Research

Multiple independent origins of a protease inhibitor resistance mutation in salvage therapy patients

Amit Kapoor[1,2], Beth Shapiro[3], Robert W Shafer[4], Nancy Shulman[4], Soo-Yon Rhee[4] and Eric L Delwart*[1,2]

Address: [1]Blood Systems Research Institute, San Francisco, CA 94118, USA, [2]University of California San Francisco, CA, USA, [3]Henry Wellcome Ancient Biomolecules Centre, Dept of Zoology, Oxford University, Oxford, UK and [4]Division of Infectious Diseases, Department of Medicine, Stanford University Medical Center, Stanford, CA, USA

Email: Amit Kapoor - Akapoor@bloodsystems.org; Beth Shapiro - beth.shapiro@zoology.oxford.ac.uk; Robert W Shafer - rshafer@stanford.edu; Nancy Shulman - nshulman@stanford.edu; Soo-Yon Rhee - syrhee@stanford.edu; Eric L Delwart* - delwarte@medicine.ucsf.edu

* Corresponding author

Abstract

Background: Combination anti-viral therapies have reduced treatment failure rates by requiring multiple specific mutations to be selected on the same viral genome to impart high-level drug resistance. To determine if the common protease inhibitor resistance mutation L90M is only selected once or repeatedly on different HIV genetic backbones during the course of failed anti-viral therapies we analyzed a linked region of the viral genome during the evolution of multi-drug resistance.

Results: Using L90M allele specific PCR we amplified and sequenced gag-pro regions linked to very early L90M containing HIV variants prior to their emergence and detection as dominant viruses in 15 failed salvage therapy patients. The early minority L90M linked sequences were then compared to those of the later L90M viruses that came to dominate the plasma quasispecies. Using Bayesian evolutionary analysis sampling trees the emergence of L90M containing viruses was seen to take place on multiple occasion in 5 patients, only once for 2 patients and an undetermined number of time for the remaining 8 patients.

Conclusion: These results indicate that early L90M mutants can frequently be displaced by viruses carrying independently selected L90M mutations rather than by descendents of the earlier mutants.

Introduction

High rates of human immunodeficiency virus (HIV) replication and mutation in vivo results in the continuous generation of genetic variation [1]. HIV within a patient is therefore present as a mixture of related but distinct genetic variants collectively referred to as a quasispecies. HIV variants in different anatomical locations of the same individual also frequently differ possibly reflecting adaptation to local cellular environments, difference in immu-

nological pressures and/or founder effects of tissue colonization [2,3]. Differences in the strength of anti-retroviral therapy selective pressure in different tissues and cell types may also contribute to the uneven distribution of drug resistance variation in vivo [4].

HIV protease inhibitors impair the maturation and resulting infectivity of viral particles leading to a rapid decline in plasma viremia as the major virus producing cells are

depleted by viral cytopathic effects and/or immune responses. Different amino acid substitutions in the viral protease region are tightly associated with reduced sensitivity to protease inhibitors and rebounding viral loads. These mutations may also emerge in a sequential order [5-8]. Primary drug resistance mutations that alone confers moderate resistance such as V82A and L90M are initially selected followed by the addition of secondary mutations often located outside of the active site of the PR, such as L10I, M36I, M46I, L63P, or A71V leading to higher levels of resistance [9,10].

In addition to protease inhibitor resistance mutations in the protease gene, HIV protease cleavage site mutations can also be selected to compensate for reduced enzymatic activity against the wild-type cleavage sites [11-13].

Evolution of protease inhibitors resistance has been studied using mathematical models as well as longitudinal sequence analysis of HIV in vivo [9,14]. Such studies confirmed the expected presence, prior to therapy, of very low level of drug resistant mutants [15]. Both secondary protease resistance mutations and protease cleavage site mutations have been detected prior to protease inhibitor selection treatment [16]. The usually negative consequence of such drug resistance mutations on viral replicative fitness (in the absence of anti-retroviral therapy) is likely to keep the pre-treatment frequency of drug resistant mutants low [17]. Early during sub-optimal anti-retroviral therapy, weakly drug resistant viruses are therefore selected followed by the accumulation of further drug resistance mutations resulting in high-level drug resistance. The genetic characteristics of selected drug resistant variants in vivo has been longitudinally analyzed after these variants have reached a significant proportion of the plasma population using direct PCR sequencing methods [18-23]. Technically more demanding has been the analysis of the early stage of drug resistant mutant selection when selected mutants are still present at a very low frequency within the dominant drug sensitive viral quasispecies. Several studies have reported the emergence of previously minority variants carrying drug resistance mutations to dominate the later quasispecies and the frequent occurrence of viral recombination [24-26]. In this study we genetically characterized protease inhibitor resistant variants carrying the protease L90M mutation before they reached readily detectable frequencies (i.e. using direct PCR population sequencing) in patients failing salvage antiretroviral therapies. L90M is one of the most common protease-inhibitor resistance mutations and is selected primarily by the protease-inhibitors saquinavir, nelfinavir, and indinavir, at least one of which was received by each of the patients in this study. The L90M mutation, which is not located near the enzyme's active site, is thought to displaces L24, which is adjacent to the catalytic residue D25, reducing the volume of the substrate cleft or protease dimer stability decreasing susceptibility to most protease inhibitors [27-29]. L90M allele specific amplification and sequencing of the resulting PCR product allowed us to compare an upstream linked region of early minority L90M variants with those of co-replicating dominant L90 viruses and to the L90M viruses which later came to dominate the plasma quasispecies [9,13,30]. Using Bayesian evolutionary analysis to sample alternative phylogenetic trees we determined that in a significant fraction of patients the early and later L90M carrying variants appear to have distinct origins, and therefore to have been originally selected on different genetic backbones.

