Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 7480–7484, August 1993
Evolution

The synonymous substitution rate of the major histocompatibility complex loci in primates

(trans-species polymorphism/allelic genealogy/mutation rate)

YOKO SATTA*, COLM O'HUIGIN*, NAOYUKI TAKAHATA†, AND JAN KLEIN*‡

*Max-Planck-Institut für Biologie, Abteilung Immungenetik, 7400 Tübingen, Germany; †National Institute of Genetics, Graduate University for Advanced Studies, Mishima 411, Japan; and ‡Department of Microbiology and Immunology, University of Miami School of Medicine, Miami, FL 33101

Communicated by Motoo Kimura, May 3, 1993

ABSTRACT **Because the divergence of many allelic lineages at the major histocompatibility complex (MHC) loci predates species divergence, standard methods of calculating synonymous substitution rates are not applicable to this system. We used three alternative methods of rate estimation: one based on the minimum number of substitutions (D_m), another on the nucleotide difference (D_{xy}), and the third on the net nucleotide difference (D_n). We applied these methods to the protein-encoding sequences of primate MHC class I (*A*, *B*, and *C*) and class II (*DRB1*) genes. To determine the reliability of the different estimates, we carried out computer simulation. The distribution of the estimates based on D_{xy} or D_n is generally much broader than that based on D_m. More importantly, the D_m-based method nearly always has the highest probability of recovering true rates, provided that D_m is not smaller than 5. Because of its desirable statistical properties, we used the D_m-based method to estimate the rate of synonymous substitutions. The rate is 1.37 ± 0.61 for *A*, 1.84 ± 0.40 for *B*, 3.87 ± 1.05 for *C*, and 1.18 ± 0.36 for *DRB1* loci, always per site per 10^9 years. Hence despite the extraordinary polymorphism, the mutation rate at the primate MHC loci is no higher than that of other loci.**

The major histocompatibility complex (MHC) is a set of genes critically involved in the initiation of immune response to parasites in vertebrate hosts (1). The functional MHC loci are characterized by high polymorphism (2) and long persistence of allelic lineages (trans-species evolution; see ref. 3). The polymorphism is maintained by balancing selection (4). The persistence of allelic lineages for periods longer than the lifetime of a species poses a problem in attempts to estimate the nucleotide substitution rate at the MHC loci. The estimation presupposes knowledge of the divergence time between orthologous genes, which is normally taken as equal to the species divergence time (e.g., see ref. 5). This assumption, however, does not apply to the MHC loci, at which many sequence differences between two alleles are known to have arisen prior to speciation. This accumulation of substitutions prior to speciation may occur both at the nonsynonymous sites in the peptide-binding region (a target of balancing selection) and at the tightly linked synonymous sites. Hence, calculations of synonymous substitution rates at MHC loci by the standard method (5) are likely to lead to a gross overestimate of the true rate. Theoretically, it should be possible to estimate MHC substitution rates by using sequence data from distantly related species that diverged before the divergence of even the oldest of the allelic MHC lineages. However, since some MHC lineages are more than 30 million years (Myr) old (6), and since extensive multiple hits at the divergent sites cannot be fully corrected, such attempts may lead to a gross underestimate of the true rate. Furthermore, the substitution rates of distantly related species may be different (7), thus adding another source of inaccuracy to the estimates. Recently, Satta and co-workers (8) developed the so-called minimum-minimum distance method, which avoids the problems associated with the standard methods of rate estimation. In this communication, we use computer simulation to define conditions under which the minimum-minimum method provides reliable estimates of substitution rates. We also examine the nucleotide difference (D_{xy}) and the net nucleotide difference (D_n) methods (9) and apply all three methods to the primate class I and class II MHC loci.

