

A Genetically Determined Insertion/Deletion Related Polymorphism in Human T Cell Receptor β Chain (TCRB) Includes Functional Variable Gene Segments

By Tong Mao Zhao, Sharon E. Whitaker, and Mary Ann Robinson

From the Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, Rockville, Maryland 20852

Summary

Polymorphism in the human T cell receptor β chain (TCRB) gene complex includes haplotypes with different numbers of TCRBV genes. An insertion/deletion related polymorphism (IDRP) in the human TCRBV region was found to involve TCRBV gene segments. Inserted TCRB haplotypes contain an additional 21.5 kb in which three TCRBV genes are encoded, members of the TCRBV7, TCRBV9, and TCRBV13 families. Two TCRBV gene segments were present only in inserted haplotypes; one of these, TCRBV7S3, is a functional gene and the other, TCRBV9S2(P), is a pseudogene because of an inframe termination colon. In addition, inserted haplotypes contain two identical copies of the TCRBV13S2 gene, whereas deleted haplotypes have only one copy. Deleted haplotypes could be subdivided into two types, deleted*1 and deleted*2, on the basis of sequence variations in TCRBV6S7 and TCRBV13S2 genes. Both deleted*1 and deleted*2 haplotypes contained the same number of TCRBV genes; both contain 60 genes of which 50 are functional, whereas, inserted haplotypes contained 63 genes of which 52 are functional. Comparisons of inserted region sequences with the homologous region in a deleted haplotype, and with sequences surrounding related TCRBV genes, revealed patterns of similarity that suggest insertion as well as deletion events have occurred in the evolution of the TCRBV gene complex. These data indicate that the genomic TCR repertoire is expanded in individuals who have inserted TCRBV haplotypes. The presence of additional TCRBV genes or, alternatively, the absence of certain TCRBV genes may have an impact upon immune responses and susceptibility to autoimmune diseases.

T lymphocytes recognize foreign antigens as peptides presented by self MHC antigens (1). The fine specificity of T cells is determined by the T cell antigen receptor (TCR) that, in the majority of T cells, is composed of an α and a β chain (2–5). For certain MHC–antigen combinations, responding T cells have been observed to use a restricted number of TCRBV genes in the formation of a TCR (6–7). Several strains of mice have major deletions in their genomic TCRB repertoire (8–11), including TCRBV genes used selectively in certain restricted responses. Mice with deletions in their TCRBV genomic repertoire fail to respond or respond very weakly to MHC–antigen combinations that require one of the deleted TCRBV genes (12). Thus, deletions in the TCRBV repertoire can create a “hole in the T cell repertoire” with respect to particular antigens.

In humans, no major deletions of germline TCRBV genes have been reported; however, two independent insertion/deletion related polymorphisms (IDRP)¹ have been identified

in the human TCRB gene complex (13, 14). One IDRP of \sim 20 kb is located near the TCRBC region and another of \sim 30 kb is located within the TCRBV region. Each IDRP has two allelic forms that are widespread in the human population, but the genes encoded within these regions have not been characterized. Understanding the role of TCR genes in human immune responses requires first characterizing the full extent of the germline repertoire including determining whether any TCRBV genes are encoded in the regions of insertion/deletion.

In the present study, three TCRBV gene segments were located in the IDRP within the TCRBV region. None was found in the IDRP located near the constant (C) region. The three inserted TCRBV genes are a BV7 gene, a BV9 pseudogene, and a second copy of a BV13 gene; thus, an inserted TCRBV IDRP haplotype contains more TCRBV genes than a deleted haplotype. Furthermore, the pattern of sequence similarity among related TCRBV family members and the inserted region suggests that insertion as well as deletion events played a role in the evolution of the TCRB gene complex.

Materials and Methods

Human DNA Samples. DNA samples were prepared from lymphoblastoid cell lines derived from donors well characterized for

¹ Abbreviations used in this paper: IDRP, insertion/deletion related polymorphisms; SSCP, single-stranded conformational polymorphism.

A preliminary report was submitted as an abstract to the American Society for Histocompatibility and Immunogenetics 19th Annual Meeting 1993, *Hum. Immunol.* 37:42.

polymorphic markers spanning the TCRB haplotype (15–19) including the TCRBV and TCRBC region IDRPs (13, 14). Donors for polymorphism studies included 126 normal unrelated Caucasian individuals, 18 North American blacks, 23 Guatemalans, 5 Chinese, and samples from the human diversity collection (National Institute of General Medical Sciences [NIGMS] Human Genetic Mutant Cell Repository, Coriell Institute for Medical Research, Camden, NJ) including 5 individuals from the Karitana tribe in Brazil, 5 individuals from the Surui tribe in Brazil, 5 Mayan, 5 Nasioi-speaking Melanesians, 5 Biaka Pygmy, and 5 Mbuti Pygmy.

Nomenclature. The nomenclature used is according to the recommendation of the World Health Organization International Union of Immunological Societies (WHO-IUIS) Subcommittee on TCR Designation (20).

Probes and Primers. Amplification and sequencing oligonucleotide primers were designed on the basis of published sequences (21) or the sequences determined in this study (sequences available upon request) and synthesized using a 394 DNA/RNA synthesizer (Applied Biosystems, Inc., Foster City, CA). Hybridization probes corresponding to TCRBC and to all 24 TCRBV families were prepared by specific PCR amplification from TCRB cDNA clones in pUC18 as described (22). A total of 100 ng of purified probe was radiolabeled using the random primer procedure.

Southern Blots. Southern blots of conventional gels (15) and of pulsed field gels (14) were performed as described.

Single-stranded Conformational Polymorphism (SSCP). SSCP was performed to screen for polymorphism (18, 19). DNA samples from 20 individuals who were selected for genetic diversity were amplified by using PCR. Reaction mixtures (20 μ l) contained DNA (200 ng), oligonucleotide primers (1 mM), dNTP (0.2 mM each) dATP, dCTP, dGTP, and dTTP; Pharmacia LKB, Piscataway, NJ), 10 mM Tris-HCl, pH 8.4, 50 mM KCl, 1.5 mM MgCl₂, gelatin (100 μ g/ml), 0.5 U DNA polymerase from *Thermus aquaticus* (Perkin Elmer-Cetus, Norwalk, CT), and 1 μ Ci of [³²P]dCTP. After a denaturation step of 2 min at 95°C, samples were subjected to 30 cycles of PCR in a DNA Thermal Cycler (Perkin-Elmer Cetus). Each cycle included 95°C for 30 s (denaturation), 60°C for 30 s (annealing), and 72°C for 1.5 min (extension). An aliquot of amplified DNA was diluted 1:10 in 0.1% SDS and 10 mM EDTA, then mixed 1:2 with 95°C formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.5% xylene cyanol. Immediately before loading, samples were heated at 95°C for 2 min. DNA samples were resolved by running 5 μ l of the reaction mixture on a 5% acrylamide gel (Long Ranger; AT Biochem, Malvern, PA) containing 10% glycerol and 0.6 \times TBE (60 mM Tris, 50 mM boric acid, 0.6 mM EDTA) at room temperature for 2.5–3 h at 30 W.

