Sequence and Characterization of the Ig Heavy Chain Constant and Partial Variable Region of the Mouse Strain 129S1

Ida Retter, Christophe Chevillard, Maren Scharfe, Ansgar Conrad, Martin Hafner, Tschong-Hun Im, Monika Ludewig, Gabriele Nordsiek, Simone Severitt, Stephanie Thies, America Mauhar, Helmut Blöcker, Werner Müller and Roy Riblet

_J Immunol_ 2007; 179:2419-2427; doi: 10.4049/jimmunol.179.4.2419

http://www.jimmunol.org/content/179/4/2419

Supplementary Material http://www.jimmunol.org/content/suppl/2008/03/14/179.4.2419.DC1

References This article cites 49 articles, 17 of which you can access for free at: http://www.jimmunol.org/content/179/4/2419.full#ref-list-1

Subscription Information about subscribing to _The Journal of Immunology_ is online at: http://jimmunol.org/subscription

Permissions Submit copyright permission requests at: http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts Receive free email-alerts when new articles cite this article. Sign up at: http://jimmunol.org/alerts
Sequence and Characterization of the Ig Heavy Chain Constant and Partial Variable Region of the Mouse Strain 129S1

Ida Retter, Christophe Chevillard, Maren Scharfe, Ansgar Conrad, Martin Hafner, Tschong-Hun Im, Monika Ludewig, Gabriele Nordsiek, Simone Severitt, Stephanie Thies, America Mauhar, Helmut Blöcker, Werner Müller, and Roy Riblet

Although the entire mouse genome has been sequenced, there remain challenges concerning the elucidation of particular complex and polymorphic genomic loci. In the murine Ig locus, different haplotypes exist in different inbred mouse strains. For example, the Igh loci haplotype sequence of the Mouse Genome Project strain C57BL/6 differs considerably from the Igh haplotype of BALB/c, which has been widely used in the analyses of Ab responses. We have sequenced and annotated the 3' half of the Igh locus of 129S1/SvImJ, covering the C_{H} region and approximately half of the V_{H} region. This sequence comprises 128 V_{H} genes, of which 49 are judged to be functional. The comparison of the Igh sequence with the homologous Igh region from C57BL/6 revealed two major expansions in the germline repertoire of Igh. In addition, we found smaller haplotype-specific differences like the duplication of five V_{H} genes in the Igh locus. We generated a V_{H} allele table by comparing the individual V_{H} genes of both haplotypes. Surprisingly, the number and position of D_{H} genes in the 129S1 strain differs not only from the sequence of C57BL/6 but also from the map published for BALB/c. Taken together, the contiguous genomic sequence of the 3' part of the Igh locus allows a detailed view of the recent evolution of this highly dynamic locus in the mouse.


The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 This research was supported by the German Bundesministerium für Bildung und Forschung through Grant PT DLR (FKZ 01KW0003) and the Bioinformatics Competence Centre “Integenomics” Grant 031U1110A/031U210A and National Institutes of Health Grant R01 AI23548.

2 I.R. and C.C. are co-first authors.

3 Current address: Institut National de la Santé et de la Recherche Médicale, Immunology and Genetics of Parasitic Diseases, Marseille, France and Laboratory of Parasitology-Myco, faculté de médecine, université de la Méditerranée, Marseille, France.

4 H.B., W.M., and R.R. are co-last authors.

5 Address correspondence and reprint requests to Prof. Werner Muller at his current address: University of Manchester, Bill Ford Chair of Cellular Immunology, Michael Smith Building, Oxford Road, Manchester, U.K.

6 Abbreviations used in this paper: RSS, recombination signal sequence; BAC, bacterial artificial chromosome; STS, sequence-tagged site; PIP, percent identity plot; YAC, yeast artificial chromosome.
and will support the elucidation of the special phenomena occurring in the Igh locus: VDJ recombination, class switch recombination, and somatic hypermutation, all of which are guided, regulated, or at least influenced by different sequence motifs (10–12).

