J. Mol. Biol. (1996) **264**, 220–232

JMB

COMMUNICATION

Sequence and Evolution of the Human Germline V_λ Repertoire

Samuel C. Williams[1]†, Jean-Pol Frippiat[1,2]*†, Ian M. Tomlinson[1] Olga Ignatovich[1], Marie-Paule Lefranc[2] and Greg Winter[1,3]

[1]*MRC Centre for Protein Engineering, Hills Road Cambridge, CB2 2QH, UK*

[2]*Laboratoire d'ImmunoGénétique Moléculaire, LIGM, Institut de Génétique Moléculaire UMR 5535 CNRS Université Montpellier II 1919 Route de Mende BP 5051, 34033 Montpellier Cedex 1, France*

[3]*MRC Laboratory of Molecular Biology Hills Road, Cambridge CB2 2QH, UK*

**Corresponding author*

We recently completed a map of the human immunoglobulin lambda (IGL) locus on chromosome 22q11.2 and showed that the V_λ genes are arranged in three distinct clusters, each containing members of different V_λ families. We have now sequenced each of these V_λ genes and determined which are functional by comparison with the expressed repertoire. Our analysis indicates that there are approximately 30 functional V_λ genes, depending on the haplotype, that belong to ten V_λ families (five $V_\lambda1$, five $V_\lambda2$, eight $V_\lambda3$, three $V_\lambda4$, three $V_\lambda5$, one $V_\lambda6$, two $V_\lambda7$, one $V_\lambda8$, one $V_\lambda9$ and one $V_\lambda10$). V_λ genes related to the major human V_λ families ($V_\lambda1$, $V_\lambda2$ and $V_\lambda3$) predominate in species that express mainly lambda light chains.

Keywords: antibody; germline; immunoglobulin; lambda; light chain

The complete variable domain of an antibody is formed by the association of the heavy and light chain variable domains (V_H and V_L). Three polypeptide loops from the V_H domain and three polypeptide loops from the V_L domain come together to form the antigen binding site. The structural diversity in this region enables antibodies to bind almost any antigen. In the primary repertoire, this diversity is created by the somatic recombination of V_H (variable), D (diversity) and J_H (joining) genes for the V_H domain and V_L and J_L genes for the V_L domain. Antibodies are selected by binding to the antigen. The affinities are then improved by somatic hypermutation (the introduction of point mutations into the rearranged V_H and V_L genes) and further rounds of selection.

In humans, the V_H, D and J_H genes are located on chromosome 14q32.3 (Cox *et al*., 1982). In contrast, the V_L and J_L genes can be derived from either the kappa (κ) locus on chromosome 2p11–12 (Malcolm *et al*., 1982) or the lambda (λ) locus on chromosome 22q11.2 (de la Chapelle *et al*., 1983). Although the sequences of the human V_H (Tomlinson *et al*., 1992; Cook *et al*., 1994), D (S. J. Corbett *et al*., unpublished results), J_H (Ravetch *et al*., 1981), V_κ (Schäble & Zachau, 1993), J_κ (Hieter *et al*., 1982) and J_λ (Vasicek & Leder, 1990) genes are known, the analysis of the germline V_λ repertoire is incomplete.

We recently constructed a physical map of the human V_λ locus using cosmid and yeast artificial chromosome (YAC) clones (Frippiat *et al*., 1995) and showed that the V_λ families are organised in three separate clusters. Here, we set out to determine the sequences of these germline V_λ genes.

Sequencing the germline V_λ genes

Cosmid and YAC clones (described by Frippiat *et al*., 1995) containing a single member of a given V_λ family were amplified with family-specific primers (see the legend to Figure 1). Where clones contained more than one V_λ family member, genes were amplified from *Eco*RI fragments extracted from an agarose gel or were amplified using gene-specific primers (Figure 1). Amplification products were then sequenced directly or sub-cloned for

Corresponding address: MRC Centre for Protein Engineering, Hills Road, Cambridge, CB2 2QH, UK.

† These authors contributed equally to this work.

0022–2836/96/000000–00 $25.00/0

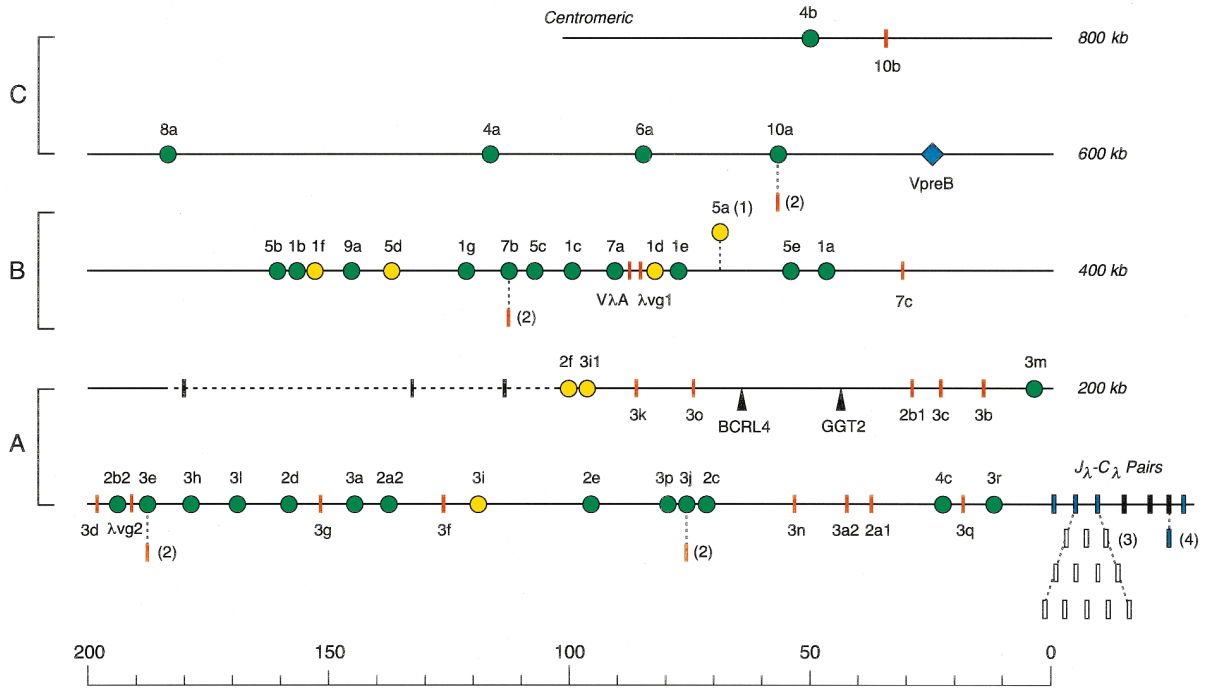

Figure 1. Map of the human immunoglobulin lambda locus. The map is based on Frippiat *et al*. (1995) (see Table 1) and runs left to right from gene 4b (top; centromeric) to the J_λ–C_λ region (bottom; telomeric), with distances from the J_λ1 gene shown in kilobases. The position of the VpreB gene is indicated by a blue diamond. The BCRL4 and GGT2 pseudogenes are indicated by arrows. Brackets A, B and C indicate the clustered organisation of the V_λ families. The transcriptional orientation of all V_λ genes so far examined is consistent with V-J joining occurring by a deletional mechanism (all are 5 to 3' with respect to the J_λ genes). Functional V_λ genes (seen expressed in functional V-J rearrangements) are depicted by green circles; V_λ genes with an open reading frame not seen in functional V-J rearrangements are depicted by yellow circles; pseudogenes (genes with stop codons or frame shifts) are depicted by vertical red bars. Four of the seven V_λ genes with open reading frames not seen in functional V-J rearrangements have genetic and/or structural defects that may prevent their rearrangement and/or functional expression. These include genes 1d (3' splice site AG replaced by AC), 2f (CACAGTG heptamer replaced by CATAGTG, see Akamatsu *et al*., 1994), 3i1 (3' splice site AG replaced by GG; no recognisable heptamer; residue 88C replaced by Y resulting in loss of disulphide bridge) and 5d (CACAGTG heptamer replaced by TACAGTG, see Akamatsu *et al*., 1994). Our failure to detect expressed counterparts of gene 5a in the λ cDNA library may be due to a polymorphic deletion of this gene in the four individuals from which the library was made (see below). The broken line between genes 7c and 2f indicates a region missing from our YAC clones due to instability (approximately 80 kb). This region contains three V_λ genes (shown as vertical black bars), with unknown sequences (Kawasaki *et al*., 1995), that are likely to be non-functional (see the text). Functional and non-functional J_λ–C_λ pairs are depicted by blue and black rectangles, respectively (Hieter *et al*., 1981; Dariavach *et al*., 1987; Udey & Blomberg, 1987; Vasicek & Leder, 1990; Combriato & Klobeck, 1991; Bauer & Blomberg, 1991). The following alternative haplotypes are shown: (1) the insertion/deletion of V_λ gene 5a (Frippiat *et al*., 1995); (2) pseudogenes replacing functional genes (see Table 1); (3) the insertion of additional unsequenced J_λ–C_λ pairs (shown as white rectangles; Taub *et al*., 1983); (4) the presence of a functional version of the J_λ6–C_λ6 pair (Dariavach *et al*., 1987). In addition to the mapped and sequenced genes shown here, we anticipate there being many diverged non-functional V_λ genes scattered throughout the IGL locus, ranging from those that only hybridise under low stringency washing conditions (the weakly hybridising genes; see the text and Frippiat *et al*., 1995) to vestigial sequences that do not hybridise at all.

