Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 4681–4685, August 1982
Genetics

Evolution of mouse immunoglobulin λ genes

(λ gene DNA sequences/pseudo λ gene/heteroduplex analysis/gene duplication/replacement and silent site mutations)

ERIK SELSING, JIM MILLER, RONALD WILSON, AND URSULA STORB

Department of Microbiology and Immunology, University of Washington, Seattle, Washington 98195

Communicated by Michael Potter, April 30, 1982

ABSTRACT **The mouse has four Cλ and two Vλ genes. We have isolated Charon 4A clones that contain all six λ genes from a BALB/c germ-line library. We present here the DNA sequences of the Cλ2, Cλ3, and Cλ4 genes and also correct what are apparently errors in previous reports of Cλ1 protein and DNA sequences. In addition, we have analyzed cloned DNAs by restriction mapping and electron microscopy to determine the relationships among the various λ genes. By heteroduplex analysis, two gene clusters containing JCλ3–JCλ1 and JCλ2–JCλ4 show homology extending from the J regions 5′ of Cλ3/Cλ2 to just 3′ of Cλ1/Cλ4. Other than the region between the genes, very little homology exists in the Cλ flanking regions. In contrast, Vλ1 and Vλ2 genes show considerable homology extending into the 5′ flanking regions. Large inverted repeats are found in the 5′ flanking regions of Vλ1 and Cλ3, as well as in the 3′ flanking regions of both Cλ gene clusters. DNA sequence divergences between the Cλ genes indicate that an ancestral JCλx–JCλy gene cluster arose at about the time of the first mammals by duplication of a primordial JCλ gene. The data further suggest that the JCλx–JCλy gene cluster duplicated after the speciation of mouse and man and subsequently diverged into the present day JCλ3–JCλ1 and JCλ2–JCλ4 gene clusters. Cλ4, a pseudogene, became inactive at about the time of duplication of the ancestral JCλx–JCλy cluster. Comparison of DNA sequence divergence between the Vλ1 and Vλ2 genes demonstrates an anomaly. The percentage of amino acid replacement changes is approximately the same for Vλ1/Vλ2 as for Cλ3/Cλ2, implying that the ancestral Vλ gene was duplicated at the same time, and possibly together with, the JCλx–JCλy cluster. However, there are fewer silent changes than amino acid replacement changes between the Vλ1/Vλ2 genes, suggesting either that a selective pressure acted on the silent sites or that Vλ genes have only recently been duplicated. We also consider the possibility of a gene conversion event subsequent to a more ancient duplication.**

Although a great deal of information about the organization of immunoglobulin genes has recently become available, the mechanisms that control the expression of immunoglobulin (Ig) genes have not been elucidated. Functional Ig genes must be rearranged before transcription (1, 2), but the enzymes, their targets, and the mechanisms involved in rearrangement are not known. Because the λ gene family in the mouse is the simplest Ig gene locus known, we are cloning the entire λ locus to study the control of its expression and evolution.

We have now isolated from a BALB/c mouse germ-line DNA library more than 40 recombinant phage containing the genes for all known λ chain constant (C) and variable (V) regions. Mouse λ genes contain four C genes (3–5) and two V genes (refs. 6 and 7 and this work). Vλ1 is rearranged to and expressed together with either Cλ1 (8) or Cλ3 (9); Vλ2 is associated with Cλ2 (10). The Cλ4 gene is a nonfunctional pseudogene due to an inactive RNA splice site (3, 11). Each of the four Cλ genes is associated with its own joining region (J) gene. We and others have found alterations of the Jλ2 and Jλ3 genes compared with Jλ1 and these alterations may be correlated with the low rates of expression of the Cλ2 and Cλ3 genes (3, 11).

In this article we present the nucleotide sequences for all the Cλ genes, thus completing the characterization of the entire λ gene family at the level of primary DNA sequence. In addition, these sequences, as well as heteroduplex analysis of cloned DNA, provide a portrait of the evolution of the λ multigene family.

EXPERIMENTAL PROCEDURES

DNA Clones. A mouse germ-line recombinant DNA library was constructed from BALB/c kidney DNA as described (4). Phage DNA preparation and restriction enzyme analysis were performed as detailed (4). Regions that contained highly repetitive DNA were identified by Southern blotting and hybridization with ^{32}P-labeled whole kidney DNA as first described by Cochet *et al.* (12).

