

José A. M. Borghans · Joost B. Beltman ·  
Rob J. De Boer

## MHC polymorphism under host-pathogen coevolution

Received: 6 June 2003 / Revised: 11 November 2003 / Published online: 13 January 2004  
© Springer-Verlag 2004

**Abstract** The genes encoding major histocompatibility (MHC) molecules are among the most polymorphic genes known for vertebrates. Since MHC molecules play an important role in the induction of immune responses, the evolution of MHC polymorphism is often explained in terms of increased protection of hosts against pathogens. Two selective pressures that are thought to be involved are (1) selection favoring MHC heterozygous hosts, and (2) selection for rare MHC alleles by host-pathogen coevolution. We have developed a computer simulation of coevolving hosts and pathogens to study the relative impact of these two mechanisms on the evolution of MHC polymorphism. We found that heterozygote advantage per se is insufficient to explain the high degree of polymorphism at the MHC, even in very large host populations. Host-pathogen coevolution, on the other hand, can easily account for realistic polymorphisms of more than 50 alleles per MHC locus. Since evolving pathogens mainly evade presentation by the most common MHC alleles in the host population, they provide a selective pressure for a large variety of rare MHC alleles. Provided that the host population is sufficiently large, a large set of MHC alleles can persist over many host generations under host-pathogen coevolution, despite the fact that allele frequencies continuously change.

**Keywords** MHC polymorphism · Evolution · Heterozygote advantage · Frequency-dependent selection · MHC diversity

**Electronic Supplementary Material** Supplementary material is available in the online version of this article at <http://dx.doi.org/10.1007/s00251-003-0630-5>

### Introduction

The population diversity of MHC molecules is extremely large: for some MHC loci, hundreds of different alleles have been identified (Parham and Ohta 1996; Vogel et al. 1999). The high rate of nonsynonymous substitutions within the MHC-peptide binding regions, compared with normal rates in other regions of the MHC (Hughes and Nei 1988, 1989; Parham et al. 1989a, 1989b), suggests that there is Darwinian selection for diversity in MHC-peptide binding. Due to the large population diversity of MHC molecules, different individuals typically mount an immune response against different subsets of the peptides of any particular pathogen (Barouch et al. 1995). Pathogens that escape presentation by the MHC molecules of one particular host may thus not be able to escape presentation in another host with different MHC molecules.

The mechanisms that have been proposed to explain the evolution of MHC polymorphism vary from MHC-dependent mate selection, and preferential abortion, to various pathogen-driven selection pressures [see (Apanius et al. 1997) for an extensive review]. A commonly held view is that the polymorphism is due to selection favoring MHC heterozygous hosts (Doherty and Zinkernagel 1975; Hughes and Nei 1988, 1989; Takahata and Nei 1990; Hughes and Nei 1992). 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 with MHC homozygous individuals (Penn et al.

---

J. A. M. Borghans · J. B. Beltman · R. J. De Boer  
Theoretical Biology, Utrecht University,  
Padualaan 8, 3584 CH Utrecht, The Netherlands

J. A. M. Borghans  
Biologie des Populations Lymphocytaires, Institut Pasteur,  
25–28 Rue du Dr Roux, 75015 Paris, France

*Present address:*

J. A. M. Borghans (✉), Clinical Viro-Immunology,  
Sanquin Research at CLB,  
Plesmanlaan 125, 1066 CX Amsterdam, The Netherlands  
e-mail: J.Borghans@sanquin.nl  
Tel.: +31-20-5123275  
Fax: +31-20-5123310

2002). It has indeed been shown that MHC heterozygosity correlates negatively with disease in HIV-infected patients (Carrington et al. 1999), with HTLV-1 infection in humans (Jeffery et al. 2000), and with LCMV infection in mice (Weidt et al. 1995).

Although there is general agreement upon the significance of heterozygote advantage, several authors have questioned whether it can explain the large MHC diversity observed in nature (Lewontin et al. 1978; Aoki 1980; Parham et al. 1989b; Lawlor et al. 1990; Wills 1991). An additional mechanism that could drive the polymorphism of MHC molecules is frequency-dependent selection by host-pathogen coevolution. Since evolution favors pathogens that avoid presentation by the most common MHC molecules in the host population, there is selection for hosts with rare, (e.g., new) MHC molecules (Trachtenberg et al. 2003). Because hosts with rare MHC alleles have the highest fitness, the frequency of rare MHC alleles will increase, and common MHC alleles will become less frequent, resulting in a dynamic polymorphism (Snell 1968; Bodmer 1972; Beck 1984; Slade and McCallum 1992). By coevolution, hosts become a “moving target” for parasites (Penn and Potts 1999), an argument that has also been used to explain the evolution of sexual reproduction (Hamilton et al. 1990). Recent studies of a snail infected by a trematode parasite provide support for host-pathogen coevolution, by demonstrating that the parasite adapts to be most virulent in the dominant host genotype (Dybdahl and Lively 1998; Lively and Dybdahl 2000). Another example is HIV-1, which reaches a higher viral load in individuals with common MHC molecules (Trachtenberg et al. 2003). HIV-1 has also evolved protein regions devoid of T-cell epitopes, because they lack immuno-proteasome cleavage sites, generate peptides that are poorly presented by MHC molecules, or evade effective immune responses (Korber et al. 2001; Moore et al. 2002; Yusim et al. 2002).

