Published in final edited form as:

Infect Genet Evol. 2011 March ; 11(2): 483–488. doi:10.1016/j.meegid.2010.12.011.

Conflicting Selection Pressures on T-cell Epitopes in HIV-1 subtype B

Stephanie Jiménez Irausquin and Austin L. Hughes*

Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, USA

Abstract

Analysis of population-level polymorphism in eight coding genes of human immunodeficiency virus type 1 (HIV-1) subtype B revealed evidence not only of past purifying selection, but also of abundant slightly deleterious nonsynonymous variants subject to ongoing purifying selection. Both CD4 and CTL epitopes showed an excess of nonsynonymous variants that were singletons (occurring in just one sequence) in our dataset. Overall, median gene diversities at polymorphic nonsynonymous sites were highest at sites located in neither CD4 nor CTL epitopes, while polymorphic nonsynonymous sites in CD4 epitopes revealed the lowest median gene diversity. Our results support the hypothesis that there is an evolutionary conflict between immune escape and functional constraint on epitopes recognized by host T-cells, and suggest that amino acid sequences of CD4 epitopes are subject to particularly strong functional constraint.

Keywords

escape mutation; evolutionary conflict; human immunodeficiency virus; natural selection; T-cell epitope

1. Introduction

Human immunodeficiency virus type 1 (HIV-1) was first identified in the early 1980's as the causative agent of acquired immunodeficiency syndrome or AIDS (Barré-Sinoussi et al., 1983; Gallo et al., 1983) and has since been characterized as a global epidemic. The main cause for AIDS progression is the decline in CD4+ T lymphocytes, which leaves the host's immune system crippled and susceptible to opportunistic infections (Sierra, 2005). According to the December 2009 AIDS epidemic update, although new HIV infections have been reduced by 17% over the past eight years, there are now more people living with HIV than ever before (WHO, 2009). At the end of 2008, it was estimated that the number of people living with HIV/AIDS worldwide was 33.4 million (WHO, 2009).

HIV-1 belongs to the *Retroviridae* family and is transmitted as enveloped, dimeric, positive-sense, single-stranded RNA of approximately 9200 nucleotides in length (ICTVdB Management, 2006). HIV-1 has an extremely high mutation rate, on the order of 3.4×10^{-5} per basepair per cycle (Mansky and Temin, 1995), which results from both an error-prone reverse transcriptase (Roberts et al., 1988) and a highly productive replication rate of $\sim 10^9$

© 2010 Elsevier B.V. All rights reserved.

*Corresponding Author: Tel.: 1 803 777 9186. Austin@biol.sc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

virions per day (Ho et al., 1995). A high mutation rate coupled with recombination events during replication (Temin, 1993), as well as selective pressures from host genes which mediate HIV-1 cellular entry and regulate the immune response (Michael, 1999), have contributed to the extensive genetic diversity characteristic of HIV-1 (Malim and Emerman, 2001).

Phylogenetic analyses of globally circulating strains of HIV-1 have identified 4 distinct groups: the major group M, of which more than 90% of HIV-1 infections belong; the outlier group O; and two new groups N and P (Plantier et al., 2009; Yang et al., 2008). Group M is further divided into 9 subtypes (A, B, C, D, F, G, H, J and K) that differ from each other by as much as 35% at the amino acid level (Gaschen et al., 2002). Subtypes A, B, C and D are highly prevalent, with the percentage of global prevalence being 12.3%, 10.2%, 49.9% and 2.5%, respectively (Hemelaar et al., 2006; Taylor et al., 2008). The most studied of the subtypes, Subtype B, predominates the Americas, Australia, Western and Central Europe, and is common in several countries of Southeast Asia, Northern Africa, and the Middle East (Buonaguro et al., 2007; Taylor et al., 2008).

The immense variation of HIV-1 has understandably hampered vaccine development (Gaschen et al., 2002) and also is believed to contribute to viral persistence (Kearney et al., 2009). Although immune control is rarely achieved (Brumme et al., 2008), cytotoxic T lymphocyte (CTL) responses are important in both the containment (Goulder and Watkins, 2004; O'Connor et al., 2010) and evolution of HIV (Klenerman et al., 2002). This is perhaps best demonstrated by selection of viral escape mutations within key CTL epitopes during both acute and chronic HIV-1 infection, which allows the virus to evade otherwise stable immunodominant CTL responses (Borrow et al., 1997; Fischer et al., 2010; Goulder et al., 1997; Miura et al., 2010; Price et al., 1997; Treurnicht et al., 2010). Specific HLA class I alleles have also been implicated in HIV disease progression (Carrington and O'Brien, 2003). For example, HIV-positive individuals that are homozygous at any of the three HLA class I loci, progress to AIDS more rapidly (Carrington and O'Brien, 2003).

