Beltman JB, Borghans JAM & de Boer RJ (2002). Major Histocompatibility Complex: Polymorphism from Coevolution. In: Adaptive Dynamics of Infectious Diseases: In Pursuit of Virulence Management, eds. Dieckmann U, Metz JAJ, Sabelis MW & Sigmund K, pp. 210–221. Cambridge University Press.

15

Major Histocompatibility Complex: Polymorphism from Coevolution

Joost B. Beltman, José A.M. Borghans, and Rob J. de Boer

15.1 Introduction

There are many examples of pathogens adapting toward evasion of immune responses. Viruses, such as influenza, rapidly alter their genetic make-up, and each year there appear to be sufficient susceptible hosts that lack memory lymphocytes from previous influenza infections to give rise to a new epidemic (Both *et al.* 1983; Smith *et al.* 1999). During human immunodeficiency virus (HIV) infection, such alterations occur at an even faster rate, enabling the virus to escape repeatedly from the immune response within a single host (Nowak *et al.* 1991). Hosts, on the other hand, are selected for counteracting immune evasive strategies by pathogens. Since the generation time of hosts is typically much longer than that of pathogens, these host adaptations are expected to evolve much more slowly.

A well-known example commonly thought to reflect adaptation of hosts to pathogens is the polymorphism of major histocompatibility complex (MHC) molecules, which play a key role in cellular immune responses. When a pathogen infects a host cell, the proteins of the pathogen are degraded intracellularly, and a subset of the resultant peptides is loaded onto MHC molecules, which are transported to the cell surface. Once the peptides of a pathogen are presented on the surface of a cell in the groove of an MHC molecule, T lymphocytes can recognize them and mount an immune response.

The population diversity of MHC molecules is extremely large: for some MHC loci, over 100 different alleles have been identified (Parham and Ohta 1996; Vogel *et al.* 1999). Nevertheless, the mutation rate of MHC genes does not differ from that of most other genes (Parham *et al.* 1989a; Satta *et al.* 1993). Studies of nucleotide substitutions at MHC class I and II loci revealed Darwinian selection for diversity at the peptide-binding regions of MHC molecules. Within the MHC peptide-binding regions, the rate of nonsynonymous substitutions is significantly higher than the rate of synonymous substitutions; in other regions of the MHC, the reverse is true (Hughes and Nei 1988, 1989; Parham *et al.* 1989a, 1989b). Compared to the enormous population diversity of MHC molecules, their diversity within any one individual is quite limited. Humans express maximally six different MHC class I genes (HLA A, B, and C), which are codominantly expressed on all nucleated body cells. Additionally, there are maximally 12 different MHC class II molecules (HLA DP, DQ, and DR), which are expressed on specialized antigen-presenting cells (Paul 1999). The complete sequence of a human MHC has been

unraveled recently (MHC Sequencing Consortium 1999). Despite the high population diversity of MHC molecules, MHC genes appear to be extremely conserved evolutionarily. Allelic MHC lineages have persisted over long evolutionary time spans, often predating the divergence of present-day species (Klein 1980; Lawlor *et al.* 1988; Mayer *et al.* 1988; Klein and Klein 1991). As a consequence, individual MHC alleles from a species tend to be more closely related to particular MHC alleles from other species than to the majority of alleles that occur within the species (Parham *et al.* 1989b).

As a result of the high population diversity of MHC molecules, different individuals typically mount an immune response against different subsets of the peptides of any particular pathogen. Pathogens that escape from presentation by the MHC molecules of a particular host may thus not be able to escape from presentation in another host with different MHC molecules. MHC polymorphism may therefore seem a good strategy by which host populations counteract escape mechanisms of pathogens. This group selection argument, however, fails to explain how such a polymorphism could have evolved (Bodmer 1972).

The mechanisms behind the selection for MHC polymorphism have been debated for over three decades. A commonly held view is that MHC polymorphism arises from selection that favors heterozygosity. Since different MHC molecules bind different peptides, MHC heterozygous hosts can present a greater variety of peptides, and hence defend themselves against a larger variety of pathogens compared to MHC homozygous individuals. This hypothesis is known as the theory of overdominance or heterozygote advantage (Doherty and Zinkernagel 1975; Hughes and Nei 1988, 1989; Takahata and Nei 1990; Hughes and Nei 1992). A recent study of patients infected with HIV-1 supports this theory. It was shown that the degree of heterozygosity of MHC class I loci correlated with a delayed onset of acquired immunodeficiency syndrome (AIDS). Individuals who are homozygous at one or more loci typically progressed more rapidly to AIDS (Carrington *et al.* 1999).

It has been argued that selection for heterozygosity alone cannot explain the large MHC diversity observed in nature (Parham *et al.* 1989b; Wills 1991). Although there is general agreement upon the significance of overdominant selection, it has been proposed that additional selection pressures must be involved in the maintenance of the MHC polymorphism (Parham *et al.* 1989b; Wills 1991). A frequently studied additional mechanism is frequency-dependent selection. The corresponding theory states that evolution favors pathogens that avoid presentation by the most common MHC molecules in the host population. Thus, there is a permanent selection force favoring hosts that carry rare (e.g., new) MHC molecules. Since hosts with rare MHC alleles have a higher fitness, the frequency of rare MHC alleles will increase, and common MHC alleles will become less frequent. The result is a dynamic equilibrium, maintaining a polymorphic population (Snell 1968; Bodmer 1972; Slade and McCallum 1992; Beck 1984).

