Epidemics xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Quantifying how MHC polymorphism prevents pathogens from adapting to the antigen presentation pathway

B.V. Schmid [a,b,*], C. Kęsmir [c,d], R.J. de Boer [d]

[a] RIVM, Bilthoven, The Netherlands
[b] Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands
[c] Academic Biomedical Centre, Utrecht, The Netherlands
[d] Faculty of Biology, Utrecht University, Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 7 September 2009
Revised 14 April 2010
Accepted 28 May 2010
Available online xxxx

Keywords:
MHC class I
Proteasome
CTL epitopes
HIV-1
Adaptation
Population level

ABSTRACT

The classical antigen presentation pathway consists of two monomorphic (proteasome and TAP) and one polymorphic components (MHC Class I). Viruses can escape CTL responses by mutating an epitope so that it is no longer correctly processed by the pathway. Whereas escape mutations that affect MHC binding are typically no longer under selection pressure in the next host of the virus (as hosts differ in their MHC alleles), escape mutations that affect the antigen processing of epitope precursors prevent the use of those epitope precursors by any of the MHC alleles in a host population. Viruses might therefore be under selection pressure to adapt to the monomorphic proteasome and TAP.

We designed an agent-based model of a host population, in which an HIV-1 like virus adapts to the antigen presentation pathway of individual hosts, as the virus spreads through the population. We studied how the polymorphism of the MHC and the monomorphism of the proteasome and TAP affected the level of adaptation to the host population that the virus could reach.

We found that due to the polymorphism and high specificity of the MHC class I molecules, the CTL epitopes that are targeted by the CTL responses of different hosts do not share many epitope precursors. Therefore, escape mutations in epitope precursors are frequently released from immune selection pressure, and can revert back to the virus wildtype sequence. As a result, the selection pressure on the virus to adapt to the proteasome and TAP is relatively small, which explains the low level of adaptation of the virus to the monomorphic steps in the antigen presentation pathway.

Introduction

The MHC-I antigen presentation pathway provides the immune system with a way to detect intracellular pathogens, by displaying intracellular protein fragments on the cell surface. One of the most remarkable features of this pathway is the high degree of polymorphism in one of its components: with over 2300 alleles known (Robinson et al., 2006; Sayers et al., 2009), the major histocompatibility (MHC) class I molecules are the most polymorphic genes in the human genome. This polymorphism is thought to have developed in response to the selection pressure exerted by pathogens in at least two ways: by means of the heterozygote advantage (Doherty and Zinkernagel, 1975; Carrington et al., 1999), which is the ability of heterozygote hosts to present a wider range of epitopes, and the rare allele advantage (Slade and McCallum, 1992; Langefors et al., 2001; Trachtenberg et al., 2003; Borghans et al., 2004), which states that pathogens encounter rare MHC alleles less frequently, and will therefore carry fewer escape mutations with them that affect these MHC alleles.

The antigen presentation pathway involves two other major molecules besides the polymorphic MHC alleles: the proteasome that cleaves proteins into small fragments, and the transporter associated with antigen processing (TAP) that transports peptide fragments into the endoplasmic reticulum, where these peptides (epitope precursors) bind to the MHC, and thus become Cytotoxic T Lymphocyte (CTL) epitopes. Surprisingly, it is only the last step of the pathway that has developed a large degree of polymorphism, even though viruses can escape CTL responses against epitopes by generating escape mutations for any of the steps in the antigen presentation pathway (Yokomaku et al., 2004; Kwun et al., 2007). It would seem that on the population level there is a fitness advantage for viruses to escape CTL epitopes by escaping the monomorphic proteasome and TAP, rather than the polymorphic MHC.

In Schmid et al. (2008) we studied 25 years of available HIV-1 sequence data, and found that HIV-1 did not appear to accumulate epitope precursor escape mutations. We postulated a mechanism why HIV-1 would not be able to adapt massively to the monomorphic components of the antigen presentation pathway, based on the

* Corresponding author. Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands.
E-mail address: boris.schmid@gmail.com (B.V. Schmid).

1755-4365/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.epidem.2010.05.003

specificity of the MHC molecules within a host, and the polymorphism of the MHC in the population. We argued that these two characteristics of the MHC would result in an intermittent exposure of epitope precursors to immune selection pressure, as HIV-1 is transmitted from one host to another. Without a constant selection pressure to maintain epitope precursor escapes, such mutations could frequently revert back into the wildtype sequence.

In this paper we quantify the ability of an HIV-1 like virus to adapt its genome to escape epitope precursor processing and MHC Class I binding in an MHC class I polymorphic host population, using an agent-based simulation model. In all simulations, the virus reaches a quasi-steady state in which the accumulation of new escape mutations is balanced by the reversion of escape mutations that are not under immune selection pressure in the current host of the virus. For HIV-1 like parameters, this quasi-steady state is reached within a few decades after the start of the epidemic, at which the virus carries pre-adaptations (i.e. the fraction of CTL epitopes for which the virus carries escape mutations at the time of infection of a new host) for approximately 35% of the dominant CTL epitopes that a new host could have responded against. 86% of these immune escape mutations are MHC-binding mutations, and 29% are epitope precursor mutations (these percentages do not sum up to 100% because a single epitope can carry an escape mutation that affects both MHC binding and epitope precursor processing at the same time).

Concluding, the presence of an MHC polymorphism in combination with the high specificity of the MHC class I molecules reduces the shared usage of epitope precursors between different hosts to a very low level. As a result, the selection pressure on the virus to adapt to the monomorphic proteasome and TAP is smaller than expected, and does not result in a massive amount of adaptation to these monomorphic molecules.

Materials and methods

Agent-based model: actors and events

The agent-based model consists of two types of actors (*hosts* and *viruses*), and five types of events (*procreation*, *death* and *new relationship* for the hosts, and *escape* and *reversion* for the virus). The time-step of the model is one month. The events that will take place in each month are determined by the predefined frequency of those events per year, and are subsequently applied in a random order to all the hosts in the population. For the default settings, each host participates (on average) in 0.25 procreation events, 1 death event, and 0.5 new relationship events per year, and each virus in 4 escape, and 2 reversion events per year. The following is a detailed description of the events and actors of the model, as well as a motivation for the used parameter values.

- Host procreation: the selected host passes on its proteasome, TAP and half of its MHC alleles (1 human leukocyte antigen A (HLA-A) allele, and 1 HLA-B allele) to its child. The other half of the MHC alleles are drawn from a constant pool of MHC alleles in order to keep the polymorphism and relative frequency of MHC alleles in the host population close to that of the initial distribution. The chance of successful childbirth decreases linearly with the population size, i.e. we have logistic growth. For simplicity, newborn children are given the age of 15, and are immediately added to the host population.

In the simulations with a realistic MHC polymorphism, the pool of MHC alleles is modeled in polymorphism and frequency after the 2-digit MHC Class I diversity in the European population (Fig. 1A), as found in the dbMHC-Anthropology database (Sayers et al., 2009). In the simulations in which we vary the MHC polymorphism, the MHC alleles have equal frequencies.

- Host death: the host is removed from the population if it fails to pass an age- and disease-dependent death chance D, which is described by the equation

$$D = e^{0.1A-10.5} + e^{-0.4A-8} + e^{0.1YV-5}. \quad (1)$$

The death chance consist of an age in years, A, specific component, which is fitted to the intrinsic death rate of North Americans (Gompertz, 1825; Carnes et al., 2006; Hallén, 2007), and a disease-specific component, which is a fitted function of the time since infection in years, Y, and the log viral load, V (Lavreys et al., 2006). The disease-specific component of Eq. (1) is only applied to infected individuals. The relation between survival and viral load is plotted in Fig. 1C.

- Host new relationship: a short-term relationship starts between two hosts. Transmission of the virus can happen in both directions between the host and a randomly selected partner if one, but not both of them are infected. The chance to transmit the virus, T, during the relationship is described by the equation

$$T = 1-\left(1-10^{-6}e^{1.8V}\right)^{C}, \quad (2)$$

which is fitted to the relationship between the log viral load, V, and the chance of infection during a single sexual contact (Chakraborty et al., 2001) (Fig. 1B, solid line), and the number of sexual contacts during a relationship, C. The default length of a short-term relationship in the model is 2 years, for which C is set to 240 sexual contacts (i.e. 10 per month, Wawer et al., 2005). The probability of transmitting the virus is higher when the relationship overlaps with the acute phase of the disease in one of the hosts (first 3 months of the infection), as during the acute phase the viral load is temporarily increased with 2 logs (Fig. 1B, dotted lines).

In simulations where we changed the number of relationships per year, we took care to adjust the number of sexual contacts per relationship such that the average number of sexual contacts per year per host stayed the same. Increasing the frequency of relationships increases the maximum number of hosts that one host can infect, and decreases the chance of transmission of the virus per relationship (Fig. 1D).

Relationship duration is not explicitly modeled: relationships are treated as point events, and are therefore non-overlapping. The resulting network can be described as a serial-monogamy dynamic sexual contact network. Mother-to-child transfer of the virus is not included in the model.

- Viral escape and reversion: the dominant within-host virus is replaced by a mutant in a selective sweep event. As a formalism for within-host evolution we assume that all single-point escape mutation variants of the pathogen exist in low numbers in the quasi-species, and that these variants compete with each other until one of the variants (which has a fitness advantage over the current dominant virus sequence) becomes the new dominant within-host variant (Asquith et al., 2006; Althaus and de Boer, 2008). At that point this new dominant variant forms the basis of a new quasi-species, and the process repeats itself. Reversions of escape mutations (Davenport et al., 2008) are implemented in the same way, but now the dominant virus sequence is replaced by a fitter mutant virus which has reverted one of the escape mutations (acquired in previous hosts) that was no longer beneficial to maintain.