Materials and Methods

Identification of Igh bacterial chromosome (BAC)

Igh locus BACs were isolated from the CitbC7T, abbreviated CT7, library (see http://www.trec.caltech.edu/). This library was prepared from the CJ7 ES cell line (13) derived from the 129S1/SvImJ mouse strain (ftp://ftp.informatics.jax.org/pub/reports/ES_CellLine.rpt). Library superpools and high-density membranes for screening and individual BAC clones were purchased from Research Genetics, now Invitrogen Life Technologies, and are also available from OpenBiosystems. Superpools were screened with a series of sequence-tagged site (STS) described in Ref. 14. These included sites in the 3′ end of the CH region (IgA exon 3), JH region, DH region (DFL16.1), and various VH region sites (VHgroupIII-VH 7183, VH11, VH107, VH GAM3-8, and VHJ606), and a yeast artificial chromosome (YAC) end (ADGC9-left arm) located in the VH10 region. Positive

FIGURE 1. Physical map of the 129S1 Igh locus, 3′ part. The locus is shown in 5′→3′ orientation displayed in segments of 200 kb each. Positions of genes are depicted as vertical lines. Short vertical lines indicate pseudogenes and “possibly functional” genes. VH genes are colored corresponding to their sequence family. Positions of the underlying BACs are indicated as gray horizontal lines below the scaling line. The unfinished sequence of the most 5′ BAC, 4K8, is marked with a gray box; preliminary gene nomenclature in this 5′ region is indicated by #. The sequence and annotation identifying each gene segment and its sequence coordinates are available from EMBL/GenBank/DDBJ (accession no. AJ851868). VH gene sequences may also be downloaded from VBASE2 (http://www.vbase2.org).
mouse BAC clones were obtained from Research Genetics, plated on agar plates containing 12.5 μg/ml chloramphenicol, and confirmed by colony PCR with the identifying primer sets. BAC ends were sequenced either directly from T7 and SP6 primers or following amplification by Vector-Hexamer PCR (15). BAC end sequences were deposited in EMBL/GenBank/DDBJ accession nos. BH021141–BH021349. Contig assembly proceeded by assessing STS content using the screening sites, D12Mit markers and others, and extensive Southern blot analysis using V_{H} probes and BAC and YAC ends, all as described in Ref. 14. Gaps were closed by developing new screening PCR assays from BAC end sequences.

### Sequencing and assembly

Sequencing of 23 BACs of the CT7 library was performed by a shotgun approach as follows: sheared fragments of either 1 or 3 kb in length (GeneMachines) were subcloned separately into a pTZ18R vector. At least 800 clones were selected from each clone library, most of the plasmid DNA was prepared following a protocol supplied by Millipore. One-third of the selected clones were amplified by TempliPhi (384-well plates containing 12.5 μ/ml chloramphenicol, and confirmed by colony PCR with the identifying primer sets. BAC ends were sequenced either directly from T7 and SP6 primers or following amplification by Vector-Hexamer PCR (15). BAC end sequences were deposited in EMBL/GenBank/DDBJ accession nos. BH021141–BH021349. Contig assembly proceeded by assessing STS content using the screening sites, D12Mit markers and others, and extensive Southern blot analysis using V_{H} probes and BAC and YAC ends, all as described in Ref. 14. Gaps were closed by developing new screening PCR assays from BAC end sequences.

### Sequence analysis

The V_{H} genes were annotated using a procedure developed for the automatic generation of the database VBASE2 (17). In this procedure, V_{H} genes are detected by a BLAST search (18) of the BAC sequences with known germline V_{H} gene sequences. BLAST hits with a minimum identity of 80% and minimum alignment length of 200 bp are analyzed with the Dnablast program (W. Müller and H. H. Althaus, unpublished data; http://www.dnaplot.de) and matched to VDJ rearrangements from the EMBL/GenBank/DDBJ. The Dnablast analysis is limited to exon 2 of the V_{H} genes, but includes detection of RSS elements. Exon 1 of each V_{H} gene, D_{H} gene, J_{H} gene, and C_{H} gene have been annotated manually by sequence comparisons with BLAST, Pipmaker (19) and other alignment programs, referring to previously published annotations; the EMBL/GenBank/DDBJ accession no. of the respective reference sequences are given in the feature table of the EMBL/GenBank/DDBJ entry AJ851868.

