Somatic Variation Precedes Extensive Diversification of Germline Sequences and Combinatorial Joining in the Evolution of Immunoglobulin Heavy Chain Diversity

By Kristin R. Hinds-Frey, Hikaru Nishikata, Ronda T. Litman, and Gary W. Litman

From the Department of Pediatrics, University of South Florida, All Children's Hospital, St. Petersburg, Florida 33701

Summary

In Heterodontus, a phylogenetically primitive shark species, the variable (V.), diversity (D.), joining (J.) segments, and constant (CH) exons are organized in individual ~18-20-kb “clusters.” A single large V. family with >90% nucleic acid homology and a monotypic second gene family are identified by extensive screening of a genomic DNA library. Little variation in the nucleotide sequences of D. segments from different germline gene clusters is evident, suggesting that the early role for D. was in promoting junctional diversity rather than contributing unique coding specificities. A gene-specific oligodeoxynucleotide screening method was used to relate specific transcription products (cDNAs) to individual gene clusters and showed that gene rearrangements are intra- rather than intercluster. This provides further evidence for restricted diversity in the immunoglobulin heavy chain of Heterodontus, from which it is inferred that combinatorial diversity is a more recently acquired means for generating diversity. The observed differences between cDNA sequences selected and the sequences of segmental elements derived from conventional genomic libraries as well as from V. segment-specific libraries generated by direct PCR amplification of genomic DNA indicate that the V. repertoire is diversified by both junctional diversity and somatic mutation. Taken together, these findings suggest a heretofore unrecognized contribution of somatic variation that preceded both extensive diversification of the germline repertoire and the combinatorial joining process in the evolution of humoral immunity.

The Ig heavy chain gene family is one of the most complex genetic systems that has been described. The organization of heavy chain genes is well established for the mouse (1-3) and human (1, 4, 5). In both cases, as well as in other mammals, variable (Vn), diversity (Dn), and joining (Jn) segments consisting of several hundred tandemly arrayed Vn and fewer numbers of Dn and Jn elements occupy a chromosomal region of 2-3 x 10⁶ base pairs. Functional Ig Vn genes arise in B lymphocytes by rearrangement of individual segmental elements, so called V-D-J-joining, a process that is mediated by recombination signal sequences (1) and is catalyzed by recombinase enzymes (1, 6, 7). The use of different combinations of Vn, Dn, and Jn segments is a significant factor in the generation of antibody diversity, as are variations in junctional joining, somatic mutation (1, 2), and, in at least one case, gene conversion (8). The mammalian type of Vn organization is not preserved in all vertebrates. Antibody diversity in Gallus, an avian, is generated from the rearrangement of single functional Vn (9) and Vn (10) genes by a process of gene conversion that uses flanking pseudogenes. A similar mechanism is used in another avian species (11).

In elasmobranchs, such as shark and skate, the Vn, Dn, Jn, and Cn (12, 13), as well as Vn, Jn, and Cn (14), elements of Ig genes are organized in a unique manner. In Heterodontus francisci (horned shark), a phylogenetically primitive species, these elements are linked in >100 “clusters,” ~18-20 kb, each containing a single Vn, two Dn, and single Jn and Cn segments (V-Dn-Dn-Jn-Cn) (13). The four recombining elements within the cluster are separated by 300-400 nucleotides, the intervening sequence (IVS) separating Jn and Cn is 8-9 kb (12) (our unpublished observation), and the Cn exons, including the transmembrane segment, encompass ~7 kb (15). The nucleotide and predicted amino acid sequences of Cn

1 Abbreviations used in this paper: FR, framework region; IVS, intervening sequence.
segments of different clusters can vary as much as the V\textsubscript{\alpha} region (15, 16). Approximately half of the clusters exhibit evidence of germline “joining” of V\textsubscript{\alpha} and D\textsubscript{\alpha} (V-D) or V\textsubscript{\alpha}, D\textsubscript{\alpha}, and J\textsubscript{\alpha} (VDJ) elements (13) in nonlymphoid cells. This type of organization is unique to eelasmbranchs and extends to a second, non-\mu-type C\textsubscript{\alpha} isotype (17). The present study focuses on the functional significance of the V-D\textsubscript{1}-D\textsubscript{2}-J-C form of gene organization. Using gene-specific oligodeoxynucleotide hybridization, the sequences of cDNA clones are related to their corresponding (parental) germline clusters. The goals of these studies were to determine: if there is combinatorial diversity, the nature of junctional diversity and whether a somatic mutation mechanism is present in this modern representative of an early point in vertebrate phylogeny.

Materials and Methods

Animals. Adult specimens of H. francisci (horned shark) were obtained from Pacific Biomarine Supply Co. (Venice, CA). After the animals were killed, all tissues were processed immediately for DNA and RNA isolation. The same animal was used to obtain material for the genomic and the cDNA \lambda libraries.