Results

Selective amplification of variant carrying L90M mutation

PCR primers were designed and tested to specifically amplify variants carrying the protease drug resistance mutation L90M present at low frequencies within quasispecies dominated by wild type viruses. We targeted the protease L90M mutation as a frequently selected primary drug resistant mutation providing measurable levels of resistance to all currently approved protease inhibitors. The L90M mutation being located near the extremity of the protease gene also allowed linkage of early L90M mutations with other upstream protease polymorphisms. PCR primers located upstream of protease were also used to amplify a region of gag including multiple protease cleavage sites [13]. A protocol was first designed and tested for the specific amplification of minority L90M mutants starting from PCR products generated using generic gag-pro nested PCR (see material and methods).

In order to measure the specificity of our L90M targeting PCR primer we first tested two plasmids (pAKL90 and pAK90M) encoding the w.t. L90 and mutant L90M codons respectively. The L90M specific primer AK90m was designed with its three 3' end nucleotides complementary to the L90M methionine codon (ATG). The 4th complementary base of that primer (relative to HIV) was deleted and a 5th base resulting in a mismatch was used (Fig. 1A). The next AK90m nucleotides were complementary to the HIV subtype B consensus sequence. Relative to other L90M targeting primers tested the deletion of the 4th base and mismatching of the 5th base significantly reduced background annealing to the w.t. L90 codon (TTG)(data not shown).

Ten-fold dilutions of generic gag-pro PCR products derived from plasmids pAKL90 and pAK90M were made using generic gag-pro 2nd PCR round primers (AKG2 and EDPR4)(see materials and methods). These PCR product log dilutions were then tested using the L90M specific primer pair (AKG3 and AK90m). Up to 10⁴ fold dilutions
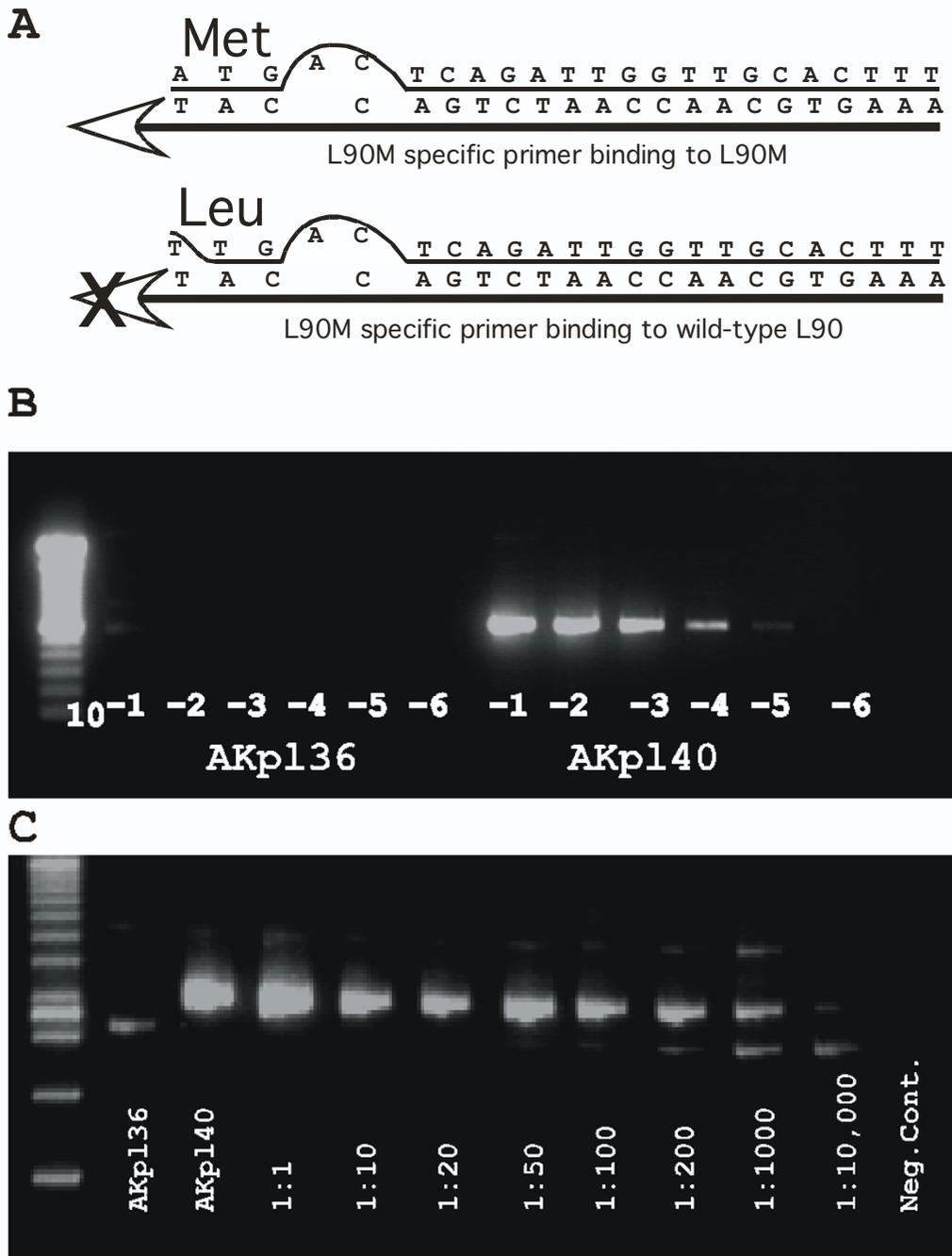
Figure 1
Design and test of the L90M allele-specific PCR protocol. A. Alignment of AK90m primer with mutant L90M and wild-type L90 region of protease gene (subtype B consensus). B. PCR amplification using L90M specific primer of serial log dilutions of L90 (AKp136) and L90M (AKp140) plasmid derived generic gag-pro amplicons. C. Sensitivity of L90M allele specific PCR using serial dilution of AKp140 derived amplicon in a constant amount of AKp136 derived L90 amplicon. Ratio of L90M to L90 amplicons DNA is shown