Viruses are implemented as a strings of 3000 letters, which represent their amino acid sequence, and parts of which (the epitopes) can be recognized by the antigen presentation pathway of the host. The fitness of a virus is expressed as its log viral load V and is described by the equation

$$V = \max\left(0, V_{\max}-0.2E-0.05M\right), \quad (3)$$

Please cite this article as: Schmid, B.V., et al., Quantifying how MHC polymorphism prevents pathogens from adapting to the antigen presentation pathway, Epidemics (2010), doi:10.1016/j.epidem.2010.05.003

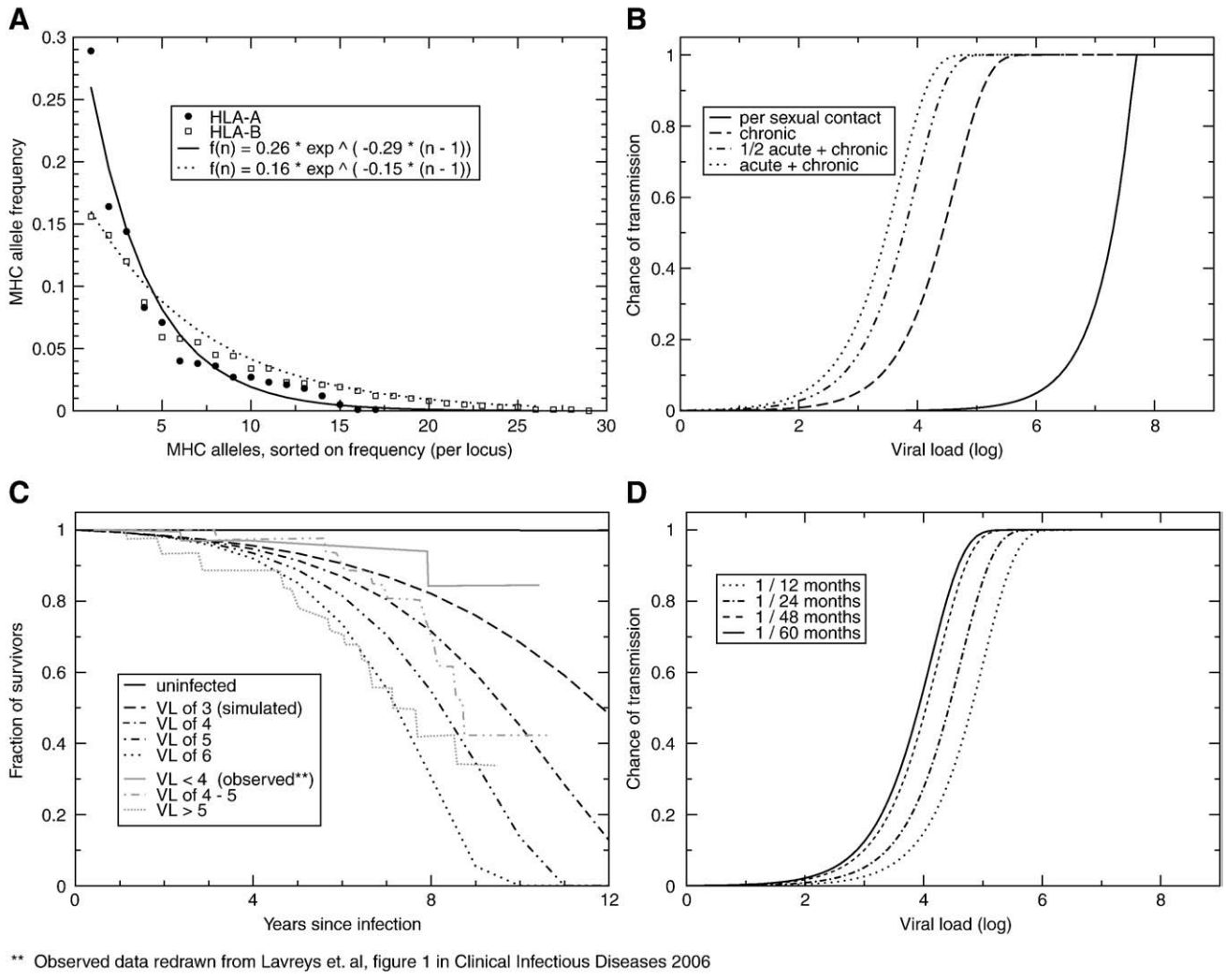

Fig. 1. A: The exponential distribution of MHC allele frequencies for both HLA-A and HLA-B, fitted to the 2-digit allele frequencies in the European population (Sayers et al., 2009). B: The relation between log viral load during the chronic phase, and the transmission chance during a relationship of 2 years (dotted lines) and the transmission chance per sexual contact (solid line). Relationships that start during the acute phase of the disease have a higher transmission chance as during the acute phase the viral load is temporarily increased by 2 logs. C: A survival curve for hosts infected at age 20 with a non-evolving virus, for different viral loads. Black lines are simulated data, and grey lines are observed data, redrawn from Lavreys et al. (2006), Fig. 1. D: The relation between log viral load (during the chronic phase), and the transmission chance during a relationship, for relationships of different length.

where E is the number of recognized CTL epitopes, and M is the number of mutations that the virus carries. For every unique CTL epitope that a host can recognize, the maximum viral load is decreased by 0.2 log (Kiepiela et al., 2007), and for every mutation that the virus carries, the viral load is decreased by 0.05 log. A virus has a minimum log viral load of 0, and a maximum log viral load V_{max} of 7 (i.e. 10^7 copies per ml blood). The viral load is increased by 2 logs during the acute phase of the disease (Piatak et al., 1993; Costin, 2007), which lasts for three months. The exact values for the costs and benefits of escape mutations only have a limited impact on the model outcome, while the escape and reversion rates of the virus have a much larger impact (see Sensitivity analysis, Results section).

By escaping the presentation of epitopes and reverting obsolete escape mutations, pathogens can increase their infectiousness during within-host evolution (Eq. (2)), but also impair their reproductive potential by increasing the death rate of their host (Eq. (1)). Limiting the within-host viral quasi-species to a single dominant virus sequence makes it possible to model within-host evolution of the virus in a computationally non-intensive way, and fits with recent observations that in 80% of the newly infected HIV-1 patients, a single viral sequence is the founder strain (Salazar-Gonzalez et al., 2009).

Antigen presentation pathway

The classical MHC-1 antigen presentation pathway can be described as three filters (proteasome, TAP and MHC class I) that are applied to intracellular proteins. The pathway tests which peptides in a protein can successfully pass through all three filters, and thus be presented as CTL epitopes on the cell surface (Tenzer et al., 2005; Groothuis et al., 2005). Although current algorithms can accurately model this pathway for a large number of MHC alleles (Larsen et al., 2005; Tenzer et al., 2005), we have opted for a simpler and computationally faster approach, and represent the proteasome, TAP and MHC alleles by regular expressions with different specificities. Regular expressions are commonly used to search for complicated text patterns, and can efficiently locate certain letter combinations in a string of text.

The specificity of the regular expressions is made to match those of proteasome, TAP and MHC molecules (Burroughs et al., 2004; Tenzer et al., 2005; Assarsson et al., 2007; Burgevin et al., 2008) by limiting the set of letter combinations that can be recognized on each position in the regular expression (Fig. 2). The regular expressions are generated by starting with a fully generic filter, and then randomly

Please cite this article as: Schmid, B.V., et al., Quantifying how MHC polymorphism prevents pathogens from adapting to the antigen presentation pathway, Epidemics (2010), doi:10.1016/j.epidem.2010.05.003

removing amino acids from random positions in the regular expression, until the desired specificity was reached. If the reduction process would result in a too low specificity of the regular expression, amino acids would be randomly added again, etc., until we were within 5% of the specificity that we were looking for (e.g., a specificity of 5% would have an acceptable range of 4.75–5.25%). This process generates regular expressions as those that can be seen in Fig. 2, in which some positions of the regular expression are more specific than the other positions, and thus somewhat resemble the anchor residues of real motifs.

Unlike the actual molecules, the regular expressions used in the model are binary: they either accept a particular amino acid on a particular position, or they do not. Therefore, the binding pattern of the proteasome, TAP and MHC regular expressions in the model cannot be modeled to exactly match the binding patterns of their biological counterparts. The specificity of the MHC molecules is thought to be in the range of 1–5% of all peptides of length 9 (Assarsson et al., 2007). In the model we use a specificity of $0.5\% \pm 0.025$, which combines both the MHC specificity as well as an unknown "dominance factor" that determines which of the presented CTL epitopes in a host will become immunodominant. The dominance factor is scaled in such a way that it reduces the average number of immunodominant CTL epitopes in the wildtype virus in a random host to approximately 15 (Assarsson et al., 2007; Kotturi et al., 2008).

No effort was made to ensure a particular degree of coevolvedness between the many MHC alleles on the one hand, and the proteasome and TAP on the other hand (Kesmir et al., 2003), but we did ensure that the monomorphic proteasome and TAP filters in the model had a combined specificity of $25\% \pm 1.25\%$, e.g. in the model 23.75–26.25% of the peptides of length 9 can be recognized by both the proteasome and TAP filters.

In our in silico antigen presentation pathway, each possible peptide of length 9 in the virus proteome of length 3000 is tested on its ability to pass through the proteasome, TAP and MHC filters (Fig. 2). Virus peptides that can pass through the proteasome, TAP and at least one of the MHC alleles of a host are collected, and the unique set of peptides in this collection is counted as the number of dominant CTL epitopes, E, that the host responds against.