DNA Libraries. The genomic library was described previously (13). The cDNA library was constructed from spleen mRNA using a kit (Pharmacia LKB Biotechnology, Inc., Piscataway, NJ) and cloned into \lambda gt11.

Subcloning in M13 mp10/11 and DNA Sequencing. DNA sequencing was performed by the dideoxy chain termination method (18) as described previously (13).

DNA Sequences in GenBank. The GenBank accession numbers for the sequences described herein are: Z11776, 1113; Z11777, 1403; Z11778, 33141; Z11779, 33141CON; Z11780, 12215; Z11781, 12271; Z11782, 12272; Z11783, 12214; Z11784, 12241; Z11785, 12221; Z11786, 12273; Z11787, 12251; Z11788, HN13; Z11790, HN84; Z11791, HN21; Z11792, HN72.

PCR Analysis of Genomic V\textsubscript{\alpha} Clones. All V\textsubscript{\alpha} plaques were selected from the plating of a genomic library (~7 \times 10\textsuperscript{5} recombinants equals approximately two to three genome equivalents), pooled, and eluted into 1x SM. The titer of the resulting pool was 5 \times 10\textsuperscript{4} PFU/ml. 1 \mu l of this material was used as template for PCR amplification. The phage were heated at 70°C for 5 min to disrupt the phage heads, then amplified using 1403-specific primers for CDR2: CCGAGCTCACCATAGTTCATCATAC-CCCTGCAGTCACCGTCACCAT with a natural PstI site (underlined), and J\textsubscript{\alpha}: CCGAGCTCACCATAGTTCATCATAC-CCCTGCAGTCACCGTCACCAT, containing a 2-nt spacer plus TGACTCAGCCAGAAGCA-3', and contain a 2-nt spacer plus SstI conserved area of framework region 1 (FR1): 5'-CCGAGCTCT-

Results

V\textsubscript{\alpha} Families. Using a homologous (Heterodontus \alpha) screening strategy employing a single probe, HXIA (19), which is equivalent to the V\textsubscript{\alpha} sequence of \lambda1403 (13), several hundred V\textsubscript{\alpha} clones were detected in a representative Heterodontus genomic \lambda library prepared from the DNA of a single animal. The complete nucleotide sequence of the V\textsubscript{\alpha} through the J\textsubscript{\alpha} elements has been determined for 24 clones. With a single exception, the V\textsubscript{\alpha} segments of all of the 24 clones are >90% identical at the nucleotide level with HXIA (\lambda1403) and thus are classified as members (defined as sequences sharing >70% nucleotide identity) of the same Ig gene family, V\textsubscript{\alpha}. The exception, \lambda1113 (13) (and below), hybridizes weakly with the selection probe HXIA and is only ~61% identical at the nucleotide sequence level to the most closely related member of the Heterodontus V\textsubscript{\alpha} gene family, although it possesses extensive sequence identity in FR3. No other weakly hybridizing clones were noted. The nucleotide sequence of \lambda1113 from 3' of V\textsubscript{\alpha} through J\textsubscript{\alpha} has been reported previously (13) and consists of a recombination signal sequence arrangement in which either the D\textsubscript{\alpha}1 and D\textsubscript{\alpha}2 segments or the D\textsubscript{\alpha}1 segment alone can be joined productively, in contrast to V\textsubscript{\alpha} genes in which D\textsubscript{\alpha}1 and D\textsubscript{\alpha}2 independently can be joined, according to the 12/23 rule of recombination (1). Despite repeated efforts to identify related genes using different genomic (and cDNA) libraries, it appears that the \lambda1113 cluster represents a monotypic V\textsubscript{\alpha} family (V\textsubscript{\alpha}II). Attempts to detect additional V\textsubscript{\alpha} families in several different genomic and cDNA libraries, using both J\textsubscript{\alpha} and C\textsubscript{\alpha} selection, have been unsuccessful, although it could be argued that unique V\textsubscript{\alpha} families might be associated with appreciably different J\textsubscript{\alpha} or C\textsubscript{\alpha} segments.

Germline Diversity of D\textsubscript{\alpha} and J\textsubscript{\alpha} Segments: C\textsubscript{\alpha} Variation. The D\textsubscript{\alpha} coding regions of the V-D\textsubscript{1}-D\textsubscript{2}-J\textsubscript{\alpha} genes, with

Published September 1, 1993
the exception of λ1113, are related very closely at the nucleotide level (Fig. 1). λ1403, the gene cluster that has been examined in greatest detail, differs at one nucleotide in D\(\mu\)1. Two of the λ1403-type genes, λ1315 and λ32004, share one nucleotide difference from λ1403 in D\(\delta\)2. These restrictions in sequence variation in the λ1403-like D\(\delta\)s not only limit germline diversity but also complicate unequivocal identification of the "parental" origin of cDNAs (see below). Of the 24 J\(\mu\) segments that have been sequenced, essentially all variation is in the 5' 11 nucleotides (13; see below). This portion typically is eliminated in the process of productive gene rearrangement. As reported previously, Ig gene clusters in *Heterodontus* are associated with unique C regions (16) that potentially can facilitate identification of the "parental" origin of cDNAs.