In spite of evidence revealing selectively favored escape mutations in CTL epitopes of infected hosts, there is also evidence that in the absence of immune pressure these mutants may confer a substantial fitness cost (Carlson and Brumme, 2008). For example, one study investigated a particular epitope within the gag gene (epitope TW10) and through in vitro experiments and structural analysis was able to demonstrate that the dominant escape mutation reduced viral replicative capacity by interfering with helix stabilization (Martinez-Picado et al., 2006). Another study by Liu et al. (2007) found that within another gag defined epitope (EW10), an escape mutation substantially diminished viral fitness in in vitro competition assays and subsequently reverted to its original epitopic form. Furthermore, there is also evidence that suggests that mutations which allow escape from host immune responses may be HLA allele-specific, reverting to a wild-type sequence in the absence of the presenting HLA allele (Moore et al., 2002). For example, in HLA-B57/5801-positive subjects, the CTL response dominating acute infection drove positive selection of an escape mutation that reverted to the wild-type after transmission to HLA-B57/5801-negative individuals; therefore providing evidence of both CTL-mediated positive selection and virus-mediated purifying selection at the population level (Leslie et al., 2004).

Selection favoring escape mutations within a host expressing the presenting HLA molecule combined with purifying selection against that mutation in the population at large is expected to give rise to an evolutionary conflict (Irausquin and Hughes, 2008, 2010). In the case of hepatitis C virus, evidence of such a conflict is provided by the existence of abundant nonsynonymous variants in CTL epitopes that are subject to ongoing purifying selection (Irausquin and Hughes, 2008, 2010). Here, we analyze coding sequence

polymorphism in 8 coding genes of HIV-1 subtype B, the most prevalent subtype in the Americas, in order to test for evidence of conflicting evolutionary pressures on CTL epitopes. Because of evidence that CD4+ T helper cell responses are critical in the maintenance of virus-specific immunity in most chronic viral infections (Gerlach et al., 1999; Matloubian et al., 1994; Walter et al., 1995), we also test for evidence of conflicting evolutionary pressures on CD4 epitopes.

2. Methods

2.1. Sequences analyzed

We downloaded from the HIV Sequence Database (<http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html>) unaligned HIV-1 subtype B nucleotide sequences with accompanying information regarding the exact sampling year and HLA information regarding the host. Only sequences explicitly annotated as drug naive were included. Sequences resulting from recombination between subtypes were excluded from the analysis, as well as sequences containing premature stop codons. Nucleotide sequences were organized into the correct reading frame using Artemis (Rutherford et al., 2000) and subsequently aligned at the amino acid level in Mega4's Alignment Explorer (Tamura et al., 2007) which utilizes a built-in ClustalW implementation; this alignment was then imposed on the nucleotide sequences by flipping from the aligned protein sequences grid to the corresponding nucleotide sequences grid in the main Alignment Explorer window (Tamura et al., 2007). The numbers of useable sequences in the alignments for each gene were as follows: *gag* 164; *pol* 46; *vif* 71; *vpr* 197; *tat-exon1* 297; *rev-exon1* 297; *vpu* 197; *tat-exon2* 586; *rev-exon2* 599; *nef* 366 (for GenBank accession numbers see Supplementary Table S1). The *env* gene was excluded from analysis because of difficulty obtaining a reliable global alignment.

2.2. Sequence analysis

Using the MEGA4 program (Tamura et al., 2007), the number of synonymous substitutions per synonymous site (d_S) and the number of nonsynonymous substitutions per nonsynonymous site (d_N) were estimated for both CTL and CD4 epitope regions as well as non-epitope regions of each gene by Nei and Gojobori's (1986) method. In preliminary studies, these quantities were also estimated by more complex methods (Li, 1993; Yang and Nielsen, 2000) which, as expected, yielded essentially identical results because the number of substitutions per site was generally low (Hughes and French, 2007); therefore, the simpler Nei and Gojobori's method was utilized because it has a lower estimation of error (Nei and Kumar, 2000). The synonymous nucleotide diversity (symbolized π_S) is defined as the mean of d_S for all pairwise comparisons among a set of sequences, while the nonsynonymous nucleotide diversity (symbolized π_N) is the mean of d_N for all pairwise comparisons among a set of sequences.

Gene diversity was estimated at individual polymorphic synonymous and nonsynonymous nucleotide sites by the formula:

$$1 - \sum_{i=1}^n x_i^2$$

where n is the number of alleles and x_i is the frequency of the i^{th} allele (Nei, 1987, p. 177). Single nucleotide polymorphisms were classified as either synonymous or nonsynonymous, based on their effect of the encoded amino acid sequence. Ambiguous sites (sites at which both synonymous and nonsynonymous variants occurred, or at which the polymorphism

could be considered synonymous or nonsynonymous depending on the pathway taken by evolution) were excluded. Polymorphic sites were also classified as singleton sites, if a difference from the most abundant nucleotide was found in only one of the sequences analyzed. All polymorphic sites considered to be in overlapping regions were excluded from analyses. For all genes combined, 1992 (35.4%) of 5635 sites were polymorphic; of these, 863 (43.3%) were synonymous, 841 (42.2%) were nonsynonymous, and 288 (14.5%) were ambiguous.