Dot plots and percent identity plots (PIPs) have been generated with the Pipmaker program, parameters: “search both strands,” “show all matches” sensitivity mode; “default,” “show all matches” for dot plots, and “chaining” for PIPs. Interspersed repeats have been detected by using the Repeatmasker web server (A. F. A. Smit, R. Hubley, and P. Green, unpublished data; http://www.repeatmasker.org). The searches were performed against the Repbase mouse dataset with the “cross_match” search engine and slow sensitivity mode. Phylogetic trees have been generated with MEGT alignments created with Dnablast (20) and visualized with MEGA version 3.1 (21). Further sequence analysis and formatting have been done with the emboss program package (22) and Perl scripts.

### Results

#### The sequence of the 3'\textsuperscript{'} half of the Igh\textsuperscript{a} locus

To elucidate the genomic sequence of the murine Igh\textsuperscript{a} haplotype, 23 BACs from the Ctb, or CT7, library of the 129S1 mouse were selected from a physical map of the locus assembled with YAC and BAC end STS, D12Mit simple sequence length polymorphism sites, and Igh-C, J, D, and V gene segments (Refs. 14 and 15; C. Chevillard and R. Riblet, unpublished observations). The BACs were sequenced with a 10-fold coverage on average. The BAC inserts represent an overall length of 2.5 Mb and were assembled to a sequence of 1.6 Mb. This sequence covers approximately one-half of the Igh locus, including the C_{H} region, J_{H} region, D_{H} region, and the J_{H}-proximal part of the V_{H} region (Fig. 1). The assembly is contiguous between the BACs 407I12 and 34H6, except for two simple repeat stretches of unknown length, one in 34H6 and one in the overlap of 459E6 and 436C3. The unfinished

### Table I. Sequence comparison of V_{H} pseudogenes and “possibly functional” V_{H} genes vs functional V_{H} genes

<table>
<thead>
<tr>
<th>Sequence Identity to the Most Similar Functional Gene (%)</th>
<th>No. of V_{H} Pseudogenes</th>
<th>No. of Possibly Functional V_{H} Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;70</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>70–79</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>80–89</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>90–99</td>
<td>6</td>
<td>12</td>
</tr>
</tbody>
</table>

### Table II. Summary of the functional V_{H} genes and V_{H} pseudogenes in the J_{H}-proximal part of the mouse Igh\textsuperscript{a} locus grouped by family

<table>
<thead>
<tr>
<th>Family</th>
<th>129S1</th>
<th>C57BL/6</th>
</tr>
</thead>
<tbody>
<tr>
<td>V_{H}7183</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>V_{H}Q52</td>
<td>14</td>
<td>(5)</td>
</tr>
<tr>
<td>V_{H}S107</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>V_{H}SM7</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>V_{H}36–60</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>V_{H}11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>V_{H}X24</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>V_{H}AM3–8</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>V_{H}3609N</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>V_{H}16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>V_{H}12</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>V_{H}606</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>V_{H}10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>V_{H}36</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>V_{H}558</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>V_{H}3609</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Unclassified relics</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>59</td>
</tr>
</tbody>
</table>

*The number of V_{H} genes in C57BL/6 published in Ref. 9 is given in parentheses when different from the number published in Ref. 7. Only V_{H} genes with 100% sequence identity to a known VDJ rearrangement are listed as functional.

\( ^{a}\) The region of this family is not covered (V_{H}10, V_{H}3609) or is incompletely covered (V_{H}J606, V_{H}15, V_{H}558) by the available sequence.

\( ^{b}\) —, not applicable.
sequence of BAC 4K8 overlaps 407I12 and extends the assembly toward the JH distal part of the VH region with 12 single fragments. The order of these fragments is hypothetical, based on sequence comparison with the homologous Ighb region in C57BL/6. The complete assembled sequence and annotation are available from EMBL/GenBank/DDBJ (accession no. AJ851868).

**Annotation of VH, D, H, JH, and CH genes in the assembled sequence**

The Igh^a^ sequence assembly was screened for VH genes using a BLAST search and a subsequent V gene analysis method based on the program DNAPLOT (17). This method annotates the V gene beginning from the coding sequence of the mature peptide chain up to the noncoding RSS sites at the 3’ end of each V gene. The results of this procedure were complemented by the annotation of exon 1. VH genes with obvious defects like stop codons, defective splice sites, and abnormal RSS elements were marked as pseudogenes. To further address the question of V gene functionality in a defined in silico approach, VH genes were matched against 6190 Ig VDJ rearrangements extracted from EMBL/GenBank/DDBJ. Those VH genes with a 100% match to a rearranged sequence are classified as functional. A physical map of the annotated region is shown in Fig. 1.

