V\textsubscript{L} Gene Segments ...... 238
6 Evolution of Immune System and Ig Diversity ....................... 239
7 Summary ....................................................... 240
References ...................................................... 241

1 Introduction

Immunoglobulins (Igs), also known as antibodies, play major roles in the vertebrate humoral immune system. They recognize and bind to foreign antigens, such as viruses, bacteria and parasites, and initiate a series of immunological responses.

---

1 Center for Human Genetics, Boston University School of Medicine, 715 Albany St., Boston, MA 02118, USA
e-mail: Ota@soken.ac.jp
2 Eisai Merrimack Valley Lab Inc., 160 Federal St., Andover, MA 01810, USA
3 Institute of Molecular Evolutionary Genetics and Department of Biology, The Pennsylvania State University, 328 Mueller, University Park, PA 16802, USA
(effector function). The dual function of the Ig protein is facilitated by its unique structure consisting of two functionally distinct domains, i.e., the variable (V) domain for antigen recognition and the constant (C) domain for effector function (Frazer and Capra 1998). Ig is generally composed of two identical heavy (IgL) and two identical light (IgL) chains, both of which contribute to the formation of V and C domains.

The diversity necessary for recognizing a variety of antigens is generated by genetic and somatic processes (Max 1998). In general, organisms maintain large multigene families with diversified members to code for numerous V domains and the diversity is further enhanced by somatic mechanisms, such as recombination, imprecise joining of gene segments, hypermutation, and gene conversion. As illustrated in Fig. 1, at least three fundamentally different systems of generating diversity have been found in vertebrates. In cartilaginous fishes clusters of gene segments are repeated many times and chromosomally dispersed in their genome, and each cluster produces either IgL or IgH (Litman et al. 1993). In mice and humans two or three multigene families of gene segments are present in separate arrays and one member of each gene segment family is selected to join together to produce a functional Ig V domain (Tonegawa 1983). In chickens only one "functional" V gene segment is present in the genome, but it is somatically modified by gene conversion after gene rearrangement of the functional gene segments (Reynaud et al. 1994).

Recent progress in comparative immunology has shown that vertebrates share some basic immunological features, such as the presence of Ig, T-cell receptor and major histocompatibility complex, but that distinct molecular and developmental immune mechanisms have been established in different evolutionary lineages (Du Pasquier and Flajnik 1998). Here we review the humoral immune system of human and other vertebrates and discuss the evolution of Ig diversity generation and its effect on the evolution of V gene segments.

2 Human Ig Genes

2.1 Gene Organization and Germline Diversity of V Gene Segments

Human and mouse are the best characterized vertebrates in terms of the molecular genetics of Ig genes and the developmental biology of B-cells in which Ig genes are expressed. In both species Ig V region is produced by the somatic rearrangement of distinct gene segments (Tonegawa 1983). The V region of IgH is encoded by variable (VH), diversity (D), and joining (JH) gene segments, whereas the V region of IgL is encoded by variable (VL) and joining (JL) gene segments. The gene organization of human Ig loci has recently been clarified (Frippiat et al. 1995; Kawasaki et al. 1997; Matsuda et al. 1998; Tomlinson and Cook 1997; Zachau 1995). The human IgH locus, located at chromosome 14q, contains 123 VH
Evolution of Vertebrate Immunoglobulin Variable Gene Segments

(a)

(b)

(c)

Fig. 1a–c. Immunoglobulin light-chain (IgL) gene organization. a Cartilaginous fishes. b Mice. c Chickens. Similar gene organizations exist at the heavy-chain loci of respective species with presence of D gene segments between V and J gene segments. Cartilaginous fishes have the cluster type gene organization and each cluster is capable of producing a unique IgL chain. Mice have the translocon type gene organization, where one V and one J gene segment are joined by gene rearrangement during B-cell lymphogenesis. Combination of different gene segments generates the primary diversity. Chickens have a single functional V gene segment and a single J gene segment. The rearranged region is later modified by somatic gene conversion with pseudogenes as donors.

(MATSUDA et al. 1998). 27 D (CORBETT et al. 1997), and nine J_H gene segments (RAVETCH et al. 1981). Of the 123 V_H gene segments, 44 are potentially functional, but the remaining genes are likely to be pseudogenes. Additional V_H and D gene segments are present on chromosomes 15 and 16 and are referred to as orphan genes (NAGAOKA et al. 1994; TOMLINSON et al. 1994). The orphan genes do not
contribute to Ig production and probably represent evolutionary relics of recent duplicated genes. Human IgL is encoded by two loci (κ and λ loci), only one of which is exclusively used in a given B-cell (isotype exclusion). The human Igκ locus located on chromosome 2p contains 76 Vκ and 5 Jκ gene segments (SCHABLE and ZACHAU 1993; ZACHAU 1995). Of the 76 Vκ gene segments, 35 are potentially functional genes. The human Igλ locus is located on chromosome 22q and contains more than 69 Vλ and 7 Jλ gene segments (FRIPIAT et al. 1995; KAWASAKI et al. 1997). Of the more than 69 Vλ gene segments approx. 30 are potentially functional (WILLIAMS et al. 1996). Orphan Vκ and Vλ gene segments have been also found on chromosomes 1, 2, and 22 (ZACHAU 1995) and on chromosome 8 (FRIPIAT et al. 1997; QUEIROZ et al. 1997), respectively.

The V_H and V_L gene segments show great diversity and are subdivided into families according to sequence similarity. Generally, V gene segments with more than approx. 80% nucleotide identity are considered to belong to the same family (BRODEUR and RIBLET 1984) and the human V gene segments are classified into seven V_H, five V_K, and ten V_L families (MATSUDA and HONJO 1996; ZACHAU 1995; WILLIAMS et al. 1996). Four V_H gene segments can be classified into two additional V_K families, but their functionality has not been well established. The number of member genes in a V gene family varies from one (V_H6, V_K4, V_K5, V_L6, V_L8, V_L9) to 65 (V_H3). Most V gene families contain pseudogenes, and in a few V gene families more than two-thirds of member genes are pseudogenes.

