Copyright © 1997 by the Genetics Society of America

A Method of Estimating the Numbers of Human and Mouse Immunoglobulin V-Genes

George Johnson and Tai T. Wu

Departments of Biochemistry, Molecular Biology and Cell Biology and of Biomedical Engineering, Northwestern University, Evanston, Illinois 60208

Manuscript received July 15, 1996
Accepted for publication November 22, 1996

ABSTRACT

Mutations in immunoglobulin V-genes can be due to gene multiplication, allelic variations, mutations induced by antigens or somatic mutations, etc., and various combinations of these. Since the number of different mouse lambda light V-gene nucleotide sequences is relatively small, a pairwise comparison between these sequences can provide a rough idea as to the contributions of the above mechanisms to the number of nucleotide differences between sequences. A plot of occurrences against the number of differences suggests that differences between one to five can be attributed to somatic mutations. Six to 12 differences can be allelic. Thirteen to 17 may be due to allelic variations together with somatic mutations. Differences >17 appear to be derived from gene multiplication. Although these numbers are most likely somewhat different in humans, they can nevertheless provide a rough guide to sort out the effect of gene multiplication. Estimations of human heavy, kappa and lambda light chain immunoglobulin V-genes are in reasonably good agreement with recent experimental studies. For mouse kappa light and heavy chains, our estimations can provide some insight to future analyses by direct sequencing of these gene segments.

SUBGROUPS were initially used to estimate the number of genes of the variable region of human kappa light chains (MILSTEIN 1967). Various subgroups have subsequently been defined for mouse kappa (POTTER 1967), human heavy (KABAT *et al.* 1976), mouse heavy (DILDROP 1984), and human lambda (KABAT *et al.* 1979) chains, using different subgroup definitions (see, for example, KABAT *et al.* 1991). Unlike the above systems, where subgrouping is used as an estimation of gene number, the mouse immunoglobulin lambda light chain V-genes is relatively small, and the gene number has reportedly been identified (SANCHER *et al.* 1987).

Several basic mechanisms are believed to be responsible for the differences between immunoglobulin V-gene sequences. Antigen stimulation, for example, causes somatic mutations in the variable region, but the number of such mutations is probably relatively small (CHEN *et al.* 1992). Divergence of strains within a species, such as mice, presumably would result in larger numbers of nucleotide differences between homologous genes (SERRANO *et al.* 1994). Finally, gene multiplication, which occurred a long time ago, would generate even larger numbers of nucleotide differences (ANDERSSON *et al.* 1991). Since these processes are probably occurring concurrently, a complicated pattern of mutations has been observed.

Corresponding author: Tai Te Wu, Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, 2145 Sheridan Rd., Evanston, IL 60208. E-mail: tt@immuno.bme.nwu.edu

To decipher the contributions from each of these mechanisms from the resulting mutation patterns in human and mouse immunoglobulin V-genes, pairwise comparisons of known complete nucleotide sequences of all human and mouse V-genes were generated and analyzed. The mouse lambda light chains provide a simple pattern of mutational differences that can be correlated with the three above-mentioned genetic processes. The resulting correlation can then be used to estimate the number of different genes for human and other mouse chains. While this approach is somewhat different from the conventional method of amino acid sequence comparison, our result, in principle, should provide a reasonable estimation of the number of different immunoglobulin V-genes.

MATERIALS AND METHODS

We have been collecting amino acid and nucleotide sequences of immunoglobulins and related proteins (KABAT *et al.* 1991). The complete nucleotide sequences of mouse and human lambda and kappa light and heavy chain variable regions coded by V-genes, *i.e.*, codons 1–95 for light chains and codons 1–94 for heavy chains, were used. Generally, cDNA sequences differ from genomic sequences in these regions by only a few nucleotides. The detailed references of these sequences are listed in the Kabat Database of Sequences of Proteins of Immunological Interest (JOHNSON *et al.* 1996). Each sequence is designated by a Kabat database identification number (KADBID), and can be retrieved by sending an e-mail to:

seqhunt2@immuno.bme.nwu.edu

with the following message:

TABLE 1

Nucleotide differences from pairwise comparison of distinct complete mouse lambda light chain V-genes

	1 007 665	2 007 671	3 007 674	4 007 675	5 007 683	6 007 684	7 007 685	8 007 687	9 007 688	10 007 689	11 007 690	12 007 691	13 007 692	14 007 695	15 007 697	16 007 700	17 007 703
1. 007665	0	1	2	1	2	5	7	2	2	2	3	8	4	7	9	8	8
2. 007671		0	3	2	3	6	8	3	3	3	3	9	5	8	10	9	8
3. 007674			0	3	4	7	9	4	4	4	5	10	6	9	11	10	10
4. 007675				0	3	6	8	3	3	3	4	9	5	8	10	9	9
5. 007683					0	7	9	4	4	4	5	10	6	7	9	8	10
6. 007684						0	12	7	7	7	8	13	9	12	14	13	13
7. 007685							0	8	9	9	8	1	11	14	14	15	15
8. 007687								0	4	4	4	9	6	9	11	10	10
9. 007688									0	4	5	10	4	9	11	10	10
10. 007689										0	5	10	6	9	11	10	10
11. 007690											0	9	7	10	12	11	10
12. 007691												0	12	15	15	16	16
13. 007692													0	11	13	12	10
14. 007695														0	2	1	15
15. 007697															0	3	17
16. 007700																0	16
17. 007703																	0
18. 007704																	
19. 007707																	
20. 007710																	
21. 007711																	
22. 007714																	
23. 007715																	
24. 007716																	
25. 007717																	
26. 007718																	
27. 007719																	
28. 007720																	
29. 007721																	
30. 007725																	
31. 007728																	
32. 007746																	
33. 007748																	
34. 007750																	
35. 019435																	
36. 020130																	

BEGIN

007665

for example. This message would retrieve entry 007665 from the database. Alternatively, the entries can be looked up through the database's WWW server at URL:

http://immuno.bme.nwu.edu

The Kabat database contains essentially all published nucleotide sequences of proteins of immunological interest currently available. To our knowledge, this is the largest and most complete aligned sequence database of these proteins.

