# A three-megabase yeast artificial chromosome Contig spanning the C57BL mouse Igh locus

Christophe Chevillard, Jennifer Ozaki, Christopher D. Herring, Roy Riblet

# A Three-Megabase Yeast Artificial Chromosome Contig Spanning the C57BL Mouse *Igh* Locus[1]

**Christophe Chevillard,* Jennifer Ozaki,† Christopher D. Herring,‡ and Roy Riblet[2§]**

**The mouse Ig H chain (*Igh*) complex locus is composed of >100 gene segments encoding the variable, diversity, joining, and constant portions of the Ab H chain protein. To advance the characterization of this locus and to identify all the $V_H$ genes, we have isolated the entire region from C57BL/6 and C57BL/10 as a yeast artificial chromosome contig. The mouse *Igh* locus extends approximately three megabases and contains at least 134 $V_H$ genes classified in 15 partially interspersed families. Two non-*Igh* pseudogenes (*Odc-rs8* and *Rpl32-rs14*) were localized in the distal part of the locus. This physical yeast artificial chromosome map will provide important structure and guidance for the sequencing of this large, complex, and highly repetitive locus.** ***The Journal of Immunology*, 2002, 168: 5659–5666.**

The Ig H chain locus, termed *Igh* in mice, encodes the H chains of Abs. It is comprised of adjacent clusters of gene segments for $V_H$, $D_H$, $J_H$, and the different isotype H chain constant regions ($C_H$). To produce a secreted or cell surface receptor Ab, a B cell must assemble an active V region gene by fusing a segment from each of the V, D, and J clusters. The overall nature of the structure and functioning of these genes is well established, but much remains to be explained. If we are to thoroughly understand, and be able to predict and manipulate, the functioning of the Ab response, we must obtain a complete description of the structural loci, both coding elements and regulatory sequences of Ab H and L chains. In the human *IGH* locus, this goal is nearly complete (1), but for experimental manipulation these loci must be characterized in the mouse. Recently, the mouse κ L chain locus was extensively described by Zachau and colleagues (2). All *Igk* constant, joining, and variable coding elements have been identified and sequenced. For the mouse H chain locus, the constant and joining gene segments were mapped and sequenced by pioneering work of Honjo and coworkers (3), and the diversity segments have been identified, mapped, and sequenced (4–6). The V region gene segments, *Igh-V* or $V_H$, have posed a larger challenge; estimates of their numbers range from hundreds to thousands (7–10). Genetic mapping experiments indicated that the locus is a centimorgan or larger (11–13), suggesting a sequence size of several million base pairs of DNA, consistent with a gene content of hundreds to thousands of coding elements. Detailed studies of the organization of $V_H$ gene families by Brodeur and colleagues (14) have provided a better understanding but much remains to be elucidated.

We have undertaken several approaches to complete the characterization of mouse Igh; these will ultimately lead to determination of the DNA sequence of the entire locus. In this study, we describe the assembly of a yeast artificial chromosome (YAC)[3] contig, or array of overlapping clones, that spans the *Igh* locus in the C57BL mouse strains, and the initial characterization of its size, physical structure, and gene content. To develop this YAC contig and resultant physical map of the mouse *Igh* locus, we screened four YAC libraries by PCR using multiple sequence-tagged sites (STSs) within the locus and identified 36 YACs. Several additional YACs were obtained from a fifth library through reference to Internet-posted data (15). YAC insert ends were isolated from all clones by vector-hexamer PCR (16) to characterize YAC overlap and assemble the contig, and to detect chimeric clones. This physical YAC map will provide important structure and guidance for the sequencing of this large, complex, and highly repetitive locus.

## Materials and Methods

### *YAC libraries*

The Princeton University mouse YAC library (Princeton, NJ) was prepared from C57BL/6J female mouse genomic DNA (17, 18). It consists of 26,000 individual pYAC4 vector clones in yeast host strain AB1380. The average clone length is 250 kb and the total estimated genome coverage is 2.2 haploid genomic equivalents. This library was screened by PCR using DNA pools from Dr. S. M. Tilghman at the Howard Hughes Medical Institute, Princeton University.

The first Whitehead Institute (WI-I) mouse YAC library (Cambridge, MA) was prepared from C57BL/6J female mouse DNA (19). It consists of two groups of pYAC4 vector clones in yeast host AB1380. The first group contains 4,100 clones with an average size of 480 kb, and the second contains 15,840 clones with an average size of 640 kb. The total estimated coverage for the whole library is 4.3 haploid genomic equivalents.

The second Whitehead Institute (WI-820) mouse YAC library was made from female C57BL/6J mouse DNA (20). It contains 38,400 clones with an average size of 820 kb, providing 10-fold coverage. All clones from this library are in the pRML vector in yeast host strain J57D. Extensive characterization of this library and identification of YACs bearing over 600 genetic loci is published (15). The Whitehead Institute libraries can be screened by pool PCR or membrane hybridization, and clones can be obtained from Research Genetics (Huntsville, AL).

The Saint Mary's Hospital Medical School (London, U.K.) RAD52 mouse YAC library was prepared from C57BL/10 female mouse DNA (21). It contains 41,568 pYAC4 vector clones in the RAD52 mutant yeast host strain 3a with an average insert size of 240 kb, providing 3.5 genome equivalents. Clones from this library are available from the Mouse Genome

*Faculty of Medicine, Immunology and Genetics of Parasitic Diseases, Institut National de la Santé et de la Recherche Médicale, Marseille, France; †Division of Genetic Medicine, Department of Medicine, Vanderbilt University, Nashville, TN 37232; ‡Department of Genetics, University of Wisconsin, Madison, WI 53706; and §Torrey Pines Institute for Molecular Studies, San Diego, CA 92121

Received for publication November 2, 2001. Accepted for publication March 28, 2002.

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 work was supported by National Institutes of Health Grant AI23548.

[2] Address correspondence and reprint requests to Dr. Roy Riblet, Torrey Pines Institute for Molecular Studies, 3550 General Atomics Court, San Diego, CA 92121. E-mail address: rriblet@tpims.org

[3] Abbreviations used in this paper: YAC, yeast artificial chromosome; STS, sequence-tagged site; PFGE, pulsed field gel electrophoresis; BLAST, basic local alignment search tool.

