Yoshikazu Ichihara, Hidenori Hayashida○, Sanzo Miyazawa○ and Yoshikazu Kurosawa

Institute for Comprehensive Medical Science, Fujita-Gakuen Health University, Toyoake, Aichi and National Institute of Genetics○, Yata, Mishima

Only D_{FL16}, D_{SP2}, and D_{Q52} gene families exist in mouse immunoglobulin heavy chain diversity gene loci, of which D_{FL16} and D_{SP2} originate from the same primordial D_H gene

In mice, 12 germ-line D_H genes belonging to three different families (D_{Q52}, D_{SP2} and D_{FL16}) have been identified. The D_H genes other than D_{Q52} are clustered in the 60 kb-long region located between V_H and J_H genes. Since there are seven D_H gene families (D_{HQ52}, D_{XP}, D_A, D_K, D_N, D_M and D_{LR}) in humans, we tried to identify new D_H gene families in the 60 kb-long region using human D_H gene probes. Mouse and human D_H genes showing the highest similarity were mouse D_{FL16} genes and human D_A genes. Southern hybridization of the mouse clones covering the 60-kb region with human D_H probes did not detect any other D_H genes. Nucleotide sequence analysis of the 4.0-kb fragment containing the $D_{FL16.1}$ gene confirmed this conclusion. Comparison of the 12 germ-line D_H genes and more than 150 somatic D_H sequences also indicated that there are not more germ-line D_H genes in the mouse genome. Moreover, comparison of nucleotide sequences of $D_{FL16.1}$ and $D_{SP2.2}$ genes and their surrounding regions suggests that both D_H gene families originate from the same primordial D_H gene. Using the flanking sequences of both D_H genes, the divergence date between D_{FL16} and D_{SP2} genes was estimated at around 37 million years ago.

1 Introduction

The V region of Ig H chain is encoded by three separate genes in the germ-line genome: V_H, D_H and J_H [1]. Both D_H-J_H and V_H-D_H joinings are necessary to complete an active V_H gene [1]. These DNA rearrangements are mediated by the recombinase which recognizes the heptamers CACTGTG and CACAGTG, and the nonamers GGTTTTTGT and ACAAAAACC [2]. The spacer length separating these oligomers is either 12 or 23 nucleotides [3]. D_H-coding sequences are bordered by two sets of 12-nucleotide spacer signals. In mouse, 12 germ-line D_H genes have been identified and they can be classified into three D_H gene families (D_{Q52}, D_{SP2} and D_{FL16} [4]). The D_H genes belonging to the D_{SP2} family are regularly spaced every 5 kb. Although human D_H genes originally identified by Siebenlist et al. [5] are also regularly spaced every 9 kb, we showed that each 9-kb repeating sequence contains six different D_H gene families (D_{XP}, D_A, D_K, D_N, D_M and D_{LR}; [6]).

In this study we tried to identify new D_H gene families in the mouse genome using human D_H gene-containing fragments as probes. Most mouse D_H genes are clustered in the 60-kb region located between V_H and J_H genes. Southern hybridization of the phage DNA covering the 60-kb region indicated that only fragments containing D_{FL16} weakly cross-hybridized with the human D_A probe. We determined the nucleotide sequence of the 4-kb DNA fragment containing $D_{FL16.1}$. This fragment does not contain any D_H gene other than $D_{FL16.1}$ itself. Comparison of nucleotide sequences of the germ-line D_H genes and more than 150 somatic D_H genes indicated that there are not more than 12 germ-line D_H genes in the mouse genome. We also discuss the evolution of the mouse D_H gene loci.

[I 7722]

* Supported by grants from the Ministries of Education, Science and Culture, Health and Welfare, and Agriculture, Forestry and Fisheries in Japan; Fujita-Gakuen Health University.

Correspondence: Yoshikazu Kurosawa, Institute for Comprehensive Medical Science, Fujita-Gakuen Health University, Toyoake, Aichi 470-11, Japan

2 Materials and methods

Six human D_H probes D_{XP}, D_A, D_K, D_N, D_M and D_{LR} were described in a previous report [6]. Three mouse D_H gene-containing clones, RI-2, RI-6, and RP13 were described by Kurosawa and Tonegawa [4]. Southern hybridization was carried out under non-stringent conditions [6, 7]. DNA sequencing was performed by the dideoxynucleotide chain termination method [8].

