Allelic Genealogy Under Overdominant and Frequency-Dependent Selection and Polymorphism of Major Histocompatibility Complex Loci

Naoyuki Takahata*,† and Masatoshi Nei*

*Center for Demographic and Population Genetics, The University of Texas Health Science Center, Houston, Texas 77225, and †National Institute of Genetics, Mishima, Shizuoka 411, Japan

ABSTRACT

To explain the long-term persistence of polymorphic alleles (trans-specific polymorphism) at the major histocompatibility complex (MHC) loci in rodents and primates, a computer simulation study was conducted about the coalescence time of different alleles sampled under various forms of selection. At the same time, average heterozygosity, the number of alleles in a sample, and the rate of codon substitution were examined to explain the mechanism of maintenance of polymorphism at the MHC loci. The results obtained are as follows. (1) The coalescence time for neutral alleles is too short to explain the trans-specific polymorphism at the MHC loci. (2) Under overdominant selection, the coalescence time can be tens of millions of years, depending on the parameter values used. The average heterozygosity and the number of alleles observed are also high enough to explain MHC polymorphism. (3) The pathogen adaptation model proposed by Snell is incapable of explaining MHC polymorphism, since the coalescence time for this model is too short and the expected heterozygosity and the expected number of alleles are too small. (4) From the mathematical point of view, the minority advantage model of frequency-dependent selection is capable of explaining a high degree of polymorphism and trans-specific polymorphism. (5) The molecular mimicry hypothesis also gives a sufficiently long coalescence time when the mutation rate is low in the host but very high in the parasite. However, the expected heterozygosity and the expected number of alleles tend to be too small. (6) Consideration of the molecular mechanism of the function of MHC molecules and other biological observations suggest that the most important factor for the maintenance of MHC polymorphism is overdominant selection. However, some experiments are necessary to distinguish between the overdominance and frequency-dependent selection hypotheses.

THE mechanism of maintenance of polymorphism at the major histocompatibility complex (MHC) loci has been debated for more than two decades, yet no consensus has been reached. Polymorphism at the MHC loci is extraordinary in several respects, and any theory of MHC polymorphism must be able to explain the following observations: (1) The level of polymorphism is extremely high, the heterozygosity in humans and mice being about 80–90% (KLEIN 1986). (2) Some pairs of polymorphic alleles appear to have coexisted in the population at least for ten million years, because they are shared by several species of mice and rats (McCONNELL et al. 1988; FIGUEROA, GUNTHER and KLEIN 1988) and humans and chimpanzees (LAWLOR et al. 1988; MAYER et al. 1988). (3) The number of amino acid differences between two alleles in a species is very high, amounting up to 20% (KLEIN and FIGUEROA 1986). (4) In the antigen-recognition site of the MHC genes, the rate of nonsynonymous (amino acid altering) nucleotide substitution is considerably higher than that of synonymous substitution (HUGHES and

NEI 1988, 1989). Using the results of MARUYAMA and NEI's (1981) study of polymorphism and codon substitution (see also NEI and ROYCHOUDHURY 1973), HUGHES and NEI (1988) pointed out that all these observations can be explained by overdominant selection in the presence of a normal rate of mutation. This and other considerations led them to conclude that the high degree of MHC polymorphism is mainly caused by overdominant selection.

However, HUGHES and NEI's conclusion is partly based on conjectures derived from MARUYAMA and NEI's study. MARUYAMA and NEI certainly showed that, under overdominant selection, the level of heterozygosity increases enormously and the rate of codon substitution is enhanced considerably compared with the neutral level, but they did not examine the persistence of polymorphic alleles quantitatively. Furthermore, HUGHES and NEI's argument against an alternative hypothesis of frequency-dependent selection is also based on conjectures. It is therefore necessary to study these problems quantitatively and examine whether all the four observations mentioned above can be explained most easily by the overdomi-
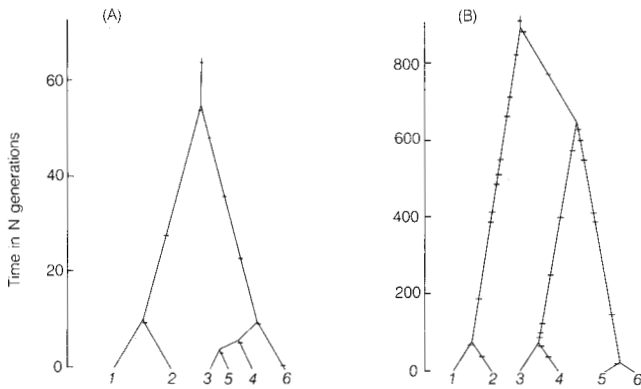
FIGURE 1.—Allelic genealogies for overdominant alleles. These genealogies were obtained by computer simulation. The - symbol denotes the occurrence of a mutation. (A) $N = 100$, $M = 0.04$, $Ns = 33$. (B) $N = 200$, $M = 0.004$, $Ns = 100$.

nance hypothesis. The purpose of this paper is to study these problems.

GENE GENEALOGY AND ALLELIC GENEALOGY

The time of persistence of neutral genes can be studied by using the theory of coalescence time, *i.e.*, the time at which n sampled genes from a locus converge to a single ancestral gene. The mean ($\bar{t}$) of this coalescence time (t) in a randomly mating population is given by

$$\bar{t} = 4N(1 - 1/n) \text{ generations,} \qquad (1)$$

where N is the effective population size (KINGMAN 1982; TAJIMA 1983; TAKAHATA and NEI 1985). Thus, two randomly chosen genes are expected to have diverged $2N$ generations ago. When N is large, this is a long time, and the two genes may have accumulated an appreciable number of mutational differences if the gene consists of a large number of nucleotides. When the population size is small or when the length of the gene is short, however, two randomly chosen genes may have the same nucleotide sequence and thus may have remained as the same allele.

In experimental studies, however, investigators are interested in the time of divergence between different alleles rather than that between randomly chosen genes. That is, what is important here is not the history of randomly chosen genes (gene genealogy) but the history of different alleles (allelic genealogy). Obviously, the history of different alleles starts from the time when a mutant allele was derived from the oldest allele existing in the population. Figure 1A shows one example of allelic genealogy generated by a computer simulation. In this case, allele *1* diverged from allele *3* by a single mutation about $54.8N$ generations ago and later produced allele *2* by another mutation about $9.6N$ generations ago. However, the allelic difference is not always by a single mutation. For example, allele *1* and allele *3* are different by 5 mutations. This occurs either because the intermediate alleles have been lost

from the population or because they are not included in the sample.

At any rate, what is important for biologists is the coalescence time of the alleles sampled rather than that of the randomly sampled genes. Unfortunately, the mathematical formulation of allelic genealogy seems to be more difficult than that of gene genealogy. Particularly, in the presence of selection this problem is quite complicated. We therefore decided to study the problem by using computer simulation.

In the following we consider two different measures of allelic divergence. One is the coalescence time of the alleles sampled (T_c), and the other the average divergence time between different alleles for all pairwise comparisons (T_d). In addition to these quantities, we will also consider heterozygosity (H), the number of alleles present in the sample (n_a), and the rate of codon substitution per locus per generation (α).

OVERDOMINANT SELECTION

Methods of computer simulation

Computer simulation was conducted by considering a diploid population of effective size N under the assumption of random mating. In each generation, mutation was introduced, and $2N$ genes were chosen at random after selection. Mutation was introduced at a rate of v per gene per generation and was assumed to produce new alleles which were always different from the existing ones. In practice, we considered the infinite-site model of mutation without recombination (KIMURA 1971; WATTERSON 1975) and assumed that all mutations occur at different sites (codon or nucleotide sites). Selection and sampling of genes were conducted at the same time. In each generation, we chose two genes at random with replacement from the parental gene pool to form a zygote, and this zygote was subjected to selection with a given probability (p), which depended on the selection scheme. We first generated a uniform random number x, and if $x < p$, the zygote was chosen to be a member of the next generation; otherwise it was discarded. We repeated this process until N zygotes were chosen. In the case of overdominant selection, symmetric overdominance was our main concern, but asymmetric overdominance was also considered.