 0022-1767/02/$02.00

Table I. *Sequence tagged sites in the* Igh *locus*[a]

| Name | STS | Primer 1 | Primer 2 | PCR Product Size (bp) | PCR Conditions (mM/°C)[b] | GenBank Accession No. | Primer Source/ Reference |
|---|---|---|---|---|---|---|---|
| $C_H$ region | | | | | | | |
| | 3′ IgA Enhancer | ctcaaggttcgagttactcattctgtgca | ctcaaggttcaggatttggagcacacctacag | 376 | 1.5/60 | X96607 | 24[c] |
| D12Rbt6 | IgA | aagaagtgctggtgcgatg | actggtcaccctgtttccag | 159 | 1.5/55 | J00475 | This study |
| D12Rbt7 | IgG1 | ctgctgcccaaactaactcc | cttgtccaccttggtgctg | 242 | 1.5/55 | J00453 | This study |
| D12Rbt8 | IgG3 | ggggccgcctattaattaaa | tcctgctgtcctttgattcc | 147 | 1.5/55 | J00451 | This study |
| D12Nds4 | IgD | agacttattgtaccccacatgttg | tatcttccaatcctagttaggc | 160 | 1.5/55 | K02138 | 25 |
| $J_H$ region | | | | | | | |
| D12Rbt9 | J3/J4 | tggtgacaatttcagggtca | gtctgactagaatcacccctgg | 205 | 1.5/55 | X63166 | This study |
| $D_H$ region | | | | | | | |
| | $D_H$FL16.1 | gagcatgttgcaggaactga | tcaaaattttccccaatagg | 310 | 1.5/60 | AF018146 | —[c] |
| $V_H$ region | | | | | | | |
| D12Rbt10 | $V_H$ group III | gaggtgcagctggtgg | gctcacagtaactttyrctcactgtg | 322 | 1.5/55 | K02890 | 26, this study |
| | $V_H$11 | atggagtgggaactgagctta | catacagaaatacgtggctgtgtcc | 461 | 1.5/55 | M22438 | 27[c] |
| D12Nds2 | $V_H$S107 | acatggtaatttatgggcaa | ctggatacctgcaatagtaga | 190 | 1.5/55 | X03253 | 25 |
| D12Rbt11 | $V_H$Ese26.1 | aagctgggtgaagcagagaa | gtcctcagatgtcaggctgc | 169 | 1.5/55 | S49538 | This study |
| D12Rbt12 | $V_H$ Group II | caggtccaactgcagcag | gatgtggttgcaacactgtg | 314 | 1.5/55 | M60252 | 26, this study |
| | $V_H$15 | caggttcacctacaacag | ccttgcacagtagtagat | 297 | 1.5/55 | U39293 | 14[c] |
| D12Rbt13 | MMIGV$_H$28 | tgcattcattttttaaaatgc | atttgcctttccacaggttg | 348 | 1.5/55 | X06862 | This study |
| D12Rbt14 | MMIGD10 | ttccatgtccaaacgcagt | tatgggtcaacagaatattgca | 147 | 1.5/55 | X04369 | This study |
| | MMIGV$_H$23 | gtgaaagtttccatcaatatc | catgtatttagagattgtgct | 103 | 1.5/55 | X06856 | 28 |
| | $V_H$G8 | tatcctggtagcattactactaac | tcttgcacaataatagaccgc | 141 | 1.5/45 | X60424 | 29[c] |
| | $V_H$186.2 | tgatgcaatattctgttgaccc | agagtcctcagatgtcaggct | 1011 | 1.5/65 | L26851 | 10 |
| D12Nds10 | MMIGHVJ2 | catgtttgcaatttaccatca | catggatttacagatggagct | 100 and 140 | 2.0/55 | J00507 | 28 |
| YAC ends | | | | | | | |
| D12Rbt1 | yADGC9 left arm | ccatggtgagagaagtttattctt | gtttcccatttgacagcatg | 103 | 1.5/55 | B07528 | This study |
| D12Rbt2 | yA8D10 left arm | tctactttcactcaacccagaca | gcagaaagacatccatgattaca | 167 | 1.5/55 | B07577 | This study |
| D12Rbt3 | yA8D10 right arm | cctaatcttttcttctccctctcc | agatgtaggaggaggaggaagg | 156 | 1.5/55 | B07576 | This study |
| D12Rbt4 | yFCDB1 right arm | agcatattcagggcatcagg | tgaattccagcattatccagc | 123 | 1.5/60 | B07536 | This study |
| D12Rbt5 | yFCLA12 right arm | attcagaaacggactcaaaagg | gctgacaggatttctaccgc | 93 | 1.5/55 | B07543 | This study |
| Mit markers | | | | | | | |
| D12Mit41 | | tgcgttaatgggtctgatagg | aattccaaaacaacagcatgc | 220 | 1.5/55 | | 33 |
| D12Mit134 | | ctatctacaacaaacttctcctggg | actcagtccaaacatatacaagatgc | 184 | 1.5/55 | | 33 |
| D12Mit150 | | cttgtcaaaatttctgttgttttaca | aaaggattttgtcactaagacatgg | 171 | 1.5/55 | | 33 |
| D12Mit208 | | tttctttctgatggaaatactttga | tcaaatgaccaaaattatagtgca | 134 | 1.5/55 | | 33 |
| D12Mit263 | | tcagatctcagcagataaatacttgg | tcccctggagcatatttgac | 113 | 1.5/55 | | 33 |
| Redesigned Mit markers[d] | | | | | | | |
| D12Mit134Rbt | | caacaaacttctcctgggttg | cccactcagtccaaacatataca | 180 | 1.5/55 | | This study |
| D12Mit150Rbt | | gctgagaatctcattgaatgtct | aacaacagaaattttgacaagaaa | 117 | 1.5/55 | | This study |
| D12Mit263Rbt | | gtcccaagttttggtctagcc | ggttccagtgcagatggaat | 114 | 1.5/55 | | This study |

[a] Indicated for each STS are the name, site, primer sequences, size of the PCR product in C57BL/6, PCR conditions (($Mg^{2+}$)/annealing temperature), GenBank accession, and literature reference. A subset of the STSs was used to screen the YAC libraries, and all were used to characterize the content of the isolated YAC clones.

[b] Magnesium concentrations and annealing temperatures are listed; cycling parameters were 1× (94°C, 30 s); 30× (94°C, 30 s; annealing temperature, 60 s; 72°C, 60 s).

[c] The following colleagues graciously provided primers: R. Lieberson, enhancer α; P. Brodeur, $V_H$15; M. Caulfield, $V_H$G8; C. Carmack, $D_H$FL16.1 and $V_H$11.

[d] These PCR markers were redesigned from MIT sequence data (http://www-genome.wi.mit.edu/) to avoid the microsatellite CA repeat (D12Mit150 and -263) for use as hybridization probes or to improve performance.

Center (Harwell, U.K.; contact Dr. P. Denny: paul@har.mrc.ac.uk. See Web server at http://www.mgc.har.mrc.ac.uk/).

The Imperial Cancer Research Fund mouse YAC library (London, U.K.) was prepared from C3H male mouse DNA (22). It contains 15,000 pYAC4 vector clones in yeast host AB1380 with an average insert size of 700 kb. It covers three haploid genomic equivalents and can be screened by high-density filter colony hybridization on membranes from the Resource Center/Primary Database (Berlin, Germany) of the German Human Genome Project. (See the information server of the Resource Center/Primary Database at http://www.rzpd.de).

### *YAC library screening*

Four YAC libraries were screened by PCR according to the protocols accompanying the DNA pools. Screening pools for the Princeton University and Whitehead Institute (WI-I) YAC libraries were obtained from Dr. S. M. Tilghman. The pools for RAD52 and the Imperial Cancer Research Fund libraries rearrayed as the "3D" library were obtained initially from Dr. S. D. M. Brown at Saint Mary's Hospital Medical School and later from Dr. E. Brundage at the Baylor College of Medicine (Houston, TX). Identified clones were obtained from these same sources.