**Identification of λ1403-related V\(\mu\)-D\(\mu\)-J\(\mu\) Clusters.** Sequence comparisons indicate that regardless of the high degree of nucleotide (and amino acid sequence) sharing, CDR2 remains the most informative region in terms of distinguishing different V\(\mu\) gene clusters, with the greatest amount of nucleotide variation being centered at nucleotides 166-174 (see Fig. 2 a, below). Under stringent conditions, 33mer probes complementing this substituted region exhibit relatively little nonspecific hybridization and allow discrimination between known phage clones that differ in the sequences of CDR2s. Genomic clone λ1403, which contains V\(\mu\), D\(\mu\)1, D\(\mu\)2, J\(\mu\) (V-D\(\mu\)-D\(\gamma\)-J\(\mu\) configuration), as well as the first two C\(\mu\) exons, was selected for the initial analyses of somatic variation at the nucleotide level (Fig. 1). M403, the gene cluster that has been examined in greatest detail, differs at one nucleotide in D\(\delta\)1.

An alternative strategy for identifying λ1403-like genes also was used. PCR products were generated, using a λ1403/λ33141 CDR2-specific and a J\(\mu\) primer, from a pool (~250 clones) of V\(\mu\)\(\lambda\) clones that were identified by cross-hybridization with the HXIA probe under conditions of moderate stringency (19). 12 of these clones were sequenced and four were found to be identical to λ33141, three identical to λ1403, four were identical to λ33141 in the coding region, but had either 1-2 bp differences in the IVSs, and one had a 1-bp deletion in the IVS. The IVS differences most likely represent PCR-induced substitutions, or cloning artifacts, as similar IVS changes have been noted when individual clones are amplified by PCR (our unpublished observations).

To investigate the possibility that other λ1403/λ33141-type genes had not been integrated in the library, genomic DNA was directly amplified using a PCR procedure that used a 5' FR1 primer complementing a highly conserved region in all sequenced shark V\(\mu\) genes and a 3' CDR2 primer. The 3' end of this latter, CDR2 primer, was complementary to the region unique to these two genes (5'-ATACCAGTACTA-183 CGCCCCCT-3'; see Fig. 2, a and b), which should prevent the amplification of other V\(\mu\) segments. After two rounds of amplification, the products were cloned in M13. Over 70 clones were sequenced and found to correspond only to the λ1403 and λ33141 germline genes. Thus, through three separate forms of analysis, it was not possible to detect other genes of the λ1403/λ33141 types that could have given rise to closely related but distinct cDNAs.

**Expression of the λ1403 and λ33141-type Genes.** The spleen is the primary immune organ in elasmobranchs. Approximately 1.6 × 10\(^4\) spleen cDNA recombinant plaques were screened with the λ1403-33mer CDR2 probe under the same hybridization wash conditions that were used in screening the genomic library, and 15 strongly hybridizing plaques were detected. 12 of these were sequenced from the 5' end of V\(\mu\) FR1 through C\(\mu\)1, and eight unique sequences were obtained. In addition a closely related clone, HH13, was recovered from the same *Heterodontus* cDNA library in the course of an unrelated experiment. The V\(\mu\) segments of the nine sequences are highly homologous to λ1403 (Fig. 2 a) and λ33141 (Fig. 2 b). Seven cDNA clones have two FR and two CDR nucleotides that are unique to λ1403 and two cDNA clones are λ33141 like.

The V\(\mu\) segments of cDNA clone 12215 and genomic clone λ1403 are identical (Fig. 2 a). The other clones typically exhibit one or two differences in the sequence segment complemented by the CDR2-specific 33mer probe and up to 11 nucleotide differences from λ1403 throughout the V\(\mu\) region. In the λ1403-related cDNAs, CDR1 and CDR2, which collectively constitute <20% of the V\(\mu\) region, exhibit 44% of the sequence differences. Overall, there are almost twice as many replacement as there are neutral substitutions. Only replacement substitutions occur in FR3; however, it should be noted that many of these occur near the junctional boundary and may represent changes that are

<table>
<thead>
<tr>
<th>D(\mu)1</th>
<th>D(\mu)2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1403</td>
<td>GGT ACA GCG GTG GGT</td>
</tr>
<tr>
<td>33141</td>
<td>GGT ACA GCA GTG GGT</td>
</tr>
<tr>
<td>2807</td>
<td>GGT ACA GCA GTG GGT</td>
</tr>
<tr>
<td>32004</td>
<td>GGT ACA GCA GTG GGT</td>
</tr>
<tr>
<td>1113</td>
<td>GGT ACT ACA GTG GGT</td>
</tr>
</tbody>
</table>