The V domain is divided into the complementarity determining regions (CDRs) and the framework regions (FRs; WU and KABAT 1970). CDRs form the hypervariable loops and correspond to antigen binding regions, whereas FRs encode the two β-pleated sheets of the Ig fold. Four FRs are separated by three CDRs at the primary nucleotide sequence, but CDRs are brought to the C-distal end of Ig domain to form the antigen binding pocket in the tertiary structure. The first three FRs (FR1, FR2, FR3), the first two CDRs (CDR1, CDR2) and a small portion of the third CDR (CDR3) are encoded by V_H or V_L gene segments, whereas the rest of CDR3 and the fourth FR (FR4) are encoded by DJ_H or J_L gene segments (MAX 1998).

CDRs are highly variable in terms of sequence length and amino acid composition. The loop structures formed by CDRs are often classified into a set of main chain structures called canonical structures (CHOTHIA et al. 1987). The number of amino acids involved in the loop formation and the presence of certain amino acids at structurally critical sites in CDR primarily determine canonical structures. The canonical structures formed at the antigen binding pocket provide valuable clues on the antigen specificity of V gene segments (VARGAS-MADRAZO et al. 1995; WEBSTER et al. 1994). For example, the canonical structures with long loop form small pocket at the antigen binding sites and are more often used in the recognition of small molecules such as hapteners, whereas the canonical structures with short loop are more adapted for recognition of large molecules such as proteins. In humans, one to six different canonical classes have been identified in CDR1 and CDR2 of V_H and V_L domains, and a structurally related set of canonical structures are often found for a given V gene family (AL-LAZIKANI et al. 1997; VARGAS-MADRAZO et al. 1996).
2.2 Polymorphism of V Gene Segments

In the V gene segment region of human Ig loci two types of polymorphism are mainly observed. One is single nucleotide polymorphism and the other is V gene segment insertion/deletion polymorphism. The extensive list of single nucleotide polymorphism is reported in VBASE (Tomlinson et al. 1996a) and summarized in Table 1. Excluding a few alleles that have deleterious mutations or small nucleotide deletion in CDRs, only 146 nucleotide differences were observed at 105 functional V gene segment loci. This level of polymorphism is much less than those of the classical MHC locus but is more comparable to the low nucleotide diversity found at many nuclear gene loci in human populations (Li and Sadler 1991). Of the 146 nucleotide differences, approximately one-third of the differences were synonymous and the rest were amino acid altering differences. Half the amino acid altering differences were observed in CDRs despite much shorter amino acid lengths in CDRs than in FRs. The higher ratio of amino acid altering/synonymous nucleotide differences is apparent in CDRs of VH gene segments, and it suggests a role of positive selection to promote the amino acid variation at the antigen binding sites (Tanaka and Nei 1989).

V gene segment insertion/deletion polymorphisms were found at both IgH and IgL loci. In the IgH locus genomic characterization of the entire V gene segment region has revealed both single- and multiple- V gene segment insertion/deletion polymorphism at multiple locations (Matsuda and Honjo 1996). In the IgK locus about 5% of individuals do not have a region called “d contig”, which contains 33 VK gene segments of recent duplicates. The insertion/deletion of V gene segments has

<table>
<thead>
<tr>
<th>Table 1. The number of codon differences observed among alleles of V gene segments reported in VBASE (from Tomlinson 1996a)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Synonymous</td>
</tr>
<tr>
<td>differencea</td>
</tr>
<tr>
<td>Amino acid</td>
</tr>
<tr>
<td>altering differencec</td>
</tr>
</tbody>
</table>

a Only the allelic differences between functional V gene segments were counted. If the same difference was observed between different pairs of alleles for a given locus, it was counted only once for the difference.
b The numbers in parentheses are the numbers of V gene segment loci at which more than two nucleotide sequences have been reported independently.
c Most codon changes are due to a single nucleotide difference but a few codon changes are due to two or three nucleotide changes. In total, 146 nucleotide differences are observed in 134 codon differences.
little effect on individual's health, indicating that the V gene repertoire is functionally redundant to some degree (PARGENT et al. 1991; SCHMIDLE et al. 1993).

2.3 Somatic Generation of Ig diversity during B-Cell Development

In order to understand the evolution of V gene segments it is important to know how the V gene segment repertoire plays a role in the recognition of foreign antigens. In the following two sections we briefly review how the V gene segments are somatically modified during B-cell development and how the diversity created is employed for antigen recognition.

The development of B-cell can be divided into two discrete stages: (a) the lymphopoietic development stage where B-cells develop from hematopoietic stem cells (see MELCHERS and ROLINK 1998 and references therein); and (b) the B-cell activation stage where Igs with high affinity against antigens are produced and B-cells differentiate into plasma and memory B-cells (DEFRANCO 1998; PICKER and SIEGELMAN 1998). In the first stage Ig loci are subjected to gene rearrangement, and the primary Ig repertoire is generated independent of antigens. It also has a mechanism built in to express only one allele of IgH and of IgL, a process known as allelic exclusion. In the second stage somatic mutation and clonal selection reshape the primary repertoire in the presence of antigens.

