Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 5893–5897, June 1994
Evolution

A uniquely high level of recombination at the *HLA-B* locus

(chimpanzee/bonobo/major histocompatibility complex class I/evolution)

STEPHEN N. MCADAM[†‡], JONATHAN E. BOYSON[†], XIAOMIN LIU[§], THEODORE L. GARBER[†], AUSTIN L. HUGHES[¶], RONALD E. BONTROP[‖], AND DAVID I. WATKINS[†**]

[†]Wisconsin Regional Primate Research Center and Department of Pathology, University of Wisconsin, 1220 Capitol Court, Madison, WI 53715; [§]University of Maryland at Baltimore, 655 West Baltimore Street, Bressler 13, Baltimore, MD 21201; [¶]Department of Biology, 208 Mueller Laboratory, Pennsylvania State University, University Park, PA 16802; and [‖]Nederlandse Centrale Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek Medical Biological Laboratories, P.O. Box 5815, 2280 HV Rijswijk, The Netherlands

Communicated by Robert Selander, February 22, 1994

ABSTRACT **Major histocompatibility complex (MHC) loci are some of the most polymorphic genes in the animal kingdom. Recently, it has been suggested that although most of the human MHC loci are relatively stable, the *HLA-B* locus can undergo rapid changes, especially in isolated populations. To investigate the mechanisms of *HLA-B* evolution we have compared the sequences of 19 *HLA-B* homologues from chimpanzees and bonobos to 65 *HLA-B* sequences. Analysis of the chimpanzee and bonobo *HLA-B* homologues revealed that despite obvious similarities between chimpanzee and human alleles in exon 2, there was little conservation of exon 3 between humans and the two chimpanzee species. This finding suggests that, unlike all other *HLA* loci, recombination has characterized the *HLA-B* locus and its homologues for over 5 million years.**

The products of the polymorphic (1) classical human major histocompatibility complex (MHC) class I loci (*HLA-A, -B, -C*) are molecules that bind peptides (2–4) and present them to $CD8^+$ T cells. The T-cell receptor of the $CD8^+$ T cell binds to the peptide/MHC complex, triggering destruction of the target cell. β_2-Microglobulin-deficient mice that do not express MHC class I molecules show delayed clearance of viruses and are susceptible to a variety of other pathogens (5–10). Selection pressure appears to have been responsible for the maintenance of diversity at the amino acid residues that line the Ag-recognition site of MHC class I molecules (11). This diversity is thought to increase the number of possible different viral or tumor antigens that can be bound and presented to cytotoxic T lymphocytes. Thus, the products of these polymorphic classical MHC class I loci are critical to the surveillance function of an intact immune system.

Despite the enormous polymorphism of the MHC loci, allelic lineages of some of these MHC loci have been maintained over millions of years (12, 13). Comparison of *HLA-A* alleles and alleles of the orthologous locus of chimpanzees have provided evidence that the origin of such lineages may predate speciation events (14–16). All known chimpanzee *A* locus alleles belong to one of the six families of alleles found at the *HLA-A* locus, indicating that this allelic lineage originated prior to the divergence of humans and chimpanzees (5–7 million years ago). By contrast, the relationship between alleles at the highly polymorphic *HLA-B* locus and its chimpanzee homologue have not been easy to determine on the basis of the limited data available to date (17). To understand how evolution of this locus has occurred, we cloned and sequenced 18 *HLA-B* homologues from nine chimpanzees and two bonobos and compared these alleles to their human counterparts (Table 1).[††]

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

PCR and Cloning of Chimpanzee and Bonobo *HLA-B* Homologues. PCR from cDNA or DNA was carried out as described (18, 19). Briefly, RNA was extracted from 2–5 × 10^6 lymphocytes using oligo(dT) bound to magnetic beads and DNA was extracted using standard techniques (20). For full-length amplification from cDNA, the LPHIII and MM2 primers were used (21). For partial-length amplification from RNA and DNA, GCBH3 [5′-GCAAGCTTGACGACAC(G/C)C(A/T)GTTCGTGA-3′] and NuA2ERI [5′-GCGAATTC-CAGC(G/T)T(G/C)TCCTTCCCGTTCTC-3′] were used. PCR products amplified from cDNA or genomic DNA were ligated into pSP65 and at least three full-length copies of each insert were sequenced to avoid PCR errors (22). At least four MHC class I-specific sequencing primers were used in each direction to generate double-stranded sequence and compressions were resolved using dITP. Sequences were analyzed using software from IBI.