3 Results

3.1 Identification of putative D_H genes by human D_H probes

Since in the mouse containing clusters of D_H genes regions consist of highly conserved 5-kb repeats [4], three mouse clones, RI-2, RI-6 and RP13 [4] were used as representatives of mouse D_H gene-containing clones (Fig. 1). DNA was digested with Eco RI. Six different human D_H gene-containing fragments, described previously [6], were used as probes for Southern hybridization. Five probes: D_{XP}, D_N, D_M, D_K and D_{LR} did not give any distinct signals (data not shown). However, the 4-kb Eco RI fragment in clone RI-2 and the 6.7-kb fragment in clones RI-6 and RP13 gave weak but distinct signals with the D_A probe as shown in Fig. 2. Southern hybridization of cellular DNA with these six human probes did not give any signal (data not shown). We concluded that if mouse D_H genes other than D_{SP2} and D_{FL16} exist, they should have been on these 4-kb and 6.7-kb fragments.

0014-2980/89/1010-1849$02.50/0

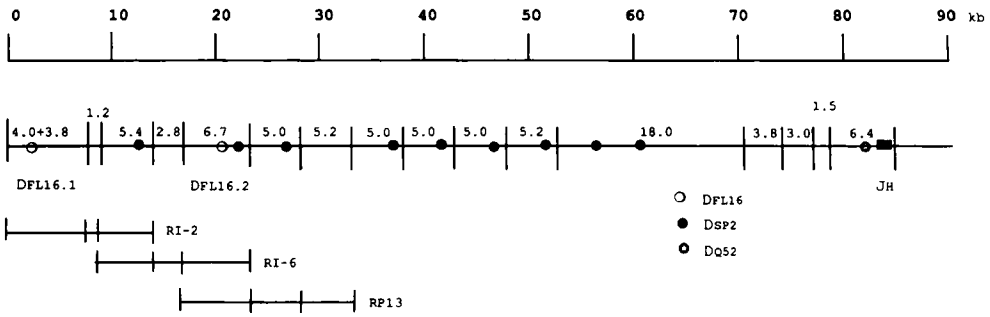

Figure 1. Organization of mouse D_H gene loci. Twelve D_H genes belonging to three families have been identified [4, 9]. Clones RI-2, RI-6, and RP13 were used in this study. Numbers on the second line indicate sizes of Eco RI fragments in kb. The 4-kb fragment containing $D_{FL16.1}$ was sequenced.

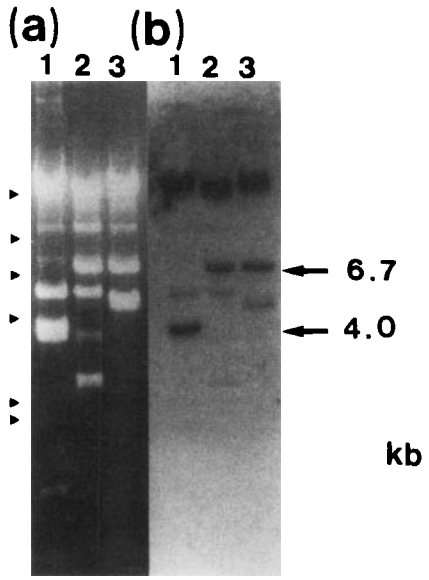

Figure 2. Southern hybridization of mouse D_H gene-containing clones [4] with human D_A probes. (a) Of each phage DNA 0.5 μg was digested with Eco RI, separated by agarose gel electrophoresis and stained with ethidium bromide. (1) RI-2 contains four Eco RI fragments: 5.4, 4.0, 3.8, 1.2 kb. (2) RI-6 contains three Eco RI fragments: 6.7, 5.4, 2.8 kb. (3) RP13 contains three Eco RI fragments: 6.7, 5.2, 5.0 kb. The origin of faint bands is not known. Closed triangles indicate the position of λ-Hind III markers. (b) Southern blots of these separated DNA which were hybridized with the human D_A probe [6]. The 4.0-kb band in clone RI-2 (1), and the 6.7-kb band in clone RI-6 (2) and clone RP13 (3) gave distinct signals.

3.2 Nucleotide sequence of the 4.0-kb fragment containing $D_{FL16.1}$

Although the 4-kb and 6.7-kb fragments contained D_{FL16} genes, we determined the total nucleotide sequence of the 4-kb fragment to find the regions giving positive signals with the D_A probe. As shown in Fig. 3, there was only one D_H gene on this fragment. Homology research between nucleotide sequences of the 15-kb human D_H-containing region [6] and this 4-kb fragment showed that homologous regions are very restricted in $D_{FL16.1}$ and D_{A4} genes themselves. Fig. 4 shows the comparison of nucleotide sequences of $D_{FL16.1}$ and D_{A4} genes. The signal and coding regions of these two genes showed 85% homology; however, the surrounding regions did not have any distinct homology. Although the 6.7-kb fragment containing $D_{FL16.2}$ was not sequenced, it is likely that the region which gave a positive signal with the D_A probe in the 6.7 kb fragment was the $D_{FL16.2}$ gene itself.