A standard triangular table listing the number of nucleotide differences was generated by pairwise comparison of all distinct mouse lambda light chain variable region sequences. A second triangular table was obtained by removing any column or row containing differences of one to 17 from the first table. For other sequences, a similar second table was also created, together with a third table where columns and rows with differences of 18–34 were also removed.

The initial cutoff of 17 nucleotide differences was obtained from plotting the occurrences against the number of nucleotide differences from the first triangular table for mouse lambda light chains. For >17 differences, the mouse sequences being compared are most likely coded by two different genes.

RESULTS

The first triangular table for mouse lambda light chain V-genes (Table 1) contains 36 sequences each of which is identified by the KADBID as discussed in MATERIALS AND METHODS. Since many of these sequences are very similar, most the differences listed in the table are small. However, it is difficult to separate the relative contributions from various genetic mechanisms generating such differences, *e.g.*, gene multiplication, allelic variation, and somatic mutations induced

18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
007	007	007	007	007	007	007	007	007	007	007	007	007	007	007	007	007	019	020
704	707	710	711	714	715	716	717	718	719	720	721	725	728	746	748	750	435	130
6	8	83	10	8	14	12	13	11	9	11	15	20	14	139	142	149	27	1
7	8	83	11	9	15	13	14	11	10	12	16	21	15	140	143	148	28	1
8	10	84	12	9	16	14	14	13	11	13	17	22	16	138	141	150	29	3
7	9	84	11	9	15	12	14	12	10	12	16	21	15	140	143	149	26	2
8	10	83	10	10	14	14	15	13	11	11	15	22	14	140	143	149	27	3
10	13	87	15	13	19	16	18	14	13	16	20	23	19	141	144	150	30	6
13	15	87	17	15	18	19	18	17	14	18	19	26	20	144	147	153	34	8
8	10	85	12	10	16	14	15	13	10	13	17	22	16	140	143	148	29	3
8	10	82	12	10	16	14	15	13	11	13	17	22	16	141	144	151	27	3
8	10	85	12	10	16	13	15	13	11	13	17	22	16	140	143	150	29	3
9	10	85	13	11	17	15	16	13	10	14	18	23	17	141	144	149	30	3
14	16	88	18	16	19	20	19	18	15	19	20	27	21	145	148	154	35	9
10	10	85	14	11	18	16	17	15	13	15	19	24	18	140	143	152	30	5
13	15	80	5	15	11	18	20	18	16	6	14	26	19	141	143	151	23	8
15	17	80	7	17	12	20	19	19	18	8	15	28	21	139	141	151	25	10
14	16	81	6	16	12	19	21	19	17	7	15	27	20	142	144	150	24	9
14	3	86	18	14	22	18	19	17	17	19	23	27	20	140	145	151	35	7
0	14	87	15	14	20	18	19	15	13	16	21	26	18	136	142	151	32	7
	0	87	18	15	22	19	19	17	17	19	23	27	20	142	145	149	35	7
		0	76	89	86	87	88	89	92	77	88	95	90	152	152	162	90	82
			0	18	14	21	23	21	19	1	17	29	22	139	142	151	26	11
				0	21	18	14	19	15	19	22	25	22	139	142	149	29	9
					0	24	25	25	23	15	4	30	24	144	145	154	30	15
						0	25	21	19	22	25	26	25	144	146	156	33	13
							0	23	20	24	26	27	23	137	139	146	36	14
								0	20	22	26	30	21	140	144	150	37	10
									0	20	24	27	23	139	142	151	33	10
										0	18	30	23	140	143	152	27	12
											0	31	25	245	146	155	33	16
												0	32	141	145	156	43	21
													0	143	146	152	37	15
														0	11	146	152	140
															0	145	150	143
																0	155	149
																	0	28
																		0

by antigens, etc. On the other hand, the origins of these sequences are well documented. Using this information, the contributions of genetic mechanisms to the observed sequence differences might be deciphered. Figure 1 illustrates a plot of occurrences against the number of differences between mouse lambda light chain sequences. Figure 1a shows all differences while Figure 1b only differences up to 40.

The peaks around 143 and 150 differences in Figure 1a are due to mouse sequence Y31 (SANCHER *et al.* 1990), a distinct lambda light chain variable region gene, and sequences CZ81 and SD 26 (REIDL *et al.* 1992) from wild mice, which show more resemblance to human lambda light chain sequences. The peak around 87 differences (Figure 1a) appears as a result of including a synthetic MOPC315 sequence (BALDWIN and SCHULTZ 1989) that is drastically different from other sequences. The peaks in the region of <40 differences are more clearly shown in Figure 1b.