A fifth library (Whitehead Institute (WI-820); Ref. 20) was produced as the vehicle for establishing a YAC physical map of the entire mouse genome; it was screened with many MIT simple-sequence repeat markers on all chromosomes (this data is available at http://www-genome.wi.mit.edu; Ref. 15). Three markers used in this screening, D12Mit263, D12Mit134, and D12Mit150, are localized in the *Igh* locus, and the identified clones were purchased from Research Genetics.

### *STS content analysis of YAC clones*

Clones identified in the screening were confirmed after selecting single yeast colonies. YAC DNAs were prepared as described (23). PCR primer sequences and the PCR conditions used are indicated in Table I (10, 14, 24–29). The PCR products were separated by electrophoresis on 10% acrylamide gel and visualized by ethidium bromide staining.

### *Pulsed field gel electrophoresis (PFGE) and Southern blot*

YAC plugs were prepared as described (23). YAC clone chromosomes were separated on 1% Seakem LE agarose gels (FMC, Rockland, Maine) using a contour-clamped homogeneous electric field apparatus (CHEF-DRIII; Bio-Rad, Richmond, California) and visualized by ethidium bromide staining and, when needed, blotted and probed with labeled total mouse DNA. For YAC content analysis, YAC DNAs were digested with *Eco*RI (Stratagene, La Jolla, CA), electrophoresed on Seakem LE agarose Gels (FMC), blotted onto Hybond-N$^+$ membranes (Amersham, Arlington Heights, IL), and probed according to Sambrook et al. (30). DNA probes were labeled by random priming with the Prime-it II kit (Stratagene). Probes included plasmid-cloned $V_H$, $D_H$, and $C_H$ gene segments and YAC end PCR products.

### *Isolation of YAC insert ends, sequencing, and generation of STSs and probes*

Both ends of each YAC were isolated by the vector-hexamer technique (16). This method consists of the PCR amplification of YAC insert ends using a nested series of vector primers in conjunction with a hexamer primer that anneals to a short arbitrary sequence randomly located in the mouse insert. Insert end PCR products were gel purified and sequenced with the ABI PRISM dye terminator cycle sequencing kit and ABI373 automated sequencer (PerkinElmer, Foster City, CA). For many ends, several products made with different hexamer primers were sequenced. Sequences were analyzed using the basic local alignment search tool (BLAST) e-mail and Web servers (at http://www.ncbi.nlm.nih.gov/BLAST; Ref. 31). The YAC end sequences were deposited in GenBank with accession numbers B07512 to B07602. PCR assays for STS markers were designed with PRIMER 0.5 (32). YAC insert ends were prepared for hybridization probes by reamplifying third-round gel purified ends with the innermost nested primer to reduce the content of the vector arm in the PCR product. Amplification products were electrophoresed on acrylamide gels and the correct size bands were isolated (16). Depending upon the brightness of the band, 1–10 microliters of gel eluate were labeled for use as probes.

### *Chromosome 12 synteny analysis*

When YAC insert ends did not hybridize to other YACs in the contig they were tested on DNAs from somatic cell hybrid lines to determine whether they mapped to chromosome 12. Cell lines Mae28 and Mae32 were kindly provided by Dr. P. D'Eustachio. These are Chinese hamster cell lines containing mouse chromosomes X and 12 and X and 16, respectively.

## Results

### *YAC library screening*

To assemble a YAC contig and establish a physical map of the mouse *Igh* locus with as much redundancy as possible, we screened all four available YAC libraries and isolated 36 clones. The libraries were screened by PCR for a series of STSs representing C region, D region, and various V region sites: Ch $\alpha$ exon 3, DhFL16.1, $V_H$ group III, $V_H$11, $V_H$Ese26.1, $V_H$ group II, $V_H$G8 (Table I). Certain of the $V_H$ PCR assays ($V_H$ group II, $V_H$ group III) were designed to detect multiple $V_H$ genes for efficient screening; although unintended, other assays performed similarly, e.g., $V_H$Ese26.1. Two additional screenings (with yADGC9 left end and yFCDB1 right end) were done later to close a gap in the YAC contig. More recently an additional library, prescreened for many "D_Mit_" microsatellite markers, has become available (15, 20). Two YAC clones bearing the D12Mit134 marker (y138G1 and y139H4) were added to this study.

### *STS content*

The YAC clones were analyzed with all PCR markers localized in the *Igh* locus. These include five microsatellites (D12Mit41, D12Mit134, D12Mit150, D12Mit208, D12Mit263) genetically mapped in the region (33). The sequences of the primers, the PCR conditions and the reference sequences for each marker are described in Table I. The results of this characterization by PCR are indicated in Fig. 1. Most YACs are positive for several loci, and the overlapping of these clones is obvious, although it is not possible to construct a completely consistent marker order from this

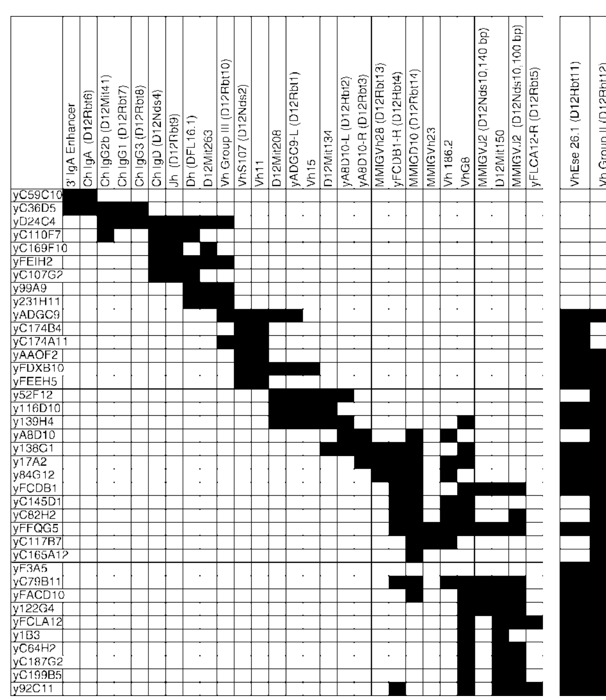


**FIGURE 1.** STS content of *Igh* YACs. The STS content of each clone was determined by PCR with all the primer sets described in Table I. STSs are listed in approximate map order, and their presence in a YAC is indicated by black boxes. Although deletions in some YACs are apparent, a continuous path linking all STS markers can be traced in these YACs. Note that the 380-kb YAC D24C4 comprises a mouse *Igh* minilocus, bears at least 4 C region genes, all $J_H$ and $D_H$ segments, and several $V_H$*7183* and $V_H$*Q52* genes (Fig. 2).

data. Some of the $V_H$ assays do not detect a single copy locus; rather, they amplify several or many sites which may be widely spread throughout the locus. In addition, some YACs have internal deletions; for example, in the C region locus (the structure of which was completely determined with phage clones; Ref. 3), y36D5 contains IgA and IgG1 constant region sequences but not the intervening D12Mit41, a microsatellite in the 3′ flank of the *IgG2b* gene.