Genomic Cloning. High molecular weight genomic DNA from

individuals with different IDRPs genotypes was digested with BamHI or BglII and then fractionated on a 10–40% sucrose gradient. Southern blots of the fractions were screened with TCRBV7 and TCRBV9 probes to determine the fractions from which to prepare selected libraries. Fractions containing the desired fragments were ligated into BamHI-digested bacteriophage lambda EMBL3 vector arms (Stratagene, La Jolla, CA). Recombinant phage were screened using a ³²P-labeled probe for TCRBV7 or TCRBV9. Positive phage clones were plaque purified then subcloned into pUC 18.

DNA Sequencing. DNA sequencing was performed using the *fmol* DNA Sequencing System (Promega Corp., Madison, WI) according to the manufacturer's instructions.

Results

Characterization of TCRB IDRPs. Two polymorphisms in the human TCRB gene complex characterized by either the insertion or the deletion of 20–30-kb stretches of DNA were identified on Southern blots of SfiI digested DNA samples resolved by pulsed field gel electrophoresis (13, 14). Each of these IDRPs has two allelic forms with allele frequencies in the Caucasian population of 46 and 54% for the deleted and inserted forms of the TCRBC IDRPs, respectively, and of 61 and 37% for the TCRBV IDRPs deleted and inserted forms, respectively (13). The IDRPs located near the TCRBC region, spans \sim 20 kb in the region between BC1 and BV4S1 (Fig. 1 A). The second IDRPs spans \sim 30 kb and is located within the TCRBV region. The SfiI restriction fragments containing the TCRBV IDRPs (\sim 280 kb in deleted haplotypes and \sim 310 kb in inserted haplotypes) are located at the 5' end of the gene complex (14). Double digestion of DNA samples from deleted and inserted TCRBV haplotypes with SfiI and SalI produced fragments of \sim 70 and \sim 100 kb, respectively, that hybridize with probes corresponding to BV7, BV9, BV13, and BV22 (Fig. 1 B) (14).

Conventional Southern blots of DNA samples from individuals with known types of TCRBV and TCRBC region IDRPs were hybridized with probes corresponding to each of the 24 TCRBV families (BV1–BV24). No correlation was observed between BV hybridization patterns and the two forms of the TCRBC IDRPs. However, there was an absolute correlation between inserted TCRBV haplotypes and certain patterns of hybridizing bands.

DNA samples digested with BamHI and hybridized with

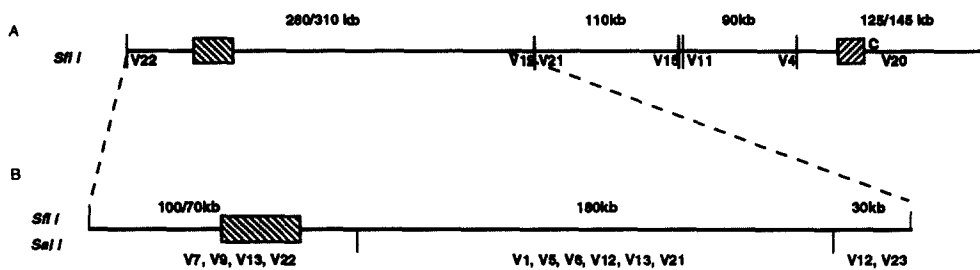


Figure 1. Localization of two IDRPs in the TCRB gene complex. (A) A map of the TCRB gene complex showing the location of the TCRBV and TCRBC IDRPs on SfiI restriction fragments. The sizes of fragments are shown above; IDRPs are marked by shaded boxes. The TCRBV IDRPs map to the 5' end of the gene complex on SfiI fragments of 280 or 310 kb for deleted or inserted TCRBV haplotypes,

respectively. The TCRBC IDRPs map to the 3' end of the gene complex on SfiI fragments of 125/145 kb for deleted/inserted TCRBC haplotypes, respectively. The TCRBV genes encoded at the borders of SfiI fragments are shown below. (B) The SfiI restriction fragment encoding the TCRBV IDRPs is divided into three regions by digestion with SalI. SfiI restriction sites are shown by vertical lines above and SalI restriction sites are shown by vertical lines below. Sizes of fragments are shown over the fragments and the TCRBV probes that hybridize with each fragments

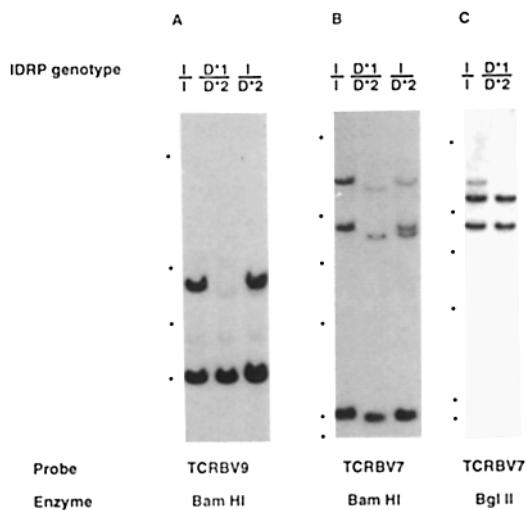


Figure 2. Conventional Southern blots of DNA samples from individuals characterized for TCRBV IDRP. The TCRBV IDRP genotype of the DNA donor is shown above each lane; I is for inserted and D is for deleted TCRBV haplotypes. (A) DNA samples are digested with BamHI and hybridized with a TCRBV9 probe. Fragments observed are 4.7 and 8.5 kb. (B) DNA samples are digested with BamHI and hybridized with a TCRBV7 probe. Fragments observed are 12.5, 11.9, 8.5, 8.0, and 2.4 kb. (C) DNA samples are digested with BglII and hybridized with a TCRBV7 probe. Fragments observed are 12.7, 10.8, and 8.5 kb. Dots to the left of blots mark fragments of λ phage DNA digested with HindIII (23.5, 9.7, 6.6, 4.3, 2.2, and 2.1 kb).

a BV9 probe revealed bands of 4.7 kb in all haplotypes, and an additional band of 8.5 kb in inserted haplotypes (Fig. 2 A). Additional bands were also observed in DNA samples from individuals with inserted TCRBV haplotypes with a BV7 probe; however, the patterns were complicated (Fig. 2 B). BV7 hybridizing fragments of 2.4 kb were observed in all DNA samples. Fragments of 8.5 and 12.5 kb were found in DNA samples from individuals with inserted TCRBV haplotypes and both of these fragments segregated with inserted TCRBV haplotypes in families. Polymorphic fragments of 8.0 or 11.9 kb were observed in DNA samples from individuals with deleted TCRBV haplotypes; only one band segregated with each deleted haplotype in family studies. Therefore, the patterns of hybridization with the BV7 probe allow detection of two forms of deleted haplotypes; those with a 11.9-kb BamHI fragment are designated deleted haplotype*1 and those with an 8.0-kb BamHI fragment are designated deleted haplotype*2. Different combinations of TCRB haplotypes are represented by the DNA samples shown in Fig. 2 B: the sample in the left lane derives from an individual homozygous for inserted haplotypes, the sample in the center lane from an individual heterozygous for deleted haplotype*1 and deleted haplotype*2, and the sample in the right lane from an individual heterozygous for an inserted haplotype and deleted haplotype*2.