PCR amplification of germline V_λ genes was performed according to Tomlinson *et al*. (1992) using the primers described by Williams & Winter (1993) and the following family-specific and gene-specific primers. **V_λ4** (VL4bckA CCC CCA AGC TTA TGA CTT GGA CCC CAC TCC T; VL4forA GGA ATT CTC ATC TGC CTG TGT CAC), **V_λ5** (VL5bckJPF GGA ATT CAA GCT TCT GCA GAT GGC CTG GAC TCC TCT C; Vλ3NON1 GGA ATT CGG TTT CTG TCT CAC TTC C), **V_λ6** (VL6bck1 GAA GGG CTG GTG GGA TCA G; Vλ3NON1 GGA ATT CGG TTT CTG TCT CAC TTC C), **V_λ10** (VL10forA GGA ATT CTT GAG TTT TTA TCT CAC TTC CCC; VL10bckA CCC CCA AGC TTC CAG TCT CCA AAC AGA GCT TCA GCA AGC ATA; VL10bckC CCC CCA AGC TTC TGG GTT GTA AAC CTG CTC TCA TGG CCC T), **2c** (VL2Vasbck CCC CCA AGC TTA GGT GAT GCC TCC AGG GAA GGG GCC ACA G; VL2046for GGA ATT CTC CTC ATT GAC TTA AAA C), **2f** (DPL14Cbck CCC CCA AGC TTG GTC TTC TGC TCC TGC TCC TCA GGG; Vλ2NON1 GGA ATT CGG TTT TGG TCT CAG TTC C), **3q** (VLN.1bck CCC CCA AGC TTG CAG GGT GTG GGG CTA TGG GAA TGA GAC CC; VLN.1for GGA ATT CAA GGG AGA GAA TGA CAT AGA TGG GAA GGG G), **3a2** (VLIII.2bck CCC CCA AGC TTA AGC CTC CTC TCC TGT CCT CTC TTG CAA GC; Vλ3NON1 GGA ATT CGG TTT CTG TCT CAC TTC C), **3a** (VL3A1bck CCC CCA AGC TTG TGG GAC TCC TGG AAA TGG GTC C; VL3A2for GGA ATT CTG TCT CCT TGT GTC ACT GTG).

All nucleotide sequencing was performed using the DyeDeoxy Terminator kit on an ABI 373A (Perkin Elmer). Sequence data were analysed using SeqEd (Perkin Elmer) and alignments were performed using the MacVector software package (Oxford Molecular).

Table 1. Assignment of germline V_λ sequences

Family[a]	Gene[b]	Sequence from this study[c]		Sequence from the literature[c]	Δ(N, P)[d]	O, P[e]
V_λ1 (B)	1a$^+$	**1a.11.2**	Z73653	DPL1(1), 1v1c2(2), 1v1L1a(2), 1v1c2c(2)	(0, 0)	O
	1e$^+$	**1e.10.2**	Z73656	**1v1042(7),** DPL8(1)	(0, 0)	O
				DPL7(1), Vλ1.2 (IGLV1S2)(13)	(2, 1)	O
				DPL6(1)	(3, 2)	O
	1d	**1d.8.3**	Z73655	**Vλ1.1(5), 1v1041(7)**, DPL4(1)	(0, 0)	O
	1c$^+$	**1c.10.2**	Z73654	DPL2(1), 1v1L1(7)	(0, 0)	O
	1g$^+$	**1g.400B5**	Z73663	DPL3(1), 1v122(7)	(0, 0)	O
	1f	**1f.366F5**	Z73662	**1v101(7),** DPL9(1)	(0, 0)	O
	1b$^+$	**1b.366F5**	Z73661	DPL5(1), 1v117d(10)	(0, 0)	O
				1v117(20), 1v119(7)	(2, 1)	O
V_λ2 (A)	2a1	**2a1.51E6**	Z73641	DPL15(1), 1v2120(22)	(0, –)	P
				ψV_λII.1(11)	(2, –)	P
	2c$^+$	**2c.118D9**	X97462			O
				1v2046(22)	(1, 1)	O
	2e$^+$	**2e.2.2**	Z73657			O
				DPL12(1)	(1, 0)	O
	2a2$^+$	**2a2.272A12**	Z73664	DPL11(1), 1v215.23(22)	(0, 0)	O
				1v2018(22)	(7, 4)	O
	2d$^+$	**2d.29D11**	Z73642	DPL13(1), 1v2132(22)	(0, 0)	O
				1v216.21(22)	(1, 1)	O
				1v2007(22)	(2, 2)	O
				1v2113(22)	(3, 2)	O
	2b2$^+$	**2b2.400B5**	Z73665			O
				Vλ2.1(IGLV2S1)(17, 18)	(1 + 1 Ins, 2)	O
				DPL10(1), 1v2066(22)	(1, 1)	O
	2b1	**2b1.22E4**	X97466			P
	2f	**2f.61E11**	Z73643	DPL14(1), 1v2011(22)	(0, 0)	O
				1v2110(22)	(1, 0)	O
				1v2031(22)	(2, 1)	O
V_λ3 (A)	3r$^+$	**3r.9C5**	Z73647	**V_λIII.1(11)**, DPL23(1), BHGL1(12), BHGL6(12)	(0, 0)	O
	3q	**3q.127E5**	X97468			P
				ψV_λN.1(11)	(1 + 1 Ins, –)	P
	3a2	**3a2.51E6**	X97465			P
				ψV_λIII.2(11)	(2 + 1 Del, –)	P
	3n	**3n.118D9**	X97470			P
	3j$^+$	**3j.118D9**	X97473			O
				IGLV3S6(18)	(6 + 5 Ins, 5)	O
				ψV_λ1(21)	(5 + 1 Del, –)	P
	3p$^+$	**3p.81A4**	X97464			O
				IGGLL295(15)	(4, 2)	O
	3i	**3i.2.2**	Z73658			O
	3f	**3f.119B4**	X97463			P
	3a$^+$	**3a.119B4**	X97471			O
	3g	**3g.29D11**	X97472	1v413(4)	(0, –)	P
	3l$^+$			**Vλ3.1(IGLV3S1)(3, 18)**, DPL16(1), 1v418(4), IGLV4BR(23)		O
	3h$^+$			**IGLV3S2(18)**		O
				1v318(4)	(4, 2)	O
	3e$^+$	3e.272A12	Z73666			O
				IGLV3S3P(18)	(6 + 5 Del + 1 Ins, –)	P
	3d			**IGLV3S4P(18)**		P
	3m$^+$	**3m.102D1**	X97474			O
				IGGLL150(15)	(3, 1)	O
	3b	**3b.57F5**	X97467			P
	3c	**3c.97H8**	Z73644			P
	3o	**3o.75H1**	Z73646			P
	3k	**3k.61E11**	X97469			P
				DPL17(1)	(1, 1)	P
	3i1	**3i1.61E11**	Z73645			O
V_λ4 (A)	4c$^+$	**4c.127E5**	Z73652	**V_λN.2(11)**, DPL24(1)	(0, 0)	O
(C)	4a$^+$	**4a.366F5**	Z73667			O
(C)	4b$^+$	**4b.68B6**	Z73648	IGLV8A1(16)	(0, 0)	O
				1v801(10)	(1, 0)	O
V_λ5 (B)	5e$^+$	**5e.366F5**	Z73672			O
	5a	**5a.366F5**	Z73668			O
	5c$^+$	**5c.366F5**	Z73670			O
		5c.400B5	Z73671		(2, 1)	O
	5d	**5d.75A1**	Z73649			O
	5b$^+$	**5b.366F5**	Z73669			O
V_λ6 (C)	6a$^+$	**6a.366F5**	Z73673	IGLV6S1(19), LV6SW-G(19), V_λVI-3.6(24)	(0, 0)	O
V_λ7 (B)	7c	**7c.11.2**	Z73660			P
	7a$^+$	**7a.2.3**	Z73659	**Vλ**7.1(5), DPL18(1), 4A(6)	(0, 0)	O
	7b$^+$	**7b.400B5**	Z73674	DPL19(1)	(0, 0)	O
				DPL20(1)	(0 + 1 Del, –)	P

continued overleaf

Table 1. *continued*

V_λ8 (C)	8a$^+$	**8a.88E1**	Z73650	DPL21(1), VL8(8), FL7(9), TL7(9), BL7(9)	(0, 0) (1, 1)	O O
V_λ9 (B)	9a$^+$	**9a.366F5**	Z73675	DPL22(1), 1v901m(10m) 1v901e(10e)	(0, 0) (1, 0)	O O
V_λ10 (C)	10a$^+$	**10a.872F9**	Z73676	 gV$_\lambda$X-4.4(14)	 (4, 2)	O O
	10b	**10b.4E7**	Z73651	 gV$_\lambda$X-5.5(14)	 (1 + 2 Del, –)	P P

[a] All genes are assigned to V_λ families: 1 to 7 (Chuchana *et al*., 1990), 8 (Winkler *et al*., 1992), 9 (Williams & Winter, 1993) or 10 (Stiernholm *et al*., 1994) (see Figure 3). The map cluster to which each family belongs is given in parentheses (see Figure 1). Note that although gene 4c (previously unassigned: Combriato & Klobeck, 1991; Williams & Winter, 1993; Frippiat *et al*., 1995) has now been assigned to the V_λ4 family by nucleotide sequence comparison (4c and 4b are 79.6% homologous; 4c and 4a are 76.4% homologous) it is not located in cluster C with the other V_λ4 genes (see Figure 3). The highly diverged pseudogene VλA (X14613; Alexandre *et al*., 1989; Chuchana *et al*., 1991) and the vestigial pseudogenes λvg1 (X99568; Alexandre *et al*., 1989) and λvg2 (X71351; Frippiat & Lefranc, 1994) cannot be assigned to one of the ten families and have therefore been excluded from this Table.

[b] Gene names are taken from Frippiat *et al*. (1995), with the addition of 2f, which resides on a 1.1 kb *Eco*RI fragment and corresponds to DPL14 (a weakly hybridising gene) and 4c, which resides on a 0.9 kb *Eco*RI fragment and corresponds to VλN.2 (Combriato & Klobeck, 1991) and DPL24 (Williams & Winter, 1993).

[+] Indicates that the gene has been seen expressed in a functional V–J rearrangement (see the text).

[c] Sequences from this study (EMBL data library accession numbers provided) give the gene name and the cosmid or YAC clone from which the sequence was isolated e.g. 1a.11.2 corresponds to the sequences of gene 1a isolated from cosmid clone 11.2. Mapped sequences from this study and the literature are shown in bold. Other sequences were assigned by homology to the closest mapped sequence (note that the sequence 3e.272A12 was amplified from YAC clone 272A12 and was assigned to 3e by homology to IGLV3S3P).