MHC promiscuity (Frahm et al., 2007), other than the promiscuity expected by chance between the randomly generated MHC molecules is not in the model. Neither are MHC supertypes in the model, other than that we estimated the degree of MHC polymorphism that we use in the simulations from 2-digit MHC allele types, which may contain multiple 4-digit MHC alleles in reality.

Implementation

The model is implemented in Clojure 1.1, a modern dialect of lisp which is especially suitable for multithreaded programs (http://clojure.org). The source code is available upon request.

Results

Model

To study the potential of viruses like HIV-1 to adapt to the antigen presentation pathway of a host population, we constructed an agent-based model of a host population infected with a chronic virus. We kept track of the level of adaptation that this virus reached to the monomorphic proteasome and TAP, and to the polymorphic MHC alleles in the host population.

The model itself is simple in design: a host population is created, from which members are randomly selected and subjected to one of four events: procreation, death, within-host adaptation of the virus, and infection of another host during a relationship. This cycle of selecting hosts and applying events to them is then repeated for as long as the simulation runs.

Hosts are defined by their age, their time since infection (if infected), and their antigen presentation pathway. The latter consists of a proteasome molecule, a TAP molecule, and four MHC molecules. Each of these molecules is implemented as a pattern filter which only recognizes a subset of peptides. The specificity of each of these filters matches the estimated specificity of their biological counterparts (Fig. 2, and Burroughs et al., 2004; Tenzer et al., 2005; Assarsson et al., 2007; Burgevin et al., 2008). A full description of how the antigen presentation pathway was implemented is available in the Materials and methods section.

Viruses are represented as a string of letters, and are tested against the antigen presentation pathway of the host to determine the number of CTL epitopes that the virus carries (i.e. the number of substrings that can pass through the filters of the current host). The number of unique CTL epitopes that the virus carries, E, as well as the number of amino acid mutations, M, that the virus carries compared to its wildtype sequence determine the viral load (Eq. (3)). The viral load in turn influences the chance that a host dies during a death event (Eq. (1), Fig. 1C), and the chance that transmission of the virus occurs during infection events (Eq. (2), Figs. 1B, D). During escape and reversion events, one of the mutants in the quasi-species within a host will

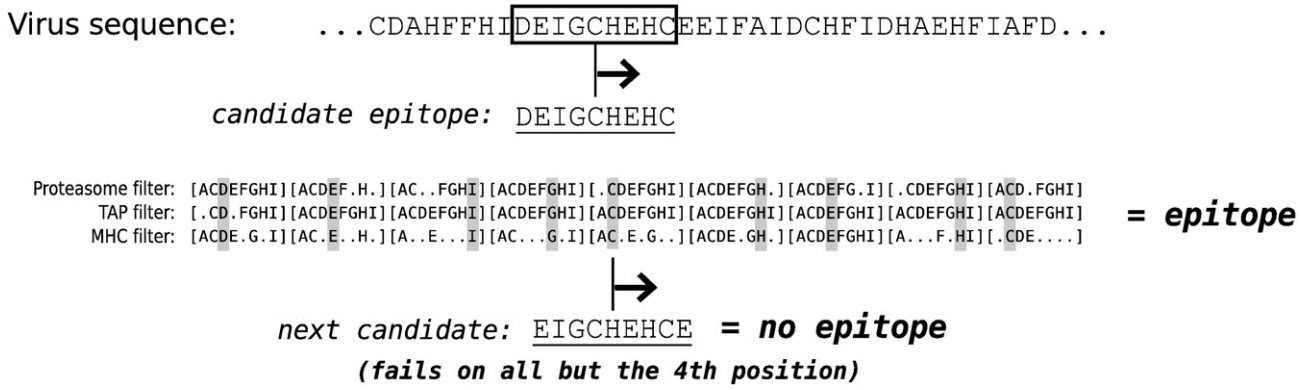

Fig. 2. Three pattern filters (implemented with regular expressions) acting as the proteasome, TAP and MHC steps of the Ag presentation pathway of a host. The filters match their biological counterparts in specificity, but are otherwise randomly generated. The proteasome and TAP filters are selected to have a combined specificity of 25%. The MHC filter has a specificity of 0.05%, which represents a combination of the specificity of the MHC molecule itself, and the MHC specific chance that a presented peptide is an immunodominant epitope. Highlighted regions indicate which amino acids on a particular position match all three filters. Peptides like DEIGCHEHC can be recognized by all three filters, and are counted as immunodominant CTL epitopes. For visual clarity, this example features an amino acid alphabet of 8 letters only. The model uses an alphabet of 20 amino acids.

Please cite this article as: Schmid, B.V., et al., Quantifying how MHC polymorphism prevents pathogens from adapting to the antigen presentation pathway, Epidemics (2010), doi:10.1016/j.epidem.2010.05.003

become the new dominant sequence, if it has a higher fitness than the resident virus sequence (Eq. (3)). In this way the virus adapts to its current host. The model is described in full detail in the Materials and methods section.

The model was initialized with a host population at its maximum population size of 5000 hosts between age 15 and 100, with monomorphic proteasome and TAP molecules, and an MHC polymorphism modeled after that of the European population (Fig. 1A). 5% of the population was inoculated with a wildtype virus sequence that was randomly generated at the start of the simulation. After 1200 timesteps (i.e. 100 years), the simulation was stopped. With this model we tracked the adaptation of a virus to its current host (Fig. 3), and distinguished between adaptation to the monomorphic proteasome and TAP, and the polymorphic MHC alleles.

Within and between-host evolution

The adaptation of a virus to the antigen presentation (MHC class I) or antigen processing (proteasome and TAP) machinery was visualized by tracking a single virus as it passes from one host to the other (Fig. 3). During these passages, the fraction of epitopes in the virus that carried escape mutations, out of all the epitopes that the host could recognize in the wildtype sequence was monitored (black line), as well as the relative number of wildtype CTL epitopes that had escaped presentation in the host due to MHC-binding escape mutations (dashed line) or due to antigen processing escape mutations (dotted line). Note that a single epitope can carry an escape mutation that affects both MHC binding and epitope precursor processing at the same time.

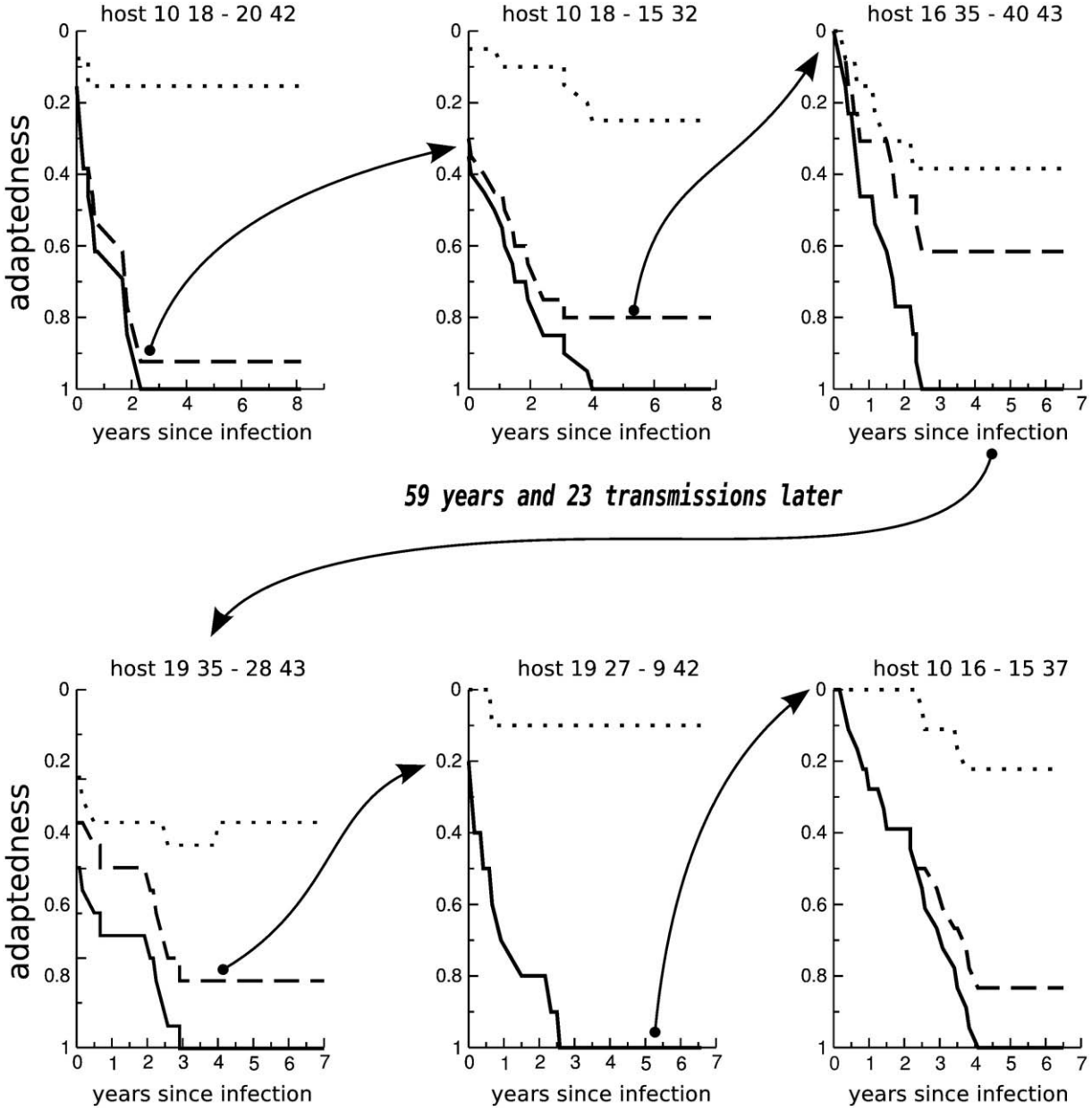

Fig. 3. Transmission chain of a virus through the host population. Each panel represents a host, and carries a label with its 4 MHC alleles, identified by a number (the first two numbers refer to the HLA-A alleles of the host, the second two to its HLA-B alleles). The time since infection is plotted on the horizontal axis, and the within-host level of adaptation to CTL epitopes (black lines) is shown on the vertical axis and ranges from 0 (virus carries as many epitopes as the wildtype virus would have in this host) to 1 (the virus is fully adapted to this host, and carries no epitopes anymore). The initial level of adaptation at time 0 indicates how pre-adapted a virus is to this host, and reflects the level of adaptation that the virus has reached to the whole host population. The dashed and dotted lines indicate what fraction of the wildtype virus' epitopes have escaped due to mutations that affected the proteasome and TAP filters (epitope precursor escapes), or the MHC filter (MHC-binding escapes). Arrows indicate the moment in the infection where the host transmitted the virus to the next host.