Figure 1. Nucleotide sequence differences in D\(\mu\) segments of V-D-J type genes. In V-D-J type genes, D\(\mu\)1 and D\(\mu\)2 sequences are shown in arbitrary triplets that are not intended to correspond to codons and are arranged to emphasize sequence identity with λ1403. Differences (including the two additional nucleotides in λ1113 D\(\mu\)1) are in lowercase bold, as is the g at position 10 in several D\(\mu\)2 sequences. λ1113 exhibits the highest degree of divergence from the λ1403 prototype.
Figure 2. (a) Nucleotide alignment of the V_{\alpha} segment of k1403 with the sequences of seven different cDNAs that represent the putative rearrangement products of this gene cluster. Nucleotide identities are indicated (−); uppercase designates neutral substitutions, and uppercase bold indicates replacement substitutions. For two nucleotide changes within a single codon, both nucleotides are indicated in bold even if one of the changes would not result in a replacement substitution. The 33mer CDR2-specific probe used to select the cDNAs is underlined. (b) Nucleotide alignment of k33141 with the sequences of two cDNAs that represent the putative rearrangement products of this germline gene. The two CDR_{\beta} nucleotides that distinguish k33141 from k1403 are underlined; all other designations are as in a.
introduced during functional rearrangements. In three instances, there are two contiguous nucleotide changes and, in one case, there are three nucleotide changes, when compared with the prototype sequence. Nucleotide 125, which is the last base in a SmaI restriction site, is highly polymorphic in the germline of other non-\(\lambda 1403/\lambda 33141\)-type clones (K. Hinds-Frey, unpublished observations) as well as in the cDNA sequences, and may represent a mutational hot spot although it does not impart a coding change.

Furthermore, \(\lambda 33141\) is distinguished from \(\lambda 1403\) by two nucleotide substitutions in CDR2 and two FR substitutions. The CDR2 regions of cDNA clone 12251 and genomic clone \(\lambda 33141\) are identical (Fig. 2 b), while HN13 shares the distinct nucleotides but differs at a single base from the CDR2 of \(\lambda 33141\). The cDNA clones exhibit several differences at the 3' of \(V\_\lambda\), which could represent somatically introduced junctional variation (Fig. 2 b; see below). The number of replacement substitutions exceeds neutral substitutions. The relatively few \(\lambda 33141\)-type cDNAs that were recovered limit further interpretation of these data.

Inter- or Intracluster Rearrangement of D. The nucleotide sequences of the putative \(D\_1\) and \(D\_2\) regions of the \(\lambda 1403\) clones are shown in Fig. 3. As illustrated in Fig. 1, \(D\_1\) and \(D\_2\) regions of the V-D-1-D-2-J genes characterized thus far are nearly identical; the \(D\_1\) segment of \(\lambda 1403\) differs from the other \(D\_1\) sequences by a single base substitution. This base difference is present in all seven of the \(\lambda 1403\)-type genes but is not found in either of the \(\lambda 33141\)-type genes, suggesting that the rearrangement of \(V\_\lambda\) and \(D\_1\) occurred within the \(\lambda 1403\) cluster (and that an equivalent intracluster rearrangement took place in \(\lambda 33141\)). The \(D\_2\) segments of \(\lambda 1403\) and \(\lambda 33141\) are identical to each other, and to other genes, precluding distinction of cluster-restricted \(D\_2\) rearrangements.

The putative \(D\_1\) regions of all of the cDNAs possess a core of bases that are identical to the germline genes. The core segments of clones 12241 and 12271 exhibit single base differences from germline \(D\_1\) that can be accounted for by point mutations. Although unlikely, these cDNAs may have been transcribed from germline clusters that are refractory to genomic cloning, but in all other regards are identical to \(\lambda 1403\) (see discussion below). Although somatic reorganization of \(Heterodontus\) Ig gene segments of the \(\lambda 1403\) type (\(V\_\gamma_2\)-D-2-D-2-J-23J) can exclude \(D\_1\), according to the 12/23 rule, from 4 to 10 of the 15 germline-encoded \(D\_1\) nucleotides are detected in the various cDNAs. It is notable that cDNA 12215, which exhibits the least change from the parental cluster in the \(V\_\lambda\) segment, also has the most extensive contribution from germline \(D\_1\) and \(D\_2\) segments.

\(D\) Region Diversity Is Generated by Extensive Junctional Modification. The overall lengths of the \(D\_1\) regions in these cDNAs range from 14 to 36 nucleotides, compared with the 28 nucleotides encoded in the genomic \(D\_1\) segments, assuming that appropriate junctional boundaries have been assigned. As these cDNAs presumably represent the rearrangement products of the \(\lambda 1403\) and \(\lambda 33141\) germline clones, it can be estimated that from 1 to 10 nucleotide additions can occur at each of the various junctions. In this series, ~65% of the junctional additions are G/Cs and at least one nucleotide addition occurs in at least one of the joining junctions, \(V\_\lambda/D\_1\), \(D\_1/D\_2\), or \(D\_2/J\_\lambda\), in all of the cDNAs. In three cDNAs, nucleotide additions occur at all three junctions. Nucleotide deletion followed by N region additions (2) are associated with the rearrangements of \(V\_\lambda\) genes in \(Heterodontus\). It appears that germline \(D\_1\) segments are used in all three reading frames, although the extensive diversity generated by the deletion and addition of bases at the junctions during segmental rearrangement compromises unequivocal alignment of germline \(D\_1\) segments and the specific cDNAs (data not shown).

