Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 9596–9600, December 1986
Genetics

Genetic polymorphism and exon changes of the constant regions of the human T-cell rearranging gene γ

(T-cell antigen receptor/gene conversion/exon duplication)

M.-P. LEFRANC*, A. FORSTER, AND T. H. RABBITTS

Medical Research Council, Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, England

Communicated by C. Milstein, August 25, 1986

ABSTRACT **The genomic nucleotide sequences of the constant-region (*C*) genes of the human T-cell rearranging gene γ are given. These sequences show considerable allelic and nonallelic variation. Allelic variants exist at both $C_\gamma 1$ and $C_\gamma 2$ loci in coding regions (as well as in restriction enzyme sites). Both C_γ genes are in the same transcriptional orientation. Moreover, the organization of the nonallelic C_γ genes reveals some interesting features: the $C_\gamma 1$ gene, like the mouse C_γ gene, has three exons, whereas the $C_\gamma 2$ gene has four exons, including a duplicated second exon that would create a putative protein with an enlarged constant region. However, these two duplicated exons in $C_\gamma 2$ have lost the cysteine residue that is thought to be involved in the interchain disulfide bridge.**

The human T-cell rearranging gene γ (TRG γ) has recently been identified (1, 2) by homology with a gene in mouse that has been shown to undergo rearrangement specifically in T cells (3, 4). So far, two human constant-region (*C*) genes have been identified (1, 2, 5, 6), as well as nine variable-region (*V*) genes (7), which apparently can rearrange independently with each of the C_γ genes (5). The productive rearrangements of the γ locus in mouse were at first thought to be restricted to single V_γ and C_γ genes (8, 9). This is not the situation in human TRG γ genes (5, 7). However, it is now apparent that the situation in the mouse is, in actuality, complex (10), and so the complexity of the system found in man may actually be paralleled by that of mouse, but the general organization seems to be different.

The two human C_γ genes are linked to each other (1) and each has at least one joining (J_γ) segment (5). We now present the complete nucleotide sequences of the genomic coding segments of the C_γ genes, which demonstrate interesting evolutionary changes in sequence and exon structure.

METHODS

λ phage clones were isolated from genomic libraries prepared in λ2001 (11), using *Sau*3A partial digestion products cloned into the *Bam*HI site (1, 7). TRG γ clones were identified by hybridization of the J_γ probe M13H60 (see Fig. 1) according to standard methods (7, 12–14). Restriction maps were produced using single and double digests of the λ phage clones or using appropriate subclones made in pUC vectors (15).

Nucleotide sequence analysis was carried out by dideoxy chain-termination procedures (16) in M13 vectors (17). Sequencing strategies were either (*i*) directed sequencing, using known restriction enzyme sites, or (*ii*) shotgun procedures (18). Data were analyzed by automatic computer methods (19).

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

RESULTS AND DISCUSSION

Comparison of the Exon Structure of $C_\gamma 1$ and $C_\gamma 2$. An overlapping map of the two human TRG γ constant-region genes was generated by comparison of the λ phage clones depicted in Fig. 1. The location of exon 1 of $C_\gamma 1$ and $C_\gamma 2$ (1, 2), as well as that of three J_γ segments (5, 7), has been described previously. The other exons of the C_γ genes have now been located in the genomic DNA by nucleotide sequencing and comparison to mouse and human cDNA sequences (3, 6). The size of each C_γ gene is large compared with other known rearranging genes. The $C_\gamma 1$ gene occupies 5.9 kb, and the $C_\gamma 2$ gene occupies 9.4 kb. Both genes have the same transcriptional orientation and, as previously shown (7), they share the same tandemly arranged pool of V_γ genes.

