

Distinct patterns of natural selection in the reverse transcriptase gene of HIV-1 in the presence and absence of antiretroviral therapy

Élcio de S. Leal,^a Edward C. Holmes,^b and Paolo M. de A. Zanotto^{a,*}

^aLaboratory of Molecular Evolution and Bioinformatics-LEMB, Department of Microbiology, University of São Paulo, São Paulo, CEP 05508-900, Brazil

^bDepartment of Zoology, University of Oxford, Oxford, OX1 3PS, UK

Received 10 December 2003; returned to author for revision 21 January 2004; accepted 1 April 2004

Available online 11 June 2004

Abstract

The emergence of human immunodeficiency virus (HIV) drug-resistant mutations during antiretroviral therapy is explained by either the preexistence of low-frequency-resistant strains before the start of therapy or by the selection of unsuppressed resistant strains during therapy. We used pairwise and maximum likelihood analyses of the ratio of nonsynonymous to synonymous substitutions per site (d_N/d_S) to study the extent of positive selection in the reverse transcriptase (RT) gene of HIV-1 from multiple data sets of drug-treated (117 sequences) and drug-naïve patients (270 sequences). In the pairwise analysis, evidence for positive selection ($d_N/d_S > 1$) was only found in drug-treated individuals and in codons conferring drug resistance. By the maximum likelihood method, a positive selection at codons conferring drug resistance was only observed in patients receiving therapy, and although positive selection was detected in drug-naïve patients, this was always at codons unrelated to drug resistance. We therefore document a striking difference in the process of allele fixation in drug resistance codons (RC) between populations of HIV-1-infected individuals naïve to treatment and those receiving therapy. Furthermore, although mutations associated with drug resistance are sometimes found in drug-naïve patients, we suggest that these are fixed because of linkage to sites experiencing immune escape. Finally, we show that compensatory changes are likely to be important in the development of HIV drug resistance by counter-acting the deleterious effects normally associated with drug resistance mutations.

© 2004 Elsevier Inc. All rights reserved.

Keywords: HIV-1; Positive selection; Antiretroviral resistance; Compensatory mutations; Drug-naïve individuals; Population genetics

Introduction

The reverse transcriptase (RT) enzyme of the human immunodeficiency virus type 1 (HIV-1) converts viral genomic single-stranded RNA into double-stranded DNA. This is an essential step in the viral life cycle such that the RT has been a target of antiretroviral therapy. Even when antiretroviral drugs drastically reduce HIV-1 replication and viral RNA is undetectable in plasma, viral suppression is not complete (Chun et al., 1997; Fleury et al., 2000; Lewin et al., 1999). Because HIV-1 has high rates of

mutation, replication, and recombination, and immense population sizes (Bebenek et al., 1989; Jung et al., 2002; Mansky and Temin, 1995), intra-host viral populations may show considerable genetic diversity. In particular, the high error rate of the RT, estimated at approximately 10^{-4} mutations per site, per replication (Bebenek et al., 1989; Mansky and Temin, 1995), means that in a viral genome of length 10 kb, approximately one mutation is introduced on average at each replication cycle during the synthesis of viral DNA. Because many of these mutations are deleterious to the virus, the rate at which mutations are fixed (substituted) in the population is approximately 2.3×10^{-5} substitutions per site, per replication (Perelson et al., 1996).

At any time, a large fraction of an intra-host HIV population represents the most adapted viral genome under the current selective pressures, such as those imposed by the human immune response or by drug therapy. Nevertheless,

* Corresponding author. Laboratory of Molecular Evolution and Bioinformatics-LEMB, Department of Microbiology, University of São Paulo, Av. Lineu Prestes, 1730, São Paulo, CEP 05508-900, Brazil. Fax: +55-11-3091-7290.

E-mail address: pzanotto@usp.br (P.M. de A. Zanotto).

the viral population as a whole may harbor a myriad of mutational variants, including those that confer drug resistance even though antiviral therapy has not been instigated (Blower et al., 2000; Gomez-Cano et al., 1998; Larder et al., 1989; Moore et al., 1991; Nájera et al., 1994; Ribeiro and Bonhoeffer, 2000; Richman et al., 1991). In many cases, drug resistance mutations are likely to be naturally deleterious to HIV in the absence of drug, but their fitness benefit during therapy is such that they are eventually fixed in the population (Larder et al., 1991; Nájera et al., 1994). Therefore, the presence of normally deleterious changes in a HIV population could reflect the transient presence of alleles that will eventually be removed by purifying selection, the action of genetic drift fixing such mutations when population sizes are small (Frost et al., 2001; Holmes and Zlotoff, 1998; Leigh Brown, 1997; Leigh Brown and Richman, 1997; Rouzine and Coffin, 1999a, 1999b), or the occurrence of compensatory changes that raise the fitness of viruses carrying normally deleterious drug resistance alleles to values near the wild type (Ribeiro and Bonhoeffer, 2000; Rouzine and Coffin, 1999a). However, despite the large body of work considering the dynamics of allele substitution during antiviral therapy, the respective roles of natural selection, genetic drift, and compensatory changes are unclear although central to predicting the future spread of drug resistance mutations. Furthermore, it is uncertain whether those drug resistance mutations that sometimes appear in drug-naïve patients are subject to the same evolutionary processes that determine their spread in drug-treated individuals.

Herein, we describe the selection pressures acting on the RT gene from many patients who are either drug-naïve or drug-treated by comparing the ratio of nonsynonymous (d_N) to synonymous (d_S) substitutions per site, with a d_N/d_S ratio >1 indicating a positive selection. To make our analyses of selection pressures as rigorous as possible, we employ both the pairwise method of Kumar et al. (2001) and a codon-based maximum likelihood method that incorporates genealogical information from the sequences under consideration (Yang et al., 2000). We show that evolutionary processes in the RT differ considerably between drug-naïve and drug-treated patients.

Results

HIV subtyping

To establish the HIV-1 subtype of the São Paulo (SP) patients, we estimated a phylogeny of their RT sequences along with a panel of subtype reference sequences from the Los Alamos HIV database (tree not shown; available from the authors upon request). Using this analysis, we found that 73 represented subtype B (82.5%) and 12 represented subtype F (14.1%).

Drug resistance mutations in drug-naïve patients

Initially, we screened for amino acid changes associated with resistance to antiretroviral drugs in drug-naïve individuals. This required the analysis of two data sets—São Paulo (SP) and Boden et al. (1999). Several mutations known to confer drug-resistance were observed in the SP data set (Table 2). Specifically, 10 of the 85 RT sequences from the SP data set harbored mutations associated with resistance to non-nucleoside reverse transcriptase inhibitors (NNRTIs), and one patient contained a substitution (M184I) associated with resistance to nucleoside reverse transcriptase inhibitors (NRTI). This latter mutation is commonly seen after 2 weeks of monotherapy with lamivudine (3TC) and drastically reduces the processivity of the RT (Back et al., 1996). Conversely, in the Boden et al. data set, drug resistance mutations were more frequent (26/80) and equally represented both classes of drugs. As well as mutations associated with drug resistance, both data sets of drug-naïve patients showed a variety of other variable codons, most notably positions 211 and 214.