### *YAC sizing and end clone characterization*

To measure the length of the physical map, we determined the size of each YAC by PFGE. The sizes ranged between 100 and 800 kb. Five clones contain two different artificial chromosomes (yA8D10, yC79B11, yC110F7, yC174B4, yFEEH5), but blotting the PFGE gel and probing with an *Igh* region probe identified the chromosome 12 YAC in each case.

For this study, we developed a new generalized strategy, called vector-hexamer PCR, for the isolation of insert ends from YACs (16). YAC insert ends were isolated from all 38 clones to characterize YAC overlap and to detect chimeric clones. Eighty-one ends were recovered from these 38 YACs. More than two ends were isolated from some of the clones which contained more than one YAC. End sequences were compared with the GenBank database with BLAST (31). A supplementary Table with GenBank accessions for the end sequences, BLAST results, YAC sizes, and end probe hybridization results is at http://www.tpims.org/research/riblet-genetics.html.

Repetitive sequences were identified in 45% of ends, and most of these, 40% of the total, belonged to the LINE1, or L1, family of high-copy number repetitive sequences. Of the 81 ends, 11 were similar to known *Igh* sequences, usually sequences flanking coding regions. With a few exceptions, these matches were only 70–90% identical to GenBank entries; these likely represent diverged duplications that are not yet present in GenBank. These sequence identifications were consistent with Southern blot and PCR characterization of the content of these YACs, and they served to anchor the developing YAC contig and physical map to the established genetic and deletion maps. Known genes other than *Igh* or repeat elements were identified by five YAC end sequences, but they appear to be the result of YAC chimerism or other artifact. For example, when yFDXB10 left (dopamine transporter), yF3A5 right (zinc finger) and left (potassium channel), and yC187G2 left ($\gamma$-aminobutyric acid receptor) ends were used as probes on panels of *Igh* YACs, they only hybridized to the YACs from which they were isolated, indicating chimerism. The first 129-bp portion of the yFDXB10 right end sequence is unique and hybridizes to other YACs in the contig; however, the remainder of this sequence matches the *Escherichia coli mdoGH* gene (accession number X64197).

Most of the PCR insert ends ($53/81 = 65\%$) were suitable for hybridization probes on YAC blots, including some of the more highly diverged L1 ends. Forty ends hybridized to at least one *Eco*RI fragment shared with other *Igh* YACs and were used to construct the contig. The majority of these cross-hybridized to multiple fragments on *Igh* YACs, reflecting the duplicated nature of this locus. The other 13 probes hybridized only to the YAC from which the probe was generated, indicating that they are not present in the *Igh* locus. Two of these probes were from clones containing more than one YAC. Thus 11 of 38 YACs are judged chimeric.

### *$V_H$ gene content of YACs*

The content of $V_H$ genes in each YAC was determined by Southern blot using probes for each of the 15 $V_H$ gene families (14).[4] Most of these families are relatively small, containing 1–10 genes and a like number of hybridizing *Eco*RI fragments on Southern blots. These were easily identified on Southern blots of digested YAC DNAs, and the content of small $V_H$ gene families in each YAC is shown in Fig. 2. It is evident that there is considerable variation in size and content of the YACs, and that one YAC, ADGC9, carries nearly all of these $V_H$ genes except the distally mapping $V_H$15, $V_H$3609P, and $V_H$J558 families. The $V_H$J558 family (the largest family) was more difficult to analyze. Seventy-seven *Eco*RI fragments were clearly mapped on the different

[4] Southern blots of YAC panels probed with 15 $V_H$ gene families and other probes may be seen at http://www.tpims.org/research/riblet-genetics.html.

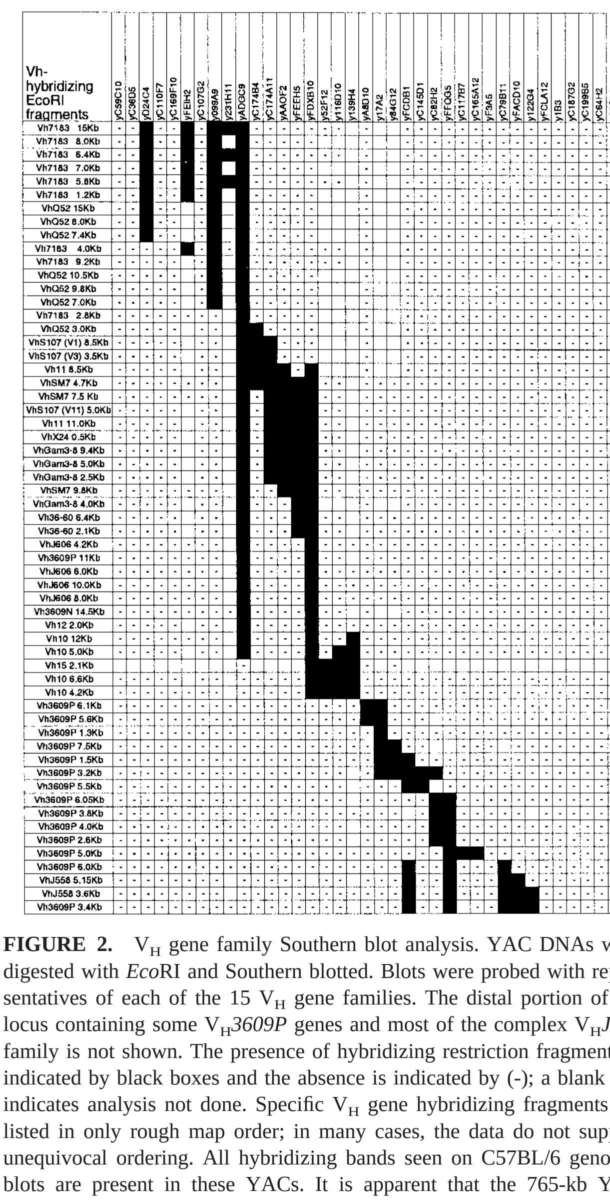


**FIGURE 2.** $V_H$ gene family Southern blot analysis. YAC DNAs were digested with *Eco*RI and Southern blotted. Blots were probed with representatives of each of the 15 $V_H$ gene families. The distal portion of the locus containing some $V_H$*3609P* genes and most of the complex $V_H$*J558* family is not shown. The presence of hybridizing restriction fragments is indicated by black boxes and the absence is indicated by (-); a blank cell indicates analysis not done. Specific $V_H$ gene hybridizing fragments are listed in only rough map order; in many cases, the data do not support unequivocal ordering. All hybridizing bands seen on C57BL/6 genomic blots are present in these YACs. It is apparent that the 765-kb YAC ADGC9 contains 40 fragments, nearly the entire proximal (3′) portion of the $V_H$ gene array including all families except $V_H$15, part of $V_H$10, and the distal $V_H$3609P and $V_H$J558 families.