The $C_\gamma 1$ gene, like the C_γ gene of mouse, has three exons [exon 1 is 330 base pairs (bp), exon 2 is 48 bp, and exon 3 is 141 bp plus a 3′ untranslated sequence of about 450 bp], but they are separated by much larger introns (3.2 kb and 2 kb) than those found in the mouse gene (4). The nucleotide sequence and derived amino acid sequence of these exons are shown in Fig. 2. To identify exon 2 (which does not cross-hybridize with the mouse γ probe), the 2.8-kb *Hin*dIII fragment between exons 1 and 3 was sequenced fully (complete data not shown), but only one exon was identified [unlike the corresponding region of the $C_\gamma 2$ gene, which has a duplication of exon 2 (see below)]. As previously noted, there are potential sites for N-linked glycosylation encoded in human C_γ (6) but only in one mouse C_γ gene (3, 10). Comparison of the $C_\gamma 1$ protein-coding sequence with that of the mouse shows that there has been conservation of the cysteine residues, including the one that is presumably involved in interchain disulfide bridges. This residue is not conserved in the human $C_\gamma 2$ gene (discussed below).

A comparison of the restriction maps of $C_\gamma 1$ and $C_\gamma 2$ showed that $C_\gamma 2$ is much larger than $C_\gamma 1$. Closer analysis of restriction sites indicated a 3.5-kb region (bar in Fig. 1) within $C_\gamma 2$ that might have resulted from a duplication or insertion of DNA. This segment includes a site for the restriction enzyme *Bam*HI. The sequence of exon 2 from $C_\gamma 1$ has an internal site for this enzyme (Fig. 2), suggesting the possibility that a duplication that included exon 2 has occurred in $C_\gamma 2$. The nucleotide sequence of a cDNA clone derived from the T-cell line HPB-MLT supports this view (6). Analysis of the genomic $C_\gamma 2$ gene indeed revealed four exons separated by introns of 4 kb, 2.7 kb, and 2 kb, with a duplicated exon 2 (Figs. 3 and 4), designated exon 2^R and exon 2. Considerable drift has occurred between exon 2 of $C_\gamma 1$ and exon 2^R of $C_\gamma 2$, as well as between exons 2^R and 2 of $C_\gamma 2$. Neither the

Abbreviations: TRG, T-cell rearranging gene; *C*, *V*, and *J*, constant-, variable-, and joining-region genes; kb, kilobase(s).

*Present address: Université des Sciences et Techniques du Languedoc, Laboratoire d'Immunogénétique, UA CNRS 1191, Place Eugène Bataillon, 34060 Montpellier, Cedex, France.