Patterns correlating with inserted and deleted TCRBV haplotypes were detected in DNA samples digested with other restriction enzymes and hybridized with BV7 and BV9 probes. For example BglII bands of 10.8 and 8.5 kb that hybridize with BV7 were found in DNA samples from all individuals

and an additional band of 12.7 kb was observed in DNA samples from individuals with inserted TCRBV haplotypes (Fig. 2 C). These correlations between TCRBV inserted or deleted haplotypes and Southern blot patterns were consistent in all individuals tested, including individuals from diverse ethnic backgrounds. The presence of additional fragments in DNA samples from individuals with inserted compared with deleted TCRBV haplotypes suggests that more TCRBV genes are encoded in inserted TCRBV haplotypes.

Cloning of TCRBV Genes from Inserted and Deleted TCRBV Haplotypes. Fragments cloned from inserted and from both types of deleted TCRBV haplotypes are shown in Fig. 3. The BamHI and BglII fragments, which hybridized with BV7 and BV9 probes, were cloned from DNA from an individual who is homozygous for inserted TCRBV haplotypes (BamHI, 12.5 and 8.5 kb [BV7] and 8.5 kb [BV9]; BglII 12.7 and 10.8 kb [BV7 and BV9], and 8.5 kb [BV7]). The polymorphic BamHI fragments hybridizing with the BV7 probe were cloned from an individual who is heterozygous for TCRBV deleted*1 and deleted*2 haplotypes; the 11.9-kb BamHI fragment is designated *del1* and the 8.0-kb BamHI fragment is designated *del2*. Sequences of the clones were determined and overlapping fragments were aligned as shown in Fig. 3.

Five of the fragments (IBg4, I7S, I9, and I7L) derived from the inserted TCRBV haplotype generated a contiguous segment of \sim 33 kb that will henceforth be designated *ins*. Eight TCRBV gene segments were present on these fragments from an inserted haplotype including three BV7 genes, two BV9 genes, two BV13 genes, and one BV6 gene. In addition, there was one separate clone (IBg3) that corresponds in size to a fragment also present in deleted haplotypes that contained one BV7 and one BV9 gene. Based upon previous estimates of the number of genes in TCRBV families (23, 24) and on information from Southern blots (Fig. 2), these clones encompass all of the BV7 and BV9 genes.

Comparison of Maps from Inserted and Deleted TCRBV Haplotypes. Maps derived from the two deleted haplotypes were aligned below the inserted haplotype, as shown in Fig. 3. The locations of TCRBV gene segments are marked, and the regions shared among the three haplotypes are shaded. This alignment shows the probable boundaries of the IDRP and indicates that inserted haplotypes contain \sim 21.5 kb of DNA that is missing from deleted haplotypes.

The TCRBV genes flanking the 5' end of the IDRP are BV7S2 and BV13S2; these are present in all haplotypes. The BV6S7 gene is present in one copy in all haplotypes and marks the 3' border of the IDRP. Inserted haplotypes contain an additional stretch of 21.5 kb in the region flanked by BV6S7 and BV13S2, which contains three gene segments including BV7S3, BV9S2(P), and a second copy of BV13S2.

The two forms of deleted haplotypes are distinguished by polymorphic BamHI restriction fragments observed on blots hybridized with the BV7 probe. The RFLP result from a polymorphic BamHI restriction site that lies within the coding region of BV6S7 (17). BV6S7*1 (previously designated VB6.7a) contains a BamHI site and is the BV6S7 allele present both in inserted and deleted*2 haplotypes. BV6S7*2 (previously designated VB6.7b) lacks this BamHI restriction site and is