Both selection for heterozygosity and frequency-dependent selection have been modeled extensively. Most models address either of the two hypotheses, and

are so-called "top-down" models. Assuming that heterozygous individuals have a higher fitness than homozygous individuals (see, for example, Takahata and Nei 1990), or assuming that individuals carrying rare alleles have a higher fitness than individuals carrying common alleles (see, for example, Takahata and Nei 1990; Wills 1991; Wills and Green 1995), it has been shown that an existing MHC polymorphism can be maintained.

Here we take a more mechanistic approach by making no assumptions about selective advantages or disadvantages. We develop a computer simulation to study the coevolution of diploid hosts with haploid pathogens. By comparing simulations in which pathogens do coevolve with simulations in which they do not, our model allows us to study the effect of selection for heterozygosity and frequency-dependent selection on the polymorphism of MHC molecules. Starting from a population diversity of only one MHC molecule, we show that a diverse set of functionally different MHC molecules is obtained. Our analysis demonstrates that selection involving rapid evolution of pathogens can account for a much larger MHC diversity than can selection for heterozygosity alone.

15.2 Simulating the Coevolution of Hosts and Pathogens

We have developed a genetic algorithm (Holland 1975) to investigate the coevolution of pathogens and MHC molecules. Genetic algorithms are frequently applied as problem-solving tools, using the principles of evolution to find solutions in, for example, optimization problems. We instead use them here as a simulation of evolution (see also Forrest 1993; Pagie and Hogeweg 1997), and thereby take them "right back to where they started from" (Huynen and Hogeweg 1989).

In our simulations, we consider a population of N_{host} diploid hosts, each represented by a series of bit strings coding for two alleles at N_L MHC loci. Pathogens are haploid and occur in N_S independent species of maximally N_G different genotypes. For simplicity, we omit the complex process of protein degradation into peptides, and model each pathogen by N_P bit strings that represent the set of peptides that can possibly be recognized by a host. Peptide presentation by an MHC molecule can occur at different positions on the MHC molecule, and is modeled by complementary matching. Peptides are L_P bits long, and MHC molecules are L_M bits long. For each peptide of a pathogen and for each MHC molecule of a host, we seek the position at which the peptide finds the maximal complementary match. If the number of complementary bits at this position is at least a predefined threshold L_T, the peptide is considered to be presented by that particular MHC molecule. In the simulations presented here, pathogens consist of $N_P = 20$ different peptides, which are $L_P = 12$ bits long. MHC molecules are $L_M = 35$ bits long, and present a peptide if, of the 12 peptide bits, at least 11 ($= L_T$) match with the MHC. Thus, the chance that a random MHC molecule presents a randomly chosen peptide is 7.3%. [The chance that a random peptide binds at a random, predefined position of an MHC molecule is $p_b = \sum_{j=L_T}^{L_P} \binom{L_P}{j} (0.5)^{L_P}$. Thus, the chance that a random MHC molecule presents a randomly chosen peptide is $1 - (1 - p_b)^{L_M - L_P + 1} = 7.3\%$.] Also, the chance that a pathogen of $N_P = 20$

peptides escapes presentation by a randomly chosen MHC molecule is $p_e = 22\%$. Hosts that carry different MHC molecules hence typically present different peptides of the pathogens.

The quality of different MHC molecules varies. Some MHC molecules may be more stably expressed on the surfaces of host cells than others, or fold into a better peptide-binding groove. To model such MHC differences, a random quality parameter $0 < Q < 1$ (drawn from a uniform distribution) is attributed to every MHC molecule in the population. These quality differences between MHC molecules prevent extensive drift in simulations with random pathogens. The fitness contribution of a host–pathogen interaction is determined by the quality of the best MHC molecule that is able to present a peptide of the pathogen. We omit the role of lymphocytes by assuming that every presented peptide is recognized by at least one functional clonotype. The role of lymphocytes, in particular the (functional) deletion of lymphocytes during self-tolerance induction, is to be reported in a follow-up paper (Borghans *et al.*, unpublished; see also Borghans *et al.* 1999).

At each generation, every host interacts with every genotypically different pathogen. To account for the shorter generation time of pathogens, we can allow for several pathogen generations per host generation. The fitness f_h of a host is proportional to the fraction of pathogens it is able to present,

$$f_h = \sum_{j=1}^{N_{\text{path}}} Q_j / N_{\text{path}} \,, \tag{15.1}$$

where N_{path} denotes the total number of different genotypes in the pathogen populations. Q_j denotes the quality of the best MHC molecule that presents at least one peptide of pathogen j; we set Q_j to zero if none of the MHC molecules of a host present pathogen j. Similarly, the fitness f_p of a pathogen is proportional to the fraction of hosts that the pathogen can infect without being presented on the host's MHC molecules,

$$f_p = 1 - \sum_{k=1}^{N_{\text{host}}} Q_k / N_{\text{host}} \,, \tag{15.2}$$

where Q_k is the quality of the best MHC molecule of host k that presents at least one peptide of the pathogen. Again, Q_k is set to zero if none of the MHC molecules of host k present the pathogen.