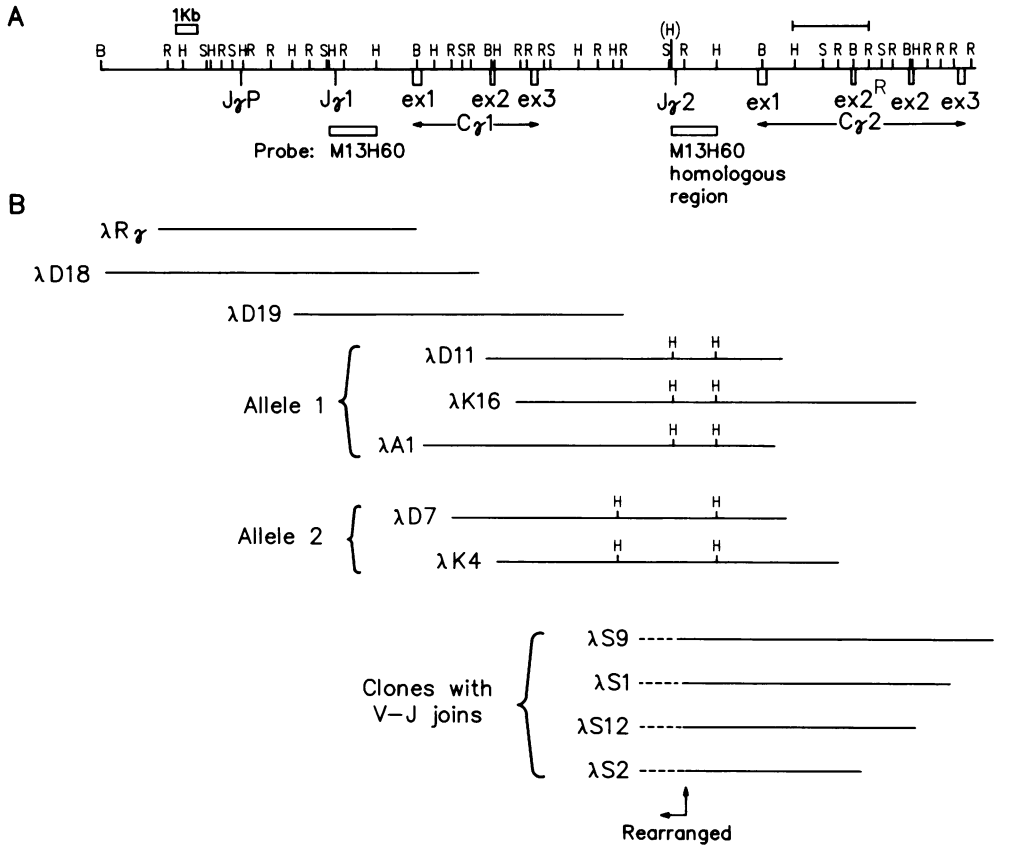

FIG. 1. (*A*) Restriction map of human TRG γ constant-region locus. The map was derived from the sites present in the various λ phage clones depicted in *B*. The location of known J_γ segments and C_γ exons (ex1, ex2, and ex3) are indicated together with the *Hin*dIII fragment that constitutes the J_γ probe M13H60. The homologous region upstream of $C_\gamma 2$ and the polymorphic *Hin*dIII site (H) are indicated. The region of putative DNA duplication including exon 2 of $C_\gamma 2$ is shown by a bar above the map. Restriction sites: B, *Bam*HI; R, *Eco*RI; H, *Hin*dIII; S, *Sac* I. kb, Kilobase. (*B*) Various λ phage clones containing genomic DNA covering the indicated areas of the C_γ locus. λR, λD, λK, and λA clones are from unrearranged loci, and λS clones are from rearranged loci. Derivation of the various clones has been described (1, 7).

small exon 2 nor 2^R in $C_\gamma 2$ retains the cysteine residue encoded in mouse genes and proposed to allow interchain bonding (3).

These results raise a number of questions. First, would a putative protein product of $C_\gamma 2$ be able to form a homo- or heterodimer through another cysteine residue, or would such a protein necessarily be a single-chain protein? Second, and related to the first question, has the exon 2 duplication and loss of the cysteine altered the function of the putative $C_\gamma 2$ gene product or converted the gene to a nonfunctional pseudogene? The latter possibility seems unlikely, as very little other nucleotide or protein drift has occurred compared with $C_\gamma 1$ (see Figs. 3 and 4). Third, what effect would introduction of an extra domain resulting from exon 2 duplication have on the folding of C_γ-encoded protein? Further mouse C_γ cDNA sequences have recently been described (10), and these have a longer putative exon 2 than do the previously described (3) mouse C_γ cDNA clones. However, there is retention of the cysteine residues, even though there are significant differences throughout the protein. The questions raised will presumably be answered by experiments which can now be designed to test the function of the various putative γ polypeptides. A putative protein product of the γ gene has been described in a population of human T cells with the surface phenotype $T3^+T4^-T8^-$ (21, 22). This polypeptide may exist in a heterodimer with an as yet undefined polypeptide, and it remains to be investigated whether this putative dimerization occurs via disulfide bonds or not (involving TRG γ1 and/or γ2 products). Further, the possibility of the existence of more distantly related C_γ genes in humans cannot be excluded, particularly since a third J_γ segment (designated $J_\gamma P$; see Fig. 1) has been identified upstream of $J_\gamma 1$ (7) and at least two other J_γ segments have been found (L. Tighe and T.H.R., unpublished data).

The human V_γ genes, which have been cloned and which map near one another (7), belong to a family of *V* genes that are shared between the two identified C_γ genes. The occurrence of J_γ segments belonging to each C_γ gene is similar to the relationship of *J* segments and *C* genes of, for example, the T-cell receptor β-chain gene and particularly resembles that of human immunoglobulin λ light chain genes (23). Similarly, the organization of mouse TRG γ genes, comprising linked *V*, *J*, and *C* genes, is like that of the mouse immunoglobulin λ genes (24). Thus, the organizations of TRG γ and immunoglobulin λ genes in man and mouse, respectively, have much in common.