Tree Construction. The trees were constructed by the neighbor-joining method (23) based on number of nucleotide substitutions per site (*d*) (24). Standard errors of branch lengths were estimated by Rzhetsky and Nei's method (25).

RESULTS

Similarity Between Chimpanzee, Bonobo, and Human α1 Domains. Comparison of 19 chimpanzee (*Pan troglodytes*) *Patr-B* and bonobo (*Pan paniscus*) *Papa-B* alleles to 65 *HLA-B* sequences demonstrated that there were striking similarities between the α1 domains of *HLA-B* sequences and their chimpanzee and bonobo orthologues (Fig. 1). For example, *Papa-B*01* and *Papa-B*04* differed from *HLA-B*0702* by only three and two residues in the α1 domain, respectively. Based on sequence similarity in the α1 domain and serological cross-reactivity it was possible to identify only four major groups of chimpanzee and bonobo *HLA-B* homologues: HLA-B15, HLA-B48, HLA-B57/58, and HLA-B27/7.

The α2 Domain Is Not Conserved Between Chimpanzees and Bonobos and Humans. The interspecies similarity of the α1 domains of the *B* locus alleles was not continued in the rest of the molecule (Figs. 1 and 2*B*). Unlike the tree of exon 2

Abbreviations: MHC, major histocompatibility complex; HLA, human leukocyte antigen.

[‡]Present address: Molecular Immunology Group, Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, U.K.
**To whom reprint requests should be addressed.

[††]The sequences reported in this paper have been deposited in the GenBank data base (accession nos. U05575–U05587).

Table 1. Origin and serological reactivity of chimpanzee and bonobo *HLA-B* homologues

Individual	Colony	Colony/ wild born	Clone	Allele	Serology		Similarity to α1*
					Human	Chimp	
Hugo	TNO	CB	ChLA-B1	*Patr-B*01*		—	B17
			ChLA-B2	*Patr-B*02*	B53	118	B48
Tank	Yerkes	CB	Ch-18	*Patr-B*04*	B40		B48
			Ch-39	*Patr-B*03*	B17		B17
Colin	Yerkes	CB	Ch-11	*Patr-B*05*			B48
Kasey	Yerkes	CB	Ch-7	*Patr-B*06*			B48
Harv	Yerkes	WB		*Patr-B*06*			B17
				*Patr-B*07*			B17
Teppie	Yerkes	WB		*Patr-B*08*			B15
				*Patr-B*09*			B17
Renee	TNO	WB		*Patr-B*10*	B17	125	B17
Victoria	TNO	WB		*Patr-B*11*	B07	106	B07
				*Patr-B*12*	B53	118	B07
Toetie	TNO	WB		*Patr-B*10*	B17	125	B17
Noel	TNO	WB		*Patr-B*10*		125	B17
Wodka	TNO	WB		*Patr-B*14*	B38/37	117	B07
				*Patr-B*13*		106	B07
1028	Duke			*Patr-B*15*	B40		B48
				*Patr-B*09*	B27		B17
Lorel	Yerkes	WB		*Papa-B*01*			B07
				*Papa-B*03*			B27
Bosondjo	Yerkes	WB		*Papa-B*02*			B27
				*Papa-B*04*			B07

*Similarity to α1 domain of *B* locus alleles indicated.