At the end of each generation, all individuals are replaced by fitness-proportional reproduction. The sizes of the host population and all pathogen species remain constant. All fitnesses are rescaled such that the highest fitness in each host population and in each pathogen species becomes one and the lowest becomes zero. The different individuals in the host population, and the different genotypes in each pathogen species, reproduce according to a fitness-dependent

reproduction function,

$$\Pr(j) = \frac{e^{\overline{f}_j}}{\sum_{k=1}^{N} e^{\overline{f}_k}} , \tag{15.3}$$

where $\Pr(j)$ is the reproduction probability of host j or pathogen genotype j, $\overline{f}_j$ denotes its rescaled fitness, and N is the total number of different individuals in the host population or genotypes in the particular pathogen species. Pathogen genotypes reproduce asexually; new-born pathogens come from parents of the same pathogen species. New-born hosts have two parents, each of which donates a randomly selected MHC allele. During reproduction, point mutations can occur. Both peptides and MHC molecules have a mutation chance of $p_{\text{mut}} = 0.1\%$ per bit per generation. The chance for a new-born host to receive a nonmutated MHC molecule is thus $(1 - p_{\text{mut}})^{L_M} = 96.6\%$, and the chance for a new-born pathogen to receive a nonmutated peptide is $(1 - p_{\text{mut}})^{L_P} = 98.8\%$. One cycle of fitness determination, reproduction, and mutation defines a generation. We study evolution over many generations.

15.3 Dynamically Maintained Polymorphism

The simulation model allows us to study the mutual influence of host and pathogen coevolution on the composition of MHC molecules in the host population, and of peptides in the pathogen species. In particular, we:

- Study whether a polymorphic set of MHC molecules can develop from an initially nondiverse host population;
- Investigate the relative roles of frequency-dependent selection and selection for heterozygosity in maintaining the polymorphism of MHC molecules.

All simulations are initialized with random pathogen genotypes, and all hosts initially carry identical MHC molecules – that is, there is neither variation between MHC molecules within the hosts, nor between the hosts. Two examples of such simulations are shown in Figure 15.1, in which the average fitnesses of the pathogens and the hosts are shown as a function of the host generation number t. To study the effect of the typically short generation time of pathogens, we consider two different cases. In one of them (Figures 15.1a and 15.1b), the pathogens evolve as fast as the hosts (i.e., one parasite generation per host generation), while in the other case (Figures 15.1c and 15.1d), the pathogens evolve 100 times faster than the hosts. Since there is no initial MHC diversity, in both simulations the pathogens immediately attain a relatively high fitness and the hosts a correspondingly low fitness. Any pathogen that is able to infect one host is able to infect all hosts, and hence it rapidly takes over the pathogen population. Under this selective pressure caused by the pathogens, the hosts develop an MHC polymorphism (as is shown in Section 15.4) and, in so doing, regain a high fitness. After about 300 host generations, a quasi-equilibrium is approached that is followed until generation $t = 1\,000$. A similar equilibrium is attained if the host population is initialized