The peak around three differences (Figure 1b) is most likely due to somatic mutations induced by antigens or by other mechanisms, *e.g.*, radiation, etc. This peak is quite distinct, covering the range of one to five differences. It overlaps with the next peak around 10 differences. As shown in Table 1, this may be the result of sequences from different strains of mice, *i.e.*, allelic variations. Other peaks are present around 15 and 19 differences, and less distinct ones around 22, 27 and 30 differences. Since the mouse lambda 2 gene has been identified as E3-19 and it has two different sequences most likely from different strains or sources (MOTOYAMA *et al.* 1991, 1994), the peak around 19 differences can be due to gene duplication or multiplication. The peak around 15 differences may be the com-

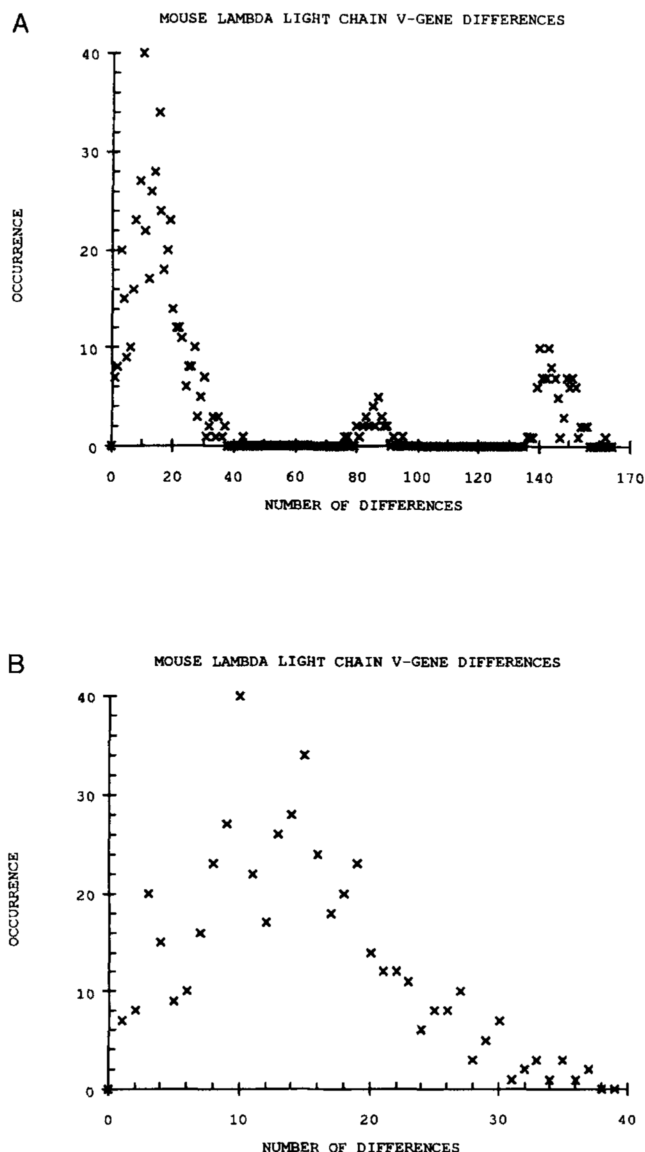

FIGURE 1.—Plot of occurrences against number of nucleotide differences listed in Table 1 for all differences (a) and for differences up to 40 (b).

bined effect of allelic variations and somatic mutations. The remaining peaks around 22, 27 and 30 differences are probably due to the various combinations of these genetic mechanisms.

Since the number of subgroups is related to gene multiplication, we are therefore able to separate its effect on nucleotide differences by constructing a second triangular table from the first one by eliminating all columns and rows containing differences between one and 17, *i.e.*, the valley between the peaks around 15 and 19 differences. This second table (Table 2) contained only six sequences. One representing the mouse lambda 1 gene, two for the two different sequences of mouse lambda 2 gene, one for the third mouse lambda gene, one from a wild mouse, and one for the synthetic MOPC315 sequence.

TABLE 2

Derived from Table 1 by eliminating all columns and rows containing one to 17 differences

	1 007 665	2 007 710	3 007 725	4 007 746	5 007 750	6 019 435
1. 007665	0	83	20	139	149	27
2. 007710		0	95	152	162	90
3. 007725			0	141	156	43
4. 007746				0	146	152
5. 007750					0	155
6. 019435						0

This simple result can then be applied to human lambda light chain sequences. A triangular table was constructed with all available complete distinct sequences for pairwise comparison. This table contains 151 sequences, which is too large to be included in this paper. The KADBIDs of these sequences are listed in the APPENDIX at the end of this paper. Unlike the mouse sequences (Table 1), the distribution of nucleotide differences is much more complicated, probably due to the large number of allelic variations and extensive gene multiplication. Although the cutoff value of 17 differences for mouse sequences may not be applicable to human sequences, it can nevertheless provide a rough estimation. Thus two additional tables are generated for human lambda light chains by eliminating all columns and rows containing one to 17 differences (Table 3) and one to 34 (*i.e.*, two times 17) differences (Table 4). These tables should give a range to the number of different human lambda light chain genes. Table 3 lists 58 sequences, while Table 4 contains 29 sequences.

Therefore, if we equate the number of subgroups to the number of distinct genes, human lambda light chains have at least 29 subgroups (Table 4), while the upper limit may be >58 (Table 3). This estimation is larger than most of previous studies (CH'ANG *et al.* 1995).