Pairwise analysis of selection pressures

The pairwise analysis of d_N/d_S provided some preliminary indication of differences in selection pressure between drug-treated and drug-naïve patients as mean d_N/d_S was higher in 117 drug-treated individuals (1.061) than in 270 drug-naïve individuals (0.564), although in both cases the null hypothesis of neutral evolution ($d_N/d_S = 1$) could not be rejected (Table 1).

Table 1

Pairwise analysis of d_N and d_S in subgroups of drug resistance codons (RC) and other codons (OC) from seven data sets of drug-treated and 15 data sets of drug-naïve individuals

Subgroups of codons					
RC (32 codons)			OC (194 codons)		
d_N (SE)	d_S (SE)	d_N/d_S (t test)	d_N (SE)	d_S (SE)	d_N/d_S (t test)
<i>Average values in 15 drug-naïve data sets</i>					
1.450 (0.442)	2.656 (0.600)	0.570 ^a ($t = 1.618$, $P < 0.05$)	9.983 (1.750)	15.826 (1.628)	0.627 ^a ($t = 2.445$, $P < 0.02$)
<i>Average values in seven drug-treated data sets</i>					
4.302 (1.438)	2.32 (0.738)	1.948 ($t = 1.227$, $P > 0.2$)	8.287 (1.788)	12.583 (1.884)	0.698 ^a ($t = 1.654$, $P < 0.05$)

^a Denotes d_N/d_S significantly different from 1 at the 95% confidence interval with a t test, where $t = (d_S - d_N) / \sqrt{[SE(d_S^2) + SE(d_N^2)]}$ and infinite degrees of freedom.

To determine whether the elevated d_N/d_S in drug-treated patients was associated with positive selection at codons related to antiretroviral resistance, we subdivided the RT gene into: (i) 32 drug-resistance codons (RC) and (ii) the other 194 codons (OC) that remained after exclusion of the drug-related codons. The average values of d_N/d_S for RC and OC subgroups from both drug-treated and drug-naïve

data sets are shown in Table 1. These results are striking in that d_N/d_S was consistently >1 for the RC codons (mean = 1.948) of the drug-treated patients and consistently <1 for the OC codons (mean = 0.698), although the null hypothesis of neutral evolution of the RC codons again could not be rejected ($P > 0.2$). Despite a lack of statistical significance, perhaps explained by the relatively small sample sizes, these

Table 2

Positively selected codons in the RT gene in drug-naïve individuals identified using a maximum likelihood method

Data set (no. of sequences)	Drug-resistance codons found ^a	Proportion of codons with $d_N/d_S \leq 0.25^b$	Proportion of codons with $d_N/d_S > 1^c$	Positively selected codons ^c (no. of haplotypes) ^d
<i>Drug-naïve individuals</i>				
São Paulo B (73)	T69N, A98G, K101E, K103R, V108I, M184I	0.826	0.040	122 (3), 162 (9), 177 (3), 200 (1), 211 (6)
São Paulo F (12)	None	0.917	0.033	None
São Paulo-Res (10)	T69N, A98G, K101E, K103R, V108I, M184I	0.962	0.038	None
São Paulo-Poly (12)	T69N	0.881	0	None
Los Alamos-B (21)	K103R	0.810	0	None
Los Alamos-C (8)	V179D	0.908	0.092	39 (3), 43 (5), 135 (4), 162 (9), 166 (10), 207 (5)
Stanford 1994 (20)	K103R	0.918	0.083	200 (1), 211 (6)
Stanford 1995 (12)	K103R, V179D	0.914	0.086	125 (3)
Stanford 1997 (10)	None	0.932	0.068	35 (2), 122 (3), 123 (3), 135 (4), 207 (5), 211 (6)
Stanford 1998 (16)	None	0.896	0.104	35 (2), 122 (3), 123 (3), 125 (3), 162 (9), 178 (3), 200 (1), 211 (6)
Stanford combined (58) ^e	K103R, V179D	0.856	0.064	35 (2), 122 (3), 123 (3), 135 (4), 177 (3), 200 (1), 207 (5), 211 (6), 214 (1)
Boden 1999 combined (80) ^f	M41L, D67N, T69D, K70R, V75M/L, K103N, V179D/E, M184V, L210W, T215F	0.864	0.062	25 (2), 66 (0), 113 (2), 125 (3), 152 (1), 197 (1), 201 (5), 205 (5)
Boden-Res (25)	M41L, D67N, T69D, K70R, V75M/L, K103N, V179D/E, M184V, L210W, T215F	0.762	0.004	None
Boden-Poly (30)	None	0.849	0.047	25 (2), 73 (1), 113 (2), 125 (3), 152 (1), 169 (1), 197 (1), 201 (5)
Brindero 1999 (18)	M41L, D67N, K70R, T215F, K219Q	0.836	0.052	None
<i>Drug-treated individuals</i>				
Loussert and Ajaka (9)	M41L, D67N, K70R, W88S ^g , A98G, V106I, Y181C, M184V, L210W, T215Y	0.809	0.019	None
Hooker and Smith (7)	M41L, V75I, A98S, K103N, Q151M, M184V, L210W, T215Y	0.809	0.191	<u>151</u> (1), 202 (5), 204 (5), <u>215</u> (1)
Eron et al. (21)	M41L, D67N, K70R, K101E, M184V, G190A, L210W, T215F/Y	0.900	0.100	<u>70</u> (0), 83 (0), 122 (3), 177 (3)
Gunthard et al. (24)	M41L, D67N, M184V, L210W, T215Y, K219Q	0.748	0.023	None
Schmit et al. (44)	M41L, D67N, T69D/S, K70R, L74V, V75I, A98G, Y115F, E138A/G, Q151M, M184V/I, L210W, T215Y/F, K219E	0.715	0.116	35 (2), <u>67</u> (0), <u>69</u> (0), <u>70</u> (0), <u>118</u> (3), <u>151</u> (1), 207 (5), <u>210</u> (6), 211 (6), 214 (1), <u>215</u> (1)
Winter et al. (5)	M41L, K70R, T69A/S, Y181C, L210W, T215Y	0.906	0.093	68 (0), 122 (3)
Stanford 1998 (7)	M41L, D67N, L74I, M184V, L210W, T215Y	0.787	0	None
All-treated combined ($n = 117$)		0.948	0.051	35 (2), <u>41</u> (5), <u>67</u> (0), <u>69</u> (0), <u>70</u> (0), <u>118</u> (3), 122 (3), 123 (3), 135 (4), <u>151</u> (1), <u>184</u> (4), 207 (5), <u>210</u> (6), <u>215</u> (1)

^a Drug-resistance codons are those listed at the Los Alamos HIV-1 drug resistance database. Positively selected codons associated with drug resistance are underlined.

^b The proportion of sites with $d_N/d_S \leq 0.25$ was estimated using model M8 of CODEML (Yang et al., 2000).

^c Only positively selected codons with statistical support of $P > 0.99$ under model M8 are listed.

^d The number of different HLA-haplotypes that recognize epitopes where the codon is located.