[d] For each gene, ΔN denotes the number of nucleotide differences between the sequences from the literature and our sequence. In addition, Del and Ins denote the total number of nucleotide deletions and insertions, respectively. ΔP denotes the number of amino acid differences for those sequences with open reading frames. In the case of gene 5c the differences between the two alleles are given. Identical sequences are listed on the same line.

[e] O denotes an open reading frame for the V exon; P denotes a pseudogene with stop codons or frame shifts in the V exon. Since the stop codons immediately preceding the heptamer sequence in genes 4c and 9a can be deleted by rearrangement they are shown as having open reading frames. Note that genes 3j, 3e and 7b have both functional and pseudogene alleles. In addition, the published sequence of 10a (gVλX-4.4; Stiernholm *et al*., 1994) has a frame shift in its 5' leader exon. In our cDNA library we have identified a functional 10a leader sequence indicating the presence of both functional and pseudogene alleles for this gene.

References are (1) Williams & Winter (1993); (2) Deftos *et al*. (1994a); (3) Frippiat *et al*. (1990); (4) Daley *et al*. (1992b); (5) Alexandre *et al*. (1989); (6) Anderson *et al*. (1984); (7) Daley *et al*. (1992a); (8) Winkler *et al*. (1992); (9) Lee *et al*. (1994); (10m) Deftos *et al*. (1994b) (manuscript version); (10e) Deftos *et al*. (1994b) (EMBL data library version); (11) Combriato & Klobeck (1991); (12) Harmer *et al*. (1995); (13) Bernard *et al*. (1990); (14) Stiernholm *et al*. (1994); (15) Fang *et al*. (1994); (16) Ch'ang *et al*. (1995); (17) Brockly *et al*. (1989); (18) Frippiat & Lefranc (1994); (19) Ch'ang *et al*. (1994); (20) Siminovitch *et al*. (1989); (21) Vasicek & Leder (1990); (22) Irigoyen *et al*. (1994); (23) Eulitz *et al*. (1995); (24) Stiernholm & Berinstein (1995).

sequencing, in which case the sequence was confirmed in two independent PCR reactions.

Some V_λ2 and V_λ3 genes could not be amplified using these primers. Here, cosmid or YAC DNA was digested with *Eco*RI, *Hin*dIII, or *Eco*RI plus *Hin*dIII, and the smallest fragments containing the gene were cloned into pBluescript II SK(+) (Stratagene). Further sub-cloning of these fragments using *Kpn*I, a conserved restriction site in framework 2, enabled the two halves of the V_λ gene to be sequenced.

When 3n1 (the most weakly hybridising band seen with the V_λ3 probe) was sub-cloned it was found to have the sequence of 2e, which resides on the same *Eco*RI fragment. 3n1 is therefore the result of cross-hybridisation of the V_λ3 probe with gene 2e and is not a genuine V_λ3 gene as previously thought. As a result, 3n1 has been deleted from our map. In addition, it was found that 3k and 3i1 had sequences identical with the adjacent genes 3k1 and 3a1, respectively (Frippiat *et al*., 1995), despite differences in the sizes of their corresponding restriction fragments. Both pairs were never found on the same cosmid or YAC clone (Frippiat *et al*., 1995) and only one pair is present in another map of the human V_λ locus (Kawasaki *et al*., 1995; and see below). It is therefore likely that they correspond to a single pair of genes (with restriction fragment length polymorphisms) rather than a local duplication. Consequently, 3k1 and 3a1 have been deleted from our map.

In total, we determined 49 different V_λ sequences from mapped clones (Table 1).

Comparison with published germline V_λ sequences and V_λ maps

Before this work, 20 germline V_λ sequences had been identified in phage or cosmid clones (Alexandre *et al*., 1989; Brockly *et al*., 1989; Frippiat *et al*., 1990; Combriato & Klobeck, 1991; Daley *et al*., 1992a; Frippiat & Lefranc, 1994) and positioned on our map (Frippiat *et al*., 1995). These sequences (with the exception of the highly diverged pseudogenes VλA, λvgl and λvg2, which could not be assigned to one of the ten V_λ families) are listed next to their corresponding V_λ genes in Table 1. An additional 67 germline V_λ sequences from the literature (including 24 of the 25 amplified from the donor DP; Williams & Winter, 1993) were assigned on the basis of sequence homology to the closest mapped gene (Table 1). In general, sequences from independent studies are identical or differ by one or two nucleotide substitutions. Thus, allelic

polymorphism is limited, as with the V_κ (Cox *et al*., 1994) and V_H (Cook & Tomlinson, 1995) genes.

The only germline sequences that could not be assigned to the map were the $V_\lambda 8$ sequence FL6/TL6 (Lee *et al*., 1994) and the $V_\lambda 2$ sequences DPL25 (Williams & Winter, 1993) and lv216.376 (Irigoyen *et al*., 1994). FL6/TL6 is now known to be located outside the functional locus (J.-P. F., unpublished results). The pseudogenes DPL25 and lv216.376 are probably alleles of a gene located in the region of instability in our YAC clones (see below). All human germline V_λ sequences are available in the V BASE sequence directory (http://www.mrc-cpe.cam.ac.uk/imt-doc/vbase-home-page.html or from I. M. T., imt@mrc-lmb.cam.ac.uk) or in IMGT, the ImMunoGeneTics database (http://imgt.cnusc.fr:8104 or http://www.ebi.ac.uk/imgt or from M.-P. L., lefranc@ligm.crbm.cnrs-mop.fr).

After publication of our map of the V_λ locus (Frippiat *et al*., 1995) another map was published independently by Kawasaki *et al*. (1995). Comparison of the two maps indicates that the dimensions and the number of V_λ genes are generally in good agreement. The main difference is a deletion of 80 kb in our map (marked as a broken line in Figure 1) that corresponds to an instability area in the YAC clones (see Frippiat *et al*., 1995). This region contains three V_λ genes (shown as vertical black bars in Figure 1) that are likely to be non-functional (see below). Other differences between the two maps relate to the composition of the probes used and the stringency of the washing conditions. We used probes from each of the ten functional V_λ families and washed at high stringency (Frippiat *et al*., 1995). Since we were interested in the functional genes, we have now mapped and sequenced all the V_λ genes that hybridised under these conditions (Table 1 and Figure 1). By contrast, Kawasaki *et al*. (1995) used four V_λ probes, and washed at low stringency, thereby identifying several V_λ genes that we also could detect only at low stringency (Frippiat *et al*., 1995). We now know that these ''weakly hybridising'' genes are not expressed or make an insignificant contribution to the expressed repertoire (see below).

Comparison of the germline and expressed repertoires reveals the presence of 30 functional V_λ genes

In total, there are 37 V_λ genes that have alleles with open reading frames (shown as circles in Figure 1). The amino acid translation and nucleotide sequences of these V_λ genes are shown in Figure 2(a) and (b), respectively.

To determine which of the 37 V_λ genes with open reading frames are functional, 335 rearranged V_λ sequences from the EMBL data library (Release 44) and 378 from a human λ cDNA library (made from four individuals, containing a total of 7600 clones; O. I. *et al*., unpublished results) were assigned to their closest germline counterparts. Expressed V-J rearrangements (with no stop codons or frame shifts in the leader peptide or in the V or C regions) were found for 30 V_λ genes with open reading frames. These functional genes are depicted as green circles in Figure 1. Probing the λ cDNA library with oligonucleotides for the remaining seven genes with open reading frames (1d, 1f, 2f, 3i, 3i1, 5a and 5d) indicated that none was expressed (see the legend to Figure 2). These genes are shown as yellow circles in Figure 1. Since four of the functional genes (3j, 3e, 7b and 10a) have non-functional alleles (see Table 1 and Figure 1), the size of the functional V_λ repertoire will differ between individuals depending on the haplotype.

To exclude the expression of additional V_λ genes, ten weakly hybridising genes from the J_λ-distal region of the locus (cluster C; see Frippiat *et al*., 1995) were cloned in pBluescript II SK(+) and used to probe the λ cDNA library. No cDNA sequence was identified that corresponds to these genes (Figure 2). In addition, we sequenced all the λ cDNA clones from two individuals that did not hybridise with probes for any of the ten V_λ families (a total of 154 clones). We were unable to identify any new genes (identical sequences in both individuals): all cDNA sequences corresponded to known V_λ genes with high levels of somatic mutation.

Evolution of the human germline V_λ repertoire

To investigate the evolution of the human V_λ repertoire we constructed the family tree shown in Figure 3. This includes the 30 functional human V_λ genes, the three BALB/c mouse V_λ genes, and the human and mouse VpreB genes (see the legend to Figure 2 for references). The tree confirms the family designations of Frippiat *et al*. (1995) with the inclusion of gene 4c in the $V_\lambda 4$ family (Table 1). The number of functional human V_λ genes in each family is therefore as follows: five $V_\lambda 1$, five $V_\lambda 2$, eight $V_\lambda 3$, three $V_\lambda 4$, three $V_\lambda 5$, one $V_\lambda 6$, two $V_\lambda 7$, one $V_\lambda 8$, one $V_\lambda 9$, one $V_\lambda 10$. These V_λ genes are arranged in three distinct clusters in the lambda locus (Frippiat *et al*., 1995): cluster A (nearest the J_λ genes) contains the $V_\lambda 2$ and $V_\lambda 3$ families and one member of the $V_\lambda 4$ family (4c), cluster B (in the middle of the V_λ locus) contains the $V_\lambda 1$, $V_\lambda 5$, $V_\lambda 7$ and $V_\lambda 9$ families and cluster C (furthest from the J_λ genes) contains the two other members of the $V_\lambda 4$ family (4a and 4b) and the $V_\lambda 6$, $V_\lambda 8$ and $V_\lambda 10$ families.

We also classified the CDR1 and CDR2 loops of each sequence according to the main-chain conformations, or ''canonical'' structures, they are likely to encode (Chothia & Lesk, 1987; see the legend to Figure 3 for details). Human V_λ genes encode nine different combinations of canonical structures for CDR1 and CDR2. In contrast, mouse V_λ genes encode just two (see Figures 2 and 3).