Recent experimental studies by direct nucleotide sequencing of the human lambda light chain gene region indicate that there are at least 52 genes (FRIPPIAT *et al.* 1995), which is in excellent agreement with our theoretical estimation.

The number of available nucleotide sequences for human and mouse, heavy and kappa light chains are much larger (JOHNSON *et al.* 1996). However, this simple method can also be used to estimate their gene numbers. Unfortunately, the triangular tables and KADBIDs listings are all too large to be included in this paper. Listings of currently available sequences can be obtained from the database mentioned in the MATERIALS AND METHODS section, and triangular table generated accordingly.

The results are listed in Table 5. The lower limits are based on a cutoff of 34 nucleotide differences, and the upper limits a cutoff of 17 nucleotide differences. For human heavy chains, we estimate that there are ~116–175 different genes. Experimental results of HONJO's group (MATSUMURA *et al.* 1994) indicate that there are >90 different human heavy chain V-genes. About one-third of them are pseudogenes. These pseudogenes can, however, be functional through the mechanism of gene conversion (BALTIMORE 1991) as in the case of chicken immunoglobulin system (REYNAUD *et al.* 1985, 1989). Thus, our estimation is only slightly higher than what has so far been found experimentally for human heavy chain V-genes.

For human kappa light chains, we estimated the gene number to be around 27–71, while experimentally this number is around 76 (SCHABLE *et al.* 1994). For mouse kappa light chains, our estimation is around 76–134, which is not an order of magnitude larger than the number of human kappa light chains as generally believed. Interestingly, the total number of human kappa and lambda light chain genes is, in our estimation, ~56–128; while that of mouse is around 79–139. Therefore, human and mouse appear to have the similar number of light chain V-genes.

For mouse heavy chains, our theoretical analysis gives around 99–199 genes, *i.e.*, again about the similar number as human heavy chain genes.

DISCUSSION

Subgroups (MILSTEIN 1967), sub-subgroups, families, etc. are used to classify variable regions sequences of human and mouse heavy and light chains, so that one can roughly estimate the numbers of V-genes coding for each of these chains. Such information can then provide some insight to the basic mechanisms of generating antibody diversity.

For human heavy chain variable region genes, HONJO's group (MATSUMURA *et al.* 1994) has sequenced the entire gene segment and concluded that there are over 90 different genes with one-third of these being pseudogenes. For human kappa light chain genes, ZACHAU's group (SCHABLE *et al.* 1994) has determined that there are ~76 different genes. Recently, the human lambda light chain gene segment has also been sequenced, and there are ~52 different genes (FRIPPIAT *et al.* 1995). Ideally, similar studies can soon be carried out experimentally for mouse heavy and kappa light immunoglobulin variable region genes as well as other multiple gene loci, *e.g.*, T cell receptors for antigens, MHC class I and II molecules, etc. However, in the absence of such studies, theoretical estimations of the numbers of these genes will be very valuable.

For mouse lambda light chains (SANCHEZ *et al.* 1987), the number of genes is very small, usually assumed to be three. In our present study, we have developed a method of analyzing available mouse lambda light chain nucleotide sequences by pairwise comparison (Table 1). The occurrences of various nucleotide differences can be used (Figure 1) to separate three possible genetic mechanisms of generating such differences, namely, gene multiplication, allelic variations, and somatic mutations due to antigen stimulation or due to other mechanisms. These effects can also be present simultaneously.

From the plots shown in Figure 1, there are several distinct peaks of nucleotide differences. Together with known origins of the mouse sequences, it is possible to assign these peaks to the effects of various genetic mechanisms. The peak around 19 differences seems to be due to gene multiplication. Thus, to estimate the number of different genes, we may use a cutoff of ~17 nucleotide differences. On the conservative side, we may use a cutoff value twice as large, *i.e.*, 34 nucleotide differences.

Our method assumes that human and mouse are under similar selective pressure so far antibody V-genes are concerned. We may underestimate the total number of immunoglobulin V-genes if after duplication, the two genes do not diverge. For example, in the case of human kappa light chains, some of the V-genes are transposed to other chromosomes and thus under different selective pressure (SCHABLE *et al.* 1994). Hoever, in general, this is unlikely since new V-genes obtained by divergence can give a selective advantage for the survival of the organism.

The very good agreements between our theoretical estimations and experimental sequencing of V-genes for human heavy, kappa and lambda light chains, based on mouse lambda light chain V-gene numbers, suggest that evolutionary dynamics of these genetic loci in human and mouse are reasonably similar. However, when more sequences become available, our theoretical numbers are expected to increase.

Our method of estimating the number of different genes has also been applied to mouse kappa light and mouse heavy chain V-genes. The results are shown in Table 5. Roughly speaking, human and mouse have similar numbers of immunoglobulin heavy chain V-genes. The sum of their kappa and lambda light chain genes are also similar. The smaller number of mouse lambda light chain V-genes seems to be compensated by the larger number of kappa light chain V-genes. Recent experimental studies (KIRSCHBAUM *et al.* 1996) suggest that there are ~140 mouse kappa light chain V-genes, also in very good agreement with our estimation of 76–134.