^e “Stanford combined” includes all 58 sequences from 1994, 1995, 1997, and 1998.

^f “Boden combined” includes all 80 sequences from Boden et al. (1999).

^g This mutation is associated with resistance to pyrophosphate analogues.

results are compatible with the idea that positive selection is directed against drug resistance codons in individuals receiving antiviral therapy. Conversely, in drug-naïve individuals, mean d_N/d_S was significantly <1 for both RC (0.570, $P < 0.05$) and OC (0.627, $P < 0.02$) codons, revealing the action of purifying selection.

Maximum likelihood analysis of selection pressures

To obtain a more refined measure of the codon-specific selection pressures acting on the RT sequences, we performed a maximum likelihood analysis of d_N/d_S in both the drug-naïve and drug-treated patient data sets (Table 2). The results for 15 drug-naïve groups indicated that although significant positive selection ($P < 0.001$ for both M3 and M8) was detectable at a variety of codons, none were associated with drug resistance. The lack of drug associated positive selection is particularly notable given that drug resistance substitutions for both NRTIs and NNRTIs drugs were found in most drug-naïve data sets (Table 2). Moreover, with the exception of codon 66, all these selected codons were located within CTL epitopes (Table 2 and Fig. 1).

A very different picture of selection pressures was observed in those patients who have received drug therapy. Positive selection was detected at various codons in four of the seven drug-treated data sets (Table 2). Most notably, a variety of these codons are associated with the development of drug resistance (Table 2 underlined). However, not all mutations known to confer drug resistance were positively selected, and many of the selected codons were unrelated to antiviral resistance. Because most of the drug-treated data sets had a few sequences, we repeated the analysis on a combined data set of the 117 sequences from the drug-treated patients (“All-Treated Combined”). This identified additional codons under positive selection at sites 41, 123, 135, 184, 207, and 210 (Table 2). Although the functional

consequences of positive selection at codons not related to drug resistance are unclear, the rest fell into known CTL or T-helper epitopes with the exception of codon 83 (Fig. 2) (Korber et al., 2001). Codon 83 was under positive selection in a drug-treated data set (“Eron et al” Table 2) and was neither associated with drug resistance nor inside CTL epitopes, thus suggesting that it may have a compensatory function to recover fitness of drug-resistant viruses.

Our analysis also revealed a strong association between positive selection and the presence of CTL epitopes. Indeed, positively selected codons in drug-naïve patients were significantly more likely to fall into regions containing CTL epitopes than other genomic regions ($\chi^2 = 5.5217$, $P < 0.0189$). In contrast, there was no difference in the location of positively selected codons in drug-treated patients with respect to the location of CTL epitopes ($\chi^2 = 0.0507$, $P < 0.8198$) (Table 3). Some precise examples of the links between drug resistance and CTL escape were also apparent. For example, the drug-resistance mutation at position 210 and the polymorphisms at positions 201, 205, 207, and 211 are all located within the HLA-B60-CTL (B*4001/B60) restricted epitope (Korber et al., 2001; McAdam and Gotch, 1999; Wainberg, 1999). Likewise, positions 35, 122, 123, 135, 177, and 211, which were selected in both drug-treated and drug-naïve individuals, are in regions with high density of epitopes (Fig. 1). Therefore, perhaps most of the positively selected codons not related with drug resistance function as targets to the immune system, as suggested in recent large-scale studies (Moore et al., 2002; Yusim et al., 2002).

In addition, we found evidence for linkage between polymorphic codons (under positive selection) and drug-related substitutions (not under positive selection) in both drug-naïve and drug-treated patients. For example, codon 177 (under positive selection in drug-naïve patients) and the drug-related substitution at position 179 (not under positive selection) are both located within the HLA-B*3501-CTL

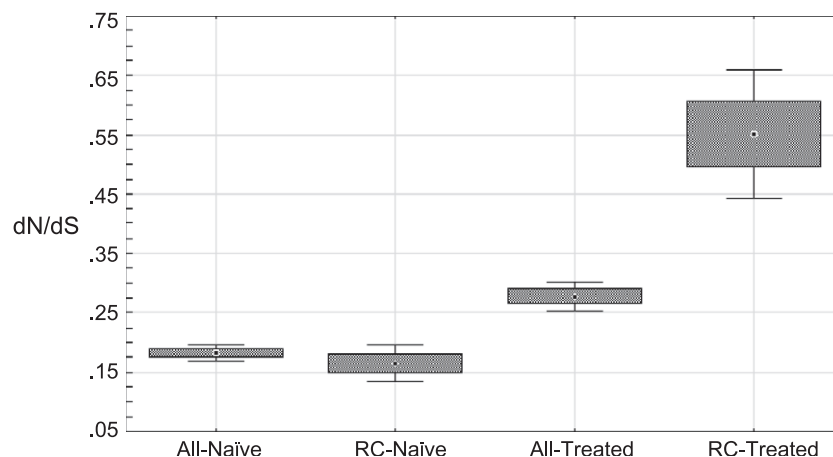


Fig. 1. Overall mean d_N/d_S ratios in RT codons estimated under model M8 in CODEML. “All-Treated” = mean d_N/d_S in all codons of the seven data sets of drug-treated patients. “All-Naïve” = mean d_N/d_S in all codons of the 15 data sets of drug-naïve patients. “RC-Treated” = mean d_N/d_S in drug-resistance codons of the seven data sets of drug-treated patients. “RC-Naïve” = mean d_N/d_S in drug-resistance codons of the 15 data sets of drug-naïve patients.

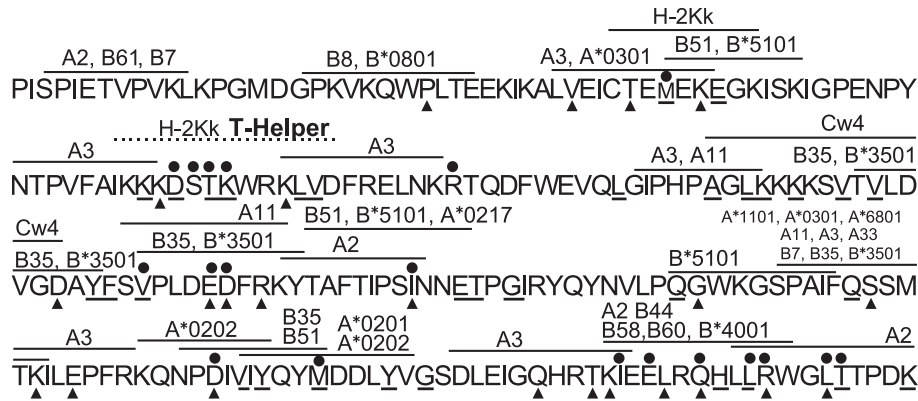


Fig. 2. Association between positively selected codons and CTL epitopes in the RT gene of HIV-1. The RT sequence corresponds to sites 1–219 of the HXB2 strain. Dots above particular residues denote the positively selected codons in drug-treated patients, while triangles below residues represent positively selected codons in drug-naïve patients. All drug-resistance codons are underlined and CTL epitopes experimentally determined according to a panel of MHC-I haplotypes are highlighted. Figure modified from HIV Molecular Immunology data base (Korber et al., 2001).

restricted epitope (RT 175–183). Moreover, in drug-treated patients, the drug-associated mutations at codon 179 (not under positive selection) and 184 (under positive selection) are located within the HLA-B51-CTL restricted epitope (RT 174–184) (Fig. 1).