Allelic Variation of Human TRG γ Constant-Region Genes. Our previous studies of the C_γ genes in humans revealed a restriction fragment length polymorphism (RFLP) associated with an allelic sequence variation in C_γ (1). The RFLP is due to a variation in *Hin*dIII sites in the region detected by the probe M13H60 (see Fig. 1). Isolation of a number of C_γ genes from different human DNA sources carrying one or the other of the polymorphic restriction fragments has shown the existence of at least two C_γ allelic chromosomes: one that carries two 2.1-kb *Hin*dIII restriction fragments (haplotype I) and one that carries a 2.1-kb fragment plus a 5-kb fragment (haplotype II) (Table 1). The Burkitt lymphoma cell line Raji is homozygous for haplotype I (1), whereas the B-cell prolymphocytic leukemia line D-PLL and the T-cell line K-1010 are heterozygous for haplotypes I and II, as judged by the forms of C_γ genes isolated in λ phage clones shown in Fig.

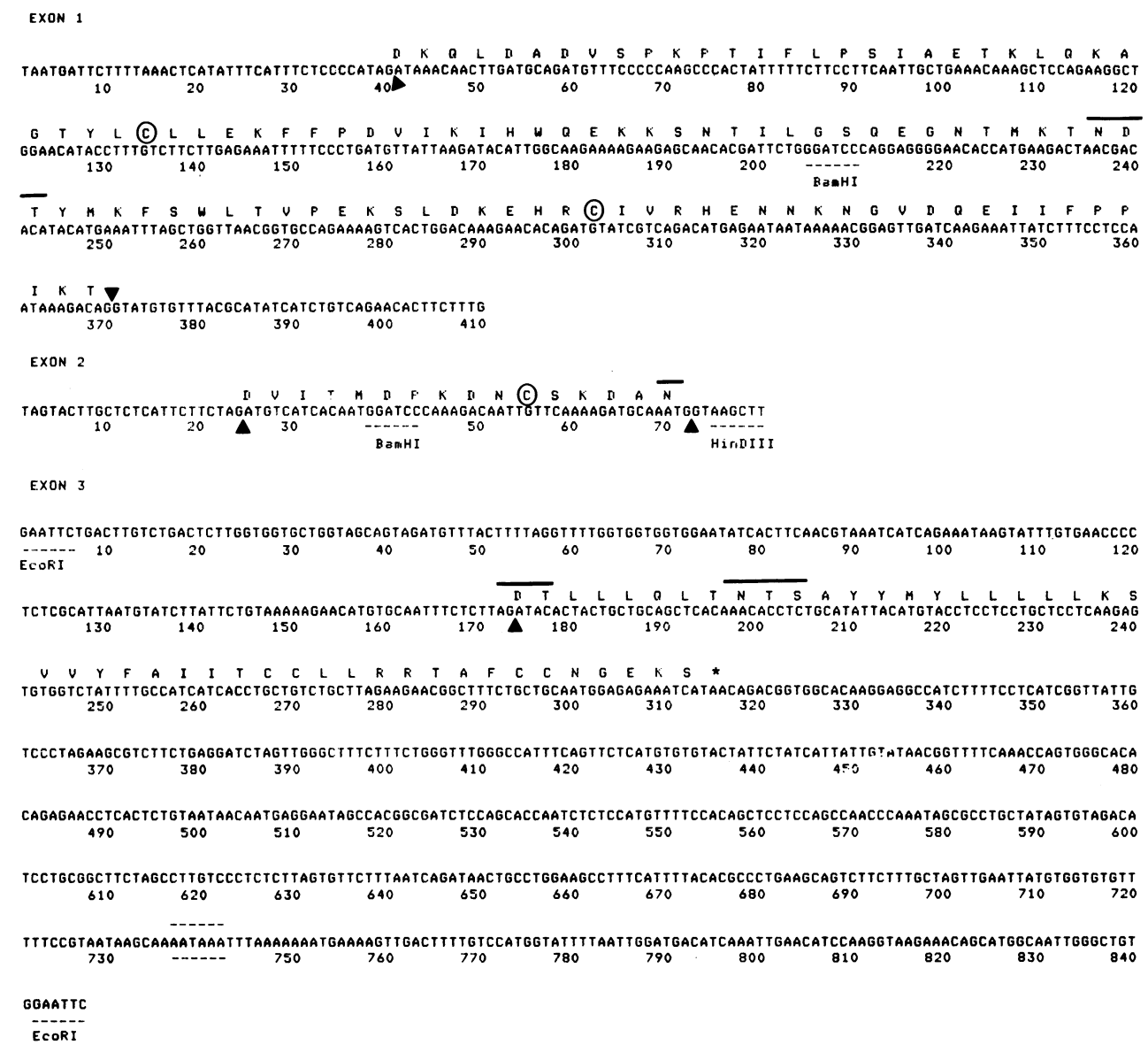

FIG. 2. Genomic exon sequences of human $C_\gamma 1$ gene. Genomic sequences from λD19 were derived from subclones made in M13 by dideoxy chain-termination procedures. The coding regions have been translated in the single-letter amino acid code, and spliced sites are shown by arrowheads. The translation termination codon is indicated by an asterisk, and the poly(A)-addition sequence AATAAA (20) is shown by broken lines above and below the sequence. Potential sites for N-linked glycosylation are overlined, and the cysteine residues probably important for disulfide bonds are circled. Restriction sites are indicated by broken underlining.

1*B*. Filter hybridization experiments with DNA from the T-cell chronic lymphocytic leukemia (T-CLL) AT5B1 showed that this DNA is also homozygous for haplotype I (data not shown).