During the lymphopoietic development the Ig diversity is furnished by combinatorial and joining diversity (see MAX 1998 and references therein). Recombination signal sequences (RSSs) juxtaposed to the VH, DJ, VL, and JL gene segments are recognized by a recombination complex and one gene segment from each gene family is selected to be joined to form the V region of IgL or IgH. Selection of gene segments is biased but the combination of functional gene segments generates substantial diversity. Despite high specificity of RSS mediated joining the recombination process is imprecise, and nucleotide insertion/deletion is commonly observed at the junction region. Nucleotide deletion is presumably due to exonuclease activity of the gene recombination complex, whereas nucleotide insertion is due either to nucleotide addition caused by terminal deoxynucleotidyl transferase (N diversity) or to palindromic nucleotide addition (P diversity) caused by repairs of asymmetrically cleaved intermediate hairpin structures generated during the gene rearrangement. Insertion/deletion of nucleotides also results in the use of the D gene segment in three different reading frames. The combination of gene segments and the joining diversity created during gene rearrangement provide extensive diversity in CDR3.

Somatic mutation observed in the secondary response is highly specific but its molecular mechanism is not yet well understood (Berek 1998; JOLLY et al. 1996). The somatic mutation targets mostly the V region irrespective to FRs or CDRs but not C region, and it is mainly of point mutation and nucleotide insertion/deletion is rarely observed. However, it is efficient to attain affinity maturation due to the fact that a few critical amino acid changes may increase its affinity tremendously and somatic mutation/selection can undergo repeatedly. It is also noteworthy that the
human virus has been known to have crossreacting epitope. The rabies virus is a
typical TD antigen and the Ig response is examined after three sequential injections
of the virus vaccine. Although only ten Ig variants are reported, no \( V_H \) gene
segment seems to be used multiple times and variable \( V_\kappa \) and \( V_\lambda \) gene segments
were used in the response. Seven of nine \( V_H \) gene segments belong to the \( V_{H3} \) gene
family, partly due to the large size of \( V_{H3} \) gene family and partly due to the
preferential expression of \( V_{H3} \) gene family. High-affinity binding capability was
found among the IgGs against various component of rabies virus vaccine. Sequence
analysis of \( V_L \) regions showed the accumulation of 2 to 14 somatic nucleotide
substitutions from candidate \( V_L \) gene segments, where the substitutions favor
amino acid replacements in CDRs in a few cases. The results suggest that mutation
and clonal selection are indeed important in TD responses, and the diversified
germline V gene repertoire can be used to generate high-affinity IgGs against various
epitopes for an antigen.

It is clear that the germline V gene segment repertoire, combinatorial, and
joining diversity provides the primary diversity and that the somatic mutation and
clonal selection reshape it to increase affinity and/or avidity. In some cases, however,
only a limited number of V gene segments have significant effect on the
antigen recognition, as seen in some TI antigen response. Therefore the individual
V gene segments and the entire repertoire encoded by V gene families are equally
important factors for the immune response.

2.5 Evolution of Human V Gene Segments

As illustrated by previous phylogenetic analyses (Matsuda and Honjo 1996;
Matsuda et al. 1998; Ota 1997; Ota and Nei 1994; Siznikova and Nei 1998),
many duplications of V gene segments have occurred during the evolution. The first
duplication of gene segments to lead to major lineages of human \( V_H \) and \( V_L \) gene
families likely occurred more than 350 and 470 million years (MY) ago, respec-
tively, and others, such as the duplication of "\( d \) contig" in IgK, occurred about
1–2MY ago. The duplication of gene segments has provided opportunities in which
mutation could create new variants that were selected for or against depending on
the nature of products. As seen in the case of A2 \( V_\kappa \) gene segment, some V gene
segments would have provided immediate selective advantage due to the high
affinity against specific antigens in germline configuration or with its little modifi-
cation. Positive selection has apparently played a role in the fixation of such var-
iants, as indicated by higher amino acid altering nucleotide substitution than one
expected from neutral evolution (Tanaka and Nei 1988; Siznikova and Nei 1998).
The extensive diversity in CDRs is thus partly attributable to the positive selection
operating on CDRs to enhance the antigen recognition against infectious agents.
The great amount of mutation is, on the other hand, deleterious because of its
random nature of mutation and has generated pseudogenes. In many instances the
V gene segments are functionally redundant, whether it is due to the presence of
recent duplicates or due to the presence of V gene segments that can be used against
similar or different epitope of same antigens. In this regard, deleterious mutations are likely subject to random genetic drift rather than purifying selection. This explains the abundance of pseudogenes among the human Ig loci. Eventually these pseudogenes were eliminated from the genome by deletion or rapid accumulation of mutations. Considering continuous generation of duplicated genes and the removal of genes from the genome, the evolution of V gene segments is a dynamic process, and their repertoire is constantly subject to change. Theoretically this type of evolutionary process results in expansion and contraction of gene families and eventually homogenization of gene family (concerted evolution). However, in reality there is no indication of concerted evolution among V gene families. In contrast, the various groups of V gene families have been present for long evolutionary time. This implies the significant role of diversifying selection to preserve the germline V gene segment repertoire. Restricted sets of canonical structures are often associated with a V gene family, and it is likely that the maintenance of V gene family results in the preservation of diverse sets of canonical structures, which in turn provides the diversity to recognize distinct classes of epitopes. In summary, the evolution of human V gene segments is characterized by two evolutionary processes: (a) the generation of gene segments by gene duplication and the elimination of gene segments by accumulation of deleterious mutations or by deletion (the evolution by “the birth and death process”), and (b) diversifying selection, which emphasize the importance of purifying selection to maintain the established V gene segment repertoire and positive selection to enhance the repertoire (Ota and Nei 1994; Nei et al. 1997; Sītnikova and Nei 1998).