4 Discussion

In the mouse, 12 D_H genes have been identified, and they can be classified into 3 D_H gene families [4]. In this study, we tried to identify new D_H gene families in the mouse genome using human D_H probes, since there are seven human D_H gene families [6]. However, the only D_H genes detected by human D_H probes were D_{FL16} genes. Most D_H genes were originally identified by using DNA fragments containing D_H-J_H joints [4, 5]. In the case of the mouse system, many D_H-J_H fragments have been sequenced, and in all cases published so far, one of the 12 D_H genes identified was involved in such joinings [4, 10, 11]. As shown in this study, the D_H genes cross-hybridizing with the available D_H probes belong to the 12 germ-line D_H genes. Therefore, it is unlikely that new D_H gene families remain to be found. If so, the 12 germ-line D_H genes should encode all somatic D_H sequences known so far.

When Kurosawa and Tonegawa [4] compared germ-line D_H sequences with somatic D_H sequences, only 16 somatic sequences were known. Now, more than 200 somatic D_H sequences are known. It is thus worth comparing once more both germ-line and somatic D_H sequences. As a source of somatic D_H sequences, we used the data book (1987) edited by Kabat et al. [12] although more data has since been published. We defined the somatic D_H segment as the region which is not encoded by either germ-line V_H or J_H genes; therefore, N regions are included in somatic D_H segments [13]. Since all of the germ-line J_H sequences are known [14], the boundaries between D_H and J_H regions can be easily assigned. We tentatively assigned the 94th amino acid residue to the germ-line V_H gene and the region after the 95th residue to the D_H region (for details see legend of Fig. 5). In the data book [12] 158 somatic D_H sequences are available. As listed in Fig. 5, one fifth of them could not be assigned to any of the three D_H gene families. Some of them are too short to be assigned. The majority of them are G-rich sequences. Does this mean that there are other germ-line D_H genes which are rich in G residues? We think that this is not the case because there is no regularity among these sequences. If these G-rich sequences were encoded by germ-line sequences, there should be sequence similarities among them. They are rich in G residues, but seem to be random sequences, and they may be the products of the activity of the terminal transferase as proposed by Alt and Baltimore [13]. The regions encoded by germ-line D_H genes would have been removed during V_H-D_H and D_H-J_H joining processes.

Fig. 5 summarizes the assignments of somatic D_H sequences to germ-line D_H genes. Classification of somatic D_H sequences was based on similarities of D_H-coding regions and coding

