Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 5380–5383, May 1996
Evolution

Ancestral major histocompatibility complex *DRB* genes beget conserved patterns of localized polymorphisms

LAKSHMI K. GAUR[†‡§] AND GERALD T. NEPOM[¶‖]

[†]Puget Sound Blood Center, 921 Terry Avenue, Seattle, WA 98104-1256; [¶]Virginia Mason Research Center, 1000 Seneca, Seattle, WA 98101; and Departments of [‡]Microbiology and [‖]Immunology, University of Washington School of Medicine, Seattle, WA 98195

Communicated by Hans Neurath, University of Washington School of Medicine, Seattle, WA, January 29, 1996 (received for review September 19, 1995)

ABSTRACT **Genes within the major histocompatibility complex (MHC) are characterized by extensive polymorphism within species and also by a remarkable conservation of contemporary human allelic sequences in evolutionarily distant primates. Mechanisms proposed to account for strict nucleotide conservation in the context of highly variable genes include the suggestion that intergenic exchange generates repeated sets of MHC *DRB* polymorphisms [Gyllensten, U. B., Sundvall, M. & Erlich, H. A. (1991) *Proc. Natl. Acad. Sci. USA* 88, 3686–3690; Lundberg, A. S. & McDevitt, H. O. (1992) *Proc. Natl. Acad. Sci. USA* 89, 6545–6549]. We analyzed over 50 primate MHC *DRB* sequences, and identified nucleotide elements within macaque and baboon *DRB6*-like sequences with deletions corresponding to specific exon 2 hypervariable regions, which encode a discrete α helical segment of the MHC antigen combining site. This precisely localized deletion provides direct evidence implicating segmental exchange of MHC-encoded *DRB* gene fragments as one of the evolutionary mechanisms both generating and maintaining MHC diversity. Intergenic exchange at this site may be fundamental to the diversification of immune protection in populations by permitting alteration in the specificity of the MHC that determines the repertoire of antigens bound.**

Extensive variation of major histocompatibility complex (MHC) genes occurs in all vertebrate species, due to a number of structural and genetic features, including gene duplication, heterodimer formation, and extensive allelic polymorphism. Notably, within MHC *DRB* genes, sequence polymorphism is nonuniform, clustered within "hypervariable" regions of the second exon (1–4), which encodes the antigen-binding portion of the MHC class II molecule. Thus, functional properties of the MHC molecule and the corresponding selective pressures that arise from immunological challenges cluster within a limited set of highly polymorphic nucleotides.

Calculations based on sequence variation have indicated that the overall mutation rate of MHC loci is not higher than that of most other genetic loci (5). However, it has been suggested that discrete segments within the second exon may accumulate mutations at different rates (4). Specific mechanisms for generating localized variability in the MHC, such as gene conversion and interlocus genetic exchange, have been suggested but they are controversial and indirect (6–8). Indeed, arguments for convergent mechanisms to account for recurrent sequences within otherwise divergent alleles have even been proposed (9). We now provide evidence from an analysis of macaque and baboon MHC *DRB*-related genes for intragenic segmental loss of specific hypervariable sequences, consistent with evolutionary mechanisms involving exchange of *DRB* gene segments generating MHC diversity.

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.

MATERIALS AND METHODS

DRBFP1 (forward) and DRBRP2 (reverse) oligonucleotide primers were used for primary amplification and sequencing of *DRB* loci in humans and nonhuman primates. Genomic DNA was amplified by PCR for 29 cycles at 55°C, 72°C, and 93°C for 30 sec each. Amplification products were ligated into the TA cloning vector (Invitrogen) before sequencing by the dideoxy termination method. As described, additional allele-specific primers were then derived from these sequences. Nested reverse primers NHP-RØ, NHP-R1, and NHP-R2 were used in conjunction with nested forward primers NHP-01, NHP-02, NHP-03, and NHP-04 to confirm the "short" MHC sequences by independent amplification from genomic DNA. Primers used included: NHP-01, TGGAGCAGGCTAAGTGTAAG; NHP-02, TTCTTGGAGTAGGCTAAGTGT; NHP-03, GGAGCAGGCTAAGTGTGAG; NHP-04, TTGGAGCAGGCTAAATATGAG; NHP-RØ, TGTAACTCTGTGACAGGCCA; NHP-R1, TGTAACTCTGTGACAAGCCG; NHP-R2, TTCCGTAATTGTAACTCTGTGA; DRBFP1, CCCCACAGCACGTTTCTTG; DRBRP2, CCGCTGCACTGTGAAGCTCT.