Electron Microscopy. Heteroduplexes between different cloned DNAs were prepared as described (13). To determine the stability of hybrids between different homologous regions, heteroduplexes that had been formed under conditions of average stringency (50% formamide/100 mM Tris, pH 7.0; 24°C) were spread at high stringency from 95% formamide/1.9 M urea/38 mM Tris. Relaxed double-stranded and single-stranded circles of plasmid pBR322 DNA were added to the DNA spreads as size markers.

DNA Sequence Analysis. The DNA sequences were determined by the dideoxynucleotide termination method (14) either as described (15) or in conjunction with M13 cloning vectors (16). The strategy used to determine the sequences of all the Cλ loci has been detailed in conjunction with Jλ DNA sequences (3).

RESULTS AND DISCUSSION

Organization of the Cλ Genes. Restriction enzyme analysis of clones containing Cλ genes demonstrated the close linkage between the Cλ3 and Cλ1 genes as well as between the Cλ2 and Cλ4 genes (Fig. 1). The similarity of these clusters together with the homology found between the Cλ2 and Cλ3 protein sequences suggested that the two clusters might be related by a gene duplication event. Heteroduplex analysis of cloned DNAs containing the Cλ3–Cλ1 and Cλ2–Cλ4 regions (Fig. 2) supported this idea. Indeed, a homology of 6.5 (±0.2) kilobases (kb) was found spanning the Cλ clusters, from the Jλ genes 5′ of Cλ3 and Cλ2 to approximately 1.0 kb 3′ of Cλ1 and Cλ4. No other homologies were detected 9.5 kb 5′ and 7.0 kb 3′ of the

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

Abbreviations: C and V, constant and variable regions of immunoglobulins; J, joining region of immunoglobulin genes; kb, kilobase(s); bp, base pair(s); Myr, million years; IR, inverted repeat.

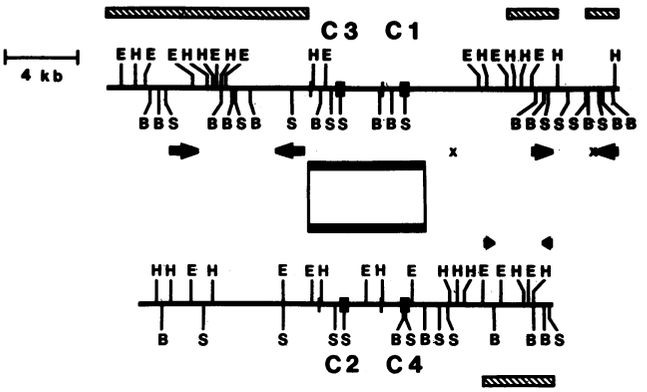

FIG. 1. Restriction enzyme maps of the two Cλ gene clusters. The four Cλ genes (black boxes) and their respective J regions (vertical lines ≈1.5 kb 5′ of each Cλ gene) are arranged on two separate stretches of cloned mouse DNA, each containing a pair of genes. Between the maps of the two Cλ clusters, the connected solid bars (homology regions) and the arrows [inverted repeat (IR) sequences] correspond to structures evident in electron microscopic analysis of heteroduplexes between cloned DNA segments (see Fig. 2). Hatched bars above and below the gene maps correspond to regions containing areas of highly repetitive DNA. Restriction enzyme abbreviations: B, *Bam*HI; E, *Eco*RI; H, *Hind*III; and S, *Sac* I.

Cλ clusters. Under the conditions used for hybridization, duplex formation required at least 80% nucleotide homology. It appears, therefore, that the two gene clusters duplicated recently. Thus, the four Cλ genes apparently arose by two sequential duplication events: first, a single primordial JCλ gene duplicated to give rise to an ancestral JCλx–JCλy cluster; then, more recently, a second gene duplication of a large segment of DNA, containing both ancestral genes, produced two separate clusters.

Electron microscopy revealed large inverted repeat (IR) sequences in both the 5′ and 3′ flanking regions of the Cλ3–Cλ1 cluster and in the 3′ flanking region of the Cλ2–Cλ4 cluster (Figs. 1 and 2). In addition, a small IR sequence (x in Fig. 1) straddles the left stem of the IR 3′ of Cλ1.