Comparative DNA sequence analysis of C_γ exon 1 from the allelic $\gamma 1$ genes, belonging to haplotypes I and II, shows that sequence variation (both nucleotide and amino acid) occurs, in addition, from differences seen between $C_\gamma 1$ and $C_\gamma 2$ (Figs. 3 and 4). Although $C_\gamma 1$ and $C_\gamma 2$ differ at the protein level only in exons 2 and 3, the two allelic forms of $C_\gamma 1$ that were sequenced (Raji $\gamma 1$ and D19$\gamma 1$) (Fig. 3) differ at three amino acid positions in exon 1 (Ile^{41}-Glu^{80}-Ile^{99} compared with Val^{41}-Lys^{80}-Val^{99}; Fig. 4). We designated these two allelic $\gamma 1$ sequences as TRG $\gamma 1^a$ and TRG $\gamma 1^b$ (Table 1). Further, although exon 1 of Raji $\gamma 1$ (TRG $\gamma 1^a$) differs at six nucleotide positions (three replacement and three silent nucleotide substitutions) from its allele D19 $\gamma 1$ (TRG $\gamma 1^b$), it is identical with the nonallelic D11 $\gamma 2$ (designated as TRG $\gamma 2^a$) (Fig. 3). These observations suggest that DNA sequence-correction mechanisms, such as gene conversion, have operated on these closely related tandem genes, like those that have occurred in the immunoglobulin α-chain *C* regions (25). In this context, the apparent drift of the duplicated exon 2 of $\gamma 2$ is more puzzling, particularly with reference to the loss of the putative interchain cysteine.

Table 1. Alleles and size of *Hin*dIII fragments in TRG γ haplotypes

Haplotype	Allele (sequence)	Size of upstream restriction fragment, kb
I	$\gamma 1^a$ (Raji $\gamma 1$)	2.1
	$\gamma 2^a$ (D11 $\gamma 2$)	2.1
II	$\gamma 1^b$ (D19 $\gamma 1$)	2.1
	$\gamma 2^b$ (D7 $\gamma 2$)	5

In haplotype I, the *Hin*dIII restriction fragments upstream of the two TRG γ genes (and containing J_γ) (1) are both 2.1 kb long, whereas in haplotype II, the fragments upstream of TRG $\gamma 1$ and TRG $\gamma 2$ are 2.1 kb and 5 kb long, respectively. Sequences corresponding to the different alleles in the coding regions are given in parentheses. These alleles (TRG $\gamma 1^a$, $\gamma 1^b$, $\gamma 2^a$, and $\gamma 2^b$) are defined by the nucleotide differences observed in exon 1 (Fig. 3).