Despite widespread evidence for positive selection, the dominant evolutionary processes in the RT from both patient groups were purifying selection; in most data sets, irrespective of treatment, the majority of codons (71.5–96.2%) have d_N/d_S values ≤ 0.25 , indicative of fairly strong structural–functional constraints (Table 2). Similarly, mean d_N/d_S values (estimated under the maximum likelihood method) showed a pattern of constraint, although the resistance codons in drug-treated patients (“RC-Treated”) have elevated values of d_N/d_S (Fig. 2).

Positive selection before and after monotherapy with didanosine (ddI)

To better understand the evolutionary effects of antiretroviral therapy, we analyzed selection pressures in 49 RT

sequences of the same group of patients before and after (1 year) monotherapy with didanosine (ddI). The maximum likelihood provided no significant evidence for positive selection at any codon, before or after therapy. Nevertheless, before therapy, all patients had leucine (L) at codon 74 while during therapy 16 of 49 patients acquired valine (V) at this site. The mutation 74V is directly associated with ddI resistance.

Phylogenetic analysis of mutations in drug-naïve and drug-treated populations

The presence of drug resistance mutations in drug-naïve patients is intriguing, especially as we found no evidence that they have been subjected to positive selection. To investigate their evolutionary history in more detail, we conducted a phylogenetic analysis of the RT sequences from the drug-naïve SP and Boden et al. data sets.

The topology of the SP tree is notable in that the sequences containing drug-resistant mutations or that are polymorphic at resistant codons are usually evenly distributed among the terminal branches (Fig. 3). This suggests that most of these mutations have occurred independently and have not been transmitted for sustained periods of time. However, a few exceptions were found in sequences that share either drug-resistance mutations or polymorphisms. In these cases, it is important to determine if the resistance mutations were transmitted together or have converged within individual hosts. In the case of sequences SP72 and SP80, the former has drug resistance mutations at position M184I and latter at position G190E, and they share the same polymorphisms (M41I and T69P) at drug-related codons, although none of these mutations were under positive selection. However, as SP72 and SP80 do not cluster together, these mutations were most likely acquired through convergent evolution, a theory supported by the observation that they are both charac-

Table 3
Association between positively selected sites and CTL epitopes^a

Drug-naïve	Positively selected sites (triangles in Fig. 2)	Other codons
Inside-epitope	22	144
Outside-epitope	1	52
$\chi^2 = 5.5217$, $df = 1$, $P < 0.0189$		
Drug-treated	Positively selected sites (dots in Fig. 2)	Other codons
Inside-epitope	14	152
Outside-epitope	5	48
$\chi^2 = 0.0507$, $df = 1$, $P < 0.8198$		

^a This test is based on the location of CTL epitopes presented in Fig. 2 and compares the significance of finding positive selected sites inside and outside CTL epitopes. The test is only significant in drug-naïve patients.

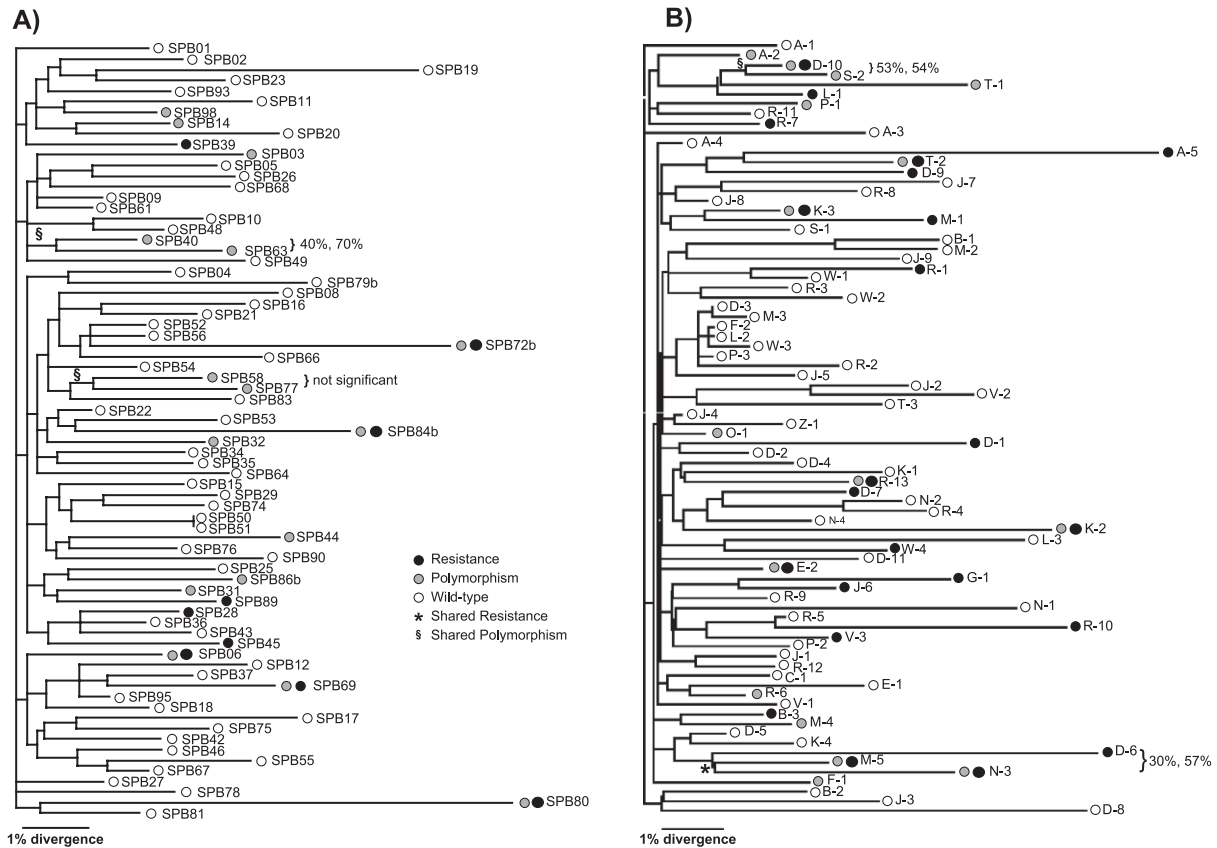


Fig. 3. Maximum likelihood phylogenetic trees for HIV-1 RT genes of drug-naïve individuals from the (A) São Paulo and (B) Boden et al. (1999) data sets. Lineages bearing resistance mutations are shown by closed circles, polymorphisms at drug resistance codons by shaded circles, and wild-type mutations by white circles, using HXB2 as a reference strain. Sister lineages having the same drug resistance mutations (denoted *) and sister lineages having the same polymorphisms in drug-resistance codons (denoted \$) are also indicated. To indicate the reliability of our phylogenetic reconstructions, the values from 500 bootstrap replicates done with the neighbor-joining method in PAUP* and the percent of the time a node was supported by the quartet-puzzling method (TREE-PUZZLE, <http://www.tree-puzzle.de/>) (Strimmer and von Haeseler, 1996) are shown, respectively, near the branches for important nodes only. All trees are unrooted with horizontal branch lengths drawn to scale.