(Fig. 2*A*), the gene tree of exons 3–8 resulted in completely different clustering among human, chimpanzee, and bonobo *B* locus sequences (Fig. 2*B*). In this tree, the *HLA-B* homologues of chimpanzees and bonobos clustered separately from their human counterparts. By contrast, in similar trees of *A* locus alleles, the great ape *HLA-A* homologues clustered with *HLA-A1, -A3, -A11* in trees of exon 2 and exons 3–8 (data not shown). Despite being very similar in the α1 domain, the *HLA-B7*-like bonobo alleles, *Papa-B*01* and *Papa-B*04*, each differed from *HLA-B*0702* in the α2 domain by more than nine amino acid substitutions. Recombination of motifs between polymorphic *B* locus alleles and a number of *Patr*- and *Papa-B*-specific amino acid substitutions differentiated the chimpanzee and bonobo sequences from their human counterparts. Many of the unique substitutions in the chimpanzee and bonobo *B* alleles (i.e., substitutions not found in any *HLA-B* allele) were present in alleles of the *HLA-A1/-A3/-A11* lineage (position 114, E and R; 151, H; 152, W and A).

The *HLA-A* and *HLA-B* Loci Evolve Differently. Analysis of human and chimpanzee MHC class I alleles reveals significant differences between mechanisms of evolution of the *A* and *B* loci. Pairwise comparisons between exons 2 and 3 of *HLA-A, Patr-A*, and *Papa-A* alleles demonstrate a correlation between the number of nucleotide substitutions in exons 2 and 3 (Fig. 3*A*)—that is, if *A* locus alleles differ by large numbers of substitutions in exon 2 they will also differ by large numbers of substitutions in exon 3, indicating that in most *A* locus alleles exons 2 and 3 share a common evolutionary history. No such correlation was seen in a similar analysis of *HLA-B, Patr-B*, and *Papa-B* alleles (Fig. 3*B*). This provides additional support for the hypothesis that intralocus recombination has occurred in exon 3 of *B* locus alleles serving to reassort polymorphic motifs between *B* locus alleles in exon 3.

DISCUSSION

Comparison of chimpanzee and human *B* locus genes suggests that *B* locus alleles may be an exception to the hypothesis of trans-species evolution (12, 13). While homologues of the *HLA-A* allelic lineages are conserved in chimpanzee, bonobos, and humans (14–16), the *HLA-B* locus allelic lineages have been scrambled by recombination. Thus, the pattern of evolution observed at the *B* locus does not support the idea that all MHC loci evolve slowly (12, 13). Although there is evidence that diversity at the class II *DRB* loci in primates has been enhanced by recombination events, *HLA-DRβ* and *Patr-DRβ* alleles of the same lineage cluster together in phylogenetic trees (20, 27, 28). Indeed, *Patr-DRβ4*0104* and *HLA-DRβ4*0104* differ at only three residues in the α1 domain (20), and the α2 domains of *DRβ* alleles are conserved between chimpanzees and humans (29). This sharing of trans-species *DRB* sequence lineages is even observed between humans and rhesus macaques (30). Furthermore, a rhesus monkey *Mamu-DRβ1*03* allele can present a peptide from the 65-kDa heat shock protein of *Mycobacterium tuberculosis/leprae* to a human T-cell clone restricted by *HLA-DR17* (31). Likewise, homologues of *HLA-A* (14–16), *-DRα*, (29), *-DRβ*, (20, 27, 28), *DQα* (32, 33), and *-DQβ*, (34, 35) alleles in great apes evolve in a trans-specific fashion.

The presence of *HLA-A1/-A3/-A11* specific substitutions in the chimpanzee and bonobo *HLA-B* homologues suggests a role for interlocus as well as intralocus recombination in the generation of diversity at the *B* locus. Thus, the great ape and human *B7*-related alleles probably derived from an *HLA-B*07*-like ancestral gene that gave rise to *HLA-B*0702* in humans and *Papa-B*01* and *Papa-B*04* in bonobos. Subsequent to their divergence from the ancestral gene, both inter- and intralocus recombination in exon 3 encoding the α2 domain served to create new *B* locus alleles in humans, chimpanzees, and bonobos. Interestingly, it was sequence changes in exon 3 that were largely responsible for the creation of new *HLA-B* alleles found in South American tribes (21, 36).

