Correction

IMMUNOLOGY AND INFLAMMATION

Correction for "Paired quantitative and qualitative assessment of the replication-competent HIV-1 reservoir and comparison with integrated proviral DNA," by Julio C. C. Lorenzi, Yehuda Z. Cohen, Lillian B. Cohn, Edward F. Kreider, John P. Barton, Gerald H. Learn, Thiago Oliveira, Christy L. Lavine, Joshua A. Horwitz, Allison Settler, Mila Jankovic, Michael S. Seaman, Arup K. Chakraborty, Beatrice H. Hahn, Marina Caskey, and Michel C. Nussenzweig, which appeared in issue 49, December 6, 2016, of *Proc Natl Acad Sci USA* (113:E7908–E7916; first published November 21, 2016; 10.1073/pnas.1617789113).

The authors note that Fig. 3 appeared incorrectly. The corrected figure and its legend appear below.

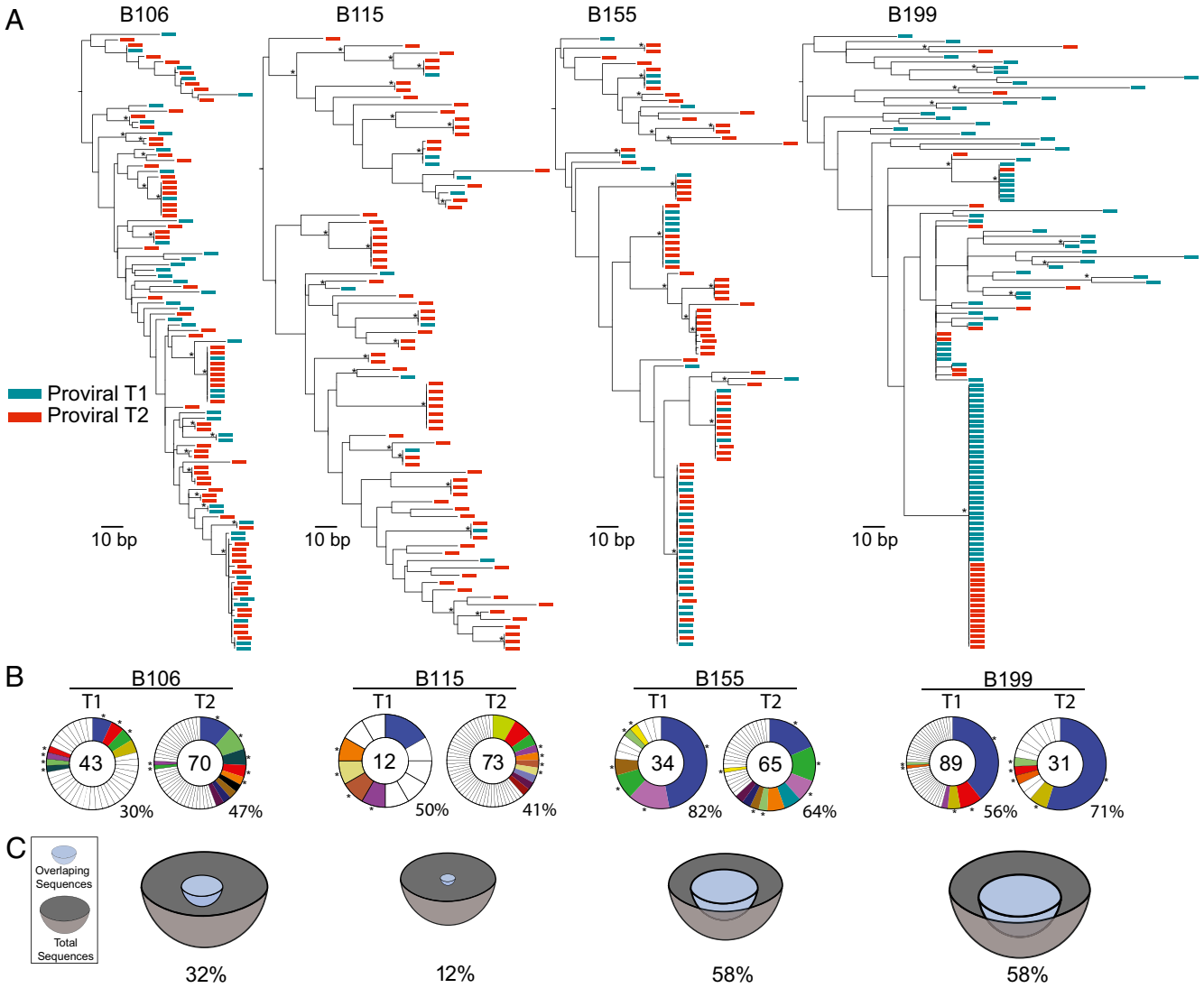

Fig. 3. *Env* sequences from archived proviral DNA. (*A*) Maximum likelihood phylogenetic trees of full-length *env* sequences derived by SGA from primary CD4$^+$ T cells from four individuals. Viruses from time point 1 are green, and viruses from time point 2 are red. Asterisks indicate nodes with significant bootstrap values (bootstrap support $\geq$ 70%). (*B*) Pie charts depict the distribution of archived *env* sequences from the two time points. The number in the inner circle indicates the total number of *env* sequences analyzed. White represents unique sequences isolated only once across both time points, and colored areas represent identical sequences that appear more than once. The size of the pie slice is proportional to the number of sequences in the clone. Clones found at both time points are the same color and denoted by asterisks. Percentages of identical sequences are displayed at the bottom right of each pie chart. (*C*) Representation of overlapping sequences between the two time points. The size of the hemisphere is proportional to the number of sequences. Light blue hemispheres represent overlapping sequences and gray hemispheres represent the total number of sequences. The percentage of overlap is indicated at the bottom of each hemisphere.

www.pnas.org/cgi/doi/10.1073/pnas.1619801114

Paired quantitative and qualitative assessment of the replication-competent HIV-1 reservoir and comparison with integrated proviral DNA