Methods

In the *minimum-minimum method* (8), a group of relatively closely related species is chosen (e.g., apes and humans). From each species, sequences of a representative sample of alleles are obtained, pairwise comparisons are made between orthologous genes, and the substitutions differentiating individual pairs are counted. In each species pair with known divergence time T, the allelic pair showing the lowest number of substitutions (D_m) is chosen, and the substitution rate is then calculated from D_m. This rate is regarded as being closest to the true rate, because the pair of alleles with the lowest D_m must have diverged closest of all allele pairs to the time of species divergence. The method may *overestimate* the substitution rate if the minimum pair of genes has diverged before the divergence of the two species, but the error can be minimized by expanding the collection of genes from different species pairs with different values of D_m and T and by choosing the minimum value among all $D_m/(2T)$ values. This minimum-minimum rate is denoted by b_{mm}. However, the method may also *underestimate* the substitution rate by focusing on the lowest number of substitutions, thereby increasing the probability that the observed number is by chance lower than the expected number of substitutions. Of course, the reliability of any method depends on the number of nucleotides compared.

In the *nucleotide difference method*, the mean number of substitutions *between* species (D_{xy}) is regarded as consisting of two components: the number of substitutions accumulated *after* speciation ($2b_{xy}T$) and the number of substitutions accumulated *before* speciation (a_{xy}). The constants a_{xy} and $2b_{xy}$ are estimated by the linear regression method for various species pairs with different T values, based on the formula, $D_{xy} = a_{xy} + 2b_{xy}T$.

The *net nucleotide difference method* assumes that the number of substitutions accumulated *within* an extant species is the same as that accumulated in the ancestral population before its split into the two species. This number (equivalent to a_{xy}) is estimated to be $(D_x + D_y)/2$, where D_x and D_y are

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.

Abbreviations: MHC, major histocompatibility complex; Myr, million years.

the number of substitutions within species x and y, respectively. This estimate is then subtracted from the number of between-species substitutions (D_{xy}) to obtain the net number (D_n) of substitutions after speciation—i.e., $D_n = D_{xy} - (D_x + D_y)/2$. The net number is used for the calculation of the substitution rate (b_n) by the regression method based on the formula $D_n = 2b_nT$. The reliability of D_{xy} and D_n methods depends on the range of the distribution of b_{xy} and b_n, as well as the accuracy in inferring the extent of ancestral polymorphism (a_{xy}). The conditions of applicability of the three methods can be determined by computer simulation.

Computer Simulation

We consider two extant species, n_1 genes sampled from one species and n_2 genes sampled from the second, and assume that the genes are under balancing selection. We consider further three variable parameters, T, b, and θ. For convenience, species divergence time T is measured in units of $2N_ef_s$ generations, where N_e is the effective population size and f_s is the scaling factor necessary to transform the neutral gene genealogy into the genealogy of genes under balancing selection (10). Hence b is a rescaled synonymous substitution rate per gene and θ is a scaled mean number of synonymous differences between alleles within a species. The θ is defined as $\theta = 4N_ef_s\mu L$, where L is the number of synonymous sites compared and μ is the mutation rate per site per generation (if N_e and f_s are fixed, different values of θ are obtained by varying the value of L). The mean number of pairwise synonymous substitutions at an MHC locus averaged over various primate species is 12.4 at the *A*, 13.8 at the *B*, and 11.4 at the *DRB1* locus, so that in the simulation we assumed θ to have these values. We considered three values of b corresponding to these three loci: the value of $b \times 10^6$ is 4.2 for the *A*, 4.7 for the *B*, and 3.0 for the *DRB1* locus. To estimate these b values, we assumed four or five different species pairs with different T values (Fig. 1). If b is correctly inferred in each replication, then the estimate, denoted by $\hat{b}$, is equal to a given value of b. In actual fact, $\hat{b}$ is a random variable and follows a certain distribution.

To determine the distribution of $\hat{b}$ values and their dependence on n_1, n_2, and θ, we generated 10^5 allelic genealogies for each combination of parameters using algorithms developed for neutral genes (11, 12). The simulations lead to three conclusions. First, the distribution of $\hat{b}$ for b_{xy} and b_n is much broader than that for b_{mm} regardless of the parameter values (Fig. 1). Second, the distribution of $\hat{b}$ is skewed in all three methods, and the most likely value is smaller than the assumed b value. Third, the mean value of $\hat{b}$ for b_{xy} and b_n is close to b, but for b_{mm}, it is smaller than b. However, if we ignore the sequence pairs with $D_m \leq 5$, not only does the mean $\hat{b}_{mm}$ value come close to b_{mm} but also the distribution of $\hat{b}_{mm}$ shows a higher peak near a given b value than the distribution of either $\hat{b}_{xy}$ or $\hat{b}_n$ (Fig. 1). This observation applies not only to the three cases shown in Fig. 1 but also to many other combinations of parameter values (data not shown). Hence, the probability of recovering the true substitution rate is highest with the D_m-based method, provided that values of $D_m \leq 5$ are ignored (for such values, the sampling error becomes too large to make an accurate estimate). Because of this provision, we could not use the large collection of sequence data compiled for the single exon encoding the peptide-binding region.