Under symmetric overdominance, the p value assigned to each homozygote was 0.5, whereas the p value for heterozygotes was $(1 + a)/2$ $(0 \le a \le 1)$. Thus, the relative fitnesses of homozygotes and heterozygotes were $1 - s$ and 1, respectively, where $s = a/(1 + a)$. All heterozygotes had the same fitness.

To reconstruct the genealogical relationships of alleles sampled from the extant population, we recorded the time of occurrence of each mutant allele and its parental allele. Each allele was specified by two vectors. One vector (U) was for keeping the genea-

logical relationships of alleles, and the other (**V**) was for recording the times when the mutations occurred. The first vector was initially set as $\mathbf{U} = (0,0, \ldots)$, and as mutation occurred, each element of the vector assumed some number. Let us illustrate this process considering the six alleles presented in Figure 1A. In this particular simulation the six alleles had the following **U** vectors.

Allele *1*	(1, 1, 6, 0, 0, 0),
Allele *2*	(1, 1, 6, 4, 0, 0),
Allele *3*	(1, 2, 4, 2, 0, 0),
Allele *4*	(1, 2, 4, 2, 2, 0),
Allele *5*	(1, 2, 4, 2, 3, 0),
Allele *6*	(1, 2, 4, 2, 1, 1).

The vector for allele *1* indicates that this allele experienced a series of mutational changes from the common ancestral allele $(0,0,0, \ldots)$ through $(1,0,0, \ldots)$, $(1,1,0, \ldots)$ and $(1,1,6,0, \ldots)$. The number for each element of this vector represents the order at which a particular mutation occurred in a parental allele. For example, $(1,0,0, \ldots)$ is the first mutant that was derived from $(0,0,0, \ldots)$, $(1,1,6,0, \ldots)$ is the sixth mutant derived from $(1,1,0, \ldots)$, and so on. (The first five mutant alleles from $(1,1,0 \ldots)$ were lost from the population before allele $(1,1,6, \ldots)$ was incorporated.) Thus, alleles *1* and *2* have one mutational difference, whereas alleles *1* and *3* have five mutational differences. The reason for the 5 mutational differences between alleles *1* and *3* is as follows. First, the numbers 1 and 2 in the second element of the vectors for alleles *1* and *3* indicate two mutational differences, whereas the numbers 6 and 4 in the third element for alleles *1* and *3* indicate two mutational differences. Similarly, the numbers 0 and 2 in the fourth element for alleles *1* and *3* indicate one mutational difference. Therefore, the sum of all mutational differences is 5 (see Figure 1A). By the same token, alleles *1* and *4* have six mutational differences, and alleles *1* and *6* have seven.

As mentioned earlier, the second vector records the times of occurrence of mutations, each element representing the time in N generations. For example, alleles *1* and *2* above had vectors $\mathbf{V} = (64.4, 54.8, 26.4, 0, \ldots)$ and $\mathbf{V} = (64.4, 54.8, 26.4, 9.6, 0, \ldots)$, respectively. Hence, allele *2* was derived $9.6N$ generations ago from allele *1*. Allele *1* was in turn derived from allele $(1,1,0, \ldots)$ $26.4N$ generations ago, though this allele was lost from the population by genetic drift or mutation (see Figure 1A). Furthermore, it is seen that allele $(1,1,0, \ldots)$ was derived from $(1,0, \ldots)$ $54.8N$ generations ago. On the other hand, allele $(1,2, \ldots)$, which also originated from $(1,0 \ldots)$, had vector $(64.4, 51.2, \ldots)$. The most recent common ancestor for allelic classes $(1,1, \ldots)$ and $(1, 2, \ldots)$ must have existed $54.8N$ generations ago so that the divergence time between alleles $(1, 1, \ldots)$ and $(1, 2, \ldots)$ is placed

at $54.8N$ generations ago. We can thus draw the genealogy of alleles *1*, *2*, and the group of alleles *3* to *6* as given in Figure 1A. Note that the present method for describing allelic genealogy also allows us to record every mutational change in all allelic lines. To examine the steady-state allelic genealogy, most of the observations were made after $1000/(4Nv)$ or 10,000 generations, whichever was larger, since the start of the simulation. This time length appeared to be sufficient for the steady state to be reached. In fact, the original allele had never been found by this time in all replications. This simulation required a large amount of computer time, so that the number of replications used was 20 in most cases. For our purpose, this was sufficient (see Table 1).

In this simulation we used the **U** and **V** vectors for every allele in the population in all generations until the allele was lost from the population. In addition, we had to use a large number of pseudorandom numbers to simulate the stochastic change of allele frequencies. Because of the large computer memory and time required, we were forced to use a relatively small population size. However, stochastic theory of population genetics tells us that the dynamics of genes in finite populations is determined mainly by $M = 4Nv$ and $S = 4Ns$. In the case of overdominant selection, the population dynamics of genes is somewhat complicated, but the steady-state distribution of allelic frequencies is determined by M and S as long as the expected heterozygosity is high as in the present case (YOKOYAMA and NEI 1979). We therefore used $N = 50, 100, 200, s = 0.048, 0.33, 0.5,$ and $v = 0.00001, 0.0001, 0.01,$ though not all the combinations of N, s and v were examined. All simulations were started with a monomorphic population fixed for a particular (original) allele.

As mentioned earlier, we examined five quantities, *i.e.*, T_c, T_d, H, n_a and α. The rate of codon substitution (α) was measured by recording the number of mutations per gene that accumulated during a given duration time, T. This number divided by T was an estimate of α. All the above quantities were first studied for neutral mutations, where analytical solutions are available for the expectations of three of the five quantities. They are:

$$E(H) = M/(1 + M),$$

$$E(n_a) = \sum_{j=1}^{n-1} M/(M + j - 1) \tag{2}$$

$$E(\alpha) = v,$$

where n is sample size (see KIMURA and CROW 1964; KIMURA 1968; EWENS 1972).

In the case of overdominant selection, there are also fairly extensive analytical studies on the allele frequency distribution (or spectrum) from which one

can compute $E(H)$ and $E(n_a)$ (WRIGHT 1939; FISHER 1958; KIMURA and CROW 1964; EWENS 1964; WRIGHT 1966; WATTERSON 1977; LI 1978; YOKOYAMA and NEI 1979). When H is high, the distribution is approximately given by

$$\Phi(x) = Ce^{Sx}(1 - x)^{B-1}x^{-1}, \qquad (3)$$

in which $S = 4Ns/(1 - sJ)$, $B = M + S(1 - J)/(1 - sJ)$, $J[=1 - E(H)]$ and C is a constant such that $\int_0^1 \Phi(x)x\,dx = 1$ (YOKOYAMA and NEI 1979). However, since J itself is a function of S and B (YOKOYAMA and NEI 1979), C and J cannot be determined directly from (3). We therefore determined C and J by using an iteration method starting with arbitrary initial values of C and J, noting $J = \int_0^1 \Phi(x)x^2dx$. We terminated this iterative computation when the difference in J between two iterations became less than 10^{-8}. Once the final values of C and J were obtained, $E(H)$ was given by $1 - J$, and $E(n_a) = \int_{1/n}^1 \Phi(x)dx$, where n is the number of genes sampled. In the present study, we sampled all genes in the population, so that n was $2N$. We used the theoretical values of n_a and H thus obtained to check the accuracy of our computer simulations.