This simple method of gene number estimation can easily be applied to T cell receptor for antigen alpha and beta chain genes, MHC class I and II molecules,

TABLE 3

Similar to Table 2 for human lambda light chain V-genes

	1 007 406	2 007 410	3 007 434	4 007 459	5 007 471	6 007 480	7 007 484	8 007 502	9 007 505	10 007 515	11 007 519	12 007 523	13 007 526	14 007 527	15 007 537	16 007 538	17 007 542	18 007 546	19 007 550
1. 007406	0	18	44	130	87	88	91	95	80	84	89	74	86	88	97	56	23	60	86
2. 007410		0	55	133	94	100	97	102	90	96	102	80	98	94	102	65	38	73	92
3. 007434			0	127	89	98	99	105	78	83	90	78	89	82	97	61	48	71	84
4. 007459				0	129	130	129	128	124	127	120	123	134	126	135	122	127	135	129
5. 007471					0	51	61	110	98	108	109	87	112	103	120	91	90	107	24
6. 007480						0	23	112	106	112	117	99	114	112	121	103	91	106	69
7. 007484							0	115	107	112	119	97	114	105	125	105	91	107	69
8. 007502								0	101	104	104	91	108	111	119	98	99	110	111
9. 007505									0	22	33	39	34	38	55	78	80	86	95
10. 007515										0	38	45	40	44	61	76	83	91	102
11. 007519											0	58	42	56	73	88	91	100	105
12. 007523												0	58	55	71	81	75	91	81
13. 007526													0	51	69	90	88	97	104
14. 007527														0	73	79	87	85	93
15. 007537															0	90	99	95	117
16. 007538																0	64	42	91
17. 007542																	0	69	89
18. 007546																		0	104
19. 007550																			0
20. 007552																			
21. 007555																			
22. 007557																			
23. 007562																			
24. 007568																			
25. 007574																			
26. 007575																			
27. 007576																			
28. 007579																			
29. 007580																			
30. 007581																			
31. 019767																			
32. 020079																			
33. 020080																			
34. 020104																			
35. 020105																			
36. 020112																			
37. 020624																			
38. 020640																			
39. 021310																			
40. 021741																			
41. 021742																			
42. 021744																			
43. 021745																			
44. 021746																			
45. 021747																			
46. 021749																			
47. 021751																			
48. 021752																			
49. 021753																			
50. 021754																			
51. 021755																			
52. 021756																			
53. 021757																			
54. 021760																			
55. 021761																			
56. 021763																			
57. 021770																			
58. 026955																			

20 007 552	21 007 555	22 007 557	23 007 562	24 007 568	25 007 574	26 007 575	27 007 576	28 007 579	29 007 580	30 007 581	31 019 767	32 020 079	33 020 080	34 020 104	35 020 105	36 020 112	37 020 624	38 020 640	39 021 310
88	32	91	98	119	136	126	122	144	124	134	145	92	136	95	91	47	100	100	79
90	44	94	103	124	139	129	126	148	133	138	148	99	138	101	99	61	100	101	90
86	61	93	102	130	132	119	120	142	133	136	139	91	133	103	96	26	102	108	86
128	125	134	134	140	126	141	142	138	107	83	136	135	161	133	132	129	129	137	123
19	95	59	29	121	135	119	124	149	135	141	143	102	138	74	73	91	51	52	98
66	98	23	61	117	144	118	122	160	129	139	151	111	154	81	76	96	67	69	111
65	101	25	70	123	145	121	126	160	130	139	151	115	155	86	82	97	74	74	111
109	103	111	113	136	142	133	136	146	133	138	149	116	138	120	118	102	116	117	101
98	94	105	115	122	134	121	122	146	131	137	140	105	136	112	109	80	108	112	24
105	99	110	118	125	139	123	123	145	132	140	144	111	142	116	114	86	115	118	37
109	99	118	123	128	139	127	131	149	137	144	144	107	143	123	122	89	121	124	47
84	88	95	98	119	128	123	121	144	126	136	139	95	129	106	101	79	101	98	39
109	101	112	124	126	139	126	129	151	134	141	148	111	147	121	118	91	118	118	45
98	96	104	117	115	137	118	116	148	134	139	145	109	140	115	110	86	104	106	44
121	108	121	131	128	147	130	133	158	135	149	153	120	146	124	122	99	128	128	59
93	76	104	101	118	131	119	116	140	133	139	137	102	143	105	102	60	105	107	81
89	35	88	99	125	135	131	127	142	127	137	141	87	133	101	95	51	103	104	85
107	76	106	118	124	144	129	123	152	136	154	154	111	144	109	107	70	111	114	87
25	96	61	42	121	135	120	126	146	135	143	141	99	134	79	78	87	66	68	96
0	98	56	39	120	139	118	123	149	138	142	142	102	135	83	79	89	66	66	98
	0	100	103	126	145	133	128	153	132	140	150	102	144	104	99	61	109	111	97
		0	65	117	146	121	127	158	133	143	149	111	146	85	80	91	75	76	114
			0	127	132	124	129	147	144	139	148	114	151	82	81	103	66	67	116
				0	133	87	86	152	128	145	141	136	160	118	115	131	127	123	119
					0	140	143	68	136	121	79	145	154	144	145	141	143	146	135
						0	23	147	141	144	130	140	152	121	118	122	128	130	120
							0	149	141	145	135	140	157	123	120	122	130	136	121
								0	143	136	62	148	158	149	150	145	157	162	147
									0	90	139	144	160	127	130	134	125	130	124
										0	133	148	167	133	133	138	140	148	134
											0	146	150	148	147	142	156	160	145
												0	93	108	102	96	111	114	105
													0	148	146	136	155	156	138
														0	19	106	76	79	112
															0	101	78	74	109
																0	104	112	88
																	0	31	110
																		0	108
																			0

