Deep Molecular Characterization of HIV-1 Dynamics under Suppressive HAART

Maria J. Buzón[1][9][¤a], **Francisco M. Codoñer**[1][9][¤b], **Simon D. W. Frost**[2], **Christian Pou**[1], **Maria C. Puertas**[1], **Marta Massanella**[1], **Judith Dalmau**[1], **Josep M. Llibre**[3], **Mario Stevenson**[4], **Julià Blanco**[1], **Bonaventura Clotet**[1,3], **Roger Paredes**[1,3], **Javier Martinez-Picado**[1,5]*

1 Institut de Recerca de la SIDA, IrsiCaixa, Hospital Germans Trias i Pujol, Universitat Autònoma de Barcelona, Badalona, Spain, 2 Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom, 3 Unitat VIH, Hospital Universitari Germans Trias i Pujol, Universitat Autònoma de Barcelona, Badalona, Spain, 4 University of Miami Miller School of Medicine, Miami, Florida, United States of America, 5 Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Abstract

In order to design strategies for eradication of HIV-1 from infected individuals, detailed insight into the HIV-1 reservoirs that persist in patients on suppressive antiretroviral therapy (ART) is required. In this regard, most studies have focused on integrated (proviral) HIV-1 DNA forms in cells circulating in blood. However, the majority of proviral DNA is replication-defective and archival, and as such, has limited ability to reveal the dynamics of the viral population that persists in patients on suppressive ART. In contrast, extrachromosomal (episomal) viral DNA is labile and as a consequence is a better surrogate for recent infection events and is able to inform on the extent to which residual replication contributes to viral reservoir maintenance. To gain insight into the diversity and compartmentalization of HIV-1 under suppressive ART, we extensively analyzed longitudinal peripheral blood mononuclear cells (PBMC) samples by deep sequencing of episomal and integrated HIV-1 DNA from patients undergoing raltegravir intensification. Reverse-transcriptase genes selectively amplified from episomal and proviral HIV-1 DNA were analyzed by deep sequencing 0, 2, 4, 12, 24 and 48 weeks after raltegravir intensification. We used maximum likelihood phylogenies and statistical tests (AMOVA and Slatkin-Maddison (SM)) in order to determine molecular compartmentalization. We observed low molecular variance (mean variability ≤ 0.042). Although phylogenies showed that both DNA forms were intermingled within the phylogenetic tree, we found a statistically significant compartmentalization between episomal and proviral DNA samples ($P < 10^{-6}$ AMOVA test; $P = 0.001$ SM test), suggesting that they belong to different viral populations. In addition, longitudinal analysis of episomal and proviral DNA by phylogeny and AMOVA showed signs of non-chronological temporal compartmentalization (all comparisons $P < 10^{-6}$) suggesting that episomal and proviral DNA forms originated from different anatomical compartments. Collectively, this suggests the presence of a chronic viral reservoir in which there is stochastic release of infectious virus and in which there are limited rounds of *de novo* infection. This could be explained by the existence of different reservoirs with unique pharmacological accessibility properties, which will require strategies that improve drug penetration/retention within these reservoirs in order to minimise maintenance of the viral reservoir by de novo infection.

Introduction

In the majority of HIV-1 infected individuals antiretroviral therapy (ART) is able to sustain suppression of plasma viral load to undetectable levels (<50 copies HIV RNA/ml plasma) for sustained intervals. However, viremia resumes if treatment is interrupted. Therefore, HIV-1 is able to persist in the face of suppressive ART. In addition low-level residual viremia has been detected with ultrasensitive assays that are able to measure down to several copies of HIV RNA/ml plasma [1,2]. It has been suggested that low level viremia in ART-suppressed patients represents release of viral particles by long-lived latently infected CD4+ T-cells [3,4,5] or virions produced as a result of low-level, residual viral replication [6,7,8,9,10,11]. The nature of this residual viremia, remains poorly understood, mainly because the very low number of virions in plasma limits its molecular characterization [3,12,13].

Intensification protocols employing integrase inhibitors have been used to probe the viral reservoirs that persist in ART-

Author Summary

In the majority of HIV-1 positive patients, antiretroviral therapy (ART) effects a sustained reduction in plasma viremia to below detectable levels. Despite this, replication competent viruses persist and fuel viremia if antiretroviral treatment is interrupted. This viral persistence stands in the way of viral eradication through ART. While this ability to persist in the face of therapy is generally considered to be attributable to a reservoir of latently infected cells, there is debate as to how this reservoir is maintained and in particular, whether there is replenishment of the reservoir by low level, residual replication. Novel antiviral agents targeting the viral integrase offer tools to explore the viral reservoirs that persist in the face of ART and we have shown that raltegravir perturbs these reservoirs as evidenced by an accumulation of episomal DNA upon rategravir intensification (Buzon et al., 2010). Through "deep sequencing" technology, we have longitudinally analyzed the genotypes of HIV episomes and integrated HIV DNA to evaluate whether they represent interrelated sequences or whether they have distinct origins. Statistical methods showed molecular compartmentalization, among and within episomal and integrated HIV-1 DNA samples, and suggest that episomal DNA in PBMC originates from a cellular/anatomic reservoir that is not revealed by sequencing of proviral DNA in PBMC in this study. These, and other data, suggest that ongoing replication, which can be blocked by adding raltegravir, occurs from proviruses that are genetically distinguishable from those detected at >1% frequency in these circulating blood cells.

suppressed patients. When viral integration is inhibited, the linear viral genome, which is the precursor to the integrated provirus, is converted to episomes [14,15]. Although sequences gleaned from episomal DNAs could be present in both productive and non-productive infections, integrase inhibition specifically results in increased episome formation and since linear cDNA is a product of reverse transcription, increases in episomal cDNA in blood cells after starting raltegravir indicates *de novo* infection and blocked integration. Because of the dynamic nature of episomes, they harbor a higher percentage of contemporary sequences as compared to proviral sequences that contain a higher percentage of archival sequences. Therefore, although episomes are dead-end products of viral replication, sequences contained within them will also be observed in functional viral genomes. As a consequence, characterization of the nature of episomal HIV-1 DNA during raltegravir intensification of a suppressive HAART regimen could provide new insights into the molecular diversity and compartmentalization of the viral reservoirs that persist in the face of suppressive therapy. We previously reported that raltegravir intensification of HAART-suppressed patients affected HIV-1 replication and immune dynamics in a large percentage of these patients [16,17]. In order to gain further insight into the molecular diversity, population structure, and compartmentalization of replicative viral forms under suppressive HAART, episomal and integrated HIV-1 DNA samples were longitudinally analyzed by deep sequencing after intensification with raltegravir. We found signs of molecular compartmentalization distinguishing episomal and integrated HIV-1 DNA populations in PBMC, suggesting that proviruses in a cellular/anatomical compartment other than those cells may give rise to stochastic release of replication-competent virus during HAART.

Results

Deep sequencing

The study sample comprised two participants from our previously reported raltegravir-intensification study [16,17] who had plasma viremia below 50 HIV-1 RNA copies per ml for two years on stable HAART. Subjects were selected based on sample availability. Reverse-transcriptase genes from episomal and integrated HIV-1 DNA were specifically amplified and analyzed at weeks 0, 2, 4, 12, 24 and 48 following raltegravir intensification. Only viral sequences present in ≥1% of the virus population were considered for further analysis. The median number and interquartile range of episomal HIV-1 DNA clonal sequences for patients 1 and 2 was 3,063 (2,017–3,473) and 4,040 (2,106–6,131), respectively. For integrated HIV-1 DNA in patients 1 and 2, the median was 2,645 (1,570–3,592) and 2,889 (2,247–4,184), respectively.