DNA Sequences of the Cλ Genes. We determined the sequence of the DNA of the Cλ genes (Figs. 3 and 4) in order to confirm assignments made previously on the basis of restriction mapping and Southern filter hybridization (4) and to compare the genes with respect to their evolution. The Cλ1 sequence (Fig. 3) showed several differences compared to published λ1 amino acid (17) and DNA (6) sequences but was in complete agreement with a recently reported λ1 cDNA sequence (18). The differences between the original amino acid sequence and our nucleotide sequence are indicated in Fig. 3. In most cases, the Cλ1 and the Cλ4 DNA sequences are identical at these positions.

The Cλ2 and Cλ3 nucleotide sequences (Fig. 4) correspond to the published protein sequences (10, 19) at all positions, except for one inversion in Cλ2. Residues 145 and 146 are serine and glycine instead of glycine and serine—i.e., the sequences here are identical in Cλ2 and Cλ3.

We had previously determined that Cλ4 represents a pseudogene due to a nonfunctional RNA donor splice site 3′ of Jλ4 (3). In addition to this defect in the J segment, the Cλ4 sequence itself exhibits many amino acid replacements relative to the Cλ1 sequence and also a deletion of 6 base pairs (bp) which eliminates amino acids 174 and 175 (Fig. 3). The deletion indicates an instability in this segment of the gene sequences because both Cλ2 and Cλ3 show 3-bp deletions in the same segment compared with Cλ1. Of the alterations in the Cλ4 gene that

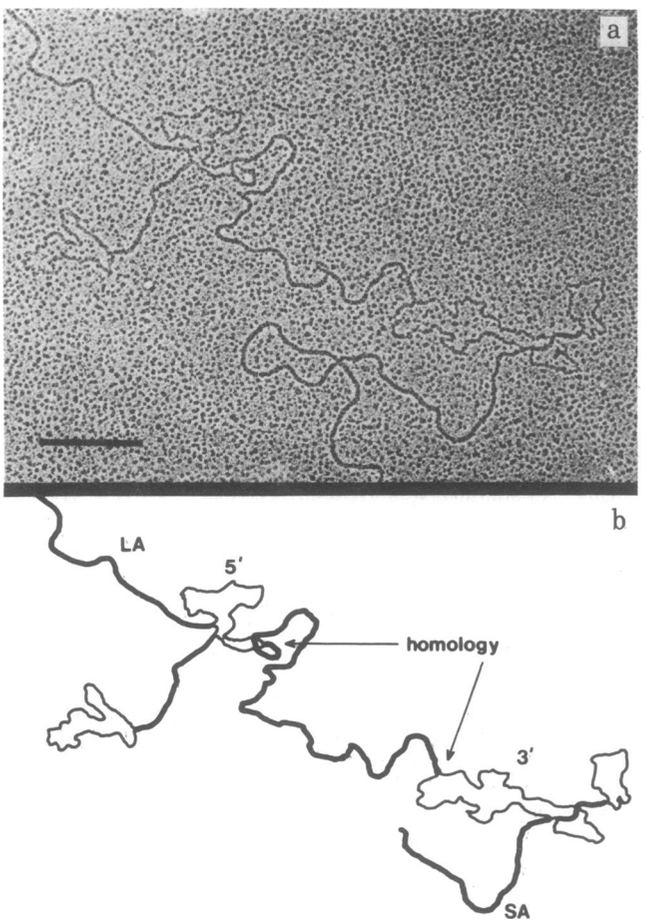

FIG. 2. Heteroduplex between the DNAs of a Cλ3–Cλ1 and a Cλ2–Cλ4 clone. LA and SA, long and short arms of Charon 4A phage; 5′ and 3′, regions where the flanking sequences of the two gene clusters are not homologous. IRs are seen 5′ of Cλ3 and 3′ of Cλ4. The IR 3′ of Cλ1 (Fig. 1) is beyond the 3′ end of the clone used for this heteroduplex. (Bar = 1 kb.)

could result in potential defects, it is not possible to assign the initial event leading to inactivation of the gene. However, in analogy to the splice site defect found in Jκ3 (1, 2), the primary event in λ4 may have been a single-base mutation in the RNA splicing recognition sequence. Subsequent evolutionary drift would then have resulted in the accumulation of additional defects.