Predicted epitopes presented by HLA class I MHC to CD8⁺ T cells and those presented by HLA class II to CD4⁺ T cells were designated as CTL and CD4 epitopes, respectively. We defined as CTL epitopes only those that were 100% conserved in at least one sequence following the Best-defined CTL/CD8⁺ Epitope Summary available from the HIV molecular immunology database

(http://www.hiv.lanl.gov/content/immunology/tables/optimal_ctl_summary.html; Llano et al., 2009; Supplementary Table S2). Similarly, we defined as CD4 epitopes only those that were 100% conserved in at least one sequence following the T-Helper/CD4⁺ Epitope Summary also available from the HIV molecular immunology database

(http://www.hiv.lanl.gov/content/immunology/tables/helper_summary.html; Supplementary Table S3). In comparing CTL and CD4 epitope and non-epitope regions only Subtype B epitopes from human with known HLA presenting antigens were included (Supplementary Tables S2 and S3). Although both CTL and CD4 epitopes presented by different alleles and loci were combined respectively for purposes of statistical analyses, it is important to mention that several amino residues formed part of epitopes presented by more than one HLA allelic product and sometimes by products of alleles at different HLA loci (Supplementary Tables S2 and S3). Also worth noting is that CTL epitopes sometimes overlapped epitopes presented by class II MHC molecules to CD4⁺ T cells (CD4 epitopes).

We used robust methods of statistical analysis in order to avoid the statistically undesirable properties of model-dependence (Hughes et al., 2006). Since gene diversity was not normally distributed at polymorphic sites, non-parametric methods were utilized for analysis (Hollander and Wolfe, 1973). So called codon-based methods of analysis were not used because of their dependence on several questionable assumptions, most particularly the unwarranted assumption that the existence of one or more codons with $d_N > d_S$ implies positive selection (Hughes, 2007).

3. Results

3.1. Synonymous and nonsynonymous polymorphism

The mean of π_S for the eight genes was significantly greater than the mean of π_N in CD4 epitopes, CTL epitopes, and non-epitope regions (Table 1). Therefore, purifying selection predominated overall in both CD4 and CTL epitopes, as well as in non-epitope regions. In order to test for ongoing purifying selection, we estimated gene diversity at individual polymorphic sites. Median gene diversity at synonymous polymorphic sites (0.0832, $N=863$) was significantly greater than that at nonsynonymous polymorphic sites (0.0359, $N=841$, Mann-Whitney test $P < 0.001$). When individual genes were analyzed separately, we found a similar pattern of median gene diversities at synonymous polymorphic sites exceeding those at nonsynonymous sites the majority of the time, although the differences for genes with this pattern were not always statistically significant (data not shown). These results support the hypothesis that there are abundant nonsynonymous variants which are subject to ongoing purifying selection, which acts to reduce gene diversity at these polymorphic nonsynonymous sites.

3.2. Polymorphic sites in epitopes

In order to compare patterns of polymorphism in CD4 epitopes to that in CTL epitopes, both synonymous and nonsynonymous polymorphic sites were categorized as follows: (i) not in either a CD4 or a CTL epitope; (ii) in a CD4 epitope only; (iii) in a CTL epitope only; and (iv) in both CD4 and CTL epitopes (Fig. 1). In the case of synonymous polymorphic sites, there was a highly significant difference ($P < 0.001$, Kruskal-Wallis test) in median gene diversity among the four categories (Fig. 1a). Median gene diversity was highest (0.0886) in sites in neither kind of epitope, followed by sites in CTL epitopes only (0.0593; Fig. 1a). Median gene diversity at nonsynonymous polymorphic sites was lowest in sites in CD4 epitopes only (0.0359) and next lowest (0.0476) in sites in both CD4 and CTL epitopes (Fig. 1a).

Similarly, there was a highly significant difference ($P < 0.001$, Kruskal-Wallis test) in median gene diversity among the four categories of polymorphic nonsynonymous sites (Fig. 1b). Median gene diversity at nonsynonymous polymorphic sites in CD4 epitopes only (0.0121) was identical to that at sites in both CD4 and CTL epitopes, and lower than median gene diversity at sites in neither kind of epitope (0.0425) or median gene diversity at sites in CTL epitopes only (0.0241; Fig. 1b). One factor contributing to low gene diversity at nonsynonymous sites in both CTL and CD4 epitopes was an excess of singleton sites. The proportion of singleton sites among nonsynonymous polymorphic sites in CTL epitopes only (104/217 or 47.9%) differed significantly from the proportion of singletons among other sites (250/624 or 40.1%; $\chi^2 = 4.08$, 1 d.f., $P < 0.05$). Likewise, the proportion of singleton sites among nonsynonymous polymorphic sites in CD4 epitopes only (98/190 or 51.6%) differed significantly from the proportion of singletons among other sites (256/651 or 39.3%; $\chi^2 = 9.06$, 1 d.f., $P = 0.003$).