Please cite this article as: Schmid, B.V., et al., Quantifying how MHC polymorphism prevents pathogens from adapting to the antigen presentation pathway, Epidemics (2010), doi:10.1016/j.epidem.2010.05.003

Within hosts, the virus predominantly adapted to the most specific step of the antigen presentation pathway, i.e. the binding of epitope precursors to any of the MHC alleles in the host (dashed lines). Escapes that affected the processing of epitope precursors were relatively rare. The first two hosts shown in Fig. 3 were the third and fourth host that the virus infected since the start of the epidemic. These two hosts had identical human leukocyte antigen A (HLA-A) molecules. As a result, many of the escape mutations that the virus accumulated in the first host were also functional escape mutations in the second host, and thus the virus started its within-host adaptation to the second host relatively pre-adapted. The virus carried no pre-adaptations for the third host shown in Fig. 3, and therefore the virus began its within-host adaptation from a fully unadapted starting point. A few decades later in the epidemic, the transmission chain continued to show this pattern (bottom half, Fig. 3), in which viruses sometimes carried many, and sometimes few or no adaptations to the new hosts they infected (sixth host, Fig. 3). This is not because the virus had never before adapted to the MHC molecules of some of these hosts, but because the virus had not maintained the adaptations to the MHC molecules of these hosts. The virus is limited in its accumulation of escape mutations for dominant CTL epitopes by its maximum escape rate of 1/3 per month, but also by its reversion rate of 1/6 per month. Furthermore, the virus has to deal with fitness constraints on how many escape mutations to immunodominant epitopes it can harbor at the same time (± 16 in this simulation), without lowering its fitness so much that it cannot be transmitted anymore (Fig. 1B, Eq. (3)).

Viral adaptation approaches a quasi-steady state

In the previous section, we followed a single virus through its subsequent hosts, and saw how it continuously adapted to its current host, yet only reached a limited level of pre-adaptation to new hosts that it infected.

In this section, we looked at the adaptation of all virus variants in the population from the start of the epidemic and onwards. As the virus adapted to the population, the average level of pre-adaptatedness to the CTL epitopes in a newly infected host increased at first, but stabilized in less than 40 years at a level of 35% of the epitopes (Fig. 4, top left panel, solid black line). Most of the escaped epitopes were due to MHC-binding escape mutations that were present in the infecting virus (86%, dashed line), but a significant fraction of the escaped epitopes was (also) affected by epitope precursor escape mutations (29%, dotted line). The ratio of epitope precursor escape mutations to MHC-binding escape mutations was on average 0.35 to 1.

A sensitivity analysis for four parameters of the model (fitness cost of mutations, frequency with which new relationships are established between hosts, and the rates at which mutations accumulate and reversions occur) showed how both the fraction of pre-adapted CTL epitopes (Fig. 4), and the time until a quasi-steady state was reached (Fig. 5), would change in response to changes in the parameter settings. A faster mutation rate (Fig. 4) did not result in a much higher level of adaptation, indicating that the virus was not limited by its mutation rate, but by the reversion rate of obsolete escape mutations. A slower

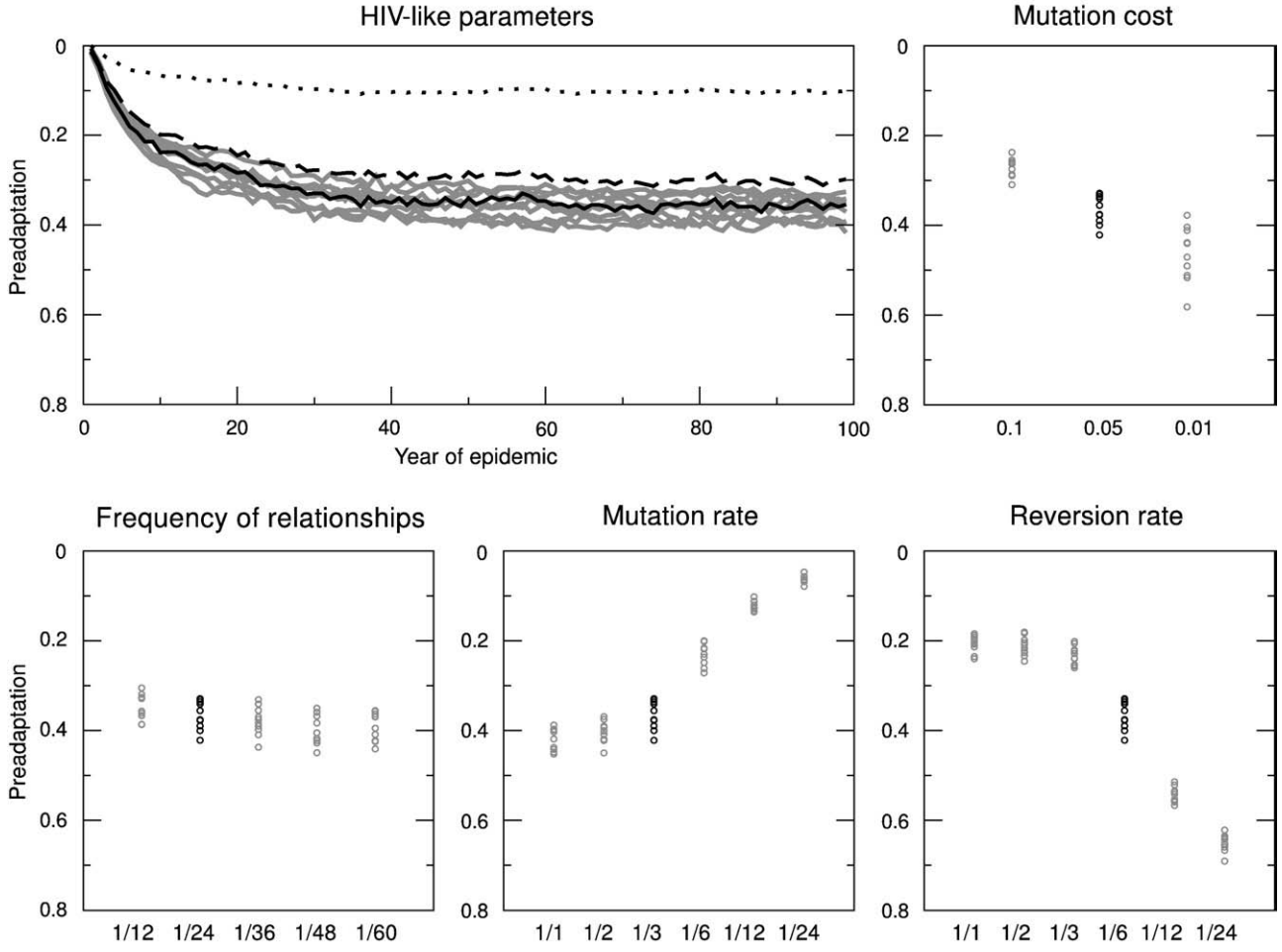

Fig. 4. Virus level of pre-adaptation to the host population reaches a quasi-steady state in less than 4 decades, and is limited to approximately 35% (top left panel, black line, with 9 duplicate runs in grey). Most of the pre-adaptations that the virus carries after reaching the quasi-steady state are mutations that escape MHC binding (dashed line), but a significant amount of mutations (also) affect epitope precursors (dotted line). Changing the parameters for mutation cost, number of new relationships per month, mutation rate per month or reversion rate per month from their defaults (black dots) shows the sensitivity of the model outcome to the parameter settings. Each dot represents a simulation.

Please cite this article as: Schmid, B.V., et al., Quantifying how MHC polymorphism prevents pathogens from adapting to the antigen presentation pathway, Epidemics (2010), doi:10.1016/j.epidem.2010.05.003

reversion rate did result in a higher level of adaptation, and at very slow reversion rates, the level of adaptation became limited by the fitness cost of accumulated escape mutations (Eq. (3)). Similarly, decreasing the fitness costs of escape mutations increased the level of adaptation that the virus reached to the host population, but the reversion process prevented the virus from accumulating a large number of escape mutations. The years it took for the virus to approach a quasi-steady state was most sensitive to the reversion rate and the fitness cost of the escape mutations (Fig. 5). However, for all parameter settings, most of the adaptation happened within the first 4 decades of the epidemic. The ratio of epitope precursor escape mutations to MHC-binding escape mutations was largely unaffected by the parameter sweeps and ranged from 0.3 to 1, to 0.4 to 1 (data not shown).