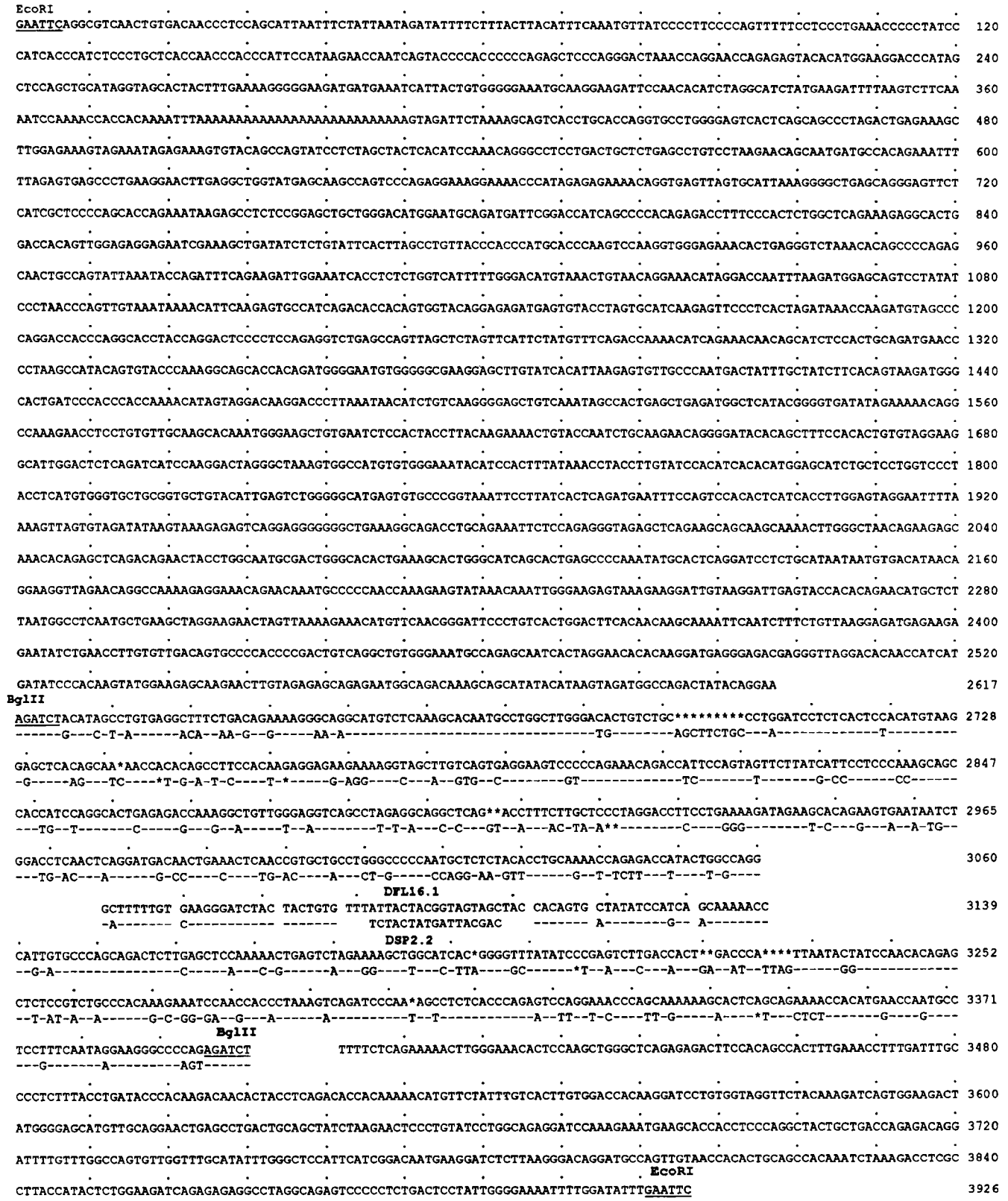

Figure 3. Nucleotide sequence of the 4-kb Eco RI fragment containing the $D_{FL16.1}$ gene. Total nucleotide sequence of the 4-kb Eco RI fragment in RI-2 was determined. For comparison, sequence of the Bgl II fragment containing the $D_{SP2.2}$ gene [17] is also shown. Bars indicate the same nucleotide as that of $D_{FL16.1}$. Stars indicate missing nucleotides.

```
DFL16.1   GGCCAGG GCTTTTTGT GAAGGGATCTAC TACTGTG TTTATTACTACGGTAGTAGCTAC CACAGTG CTATATCCATCA GCAAAAACC CATTGTG
            .  ..  .........  ......  ..  .  ........     .  .....    ....  ....  ........  .  .  .  ...  ..  .  .....    .  .
DA4       CTCAGGG GCTTTTTGT GAAGGGTCCTCC TACTGTG     TGACTAC   AGTAACTAC CACAGTG ATGAACCCAGCA GCAAAAACT GACCGGA
```

Figure 4. Comparison of nucleotide sequences between mouse $D_{FL16.1}$ and human D_{A4}. Combinations of mouse and human D_H genes showing the highest similarity were mouse D_{FL16} genes and human D_A genes. The sequence of D_{A4} gene was published in a previous study [6]. Positive signals, with human D_A probes, from the D_H gene-containing clones (Fig. 2) should be due to the above homology.