Episomal and integrated HIV-1 DNA belong to different viral populations

We constructed a phylogenetic tree for each patient to assess if episomal and integrated HIV-1 DNA sequences belonged to different genetic populations or to one intermixed population. We used a neighbor-joining approach, as implemented in MEGA4 [18], to construct a phylogenetic tree for each patient with the best evolutionary model found in jModeltest v0.1.1. Phylogenetic trees did not show a clear cluster differentiation of both DNA forms (Figs. 1 and 2), suggesting a lack of population structure between both DNA samples, at least at the sequence composition level. Even when sequences do not clearly group into separate branches, statistical analysis can still reveal differences in sequence diversity between different HIV populations [12,19,20]. Therefore, to better assess possible compartmentalization, we performed a population structure test based on the analysis of molecular variance (AMOVA) on pairwise genetic distances and percentages of the presence of each clone in each population [21]. This test showed different ratios of population structure (FST) between episomal and integrated sequences (Table 1), which all were statistically significant ($P<10^{-6}$) in both patients. As different tests of population structure can yield contradictory results, we performed a recommended conservative analysis [22] using a complementary compartmentalization test. We applied the Slatkin-Maddison test [23], which is based only on tree topology comparison. Consistent with the AMOVA test, the results of the Slatkin-Maddison analysis showed only a few migration events between episomal and integrated samples: 12 out of 55 possible migration events, and 9 out of 46 possible migration events in patients 1 and 2 respectively ($P = 0.001$) (Table 1). In addition, we obtained similar results when a longitudinal point-by-point comparison was performed between both DNA viral forms, except for three samples where the migration events had a high probability of being random (Table 1). Interestingly, two of these three samples were taken at the study baseline, i.e. right before HAART intensification with raltegravir. Overall, these results point to statistically significant compartmentalization between episomal and integrated DNA samples and suggest that they belong to different viral populations.

Non-chronological temporal compartmentalization within episomal and integrated HIV-1 viral forms

We next assessed whether longitudinal episomal and integrated DNA sequences had a temporal structure. Firstly, we constructed a separate neighbor-joining phylogenetic tree for each patient and each viral DNA form (Fig. 3 and 4). Phylogenies showed evidence of a temporal structure within episomal and integrated viral DNA forms across different time-points. Temporal structure was more
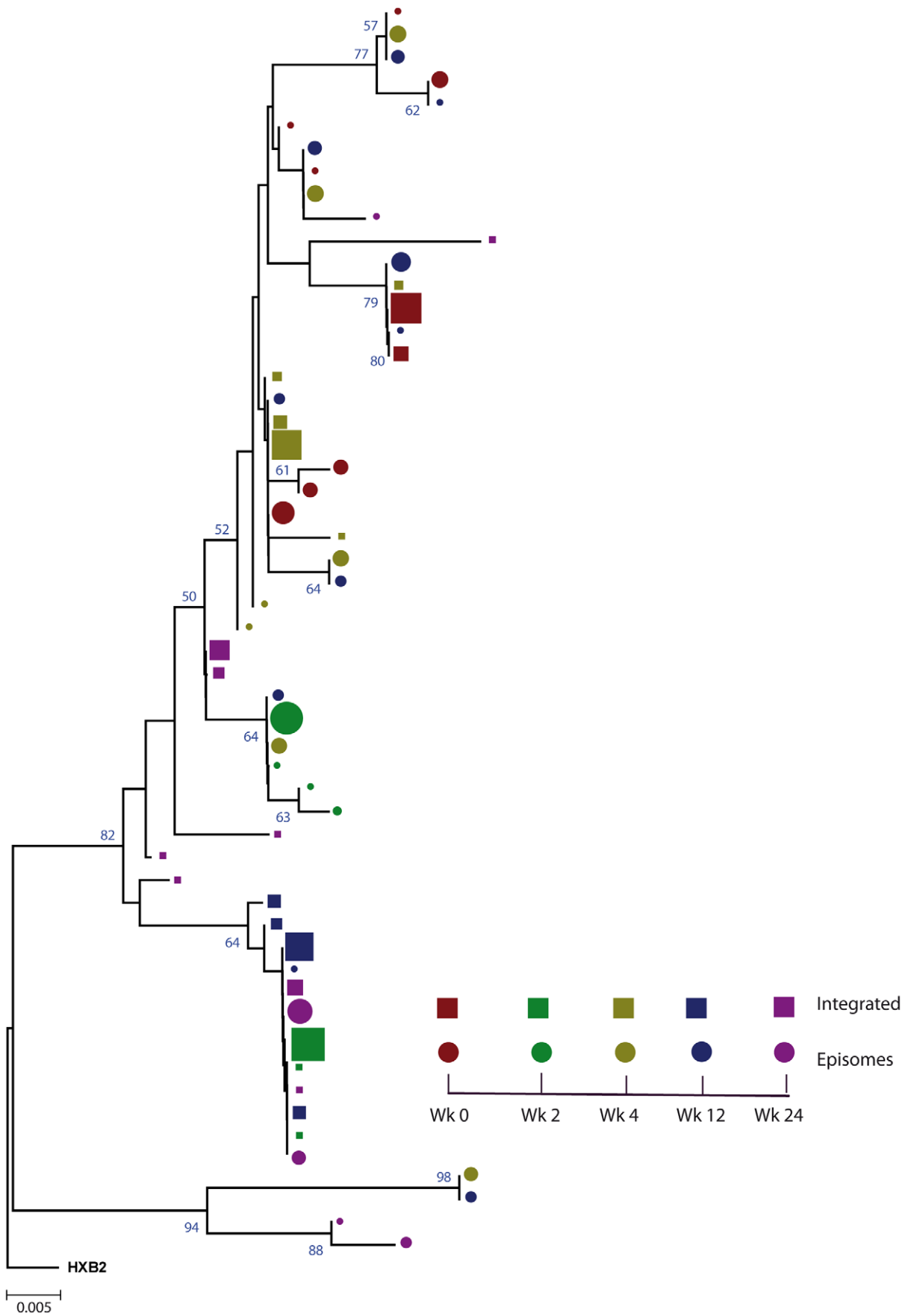
Figure 1. Phylogenetic tree of patient 1. A neighbor-joining approach, as implemented in MEGA4, was used to construct a phylogenetic tree with the best evolutionary model found in jModeltest v0.1.1. Circles and squares represent longitudinal episomal and integrated DNA sequences, respectively. Legends of phylogenetic trees represent weeks available for further analysis. Sizes of the symbols represent the different percentages of clonal sequences. 1,000 bootstrap replicates were performed; only values greater than 50% are shown at tree nodes.
doi:10.1371/journal.ppat.1002314.g001

evident in the episomal samples of patient 2 (Fig. 4a) and the integrated samples of patient 1 (Fig. 3b). However, a clear sign of temporal structure was difficult to observe in the remaining phylogenetic trees (Fig. 3a and 4b). Therefore, we used the AMOVA and Slatkin-Maddison tests to again assess the presence of temporal population structure within each viral DNA form. AMOVA showed that all longitudinal comparisons within episomal and integrated samples were significantly different ($P<10^{-6}$) indicating different temporal population structures (Tables 2–5). Statistically significant Slatkin-Maddison results were partly consistent with those detected by AMOVA (Tables 2–5). Discrepancies between both tests might be due to the large number of sequences with the same haplotype (sometimes present in both compartments). In fact, the performance of the Slatkin-Maddison test might be limited when there is a combination of relatively short sequence, high depth and low within-patient diversity [23], as in this case.