Comparison of the nucleotide sequences of the Cλ genes (Figs. 3 and 4) supports the close evolutionary relationship found by heteroduplex analysis. We have determined the divergence between Cλ genes with respect to nucleotide changes which alter the amino acid sequence (replacement sites) and those which do not (silent sites) (20). The percentage of replacement site divergence was converted into million years (Myr) based on the evolutionary clock for globin genes (21). We have arbitrarily chosen the globin clock for lack of an accurate immunoglobulin clock, although there is evidence that immunoglobulin genes have evolved faster than globin genes (22). This was done simply to give some time perspective; the actual percentage of replacement and silent site divergences are unaffected by the choice of a clock. It thus was estimated that the duplication of the ancestral gene cluster occurred about 20 Myr ago [i.e., after speciation of mouse and man (23)] whereas the original gene duplication, which produced the gene cluster, occurred about 240 Myr ago [i.e., at about the time of the first mammals (23)] (Table 1).

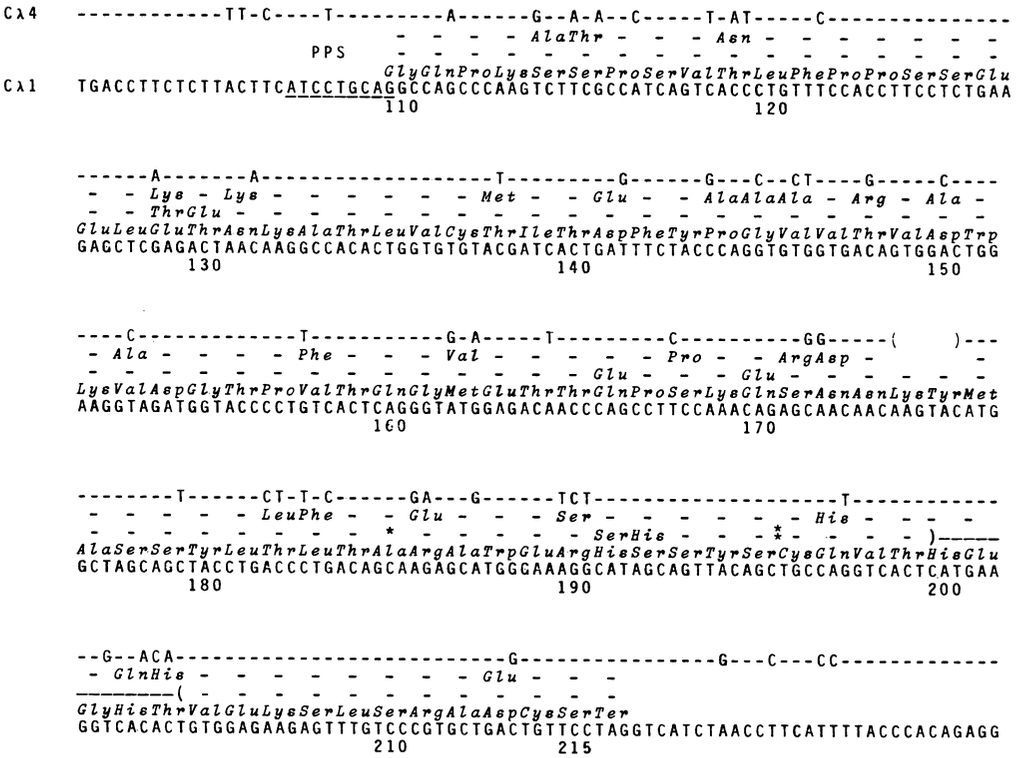

FIG. 3. Nucleotide and amino acid sequences for the entire Cλ1 gene and for Cλ4 at positions that differ from the Cλ1 sequences. Recognition sequences for RNA splicing are indicated by an underline. Differences between the nucleotide sequence for Cλ1 and the previously published Cλ1 protein sequence (PPS) (17) are also shown. Relative to the Cλ1 nucleotide sequence, the published protein sequence did not include the alanine at position 185 (*) and included an additional serine between residues 195 and 196 (‡). The region of the Cλ1 protein that had not been analyzed previously is indicated by —. A deletion of six nucleotides in the Cλ4 DNA sequence relative to Cλ1 (see text) is shown by parentheses.