\(J\_\lambda\) and \(C\_\lambda\) Show Evidence for Intracluster Rearrangement. The complete (including 5') sequences of the \(J\_\lambda\) segments of \(\lambda 1403\) and \(\lambda 33141\) are identical to each other but are unique at several positions from the other \(J\_\lambda\) segments characterized in the course of these (Fig. 4) and earlier studies (13). The sequences of the 5' \(J\_\lambda\) segments of \(\lambda 1403\)- and \(\lambda 33141\)-selected cDNAs, with the exception of a single difference in cDNA 12214 and a three nucleotide substitution in 12272,

### Table 3

<table>
<thead>
<tr>
<th>CLONE</th>
<th>(V_\lambda/D_1)</th>
<th>(D_1)</th>
<th>(D_1/D_2)</th>
<th>(D_2)</th>
<th>(D_2/J_\lambda)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1403</td>
<td>CGTA</td>
<td>AACG</td>
<td>ATATACG</td>
<td>ATATACG</td>
<td>ATATACG</td>
</tr>
<tr>
<td>12215</td>
<td>CCGGGC</td>
<td>GTACAG</td>
<td>ATATACG</td>
<td>ATATACG</td>
<td>ATATACG</td>
</tr>
<tr>
<td>12271</td>
<td>AA</td>
<td>GTACAG</td>
<td>ATATACG</td>
<td>ATATACG</td>
<td>ATATACG</td>
</tr>
<tr>
<td>12214</td>
<td>TGGT</td>
<td>GGCA</td>
<td>ATATACG</td>
<td>ATATACG</td>
<td>ATATACG</td>
</tr>
<tr>
<td>12241</td>
<td>TGGG</td>
<td>GGCA</td>
<td>ATATACG</td>
<td>ATATACG</td>
<td>ATATACG</td>
</tr>
<tr>
<td>12221</td>
<td>TGGG</td>
<td>GGCA</td>
<td>ATATACG</td>
<td>ATATACG</td>
<td>ATATACG</td>
</tr>
<tr>
<td>12273</td>
<td>A</td>
<td>GTACAG</td>
<td>ATATACG</td>
<td>ATATACG</td>
<td>ATATACG</td>
</tr>
<tr>
<td>12251</td>
<td>CCGGGC</td>
<td>ACAGCA</td>
<td>ATATACG</td>
<td>ATATACG</td>
<td>ATATACG</td>
</tr>
<tr>
<td>HN13</td>
<td>T</td>
<td>GTACAG</td>
<td>ATATACG</td>
<td>ATATACG</td>
<td>ATATACG</td>
</tr>
<tr>
<td>33141</td>
<td>T</td>
<td>GTACAG</td>
<td>ATATACG</td>
<td>ATATACG</td>
<td>ATATACG</td>
</tr>
</tbody>
</table>

at \(V\_\lambda/D\_1\), \(D\_1/D\_2\), and \(D\_2/J\_\lambda\) junctions are indicated in upper case. The two differences between the germline \(D\_1\) segments and their corresponding cDNAs are in lower case. The most extensive \(D\_1\) and \(D\_2\) contiguous contribution is in cDNA 12215. The AT dinucleotide in the \(D\_2\) homology alignment of \(\lambda 1403\) has alternative placements or may represent a nucleotide addition that arose in junctional joining. In 12271, an alignment is suggested in which the \(D\_1\) contribution is maximized; however, it is noted that the sequence GCG at the 3' of the \(V\_\lambda/D\_1\) boundary is identical with the GCG \(D\_1\) core that has been preserved by all other cDNAs.

819  Hinds-Frey et al.
are identical to the sequences of \( \lambda 1403 \) and \( \lambda 33141 \) genes. This finding is consistent with intracluster joining of \( J_n \) segmental elements, as has been suggested above for \( V_n \) and \( D_n1 \) joining. The single difference in clone 12214 and the differences at the 3' of several \( J_n \)s of the \( \lambda 1403 \) and \( \lambda 33141 \) types most likely represent somatic substitutions.