It should be noted that when there are k alleles at a locus and symmetric overdominance operates, the deterministic change of the frequency (x_i) of the ith allele per generation can be written as

$$\Delta x_i = - \frac{sx_i(x_i - J)}{1 - sJ}, \qquad (4)$$

where $J = \sum_{j=1}^k x_j^2$ and the effect of mutation is neglected (WRIGHT 1969). This indicates that whenever x_i is smaller than homozygosity J, it increases and eventually reaches J.

Results of simulations

Neutral mutations: Before performing simulations for the selection models, we checked our computer program by comparing results with the theoretical expectations for neutral alleles and at the same time generated the allelic genealogy for neutral alleles (not gene genealogy). The means of H, n_a, T_c, T_d, and α for all replications are presented in the rows of $Ns = 0$ in Table 1. In this table, $\bar{T}_c$ and $\bar{T}_d$ represent the conditional mean coalescence time and the average pairwise divergence time, respectively, excluding the case of population monomorphism, and are expressed in units of N generations, whereas $\bar{\alpha}$ is in units of v so that $\bar{\alpha} = 1$ for neutral mutations. Table 1 indicates that $\bar{H}$, $\bar{n}_a$, and $\bar{\alpha}$ for neutral alleles are in fairly good agreement with the theoretical predictions from (2), though the number of replications used is rather small.

Table 1 shows that when M is small the mean coalescence time of different alleles $(\bar{T}_c)$ is much

smaller than the coalescence time $(\bar{t})$ of randomly sampled genes. Note that in the present case $n = 2N$, and thus $\bar{t} \cong 4N$ generations from Equation 1. (We did not evaluate T_c for neutral alleles for the case of $M = 0.004$ because the population was monomorphic most of the time and it required a large amount of computer time to obtain a reliable estimate.) The small value of T_c when M is small is of course due to the fact that many new mutant alleles are quickly eliminated from the population and only one out of $2N$ mutants is fixed in the population.

As M increases, however, T_c increases, and thus we obtain $\bar{T}_c = 3.1N$ generations when $M = 4$. This is still smaller than $\bar{t}$ but indicates that the coalescence time of neutral alleles can be quite long when M is large. Nevertheless, $\bar{T}_c$ is always smaller than $\bar{t}$. The reason for this is that $\bar{T}_c$ refers to the time at which the second oldest allele was derived by mutation from the oldest allele existing in a sample, whereas $\bar{t}$ refers to the oldest event of gene splitting by reproduction whether the resulting two genes are the same allele or not.

Figure 2 shows one of the allelic genealogies obtained in our computer simulation for the case of $M = 4$. Since the mutation rate is very high in this case, many alleles are maintained in the population without selection. The number of codon differences between different alleles is also substantial. However, the $\bar{T}_c$ value is about $3.1N$ generations and thus is still smaller than the $\bar{t}$ value.

As mentioned above, $\bar{T}_c$ is a conditional coalescence time, given that the population is polymorphic. By contrast, $\bar{t}$ is the absolute coalescence time. Therefore, if we consider the absolute coalescence time including the events of monomorphism, it becomes much lower than the value in Table 1 when M is low since the probability of monomorphism (P_M) is given by q^M (KIMURA 1971), where q is $1/n$. For example, when $M = 0.04$ and $n = 200$, we have $P_M = 0.81$. Therefore, the absolute coalescence time will be $0.29N$ generations, since $(1 - P_M) \times \bar{T}_c = 0.19 \times 1.5 = 0.29$.

Another important aspect of T_c is its distribution. Although $\bar{T}_c$ is quite large when M is large, T_c widely varies from replication to replication. In the case of $M = 4$, T_c ranged from a small value to $8.8N$ generations (Figure 3A), though only 100 replications were examined. Comparison of this distribution with that of the coalescence time of random genes (t) (Figure 3B) indicates that both distributions are somewhat similar to each other, though the former has a smaller mean and a smaller variance.

$\bar{T}_d$ in Table 1 is the average divergence time between alleles for all pairwise comparisons, excluding the case of monomorphic populations. Interestingly, $\bar{T}_d$ is nearly independent of M. Of course, if we compute $\bar{T}_d$ including monomorphic loci, it will be an increasing function of M.

TABLE 1

Stimulation results and theoretical values (in parentheses) of the quantities studied for the cases of neutral alleles and symmetric overdominance

M	Ns	$\bar{H}$	$\bar{n}_a$	$\bar{T}_c$	$\bar{T}_d$	$\bar{\alpha}$
0.004	0	— (0.004)	— (1.0)	—	—	— (1.0)
	33	0.698 ± 0.016 (0.708)	3.6 ± 0.2 (3.7)	387.0 ± 68.7	266.0 ± 43.4	10.8 ± 1.0
	100*	0.813 ± 0.005 (0.820)	5.9 ± 0.2 (6.0)	468.0 ± 55.3	311.0 ± 36.4	10.2 ± 0.5
0.04	0	0.056 ± 0.033 (0.038)	1.3 ± 0.1 (1.2)	1.5 ± 0.5	1.4 ± 0.5	0.8 ± 0.2 (1.0)
	4.8	0.395 ± 0.045 (0.469)	2.2 ± 0.2 (2.3)	10.2 ± 2.2	8.9 ± 2.0	3.7 ± 0.2
	33	0.755 ± 0.009 (0.756)	4.8 ± 0.3 (4.7)	50.9 ± 6.4	32.3 ± 4.1	6.2 ± 0.3
	100*	0.843 ± 0.003 (0.848)	7.2 ± 0.2 (7.5)	80.6 ± 7.7	51.6 ± 4.7	7.7 ± 0.4
0.4	0	0.216 ± 0.052 (0.286)	2.5 ± 0.3 (3.1)	1.4 ± 0.4	1.1 ± 0.3	0.9 ± 0.1 (1.0)
	4.8	0.569 ± 0.045 (0.635)	4.4 ± 0.3 (4.4)	10.0 ± 2.2	5.9 ± 1.2	2.1 ± 0.1
	33	0.814 ± 0.005 (0.818)	7.3 ± 0.2 (7.5)	13.2 ± 1.4	7.9 ± 0.8	3.3 ± 0.2
	100*	0.877 ± 0.002 (0.883)	10.9 ± 0.2 (11.2)	19.3 ± 2.5	10.8 ± 1.3	4.3 ± 0.2
4.0	0	0.801 ± 0.014 (0.800)	15.0 ± 0.7 (16.2)	3.1 ± 0.5	1.6 ± 0.3	1.0 ± 0.1 (1.0)
	100*	0.924 ± 0.002 (0.929)	25.0 ± 0.6 (24.2)	4.9 ± 0.02	2.8 ± 0.1	2.0 ± 0.1

$\bar{H}$, mean heterozygosity; $\bar{n}_a$, mean number of alleles in the sample; $\bar{T}_c$, mean coalescence time for different alleles sampled, excluding monomorphic cases; $\bar{T}_d$, mean of the average divergence time between alleles for all pairwise comparisons, excluding the monomorphic cases. T_c and T_d are measured in N generations. $\bar{\alpha}$, mean rate of condon substitution. The numbers after the ± sign are the standard errors of the mean values. $M = 4Nv$. $N = 100$ except for the case with * where $N = 200$. Data are from 20 replications.
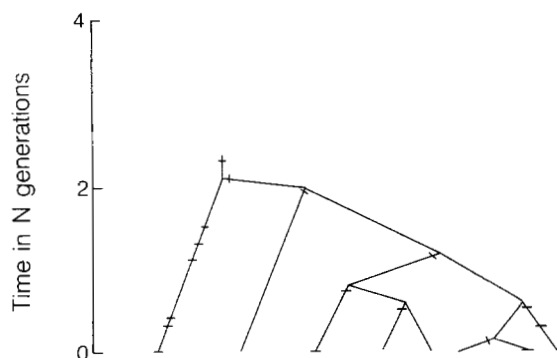

FIGURE 2.—Allelic genealogy for neutral alleles. The - symbol denotes the occurrence of a mutation. $N = 100$, $M = 4$, $Ns = 0$.