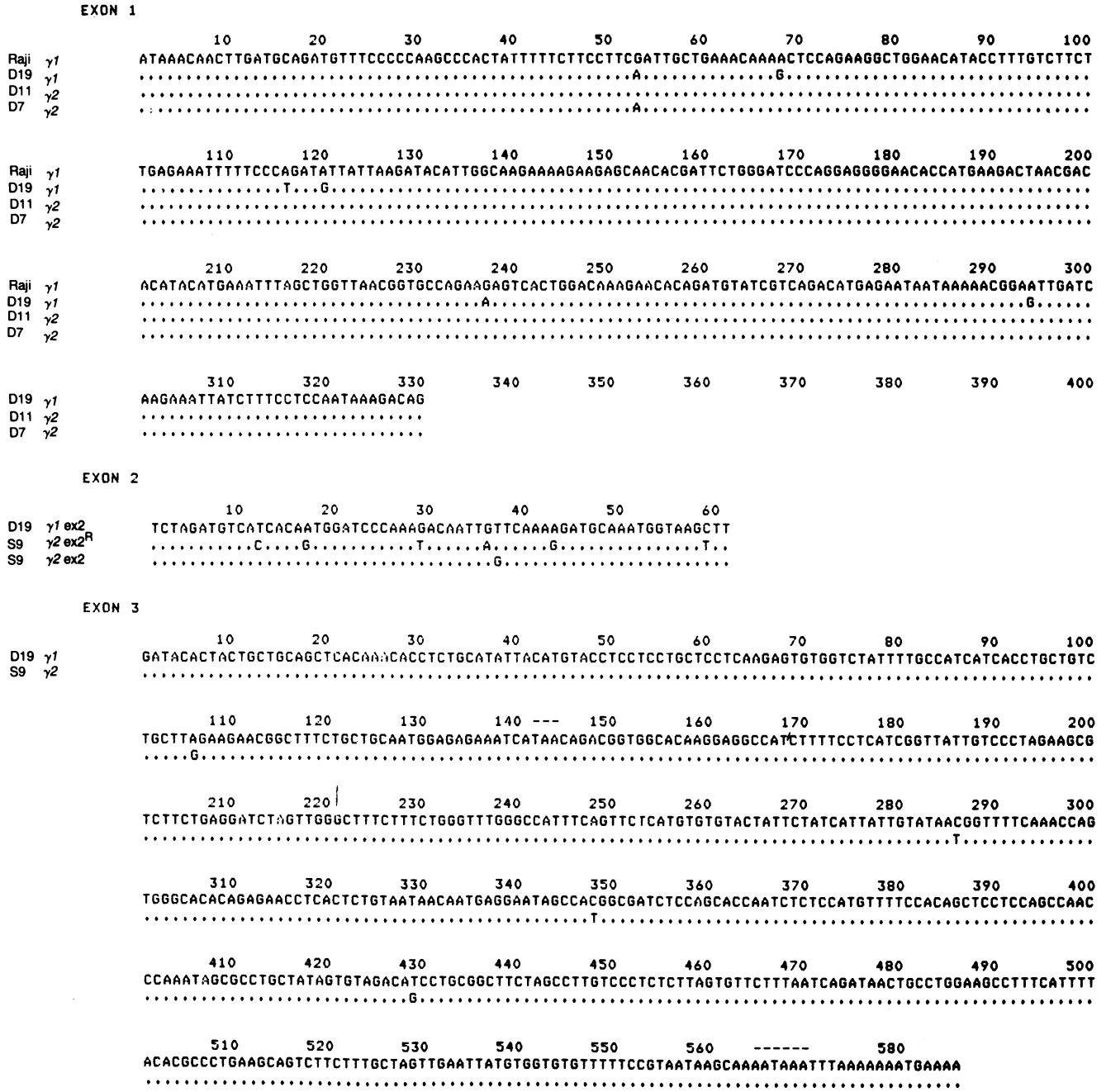

FIG. 3. Comparison of the nucleotide sequences of the allelic and nonallelic human C_γ genes. Nucleotide sequences of the gene segments are compared with each other; dots represent identities, and nucleotide changes are shown appropriately. The translation stop codon and polyadenylylation signal are shown by broken overlining. Partial sequences have been obtained in some cases (e.g., Raji, because of the content of the λ clone λRγ), and only exon 1 has been fully compared for the various allelic forms (see text). Derivation of the various clones has been described (1, 7). D19, D11, and D7 are λ clones from the B-cell prolymphocytic leukemia line D-PLL; S9 is from the T-cell line SUP-T1.