Comparison of those Cλ sequences that were separated since the primordial Cλ gene duplication (Table 1) gives an average of 24.4% replacement site divergence and 67.1% silent site divergence. This differential change indicates that functional restraints have acted upon the replacement sites, as previously indicated for other gene systems (20, 21). A differential of replacement versus silent changes also occurred for the Cλ2 and Cλ3 genes which emerged after the recent gene cluster dupli-

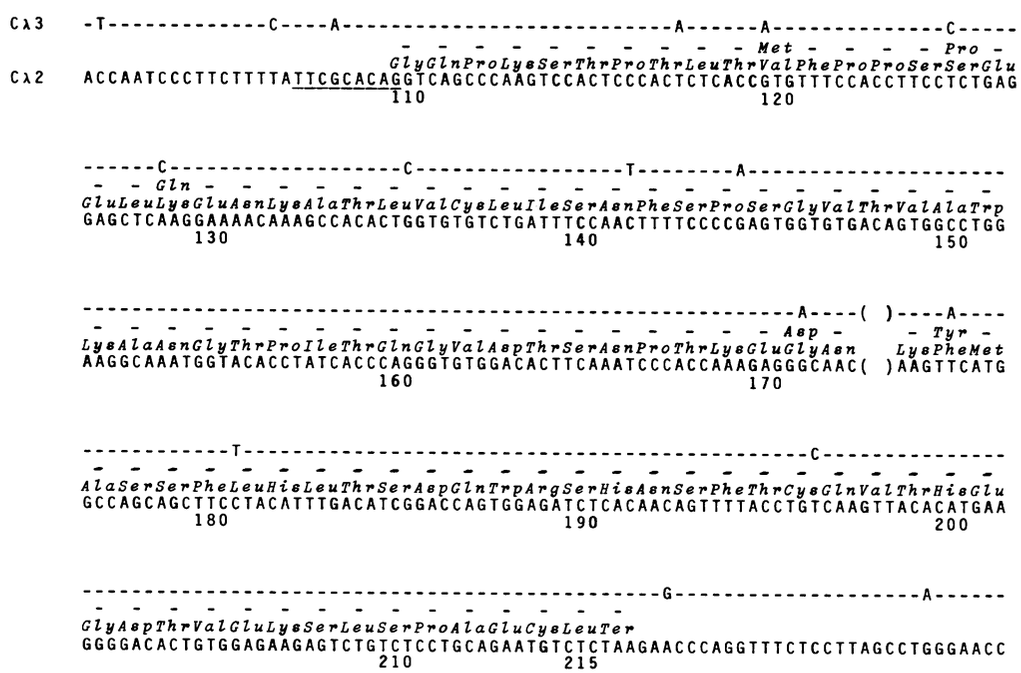

FIG. 4. Nucleotide and amino acid sequences for the entire Cλ2 gene and for Cλ3 at positions that differ from the Cλ2 sequences. Recognition sequences for RNA splicing are indicated by an underline. Deletions of three nucleotides in the Cλ2 and Cλ3 DNA sequences relative to Cλ1 (see text) are indicated by parentheses.

Table 1. Evolutionary relationship of the λ genes

Gene pair comparisons	Site divergence, %* Replacement	Silent	Time, Myr†
Primary duplication			
Cλ1/Cλ2	24.0	68.9	
Cλ1/Cλ3	24.4	64.5	240
Cλ2/Cλ4	23.6	68.0	
Cλ3/Cλ4	25.4	67.0	
Secondary duplication			
Cλ2/Cλ3	2.2	12.0	20
Cλ1/Cλ4	15.5‡	17.3	
Vλ1/Vλ2	4.5§	2.3	

* The percentage corrected divergence of the gene sequences was calculated as described (20).

† The time, in Myr, is based on the evolutionary clock for globin genes (21) (see text). The average time estimate is given for each of two duplication events, which produced the four Cλ genes. The time estimates of the primary duplication are based solely on the replacement site divergence because estimates based on silent site divergence are inaccurate for genes that have diverged >85 Myr ago (see refs. 20 and 21 for details).

‡ This high value indicates that Cλ4, pseudogene, has been evolving without selective pressure (see text for discussion).