The fact that an unusually high proportion of alleles at the *B* locus are products of recombination need not imply that this locus is a "hotspot" for recombination. If recombinant alleles have been selectively favored at the *B* locus to a

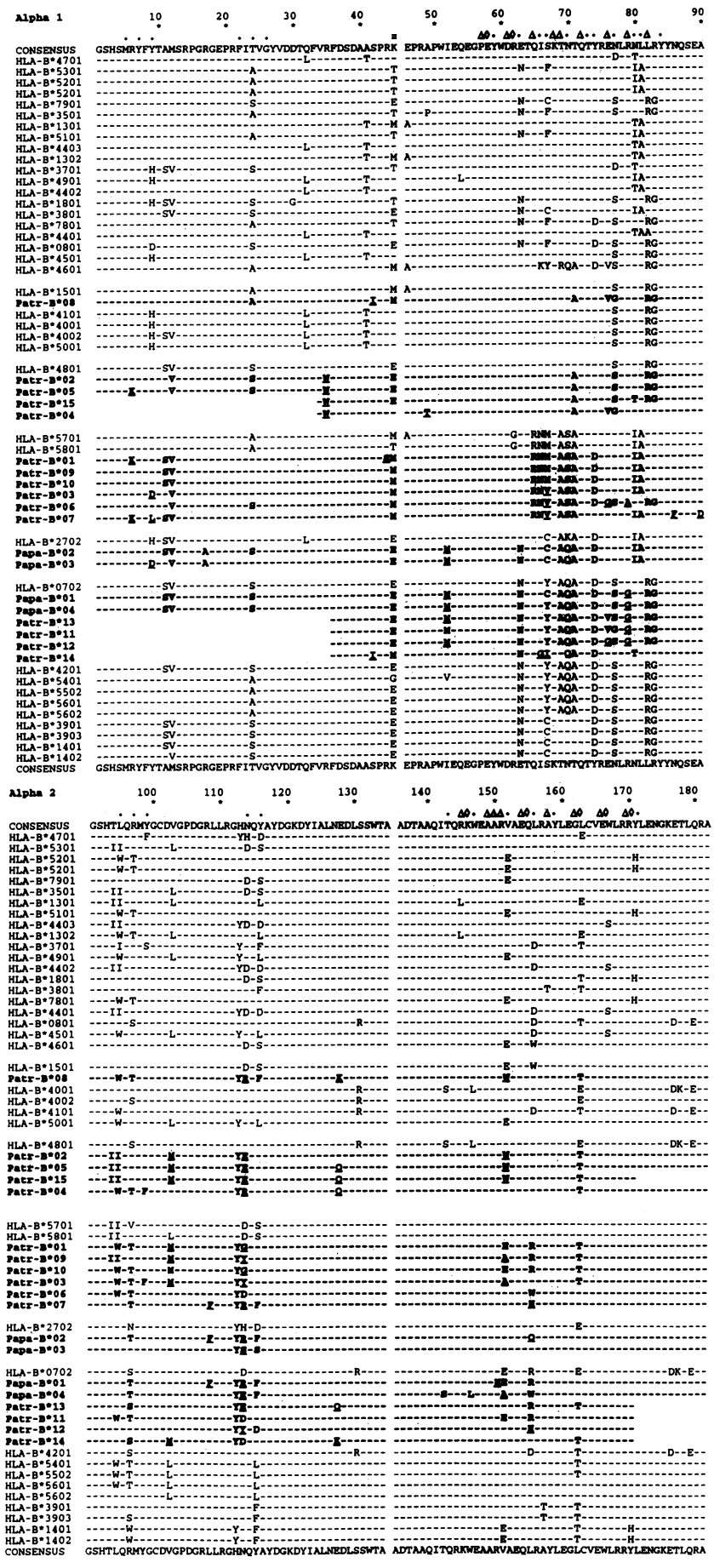

FIG. 1. Predicted amino acid sequences of the chimpanzee and bonobo *HLA-B* homologues compared to *HLA-B* sequences. A consensus sequence of the α1 and α2 domain is shown and the chimpanzee and bonobo *HLA-B* homologues (in boldface type) and *HLA-B* alleles are compared; identity is indicated with a dash. Any unique amino acid substitutions in the bonobo and chimpanzee *B* alleles not found in previously sequenced human *HLA-B* alleles are indicated by underlining them. Amino acids that point into the Ag-recognition site (●), point both into the Ag-recognition site and up toward the T-cell receptor (◇), and point up toward the T-cell receptor (△) are indicated (3). Residue 45 of the α1 domain, which affects peptide binding in the B pocket, is also indicated (□).