Data Analysis

We examined the entire protein-coding sequence of genes at four MHC loci, *A*, *B*, *C* (all three are class I loci), and *DRB1* (a class II locus) from seven primate species (Table 1; we treat the loci separately, in case they evolve at different

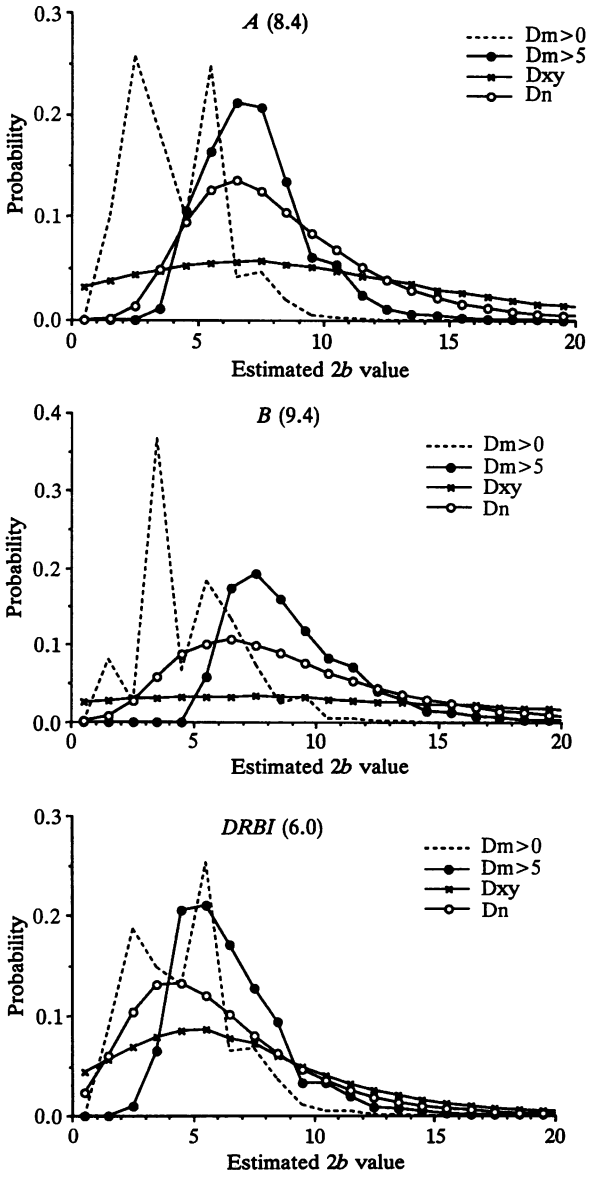

FIG. 1. Probability distribution of estimated $2b$ values based on D_m, D_{xy}, and D_n and obtained by computer simulation for particular values of T. The $2b$ values are measured in units of 10^{-6}. We assumed that $T = 0.23$, 0.36, 0.57, 0.87, and 1.15 and $n_1 = 19$ and $n_2 = 4$ for the *A* locus; $T = 0.22$, 0.35, 0.57, and 0.79 and $n_1 = 26$ and $n_2 = 4$ for the *B* locus; and $T = 0.17$, 0.28, 0.88, and 1.30 and $n_1 = 19$ and $n_2 = 1$ for the *DRB1* locus. We assumed further that $D_m > 0$ and $D_m > 5$. The assumed $2b$ value for each simulation is given in parentheses.