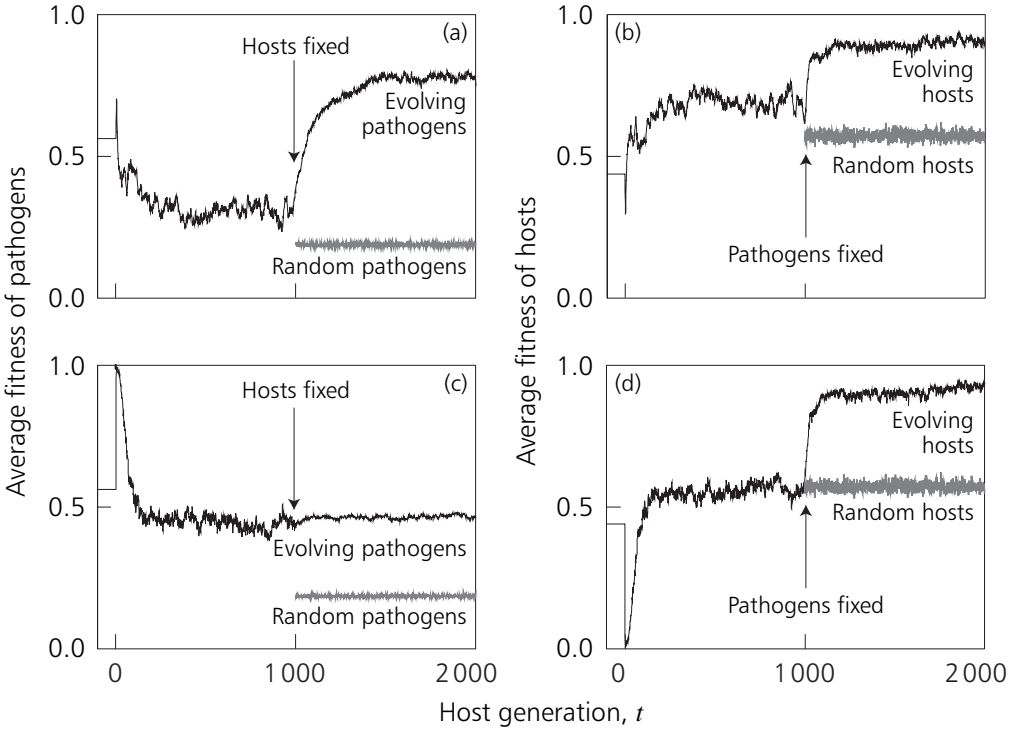

Figure 15.1 Fitness of hosts and pathogens. The average fitnesses of pathogens (a, c) and hosts (b, d) in a simulation in which the generation time of the pathogens is equal to that of the hosts (a, b), and in a simulation in which the generation time of the pathogens is 0.01 times that of the hosts (c, d) are plotted against the host generation t. Note that, by Equations (15.1) and (15.2), the average host and pathogen fitnesses in a single simulation always totals one. The simulations are initialized with MHC-identical hosts and random pathogens. Coevolution is stopped at host generation $t = 1\,000$. We either stop the evolution of the hosts and let only the pathogens carry on evolving (a, c), or we stop the evolution of the pathogens and let only the hosts carry on evolving (b, d). The gray curves denote the average fitness of randomly created pathogens evaluated against the fixed host populations of generation $t = 1\,000$ (a, c), and the average fitness of random, heterozygous hosts evaluated against the fixed pathogen populations of generation $t = 1\,000$ (b, d). Other parameters: $N_{\text{host}} = 200$, $N_L = 1$, $N_S = 50$, $N_G = 10$, $N_P = 20$, $L_P = 12$, $L_M = 35$, and $L_T = 11$.

with random MHC molecules (not shown here). The average fitnesses during the quasi-equilibrium depend on the relative generation time of the pathogens. The faster the pathogens evolve, the higher their average fitness, and the lower the average fitness of the hosts (Figure 15.2). Once the pathogens evolve 100 times faster than the hosts, the average pathogen fitness saturates.

The quasi-equilibrium that is approached is a dynamic one. As in a Red Queen situation, hosts and pathogens continually counteract each other by adaptation. This follows from additional simulations in which, from $t = 1\,000$ onward, further evolution of either the hosts or the pathogens is prevented. If the pathogens and the hosts evolve equally fast, and the evolution of the hosts is subsequently halted, the pathogens markedly increase their fitness (Figure 15.1a). Such an increase of the average pathogen fitness is not observed, however, if the pathogens

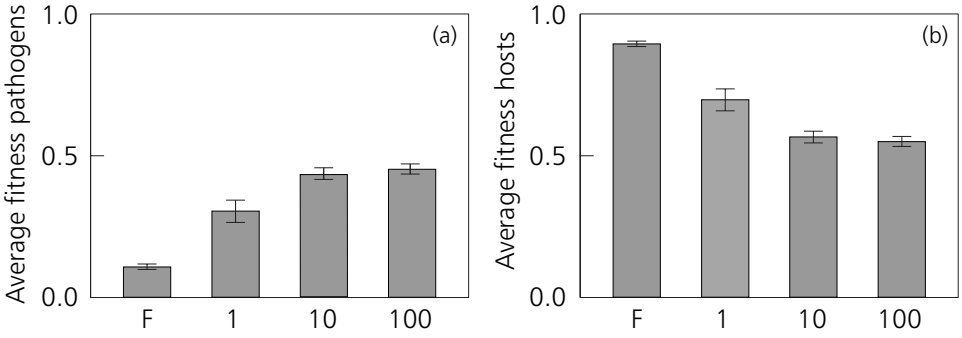

Figure 15.2 The average fitnesses of pathogens (a) and hosts (b) over the final 100 generations of the coevolution (i.e., between $t = 900$ and $t = 1\,000$). Results are shown for four different simulation types: F = fixed (nonevolving) pathogens, 1 = pathogens evolving as fast as the hosts, 10 = pathogens evolving 10 times faster than the hosts, 100 = pathogens evolving 100 times faster than the hosts. In the coevolutionary simulations, there are typically two different genotypes per pathogen species (not shown). We therefore initialized the F simulation with two randomly chosen pathogen genotypes per species. The error bars denote the standard deviations of the average host and pathogen fitnesses in time. Parameters are set as in Figure 15.1.

were evolving 100 times faster than the hosts before the evolution of the hosts was stopped (Figure 15.1c). Their short generation time apparently enables the pathogens to adapt "completely" during each host generation even before the host population is frozen. Stopping the evolution of the hosts then hardly makes a difference. Remarkably, once the evolution of the hosts is stopped, the pathogens that used to evolve as fast as the hosts attain a significantly higher average fitness (Figure 15.1a) than the pathogens that used to evolve faster than the hosts (Figure 15.1c). The reason for this difference is addressed in Section 15.4. Likewise, if the evolution of the pathogens is stopped and only the hosts carry on evolving, they evolve such that they can resist almost all pathogens – that is, they approach a fitness of one (Figures 15.1b and 15.1d). Pathogens that evolve in a nonevolving host population attain a larger average fitness than random pathogens (see the gray curves in Figures 15.1a and 15.1c). Similarly, evolving hosts in the presence of a nonevolving pathogen population attain a higher fitness than random, heterozygous hosts (see the gray curves in Figures 15.1b and 15.1d). Thus, evolving hosts and pathogens have the capacity to adapt to nonevolving populations of pathogens or hosts, respectively.

15.4 Host and Pathogen Evolution

As soon as a coevolutionary simulation is started, the number of different MHC molecules in the host population rapidly increases to reach a high quasi-equilibrium diversity (Figure 15.3). This diversification also occurs if the pathogens do not evolve at all. In that case, the high population diversity of MHC molecules results from selection that favors heterozygous hosts. The faster the pathogens evolve, however, the larger the MHC population diversity becomes (Figure 15.4a).