YACs (Fig. 3), but this is unlikely to be a complete analysis due to problems with resolving small differences in mobility, potential deletions in some or all YACs, and judging weak cross-hybridizations on these YAC clone blots.

*The YAC contig of* Igh

The YAC clones were arranged in overlapping order using the patterns of hybridization obtained when YAC panel blots were probed with each useful insert end. This is represented in the lower

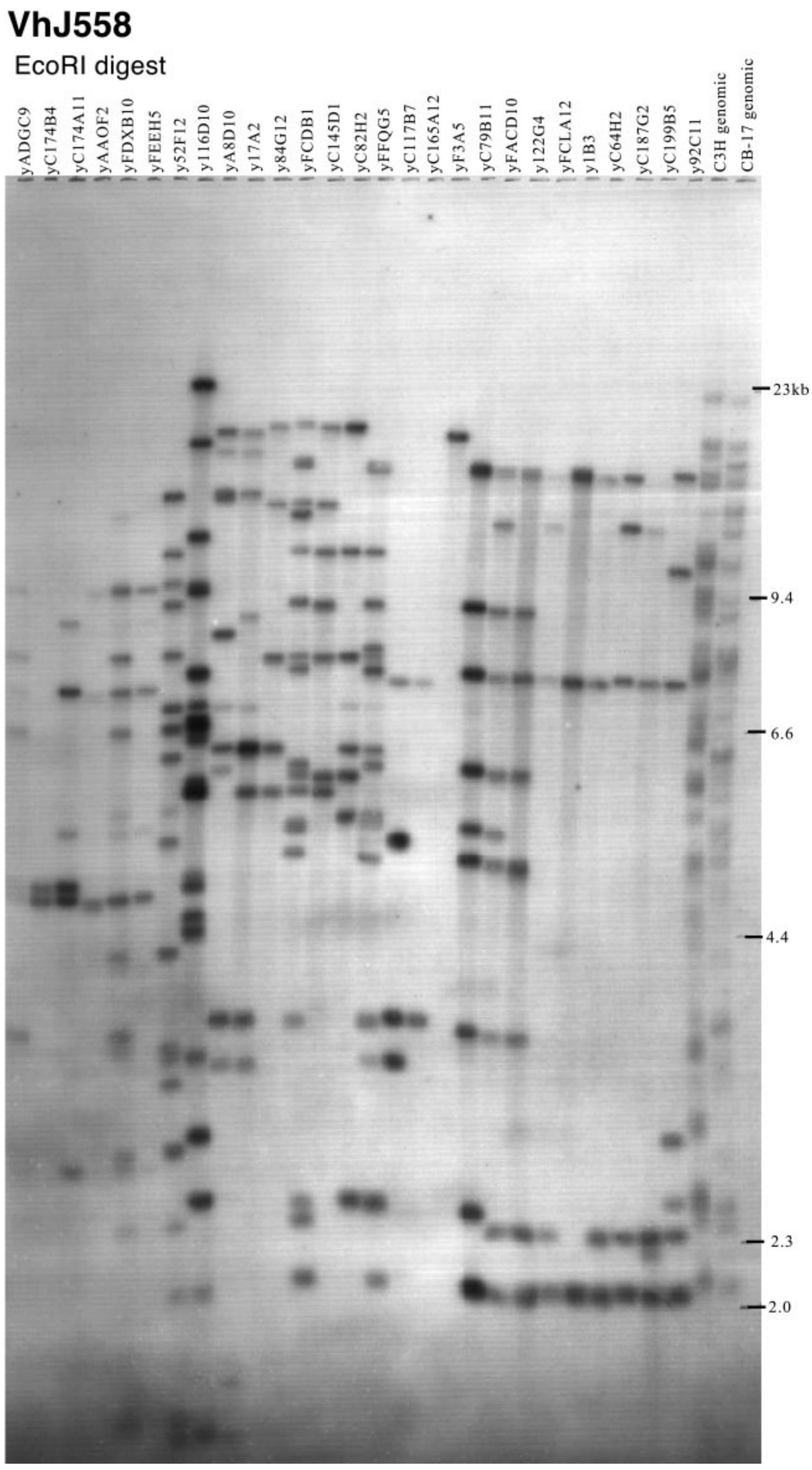


**FIGURE 3.** $V_H$*J558* gene content of YACs. YAC DNAs, C3H, and C.B-17 genomic DNAs were digested with *Eco*RI and Southern blotted. This blot was hybridized with the $V_H$J558 family probe $V_H$A1, washed under stringent conditions (membranes were washed at 65°C for 20 min with each of: solution 1, 2× SSC, SDS 0.1%; solution 2, 1× SSC, SDS 0.1%; and solution 3, 0.1× SSC, SDS 0.1%). Seventy-seven *Eco*RI bands ranging from 2 to 23 kb were detected and mapped on our YAC contig (see Fig. 4).

portion of Fig. 4. Note that the map is presented in chromosomal orientation, centromere to the left, telomere to the right, rather than the transcriptional V→D→J→C orientation usually seen (14). Note also that only three of the YACs are derived from C3H and none are crucial to the construction of the contig. The overlaps yielded a continuous path of YACs from a point midway in the constant region gene cluster at the left of the figure through all known $V_H$ gene families. All hybridizing bands visible on genomic Southern blots of C57BL/6 DNA are present in this YAC contig indicating that we have recovered the entire *Igh* locus. We estimate that the overall length of the mouse *Igh* locus is at least 2.5 million base pairs as follows: the 3′ portion of the C region gene cluster containing the 3′ enhancer and *IgA*, *IgE*, and *IgG2a* $C_H$ genes comprises ~100 kb and is proximal (3′, left in Fig. 4) to YAC yD24C4, 380 kb. yD24C4 overlaps yADGC9, 765 kb, and their combined length is ~1100 kb. At the 5′ end of the locus is yFFQG5, which is 540 kb and contains the most distal $V_H$*J558* genes, lacking only the most terminal one. Between yADGC9 and yFFQG5, but not overlapping them, is the large YAC y138G1. Its length was not determined, but it is from the Whitehead Institute (WI-820) library of large YACs, and its $V_H$ gene content and its overlaps with other YACs indicate a length of ~1 Mb. Thus, we conclude that *Igh* spans 2.5- 3 Mb.

The pattern of overlaps of YACs generated a series of sequential segments or bins along the contig ranging in size from 20 to 200 kb or more. Each PCR marker or hybridizing Southern blot band was assigned to one of these bins producing a detailed physical map of the entire locus. The bins are indicated in the upper portion of Fig. 4 where we have summarized the blot hybridization results from Fig. 2 listing the number of hybridizing bands from each $V_H$ gene family that reside in each bin. This physical dissection of the $V_H$ gene array provides a more accurate count of the number of hybridizing restriction fragments and minimum estimate of the number of $V_H$ genes (and pseudogenes). Hybridization of YAC panel Southern blots with probes from all 15 $V_H$ gene families reveals 134 bands. This is likely to be an underestimate of the number of $V_H$ genes due to multiple genes on some restriction fragments and multiple unresolved fragments appearing as single bands.