of the L90M generic gag-pro PCR products could be amplified while none of the L90 generic gag-pro PCR products was amplified (Fig 1B). To further evaluate the discriminatory capability of the L90M specific primer AK90m and more closely mimic conditions using clinical samples serial dilutions of the L90M PCR products were mixed with a constant amount of L90 PCR products. The presence of L90M DNA could still be detected after a 1000 fold dilution while the L90 DNA target generated only a shorter non-specific fragment at the highest L90M dilutions (Fig. 1C). To confirm that the amplicons generated from the highest pAK90M PCR DNA dilutions (1:200 and 1:1000) were indeed derived from the minority L90M variant they were purified and directly sequenced. Sequencing confirmed, through the detection of all 15 internal mutations distinguishing pAK90M from pAKL90, that the L90M specific primers had specifically amplified the L90M variant at the highest dilutions.

Sequence analysis of clones for confirmation of minority population genotype

To further substantiate the specificity of our L90M specific amplifications we analyzed the second PCR round generic gag-pro PCR products from two plasma samples in which the L90M variants were detected only using L90M specific primers (i.e. not by direct sequencing of the generic PCR products). Generic gag-pro amplicons from patient samples 608 (03/98) and 1391 (05/99) were subcloned into E. coli plasmids and the presence of the L90M mutation determined using the L90M specific primers. 150 insert containing E. coli colonies (i.e. white) from each subcloning were replica plated and the presence of L90M in their inserts tested by colony PCR (see materials and methods). Two colony PCRs from 608 and one from 1391 were L90M positive. These plasmids were purified and the presence of the L90M mutations confirmed by plasmid insert sequencing. The sequence of the L90M plasmid subclones were also compared to those derived by sequencing the L90M specific amplicons derived from the generic gag-pol PCR. The sequence polymorphisms distinguishing the L90M specific amplicons from their respective majority population sequences were similar to those detected in the L90M positive plasmid subclones (data not shown). These results further support the specificity of detection of minority L90M variants using our allele specific PCR protocol.

Protease gene evolution

To analyze the early evolution of L90M longitudinally collected plasma samples from 45 patients who developed L90M (as detected by direct PCR population sequencing) were obtained. Generic gag-pol amplification products from plasma samples available from earlier time points (negative for L90M by direct PCR population sequencing) were then tested using the L90M specific primers. The

early presence of low frequency L90M variants was detected in fifteen of the forty-five patients. For nine patients only a single time point showed the presence of low-frequency L90M mutants. From these nine patients three sequences were derived. Direct PCR population sequencing of the generic gag-pol amplicons was used to determine the consensus sequences of both the early dominant w.t. and the later dominant L90M populations. The later time point generic gag-pro sequences invariably confirmed the dominance of the L90M mutation which was detectable by direct population sequencing while the earlier time point only showed w.t. codon 90 by direct population sequencing as expected from the patient selection criteria (see materials and methods). The early time point amplicons generated using the L90M specific primer were also directly sequenced to determine the consensus sequence of these minority L90M populations. Intermediate time points were available for another 6 patients and the generic gag-pro and L90M specific PCR products, when generated from these samples, were also directly sequenced.

Bayesian analysis for monophyletic origin of L90M carrying variants

To determine whether it was appropriate to include all sequences in a single analysis and to test for recombination in the data set, we conducted a Phi test [31], as implemented in SplitsTree4 [32]. The Phi statistic tests for genealogical correlation between neighboring sites (which is negatively correlated with rate of recombination) using a modified pair-wise compatibility approach [31]. Use of SplitsTree v 4 found no evidence for recombination in the data set (p = 0.17).