3 Cartilaginous Fish Ig Genes

3.1 Gene Organization and Generation of Diversity

Cartilaginous fishes possess a distinct Ig gene organization (Litman et al. 1993; Machalonius et al. 1998). Four or three types of gene segments are used as in the case of higher vertebrate but hundreds of clusters of V, (D.) J, and C are present at multiple sites. Somatic gene rearrangement occurs within a cluster, and in many instances gene segments in a cluster are joined at germline level. Overall combinatorial and joining diversity is thus limited. Somatic mutation is known to occur at the Ig loci of cartilaginous fishes, but the significance of somatic mutation remains uncertain, since it is unknown whether allelic exclusion is present in the organism (Hsu 1998). Clonal selection without allelic exclusion would be hardly effective to attune affinity maturation and potentially increase the risks of producing autoantibodies. The presence of germline joined functional genes specifically rejects the allelic exclusion models proposed for mammals (Alt et al. 1981; Coleclough et al. 1981).
similar or different epitope of same antigens. In this regard, deleterious mutations are likely subject to random genetic drift rather than purifying selection. This explains the abundance of pseudogenes among the human Ig loci. Eventually these pseudogenes were eliminated from the genome by deletion or rapid accumulation of mutations. Considering continuous generation of duplicated genes and the removal of genes from the genome, the evolution of V gene segments is a dynamic process, and their repertoire is constantly subject to change. Theoretically this type of evolutionary process results in expansion and contraction of gene families and eventually homogenization of gene family (concerted evolution). However, in reality there is no indication of concerted evolution among V gene families. In contrast, the various groups of V gene families have been present for long evolutionary time. This implies the significant role of diversifying selection to preserve the germline V gene segment repertoire. Restricted sets of canonical structures are often associated with a V gene family, and it is likely that the maintenance of V gene family results in the preservation of diverse sets of canonical structures, which in turn provides the diversity to recognize distinct classes of epitopes. In summary, the evolution of human V gene segment is characterized by two evolutionary processes: (a) the generation of gene segments by gene duplication and the elimination of gene segments by accumulation of deleterious mutations or by deletion (the evolution by “the birth and death process”), and (b) diversifying selection, which emphasize the importance of purifying selection to maintain the established V gene segment repertoire and positive selection to enhance the repertoire (OTA and NEI 1994; NEI et al. 1997; SITNIKOVA and NEI 1998).

3 Cartilaginous Fish Ig Genes

3.1 Gene Organization and Generation of Diversity

Cartilaginous fishes possess a distinct Ig gene organization (LITMAN et al. 1993; MACHALONIS et al. 1998). Four or three types of gene segments are used as in the case of higher vertebrate but hundreds of clusters of V, (D), J, and C are present at multiple sites. Somatic gene rearrangement occurs within a cluster, and in many instances gene segments in a cluster are joined at germline level. Overall combinatorial and joining diversity is thus limited. Somatic mutation is known to occur at the Ig loci of cartilaginous fishes, but the significance of somatic mutation remains uncertain, since it is unknown whether allelic exclusion is present in the organism (HSU 1998). Clonal selection without allelic exclusion would be hardly effective to attain affinity maturation and potentially increase the risks of producing autoantibodies. The presence of germline joined functional genes specifically rejects the allelic exclusion models proposed for mammals (ALT et al. 1981; COLECLOUGH et al. 1981).
3.2 Evolution of Cartilaginous Fish V Gene Segments

Variable numbers of \( V_H \) and \( V_L \) gene families are found among cartilaginous fishes (Table 2). Figure 2 presents a phylogenetic tree of cartilaginous fish \( V_H \) gene segments. There are three distinct major clusters of \( V_H \) gene segments, one (group E) associated with \( C_\mu \) gene segment and the others (groups F and G) mostly associated with a second C gene segment (\( C_\omega/\text{IgNARC}/C_\kappa \)), as noted by Shen et al. (1996). The only exception to this pattern of association of \( V_H \) gene segments with C gene segments is the monotypic horned shark \( V_H \) gene family represented by Hfr1113. The C gene segment associated with Hfr1113 is partly characterized and is only slightly different from other \( C_\mu \) gene segments (Hinds-Frey et al. 1993). However, excluding this exception, the coevolution of the \( V \) gene family and C gene segments generally exists and the similar association is also found in cartilaginous fish IgL (Rast et al. 1994). These results suggest that the gene segments have been

<table>
<thead>
<tr>
<th>Species</th>
<th>( V_H ) gene family</th>
<th>( V_L ) gene family</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horned shark</td>
<td>2</td>
<td>3</td>
<td>Litman et al. 1993; Rast et al. 1994</td>
</tr>
<tr>
<td>Carcharhinine shark</td>
<td>7</td>
<td>1</td>
<td>Berstein et al. 1996; Hohman et al. 1993</td>
</tr>
<tr>
<td>Little skate</td>
<td>2</td>
<td>2</td>
<td>Shen et al. 1996</td>
</tr>
<tr>
<td>Spotted ratfish</td>
<td>4</td>
<td>2</td>
<td>Rast et al. 1994; Rast et al. 1994, 1998</td>
</tr>
</tbody>
</table>