rates). Since inclusion in the analysis of sequences that may be the result of intragenic recombination could lead to under- or overestimation of nucleotide differences, we attempted to identify such sequences in our data by the method described in ref. 24. The method is based on the assumption that the number of nucleotide substitutions in a particular region, relative to that in the entire gene, is binomially distributed. We applied this method to the synonymous sites and substitutions only. Sequences that show significantly ($P < 0.01$) more or fewer synonymous substitutions in a particular region of a gene than expected are presumed to be anomalous, possibly the result of recombination. We therefore divided the protein-coding sequences of the class I genes into three regions, E1 plus E2, E3, and E4–E8, and those of the class II genes into two regions, E1 plus E2 and E3–E6 (where E = exon). We then compared the number of synonymous

Table 1. MHC DNA sequences used for the rate calculation and the list of anomalous alleles excluded from the data set

MHC locus	Species (ref.)	*n*	Allele(s)
A	Human (13)	19	*0101, 0205, 0206, 0210, 0301, 2501, 2601, 2901, 3101, 3201, 3301, 0201*, 0203*, 1101*, 2401*, 3001*, 6801*, 6802*, 6901**
	Chimpanzee (14, 15)	4	*01, 02, 03, 04*
	Gorilla (16)	4	*0101, 0201, 0401, 0501*
	Orangutan (17)	3	*01, 02, 03*
	Gibbon (17)	2	*01, 02*
	Macaque (18)	1	*01*
B	Human (13)	26	*0702, 1301, 1402, 2702, 2705, 3501, 3502, 3701, 4002, 4901, 5301, 5701, 0801*, 1302*, 1801*, 4001*, 4002*, 4101*, 4201*, 4401*, 4601*, 4701*, 5101*, 5201*, 5801*, 7801**
	Chimpanzee (15)	4	*01, 03, 02*, 05**
	Gorilla (16)	4	*0101, 0102, 0103, 0201*
	Orangutan (17)	3	*01, 02, 03*
	Gibbon (17)	1	*01*
C	Human (13)	4	*0101, 1401, 0201*, 1202*, 0301**, 0601***
	Chimpanzee (15)	0	*01***
	Gorilla (16)	4	*0101, 0102, 0201*, 0203*, 0202***
DRB1	Human (19)	19	*0101, 0102, 0301, 0401, 0404, 1001, 1101, 1103, 1301, 1501, 1502, 1601, 1602, 0302*, 0701*, 0802*, 0901*, 1102*, 1201**
	Chimpanzee (20)	1	*02*
	Gorilla (21)	1	*08*
	Macaque (22)	1	*01*
	Tamarin (23)	1	*0101**

n, Number of sequences used for the estimation of synonymous substitution rates. Anomalous alleles excluded at the $P < 0.05$ and 0.01 levels are indicated by * and **, respectively.

substitutions in each of these regions to that in the entire sequence for each gene pair. Both alleles of a pair with significant ($P < 0.01$) excess or shortage of substitutions in a particular region were excluded from further analysis (Table 1). When, by this exclusion, the number of sequences was reduced to one in some species, within-species variability could not be calculated for the estimation of b_n; instead, it was assumed that this variability was the same as in a partner species for which more than one sequence was available.

After the exclusion of the anomalous sequences, D_m, D_{xy}, and D_n values were calculated for the individual species pairs and plotted against species divergence times (Fig. 2). The slope of the line in the D_m-based method indicates $2b_{mm}$. In the two other methods, we computed the regression lines based on $D_{xy} = a_{xy} + 2b_{xy}T$ and $D_n = 2b_nT$. The b values thus estimated are shown in Table 2, which also provides the correlation coefficient (r) between D and T. The correlation is generally high for D_m, but it is relatively low for D_{xy} or D_n for some loci.

The b_{mm} value at the *C* locus is quite high. This is probably due to the small number of sampled genes and species. When the number of samples is small, the probability of choosing as the minimum a pair of orthologous genes that diverged long before speciation is high. Furthermore, because there is only one species pair at this locus, b_{xy} could not be calculated. Therefore we excluded this locus from subsequent consideration.

