Brief Definitive Report

GENE CORRECTION IN THE EVOLUTION OF THE T CELL RECEPTOR β CHAIN

BY EVELYNE JOUVIN-MARCHE, MARY HELLER, AND STUART RUDIKOFF

From the Laboratory of Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892

The recent isolation of cDNA and genomic clones corresponding to the α and β chains of the T cell antigen receptor (TCR) of both mice and humans has permitted a number of studies regarding evolution of these genes (1–3). Both the α and β chain genes show organizational and structural homology to immunoglobulins (1–4) in that there are variable (V), diversity (D), joining (J), and constant (C) region gene segments that are separated in germline DNA. These segments undergo rearrangement during T cell ontogeny to assemble a functional gene capable of encoding the corresponding protein product (5, 6). The β gene complex contains two highly similar C region genes ($C_T\beta_1$ and $C_T\beta_2$), each of which is associated at its 5′ end with a cluster of $J_T\beta$ and $D_T\beta$ gene segments (for $C_T\beta_2$, there is only one $D_T\beta_2$ gene segment [7–10]). Both the $C_T\beta$ genes are divided into four exons separated by corresponding introns. The two $C_T\beta$ constant genes are present and highly conserved in all inbred mice (11), with the marked exception of NZW, which has lost $C_T\beta_1$, $D_T\beta_2$, and $J_T\beta_2$ regions (12), and SJL, which displays a restriction fragment polymorphism for the $C_T\beta_1$ gene only (13). A similar locus organization has been found in humans, indicating that the presumed duplication of a primordial $C_T\beta$ gene giving rise to $C_T\beta_1$ and $C_T\beta_2$ occurred before the divergence of mice and humans. Surprisingly, comparison of $C_T\beta_1$ and $C_T\beta_2$ sequences in either mouse or humans revealed only minimal divergence in the exon 1 region, suggesting the occurrence of some form of gene correction. Because the apparent correction of $C_T\beta$ sequences, possibly by a mechanism such as gene conversion, is of such potential biological import, and because this putative correction has occurred in such a pauci-gene family in contrast to its more usual association with larger multigene families, we have further examined evolutionary mechanisms operating at this locus by cloning and sequencing the $C_T\beta$ genes from *Mus pahari* (Pa), believed to be the most ancient mouse species existing today. Our results indicate that a correctional event independent of that seen in inbred mice has occurred in this natural population to similarly maintain exon 1 homology.

Materials and Methods

Genomic Clones and Sequence Determination. A Pa recombinant library was constructed by ligating fragments of ~20 kb length derived from partial Mbo I digestion of liver DNA from two individuals into λJ_1 vector (14). 2×10^6 phage were screened by filter hybridization under high-stringency conditions (0.1 × SSC, 0.1% SDS, 65°C) with 86T5, a cDNA probe containing constant region sequences of the T cell receptor β chain (2).

Sequencing was performed by the chain termination procedure of Sanger (15) using M13 phage vectors (16).

Results and Discussion

We have chosen to approach the question of mutational mechanisms operating on $C_T\beta$ genes by an examination of this locus in our wild mouse colony that contains representatives of all four subgenera of *Mus*. Two $C_T\beta$ genes are readily detectable in all wild species from our colony, but the pattern of hybridizable fragments is very different from that observed in *M. musculus domesticus* (17). We have begun analysis of these restriction polymorphisms by characterizing the $C_T\beta$ genes from Pa, a member of the subgenus *coelomys*, believed to be the most evolutionary distant species from inbred laboratory mice. Two hybridizing fragments are detectable by Southern blot analysis (not shown) in Pa genomic DNA, suggesting the presence of two $C_T\beta$ genes, although both fragments are polymorphic in comparison to those of inbred mice (17). To further characterize these genes, a genomic library was screened and a phage clone isolated that contained two hybridizing fragments separated by ~6 kb. Because both hybridizing fragments are present in a single phage, the insert presumably contained $C_T\beta_1$ and $C_T\beta_2$ from a single haplotype. This assessment was subsequently confirmed by sequence comparison with B10.A $C_T\beta_1$ and $C_T\beta_2$ genes. Sequences obtained for the two genes, designated $PaC_T\beta_1$ and $PaC_T\beta_2$, are presented in Fig. 1. As has been found for the inbred mouse and for humans, both $C_T\beta$ genes have the same transcriptional orientation and consist of four exons separated by introns of different sizes. In the $PaC_T\beta_1$ gene, the intron sizes are 491, 96, and 312 bp, and in $PaC_T\beta_2$, they are 487, 144, and 285 bp. In comparison, introns from the $C_T\beta_2$ gene of the inbred B10.A mouse are 506, 145, and 383 nucleotides, respectively. Complete nucleotide sequences are not available for the $C_T\beta_1$ introns. Consensus donor and acceptor RNA splice signals are located adjacent to each exon, and a termination codon is found at the 3′ end of exon 4 that is separated by 199 and 153 bp, respectively, from the $C_T\beta_1$ and $C_T\beta_2$ polyadenylation signals. The three potential glycosylation sites present in B10.A genes are also conserved in Pa. Thus, the two $PaC_T\beta$ genes have all characteristics of functional $C_T\beta$ genes.