(A) FL16 family

Frame I	N_L	TT TAT TAC TAC GGT AGT AGC TAC	N_R	J_H	ref.
	TT	C TAC GGT AGT AGC T	GG	J4	23
		TAC GGT	GGGGCCT	J4	24
	A	AC TAC GGT AGT		J2	27
	GGC	TAC TAC GGT AGT AG	A	J2	30
	AGGG	AC TAC GTT AGT AGG TAC	GACCC	J3	55,57
	GTCTCAA TT TAT TAC TAC GGT CGT AGC GAC		AAATACTTCACTT	J4	60
	GC T TAT TAC TAC GGT		CCT	J3	69
	G AT TAC TAC GGT AGT			J2	76
	C AT TAC TAC GGT AGT AGC T		CC	J3	83
	TAT TAC TAC GGT AGT AGC		CAT	J3	84
	(CGC) TAT TAC TAC GGT AGT AGC		CTA	J1	85
	GGT AGT AGC TAC G			J4	90
	TACG AT TAC TAC GGT AGT AGC *TAC*			J2	92,102
	TAT TAC TAC GGT AGT AGC			J2	95
	TACA AT TAC TAC GGT AGT ATC *TA*			J4	97
	TAT TAC TAC GGT AGT AGC *TAC*			J2	98
	CG T TAT TAC TAC GGT AGT AGC		CCTTG	J2	109
	TCGGATGG T TAC TAC G		ACTGGTTTG	J3	110
	GGGCAGA TT TAT TAC TAC GGT AGT ACC *T*			J2	112
	TAT TAC TAC GGT AGT AGC			J2	121
	TCG GGT AGT AAC TAC C			J1	123
	CCCCACCCAT C TAC GGT AGT AGC TAC			J4	130
	GGG TAC TAC GGT AGT AGC			J2	132
	TAT TAC TAC GGT AGT AGC *TAC*			J2	133
	TATG AT TAC TAC GGT AGT AGC *TAC*			J2	134
	TAT TAC TAC GGT AGT AGC *TAC*			J2	139
	AG T TAT TAC TAC GGT AGT A		CGTCCG	J3	141
	TAC TAC GGT AGT AGC TAC T			J2	145
	TACCTC TAT TAC TAC GGT AGT AGC *TAC*			J2	148
	TACGA T TAT TAC TAC GG		GT	J3	149
	TAT TAC TAC GGT		GCTG	J2	154
	TAT TAC TAC GGT AGT AGC T		T	J3	155
	TAT TAC TAC GAG AGT AGC		CT	J3	156
	(CTC) TAT TAC TAC GTT AGT AGC TAC G			J3	157
	(CCT) TAT TAC TAC GG		GGGGG	J1	158
	(CCT) TAT TAC TAC GG		GG	J1	159
	▶GGG TAC GGC T		TAT	J4	163
	G AT TAC TAC GGT AGT AGC *TAC*			J1	167
	A AT TAC TAC GGT AGT ACC *T*			J1	172
	GGA TAC TAC GG		AGGAG	J3	173,174
	TC T TAC TCC GGT AGT AGC		C	J3	175,176
	GATGCGG AC TAC GGT AAT AGC TAC		TTTG	J1	181
	GGG AGT AGC TAC		GGAG	J3	184
	GG AT TAC TAC GGT		G	J3	185,187,188
	GATGCAGAGGT T TAC TAT GGT GGT AGC T		CT	J1	199
	TC T TAC TAC GGT AG		CC	J3	203
	A AC GAC GGT AGT AGC TAC		GG	J2	204
	TCCC AT TAC TAT GGT GGT AGC TAC G			J2	217
	TCGGTC TAC TAT GGT GGT AGT TAC			J2	218
	TCGA AT TAC TAT GGT GGT AGC TAC TC			J2	219
	GGGAT T TAC TAC AAT AGT AGC		CC	J2	221
	▶GGA TT CAT TAC TAC GTC CAC			J2	237,238,239,240
	GAT TAC GG		G	J3	244,245
	GAT TAC GG		G	J2	246
	GA T TAT TAC TAC GGT AG		GG	J3	255

Frame II	TTT ATT ACT ACG GTA GTA GCT AC			
	TTC ATT ACT ACG GCT AC			
GATGCGCTC	ATT ACT ACG GTA G	GGAGGGGGT	J2	124
▶(CCT)	ACT ACG GC	CTTAGAGGGG	J1	138
GGA	ACT ACG GTG G	GGAGA	J2	146
CCCCCTC	TT ATT TCG TTA GTA GC	GG	J4	241
	ATT ACT ACG GTA G		J4	251

Frame III	T TTA TTA CTA CGG TAG TAG CTA C			
	T TCA TTA CTA CGG CTA C			
(ACC)GA	A CGG T	TAGGGG	J1	122
CG	A CGA CGG GAG T	CGA	J2	140
GGGATT	TTT CGA CGG	G	J4	147
TC	A CTA CGG	GT	J4	162
A	GG CTA	ATGG	J4	165
AAGGGAC	TA CTA CGG T		J4	202
▶CAA	CTA CGG CT	C	J3	249