Over the course of the epidemic, the population size dropped to ±60% of the original population size, and the average age of individuals drops from 50 to 21 years. In the quasi-steady state, ±66% of the population was infected with the virus. This is a higher prevalence than the maximum prevalence that has been observed thus far for HIV-1 (42% in Swaziland; Mathunjwa and Gary, 2006), and close to an earlier estimate of the quasi-steady state prevalence of HIV-1 (70%; van Ballegooijen et al., 2003). In 92% of the 170 simulations shown in Fig. 4, the prevalence fell between 45% and 73%. Prevalences that were lower than 45% or higher than 73% occurred only at the lowest (1/60) and highest (1/12) number of new relationships per host per month, respectively, which is an intuitive result.

Effect of MHC polymorphism on adaptation

When a virus is transmitted to a new host, the immune selection pressure on the virus shifts to a new subset of epitope precursors, due to the MHC polymorphism and relatively high specificity of MHC class I molecules. Epitope precursor escape mutations that are no longer under selection pressure can revert to the wildtype sequence (Friedrich et al., 2004; Barouch et al., 2005; Herbeck et al., 2006; Goonetilleke et al., 2009). In Schmid et al. (2008) we postulated that this intermittent exposure of epitope precursors is what prevented HIV-1 from efficiently exploiting the monomorphic proteasome and TAP molecules.

The proteasome and TAP filters in the model have a combined specificity of 25%, which means that on average 3 out of 4 mutations in an epitope will result in an epitope precursors escape mutation. In simulations with a realistic MHC allele frequency distribution (Fig. 4), we found that after 4 decades of viral evolution, a slightly higher fraction (29%) of the escaped epitopes were epitope precursor escape mutations. The small difference between the expected and observed fraction suggested that in this host population, the immune selection pressure

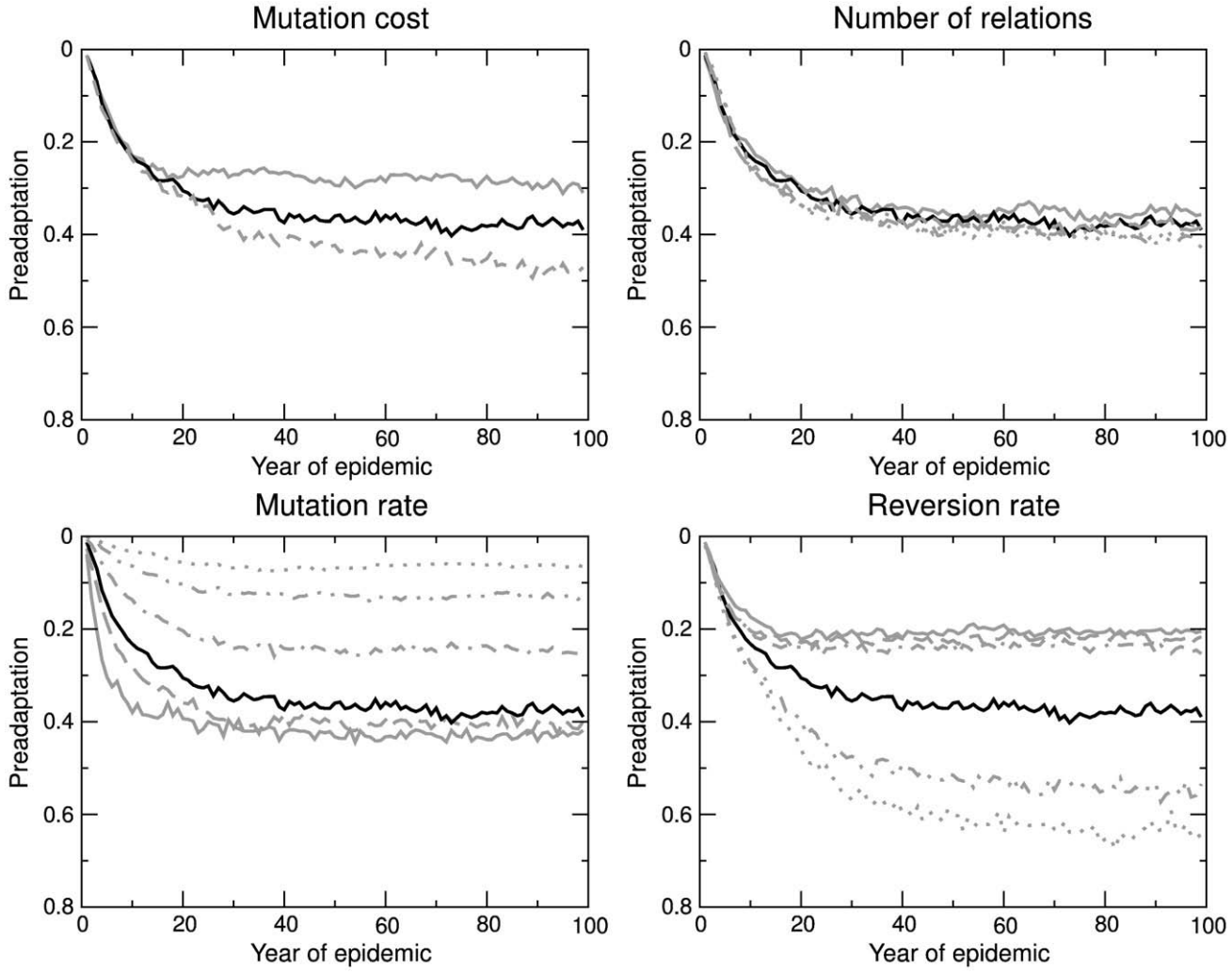

Fig. 5. The time until the level of adaptation of the virus reaches a quasi-steady state. Four different parameters were changed from HIV-like settings (black lines) to determine the sensitivity of the model outcome to parameter changes. The solid grey lines represent the highest parameter values for the four tested parameters (0.1 for mutation cost, 1/12 for new relationship frequency, 1/1 for mutation rate, 1/1 for reversion rate). As the lines become more dotted, they represent lower parameter values (as listed in Fig. 4), with the fully dotted line representing the parameter values of 0.01, 1/60, 1/24 and 1/24, respectively.

Please cite this article as: Schmid, B.V., et al., Quantifying how MHC polymorphism prevents pathogens from adapting to the antigen presentation pathway, Epidemics (2010), doi:10.1016/j.epidem.2010.05.003

on the virus to adapt to the monomorphic proteasome and TAP was relatively small.

By altering the degree of MHC polymorphism in the host population, we could manipulate the selection pressure on the virus to adapt to the monomorphic components of the antigen presentation pathway, and determine how the total level of adaptation that a virus could reach to the population was affected by the MHC polymorphism (Fig. 6). In these simulations in which we varied the MHC polymorphism, the MHC alleles had equal frequencies in the host population.

We express the immune selection pressure on the virus to adapt to the proteasome and TAP as the ratio of epitope precursor escape mutations to MHC-binding escape mutations. At the lowest degrees of MHC polymorphism, the virus was very well adapted to the host population, and carried escape mutations for most of the dominant CTL epitopes in the host population (Fig. 6, white bars). Its ratio of epitope precursor escape mutations to MHC-binding escape mutations was 0.32 to 1 (Fig. 6, dashed line). Increasing the degree of MHC polymorphism resulted in a lower overlap in shared CTL epitopes between subsequent hosts (Fig. 6, solid line), and the virus could only reach a limited level of adaptation to the host population. Unexpectedly, the ratio of epitope precursor escape mutations to MHC-binding mutations saturated at a ratio of 0.38 to 1 with an increasing degree of MHC polymorphism, indicating that the MHC polymorphism had little effect on the selection pressure on the virus to adapt to the monomorphic components of the antigen presentation pathway. Apparently, as a result of the high specificity of MHC molecules in binding dominant CTL epitopes (0.5%), the chance that the MHC molecules of two subsequent hosts were using the same epitope precursors was a limiting factor for the virus in its ability to adapt to the proteasome and TAP filters.

To study a parameter regime in which MHC molecules were more frequently sharing epitope precursors, we reduced the length of the virus genome sixfold to 500aa, and increased the specificity of MHC molecules from 0.5% to 3.0%. In this parameter regime, the ratio of epitope precursor to MHC binding escapes ranged from 0.46 to 1 at low MHC polymorphism to 0.74 to 1 at a high degree of MHC polymorphism. The difference between the two parameter regimes was most pronounced in the level of pre-adaptation at high degrees of MHC polymorphism. In the original parameter regime the fraction of epitopes in the virus that was pre-adapted was 17% at an MHC polymorphism of 40 alleles per locus (Fig. 6, white bars), whereas in this parameter regime the fraction of pre-adapted epitopes reached 33% (data not shown).

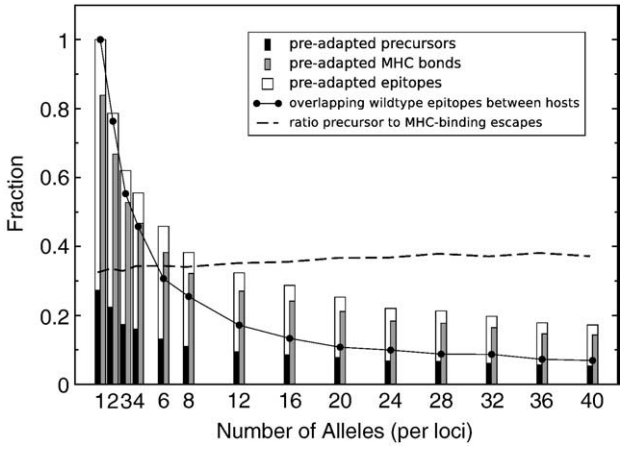

Fig. 6. The degree of MHC polymorphism affects the quasi-steady state level of pre-adaptation that the virus reaches (i.e. the fraction of CTL epitopes for which the virus carries escape mutations at the time of infection of a new host). Increasing the number of MHC alleles in the host population increases the fraction of pre-adapted CTL epitopes (white bars). These adaptations can be either due to precursor escape mutations, that escape the proteasome or TAP filter (black bars), MHC-binding escape mutations, that escape the MHC filter (grey bars), or both. Each bar is the average of 10 simulations. The ratio of epitope precursors to MHC-binding escape is plotted as a dashed line. The average overlap in CTL epitopes in the wildtype virus that subsequent hosts can recognize is plotted as a solid line, and shows how an increase in MHC polymorphism affects the continuity of the immune selection pressure on epitopes.