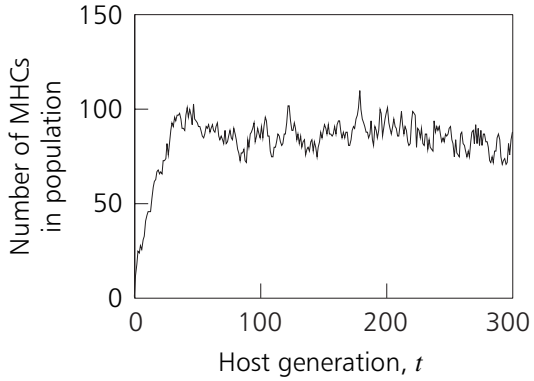

Figure 15.3 Evolution of MHC polymorphism. The number of different MHC molecules in the host population is shown from the start of the coevolution ($t = 0$) until host generation $t = 300$. The generation time of the pathogens is 100 times shorter than that of the hosts. Parameters are set as in Figure 15.1.

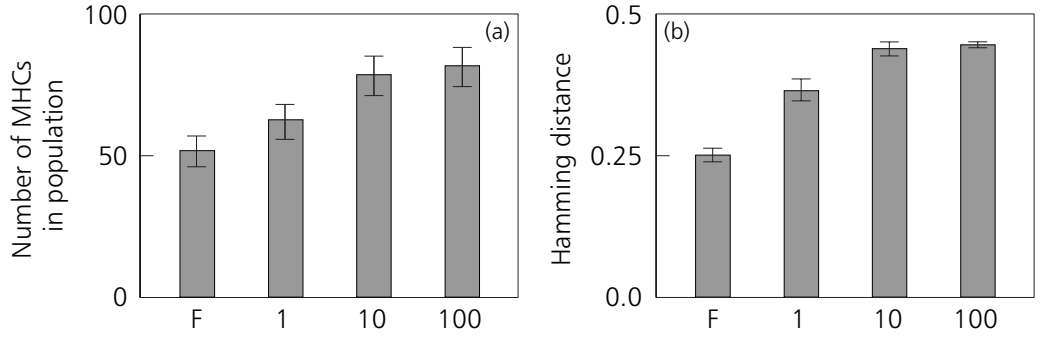

Figure 15.4 MHC molecules become functionally polymorphic. (a) The average number of different MHC molecules in the host population. (b) The average of the Hamming distances between all possible pairs of different MHC molecules in the host population. Parameters are set as in Figure 15.1. Horizontal axis labels are as explained in Figure 15.2.

To check if the MHC molecules that arise in a host population are really different from each other, and do not differ at a few mutations only, we have calculated the average genetic distance (Hamming distance) between all different MHC molecules in the host population (Figure 15.4b). Evolution of the pathogens appears to increase MHC diversity; the shorter the generation time of the pathogens, the larger the genetic distance between the MHC molecules of the hosts. Thus, rapidly coevolving pathogens trigger selection for a functionally diverse set of MHC molecules.

To measure the extent to which the pathogens evade presentation on the MHC molecules of the hosts, we calculated the average fraction of peptides from the pathogen genotypes presented by the MHC molecules in the host population. The faster the pathogens evolve, the better their evasion of presentation by the hosts' MHC molecules (see the gray bars in Figure 15.5). If the pathogens evolve, the average fraction of peptides presented by the MHC molecules of the hosts is smaller

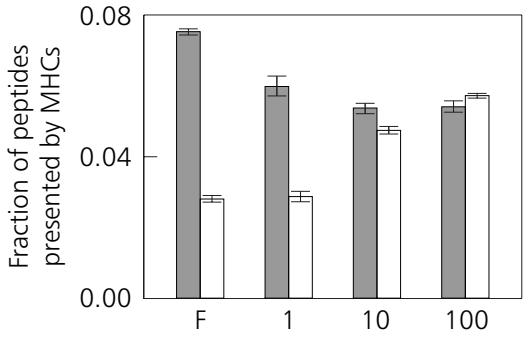

Figure 15.5 Pathogens evolve toward evasion of presentation by the particular MHC molecules present in the host population. The average presentation efficiency of the MHC molecules (i.e., the average fraction of peptides from the pathogen genotypes presented by the MHC molecules) is plotted for different pathogen generation times. The gray bars show the average presentation efficiency of the MHC molecules of coevolving hosts – that is, between host generation $t = 900$ and $t = 1\,000$ in Figure 15.1. The white bars denote the average presentation efficiency of the MHC molecules that have been frozen at host generation $t = 1\,000$ in Figure 15.1, after the pathogens have been allowed to evolve for 1 000 generations – that is, between host generation $t = 1\,900$ and $t = 2\,000$ in Figure 15.1. Parameters are set as in Figure 15.1. Horizontal axis labels are as explained in Figure 15.2.

than the expected 7.3% calculated above for MHC molecules binding random peptides. Thus, the pathogens in our simulations indeed evolve toward evasion of presentation by the particular MHC molecules present in the host population.