Fig 2 shows a summary of the phylogenetic relationships between the sequences derived in this study, in which each patient forms a well-supported monophyletic clade thereby excluding the possibility of contamination or sample mix-up. In order to investigate whether the two or more sequences carrying the L90M mutation (the early minority and the later dominant variants) were originally derived from a single virus within each patient, we investigated the probability that the L90M associated sequences formed a monophyletic clade with respect to sequences lacking the L90M mutation within each patient cluster. The probability of monophyly was calculated over 9000 trees drawn from the posterior distribution of the MCMC analysis, and is listed for each patient in Table 1. Our analysis indicated that the L90M mutations evolved on more than one occasion in at least 5 patients (p ≤ 0.05; patients 1174, 1317, 1125, 6501, 608). Alternatively, evidence for a single mutation event leading to both the early and the late L90M mutation was seen for two patients (p ≥ 0.95; patients 4334, 1124) while the phylogenetic evidence was inconclusive for the remaining eight patients

Table 1: Probability of monophyletic origin for L90M sequences

Patient	p(monophyly)
Multiple origins L90M	
608	0.05
6501	0.00
1125	0.00
1174	0.00
1317	0.00
Single origin L90M	
1124	0.99
4334	1.00
Inconclusive	
597	0.34
7071	0.29
1134	0.23
1329	0.14
1391	0.87
1527	0.30
1834	0.58
2091	0.18

The probability that sequences containing the L90M mutation within each patient were derived from a single ($p \geq 0.95$) or multiple ($p \leq 0.05$) mutation events, given as the proportion of 9000 trees drawn from the posterior distribution of the Bayesian MCMC analysis in which the sequences with the L90M mutations formed a monophyletic clade.

(p: >0.05 to <0.95; patients 597, 1329, 7071, 1834, 1527, 1391, 2091, 1134).

Discussion

The high level of genetic diversity of HIV in vivo provides it with multiple strategies for evolving high levels of antiviral drug resistance. The originally selected drug resistant mutants may accumulate further drug resistance protease and RT mutations in a step-wise fashion to increase their level of drug resistance, develop resistance to new antiviral drugs and attenuate fitness costs imposed by these mutations. Longitudinally observed drug resistant populations may also have independent origins, having been initially selected on different genetic backbones. In order to further investigate this phenomenon we selectively amplified and then sequenced HIV variants carrying the frequently detected protease inhibitor resistance mutations L90M when these viruses were still only present as minority variants. We then compared these viruses with the L90M viruses from the same patients that later came to dominate their plasma quasispecies. Evidence was obtained that in a substantial subset of patients (5/15) the L90M mutations were selected on multiple occasions. The L90M variant that eventually emerged to readily detectable levels (i.e. using direct PCR population sequencing)

could therefore be of different origin than the earlier replicating minority L90M viruses. In another subset of patients our analysis indicated that the later dominant L90M variants descended from the earlier minority L90M variant in a manner consistent with step-wise drug resistance mutations gradually accumulating on descendents of the original L90M virus. The remaining patients could not be confidently categorized into either group.

In vivo and in vitro studies have suggested an important role for protease cleavage site mutations in restoring replicative fitness of drug resistant HIV variants [33,34]. We found no significant evolution of these cleavage sites among the 15 patients analyzed here.

While the phylogenetic analysis was able to differentiate between monophyletic or multiple origins of the L90M variants in many of the patients, it appeared most powerful when multiple sequences were available from several time points. Four out of five of the subjects with multiple L90M origins four were sampled at ≥4 time points (the fifth patient 6501 was sampled twice). All patients with apparently monophyletic L90M origins were sampled ≤3 times. Increasing the number of sampling over the course of infection may therefore significantly improve the capacity of such phylogenetic analyses to detect multiple origins of a drug resistance mutation.