RESULTS AND DISCUSSION

Over 50 *DRB*-like sequences from a wide variety of nonhuman primates were derived by amplification of the second exon homologs using *DRB1* consensus primers. Most of the sequences gave the expected 270-bp product that aligned with *DRB* second exon sequences. However, four short variants of 208 bp were identified in macaques; these sequences were also homologous to portions of the *DRB* second exon with a 62-bp gap and were observed in two species of macaques, *Macaca fascicularis* and *Macaca nemestrina* (Fig. 1). The deleted nucleotide sequences corresponded to codons 60–80 of the *DRB* second exon, also creating a frame shift at the junction flanking this gap. Additional polymorphisms distinguished the four short variants, consistent with continued divergence of these sequences subsequent to a common ancestral deletional event.

A very similar sequence was also identified in a baboon *DRB*-like gene shown in Fig. 1, designated *Paca-DRB6*pss02*. Again, codons 60–80 were specifically deleted with the same frame-shift mutation at this site. Additional polymorphisms within the baboon-derived sequence, particularly at codon 35–37, indicate that this gene also contains additional mutations, consistent with continued accumulation of deleterious nucleotide changes.

Abbreviations: MHC, major histocompatibility complex; HVR, variable region.

Data deposition: The sequences reported in the paper have been deposited in the GenBank data base (accession nos. L76641–L76651, L76675–L76695, L76720–L76725, L76980–L76981, L77100–L77105, and L77110–L77112).

[§]To whom reprint requests should be addressed at: Puget Sound Blood Center, 921 Terry Avenue, Seattle, WA 98104-1256.