Symmetric overdominance: In the presence of overdominant selection the average heterozygosity and the number of alleles per locus increases substantially, as was shown by the previous authors (e.g., KIMURA and CROW 1964; EWENS 1964; MARUYAMA and NEI 1981). As expected, $\bar{T}_c$ also increases dramatically as Ns increases. For example, when $M = 0.04$, the $\bar{T}_c$ value for $Ns = 100$ is 54 times higher than that for $Ns = 0$. Note that when $Ns \geq 33$, the probability of the population being monomorphic is virtually 0. Therefore, $\bar{T}_c$ is essentially the absolute coalescence time. In general, for a given value of Ns, $\bar{T}_c$ is higher when M is small than when M is large. This is because in the presence of a high mutation rate the rate of allelic turnover increases. When the mutation rate is low, a particular set of alleles may be maintained in the population for a long time even if genetic drift operates in each generation.

In general, however, the number of allelic lineages maintained for a long evolutionary time is relatively small, as will be seen from the examples in Figure 1, A and B. The other allelic lineages are relatively short-lived. In the example of Figure 1B, the coalescence time of the six alleles sampled is about $900N$ generations. Even the second and third oldest lineages diverged more than $600N$ generations ago. By contrast, the remaining three allelic divergences occurred within the last $35N$ generations.

Figure 1B shows that in the presence of overdominant selection polymorphic alleles (allelic lineages) may persist in the population for an extremely long time. The average heterozygosity of protein loci in rodents suggests that the long-time effective population size is probably of the order of 10^5 (NEI and GRAUR 1984). If there are two generations in a year in rodent populations in nature, $600N$ generations would correspond to about 30 million years. One can therefore easily explain the shared polymorphism at the MHC loci between mice and rats mentioned earlier. Of course, T_c varies greatly from replication to replication, as in the case of neutral alleles. However, the mean T_c is sufficiently large to explain the persistence of polymorphic alleles observed in rodents and primates if M is relatively small and Ns is large (Table 1).

One might notice that when T_c is large (e.g., the case of $M = 0.004$ and $Ns = 100$) $\bar{n}_a$ tends to be small in Table 1 and that the number of alleles observed at an MHC locus is generally about 10 or larger in human and mouse populations. However, $\bar{n}_a$ and $\bar{T}_c$ can both be increased if we increase N and Ns in our computer simulation. For example, in a separate computer simulation with $N = 500$, $M = 0.004$, and $Ns = 250$ (one replication), we obtained $n_a = 10$ and $T_c =$
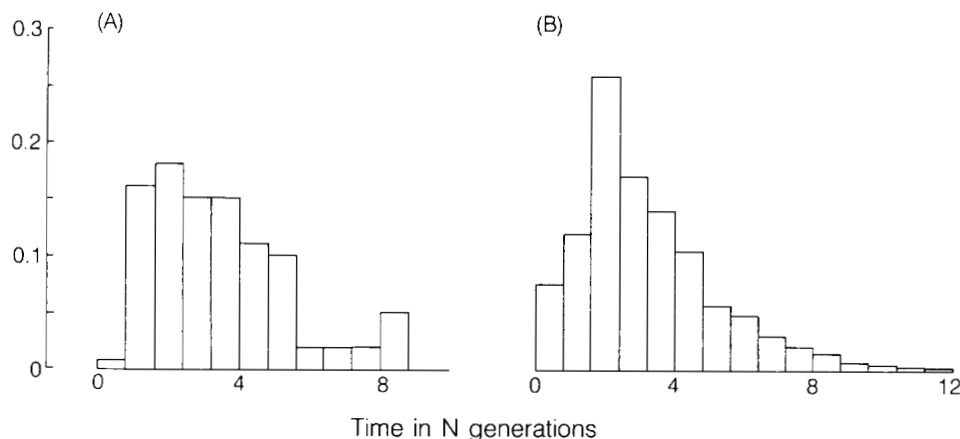
FIGURE 3.—Distributions of the coalescence time for different alleles (T_c) and the coalescence time for sampled genes (t) for the case of neutral genes. (A) is for $\bar{T}_c$ and (B) is for t. $N = 100$, $M = 4$. All genes in the population were sampled.

Time in N generations

$1052N$. Therefore, there is no problem in explaining the observed levels of n_a and T_c by overdominant selection. In practice, N is probably of the order of 10^5, as mentioned above. Therefore, even if s is as small as 0.01, Ns can be 1000.

The average divergence time between alleles ($\bar{T}_d$) is always a little smaller than $\bar{T}_c$, but the two quantities are highly correlated. $\bar{T}_d$ also varies greatly from replication to replication, but the variance of this distribution is smaller than that of T_c (data not shown).

Our results concerning the rate of codon substitution (α) are similar to those of MARUYAMA and NEI (1981). (Overdominant selection enhances the rate of codon substitution, because a new mutant allele is almost always in heterozygous condition and thus enjoys selective advantage over more common alleles which would exist in both heterozygous and homozygous condition.) When M is small, $\bar{\alpha}$ is much greater than that for neutral mutations, but as M increases, the substitution rate for a given Ns value declines. Part of the reason is that as M increases heterozygosity (H) increases and that when H is very high, most individuals are nearly equally fit so that the rate of gene substitution becomes similar to that of neutral mutations.

Asymmetric overdominance: MARUYAMA and NEI (1981) already showed that when the fitness of a homozygote varies from genotype (A_iA_i) to genotype (A_jA_j) but the mean s value remains the same, the average heterozygosity is a little smaller than that for the case of symmetric overdominance (the same selection coefficient for all homozygotes) but that the rate of codon substitution remains nearly the same. We conducted a small-scale simulation for this case to examine how $\bar{T}_c$ and $\bar{T}_d$ are affected by asymmetry of overdominance. Our scheme of selection was somewhat different from MARUYAMA and NEI's. We assumed that all heterozygotes have fitness 1, whereas the fitness of a given homozygote (A_iA_i) is given by $1 - s_i$. We determined s_i by choosing a random number that was uniformly distributed between a and b ($0 \leq$

TABLE 2

Simulation results for asymmetric overdominant selection

s^a	$\bar{H}$	$\tilde{n}_a$	$\bar{T}_c$	$\bar{T}_d$	$\tilde{\alpha}$
[0.5, 0.5]	0.790	5.4	61.6	39.2	6.5
(a)	0.734	4	17.7	17.2	3.9
(b)	0.844	8	123.0	75.2	10.6
[0.2, 0.8]	0.680	3.7	88.7	52.4	2.6
(a)	0.531	3	32.0	24.0	1.1
(b)	0.750	5	123.0	90.4	4.3
[0.1, 0.9]	0.489	2.5	68.8	58.8	2.4
(a)	0.238	2	16.6	16.6	1.2
(b)	0.716	5	117.0	117.0	4.7

The rows (a) and (b) represent the minimum and maximum values observed in a simulation with 10 replications. $M \equiv 4Nv = 0.04$; $N = 100$.

[a] Selective disadvantage of homozygotes of a particular allele was determined by choosing a uniform random number which is distributed between the range specified in this column.

a, $b \leq 1$), so that the mean s value was $\bar{s} = (a + b)/2$. In the present paper we considered only the case of $\bar{s} = 0.5$, and the number of replications examined was 10 for each case.