Julio C. C. Lorenzi[a,1], Yehuda Z. Cohen[a,1], Lillian B. Cohn[a], Edward F. Kreider[b,c], John P. Barton[d,e,f,g,h], Gerald H. Learn[b,c], Thiago Oliveira[a], Christy L. Lavine[i], Joshua A. Horwitz[a], Allison Settler[a], Mila Jankovic[a], Michael S. Seaman[i], Arup K. Chakraborty[d,e,f,g,h], Beatrice H. Hahn[b,c], Marina Caskey[a], and Michel C. Nussenzweig[a,j,2]

[a]Laboratory of Molecular Immunology, The Rockefeller University, New York, NY 10065; [b]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104; [c]Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104; [d]Department of Chemical Engineering, Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA 02139; [e]Department of Physics, Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA 02139; [f]Department of Chemistry, Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA 02139; [g]Department of Biological Engineering, Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA 02139; [h]Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA 02139; [i]Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215; and [j]Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10065

Contributed by Michel C. Nussenzweig, October 27, 2016 (sent for review October 17, 2016; reviewed by Sharon R. Lewin and Douglas D. Richman)

HIV-1–infected individuals harbor a latent reservoir of infected CD4$^+$ T cells that is not eradicated by antiretroviral therapy (ART). This reservoir presents the greatest barrier to an HIV-1 cure and has remained difficult to characterize, in part, because the vast majority of integrated sequences are defective and incapable of reactivation. To characterize the replication-competent reservoir, we have combined two techniques, quantitative viral outgrowth and qualitative sequence analysis of clonal outgrowth viruses. Leukapheresis samples from four fully ART-suppressed, chronically infected individuals were assayed at two time points separated by a 4- to 6-mo interval. Overall, 54% of the viruses emerging from the latent reservoir showed gp160 *env* sequences that were identical to at least one other virus. Moreover, 43% of the *env* sequences from viruses emerging from the reservoir were part of identical groups at the two time points. Groups of identical expanded sequences made up 54% of proviral DNA, and, as might be expected, the sequences of replication-competent viruses in the active reservoir showed limited overlap with integrated proviral DNA, most of which is known to represent defective viruses. Finally, there was an inverse correlation between proviral DNA clone size and the probability of reactivation, suggesting that replication-competent viruses are less likely to be found among highly expanded provirus-containing cell clones.

HIV | reservoir | replication-competent | culture | method

A reservoir of latently infected cells persists in HIV-1–infected individuals treated with antiretroviral therapy (ART) (1). This reservoir endures for the lifetime of the individual and presents the greatest barrier to an HIV-1 cure (2, 3). Although there is a growing understanding of the cellular and molecular nature of this compartment, many questions remain about the composition of the latent reservoir and the ability of current techniques to characterize it accurately (4, 5).

One of the challenges in studying the reservoir is that the majority (>90%) of integrated proviruses in CD4$^+$ T-cell DNA are defective and cannot produce infectious virions (6–9). Comparatively few cells harbor the replication-competent proviruses that constitute the clinically relevant reservoir, and there are currently no markers to distinguish these cells from those cells bearing defective proviruses. An additional problem is that it is difficult to measure the size of the replication-competent reservoir accurately. The best available assay measures the size of the reservoir by limiting dilution cultures under conditions that favor latent virus outgrowth [quantitative viral outgrowth assay (QVOA)] (10, 11). However, this assay is typically performed on peripheral blood and can significantly underestimate the size of the replication-competent reservoir even within this compartment (7, 9).

Finally, the genetic characteristics of the reservoir have been studied primarily by sequencing integrated proviral DNA or cell-associated RNAs, many of which are defective and thus not representative of the replication-competent reservoir (12–16). Phylogenetic analyses of the replication-competent reservoir are usually limited because bulk outgrowth cultures produce species with little diversity, even when analyzed by ultra-deep sequencing (17–20).

Here, we report on a modified QVOA that includes a qualitative measure of the reservoir [qualitative and quantitative viral outgrowth assay (Q^{2}VOA)]. We use the assay to describe the genetic and biological diversity of the replication-competent reservoir and examine the relationship between replicating and archived proviruses in CD4$^+$ T cells from the same ART-suppressed individuals at two time points separated by 4–6 mo.

Results

To investigate the genetic and phenotypic complexity of the replication-competent reservoir, we modified the QVOA protocol to

Significance

A reservoir of latently infected cells poses the greatest challenge to HIV-1 eradication. Efforts to develop strategies to eliminate the reservoir have been hampered, in part, by the lack of a precise understanding of the cellular and molecular nature of this reservoir. We describe a new method to analyze the replication-competent latent reservoir quantitatively and qualitatively. We find that over 50% of the replication-competent viruses in the reservoir form part of groups with identical *env* sequences. However, a negative correlation exists between integrated proviral clones and replication-competent viruses, such that the larger the proviral clone, the lower is its probability of representing a replication-competent virus.

Author contributions: J.C.C.L., Y.Z.C., L.B.C., J.A.H., M.C., and M.C.N. designed research; J.C.C.L., Y.Z.C., C.L.L., and A.S. performed research; E.F.K., J.P.B., G.H.L., and A.K.C. contributed new reagents/analytic tools; J.C.C.L., Y.Z.C., E.F.K., J.P.B., G.H.L., T.O., M.J., M.S.S., A.K.C., B.H.H., and M.C.N. analyzed data; and J.C.C.L., Y.Z.C., L.B.C., and M.C.N. wrote the paper.

Reviewers: S.R.L., University of Melbourne; and D.D.R., University of California, San Diego.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. KY113379–KY114054).

[1]J.C.C.L. and Y.Z.C. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: nussen@rockefeller.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1617789113/-/DCSupplemental.