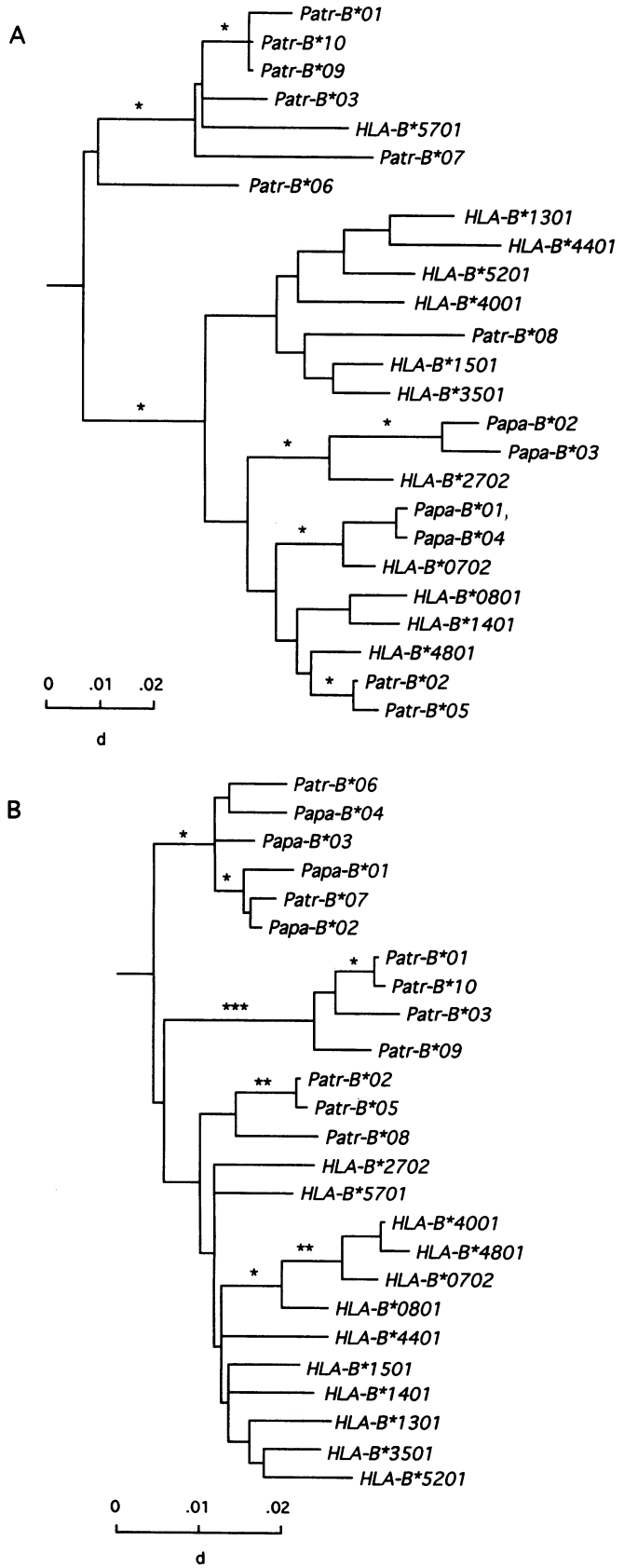

FIG. 2. Phylogenetic trees for exon 2 (*A*) and exons 3–8 (*B*) of *B* locus alleles from human (*HLA-*), chimpanzee (*Patr-*), and bonobo (*Papa-*). Statistically significant internal branches are indicated as follows: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. Trees were rooted using *HLA-A*, *Patr-A*, and *Papa-A* sequences.