§ This analysis was based on published sequences for Vλ1 (6, 24) and Vλ2 (7). The apparent anomalies of the Vλ gene divergences are discussed in the text.

cation. However, both the replacement and silent changes between the Cλ1 and Cλ4 genes are of about the same order of magnitude as the silent changes between Cλ2 and Cλ3. This suggests that the Cλ4 gene has been evolving without selection since the Cλ1/Cλ4 gene duplication event (25). Thus, it seems that the Cλ4 gene became nonfunctional close to the time of the actual duplication.

The majority of the intron and flanking DNA sequences (not shown) that are directly comparable between the two gene clusters have diverged 17.5%, when homologies are maximized and the divergence is corrected for multiple mutations (20). This is approximately the same percentage as the silent changes within the genes themselves, suggesting that there may be few constraints on the majority of these intron sequences. However, the 200-nucleotide segment immediately 5′ of the Cλ2 and Cλ3 genes has diverged only 8.3% (not shown). Within this portion of the intron, therefore, there may be some functional sequences, possibly related to RNA splicing, which are not directly obvious. Nevertheless, the 5′ portion of this intron near the J segment exhibits the same high divergence as the remainder of the intron and the J segment itself shows 10% divergence.

In order to determine whether other portions of the gene clusters that we had not subjected to sequence analyses show a high degree of conservation, heteroduplexes between the two Cλ gene clusters were subjected to denaturing conditions to melt less-matched duplex regions. With heteroduplexes of closely related sequences, such as the homology region between the Cλ gene clusters, A+T-rich regions and partially mismatched regions cannot be distinguished by this method. As controls, therefore, homoduplexes of cloned DNAs of the Cλ3–Cλ1 and the Cλ2–Cλ4 gene clusters were melted in the same manner. It was found that, when homoduplexes were melted, 35% (Cλ3–Cλ1) and 46% (Cλ2–Cλ4) of the 6.5-kb intercluster homology region remained duplexed (not shown). In heteroduplexes, however, only 360 bp (5.6%) of the cluster remained double stranded (not shown). This region corresponds to the coding sequences for Cλ3/Cλ2 which exhibit only 3% divergence (see Fig. 3). Thus, these coding sequences are the only sequences of average or high G+C content within the clusters that have been highly conserved.

Overall Structure of the Vλ Genes. We have isolated four recombinant phage containing the Vλ1 gene and 12 clones of the Vλ2 gene from a BALB/c mouse germ-line DNA library. All Vλ clones exhibit restriction maps consistent with either the Vλ1 or Vλ2 loci (Fig. 5). This strongly indicates that no other Vλ genes are found in the BALB/c mouse genome. Thus, in keeping with the conclusions derived from extensive studies of λ light chains (8), much of the diversity found among λ chains apparently arises by a somatic mutation process (15, 26–28).

Cloned DNAs of Vλ1 and Vλ2 genes were heteroduplexed (not shown). In support of gene duplication from an ancestral Vλ gene, several regions of homology are seen (Fig. 5). The major 4.5-kb homology spans the V genes, extending only 0.6 kb 3′ of the V genes. Three additional short homologies are observed 5′ of the Vλ genes. These are separated by nonhomologous regions having variable lengths within each locus and also between the two loci. The most striking difference between the loci is the presence of a large IR in the 5′ flanking region of Vλ1. In this region, Vλ2 shows only a short sequence of about 500 bp interspersed between two regions of Vλ1/Vλ2 homology.

Flanking Regions. What is the origin of the sequences that are not homologous in the flanking regions of Vλ1/Vλ2 and Cλ3–Cλ1/Cλ2–Cλ4? Under the conditions used for heteroduplex formation, these regions must differ by more than 20%. Comparing the portions of the Vλ and Cλ regions of known sequence, we find that random genetic drift would have produced less than 20% divergence between homologous sequences (Table 1). It therefore is likely that the highly divergent flanking sequences of Vλ and Cλ genes did not arise by random mutation. Instead, they may reflect deletions or insertions or, alternatively, selection for increased mutation rates. Although there is some precedent for the last in the case of structural genes, in which some coding sequences seem to undergo more rapid evolution (20), there is no such case known to us involving noncoding regions. It is notable that most of the nonhomologous flanking regions are composed of highly repetitive DNA in at least one of the two partners (Figs. 1 and 5). Furthermore, in many cases IRs are involved. It is possible that repetitive DNA in general, and IRs in particular, comprise DNA sequences that are mobile in the genome. These may be related to transposable