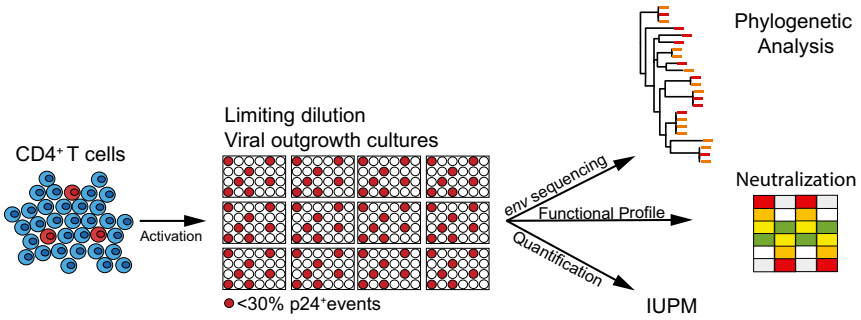

Fig. 1. Quantitative and qualitative analysis of the replication-competent reservoir. Diagrammatic representation of the assay. $CD4^+$ T cells are cultured at a limiting dilution under conditions whereby a single virus emerges from the latent reservoir in each positive well (red). The number of infectious units per million (IUPM) is determined directly from the number of p24-positive wells. Virus-containing supernatants from positive cultures are harvested for *env* sequencing and neutralization assays.

increase the number of unique outgrowth cultures and sequenced the emerging viruses. Unlike QVOA, where multiple dilutions are assayed, Q^2VOA is performed using a single predetermined dilution that produces less than 30% positive wells to maximize the total number of individual viruses that can be sequenced. Based on Poisson distribution, this technique produces cultures that are likely to contain single replication-competent proviruses (Fig. 1).

$CD4^+$ T lymphocytes were isolated from each of four chronically infected individuals who had been virologically suppressed by combination ART for 4–22 y at two time points 4–6 mo apart (Table S1). We tested 0.40–1.44 × 10^8 $CD4^+$ T lymphocytes from each ART-treated individual at each time point. On average, 13.5% of cultures were positive for p24. The number of cells yielding replication-competent viruses varied across individuals from 0.19 to 1.07 infectious units per million, which is similar to values obtained by others (3, 6) (Table 1).

To characterize the cultured viruses molecularly, we produced cDNA from culture supernatants and sequenced the *env* gene using primers that resulted in a clonal prediction score of 94 of 100 (silicianolab.johnshopkins.edu/cps) (21). Thus, there was a high probability that identical *env* sequences represented identical full-length genomes. We obtained a total of 234 *env* sequences from Q^2VOA, of which 13.7% were excluded from further analysis due to the presence of short reads (3.8%) or the presence of reads producing an inconclusive consensus (9.8%). The phylogenetic analysis of the remaining 202 *env* sequences showed that the four individuals were infected with epidemiologically unrelated clade B viruses (Fig. S1). Phylogenetic analysis of individual sequences revealed the existence of a diverse viral population composed of multiple (bootstrap-supported) clusters for each of the four individuals (Fig. 2*A*).

To compare the diversity of viruses obtained from a "bulk" culture with the diversity of viruses derived by Q^2VOA, we performed single genome analysis (SGA) on bulk culture supernatants established at the same time from the same four individuals. In contrast to Q^2VOA, bulk culture supernatants were mainly monotypic and showed much reduced overall diversity (Fig. 2*A*). Thus, when multiple infected cells are reactivated in a single culture, the strain or strains with the fastest growth kinetics, or the greatest fitness in culture, dominate.

Individual replication-competent viruses obtained from different Q^2VOA cultures frequently encoded identical *env* sequences. When Q^2VOA-derived viruses obtained at the two time points were compared for each subject, typically less than half (40–52%) of their *env* sequences were unique (Table 2). The majority of sequences were identical to at least one other independently derived replication-competent virus obtained from the same subject. For example, of a total of 49 *env* sequences isolated from the two time points from B106, only 21 were unique. The majority (22) were identical to at least one other sequence that appeared at one of the two time points. We will refer to these repeated sequences irrespective of whether they appear at one or both time points as "clones" because they must originate from at least two different $CD4^+$ T cells. This finding does not necessarily imply that the viruses are integrated in the same location in the genome because it is possible that identical viruses can infect different $CD4^+$ T cells. The size of these clones ranged from two to 10 members, with a mean of 3.69 when all four individuals and both time points were considered (Fig. 2 *B* and *C* and Table 2). Fifty-four percent of all replication-competent viruses emerging in Q^2VOA cultures were derived from expanded clones (Fig. 2 *B* and *C* and Table 2).

To determine whether the viral sequences obtained from the replication-competent reservoir remained stable over time, we compared the sequences from the two time points for each individual. Many branches in the phylogenetic trees contained sequences derived from both time points (Fig. 2 *A*–*C* and Table 2). The relationship between the sequences from both time points was formally assessed by determining their Genealogical Sorting Index (GSI), which quantitates the degree of phylogenetic association between sequences (23). GSI values, which range between 0 (complete interspersion) and 1 (complete monophyly), showed that for each of the four individuals, the sequences obtained by Q^2VOA from both time points could not be segregated as distinct groups (Table 3). This finding demonstrates that the viral population

Table 1. Q^2VOA overall results and IUPM

	Months between time points	Total $CD4^+$ cells tested	Wells tested	Positive wells (%)	IUPM	Total $CD4^+$ cells tested	Wells tested	Positive wells (%)	IUPM
Study ID		Time point 1				Time point 2			
B106	4	39.6 × 10^6	132	31 (23.5)	0.89	75.6 × 10^6	252	28 (11.1)	0.39
B115	4	57.6 × 10^6	192	40 (20.8)	0.57	140 × 10^6	468	50 (10.7)	0.38
B155	6	75.6 × 10^6	252	69 (27.4)	1.07	72 × 10^6	240	24 (16.7)	0.35
B199	4	43.2 × 10^6	144	24 (16.7)	0.61	144 × 10^6	480	26 (5.4)	0.19

ID, identification; IUPM, infectious units per million cells.