The \( C_n1 \) exons of \( Heterodontus \) Ig genes differ significantly from each other at the 5' (16). Lesser degrees of variation occur in the other \( C_n \) exons (15). The sequences of \( C_n1 \) of \( \lambda 1403 \) and \( \lambda 33141 \) differ at 10 positions from the \( C_n1 \) consensus sequence of the ~25 germline \( C_n \) genes that have been characterized to date (15, 16; and our unpublished observations) but are identical to each other. The \( C_n1 \) sequences of all nine cDNAs are identical to the corresponding germline \( \lambda 1403/33141 \) sequence (not shown), indicating that the \( J_n/C_n1 \) splice is of an intra- vs. intercluster nature. In addition to establishing that rearrangements are intra- vs. intercluster, the complete sequence identity between germline and cDNA \( C_n1 \) sequences indicates that variation primarily is restricted to the \( V_D \) and \( J \) regions of the gene, a finding entirely consistent with the \( V \) region-restricted nature of somatic mutation.

\( \lambda 1113 \)-restricted Clusters and Expression of the \( \lambda 1113 \)-type Gene. As indicated above, \( \lambda 1113 \) is a unique (presumably monotypic) gene family in which an inverted \( D_n1-D_n2 \) IVS changes the orientation of the functional recombination signal sequence spacers, i.e., V23-12D12-23D212-23J (\( \lambda 1113 \)-type) vs. V23-12D212-23D212-23J (\( \lambda 1403 \)-type gene). In addition, \( \lambda 1113 \) differs from all other \( V-D_J \)-type gene families characterized to date by a three-nucleotide insertion in CDR2 (Fig.

**Figure 4.** Nucleotide sequence identities in the \( J_n \) region of cDNAs that derive from germline clusters \( \lambda 1403 \) and \( \lambda 33141 \). The nucleotides that appear to be specific for the \( \lambda 1113 \) and \( \lambda 33141 \) (\( \lambda 1403 \)-type) sequences are underlined. Sequence identities are shown (-). Neutral substitutions are indicated in upper case and replacement substitutions are in bold.

**Figure 5.** (a) Nucleotide alignment of \( \lambda 1113 \) with the sequences of three different cDNAs that represent the putative rearrangement products of this gene cluster. For the \( V_n \) and \( J_n \) segments, nucleotide identities are indicated (-); upper case designates neutral substitutions and upper case bold indicates replacement substitutions. The probe sequence used to select the cDNAs is underlined. (b) \( D_n1, D_n2, \) and corresponding cDNA nucleotides are in upper case bold, except for nucleotides that appear to be unique to the \( \lambda 1113 \) gene (see Fig. 1), which are shown in lower case bold, as is the \( G \) at position 10 of \( D_n2 \) that is found in half of the \( D_n2 \) segments compared in Fig. 1, including \( \lambda 1113 \); additional nucleotides at the \( V_n/D_n1, D_n1/D_n2, \) and \( D_n2/J_n \) boundaries are indicated in upper case bold. Potential homologies in these regions are indicated by inspection of vertical alignments.
1) and a six-nucleotide deletion in FR3 (Fig. 5). The Dα1 segment of λ1113 (Fig. 1) differs at two positions and also is two nucleotides longer than the Dα1 segments of the other V-D1-D2-J genes characterized thus far. The Dα2 differs by a single 5′ nucleotide from the V-D1-D2-J-type gene (Fig. 1); however, this potentially informative base difference is not present in the three cDNAs. The Js cannot be distinguished readily from other V-D1-D2-J-type genes. The Cα1 exon of λ1113 contains a three-nucleotide addition and 18 additional nucleotide changes that distinguish it from the Cα1 exons of all other V-D1-D2-J genes characterized to date.

A 33mer probe specific for the CDR2 of λ1113 was used to screen a genomic λ library as described above for λ1403. Only three positive clones were identified in this screening and these had identical or overlapping restriction maps as well as identical V-D1-D2-J sequences, including IVSs.

The 33mer probe then was used to select λ1113-like cDNAs. Only three positive clones (HN21, HN72, and HN84) were identified (Fig. 5). All three cDNAs possess the three additional λ1113-type nucleotides in CDR2 and the six nucleotide deletion in FR3 identified in λ1113, compelling evidence that these cDNAs derived from λ1113 or a closely related (indistinguishable) gene (Fig. 5 a). HN84 is identical to λ1113 in all coding regions; the other two cDNAs exhibit limited differences. With all three λ1113-type cDNAs, junctional diversity resembles that seen with the λ1403- and λ33141-type genes (Fig. 5 b). The contribution of Dα1 in HN72 is considerable, although, the nature of the TAC deletion (by homology to λ1113 Dα1) in the D region of HN72 is uncertain (similar effects have been noted previously [20]). As indicated in Fig. 1, the Dα1 region of λ1113 has distinct nucleotides (TαAβ), as well as two additional nucleotides (AαTβ) that are not seen in other Dα1 segments. The two unique nucleotides, TαAβ, are present in HN21. The absence of these in HN72 and HN84 is uninformative owing to extensive Dα1 deletion (HN21) and the TAC deletion (HN72) indicated above. HN72 contains two nucleotides that correspond to AαTβ of λ1113. All three cDNAs share the G0 identified in the Dα2 of λ1113 (Fig. 1), further supporting the conclusion that these elements originate in the λ1113 cluster. As in the case with the λ1403- and λ33141-type gene clusters, germline Dα sequences in λ1113 transcripts are used in all three reading frames (not shown).