In the 3’ half of the Igh^a^ V region, 128 VH genes were annotated, of which 49 have been found in rearrangements. Of 79 VH genes without clear evidence of functionality, 63 are obvious pseudogenes. The remaining 16 VH genes are designated as “possibly functional.” To further analyze the potential functionality of the VH genes, we performed a pairwise sequence comparison of all 79 possibly functional and pseudogenes with the 49 functional VH genes. The result (summarized in Table I) shows that the majority of pseudogenes have major changes in their sequence, whereas some of them are very similar to functional genes. Although 12 possibly functional VH genes have at least 90% sequence identity to a functional gene, this is also the case for 6 pseudogenes. This indicates that a high sequence similarity to a functional gene cannot be taken as evidence for functionality, and the functionality of the 16 possibly functional genes cannot be proven based on the available data. However, because somatic mutations in the VDJ rearrangements prevent a 100% match to a germline sequence, it is likely that at least some of the possibly functional sequences are indeed functional.

One functional and four pseudogenes have an exact duplicated counterpart. The VH gene sequence family assignment revealed that the currently elucidated part of the locus comprises members from 12 of 15 known VH gene families, lacking representation from only the most 5’ VH3609P, VH10, and VHJ558 families (Table II), with the exception of the pseudogene VHJ558.a1psi.119. Seven VH^a^ pseudogenes could not be assigned to a certain VH gene family and were designated as VH relics, of which four are located in the V^a^ region and closely related. Functional RSS elements were found at the 3’ end of 72 VH genes. All RSS elements from VH^a^ genes of the families VHGAM-8 and VHGAM7, and 2 RSS elements of the VH36-60 family exhibit a 22-bp spacer. Twenty VH pseudogenes show abnormal RSS elements. A table listing all VH genes with their names, properties, and positions within the assembly is available as supplemental data.7

The VH genes have been named according to previously published nomenclature rules (4, 9). Each name consists of three parts separated by dots: the first part gives the VH gene family. The second part gives one letter for the Igh haplotype, in this case “a,” and the number of the VH gene within this particular family followed by the psi tag for pseudogenes. The third part gives a number for the position of the VH gene within the locus, where both the absolute and the family numbering run from the JH proximal to the distal side of the locus. Names for VH genes within the 12 fragments of BAC 4K8 are given provisionally and therefore are marked with # at the end. Because other nomenclature schemata for VH genes have been published (8, 23) and numerous

---

7 The online version of this article contains supplemental material.
The distinct polymorphisms in the murine Ig receptor loci are due to the strong evolutionary dynamics of these loci. The divergent evolution of V_{H} genes has been described as resulting from diversifying selection and evolution by birth and death process (24). The inbred strains provide something like a snapshot of these loci, so that haplotype and even allele assignment is possible. We compared the Igh^{a} and Igh^{b} haplotype by a dot plot of the homologous V_{H} region sequences from 129S1 and C57BL/6 (Fig. 2A). This plot indicates that the overall structure, namely, the position of the discrete V_{H} gene family clusters, is very similar in both haplotypes, a finding that confirms previous experimental results based on Southern blot analyses (25). Concerning individual V_{H} gene family clusters, the dot plot shows two regions of expansion in the Igh^{b} locus: the V_{H}7183/V_{H}Q52 region (previously described in Ref. 26) and the interspersed V_{H}GAM3-8/V_{H}12 families. The V_{H}7183/V_{H}Q52 cluster is well conserved between both strains at the 3’ end, followed by a sequence region which is much larger in Igh^{b} than in Igh^{a}. This local expansion results in approximately twice as many functional V_{H}7183/V_{H}Q52 genes in the Igh^{b} repertoire compared with Igh^{a} (Table II). Similarly, the V_{H}GAM3-8/V_{H}12 region is larger and comprises more V_{H} genes in Igh^{b} compared with Igh^{a}. A detailed view of the differences between Igh^{b} and Igh^{a} is provided by a percent identity plot of the homologous regions (supplemental data). The plot displays annotations and interspersed repeats in the 129S1 sequence along with the percent identity detected in the C57BL/6 sequence.