terized by long branches. In contrast, sequences SP40 and SP63 share the drug-resistance mutation A98G and the polymorphism at R211K, the latter being selected in several data sets, and group together on the tree. The genetic distance between these two sequences ($d = 0.034$) implies that they diverged at least 5 years before sampling (assuming a substitution rate of 2.3×10^{-5} substitutions per site, per year), so that this mutation has persisted in the population for at least 1403 viral generations in the absence of continuous drug pressure (under an average viral generation time of 2.6 days). Likewise, sequences M5 and N3 from the Boden data set share the drug-resistance mutation V75L ($d = 0.057$) and diverged at least 9 years before sampling, while sequences D10 and S2 share the V179I polymorphism ($d = 0.019$) and diverged at least 3 years before sampling. If these drug-related mutations were indeed in the common ancestor of each sequence pair, then their persistence in the population despite their deleterious effects in the absence of drug suggests that compensatory mutations may be required to recover fitness.

Discussion

Natural selection and drug resistance

There is compelling evidence that the positive selection of advantageous mutations is a key process in HIV evolution, both for drug resistance (Crandall et al., 1999; Frost et al., 2001) and immune escape (Ross and Rodrigo, 2002; Williamson, 2003; Yang et al., 2000; Zanotto et al., 1999). Despite this, some authors have suggested that allele fixation in HIV populations is often a stochastic process, with genetic drift the dominant force (Frost et al., 2000; Leigh Brown and Richman, 1997). This is due to a complex interaction of factors, including small and fluctuating effective population sizes, changes in cellular conditions (target cells), nonuniform distribution of viruses, variation in host immune responses (namely levels of CD8⁺ and CD4⁺), differing combinations of antiretroviral drugs, and co-infection with multiple strains (Ribeiro and Bonhoeffer, 2000; Rouzine and Coffin, 1999a, 1999b).

Although it is important to remember that all alleles, whatever their fitness, are subject to drift when they first appear and are at low frequency in viral populations, our study strongly supports the idea that the evolution of drug resistance in HIV is a selectively driven process. In particular, the only evidence for positive selection at known resistance codons was found in patients undergoing antiviral therapy. A similar result was recently obtained in a study of RT sequences from African drug-naïve individuals infected with subtype C viruses (Gordon et al., 2003). Consequently, we conclude that drug resistance mutations in drug-naïve and drug-treated patients are subject to different evolutionary processes, with positive selection predominant in the latter.

Of particular note was our observation that drug resistance mutations in drug-naïve patients have not been fixed as a result of their individual fitness advantage. Moreover, many of these mutations appear to be deleterious for the virus in the absence of drug pressure. For example, M184I and G190E (found in sequences SP72 and SP80) are known to cause a drastic loss of viral fitness in the absence of drug (Back et al., 1996; Boyer et al., 1998; Fan et al., 1996). In particular, M184I is located at the YMDD catalytic motif of the polymerase active site of RT and confers high level of resistance to 3TC (lamivudine). However, viruses with this mutation have lower fitness relative to wild-type viruses and are frequently replaced by a fitter 3TC-resistant mutant (M184V). Likewise, G190E causes a drastic reduction in catalytic and RNaseH activities, and polymorphisms at codons 74 and 75 have been shown to compensate for the effects of G190E on RT processivity (Boyer et al., 1998). The inviability of these mutants in the absence of drug implies that compensatory mutations rescue the fitness of the mutant to near that of wild type. In this context, it is noteworthy that the majority of drug-resistance codons are invariant in most of the HIV subtype sequences in the Los Alamos database. This implies that these sites are highly conserved because of structural–functional constraints so that any mutations at these sites will normally be deleterious. This is confirmed by our observation that most RT codons are under strong purifying selection.

How then to explain the long-term persistence of deleterious drug-related mutations in the absence of drug? In some cases, it is possible that we have by chance sampled deleterious mutations shortly before their selective removal from the population. Although this may be true for a subset of the resistance mutations, particularly those that fall on the terminal branches of the phylogenetic tree, suggesting that they appeared recently, it is unlikely to explain every occurrence of a drug resistance mutation in a drug-naïve patient. Another possibility is that their spread is entirely due to the random sampling associated with genetic drift. However, we consider this scenario to be unlikely because of the extended time it takes genetic drift to fix neutral mutations; if we conservatively assume an effective population size (N_e) of 10^4 , genetic drift would take an average

of 42 years (i.e., $2N_e$ viral generations) to fix these mutations. Consequently, even though these drug resistance mutations may not have achieved complete fixation, it is difficult to imagine how their presence can be entirely due to random sampling. As such, their prolonged presence in the viral population again suggests that associated mutations play a compensatory role to recover viral fitness. The importance of compensatory mutations has previously been observed in relation to CTL escape in HIV. In this case, a key CTL escape mutation (R264K) in p17 *gag* from HLA-B27+ patients could only spread to fixation if accompanied by one or two nearby compensatory mutations that restored the function of the viral capsid (Kelleher et al., 2001).

If drug-resistant mutations in drug-naïve patients generally have low individual fitness such that compensatory mutations are essential, then their presence in the population is most likely due to linkage with mutations and that have been selected because they confer immune escape. Indeed, most resistance codons fall within known CTL epitopes (Korber et al., 2001), and in the drug-naïve data sets, significant evidence for positive selection was always found in non-resistance codons, many of which reside in CTL epitopes. This idea is also supported by the presence of positively selected codons and drug-related mutations within the same CTL epitopes. In these circumstances, the strength of both linkage and selection pressure for CTL escape would need to be strong enough to counteract the affect of recombination, which may be extremely frequent within HIV populations (Jung et al., 2002). This is clearly a model that needs to be explored further. Indeed, to fully understand the importance of immune selection and compensatory changes in the evolution of drug resistance mutations in drug-naïve patients will require longitudinal studies of the substitution process of mutant HIV alleles.

Antiretroviral therapy and HIV evolution at the population level

Because HIV replication may not be completely suppressed under antiretroviral therapy, even with good adherence (Chun et al., 1997; Fleury et al., 2000), it has been proposed that drug-resistant strains will attain high frequencies in infected populations, as has been observed (Angarano et al., 1994; Boden et al., 1999; Conlon et al., 1994; Perrin et al., 1994; Salomon et al., 2000; Yerly et al., 1999). Our observation of an increase in frequency of L74V in the presence of ddI may be such a case of transmission of drug-resistant strains at the population level. Conversely, the observation that most amino acid sites associated with resistance are conserved in the absence of drug therapy suggests that they are generally deleterious and may require compensatory changes to recover fitness (Richman et al., 1991). Therefore, it can be argued that in the absence of compensatory changes and drug pressure, most resistant strains are not able to spread far in drug-naïve individuals unless they are by chance linked to immune system escape

mutants. Furthermore, natural selection might be a relatively weak force at the population level in HIV, so that mutants with high fitness within hosts may not attain high frequencies. The relative weakness of selection as a determinant of inter-host evolution is due to several factors; that rates of partner exchange vary enormously between hosts so that genetic drift dominates substitution dynamics, that the large population bottleneck at transmission counters the inheritance of favorable mutations, and that many advantageous mutations, such as those involved in drug resistance or CTL escape, arise relatively late during the life history of an individual HIV infection and so are unlikely to be transmitted to new hosts (Rambaut et al., 2004). Consequently, mutations that are selectively favored within individuals may only have limited reproductive success at the population level.