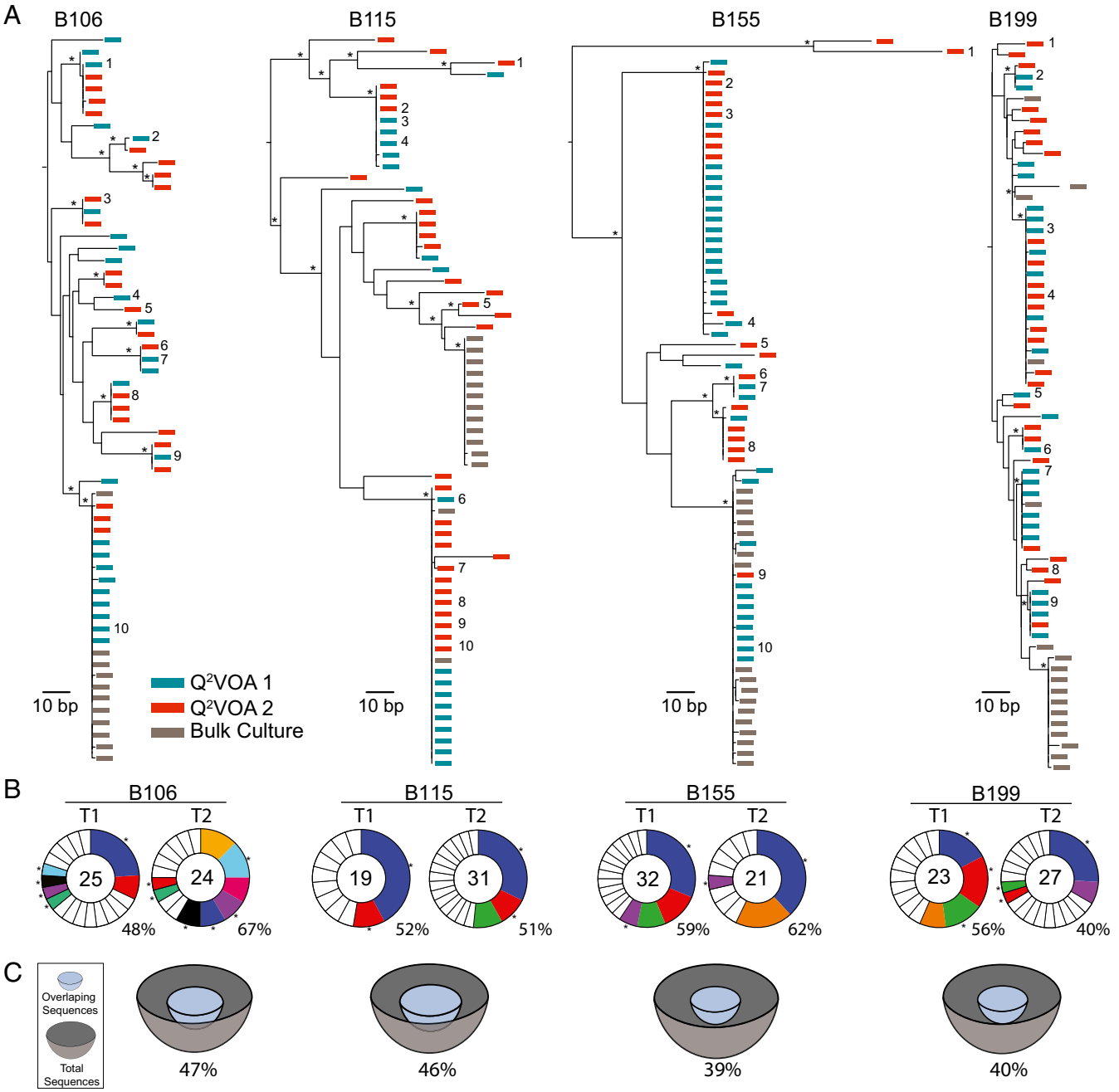

Fig. 2. *Env* sequences from outgrowth cultures. (*A*) Maximum likelihood phylogenetic trees of full-length *env* sequences of viruses from Q^{2}VOA outgrowth cultures from four individuals. Viruses from time point 1 are green, viruses from time point 2 are red, and bulk culture SGA is gray. Asterisks indicate nodes with significant bootstrap values (bootstrap support ≥ 70%). Numbers next to sequences correspond to viruses assayed for neutralization in Fig. 6. (*B*) Pie charts depict the distribution of culture-derived *env* sequences from the two time points. The number in the inner circle indicates the total number of *env* sequences analyzed. White represents sequences isolated only once across both time points, and colored areas represent identical sequences that appear more than once. The size of the pie slice is proportional to the number of sequences in the clone. Clones found at both time points are the same color and denoted by asterisks. Percentages of identical sequences are displayed at the bottom right of each pie chart. (*C*) Representation of overlapping sequences between the two time points. The size of the hemisphere is proportional to the number of sequences. Light blue hemispheres represent overlapping sequences and gray hemispheres represent the total number of sequences. The percentage of overlap is indicated at the bottom of each hemisphere.

emerging from the latent reservoir in four individuals was stable over the 4- to 6-mo time interval analyzed.

To examine the relationship between proviruses integrated into CD4$^+$ T-cell DNA and the replication-competent viruses obtained by Q^{2}VOA, we performed SGA on DNA isolated from primary CD4$^+$ T lymphocytes from the same individuals at both time points. We obtained a total of 498 *env* sequences, of which 16.3% were excluded from further analysis due to the presence of hypermutated regions (2.4%), short reads (5.6%), or reads producing an inconclusive consensus (8.2%). The remaining 417 full-length *env* sequences, or 85–113 per individual, fell within the same four patient-specific clades as the Q^{2}VOA-derived *env* sequences, indicating absence of sample mix-up or contamination (Fig. S1).

Table 2. Distribution of observed sequences in Q²VOA

	Nonclonal			Clonal			
Study ID	T1	T2	Total (%)	T1	T2	Total (%)	Overlapping sequences between T1 and T2 (%)
B106	13	8	21/49 (42.9)	12	16	28/49 (57.1)	23/49 (46.9)
B115	9	15	24/50 (48.0)	10	16	26/50 (52.0)	23/50 (46.0)
B155	13	8	21/53 (39.6)	19	13	32/53 (60.4)	21/53 (39.6)
B199	10	16	26/50 (52.0)	13	11	24/50 (48.0)	20/50 (40.0)

As previously reported for archived proviral DNA amplified from primary CD4⁺ T cells, we found unique sequences, as well as large groups of identical sequences, marking expanded cell clones (14, 15, 24, 25). For example, in B106, 67 of 113 *env* sequences obtained from the two time points were unique and the remaining 46 were members of clones. In addition, 36 of the 46 identical sequences overlapped between the two time points (Fig. 3 *A* and *B* and Table 4). Thus, like the Q²VOA-derived sequences, the archived proviral population was stable over the time interval analyzed.