(B) SP2 family

Frame I	N_L	TC TAC TAT GGT TAC GAC TC TAC TAT GGT AAC TAC CC TAC TAT GGT TAC GAC CC TAC TAT GGT AAC TAC CC TAG TAT GGT AAC TAC TC TAC TAT GAT TAC GAC	N_R	J_H	ref.
	GGT	TAT GGT	GGCCT	J2	25
	GATA	AT GAT CAC	CTCT	J2	28
		TAC TAT GAT TAC T		J3	29
	GGG	AT GGT AAT TAC	AGGAATTTG	J2	31
	G	AC TAT GG	G	J3	54
		TAC TAT GGT AAC *TAC*		J2	86
	AGCG	AG TAT GGT AA	T	J2	108
	GAGGAA	TAC TAT GAT TAC G	CTT	J2	129
	(ATT)	TAT AGT AAC TAC	AAGT	J2	136
		TAC TAT ACG TAC	CCT	J4	150
		TAC TAT AGG TAC	CCT	J4	151,152,
		TAT AGG TAC	GAGAGGG	J3	153
	TCGG	AT GGT *TAC TA*		J4	161
	GGGGGA	TAT GG	GGGAGTA	J4	166
	G	CC TAC TAT GG	AG	J3	171
	GG	TAT AGG TAC G	G	J3	191
	GG	TAT GCT TAC G	GGG	J3	192
	GG	GAT TAC GAC	GGGG	J3	193
	GATG	CC TAC TAT AGT AAC *TAC*		J1	194
	CCG	TAT GGT AAC *TAC*		J4	200,201
	C	CC TAC TAT AGG TAC	CCT	J1	205
	TGG	GGT AAC TAC	CC	J4	234
	(ATC)	TAC TAT AGG TA	T	J4	252
	TGG	GGT AAC TAC	CCTC	J4	253
		TAC TAT AGG TAC	CCT	J4	254
		TAC TAT AGT AAC TAC	CC	J4	256

Frame II	TCT ACT ATG GTT ACG AC			
	TCT ACT ATG GTA ACT AC			
	CCT ACT ATG GTT ACG AC			
	CCT ACT ATG GTA ACT AC			
	CCT AGT ATG GTA ACT AC			
	TCT ACT ATG ATT ACG AC			
ATGGGGC	CT ATG GTG GTT AC	ACCC	J3	59
TCG	TCT AC	CTC	J4	164

Frame III	T CTA CTA TGG TTA CGA C			
	T CTA CTA TGG TAA CTA C			
	C CTA CTA TGG TTA CGA C			
	C CTA CTA TGG TAA CTA C			
	C CTA GTA TGG TAA CTA C			
	T CTA CTA TGA TTA CGA C			
GG	GA TTA C	A	J3	68,70,72
TACAGG	TGG TTA	GAGG	J3	135
GG	A TGG TTA CGA C	GTG	J4	143,144
	TGG GAA	TC	J4	160

(C) D_{Q52}

CAACTGGGAC			
GAT CTGG		J4	61
GATC CTGGG G		J3	63
TAT CGACTGGGAC GG		J2	104
G ACTGGG TC		J3	128
TCA AACTGGG		J2	131
GG TGGGAC		J1	178,179,
GCT AACTGGGA T		J3	189

(D) not classified

GATAGGGG	J3 32	GATCATGGG	J3 49
GATGGGGG	J3 33	GATTGG	J4 50
GATCGTGGG	J3 34	GATCAGGGG	J2 51
GATCGGGGGG	J3 35	GATCAGGGG	J4 52
GATCGAGGGGTG	J3 36	AACGGAGGG	J4 56
GACAGA	J4 37	GTAGCTCCGGGG	J2 58
GATCGGGG	J3 38	GATAGG	J1 65
GATGGGTT	J4 39	GGG	J3 89
GATGGGGA	J4 40	TATTG	J4 96
GAAGGGG	J4 41	GATTGGGGCT	J3 117
GATAGCGGA	J3 42	---	J3 126
GATCGGG	J2 43	TATTT	J3 127
GATCATGGG	J2 44	AGGGATCTCAGGG	J1 137
GATCGGGG	J3 45	CCGGGGGTCCC	J2 206
GATGGGGG	J3 46	GACGGGGGGA	J2 247
GATGGGG	J2 47	(CCC)---	J2 248
GATGGG	J2 48	TTAGACACCTCCG	J3 250