Exon 3 of $C_\gamma 1$ and $C_\gamma 2$ is highly conserved in both coding and noncoding regions (Figs. 3 and 4). A single amino acid substitution occurs near the carboxyl terminus, resulting from the only base change in the coding portion of this exon. The 3′ noncoding region is also remarkably well conserved, with only three base changes. This high conservation is further indication that $C_\gamma 2$ is not a pseudogene, even though the putative interchain cysteine has been lost, since 3′ noncoding fragments of tandem genes can be highly divergent, as, for instance, in the human T-cell receptor β-chain *C* gene (26, 27).

The major question concerning the TRG γ gene is the function of the putative protein product. Although the γ mRNA appears very early in thymus differentiation (28), there is some evidence that γ may not be an absolute requirement. For example, AT5B1 cells have rearranged both α and β chains (R. Baer and T.H.R., unpublished work) and express receptor (as judged by T3 surface expression). However, we did not find productive γ rearrangement in these cells (7), at least not with the two C_γ loci analyzed so far. Thus, a role for γ in differentiation of this T-cell is hard to understand.

We thank Drs. J. Strominger and J. Seidman for communication of data prior to publication. The cost of publication was financed by the Association pour le développement de la Recherche sur le Cancer.

1. Lefranc, M.-P. & Rabbitts, T. H. (1985) *Nature (London)* **316**, 464–466.
2. Murre, C., Waldmann, R. A., Morton, C. C., Bongiovanni,

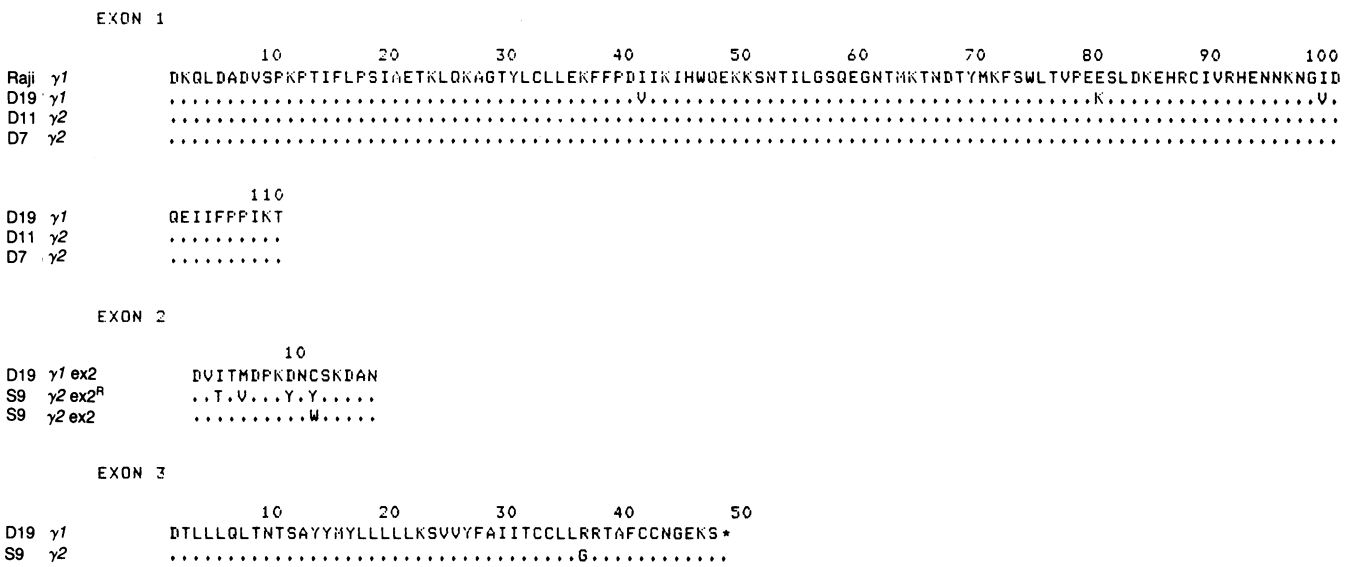

FIG. 4. Comparison of C_γ amino acid sequences. The nucleotide sequences shown in Fig. 3 (see legend for explanation of origin) have been converted to single-letter amino acid sequences. Dots represent identities, and the asterisk at the carboxyl terminus represents the translation termination codon.