Median gene diversity at synonymous polymorphic sites was significantly greater than that at nonsynonymous polymorphic sites in all four categories of sites: (1) sites in neither CTL nor CD4 epitopes; (2) sites in CTL epitopes only; (3) sites in CD4 epitopes only; and (4) sites in both CTL and CD4 epitopes ($P < 0.001$ in each case, Mann-Whitney test, Fig. 1). Therefore, comparisons between synonymous and nonsynonymous polymorphic sites among the four categories of sites also revealed evidence of ongoing purifying selection.

4. Discussion

Analysis of 8 HIV-1 subtype B coding genes revealed strong evidence of purifying selection on protein-coding regions, including those that encode CTL and CD4 epitopes. In comparisons between CTL epitope, CD4 epitope, and Non-Epitope regions, synonymous nucleotide diversity (π_S) significantly exceeded nonsynonymous nucleotide diversity (π_N) for all 8 genes within each of the three categories. Furthermore, gene diversity at nonsynonymous polymorphic sites was significantly lower than that at synonymous polymorphic sites; providing evidence that there are abundant slightly deleterious nonsynonymous variants subject to ongoing purifying selection.

In examining nucleotide sequence polymorphism among different categories of polymorphic sites, although consistent with the pattern of purifying selection, there were certain differences. In both synonymous and nonsynonymous polymorphic sites, median gene diversity was highest in non-epitope regions and lowest in sites in CD4 epitopes only (Fig. 1). These results identified polymorphic sites outside of epitopes as being subject to less stringent purifying selection, and conversely identified polymorphic sites located within CD4 epitopes only as being subject to the strongest purifying selection. Sites in both CTL and CD4 epitopes showed an excess of nonsynonymous singleton sites, consistent with strong ongoing purifying selection on these regions.

Studies suggesting an evolutionary conflict between immune escape and functional constraint in protein-coding genes of HIV have focused mainly on CTL epitopes, most of which are located within the Gag protein. The reason for this focus is probably that Gag-specific CTL responses have been strongly associated with effective control of HIV infection (Edwards et al., 2002; Zuniga et al., 2006). For example, two *in vivo* studies described reversion of escape mutations in two Gag B*57 epitopes (1) escape mutation T242N in epitope TSTLQEQIAW (Gag₂₄₀₋₂₄₉, TW10; Leslie et al., 2004) and (2) escape mutation A163X (X=G, N, D, or S) in epitope KAFSPEVIPMF (Gag₁₆₂₋₁₇₂, KF11; Crawford et al., 2007), following transmission to MHC-mismatched recipients (Prado et al., 2009). Moreover, another study demonstrated that the R264K escape mutation in epitope KRWILGLNK (Gag₇₈₇₋₈₁₆, KK10) dramatically compromised *in vitro* viral replication capacity (Schneidewind et al., 2007). Interestingly, the sites corresponding to all three of these escape mutations were not polymorphic in our data set. An escape mutation (E75D) within the ETINEEAAEW epitope (Gag₇₁₋₈₀, EW10) that substantially diminishes viral fitness in *in vitro* competition assays and subsequently reverted to its original epitopic form (Liu et al., 2007), was classified as an ambiguous polymorphic site in our dataset and thus not included in our analyses. Although the EW10 epitope was not defined in our study, since it is not considered a Best-defined CTL/CD8+ epitope, the polymorphic site fell within a defined CD4 epitope (ETINEEAAEWDRVHPVHA, Gag 71-88). Gene diversity at this site in our data set was 0.0478, substantially greater than median gene diversity in either CTL or CD4 epitopes, perhaps reflecting the occurrence of escape mutations at this site.

One factor contributing to the observed strong sequence conservation of CTL and CD4 epitopes might be that, soon after HIV-1 began to infect humans, escape mutations quickly became fixed in all epitopes where such mutations were not deleterious to the virus (Moore et al., 2002). Thus, the remaining CTL and CD4 epitopes would be those at which most amino acid replacements are deleterious to the virus. Consistent with this hypothesis are data showing numerous forward and backward amino acid changes in CTL epitopes of HIV-1 (Moore et al., 2002; Piontkivska and Hughes, 2004, 2006). The repeated occurrence of the same amino acid substitutions at numerous points in a phylogeny is consistent with a failure to fix that substitution, which in turn suggests the presence of selection against that substitution at the population level.