This physical map is consistent with the detailed deletion map (14); the $V_H$7183 and $V_H$Q52 families are nearest the $D_H$ segments and are thoroughly interspersed. Following the first two $V_H$ families are the $V_H$*S107 V1* and *V3* genes and then a region of complex interspersion of the remaining small $V_H$ families. The $V_H$J558 family begins in this region, as well, and extends to the distal end of the locus. The entire distal half or more of the $V_H$ gene array is occupied by the $V_H$*J558* gene family. Also in the distal region are the $V_H$*3609P* genes interspersed among the $V_H$*J558* genes. The $V_H$*J558* family is shown extending proximally (leftward) approaching the $V_H$7183/$V_H$Q52 region. The identities of these most proximal YAC bands that hybridize with a J558 probe are not known; because the effective stringency of these YAC clone blots

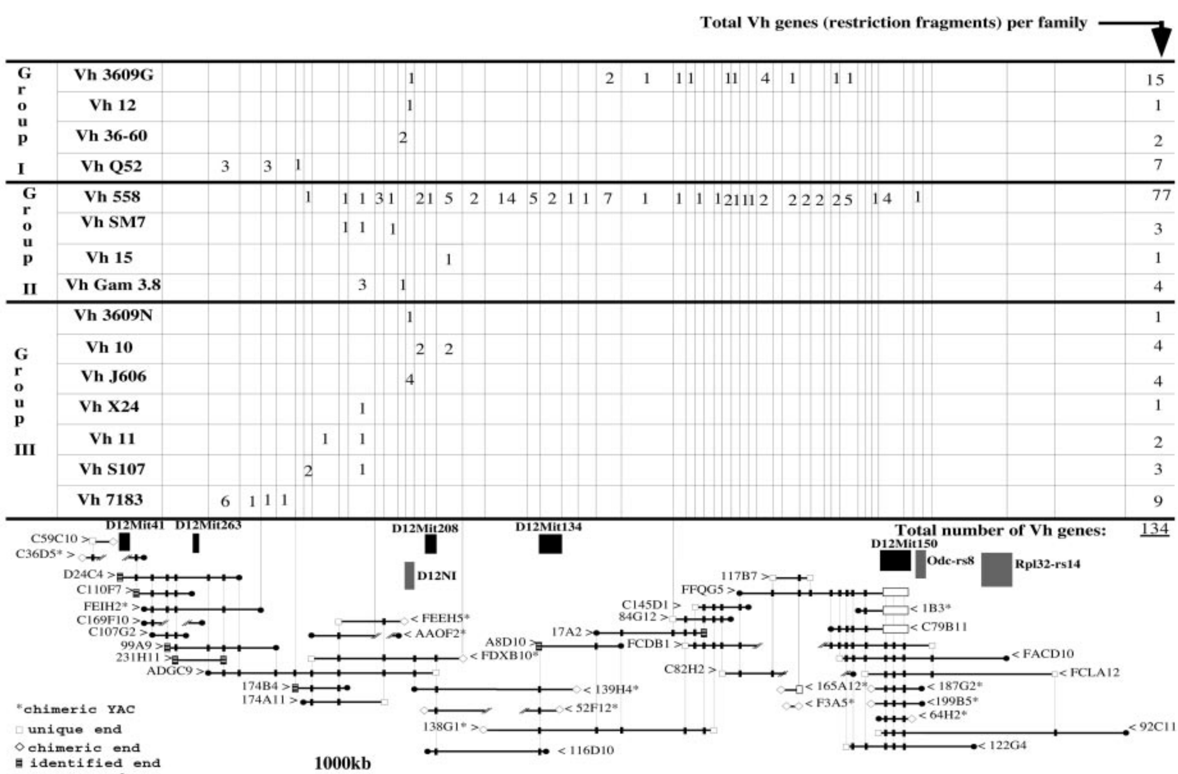


**FIGURE 4.** YAC contig of the mouse *Igh* locus. In the lower part of the figure, YACs are represented by horizontal bold lines: the YAC name is at either end with an arrow (>, <) pointing to the line. The map is shown in chromosomal orientation, left is toward the centromere; this is opposite to the transcriptional orientation, 5′-$V_H$→$D_H$→$J_H$→$C_H$. YAC insert ends are represented as unique (no GenBank match), chimeric, identified (GenBank match), or repeat sequences according to the key at the lower left. YACs identified as chimeric by end sequence or gene content are marked with an asterisk. Internal deletions are indicated by broken lines. The three C3H YACs, 116D10, 122G4, and 92C11, are at center bottom and bottom right of the contig; they agree with the C57BL YACs and suggest that the C3H haplotype may have the same structure. YAC overlaps are identified by hybridizations with nonrepetitive YAC end probes. These are indicated by connecting vertical lines and small boxes on the positive YACs. The positions of YAC ends divide the locus into a series of "bins", indicated by vertical lines that are carried into the upper part of the figure where we summarize the hybridization data from Fig. 2. For each of the 15 $V_H$ gene families, the number of hybridizing restriction fragments residing in each bin is shown. Below the $V_H$ gene families, the locations of various STS and pseudogene loci are also indicated.

is reduced they may, in fact, be members of a related family, e.g., $V_H$SM7. The more stringent genomic blots in the deletion mapping studies (14) did not suggest any $V_H$*J558* genes in this region. There are disagreements in gene placements between this map and the deletion map; for instance, we place a $V_H$*11* gene immediately distal (5′) to a bin occupied by the $V_H$S107 V1/V3 pair and a $V_H$*J558* or $V_H$*SM7* gene, while the deletion map inserts a $V_H$*X24* gene in this interval. This could reflect haplotype differences, i.e., the YAC map is of C57BL/6 and C57BL/10 (the 3 C3H YACs in our study do not contribute to the binning structure) while the deletion map is a composite of C57BL/6 and BALB/c, and in the region just discussed it is dominated by BALB/c information. Overall, this physical map confirms and provides additional detail and physical scale to the previous deletion and genetic maps of *Igh*.

Five D12Mit markers (33) that map in the *Igh* locus are clearly localized on the contig. D12Mit41 marks the *IgG2b* gene and is present on YACs that hybridize with $C_H$ region probes. D12Mit263 maps in the interval between the last $D_H$ segment, DFL16.1, and the first $V_H$ gene, $V_H$*81X*. D12Mit208 marks the region of interspersed small $V_H$ families. D12Mit134 marks the proximal (3′) part of the $V_H$J558 region, and D12Mit150 marks the end of the $V_H$ array. We localized on the contig several additional genetic markers previously mapped in or near *Igh*. D12N1 is an anonymous DNA segment identified by an aberrant H chain rearrangement (34); it is located near D12Mit208 in the bin containing all $V_H$*J606* and $V_H$*3609N* genes. Two pseudogenes, *Odc-rs8* and *Rpl32-rs14*, map at the end of the $V_H$ array just beyond D12Mit150.