The relative roles of heterozygote advantage and frequency-dependent selection in the evolution of MHC polymorphism have been debated for decades (Lewontin et al. 1978; Aoki 1980; Hughes and Nei 1988, 1989; Parham et al. 1989b; Takahata and Nei 1990; Lawlor et al. 1990; Wills 1991). In our model simulations, we found that host-pathogen coevolution readily accounts for a high degree of MHC polymorphism, whereas heterozygote advantage per se gives rise to only low degrees of polymorphism. Apparently, the frequency-dependent selection, which is added to heterozygote selection when hosts and pathogens coevolve, considerably enlarges the MHC polymorphism. Our simulations also support our finding that heterozygote advantage can only account for a high degree of polymorphism if all MHC alleles confer unrealistically similar fitness contributions to their hosts (see our companion paper: De Boer et al. 2004, DOI 10.1007/s00251-003-0629-y). The mechanism of host-pathogen coevolution has been criticized because it would give rise to a too dynamic polymorphism, in which MHC allele frequencies fluctuate over time (Hughes and Nei 1992; Slade and McCallum 1992), which seems at

variance with the long-term persistence of MHC alleles (Klein 1980; Lawlor et al. 1988; Mayer et al. 1988; Klein and Klein 1991). Here, we confirm this dynamic behavior, but show that a high degree of polymorphism of long-lived MHC alleles can nevertheless be obtained. If the host population is sufficiently large, MHC alleles need not go extinct, because pathogens rapidly adapt to common MHC alleles and no longer adapt to rare MHC alleles.

## Materials and methods

We have developed a model resembling a genetic algorithm (Holland 1975) to investigate the coevolution of pathogens and hosts expressing MHC molecules (see also Borghans and De Boer 2001; Beltman et al. 2002). In the simulations presented here, we consider a population of  $N_{\text{host}}$  diploid hosts, each represented by a series of bit strings (i.e., sequences of zeros and ones, used as a simple way to model amino acid sequences) coding for two alleles at  $N_{\text{loci}}$  MHC loci. Since MHC loci are closely linked we chose to let them be inherited as haplotypes. Pathogens are haploid and occur in  $N_S$  independent species of maximally  $N_G$  different genotypes. This subdivision of pathogens into species avoids too strong a reduction in pathogen diversity under fierce selection. For simplicity, we omit the complex process of protein degradation into peptides, and model each pathogen by  $N_{\text{pep}}$  bit strings representing its peptides. In reality, large differences exist in binding efficiency of different peptides with different MHC molecules. This causes some peptides to be immunodominantly presented, while others remain subdominant and typically do not evoke a protective immune response. Here we only consider the limited set of pathogen peptides that can possibly lead to protective immunity. Peptides and MHC molecules are  $L=16$  bits long. Peptide presentation by an MHC molecule is modeled by complementary matching. If the longest stretch of adjacent complementary bits is at least  $L_T=7$  bits long, the peptide is considered to be presented by the MHC molecule (see Fig. 1). With these parameters the chance that a random MHC molecule presents a randomly chosen peptide is about 5%, which seems a realistic value (Kast et al. 1994). Hosts with different MHC molecules thus tend to present different peptides of the same pathogen.

To account for the short generation time of pathogens, we allow for ten pathogen generations per host generation. At each pathogen generation, every host interacts with one randomly chosen member of each pathogen species. The fitness of a host is defined as the fraction of pathogens it presents during one host generation. The fitness of a pathogen is the fraction of hosts that the pathogen can infect during one pathogen generation without being immunodominantly presented by the MHC molecules of the host. At the end of each generation, all individuals are replaced by fitness-proportional reproduction. In the simulations presented here, the chance that an individual reproduces is proportional to its squared fitness divided by the sum of the squared fitnesses in the pathogen species or host population. Pathogens reproduce asexually; newborn pathogens come from parents of the same pathogen species. Newborn hosts have two parents, each of which donates a randomly selected MHC haplotype. The size of the host population and the number of pathogens remain constant.

```
MHC:           1010010101011100
peptide:       0110101010100110
match:         **  ***** *
```

**Fig. 1** An example of bitstring matching. The longest complementary match between the MHC molecule and the peptide is nine bits long. The peptide is hence presented by the MHC molecule