Our results support the hypothesis that many nonsynonymous mutations in both CTL and CD4 epitopes of HIV-1 subtype B are subjected to conflicting evolutionary pressures: positive selection favoring escape mutations within host expressing the respective presenting HLA molecule and purifying selection acting to remove them in the population at large. The fact that sites in CD4 epitopes are more constrained than sites in CTL epitopes may be explained by the fact that in the majority of HIV-1 infected individuals, the CD4+ T helper cell response are functionally impaired in all stages of disease (Rosenberg et al., 1997) despite evidence that these cellular responses are critical in the maintenance of virus-specific immunity in most chronic viral infections (Gerlach et al., 1999; Matloubian et al., 1994; Walter et al., 1995). One notable exception is the observation of robust T helper cell responses in persons with long-term non-progressive infection (Rosenberg et al., 1997). However, in most HIV-1 affected individuals, the absence of a strong CD4+ T helper cell response is likely to reduce the selective pressure favoring escape in these regions, and thus to alleviate the evolutionary conflict between escape and amino acid sequence conservation. In future studies, the hypothesis should be tested further, using population data from other HIV-1 subtypes, particularly subtypes A and C. Pyrosequencing dataset may also prove beneficial in further testing the hypothesis, because they can provide evidence on the evolutionary dynamics of epitopes within hosts (Hughes et al. 2010).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by grant GM43940 from the National Institutes of Health to A.L.H.

References

- Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dautet C, Axler-Blin C, Vézinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*. 1983; 220:868–871. [PubMed: 6189183]
- Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, Meyers H, Nelson JA, Gairin JE, Hahn BH, Oldstone MB, Shaw GM. Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. *Nat Med*. 1997; 3:205–11. [PubMed: 9018240]
- Brumme ZL, Brumme CJ, Carlson J, Streeck H, John M, Eichbaum Q, Block BL, Baker Y, Carrington M, Allen TM, Mallal S, Altfeld M, Heckerman D, Walker BD. Marked epitope- and allele-specific differences in rates of mutation in human immunodeficiency type 1 (HIV-1) Gag, Pol, and Nef cytotoxic T-lymphocyte epitopes in acute/early HIV-1 infection. *J Virol*. 2008; 82:9216–27. [PubMed: 18614631]
- Buonaguro L, Tornesello ML, Buonaguro FM. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *J Virol*. 2007; 81:10209–19. Review. [PubMed: 17634242]
- Carlson JM, Brumme ZL. HIV evolution in response to HLA-restricted CTL selection pressures: a population-based perspective. *Microbes Infect*. 2008; 10:455–61. [PubMed: 18407775]
- Carrington M, O'Brien SJ. The influence of HLA genotype on AIDS. *Annu Rev Med*. 2003; 54:535–51. Review. [PubMed: 12525683]
- Crawford H, Prado JG, Leslie A, Hué S, Honeyborne I, Reddy S, van der Stok M, Mncube Z, Brander C, Rousseau C, Mullins JJ, Kaslow R, Goepfert P, Allen S, Hunter E, Mulenga J, Kiepiela P, Walker BD, Goulder PJ. Compensatory mutation partially restores fitness and delays reversion of escape mutation within the immunodominant HLA-B*5703-restricted Gag epitope in chronic human immunodeficiency virus type 1 infection. *J Virol*. 2007; 81:8346–51. [PubMed: 17507468]
- Edwards BH, Bansal A, Sabbaj S, Bakari J, Mulligan MJ, Goepfert PA. Magnitude of functional CD8+ T-cell responses to the gag protein of human immunodeficiency virus type 1 correlates inversely with viral load in plasma. *J Virol*. 2002; 76:2298–305. [PubMed: 11836408]
- Fischer W, Ganusov VV, Giorgi EE, Hraber PT, Keele BF, Leitner T, Han CS, Gleasner CD, Green L, Lo CC, Nag A, Wallstrom TC, Wang S, McMichael AJ, Haynes BF, Hahn BH, Perelson AS, Borrow P, Shaw GM, Bhattacharya T, Korber BT. Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. *PLoS One*. 2010; 5:e12303. [PubMed: 20808830]
- Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, Kalyanaraman VS, Mann D, Sidhu GD, Stahl RE, Zolla-Pazner S, Leibowitch J, Popovic M. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science*. 1983; 220:865–867. [PubMed: 6601823]
- Gaschen B, Taylor J, Yusim K, Foley B, Gao F, Lang D, Novitsky V, Haynes B, Hahn BH, Bhattacharya T, Korber B. Diversity considerations in HIV-1 vaccine selection. *Science*. 2002; 296:2354–60. Review. [PubMed: 12089434]
- Gerlach JT, Diepolder HM, Jung MC, Gruener NH, Schraut WW, Zachoval R, Hoffmann R, Schirren CA, Santantonio T, Pape GR. Recurrence of hepatitis C virus after loss of virus-specific CD4(+) T-cell response in acute hepatitis C. *Gastroenterology*. 1999; 117:933–41. [PubMed: 10500077]
- Goulder PJ, Phillips RE, Colbert RA, McAdam S, Ogg G, Nowak MA, Giangrande P, Luzzi G, Morgan B, Edwards A, McMichael AJ, Rowland-Jones S. Late escape from an immunodominant