TABLE 3

Continued

	40 021 741	41 021 742	42 021 744	43 021 745	44 021 746	45 021 747	46 021 749	47 021 751	48 021 752	49 021 753	50 021 754	51 021 755	52 021 756	53 021 757	54 021 760	55 021 761	56 021 763	57 021 770	58 026 955
1. 007406	92	44	96	18	51	49	79	92	133	79	92	97	104	100	90	91	93	34	96
2. 007410	103	59	103	32	64	62	89	100	136	89	103	103	111	109	103	103	102	49	105
3. 007434	85	58	91	51	57	31	83	93	127	86	102	104	111	103	84	88	90	65	106
4. 007459	127	134	133	134	126	129	125	134	144	127	135	135	130	134	123	127	132	133	135
5. 007471	105	98	115	96	93	91	98	108	131	101	66	74	55	71	104	109	113	94	68
6. 007480	115	101	118	98	97	99	108	108	126	112	24	40	83	39	110	113	115	97	33
7. 007484	119	105	120	98	103	101	112	112	133	116	28	44	84	46	115	114	117	98	43
8. 007502	103	103	111	99	97	105	96	116	142	96	114	114	119	120	104	110	110	102	116
9. 007505	28	86	43	89	73	79	31	35	134	35	113	115	115	112	32	23	34	94	110
10. 007515	37	91	44	91	77	84	39	42	137	41	115	114	118	115	35	33	44	94	108
11. 007519	44	98	55	97	81	89	47	52	135	49	122	121	115	120	47	38	47	95	117
12. 007523	50	76	67	82	77	78	42	63	133	46	103	107	108	106	52	51	61	88	100
13. 007526	50	95	51	96	85	90	47	54	132	50	116	120	124	119	48	47	49	96	115
14. 007527	58	91	46	97	80	87	47	50	126	52	113	119	118	116	34	41	54	96	114
15. 007537	61	104	76	106	88	105	61	77	141	67	129	120	129	122	67	68	70	111	126
16. 007538	87	62	88	65	28	71	83	93	134	83	108	102	109	110	83	84	83	70	110
17. 007542	91	50	95	27	56	54	84	91	136	83	94	94	108	102	89	90	90	49	96
18. 007546	98	66	102	69	38	80	92	102	135	92	110	109	123	109	96	94	93	79	114
19. 007550	105	100	107	96	93	89	94	104	132	99	77	80	60	81	99	103	107	98	83
20. 007552	106	100	112	98	96	92	100	109	128	102	74	78	60	78	105	111	111	96	80
21. 007555	100	51	99	34	67	63	92	100	141	91	101	97	112	107	98	100	104	51	98
22. 007557	117	101	118	98	101	97	113	110	129	116	33	45	86	45	113	116	115	101	47
23. 007562	115	107	122	102	104	104	113	123	132	111	74	77	65	75	119	116	124	104	74
24. 007568	127	130	132	123	113	134	117	120	101	118	120	119	131	128	124	122	125	132	124
25. 007574	139	143	142	142	134	141	136	146	143	136	144	148	143	142	137	140	137	150	154
26. 007575	126	135	124	129	117	124	119	122	34	124	123	116	133	123	121	125	131	130	127
27. 007576	128	129	126	127	112	124	120	121	54	126	127	124	135	129	119	126	130	131	130
28. 007579	154	152	153	147	141	146	147	159	154	145	156	162	156	154	148	149	152	157	161
29. 007580	139	136	142	131	133	136	125	137	147	134	130	128	134	132	127	132	133	134	134
30. 007581	141	140	141	135	144	139	137	145	151	138	144	140	147	142	136	141	146	146	144
31. 019767	150	153	148	148	138	145	146	154	142	146	153	149	153	150	148	146	148	154	156
32. 020079	111	105	119	102	94	98	102	116	146	111	112	117	117	122	109	107	116	106	122
33. 020080	144	147	146	145	138	137	135	147	156	142	153	154	147	154	139	145	148	148	160
34. 020104	123	104	120	99	98	109	114	125	132	117	89	98	97	98	119	118	118	108	92
35. 020105	116	96	119	92	92	103	109	120	127	114	86	91	98	94	114	114	115	108	87
36. 020112	91	59	93	55	56	24	85	97	129	87	101	100	109	101	90	89	93	67	100
37. 020624	119	108	118	110	109	104	111	116	137	114	78	87	82	86	110	116	118	109	78
38. 020640	121	110	122	111	110	113	114	118	136	116	74	91	82	88	114	118	120	111	78
39. 021310	35	89	53	88	77	86	23	50	133	24	114	118	115	116	38	41	51	92	114
40. 021741	0	83	48	87	73	79	31	41	131	29	113	109	109	114	38	28	39	92	117
41. 021742		0	92	40	45	52	78	89	135	81	99	95	110	100	83	84	89	56	106
42. 021744			0	92	81	85	44	42	132	47	116	109	111	110	32	42	48	90	125
43. 021745				0	51	47	79	89	129	75	93	90	105	101	88	90	90	39	103
44. 021746					0	55	65	82	124	69	96	94	102	98	75	70	73	60	106
45. 021747						0	77	86	123	78	95	95	103	96	78	79	87	59	102
46. 021749							0	40	123	19	103	108	104	106	31	34	44	80	112
47. 021751								0	126	46	105	102	114	104	34	35	47	92	119
48. 021752									0	128	124	123	144	125	126	127	133	135	136
49. 021753										0	108	109	105	108	36	38	44	81	117
50. 021754											0	41	79	37	102	112	111	93	44
51. 021755												0	77	41	109	110	112	96	57
52. 021756													0	82	106	112	110	96	96
53. 021757														0	107	110	108	99	61
54. 021760															0	33	41	88	117
55. 021761																0	34	89	115
56. 021763																	0	91	119
57. 021770																		0	99
58. 026955																			0