We applied a similar analysis to the simulations in which either the hosts or the pathogens are prevented from evolving. This analysis partially explains our earlier observation that pathogens evolving in a frozen host population stringently selected by rapidly coevolving pathogens (Figure 15.1c) attain a lower fitness than pathogens evolving in a host population selected only moderately (Figure 15.1a). If the pathogens do not evolve faster than the hosts, the fraction of pathogen peptides recognized by the hosts' MHC molecules decreases dramatically when the evolution of the hosts is stopped (see the white bars denoted by F and 1 in Figure 15.5). Apparently, during the coevolution the hosts specialize on the particular pathogens present in the population. This specialization enables the pathogens to escape immune recognition once the evolution of the hosts is stopped. In contrast, if the pathogens evolve faster than the hosts during the coevolution, the hosts cannot specialize on the particular pathogens present in the population. As a consequence, the pathogens fail to escape immune recognition once the evolution of the hosts is stopped (see the white bars denoted by 10 and 100 in Figure 15.5). Another reason why the evolutionary history of a frozen host population influences the escape possibilities of a pathogen lies in the polymorphism of the hosts' MHC molecules. As discussed above, the faster the evolution of the pathogens is, the more polymorphic the MHC molecules of the hosts become. Thus, pathogens evolving in a frozen host population that used to be stringently selected by rapidly coevolving pathogens have more difficulty in escaping presentation by the highly polymorphic MHC molecules of the hosts.

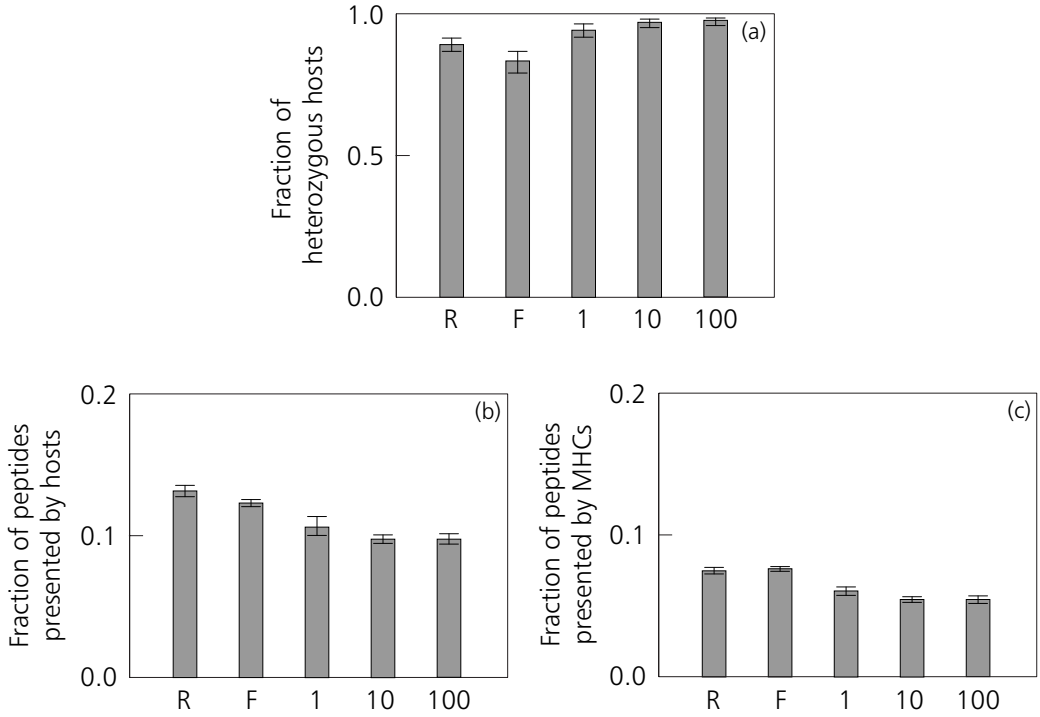

Figure 15.6 Hosts become functionally heterozygous. (a) The average fraction of heterozygous hosts. (b) The average fraction of peptides from the pathogens presented by the hosts. (c) The average fraction of peptides from the pathogens presented by the individual MHC molecules of the hosts. R denotes the simulation in which pathogens are introduced randomly at every host generation. Like the fixed pathogen population denoted by F, randomly introduced pathogen species consist of two randomly created pathogen genotypes per species. Parameters are set as in Figure 15.1. Horizontal axis labels are as explained in Figure 15.2.

15.5 Heterozygosity versus Frequency-dependent Selection

Since the evolution of pathogens can be switched off in our model, we can separately study the effect of selection for heterozygosity. In coevolutionary simulations, there is selection for heterozygosity as well as frequency-dependent selection. To exclude evolution of the pathogens, one possibility is to let the hosts evolve in response to a fixed pathogen population. As we have seen, in that case hosts adapt to the specific pathogens that are present (Figure 15.5). To exclude this specialization, we have also performed simulations in which at every host generation all pathogens are replaced by random ones (R in Figure 15.6).

The role of selection for heterozygosity appears to be strong under all conditions. During the quasi-equilibrium, the fraction of heterozygous hosts is always close to one (Figure 15.6a). To check if this heterozygosity is also functional (i.e., if the two MHC molecules of a host generally present different peptides), we compare the average fraction of peptides from the pathogens that are presented by the hosts (Figure 15.6b) with the average fraction of peptides from the pathogens presented by their individual MHC molecules (Figure 15.6c). It appears that in all simulations, the hosts (with their two MHC molecules) present nearly twice as