Summarizing, even in conditions that strongly favored the virus to adapt to the proteasome and TAP, we find no massive adaptation to these monomorphic molecules. Under all conditions tested, the majority of the adaptations in the virus were escape mutations that affected the MHC-binding filter.

Discussion

Viruses can escape CTL responses by acquiring mutations that abrogate the binding of CTL epitopes to the MHC class I molecules of the host. Due to the high degree of MHC polymorphism in the human population, such mutations typically no longer have a selective advantage for the virus when it is transmitted to a new host. However, efficient CTL epitope presentation also depends on two monomorphic components of the antigen presentation pathway, the proteasome and TAP. In an MHC polymorphic host population, viruses could be under selection pressure to adapt to these monomorphic components (Brander et al., 1999; Yokomaku et al., 2004), as a single epitope precursor escape mutation would prevent the formation of epitope–MHC complexes that depend on that particular epitope precursor in all hosts.

In this paper we tested the ability of an HIV-1 like virus to adapt its genome to escape the presentation of CTL epitopes in a host population, using an agent-based simulation model. We were especially interested in the ratio of epitope precursor escape mutations to MHC-binding escape mutations, to understand whether viruses were under strong selection pressure to accumulate proteasome and TAP escape mutations in the presence of a high MHC polymorphism. In our simulations, the virus approached a quasi-steady state level of adaptation to its host population in a few decades, which was limited by the reversion of escape mutations that were no longer under selection pressure in its current hosts, and by the fitness cost of the escape mutations that it had accumulated (Figs. 4 and 5). In this quasi-steady state, the ratio of epitope precursor escapes to MHC-binding escapes in the virus remained close to their expected values (based on the specificity of the proteasome, TAP and MHC molecules of the hosts), indicating that there was only a small amount of selection pressure on the virus to adapt to the monomorphic proteasome and TAP. Even when we increased the degree of MHC polymorphism in the host population (Fig. 6), the selection pressure on the virus to adapt to the monomorphic components of the antigen presentation pathway remained limited: the presence of an MHC polymorphism in combination with the high specificity of the MHC class I molecules reduced the shared usage of the epitope precursors between different hosts to a too low level for the virus to exploit the monomorphic properties of the proteasome and TAP alleles.

Recently, there have been reports that the number of epitopes in HIV-1 is gradually decreasing over time (Kawashima et al., 2009; Vider-Shalit et al., 2009), suggesting that HIV-1 has not yet reached a quasi-steady state. Kawashima et al. (2009) reported that between 1983 and 2008, the frequency of a particular mutation in the HIV-1 protein reverse transcriptase (RT I135X), which affects the presentation of an HLA-B*51 CTL epitope, increased from 21% to 70% in HLA-B*51-negative patients. It could be that the observed accumulation of escape mutations is not due to an ongoing process of accumulation in HIV-1 since it was introduced in the human species, but reflects a more recent change in the MHC allele frequencies that the virus encountered. The HIV-1 virus was introduced in Japan in 1983 through Japanese hemophiliacs that were infected with HIV-1 from blood plasma imported from the USA (Shimizu et al., 1992), where the prevalence of HLA-B*51 is lower (12%) than in Japan (22%) (Kawashima et al., 2009; Robinson et al., 2006). By 2008, the HIV-1 virus would have had 25 years to adapt to the MHC allele frequency distribution of the Japanese population, and therefore an increase in the I135X escape variant mutation is to be expected, even if the virus would have reached a quasi-steady state level of adaptation in 1983 in the USA. Whether this change in local MHC frequencies is

Please cite this article as: Schmid, B.V., et al., Quantifying how MHC polymorphism prevents pathogens from adapting to the antigen presentation pathway, Epidemics (2010), doi:10.1016/j.epidem.2010.05.003

sufficient to explain the observed increase in the RT I135X escape variant is unclear. Based on the above explanation for the increase of the RT I135X escape variant, we would predict that epitope escape variants for MHC alleles that are common in the USA, but are rare in Japan, would have decreased in frequency between 1983 and 2008 in Japanese HIV-1 infected patients.

Our simulations showed that an HIV-1 like virus could reach a quasi-steady state level of adaptation to a host population with a realistic MHC allele frequency distribution in a few decades (Fig. 5, <40 years), and outlines the parameters that this time till quasi-steady state depends on. In Schmid et al. (2008), we observed no significant decrease in the number of epitope precursors or CTL epitopes in HIV-1 class B during the last 30 years, which fits the simulation results, as the introduction of HIV-1 group M is estimated to have occurred around 1910 (Worobey et al., 2008) or 1930 (Korber et al., 2000). Moreover, the simulations in this paper start with a wildtype virus that is randomly generated and fully unadapted to the host population, whereas HIV-1 is related to and likely originated from the SIVcpz virus in the chimpanzee subspecies *Pan troglodytes troglodytes* (Keele et al., 2006), to which it has adapted over a few centuries (Wertheim and Worobey, 2009). Humans and chimpanzee are closely related, and chimpanzee orthologs of the active subunits of proteasome and TAP1 have an amino acid sequence similarity of >98% (although for three proteasomal subunits and TAP2 we have limited information, as their annotation is not complete). Furthermore, the chimpanzee proteasome and TAP appear to be functionally equivalent to their human counterparts in experimental settings (Kimura et al., 2005; Dugan and Hewitt, 2008). The amino acid sequence similarity of MHC class I molecules between humans and chimpanzee orthologs is also over 98% (Anzai et al., Jun, 2003), and share similar binding motifs (Hoof et al., 2008). It therefore is likely that at least some of the adaptations that SIVcpz made to the chimpanzee host were carried over when the virus jumped to the human species, giving HIV-1 a head start in adapting to the human population, compared to the simulations in the model.

One concept that is missing in the model is the possibility for the virus to create compensatory mutations, i.e., the ability to lower the mutation cost of escape mutations by additional mutations (Nijhuis et al., 1999; Leslie et al., 2005). This type of mutation could decrease the reversion rate of the virus, as the virus might be locked into a local fitness optimum, which slows down or prevents reversion of the escape mutation (Schneidewind et al., 2009). In the simulation, a lower reversion rate of escape mutations predominantly affected the level of pre-adaptation that the virus could reach (Fig. 4). However, if epitope precursor escape mutations and MHC-binding escape mutations differ in their ability to be locked in by compensatory mutations, it could skew the ratio of epitope precursors escape mutations to MHC-binding escape mutations, promoting the accumulation of one over the other.

Summarizing, in the presence of an MHC polymorphism, the high specificity of the MHC class I molecules reduces the shared usage of epitope precursors between different hosts to a low level. As a result, the selection pressure on the virus to adapt to the monomorphic proteasome and TAP is relatively small, which explains the low level of adaptation of viruses to the proteasome and TAP, despite the fact that these two steps of the antigen presentation pathway are monomorphic.

Acknowledgments

The authors would like to thank Dr. Jonathan M. Carlson for his thought-provoking and outstanding review of the initial manuscript. This study was financially supported by the Netherlands Organisation for Scientific Research (NWO) (VICI grant 016.048.603, Open Program 812.07.003), and a High Potential grant (2007) from Utrecht University.

References

Althaus, C.L., de Boer, R.J., 2008. Dynamics of immune escape during HIV/SIV infection. PLoS Comput. Biol. 4 (7), e1000103 URL http://dx.doi.org/10.1371/journal.pcbi.1000103.

Anzai, T., Shiina, T., Kimura, N., Yanagiya, K., Kohara, S., Shigenari, A., Yamagata, T., Kulski, J. K., Naruse, T.K., Fujimori, Y., Fukuzumi, Y., Yamazaki, M., Tashiro, H., Iwamoto, C., Umehara, Y., Imanishi, T., Meyer, A., Ikeo, K., Gojobori, T., Bahram, S., Inoko, H., 2003. Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. Proc. Natl. Acad. Sci. U S A 100 (13), 7708–7713 Jun, URL http://dx.doi.org/10.1073/pnas.1230533100.

Asquith, B., Edwards, C.T.T., Lipsitch, M., McLean, A.R., 2006. Inefficient cytotoxic T lymphocyte-mediated killing of HIV-1-infected cells in vivo. PLoS Biol. 4 (4), e90 Apr, URL http://dx.doi.org/10.1371/journal.pbio.0040090.

Assarsson, E., Sidney, J., Oseroff, C., Pasquetto, V., Bui, H.-H., Frahm, N., Brander, C., Peters, B., Grey, H., Sette, A., 2007. A quantitative analysis of the variables affecting the repertoire of T cell specificities recognized after vaccinia virus infection. J. Immunol. 178 (12), 7890–7901 Jun.

Barouch, D.H., Powers, J., Truitt, D.M., Kishko, M.G., Arthur, J.C., Peyerl, F.W., Kuroda, M. J., Gorgone, D.A., Lifton, M.A., Lord, C.I., Hirsch, V.M., Montefiori, D.C., Carville, A., Mansfield, K.G., Kunstman, K.J., Wolinsky, S.M., Letvin, N.L., 2005. Dynamic immune responses maintain cytotoxic T lymphocyte epitope mutations in transmitted simian immunodeficiency virus variants. Nat. Immunol. 6 (3), 247–252 Mar, URL http://dx.doi.org/10.1038/ni1167.