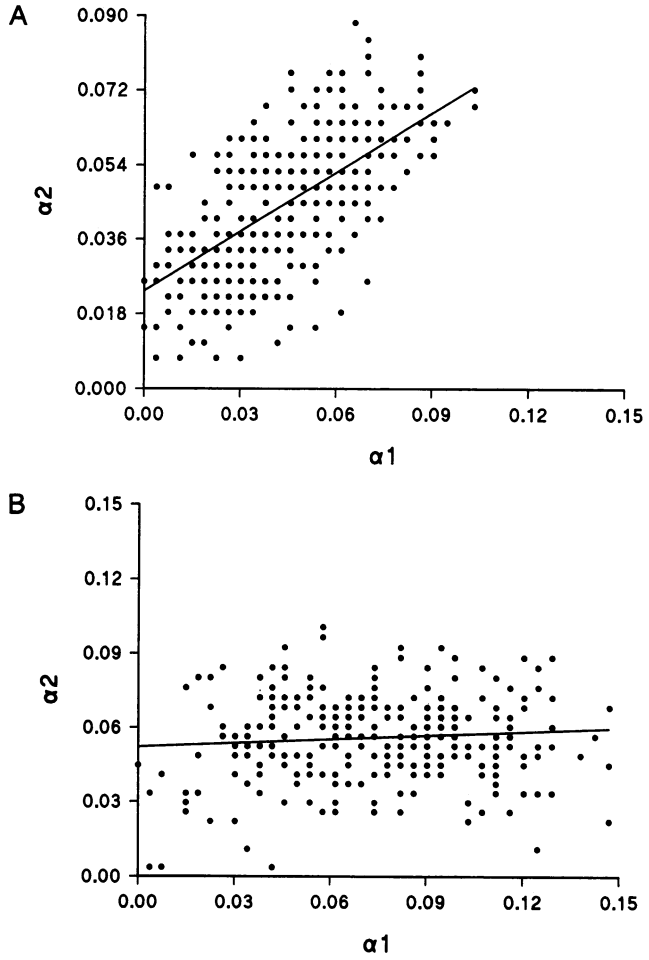

FIG. 3. Number of nucleotide substitutions per site (24) in exons encoding the α2 domain as a function of that of exons encoding the α1 domain for all pairwise comparisons among human, chimpanzee, and bonobo *A* locus alleles (*A*) and *B* locus alleles (*B*); human sequences used are as in Fig. 2. The lines shown are linear regression lines: (*A*) $y = 0.024 + 0.479x$; (*B*) $y = 0.057 + 0.050x$. Pairwise comparisons of the number of nucleotide substitutions per site, *d* (24), in exons encoding the α2 domain were plotted as a function of the number of nucleotide substitutions per site in exon 2 (encoding the α1 domain). Because pairwise comparisons are not strictly independent, in order to have a conservative test of the null hypothesis that the correlation coefficient is equal to zero, we tested with $n - 2$ df, where n is the number of sequences compared rather than the number of pairwise comparisons. In *A*, $r = 0.612$ ($P < 0.001$); in *B*, $r = 0.094$ [not significant (NS)]. Similar values were obtained when the number of synonymous substitutions per site (26) in exons encoding the α2 domain was correlated with that of exons encoding in the α1 domain—namely, for the *A* locus $r = 0.562$ ($P < 0.01$), and for the *B* locus $r = 0.056$ (NS). Similar values were also obtained for the correlation of the number of nonsynonymous nucleotide substitutions per site in exon 2 with that in exon 3—namely, for the *A* locus $r = 0.551$ ($P < 0.01$) and for the *B* locus $r = 0.075$ (NS).

greater extent than they have at other MHC loci, recombinant alleles are expected to be observed more frequently at the *B* locus than at other loci even if the rate of recombination at the *B* locus is similar to that at other MHC loci. An analogous phenomenon has been observed in bacteria and other microorganisms (37–39), in which recombinants are much more likely to be observed at loci where diversity is selectively favored.