In sum, we have shown that the evolution of drug-resistant mutations differs dramatically between drug-treated and drug-naïve populations, with the former dominated by positive selection and the latter most likely spreading in linkage with immune-selected changes. We also suggest that because many HIV drug resistance mutations are naturally deleterious, compensatory changes are critical to understanding their evolutionary dynamics.

Methods

Samples

The patients enrolled in this study were monitored at the Infectious Diseases Discipline (DIPA) of the Federal University of São Paulo (UNIFESP) and at the AIDS Clinics of the Department/Division of Infectious Diseases, School of Medicine of the University of São Paulo. Patients were classified as “drug-naïve” and selected according to the following criteria: (i) HIV-1 positive (as determined by ELISA and Western blotting), (ii) no previous treatment with any antiretroviral drug, and (iii) at least 18 years of age. Patients under any other kind of drug regimen were also excluded from the study. Based on the clinical records (data not shown), most of the individuals enrolled in this study were infected through heterosexual contact and were asymptomatic when sampled.

Viral isolation and sequencing

DNA was extracted from whole blood frozen samples using QIAamp viral DNA kit (Qiagen Inc., Chatsworth, CA) according to the manufacture’s instructions. PCR reactions were made using a nested approach to generate fragments of the RT gene of HIV-1. The following primers were used in the amplifications. For the first round of amplification, we used the Kozal-1 primer (Kozal et al., 1996), 5′-CAG AGC CAA CAG CCC CAC CA-3′ (nucleotide position of HIV-1 HXB2: 2146–2165); and Kozal-2,

5′-TTT CCC CAC TAA CTT CTG TAT GTC ATT GAC A-3′ (3307–3337). For the second round, we used both the Frenkel-1 primer (Frenkel et al., 1995), 5′-GTT GAC TCA GAT TGG TTG CAC-3′ (2518–2538); and Frenkel-2, 5′-GTA TGT CAT TGA CAG TCC AGC-3′ (3300–3320).

PCR reactions had a final volume of 50 l, containing 10 mM Tris–HCl, pH 8.0, 50 mM KCl, 3.5 mM MgCl₂, 0.2 mM of each dNTP 0.2 M of each primer and 2.5 U of Taq DNA polymerase (Gibco BRL). First and second round reactions were incubated at 94 °C for 10 min, then 30 cycles of 94 °C for 30 s; 58 °C for 45 s; 72 °C for 1–2 min, and a final extension of 72 °C for 10 min. All reactions were performed on a Perkin-Elmer thermal cycler (model GeneAmp 2400). Second round PCR products were inspected by electrophoresis on 1.5% agarose gels stained with ethidium bromide using 0.5× TBE buffer.

Bulk PCR amplicons (second round) were purified with the QIAquick PCR purification kit (Qiagen Inc.) and directly sequenced. Cycle sequencing was done on both strands with the same primers used in the second round of the PCR using the Big Dye Terminator Cycle Sequencing Ready Reaction kit (Perkin-Elmer Applied Biosystems) and AmpliTaq DNA polymerase, according with the manufacturers instructions. Electrophoresis of the sequencing reactions was done with an automated capillary sequencer (ABI PRISM 310 Perkin-Elmer Applied Biosystems). Nucleic acid sequences were edited utilizing Sequence Navigator software (Applied Biosystems). The electropherograms of both strands were aligned and a consensus sequence was obtained. Consensus sequences were translated and edited manually, resulting in a data set containing 85 nucleic acid sequences of RT gene with length of 678 bp (226 codons). These sequences have been deposited on GenBank (accession numbers AF480743–AF480832).

Drug-naïve and drug-treated data sets

For comparative purposes, sequences of the RT gene from other data sets of both drug-naïve or drug-treated HIV-infected patients were obtained from the Los Alamos HIV data base (<http://hiv-web.lanl.gov>), the Stanford HIV data base (<http://hivdb.stanford.edu/hiv/>), GenBank, or directly from the authors of previous studies. These included 18 sequences from Rio de Janeiro (Brindeiro et al., 1999), 80 sequences from United States (Boden et al., 1999), and a further 58 taken from the Stanford and Los Alamos databases. This last data set included representatives from other subtypes to allow HIV subtype assignment. Overall, we collected 15 data sets of drug-naïve patients and seven data sets of drug-treated patients, and where possible the sequences were analyzed according to the year of sampling. As shown in Table 2, the data sets were also designated according to their component subtypes when necessary (e.g., SP-F, Los Alamos C). Finally, we analyzed 49 sequences sampled from patients before and after (1 year) therapy with didanosine (ddl).

Sequence alignment and phylogenetic inference

Sequences were aligned using the ClustalW package (Thompson et al., 1994) and the SE-AL program (Rambaut, 1999). All insertions, deletions, and stop codons were removed before analysis. The subtype classification of the São Paulo RT sequences was established by comparing our sequences to those from the HIV subtype references through maximum likelihood (ML) phylogenetic analysis using the PAUP* package (Swofford, 2002) (all parameter values available from the authors upon request). To estimate the divergence time (T) between any two HIV sequences, we assumed a viral generation time (g) of 2.6 days and a nucleotide substitution rate (r) of 2.3×10^{-5} per generation (Perelson et al., 1996). Using corrected pairwise distances (d) between sequences, we can then estimate T by the relation $T = d/2r$.

For the maximum likelihood analysis of selection pressures (see below), phylogenetic trees are required for each data set. We first determined the most appropriate model of nucleotide substitution for each data set using the program Modeltest v3.06 (Posada and Crandall, 1998). A starting tree under this model was then obtained through the quartet-puzzling method available in PAUP*, and these trees were then swapped using the nearest-neighbor interchange search algorithm until the tree of highest likelihood was obtained.