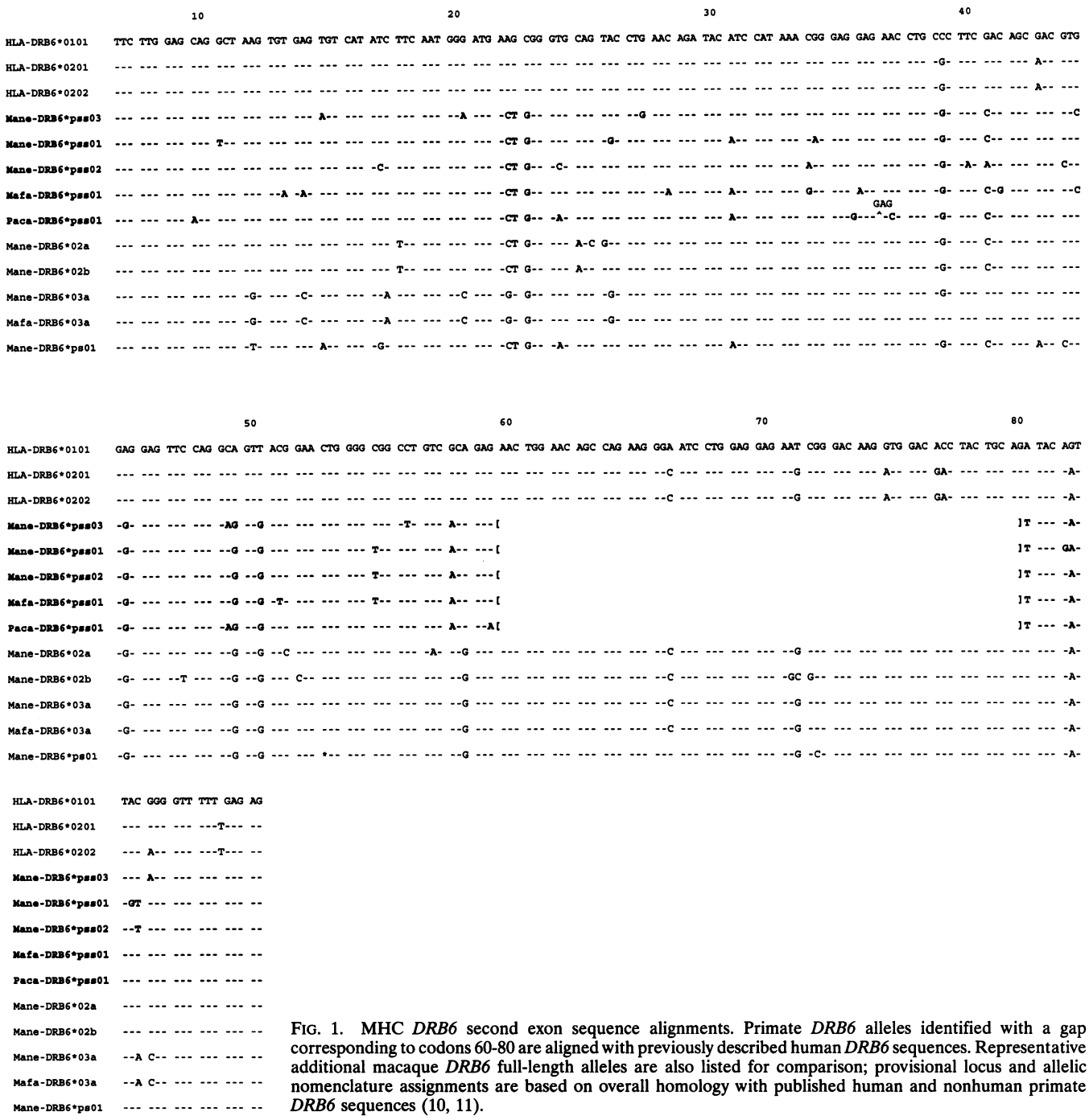

FIG. 1. MHC *DRB6* second exon sequence alignments. Primate *DRB6* alleles identified with a gap corresponding to codons 60-80 are aligned with previously described human *DRB6* sequences. Representative additional macaque *DRB6* full-length alleles are also listed for comparison; provisional locus and allelic nomenclature assignments are based on overall homology with published human and nonhuman primate *DRB6* sequences (10, 11).

To confirm the presence of the short sequences in genomic DNA, additional primer oligonucleotides (NHP-RØ, NHP-R1, NHP-R2) were constructed that annealed to the predicted junctional sequence linking codons 59 and 81 when codons 60–80 are deleted. Specifically amplified DNA of the predicted length was obtained from all four macaque samples and nucleotide sequences confirmed the previously identified genes.

Because the deleted gene segments corresponded to most of the distal portion of the second exon, sequences in the proximal portion of the exon were used for alignment comparisons. Conserved lineage motifs in this region for each of the five deleted *DRB* sequences corresponded to homology with the *DRB6* locus. This alignment is shown in Fig. 1. *DRB6* genes have been identified in prosimians and most primate species, suggesting that it is one of the oldest ancestral *DRB* homologs (11). In modern primate and human lineages, *DRB6* analogs are pseudogenes with, for example, deletions of the first exon commonly found (12). Indeed, one of the short *DRB6* sequences observed in the macaques contains an additional sequence polymorphism at codon 10 that encodes a stop codon, consistent with the interpretation that these represent ancestral pseudogenes.

DNA from a variety of human MHC haplotypes was also studied, using both the amplification primers originally used in this study as well as primers corresponding to the predicted junctional sequences; no short human homologs were identified.

The boundaries of the deleted DNA segment from codon 60–80 are of interest because they have a specific topological alignment within the *DRB* molecule. Codons 7–54 encode a β-pleated sheet platform structure underlying the MHC-antigen binding groove, and codons 60–80 encode a long α

helical loop that forms the lateral boundary of the antigen binding MHC groove (13). Previous sequence comparisons of full-length *DRB* genes have suggested that these two regions have distinct evolutionary histories. Statistical analyses of the relative frequencies of replacement and silent substitutions (14), as well as phylogenetic analysis (4, 15), have been used to argue that the proximal (β sheet portion) second exon contains most of the stable allelic lineage-related sequence motifs, with a slow conventional accumulation of additional mutations over time. In contrast, the distal (α helical portion) second exon represents a fixed sequence element that occurs in the context of multiple different allelic lineages. We believe we have identified in the short *DRB6* macaque and baboon sequences direct evidence consistent with a mechanism of intergenic