The exactness of both the D_{xy} and D_n methods depends on the accuracy of inferring the extent of ancestral polymorphism (a_{xy}). The value of a_{xy} in the D_{xy}-based method is largely determined by D_{xy} values of closely related species and it is rather insensitive to D_{xy} values for distantly related species. The method assumes constancy of a_{xy} values during evolution, but since a_{xy} depends on population size, which is likely to vary, this assumption is not justified. The substitution rate can therefore be underestimated or overestimated depending on the estimated value of a_{xy}. The D_n-based method predicts that the a_{xy} values will be the same for the pairs of species that diverged at approximately the same time. This prediction is not borne out by observations. For example, the mean number of synonymous substitutions at the *A* locus for the ancestral population of the human and gorilla pair is 2.3, while that of the chimpanzee and gorilla pair is 9.5. The reason for this discrepancy is apparently the fact that the estimates in the D_n-based method are greatly influenced by the number of sampled genes. In contrast to the D_{xy}- and D_n-based methods, the minimum-minimum method is free from troublesome estimates of ancestral polymorphism.

The synonymous substitution rate estimated as b_{mm} ranges from 1.18 to 1.84 per site per 10^9 years, depending on the locus (Table 2). As the number of sampled sequences and species increases, the estimates of b_{mm} can be expected to converge onto the lower values of the range. The synonymous substitution rate estimated as b_{xy} or b_n ranges from 1.25 to 1.66 for b_{xy} and from 0.78 to 1.55 for b_n, always per site per 10^9 years (Table 2). Although the ranges of these two estimates overlap with that of the b_{mm}, the estimate obtained by the D_m-based method is better than that based on the other two methods. The reason for this is that increased sampling can be expected to narrow the range for the D_m-based method (focus it on the lowest value of the range); no such reduction can be expected with the other two methods. Furthermore, the standard deviation of D_{xy}, which is generally larger than that of D_n, cannot be substantially reduced by increasing sample size (28). In the initial analysis, we excluded sequences that showed more or fewer substitutions than expected at the significance level of 1%. If we increase the level to 5%, we obtain b_{xy} and b_n estimates that are different from those in the initial analysis (data not shown). This result indicates that the D_{xy}- and D_n-based methods provide different rate estimates depending on the number of sampled alleles. The D_m-based method, on the other hand, is less sample size-dependent and hence more reliable.

Our estimate of the synonymous substitution rate at each of the MHC loci is lower than the previous estimates. Hayashida and Miyata (29) calculated a value of 3.0 per site per 10^9 years