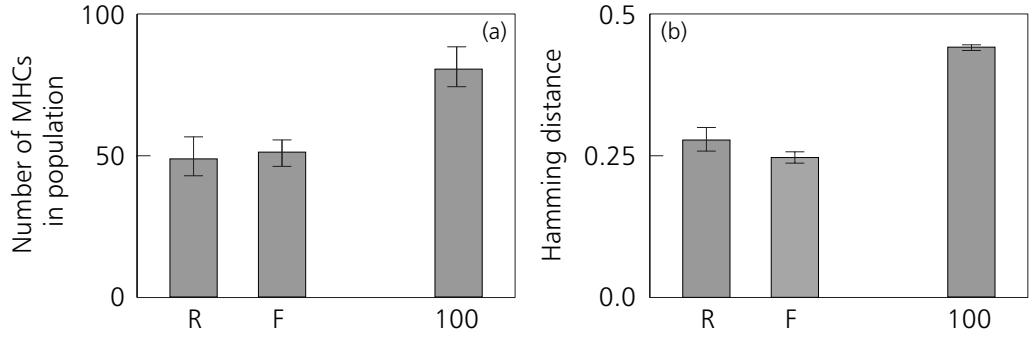

Figure 15.7 Selection for heterozygosity versus frequency-dependent selection. (a) The average number of different MHC molecules in the host population, and (b) the average Hamming distance between the different MHC molecules. We show a coevolutionary simulation in which the pathogens evolve 100 times faster than the hosts (100), and two simulations in which the pathogens do not evolve (R and F). The coevolutionary simulation represents the MHC diversity that evolves in the presence of both frequency-dependent selection and selection for heterozygosity, while the two latter simulations (R and F) represent the MHC diversity that evolves under selection for heterozygosity only.

many peptides as their individual MHC molecules. Thus, the hosts in our simulations indeed typically carry functionally different MHC molecules.

To study the relative roles of selection for heterozygosity and frequency-dependent selection, we compare the MHC polymorphism arising in the absence and presence of frequency-dependent selection. Figure 15.7a shows that heterozygosity plus frequency-dependent selection (i.e., a simulation with evolving pathogens, denoted by 100) results in a much higher degree of polymorphism than selection for heterozygosity alone (i.e., simulations with nonevolving pathogens, R and F). The average genetic differences between the MHC molecules that arise support this notion (see Figure 15.7b). Summarizing, our simulations show that a polymorphic set of MHC molecules rapidly develops in an initially nondiverse host population, and that selection by coevolving pathogens can account for a much larger population diversity of MHC molecules than mere selection for heterozygosity can.

15.6 Discussion

We have shown that both the origin and the maintenance of MHC polymorphism can be understood in a model that does not assume any *a priori* selective advantage of heterozygous hosts or hosts with rare MHC molecules. By starting our simulations with MHC-identical hosts, we have studied a "worst-case" scenario. Polymorphisms of MHC-like molecules seem to have been present since colonial or multicellular life (Buss and Green 1985). Thus, the *origin* of MHC polymorphism may not lie in immune function. For example, de Boer (1995) showed that in primitive colonial organisms the preservation of "genetic identity" is sufficient to account for highly polymorphic histocompatibility molecules.

Our simulation model demonstrates that coevolution of hosts and pathogens yields a larger MHC polymorphism than merely selection for heterozygosity. Our

analysis thus supports the view that additional selection pressures on top of overdominant selection do play a role in the evolution of the MHC polymorphism (Parham *et al.* 1989b; Wills 1991). It has been shown experimentally that many MHC alleles have persisted for significant evolutionary periods of time (Klein 1980; Lawlor *et al.* 1988; Mayer *et al.* 1988; Klein and Klein 1991). This has been used as an argument against frequency-dependent selection (Hughes and Nei 1988), but was later demonstrated to be compatible with selection for rare MHC molecules (Takahata and Nei 1990). Analysis of the persistence of particular MHC alleles in our simulations would allow us to study this in more detail.

To increase the speed of our simulations, we used a rather high mutation frequency for the hosts' MHC molecules: $p_{\text{mut}} = 0.001$ per bit per generation. Indeed, decreasing this mutation frequency resulted in a lower MHC population diversity. Increasing the host population size in our simulations, on the other hand, increased the MHC polymorphism. Using a mutation frequency for MHC molecules of $p_{\text{mut}} = 10^{-6}$ and a host population size of $N_{\text{host}} = 1\,000$ hosts, we still found a population diversity of approximately 30 different MHC molecules. Independently of the choices of p_{mut} and N_{host}, the MHC polymorphism attained in coevolutionary simulations was always considerably (e.g., fivefold) higher than the polymorphism arising under overdominant selection only (results not shown).

Regarding the enormous population diversity of MHC molecules observed in nature (Parham and Ohta 1996; Vogel *et al.* 1999), it is surprising that the number of different MHC molecules expressed per individual is quite limited (Paul 1999). In our simulations, hosts carry only one MHC gene. What would change if this number of MHC genes per individual increased? Individuals expressing more MHC genes are expected to have a selective advantage, in that more pathogens would be presented. This selective advantage vanishes, however, once the chance to present (at least one peptide from) any pathogen approaches 100%. For the parameter setting used here, the chance that a random pathogen consisting of 20 peptides evades presentation by a single MHC molecule is $p_e = 22\%$. In the absence of pathogen evolution, expression of about 10 different MHC molecules would thus be sufficient to ensure the presentation of virtually any pathogen. In coevolutionary situations, however, the selection for expression of MHC molecules that are different from the other MHC molecules in the population would remain. This selection only disappears when the number of different MHC molecules per individual becomes so large that every host is expected to present all pathogen peptides. If individuals no longer draw different "samples" from the pool of peptides from each pathogen, pathogens may be expected to exploit this "predictability" of the hosts' immune responses (Wills and Green 1995). We have demonstrated, for instance, that increasing the number of MHC loci increases the likelihood of autoimmunity (Borghans and de Boer 2001). Extension of the current model with host self-molecules and a variable number of MHC genes sheds light on the role of such mechanisms in the maintenance of the MHC polymorphism (Borghans *et al.*, unpublished).