A comparison of the $PaC_T\beta_1$ and $PaC_T\beta_2$ sequences reveals striking variation in the degree of homology in different regions of the two genes. As might be expected, the exons are in general quite conserved and the introns are highly divergent. However, only four mutations are observed in the exon 1 sequence, of which three are silent. The transmembrane region is more divergent, with 18 mutations and six amino acid changes between the two genes. The smaller exons 2 and 4 are also slightly more divergent than their B10.A counterparts. A total of 30 mutations are observed between exons of $PaC_T\beta_1$ and $PaC_T\beta_2$, encoding 10 amino acid substitutions, in comparison to 23 mutations and 4 amino acid substitutions in B10.A. A predominance of silent substitutions is seen throughout these genes, indicating selection against replacement changes.

The observed four substitutions in the exon 1 region between $PaC_T\beta_1$ and $PaC_T\beta_2$ is much less than would be expected considering the divergence in other regions of the molecule and the presumed separation time between Pa and *M.*

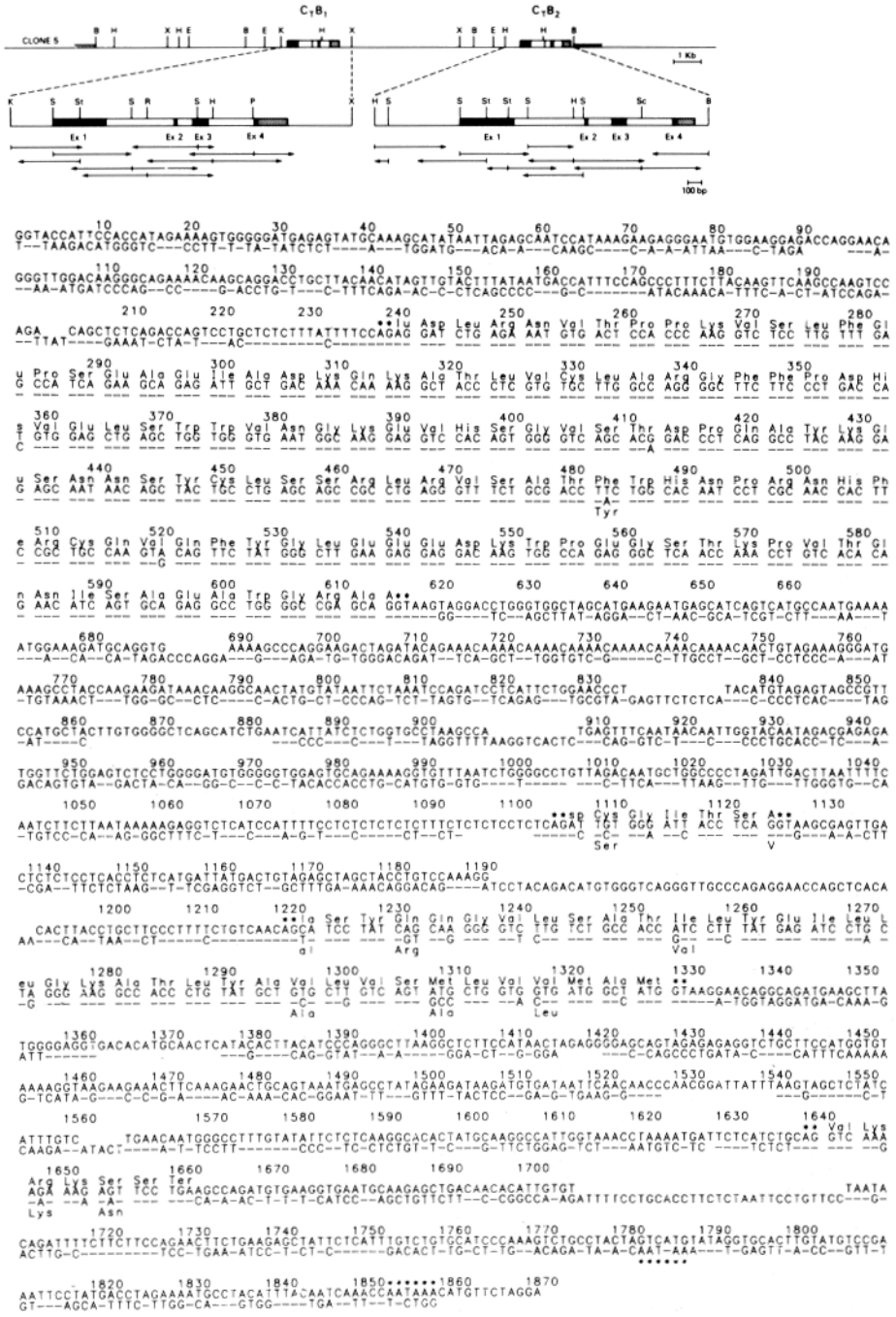

FIGURE 1. Genomic clone and sequence determination of Pa $C_T\beta$ genes. Restriction maps and sequences were generated using the following enzymes: B, Bam HI; E, Eco RI; K, Kpn; H, Hae III; P, Pst I; R, Rsa I; S, Sau 3A; Sc, Sac I; St, Stu I; X, Xba I. Sequencing was performed by the chain-termination procedure of Sanger et al. (15), as indicated by the