Our results revealed that in both episomal and integrated HIV-1 DNA samples, distinct genetic populations appeared at different time-points, suggesting that the appearance of each viral DNA form in blood could be the result of stochastic mobilization of different HIV-infected cells. This effect has been observed for residual viremia [12] and for different populations of CD4$^+$ T-cells [7]. Furthermore, we did not observe any signs of evolution across longitudinal samples within episomal DNA or within integrated viral forms (Fig. 3 and 4). We found temporal variation but no evidence of continued evolution. Of note, patient 1's viruses harbor the mutation M184I in the reverse transcriptase of the integrated HIV sequences from weeks 0 and 2 (but not from weeks 4 or 24), which is associated to resistance to lamivudine and emtricitabine. This patient was under a regimen containing tenofovir, lamivudine, lopinavir, ritonavir and raltegravir. In contrast episomal sequences from weeks 0, 2 and 4 were wild type, which suggest that 2LTR circles were generated in a different cellular/anatomical compartment that is possibly less accessible to lamivudine.

Different proportions of clonal sequences, rather than different sequences, determine population structure in patients under suppressive HAART

A population structure can occur at two levels: (i) different composition at the sequence level and (ii) different proportions of specific haplotypes. Therefore, the percentage of each clonal sequence of each DNA sample was represented (Fig. 3c–d and Fig. 4c–d). The results show that some haplotypes were shared among and within HIV-1 viral DNA forms, but unique haplotypes were also found. Moreover, when sequences were shared, the percentage of each haplotype was different between samples. This observation, together with the low molecular variability found in each sample (Table 1), suggests that patients under suppressive HAART have a limited variability in viral DNA sequences and that the presence and relative proportion of each clonal sequence might determine whether a population structure exists.

Discussion

Previous reports have shown evidence of compartmentalization between residual plasma viremia and proviruses in fractionated

and unfractionated PBMC [12]. However, this is the first time that episomal cDNA and integrated HIV-1 DNA genomes have been extensively compared. We found a statistically significant compartmentalization between episomal and integrated DNA samples in PBMC suggesting that, as with residual plasma viral RNA, episomal HIV-1 DNA forms are genetically distinct from proviral genomes and that they encompass two different genetic populations. In addition we have shown that in both episomal and integrated HIV-1 DNA samples, distinct genetic populations appeared, in a non-chronologic manner, at different time points. Longitudinal, non-chronological population structure between and within samples was detected in both circular episomal and proviral HIV-1 DNA viral forms. One explanation for our findings is that episomal and proviral sequences were generated in distinct cell types or anatomical compartments, possibly with different pharmacological penetration profiles. The detection of both DNA forms could result from stochastic mobilization from tissues to blood of a few infected cells, as their low molecular variance suggests. However, the labile nature of episomal DNA and its specific dynamics after raltegravir intensification [16] implies that the infection events that generated them occur in a pharmacologically privileged site (because the infections that generated them are occurring in the face of RT inhibitors) yet that is still accessible to raltegravir. A recent report shows that ileum may support ongoing productive infection in some patients on HAART, even if the contribution to plasma RNA is not discernible. In fact, raltegravir intensification contributed to a decrease in the cell-associated HIV RNA in this anatomic site relative to other gut sites or PBMC, suggesting that gut sites differ with respect to penetration by antiretroviral drugs, immunologic environments or the composition of CD4$^+$ T cell populations [24]. Alternatively, our results might also suggest that cells containing these transiently-increased episomes might result from new infections with replication-competent viruses originating from rare proviruses not detectable in peripheral blood mononuclear cells, and that these episomes have had their integration blocked by raltegravir.

The lack of population structure or evolution in integrated HIV-1 genomes is best explained by the fact that proviruses are predominantly defective and archival. Although the provirus is the molecular precursor for all virions, only a very small percentage of proviruses are replication competent and only a small percentage of these would exist in a latent state- one capable of producing replication-competent virions. Therefore, temporal structure might simply be a result of continuous seeding of new cells over time. In this regard, multiple monotypic HIV-1 sequences have been observed across the uterine cervix and in blood, presumably as a result of the proliferation of cells harboring proviruses [25]. Although the origin of these cells with integrated HIV-1 DNA in our study remains unknown, forthcoming genotypic analysis across cell subpopulations in blood and tissues may cast light on this issue. Memory CD4$^+$ T cells are thought to be a stable reservoir of HIV infection [26,27,28,29,30,31]. The transient increases in 2LTR circles in PBMC from patients during early HAART in the absence of raltegravir, has recently been associated with the redistribution of 2LTR-enriched memory CD4+ T-cells from lymphoid tissues due to short-term decreases in immune activation [32]. In our study, changes in immune activation occurred more slowly. As such, distinct mechanisms appear to account for the
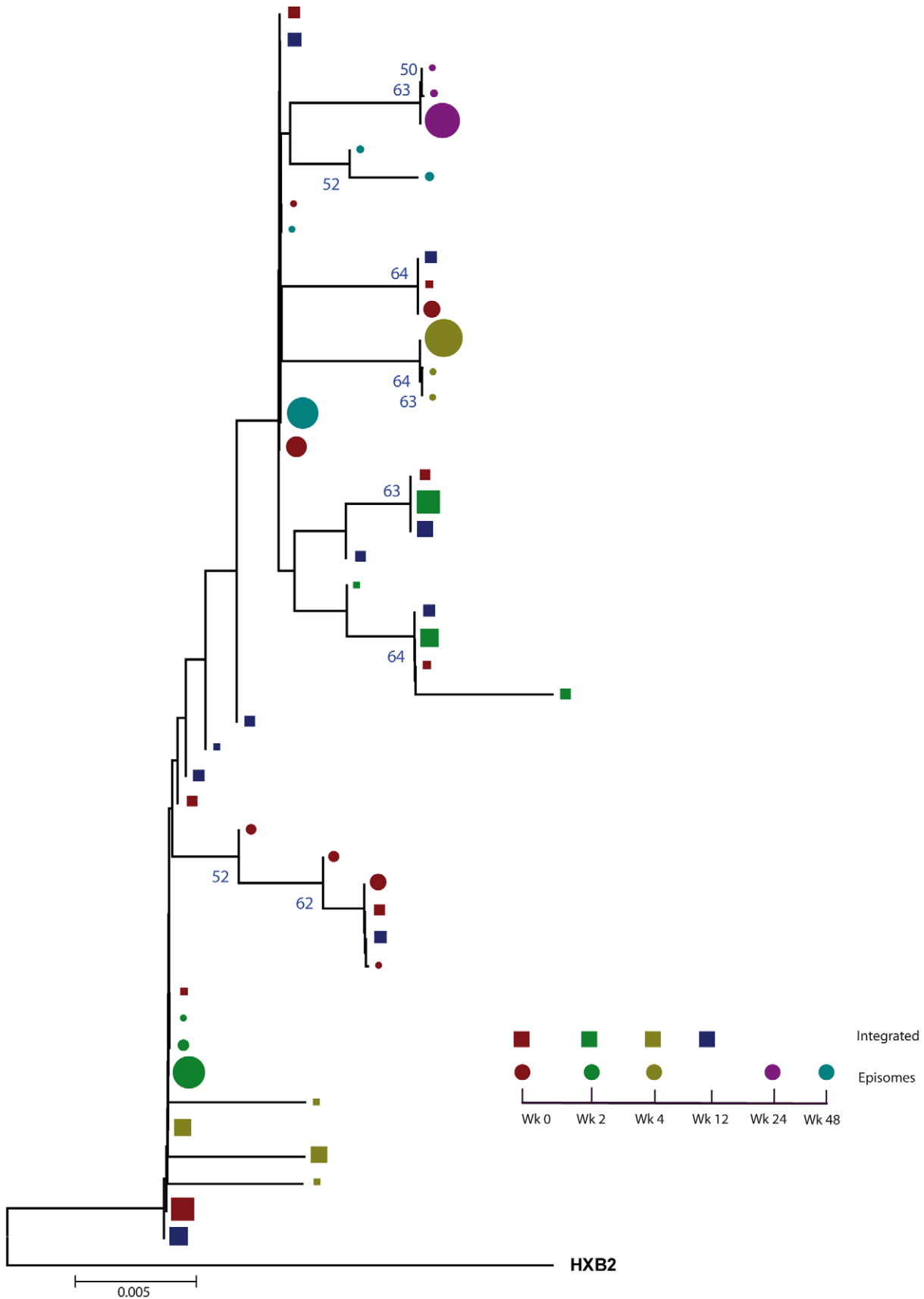
Figure 2. Phylogenetic tree of patient 2. A neighbor-joining approach, as implemented in MEGA4, was used to construct a phylogenetic tree with the best evolutionary model found in jModeltest v0.1.1. Circles and squares represent longitudinal episomal and integrated DNA sequences,