We find that the human V_λ genes most closely related to mouse genes are from the least frequently expressed families ($V_\lambda 4$, $V_\lambda 5$, $V_\lambda 7$, $V_\lambda 8$ and $V_\lambda 9$; Combriato & Klobeck, 1991; O. I. *et al*., unpublished results). They also encode the same canonical

			FR1	CDR1	FR2	CDR2	FR3	CDR3
			2	3	4	5	6 7 8	9
	CDR1-2	Gene	123456789123456789012 3	45678901abc234	567890123456789	01abcde23456	789012345678ab901234567890123456789	9012345abcde
Vλ1	13-7(A)	1a+	QSVLTQPPSVSEAPRQRVTISC	SGSSSNIGNN-AVN	WYQQLPGKAPKLLIY	YD-----DLLPS	GVSDRFSGSKSG--TSASLAISGLQSEDEADYYC	AAWDDSLNG
	14-7(A)	1e+	QSVLTQPPSVSGAPGQRVTISC	TGSSSNIGAGYDVH	WYQQLPGTAPKLLIY	GN-----SNRPS	GVPDRFSGSKSG--TSASLAITGLQAEDEADYYC	QSYDSSLSG
	13-7(A)	1d	QSVLTQPPSVSAAPGQKVTISC	SGSSSDMGNY-AVS	WYQQLPGTAPKLLIY	EN-----NKRPS	GIPDRFSGSKSG--TSATLGITGLWPEDEADYYC	LAWDTSPRA
	13-7(A)	1c+	QSVLTQPPSASGTPGQRVTISC	SGSSSNIGSN-TVN	WYQQLPGTAPKLLIY	SN-----NQRPS	GVPDRFSGSKSG--TSASLAISGLQSEDEADYYC	AAWDDSLNG
	13-7(A)	1g+	QSVLTQPPSASGTPGQRVTISC	SGSSSNIGSN-YVY	WYQQLPGTAPKLLIY	RN-----NQRPS	GVPDRFSGSKSG--TSASLAISGLRSEDEADYYC	AAWDDSLSG
	14-7(A)	1f	QSVLTQPPSVSGAPGQRVTISC	TGSSSNIGAGYVVH	WYQQLPGTAPKLLIY	GN-----SNRPS	GVPDQFSGSKSG--TSASLAITGLQSEDEADYYC	KAWDNSLNA
	13-7(A)	1b+	QSVLTQPPSVSAAPGQKVTISC	SGSSSNIGNN-YVS	WYQQLPGTAPKLLIY	DN-----NKRPS	GIPDRFSGSKSG--TSATLGITGLQTGDEADYYC	GTWDSSLSA
Vλ2	14-7(A)	2c+	QSALTQPPSASGSPGQSVTISC	TGTSSDVGGYNYVS	WYQQHPGKAPKLMIY	EV-----SKRPS	GVPDRFSGSKSG--NTASLTVSGLQAEDEADYYC	SSYAGSNNF
	14-7(A)	2e+	QSALTQPRSVSGSPGQSVTISC	TGTSSDVGGYNYVS	WYQQHPGKAPKLMIY	DV-----SKRPS	GVPDRFSGSKSG--NTASLTISGLQAEDEADYYC	CSYAGSYTF
	14-7(A)	2a2+	QSALTQPASVSGSPGQSITISC	TGTSSDVGGYNYVS	WYQQHPGKAPKLMIY	EV-----SNRPS	GVSNRFSGSKSG--NTASLTISGLQAEDEADYYC	SSYTSSSTL
	14-7(A)	2d+	QSALTQPPSVSGSPGQSVTISC	TGTSSDVGSYNRVS	WYQQPPGTAPKLMIY	EV-----SNRPS	GVPDRFSGSKSG--NTASLTISGLQAEDEADYYC	SLYTSSSTF
	14-7(A)	2b2+	QSALTQPASVSGSPGQSITISC	TGTSSDVGSYNLVS	WYQQHPGKAPKLMIY	EV-----SKRPS	GVSNRFSGSKSG--NTASLTISGLQAEDEADYYC	CSYAGSSTF
	14-7(A)	2f	QSALTQPPFVSGAPGQSVTISC	TGTSSDVGDYDHVF	WYQKRLSTTSRLLIY	NV-----NTRPS	GISDLFSGSKSG--NMASLTISGLKSEVEANYHC	SLYSSSYTF
Vλ3	11-7	3r+	SYELTQPPSVSVSPGQTASITC	SG-DK-LGDK-YAC	WYQQKPGQSPVLVIY	QD-----SKRPS	GIPERFSGSNSG--NTATLTISGTQAMDEADYYC	QAWDSSTA
	11-7	3j+	SYELTQPLSVSVALGQTARITC	GG-NN-IGSK-NVH	WYQQKPGQAPVLVIY	RD-----SNRPS	GIPERFSGSNSG--NTATLTISRAQAGDEADYYC	QVWDSSTA
	11-7	3p+	SYELTQPPSVSVSPGQTARITC	SG-DA-LPKK-YAY	WYQQKSGQAPVLVIY	ED-----SKRPS	GIPERFSGSSSG--TMATLTISGAQVEDEADYYC	YSTDSSGNH
	11-7	3i	SYELTQPHSVSVATAQMARITC	GG-NN-IGSK-AVH	WYQQKPGQDPVLVIY	SD-----SNRPS	GIPERFSGSNPG--NTTTLTISRIEAGDEADYYC	QVWDSSSDH
	11-7	3a+	SYELTQPPSVSVSLGQMARITC	SG-EA-LPKK-YAY	WYQQKPGQFPVLVIY	KD-----SERPS	GIPERFSGSSSG--TIVTLTISGVQAEDEADYYC	LSADSSGTY
	11-7	3l+	SSELTQDPAVSVALGQTVRITC	QG-DS-LRSY-YAS	WYQQKPGQAPVLVIY	GK-----NNRPS	GIPDRFSGSSSG--NTASLTITGAQAEDEADYYC	NSRDSSGNH
	11-7	3h+	SYVLTQPPSVSVAPGKTARITC	GG-NN-IGSK-SVH	WYQQKPGQAPVLVIY	YD-----SDRPS	GIPERFSGSNSG--NTATLTISRVEAGDEADYYC	QVWDSSSDH
	11-7	3e+	SYELTQLPSVSVSPGQTARITC	SG-DV-LGEN-YAD	WYQQKPGQAPELVIY	ED-----SERYP	GIPERFSGSTSG--NTTTLTISRVLTEDEADYYC	LSGDEDN
	11-7	3m+	SYELMQPPSVSVSPGQTARITC	SG-DA-LPKQ-YAY	WYQQKPGQAPVLVIY	KD-----SERPS	GIPERFSGSSSG--TTVTLTISGVQAEDEADYYC	QSADSSGTY
	11-7	3i1	SSGPTQVPAVSVALGQMARITC	QG-DS-MEGS-YEH	WYQQKPGQAPVLVIY	DS-----SDRPS	RIPERFSGSKSG--NTTTLTITGAQAEDEADYYY	QLIDNHA
Vλ4	12-11	4c+	LPVLTQPPSASALLGASIKLTC	TLSSEHSTY--TIE	WYQQRPGRSPQYIMK	VKS-DGSHSKGD	GIPDRFMGSSSG--ADRYLTFSNLQSDDEAEYHC	GESHTIDGQVG*
	12-11	4a+	QPVLTQSSSASASLGSSVKLTC	TLSSGHSSY--IIA	WHQQQPGKAPRYLMK	LEG-SGSYNKGS	GVPDRFSGSSSG--ADRYLTISNLQLEDEADYYC	ETWDSNT
	12-11	4b+	QLVLTQSPSASASLGASVKLTC	TLSSGHSSY--AIA	WHQQQPEKGPRYLMK	LNS-DGSHSKGD	GIPDRFSGSSSG--AERYLTISSLQSEDEADYYC	QTWGTGI
Vλ5	14-11	5e+	QPVLTQPPSSSASPGESARLTC	TLPSDINVGSYNIY	WYQQKPGSPPRYLLY	YYS-DSDKGQGS	GVPSRFSGSKDASANTGILLISGLQSEDEADYYC	MIWPSNAS
	14-11	5a	QPVLTQPTSLSASPGASARFTC	TLRSGINVGTYRIY	WYQQKPGSLPRYLLR	YKS-DSDKQQGS	GVPSRFSGSKDASTNAGLLLISGLQSEDEADYYC	AIWYSSTS
	14-11	5c+	QAVLTQPASLSASPGASASLTC	TLRSGINVGTYRIY	WYQQKPGSPPQYLLR	YKS-DSDKQQGS	GVPSRFSGSKDASANAGILLISGLQSEDEADYYC	MIWHSSAS
	14-11	5d	QPVLTQPTSLSASPGASARLTC	TLRSGINLGSYRIF	WYQQKPESPPRYLLS	YYS-DSSKHQGS	GVPSRFSGSKDASSNAGILVISGLQSEDEADYYC	MIWHSSAS
	14-11	5b+	QPVLTQPSSHSASSGASVRLTC	MLSSGFSVGDFWIR	WYQQKPGNPPRYLLY	YHS-DSNKGQGS	GVPSRFSGSNDASANAGILRISGLQPEDEADYYC	GTWHSNSKT
Vλ6	13-7(B)	6a+	NFMLTQPHSVSESPGKTVTISC	TRSSGSIASN-YVQ	WYQQRPGSSPTTVIY	ED-----NQRPS	GVPDRFSGSIDSSSNSASLTISGLKTEDEADYYC	QSYDSSN
Vλ7	14-7(B)	7a+	QTVVTQEPSLTVSPGGTVTLTC	ASSTGAVTSGYYPN	WFQQKPGQAPRALIY	ST-----SNKHS	WTPARFSGSLLG--GKAALTLSGVQPEDEAEYYC	LLYYGGAQ
	14-7(B)	7b+	QAVVTQEPSLTVSPGGTVTLTC	GSSTGAVTSGHYPY	WFQQKPGQAPRTLIY	DT-----SNKHS	WTPARFSGSLLG--GKAALTLSGAQPEDEAEYYC	LLSYSGAR
Vλ8	14-7(B)	8a+	QTVVTQEPSFSVSPGGTVTLTC	GLSSGSVSTSYYPS	WYQQTPGQAPRTLIY	ST-----NTRSS	GVPDRFSGSILG--NKAALTITGAQADDESDYYC	VLYMGSGI
Vλ9	12-12	9a+	QPVLTQPPSASASLGASVTLTC	TLSSGYSNY--KVD	WYQQRPGKGPRFVMR	VGTGGIVGSKGD	GIPDRFSVLGSG--LNRYLTIKNIQEEDESDYHC	GADHGSGSNFV*
Vλ10	13-7(C)	10a+	QAGLTQPPSVSKGLRQTATLTC	TGNSNNVGNQ-GAA	WLQQHQGHPPKLLSY	RN-----NNRPS	GISERLSASRSG--NTASLTITGLQPEDEADYYC	SAWDSSLSA
Mouse	12-11	Vλx	QLVLTQSSSASFSLGASAKLTC	TLSSQHSTY--TIE	WYQQQPLKPPKYVME	LKK-DGSHSTGD	GIPDRFSGSSSG--ADRYLSISNIQPEDEAIYIC	GVGDTIKEQFV*
	14-7(B)	Vλ1	QAVVTQESALTTSPGETVTLTC	RSSTGAVTTSNYAN	WVQEKPDHLFTGLIG	GT-----NNRAP	GVPARFSGSLIG--DKAALTITGAQTEDEAIYFC	ALWYSNHF
	14-7(B)	Vλ2	QAVVTQESALTTSPGGTVILTC	RSSTGAVTTSNYAN	WVQEKPDHLFTGLIG	GT-----SNRAP	GVPVRFSGSLIG--DKAALTITGAQTEDDAMYFC	ALWYSTHF
VpreB	14-11	Human VpreB	QPVLHQPPAMSSALGTTIRLTC	TLRNDHDIGVYSVY	WYQQRPGHPPRFLLR	YFS-QSDKSQGP	QVPPRFSGSKDVARNRGYLSISELQPEDEAMYYC	AMGARSSEKEER
	14-11	Mouse VpreB1	QPMVHQPPLASSSLGATIRLSC	TLSNDHNIGIYSIY	WYQQRPGHPPRFLLR	YFS-HSDKHQGP	DIPPRFSGSKDTTRNLGYLSISELQPEDEAVYYC	AVGLRSQEKKRM
	14-11	Mouse VpreB2	QPMVHQPPSASSSLGATIRLSC	TLSNDHNIGIYSIY	WYQQRPGHPPRFLLR	YFS-HSDKHQGP	DIPPRFSGSKDTARNLGYLSISELQPEDEAVYYC	AVGLRSHEKKRM