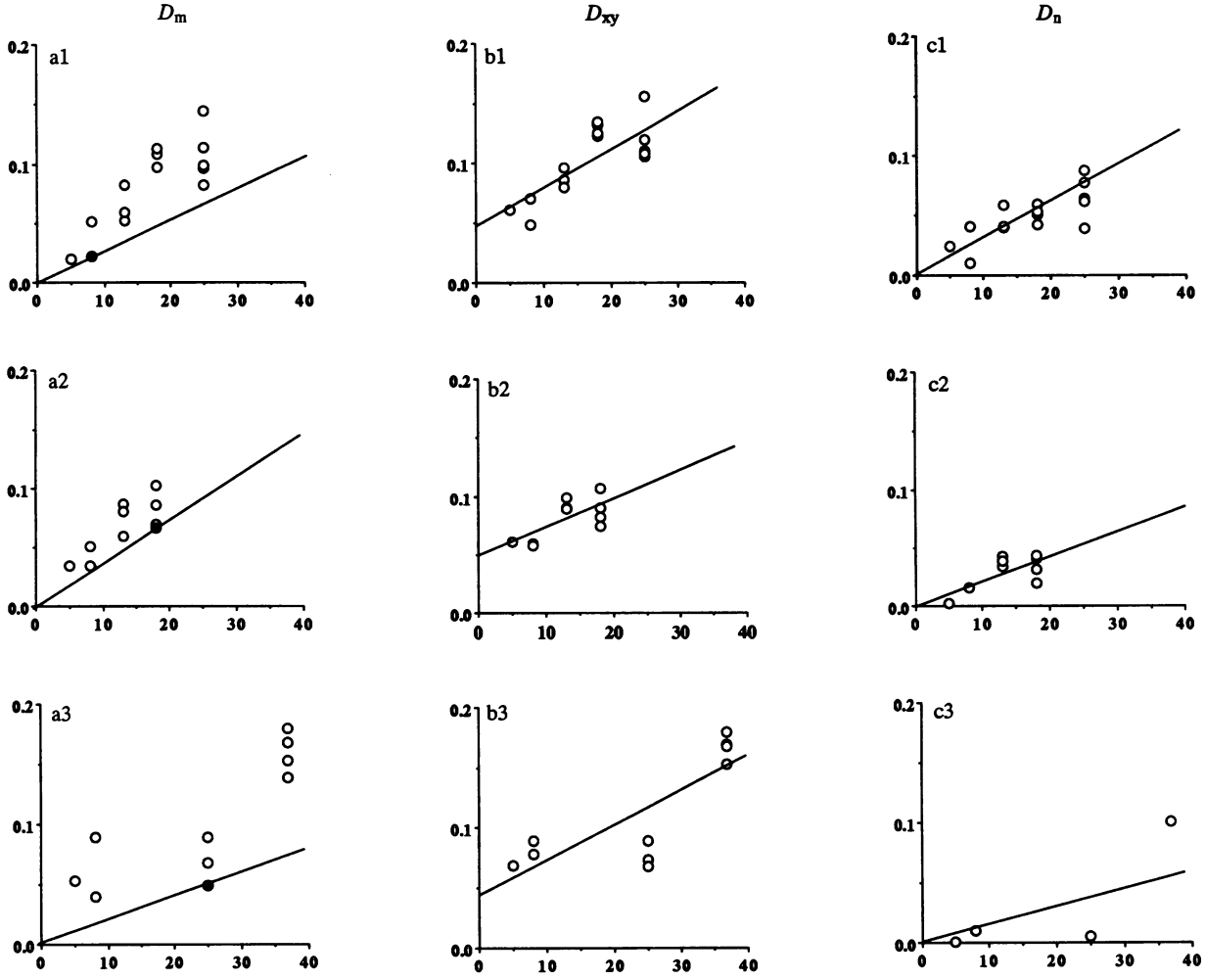

FIG. 2. Relationship between the number of synonymous substitutions per site [ordinate; D_m (a1, a2, a3), D_{xy} (b1, b2, b3), or D_n (c1, c2, c3)] and species divergence time measured in units of Myr (abscissa). From the top row to the bottom: *HLA-A, -B,* and *DRB1*. Species divergence times are human vs. chimpanzee, 5 Myr; gorilla vs. human and chimpanzee, 8 Myr; orangutan vs. human, chimpanzee, and gorilla, 14 Myr; gibbon vs. human, chimpanzee, gorilla, and orangutan, 18 Myr; macaque vs. human, chimpanzee, gorilla, orangutan, and gibbon, 25 Myr; tamarin vs. human, chimpanzee, gorilla, orangutan, gibbon, and macaque, 37 Myr (25). The slope of the line in the D_m-based method indicates $2b_{mm}$. The filled circle indicates the minimum-minimum number of substitutions that gives the estimated rate. In the two other methods, the slopes represent the regression coefficients based on $D_{xy} = a_{xy} + 2b_{xy}T$ and $D_n = 2b_nT$, respectively. The b values thus estimated are shown in Table 2, which also provides the correlation coefficient between D and T. The correlation is generally high for D_m, but it is relatively low for D_{xy} and D_n for some loci. The number of substitutions in each pair of alleles was estimated using the Jukes and Cantor correction method (26). The G+C content has an effect on the correction (27) when the number of synonymous differences per site is large because the G+C content determines the level of saturation of transitional substitutions. In fact, the G+C content at the third positions of codons at the MHC loci is about 80% and there is bias toward transitional substitutions. Details of the effect of these biases on correction of multiple hits at the synonymous sites of MHC genes will be presented elsewhere.

for class I genes, but their estimate was based on the comparison of human and mouse genes. The reason for the high estimate appears to be either the acceleration of rates due to the shorter generation time of rodents (30) or some uncertainty in the assumed species divergence times (31). Chen and co-workers (17) estimated the rate of the class I *A* locus to be

Table 2. Synonymous substitution rates (per site per 10^9 years) estimated as b_{mm} in the D_m-based method, b_{xy} in the D_{xy}-based method, and b_n in the D_n-based method