## Discussion

We have screened four YAC libraries representing ~12-fold coverage of the mouse genome and have identified and analyzed 38 YACs bearing portions of the *Igh* locus. We isolated the ends of each YAC insert and used these to identify overlaps among the YACs and assemble a continuous path of YAC clones through *Igh*. Mouse *Igh* is large, 2.5–3 Mb, and contains a minimum of 134 $V_H$ gene (or pseudogene) segments. This contrasts with the recently sequenced human *IGH* locus (1) which is more compact, only 1 Mb, but contains nearly as many $V_H$ segments, independently counted as 95 (35) and 123 (1). At least half of the human $V_H$ segments are pseudogenes. The mouse Ig κ L chain locus, *Igk*, is also large, 3.5 Mb, and contains ~140 Vk segments of which as many as two-thirds are functional (2). The total complement of mouse $V_H$ genes and its proportion of expressed vs pseudogenes remain to be determined by more detailed physical characterization and sequencing. Relatively little analysis of $V_H$ genes has been done in the C57BL/6 or other $Igh^b$ mouse strains relevant to this work with the notable exceptions of the $V_H$J558 family in C.B-20 (9), the $V_H$S107 family in B10.P (36), and the $V_H$10 (37) and $V_H$7183 families (38) in C57BL/6. A thorough analysis of germline and expressed $V_H$*J558* genes identified 67 candidates considered to be expressible germline genes (9). This estimate is surprisingly close to our band count of 77 and suggests that this gene family may have relatively few pseudogenes . All four germline $V_H$*S107* genes were cloned and sequenced from the closely related B10.P strain (36); one of these, V3, was found to be a pseudogene in B10.P as it is in BALB/c. We were unable to consistently detect the V13 gene in our analysis and have omitted it. Langdon et al. (38) isolated 13–15 germline $V_H$*7183* genes by PCR from C57BL/6 liver DNA; we count only nine hybridizing fragments. If the PCR experiment is correct then either some restriction fragments carry several $V_H$*7183* genes or some hybridization bands conceal multiple similar-sized fragments.

Have we recovered the entire *Igh* locus in this YAC contig? Gaps remain in the representation of the constant region; presumably the isotype switch sequences or other repetitive elements render this portion of the *Igh* locus especially unstable in yeast. The human IGH constant region locus was also resistant to YAC cloning (39). In the $V_H$ gene array, we found another potential impediment to YAC stability, an extraordinary frequency of LINE1 elements. In our sequencing of YAC insert ends, we found that 40% belonged to the L1 family of high copy number repetitive sequences (16); this frequency is surprisingly higher than the 5% incidence observed in another YAC contig on chromosome 11 (40), but more similar to the 25% incidence in the *Igk* locus (41) and a probably higher frequency in the MHC (42). Interestingly, the human *IGH* locus also has a 40% content of L1 sequence (1). This density of repetitive sequences makes these regions potentially unstable in YACs because of the high activity of the homologous recombination machinery in yeast cells; this and the repetitive nature of the $V_H$ gene families themselves likely are responsible for the many deletions we detected in these YACs and perhaps additional deletions that we cannot detect at this level of analysis. Although we are aware of deletions in many YACs in the contig, other clones appear stable and intact, and the redundancy, or depth of coverage, that we sought in screening all available YAC libraries has resulted in apparently complete representation of all $V_H$ genes. Every hybridizing fragment seen on C57BL/6 Southern blots probed with all 15 $V_H$ families is present in the YAC contig.

Eleven of the YAC insert ends matched or were similar to known *Igh* sequences. The orientation of these 11 sequences relative to chromosome 12 was consistent with all other known elements in the H chain locus, supporting the generality that all gene segments of the *Igh* locus are in the same transcriptional orientation and that rearrangement in *Igh* occurs exclusively by deletion. The binning structure provided by the YAC overlaps has added a great deal of detail to the physical map of the $V_H$ gene array and reveals the relatively massive scale of the $V_H$*J558* gene family.

## Acknowledgments

We thank the following for generous assistance and access to YAC libraries: S. Tilghman, D. Koos, and G. Guan, Princeton University; S. D. M. Brown, F. Chartier, and G. Argyropoulos, St. Mary's Hospital Medical School; and E. Brundage, Baylor College of Medicine. We thank R. Lieberson, L. Eckhardt, P. Brodeur, M. Caulfield, and C. Carmack for primers and R. D. Miller, C. Carmack, and R. Epstein-Baak for early help and encouragement.

## References

1. Matsuda, F., K. Ishii, P. Bourvagnet, K. Kuma, H. Hayashida, T. Miyata, and T. Honjo. 1998. The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. *J. Exp. Med. 188:2151.*
2. Thiebe, R., K. F. Schable, A. Bensch, J. Brensing-Kuppers, V. Heim, T. Kirschbaum, H. Mitlohner, M. Ohnrich, S. Pourrajabi, F. Roschenthaler, et al. 1999. The variable genes and gene families of the mouse immunoglobulin κ locus. *Eur. J. Immunol. 29:2072.*
3. Shimizu, A., N. Takahashi, Y. Yaoita, and T. Honjo. 1982. Organization of the constant region gene family of the mouse immunoglobulin heavy chain. *Cell 28:499.*
4. Kurosawa, Y., and S. Tonegawa. 1982. Organization, structure, and assembly of immunoglobulin heavy chain diversity segments. *J. Exp. Med. 155:201.*
5. Wood, C., and S. Tonegawa. 1983. Diversity and joining segments of mouse immunoglobulin heavy chain genes are closely linked and in the same orientation: implications for the joining mechanism. *Proc. Natl. Acad. Sci. USA 80: 3030.*
6. Feeney, A. J., and R. Riblet. 1993. Dst4: a new, and probably the last, functional Dh gene in the BALB/c mouse. *Immunogenetics 37:217.*
7. Brodeur, P. H., and R. Riblet. 1984. The immunoglobulin heavy chain variable region (Igh-V) locus in the mouse. I. One hundred Igh-V genes comprise seven families of homologous genes. *Eur. J. Immunol. 14:922.*
8. Livant, D., C. Blatt, and L. Hood. 1986. One heavy chain variable region gene segment subfamily in the BALB/c mouse contains 500–1000 or more members. *Cell 47:461.*