Borghans, J.A.M., Beltman, J.B., de Boer, R.J., 2004. MHC polymorphism under host–pathogen coevolution. Immunogenetics 55 (11), 732–739 Feb, URL http://dx.doi.org/10.1007/s00251-003-0630-5.

Brander, C., Yang, O.O., Jones, N.G., Lee, Y., Goulder, P., Johnson, R.P., Trocha, A., Colbert, D., Hay, C., Buchbinder, S., Bergmann, C.C., Zweerink, H.J., Wolinsky, S., Blattner, W. A., Kalams, S.A., Walker, B.D., 1999. Efficient processing of the immunodominant, HLA-A*0201-restricted human immunodeficiency virus type 1 cytotoxic T-lymphocyte epitope despite multiple variations in the epitope flanking sequences. J. Virol. 73 (12), 10191–10198 Dec.

Burgevin, A., Saveanu, L., Kim, Y., Barilleau, E., Kotturi, M., Sette, A., van Endert, P., Peters, B., 2008. A detailed analysis of the murine TAP transporter substrate specificity. PLoS ONE 3 (6), e2402 URL http://dx.doi.org/10.1371/journal.pone.0002402.

Burroughs, N.J., de Boer, R.J., Keşmir, C., 2004. Discriminating self from nonself with short peptides from large proteomes. Immunogenetics 56 (5), 311–320 Aug, URL http://dx.doi.org/10.1007/s00251-004-0691-0.

Carnes, B.A., Holden, L.R., Olshansky, S.J., Witten, M.T., Siegel, J.S., 2006. Mortality partitions and their relevance to research on senescence. Biogerontology 7 (4), 183–198 Aug, URL http://dx.doi.org/10.1007/s10522-006-9020-3.

Carrington, M., Nelson, G.W., Martin, M.P., Kissner, T., Vlahov, D., Goedert, J.J., Kaslow, R., Buchbinder, S., Hoots, K., O'Brien, S.J., 1999. HLA and HIV-1: heterozygote advantage and B*35–Cw*04 disadvantage. Science 283 (5408), 1748–1752 Mar.

Chakraborty, H., Sen, P.K., Helms, R.W., Vernazza, P.L., Fiscus, S.A., Eron, J.J., Patterson, B. K., Coombs, R.W., Krieger, J.N., Cohen, M.S., 2001. Viral burden in genital secretions determines male-to-female sexual transmission of HIV-1: a probabilistic empiric model. AIDS 15 (5), 621–627 Mar.

Costin, J.M., 2007. Cytopathic mechanisms of HIV-1. Virol. J. 4, 100 URL http://dx.doi.org/10.1186/1743-422X-4-100.

Davenport, M.P., Loh, L., Petravic, J., Kent, S.J., 2008. Rates of HIV immune escape and reversion: implications for vaccination. Trends Microbiol. 16 (12), 561–566 Dec, URL http://dx.doi.org/10.1016/j.tim.2008.09.001.

Doherty, P.C., Zinkernagel, R.M., 1975. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature 256 (5512), 50–52 Jul.

Dugan, G.E., Hewitt, E.W., 2008. Structural and functional dissection of the human cytomegalovirus immune evasion protein us6. J. Virol. 82 (7), 3271–3282 Apr, URL http://dx.doi.org/10.1128/JVI.01705-07.

Frahm, N., Yusim, K., Suscovich, T.J., Adams, S., Sidney, J., Hraber, P., Hewitt, H.S., Linde, C.H., Kavanagh, D.G., Woodberry, T., Henry, L.M., Faircloth, K., Listgarten, J., Kadie, C., Jojic, N., Sango, K., Brown, N.V., Pae, E., Zaman, M.T., Bihl, F., Khatri, A., John, M., Mallal, S., Marincola, F.M., Walker, B.D., Sette, A., Heckerman, D., Korber, B.T., Brander, C., 2007. Extensive HLA class I allele promiscuity among viral CTL epitopes. Eur. J. Immunol. 37 (9), 2419–2433 Sep, URL http://dx.doi.org/10.1002/eji.200737365.

Friedrich, T.C., Dodds, E.J., Yant, L.J., Vojnov, L., Rudersdorf, R., Cullen, C., Evans, D.T., Desrosiers, R.C., Moth, B.R., Sidney, J., Sette, A., Kunstman, K., Wolinsky, S., Piatak, M., Lifson, J., Hughes, A.L., Wilson, N., O'Connor, D.H., Watkins, D.I., 2004. Reversion of CTL escape-variant immunodeficiency viruses in vivo. Nat. Med. 10 (3), 275–281 Mar, URL http://dx.doi.org/10.1038/nm998.

Gompertz, B., 1825. On the nature of the function expressive of the law of human mortality. Philos. Trans. R. Soc. Lond. 36, 513–585.

Goonetilleke, N., Liu, M.K.P., Salazar-Gonzalez, J.F., Ferrari, G., Giorgi, E., Ganusov, V.V., Keele, B.F., Learn, G.H., Turnbull, E.L., Salazar, M.G., Weinhold, K.J., Moore, S., B. C. H. A. V. I. C. C., Letvin, N., Haynes, B.F., Cohen, M.S., Hraber, P., Bhattacharya, T., Borrow, P., Perelson, A.S., Hahn, B.H., Shaw, G.M., Korber, B.T., McMichael, A.J., 2009. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. J. Exp. Med. 206 (6), 1253–1272 Jun, URL http://dx.doi.org/10.1084/jem.20090365.

Groothuis, T.A.M., Griekspoor, A.C., Neijssen, J.J., Herberts, C.A., Neefjes, J.J., 2005. MHC class I alleles and their exploration of the antigen-processing machinery. Immunol. Rev. 207, 60–76 Oct, URL http://dx.doi.org/10.1111/j.0105-2896.2005.00305.x.

Hallén, A., 2007. Gompertz law and aging as exclusion effects. Biogerontology 8 (5), 605–612 Oct, URL http://dx.doi.org/10.1007/s10522-007-9087-5.

Herbeck, J.T., Nickle, D.C., Learn, G.H., Gottlieb, G.S., Curlin, M.E., Heath, L., Mullins, J.I., 2006. Human immunodeficiency virus type 1 ENV evolves toward ancestral states

Please cite this article as: Schmid, B.V., et al., Quantifying how MHC polymorphism prevents pathogens from adapting to the antigen presentation pathway, Epidemics (2010), doi:10.1016/j.epidem.2010.05.003

upon transmission to a new host. J. Virol. 80 (4), 1637–1644 Feb, URL http://dx.doi.org/10.1128/JVI.80.4.1637-1644.2006.

Hoof, I., Keşmir, C., Lund, O., Nielsen, M., 2008. Humans with chimpanzee-like major histocompatibility complex-specificities control HIV-1 infection. AIDS 22 (11), 1299–1303 Jul, URL http://dx.doi.org/10.1097/QAD.0b013e328302f39f.

Kawashima, Y., Pfafferott, K., Frater, J., Matthews, P., Payne, R., Addo, M., Gatanaga, H., Fujiwara, M., Hachiya, A., Koizumi, H., Kuse, N., Oka, S., Duda, A., Prendergast, A., Crawford, H., Leslie, A., Brumme, Z., Brumme, C., Allen, T., Brander, C., Kaslow, R., Tang, J., Hunter, E., Allen, S., Mulenga, J., Branch, S., Roach, T., John, M., Mallal, S., Ogwu, A., Shapiro, R., Prado, J.G., Fidler, S., Weber, J., Pybus, O.G., Klenerman, P., Ndung'u, T., Phillips, R., Heckerman, D., Harrigan, P.R., Walker, B.D., Takiguchi, M., Goulder, P., 2009. Adaptation of HIV-1 to human leukocyte antigen class I. Nature 458 (7238), 641–645 Apr.

Keele, B.F., Heuverswyn, F.V., Li, Y., Bailes, E., Takehisa, J., Santiago, M.L., Bibollet-Ruche, F., Chen, Y., Wain, L.V., Liegeois, F., Loul, S., Ngole, E.M., Bienvenue, Y., Delaporte, E., Brookfield, J.F.Y., Sharp, P.M., Shaw, G.M., Peeters, M., Hahn, B.H., 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science 313 (5786), 523–526 Jul, URL http://dx.doi.org/10.1126/science.1126531.

Kesmir, C., van Noort, V., de Boer, R.J., Hogeweg, P., 2003. Bioinformatic analysis of functional differences between the immunoproteasome and the constitutive proteasome. Immunogenetics 55 (7), 437–449 Oct, URL http://dx.doi.org/10.1007/s00251-003-0585-6.

Kiepiela, P., Ngumbela, K., Thobakgale, C., Ramduth, D., Honeyborne, I., Moodley, E., Reddy, S., de Pierres, C., Mncube, Z., Mkhwanazi, N., Bishop, K., van der Stok, M., Nair, K., Khan, N., Crawford, H., Payne, R., Leslie, A., Prado, J., Prendergast, A., Frater, J., McCarthy, N., Brander, C., Learn, G.H., Nickle, D., Rousseau, C., Coovadia, H., Mullins, J.I., Heckerman, D., Walker, B.D., Goulder, P., 2007. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. Nat. Med. 13 (1), 46–53 Jan, URL http://dx.doi.org/10.1038/nm1520.

Kimura, Y., Gushima, T., Rawale, S., Kaumaya, P., Walker, C.M., 2005. Escape mutations alter proteasome processing of major histocompatibility complex class I-restricted epitopes in persistent hepatitis C virus infection. J. Virol. 79 (8), 4870–4876 Apr, URL http://dx.doi.org/10.1128/JVI.79.8.4870-4876.2005.