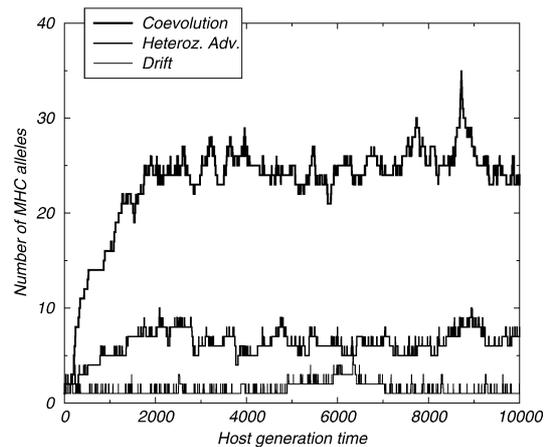
During reproduction mutations can occur. It is thought that a single amino-acid replacement in a peptide or in an MHC molecule can dramatically change its binding characteristics, and that most mutations in the human MHC region are gene conversions rather than point mutations (Parham and Ohta 1996). To model such drastic effects of mutations, we implemented mutation events by generating new random bit strings. Although this is not a very realistic choice, it helps to speed up the evolutionary selection in our simulations, and does not affect our results; we have done extensive simulations using different mutation schedules instead and all results remained similar (not shown). MHC alleles mutate at a frequency  $\mu_{\text{host}}=10^{-5}$  per allele per host generation. Different pathogen species have different mutation rates: one pathogen species has a mutation frequency of  $\mu_{\text{path}}=10^{-1}$  per peptide per pathogen generation, and another pathogen species  $\mu_{\text{path}}=\mu_{\text{host}}$ . The mutation frequencies of all other pathogen species are scaled logarithmically between these two extremes. Thus, most pathogen species mutate much faster than the MHC molecules of the hosts. One cycle of fitness determination, reproduction, and mutation defines a generation. We have studied evolution over many generations. The program was coded in the C-programming language and is available as Electronic Supplementary Material in the online version of this article at <http://dx.doi.org/10.1007/s00251-003-0630-5> Simulations with thousands of hosts took several days of CPU-time on Pentium 800 Mhz PCs running Linux.

Importantly, we have studied many variations of the simulation model presented here, and found the results to be very insensitive to the specific model assumptions and parameter choices made. For example, using a different reproduction function, a different mutation schedule (e.g., point mutations or MHC recombination), or different rules for MHC-peptide binding, did not qualitatively affect the results. To keep the computations feasible, a number of parameters were given low values. Extensive tests with more realistic parameter values — e.g. larger population sizes, more pathogen generations within a host generation, more peptides per pathogen, more pathogen species, a larger potential diversity of peptides, different MHC lengths and multiple MHC loci per host — did not lead to any qualitatively different results. Since all these changes did not affect our results, we chose to run the main simulations using a simple, transparent — albeit somewhat unrealistic — and fast model. The only parameter that played a crucial role in our simulations was the chance that a random MHC molecule presents a random peptide. To make sure that different hosts sample different peptides from pathogens, this parameter should not be much larger than the experimental estimate of 5%.

## Results and discussion

### MHC polymorphism

To study the evolution of MHC polymorphism, all hosts were initialized with the same randomly chosen MHC allele. Pathogens were initialized fully randomly; i.e., each pathogen species started with  $N_G=10$  genotypes each consisting of  $N_{\text{pep}}=20$  randomly generated peptides. In a population of 1,000 diploid hosts coevolving with 50 pathogen species, a stable MHC polymorphism of approximately 27 alleles on a single MHC locus developed in about 2,000 host generations (see the thick solid curve in Fig. 2). At the end of the simulation, approximately 95% of the hosts were MHC heterozygous (see Table 1). The pathogens could escape from presentation by the MHC molecules of 45% of the heterozygous hosts, which is much more than the 16% expected for a random pathogen (compare the heterozygote fitness of Coevolution with Control in Table 1). Apparently, most of the



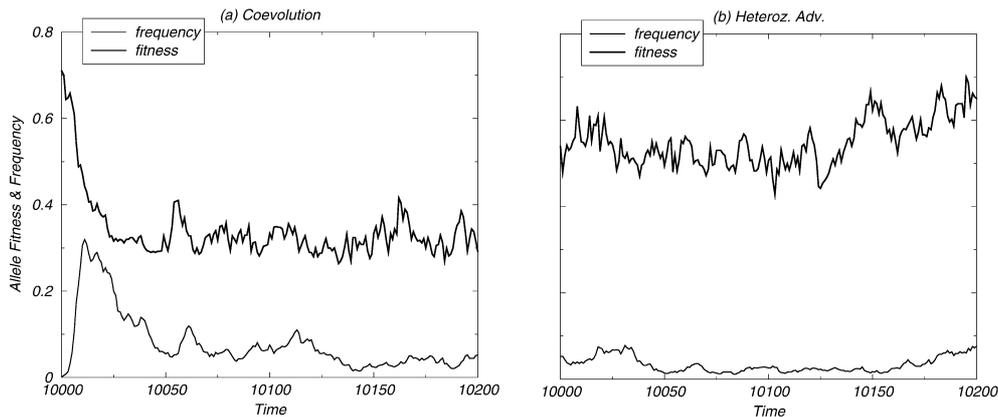
**Fig. 2** The evolution of MHC polymorphism in host populations starting with a single MHC allele at one locus ( $N_{\text{loci}}=1$ ). The *thick solid curve (Coevolution)* depicts a simulation in which hosts and pathogens coevolved (hence including both heterozygote advantage and frequency-dependent selection). The *dashed curve (Heteroz. Adv.)* represents a simulation in which there was only selection for heterozygous hosts: only the hosts were selected, pathogens mutated but were not selected because they were all given the same, fixed fitness. The *dotted curve (Drift)* represents evolution in the absence of selection: both hosts and pathogens mutated, but their selection was avoided by giving all hosts and all pathogens the same fitness. Simulations were run with 1,000 hosts and  $N_S=50$  pathogen species, each with maximally  $N_G=10$  different genotypes. Each pathogen genotype consisted of  $N_{\text{pep}}=20$  peptides, each  $L=16$  bits long. MHC molecules were also  $L=16$  bits long, and were considered to present a peptide if the longest adjacent complementary match between a peptide and the MHC was at least  $L_T=7$  bits long (which occurs with approximately 5% chance). MHC alleles mutated at a rate  $\mu_{\text{host}}=10^{-5}$  per allele per host generation. Pathogen peptides mutated at rates varying between  $\mu_{\text{path}}=10^{-1}$  and  $\mu_{\text{path}}=10^{-5}$  per peptide per pathogen generation. There were ten pathogen generations per host generation