Figure 2(a) *legend on page* 229

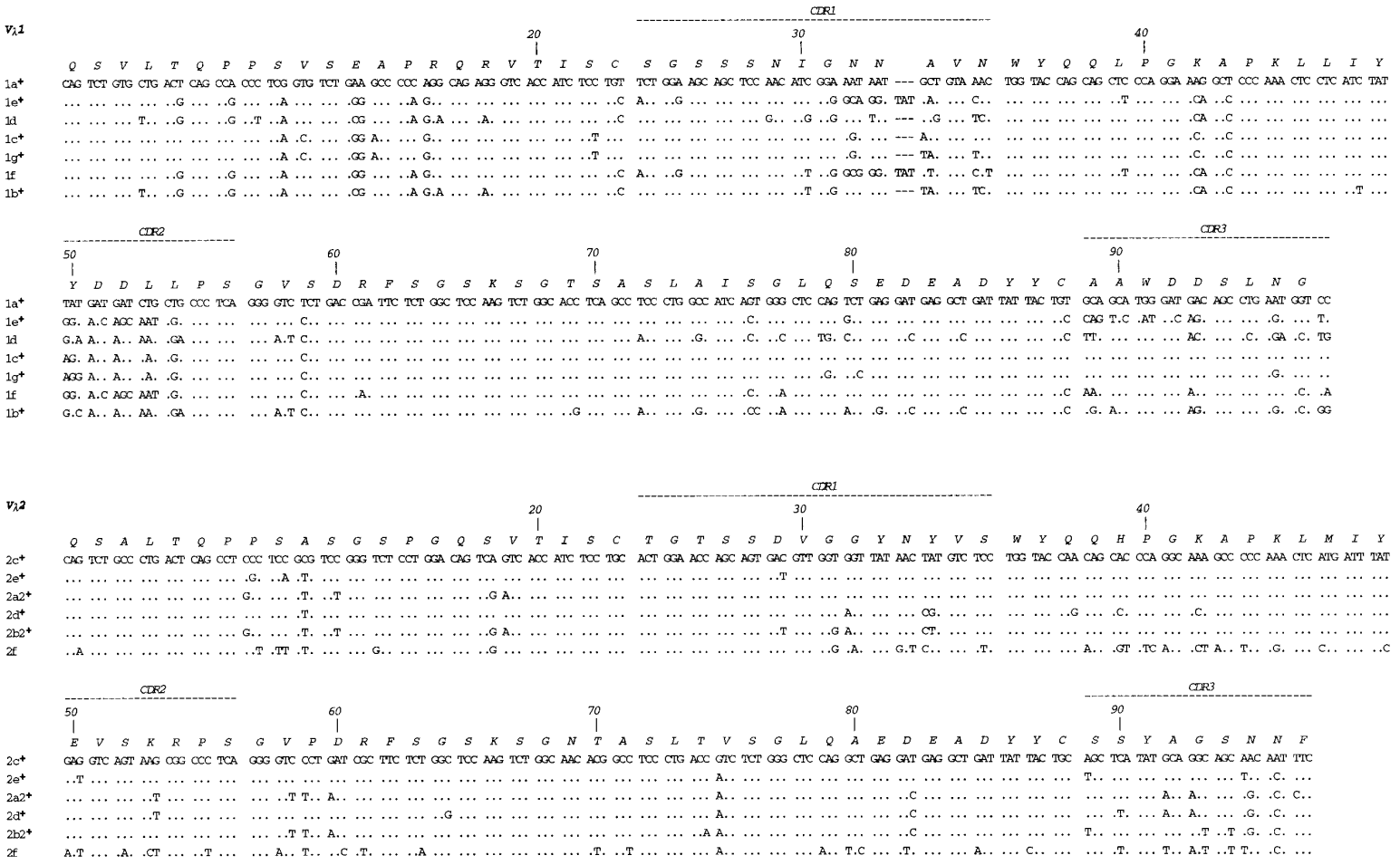

Figure 2(b) *legend on page* 229

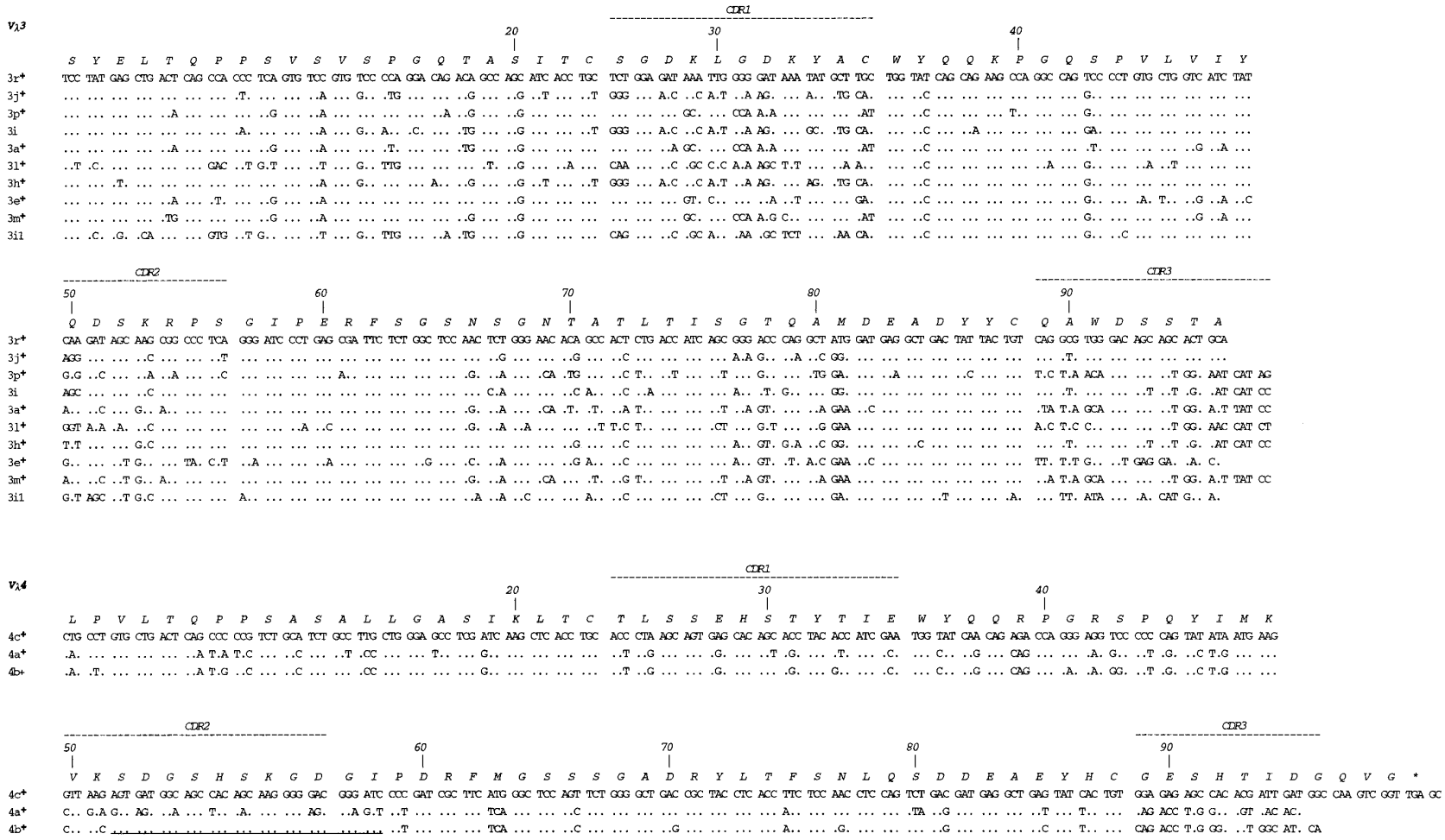

Figure 2(b) *legend on page* 229

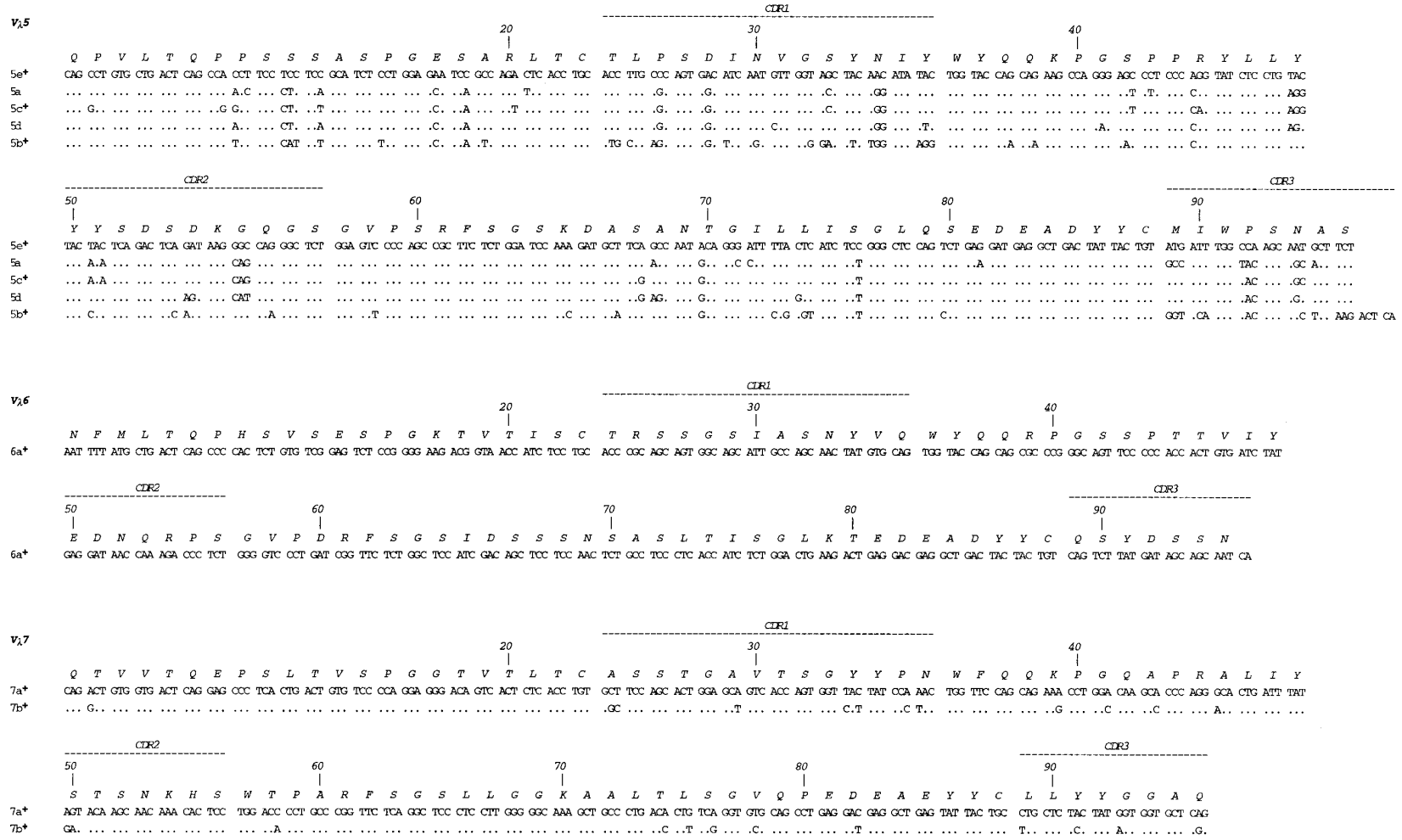

Figure 2(b) *legend on page* 229

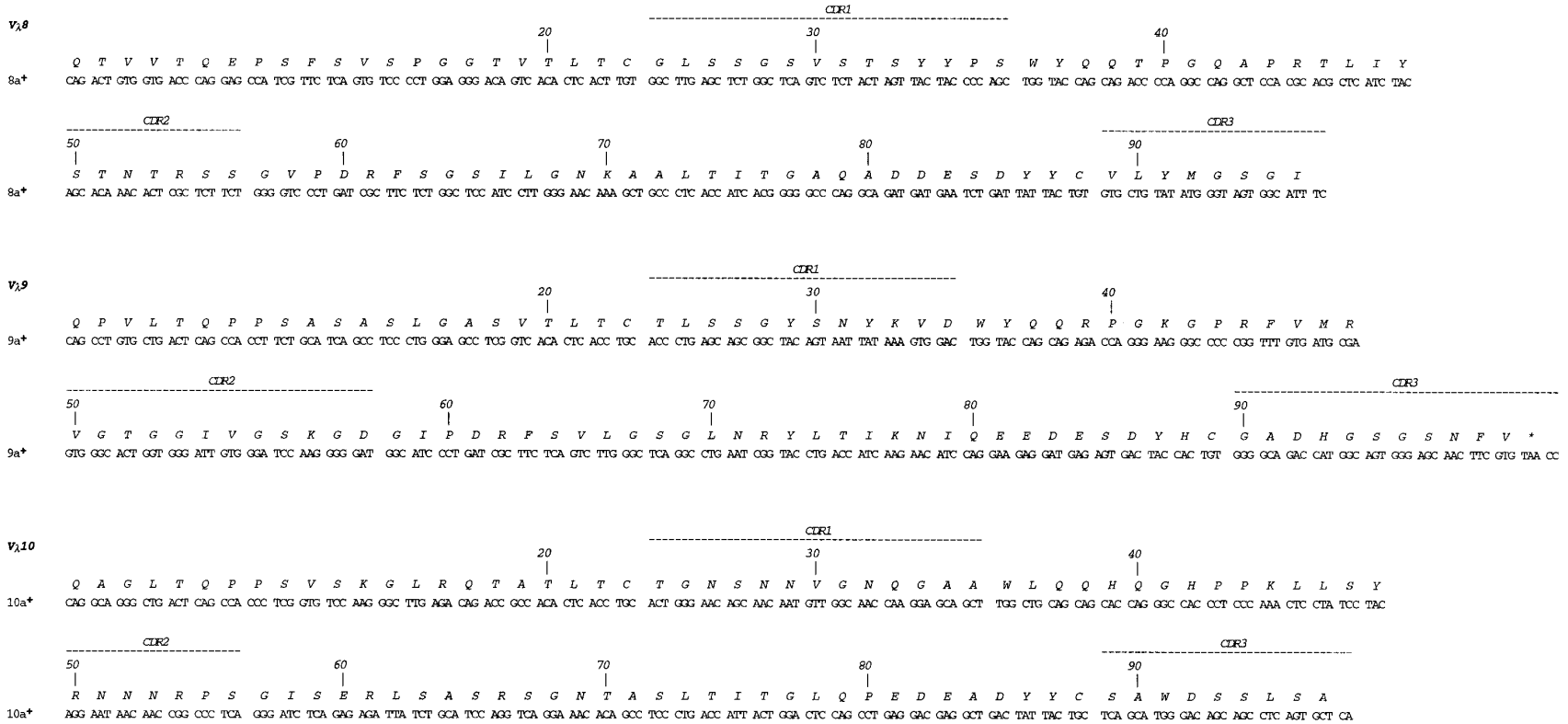

Figure 2(b)

Figure 2. Sequences of the V_λ genes with open reading frames. (a) Amino acid sequences. For each V_λ gene, the first named sequence in Table 1 is given: $^+$ next to the gene name indicates that it has been seen in the expressed repertoire. In addition, the sequences of the three BALB/c mouse V_λ genes and human and mouse VpreB genes are shown. EMBL accession numbers and references for these sequences are: Mouse Vλ1 (X58409), Bernard *et al*. (1978); Mouse Vλ2 (X58412), Tonegawa *et al*. (1978); Mouse Vλx (D38129), Sanchez *et al*. (1990); Human VpreB (M34927), Bauer *et al*. (1988); Mouse VpreB1 (X05557) and VpreB2 (X05563), Kudo & Melchers (1987). CDRs are labelled according to Kabat *et al*. (1991). The column *CDR1-2* gives the canonical structure class for each sequence (see the legend to Figure 3 for details). Sequences are numbered according to Kabat *et al*. (1991), except for CDR1, CDR2 and FR3, where pad characters have been inserted according to comparisons of known structures.