TABLE 4

Derived from Table 3 by eliminating all columns and rows containing 18–34 differences

	1 007 459	2 007 471	3 007 502	4 007 537	5 007 568	6 007 574	7 007 576	8 007 579	9 007 580	10 007 581	11 019 767	12 020 079	13 020 080	14 020 104	15 020 112	16 020 624	17 021 310	18 021 741	19 021 742	20 021 744	21 021 745	22 021 746	23 021 751	24 021 752	25 021 754	26 021 755	27 021 756	28 021 757	29 021 770
1. 007459	0	129	128	135	140	126	142	138	107	83	136	135	161	133	129	129	123	127	134	133	134	126	134	144	135	135	130	134	133
2. 007471		0	110	120	121	135	124	149	135	141	143	102	138	74	91	51	98	105	98	115	96	93	108	131	66	74	55	71	94
3. 007502			0	119	136	142	136	146	133	138	149	116	138	120	102	116	101	103	103	111	99	97	116	142	114	114	119	120	102
4. 007537				0	128	147	133	158	135	149	153	120	146	124	99	128	59	61	104	76	106	88	77	141	129	120	129	122	111
5. 007568					0	133	86	152	128	145	141	136	160	118	131	127	119	127	130	132	123	113	120	101	120	119	131	128	132
6. 007574						0	143	68	136	121	79	145	154	144	141	143	135	139	143	142	142	134	146	143	144	148	143	142	150
7. 007576							0	149	141	145	135	140	157	123	122	130	121	128	129	126	127	112	121	54	127	124	135	129	131
8. 007579								0	143	136	62	148	158	149	145	157	147	154	152	153	147	141	159	154	156	162	156	154	157
9. 007580									0	90	139	144	160	127	134	125	124	139	136	142	131	133	137	147	130	128	134	132	134
10. 007581										0	133	148	167	133	138	140	134	141	140	141	135	144	145	151	144	140	147	142	146
11. 019767											0	146	150	148	142	156	145	150	153	148	148	138	154	142	153	149	153	150	154
12. 020079												0	93	108	96	111	105	111	105	119	102	94	116	146	112	117	117	122	106
13. 020080													0	148	136	155	138	144	147	146	145	138	147	156	153	154	147	154	148
14. 020104														0	106	76	112	123	104	120	99	98	125	132	89	98	97	98	108
15. 020112															0	104	88	91	59	93	55	56	97	129	101	100	109	101	67
16. 020624																0	110	119	108	118	110	109	116	137	78	87	82	86	109
17. 021310																	0	35	89	53	88	77	50	133	114	118	115	116	92
18. 021741																		0	83	48	87	73	41	131	113	109	109	114	92
19. 021742																			0	92	40	45	89	135	99	95	110	100	56
20. 021744																				0	92	81	42	132	116	109	111	110	90
21. 021745																					0	51	89	129	93	90	105	101	39
22. 021746																						0	82	124	96	94	102	98	60
23. 021751																							0	126	105	102	114	104	92
24. 021752																								0	124	123	144	125	135
25. 021754																									0	41	79	37	93
26. 021755																										0	77	41	96
27. 021756																											0	82	96
28. 021757																												0	99
29. 021770																													0

TABLE 5

Estimated numbers of human and mouse immunoglobulin V-genes, based on these of mouse lambda light chain V-genes

	Mouse	Human
Lambda light chain	3–5	29–58
Kappa light chain	76–134	27–71
Heavy chain	99–199	116–175

and other proteins from multi-gene families. We are currently analyzing these sequences.

LITERATURE CITED

ANDERSSON, E., V. TORMAREN and T. MATSUMAGA, 1991 Evolution of a VH gene family in low vertebrates. Intl. Immunol. **3:** 527–533.

BALDWIN, E., and P. G. SCHULTZ, 1989 Generation of a catalytic antibody by site directed mutagenesis. Science **245:** 1104–1107.

BALTIMORE, D., 1981 Gene conversion: some implications for immunoglobulin genes. Cell **24:** 592–594.

CH'ANG, L. Y., M. SCHELL, C. RINGELBERG, D. T. WEISS and A. SOLOMON, 1995 Molecular characterization of human V lambda VIII germline gene. Mol. Immunol. **32:** 49–55.

CHEN, C., V. A. ROBERTS and M. B. RITTENBERG, 1992 Generation and analysis of random point mutations in an antibody CDR2 sequence: many mutated antibodies loss their ability to bind antigen. J. Exp. Med. **176:** 855–866.

DILDROP, R., 1984 A new classification of mouse VH sequences. Immunol. Today **5:** 84–86.

FRIPPIAT, J.-P., S. C. WILLIAMS, I. M. TOMLINSON, G. P. COOK, D. CHERIF *et al.*, 1995 Organization of the human immunoglobulin lambda light-chain locus on chromosome 22q11.2. Hum. Mol. Genet. **4:** 983–991.