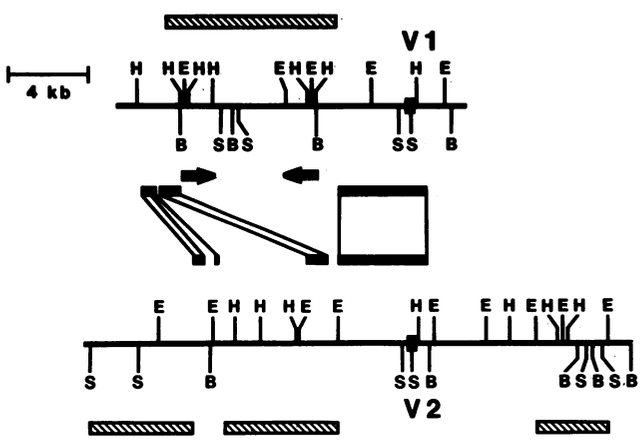

FIG. 5. Restriction enzyme maps of the two Vλ genes. The two Vλ genes (black boxes) are arranged on two separate stretches of cloned mouse DNA. Between the maps of the two Vλ genes, the four connected solid bars (homology regions) and the arrows (IR sequences) correspond to structures evident in electron microscopic analysis of heteroduplexes between cloned DNA segments. The hatched bars above and below the gene maps correspond to regions that contain areas of highly repetitive DNA. Restriction enzyme abbreviations are in Fig. 1.

elements (29). It remains to be determined if the different compositions of flanking sequences surrounding the λ genes have any functional significance.

Were the Vλ and Cλ Genes Duplicated Together? Vλ1 and Vλ2 DNA sequences (6, 7, 24) were compared to calculate their divergence (Table 1). Sites that change an amino acid (replacement sites) were found to differ by 4.5%; silent sites differed by 2.3%. The 4.5% replacement site divergence is not significantly different from the 2.2% replacement site divergence between Cλ2 and Cλ3 (Table 1). However, silent sites in Cλ2/Cλ3 diverged by 12.0% whereas silent sites of the Vλ genes diverged even less than the replacement sites.

The low value of the silent site/replacement site divergence ratio for Vλ1/Vλ2 presents a difficulty in determining the time of duplication for these two genes. Although several possibilities can be proposed, the crucial question is whether the time of Vλ duplication is accurately represented by the replacement site divergence or by the silent site divergence. The low percentage of silent site changes suggests a separate, more recent, event than the duplication of the Cλx–Cλy cluster. This would imply that amino acid replacements in the Vλ genes are not constrained and possibly are selected for (see Table 1).

On the other hand, the frequency of replacement site changes suggests that the Vλ genes may have been duplicated together with the Cλx–Cλy cluster—i.e.,

$$\text{Vλ–JCλx–JCλy} \begin{cases} \rightarrow \text{Vλ1–JCλ3–JCλ1} \\ \rightarrow \text{Vλ2–JCλ2–JCλ4} \end{cases}$$

This scenario would require that some functional constraints must act on the nucleotide sequences of Vλ genes in order to explain the low silent site divergence observed. Alternatively, a recent gene conversion event (30, 31) may have acted upon the Vλ1 and Vλ2 genes. This gene conversion would have eliminated any divergence between the Vλ genes that had accumulated. Subsequent to the gene conversion, the Vλ genes would have resumed their divergence. This model must also assume little selective pressure against amino acid changes in the Vλ genes to account for the relatively low ratio of silent site/replacement site divergence observed. It may be that, during evolution, V genes are allowed to drift and only become fixed when a sequence essential for binding to a particular antigen has evolved (32). An increase in the ratio of silent to replacement substitutions would only be expected after the time of fixation. The duplication of an original Vλ–Jλx–Jλy cluster into two separate expression units, as outlined above, possibly could explain the preferential association of Vλ1 with Cλ1 and Cλ3 (8, 9) and of Vλ2 with Cλ2 (10) on the basis of physical organization.

The heteroduplex and sequence analyses presented here indicate that several gene duplications have given rise to the λ gene family in the mouse. Our data all are consistent with a two-step process of duplication giving rise to the four present-day Cλ genes. Due to the anomalous divergence between Vλ1/Vλ2, it is not possible at this stage to define the relationships between the Vλ and Cλ gene duplications unambiguously. When we have completed the linkage between all λ genes, the evolution of the Vλ genes and the constraints acting on the rearrangements of λ genes may be clarified.