(b) Nucleotide sequences. Numbering is as described in (a). A translation of the first sequence in each family is given in italics. Dots represent identical nucleotides and dashes correspond to deletions. The underlined region in gene 4b is likely to be the result of a recent germline gene conversion event from gene 4c.

To exclude the expression of genes 1d, 1f, 2f, 3i, 3i1, 5a and 5d, the human λ cDNA library (O. I. *et al*., unpublished results) was probed with specific oligonucleotides according to the protocol described by Tomlinson *et al*. (1992). Oligonucleotide sequences (5' to 3') were 1d (TCA GGC CAG AGG CCA GT), 1f (TCA GAC TGG AGT CCA GT), 2f (GCC CCG GAC ACA AAA GG), 3i (GGC CAT CTG TGC TGT GG), 3i1 (TGA GTT GGC CCA GAG GA), 5a/5d (CTG GCT GAT GCT CCA GGA GAT, this oligonucleotide also hybridises to the functional gene 5c). All potentially positive cDNA clones were sequenced and none was found to correspond to these genes. Probing with the weakly hybridising genes was performed as follows: *Eco*RI, *Hin*dIII or *Eco*RI/*Hin*dIII restriction fragments from pBluescript II SK(+) were used, in turn, to probe the λ cDNA library according to the hybridisation protocol and high stringency washing conditions described by Frippiat *et al*. (1995). No positive cDNA clone was identified.