Fig. 2. A phylogenetic tree of 69 cartilaginous fish \( V_H \) gene segments and 2 outgroup \( V_L \) gene segments. The phylogenetic tree was estimated by neighbor-joining method (Saitou and Nei 1986) with Poisson correction distances. Only amino acid sequences in FRs were used in the analysis. The branch lengths are measured in terms of the number of amino acid substitutions per site, with the scale given below the tree. Asterisks, probabilities at which the branch length estimation is different from zero (confidence probabilities): *, > 90%; **, > 95%; ***; 99%. Most sequences are retrieved from GenBank (accession numbers: Horned shark (Heterodontus francisci): HfrF101 (X13449), HfrH101 (Z11788), HfrH113 (Z11791), HfrH115 (Z11782), HfrH116 (Z11790), HfrH117 (M12158), HfrX (Z11789), HfrH113 (Z11776), Hfr1135 (X13447), Hfr11320 (X13448), Hfr1245 (Z11777), Hfr1224 (Z11783), Hfr1221 (Z11780), Hfr11221 (Z11785), Hfr1224 (Z11784), Hfr11221 (Z11787), Hfr11221 (Z11781), Hfr11221 (Z11802), Hfr11223 (Z11860), Hfr11241 (Z11778). Nurse shark (Ginglymostoma cirratum): GoA4-2 (M92851), Go1G2NARC (U51450). Bull shark: (Carcharhinus leucas): CluT2U (U50613), CluT3U (U50612). Carcharhinus plumbeus: ClpRACE (U50606), Cpl1spleen (U50609), Cpl2spleen (U50607), Cpl4RACE (U40560), Cpl6spleen (U50604), Cpl7RACE (U50610), Cpl7spleen (U50608), Clp8RACE (U50605). Clearnose skate (Raja erinacea): RegVCC2 (U08010), Little skate (Raja erinacea): RcrRe20 (M92672), RcrRe102 (X16146), RcrRe107 (X15124), Spotted ratfish (Hydrolagus colliei): HcoCos5 (AF003841), HcoCos7 (AF003964), HcoCos441 (AF003861), HcoCos442 (AF003839), HcoCos459 (AF003860), HcoCos46 (AF003853), HcoCos431 (AF003835), HcoCos1432 (AF003855), HcoCos1433 (AF003856), HcoCos1434 (AF003857), HcoCos1435 (AF003858), HcoCos1436 (AF003859). Other carcharhinine shark IgM sequences are taken from Shen et al. (1996). Two cartilaginous fish \( V_L \) sequences, HfrII (L25560), RerII (L25556), were used as outgroup.
rarely exchanged between clusters, and that the duplication of entire cluster rather than duplication of individual gene segments have been more important during cartilaginous fish evolution. Duplication of individual gene segments has apparently occurred in the ratfishes, but the V_H gene segments unlinked with C gene segment appear to be pseudogenes (Rast et al. 1998).

In general, the number of V gene families for a given C gene segment isotype is limited. So far only one V gene family has been found for *Raja* type I and type II IgLs, carcharhine shark type II IgL, nurse shark type III IgL, and ratfish type II IgL, with exception of Hco712 (Anderson et al. 1995; Barre et al. 1994; Hohman et al. 1995; Rast et al. 1994). In horned shark a single major and second monotypic
$V_H$ gene families were found in association with $C_m$ (Litman et al. 1993). Even though carcharhine shark $V_H$ gene segments can be classified into seven $V_H$ gene families, they are basically separated into two groups, one group consisting of six closely related $C_m$-associated $V_H$ gene families and the other group of one $C_m$/IgNARC/C$_s$-associated $V_H$ gene family (Fig. 2). Despite of the limited diversity among $V$ gene families, however, the amount of variation within a $V$ gene family is not negligible (Kokubu et al. 1988a; Anderson et al. 1995; Shen et al. 1996). A high frequency of amino acid altering nucleotide differences are found in CDRs, and the positive selection seems to have operated on the diversification of antigen recognition sites in cartilaginous fishes as well (Anderson et al. 1995; Ota et al. 1995; Matsunaga and Anderson 1997).

One intriguing question in term of cartilaginous fish Ig evolution is how the homology of C gene segments has been maintained among different clusters to sustain its capacity for effective signaling to downstream host defense molecules. Nucleotide sequences among $C_m$ gene segments of different clusters show an average of approx. 90% identity in horned sharks (Kokubu et al. 1988b). The fact that a single $V_H$ gene family or closely related $V_H$ gene families are generally associated with an isotype suggests that the homology of C gene segments has been gained by the continuous and gradual turnover of Ig clusters. That is to say, new cluster of gene segments have been constantly created by gene duplication, while other clusters have been lost by deletion or by accumulating by deleterious mutations (the evolution by the “birth and death process”). Although positive selection has been operating at $V$ gene segments to diversify CDRs to some degree, extensive diversity would have been hardly accumulated to establish extensive $V$ gene families due to turnover of clusters.

4 Chicken Ig Genes

4.1 Gene Organization and Generation of Diversity

Chicken Ig gene organization is unique and contains single functional $V$ and $J$ gene segments at the IgL and IgH loci (McCormack et al. 1991; Reynaud et al. 1994). Many “pseudo”-$V$ gene segments are present upstream of the functional $V$ gene segment, but the lack of proper RSSs makes it impossible for them to be rearranged with D or $J$ gene segments. These pseudo-$V$ gene segments often contain no leader peptide exon, are truncated at 5' and 3' coding region, and carry nonsense mutation. The use of a single $V$ and a single $J$ gene segment originally results in the production of almost identical Igs, which may play a role in proliferation of B-cells at early development. The fusion of multiple D segments is known to occur (Reynaud et al. 1989; Mankikka and Toivanen 1991) but highly homologous D gene segments, the exclusive use of IgL, and the absence of N diversity limit the diversity of the primary antibody repertoire. Although the humoral immune system
in chicken may be inefficient against T1-2 antigens (Jeurissen et al. 1998), it seems to be effective against TD antigens. As it turns out, in chicken, the Ig diversity is mainly generated by somatic mechanisms, i.e., somatic gene conversion and somatic point mutation, in the specific organ called the bursa of Fabricius. Somatic gene conversion utilizes pseudogenes located upstream of the functional gene and copies small piece of the pseudogenes to the rearranged gene segment unidirectionally and intrachromosomally. There are 25 or 26 pseudo-$V_{\lambda}$ genes (Reynaud et al. 1987; McCormack et al. 1993) and approximately 80 pseudo-$V_{H}$ genes (Reynaud et al. 1989). Some pseudo-$V_{H}$ gene segments also contain D- or J-like sequences that contribute to the diversity in CDR3. Among pseudogenes extensive nucleotide differences as well as length variation are observed in CDRs. Repeated somatic gene conversions at multiple positions by different pseudogenes therefore generate enormous diversity, which can be enhanced later by somatic point mutations (Parvari et al. 1990; Arakawa et al. 1998).