The nucleotide sequence of the Jα region of HN84 is identical to the corresponding sequence of λ1113. Both HN21 and HN72 exhibit replacement substitutions at the 3′ of Jα, similar to findings for Jα regions in several λ1403- and λ33141-type genes (Fig. 5 b). HN21 differs from the putative parental prototype sequence at an additional 5′ position. Because the sequence of the Jα segment of λ1113 does not vary significantly from the Jα consensus sequence, it is not possible, in this case, to provide unequivocal evidence for intra- vs. intercluster Jα use. The sequences of the Cα1 exons of cDNA HN21, HN72, and HN84 are identical to the unique (see above) corresponding sequence of λ1113 (not shown). Intracluster rearrangement of elasmobranch Vα genes is entirely consistent with these observations.

Discussion

Heterodontus represents the most phylogenetically distant form in which information is available about antibody gene structure and organization. A remarkable degree of sequence homology exists between the coding segments and recombination signal sequence elements of Ig genes in Heterodontus and all vertebrate species characterized to date [13, 15, 19]. Although strong Vα homology and characteristic χ-type heavy chain organization and regulation resemble mammalian Ig, close linkage of segmental elements (cluster organization), presence of both Dα1 and Dα2 elements, and absence of an invariant Ig gene promoter are characteristic of TCRs. The dramatic changes in gene organization that have occurred during vertebrate phylogeny raise important questions about the functional regulation of Ig genes that possess alternative patterns of segmental arrangement as well as the somatic mechanisms that expand germline Ig gene repertoires.

In addition to the unique gene organization, the Ig heavy chain gene system of Heterodontus is distinguished from that of higher vertebrates by a high degree of identity in the individual members of the primary Vα family (Vα1) as well as by the extremely high degree of sequence identity in both the Dα and Jα segmental elements, as indicated above. At the nucleotide level, the individual members of the Vα1 family are >90% related, in marked contrast to the extensively diversified Vα families in mammals [3, 4, 21] and the amphibian Xenopus laevis [22, 23]. As there probably are ∼200 gene clusters and ∼50% of these are joined in the germline, the entire repertoire of Heterodontus heavy chain genes apparently arises from ∼100 different clusters containing closely related Vα genes. Presumably, this restricts the potential of the antibody repertoire. The only non-Vα1 gene cluster, λ1113, would contribute only minimally to Vα diversity. As indicated above, this gene most likely arose through inversion of the Dα1-Dα2 IVS (13), an effect that has been observed in vitro (24) and has been proposed to represent a significant component in the evolution of Ig genes (13, 24). The inversion may relieve this particular cluster of the gene correction effect that we suggest could account for the high degree of sequence identity in both coding and noncoding regions, and indeed, the IVS regions of this gene are related only distantly to other Vα genes (13).

To address the role of somatic mechanisms in the generation of Ig diversity, a strategy has been devised that permits the identification of the rearrangement products of specific gene clusters. In these experiments, CDR2-specific probes have been used both to select cDNAs related to a specific germline gene cluster and to determine how many other candidate (parental) gene clusters are present in the germline. Regarding the latter, identification of all the individual members of a multigene family is confounded by the requirement that the relevant gene sequences be clonable and that the specific replication of the clones containing these genes (sequences) is not compromised relative to the replication of clones containing other genes. Because of these uncertainties, which are not easily amenable to experimental resolution, it is difficult to establish that all members of a germline...
gene family have been detected. While approaches such as PCR amplification of *Heterodontus* genomic DNA (complete clusters) and direct sequencing of PCR-generated templates are attractive, it is not possible to efficiently derive V-D1-D2-J sequences due to the near identity of joined VDJ and V-D-J genes, which possesses shorter IVS sequences and are amplified to the exclusion of the V-D1-D2-J forms (K. Hinds-Frey, unpublished observations). Nevertheless: (a) the detection in this study of identical \(10^3\) genes in the libraries of two different individuals; (b) the isolation of the closely related \(\lambda 33141\) gene, which may represent an allelic form; (c) the inability to detect additional, related gene clusters by screening of conventional and PCR-generated (FR1-CDR2) libraries; (d) the inability to detect additional \(10^3/\lambda 33141\)-type genes by direct amplification of FR1-CDR2 sequences; and (e) the high degree of concordance between germline (V, D1, D2, J, and C\(_n\)) and cDNA sequences with \(10^3/\lambda 33141\), \(\lambda 33141\), the monotypic \(\lambda 1113\), and a fourth, unique cluster (\(\lambda 2807\); results not shown) are consistent with the identification of the rearrangement products of an individual gene cluster. Furthermore, the findings that the \(C_n\)1 sequences of the \(10^3/\lambda 33141\) and \(\lambda 1113\) cDNAs are identical with their respective germline counterparts would be inconsistent with the origin of these transcripts from other clusters. While hypotheses could be invoked that would explain all somatic changes as having arisen from as yet undetected gene clusters that also cannot be amplified by direct PCR, such claims are difficult to substantiate based on these observations. Finally, such hypotheses would be entirely inconsistent with the various estimates of the number of gene clusters in this species, some of which we based on direct, quantitative Southern blot analysis of C region exons (13, 16).