Amino acid and nucleotide alignments of the leader sequences for the functional Vλ genes are available in the VBASE directory (see the text for address).

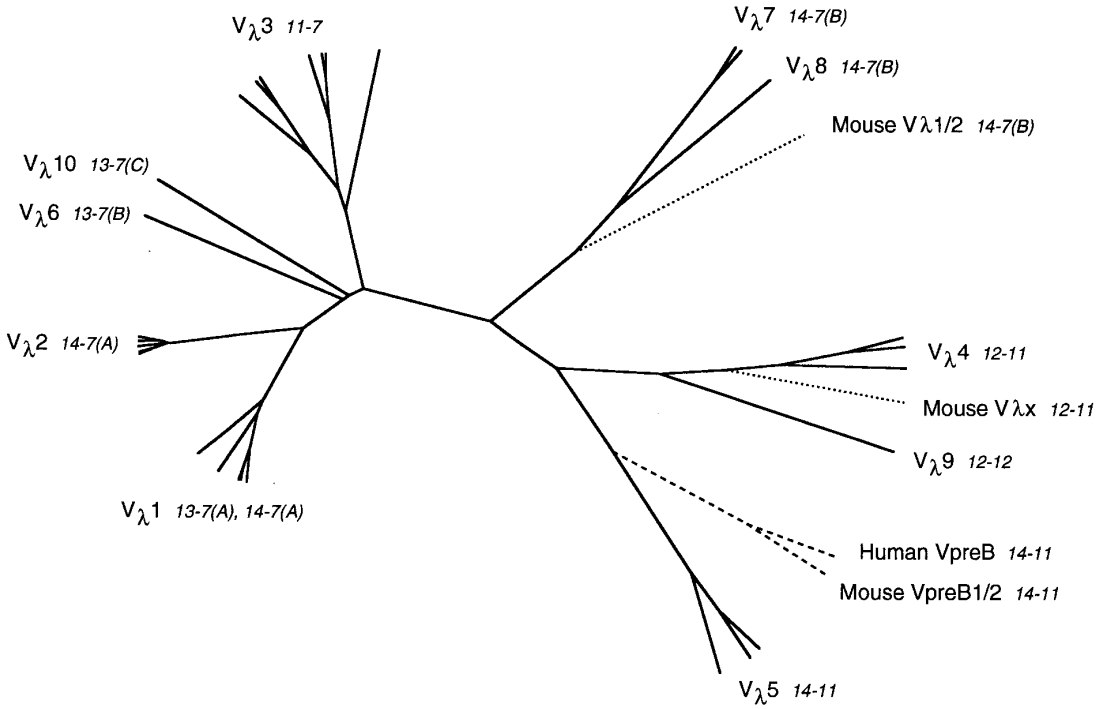

Figure 3. Evolution of the human V_λ repertoire. The nucleotide sequences of the 30 functional human V_λ genes (continuous black lines), the BALB/c mouse V_λ genes (dotted lines) and the human and mouse VpreB genes (broken lines) were used with the software PHYLIP (Felsenstein, 1989) to construct an unrooted family tree (see Figure 2 for references). CDR1 and CDR2 canonical structure classes are given in italics next to family names. Canonical structures are determined by the loop length and the identity of key structural determining residues involved in the packing of the loop (Chothia & Lesk, 1987; Williams & Winter, 1993; Wu & Cygler, 1993). The canonical structure classes are given in italics: this gives the length of CDR1 and CDR2 separated by a hyphen. Combinations that have the same loop lengths but encode different structures are distinguished by the letters A, B or C in parentheses. For example, comparison of key residues involved in 14 amino acid CDR1 loops (Chothia & Lesk, 1987; Wu & Cygler, 1993) indicates that there are two distinct canonical structures for this length: one encoded by the human $V_\lambda1$ and $V_\lambda2$ families, labelled 14-7(A), the other encoded by the mouse Vλ1 and Vλ2 genes and human $V_\lambda7$ and $V_\lambda8$ families, labelled 14-7(B). In addition, there are three different canonical structure combinations for the CDR1-CDR2 length 13-7: 13-7(B) differs from 13-7(A) in the L1 loop, whereas 13-7(C) differs from 13-7(A) in the L2 loop.

structures (14–7(B) by the human $V_\lambda7$ and $V_\lambda8$ families and the mouse Vλ1 and Vλ2 genes; 12–11 by the human $V_\lambda4$ family and mouse Vλx gene; 14–11 by the human $V_\lambda5$ family and the human and mouse VpreB genes). In contrast, the largest and most frequently expressed human V_λ families, $V_\lambda1$, $V_\lambda2$ and $V_\lambda3$ (Combriato & Klobeck, 1991; O. I. *et al*., unpublished results), are very different from the mouse genes in both sequence and structure (Figure 3). Comparisons with species that express predominantly lambda light chains (chicken, Reynaud *et al*., 1987; cow, Sinclair *et al*., 1995; horse, Home *et al*., 1992; and sheep, Reynaud *et al*., 1991) indicate that their V_λ genes are most closely related to members of the human $V_\lambda1$, $V_\lambda2$ or $V_\lambda3$ families. Furthermore, cow, horse and sheep V_λ genes appear to encode the same loop structures as these families (chicken and horse encode several new structures not seen in humans or mice). It therefore seems that BALB/c mice lack the major V_λ families and are forced to use a minor set of V_λ genes. This may explain the overwhelming presence of kappa chains in the mouse antibody repertoire (>95%; McIntire & Rouse, 1970).

Acknowledgements

We thank Fiona Sait, Samantha Barre and Erik Sonnhammer for their help with parts of this project. This work was supported by the UK Medical Research Council, the Centre National de la Recherche Scientifique, the Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche and the Groupement de Recherche et d'Etude sur les Génomes. J.-P. F. was supported by the Association pour la Recherche sur le Cancer and by the Société Française d'Hématologie. I. M. T. was supported by Trinity College, Cambridge, UK. O. I. was supported by Darwin Trust of Edinburgh University, UK and an Overseas Research Students Award.

References

Akamatsu, Y., Tsurushita, N., Nagawa, F., Matsuoka, M., Okazaki, K., Imai, M. & Sakano, H. (1994). Essential residues in V(D)J recombination signals. *J. Immunol.* **153**, 4520–4529.

Alexandre, D., Chuchana, P., Brockly, F., Blancher, A., Lefranc, G. & Lefranc, M.-P. (1989). First genomic sequence of a human Ig variable λ gene belonging to subgroup I. Functional genes, pseudogenes and

vestigial sequences are interspersed in the IGLV locus. *Nucl. Acids Res.* **17**, 3975.

Anderson, M. L. M., Szajnert, M. F., Kaplan, J. C., McColl, L. & Young, B. D. (1984). The isolation of a human Ig V_λ lambda gene from a recombinant library of chromosome 22 and estimation of its copy number. *Nucl. Acids Res.* **12**, 6647–6661.

Bauer, S. R., Kudo, A. & Melchers, F. (1988). Structure and pre-B lymphocyte restricted expression of the VpreB gene in humans and conservation of its structure in other mammalian species. *EMBO J.* **7**, 111–116.

Bauer, T. R., Jr & Blomberg, B. (1991). Recharacterization of JC lambda 6 and identification of a functional JC lambda 7. *J. Immunol.* **146**, 2813–2820.

Bernard, F., Chuchana, P., Frippiat, J.-P., Buluwela, L. & Lefranc, M.-P. (1990). Genomic sequence of IGLV1S2, a human immunoglobulin variable lambda gene belonging to subgroup I. *Nucl. Acids Res.* **18**, 7139.

Bernard, O., Hozumi, N. & Tonegawa, S. (1978). Sequences of mouse immunoglobulin light chain genes before and after somatic changes. *Cell*, **15**, 1133–1144.

Brockly, F., Alexandre, D., Chuchana, P., Huck, S., Lefranc, G. & Lefranc, M.-P. (1989). First nucleotide sequence of a human immunoglobulin variable λ gene belonging to subgroup II. *Nucl. Acids Res.* **17**, 3976.

Ch'ang, L.-Y., Yen, C.-P., Besl, L., Schell, M. & Solomon, A. (1994). Identification and characterisation of a functional human Ig $V_{\lambda VI}$ germline gene. *Mol. Immunol.* **31**, 531–536.

Ch'ang, L.-Y., Schnell, M., Ringelberg, C., Weiss, D. T. & Solomon, A. (1995). Molecular characterisation of a human $V_{\lambda VIII}$ germline gene. *Mol. Immunol.* **32**, 49–55.

Chothia, C. & Lesk, A. M. (1987). Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* **196**, 901–917.

Chuchana, P., Blancher, A., Brockly, F., Alexandre, D., Lefranc, G. & Lefranc, M.-P. (1990). Definition of the human immunoglobulin variable lambda (IGLV) gene subgroups. *Eur. J. Immunol.* **20**, 1317–1325.

Chuchana, P., Alexandre, D., Lefranc, G. & Lefranc, M.-P. (1991). The IGLVA gene as a STS in the human immunoglobulin lambda light chain locus (located at 22q11). *Nucl. Acids Res.* **19**, 4786.

Combriato, G. & Klobeck, H.-G. (1991). V_λ and J_λ–C_λ gene genes of the human immunoglobulin λ light chain locus are separated by 14 kb and rearrange by a deletion mechanism. *Eur. J. Immunol.* **21**, 1513–1522.

Cook, G. P. & Tomlinson, I. M. (1995). The human immunoglobulin V_H repertoire. *Immunol. Today*, **16**, 237–242.

Cook, G. P., Tomlinson, I. M., Walter, G., Riethman, H., Carter, N. P., Buluwela, L., Winter, G. & Rabbitts, T. H. (1994). A map of the human immunoglobulin V_H locus completed by analysis of the telomeric region of chromosome 14q. *Nature Genet.* **7**, 162–168.