respectively. Legends of phylogenetic trees represents weeks available for further analysis. Size of the symbols represent the different percentages of clonal sequences. 1,000 bootstrap replicates were performed; only values greater than 50% are shown at tree nodes.
doi:10.1371/journal.ppat.1002314.g002

changes in 2LTR circles. Interestingly, it has also recently been shown that the majority of a highly specialized subset of antigen-specific memory $CD4^+$ T cells in mice were found to reside in a resting state in the bone marrow, surviving in close proximity to IL-7-secreting stromal cells [33]. Coincidently, IL-7 also induces homeostatic proliferation of human central memory $CD4^+$ T cells without causing viral reactivation [34]. Therefore, the different population structure within integrated DNA, which is not coincident with the episomal DNA population, could be explained by the re-activation and mobilization of memory $CD4^+$ T-cells from different niches in bone marrow without subsequent viral production, and could be a consequence of cellular rather than viral dynamics. Previous studies have examined rebounding viremia following treatment interruption and suggested that emerging viral variants result from the stochastic reactivation of different HIV-1 infected cells [4,35,36,37]. Finally, infectious events during HAART can occur in multiple, temporary, small and locally scattered bursts [38] consistent with our observed genotypic compartmentalization. In addition, we observed that the cell-associated HIV proviral DNA remained relatively stable (Fig. S1) despite the large shifts in sequence populations observed during the study. This would not only mean that newly infected cell populations had undergone expansion but that other populations contracted at the same time. It is tempting to speculate that these dynamics might be consistent with the continuous trafficking between blood and tissues (preferentially lymphoid tissue) of stochastically activated antigen-specific memory CD4+ T-cells. We believe that it is unlikely that the results

obtained reflect limited sampling, because we detected different proportions of shared haplotypes at different longitudinal time points (Fig. 1c,d and Fig. 2c,d). Sequence diversity restrictions due to sampling limitations would be reflected by either a completely different population structure or by invariable shared haplotypes in different samples. Moreover, at least in some time points, there is quite a large number of unique haplotypes (>1%), suggesting good depth.

Proviral HIV sequences are currently thought to be representative of archival HIV infection in an infected patient. Based on this hypothesis, sources of residual viremia other than $CD4^+$ T-cells have been postulated as long-lived viral reservoirs [3]. Our observation that longitudinal detection of proviral genomes is dynamic in patients on HAART is important because it points to some limitations in the conclusions drawn from cross-sectional studies comparing HIV sequences in plasma and circulating T-cells.

Our results collectively suggest the presence of a chronic viral reservoir in which there is stochastic release of infectious virus and in which there are limited rounds of de novo infection. This could be explained by the existence of a limited cellular/anatomic reservoir in which de novo infection continues during HAART because some antiretroviral drugs do not effectively inhibit replication in this compartment. However, evidence that episomes transiently increase after raltegravir intensification suggests that this cellular/anatomic reservoir may be accessible to raltegravir, in contrast to other drugs. If proven in future work, the concept that ongoing replication during successful HAART originates from proviruses that are not detectable in peripheral blood mononuclear cells has important

Table 1. Population structure analysis.

Patient	Time	Mean Diversity		AMOVA		Slatkin-Maddison Test	
		Π_{Epi}	Π_{Int}	F_{ST}	P-value	Migrations	Prob[*]
1	W0	0.0091	<0.0001	0.7287	$<10^{-6}$	1	0.105
	W2	0.0019	<0.0001	0.9889	$<10^{-6}$	1	0.056
	W4	0.0265	0.0063	0.1891	$<10^{-6}$	2	0.026
	W12	0.0313	0.0019	0.8412	$<10^{-6}$	2	0.033
	W24	0.0422	0.0198	0.5160	$<10^{-6}$	4	0.578
	Total	0.0278	0.0201	0.1684	$<10^{-6}$	12	0.001
2	W0	0.0057	0.0081	0.1679	$<10^{-6}$	5	0.584
	W2	<0.0001	0.0067	0.1092	$<10^{-6}$	1	0.058
	W4	<0.0001	0.0084	0.7680	$<10^{-6}$	1	0.057
	W12	NA	0.0054	NA	NA	NA	NA
	W24	<0.0001	NA	NA	NA	NA	NA
	W48	0.0019	NA	NA	NA	NA	NA
	Total	0.0076	0.0079	0.1457	$<10^{-6}$	9	0.001