Archived proviral DNA sequences were then compared with replication-competent Q²VOA-derived sequences. Because both the proviral DNA and the replication-competent viral sequences for any given individual were overlapping at the two time points, we combined each of the two sets of sequences (Fig. 4). As might be expected, given that the sample size is limited, we found relatively limited overlap between the archived proviral sequences and Q²VOA culture-derived sequences in phylogenetic trees, with multiple large clusters composed of sequences isolated from only a single source (Fig. 4*A*). The 417 archived proviral sequences contained 50 expanded clones, 12 of which were also found in the replication-competent outgrowth cultures. The relative frequency of expanded clones from these two sources differed: Some greatly expanded clones identified in primary CD4⁺ T-cell DNA represented only a small fraction of clones identified in the outgrowth cultures, and rare archived clones were disproportionally abundant among the reactivated latent replication-competent viruses (Fig. 4*B*). This finding is best illustrated in subject B155, where the largest archived proviral clone, which comprised 28 of 99 total sequences, was found only once among 52 replication-competent viruses (Fig. 4 *B* and *C*). In contrast, in B199, one archived proviral sequence, which appeared only once in a total of 120 sequences derived from primary CD4⁺ T-cell DNA, was found in 11 of 50 Q²VOA culture-derived sequences emerging from the latent reservoir. B115 provided the clearest example of the discrepancy between proviral DNA and cultured viruses, with no instances of matching sequences between 85 proviral sequences and 50 outgrowth viruses (Fig. 4 *B* and *C*). Although this discrepancy is most likely due to the prevalence of defective archived proviral sequences (6–9), differences in proviral accessibility to polymerase may also contribute.

To determine the extent to which sequences from replication-competent viruses and archived proviruses were compartmentalized, we calculated their GSI. This analysis demonstrated that archived proviral and Q²VOA culture-derived sequences were significantly segregated for every single individual analyzed (Table 5). Thus, proviral sequences derived from primary CD4⁺ T-cell DNA do not provide an accurate representation of viruses comprising the replication-competent reservoir.

Table 3. GSI and probability values for HIV *env* trees under the null hypothesis that Q²VOA-derived sequences from both visits are a single mixed group

	Q²VOA			
	Visit 1		Visit 2	
Study ID	GSI	*P* value	GSI	*P* value
B106	0.067	0.45	0.02	0.92
B115	0.066	0.14	0	1
B155	0.059	0.26	0.012	0.82
B199	0.053	0.3	0	1

To understand the relationship between the two groups of sequences better, we analyzed them by a mathematical model that describes every sequence by two variables (*SI Methods*). The first (p) is the frequency of a sequence in the proviral compartment, and the second (r) is its probability of reactivation in the active viral culture. These parameters were extracted from the experimental data using Bayesian inference methods. The data show that clone size is negatively correlated with the activation probability ($r = -0.94$, $p = 3.4 \times 10^{-34}$) (Fig. 5). These results indicate that the larger the clone of archived proviral sequences, the lower is its probability of representing a replication-competent virus in the Q²VOA outgrowth culture.

Finally, to characterize the viruses obtained by Q²VOA further, we tested virus-containing supernatants selected to be representative of the diversity in the phylogenetic trees for sensitivity to a panel of broadly neutralizing antibodies (bNAbs) that are currently in clinical development. As might be expected, there was significant variation in the sensitivity profile of each individual's replication-competent reservoir. In each case, there was at least one example of coexisting sensitive and resistant viruses for any one tested bNAb (Fig. 6*A*). For example, in B155, all but one of the 10 viruses tested were resistant to 10-1074, a bNAb that targets the base of the V3 loop and surrounding glycans (26). As might be expected, the 10-1074–susceptible virus segregated from the others in the phylogenetic tree (Fig. 6*A* and Fig. S2). When results from all individuals were combined, a large range in sensitivities was observed against the panel of bNAbs (Fig. 6*B*). In contrast, supernatants obtained from the bulk cultures showed bNAb sensitivity that appeared to be representative of a single viral cluster in any individual (Fig. 6*A*). Thus, Q²VOA differs from bulk cultures in that it enables measurement of the genetic and phenotypic diversity of the replication-competent reservoir. Finally, the observation that bNAb neutralization-sensitive and -resistant clones coexist in the reservoir provides an explanation for the observation that combination therapy will likely be required to maintain suppression (27, 28).

Discussion

A long-lived reservoir of HIV-1–infected cells that remain silent in infected individuals receiving ART is the primary impediment to HIV-1 cure (22, 29, 30). The reservoir is established early in infection (31) and has an estimated $t_{1/2}$ of 44 mo (2, 3). Several mechanisms have been proposed to account for the persistence of this reservoir, including low-level viral replication, the long $t_{1/2}$ of latently infected cells, and the proliferative expansion of these latently infected cells (5, 32, 33).

Arguments against persistent low-level viral replication include the observation that further intensification of ART has no measurable effect (34, 35). In addition, there is little evidence for viral evolution even after prolonged periods of ART (36). In contrast,