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B.H., Wolinsky, S., Bhattacharya, T., 2000. Timing the ancestor of the HIV-1 pandemic strains. Science 288 (5472), 1789–1796 Jun.

Kotturi, M.F., Scott, I., Wolfe, T., Peters, B., Sidney, J., Cheroutre, H., von Herrath, M.G., Buchmeier, M.J., Grey, H., Sette, A., 2008. Naive precursor frequencies and MHC binding rather than the degree of epitope diversity shape CD8+ T cell immunodominance. J. Immunol. 181 (3), 2124–2133 Aug.

Kwun, H.J., da Silva, S.R., Shah, I.M., Blake, N., Moore, P.S., Chang, Y., 2007. Kaposi's sarcoma-associated herpesvirus latency-associated nuclear antigen 1 mimics Epstein–Barr virus EBNA1 immune evasion through central repeat domain effects on protein processing. J. Virol. 81 (15), 8225–8235 Aug, URL http://dx.doi.org/10.1128/JVI.00411-07.

Langefors, A., Lohm, J., Grahn, M., Andersen, O., von Schantz, T., 2001. Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. Proc. Biol. Sci. 268 (1466), 479–485 Mar.

Larsen, M.V., Lundegaard, C., Lamberth, K., Buus, S., Brunak, S., Lund, O., Nielsen, M., 2005. An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. Eur. J. Immunol. 35 (8), 2295–2303 Aug, URL http://dx.doi.org/10.1002/eji.200425811.

Lavreys, L., Baeten, J.M., Chohan, V., McClelland, R.S., Hassan, W.M., Richardson, B.A., Mandaliya, K., Ndinya-Achola, J.O., Overbaugh, J., 2006. Higher set point plasma viral load and more-severe acute HIV type 1 (HIV-1) illness predict mortality among high-risk HIV-1-infected African women. Clin. Infect. Dis. 42 (9), 1333–1339 May, URL http://dx.doi.org/10.1086/503258.

Leslie, A., Kavanagh, D., Honeyborne, I., Pfafferott, K., Edwards, C., Pillay, T., Hilton, L., Thobakgale, C., Ramduth, D., Draenert, R., Gall, S.L., Luzzi, G., Edwards, A., Brander, C., Sewell, A.K., Moore, S., Mullins, J., Moore, C., Mallal, S., Bhardwaj, N., Yusim, K., Phillips, R., Klenerman, P., Korber, B., Kiepiela, P., Walker, B., Goulder, P., 2005. Transmission and accumulation of CTL escape variants drive negative associations between HIV polymorphisms and HLA. J. Exp. Med. 201 (6), 891–902 Mar, URL http://dx.doi.org/10.1084/jem.20041455.

Mathunjwa, T.R., Gary, F.A., 2006. Women and HIV/AIDS in the kingdom of Swaziland: culture and risks. J. Natl. Black Nurses Assoc. 17 (2), 39–46 Dec.

Nijhuis, M., Schuurman, R., de Jong, D., Erickson, J., Gustchina, E., Albert, J., Schipper, P., Gulnik, S., Boucher, C.A., 1999. Increased fitness of drug resistant hiv-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. AIDS 13 (17), 2349–2359 Dec.

Piatak, M., Saag, M.S., Yang, L.C., Clark, S.J., Kappes, J.C., Luk, K.C., Hahn, B.H., Shaw, G.M., Lifson, J.D., 1993. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. Science 259 (5102), 1749–1754 Mar.

Robinson, J., Waller, M.J., Fail, S.C., Marsh, S.G.E., 2006. The IMGT/HLA and IPD databases. Hum. Mutat. 27 (12), 1192–1199 Dec, URL http://dx.doi.org/10.1002/humu.20406.

Salazar-Gonzalez, J.F., Salazar, M.G., Keele, B.F., Learn, G.H., Giorgi, E.E., Li, H., Decker, J.M., Wang, S., Baalwa, J., Kraus, M.H., Parrish, N.F., Shaw, K.S., Guffey, M.B., Bar, K.J., Davis, K.L., Ochsenbauer-Jambor, C., Kappes, J.C., Saag, M.S., Cohen, M.S., Mulenga, J., Derdeyn, C.A., Allen, S., Hunter, E., Markowitz, M., Hraber, P., Perelson, A.S., Bhattacharya, T., Haynes, B.F., Korber, B.T., Hahn, B.H., Shaw, G.M., 2009. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. J. Exp. Med. 206 (6), 1273–1289 Jun, URL http://dx.doi.org/10.1084/jem.20090378.

Sayers, E.W., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S., Feolo, M., Geer, L.Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D.J., Madden, T.L., Maglott, D.R., Miller, V., Mizrachi, I., Ostell, J., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Shumway, M., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusova, T.A., Wagner, L., Yaschenko, E., Ye, J., 2009. Database resources of the national center for biotechnology information. Nucleic Acids Res. 37, D5–D16 Jan, (Database issue) URL http://dx.doi.org/10.1093/nar/gkn741.

Schmid, B.V., Keşmir, C., de Boer, R.J., 2008. The specificity and polymorphism of the MHC Class I prevents the global adaptation of HIV-1 to the monomorphic proteasome and TAP. PLoS ONE 3 (10), e3525 URL http://dx.doi.org/10.1371/journal.pone.0003525.

Schneidewind, A., Brumme, Z.L., Brumme, C.J., Power, K.A., Reyor, L.L., O'Sullivan, K., Gladden, A., Hempel, U., Kuntzen, T., Wang, Y.E., Oniangue-Ndza, C., Jessen, H., Markowitz, M., Rosenberg, E.S., Skaly, R.-P., Kelleher, A.D., Walker, B.D., Allen, T.M., 2009. Transmission and long-term stability of compensated CD8 escape mutations. J. Virol. 83 (8), 3993–3997 Apr, URL http://dx.doi.org/10.1128/JVI.01108-08.

Shimizu, N., Takeuchi, Y., Naruse, T., Inagaki, M., Moriyama, E., Gojobori, T., Hoshino, H., 1992. Six strains of human immunodeficiency virus type 1 isolated in Japan and their molecular phylogeny. J. Mol. Evol. 35 (4), 329–336 Oct.

Slade, R.W., McCallum, H.I., 1992. Overdominant vs. frequency-dependent selection at MHC loci. Genetics 132 (3), 861–864 Nov.

Tenzer, S., Peters, B., Bulik, S., Schoor, O., Lemmel, C., Schatz, M.M., Kloetzel, P.-M., Rammensee, H.-G., Schild, H., Holzhütter, H.-G., 2005. Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. Cell. Mol. Life Sci. 62 (9), 1025–1037 May, URL http://dx.doi.org/10.1007/s00018-005-4528-2.

Trachtenberg, E., Korber, B., Sollars, C., Kepler, T.B., Hraber, P.T., Hayes, E., Funkhouser, R., Fugate, M., Theiler, J., Hsu, Y.S., Kunstman, K., Wu, S., Phair, J., Erlich, H., Wolinsky, S., 2003. Advantage of rare HLA supertype in HIV disease progression. Nat. Med. 9 (7), 928–935 Jul, URL http://dx.doi.org/10.1038/nm893.

van Ballegooijen, M., Bogaards, J.A., Weverling, G.-J., Boerlijst, M.C., Goudsmit, J., 2003. AIDS vaccines that allow HIV-1 to infect and escape immunologic control: a mathematic analysis of mass vaccination. J. Acquir. Immune Defic. Syndr. 34 (2), 214–220 Oct.

Vider-Shalit, T., Almani, M., Sarid, R., Louzoun, Y., 2009. The HIV hide and seek game: an immunogenomic analysis of the HIV epitope repertoire. AIDS 23 (11), 1311–1318 Jul, URL http://dx.doi.org/10.1097/QAD.0b013e32832c492a.

Wawer, M.J., Gray, R.H., Sewankambo, N.K., Serwadda, D., Li, X., Laeyendecker, O., Kiwanuka, N., Kigozi, G., Kiddugavu, M., Lutalo, T., Nalugoda, F., Wabwire-Mangen, F., Meehan, M.P., Quinn, T.C., 2005. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. J. Infect. Dis. 191 (9), 1403–1409 May, URL http://dx.doi.org/10.1086/429411.

Wertheim, J.O., Worobey, M., 2009. Dating the age of the SIV lineages that gave rise to HIV-1 and HIV-2. PLoS Comput. Biol. 5 (5), e1000377 May, URL http://dx.doi.org/10.1371/journal.pcbi.1000377.

Worobey, M., Gemmel, M., Teuwen, D.E., Haselkorn, T., Kunstman, K., Bunce, M., Muyembe, J.-J., Kabongo, J.-M.M., Kalengayi, R.M., Marck, E.V., Gilbert, M.T.P., Wolinsky, S.M., 2008. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature 455 (7213), 661–664 Oct, URL http://dx.doi.org/10.1038/nature07390.

Yokomaku, Y., Miura, H., Tomiyama, H., Kawana-Tachikawa, A., Takiguchi, M., Kojima, A., Nagai, Y., Iwamoto, A., Matsuda, Z., Ariyoshi, K., 2004. Impaired processing and presentation of cytotoxic-T-lymphocyte (CTL) epitopes are major escape mechanisms from CTL immune pressure in human immunodeficiency virus type 1 infection. J. Virol. 78 (3), 1324–1332 Feb.

Please cite this article as: Schmid, B.V., et al., Quantifying how MHC polymorphism prevents pathogens from adapting to the antigen presentation pathway, Epidemics (2010), doi:10.1016/j.epidem.2010.05.003