Mean internal diversity for each time point and DNA source is shown for each patient. Genetic diversity (π), defined as the average number of nucleotide differences per site between any two sequences chosen randomly from the sample population, is calculated with the best evolutionary model found in the jmodeltest Tamura and Nei model [45]. The fixation index, F_{ST}, from the AMOVA analysis is a measure of the diversity of randomly chosen sequences within the same sub-population relative to that found in the entire population. A zero value implies that the two populations mix freely, while a value of one implies that the two subpopulations are completely separate. The Slatkin Maddison test uses an estimate of the number of 'migrations' between subpopulations to assess population structure, with smaller numbers of migrations (for a given number of sequences) indicating more structure. Both AMOVA and the Slatkin-Maddison test calculate a p value for the null hypothesis of no population structure by randomly permuting sequences between subpopulations; they differ as the AMOVA takes into account the frequency of each unique sequence or haplotype, whereas the Slatkin Maddison test does not.
*Probability that a migration event is random after 10,000 resampling replicates.
NA, indicates that the test or the data is not applicable or available.
doi:10.1371/journal.ppat.1002314.t001
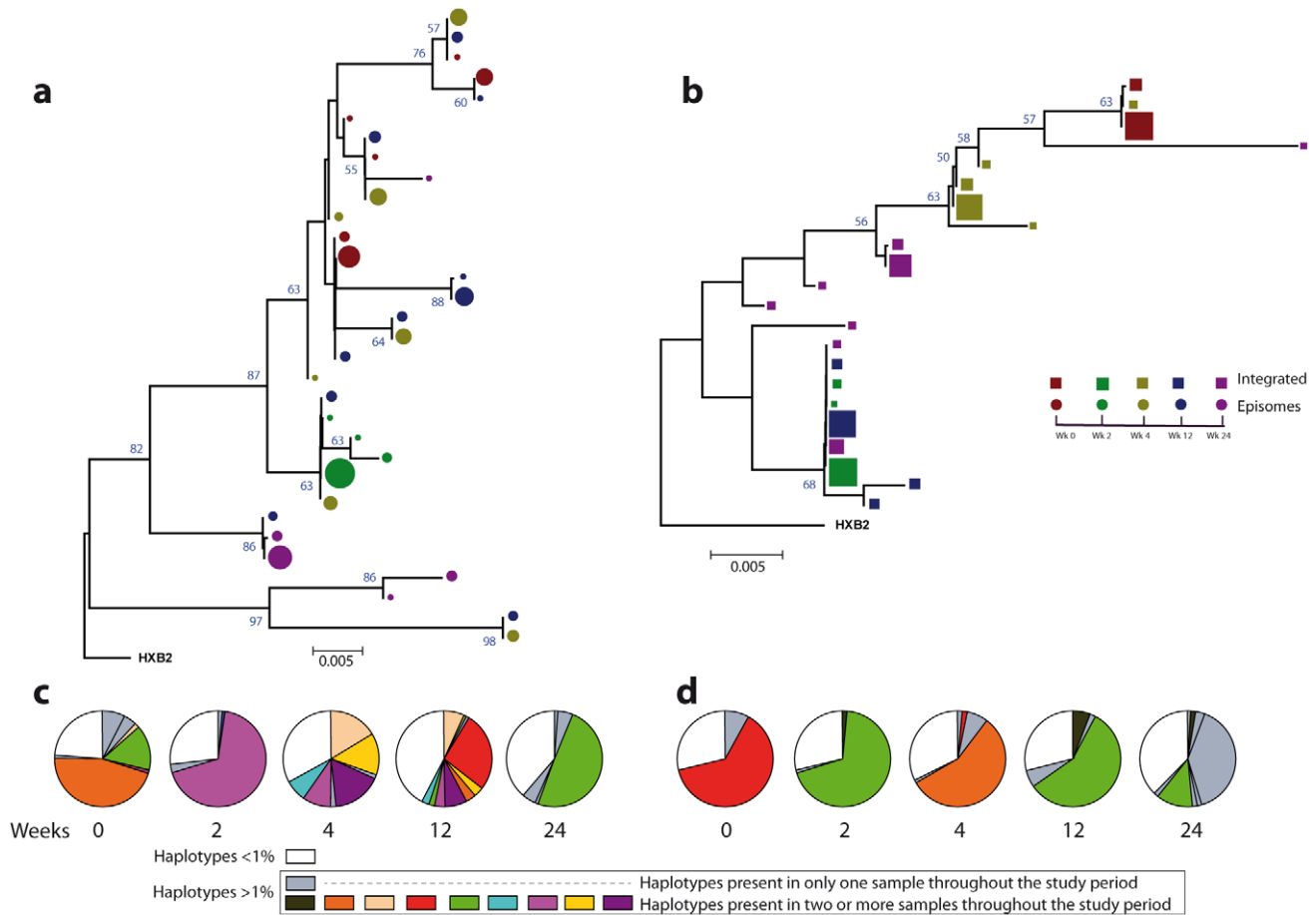
Figure 3. Phylogenetic tree and proportions of episomal and integrated HIV-1 sequences in patient 1. A neighbor-joining approach, as implemented in MEGA4, was used to construct a phylogenetic tree with the best evolutionary model found in jModeltest v0.1.1. **a.** Circles represent longitudinal episomal DNA sequences. **b.** Squares represent longitudinal integrated DNA sequences. Legend of phylogenetic trees represents weeks available for further analysis. Sizes of the symbols represent the different percentages of clonal sequences. **c.** Longitudinal representation of the clonal variability of each episomal sample. **d.** Longitudinal representation of the clonal variability of each integrated sample. Areas of pie charts in white shading indicate sequences present with a frequency below 1%; gray shading indicates sequences with a frequency above 1% present in only one sample throughout the study period; colors represent sequences with a frequency above 1% present in two or more samples throughout the study period. 1,000 bootstrap replicates were performed; only values greater than 50% are shown at tree nodes.
doi:10.1371/journal.ppat.1002314.g003

implications for the design of strategies aimed at viral eradication or functional cure. It indicates the need to further define the limited and covert cellular/anatomic reservoir in which ongoing HIV replication may occur during suppressive HAART.

Materials and Methods

Ethics statement

The study was approved by the Germans Trias i Pujol hospital review board and informed consent was obtained in writing from study participants.

Study patients

We extensively analyzed longitudinal samples from 2 HIV-infected patients whose plasma viral load had been suppressed to <50 HIV-1 RNA copies/ml for 2 years on a stable HAART regimen. Both patients had participated in a previously reported raltegravir-intensification study [16,17] were intensification of a three-drug suppressive HAART regimen resulted in a specific and transient increase in episomal DNA in a large percentage of patients. The original study was designed to compare populations of

episomal and integrated HIV-1 DNA and plasma viral RNA in 5 patients with detectable episomal DNA before raltegravir intensification. Although plasma viral load assays employed 7 ml of plasma, we were unable to amplify viral RNA sequences for the majority of time points nor amplify episomal and integrated HIV-1 DNA from the majority of longitudinal samples in 3 patients. For this reason, structure comparisons between episomes and integrated HIV-1 DNA were only possible for the 2 subjects shown in this study. Episome and proviral DNA dynamics for both patients are shown in Fig. S1. HAART regimens included lopinavir, ritonavir, lamivudine, tenofovir and raltegravir for patient 1 and efavirenz, emtricitabine, tenofovir and raltegravir for patient 2. None of the included patients had previously been exposed to integrase inhibitors. Peripheral blood mononuclear cells (PBMC) and plasma samples included in this study encompassed weeks 0, 2, 4, 12, 24 and 48 after raltegravir intensification.