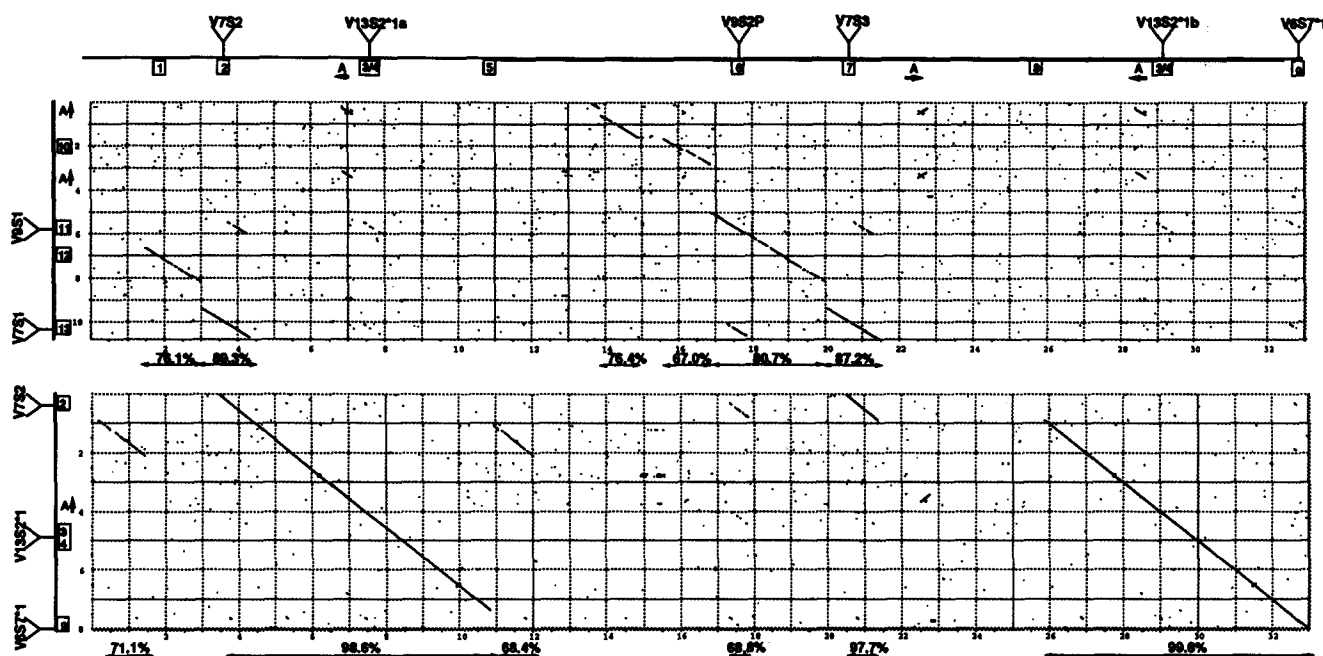


Figure 5. Dot matrix comparison of regions encoding TCRB7 and TCRBV9 genes. A map spanning the inserted region from an inserted TCRBV haplotype (*ins*) is shown across the top of the graph oriented with the 5' end in the upper left hand corner. The *darker line* indicates the inserted region and the *fine line* marks the regions that are also present in deleted*2 haplotypes. Positions of TCRBV genes are shown above. Sizes are shown in kb. Positions and orientation of *Alu* repeats are indicated by an *A* and an *arrow*. Boxed numbers show locations of sequences analyzed by SSCP. The top plot shows comparison of the similar regions from inserted with deleted*2 haplotype (clone D7S) oriented with the 5' end at the top. The bottom plot shows comparison with clone IBg3 (oriented with the 5' end at the top) that carries TCRBV7S1 and TCRBV9S1 genes and is common to all TCRBV haplotypes. Regions of similarity (>65%) between sequences appear as diagonal lines. The analysis was performed using MacVector comparing both strands, with a window size of 30 bases and a hash value of 6. The percentage sequence identity for regions showing significant similarity over a distance of >450 bp are shown below the plot with the *arrows* marking the limits of sequence comparison. Percentage identity was determined by dividing the number of exact matches by the total number of bases compared. The regions examined by SSCP are listed by the first and last positions and numbered according to the sequences in Genbank; (1) 1857-2269 (U07978); (2) 4094-4302 (U07978); (3) 7351-7748 (U07978); (4) 7630-7869 (U07978); (5) 10948-11206 (U07978); (6) 17645-17757 (U07978); (7) 21148-21369 (U07978); (8) 25685-26031 (U07978); (9) 7864-8053 (U07975); (10) 1659-1977 (U07977); (11) 5692-5902 (U07977); (12) 7487-7795 (U07977); and (13) 10410-10669 (U07977).

different haplotypes, inserted, deleted*1, and deleted*2, were identical.

The BV7 gene most similar to BV9 genes is BV7S1, which shares 70.8 and 69.6% nucleotide identity with BV9S1 and BV9S2, respectively. BV9S1 encodes a functional gene but BV9S2(P) is a pseudogene because of an inframe stop codon at the position of amino acid 24 which is a conserved Cys residue in functional genes.

Inserted haplotypes contain two BV13S2 genes that have identical sequences, BV13S2*1a and BV13S2*1b. (The nomenclature for TCR genes has been established by the WHO-IUIS Sub-Committee on TCR Designation (20). To distinguish two identical sequences present at the different loci in the same haplotype, a letter (i.e., a or b) suffix is added to the name.) The BV13S2 genes from inserted haplotypes are also identical to the V13S2 gene from the deleted*2 haplotypes (BV13S2*1). In contrast, the BV13S2 gene from deleted*1 haplotypes (BV13S2*2) contains two substitutions resulting in an amino acid replacement in the leader sequence and nine substitutions plus a 2-bp deletion in the intron (data not shown). The BV13S2*2 sequence was reported previously (21).

Sequence comparisons of TCRBV gene segments present

in the inserted region with genes present in all haplotypes allow for certain conclusions but also reveal puzzling findings. Inserted haplotypes appear to have derived from deleted*2 haplotypes. The inserted and the deleted*2 haplotypes have the same allelic variants of BV6S7 and BV13S2 genes, whereas deleted*1 haplotypes have different alleles. The inserted BV7S3 gene is most similar to BV7S2, which would suggest that BV7S3 derived from BV7S2. However, the BV7S3 gene in the inserted region is linked to BV9S2(P) and the functional BV9 gene is linked to BV7S1 and there is no BV9 gene linked to BV7S2. Examination of the genomic regions flanking the TCRBV gene segments reveals the extent of duplicated regions.

Dot Matrix Comparisons To Identify Related Genomic Regions. Sequences from inserted and deleted*2 haplotypes were compared by dot matrix analysis (Fig. 5, *bottom*). The region from the inserted haplotype analyzed was *ins* which includes a contiguous stretch of 33 kb containing the inserted TCRBV region along with flanking genes shared with deleted haplotypes. The *ins* sequence is written along the x-axis and compared with the sequence *del 2* (clone D7S), which is written along the y-axis. Two major regions of similarity span virtually the full length of the *del 2* sequence (appearing as diagonal lines on the plot). The region of similarity map-

ping to the 5' end of *ins* (positions 3.5–11 kb) spans >7 kb and shows 98.6% sequence identity with most of *del 2*, excluding a region containing the BV6S7 gene. The second region of similarity, which maps to the 3' end of *ins* (positions 26–33 kb), spans >7 kb, and shows 99.6% identity with *del 2* excluding the BV7S2 gene and flanking region. Thus a large segment of DNA which includes the BV13S2 gene and ~6 kb of flanking sequence is duplicated in inserted haplotypes and its sequence is highly conserved.

The *ins* sequence includes two additional regions that contain no functional TCRBV genes, but that show significant identity with a portion (~1.1 kb) of this large duplication. One is located at the 5' end of the insertion and shows 68.4% identity (positions 11–12 kb of *ins*). The other region located 5' of the BV7S2 gene show 71.1% identity (positions 0.1–1.5 kb of *ins*). This region is likely to also be present both in deleted*1 and deleted*2 haplotypes because the sequence derives from a cloned restriction present in all haplotypes. The comparison shown here is for sequence determined for a clone isolated from DNA from a donor with inserted haplotypes. Although there is no functional TCRBV gene segment encoded in this region, both segments appear to contain a remnant of a TCRBV gene. Weak, yet significant, sequence identity (50–64%) is observed with portions of several TCRBV genes but sequence elements critical for TCRBV genes have not been conserved and an open reading frame has not been maintained.

Two additional regions of similarity were detected in the comparison of *ins* and *del 2*. As discussed above, BV7S3 and BV9S2(P) are similar to BV7S2 in coding region sequences and the dot matrix comparison reveals that similarities extending to flanking sequences as well. BV7S2 and its flanking regions show 97.7% identity to BV7S3 (position 21 kb of *ins*) and its flanking regions and 68.8% identity to BV9S2(P) (position 18.5 kb of *ins*) and flanking regions. The regions of similarity, however, are limited and are amidst sequences that show no significant identity. The comparison of *ins* with *del 2* reveals the likely origin for duplicated regions flanking the BV13S2 gene; however, a large stretch of DNA in the insertion shows no significant identity with the same region of a deleted haplotype. Therefore, the inserted region was compared with the upstream region encoding BV7 and BV9 family members.

Comparison of *ins* with clone IBg3 containing BV7S1 and BV9S1 is shown in Fig. 5 (top). The restriction fragment represented by IBg3 is present in all individuals; however, the sequenced clone was isolated from a donor with inserted TCRB haplotypes. Thus, this comparison involves two stretches of DNA that are present in the same haplotype. Diagonal lines show identities in the regions flanking BV7 genes (positions 4 and 21 kb of *ins*) and BV9 genes (position 18.5 kb of *ins*). As expected, comparisons between genes within the same TCRBV family are stronger and extend over a greater distance than comparisons between different TCRBV gene families.