4.2 Evolution of Chicken V Gene Segments

All chicken $V_{H}$ and $V_{\lambda}$ gene segments identified belong to a single V gene family and show little variation in FRs, specifically among $V_{H}$ gene segments (Reynaud et al. 1987, 1989). Nucleotide differences in CDRs, however, extensive and are capable of generating high-affinity Igs (Michael et al. 1998). High homology in FRs indicates that the pseudo-$V_{H}$ gene segments have been derived from recent gene duplication or are subject to gene conversion. High homology of FRs would increase the efficiency of somatic gene conversion, since V gene segments with high similarity to the functional gene segment undergo somatic gene conversion more frequently. Analyses of pseudo-V gene segments have provided evidence of germ-line gene conversion (McCormack et al. 1993; Benatar and Ratcliffe 1993). Frequent germ-line gene conversion without selection generally results in homogenization of member genes, but it can create diversity in the presence of positive selection. Indeed, positive selection has been shown to operate in CDRs of pseudo-$V_{H}$ gene segments (Ota and Nei 1995).

Almost all vertebrates except cartilaginous fishes have human-type gene organization (Vn-Dn-Jn-Cn), or so-called “translocon” gene organization (Litman et al. 1993; Marchalonis et al. 1998). Therefore the chicken gene organization has likely been derived from the translocon gene organization. It remains to be examined, however, what caused the loss of functional genes in chickens (Ota and Nei 1995). It is possible that the utilization of somatic mutation has reduced the selection pressure to maintain functional genes and all V gene segments except one became nonfunctional owing to accumulation of deleterious mutations in the promoter region or coding region. Alternatively, population size bottlenecks or an extended period of population size reduction would have caused accumulation of deleterious mutations by genetic drift, as suggested by Ota and Nei (1995).

One unique feature of chicken pseudo-$V_{H}$ gene segments is the presence of germline joined $V_{H}$D and $V_{H}$DJ$_{H}$ gene segments. How are D- or DJ$_{H}$-like
sequences brought to V gene segments, while keeping functional D or V gene segments intact? This might be attributable to another unique feature of chicken Ig gene organization, i.e., the mutually inverted orientation of pseudo-V gene segments (Reynaud et al. 1987, 1989). Since the gene rearrangement of a D or DJH gene segment with a V gene segment in opposite transcriptional polarity would result in only inversion of the fragment but not deletion between them, incidences of gene rearrangement in germline cells by leaky expression of gene rearrangement complex can join D- or DJH-like gene segments to V gene segments while maintaining D gene segment diversity and D-proximal V gene segments intact. Alternatively, all features observed in chicken pseudogenes are also consistent with the hypothesis that the reverse-transcribed products of rearranged genes were integrated into genome (Ota and Nets 1995; see also Blanden and Steel (1998) for the possible role of reverse transcription on somatic mutations in higher vertebrates). Further studies are necessary for testing the evolutionary hypotheses on the origin of chicken Ig genes.

5 Evolution of V Gene Families

As mentioned above, variable number of V gene families have been detected among vertebrates (see also Table 3). How did this large number of gene families originate? In order to understand the evolution of V gene families their phylogenetic rela-

<table>
<thead>
<tr>
<th>Species</th>
<th>V&lt;sub&gt;H&lt;/sub&gt; gene family</th>
<th>V&lt;sub&gt;L&lt;/sub&gt; gene family</th>
<th>V&lt;sub&gt;nuvs&lt;/sub&gt; gene family</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>7</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Chicken</td>
<td>1</td>
<td>–</td>
<td>1</td>
</tr>
<tr>
<td>Xenopus laevis</td>
<td>11</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Rainbow trout</td>
<td>11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Horned shark</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Fig. 3. Neighbor-joining tree of 92 V<sub>H</sub> gene segments and 2 outgroup V<sub>L</sub> gene segments. The first three letters of genes were designated the first letter of species genus name plus the first two letters of species name: bull shark (Cio), cattle (Bau), channel catfish (Ipnot), chicken (Galo), horned shark (Hpr), human (Hpu), little skate (Rer), mexican axolotl (Ane), mouse (Mmu), nurse shark (Grci), pig (Scp), rabbit (Ore), rainbow trout (Omy), sandbar shark (Cpf), sheep (Oar), short-tailed opossum (Mdo), spotted ratfish (Hco), sturgeon (Ahu), and Xenopus laevis (Xlu). Please refer to Ota (1997) and Roman et al. (1996) as well as the legend of Fig. 1 for the source of most sequences. Additional sequences were retrieved from Genbank (accession numbers): Aba18.1 (Y13261), Aba205 (AJ222823), Aba33 (Y13255), Ame1.182M (AF027252), Ame2.259M (AF027253), Ame3.73M (AF027256), Ame4.43M (AF027257), Ame5.43M (AF027259), Ame6.65Y (AF027260), Ame7.15M (AF027261), Ame8.46M (AF027262), Ame9.103M (AF027267), Ame10.33Y (AF027268), Ame11.5M (AF027269), Ame11.182M (AF027252), Bua8 (U5164), IpnotVH7a (U3410), MdoVH2 (AF012124), Mdo257 (AF007070), Mmu15A (U39293), OarVIB (Z40188), and ScpVHA (AF064688).
tionships are investigated in this section. The analyses provide further insight on the evolution of \( V \) gene segments, since multigene families are subject to different modes of evolution depending on underlying evolutionary forces (NEI et al. 1997).