We are grateful to M. McMullen and J. Voss for some of the clones, to C. Laird for critical reading of the manuscript, and to K. Spangler for typing the manuscript. This work was supported by Grants CA/AI-25754 and DE-02600 from the National Institutes of Health and Grant PCM78-13205 from the National Science Foundation. J.M. is supported by National Institutes of Health Training Grant GM07270.

1. Max, E., Seidman, J. & Leder, P. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3450–3454.
2. Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. (1979) *Nature (London)* **280**, 288–294.
3. Miller, J., Selsing, E. & Storb, U. (1982) *Nature (London)* **295**, 428–430.
4. Miller, J., Bothwell, A. & Storb, U. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3829–3833.
5. Blomberg, B., Traunecker, A., Eisen, H. & Tonegawa, S. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3765–3769.
6. Bernard, O., Hozumi, N. & Tonegawa, S. (1978) *Cell* **15**, 1133–1144.
7. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. & Gilbert W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1485–1489.
8. Weigert, M. & Riblet, R. (1977) *Cold Spring Harbor Symp. Quant. Biol.* **41**, 837–846.
9. Breyer, R., Sauer, R. & Eisen, H. (1981) *ICN–UCLA Symp. Mol. Cell. Biol.* **20**, 105–110.
10. Dugan, E. S., Bradshaw, R. A., Simms, E. S. & Eisen, H. N. (1973) *Biochemistry* **12**, 5400–5416.
11. Blomberg, B. & Tonegawa, S. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 530–533.
12. Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F. & Chambon, P. (1979) *Nature (London)* **282**, 567–574.
13. Walfield, A., Storb, U., Selsing, E. & Zentgraf, H. (1980) *Nucleic Acids Res.* **8**, 4689–4707.
14. Sanger, F., Nicklen, S. & Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
15. Selsing, E. & Storb, U. (1981) *Cell* **25**, 47–58.
16. Messing, J., Crea, R. & Seeburg, P. (1981) *Nucleic Acids Res.* **9**, 309–321.
17. Appella, E. (1971) *Proc. Natl. Acad. Sci. USA* **68**, 590–594.
18. Bothwell, A., Paskind, M., Schwartz, R., Sonenshein, G., Gefter, M. & Baltimore, D. (1981) *Nature (London)* **290**, 65–67.
19. Azuma, T., Steiner, L. & Eisen, H. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 569–573.
20. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. & Dodgson, J. (1980) *Cell* **20**, 555–566.
21. Efstratiadis, A., Posakony, J., Maniatis, T., Lawn, R., O'Connell, C., Spritz, R., DeRiel, J., Forget, B., Weissman, S., Slightom, J., Blechl, A., Smithies, O., Baralle, F., Shoulders, C. & Proudfoot, N. (1980) *Cell* **21**, 653–668.
22. Fudenberg, H., Pink, J., Wang, A.-C. & Douglas, S. (1978) *Basic Immunogenetics* (Oxford Univ. Press, New York).
23. Grant, J. A., Sanders, B. & Hood, L. (1971) *Biochemistry* **10**, 3123–3132.
24. Arp, B., McMullen, M. D. & Storb, U. (1982) *Nature (London)*, in press.
25. Lacy, E. & Maniatis, T. (1980) *Cell* **21**, 545–553.
26. Pech, M., Höchtl, J., Schnell, H. & Zachau, H. G. (1981) *Nature (London)* **291**, 668–670.
27. Bothwell, A. L. M., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K. & Baltimore, D. (1981) *Cell* **24**, 625–637.
28. Crews, S., Griffin, J., Huang, H., Calame, K. & Hood, L. (1981) *Cell* **25**, 59–66.
29. Calos, M. P. & Miller, J. H. (1980) *Cell* **20**, 579–595.
30. Roman, H. L. (1963) in *Methodology in Basic Genetics*, ed. Burdette, W. J. (Holden-Day, San Francisco), pp. 209–221.
31. Baltimore, D. (1981) *Cell* **24**, 592–594.
32. Potter, M., Rudikoff, S., Vrana, M., Rao, D. N. & Mushinski, E. B. (1976) *Cold Spring Harbor Symp. Quant. Biol.* **41**, 661–666.