Nucleic acid purification

A median of 6×10^7 PBMC were obtained at weeks 0, 2, 4, 12, 24 and 48 after intensification and purified by Ficoll centrifugation and resuspended in 350 μl of P1 buffer (Qiaprep miniprep kit,
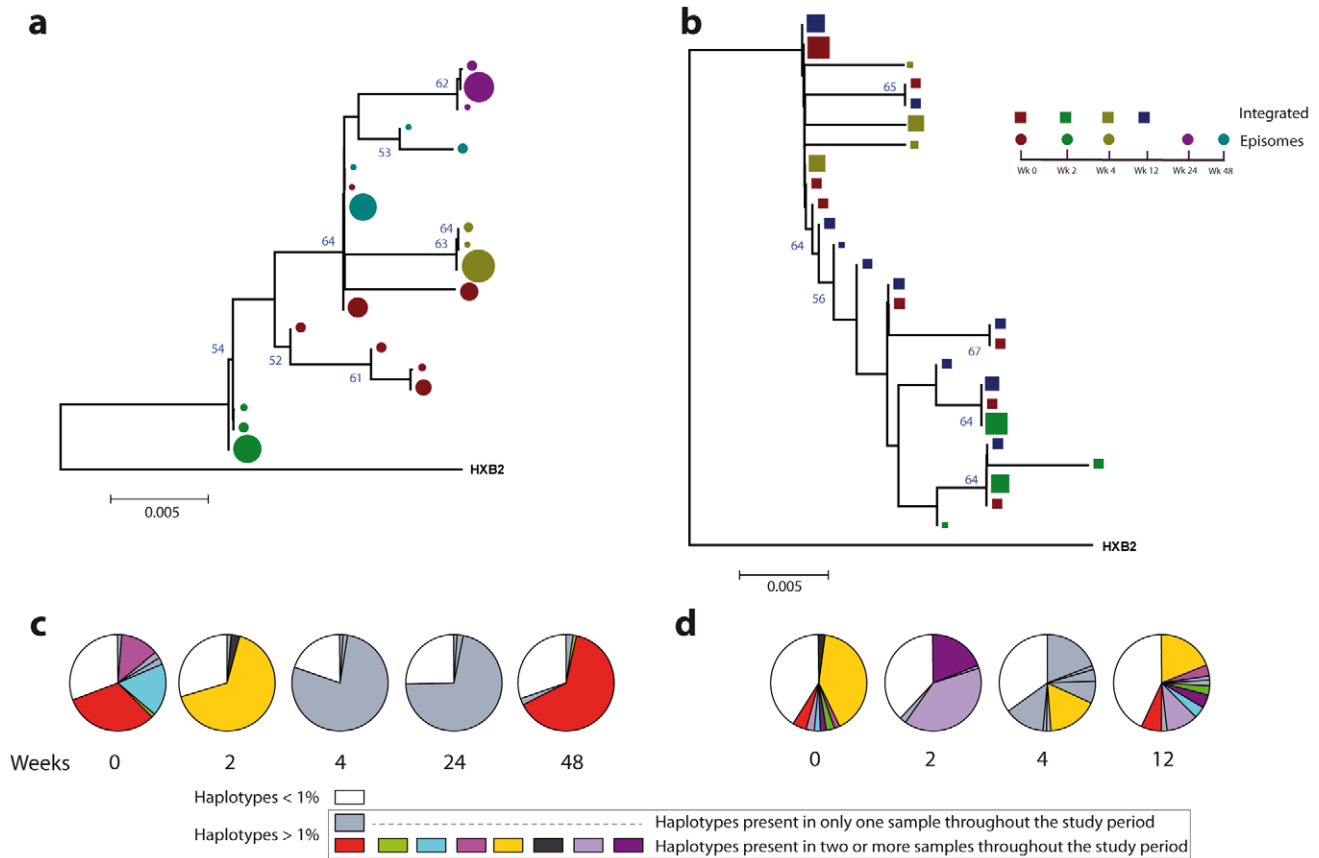
Figure 4. Phylogenetic tree and proportions of episomal and integrated HIV-1 sequences in patient 2. A neighbor-joining approach as implemented in MEGA4 was used to construct a phylogenetic tree with the best evolutionary model found in jModelTest v0.1.1. **a.** Circles represent longitudinal episomal DNA sequences. **b.** Squares represent longitudinal integrated DNA sequences. Legend of phylogenetic trees represents weeks available for further analysis. Sizes of the symbols represent different percentage of clonal sequences. **c.** Longitudinal representation of clonal variability of each episomal sample. **d.** Longitudinal representation of clonal variability of each integrated sample. Areas of pie charts in white shading indicate sequences present with a frequency below 1%; gray shading indicates sequences with a frequency above 1% present in only one sample throughout the study period; colors represent sequences with a frequency above 1% present in two or more samples throughout the study period. 1,000 bootstrap replicates were performed; only values greater than 50% are shown at tree nodes.
doi:10.1371/journal.ppat.1002314.g004

Qiagen). 250 µl of cell suspensions were used for extrachromosomal HIV-1 DNA extraction (Qiaprep miniprep kit, Qiagen) using a modification for the isolation of low-copy-number plasmids. Total cellular DNA was purified from 100 µl of cell resuspension with a standard protocol (QIAamp DNA Blood Kit, Qiagen) as previously described [16].

Table 2. Temporal population structure: AMOVA test for comparison between longitudinal episomal HIV-1 DNA.

	AMOVA test - Episomal HIV-1 DNA					
	F_{st} (*P*-value)					
	W0	**W2**	**W4**	**W12**	**W24**	**W48**
W0	–	*0.787 ($<10^{-6}$)*	*0.107 ($<10^{-6}$)*	*0.238 ($<10^{-6}$)*	*0.764 ($<10^{-6}$)*	*NA*
W2	**0.584 ($<10^{-6}$)**	–	*0.616 ($<10^{-6}$)*	*0.762 ($<10^{-6}$)*	*0.873 ($<10^{-6}$)*	*NA*
W4	**0.664 ($<10^{-6}$)**	**0.998 ($<10^{-6}$)**	–	*0.133 ($<10^{-6}$)*	*0.592 ($<10^{-6}$)*	*NA*
W12	**NA**	**NA**	**NA**	–	*0.655 ($<10^{-6}$)*	*NA*
W24	**0.669 ($<10^{-6}$)**	**0.997 ($<10^{-6}$)**	**0.989 ($<10^{-6}$)**	**NA**	–	*NA*
W48	**0.273 ($<10^{-6}$)**	**0.969 ($<10^{-6}$)**	**0.964 ($<10^{-6}$)**	**NA**	**0.965 ($<10^{-6}$)**	–

The longitudinal comparisons of patient 1 are indicated in the upper-right hemi-matrix of the table (*italics*); those of patient 2 are shown in the lower-left hemi-matrix of the table (**bold**).
doi:10.1371/journal.ppat.1002314.t002

Table 3. Temporal population structure: AMOVA test for comparison between longitudinal integrated HIV-1 DNA.

	AMOVA test - Integrated HIV-1 DNA					
	F_{st} (*P*-value)					
	W0	W2	W4	W12	W24	W48
W0	–	*0.997 ($<10^{-6}$)*	*0.969 ($<10^{-6}$)*	*0.986 $<10^{-6}$*	*0.794 ($<10^{-6}$)*	*NA*
W2	**0.553 ($<10^{-6}$)**	–	*0.985 ($<10^{-6}$)*	*0.077 ($<10^{-6}$)*	*0.681 ($<10^{-6}$)*	*NA*
W4	**0.270 ($<10^{-6}$)**	**0.654 ($<10^{-6}$)**	–	*0.972 ($<10^{-6}$)*	*0.564 ($<10^{-6}$)*	*NA*
W12	**0.127 ($<10^{-6}$)**	**0.316 ($<10^{-6}$)**	**0.350 ($<10^{-6}$)**	–	*0.763 ($<10^{-6}$)*	*NA*
W24	**NA**	**NA**	**NA**	**NA**	–	*NA*
W48	**NA**	**NA**	**NA**	**NA**	**NA**	–

The longitudinal comparisons of patient 1 are indicated in the upper-right hemi-matrix of the table (*italics*); those of patient 2 are shown in the lower-left hemi-matrix of the table (**bold**).
doi:10.1371/journal.ppat.1002314.t003

Amplification of integrated and episomal HIV-1 DNA

Analysis of HIV genomes from a sample containing a low copy number of HIV, such as PBMC from patients with undetectable viral load, can result in a high probability of resampling. The probability of resampling is related to both the number of target molecules during the amplification step and the number of sequenced clones. Therefore the higher the input of target molecules in the PCR and the higher the number of sequenced clones, the less likely the probability of resampling [39]. In order to avoid resampling, we extracted episomal and integrated DNA from a median of 6×10^7 PBMC to increase the number of input molecules during the first PCR. In addition, only samples with individual clonal sequences higher than 1,500 after deep sequencing were considered for further analysis.

We used a two-step PCR to amplify the RT region of episomal and integrated HIV-1 DNA. Primers *Aluf* and *LA7* for integrated DNA and *Jct f* and *LA7* for episomal DNA were used as previously described [10]. Nested PCR amplification of the RT region (codons 150 to 250) was performed as part of the DS protocol (see below). Nested PCR of background controls with primers *DR pol f* and *DR pol r* [10], in parallel to 454 amplification, was carried out to ensure that nested PCR was specific for integrated and episomal DNA viral forms.