Analysis of selection pressures

We used both pairwise and codon-specific methods to estimate selection pressures. Initially, we compiled data sets of all the RT sequences from drug-naïve patients ($n = 270$) and drug-treated patients ($n = 117$) and determined the mean d_N/d_S in these two groups using the pairwise method of Kumar et al. (2001) as implemented in the MEGA v2.1 package. Next, we estimated d_N/d_S separately for each of the 15 data sets of drug-naïve patients and seven data sets of drug-treated patients. Rather than estimating mean d_N/d_S values across the whole sequence, we analyzed separately codons associated with drug resistance (denoted RC) and other codons not associated with drug resistance (denoted OC). The RC subgroup was composed of the following 32 codons, according to the Stanford HIV database: M41, E44, K65, D67, T69, K70, L74, V75, L92, A98, L100, K101, K103, V106, V108, Y115, V118, E138, T139, G141, Q151, Q161, V179, Y181, M184, Y187, V188, G190, H208, L210, T215, and K219. We also used a maximum likelihood method to determine codon-specific selection pressures, as implemented in the CODEML program from the PAML v. 3.14 package (Yang et al., 2000). This compares the fit to the data of various models of codon evolution, which differ in the distribution of d_N/d_S among sites and take into account the phylogenetic relationships of the sequences. Model 0 (M0) assumes a single d_N/d_S for all sites in the alignment and hence is the simplest model specified. The “neutral” model (M1) allows different proportions of con-

served sites ($d_N/d_S = 0$) and neutral sites ($d_N/d_S = 1$), both estimated from the data. M2 adds an additional class of sites with its d_N/d_S ratio (which can be >1) estimated from the data. M3 also allows positive selection by incorporating three categories of codon sites with the d_N/d_S value at each estimated from the data. M7 specifies 10 categories of d_N/d_S , none of which may be <1 , so the model only allows neutral evolution. Finally, M8 incorporates an extra (11th) class of sites to M7 that can take on any value of d_N/d_S , including those supporting positive selection. Nested models can be compared using a standard likelihood ratio test (LRT). Significant evidence for positive selection is provided if M2, or more normally M3, significantly rejects M0 and M1, if M8 significantly rejects M7, and if the favored models contain a class of codons where $d_N/d_S > 1$. If positive selection is detected, individual codons subject to positive selection can be determined using the Bayesian methods available in CODEML.

Acknowledgments

ESL was funded by FAPESP grant 00/04830-8. PMAZ has a PQ scholarship from CNPq and funded by FAPESP grant 00/04205-6. This research was partially supported by Bristol-Myers Squibb.

References

- Angarano, G., Monno, L., Appice, A., Gianelli, A., Romanelli, C., Fico, C., Pastore, G., 1994. Transmission of zidovudine-resistant HIV-1 through heterosexual contact. *AIDS* 8, 1013–1014.
- Back, N.K.T., Nijhuis, M., Keulen, W., Boucher, C.A.B., Essink, B.B.O., van Kuilenburg, A.B.P., van Gennip, A.H., Berkhout, B., 1996. Reduced replication of 3TC-resistant HIV-1 variants in primary cells due to a processivity defect of the reverse transcriptase enzyme. *EMBO J.* 15, 4040–4049.
- Bebenek, K., Abbots, J., Roberts, J.D., Wilson, S.H., Kunkel, T.A., 1989. Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase. *J. Biol. Chem.* 264, 16948–16956.
- Blower, S.M., Gershengorn, H.B., Grant, R.M., 2000. A tale of two futures: HIV and antiretroviral therapy in San Francisco. *Science* 287, 650–654.
- Boden, D., Hurley, A., Zhang, L., Cao, Y., Guo, Y., Jones, E., Tsay, J., Ip, J., Farthing, C., Limoli, K., Parkin, N., Markowitz, M., 1999. HIV-1 drug resistance in newly infected individuals. *JAMA* 282, 1135–1141.
- Boyer, P.L., Gao, H.-Q., Hughes, S.H., 1998. A mutation at position 190 of human immunodeficiency virus type 1 reverse transcriptase interacts with mutations at positions 74 and 75 via template primer. *Antimicrob. Agents Chemother.* 42, 447–452.
- Brindeiro, R., Vanderborght, B., Caride, E., Correa, L., Oravec, R.M., Berro, O., Stuyver, L., Tanuri, A., 1999. Sequence diversity of the reverse transcriptase of human immunodeficiency virus type 1 from untreated Brazilian individuals. *Antimicrob. Agents Chemother.* 43, 1674–1680.
- Chun, T.W., Stuyver, L., Mizel, S., Ehler, L., Mican, J., Baseler, M., Lloyd, A., Nowak, M., Fauci, A., 1997. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc. Natl. Acad. Sci. U.S.A.* 94, 13193–13197.
- Conlon, C.P., Klennerman, P., Edwards, A., Larder, B.A., Phillips, R.E., 1994. Heterosexual transmission of human immunodeficiency virus