Cox, D. W., Markovic, V. D. & Teshima, I. E. (1982). Genes for immunoglobulin heavy chains and for α_1-antitrypsin are localized to specific regions of chromosome 14q. *Nature*, **297**, 428–430.

Cox, J. P. L., Tomlinson, I. M. & Winter, G. (1994). A directory of human germ-line V_κ genes reveals a strong bias in their usage. *Eur. J. Immunol.* **24**, 827–836.

Daley, M. D., Olee, T., Peng, H.-Q., Soto-Gil, R. W., Chen, P. P. & Siminovitch, K. A. (1992a). Molecular characterisation of the human immunoglobulin VλI germline gene repertoire. *Mol. Immunol.* **29**, 1031–1042.

Daley, M. D., Peng, H.-Q., Misener, V., Liu, X.-Y., Chen, P. P. & Siminovitch, K. A. (1992b). Molecular analysis of human immunoglobulin Vλ germline genes: subgroups VλIII and VλIV. *Mol. Immunol.* **29**, 1515–1518.

Dariavach, P., Lefranc, G. & Lefranc, M.-P. (1987). Human immunoglobulin $C_\lambda 6$ gene encodes the $Kern^+Oz^-$ λ chain and $C_\lambda 4$ and $C_\lambda 5$ are pseudogenes. *Proc. Natl Acad. Sci. USA*, **84**, 9074–9078.

de la Chapelle, A., Lenoir, A., Boue, G., Gallano, P., Huerre, C., Szajnert, M. F., Jeanpierre, M., Lalouel, M. & Kaplan, J. C. (1983). Lambda Ig constant region genes are translocated to chromosome 8 in a Burkitt's lymphoma (8, 22). *Nucl. Acids Res.* **11**, 1133–1142.

Deftos, M., Olee, T., Carson, D. A. & Chen, P. P. (1994a). Defining the genetic origins of three rheumatoid synovium-derived IgG rheumatoid factors. *J. Clin. Invest.* **93**, 2545–2553.

Deftos, M., Soto-Gil, R., Quan, M., Olee, T. & Chen, P. P. (1994b). Utilization of a potentially universal downstream primer in the rapid identification and characterization of Vλ genes from two new human Vλ gene families. *Scand. J. Immunol.* **39**, 95–103.

Eulitz, M., Ch'ang, L.-Y., Zirkel, C., Schell, M., Weiss, D. T. & Solomon, A. (1995). Distinctive serologic, chemical, and molecular properties of human λIV light chains. *J. Immunol.* **154**, 3256–3265.

Fang, Q., Kannapell, C. C., Gaskin, F., Solomon, A., Koopman, W. J. & Fu, S. M. (1994). Human rheumatoid factors with restrictive specificity for rabbit immunoglobulin G: auto- and multi-reactivity, diverse V_H gene usage and preferential usage of V_λIIIb. *J. Exp. Med.* **179**, 1445–1456.

Felsenstein, J. (1989). PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics*, **5**, 164–166.

Frippiat, J.-P. & Lefranc, M.-P. (1994). Genomic organisation of 34 kb of the human immunoglobulin lambda locus (IGLV): restriction map and sequences of new VλIII genes. *Mol. Immunol.* **31**, 657–670.

Frippiat, J.-P., Chuchana, P., Bernard, F., Buluwela, L., Lefranc, G. & Lefranc, M.-P. (1990). First genomic sequence of a human Ig variable lambda gene belonging to subgroup III. *Nucl. Acids Res.* **18**, 7134.

Frippiat, J.-P., Williams, S. C., Tomlinson, I. M., Cook, G. P., Cherif, D., Le Paslier, D., Collins, J. E., Dunham, I., Winter, G. & Lefranc, M.-P., (1995). Organisation of the human immunoglobulin lambda light-chain locus on chromosome 22q11.2. *Hum. Mol. Genet.* **4**, 983–981.

Harmer, I. J., Loizou, S., Thompson, K. M., So, A. K. L., Walport, M. J. & Mackworth-Young, C. (1995). A human monoclonal anti-phospholipid antibody that is representative of serum antibodies and is germline encoded. *Arthritis Rheum.* **38**, 1068–1076.

Hieter, P. A., Hollis, G. F., Korsmeyer, S. J., Waldmann, T. A. & Leder, P. (1981). Clustered arrangement of immunoglobulin λ constant region genes in man. *Nature*, **294**, 536–540.

Hieter, P. A., Maizel, J. V. & Leder, P. (1982). Evolution of human immunoglobulin κ J region genes. *J. Biol. Chem.* **257**, 1516–1522.

Home, W. A., Ford, J. E. & Gibson, D. M. (1992). L chain isotype regulation in horse. I. Characterisation of Ig λ genes. *J. Immunol.* **149**, 3927–3936.

Irigoyen, M., Manheimer-Lory, A., Gaynor, B. & Diamond, B. (1994). Molecular analysis of the human

immunoglobulin VλII gene family. *J. Clin. Invest.* **94**, 532–538.

Kabat, E. A., Wu, T. T., Perry, H. M., Gottesman, K. S. & Foeller, C. (1991). *Sequences of Proteins of Immunological Interest*. U.S. Department of Health and Human Services, Bethesda, MD, USA.

Kawasaki, K., Minoshima, S., Schooler, K., Kudoh, J., Asakawa, S., de Jong, P. J. & Shimizu, N. (1995). The organisation of the human immunoglobulin λ gene locus. *Genome Res.* **5**, 125–135.

Kudo, A. & Melchers, F. (1987). A second gene, VpreB in the lambda 5 locus of the mouse, which appears to be selectively expressed in pre-B lymphocytes. *EMBO J.* **6**, 2267–2272.

Lee, G., Ware, R. R. & Latov, N. (1994). Somatically mutated member of the human $V_{\lambda VIII}$ gene family encodes anti-myelin associated glycoprotein (MAG) activity. *J. Neuroimmunol.* **51**, 45–52.

Malcolm, S., Barton, P., Murphy, C., Ferguson-Smith, M. A., Bentley, D. L. & Rabbitts, T. H. (1982). Localization of human immunoglobulin kappa light chain variable region genes to the short arm of chromosome 2 by in situ hybridization. *Proc. Natl Acad. Sci. USA*, **79**, 4957–4961.

McIntire, K. R. & Rouse, A. M. (1970). Mouse immunoglobulin light chains: alterations of κ:λ ratio. *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **29**, 704.

Ravetch, J. V., Siebenlist, U., Korsmeyer, S., Waldmann, T. & Leder, P. (1981). Structure of the human immunoglobulin μ locus: characterization of embryonic and rearranged J and D genes. *Cell*, **27**, 583–591.

Reynaud, C.-A., Anquez, V., Grimal, H. & Weill, J.-C. (1987). A hyperconversion mechanism generates the chicken light chain preimmune repertoire. *Cell*, **48**, 379–388.

Reynaud, C.-A., Mackay, C. R., Müller, R. G. & Weill, J.-C. (1991). Somatic generation of diversity in a mammalian primary lymphoid organ: the sheep ileal Peyer's patches. *Cell*, **64**, 995–1005.

Sanchez, P., Marche, P. N., Rueff-Juy, D. & Cazenave, P. A. (1990). Mouse Vλx gene sequence generates no junctional diversity and is conserved in mammalian species. *J. Immunol.* **144**, 2816–2820.

Schäble, K. F. & Zachau, H. G. (1993). The variable genes of the human immunoglobulin κ locus. *Biol. Chem. Hoppe-Seyler*, **374**, 1001–1022.

Siminovitch, K. A., Misener, V., Kwong, P. C., Song, Q.-L. & Chen, P. P. (1989). A natural autoantibody is encoded by germline heavy and lambda light chain variable region genes without somatic mutation. *J. Clin. Invest.* **84**, 1675–1678.

Sinclair, M. C., Gilchrist, J. & Aitken, R. (1995). Molecular characterisation of bovine Vλ regions. *J. Immunol.* **155**, 3068–3078.

Stiernholm, N. B. J. & Berinstein, N. L. (1995). A mutated promotor of a human Ig Vλ gene is associated with reduced germ-line transcription and a low frequency of rearrangement. *J. Immunol.* **154**, 1748–1761.

Stiernholm, N. B. J., Kuzniar, B. & Berinstein, N. L. (1994). Identification of a new human Vλ gene family-VλX. *J. Immunol.* **152**, 4969–4975.

Taub, R. A., Hollis, G. F., Hieter, P. A., Korsmeyer, S., Waldmann, T. A. & Leder, P. (1983). Variable amplification of immunoglobulin λ light-chain genes in human populations. *Nature*, **304**, 172–174.

Tomlinson, I. M., Walter, G., Marks, J. D., Llewelyn, M. B. & Winter, G. (1992). The repertoire of human germline V_H sequences reveals about fifty groups of V_H genes with different hypervariable loops. *J. Mol. Biol.* **227**, 776–798.

Tonegawa, S., Maxam, A M., Tizard, R., Bernard, O. & Gilbert, W. (1978). Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain. *Proc. Natl Acad. Sci. USA*, **75**, 1485–1489.

Udey, J. A. & Blomberg, B. (1987). Human lambda light chain locus: Organisation and DNA sequences of three genomic J regions. *Immunogenetics*, **25**, 63–70.

Vasicek, T. J. & Leder, P. (1990). Structure and expression of the human immunoglobulin λ genes. *J. Exp. Med.* **172**, 609–620.

Williams, S. C. & Winter, G. (1993). Cloning and sequencing of human immunoglobulin V_λ genes. *Eur. J. Immunol.* **23**, 1456–1461.

Winkler, T. H., Fehr, H. & Kalden, J. R. (1992). Analysis of immunoglobulin variable region genes from human IgG anti-DNA hybridomas. *Eur. J. Immunol.* **22**, 1719–1728.

Wu, S. & Cygler, M. (1993). Conformation of complementarity determining region L1 loop in murine IgG λ light chain extends the repertoire of canonical forms. *J. Mol. Biol.* **229**, 597–601.

Edited by J. Karn

(Received 29 May 1996; received in revised form 14 August 1996; accepted 23 August 1996)