Deep HIV-1 sequencing

Pooled, purified PCR products were used as template to generate a single amplicon covering codons 150 to 250 from the RT region. The amplicon library was generated in triplicate during 20 cycles of PCR amplification (Platinum Taq DNA Polymerase High Fidelity, Invitrogen, Carlsbad, CA) followed by pooling and purification of triplicate PCR products using magnetic beads (Agencourt AMPure Kit (Beckman Coulter, Benried, Germany) to eliminate primer-dimers. The number of molecules was quantified by fluorometry (Quant-iT PicoGreen dsDNA assay kit, Invitrogen, Carlsbad, CA). The quality of each amplicon was analyzed by spectrometry using a BioAnalyzer (Agilent Technologies Inc., Santa Clara, CA). Deep Sequencing (DS) was performed in-house on a 454 Life Science/Roche platform. The error rate of the *in-house* DS technique, as inspected with 992 pNL43 clonal sequences obtained with DS under the same conditions as those used for patient samples, was 0.07% (0.13%), which is close to previous reports [40]. This mismatch rate corresponds to a variability rate of 1.69×10^{-5}, within the range of expected PCR error. The 99th percentile of mismatches would establish the threshold for nucleotide errors in 0.61%. Therefore, we decided to include for further analysis only patient clonal sequences present at $\geq1\%$ of the viral population.

Table 4. Temporal population structure: Slatkin-Maddison test for comparison between longitudinal episomal HIV-1 DNA.

	Slatkin-Maddison test - Episomal HIV-1 DNA					
	Migrations (Probability)					
	W0	W2	W4	W12	W24	W48
W0	–	*1/4 (0.006)*	*5/5 (0.698)*	*6/6 (0.810)*	*2/4 (0.028)*	*NA*
W2	**1/3 (0.025)**	–	*1/4 (0.008)*	*1/4 (0.001)*	*1/3 (0.015)*	*NA*
W4	**1/3 (0.017)**	**1/2 (0.097)**	–	*6/7 (0.999)*	*3/4 (0.231)*	*NA*
W12	**NA**	**NA**	**NA**	–	*4/5 (0.496)*	*NA*
W24	**1/3 (0.018)**	**1/2 (0.099)**	**1/2 (0.099)**	**NA**	–	*NA*
W48	**3/4 (0.445)**	**1/3 (0.058)**	**1/3 (0.058)**	**NA**	**1/3 (0.029)**	–

The longitudinal comparisons of patient 1 are indicated in the upper-right hemi-matrix of the table (*italics*); those of patient 2 are shown in the lower-left hemi-matrix of the table (**bold**).
doi:10.1371/journal.ppat.1002314.t004

Table 5. Temporal population structure: Slatkin-Maddison test for comparison between longitudinal integrated HIV-1 DNA.

	Slatkin-Maddison test - Integrated HIV-1 DNA					
	Migrations (Probability)					
	W0	W2	W4	W12	W24	W48
W0	–	*1/2 (0.204)*	*1/2 (0.098)*	*1/2 (0.137)*	*1/2 (0.046)*	*NA*
W2	**3/4 (0.360)**	–	*1/3 (0.034)*	*2/3 (0.425)*	*3/3 (0.999)*	*NA*
W4	**3/4 (0.362)**	**1/3 (0.029)**	–	*1/3 (0.014)*	*2/4 (0.027)*	*NA*
W12	**6/7 (0.909)**	**3/4 (0.999)**	**3/4 (0.333)**	–	*3/4 (0.374)*	*NA*
W24	**NA**	**NA**	**NA**	**NA**	–	*NA*
W48	**NA**	**NA**	**NA**	**NA**	**NA**	–

The longitudinal comparisons of patient 1 are indicated in the upper-right hemi-matrix of the table (*italics*); those of patient 2 are shown in the lower-left hemi-matrix of the table (**bold**).
doi:10.1371/journal.ppat.1002314.t005

Sequence analysis

454 DNA amplicon sequences were aligned with an HXB2 reference sequence using Muscle v3.7 [41] and an independent alignment for each DNA, time-point and patient was built. In order to increase the number of sequences for further analysis and to avoid sequencing errors produced at the end of the sequencing run, we extracted from codon 50 to 209 from each of the sequences obtained with DS. Technical errors of the DS technique drive the introduction of indeterminations (introducing N instead of A, T, G or C) into the sequences. These indetermination were substitute by gaps. An *in-house* method was used to merge clonal sequences into unique sequences. Only clonal sequences present at ≥1% of the clonal population were used. Alignments are available upon request.

Phylogenetic analysis

jModeltest v0.1.1 [42] was used to infer the best phylogenetic model to explain the alignment sequence evolution. This program is able to implement a discrete gamma distribution (Γ) which models the heterogeneity rate among sites. A neighbour-joining approach, as implemented in MEGA4 [18], was used to construct a phylogenetic tree with the best evolutionary model found in jModeltest v0.1.1.

Population structure analysis

In order to detect differences in sequence composition between episomal and integrated DNA at different time-points, we performed two analyses, (i) an analysis of molecular variance (AMOVA) as implemented in the Arlequin software package [21], which is based on pairwise genetic distances and percentage of presence of each clone at each population, and (ii) a tree based topology method, the Slatkin-Maddison test [23] as implemented in HYPHY [43]. The Slatkin-Maddison tests involve estimating the number of migrations between populations, and determining whether the estimated number of migrations is less than expected if there were no compartmentalization. As the maximum possible number of migrations depends on the number of sequences analyzed from each compartment, a randomization test is performed to estimate a p value to assess the significance of compartmentalization. This approach has previously been used to study differences in sequence diversity between different HIV populations [12,19,20]. AMOVA is a genetic distance-based test where the frequency of a sequence variant (haplotype) i in the organ j, x_{ij}, can be expressed as $x_{ij} = x + a_i + b_{ij}$, where a_i and b_{ij} are episomal or integrated DNA and the haplotype within-DNA specific effects, respectively. These two factors have associated variances σ_a^2 and σ_b^2 that can be described as total variance among haplotypes as σ_T^2. The F_{ST} index measures the population differentiation. This value is defined as the ratio between σ_a^2/σ_T^2, and it can be estimated from the usual partition of total variance into its components in a nested analysis of variance (ANOVA) [44]. We carried out AMOVA analysis by computing F_{ST} with a distance matrix obtained from Arlequin program using the best evolutionary model found by jModeltest.

Supporting Information

Figure S1. Longitudinal dynamics of 2-LTR circles and total HIV-1 DNA during the study.
(TIF)

References

1. Palmer S, Wiegand AP, Maldarelli F, Bazmi H, Mican JM, et al. (2003) New real-time reverse transcriptase-initiated PCR assay with single-copy sensitivity for human immunodeficiency virus type 1 RNA in plasma. J Clin Microbiol 41: 4531–4536.

2. Palmer S, Maldarelli F, Wiegand A, Bernstein B, Hanna GJ, et al. (2008) Low-level viremia persists for at least 7 years in patients on suppressive antiretroviral therapy. Proc Natl Acad Sci U S A 105: 3879–3884.

3. Bailey JR, Sedaghat AR, Kieffer T, Brennan T, Lee PK, et al. (2006) Residual human immunodeficiency virus type 1 viremia in some patients on antiretroviral therapy is dominated by a small number of invariant clones rarely found in circulating CD4+ T cells. J Virol 80: 6441–6457.

4. Joos B, Fischer M, Kuster H, Pillai SK, Wong JK, et al. (2008) HIV rebounds from latently infected cells, rather than from continuing low-level replication. Proc Natl Acad Sci U S A 105: 16725–16730.

5. Kieffer TL, Finucane MM, Nettles RE, Quinn TC, Broman KW, et al. (2004) Genotypic analysis of HIV-1 drug resistance at the limit of detection: virus production without evolution in treated adults with undetectable HIV loads. J Infect Dis 189: 1452–1465.