Portions of the inserted segment that did not derive from *del 2* showed significant levels of identity with clone IBg3. The region of *ins* flanking BV9S2(P) and including BV7S3

(from 14 to 21.5 kb of *ins*) was similar to IBg3 in four segments with levels of identity ranging from 67 to 87.2%. Gaps in the regions of identity indicate that there are insertions or deletions in one of the sequences. A horizontal gap in the diagonal line (at position 15 kb of *ins*) indicates the presence of additional sequences in *ins* missing from IBg3. The two vertical gaps in the diagonal line (at positions 17 and 20 kb of *ins*) indicate that *ins* contains deletions of 2 and 1 kb, respectively, compared with clone IBg3. The 5' flanking regions of BV7 genes again show a closer relationship between the inserted BV7S3 and BV7S2. Gaps of the same size in the regions 5' of BV7S2 (at position 3 kb of *ins*) and of BV7S3 (at position 20 kb of *ins*) were revealed when compared with BV7S1.

Repetitive DNA Sequences in the Region of Insertion/Deletion. Additional short stretches of sequence similarity were observed in comparisons of *ins* with the homologous region from *del 2* and with the upstream region on clone IBg3. These regions of similarity correspond to single copies of Alu repetitive elements and are marked on Fig. 5 with an "A". In *ins*, there are two Alu repeats within the inserted region (at positions of 22.5 and 28.5 kb of *ins*) and one in the region 5' of BV13S2 (at position 7 kb of *ins*). The Alu repeat that is located 5' of BV13S2 shows 99.3% identity to those located at 28.5 kb *ins*, at the corresponding position in *del 2*, and to that at the corresponding position in *del 1*. Two additional Alu repeats are found in clone IBg3 located upstream of BV9S1. These two Alu repeats and the repeat found at 22.5 kb of *ins* each show only ~75% identity to other Alu sequences in this region.

Comparisons of the Region of the Insertion/Deletion in Different Individuals. Because the sequences determined in the present study were from clones each derived from a single representative haplotype, the extent of sequence variation of different individuals was examined by using SSCP analysis. SSCP can detect sequence differences in DNA fragments that are detected as bands of single-stranded DNA with distinct migration patterns. 13 separate pairs of oligonucleotides flanking segments of ~300 bp were used to amplify sequences within the region of the insertion and in regions shared by all haplotypes for SSCP analysis. These regions are marked with a number below the line representing *ins* and to the right of the line representing *del 2* and IBg3 on Fig. 5. A minimum of 20 DNA donors were treated with each pair of oligonucleotides; DNA donors were selected for representing combinations of inserted, deleted*1, and deleted*2 haplotypes and for diverse ethnic backgrounds in order to maximize the opportunity for observing genetic heterogeneity. SSCP patterns from genomic DNA were compared with the patterns obtained from DNA from the molecular clones that have known sequences.

Primers for SSCP analyses of BV coding regions were designed to specifically amplify individual BV gene segments and not other family members (SSCP regions 2, 3, 4, 6, 7, 9, 11, and 13). In addition to gene-specific primers, primers that amplified all BV7 genes were used. The two allelic forms of BV6S7 and of BV13S2 were detected by SSCP (SSCP regions 3 and 9). Genes within the insertion BV7S3 and BV9S2(P)

were only amplified from DNA from donors with inserted haplotypes. Two patterns were observed for BV9S2. No additional polymorphism was detected in the regions of BV gene segments.

Five noncoding regions examined for polymorphism (SSCP region 1, 5, 8, 10, and 12). Three of these regions had simple SSCP patterns (1, 10, and 12) and no polymorphism was detected. The regions within or at the border of the IDRPs (SSCP regions 8 and 5, respectively) had more complex patterns. For both regions, one set of bands was detected only in DNA samples from individuals with inserted haplotypes and additional bands were detected in all DNA samples suggesting that more than one sequence from genomic DNA was amplified. The pattern observed for region 5 showed no differences for the bands associated with the inserted haplotype but some minor migration differences among individuals were observed for additional bands. The predominant pattern for region 8 was the same as for cloned DNA; polymorphism in this region appeared to involve length differences in 3 of the 13 individuals with inserted haplotypes who were tested.

SSCP analysis can distinguish DNA fragments that differ by only a single base substitution, thus it is a powerful technique for screening for polymorphism. However, there may be substitutions that are not detected. Data from this survey indicates that sequences in and around the region of insertion/deletion are highly conserved and suggest that genetic polymorphism is limited. This result is consistent with previous studies examining the extent of polymorphism elsewhere in the TCRB gene complex (reviewed in 26).

Discussion

The human TCRB gene complex contains two regions of genetically determined insertion/deletion, one among the variable genes and the other adjacent to the TCRBC region genes (13, 14). The present data reveal that the TCRBV region IDRPs involves the presence of additional TCRBV genes in the inserted form; no identifiable TCRB genes were found in the TCRBC region IDRPs. The BV gene segments encoded within the 21.5 kb (In the original report of the IDRPs, the size of the inserted region for the TCRBV IDRPs was estimated to be 30 kb by using pulsed field electrophoresis. The size discrepancy is likely to be because of difficulty in estimating sizes for large fragments.) inserted region include at least one unique functional gene, a pseudogene, and an exact replicate of a gene found elsewhere in the haplotype. Deleted haplotypes could be subdivided based on allelic differences within a BV gene, but both deleted forms contain the same number of BV genes. Detailed comparisons of sequence in the TCRBV region from individuals with inserted or deleted haplotypes revealed areas of extensive duplication in and around the IDRPs. The presence of the IDRPs could have functional consequences and may give clues concerning evolutionary events shaping the TCR complex.

Certain mouse strains were found to have different numbers of TCRBV genes (8–11), however, this is the first report of such differences in humans. Some strains of mice are missing 50–70% of the wild-type complement of TCRBV gene seg-

ments. While gross differences in immune function were not observed in these strains, functional consequences of these deletions have been reported. Some studies have concluded that the TCR deletions correlate with susceptibility or resistance to autoimmune diseases (27, 28). Another study showed that deletions in TCR have an impact in specific immune responses (12). Certain peptide–MHC combinations elicit T cells that express TCR composed of specific TCRB gene segments (6, 7). Mice missing the relevant TCRB chain from their repertoire were found to be unable to respond to the antigen (12). Although the mice are missing a significant portion of their TCRB repertoire, their failure to respond is associated with the loss of a specific TCRBV gene. There are fewer genes involved in the human TCRBV insertion/deletion than are involved in BV deletions in mice, however, the insertion/deletion in human TCRB may have an impact on specific immune responses that use the TCRBV genes present in the insertion.