To resolve the V region comparison to the level of single V_{H} genes, we performed a multiple sequence alignment of the combined V_{H} gene repertoire of both strains. Whenever an Igh^{b} V_{H} gene had its best match within the Igh^{a} repertoire and inversely the best match of this Igh^{a} V_{H} gene was the referring Igh^{b} V gene, we designated this V_{H} gene pair as alleles (Table III). In case of the duplicated Igh^{b} V genes, one Igh^{b} allele refers to the pair of Igh^{a} alleles. For all alleles a minimum sequence identity of 90% was combined VH gene repertoire of both strains. Whenever an Igh^{a} V_{H} gene had its best match within the Igh^{b} repertoire and inversely the best match of this Igh^{b} V_{H} gene was the referring Igh^{a} V gene, we designated this V_{H} gene pair as alleles (Table III). In case of the duplicated Igh^{b} V genes, one Igh^{b} allele refers to the pair of Igh^{a} alleles. For all alleles a minimum sequence identity of 90% was postulated. With this method, 53 pairs of alleles were assigned. The genes of the VHJ606 family were excluded from the allele assignment, as both the order of these genes is hypothetical and it is unlikely that all VHJ606 genes are included in the current assignment, as both the order of these genes is hypothetical and it is unlikely that all VHJ606 genes are included in the current assignment.

The Igh-D region of 129S1 consists of two separate homology blocks

In the Igh-D region of 129S1, 15 D_{H} genes are annotated and assigned to one of the sequence families DSP2, DFL16, DST4, or trivial names are in use, we collected all names to our knowledge for each V_{H} gene and make them available in the database VBASE2 (17).

The Journal of Immunology

FIGURE 3. Alignment of the D_{H} gene families DST4 and DFL16. RSS elements are marked as gray boxes.
DSP2.2a and DSP2.2b. Two new DH genes are seen, DFL16.3 and DST4.3, which are probably not functional (see below). The dot plot of the DH region (Fig. 2B) shows that the DH genes are arranged in two separate homology blocks: the main block, including all DSP2 genes, starts before DFL16.1 and ends after DST4. Upstream of this major DH cluster there is an additional small block consisting of the genes DST4.2, DMB1, and DFL16.3. The DH gene DQ52 constitutes the 3' end of the DH region and is separate from the main DH gene block. Both the main and the small blocks are composed of multiple related repeats of a 3-kb sequence containing one DH gene.

The sequences of the DH genes DFL16.3, DST4.2, and DST4.3 (Fig. 3) were checked in silico for potential functionality. A BLAST search of the DH gene sequences was performed against a database comprised of the junctional regions of 6190 Ig VDJ rearrangements extracted from EMBL/GenBank/DDBJ. For DFL16.3 and DST4.3, no specifically matching rearrangements were found, so both DH genes are classified as nonfunctional. For DST4.2, six rearrangements were found that shared the DST4.2-specific insertion of a cytidine nucleotide in the second position. Interestingly, three rearrangements stem from a mouse line where the main block of DH genes has been deleted (27). Thus, it seems that the lack of DH genes from the main block enforces the usage of the otherwise seldom or nonused DH genes of the small block.

Comparison of the DH genes of 129S1, BALB/c, and C57BL/6

Although the IgV regions of different inbred strains of mice allow a clear assignment to a relatively small number of haplotypes, the IgD regions show even more polymorphism (28). Although DH gene usage and the role of the DH region within the process of VDJ rearrangement have been intensively studied (29–33), a complete nucleotide sequence of the DH region of BALB/c has never been published. The existing map is based on Southern blot hybridization experiments (34, 35). The Mouse Genome Sequencing Consortium provided a complete sequence of the DH region of C57BL/6, which has recently been annotated (7, 36). Fig. 4 compares the DH regions of 129S1, BALB/c, and C57BL/6. The three maps have a common feature, namely, the main DH gene block is separate from the downstream DH gene DQ52. The small DH gene block is present in 129S1 and C57BL/6, but it is unknown whether it exists in BALB/c. Compared with 129S1 and BALB/c, the main DH gene block is smaller in C57BL/6 and contains fewer genes. The main blocks of 129S1 and BALB/c contain different DSP2 genes. Differences in the size of the main DH blocks may be due to the fact that the BALB/c map is of limited resolution. However, the available data show notable differences in the DH region of both IghR haplotype strains 129S1 and BALB/c.

