Figure legend for sequence data from 118 patients from the Bacheler et al (2000) dataset

Data are from:
Bacheler, L. T. et al. Human immunodeficiency virus type 1 mutations selected in patients failing efavirenz combination therapy. Antimicrobial Agents and Chemotherapy 44, 2475–2484 (2000).
http://www.ncbi.nlm.nih.gov/pubmed/10952598

Data were downloaded from Genbank. For each patient, we removed time points for which there are only one clone, and then we removed patients for which there is only one time point. There are 118 patients left.

Figure legends
The attached figures show a summary of the sequences of all 118 patients. Mutations that were only observed once in a patient (singletons) are removed.
Each row represents a sequenced viral isolate. Each column represents a polymorphic site, with the derived synonymous and non-synonymous polymorphisms shown in black and red respectively.
Codons that are relevant for resistance and polymorphic in the patient are indicated in grey (if codon in susceptible state) or blue (if resistant). Mutations in resistance codons are in purple, except the A to C mutation at the third position of the 103 codon of RT, which is colored yellow and the synonymous A to G mutation at the same site, which is colored black.

30 Sweep intervals
To study selective sweeps in HIV, we used viral sequences from patients (out of the larger dataset of 118 patients) for which we had samples at two consecutive time-points that satisfy the following two criteria. (1) At the first time-point, no known resistance mutations to any of the drugs used in the trial was present at more than 30% frequency in the sample. (2) At the next time-point, at least one drug resistant allele increases in frequency by at least 70%. **There were 30 such patients.** In most cases (26 out of 30) the frequency of the mutations changes from 0 to 100% in the samples. The four exceptions are patient 89 (mutation G190S changed from 0 to 75%), patient 168 (mutation K103N changed from 0 to 83%), patient 22 (mutation K103N changed from 13 to 100%, while mutation V82A changed from 0 to 100%,), and patient 81 (mutation K103N changed from 14 to 100%). In most cases (24 out of 30) only a single drug-resistance mutation went to fixation (2 mutations in patients 22, 56, 87, 91, 154 and 166).

The patients were treated with Zidovudine, Lamivudine, Efavirenz and Indinavir. There are many known mutations that confer resistance to one or more of these drugs. We used the following list of major drug-resistance mutations (the number is the codon and the letter is the amino-acid that confers resistance). Protease: 46IL, 82AFT, 84V; Reverse Transcriptase: 41L, 62V, 65R, 67N, 70R, 75I, 77L, 100I, 101P, 103N, 106MA, 108I, 116Y, 151M, 181CI, 184VI, 188LCH, 190SA, 210W, 215YF, 219QE, 225H (Reference: Johnson V, Brun-V´ezinet F, Clotet B, Günthard H, Kuritzkes D, et al. (2010) Update of the drug resistance mutations in HIV-1: December 2010. Top HIV Med 18: 156-163).

Most of the excluded patients (76 patients) were excluded because they already had drug-resistant virus when the first sample was taken.

Details on 30 sweep-intervals

Patient	Start interval (day)	End interval (day)	K103N fixed?	Num res. muts fixed	Sample size	Num AAC allele	Num AAT allle	Soft sweep
"P00005"	0	330	TRUE	1	4	3	1	1
"P00007"	502	658	TRUE	1	6	6	0	0
"P00012"	0	88	TRUE	1	2	1	1	1
"P00013"	0	161	TRUE	1	7	0	7	0
"P00020"	0	167	FALSE	1	6	-1	-1	-1
"P00022"	0	335	TRUE	2	4	4	0	0
"P00024"	0	284	TRUE	1	7	5	2	1
"P00032"	7	308	TRUE	1	8	8	0	0
"P00044"	0	57	TRUE	1	3	3	0	0
"P00056"	6	335	TRUE	2	2	2	0	0
"P00057"	6	85	TRUE	1	6	3	3	1
"P00058"	0	28	TRUE	1	7	4	3	1
"P00063"	0	56	TRUE	1	7	7	0	0
"P00066"	0	277	TRUE	1	5	5	0	0
"P00070"	56	169	FALSE	1	6	-1	-1	-1
"P00071"	0	112	TRUE	1	7	0	7	0
"P00077"	0	84	TRUE	1	7	0	7	0
"P00081"	56	224	TRUE	1	8	7	1	1
"P00083"	14	462	FALSE	1	8	-1	-1	-1
"P00086"	13	84	TRUE	1	6	0	6	0
"P00087"	0	193	TRUE	2	8	7	1	1
"P00089"	28	56	FALSE	1	8	-1	-1	-1
"P00091"	0	424	TRUE	2	2	2	0	0
"P00095"	0	308	TRUE	1	6	6	0	0
"P00154"	0	356	TRUE	2	7	7	0	0
"P00159"	0	225	FALSE	1	6	-1	-1	-1
"P00166"	0	374	TRUE	2	4	4	0	0
"P00167"	0	117	FALSE	1	7	-1	-1	-1
"P00168"	15	78	TRUE	1	6	2	3	1
"P00171"	112	375	FALSE	1	2	-1	-1	-1

27 No-sweep intervals:

As controls, we used 27 intervals where no resistance mutation swept to high frequency. These intervals are characterized by: 1) No change in frequency of a resistance mutation of more than 30% during the interval, 2) No resistance mutation may be present at more than 30% frequency at the

beginning of the interval, and 3) The interval had to be longer than 10 days.
Note: some patients occur more than once in the no-sweep intervals and some patients occur in the sweep intervals and in the no-sweep intervals.

Details on the no-sweep intervals:

Patient	Start interval (day)	End interval (day)
"P00004"	110	236
"P00006"	0	14
"P00006"	14	70
"P00007"	0	502
"P00008"	0	217
"P00008"	217	506
"P00043"	0	167
"P00062"	0	112
"P00070"	0	56
"P00081"	0	56
"P00083"	0	14
"P00085"	0	113
"P00086"	0	13
"P00089"	7	28
"P00155"	0	32
"P00155"	32	59
"P00155"	59	111
"P00156"	0	51
"P00156"	51	82
"P00157"	0	110
"P00162"	6	103
"P00162"	103	148
"P00168"	0	15
"P00169"	0	56
"P00169"	56	84
"P00171"	0	112
"P00172"	0	84

Analysis in:
Pleuni Pennings, Sergey Kryazhimskiy, John Wakeley
Loss and Recovery of Genetic Diversity in Adapting Populations of HIV
(Submitted on 15 Mar 2013)
http://arxiv.org/abs/1303.3666