It is also noted that HIV recombination within a patient, which is known to be commonplace in HIV-1 and is currently not possible to incorporate in a BEAST analysis, may also affect the resulting reconstructions. Recombination may rapidly shuffle neighboring sequence fragments obscuring the origin of different drug resistant variants. Indeed such recombination has been recently demonstrated to occur in vivo [25,26] as anticipated from the common detection of cell co-infected with multiple variants [35]. The Phi test used here works on the principle that in the presence of recombination, sites that are nearer to each other in sequence space should have greater compatibility with each other than will sites that are distant from each other [31]. The statistical significance of genealogical correlations of adjacent sites is then evaluated using a permutation test, in which the null hypothesis of no recombination would result in no effect on correlation of adjacent sites after permutation (as all sites share the same history). The test is particularly appropriate for this data set as it does not assume a single population, is powerful to detect recombination regardless of population and demographic history, mutation rate and rate of recombination, and has been shown to accurately distinguish between recent mutations and recombination even with closely related sequences. While recombination was not detected it remains possible that more frequent sampling or the analysis of clonal sequence data rather than
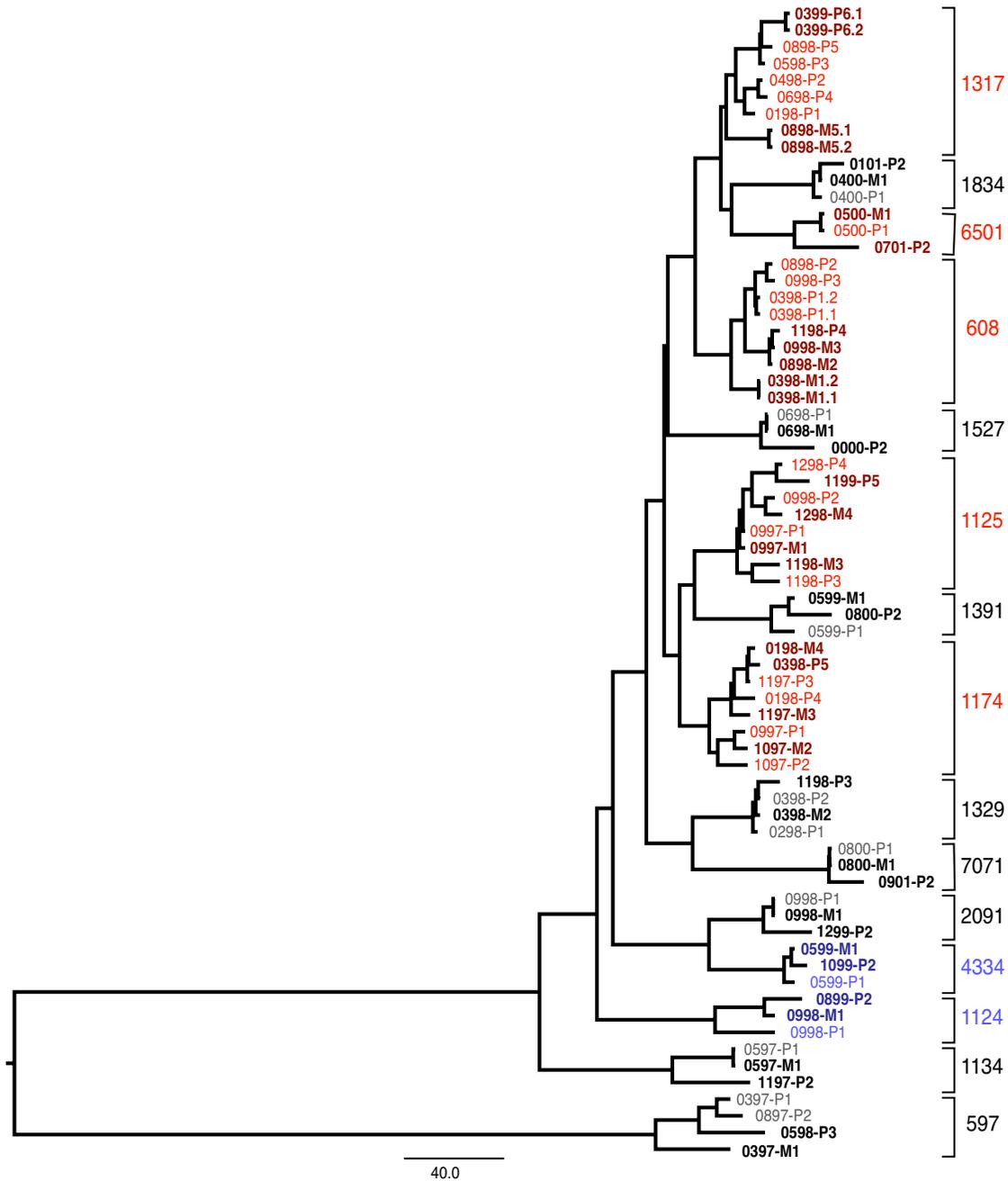
Figure 2
Summary tree showing the evolutionary relationships between all of the sequences used in this analysis, showing median node heights derived from 9000 posterior trees. Labels on the tips first show the sampling date in month and year and whether the generic (P for population) or L90M specific (M for mutant) PCR primers were used to generate the sequenced amplicons. The following number (1 through 6) represents the order of the bleeds analyzed. Patient numbers are given in the margin. Sequences that have the L90M mutation are indicated in darker bold fonts. Each patient forms a monophyletic cluster with 100% Bayesian posterior probability support (in 100% of posterior trees). Posterior support values for the monophyletic relationships of the sequences with the L90M mutation are given in Table 1. Patients with sequences showing evidence for multiple origins of L90M are shown in red, those with a single L90M origin in blue and those with an inconclusive number of L90M origins in black.

population consensus sequences may have revealed a greater level of recombination.

The detection of a polyclonal origin for the commonly selected L90M protease mutation was anticipated given the constant generation of HIV mutants. Selection of some drug resistance mutation has been shown as early as the initial phase of viremia control in patients undergoing suppressive therapy [36]. This study shows that a common drug resistance mutation can frequently have multiple origins.

Methods
Patient Samples
The plasma samples of HIV infected persons undergoing direct PCR sequencing for drug resistant genotyping at Stanford University Hospital Diagnostic Virology Laboratory meeting the following criteria were selected for this study: (i) Two drug resistance genotypes were performed before or following virologic failure on a protease inhibitor containing regimen; and (ii) the first and second genotype performed showed complete absence followed by the presence of the primary protease inhibitor resistance mutationL90M using direct PCR population sequencing. Available viral loads and antiretroviral drug regimen are shown (see Additional file 1). The study was approved by Stanford University Hospital and the University of California San Francisco committees on human research.

Viral RNA isolation and RT-nPCR
Viral RNA was extracted from 280 μl of patient's plasma using QIAGEN viral RNA extraction kit and eluted in 50 μl of RNase-free water. 15 μl of extracted RNA was reverse transcribed using 5 pmol of primer EDPR2 (TTGTT-TAACTTTTGGGCCATCC [HXB2 positions 2597 to 2618]) in a solution containing 2 μl of 10 mM dNTP at 95°C for 10 sec, followed by 65C for 5 min. 200 U of SuperScript II RNase H- Reverse transcripatse, 5× first strand buffer, 30 U of RNase Inhibitor in final volume of 25 μl was kept on 42°C for 60 min, followed by 75°C for 15 min to inactivate the enzyme. Nested PCR using HIV-1 subtype B primers was then used to amplify a fragment encompassing seven HIV protease cleavage sites (CA/p2, p2/NC, NC/p1, NC/TFP, p1/p6, TFP/p6 and p6/protease) and all 99 amino acids of HIV protease gene (from HXB2 positions 1829 to 2577). First round of PCR was done using 10 μl of cDNA and primers AKG1 (GATGACAGAAACCTTGTT-GGTCCA HXB2 positions 1736 to 2163]) and EDPR2 (TTGTTTAACTTTTGGGCCATCC HXB2 positions 2597–2618) generating a 882 bp product. Second round PCR used 1 μl of first round PCR product and primers AKG2 (GACAGCATGTCAGGGAGTAGG HXB2 positions 1829–1850) and EDPR4 (CTGGTACAGTTTCAATAG-GACTAATGG HXB2 positions 2551 to 2577) generating a 748 bp product [22]. In text we refer to these primers as