MHC loci	L	b_{mm}	b_{xy}	b_n
		Class I		
A	232	1.37 ± 0.61 (0.823)	1.66 ± 0.56 (0.771)	1.55 ± 0.60 (0.752)
B	255	1.84 ± 0.40 (0.789)	1.25 ± 0.56 (0.704)	1.09 ± 0.60 (0.680)
C	219	3.87 ± 1.05 (NA)	NA (NA)	0.77 ± 0.53 (NA)
		Class II		
DRB1	166	1.18 ± 0.36 (0.812)	1.45 ± 0.57 (0.806)	0.78 ± 2.39 (0.815)

L, Number of synonymous sites. Pairs of alleles that give the minimum D_m values are *Hosa-A*3301* vs. *Gogo-A*0401*, *Popy-B*03* vs. *Hyla-B*01*, *Hosa-C*0601* vs. *Gogo-C*0203*, and *Hosa-DRB1*0401* vs. *Mane-DRB*01* for *A, B, C* and *DRB1*, respectively. The values in parentheses are the correlation coefficients between D_m (D_{xy}, D_n) values and species divergence times. NA, not applicable. *Hosa-*, *Gogo-*, *Popy-*, *Hyla-*, and *Mane-* indicate the MHC loci of human, gorilla, orangutan, gibbon, and pig-tail macaque, respectively.

Table 3. Estimated mutation rates of humans

Rate $\times 10^8$ per site per generation	Data used for estimation	Source of data
3.9	Mean synonymous substitution rate*	Modified from ref. 34
1.0	Electrophoretic variants (36 loci)	Ref. 35
2.4	MHC *DRB1*	This paper

*This rate was calculated based on comparisons of 10 loci between Old World monkeys and humans. It was assumed that there were no substantial ancestral polymorphisms at these loci; therefore, the standard method for calculating the substitution rate was used.

3.4 per site per 10^9 years, based on the comparison of human and orangutan sequences, but they ignored the possibility that the orthologous genes diverged long before the divergence of the two species. Finally, Hughes and Nei (32) used several pairs of species, including primates, rabbit, mouse, and domestic fowl, to obtain the synonymous substitution rate of 2.0 per site per 10^9 years for class II genes by the D_{xy}-based method. Although this estimate does not differ greatly from ours, it is rendered unreliable by rate variation in the different taxonomic groups and by the inefficacy of correcting multiple hits in distantly related species. Our estimate therefore avoids the pitfalls that marred previous estimates.

Estimates of Mutation Rates

According to the neutral theory of molecular evolution, the nucleotide substitution rate is proportional to the neutral mutation rate (33). Since synonymous changes are by and large neutral (5), our estimates may be used to calculate the neutral mutation rate of the MHC genes. If we use the b_{mm} value for the *DRB1* locus and take a generation time of 20 years for modern humans, the neutral mutation rate becomes 2.36×10^{-8} per site per generation, which is comparable to the estimated mutation rate of other genes (Table 3). Assuming $N_e = 10^4$ (36, 37), we can also compute the expected nucleotide diversity (average number of nucleotide differences per site between two DNA sequences). This value becomes 0.09%, which is in good agreement with the observation that the nucleotide diversity of 49 human (non-MHC) loci ranges from 0.03% to 0.11% (38). These agreements strengthen our confidence in our estimate of the synonymous substitution rate at the MHC loci.

We thank anonymous referees for their comments. This work was supported in part by grants from the Ministry of Education, Science and Culture (Japan) and the National Institutes of Health (Bethesda, MD).