The conservation of the α1 domain between chimpanzee, bonobo, and human *B* locus alleles implies that the peptide binding pockets encoded by this domain may be under strong structural and functional constraints. In many of the peptide binding motifs analyzed, it appears that the anchoring amino acid is bound by the second, or B, pocket (40–42). The amino acids lining this pocket are largely made up of residues

encoded for by exon 2. Thus, the peptide binding motifs of the chimpanzee, bonobo, and human *B* locus alleles may be very similar, with likely conservation of the anchor residues between these two species. The difference in the $\alpha 2$ domains may, however, change some of the amino acids accommodated by pockets D–F but not change the anchor amino acid of the peptides bound by chimpanzee, bonobo, and human *B* locus alleles. Perhaps the advantage of being able to adapt rapidly to changing pathogens selects for the binding of slightly different peptides. An example of this may be the differential peptide binding abilities of *HLA-B35* and *HLA-B53*, which differ by only five amino acids at the end of the $\alpha 1$ domain. The anchor residue of proline is conserved between the peptides bound by these two alleles, yet *HLA-B53* does not require a tyrosine residue at position 9 of the bound peptide, unlike *HLA-B35*, which does. *HLA-B53*, however, can now bind to a different malarial peptide and confers selective advantage to individuals with this allele (42, 43).

We thank Peter Parham for critical review of this manuscript and the Yerkes Primate Center and Barton Haynes (Duke University, Durham, NC) for blood samples from chimpanzees and bonobos. We also thank Riet Noort for serologic typing and Nel Otting for establishing B-cell lines. We thank Lori Geurts for preparation of this manuscript. This work was supported by grants from the National Institutes of Health (RR00167 and AI32426 to D.I.W. and GM43940 to A.L.H.).