The results of our simulation are presented in Table 2. It is clear that as the difference $b - a$ increases, both $\bar{H}$ and $\tilde{n}_a$ tend to decrease in conformity with MARUYAMA and NEI's earlier results but not to a great extent. By contrast, $\bar{T}_c$ and $\bar{T}_d$ tend to increase with increasing value of $b - a$. This increase is of course expected to occur, because with asymmetric overdominance a particular set of alleles may maintain a highly balanced polymorphism, whereas others are eliminated quickly from the population (ROBERTSON 1962).

FREQUENCY-DEPENDENT SELECTION

Pathogen adaptation model: Before the function of MHC molecules was clarified at the molecular level, several authors (e.g., SNELL 1968; BODMER 1972) speculated that MHC alleles generate heterozygote disadvantage in association with infectious diseases and that in order to maintain a high level of polymorphism some kind of frequency-dependent selection is neces-

sary. One of the frequency-dependent selection models suggested at that time was the pathogen adaptation model (SNELL 1968; BODMER 1972). This model is based on the assumption that host individuals carrying new antigens, which have arisen recently by mutation, will be at an advantage because viruses will not yet have had the time to adapt to infecting the cells carrying a new antigen. This will therefore generate a form of frequency-dependent selection (or time-dependent selection), in which a mutant MHC allele initially has a selective advantage compared with an old allele but the advantage gradually declines with time.

We simulated this form of selection assuming that the selective advantage (s) over an old allele declines exponentially in each generation (t) with a rate of r, i.e., $s = s_0 e^{-rt}$, where $s_0 = 1$ was assumed. This form of selection was assumed to occur for each allele independently. Thus, the fitness of a zygote was the product of the fitnesses of the two genes involved. The other parts of the simulation method were identical with those for overdominant selection. Strictly speaking, this model is not frequency-dependent selection, because s does not depend on the frequency of the mutant allele. However, the selective advantage of a mutant allele gradually declines with time, as many immunologists suggested.

In this pathogen adaptation model, a new mutant allele at an MHC locus initially has a selective advantage over old alleles, so that the rate of incorporation of mutant alleles into the population increases. This suggests that in the presence of pathogen adaptation average heterozygosity, the number of alleles, and the rate of codon substitution will increase compared with those for neutral alleles. Table 3 shows that $\bar{H}$ and $\bar{n}_a$ indeed tend to be larger for this model than for neutral alleles (compare the results with the neutral expectations with the same M value in Table 1), but the extent of the increase is small. The $\bar{\alpha}$ value for the pathogen adaptation model is also higher than that for neutral alleles, but the value is again not as high as in the case of overdominant selection. The effects of pathogen adaptation on $\bar{T}_c$ and $\bar{T}_d$ are somewhat different from those on $\bar{H}$ and $\bar{n}_a$. The $\bar{T}_c$ and $\bar{T}_d$ values are slightly higher than those for neutral alleles when the rate of change (r) in selection coefficient is large but become smaller when r is small. This reflects the fact that when r is small, the mutant allele retains the selective advantage for a long time and thus the rate of turnover of alleles increases. In any event, the pathogen adaptation model is not an appropriate model to explain MHC polymorphism, as predicted by HUGHES and NEI (1988).

Minority advantage: A popular model of frequency-dependent selection is minority advantage, in which a genotype has a selective advantage over others

whenever it becomes rare in the population (WRIGHT and DOBZHANSKY 1946; CLARKE 1976). This model has been developed primarily for a pair of alleles at a locus but can be extended to the case of multiple alleles. One way to extend the model is to assume that the fitness of genotype A_iA_j is given by $(1 - sx_i)(1 - sx_j)$, where x_i and x_j are the frequencies of alleles A_i and A_j. In a randomly mating population, the frequencies of genotypes A_iA_i and A_iA_j are given by x_i^2 and $2x_ix_j$, respectively. Therefore, in the absence of mutation and genetic drift, the frequency (x_i') of allele A_i in the next generation is given by

$$x_i' = (x_i - sx_i^2 - sx_iJ)/\bar{W}, \qquad (5)$$

where $J = \Sigma_i x_i^2$, and $\bar{W}$ is the mean fitness given by $\bar{W} = 1 - sJ$. Therefore, the expected change of allele frequency per generation is

$$\Delta x = x_i' - x_i = -\frac{sx_i(x_i - J)}{1 - sJ}. \qquad (6)$$

Equation 6 is identical with the gene frequency change [Equation 4] under overdominant selection. This means that the model of minority advantage considered here produces the same effect on the population dynamics of alleles as that of overdominant selection and that it is capable of explaining the high degree of polymorphism and the long persistence of polymorphic alleles at the MHC loci. Therefore, mathematical study alone cannot distinguish between this model and the overdominance model. To distinguish between these models, we must consider biological aspects (see DISCUSSION).

Molecular mimicry: In the above two models, we considered polymorphism only for the host population. However, if we assume that pathogen adaptation to a new host allele is accomplished by mutations that occur at a parasite locus, we must consider the allelic variation on both host and parasite sides. This is particularly so in DAMIAN's (1964, 1987) model of molecular mimicry, in which a parasite gains resistance to the host immune system by generating a parasitic peptide that mimicks the host MHC molecule. DAMIAN (1964) considered a polysaccharide as an antigenic substance, but in the context of MHC polymorphism we must consider the antigen recognition site (ARS) of the MHC molecule. One way to study this problem is to use the model given in Table 4. In this model we consider a series of MHC antigenic states (A_1, A_2, ..., A_K) that are generated by changes in amino acids in the ARS. For simplicity, we assume that mutation occurs stepwise from A_i to $A_i + 1$ or $A_i + 1$ to A_i with a rate of v per generation in the same manner as that of OHTA and KIMURA's (1973) stepwise mutation model. This model seems to be reasonable if the structure of the ARS changes slightly by each amino acid substitution and the original structure can be restored by back mutation.

M	$2Nr^a$	$\bar{H}$	$\bar{n}_a$	$\bar{T}_c$	$\bar{T}_d$	$\bar{\alpha}$
0.04	10	0.036 ± 0.025	1.2 ± 0.1	1.9 ± 1.2	1.9 ± 1.2	1.4 ± 0.1
	1	0.070 ± 0.030	1.4 ± 0.1	1.4 ± 0.8	1.4 ± 0.8	1.7 ± 0.1
	0.1	0.035 ± 0.021	1.2 ± 0.1	0.4 ± 0.2	0.4 ± 0.2	1.5 ± 0.1
0.4	10	0.340 ± 0.048	3.6 ± 0.3	4.2 ± 1.4	2.6 ± 0.7	1.4 ± 0.2
	1	0.360 ± 0.051	3.1 ± 0.4	2.2 ± 0.6	1.6 ± 0.4	1.7 ± 0.2

[a] $s = e^{-rt}$: Selective advantage of a new mutant allele that appeared t generations ago. $M = 4Nv$, where N and v are the effective population size and the mutation rate, respectively. $N = 50$. Data from 20 replications.

TABLE 4

Fitnesses of host and parasite genotypes (haploid model) in the molecular mimicry model

Host Parasite	A_1 x_1	A_2 x_2	A_3 x_3	A_4 x_4	—	Mean (P)
$B_1\ y_1\ (H)$	$1 - sy_1$	1	1	1	—	
(P)	$1 + tx_1$	1	1	1	—	$1 + tx_1^2$
$B_2\ y_2\ (H)$	1	$1 - sy_2$	1	1	—	
(P)	1	$1 + tx_2$	1	1	—	$1 + tx_2^2$
$B_3\ y_3\ (H)$	1	1	$1 - sy_3$	1	—	
(P)	1	1	$1 + tx_3$	1	—	$1 + tx_3^2$
$B_4\ y_4\ (H)$	1	1	1	$1 - sy_4$	—	
(P)	1	1	1	$1 + tx_4$	—	$1 + tx_4^2$
—	—	—	—	—	—	—
Mean (H)	$1 - sy_1^2$	$1 - sy_2^2$	$1 - sy_3^2$	$1 - sy_4^2$	—	—

x_i and y_i are the frequencies of the ith host and ith parasite alleles, respectively.