pathogens were well adapted by being poorly presented by the 27 different MHC molecules present in the host population. [The 27 MHC molecules that evolved when hosts and pathogens coevolved seemed to be randomly distributed over the bit string shape space: when the evolved host population of Fig. 2 was exposed to random pathogens, the average fitnesses of homozygous and heterozygous hosts were very similar to those in a control host population with two random MHC alleles (see Control in Table 1). We found similar results for the peptides of the pathogens: when the evolved pathogen population was exposed to a host population with random MHC molecules, the average fitness of the pathogens was similar to that of a random pathogen population (not shown).] The parameter setting strongly favored MHC heterozygous hosts (by a factor 1.4): a pathogen with 20 peptides had an 84% chance of becoming presented in a host expressing two different MHC molecules, and only 59% in a host expressing a single MHC molecule (see Control in Table 1). In the coevolutionary simulations, there was both heterozygote selection and frequency-dependent selection. To investigate whether the evolved polymorphism mainly resulted from frequency-dependent

**Table 1** Characteristics of three different simulations. The columns give the number of MHC alleles that evolved, the percentage of heterozygous hosts in the population, the average fitness of homozygous and heterozygous hosts (with standard deviations), and the ratio of the average fitness of the heterozygotes over the average fitness of the homozygotes. *Coevolution* corresponds to

	No. of MHC alleles	Percentage of heterozygotes	Fitness of homozygotes	Fitness of heterozygotes	Ratio
Coevolution	27	95%	0.33 ( $\pm 0.05$ )	0.55 ( $\pm 0.05$ )	1.7
Heteroz. adv.	7	83%	0.61 ( $\pm 0.05$ )	0.84 ( $\pm 0.03$ )	1.4
Control	2	50%	0.59 ( $\pm 0.07$ )	0.84 ( $\pm 0.05$ )	1.4

simulations in which pathogens coevolved with diploid hosts (see Fig. 2). *Heteroz. adv.* corresponds to simulations with heterozygote advantage only (see Fig. 2). *Control* represents an initial condition of a host population with two random MHC alleles and 500 fully random pathogens. Parameter values used are given in the legend of Fig. 2



**Fig. 3a, b** Fluctuations in the MHC allele frequencies and their absolute fitnesses. The absolute fitness of the MHC alleles represents the fraction of pathogens in the population that an MHC molecule could present. Panels **a** and **b** show the oscillations

of the rarest MHC allele at the end of the simulations in Fig. 2 under host-pathogen coevolution and under heterozygote advantage, respectively

selection or from heterozygote advantage, simulations were run in which only heterozygote advantage could play a role. To this end, pathogen selection was avoided by giving all pathogens the same, fixed fitness. Hence, pathogens were allowed to mutate, but not to be selected. Since this yielded a polymorphism of only seven MHC alleles (see the dashed curve in Fig. 2), frequency-dependent selection by host-pathogen coevolution must have been the main factor contributing to the MHC polymorphism in our coevolutionary simulations. Yet, even the low degree of polymorphism arising under heterozygote advantage was truly due to selection for MHC diversity, because when host and pathogen phenotypes only changed by genetic drift an even lower degree of polymorphism was attained (see the thin curve in Fig. 2).

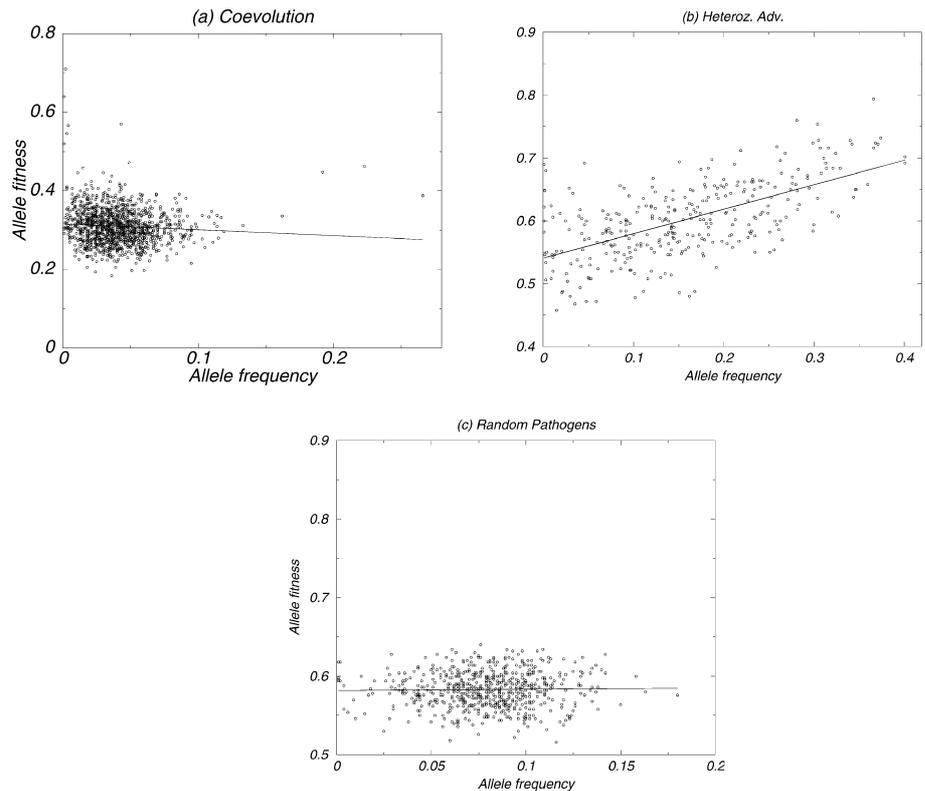
In the simulations presented in Fig. 2, we considered only one MHC locus per individual. In reality, however, a large degree of polymorphism is found at multiple MHC loci. We tested how much MHC diversity would arise at multiple loci, and found that a high MHC polymorphism at up to six loci was easily obtained when hosts and pathogens coevolved. Under heterozygote advantage, in contrast, a realistically high MHC polymorphism was never obtained, because the fitness of both homozygous and heterozygous hosts rapidly approached one when the number of loci was increased (results not shown).