References

References in the book in which this chapter is published are integrated in a single list, which appears on pp. 465–514. For the purpose of this reprint, references cited in the chapter have been assembled below.

Beck K (1984). Coevolution: Mathematical analysis of host–parasite interactions. *Journal of Mathematical Biology* **19**:63–77

Bodmer WF (1972). Evolutionary significance of the HL-A system. *Nature* **237**:139–145

Borghans JAM, Noest AJ & De Boer RJ (1999). How specific should immunological memory be? *Journal of Immunology* **163**:569–575

Both GW, Sleigh MJ, Cox NJ & Kendal AP (1983). Antigenic drift in influenza virus H3 hemagglutinin from 1968 to 1980: Multiple evolutionary pathways and sequential amino acid changes at key antigenic sites. *Journal of Virology* **48**:52–60

Buss LW & Green DR (1985). Histoincompatibility in vertebrates: The relict hypothesis. *Developmental and Comparative Immunology* **9**:191–201

Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K & O'Brien SJ (1999). HLA and HIV-1: Heterozygote advantage and B*35-Cw*04 disadvantage. *Science* **283**:1748–1752

De Boer RJ (1995). The evolution of polymorphic compatibility molecules. *Molecular Biology and Evolution* **12**:494–502

Doherty PC & Zinkernagel RM (1975). Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* **256**:50–52

Forrest S (1993). Genetic algorithms: Principles of natural selection applied to computation. *Science* **261**:872–878

Holland JH (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI, USA: University of Michigan Press

Hughes AL & Nei M (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**:167–170

Hughes AL & Nei M (1989). Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection. *Proceedings of the National Academy of Sciences of the USA* **86**:958–962

Hughes AL & Nei M (1992). Models of host–parasite interaction and MHC polymorphism. *Genetics* **132**:863–864

Huynen MA & Hogeweg P (1989). Genetic algorithms and information accumulation during the evolution of gene regulation. In *Proceedings of the Third International Conference on Genetic Algorithms*, ed. Schaffer JD. San Mateo, CA, USA: Morgan Kaufmann Publishers

Klein J (1980). Generation of diversity to MHC loci: Implications for T cell receptor repertoires. In *Immunology 80*, eds. Fougereau M & Dausset J. London, UK: Academic Press

Klein J & Klein D (1991). *Molecular Evolution of the MHC Complex*. Heidelberg, Germany: Springer

Lawlor DA, Ward FE, Ennis PD, Jackson AP & Parham P (1988). HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. *Nature* **335**:268–271

Mayer WE, Jonker M, Klein D, Ivanyi P, Van Seventer, G. & Klein J (1988). Nucleotide sequences of chimpanzee MHC class I alleles: Evidence for trans-species mode of evolution. *The EMBO Journal* **7**:2765–2774

MHC Sequencing Consortium (1999). Complete sequence and gene map of a human major histocompatibility complex. *Nature* **401**:921–923

Nowak MA, Anderson RM, McLean AR, Wolfs TF, Goudsmit J & May RM (1991). Antigenic diversity thresholds and the development of AIDS. *Science* **254**:963–969

Pagie L & Hogeweg P (1997). Evolutionary consequences of coevolving targets. *Evolutionary Computation* **5**:401–418

Parham P & Ohta T (1996). Population biology of antigen presentation by MHC class I molecules. *Science* **272**:67–74

Parham P, Benjamin RJ, Chen BP, Clayberger C, Ennis PD, Krensky AM, Lawlor DA, Littman DR, Norment AM, Orr HT, Salter RD & Zemmour J (1989a). Diversity of class I HLA molecules: Functional and evolutionary interactions with T cells. *Cold Spring Harbor Symposia on Quantitative Biology* **54**:529–543

Parham P, Lawlor DA, Lomen CE & Ennis PD (1989b). Diversity and diversification of HLA-A,B,C alleles. *Journal of Immunology* **142**:3937–3950

Paul WE (1999). *Fundamental Immunology*. New York, NY, USA: Raven Press

Satta Y, O'hUigin C, Takahata N & Klein J (1993). The synonymous substitution rate of the major histocompatibility complex loci in primates. *Proceedings of the National Academy of Sciences of the USA* **90**:7480–7484

Slade RW & McCallum HI (1992). Overdominant versus frequency-dependent selection at MHC loci. *Genetics* **132**:861–864

Smith DJ, Forrest S, Ackley DH & Perelson AS (1999). Variable efficacy of repeated annual influenza vaccination. *Proceedings of the National Academy of Sciences of the USA* **96**:14001–14006

Snell GD (1968). The H-2 locus of the mouse: Observations and speculations concerning its comparative genetics and its polymorphism. *Folia Biologica (Prague)* **14**:335–358

Takahata N & Nei M (1990). Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* **124**:967–978

Vogel TU, Evans DT, Urvater JA, O'Connor DH, Hughes AL & Watkins DI (1999). Major histocompatibility complex class I genes in primates: Coevolution with pathogens. *Immunological Reviews* **167**:327–337

Wills C (1991). Maintenance of multiallelic polymorphism at the MHC region. *Immunological Reviews* **124**:165–220

Wills C & Green DR (1995). A genetic herd-immunity model for the maintenance of MHC polymorphism. *Immunological Reviews* **143**:263–292