The unique sequence characteristics of \(D_n\)1's, the 5' \(J_n\) segments, as well as the \(C_n\)1 exons of genes \(10^3/\lambda 33141\) and \(\lambda 33141\), relative to the other genes that have been described, provide a means whereby intra- vs. intercluster gene rearrangements potentially can be distinguished. It is most likely that rearrangements occur within clusters, as is suggested from the analyses of \(V_n/D_n\1, V_n/J_n,\) and \(V_n/C_n\ association patterns. This restriction may relate to the remarkably close linkage of V-D1-D2-J (300–350 nucleotides separate the segmental elements), and is consistent with the presence of \(V_n\) clusters (of uncharacterized organization) on multiple chromosomes in *Heterodontus* (Amemiya et al., unpublished observations). Whereas spatial proximity is not an obligatory requirement of segmental rearrangements, certain components of the joining process are influenced by the positions of the recombining elements, e.g., close proximity may be the primary factor in the selective joining of \(D_n\) and \(J_n\) segments (2). The intersegmental distances of the *Heterodontus* elements are several orders of magnitude less than the minimal linkage distances (most proximal \(V_n, D_n,\) and \(J_n\) elements) in the mammalian Ig gene loci and may drive recombination of immediately adjacent elements. In terms of the evolution of this highly diversified multigene family, selective pressures may operate on the entire cluster vs. individual elements. Despite the near homogeneity in the sequences of germline \(D_n\) segments, the high degree of diversity in the D regions of cDNAs that derive from \(10^3/\lambda 33141\) is remarkable. The D region, which includes both \(D_n\) and \(N\) additions, ranges from 5 to 13 amino acids; however, no D region sequence in these comparison groups shares more than three amino acids with any other cDNA and typically show only two amino acids in common. An average of six unique amino acids (range, 4–9) are introduced in the D regions of the seven cDNAs in the \(10^3/\lambda 33141\) comparison group, whereas the \(V\) segments of these same genes, which are 10 times longer, average only four replacements, with a range of 0–7 changes. In \(10^3/\lambda 33141\)-type genes, it is neither germline diversity nor somatic mutation but D region variation that contributes extensive sequence diversity. This diversity arises from junctional N diversity and use of \(D_n\) sequences in all three reading frames. These findings in a lower vertebrate are in marked contrast to the human \(D_n\) region, in which \(\sim 30\) germline \(D_n\) gene segments have been characterized (25–27). There is no \(D_n\) consensus sequence, and these D's range in length from 11 to 44 nucleotides (23). The murine D locus is similarly diversified (23); however, two groups of three identical D segments are present among the 15 avian \(D_n\) elements that have been described previously (28). Without exception, all relevant germline cDNA comparisons suggest that somatic mutation or a highly restricted form of gene conversion is giving rise to the observed sequence variation. Mutation also is observed in \(D_1\) and \(J_1\) segments; however, no mutations (changes) are noted in the \(C_n\) segments of any of the cDNAs in this study. Thus, the \(C_n\)1 exon, which typically is localized 8–9 kb from \(J_n,\) does not appear to be targeted for mutation, consistent with the previously described restriction of the Ig mutational (hypermutation) mechanism to the V region, and immediately flanking sequences (29, 30).

The existence of a somatic mutation mechanism seems contrary to reports that the hapten-specific immune response of elasmobranchs is not associated with affinity maturation or interindividual variation (31), even after prolonged periods of antigenic stimulation (32). As indicated above, the cDNAs were not selected as components of an antigen-driven selection process. Rather, cDNAs in these unimmunized animals are being related to parental clusters, without any knowledge of their specificity. The only real requirement for cDNA detection was imposed at the screening stage and would eliminate from study only those clones in which the CDR2 had been so mutated as to render the clones unreactive with the CDR2-specific oligonucleotide probe. While presumed mutation effects are apparent and occur preferentially in CDRs, caution is warranted against any conclusions that such a process either alters primary antibody recognition or occurs with sufficient frequency to effectively alter the specificity and/or affinity of the antibody response. The absence of germlinal centers in this (and other lower vertebrate) species may result in a failure to select somatic mutations. Alternatively, this (and other lower vertebrate) species may use other mechanisms of selection.

The sequence of joining and mutational events in the *Het*
**References**