1. Klein, J. (1986) *Natural History of the Major Histocompatibility Complex* (Wiley, New York).
2. Nei, M. & Hughes, A. L. (1991) in *Evolution at the Molecular Level*, eds. Selander, R. K., Clark, A. G. & Whittam, T. S. (Sinauer, Sunderland, MA), pp. 222–247.
3. Klein, J. (1980) in *Immunology 80*, eds. Fougereau, M. & Dausset, J. (Academic, London), pp. 239–253.
4. Hughes, A. L. & Nei, M. (1988) *Nature (London)* **335**, 167–170.
5. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, U.K.).
6. Klein, J., Kasahara, M., Gutknecht, J. & Figueroa, F. (1990) *Chem. Immunol.* **49**, 35–50.
7. Britten, R. J. (1986) *Science* **231**, 1393–1398.
8. Satta, Y., Takahata, N., Schönbach, C., Gutknecht, J. & Klein, J. (1991) in *Molecular Evolution of the Major Histocompatibility Complex*, eds. Klein, D. & Klein, J. (Springer, Heidelberg), pp. 51–62.
9. Nei, M. & Li, W.-H. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5269–5273.
10. Takahata, N. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2419–2423.
11. Kingman, J. F. C. (1982) *Stochastic Processes Appl.* **13**, 235–248.
12. Tavaré, S. (1984) *Theor. Pop. Biol.* **26**, 119–164.
13. Zemmour, J. & Parham, P. (1991) *Immunogenetics* **33**, 310–320.
14. Mayer, W. E., Jonker, M., Klein, D., Ivanyi, P., van Seventer, G. & Klein, J. (1988) *EMBO J.* **7**, 2765–2774.
15. Lawlor, D. A., Warren, E., Ward, F. E. & Parham, P. (1990) *Immunol. Rev.* **113**, 147–185.
16. Lawlor, D. A., Warren, E., Taylor, P. & Parham, P. (1991) *J. Exp. Med.* **174**, 1491–1509.
17. Chen, Z. W., McAdam, S. N., Hughes, A. L., Dogon, A. L., Letvin, N. L. & Watkins, D. I. (1992) *J. Immunol.* **148**, 2547–2554.
18. Miller, M. D., Yamamoto, H., Hughes, A. L., Watkins, D. I. & Letvin, N. L. (1991) *J. Immunol.* **147**, 320–329.
19. Marsh, S. G. E. & Bodmer, J. G. (1991) *Immunogenetics* **33**, 321–334.
20. Fan, W., Kasahara, M., Gutknecht, J., Klein, D., Mayer, W. E., Jonker, M. & Klein, J. (1989) *Hum. Immunol.* **26**, 107–121.
21. Kasahara, M., Klein, D., Vincek, V., Sarapata, D. E. & Klein, J. (1992) *Genomics* **4**, 340–349.
22. Zhu, Z., Vincek, V., Figueroa, F., Schönbach, C. & Klein, J. (1991) *Mol. Biol. Evol.* **8**, 563–578.
23. Grahovac, B., Mayer, W. E., Vincek, V., Figueroa, F., O'hUigin, C., Tichy, H. & Klein, J. (1992) *Mol. Biol. Evol.* **9**, 403–416.
24. Satta, Y. (1992) in *The Proceedings of the 17th Taniguchi Symposium*, ed. Takahata, N. (Japan Sci. Soc. Press, Tokyo), pp. 111–131.
25. Martin, R. D. (1990) *Primate Origins and Evolution: A Phylogenetic Reconstruction* (Chapman and Hall, London).
26. Jukes, T. H. & Cantor, C. R. (1969) in *Mammalian Protein Metabolism III*, ed. Munro, H. N. (Academic, New York), pp. 21–132.
27. Kondo, R., Horai, S., Satta, Y. & Takahata, N. (1993) *J. Mol. Evol.*, in press.
28. Takahata, N. & Nei, M. (1985) *Genetics* **110**, 325–344.
29. Hayashida, H. & Miyata, T. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2671–2675.
30. Wu, C.-I. & Li, W.-H. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1741–1745.
31. Easteal, S. (1991) *Mol. Biol. Evol.* **8**, 115–127.
32. Hughes, A. L. & Nei, M. (1990) *Mol. Biol. Evol.* **7**, 491–514.
33. Kimura, M. (1968) *Nature (London)* **217**, 624–626.
34. Klein, J., Satta, Y., O'hUigin, C. & Takahata, N. (1993) *Annu. Rev. Immunol.* **11**, 269–295.
35. Neel, J. V., Satoh, C., Goriki, K., Fujita, M., Takahashi, N., Asakawa, J.-I. & Hazama, R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 389–393.
36. Nei, M. & Grauer, D. (1984) *Evol. Biol.* **17**, 73–118.
37. Takahata, N. (1993) *Mol. Biol. Evol.* **10**, 2–22.
38. Li, W.-H. & Sadler, L. A. (1991) *Genetics* **129**, 513–523.