arrows. Exons and introns are depicted as solid and open boxes, respectively, and hatched areas correspond to 3′ untranslated regions. Nucleotide and predicted amino acid sequences were aligned to maximize homology, with insertion of gaps where necessary. Dashed lines in the $PaC_T\beta_2$ sequence indicate homology to $PaC_T\beta_1$. Splice signals and polyadenylation sites are marked with asterisks.

TABLE I

Exon 1 Sequence Differences between M. pahari and M. m. domesticus

	Codon number								
	17	21	22	67	77	79	96	99	109
$PaC_T\beta_1$	GAA	GCT	GAC	AAC	GTT	GCG	TAT	GAA	ACC
$PaC_T\beta_2$	GAA	GCT	GAC	AAC	GTT	GCG	TAT	GAA	ACC
$B10.AC_T\beta_1$	AAA	GCA	AAC	TAT	GTC	GCT	CAT	TCA	CCC
$B10.AC_T\beta_2$	AAA	GCA	AAC	TAT	GTC	GCT	CAT	TCA	CCC

Substitutions shared between the two $C_T\beta$ genes in a species-specific manner are underlined.

TABLE II

Percent Sequence Divergence between M. pahari, B10.A, and Human $C_T\beta$ Genes

	$PaC_T\beta_1$ $PaC_T\beta_2$	$B.AC_T\beta_1$ $B.AC_T\beta_2$	$huC_T\beta_1$ $huC_T\beta_2$	$PaC_T\beta_1$ $B.AC_T\beta_1$	$PaC_T\beta_1$ $huC_T\beta_1$	$PaC_T\beta_2$ $B.AC_T\beta_2$	$PaC_T\beta_2$ $huC_T\beta_2$
5′ flanking region	59.2	31.0	19.1	14.7	46.6	11.5	46.6
Exon 1	1.1	0.2	1.3	3.5	19.4	3.7	19.4
Intron 1*	58.2	34.3		9.3	46.0	7.7	44.4
5′ Region	62.8		1.1		37.5		
Remainder	56.9		42.6		48.5		
Intron 2	57.7	25.0	28.1	10.3	39.7	10.3	44.6
Exon 3	16.7	14.9	6.5	4.6	11.2	7.4	14.9
Intron 3	55.5	34.6	63.6	8.4	47.6	5.6	47.7
3′ Noncoding region	61.7	47.5	69.7	6.2	36.4	3.7	44.7