the first and second PCR rounds generic gag-pro primers. The final volume of each PCR reaction was 50 μl containing 10 pmol of each primers, 10 mM Tris-HCl pH 9.0, 50 mM KCl, 2.5 mM MgCl$_2$, 0.1% triton-X100, 2.5 mM of each dNTP and 3.5 U *Taq* polymerase. PCR program consisted of 3 cycles at 94°C for 45 s, 57°C for 45 s, and 72°C for 1 min, followed by 27 cycles at 94°C for 30 s, 57°C for 30 s, and 72°C for 1 min, and final extension at 72°C for 5 min. Using pNL4-3 plasmid dilutions the sensitivity of this nested PCR was determined to be between 1 and 10 copies (data not shown). Second round PCR products generated with AKG2 and EDPR4 were purified using Quiagen PCR product purification kit and sequenced using primer EDPR4. Sequencing of the entire PCR products was on occasion prevented by the presence of variants of different length in the PCR product (due to the co-amplification of indel variants). When sequencing such PCRs the sequencing electrophoregram abruptly shows the presence of mixed bases at every nucleotide position following an insertion-deletion (indel) containing region. To allow sequencing beyond the region of length polymorphism these PCR fragment were sequenced from the other direction using primer AKG2 allowing us to generate the consensus sequence on both side of the indel. For patients 608 and 1317 both consensus sequences (with and without extra indel codon) are used in Figure 2 (taxa labels ending in .1 and .2).

Selective amplification of minority variant carrying L90M protease mutation
The pGEM-T vector backbone plasmids pAKL90 and pAK90M each have a 748 bases long HIV fragment insert of the second round PCR product generated by the second PCR round generic gag-pro primers AKG2 and EDPR4. pAK90M has the L90M mutation (methionine codon 90 = ATG) while pAKL90 is wild type L90 (leucine codon 90 = TTG). pAKL90 and pAK90M also differ at 15 nucleotides positions within the fragment amplified by the L90M specific PCR.

Input DNA into the L90M specific PCR consisted of PCR DNA generated from either plasmids or from cDNA from clinical samples using the nested PCR generic gag-pro primers. Second round PCR DNA was purified using Quiagen PCR product purification columns, diluted 1:100 in water and 10 μl of dilution was used as input for the L90M selective PCR. The L90M specific primer was AK90m (AAAGTGCAACCAATCTGACCAT HXB2 positions 2520–2542) used together with AKG3 (ACCCGGCCATAAAG-CAAG HXB2 positions 1853–1871) in a final PCR reaction volume of 50 μl containing 10 mM Tris-HCl pH 9.0, 50 mM KCl, 2.5 mM MgCl$_2$, 0.1% triton-X100, 2.5 mM of each dNTP, 15 pmol of each primer and 3.5 U Hot start *Taq* polymerase. PCR cycling consisted of polymerase activation at 95°C for 15 min followed by 5 cycles at 95°C for

45 s, 50°C for 45 s, and 72°C for 1 min, and 10 cycles at 95°C for 30 s, 50°C for 30 s, and 72°C for 1 min, and final extension at 72°C for 5 min. L90M specific amplification products generated with AK90m and AKG3 were purified using Quiagen PCR product purification kit and sequenced using primer AKG3. When indels variants of different length were co-amplified, preventing complete sequencing from AKG3, the primer AK90m was also used to sequence from the other direction. Amplification of L90M variants from E. coli colonies was performed by touching a sterile toothpick to each colony and the toothpick then rubbed on the inside of a L90M specific PCR reaction mixture followed by only eight cycle of L90M specific PCR amplification (using AK90m and AKG3). The PCR products were analyzed by PAGE.

Sequence analysis

PCR products of second round generic gag-pro amplicons were 748 nucleotides. The L90M specific amplicons were 689 nucleotides. 611 nucleotides (HXB2 position 1903 to 2513) were included in the phylogenetic analysis to accommodate nucleotide sequence unavailable from the shorter L90M specific amplicons and low quality sequence data immediately adjacent to the sequencing primer. Mixed bases at nucleotide position were recorded in direct sequencing when one minority electrophoregram peak was present at ≥40% of the total peak height. Genbank accession numbers are: <u>EU380596</u>–<u>EU380664</u>

Sequence data was generated from 15 patients. The number of sequences collected for each patient varied from three to nine depending on the number of time points available. Sequences were aligned using CLUSTAL W with all settings set to the default values, and checked by eye. Bayesian phylogenetic analysis was performed using the software BEAST v1.4 (Drummond AJ & Rambaut A (2006) BEAST v1.4, Available from [37]) that allows the incorporation of sampling times into the evolutionary model to inform the molecular rate. A relaxed molecular clock [38] was assumed, with a constant-size coalescent prior on the tree. The HKY+G_4 model of nucleotide evolution was used, which allows for different rates of transitions and transversions and for rate heterogeneity along the sequence (four categories). Posterior distributions of parameters and trees were investigated using Markov chain Monte Carlo (MCMC) analysis, with samples from the posterior drawn every 2000 steps over a total of 20,000,000 steps, and the first 10% discarded as burn-in.