### 5.1 Evolution of \( V_H \) Gene Families

A phylogenetic tree for 92 \( V_H \) gene families from various organisms is shown in Fig. 3. The \( V_H \) gene tree shows that vertebrate \( V_H \) gene families are classified into eight groups, A–H, and a few unassigned \( V_H \) gene families. Three additional groups were added to five previous groups A–E (OTA and NEI 1994) due to the recent accumulation of \( V_H \) gene segment from lower vertebrates. Two *Xenopus* (XlaVH8; XlaVH10), two Mexican axolotl (Amc9.103M, Amc10.33Y), and one rainbow trout (OmyVIII) are tentatively not assigned to any group due to inconsistent results among analyses.

As shown by earlier studies (OTA and NEI 1994; SCHROEDER et al. 1990; TUTTER and RIBLET 1989), the mammalian \( V_H \) gene families are classified into three groups A–C or three clans I–III. Both humans and mice have \( V_H \) gene families belonging to all three groups, indicating that mammalian ancestors possessed \( V_H \) gene segments of the three groups. Since *Xenopus* \( V_H \) gene families cluster with mammalian \( V_H \) gene families within the three groups and even teleost fish (IpuVH1, OmyVH1, OmyVH4), coelacanth (LcuVH) and sturgeon (Ab3.3) \( V_H \) gene families are found in group C, the divergence of these groups must have occurred before the emergence of amphibians. Despite of the early divergence of these three groups, the avian and mammalian species that use gene conversion and/or hypermutation as the primary source of diversification, such as chicken, rabbit, sheep, cattle, and pig, have only a single or two \( V_H \) gene families belong to group B and/or group C. Meanwhile, most extensive diversified \( V_H \) gene families are observed for amphibian species, where many \( V_H \) gene segments belong to the four groups A–C and H and a few unassigned gene families. Most teleost fish \( V_H \) gene families are classified into groups C and D. As mentioned above, cartilaginous fishes \( V_H \) gene families are classified into the three groups E–G.

---

**Fig. 4.** Neighbor-joining tree of 54 \( V_L \) gene segments. See Fig. 3 for designation of genes: cattle (Bia), channel catfish (Ipu), chicken (Gga), horned shark (His), horse (Eca), human (Hsa), little skate (Rea), mouse (Mus), muscovy duck (Cmo), nurse shark (Glu), rabbit (Ore), rainbow trout (Omy), sandbar shark (Cpi), sheep (Osa), spotted ratfish (Heri), sturgeon (Ahu), and *Xenopus laevis* (Xla). Please refer to HAIK et al. (1996) and STRIKOV and SC (1998) for the source of most sequences. Additional sequences were retrieved from Genbank (accession number: Bia106.1(X090552), Bia2 (U322585), Esc (X75611), IpuL5(F125705), and OmyL51(U69987). Cartilaginous fish type 1 \( V_L \) gene segments were used as outgroup instead of \( V_H \) gene segments to increase the number of amino acids used in the analysis. Note that the phylogenetic analysis including \( V_H \) gene segments (data not shown) and previous phylogenetic analyses (HAIKE et al. 1996; RAST et al. 1994) have shown that the cartilaginous fish type 1 \( V_L \) gene segments were the first groups diverged from the rest of \( V_L \) gene segments.
5.2 Evolution of $V_L$ Gene Families

A phylogenetic tree for $V_L$ gene families from various organisms is shown in Fig. 4. Here vertebrate $V_L$ gene families are classified into nine groups and one unassigned $V_L$ gene family. One monophyletic group consists of $V_k$ gene families from cartilaginous fishes to mammals. On the other hand, $V_A$ gene families are polyphyletic with respect to $V_k$ gene families and separated into eight different groups. These $V_L$ groups diverged at an early stage of vertebrate evolution (more than 470MY ago), since both groups $B$ and $k$ have $V_L$ gene families from cartilaginous fish to mammals.

Among species studied here, human $V_L$ gene families are by far the most extensively diverged and are classified into six groups, A–E and $k$. One $V_L$ gene family Hsp42 is not included because it may have originated by recombination (Strynikova and Su 1998). Only one $V_A$ gene family known for chicken or rabbit belongs to group B or D, respectively. Rabbits are also known to use IgG and so far a single $V_A$ gene family has been detected. All sheep and cattle $V_L$ gene families belong to group A. As in the case of $V_H$ gene family evolution, avian and mammalian species that use gene conversion and/or hypermutation as the primary source of diversification show limited diversity of $V_L$ gene families (see below also). *Xenopus* $V_L$ gene families are classified into the four groups C, F, G, and $k$. Teleost fish $V_L$ gene families are mostly of the $V_b$ group with the exception of one $V_L$ gene family reported in rainbow trout (Partula et al. 1996). Cartilaginous fish $V_L$ gene families are classified into the three groups, B, H and $k$, and the unassigned $V_L$ gene segment Hco712.

5.3 Coevolution of $V_H$ and $V_L$ Gene Segments

From the phylogenetic analyses conducted above the following evolutionary patterns are commonly observed in the evolution of $V_H$ and $V_L$ gene segments. (a) One dominant $V$ gene family or closely related $V$ gene families are generally associated with each C isotype in cartilaginous fishes. (b) More diversified $V$ gene families have been maintained under "translocon" type gene organization than "cluster" type gene organization for a given species, except for those that utilize gene conversion/hypermutation to generate the primary diversity (see also Table 3). (c) Amphibians, both *Xenopus* and Mexican axolotl, show one of the most extensive $V$ gene family diversity. (d) Avian and mammalian species that use somatic gene conversion/hypermutations to generate the primary diversity have a reduced number of $V$ gene families. (e) The major lineages of $V$ gene families have been separated for more than 350MY for $V_H$ gene families and more than 470MY for $V_L$ gene families.