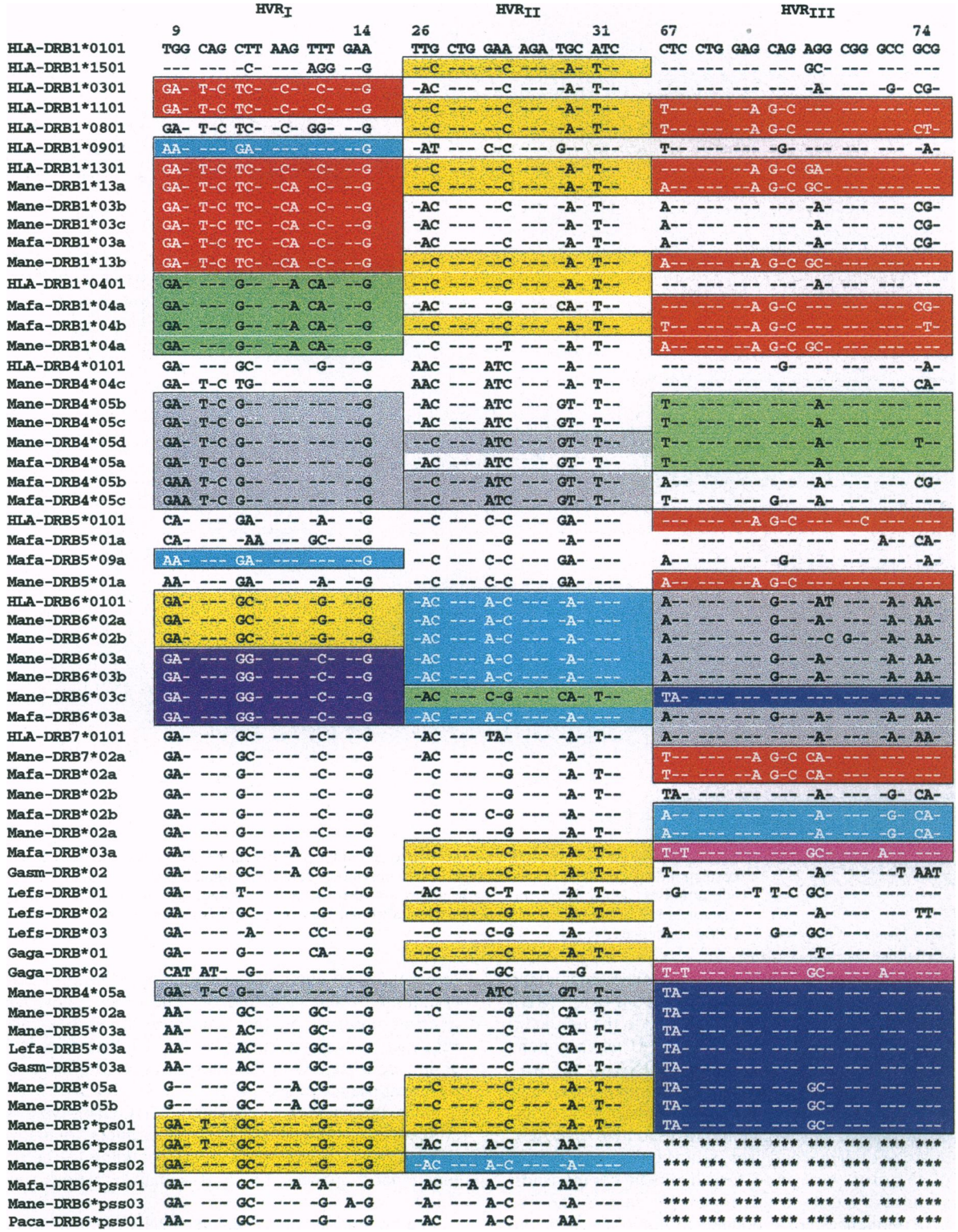

FIG. 2. Selected human (HLA), macaque, baboon, and prosimian (lemur and galago) sequence alignment representing the three variable regions (HVR) within the *DRB* second exon. Mane, *M. nemestrina*; Mafa, *M. fascicularis*; Lefa, *Lemur fulvus albifons*; Lefs, *L. fulvus sanfordi*; Gaga, *Galago garnetti*; Gasm, *Galago senegalensis moholi*; Paca, *Papio cynocephalus anubis*. The allelic designation for the primates is provisional; allelic designation "ps" indicates a pseudogene with a stop codon and "pss" indicates pseudogenes with a short sequence due to the 62-bp gap in the α helical region.

segmental interchange for the distal (α helical portion) *DRB* second exon. The short macaque *DRB6* pseudogenes appear to represent the donor sites from such an exchange, which in these cases was apparently nonreciprocal, leading to the evolutionary remnants found.

Segmental recombination at this site in *DRB* second exons is presumably not limited to the *DRB6* pseudogene. Sequence identities between the α helical portion of the *DRB* molecule among different alleles have been noted by many investigators and are present in most of the functional allelic lineages in modern primates and man. For example, the human *DRB1*1414* gene sequence is consistent with a recombination between *DRB1*1404* and *DRB1*0802* or *DRB1*0804* alleles at this second exon site (16), and similarly the *DRB1*0415* sequence may derive from a *DRB1*0401/DRB1*11* recombinant (17). Analysis of additional sequences derived from nonhuman primates identified some striking examples that also appear to represent this diversification mechanism. As shown in Fig. 2, polymorphism among *DRB* genes can be viewed as a patchwork of distinct clusters of variable sequence elements (HVR) that occur in three sites, including the HVR_{III} cluster encoding residues 67–74 within the α helical loop region. Conserved sequence motifs from this region occur interspersed among loci, alleles, and even between species, as highlighted in Fig. 2.