6. Chun TW, Justement JS, Moir S, Hallahan CW, Maenza J, et al. (2007) Decay of the HIV reservoir in patients receiving antiretroviral therapy for extended periods: implications for eradication of virus. J Infect Dis 195: 1762–1764.

7. Chun TW, Nickle DC, Justement JS, Large D, Semerjian A, et al. (2005) HIV-infected individuals receiving effective antiviral therapy for extended periods of time continually replenish their viral reservoir. J Clin Invest 115: 3250–3255.

8. Gunthard HF, Frost SD, Leigh-Brown AJ, Ignacio CC, Kee K, et al. (1999) Evolution of envelope sequences of human immunodeficiency virus type 1 in cellular reservoirs in the setting of potent antiviral therapy. J Virol 73: 9404–9412.

9. Martinez-Picado J, Frost SD, Izquierdo N, Morales-Lopetegi K, Marfil S, et al. (2002) Viral evolution during structured treatment interruptions in chronically human immunodeficiency virus-infected individuals. J Virol 76: 12344–12348.

10. Sharkey M, Triques K, Kuritzkes DR, Stevenson M (2005) In vivo evidence for instability of episomal human immunodeficiency virus type 1 cDNA. J Virol 79: 5203–5210.

11. Sharkey ME, Teo I, Greenough T, Sharova N, Luzuriaga K, et al. (2000) Persistence of episomal HIV-1 infection intermediates in patients on highly active anti-retroviral therapy. Nat Med 6: 76–81.

12. Brennan TP, Woods JO, Sedaghat AR, Siliciano JD, Siliciano RF, et al. (2009) Analysis of human immunodeficiency virus type 1 viremia and provirus in resting CD4+ T cells reveals a novel source of residual viremia in patients on antiretroviral therapy. J Virol 83: 8470–8481.

13. Sahu GK, Sarria JC, Cloyd MW (2010) Recovery of replication-competent residual HIV-1 from plasma of a patient receiving prolonged, suppressive highly active antiretroviral therapy. J Virol 84: 8348–8352.

14. Middleton T, Lim HB, Montgomery D, Rockway T, Tang H, et al. (2004) Inhibition of human immunodeficiency virus type I integrase by naphthamidines and 2-aminobenzimidazoles. Antiviral Res 64: 35–45.

15. Svarovskaia ES, Barr R, Zhang X, Pais GC, Marchand C, et al. (2004) Azido-containing diketo acid derivatives inhibit human immunodeficiency virus type 1 integrase in vivo and influence the frequency of deletions at two-long-terminal-repeat-circle junctions. J Virol 78: 3210–3222.

16. Buzon MJ, Massanella M, Llibre JM, Esteve A, Dahl V, et al. (2010) HIV-1 replication and immune dynamics are affected by raltegravir intensification of HAART-suppressed subjects. Nat Med 16: 460–465.

17. Llibre J, Buzón M, Massanella M, Esteve A, Dahl V, et al. (2011) Treatment intensification with raltegravir in subjects with sustained HIV-1 viremia suppression: a randomized 48 weeks study. Antivir Ther: In press.

18. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596–1599.

19. Borderia AV, Codoner FM, Sanjuan R (2007) Selection promotes organ compartmentalization in HIV-1: evidence from gag and pol genes. Evolution 61: 272–279.

20. Sanjuan R, Codoner FM, Moya A, Elena SF (2004) Natural selection and the organ-specific differentiation of HIV-1 V3 hypervariable region. Evolution 58: 1185–1194.

21. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131: 479–491.

22. Zarate S, Pond SL, Shapshak P, Frost SD (2007) Comparative study of methods for detecting sequence compartmentalization in human immunodeficiency virus type 1. J Virol 81: 6643–6651.

23. Slatkin M, Maddison WP (1989) A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics 123: 603–613.

24. Yukl SA, Shergill AK, McQuaid K, Gianella S, Lampiris H, et al. (2010) Effect of raltegravir-containing intensification on HIV burden and T-cell activation in multiple gut sites of HIV-positive adults on suppressive antiretroviral therapy. Aids 24: 2451–2460.

25. Bull ME, Learn GH, McElhone S, Hitti J, Lockhart D, et al. (2009) Monotypic human immunodeficiency virus type 1 genotypes across the uterine cervix and in blood suggest proliferation of cells with provirus. J Virol 83: 6020–6028.

26. Chun TW, Carruth L, Finzi D, Shen X, DiGiuseppe JA, et al. (1997) Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. Nature 387: 183–188.

27. Chun TW, Finzi D, Margolick J, Chadwick K, Schwartz D, et al. (1995) In vivo fate of HIV-1-infected T cells: quantitative analysis of the transition to stable latency. Nat Med 1: 1284–1290.

28. Chun TW, Stuyver L, Mizell SB, Ehler LA, Mican JA, et al. (1997) Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. Proc Natl Acad Sci U S A 94: 13193–13197.

29. Finzi D, Blankson J, Siliciano JD, Margolick JB, Chadwick K, et al. (1999) Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. Nat Med 5: 512–517.

30. Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, et al. (1997) Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science 278: 1295–1300.

31. Wong JK, Hezareh M, Gunthard HF, Havlir DV, Ignacio CC, et al. (1997) Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. Science 278: 1291–1295.

32. Zhu W, Jiao Y, Lei R, Hua W, Wang R, et al. (2011) Rapid Turnover of 2-LTR HIV-1 DNA during Early Stage of Highly Active Antiretroviral Therapy. PLoS One 6: e21081.

33. Tokoyoda K, Zehentmeier S, Hegazy AN, Albrecht I, Grun JR, et al. (2009) Professional memory CD4+ T lymphocytes preferentially reside and rest in the bone marrow. Immunity 30: 721–730.

34. Chomont N, El-Far M, Ancuta P, Trautmann L, Procopio FA, et al. (2009) HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. Nat Med 15: 893–900.

35. Chun TW, Davey RT, Jr., Ostrowski M, Shawn Justement J, Engel D, et al. (2000) Relationship between pre-existing viral reservoirs and the re-emergence of plasma viremia after discontinuation of highly active anti-retroviral therapy. Nat Med 6: 757–761.

36. Noe A, Plum J, Verhofstede C (2005) The latent HIV-1 reservoir in patients undergoing HAART: an archive of pre-HAART drug resistance. J Antimicrob Chemother 55: 410–412.

37. Verhofstede C, Noe A, Demecheleer E, De Cabooter N, Van Wanzeele F, et al. (2004) Drug-resistant variants that evolve during nonsuppressive therapy persist in HIV-1-infected peripheral blood mononuclear cells after long-term highly active antiretroviral therapy. J Acquir Immune Defic Syndr 35: 473–483.

38. Grossman Z, Polis M, Feinberg MB, Levi I, Jankelevich S, et al. (1999) Ongoing HIV dissemination during HAART. Nat Med 5: 1099–1104.

39. Liu SL, Rodrigo AG, Shankarappa R, Learn GH, Hsu L, et al. (1996) HIV quasispecies and resampling. Science 273: 415–416.

40. Varghese V, Shahriar R, Rhee SY, Liu T, Simen BB, et al. (2009) Minority variants associated with transmitted and acquired HIV-1 nonnucleoside reverse transcriptase inhibitor resistance: implications for the use of second-generation nonnucleoside reverse transcriptase inhibitors. J Acquir Immune Defic Syndr 52: 309–315.

41. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797.

42. Posada D (2008) jModelTest: phylogenetic model averaging. Mol Biol Evol 25: 1253–1256.

43. Kosakovsky Pond SL, Frost SD (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol 22: 1208–1222.

44. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 1: 47–50.

45. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10: 512–526.