1. Parham, P., Lomen, C. E., Lawlor, D. A., Ways, J. P., Holmes, N., Coppin, H. L., Salter, R. D., Wan, A. M. & Ennis, P. D. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4005–4009.
2. Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennet, W. S., Strominger, J. L. & Wiley, D. C. (1987) *Nature (London)* **329**, 506–512.
3. Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennet, W. S., Strominger, J. L. & Wiley, D. C. (1987) *Nature (London)* **329**, 512–518.
4. Garrett, T. P. J., Saper, M. A., Bjorkman, P. J., Strominger, J. L. & Wiley, D. C. (1989) *Nature (London)* **342**, 692–696.
5. Zijlstra, M., Bix, M., Simister, N. E., Loring, J. M., Raulet, D. H. & Jaenisch, R. (1990) *Nature (London)* **344**, 742–746.
6. Bender, B. S., Croghan, T., Zhang, L. & Small, P. A., Jr. (1992) *J. Exp. Med.* **175**, 1143–1145.
7. Hou, S., Doherty, P. C., Zijlstra, M., Jaenisch, R. & Katz, J. M. (1992) *J. Immunol.* **149**, 1319–1325.
8. Muller, D., Koller, B. H., Whitton, J. L., LaPan, K. E., Brigman, K. K. & Frelinger, J. A. (1992) *Science* **255**, 1576–1578.
9. Tarelton, R. L., Koller, B. H., Latour, A. & Postan, M. (1992) *Nature (London)* **356**, 338–340.
10. Flynn, J. L., Goldstein, M. M., Triebold, K. J., Koller, B. & Bloon, B. R. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 12013–12017.
11. Hughes, A. L. & Nei, M. (1988) *Nature (London)* **335**, 167–170.
12. Klein, J. (1987) *Hum. Immunol.* **19**, 155–162.
13. Klein, J., Satta, Y., O'hUigin, C. & Takahata, N. (1993) *Annu. Rev. Immunol.* **11**, 269–295.
14. Lawlor, D. A., Ward, F. E., Ennis, P. D., Jackson, A. P. & Parham, P. (1988) *Nature (London)* **335**, 268–271.
15. Mayer, W. E., Jonker, M., Klein, D., Ivanyi, P., van Seventer, G. & Klein, J. (1988) *EMBO. J.* **7**, 2765–2774.
16. Chen, Z. W., Hughes, A. L., Ghim, S. H., Letvin, N. L. & Watkins, D. I. (1993) *Immunogenetics* **38**, 238–240.
17. Lawlor, D. A., Warren, E., Ward, F. E. & Parham, P. (1990) *Immunol. Rev.* **113**, 147-185.
18. Newman, P. J., Gorski, J., White, G. C., Gidwitz, S., Cretney, C. J. & Aster, R. H. (1988) *J. Clin. Invest.* **82**, 739–743.
19. Watkins, D. I., Garber, T. L., Chen, Z. W., Toukatly, G., Hughes, A. L. & Letvin, N. L. (1991) *Immunogenetics* **33**, 79–89.
20. Kenter, M., Otting, N., Anholts, J., Jonker, M., Schipper, R. & Bontrop, R. E. (1992) *Immunogenetics* **37**, 1–11.
21. Watkins, D. I., McAdam, S. N., Liu, X., Strang, C. R., Milford, E. L., Levine, C. G., Garber, T. L., Dogon, A. L., Lord, C. I., Ghim, S. H., Troup, G. M., Hughes, A. L. & Letvin, N. L. (1992) *Nature (London)* **357**, 329–333.
22. Ennis, P. D., Zemmour, J., Salter, R. D. & Parham, P. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2833–2837.
23. Saitou, N. & Nei, M. (1986) *Mol. Biol. Evol.* **4**, 406–425.
24. Jukes, T. H. & Cantor, C. R. (1969) in *Mammalian Protein Metabolism*, ed. Munro, H. N. (Academic, New York), pp. 21–132.
25. Rzhetsky, A. & Nei, M. (1992) *Mol. Biol. Evol.* **9**, 367–375.
26. Nei, M. & Gojobori, T. (1986) *Mol. Biol. Evol.* **3**, 418–426.
27. Gyllensten, U. B., Sundvall, M. & Erlich, H. A. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 3686–3690.
28. Mayer, W. E., O'hUigin, C., Zaleska-Rutczynska, Z. & Klein, J. (1992) *Immunogenetics* **37**, 12–23.
29. Fan, W., Kasahara, M., Gutknecht, J., Klein, D., Mayer, W. E., Jonker, M. & Klein, J. (1989) *Hum. Immunol.* **26**, 107–121.
30. Slierendregt, B. L., van Noort, J. T., Bakas, R. M., Otting, N., Jonker, M. & Bontrop, R. E. (1992) *Hum. Immunol.* **35**, 29–39.
31. Geluk, A., Elferink, D. G., Slierendregt, B. L., van Meijgaarden, K. E., de Vries, R. R. P., Ottenhoff, T. H. M. & Bontrop, R. E. (1993) *J. Exp. Med.* **177**, 979–987.
32. Gyllensten, U. B. & Erlich, H. A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9986–9990.
33. Kenter, M., Otting, N., Anholts, J., Leunissen, J., Jonker, M. & Bontrop, R. E. (1992) *Immunogenetics* **36**, 71–78.
34. Gyllensten, U. B., Lashkari, D. & Erlich, H. A. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1835–1839.
35. Otting, N., Kenter, M., van Weeren, P., Jonker, M. & Bontrop, R. E. (1992) *J. Immunol.* **149**, 461–470.
36. Belich, M. P., Madrigal, A., Hildebrand, W. H., Zemmour, J., Williams, R. C., Luz, R., Petzl-Erler, M. L. & Parham, P. (1992) *Nature (London)* **357**, 326–329.
37. Hughes, A. L. (1992) *Mol. Biol. Evol.* **9**, 381–393.
38. Nelson, K. & Selander, R. K. (1992) *J. Bacteriol.* **174**, 6886–6895.
39. Smith, N. H., Beltran, P. & Selander, R. K. (1990) *J. Bacteriol.* **172**, 2209–2216.
40. Falk, K., Rotzschke, O., Stevanovic, S., Jung, G. & Rammensee, H.-G. (1991) *Nature (London)* **351**, 290–296.
41. Jardetzky, T. S., Lane, W. S., Robinson, R. A., Madden, D. R. & Wiley, D. C. (1991) *Nature (London)* **353**, 326–329.
42. Hill, A. V. S., Elvin, J., Willis, A. C., Aidoo, M., Allsopp, C. E. M., Gotch, F. M., Gao, X. M., Takiguchi, M., Greenwood, B. M., Townsend, A. R. M., McMichael, A. J. & Whittle, H. C. (1992) *Nature (London)* **360**, 434–439.
43. Hill, A. V. S., Allsopp, C. E. M., Kwiatkowski, D., Anstey, N. N., Twumasi, P., Rowe, P. A., Bennett, S., Brewster, D., McMichael, A. J. & Greenwood, B. M. (1991) *Nature (London)* **352**, 595–600.