An example of this segmental conservation is shown for the HVR_{III} sequence TAC CTG GAG CAG AGG CGG GCC GCG (Fig. 2, dark blue), which is found in *DRB4*-like, *DRB5*-like, and *DRB6*-like loci in several different alleles distributed across macaque, lemur, and galago species, and is present in the context of a least four different HVR_I and HVR_{II} sequences. Although it is not possible to definitively identify the codon 60–80 segment currently "missing" from the short macaque *DRB6* sequences, it may be noteworthy that some contemporary *DRB6* genes have identical HVR_I and HVR_{II} segments (i.e., compare *Mane-DRB6*pss02* with *Mane-DRB6*02a* and *Mane-DRB6*02b*), suggesting the possibility that the *DRB6*02* HVR_{III} sequence ATC CTG GAG GAG AAG CGG GAC AAG (Fig. 2, grey) could correspond to the hypervariable segment deleted in the short sequences. If this is the case, we can speculate that potential acceptor genes for this sequence, if indeed it was donated in a nonreciprocal exchange, might be the macaque *DRB6*03a* or even the human *HLA-DRB7* pseudogene sequences, which share contemporary HVR_{III} homologs. Most likely, this form of segmental exchange is frequently reciprocal, yielding gene products of normal length; the existence of the short *DRB6* homologs may represent an exception to this reciprocal recombination that, nevertheless, provides direct evidence for this form of diversification mechanism.

Strong selective pressures encourage both diversification of MHC *DRB* alleles within a species as well as conservation of functionally successful motifs, not only within a species but also between species (18, 19). It is possible that a mechanism of segmental recombination for the α helical portion of the second exon represents an evolutionary adaptation favoring the exchange of successful structural motifs in this region among different alleles. Indeed, Gyllensten *et al.* (15) have noted that *DRB* sequences between codons 51–55 are partially homologous to bacterial χ-like recombination signals and have suggested this as a basis for the distinct evolutionary histories of the β-sheet region compared with the α helical portions of the second exon (4). Because this χ-like sequence occurs in a nonpolymorphic portion of the exon, it raises the possibility that conservation of a mechanism for localized segmental exchange is one of the underlying evolutionary features accounting for MHC *DRB* diversity.