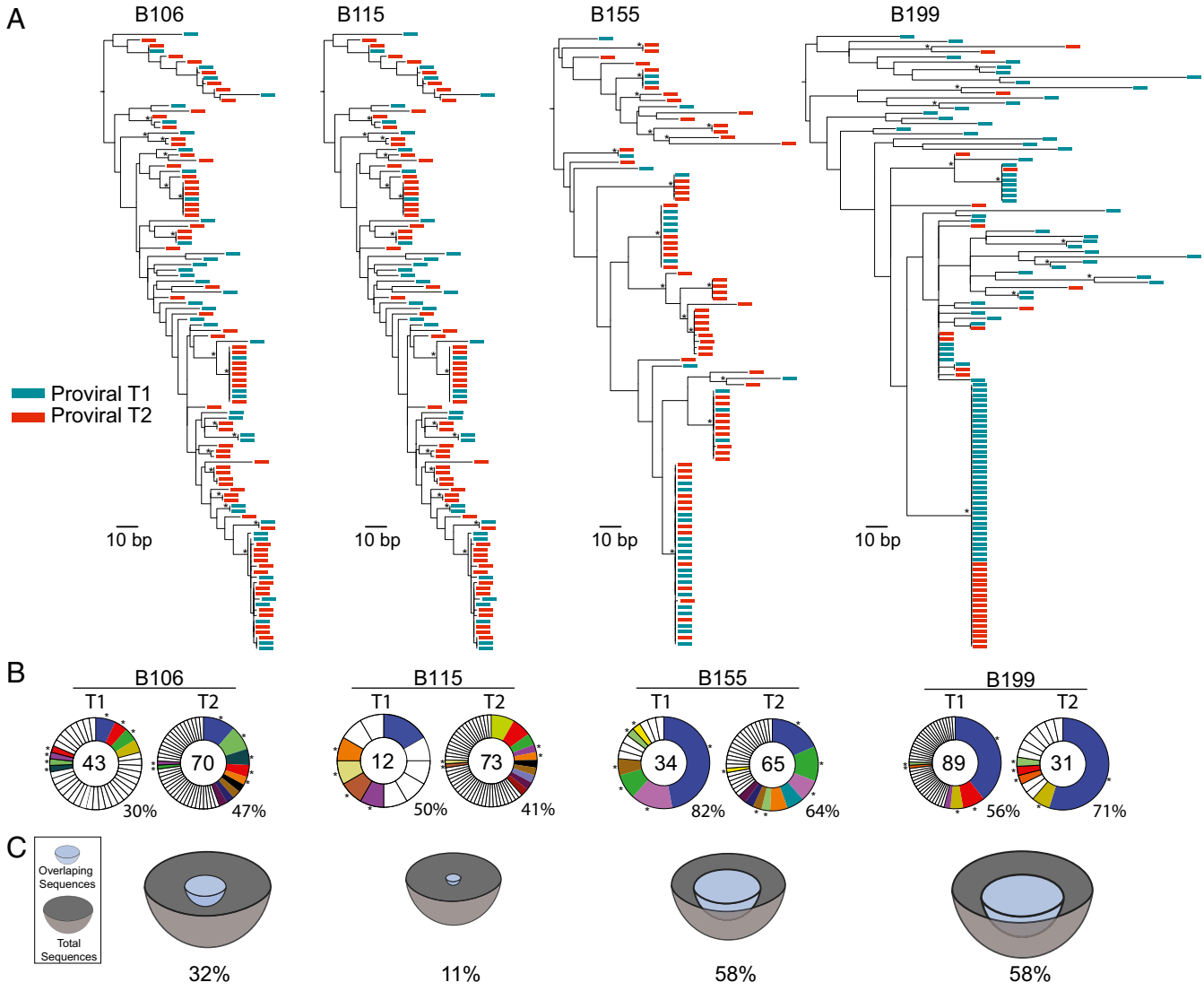

Fig. 3. *Env* sequences from archived proviral DNA. (*A*) Maximum likelihood phylogenetic trees of full-length *env* sequences derived by SGA from primary CD4$^+$ T cells from four individuals. Viruses from time point 1 are green, and viruses from time point 2 are red. Asterisks indicate nodes with significant bootstrap values (bootstrap support $\geq$ 70%). (*B*) Pie charts depict the distribution of archived *env* sequences from the two time points. The number in the inner circle indicates the total number of *env* sequences analyzed. White represents unique sequences isolated only once across both time points, and colored areas represent identical sequences that appear more than once. The size of the pie slice is proportional to the number of sequences in the clone. Clones found at both time points are the same color and denoted by asterisks. Percentages of identical sequences are displayed at the bottom right of each pie chart. (*C*) Representation of overlapping sequences between the two time points. The size of the hemisphere is proportional to the number of sequences. Light blue hemispheres represent overlapping sequences and gray hemispheres represent the total number of sequences. The percentage of overlap is indicated at the bottom of each hemisphere.

the idea that the reservoir is maintained, at least in part, by the proliferation of latently infected cells is supported by the observation of increasing proportions of identical proviral sequences in circulating CD4$^+$ T cells (15).

Archived proviral integration site analysis provided further support for clonal expansion of infected T cells (8, 24, 25). In all cases examined, a large proportion of the archived integrated proviruses were found to belong to expanded clones of CD4$^+$ T cells that

Table 4. Distribution of observed sequences in proviral DNA

	Nonclonal			Clonal			Overlapping
Study ID	T1	T2	Total (%)	T1	T2	Total (%)	Sequences between T1 and T2 (%)
B106	30	37	67/113 (59.3)	13	33	46/113 (40.7)	36/113 (31.9)
B115	6	43	49/85 (57.6)	6	30	36/85 (42.4)	10/85 (11.8)
B155	6	23	29/99 (29.3)	28	42	70/99 (70.7)	58/99 (58.6)
B199	39	9	48/120 (40.0)	50	22	72/120 (60.0)	70/120 (58.3)