### Dynamics of allele frequencies

The fact that many MHC alleles have persisted for significant evolutionary periods of time, predating the divergence of present-day species (Klein 1980; Lawlor et al. 1988; Mayer et al. 1988; Klein and Klein 1991), has been used as an argument against frequency-dependent selection (Hughes and Nei 1988). Data from South Amerindians revealed, however, that 23 out of 28 MHC alleles found at the *HLA-B* locus were created de novo by recombination (Parham and Ohta 1996). Because this population has existed for only a thousand generations, this suggests rapid evolutionary dynamics of MHC alleles in this population. A possible factor in the rapid allele turnover could be the strong selection imposed by the introduction of novel diseases by the settlement of Europeans (Black 1992). Interestingly, in our simulations, large fluctuations of allele frequencies were found over time both in the presence and in the absence of host-pathogen coevolution (Fig. 3). The largest fluctuations were found in coevolutionary runs, because each novel rare allele had a much higher fitness than the established alleles (i.e., 60% vs 33% in Table 1). When a so much fitter new allele spreads through the population, other alleles may go extinct (Slade and McCallum 1992).

In the presence of host-pathogen coevolution, most resident MHC molecules had very similar absolute

**Fig. 4a–c** The fitness of MHC alleles as a function of their frequency in the population. The absolute fitness of the MHC alleles, defined as the fraction of pathogens in the population that an MHC molecule could present, is given in *black*. The marginal fitness of the MHC alleles, defined as the average fitness of all hosts in which an MHC allele occurred, is given in *gray*. Both fitness measures were plotted for each MHC allele in the host population, for every 100 host generations between generation 5,000 and 10,000 of Fig. 2. The lines show the regression lines through the absolute and marginal fitness values. Panel **a** represents simulations in which both hosts and pathogens evolved, in **b** there was only heterozygote advantage, while in **c** pathogens were introduced randomly at each host generation



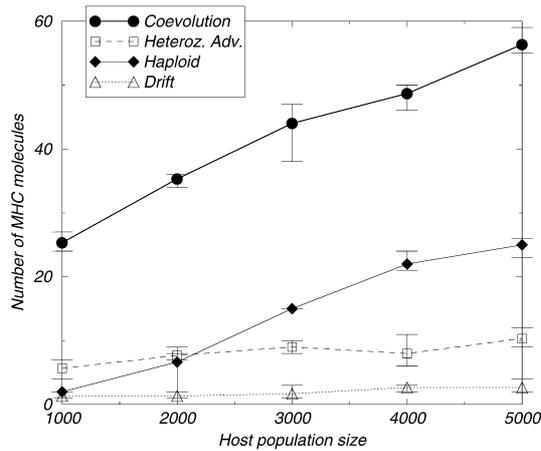
fitnesses; i.e., most alleles presented about the same number of pathogens in the population, and hardly any MHC allele was present at high frequency (Fig. 4a). The absolute fitness of the MHC alleles fluctuated by pathogen adaptation. Novel alleles tended to have the highest fitness (i.e., 0.6, see Table 1), and their frequency rapidly increased. Once they became frequent the pathogens adapted to them, which decreased their fitness contribution and their frequency (Fig. 3a). As a consequence, such MHC alleles were released from pathogen adaptation, which allowed their fitness values to increase. The resulting dynamical process of selection against common MHC alleles caused the persistence of a large variety of MHC alleles.