- cytotoxic T-lymphocyte response associated with progression to AIDS. *Nat Med.* 1997; 3:212–7. [PubMed: 9018241]
- Goulder PJ, Watkins DI. HIV and SIV CTL escape: implications for vaccine design. *Nat Rev Immunol.* 2004; 4:630–40. Review. [PubMed: 15286729]
- Hemelaar J, Gouws E, Ghys PD, Osmanov S. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS.* 2006; 20:W13–23. [PubMed: 17053344]
- Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature.* 1995; 373:123–6. [PubMed: 7816094]
- Hollander, M.; Wolfe, DA. Nonparametric statistical methods. Wiley & Sons; New York: 1973.
- Hughes AL. Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity.* 2007; 99:364–373. [PubMed: 17622265]
- Hughes AL, French JO. Homologous recombination and the pattern of nucleotide substitution in *Ehrlichia ruminantium*. *Gene.* 2007; 387:31–37. [PubMed: 17005333]
- Hughes AL, Friedman R, Glenn NL. The future of data analysis in evolutionary genomics. *Curr Genomics.* 2006; 7:227–234.
- Hughes AL, O'Connor S, Dudley DM, Burwitz BJ, Bimber BN, O'Connor D. Dynamics of haplotype frequency change in a CD8+TL epitope of simian immunodeficiency virus. *Infect Genet Evol.* 2010:555–560. [PubMed: 20149896]
- ICTVdB Management. Human immunodeficiency virus 1. In: Büchen-Osmond, C., editor. ICTVdB - The Universal Virus Database, version 4. Columbia University; New York, USA: 2006. 00.061.1.06.009
- Irausquin SJ, Hughes AL. Distinctive pattern of sequence polymorphism in the NS3 protein of hepatitis C virus type 1b reflects conflicting evolutionary pressures. *J Gen Virol.* 2008; 89:1921–1929. [PubMed: 18632963]
- Irausquin SJ, Hughes AL. Conflicting selection pressures target the NS3 protein in hepatitis C virus genotypes 1a and 1b. *Virus Research.* 2010; 147:202–207. [PubMed: 19896990]
- Kearney M, Maldarelli F, Shao W, Margolick JB, Daar ES, Mellors JW, Rao V, Coffin JM, Palmer S. Human immunodeficiency virus type 1 population genetics and adaptation in newly infected individuals. *J Virol.* 2009; 83:2715–27. [PubMed: 19116249]
- Klenerman P, Wu Y, Phillips R. HIV: current opinion in escapology. *Curr Opin Microbiol.* 2002; 5:408–13. [PubMed: 12160861]
- Leslie AJ, Pfafferott KJ, Chetty P, Draenert R, Addo MM, Feeney M, Tang Y, Holmes EC, Allen T, Prado JG, Altfeld M, Brander C, Dixon C, Ramduth D, Jeena P, Thomas SA, St John A, Roach TA, Kupfer B, Luzzi G, Edwards A, Taylor G, Lyall H, Tudor-Williams G, Novelli V, Martinez-Picado J, Kiepiela P, Walker BD, Goulder PJ. HIV evolution: CTL escape mutation and reversion after transmission. *Nat Med.* 2004; 10:282–9. [PubMed: 14770175]
- Li WH. Unbiased estimates of the rates of synonymous and nonsynonymous substitution. *J Mol Evol.* 1993; 36:96–99. [PubMed: 8433381]
- Liu Y, McNeven J, Zhao H, Tebit DM, Troyer RM, McSweyn M, Ghosh AK, Shriner D, Arts EJ, McElrath MJ, Mullins JI. Evolution of human immunodeficiency virus type 1 cytotoxic T-lymphocyte epitopes: fitness-balanced escape. *J Virol.* 2007; 81:12179–88. [PubMed: 17728222]
- Llano A, Frahm N, Brander C. How to Optimally Define Optimal Cytotoxic T Lymphocyte Epitopes in HIV Infection? *HIV Molecular Immunology.* 2009; 2009:3–24.
- Malim MH, Emerman M. HIV-1 sequence variation: drift, shift, and attenuation. *Cell.* 2001; 104:469–72. Review. [PubMed: 11239404]
- Mansky LM, Temin HM. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol.* 1995; 69:5087–94. [PubMed: 7541846]
- Martinez-Picado J, Prado JG, Fry EE, Pfafferott K, Leslie A, Chetty S, Thobakgale C, Honeyborne I, Crawford H, Matthews P, Pillay T, Rousseau C, Mullins JI, Brander C, Walker BD, Stuart DI, Kiepiela P, Goulder P. Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. *J Virol.* 2006; 80:3617–23. [PubMed: 16537629]
- Matloubian M, Concepcion RJ, Ahmed R. CD4+ T cells are required to sustain CD8+ cytotoxic T-cell responses during chronic viral infection. *J Virol.* 1994; 68:8056–63. [PubMed: 7966595]