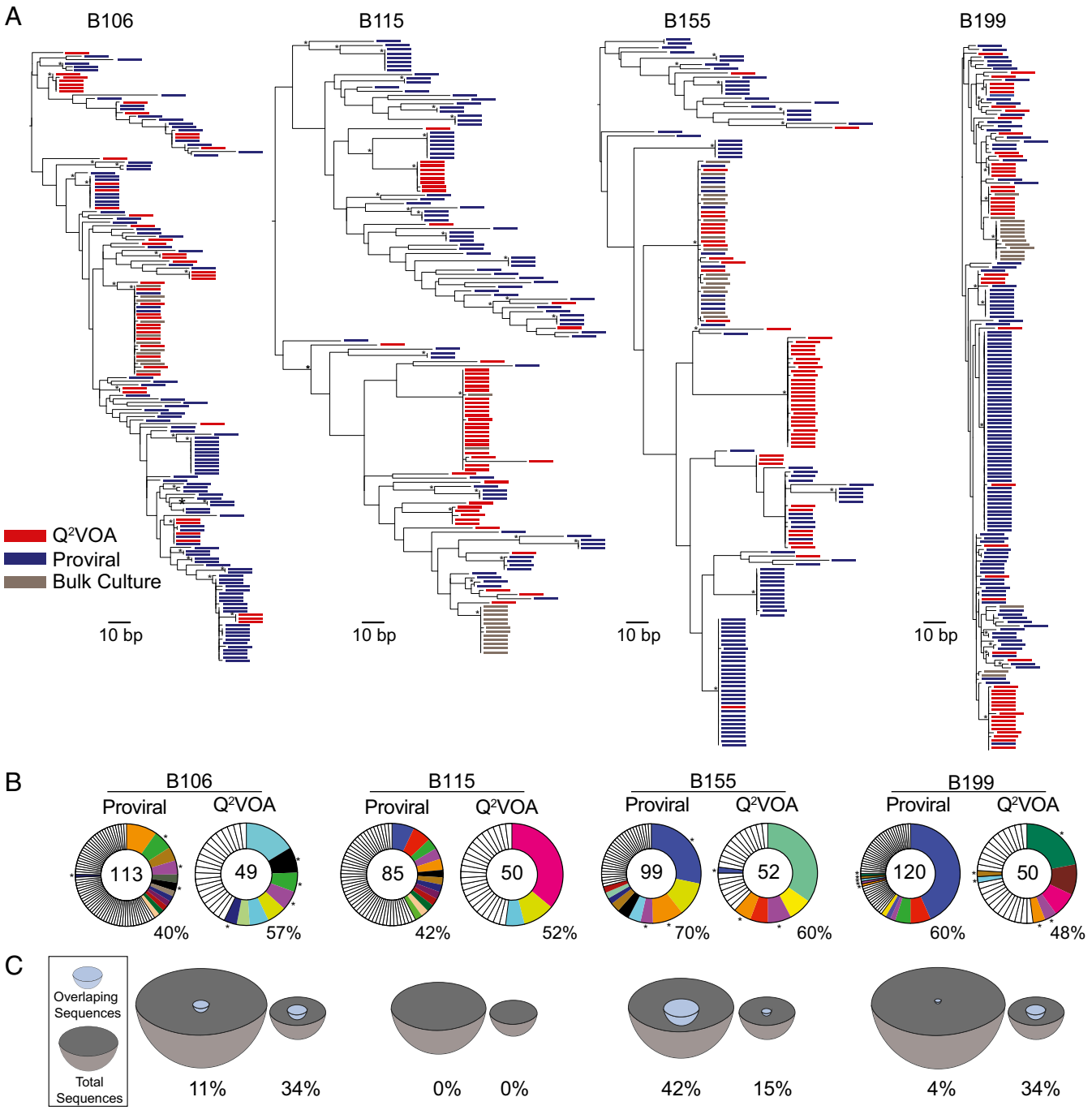

Fig. 4. Comparison of *env* sequences from archived proviruses and replication-competent viruses. Sequences from the two time points were pooled for each patient. (*A*) Maximum likelihood phylogenetic trees of *env* sequences. Limiting dilution outgrowth viruses are red, bulk culture viruses are gray, and viral sequences amplified from primary $CD4^+$ T cells are blue. Asterisks indicate nodes with significant bootstrap values (bootstrap support ≥ 70%). (*B*) Pie charts depicting the distribution of archived proviruses and culture-derived sequences. The numbers in the inner circles indicate the total number of *env* sequences analyzed. White represents sequences isolated only once, and colored areas represent identical sequences. The size of the pie slice is proportional to the number of sequences in the clone. Clones found in proviral DNA and outgrowth cultures are the same color and denoted by asterisks. Percentages of identical groups of sequences are displayed at the bottom right of each pie chart. (*C*) Representation of overlapping sequences between the two sources. The size of the hemisphere is proportional to the number of sequences. Light blue hemispheres represent overlapping sequences and gray hemispheres represent the total number of sequences. The percentage of overlap is indicated at the bottom of each hemisphere.

share a unique integration site. However, as expected from the observation that the vast majority of integrated proviruses are defective (7, 9), many of the proviruses found in clonally expanded T cells are also defective (8). The only exception was found in a singular individual with metastatic squamous cell carcinoma, which showed a large expanded clone of replication-competent virus (37). However, the integration site of the provirus in this individual was ambiguous and could not be mapped with certainty. Thus, the precise contribution of expanded cell clones bearing replication-competent proviruses to maintaining the latent reservoir is not defined well.

Table 5. GSI and probability values for HIV *env* trees under the null hypothesis that Q²VOA- derived sequences and proviral-derived sequences are a single mixed group

	Q²VOA		Proviral	
Study ID	GSI	*P* value	GSI	*P* value
B106	0.184	<0.01	0.114	<0.001
B115	0.61	<0.001	0.43	<0.001
B155	0.421	<0.001	0.125	<0.001
B199	0.15	<0.001	0.071	<0.05

To examine the diversity of the latent but replication-competent reservoir, we developed a modification of current quantitative viral outgrowth protocols to include a qualitative measure of the reservoir of replication-competent proviruses. Q²VOA maximizes the yield of limiting dilution cultures to obtain viruses originating from a single cell. Consistent with the absence of viral evolution in individuals on suppressive ART, Q²VOA revealed that the latent replication-competent viral population is similar between time points 4–6 mo apart (12, 13, 38). This interval is relatively short in the life of an infected individual, and it will be interesting to evaluate the reservoir by Q²VOA over longer time intervals.

Fifty-four percent of the viruses emerging in Q²VOA belonged to expanded cell clones bearing the same *env* sequences, and these viruses have a high probability of also being identical in the rest of their genomes [prediction score of 94 of 100; available at silicianolab.johnshopkins.edu/cps (21)]. The isolation of identical sequences from multiple wells on different plates across two time points suggests that these replicating viruses do not accumulate significant mutations during 14 d in culture. The lack of immunological pressure under culture conditions may account for this finding.

There are several nonmutually exclusive explanations for the prevalence of expanded CD4⁺ T-cell clones carrying replication-competent viruses in the latent reservoir. Such cells could be clonally expanded as part of normal homeostatic processes, in response to antigen, or as a result of HIV-1 integration into and disruption or activation of genes that regulate cell division (24, 25). Alternatively, identical viral sequences could represent a burst of independent viral integration events into unrelated CD4⁺ T cells. Mapping the precise integration sites of the replication-competent proviruses is required to distinguish definitively between viral clones produced by CD4⁺ T-cell clonal expansion and by bursts of proviral integration.