In contrast, under heterozygote advantage the absolute fitness contributions of the MHC alleles differed considerably (Fig. 4b). Alleles giving a high fitness contribution were present at higher frequencies than MHC alleles with low absolute fitnesses. These fitness and frequency differences between MHC alleles were due to successful adaptation of some MHC molecules to the pathogen population. When the hosts were exposed to a new, random pathogen population at each host generation, these fitness and frequency differences between the MHC alleles were much smaller (Fig. 4c). The large fitness differences in simulations of heterozygote advantage hampered the evolution of a large degree of polymorphism: as we have shown in our companion paper, coexistence of many MHC alleles requires sufficiently small fitness differences between the alleles (De Boer et al. 2004, DOI 10.1007/s00251-003-0629-y). Once the

system had approached equilibrium, all MHC alleles should have the same marginal fitness, however (Nagylaki 1992). That is, the average of the fitnesses of all hosts in which a particular MHC allele occurred should be similar for all MHC alleles. Indeed, alleles with low and high fitness contributions obtained similar marginal fitnesses: alleles with low absolute fitnesses were present at low frequencies (Fig. 4b) and hence occurred more often in heterozygous hosts, which on average had a 1.4-fold higher fitness than homozygous hosts (see Table 1) (Apanius et al. 1997).

#### Longevity of MHC alleles

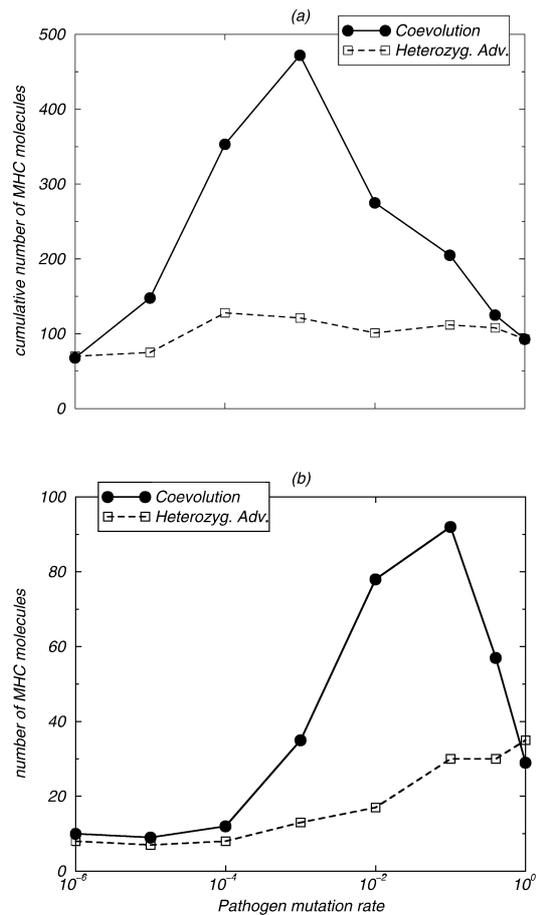
In our model, two parameters determine the longevity of MHC alleles. The first one is the host population size, affecting the chance that rare alleles are driven to extinction. Increasing the host population size from 1,000 to 5,000 hosts more or less linearly increased the degree of MHC polymorphism when hosts and pathogens coevolved (see Fig. 5). Since a host population of 5,000 individuals coevolving with pathogens developed a polymorphism of more than 50 alleles, host-pathogen coevolution should be able to account for the naturally observed polymorphism of more than 100 alleles per locus in natural populations (Parham and Ohta 1996). Conversely, under heterozygote advantage, the degree of polymorphism remained small when the population size was increased. Apparently, and in line with what we show in our companion paper (De Boer et al. 2004, DOI



**Fig. 5** When hosts and pathogens coevolved (see *Coevolution* and *Haploid*), the MHC polymorphism arising after 10,000 host generations increased with the host population size. In the absence of host-pathogen coevolution (see *Heteroz. Adv.* and *Drift*) the degree of polymorphism was not dependent on the host population size. The data points give the average of three runs differing in random seed; the *error bars* denote the range of the three runs. Parameters and legend as in Fig. 2

10.1007/s00251-003-0629-y), the number of MHC alleles evolving under heterozygote advantage depends largely on the fitness distribution of the alleles, and much less on the host population size. Simulations with haploid hosts showed that even in the absence of heterozygote advantage, host-pathogen coevolution could account for a high degree of polymorphism, increasing with the host population size (see the diamonds in Fig. 5).

The second parameter affecting the longevity of MHC alleles is the mutation rate of the pathogens. In the simulations presented so far, different pathogen species had different mutation rates. To investigate the consequences of pathogen mutation rates, we have also run simulations in which every pathogen species has the same mutation rate. First, as a measure of the MHC diversity developing over time, we followed the total number of MHC molecules that were present during the last 8,000 generations of the simulations (see Fig. 6a). At low pathogen mutation rates, e.g.,  $\mu_{\text{path}}=10^{-6}$ , the pathogens evolved so slowly that the difference between host-pathogen coevolution and heterozygote advantage vanished. The hosts evolved a set of MHC molecules specialized at presenting the peptides present in the pathogen population. The fitness of the hosts was therefore somewhat higher than when they were exposed to random pathogens (not shown). Intermediate mutation rates (i.e., around  $\mu_{\text{path}}=10^{-3}$ ) allowed coevolving pathogens to evade presentation by the MHC molecules of the hosts, causing a selection pressure for more MHC diversity. Higher pathogen mutation rates hampered the adaptation of coevolving pathogens to the MHC molecules in the host population, because pathogens with well-adapted peptides were strongly affected by a single mutation: any novel mutant peptide provides an “Achilles heel” by which a pathogen can be presented to the



**Fig. 6a, b** MHC polymorphism as a function of the pathogen mutation rate. The curves marked with *circles* represent host-pathogen coevolutionary simulations, the curves marked with *squares* give simulations with heterozygote advantage. **a** The total number of different MHC molecules that evolved between host generation 2,000 and 10,000. **b** The degree of MHC polymorphism at generation 10,000. Parameters as in Fig. 2 but with 5,000 hosts

immune system by 5% of the MHC alleles in the host population. When the pathogens had a mutation rate of 100%, coevolution and heterozygote advantage became identical because the pathogen population was completely random. Even in the coevolutionary runs, the evolution of MHC diversity could hence only be due to heterozygote advantage.