* Intron comparisons involving the $B10.AC_T\beta_1$ gene are based on limited sequences available in the literature (2) and constitute 61 bp of the 5′ flanking region and a total of 336 bp in introns. Regions for comparison were defined by exon/intron boundaries, except for the first 105 bp of intron 1, which have been compared independently. Coding regions of exons 2 and 4 were not included due to their small size.

m. domesticus. The similar lack of exon 1 mutations between $C_T\beta_1$ and $C_T\beta_2$ genes in both B10.A and humans has led to the speculation that some form of gene correction (i.e., double crossover or gene conversion) has operated to prevent divergence of this segment. When exon 1 sequences of both $C_T\beta_1$ and $C_T\beta_2$ are compared between Pa and *M. m. domesticus*, 16 positions are noted at which sequence differences occur. At 11 of these (Table I), the substitutions are shared between $C_T\beta_1$ and $C_T\beta_2$ in a species-specific manner. There appears to be some tendency toward clustering of these substitutions, although the data is too limited to permit a clear determination. Several of the interchanges result in nonconservative replacements, such as glutamic acid for lysine (codon 17), asparagine for tyrosine (codon 67) and glutamic acid for serine (codon 99). For each of the other five positions, three sequences out of four are identical, with no pattern as to which sequence contains the mutation. It therefore appears that, in each species, the $C_T\beta_1$ and $C_T\beta_2$ sequences have been independently corrected in a *cis* manner. This correction appears to be localized specifically to exon 1, as sequences 5′ of the exon and in the intron immediately 3′ are highly divergent (Table II). In contrast, the 5′ region of the intron in the human $C_T\beta$ genes also appears to have been corrected (6). Where comparisons are possible, other regions of the $PaC_T\beta_1$ and $C_T\beta_2$ genes are as divergent from each other as the B10.A genes. Interestingly, the human genes are considerably less divergent from each other in all segments from the 5′ flanking region through exon 3 (6). The extreme sequence divergence between $PaC_T\beta_1$ and $C_T\beta_2$ in flanking and intron regions, compared with the high homology between $PaC_T\beta_1$-$B10.AC_T\beta_1$

and Pa$C_T\beta_2$-B10.A$C_T\beta_2$ (Table II), clearly indicates that neither of these species has deleted either $C_T\beta_1$ or $C_T\beta_2$ and subsequently duplicated the remaining gene. Our results suggest that gene correction is a dynamic process occurring during the modern evolution of a natural population, and that such correction may not be a rare occurrence even within pauci-gene families consisting of as few as two genes. Thus, the observed independent gene correction in these two species appears to reflect a strong selective pressure to maintain homogeneous $C_T\beta_1$ and $C_T\beta_2$ exon 1 sequences that presumably are important in some aspect of functional recognition. The above results, in conjunction with similar observations in other mammalian gene families such as immunoglobulins (18) and major histocompatibility antigens (19, 20), provide an increasing body of evidence implicating corrective processes such as gene conversion as being of fundamental importance in the evolution of gene families in higher vertebrates.

Summary

Mutational mechanisms operating at the T cell receptor β chain locus have been examined by comparison of the $C_T\beta_1$ and $C_T\beta_2$ gene sequences from *Mus pahari*, believed to be the oldest living species in the genus *Mus*, with those of inbred mice. Results indicate that a gene correction event independent of that suggested to have occurred in inbred mice has homogenized the *M. pahari* $C_T\beta$ exon 1 sequences, minimizing diversity in this region of the molecule. These observations suggest that correction events such as gene conversion may occur frequently, even in pauci-gene families with as few as two members, and therefore play a significant role in gene diversification or homogenization of small as well as large gene families.

We thank Dr. M. Davis for the 86T5 cDNA clone, and Ms. M. Millison for preparation of the manuscript.

Received for publication 18 June 1986 and in revised form 17 September 1986.