The JH and Cγ regions of 129S1

Close to the most downstream DH gene DQ52, four JH genes are located within a short stretch of 1.5 kb. The adjacent Cγ region contains eight genes coding for eight different isotypes of the H chain of the Ab. Each Cγ gene consists of a group of three to five exons coding for different domains in the Ab molecule. B cells can switch the isotype of the produced Ab by class switch recombination. This is mediated by repetitive sequences upstream of the particular isotype exon group, the switch (S) regions. JH genes, Cγ region exons, and S regions were annotated referring to previously published annotation of these sequences from BALB/c mice (for references, see EMBL/GenBank/DDBJ AJ851868). In addition to the previously known CγψY pseudogenes (37), one new CγψY pseudogene was found, consisting of CH2, CH3, and a truncated M1 exon. The new pseudogene was designated as Cγψpψψψ0 and is in an inverted orientation located upstream of the Cγψψψψ3 group, thereby heading the Cγψψψψ cluster.

To depict the genomic structure of the Cγ region, comprising internal sequence homologies as well as location and size of repetitive sequences, we performed a dot plot of the genomic sequence (Fig. 5). In the dot plot, S regions appear as black boxes adjacent to the I exon. The S region of Cγψψψψ1 is noticeably enlarged. Four sequence blocks comprising the Cγψψψψ exon groups show strong conservation, starting upstream of the I exons and...
extending beyond the membrane exons of each group. Parts of this conserved block are inversely inserted upstream of C\(_{\text{H}}\)\(^{3}\), C\(_{\text{H}}\)\(^{2b}\), and C\(_{\text{H}}\)\(^{2a}\). These findings about the overall C\(_{\text{H}}\) region structure are in accordance with data published on the C\(_{\text{H}}\) region of BALB/c (37, 38). A refined view of the relationship between the J\(_{\text{H}}\) and C\(_{\text{H}}\) loci of both strains is gained by comparison of the coding sequences. The sequences of the four J\(_{\text{H}}\) genes of 129S1 and BALB/c are identical, except for a silent point mutation in the third valine codon of J\(_{\text{H}}\)\(^{1}\) (C vs T). Remarkably, the C\(_{\text{H}}\) region coding sequences from 129S1 and BALB/c show several amino acid exchanges concerning the isotypes IgD, IgG1, IgG2b, IgE, and IgA (Table IV). However, the CDS of IgM, IgG3, and IgG2a are identical.

**Discussion**

**Nomenclature of the mouse Igh-V genes**

For a sensible gene nomenclature, the gene name should provide a reasonable amount of information about the gene. In particular for the Igh locus of mouse, numerous nomenclature suggestions have been made, each representing different aspects of the genes. The IMGT (http://imgt.cines.fr/) established a systematic nomenclature by assigning a number to each V gene family and numbering V genes within each family (39). This nomenclature regards genes independent of their chromosomal position and haplotype, but it accounts for V gene alleles. de Bono (8) applied another rule set to genes within each family (39). This nomenclature regards genes IMGT (http://imgt.cines.fr/) established a systematic nomenclature been made, each representing different aspects of the genes. The nomenclature rules that Johnston et al. (9) and we applied (Table IV). However, the CDS of IgM, IgG3, and IgG2a are identical.

**Heterogeneity within the Igh\(^{h}\) haplotype**

Serological studies using polyclonal antiallotype sera classified 129/Sv and BALB/c together as Igh\(^{h}\) (40), and later studies with mAbs to BALB/c (i.e., IgM, IgD, and IgG1) bound 129/Sv Ig as well (41, 42). RFLP analyses of the V\(_{\text{H}}\) region showed a strong correspondence between the C\(_{\text{H}}\) region and V\(_{\text{H}}\) region haplotypes (4, 43), meaning that strains with identical C\(_{\text{H}}\) region patterns often also exhibit the same V\(_{\text{H}}\) region patterns. For 129/Sv, comprehensive Southern blot analyses revealed restriction patterns identical to the V\(_{\text{H}}\) region of BALB/c, with the exception of the pattern for the VH\(_{\text{H}}\)309P family at the distal part of the locus (5).