The degree of MHC polymorphism at any point in time showed a roughly similar picture as the total number of MHC molecules developing over time (see Fig. 6b). The peaks of the two curves, however, did not coincide. Around  $\mu_{\text{path}}=10^{-3}$ , when a maximum of almost 500 MHC molecules developed in 8,000 generations, the actual number of MHC alleles present at any point in time was relatively low. At these intermediate mutation rates, the pathogens rapidly evolved peptides that were poorly presented by the MHC molecules in the host population. Novel MHC alleles therefore had a much higher fitness than common MHC alleles, and tended to replace the established MHC alleles. As this replacement effect

decreased at higher pathogen mutation rates, the largest degree of MHC polymorphism — i.e., almost 100 MHC alleles at a single locus — was found around  $\mu_{\text{path}}=10^{-1}$ , when only 200 MHC alleles developed in 8,000 generations. The MHC polymorphism developing under heterozygote advantage was largest at a pathogen mutation rate of 100%, because the fitness contributions of the MHC alleles were most similar when the pathogens were fully random (De Boer et al. 2004, DOI 10.1007/s00251-003-0629-y). The largest difference between host-pathogen coevolution and heterozygote advantage was found at intermediate mutation rates of the pathogens (see Fig. 6) and when pathogens had different mutation rates (see Fig. 5). In the latter case, the fitness differences between the MHC alleles were dampened by the pathogens mutating too fast or too slow for efficient adaptation to the MHC molecules present in the host population.

## Conclusions

Our simulations show that the degree of MHC polymorphism developing under host-pathogen coevolution is much larger than under mere heterozygote advantage, and increases with the host population size. Because of the fitness differences between MHC alleles, heterozygote advantage can only account for a modest degree of MHC polymorphism even in large host populations (see also De Boer et al. 2004, DOI 10.1007/s00251-003-0629-y). The presence of coevolving pathogens provides an additive frequency-dependent selection mechanism for MHC diversity, because pathogens mainly adapt to the most common (i.e., up to that point in time the fittest) MHC alleles in the population (Trachtenberg et al. 2003). Thereby, host-pathogen coevolution can account for realistic degrees of MHC polymorphism, while mere heterozygote advantage cannot. Both frequency-dependent selection by host-pathogen coevolution and heterozygote advantage give rise to a dynamic equilibrium with fluctuating allele/haplotype frequencies. Despite these oscillations, many MHC alleles can survive in large host populations under host-pathogen coevolution.

Many other factors, such as geographic or social isolation and strong selective pressures by severe infections, may influence the degree of MHC polymorphism arising in a population (Hill et al. 1991; Hedrick 2002). We fully acknowledge the importance of such factors, but have nevertheless omitted them in our simulations to be able to compare the two most frequently used explanations for the large MHC polymorphism. Only by omitting all other factors could we conclude that heterozygote advantage per se is insufficient to explain why the MHC loci evolved such a large polymorphism, and that other factors such as geographic heterogeneity or host-pathogen coevolution are required. The current simulations show that host-pathogen coevolution suffices to explain the large degree of MHC polymorphism.

**Acknowledgements** We thank Wayne Potts, Dustin Penn, Can Keşmir, Paulien Hogeweg, and Ludo Pagie for useful discussions and comments on earlier versions of this manuscript. J.A.M.B. acknowledges financial support by the EC (Marie Curie Fellowship, Quality of Life, contract 1999-01548).