Figure 5. Assignments of somatic D_H sequences to germ-line D_H genes. The data book (page 508 to 519) edited by Kabat et al. [12] was used as the source of somatic D_H segments. Ref. indicates the number used in this book. Classification of somatic D_H sequences was based on similarity of coding regions and coding frames. Boundaries between V_H and J_H genes were tentatively fixed at the 94th and 95th amino acid residues. N sequences (N_L at the boundaries between V_H and D_H, N_R at the boundaries between D_H and J_H) are also written. Since GG, GA, GAT and CC sequences for the 95th residue might be encoded by germ-line V_H genes [22], they are shown in italics. Since there have been no reports showing CGC, CTG, CCT, or ACC at the 94th residue in germ-line V_H genes [12, 22], they are shown in parentheses. Boundaries between D_H and J_H genes were assigned based on germ-line J_H sequences [14]. When nucleotides at the boundaries can be encoded by germ-line D_H and J_H genes, they are indicated in italics. When nucleotides possibly encoded by germ-line D_H genes are different from the corresponding germ-line D_H genes, they are underlined. Black triangles in the D_{FL16} family indicate that $D_{FL16.2}$ was used; in the other cases, $D_{FL16.1}$ was used. Bars in (D) indicate that there is no sequence in the somatic D_H region.

Table 1. Nucleotide difference of flanking regions between D_{FL16} and D_{SP2}[a)]

	Position from	to	N	M	K	K^c
5′-Flanking	2618	3060	445	121	0.2719	0.3377
3′-Flanking	3140	3401	266	65	0.2443	0.2956
Total			711	186	0.2616	0.3217

a) 5′ and 3′-Flanking sequences of D_H genes were compared. N is the number of sites compared between $D_{FL16.1}$ and $D_{SP2.2}$ [17]. Deletion of continuous two to nine nucleotides was assumed to have occurred as a single event. M is the number of sites showing a difference between D_{FL16} and D_{SP2}. K and K^c indicate nucleotide difference per site and difference corrected for multiple substitutions $K^c = -\frac{3}{4}\ln\left(1-\frac{4}{3}K\right)$ [18, 19], respectively. Using $K^c =$ 0.3217 for D_{FL16} and D_{SP2}, the divergence between rat and mouse would have occurred 17 million years ago [20], and using a K^c value for rat and mouse of 0.148 [21], the divergence date between D_{FL16} and D_{SP2} was estimated to be about 37 million years $\left(17 \times \frac{0.322}{0.148} = 37\right)$.

frames. The following characteristics were observed (a) $D_{FL16.1}$ is the most frequently (73/158) used D_H gene, (b) the codon frame I (TAC TAC GGT and TAC TAT GGT) encoding Tyr-Tyr-Gly is predominantly used in both D_{FL16} and D_{SP2} genes, 65/77 and 29/38, respectively; (c) in the cases where N sequences were not observed at the boundaries between D_H and J_H genes, 1 to 6 nucleotide-long redundancy frequently existed, that is, a few nucleotides such as CTAC can derive either from germ-line D_H or J_H. The third point may reflect the repair mechanism taking place after digestion of the ends of D_H and J_H genes with exonuclease. DNA polymerase and ligase might be involved in the joining process of the processed ends. Since DNA polymerase requires a primer for polymerization [15], the ends of the joined fragments should have complementary nucleotides to supply template and primer. These characteristics were already observed in Kurosawa and Tonegawas's study [4], although only 16 somatic sequences were available; now, they can be generalized in mouse somatic D_H sequences. Since virtually all of the somatic D_H sequences can be encoded by the 12 D_H genes, we concluded that there are only three D_H gene families in mouse genome.

D_{FL16} family has two members and D_{SP2} family has nine members [4]. It is quite obvious that the members belonging to each family were created by a gene duplication mechanism. Moreover, sequences of D_{FL16} and D_{SP2} are also homologous to each other, as shown in Fig. 3. The sequence similarity has been found not only in D_H genes themselves but also in the surrounding regions; therefore, it is likely that both gene families orginate from the same primordial D_H gene. Using the flanking sequences of both genes, we calculated the divergence date beteen D_{FL16} and D_{SP2} genes as described in Table 1, and concluded that D_{FL16} and D_{SP2} genes had diverged around 37 million years ago. Fig. 6 schematically shows the evolutional pathway that created a set of D_H genes in the mouse genome. A primordial D_H gene was duplicated around 37 million years ago. Mutations were introduced into both DNA fragments, resulting in D_{FL16} and D_{SP2} genes. Both

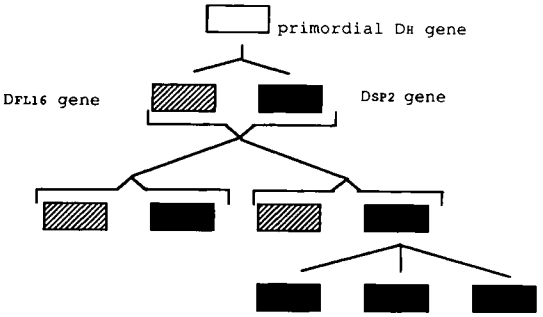

Figure 6. Phylogenetic relationship between D_{FL16} and D_{SP2} gene families. D_{FL16} and D_{SP2} genes diverged from a primordial D_H gene around 37 million years ago. Both genes were duplicated once more. After that, only the D_{SP2} gene was multiplied.

genes were duplicated once more. After that, only 5-kb fragments containing the D_{SP2} gene were multiplied several times.