Based on this, it was unexpected to find the remarkable differences between the C\(_{\text{H}}\) coding sequences of 129S1 and BALB/c that are listed in Table IV. The D\(_{\text{H}}\) region sequence of 129S1 shows a mixed haplotype, represented by a DST4-Igh\(^{a}\) allele and a DQ52-Igh\(^{a}\) allele. A detailed comparison with the D\(_{\text{H}}\) region of BALB/c is limited by the lack of BALB/c genomic sequence. However, based on the available sequence information, there are obvious differences between the D\(_{\text{H}}\) regions of BALB/c and 129S1. Taken together, our results point to a distinct heterogeneity within the Igh\(^{h}\) haplotype that had not been detected in previous experiments and which might affect, to a minor degree, also the V\(_{\text{H}}\) region of 129S1.

**Interspersed repeats in the murine Igh locus**

An unusually high content of interspersed repeats has been repeatedly noted for the murine Igh and Igd loci (14, 15, 44). In the elucidated V\(_{\text{H}}\) region of 129S1, Repeatmasker analysis shows that interspersed repeats occupy 54% of the sequence: the content of LINE-1 (L1) elements is 36%, whereas the SINE content is <2%. This considerably deviates from the average for the mouse genome, where 19% L1 and 8% SINE content have been reported (6). The distribution of repetitive elements in the V\(_{\text{H}}\) region of 129S1 is very similar to the distribution that was reported for the V\(_{\text{H}}\) region of C57BL/6 (9). This unusually high density of LINE elements has several interesting parallels; a high density has been noted around other monoafllelically expressed genes (45), and the X chromosome has a high LINE density. Lyon (46) has proposed that this L1 density is a factor in the heterochromatization of the inactive X. To visualize interspersed repeats in relation to the V, D, J, and C gene positions, we performed a percent identity plot (PIP) of the Igh\(^{a}\) vs Igh\(^{h}\) sequence with the PipMaker program (supplemental data). The PIP shows not only differences between the haplotypes, but also the position and class of interspersed repeats in the 129S1 sequence and can therefore be taken as a high-resolution map of the region. It graphically displays the relatively uniform distribution of L1 elements throughout the V\(_{\text{H}}\) region and rarity in the C\(_{\text{H}}\) region; also apparent are insertions of L1 elements in 129S1 that are absent in C57BL/6, consistent with the continuing evolution of this complex locus.

**Expansion of functional and nonfunctional V\(_{\text{H}}\) genes by block duplications**

Duplications of both large and small sequence blocks are a common phenomenon in the Ig loci (9, 47–49) and are assumed to be an essential force in the generation of multiple gene copies in these loci. Ota and Nei (24) explain the maintenance of diversity within the large V\(_{\text{H}}\) gene family by selective forces favoring diversification of the duplicated genes. To explain the high number of pseudogenes within the Ig loci, Kawasaki et al. (48) suggested the coamplification and fixation of pseudogenes along with adjacent functional V genes. We tested this hypothesis on the VH\(_{\text{H}}\)7183/ VH\(_{\text{H}}\)Q52 region of the murine Igh-V locus, which is expanded in the 129S1 strain compared with C57BL/6 (Fig. 2A). We generated a phylogenetic tree of the VH\(_{\text{H}}\)7183 Ig\(^{h}\) family (Fig. 6A) and noticed that there is a clear separation of nonfunctional and functional VH\(_{\text{H}}\)7183 sequences, indicating that the nonfunctional