The three TCRBV genes encoded within the IDRPs, include BV7S3, a second copy of BV13S2 and a pseudogene, BV9S2(P). Comparison of the inserted BV7S3 gene with BV7S2, the most similar family member present in all individuals, revealed five amino acid substitutions that all localize to CDR2. Strains of mice with different allelic variants of a BV gene, which have a single amino acid substitution in CDR1, were found to have distinct functions (29). The impact of a single gene difference between haplotypes would be observed in a restricted immune response that preferentially uses the relevant TCRBV gene. Thus, the unique sequence features of BV7S3 may confer unique functional properties that lead to protective responses or, alternatively, to an autoimmune reaction.

The second copy of BV13S2 present in inserted haplotypes may also have functional ramifications. Individuals homozygous for deleted haplotypes have two copies of the BV13S2 gene; whereas, individuals homozygous for inserted haplotypes have four copies. The level of expression, and, thus, the used repertoire of BV13S2 may be more extensive in individuals who have additional copies of the gene. The present analyses indicate that BV9S2(P) is a pseudogene in all individuals; because it is nonfunctional its presence is not likely to have an impact on the expressed repertoire.

The extent of the human germline TCRBV repertoire has been estimated by counting numbers of novel TCRB cDNA sequences and numbers of bands observed on Southern blots hybridized with TCRBV probes. Recently these estimates were reevaluated by identification of orphan TCRBV genes on chromosome 9 (14), and by sequence analysis of TCRBV genes in genomic DNA samples derived from multiple unrelated individuals including cosmid clones (23, 24). These studies concluded that there are at least 63 TCRBV gene segments of which 52 encode functional genes. (The gene designated BV13S9 in the report by Wei et al. (24) is identical to what has been designated BV13S2*2 in the present report. The present studies show that BV13S2*1 and BV13S2*2 sequences are allelic, thus, the nomenclature has been corrected.) These estimates indicate the numbers of genes present in inserted TCRBV haplotypes. Deleted*1 and deleted*2 haplo-

types contain three fewer genes and, thus, have 60 TCRBV gene segments of which 50 are functional. The two different forms of deleted TCRBV haplotypes are highly similar in organization and in gene number; they were distinguished by nucleotide sequencing data which suggests that they diverged by mutation. The inserted haplotype is more similar to one of the deleted haplotypes, deleted*2, thus it appears that inserted haplotypes may have derived from deleted haplotypes rather than the reverse.

Similar to many other clustered gene families, the TCRB gene complex is thought to have evolved by a series of gene duplications and subsequent mutation events resulting in TCRBV families. TCRBV gene families are defined by >75% sequence identity and the largest TCRBV families, BV5, BV6, and BV13, occur in the TCRB gene complex in duplicated cassettes (21). Because the inserted TCRBV region contains TCRBV genes that are members of TCRBV families encoded elsewhere in the TCRB gene complex, the origin of the inserted TCRBV region was investigated by performing sequence comparisons of the inserted region with the regions flanking related TCRBV genes.

The inserted region, appears to have arisen by duplication of regions within TCRB because nearly identical sequences were found in the clones isolated from deleted haplotypes. Based upon sequence comparisons, the inserted haplotype appears to have derived from the deleted*2 haplotype. Inserted and deleted*2 haplotypes share allelic variants of BV6S7 and BV13S2 genes and have highly conserved noncoding regions. The 5' end of the inserted region is similar to sequences upstream of the BV7S1 gene and the 3' end of the inserted region is similar to the region encoding BV7S2. Regions in the insertion, which do not show similarities, occur at the boundaries of clones that have been isolated. Further extension of these sequences may reveal additional similarities.

A possible sequence of events to account for the origin

of the insertion is shown in Fig. 6. This hypothesis is based on the occurrence of an unequal crossover agent between two deleted*2 haplotypes causing the duplication of a large region of DNA spanning from BV9S1 to BV13S2. This large duplicated region was then subject to further insertions and deletions and to gene conversion-like events before stabilization. Insertions and deletions are observed in the inserted region as gaps in the lines of similarity on the dot matrix plot (Fig. 5). According to the present hypothesis, a large deletion of the region spanning from the duplicated BV7S2 to the region 3' of BV7S1 would have been necessary to result in the inserted haplotype. The regions of similarity suggest putative origins of the insertion and support this hypothesis, with one exception. The region flanking BV7S3 shows the greatest primary sequence identity (97.7%) with BV7S2 which according to this model was deleted from the insertion. This region of high sequence identity ends abruptly at the 3' edge and the noncoding regions show significant sequence identity to the 3' region flanking BV7S1. The 2.8-kb flanking BV7S3 shows greater sequence identity to BV7S2 (one stretch of 92%) than to BV7S1 (two stretches of 76.1 and 89.3%). However the region of identity with BV7S2 ends abruptly at the 5' and 3' edges and identity with BV7S1 extends. A gene conversion-like process could account for the greater sequence identity of the 2.8-kb region flanking BV7S3 to the same region of BV7S2 in the midst of sequence similar to BV7S1. Within TCRB, gene correction (gene conversion like or double crossing over) has been proposed as the mechanism by which sequence identity of two regions within the gene complex is maintained. For example, gene conversion was proposed to account for the high degree of sequence identity observed for exon 1 of TCRBC1 and BC2 in a number of different species (30, 31). There is also evidence for gene conversion-like events in the immunoglobulin (32, 33) and histocompatibility gene families (34). These examples involve