9. Gu, H., D. Tarlinton, W. Muller, K. Rajewsky, and I. Forster. 1991. Most peripheral B cells in mice are ligand selected. *J. Exp. Med. 173:1357.*
10. Rothenfluth, H. S., A. J. Gibbs, R. V. Blanden, and E. J. Steele. 1994. Analysis of patterns of DNA sequence variation in flanking and coding regions of murine germ-line immunoglobulin variable genes-evolutionary implications. *Proc. Natl. Acad. Sci. USA 91:12163.*
11. Riblet, R., M. Weigert, and O. Makela. 1975. Genetics of mouse antibodies. II. Recombination between $V_H$ genes and allotype. *Eur. J. Immunol. 5:778.*
12. Weigert, M., and R. Riblet. 1978. The genetic control of antibody variable regions in the mouse. *Springer Semin. Immunopathol. 1:133.*
13. Riblet, R., A. Tutter, and P. Brodeur. 1986. Polymorphism and evolution of Igh-V gene families. *Curr. Top. Microbiol. Immunol. 127:168.*
14. Mainville, C., K. Sheehan, L. D. Klaman, C. A. Giorgetti, J. L. Press, and P. H. Brodeur. 1996. Deletional mapping of fifteen mouse $V_H$ gene families reveals a common organization for three Igh haplotypes. *J. Immunol. 156:1038.*
15. Nusbaum, C., D. K. Slonim, K. L. Harris, B. W. Birren, R. G. Steen, L. D. Stein, J. Miller, W. F. Dietrich, R. Nahf, V. Wang, et al. 1999. A YAC-based physical map of the mouse genome. *Nat. Genet. 22:388.*
16. Herring, C. D., C. Chevillard, S. L. Johnston, P. J. Wettstein, and R. Riblet. 1998. Vector-hexamer PCR isolation of all insert ends from a YAC contig of the mouse Igh locus. *Genome Res. 8:673.*
17. Burke, D. T., G. F. Carle, and M. V. Olson. 1987. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science 236:806.*
18. Rossi, J. M., D. T. Burke, J. C. Leung, D. S. Koos, H. Chen, and S. M. Tilghman. 1992. Genomic analysis using yeast artificial chromosome library with mouse DNA inserts. *Proc. Natl. Acad. Sci. USA 89:2456.*
19. Kusumi, K., J. S. Smith, J. A. Segre, D. S. Koos, and E. S. Lander. 1993. Construction of a large-insert yeast artificial chromosome library of the mouse genome. *Mamm. Genome 4:391.*
20. Haldi, M., C. Strickland, P. Lim, V. VanBerkel, X.-N. Chen, D. Noya, J. R. Korenberg, Z. Husain, J. Miller, and E. S. Lander. 1996. A comprehensive large-insert yeast artificial chromosome library for physical mapping of the mouse genome. *Mamm. Genome 7:767.*
21. Chartier, F. L., J. T. Keer, M. J. Sutcliffe, D. A. Henriques, P. Mileham, and S. D. M. Brown. 1992. Construction of a mouse yeast artificial chromosome library in a recombination-deficient strain of yeast. *Nat. Genet. 1:132.*
22. Larin, Z., A. P. Monaco, and H. Lehrach. 1991. Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl. Acad. Sci. USA 88:4123.*
23. Ausubel, F. M., R. Brent, R. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl. 1995. *Current Protocols in Molecular Biology*. John Wiley & Sons, New York.
24. Lieberson, R., S. L. Giannini, B. K. Birshtein, and L. A. Eckhardt. 1991. An enhancer at the 3′ end of the mouse immunoglobulin heavy chain locus. *Nucleic Acids Res. 19:933.*
25. Love, J. M., A. M. Knight, M. A. McAleer, and J. A. Todd. 1990. Towards construction of a high resolution map of the mouse genome using PCR-analysed microsatellites. *Nucleic Acids Res. 18:4123.*
26. Schroeder, H., Jr., and J. Y. Wang. 1990. Preferential utilization of conserved immunoglobulin heavy chain variable gene segments during human fetal life. *Proc. Natl. Acad. Sci. USA 87:6146.*
27. Carmack, C. E., S. A. Shinton, K. Hayakawa, and R. R. Hardy. 1990. Rearrangement and selection of $V_H11$ in the Ly-1 B cell lineage. *J. Exp. Med. 172:371.*
28. Hearne, C. M., M. A. McAleer, J. M. Love, T. J. Aitman, R. J. Cornall, S. Ghosh, A. M. Knight, J. B. Prins, and J. A. Todd. 1991. Additional microsatellite markers for mouse genome mapping. *Mamm. Genome 1:273.*
29. Caulfield, M. J., and D. Stanko. 1992. A pathogenic monoclonal antibody, G8, is characteristic of antierythrocyte autoantibodies from Coombs'-positive NZB mice. *J. Immunol. 148:2068.*
30. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
31. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol. 215:403.*
32. Lincoln, S. E., M. J. Daly, and E. S. Lander. 1991. *PRIMER: A Computer Program for Automatically Selecting PCR Primers.* MIT Center for Genome Research and Whitehead Institute for Biomedical Research, Cambridge.
33. Dietrich, W., J. Miller, R. Steen, M. A. Merchant, D. Damron-Boles, Z. Husain, R. Dredge, M. J. Daly, K. A. Ingalls, T. J. O'Connor, et al. 1996. A comprehensive map of the mouse genome. *Nature 380:149.*
34. Bauer, S. R., L. A. D'Hoostelaere, and K. Huppi. 1988. Restriction fragment length polymorphism near the IgH locus on mouse chromosome 12. *Nucleic Acids Res. 16:8200.*
35. Cook, G. P., I. M. Tomlinson, G. Walter, H. Riethman, N. P. Carter, L. Buluwela, G. Winter, and T. H. Rabbitts. 1994. A map of the human immunoglobulin $V_H$ locus completed by analysis of the telomeric region of chromosome 14q. *Nat. Genet. 7:162.*
36. Perlmutter, R. M., B. Berson, J. A. Griffin, and L. Hood. 1985. Diversity in the germline antibody repertoire: molecular evolution of the T15 $V_H$ gene family. *J. Exp. Med. 162:1998.*
37. Whitcomb, E. A., B. B. Haines, A. P. Parmelee, A. M. Pearlman, and P. H. Brodeur. 1999. Germline structure and differential utilization of $Igh^a$ and $Igh^b$ $V_H10$ genes. *J. Immunol. 162:1541.*
38. Langdon, S. D., M. Inaioki, G. Kelsoe, and T. F. Tedder. 2000. Germline sequences of $V_H7183$ gene family members in C57BL/6 mice demonstrate natural selection of particular sequences during recent evolution. *Immunogenetics 51: 241.*
39. Mendez, M. J., H. Abderrahim, M. Noguchi, N. E. David, M. C. Hardy, L. L. Green, H. Tsuda, S. Yoast, C. E. Maynard-Currie, D. Garza, et al. 1995. Analysis of the structural integrity of YACs comprising human immunoglobulin genes in yeast and in embryonic stem cells. *Genomics 26:294.*
40. Nehls, M., K. Lüno, M. Schorpp, D. Pfeifer, S. Krause, U. Matysiak-Scholze, H. Dierbach, and T. Boehm. 1995. YAC/P1 contigs defining the location of 56 microsatellite markers and several genes across a 3.4 cM interval on mouse chromosome 11. *Mamm. Genome 6:321.*
41. Schupp, I. W., T. Schlake, T. Kirschbaum, H. G. Zachau, and T. Boehm. 1997. A yeast artificial chromosome contig spanning the mouse immunoglobulin $\kappa$ light chain locus. *Immunogenetics 45:180.*
42. Jones, E. P., H. Xiao, R. A. Schultz, L. Flaherty, Z. Trachtulec, V. Vincek, Z. Larin, H. Lehrach, and K. F. Lindahl. 1995. MHC class I gene organization in >1.5-Mb YAC contigs from the H2-M region. *Genomics 27:40.*