K. F., Waldman, T. A., Shows, T. B. & Seidman, J. G. (1985) *Nature (London)* **316,** 549–552.

3. Saito, H., Kranz, D. M., Takagaki, Y., Hayday, A. C., Eisen, H. N. & Tonegawa, S. (1984) *Nature (London)* **309,** 757–762.
4. Hayday, A. C., Saito, H., Gillies, S. D., Kranz, D. M., Tanigawa, G., Eisen, H. N. & Tonegawa, S. (1985) *Cell* **40,** 259–269.
5. Lefranc, M.-P., Forster, A. & Rabbitts, T. H. (1986) *Nature (London)* **319,** 420–422.
6. Dialynas, D. P., Murre, C., Quertermous, T., Boss, J. M., Leiden, J. M., Seidman, J. G. & Strominger, J. L. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 2619–2623.
7. Lefranc, M.-P., Forster, A., Baer, R., Stinson, M. A. & Rabbitts, T. H. (1986) *Cell* **45,** 237–246.
8. Kranz, D. M., Saito, H., Heller, M., Takagaki, Y., Haas, W., Eisen, H. N. & Tonegawa, S. (1985) *Nature (London)* **313,** 752–755.
9. Heilig, J. S., Glimcher, L. H., Kranz, D. M., Clayton, L. K., Greenstein, J. L., Saito, H., Maxam, A. M., Burakoff, S. J., Eisen, H. N. & Tonegawa, S. (1985) *Nature (London)* **317,** 68–70.
10. Iwamoto, A., Rupp, F., Ohashi, P. S., Walker, C. L., Pircher, H., Joho, R., Hengartner, H. & Mak, T. W. (1986) *J. Exp. Med.* **163,** 1203–1212.
11. Karn, J., Matthes, H. W. D., Gait, M. J. & Brenner, S. (1984) *Gene* **32,** 219–224.
12. Denhardt, D. T. (1966) *Biochem. Biophys. Res. Commun.* **23,** 641–646.
13. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251.
14. Benton, W. D. & Davis, R. W. (1977) *Science* **196,** 180–182.
15. Vieira, J. & Messing, J. (1982) *Gene* **19,** 259–268.
16. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
17. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143,** 161–178.
18. Bankier, A. T. & Barrell, B. G. (1983) *Nucleic Acid Biochem. B* **508,** 1–34.
19. Staden, R. (1986) *Nucleic Acids Res.* **14,** 217–231.
20. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263,** 211–214.
21. Brenner, M. B., McLean, J., Dialynas, D. P., Strominger, J. L., Smith, J. A., Owen, F. L., Seidman, J. G., Ip, S., Rosen, F., Krangel, M. S. (1986) *Nature (London)* **322,** 145–149.
22. Bank, I., DePinho, R. A., Brenner, M. A., Cassimeris, J., Alt, F. W. & Chess, L. (1986) *Nature (London)* **322,** 179–181.
23. Hieter, P. A., Hollis, G. F., Korsmeyer, S. J., Waldman, T. A. & Leder, P. (1981) *Nature (London)* **294,** 536–540.
24. Tonegawa, S. (1983) *Nature (London)* **302,** 575–581.
25. Flanagan, J., Lefranc, M.-P. & Rabbitts, T. H. (1984) *Cell* **36,** 681–688.
26. Tunnacliffe, A., Kefford, R., Milstein, C., Forster, A. & Rabbitts, T. H. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 5068–5072.
27. Toyonaga, B., Yoshikai, Y., Vadasz, V., Chin, B. & Mak, T. W. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 8624–8628.
28. Raulet, D. H., Garman, R. D., Saito, H. & Tonegawa, S. (1985) *Nature (London)* **314,** 103–107.