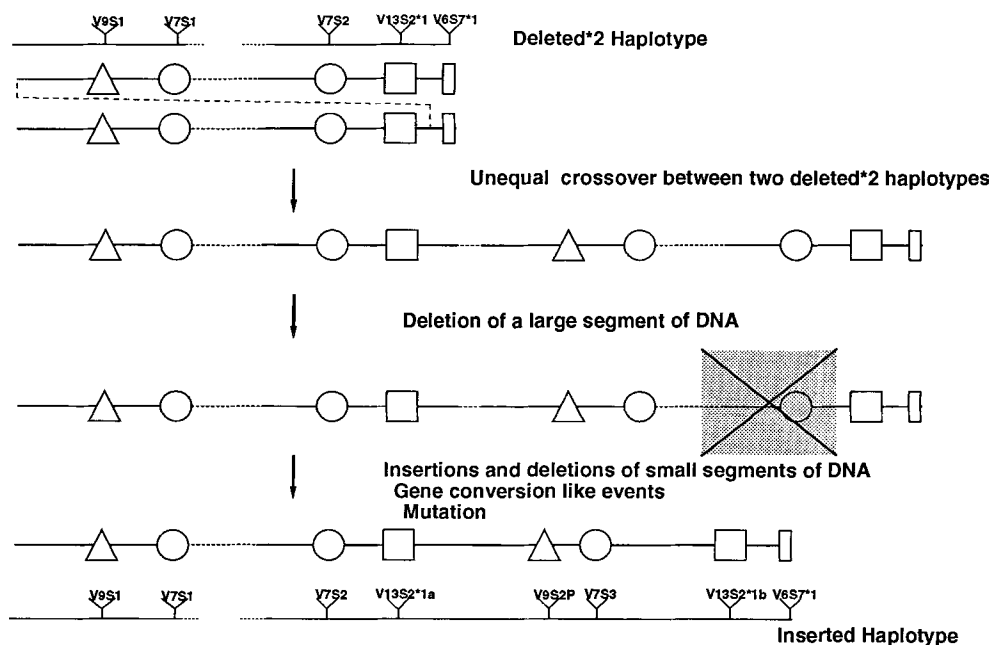


Figure 6. Schematic representation of one hypothesis to explain the derivation of the inserted TCRB haplotype. The map of deleted*2 haplotypes is shown at the top and the map of the inserted haplotype is shown at the bottom of the figure. Between the maps TCRBV gene segments are represented by triangles (BV9), circles (BV7), squares (BV13), and rectangles (BV6). Different family members are represented by shading. It is hypothesized that an unequal crossover event occurred between two deleted*2 haplotypes duplicating a region from BV13S2 to BV9S1. In a subsequent event, a large segment of DNA was deleted removing a region including the duplicated copy of BV7S2. Insertions and deletions of shorter DNA segments, gene conversion like events and mutations operated in the evolution of the region to give rise to the inserted TCRBV haplotype.

gene conversion like events of regions that are smaller than the 2.8-kb region in TCRB. However, gene conversion like mechanisms involving longer stretches of DNA are possible and such mechanisms may account for the observation that there are regions of greater similarity flanked by regions of lesser identity in TCRB.

Alternatively, the insertion may have derived from a series of duplication events. According to this hypothesis, the region encoding BV9S1-BV7S1 is likely to have duplicated first. Sequence comparisons indicate that the 3' end of the inserted region is most similar to the BV9S1-BV7S1 region, however, the percentage sequence identities observed between these two regions are lower than those observed between the remaining portion of the insertion and its counterpart. In a second event, the region flanking BV13S2 was copied which accounts for the high level of sequence identity between the insertion and the regions present both in deleted haplotypes and at another site in inserted haplotypes. Multiple duplication events would result in the discontinuities in sequence that are observed in this region. There is, however, one final observation to be explained. The sequence of BV7S3 is more like BV7S2 but the extended region flanking BV7S3 is more like BV7S1. In this hypothetical model, gene conversion like events may be invoked to describe the mechanism by which BV7S3 arose; the coding and immediate flanking sequence of the duplicated BV7S1 (the present BV7S3 gene) gene was changed to be more similar to BV7S2 retaining the extended BV7S1 flanking region.

The high degree of conservation between the inserted sequences and other regions in TCRB is surprising, particularly for flanking regions. In inserted haplotypes there are two regions over 6 kb in length that share >98% sequence identity. The insertion event is not likely to have been recent because of the frequency and distribution of the different allelic forms of the TCRBV IDRP in the human population, therefore, limited time since duplication is not a plausible explanation for this extraordinary degree of sequence identity. Coding regions of TCRBC and TCRBV genes are highly conserved. Allelic polymorphism has been detected in several BV genes, but it is limited to a few alleles even in surveys of individuals from diverse ethnic backgrounds. Coding regions may be conserved due to selective pressures related to TCR function, such conservation may serve to limit self reactivity. However, this does not provide a rationale for conservation of the noncoding regions. In addition to the present data, other reports suggest that noncoding regions are conserved throughout the gene complex. For example, few re-

striction enzymes reveal polymorphism in TCR genes on Southern blots (26) and surveys using SSCP detect only limited polymorphism in TCR sequences (18, 19, and data not shown). Perhaps, there are sequence features within the TCRB gene complex that limit mutations in the whole region. It is possible that conserved sequences contribute to the frequent occurrence of gene duplication in the TCRB gene complex by providing regions of sequence similarity to allow alignment of discontinuous segments leading to unequal crossing over.

The Alu family of repetitive sequences have been reported to be sites of crossover events in the human genome (35, 36). Alu repeat sequences are thought to provide regions of similarity necessary for recombination, and since Alu repeats are widely dispersed throughout the genome, the recombination events may be nonhomologous. Sites of unequal crossovers that lead to gene duplications are distinguished by high sequence identity on one side of the Alu repeat and the absence of sequence identity on the opposite side. Although Alu sequences are present they do not appear to have been involved in the expansion of the TCRB gene complex in the region of insertion/deletion. There are two Alu repeats that are located at positions bordering the insertion (positions 7 and 28.5 kb of *ins*, Fig. 5) but these Alu repeats display significant sequence identity on both sides and thus appear to be duplicated copies rather than sites of recombination. The remaining Alu repeats do not lie at a border of a putative insertion or a deletion in the sequence, do not show significant sequence identity to another Alu sequence in the region, and do not show significant sequence identity to the flanking regions other Alu repeats. Similar conclusions were reached in a recent study of the human Ig gene complex; evidence was obtained for duplication and translocation events that were apparently independent of Alu repeats (37).

It is surprising that there is not more variation in the number of genes present in TCRB haplotypes. The present study supports the supposition that multiple duplication events have generated TCRBV families and suggests that TCRB evolution has been a dynamic process. The TCRBV and TCRBC IDRP represent variations generated by this process at two different regions of the TCRB gene complex. In the Caucasian population, an additional rare variant has been detected that involves the further deletion of ~30 kb from deleted*1 haplotypes (38 and data not shown). Outside of these observed regions of insertion/deletion, the TCRB gene complex appears to be highly conserved in all individuals studied.

The authors thank Dr. T. J. Kindt and Dr. B. Lynch for critical comments on the manuscript.

Address correspondence to Dr. Mary Ann Robinson, Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, Twinbrook II, 12441 Parklawn Drive, Rockville, MD 20852.

Received for publication 29 March 1994 and in revised form 8 July 1994.

References

1. Schwartz, R.H. 1985. T-lymphocyte recognition of antigen in association with gene products of the major histocompatibility complex. *Annu. Rev. Immunol.* 3:237.
2. Davis, M.M., and P.J. Bjorkman. 1988. T-cell antigen receptor