## References

- Aoki K (1980) A criterion for the establishment of a stable polymorphism of higher order with an application to the evolution of polymorphism. *J Math Biol* 9:133–146
- Apanius V, Penn D, Slev PR, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. *Crit Rev Immunol* 17:179–224
- Barouch D, Friede T, Stevanovic S, Tussey L, Smith K, Rowland-Jones S, Braud V, McMichael A, Rammensee HG (1995) HLA-A2 subtypes are functionally distinct in peptide binding and presentation. *J Exp Med* 182:1847–1856
- Beck K (1984) Coevolution: mathematical analysis of host-parasite interactions. *J Math Biol* 19:63–77
- Beltman JB, Borghans JAM, De Boer RJ (2002) Major histocompatibility complex: polymorphism from coevolution. In: Dieckmann U, Metz JAJ, Sabelis MW, Sigmund K (eds) *Adaptive dynamics of infectious diseases. In pursuit of virulence management*. Cambridge University Press, Cambridge, pp 210–221
- Black FL (1992) Why did they die? *Science* 258:1739–1740
- Bodmer WF (1972) Evolutionary significance of the HL-A system. *Nature* 237:139–145
- Borghans JAM, De Boer RJ (2001) Diversity in the immune system. In: Segel LA, Cohen IR (eds) *Design principles for the immune system and other distributed autonomous systems*. Oxford University Press, Oxford, pp 161–183
- 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, Borghans JAM, Van Boven M, Keşmir C, Weissing FJ (2004) Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. *Immunogenetics*. DOI 10.1007/s00251-003-0629-y
- Doherty PC, Zinkernagel RM (1975) Enhanced immunological surveillance in mice heterozygous at the *H-2* gene complex. *Nature* 256:50–52
- Dybdahl MF, Lively CM (1998) Host-parasite coevolution: evidence for rare advantage and time-lagged selection in a natural population. *Evolution* 52:1057–1066
- Hamilton WD, Axelrod R, Tanese R (1990) Sexual reproduction as an adaptation to resist parasites (a review). *Proc Natl Acad Sci USA* 87:3566–3573
- Hedrick PW (2002) Pathogen resistance and genetic variation at MHC loci. *Evolution* 56:1902–1908
- Hill AV, Allsopp CE, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, McMichael AJ, Greenwood BM (1991) Common west African HLA antigens are associated with protection from severe malaria. *Nature* 352:595–600
- Holland JH (1975) *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor
- 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. *Proc Natl Acad Sci USA* 86:958–962
- Hughes AL, Nei M (1992) Models of host-parasite interaction and MHC polymorphism. *Genetics* 132:863–864
- Jeffery KJ, Siddiqui AA, Bunce M, Lloyd AL, Vine AM, Witkover AD, Izumo S, Usuku K, Welsh KI, Osame M, Bangham CR (2000) The influence of HLA class I alleles and heterozygosity

- on the outcome of human T cell lymphotropic virus type I infection. *J Immunol* 165:7278–7284
- Kast WM, Brandt RM, Sidney J, Drijfhout JW, Kubo RT, Grey HM, Melief CJ, Sette A (1994) Role of HLA-A motifs in identification of potential CTL epitopes in human papillomavirus type 16 E6 and E7 proteins. *J Immunol* 152:3904–3912
- Klein J (1980) Generation of diversity to MHC loci: implications for T cell receptor repertoires. In: Fougereau M, Dausset J (eds) *Immunology* 80. Academic Press, London
- Klein J, Klein D (1991) *Molecular evolution of the MHC complex*. Springer, Berlin Heidelberg New York
- Korber B, Gaschen B, Yusim K, Thakallapally R, Keşmir C, Detours V (2001) Evolutionary and immunological implications of contemporary HIV-1 variation. *Br Med Bull* 58:19–42
- 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
- Lawlor DA, Zemmour J, Ennis PD, Parham P (1990) Evolution of class-I MHC genes and proteins: From natural selection to thymic selection. *Annu Rev Immunol* 8:23–63
- Lewontin RC, Ginzburg LR, Tuljapurkar SD (1978) Heterosis as an explanation for large amounts of genic polymorphism. *Genetics* 88:149–170
- Lively CM, Dybdahl MF (2000) Parasite adaptation to locally common host genotypes. *Nature* 405:679–681
- Mayer WE, Jonker M, Klein D, Ivanyi P, Van Severter G, Klein J (1988) Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. *EMBO J* 7:2765–2774
- Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA (2002) Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* 296:1439–1443
- Nagylaki T (1992) *Introduction to Theoretical Population Genetics*. Springer, Berlin Heidelberg New York
- 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 Symp Quant Biol* 54:529–543
- Parham P, Lawlor DA, Lomen CE, Ennis PD (1989b) Diversity and diversification of HLA-A, B, C alleles. *J Immunol* 142:3937–3950
- Penn DJ, Damjanovich K, Potts WK (2002) MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proc Natl Acad Sci USA* 99:11260–11264
- Penn DJ, Potts WK (1999) The evolution of mating preferences and major histocompatibility complex genes. *Am Nat* 153:145–164
- Slade RW, McCallum HI (1992) Overdominant versus frequency-dependent selection at MHC loci. *Genetics* 132:861–864
- Snell GD (1968) The *H-2* locus of the mouse: observations and speculations concerning its comparative genetics and its polymorphism. *Folia Biol Praha* 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
- Trachtenberg E, Korber B, Sollars C, Kepler TB, Hraber PT, Hayes E, Funkhouser R, Fugate M, Theiler J, Hsu YS, Kunstman K, Wu S, Phair J, Erlich H, Wolinsky S (2003) Advantage of rare HLA supertype in HIV disease progression. *Nat Med* 9:928–935
- 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. *Immunol Rev* 167:327–337
- Weidt G, Deppert W, Buchhop S, Dralle H, Lehmann-Grube F (1995) Antiviral protective immunity induced by major histocompatibility complex class I molecule-restricted viral T-lymphocyte epitopes inserted in various positions in immunologically self and nonself proteins. *J Virol* 69:2654–2658
- Wills C (1991) Maintenance of multiallelic polymorphism at the MHC region. *Immunol Rev* 124:165–220
- Yusim K, Keşmir C, Gaschen B, Addo MM, Altfeld M, Brunak S, Chigaev A, Detours V, Korber BT (2002) Clustering patterns of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. *J Virol* 76:8757–8768