On the parasite side, we also consider a series of alleles $(B_1, B_2,..., B_K)$, whose products are responsible for molecular mimicry. We assume that molecular mimicry occurs when the product of a particular allele mimicks the ARS of the MHC molecule of a host allele and that there is one to one correspondence between the host and parasite alleles when mimicry occurs (Table 4). The parasite allele participating in the mimicry has a selective advantage over other parasite alleles, whereas the host allele (antigenic state) has a selective disadvantage in comparison to other host alleles. We assume that the selective advantage in the parasite and the selective disadvantage in the host are frequency-dependent and are given by tx_i and $- sy_i$ respectively for the ith pair of alleles, where t and s are constant $(0 \leq t, s \leq 1)$, and x_i and y_i are the relative frequencies of the ith host allele and the ith parasite allele, respectively. We also assume that mutation occurs stepwise between B_i and $B_i + 1$ with a rate of μ per generation as in the case of host alleles (antigenic states) and that the number of allelic states is the same (K) as that for the host. In practice, the generation time for the parasite is generally much shorter than that for the host. So, we measure both selection coefficient and mutation rate per host generation.

Note that the above model is a haploid model and

is similar to SEGER's (1988) for the polymorphism generated by host-pathogen interaction. SEGER, however, studied the dynamics of allelic frequency changes in infinite host and parasite populations, neglecting the mutation in the host. In this paper we are interested in $\bar{H}$, $\bar{n}_a$, $\bar{T}_c$, $\bar{T}_d$, and $\bar{\alpha}$ in finite populations. However, it is instructive to know the deterministic changes of allele frequencies in the host and parasite populations. In the absence of mutation, the frequency changes of the ith host and the ith parasite alleles (Δx_i and Δy_i, respectively) are given by

$$\Delta x_i = \frac{sx_i(\Sigma\ jx_jy_j^2 - y_i^2)}{1 - s\ \Sigma\ jx_jy_j^2}, \tag{7}$$

$$\Delta y_i = \frac{ty_i(x_i^2 - \Sigma\ jx_j^2y_j)}{1 + t\ \Sigma\ jy_j^2}. \tag{8}$$

An analysis of local stability similar to that conducted by SEGER (1988) shows that the above genetic system is unstable. However, SEGER's study suggests that in the presence of mutation the equilibrium allele frequencies may become locally stable depending on the combination of parameter values (t, s, K, u and v), though the allele frequency changes in the host population are very chaotic for $K \geq 3$. In finite populations, such a local stability is disturbed by genetic drift, but the existence of the local stability suggests that the extent of polymorphism is enhanced by the above genetic system.

It should be noted that the stepwise mutation model representing the antigenic state of an allele was used only for selection (molecular mimicry), and the history of alleles was recorded by the infinite-site model, as mentioned earlier. The antigenic state of a host allele was represented by vector $\mathbf{H}$, which, for example, had a form of $(0, 1, 0, 0, ...)$ for allele A_2. That is, when the antigenic state of an allele was i, the ith element of this vector had value 1 and all other elements had value 0. When a mutation occurred from A_2 to A_3, $\mathbf{H}$ changed from $(0, 1, 0, 0, ...)$ to $(0, 0, 1, 0, 0, ...)$. In this way, the antigenic states of all alleles were recorded. The number of antigenic states considered was either 20 or 50, and a circular state model was used.

Vector $\mathbf{H}$ is clearly insufficient for recording the

TABLE 5

Simulation results for frequency-dependent selection—molecular mimicry model

v	u	K	$\bar{H}$	$\bar{n}_a$	$\bar{T}_c$	$\bar{T}_d$	$\bar{\alpha}$
5×10^{-5}	5×10^{-5}	20	0.011 ± 0.007	1.2 ± 0.1	0.02 ± 0.04	0.02 ± 0.04	115.0 ± 0.5
	5×10^{-2}	20	0.613 ± 0.014	3.0 ± 0.1	92.3 ± 11.8	67.1 ± 8.6	2.8 ± 0.1
	5×10^{-2}	50	0.625 ± 0.096	3.3 ± 0.3	99.8 ± 16.5	71.7 ± 11.2	3.1 ± 0.2
5×10^{-4}	5×10^{-3}	20	0.586 ± 0.024	4.6 ± 0.2	4.0 ± 0.6	2.7 ± 0.4	7.4 ± 0.3
	5×10^{-2}	20	0.670 ± 0.015	5.1 ± 0.2	20.5 ± 2.7	12.1 ± 1.4	2.4 ± 0.2
	5×10^{-2}	50	0.715 ± 0.011	5.5 ± 0.3	26.0 ± 3.2	15.2 ± 2.0	2.3 ± 0.2
5×10^{-3}	5×10^{-2}	20	0.839 ± 0.011	18.5 ± 0.8	12.0 ± 1.4	5.2 ± 0.6	1.4 ± 0.0
	5×10^{-2}	50	0.816 ± 0.013	18.2 ± 0.7	12.1 ± 1.9	6.4 ± 1.0	1.4 ± 0.0

v, mutation rate for the host locus; u, mutation rate for the parasite locus; K, number of antigenic states (circular model); $N = 200$; $Ns = Nt = 100$. Data from 20 replications.

history of each allele, because different alleles with the same antigenic state may have different histories. Since we are interested in the persistence of allelic lineages, we must record the history of each allele. Distinction between different alleles within the same antigenic state is also necessary for computing $\bar{H}$, $\bar{n}_a$, and $\bar{\alpha}$. Recording the history of alleles was conducted by using vectors **U** and **V** previously mentioned. Therefore, the property of a host allele was described by three vectors, **U**, **V** and **H**.

To facilitate computer simulation, we assumed that the population dynamics of parasite alleles are deterministic. We represented the frequency of the ith allele by the ith element of vector **P**, the sum of all elements in the vector being 1. We did not use vectors **U** and **V** for the parasite population because we were not interested in the history of parasite alleles.

The diploid-equivalent host population size was assumed to be 200 (400 genes), and $s = t = 0.5$ was used. Since the frequency of antigenic state i in the host population could easily be computed from **H** vectors and the allele frequencies in the parasite population were known, selection and sampling of host individuals for the next generation were conducted as described before by using the selection scheme in Table 4. The parasite population in the next generation was determined deterministically as mentioned above by considering both mutation and selection.

Table 5 shows the results of computer simulation for the molecular mimicry model. As expected from SEGER's (1988) deterministic analysis of allele frequency changes, the extent of polymorphism ($\bar{H}$) is high when u > v and u is high. Particularly when $v = 5 \times 10^{-3}$ and $u = 5 \times 10^{-2}$, both H and n_a are very high. In this case, however, $\bar{T}_c$ and $\bar{\alpha}$ are too low to explain the general pattern of MHC polymorphism. When $v = 5 \times 10^{-4}$ and $u = 5 \times 10^{-2}$, $\bar{\alpha}$ is sufficiently high to explain the accelerated nonsynonymous nucleotide substitution in the ARS. However, $\bar{T}_c$ and $\bar{T}_d$ are still too small to explain the long-persistence of polymorphic allelic lineages at MHC loci. When $v = 5 \times 10^{-5}$ and $u = 5 \times 10^{-2}$, $\bar{T}_c$ and $\bar{\alpha}$ are both quite high, but $\bar{n}_a$ and $\bar{H}$ are now too small.