JOHNSON, G., E. A. KABAT and T. T. WU, 1996 Kabat database of proteins of immunological interest, pp. 6.1–6.21 in *Handbook of Experimental Immunology, I. Immunochemistry,* edited by D. M. WEIR. Blackwell Science, Inc., Cambridge.

KABAT, E. A., T. T. WU and H. BILOFSKY, 1976 *Variable Regions of Immunoglobulins,* National Institutes of Health, Bethesda, MD.

KABAT, E. A., T. T. WU and H. BILOFSKY, 1979 *Sequences of Immunoglobulin Chains,* National Institutes of Health, Bethesda, MD.

KABAT, E. A., T. T. WU, H. M. PERRY, K. S. GOTTESMAN and C. FOELLER, 1991 *Sequences of Proteins of Immunological Interest,* Ed. 5, National Institutes of Health, Bethesda, MD.

KIRSCHBAUM, T., R. JAENICHEN and H. G. ZACHAU, 1996 The mouse immunoglobulin kappa locus contains about 140 variable gene segments. Eur. J. Immunol. **26:** 1613–1620.

MATSUMURA, R., F. MATSUDA, H. NAGAOKA, J. FUJIKURA, E. K. SHIN *et al.*, 1994 Structural analysis of the human VH locus using nonrepetitive intergenic probes and repetitive sequence probes. J. Immunol. **152:** 660–666.

MILSTEIN, C., 1967 Linked groups of residues in immunoglobulin chains. Nature **216:** 330–332.

MOTOYAMA, N., H. OKADA and T. AZUMA, 1991 Somatic mutations in constant regions of mouse lambda 1 light chains. Proc. Natl. Acad. Sci. USA **88:** 7933–7937.

MOTOYAMA, N., T. MIWA, Y. SUGUKI, H. OKADA and T. AZUMA, 1994 Comparison of somatic mutation frequency among immunoglobulin genes. J. Exp. Med. **179:** 395–403.

POTTER, M., 1977 Antigen-binding myeloma proteins of mice. Adv. Immunol. **5:** 141–211.

REIDL, L. S., C. M. KINOSHITA and L. A. STEINER, 1992 Wild mice express an Ig V lambda gene that differs from all V lambda in BALB/c but resembles a human V lambda subgroup. J. Immunol. **149:** 471–480.

REYNAUD, C.-A., V. ANQUEZ, A. DAHAN and J.-C. WEILL, 1985 A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. Cell **40:** 283–291.

REYNAUD, C.-A., A. DAHAN, V. ANQUEZ and J.-C. WEILL, 1989 Somatic hyperconversion diversifies the single VH gene of the chicken with a high incidence in the D region. Cell **88:** 171-183.

SANCHEZ, P., P. N. MARCHE, C. LE GUERN and P. A. CAVENAVE, 1987 Structure of the third murine immunoglobulin lambda light chain variable region that is expressed in laboratory mice. Proc. Natl. Acad. Sci. USA **84:** 9185–9188.

SANCHEZ, P., P. N. MARCHE, D. RUEFF-JUY and P. A. CAVENAVE, 1990 Mouse V lambda x gene sequence generates no junctional diversity is conserved in mammalian species. J. Immunol. **144:** 2816–2820.

SCHABLE, K., R. THIEBE, A. FLUGEL, A. MEINDL and H. G. ZACHAU, 1994 The human immunoglobulin kappa locus: pseudogenes, unique and repetitive sequences. Biol. Chem. Hoppe-Seyler **375:** 189–199.

SERRANO, M. P., A. CARDONA, B. VERNET DER GACABEDIAN, J. F. BUCH and J. M. PLEANS, 1994 Nucleotide sequences of variable regions of a human anti-acetylcholine receptor antibody derived from a mysthenia patient. Mol. Immunol. **31:** 413–417.

Communicating editor: G. B. GOLDING

APPENDIX

The 151 distinct complete human lambda light chain V-gene sequences, positions 1–95, are (in KADBID) as follows:

007406; 007407; 007410; 007413; 007417; 007421; 007422; 007424; 007425; 007426; 007427; 007428; 007430; 007433; 007434; 007438; 007440; 007441; 007444; 007447; 007449; 007450; 007453; 007459; 007460; 007461; 007462; 007463; 007464; 007468; 007471; 007473; 007474; 007480; 007484; 007488; 007489; 007492; 007493; 007494; 007495; 007498; 007502; 007505; 007508; 007509; 007510; 007511; 007512; 007513; 007514; 007515; 007516; 007517; 007518; 007519; 007520; 007522; 007523; 007526; 007527; 007533; 007534; 007535; 007536; 007537; 007538; 007539; 007540; 007542; 007544; 007546; 007550; 007552; 007555; 007557; 007560; 007562; 007568; 007572; 007574; 007575; 007576; 007678; 007579; 007580; 007581; 019745; 019767; 020079; 020080; 020081; 020098; 020102; 020104; 010205; 020112; 020136; 020138; 020522; 020623; 020624; 020625; 020634; 020640; 020642; 020644; 020646; 020650; 021310; 021311; 021312; 021455; 021741; 021742; 021744; 021745; 021746; 021747; 021748; 021749; 021750; 021751; 021752; 021753; 021754; 021755; 021756; 021757; 021758; 021759; 021760; 021761; 021762; 021763; 021764; 021765; 021766; 021767; 021768; 021769; 021770; 021771; 022264; 022265; 022266; 022682; 023194; 023195; 026955; 026956.