Localization of *DRB1* segmental variation to the second exon implies functional selection mechanisms, because this segment encodes the class II MHC domain that interacts with highly variable peptide antigens. Indeed, the comparable functional site in the third exon of the MHC class I *HLA-B* locus has also been suggested to arise by similar mechanisms, based on sequence comparisons that suggest recombinant origins (20). As pointed out by others (21), there are also examples of sequence variation in *DRB* genes consistent with concepts of convergent evolution. However, the specific localization of the deleted *DRB6* gene segment described here is most consistent with the interpretation that the evolutionary history of the α helical portion of the second exon likely involves specialized recombinational events.

We thank Kathy Snyder, Jeanne Anderson, and Emily Schultz for their excellent technical assistance, and Nicolette Ducommun for preparing the manuscript. We also thank the Washington Regional Primate Center (Seattle) for the old world monkey blood samples and Duke Primate Center (Durham, NC), for the prosimian blood samples. This work is supported in part by a grant to the Washington Regional Primate Research Center from the National Institutes of Health (RR00116).

1. Lundberg, A. S. & McDevitt, H. O. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 6545–6549.
2. Marsh, S. G. E. & Bodmer, J. G. (1992) *Tissue Antigens* **40,** 229–243.
3. Fan, W., Kasahara, M., Gutknecht, J., Klein, D., Mayer, W. E., Jonker, M. & Klein, J. (1989) *Human Immunol.* **26,** 107–121.
4. Gyllensten, U. B., Sundvall, M. & Erlich, H. A. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 3686–3690.
5. Hayashida, H. & Miyata, T. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 2671–2675.
6. Lopez de Castro, J. A., Strominger, J. L., Strong, D. M. & Orr, H. T. (1992) *Tissue Antigens* **79,** 3813–3817.
7. Ohta, N., Nishimura, Y. K., Tanimoto, K., Horiuchi, Y., Abe, C., Shiokawa, Y., Abe, T., Katagiri, M., Yoshiki, T. & Sasazuki, T. (1982) *Human Immunol.* **5,** 123–132.
8. Mellor, A. L., Weiss, E. H., Ramachandran, K. & Flavell, R. A. (1985) *Nature (London)* **306,** 792–795.
9. Klein, J. & O'hUigin, C. (1995) *Immunol. Rev.* **143,** 89–111.
10. Marsh, S. G. E. & Bodmer, J. G. (1995) gopher://sta08.swmed.edu:70/11c:/seq.
11. Figueroa, F., O'hUigin, C., Inoki, H. & Klein, J. (1991) *Immunogenetics* **34,** 324–337.
12. Corell, A., Martin-Villa, J. M., Morales, P., deJuan, M. D., Varela, P., Vicario, J. L., Martinez-Laso, J. & Arnaiz Villena, A. (1991) *Mol. Immunol.* **28,** 533–543.
13. Chicz, R. M., Urban, R. G., Lane, W. S., Gorga, J. C., Stern, L. J., Vignali, D. A. A. & Strominger, J. L. (1992) *Nature (London)* **358,** 764–768.
14. Sigurdardottir, S., Borsch, C., Gustafsson, K. & Andersson, L. (1992) *J. Immunol.* **148,** 968–973.
15. Gyllensten, U., Sundvall, M., Ezcurra, I. & Erlich, H. A. (1991) *J. Immunol.* **146,** 4368–4376.
16. Fogdell, A. & Olerup, O. (1994) *Tissue Antigens* **43,** 327–239.
17. Tiercy, J.-M., Gebuhrer, L., Betuel, H., Mach, B. & Jeannet, M. (1993) *Tissue Antigens* **41,** 97–101.
18. Klein, J. (1987) *Human Immunol.* **19,** 155–162.
19. Hughes, A. L. & Nei, M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 958–962.
20. McAdam, S. N., Boyson, J. E., Liu, X., Garber, T. L., Hughes, A. L., Bontrop, R. E. & Watkins, D. I. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 5893–5897.
21. Hughes, A. L. & Nei, M. (1989) *Genetics* **122,** 681–686.