These results indicate that the molecular mimicry model is less satisfactory than the overdominance model for explaining MHC polymorphism. However, it might be possible to have more appropriate values of $\bar{H}$, $\bar{n}_a$, $\bar{T}_c$, $\bar{T}_d$ and $\bar{\alpha}$ by exploring different sets of parameter values. For example, $\bar{H}$ and $\bar{n}_a$ can probably be increased by increasing N, Nv, and Ns, though it is not easy to simulate this case. Thus, from the mathematical point of view alone, the molecular mimicry model cannot be ruled out. The problem is therefore whether this model is realistic or not from the biological point of view. This problem will be discussed in the next section.

DISCUSSION

We have shown that overdominant selection can maintain polymorphic allelic lineages for a long time and thus it is a sufficient explanation for the *trans*-specific polymorphism (KLEIN 1987) observed in rodents and primates. When polymorphic alleles are maintained in a population for a long time, allelic divergence naturally occurs by accumulation of mutations. Therefore, the large number of amino acid (nucleotide) differences observed between different alleles in mice and humans can also be explained by overdominant selection. Furthermore, overdominant selection is capable of explaining the high level of heterozygosity and the rate of nonsynonymous nucleotide substitution higher than that of synonymous substitution at the ARS of MHC genes (MARUYAMA and NEI 1981). Therefore, the overdominance hypothesis is a legitimate explanation, as emphasized by HUGHES and NEI (1988, 1989).

Previously, KLEIN (1987) suggested that the trans-specific polymorphism observed in rodents and primates might be explained by neutral mutations. The present study clearly indicates that this explanation is not valid. The mean coalescence time for neutral alleles is too small to explain the trans-specific polymorphism in mice and primates, since it is always smaller than $4N$ generations. The long-term effective

population size in rodents is probably of the order of 10^5, as mentioned earlier. Therefore, the expected allelic coalescence time is less than about 4×10^5 generations or 2×10^5 years. This is much smaller than the observed value (at least 10 million years). In humans and apes, the long-term N and the long-term generation time seem to be about 10^4 and 15 years, respectively (NEI and GRAUR 1984). Therefore, the expected coalescence time for neutral alleles is less than about 600,000 years. This is again much smaller than the time of divergence (about 7 million years ago; see SIBLEY and AHLQUIST 1984) between humans and chimpanzees, which share the same pairs of MHC allelic lineages.

The possibility of overdominant selection at MHC loci was first suggested by DOHERTY and ZINKERNA-GEL (1975). They speculated that since different MHC molecules recognize different foreign antigens, a heterozygote having two different MHC molecules should be more resistant to infectious diseases than a homozygote having one type of MHC molecule. Using X-ray crystallography, BJORKMAN et al. (1987) showed that the antigen recognition site (ARS) of MHC class I molecules forms a groove composed of 57 amino acid residues. This groove is believed to hold a foreign processed peptide, which is generally 10 to 30 amino acids long (e.g., REDDEHASE, ROTHBARD and KOSZIN-NOWSKI 1989; BRACIALE et al. 1989). It is known that MHC polymorphism is mainly caused by variation in the amino acid sequence of the ARS and that different allelic products recognize different foreign peptides. (Strictly speaking, an allelic product recognizes a group of foreign peptides sharing the same structural motif; SETTE et al. 1988). Examining the rates of synonymous and nonsynonymous nucleotide substitution, HUGHES and NEI (1988) showed that positive Darwinian selection occurs at the ARS. Therefore, there are ample biological data to support the over-dominance hypothesis.

We have seen that the model of minority advantage considered here produces essentially the same pattern of genetic polymorphism as that for overdominant selection. Therefore, from the mathematical property alone, we cannot distinguish between the two models. However, the model of minority advantage has some unreasonable features. As mentioned earlier, it is possible that a newly arisen mutant allele has a selective advantage over old ones because there are no pathogens adapted to it. However, if an allele becomes old and again rare in the population, why does the allele regain a selective advantage, as in the model of minority advantage? It seems to us that as the frequency of the mutant allele increases, its selective advantage disappears and its final fate is to disappear from the population by genetic drift or selection (HOWARD 1987). Nevertheless, this hypothesis cannot be ruled

out at the present time, because there are no experimental data against it.

NEI and HUGHES (1990) recently suggested an experiment by which the two hypotheses may be distinguished. It is to examine the fitness of a homozygote for a rare allele at a locus. If minority advantage really occurs, this genotype should have a higher fitness than a heterozygote for a pair of common alleles. By contrast, if the overdominance hypothesis is correct, the reverse would be true. It is hoped that this type of experiment will be conducted in the near future.

The biological basis of molecular mimicry at MHC loci is quite vague and highly speculative. According to this hypothesis, a parasite antigen may mimick a host MHC antigen so that the parasite carrying the antigen escapes the attack by the immune system of the host. This idea came from the finding that some microorganisms carry antigenic substances that react to the antisera of higher organisms. A good example is the presence of the antigenic substances (polysaccharides) of human ABO blood groups in bacteria (DAMIAN 1964). In the case of MHC, however, there is no evidence that molecular mimicry has occurred (DAMIAN 1987). Some authors (e.g., SHER, HALL and VADAS 1978; SIMPSON et al. 1983) reported that parasites may acquire host MHC antigens, but this is not the same phenomenon as molecular mimicry. Furthermore, it is not clear how often the acquisition of host MHC antigens occurs.

If molecular mimicry really occurs for MHC molecules, its mechanism must be entirely different from that of overdominant selection; it should be related to the tolerance induction of T cells. It is known that the immune system of a host individual does not recognize its own MHC antigens as foreign molecules, because the T cells that react to them are eliminated in an early stage of development (MARRACK and KAPPLER 1987). Therefore, if a parasite antigen mimicks a host MHC antigen, it may escape the attack by the host immune system. In practice, however, mimicking of a MHC molecule would not be a simple matter even if we consider only the ARS. Although the ARS is composed of 57 amino acid residues in MHC class I molecules, these residues are distributed over the α_1 and α_2 domains of the molecule, each of which has about 90 amino acids (BJORKMAN et al. 1987). It seems therefore extremely difficult for a parasitic antigen to mimick a MHC molecule and still to have its own function.

It should also be noted that MHC antigens are expressed only on cell surfaces and that they recognize intracellularly processed peptides (KLEIN 1986). Since parasites (mostly viruses) enter into a cell, their antigens may be subject to intercellular processing even if the antigen mimicks some MHC molecules. If this happens, some of the processed peptides may now be

recognized by MHC molecules. (In general, the number of antigenic sites of a parasitic protein is small; BRACIALE et al. 1989). In this case, molecular mimicry will be no use for preventing the attack by the host immune system.

At any rate, in light of the recent findings of the mechanism of the function of MHC molecules, the molecular mimicry hypothesis does not seem to be very realistic. Note also that our mathematical model of molecular mimicry is unrealistic in one respect. According to our model, the MHC polymorphism in host organisms is maintained by the presence of an equally high degree of antigenic polymorphism in parasites. In practice, however, parasitic viruses such as the influenza virus A causing an epidemic in human populations are not really so polymorphic. Rather, an epidemic occurs because the invading virus is a new mutant and is tolerant to the host immune system. Once an antibody to the virus is developed in most individuals, the epidemic usually retreats quickly. This indicates that the extent of antigenic polymorphism in parasites in a given host population is very low at any given moment of time and thus the degree of MHC polymorphism maintained by molecular mimicry is likely to be much lower than that in our computation even if molecular mimicry really exists.