- type 1 variants associated with zidovudine resistance. *J. Infect. Dis.* 169, 411–415.
- Crandall, K.A., Kelsey, C.R., Imamich, H., Clifford Lane, H., Salzman, N.P., 1999. Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detected selection. *Mol. Biol. Evol.* 16, 371–382.
- Fan, N., Rank, K.B., Slade, D.E., Poppe, S.M., Evans, D.B., Kopta, L.A., Olmsted, R.A., Thomas, R.C., Tarpley, W.G., Sharma, S.K., 1996. A drug resistance mutation in the inhibitor binding pocket of human immunodeficiency virus type 1 reverse transcriptase impairs DNA synthesis and RNA degradation. *Biochemistry* 35, 9737–9745.
- Fleury, S., Rizzardi, G.P., Chapuis, A., Tambussi, G., Knabenhans, C., Simeoni, E., Meuwly, J.Y., Corpataux, J.-M., Lazzarin, A., Miedema, F., Pantaleo, G., 2000. Long-term kinetics of T cell production in HIV-infected subjects treated with highly active antiretroviral therapy. *Proc. Natl. Acad. Sci. U.S.A.* 97, 5393–5398.
- Frenkel, L.M., Wagner II, L.E., Atwood, S.M., Cummins, T.J., Dewhurst, S., 1995. Specific, sensitive, and rapid assay for human immunodeficiency virus type 1 pol mutations associated with resistance to zidovudine and didanosine. *J. Clin. Microbiol.* 33, 342–347.
- Frost, S.D.W., Nijhuis, M., Schuurman, R., Boucher, C.A.B., Leigh Brown, A.J., 2000. Evolution of lamivudine resistance in human immunodeficiency virus type 1-infected individuals: the relative roles of drift and selection. *J. Virol.* 74, 6262–6268.
- Frost, S.D.W., Günthard, H.F., Wong, J.K., Havlir, D., Richman, D.D., Leigh Brown, A.J., 2001. Evidence for positive selection driving the evolution of HIV-1 env under potent therapy. *Virology* 284, 250–258.
- Gomez-Cano, M., Rubio, A., Puig, T., Perez-Olmeda, M., Ruiz, L., Leal, M., Clotet, B., Soriano, V., 1998. Prevalence of drug resistance mutations over time in naive HIV-1 positive subjects living in Spain. *AIDS* 12, 1015–1020.
- Gordon, M., De Oliveira, T., Bishop, K., Coovadia, H.M., Madurai, L., Engelbrecht, S., Janse van Rensburg, E., Mosam, A., Smith, A., Cassol, S., 2003. Molecular characteristics of human immunodeficiency virus type 1 subtype C viruses from KwaZulu-Natal, South Africa: implications for vaccine and antiretroviral control strategies. *J. Virol.* 77, 2597–2599.
- Holmes, E.C., Zotto, P.M. de A., 1998. Genetic drift of human immunodeficiency virus type 1? *J. Virol.* 72, 886–887.
- Jung, A., Maier, R., Vartanian, J.P., Bocharov, G., Jung, V., Fischer, U., Meese, E., Wain-Hobson, S., Meyerhans, A., 2002. Recombination—Multiply infected spleen cells in HIV patients. *Nature* 418, 144.
- Kelleher, A.D., Long, C., Holmes, E.C., Allen, R.L., Wilson, J., Conlon, C., Workman, C., Shaunak, S., Wulfestieg, K., Goulder, P., Brander, C., Ogg, G., Sullivan, J.S., Dyer, W., Jones, I., McMichael, A.J., Rowland-Jones, S., Philips, R.E., 2001. Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27 restricted CTL responses. *J. Exp. Med.* 193, 375–385.
- Korber, B., Brander, C., Haynes, B., Koup, R., Kuiken, C., Moore, J., Walker, B., Watkins, D., 2001. HIV Molecular Immunology Database. Los Alamos, NM, <http://hiv-web.lanl.gov>.
- Kozal, M.J., Shah, N., Shen, N., Yang, R., Fucini, R., Merigan, T.C., Richman, D.D., Morris, D., Hubbel, E., Chee, M., Gingeras, T.R., 1996. Extensive polymorphism observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. *Nat. Med.* 2, 753–759.
- Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17, 1244–1245.
- Larder, B.A., Darby, G., Richman, D.D., 1989. HIV with reduced sensitivity to zidovudine isolated during prolonged therapy. *Science* 243, 1731–1734.
- Larder, B.A., Coates, K.E., Kemp, S.D., 1991. Zidovudine-resistant human immunodeficiency virus selected by passage in cell culture. *J. Virol.* 65, 5232–5236.
- Leigh Brown, A.J., 1997. Analysis of HIV-1 env gene sequences reveals evidence for a low effective number in the viral population. *Proc. Natl. Acad. Sci. U.S.A.* 94, 1862–1865.
- Leigh Brown, A.J., Richman, D.D., 1997. HIV-1: gambling on the evolution of drug resistance? *Nat. Med.* 3, 268–271.
- Lewin, S.R., Vesonen, M., Kostrikis, L., Hourley, A., Duran, M., Zhang, D.D., Ho, D.D., Markowitz, M., 1999. Use of real-time PCR and molecular beacons to detect virus replication in human immunodeficiency virus type 1-infected individuals on prolonged effective antiretroviral therapy. *J. Virol.* 73, 6099–6103.
- Mansky, L.M., Temin, H.M., 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J. Virol.* 69, 5087–5094.
- McAdam, S., Gotch, F., 1999. The cytotoxic T lymphocytes response to the immunodeficiency viruses. In: Dalgleish, A., Weiss, R. (Eds.), *HIV and the New Viruses*. Academic Press, London, pp. 75–87.
- Moore, R.D., Hidalgo, J., Sugland, B.W., Chaisson, R.E., 1991. Zidovudine and the natural history of the acquired immunodeficiency syndrome. *N. Engl. J. Med.* 324, 1412.
- Moore, C.B., John, M., James, I.R., Christiansen, F.T., Witt, C.S., Mallal, S.A., 2002. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* 296, 1439–1443.
- Nájera, I., Richman, D.D., Olivares, I., Roja, J.M., Peinado, M.A., Peruchó, M., Nájera, R., Lopez-Galindez, C., 1994. Natural occurrence of drug resistance mutations in the reverse transcriptase of human immunodeficiency virus type 1 isolate. *AIDS Res. Hum. Retroviruses* 10, 1479–1488.
- Perelson, A.S., Neumann, A.U., Markowitz, M., Leonard, J.M., Ho, D.D., 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell lifespan, and viral generation time. *Science* 271, 1582–1586.
- Perrin, L., Yerly, S., Rakik, A., Kinlock, S., Hirschel, B., 1994. Transmission of 215 mutants in primary HIV infection and analysis after 6 months of zidovudine. *AIDS* 8, S3.
- Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817–818.
- Rambaut, A., 1999. Sequence Alignment Editor (SE-AL) Software Version 1.0 Alpha 1. Department of Zoology, University of Oxford. <http://evolve.zoo.ox.ac.uk/>.
- Rambaut, A., Crandall, K.A., Posada, D., Holmes, E.C., 2004. The causes and consequences of HIV evolution. *Nat. Rev. Genet.* 5, 52–61.
- Ribeiro, R.M., Bonhoeffer, S., 2000. Production of resistant HIV mutants during antiretroviral therapy. *Proc. Natl. Acad. Sci. U.S.A.* 97, 7681–7686.
- Richman, D., Shih, C.K., Lowy, I., Rose, J., Prodanovich, P., Goff, S., Griffin, J., 1991. Human immunodeficiency virus type 1 mutants resistant to nonnucleoside inhibitors of reverse transcriptase arise in tissue culture. *Proc. Natl. Acad. Sci. U.S.A.* 88, 11241–11245.
- Ross, H.A., Rodrigo, A., 2002. Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. *J. Virol.* 76, 11715–11719.
- Rouzine, I., Coffin, J.M., 1999a. Interplay between experimental and theory in development of a working model for HIV-1 population dynamics. In: Domingo, E., Webster, R., Holland, J. (Eds.), *Origin and Evolution of Viruses*. Academic Press, London, pp. 225–262.
- Rouzine, I., Coffin, J.M., 1999b. Linkage disequilibrium test implies a large effective population number for HIV in vivo. *Proc. Natl. Acad. Sci. U.S.A.* 96, 10758–10763.
- Salomon, H., Wainberg, M.A., Brenner, B., Quan, Y., Rouleau, D., Cote, P., LeBlanc, R., Lefebvre, E., Spira, B., Tsoukas, C., Sekaly, R.P., Conway, B., Mayers, D., Routy, J.-P., Investigators of the Quebec Primary Infection Study, 2000. Prevalence of HIV-1 resistant to antiretroviral drugs in 81 individuals newly infected by sexual contact or injecting drug use. *AIDS* 14, 17–23.
- Strimmer, K., von Haeseler, A., 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* 13, 964–969.
- Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through

- sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- Wainberg, M.A., 1999. HIV resistance to antagonists of viral reverse transcriptase. In: Dalglish, A., Weiss, R. (Eds.), *HIV and the New Viruses*. Academic Press, London, pp. 223–249.
- Williamson, S., 2003. Adaptation in the *env* gene of HIV-1 and evolutionary theories of disease progression. *Mol. Biol. Evol.* 20, 1318–1325.
- Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.M.K., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155, 431–449.
- Yerly, S., Kaiser, L., Race, E., Bru, J.P., Clavel, F., Perrin, L., 1999. Transmission of antiretroviral-drug-resistant HIV-1 variants. *Lancet* 354, 729–733.
- Yusim, K., Kesmir, C., Gaschen, B., Addo, M.M., Altfeld, M., Brunak, S., Chigaev, A., Detours, V., Korber, B.T., 2002. Clustering patterns of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. *J. Virol.* 76, 8757–8768.
- Zanotto, P.M. de A., Kallas, E.G., Souza, R.F., Holmes, E.C., 1999. Genealogical evidence for positive selection in the *nef* gene of HIV-1. *Genetics* 153, 1077–1089.