- genes and T-cell recognition, *Nature (Lond.)*. 334:395.
3. Marrack, P., and J. Kappler. 1987. The T cell receptor. *Science (Wash. DC)*. 238:1073.
 4. Toyonaga, B., and T.W. Mak. 1987. Genes of the T-cell antigen receptor in normal and malignant T cells. *Annu. Rev. Immunol.* 5:585.
 5. Wilson, R.K., E. Lai, P. Concannon, R.K. Barth, and L.E. Hood. 1988. Structure, organization and polymorphism of murine and human T-cell receptor alpha and beta chain gene families. *Immunol. Rev.* 101:149.
 6. Casanova, J.-L., and J.L. Maryanski. 1993. Antigen-selected T-cell receptor diversity and self-nonsel self homology. *Immunol. Today*. 14:391.
 7. Dillon, S.R., S.C. Jameson, and P.J. Fink. 1994. $V\beta 5^+$ T cell receptors skew toward OVA + H-2K^b recognition. *J. Immunol.* 152:1790.
 8. Behlke, M.A., H.S. Chou, K. Huppi, and D.Y. Loh. 1986. Murine T-cell receptor mutants with deletions of β chain variable region genes. *Proc. Natl. Acad. Sci. USA*. 83:767.
 9. Haqqi, T.M., S. Banerjee, W.L. Jones, G. Anderson, M.A. Behlke, D.Y. Loh, H.S. Luthra, and C.S. David. 1989. Identification of T-cell receptor $V\beta$ deletion mutant mouse strain AU/ssj (H-2^g) which is resistant to collagen induced arthritis. *Immunogenetics*. 29:180.
 10. Haqqi, T.M., S. Banerjee, G.D. Anderson, and C.S. David. 1989. RIII S/J (H-2^r). An inbred mouse strain with a massive deletion of T cell receptor $V\beta$ genes. *J. Exp. Med.* 169:1903.
 11. Jouvin-Marche, E., N.S. Trede, A. Bandeira, A. Tomas, D.Y. Loh, and P.-A. Cazenave. 1989. Different large deletions of T cell receptor $V\beta$ genes in natural populations of mice. *Eur. J. Immunol.* 19:1921.
 12. Nanda, N.K., R. Apple, and E. Sercarz. 1991. Limitations in plasticity of the T-cell receptor repertoire. *Proc. Natl. Acad. Sci. USA*. 88:9503.
 13. Seboun, E., M.A. Robinson, T.J. Kindt, and S.L. Hauser. 1989. Insertion/deletion-related polymorphisms in the human T cell receptor beta gene complex. *J. Exp. Med.* 170:1263.
 14. Robinson, M.A., M.P. Mitchell, S. Wei, C.E. Day, T.M. Zhao, and P. Concannon. 1993. Organization of human T-cell receptor beta-chain genes: clusters of V beta genes are present on chromosomes 7 and 9. *Proc. Natl. Acad. Sci. USA*. 90:2433.
 15. Robinson, M.A., and T.J. Kindt. 1985. Segregation of polymorphic T-cell receptor genes in human families. *Proc. Natl. Acad. Sci. USA*. 82:3804.
 16. Robinson, M.A. 1989. Allelic sequence variations in the hyper-variable region of a T-cell receptor β chain: correlation with restriction fragment length polymorphism in human families and populations. *Proc. Natl. Acad. Sci. USA*. 86:9422.
 17. Li, Y., P. Szabo, M.A. Robinson, B. Dong, and D.N. Posnett. 1990. Allelic variations in the human T cell receptor $V\beta 6.7$ gene products. *J. Exp. Med.* 171:221.
 18. Day, C.E., T. Zhao, and M.A. Robinson. 1992. Silent allelic variants of a T-cell receptor $V\beta 12$ gene are present in diverse human populations. *Hum. Immunol.* 34:196.
 19. Barron, K.S., and M.A. Robinson. 1994. The human T cell receptor variable gene segment, TCRBV6S1 has two null alleles. *Hum. Immunol.* 40:17.
 20. WHO-IUIS Nomenclature Sub-Committee on TCR Designation. 1993. Nomenclature for T-cell receptor (TCR) gene segments of the immune system. *Bull. WHO*. 71:113.
 21. Li, Y., P. Szabo, and D.N. Posnett. 1991. The genomic structure of human $V\beta 6$ T cell antigen receptor genes. *J. Exp. Med.* 174:1537.
 22. Robinson, M.A. 1992. Usage of human T-cell receptor $V\beta$, $J\beta$, $C\beta$, and $V\alpha$ gene segments is not proportional to gene number. *Hum. Immunol.* 35:60.
 23. Robinson, M.A., and P. Concannon. 1994. Organization of the human TCRB gene complex. In *T Cell Receptor Genes*. J. Bell, editor. Oxford Press. In press.
 24. Wei, S., P. Charmley, M.A. Robinson, and P. Concannon. 1994. The extent of the human germline T cell receptor germline T cell receptor V beta gene segment repertoire. *Immunogenetics*. 40:27.
 25. Chothia, C., D.R. Boswell, and R.M. Lesk. 1988. The outline structure of the T-cell alpha beta receptor. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:3745.
 26. Day, C., K. Schmitt, and M.A. Robinson. 1994. Frequent recombination in the human T-cell receptor beta gene complex. *Immunogenetics*. 39:335.
 27. Banerjee, S., T.M. Haqqi, H.S. Luthra, J.M. Stuart, and C.S. David. 1988. Possible role of $V\beta$ T cell receptor genes in susceptibility to collagen-induced arthritis in mice. *J. Exp. Med.* 167:832.
 28. Noonan, D.J., P.J. McConahey, and G.J. Cardenas. 1990. Correlations of autoimmunity with H-2 and T cell receptor β chain genotypes in (NZB \times NZW) F2 mice. *Eur. J. Immunol.* 20:1105.
 29. Gahn, S.-J., B.J. Fowlkes, S.D. Jameson, N.R.J. Gascoigne, M.M. Cotterman, O. Kanagawa, R.H. Schwartz, and L.A. Matis. 1991. Profound alteration in an $\alpha\beta$ T-cell antigen receptor repertoire due to polymorphism in the first complementarity-determining region of the β chain. *Proc. Natl. Acad. Sci. USA*. 88:10267.
 30. Jouvin-Marche, E., M. Heller, and S. Rudikoff. 1986. Gene correction in the evolution of the T cell receptor β chain. *J. Exp. Med.* 164:2083.
 31. Mami-Chouaib, F., and T.J. Kindt. 1989. Distinct pattern of conservation of exons and introns of T cell receptor β chain constant region genes in subspecies of *Mus*. *Immunogenetics*. 30:42.
 32. K.L. Knight, and R.S. Becker. 1990. Molecular basis of allelic inheritance of rabbit immunoglobulin VH allotypes: implications for the generation of antibody diversity. *Cell*. 60:963.
 33. Clarke, S.H., and S. Rudikoff. 1984. Evidence for gene conversion among immunoglobulin heavy chain variable region genes. *J. Exp. Med.* 159:773.
 34. 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.
 35. Nicholls, R.D., N. Fischel-Ghodsian, and D.R. Higgs. 1987. Recombination at the human α -globin gene cluster: sequence features and topological constraints. *Cell*. 49:369.
 36. Rouyer, F., M.-C. Simmler, D.C. Page, and J. Weissenbach. 1987. A sex chromosome rearrangement in a human XX male caused by Alu-Alu recombination. *Cell*. 51:417.
 37. Matsumura, R., F. Matsuda, H. Nagaoka, J. Fujikura, E.K. Shin, Y. Fukita, M. Haino, and T. Honjo. 1994. Structural analysis of the human VH locus using nonrepetitive intergenic probes and repetitive sequence probes. *J. Immunol.* 152:660.
 38. Seboun, E., L. Houghton, C.J. Hatem, Jr., R. Lincoln, and Hauser S.L. 1993. Unusual organization of the human T-cell receptor beta-chain gene complex is linked to recombination hotspots. *Proc. Natl. Acad. Sci. USA*. 90:5026.