<table>
<thead>
<tr>
<th>C Region Gene</th>
<th>Exon/Domain</th>
<th>Nucleotide Position within Exon</th>
<th>Amino Acid Difference 129S1 vs BALB/c</th>
</tr>
</thead>
<tbody>
<tr>
<td>C(_{\text{H}})(^{\delta})</td>
<td>CH3</td>
<td>125..127</td>
<td>S vs G [J00475]</td>
</tr>
<tr>
<td>C(_{\text{H}})</td>
<td>CH1</td>
<td>219..224</td>
<td>TW vs PR [J00453]</td>
</tr>
<tr>
<td>C(_{\text{H}})</td>
<td>CH3</td>
<td>174..176</td>
<td>D vs N [J00453]</td>
</tr>
<tr>
<td>C(_{\text{H}})</td>
<td>CH3</td>
<td>180..182</td>
<td>D vs N [J00453]</td>
</tr>
<tr>
<td>C(_{\text{H}})</td>
<td>CH1</td>
<td>72..74</td>
<td>L vs S [J00763]</td>
</tr>
<tr>
<td>C(_{\text{H}})</td>
<td>CH1</td>
<td>105..107</td>
<td>S vs P [J00763]</td>
</tr>
<tr>
<td>C(_{\text{H}})</td>
<td>CH3</td>
<td>27..29</td>
<td>I vs T [J00763]</td>
</tr>
<tr>
<td>C(_{\text{C}})</td>
<td>CH1</td>
<td>93..95</td>
<td>G vs N [X01857]</td>
</tr>
<tr>
<td>C(_{\alpha})</td>
<td>CH1</td>
<td>45..47</td>
<td>A vs V [J00475]</td>
</tr>
<tr>
<td>C(_{\alpha})</td>
<td>CH1</td>
<td>51..53</td>
<td>S vs C [J00475]</td>
</tr>
<tr>
<td>C(_{\alpha})</td>
<td>Hinge-CH2</td>
<td>195..197</td>
<td>V vs A [J00475]</td>
</tr>
</tbody>
</table>

EMBL/GenBank/DDBJ accession no. of BALB/c reference sequences are given in brackets.
sequences have evolved independently from the functional sequences and there are, with the exception of V_{H}7183.a1psi.1 and V_{H}7183.a43psi.70, no pseudogenes related to the functional sequences. Looking at the physical map of the V_{H}7183 region, one can see repeated patterns where a certain functional gene is close to a pseudogene in a conserved distance and order (Fig. 6B). The first, most obvious example of such a block consists of a V_{H} relic sequence next to a functional V_{H}Q52 gene. This block, indicated by a noncolored box in Fig. 6B, appears four times within the V_{H}7183/V_{H}Q52 region. In addition, we find other blocks involving pseudogenes of the V_{H}7183 family next to functional genes of the V_{H}Q52 or the V_{H}7183 family, indicated as colored boxes in Fig. 6B. When we superimpose these blocks on the phylogenetic tree shown in Fig. 6A, we can nicely see clustering of the pseudogenes within the phylogenetic tree. From our analysis, we conclude that functional V_{H} genes are duplicated as large stretches of DNA containing more flanking sequences than necessary to encode for a functional V gene and by this “blockwise” duplication pseudogenes are expanded as well. In Fig. 6A, we further show the relationship to corresponding alleles of the C57BL/6 locus, indicating that the phylogenetic distribution between the functional and nonfunctional V_{H} genes remains valid also in case of the C57BL/6 strain. When we analyzed the order of functional and nonfunctional sequences within the entire Igh region, we could see an underrepresentation of two functional V_{H} genes next to each other (data not shown). In case of a random distribution, we would have expected a higher fraction of neighboring functional V_{H} genes. This finding supports the idea of blockwise expansion of large segments of DNA containing at least one functional V_{H} gene. We cannot rule out that the adjacent pseudogenes comprise functional regulatory elements that are used, for example, for the opening of the locus during B lymphocyte development or later by the functional V_{H} genes and are thereby positively selected. However, the latter possibility seems unlikely since there are examples of pseudogenes located both upstream and downstream of a functional gene within the indicated homology blocks. In addition, the fact that we find only one orientation of V_{H} genes, irrespective if these are functional, nonfunctional, or relics, points to a directed mechanism in the evolution of this locus. Currently, we have no conclusive explanation for the driving forces underlying the evolution of V_{H} genes. Given the complexity of the genealogical trees of mouse inbred strains (50), we cannot state when and how modifications of the V_{H}Q genes happened. Either they occurred during the generation of inbred strains within the last 100 years or these differences represent allelic variants of wild mice selected over a time span of 100,000 years or more that have been fixed during the generation of inbred lines. As BALB/c and 129S1 mice, although distantly related in the genealogical tree of mouse inbred strains, have similar Igh haplotypes, we favor the latter possibility.

Acknowledgments
We thank Klaus Rajewsky and Blair Prochnow for critical review of this manuscript.

Disclosures
The authors have no financial conflict of interest.

References