Additional material

Additional file 1

Longitudinal viral loads, anti-HIV treatments, and direct PCR population sequencing drug resistance genotypes of salvage patients. Salvage patient summaries show plasma HIV-1 RNA levels (on a log scale), direct PCR population sequencing drug resistance genotypes, and anti-retroviral treatment histories. Large square and circle symbols indicate cryopreserved samples available for the generic protease and L90M specific amplifications described in this study. Solid squares indicate samples in which L90M was detected as minority variants while no L90M variants were detected in empty square samples. Empty circle indicate samples in which L90M was the dominant variant as determined by direct PCR population sequencing. Small closed circles indicate samples for which only plasma HIV-1 RNA level determination and/or population-based drug resistance genotypic testing was done.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1742-4690-5-7-S1.pdf]

References

1. Coffin JM: **HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy.** *Science* 1995, **267**:483-489.
2. Paolucci S, Baldanti F, Campanini G, Zavattoni M, Cattaneo E, Dossena L, Gerna G: **Analysis of HIV drug-resistant quasispecies in plasma, peripheral blood mononuclear cells and viral isolates from treatment-naive and HAART patients.** *J Med Virol* 2001, **65**:207-217.
3. Kepler TB, Perelson AS: **Drug concentration heterogeneity facilitates the evolution of drug resistance.** *Proc Natl Acad Sci USA* 1998, **95**:11514-11519.
4. Magierowska M, Bernardin F, Garg S, Staprans S, Miller MD, Van Rompay KK, Delwart EL: **Highly uneven distribution of tenofovir-selected simian immunodeficiency virus in different anatomical sites of rhesus macaques.** *J Virol* 2004, **78**:2434-2444.
5. Molla A, Korneyeva M, Gao Q, Vasavanonda S, Schipper PJ, Mo HM, Markowitz M, Chernyavskiy T, Niu P, Lyons N, *et al.*: **Ordered accumulation of mutations in HIV protease confers resistance to ritonavir.** *Nat Med* 1996, **2**:760-766.
6. Shafer RW: **Genotypic testing for human immunodeficiency virus type 1 drug resistance.** *Clin Microbiol Rev* 2002, **15**:247-277.
7. Condra JH, Holder DJ, Schleif WA, Blahy OM, Danovich RM, Gabryelski LJ, Graham DJ, Laird D, Quintero JC, Rhodes A, *et al.*: **Genetic correlates of in vivo viral resistance to indinavir, a human immunodeficiency virus type 1 protease inhibitor.** *J Virol* 1996, **70**:8270-8276.
8. Condra JH, Schleif WA, Blahy OM, Gabryelski LJ, Graham DJ, Quintero JC, Rhodes A, Robbins HL, Roth E, Shivaprakash M, *et al.*: **In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors [see comments].** *Nature* 1995, **374**:569-571.
9. Kijak GH, Simon V, Balfe P, Vanderhoeven J, Pampuro SE, Zala C, Ochoa C, Cahn P, Markowitz M, Salomon H: **Origin of human immunodeficiency virus type 1 quasispecies emerging after antiretroviral treatment interruption in patients with therapeutic failure.** *J Virol* 2002, **76**:7000-7009.
10. Nijhuis M, Schuurman R, de Jong D, Erickson J, Gustchina E, Albert J, Schipper P, Gulnik S, Boucher CA: **Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy.** *Aids* 1999, **13**:2349-2359.
11. Mammano F, Petit C, Clavel F: **Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients.** *J Virol* 1998, **72**:7632-7637.
12. Gatanaga H, Suzuki Y, Tsang H, Yoshimura K, Kavlick MF, Nagashima K, Gorelick RJ, Mardy S, Tang C, Summers MF, Mitsuya H: **Amino acid substitutions in Gag protein at non-cleavage sites are**

indispensable for the development of a high multitude of HIV-1 resistance against protease inhibitors. *J Biol Chem* 2002, **277:**5952-5961.

13. Cote HC, Brumme ZL, Harrigan PR: **Human immunodeficiency virus type 1 protease cleavage site mutations associated with protease inhibitor cross-resistance selected by indinavir, ritonavir, and/or saquinavir.** *J Virol* 2001, **75:**589-594.

14. Nijhuis M, Boucher CA, Schipper P, Leitner T, Schuurman R, Albert J: **Stochastic processes strongly influence HIV-1 evolution during suboptimal protease-inhibitor therapy.** *Proc Natl Acad Sci USA* 1998, **95:**14441-14446.

15. Coffin JM: **HIV viral dynamics.** *Aids* 1996, **10(Suppl 3):**S75-84.

16. Barrie KA, Perez EE, Lamers SL, Farmerie WG, Dunn BM, Sleasman JW, Goodenow MM: **Natural variation in HIV-1 protease, Gag p7 and p6, and protease cleavage sites within gag/pol polyproteins: amino acid substitutions in the absence of protease inhibitors in mothers and children infected by human immunodeficiency virus type 1.** *Virology* 1996, **219:**407-416.

17. Croteau G, Doyon L, Thibeault D, McKercher G, Pilote L, Lamarre D: **Impaired fitness of human immunodeficiency virus type 1 variants with high-level resistance to protease inhibitors.** *J Virol* 1997, **71:**1089-1096.