The reason why $D_{FL16.1}$ is the most frequently used D_H gene is not clear. As long as the usage frequency of D_{FL16} and D_{SP2} was observed in D_H-J_H joinings, D_{SP2} and D_{FL16} genes were equally used [4, 10, 11]. Moreover, judging from the sequence observed in D_H-J_H joints [4, 11], not only the codon frame encoding Tyr-Tyr-Gly, but also the other codon frames were used. Selection might have occurred at the cellular level, not at the joining process. The reading frame of D_H regions has also been discussed by others [16], leading to essentially the same conclusion as ours.

We thank Drs. Y. Takagi, I. Ishiguro and K. Fujita for their encouragement. We are also grateful to M. Yasuda, C. Kato and T. Inoue for their technical assistance, and to Ms A. Nagata for preparing the manuscript.

Received June 11, 1989.

5 References

1 Tonegawa, S., *Nature* 1983. *302:* 575.
2 Sakano, H., Huppi, K., Heinrich, G. and Tonegawa, S., *Nature* 1979. *280:* 288.
3 Early, P., Huang, H., Davis, M., Calame, K. and Hood, L., *Cell* 1980. *19:* 981.
4 Kurosawa, Y. and Tonegawa, S., *J. Exp. Med.* 1982. *155:* 201.
5 Siebenlist, U., Ravetch, J. V., Korsmeyer, S., Waldmann, T. and Leder, P., *Nature* 1981. *294:* 631.
6 Ichihara, Y., Matsuoka, H. and Kurosawa, Y., *EMBO J.* 1988. *7:* 4141.
7 Southern, E. M., *J. Mol. Biol.* 1975. *98:* 503.
8 Sanger, F., Nicklen, S. and Coulson, A. R., *Proc. Natl. Acad. Sci. USA* 1977. *74:* 5463.
9 Wood, C. and Tonegawa, S., *Proc. Natl. Acad. Sci. USA* 1983. *80:* 3030.
10 Alt, F., Yancopoulos, G. D., Blackwell, T. K., Wood, C., Thomas, E., Boss, M., Coffman, R., Rosenberg, N., Tonegawa, S. and Baltimore, D., *EMBO J.* 1984. *3:* 1209.
11 Reth, M. G. and Alt, F. W., *Nature* 1984. *312:* 418.
12 Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. M. and Gottesman, K. S., 1987. *Sequences of Proteins of Immunological Interest.* US Dept Health and Human Services, Washington, DC.
13 Alt, F. W. and Baltimore, D., *Proc. Natl. Acad. Sci. USA* 1982. *79:* 418.

14 Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. and Tonegawa, S., *Nature* 1980. *286:* 676.
15 Goulian, M., *Proc. Natl. Acad. Sci. USA* 1968. *61:* 284.
16 Kaartinen, M. and Makela, O., *Immunol. Today* 1985. *6:* 324..
17 Kurosawa, Y., Von Boehmer, H., Haas, W., Sakano, H., Traunecker, A. and Tonegawa, S., *Nature* 1981. *290:* 565.
18 Jukes, T. H. and Cantor, C. R. in Munro, H. N. and Allison, J. B., (Eds), *Mammalian Protein Metabolism,* Academic Press, New York 1969, Vol. 2, pp. 21–132.
19 Kimura, M. and Ohta, T., *J. Mol. Evol.* 1972. *2:* 87.
20 Miyata, T., Hayashida, H., Kikuno, R., Hasegawa, M., Kobayashi, M. and Koike, K., *J. Mol. Evol.* 1982. *19:* 28.
21 Hayashida, H. and Miyata, T., *Proc. Natl. Acad. Sci. USA* 1983. *80:* 2671.
22 Rechavi, G., Bienz, B., Ram, D., Ben-Neriah, Y., Cohen, J. B., Zakut, R. and Givol, D., *Proc. Natl. Acad. Sci. USA* 1982. *79:* 4405.