It should be noted that there are several other hypotheses for explaining MHC polymorphism (gene conversion, maternal-fetal incompatibility, mating preference, etc.; see NEI and HUGHES 1990 for a comprehensive discussion on this subject). In this paper we considered only the overdominance and frequency-dependent selection hypotheses, because these are the most contested hypotheses. As mentioned above, there is ample evidence for overdominant selection, but we cannot eliminate the hypothesis of frequency-dependent selection at the present time.

LITERATURE CITED

BRACIALE, T. J., M. T. SWEETSER, L. A. MORRISON, D. J. KITTLESEN and V. L. BRACIALE, 1989 Class I major histocompatibility complex-restricted cytolytic T lymphocytes recognize a limited number of sites on the influenza hemagglutinin. Proc. Natl. Acad. Sci. USA 86: 277–281.

BJORKMAN, P. J., M. A. SAPER, B. SAMRAOUI, W. S. BENNETT, J. L. STROMINGER and D. C. WILEY, 1987 The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. Nature 329: 512–518.

BODMER, W. F., 1972 Evolutionary significance of the HLA system. Nature 237: 139–145.

CLARKE, B., 1976 The ecological genetics of host-parasite relationships, pp. 87–103 in Genetic Aspects of Host-Parasite Relationships, edited by A. E. R. TAYLOR and R. M. MULLER. Blackwell, Oxford.

DAMIAN, R. T., 1964 Molecular mimicry: antigen sharing by parasite and host and its consequences. Am. Nat. 98: 129–149.

DAMIAN, R. T., 1987 Molecular mimicry revisited. Parasitol. Today 3: 263–266.

DOHERTY, P. C., and R. M. ZINKERNAGEL, 1975 Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature 256: 50–52.

EWENS, W. J., 1964 The maintenance of alleles by mutation. Genetics 50: 891–898.

EWENS, W. J., 1972 The sampling theory of selectively neutral alleles. Theor. Popul. Biol. 3: 87–112.

FIGUEROA, F., E. GUNTHER and J. KLEIN, 1988 MHC polymorphism predating speciation. Nature 335: 265–267.

FISHER, R. A., 1958 The Genetical Theory of Natural Selection, Second Revised Edition. Dover, New York.

HOWARD, J. C., 1987 MHC organization of the rat: evolutionary considerations, pp. 397–411 in Evolution and Vertebrate Immunity, edited by G. KELSOE and D. H. SCHULZE. University of Texas Press, Austin.

HUGHES, A. L., and M. NEI, 1988 Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335: 167–170.

HUGHES, A. L., and M. NEI, 1989 Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. USA 86: 958–962.

KIMURA, M., 1968 Evolutionary rate at the molecular level. Nature 217: 624–626.

KIMURA, M., 1971 Theoretical foundation of population genetics at the molecular level. Theor. Popul. Biol. 2: 174–208.

KIMURA, M., and J. F. CROW, 1964 The number of alleles that can be maintained in a finite population. Genetics 49: 725–738.

KINGMAN, J. F. C., 1982 On the genealogy of large populations. J. Appl. Probab. 19A: 27–43.

KLEIN, J., 1986 Natural History of the Major Histocompatibility Complex. Wiley, New York.

KLEIN, J., 1987 Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. Hum. Immunol. 19: 155–162.

KLEIN, J., AND F. FIGUEROA, 1986 Evolution of the major histocompatibility complex. CRC Crit. Rev. Immunol. 6: 295–389.

LAWLOR, D. A., F. E. WARD, P. D. ENNIS, A. P. JACKSON and P. PARHAM, 1988 HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. Nature 335: 268–271.

LI, W.-H., 1978 Maintenance of genetic variability under the joint effect of mutation, selection, and random drift. Genetics 85: 331–337.

MARRACK, P., and J. KAPPLER, 1987 The T Cell Receptor. Science 238: 1073–1078.

MARUYAMA, T., and M. NEI, 1981 Genetic variability maintained by mutation and overdominant selection in finite populations. Genetics 98: 441–459.

MAYER, W. E., M. JONKER, D. KLEIN, P. IVANYI, G. VAN SEVENTER and J. KLEIN, 1988 Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. EMBO J. 7: 2765–2774.

McCONNELL, T. J., W. S. TALBOT, R. A. McINDOE and E. K. WAKELAND, 1988 The origin of MHC class II gene polymorphism within the genus Mus. Nature 332: 651–654.

NEI, M., and D. GRAUR, 1984 Extent of protein polymorphism and the neutral mutation theory. Evol. Biol. 17: 73–118.

NEI, M., and A. L. HUGHES, 1990 Polymorphism and evolution of the major histocompatibility complex loci in mammals, in Evolution at the Molecular Level, edited by R. K. SELANDER, A. G. CLARK and T. S. WHITTAM. Sinauer Associates (in press).

NEI, M., and A. K. ROYCHOUDHURY, 1973 Probability of fixation and mean fixation time of an overdominant mutation. Genetics 74: 371–380.

OHTA, T., and M. KIMURA, 1973 A model of mutation appropriate to estimate the number of electrophoretically detectable

alleles in a finite population. Genet. Res. **22:** 201–204.

REDDEHASE, M. J., J. B. ROTHBARD and U. H. KOSZINOWSKI, 1989 A pentapeptide as minimal antigenic determinant for MHC class I-restricted T lymphocytes. Nature **337:** 651–653.

ROBERTSON, A., 1962 Selection for heterozygotes in small populations. Genetics **47:** 1291–1300.

SEGER, J., 1988 Dynamics of some simple host-parasite models with more than two genotypes in each species. Phil. Trans. R. Soc. Lond. B **319:** 541–555.

SETTE, A., S. BUUS, S. COLON, C. MILES and H. M. GREY, 1988 I-A^d-binding peptides derived from unrelated protein antigens share a common structural motif. J. Immunol. **141:** 45–48.

SHER, A., B. F. HALL and M. A. VADAS, 1978 Acquisition of murine major histocompatibility complex gene products by schistosomula of *Schistosoma mansoni.* J. Exp. Med. **148:** 46–57.

SIBLEY, C. G., and J. E. AHLQUIST, 1984 The phylogeny of the hominoid primates, as indicated by DNA-DNA hybridization. J. Mol. Evol. **20:** 2–15.

SIMPSON, A. J. G., D. SINGER, T. F. McCUTCHAN, D. L. SACKS and A. SHER, 1983 Evidence that schistosome MHC antigens are not synthesized by the parasite but are acquired from the host as intact glycoproteins. J. Immunol. **131:** 962–965.

SNELL, G. D., 1968 The H-2 locus of the mouse: observations and speculations concerning its comparative genetics and its polymorphism. Folia Biol. **14:** 335–358.

TAJIMA, F., 1983 Evolutionary relationship of DNA sequences in finite populations. Genetics **105:** 437–460.

TAKAHATA, N., and M. NEI, 1985 Gene genealogy and variance of interpopulational nucleotide differences. Genetics **110:** 325–344.

WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **6:** 217–250.

WATTERSON, G. A., 1977 Heterosis or neutrality? Genetics **85:** 789–814.

WRIGHT, S., 1939 The distribution of self-fertility alleles in populations. Genetics **24:** 538–552.

WRIGHT, S., 1966 Polyallelic random drift in relation to evolution. Proc. Natl. Acad. Sci. USA **55:** 1074–1081.

WRIGHT, S., 1969 Evolution and the genetics of populations. Vol. 2: The Theory of Gene Frequencies. University of Chicago Press, Chicago.

WRIGHT, S., and T. DOBZHANSKY, 1946 Genetics of natural populations XII. Experimental reproduction of some of the changes caused by natural selection in certain populations of *Drosophila pseudoobscura.* Genetics **31:** 125–145.

YOKOYAMA, S., and M. NEI, 1979 Population dynamics of sex-determining alleles in honey bees and self-incompatibility alleles in plants. Genetics **91:** 609–626.