18. Barbour JD, Wrin T, Grant RM, Martin JN, Segal MR, Petropoulos CJ, Deeks SG: **Evolution of phenotypic drug susceptibility and viral replication capacity during long-term virologic failure of protease inhibitor therapy in human immunodeficiency virus-infected adults.** *J Virol* 2002, **76:**11104-11112.

19. Borman AM, Paulous S, Clavel F: **Resistance of human immunodeficiency virus type 1 to protease inhibitors: selection of resistance mutations in the presence and absence of the drug.** *J Gen Virol* 1996, **77(Pt 3):**419-426.

20. Condra JH, Schleif WA, Blahy OM, Gabryelski LJ, Graham DJ, Quintero JC, Rhodes A, Robbins HL, Roth E, Shivaprakash M, *et al.*: **In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors.** *Nature* 1995, **374:**569-571.

21. Deeks SG, Wrin T, Liegler T, Hoh R, Hayden M, Barbour JD, Hellmann NS, Petropoulos CJ, McCune JM, Hellerstein MK, Grant RM: **Virologic and immunologic consequences of discontinuing combination antiretroviral-drug therapy in HIV-infected patients with detectable viremia.** *N Engl J Med* 2001, **344:**472-480.

22. Doukhan L, Delwart E: **Population genetic analysis of the protease locus of human immunodeficiency virus type 1 quasispecies undergoing drug selection, using a denaturing gradient-heteroduplex tracking assay.** *J Virol* 2001, **75:**6729-6736.

23. Gandhi RT, Wurcel A, Rosenberg ES, Johnston MN, Hellmann N, Bates M, Hirsch MS, Walker BD: **Progressive reversion of human immunodeficiency virus type 1 resistance mutations in vivo after transmission of a multiply drug-resistant virus.** *Clin Infect Dis* 2003, **37:**1693-1698.

24. Charpentier C, Dwyer DE, Mammano F, Lecossier D, Clavel F, Hance AJ: **Role of minority populations of human immunodeficiency virus type 1 in the evolution of viral resistance to protease inhibitors.** *J Virol* 2004, **78:**4234-4247.

25. Nora T, Charpentier C, Tenaillon O, Hoede C, Clavel F, Hance AJ: **Contribution of Recombination to the Evolution of Human Immunodeficiency Viruses Expressing Resistance to Antiretroviral Treatment.** *J Virol* 2007.

26. Charpentier C, Nora T, Tenaillon O, Clavel F, Hance AJ: **Extensive recombination among human immunodeficiency virus type 1 quasispecies makes an important contribution to viral diversity in individual patients.** *J Virol* 2006, **80:**2472-2482.

27. Hong L, Zhang XC, Hartsuck JA, Tang J: **Crystal structure of an in vivo HIV-1 protease mutant in complex with saquinavir: insights into the mechanisms of drug resistance.** *Protein Sci* 2000, **9:**1898-1904.

28. Ode H, Ota M, Neya S, Hata M, Sugiura W, Hoshino T: **Resistant mechanism against nelfinavir of human immunodeficiency virus type 1 proteases.** *J Phys Chem B* 2005, **109:**565-574.

29. Ode H, Neya S, Hata M, Sugiura W, Hoshino T: **Computational simulations of HIV-1 proteases--multi-drug resistance due to nonactive site mutation L90M.** *J Am Chem Soc* 2006, **128:**7887-7895.

30. Bally F, Martinez R, Peters S, Sudre P, Telenti A: **Polymorphism of HIV type 1 gag p7/p1 and p1/p6 cleavage sites: clinical signif-
icance and implications for resistance to protease inhibitors.** *AIDS Res Hum Retroviruses* 2000, **16:**1209-1213.

31. Bruen TC, Philippe H, Bryant D: **A simple and robust statistical test for detecting the presence of recombination.** *Genetics* 2006, **172:**2665-2681.

32. Huson DH, Bryant D: **Application of phylogenetic networks in evolutionary studies.** *Mol Biol Evol* 2006, **23:**254-267.

33. Watkins T, Resch W, Irlbeck D, Swanstrom R: **Selection of high-level resistance to human immunodeficiency virus type 1 protease inhibitors.** *Antimicrob Agents Chemother* 2003, **47:**759-769.

34. Feher A, Weber IT, Bagossi P, Boross P, Mahalingam B, Louis JM, Copeland TD, Torshin IY, Harrison RW, Tozser J: **Effect of sequence polymorphism and drug resistance on two HIV-1 Gag processing sites.** *Eur J Biochem* 2002, **269:**4114-4120.

35. Jung A, Maier R, Vartanian JP, Bocharov G, Jung V, Fischer U, Meese E, Wain-Hobson S, Meyerhans A: **Multiply infected spleen cells in HIV patients.** *Nature* 2002, **418:**144.

36. Metzner KJ, Allers K, Rauch P, Harrer T: **Rapid selection of drug-resistant HIV-1 during the first months of suppressive ART in treatment-naive patients.** *Aids* 2007, **21:**703-711.

37. **Bayesian Evolutionary Analysis Sampling Trees** [http://beast.bio.ed.ac.uk]

38. Drummond AJ, Ho SY, Phillips MJ, Rambaut A: **Relaxed phylogenetics and dating with confidence.** *PLoS Biol* 2006, **4:**e88.