- Michael NL. Host genetic influences on HIV-1 pathogenesis. *Curr Opin Immunol.* 1999; 11:466–74. Review. [PubMed: 10448150]
- Miura T, Brumme ZL, Brockman MA, Rosato P, Sela J, Brumme CJ, Pereyra F, Kaufmann DE, Trocha A, Block BL, Daar ES, Connick E, Jessen H, Kelleher AD, Rosenberg E, Markowitz M, Schafer K, Vaida F, Iwamoto A, Little S, Walker BD. Impaired replication capacity of acute/early viruses in persons who become HIV controllers. *J Virol.* 2010; 84:7581–91. [PubMed: 20504921]
- Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science.* 2002; 296:1439–43. [PubMed: 12029127]
- Nei, M. *Molecular evolutionary genetics.* Columbia University Press; New York: 1987.
- Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* 1986; 3:418–426. [PubMed: 3444411]
- Nei, M.; Kumar, S. *Molecular Evolution and Phylogenetics.* Oxford University Press; New York: 2000.
- O'Connor SL, Lhost JJ, Becker EA, Detmer AM, Johnson RC, Macnair CE, Wiseman RW, Karl JA, Greene JM, Burwitz BJ, Bimber BN, Lank SM, Tuscher JJ, Mee ET, Rose NJ, Desrosiers RC, Hughes AL, Friedrich TC, Carrington M, O'Connor DH. MHC heterozygote advantage in simian immunodeficiency virus-infected mauritian cynomolgus macaques. *Sci Transl Med.* 2010; 2:22ra18.
- Piontkivska H, Hughes AL. Between-host evolution of CTL epitopes in Human Immunodeficiency Virus Type 1 (HIV-1): an approach based on phylogenetically independent comparisons. *J Virol.* 2004; 78:11758–11765. [PubMed: 15479817]
- Piontkivska H, Hughes AL. Patterns of sequence evolution at epitopes for host antibodies and cytotoxic T-lymphocytes in Human Immunodeficiency Virus Type 1. *Virus Research.* 2006; 116:98–105. [PubMed: 16214253]
- Plantier JC, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemée V, Damond F, Robertson DL, Simon F. A new human immunodeficiency virus derived from gorillas. *Nat Med.* 2009; 15:871–2. [PubMed: 19648927]
- Prado JG, Honeyborne I, Brierley I, Puertas MC, Martinez-Picado J, Goulder PJ. Functional consequences of human immunodeficiency virus escape from an HLA-B*13-restricted CD8+ T-cell epitope in p1 Gag protein. *J Virol.* 2009; 83:1018–25. [PubMed: 18945768]
- Price DA, Goulder PJ, Klennerman P, Sewell AK, Easterbrook PJ, Troop M, Bangham CR, Phillips RE. Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. *Proc Natl Acad Sci U S A.* 1997; 94:1890–5. [PubMed: 9050875]
- Roberts JD, Bebenek K, Kunkel TA. The accuracy of reverse transcriptase from HIV-1. *Science.* 1988; 242:1171–3. [PubMed: 2460925]
- Rosenberg ES, Billingsley JM, Caliendo AM, Boswell SL, Sax PE, Kalams SA, Walker BD. Vigorous HIV-1-specific CD4+ T cell responses associated with control of viremia. *Science.* 1997; 278:1447–50. [PubMed: 9367954]
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. Artemis: sequence visualization and annotation. *Bioinformatics.* 2000; 16:944–945. [PubMed: 11120685]
- Schneidewind A, Brockman MA, Yang R, Adam RI, Li B, Le Gall S, Rinaldo CR, Craggs SL, Allgaier RL, Power KA, Kuntzen T, Tung CS, LaBute MX, Mueller SM, Harrer T, McMichael AJ, Goulder PJ, Aiken C, Brander C, Kelleher AD, Allen TM. Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication. *J Virol.* 2007; 81:12382–93. [PubMed: 17804494]
- Sierra S, Kupfer B, Kaiser R. Basics of the virology of HIV-1 and its replication. *J Clin Virol.* 2005; 34:233–44. Review. [PubMed: 16198625]
- Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol.* 2007; 24:1596–1599. [PubMed: 17488738]
- Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM. The challenge of HIV-1 subtype diversity. *N Engl J Med.* 2008; 358:1590–602. Review. [PubMed: 18403767]