The above features of the coevolution of $V_H$ and $V_L$ gene segments are striking and likely attributable to some common evolutionary factors acting on the $V$ gene family but not to random genetic events (Strynikova and Su 1998). The next section reviews the change in diversity generation that occurred during the vertebrate evolution and examine the underlying evolutionary factors affecting the evolution of $V$ gene family.
6 Evolution of Immune System and Ig Diversity

In terms of the generation of Ig diversity, three main changes have occurred during the vertebrate evolution: (a) the establishment of translocon gene organization, (b) the establishment of an organismal system for efficient clonal selection, and (c) the use of somatic gene conversion/hypermutation to generate the primary diversity.

The establishment of translocon gene organization likely occurred before the emergence of teleost fishes at IgH (Litman et al. 1993; Marchalonis et al. 1993). The establishment of translocon gene organization was very important since it would enable the combinatorial diversity and increase the diversity in CDR3 tremendously. Furthermore, it also has facilitated the accumulation of V gene family, since the V gene segments have evolved independently from the evolution of C gene segments and diversified V gene segments are more productively utilized to create Ig diversity by combinatorial diversity. It is still questionable if it occurred at IgL loci around the same time, since teleost fish Igk locus has cluster type gene organization (V_{1,2}-J-C)_n, similar to cartilaginous fishes (Marchalonis et al. 1998 and references therein). The recent report by Lundqvist et al. (1998), however, suggests that the sturgeon has the translocon type gene organization at the IgL locus and raises a possibility that the teleost fish cluster type gene organization might have derived secondarily.

The lower vertebrates are known for relatively “poor” humoral immune response (Du Pasquier 1980) despite the diversified V gene families seen among them. A high rate of somatic mutation is present in Xenopus and the poor immune response has been attributed to the lack of system to select B-cells that generate higher affinity Igs (Wilson et al. 1992). Germinal centers, where the clonal selection of B-cells occurs during secondary response in avian and mammalian species, are not found among lower vertebrates and the invention of such apparatus would have had significant impact on the generation of secondary Ig diversity. Specifically, it has solved one problem which “anticipatory” immune system encounters. In most vertebrates each B-cell expresses only one kind of Ig (allelic exclusion and isotypic exclusion). Consequently, the antigen recognition capacity is dictated not only by the ability to generate diversity but also by the number of cells to hold the diversity. By integrating efficient evolutionary systems, i.e., mutation and selection, into its somatic processes, B-cells can expand Ig repertoire temporarily, select B-cells producing high-affinity Igs and eliminate unnecessary B-cells under the limited resources in its secondary response.

Lastly, among many avian and mammalian species the mechanism to generate the primary repertoire by hypermutation and/or somatic gene conversion has been found (Butler 1997). In chickens, rabbits, cattle, and pigs somatic gene conversion and hypermutation play major roles in the primary diversification (Knight 1992; Lucier et al. 1998; McCormack 1993; Parng et al. 1996; Reynaud et al. 1994; Sun et al. 1996). In sheep hypermutation is the primary source of diversity (Reynaud et al. 1991). In these organisms a few closely related V gene families are expressed (Knight 1992; Lopez et al. 1998; Reynaud et al. 1987, 1989). This is
partly due to the fact that the gene rearrangement occurs only at the early stage of organismal development in some species (McCormack et al. 1993; Knight and Winstead 1997). However, genetic data provided so far have indicated the absence of a diverse array of V gene family in their genome. Since the various lineages of V gene families were present at the emergence of tetrapods (see Figs. 3, 4), a reduction in germline V gene repertoire must have occurred during their evolution. The loss of germline diversity have apparently occurred independently in several lineages of evolution (Nei et al. 1997; Sitnikova and Su 1998). For example, swine and sheep are phylogenetically closely related to each other but swine have \( V_H \) gene families belonging to group C (Sun et al. 1994), whereas sheep have \( V_H \) gene families belonging to group B (Dufour et al. 1996). This can be explained only if ancestors of these organisms had \( V_H \) gene families belonging to groups B and C and sheep and swine lost or are losing one of them (Sitnikova and Su 1998). Consistent with this hypothesis, cattle, another artiodactyl species, have two gene families similar to mouse Q52 (group B) and X81 (group C) but express only the \( V_H \) gene family belonging to group B (Lopez et al. 1998). In these artiodactyl species, the utilization of somatic gene conversion or hypermutation for the generation of primary diversity has likely reduced the selection pressure to maintain diversified V gene families in germline. In this respect, it is interesting to note that only two \( V_H \) gene families belonging to group C are expressed and no other \( V_H \) gene family seems to be present in short-tailed opossum (Miller et al. 1998; see also Fig. 3). Since the marsupials are the most closely related animals to placental mammals, further characterization of their immune system will shed light on the evolution of somatic gene conversion and hypermutation system in placental mammals.

7 Summary

Evolution of Ig V gene segments are generally characterized by (a) evolution by “the birth and death process” and (b) diversifying selection. However, the detailed evolutionary pattern of V gene segments varies among species due to the fact that the humoral immune system itself has changed during vertebrate evolution. The change in somatic diversification system coupled with the change in lymphocyte development has imposed a significant impact on the evolution of Ig genes. In order to understand the evolution of immunological genes it is important to view it in the context of the evolution of the entire immune system itself.

Acknowledgements. We thank Chris T. Amemiya, David Garrity, and Su Chen for critical reading of the manuscript.
References


Kirkham PM, Mortani F, Newton JA, Schroder HW Jr (1992) Immunoglobulin \( V_{H} \) clan and family identity predicts variable domain structure and may influences antigen binding. EMBO J 11:603–609


Matsunaga T, Andersson E (1997) Analysis of VH gene diversity in rainbow trout (Oncorhynchus mykiss): both nonsynonymous and synonymous nucleotide changes are more frequent in CDRs than in FRs. Immunogenetics 45:201–208