References

1. Yanagi, Y., Y. Yoshikai, K. Leggett, S. P. Clark, I. Aleksander, and T. W. Mak. 1984. A human T cell specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. *Nature* (*Lond.*). 308:145.
2. Hedrick. S. M., D. I. Cohen, E. A. Nielsen, and M. M. Davis. 1984. Isolation of cDNA clones encoding T cell–specific membrane-associated proteins. *Nature* (*Lond.*). 308:149.
3. Sim, G. K., J. Yague, J. Nelson, P. Marrack, E. Palmer, A. Augustin, and J. Kappler. 1984. Primary structure of human T-cell receptor α chain. *Nature* (*Lond.*). 312:771.
4. Hedrick, S. M., E. A. Nielsen, J. Kavaler, D. I. Cohen, and M. M. Davis. 1984. Sequence relationships between putative T-cell receptor polypeptides and immunoglobulins. *Nature* (*Lond.*). 308:153.
5. Chien, Y., N. R. J. Gascoigne, J. Kavaler, N. E. Lee, and M. M. Davis. 1984. Somatic recombination in a murine T-cell receptor gene. *Nature* (*Lond.*). 309:322.
6. Tunnacliffe, A., R. Kefford, C. Milstein, A. Forster, and T. H. Rabbitts. 1985. Sequence and evolution of the human T-cell antigen receptor β chain genes. *Proc. Natl. Acad. Sci.USA.* 82:5068.

7. Sims, J. E., A. Tunnacliffe, W. J. Smith, and T. H. Rabbitts. 1984. Complexity of human T-cell antigen receptor β chain constant- and variable-region genes. *Nature* (*Lond.*). 312:541.
8. Malissen, M., K. Minard, S. Mjolsness, M. Kronenberg, J. Governman, T. Hunkapiller, M. B. Prystowsky, Y. Yoshikai, F. Fitch, T. W. Mak, and L. Hood. 1984. Mouse T-cell antigen receptor: Structure and organization of constant and joining gene segments encoding the β polypeptide. *Cell.* 37:1101.
9. Gascoigne, N. R. J., H.-Y. Chien, D. M. Becker, J. Kavaler, and M. M. Davis. 1984. Genomic organization and sequence of T-cell receptor β-chain constant- and joining-region genes. *Nature* (*Lond.*). 310:387.
10. Siu, G., S. Clark, Y. Yoshikai, M. Malissen, Y. Yanagi, E. Strauss, T. W. Mak, and L. Hood. 1984. The human T-cell antigen receptor is encoded by variable, diversity and joining gene segments that rearrange to generate a complete V gene. *Cell.* 37:393.
11. Caccia, N., M. Kronenberg, D. Saxe, R. Haars, G. Bruns, J. Goverman, M. Malissen, H. Willard, Y. Yoshikai, M. Simon, L. Hood, and T. W. Mak. 1984. The T-cell receptor β chain genes are located on chromosome 6 in mice and chromosome 7 in humans. *Cell.* 37:1091.
12. Kotzin, B. L., V. L. Barr, and E. Palmer. 1985. A large deletion within the T-cell receptor beta-chain gene complex in New Zealand white mice. *Science* (*Wash. DC*). 229:167.
13. D'Hoostelaere, L. A., E. Jouvin-Marche, and K. Huppi. 1985. Localization of $C_T\beta$ and Cκ on mouse chromosome 6. *Immunogenetics.* 22:277.
14. Mullins, J. I., D. S. Brody, R. C. Binai, and S. M. Cotter. 1984. Viral transduction of c-*myc* gene in naturally occurring feline leukemias. *Nature* (*Lond.*). 308:856.
15. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463.
16. Messing, J., and J. Vieira. 1982. A new pair of M13 vectors for selecting either DNA strand of double digest restriction fragments. *Gene* (*Amst.*). 19:269.
17. Huppi, K., L. A. D'Hoostelaere, and E. Jouvin-Marche. 1986. The context of T-cell receptor β chain genes among wild and inbred mouse species. *Curr. Topics Microbiol. Immunol.* 127:291.
18. Clarke, S., and S. Rudikoff. 1984. Evidence for gene conversion among immunoglobulin heavy chain variable region genes. *J. Exp. Med.* 159:773.
19. Pease, L. R., D. H. Schulze, G. M. Pfaffenbach, and S. G. Nathenson. 1983. Spontaneous H-2 mutants provide evidence that a copy mechanism analogous to gene conversion generates polymorphism in the major histocompatibility complex. *Proc. Natl. Acad. Sci. USA.* 80:242.
20. Weiss, E., L. Golden, R. Zakut, A. Mellor, K. Fahrner, S. Kvist, and R. Flavell. 1983. DNA sequence of the H2-K^b gene: evidence for gene conversion as a mechanism for the generation of polymorphism in histocompatibility antigens. *EMBO* (*Eur. Mol. Biol. Organ.*) *J.* 2:453.