- Temin HM. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proc Natl Acad Sci U S A*. 1993; 90:6900–3. Review. [PubMed: 7688465]
- Treurnicht FK, Seoighe C, Martin DP, Wood N, Abrahams MR, de Assis Rosa D, Bredell H, Woodman Z, Hide W, Mlisana K, Karim SA, Gray CM, Williamson C. Adaptive changes in HIV-1 subtype C proteins during early infection are driven by changes in HLA-associated immune pressure. *Virology*. 2010; 396:213–25. [PubMed: 19913270]
- Walter EA, Greenberg PD, Gilbert MJ, Finch RJ, Watanabe KS, Thomas ED, Riddell SR. Reconstitution of cellular immunity against cytomegalovirus in recipients of allogeneic bone marrow by transfer of T-cell clones from the donor. *N Engl J Med*. 1995; 333:1038–44. [PubMed: 7675046]
- World Health Organization. [10 June 10] AIDS epidemic update. 2009. [Internet]. Published [09 November 30] Available at: <http://www.who.int/hiv/pub/epidemiology/epidemic/en/index.html>
- Yang X, Yang H, Zhou G, Zhao GP. Infectious disease in the genomic era. *Annu Rev Genomics Hum Genet*. 2008; 9:21–48. Review. [PubMed: 18767959]
- Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol*. 2000; 17:32–43. [PubMed: 10666704]
- Zuñiga R, Lucchetti A, Galvan P, Sanchez S, Sanchez C, Hernandez A, Sanchez H, Frahm N, Linde CH, Hewitt HS, Hildebrand W, Altfeld M, Allen TM, Walker BD, Korber BT, Leitner T, Sanchez J, Brander C. Relative dominance of Gag p24-specific cytotoxic T lymphocytes is associated with human immunodeficiency virus control. *J Virol*. 2006; 80:3122–5. [PubMed: 16501126]

Fig. 1.

Median gene diversity in non-epitope (NonEp) and in CD4 epitope and CTL epitope regions of (A) polymorphic synonymous and (B) polymorphic nonsynonymous sites for all genes. Numbers of sites in each category are shown. For both synonymous and nonsynonymous polymorphic sites, median gene diversity differed significantly among categories of sites (Kruskal-Wallis test, $P < 0.001$). Mann-Whitney tests of the hypothesis that median gene diversity in a given category for synonymous polymorphic sites equals the corresponding value for nonsynonymous polymorphic sites: * $P < 0.001$, ** $P < 0.01$.

Table 1

Synonymous (π_S) and non-synonymous (π_N) nucleotide diversity ($\pm$ S.E) in CD4 epitope, CTL epitope, and Non-Epitope regions of 8 HIV-1 subtype B genes

Gene	Number of Sequences	CD4		CTL		Non-Epitope	
		$\pi_S \pm SE$	$\pi_N \pm SE$	$\pi_S \pm SE$	$\pi_N \pm SE$	$\pi_S \pm SE$	$\pi_N \pm SE$
gag	164	0.1298 $\pm$ 0.0191	0.0267 $\pm$ 0.0055	0.1553 $\pm$ 0.0233	0.0233 $\pm$ 0.0046	0.1781 $\pm$ 0.0207	0.0418 $\pm$ 0.0054
pol	46	0.1151 $\pm$ 0.0509	0.0026 $\pm$ 0.0017	0.1345 $\pm$ 0.0214	0.0163 $\pm$ 0.0036	0.1425 $\pm$ 0.0082	0.0246 $\pm$ 0.0021
vif	71	N/A	N/A	0.1499 $\pm$ 0.1052	0.0402 $\pm$ 0.0233	0.1226 $\pm$ 0.0199	0.0432 $\pm$ 0.0069
vpr	197	0.0784 $\pm$ 0.0163	0.0317 $\pm$ 0.0120	0.0703 $\pm$ 0.0329	0.0453 $\pm$ 0.0284	0.0578 $\pm$ 0.0245	0.0195 $\pm$ 0.0053
tat-ex1	297	0.0911 $\pm$ 0.0512	0.0359 $\pm$ 0.0148	0.0596 $\pm$ 0.0545	0.0617 $\pm$ 0.0470	0.0894 $\pm$ 0.0513	0.0683 $\pm$ 0.0254
rev-ex1	297	N/A	N/A	0.4460 $\pm$ 0.1498	0.1196 $\pm$ 0.0543	0.1425 $\pm$ 0.0952	0.0225 $\pm$ 0.0149
vpu	197	N/A	N/A	N/A	N/A	0.0739 $\pm$ 0.0216	0.0405 $\pm$ 0.0073
tat-ex2	586	N/A	N/A	N/A	N/A	0.0830 $\pm$ 0.0387	0.0907 $\pm$ 0.0283
rev-ex2	599	N/A	N/A	0.1000 $\pm$ 0.0636	0.0548 $\pm$ 0.0329	0.1314 $\pm$ 0.0296	0.0595 $\pm$ 0.0157
nef	366	0.2136 $\pm$ 0.0686	0.0788 $\pm$ 0.0212	0.1745 $\pm$ 0.0390	0.0405 $\pm$ 0.0102	0.1483 $\pm$ 0.0316	0.0813 $\pm$ 0.0129
Overall mean		0.1256 $\pm$ 0.0238	0.0351 $\pm$ 0.0123 [*]	0.1613 $\pm$ 0.0432	0.0502 $\pm$ 0.0112 ^{**}	0.1169 $\pm$ 0.0123	0.0492 $\pm$ 0.0079 [*]

Paired t-tests rejected the hypothesis that overall mean π_S = overall mean π_N for CD4 epitope, CTL epitope, and Non-Epitope regions of each gene.

^{*} $P < 0.01$;

^{**} $P < 0.05$.