Clonally expanded CD4⁺ T cells with shared proviral integration sites were detected in samples of 1–2 × 10^6 cells (8, 24, 25), which are far smaller than the samples assayed in Q²VOA. Only the largest of these previously described expanded cell clones were amenable to further characterization, and all 75 proviruses that were examined molecularly showed gross defects that rendered them replication-incompetent (8). The finding that large proviral clones are typically defective is entirely consistent with the observation that proviral clone size is inversely related to the probability of reactivation. The larger the archival proviral clone, the less likely it was to be represented in the replication-competent latent reservoir. This observation is also in agreement with the finding that most clonally expanded proviruses integrated into CD4⁺ T-cell DNA are defective (6–9).

The relative stability of clones in the replication-competent reservoir has significant implications for assessing efforts aimed at HIV-1 eradication. In particular, Q²VOA may be a sensitive assay with which to measure changes in the reservoir that result from immune-targeted interventions (39–42). Viruses isolated from the replication-competent latent reservoir demonstrated a broad range of neutralization sensitivity to bNAbs, and may also show differing levels of sensitivity to immune-based interventions. The coexistence of neutralization-sensitive and -resistant clones in the same patient may be important to consider in development of clinical strategies using bNAbs in the future. However, whether the viruses detected in Q²VOA represent species that are preferentially reactivated upon termination of ART in infected individuals remains to be determined.

In conclusion, we find clones of replication-competent viruses in the latent reservoir. The viral sequences of these clones show limited overlap with archived proviral sequences in primary CD4⁺ T cells, suggesting that the two compartments may be under different types of selective pressure in vivo. Finally, if the reservoir is maintained by CD4⁺ T-cell division, it may facilitate immunotherapy-directed efforts for HIV-1 cure because T-cell activation is thought to be associated with expression of HIV-1 antigens (39).

Methods

Study Participants. The study was conducted under the approval of The Rockefeller University Institutional Review Board. Written informed consent was obtained from all participants under study protocol MNU-0628 at The Rockefeller University. Eligible participants were adults aged 18–65 y with HIV-1 infection and undetectable plasma HIV-1 RNA levels (<20 copies per milliliter) while on ART. Each participant underwent leukapheresis at two time points 4–6 mo apart.

Q²VOA. Viral outgrowth cultures were performed as previously described (43, 44). Peripheral blood mononuclear cells (PBMCs) from virologically suppressed, chronically HIV-infected individuals on ART were obtained by leukapheresis. Briefly, leukapheresis products from each study participant at each time point were processed, and PBMCs were isolated by density centrifugation on Ficoll (Thermo Scientific). Cryopreserved PBMCs were then used to isolate total CD4⁺ T lymphocytes using negative selection by magnetic beads (Miltenyi). Purified CD4⁺ T lymphocytes were cultured at 0.3 × 10^6 cells per well in 24-well plates in 1 mL of RPMI 1640 (Gibco) supplemented with 10% (vol/vol) FBS (HyClone; Thermo Scientific), 1% penicillin/streptomycin (Gibco), 1 μg/mL phytohemagglutinin (Life Technologies), and 100 U/mL IL-2 (Peprotech) at 37 °C and 5% CO_2. Each well also received 1 mL of medium containing 2.5 × 10^6 irradiated heterologous PBMCs. After 24 h, 1.5 mL of medium was discarded and 10^6

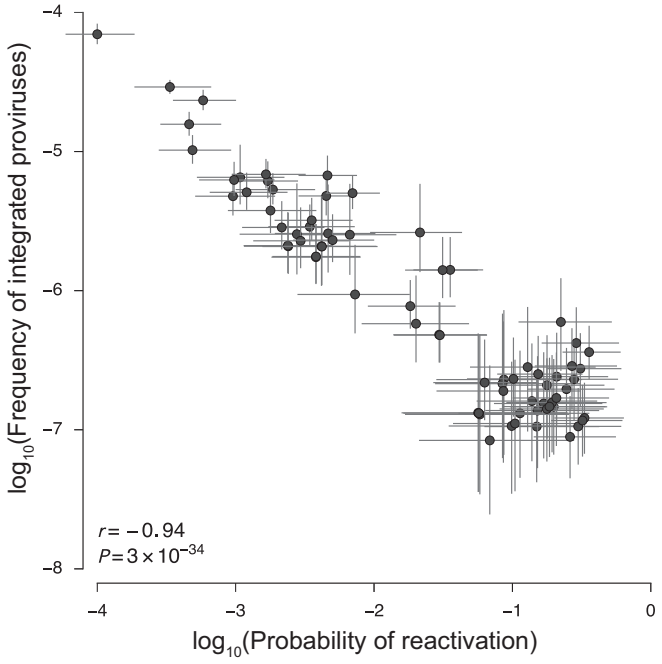

Fig. 5. Negative correlation between proviral clone size and probability of reactivation in culture. The frequency of integrated provirus (*p*) is negatively correlated with its probability of reactivation in culture (*r*). Bars denote interquartile ranges of posterior parameter estimates. The Pearson correlation is computed using median values for large clones. Clones with diverse frequencies and reactivation probabilities are observed in each patient.

A

	Virus ID	3BNC117	VRC01	10-1074	PGT121	PGDM1400	10E8
B106	1	1.01	14.1	4.32	1.14	1.42	>50
	2	1.87	>50	0.07	0.27	0.87	>50
	3	1.30	13.1	0.56	0.93	1.38	>50
	4	1.09	13.6	0.04	0.40	0.47	>50
	5	0.37	4.88	0.02	0.20	NT	>50
	6	0.64	6.46	0.03	0.33	0.15	>50
	7	0.43	8.10	0.03	0.31	0.17	>50
	8	0.58	4.24	0.47	0.84	0.07	>50
	9	5.40	26.5	0.28	7.32	0.19	>50
	10	1.52	22.8	1.10	0.90	0.22	>50
	Bulk	1.61	14.2	0.84	3.46	0.31	>50
B115	1	18.4	NT	0.34	NT	NT	NT
	2	2.87	NT	0.59	NT	NT	NT
	3	2.65	NT	0.49	NT	NT	NT
	4	2.72	NT	0.41	NT	NT	NT
	5	7.13	NT	0.42	NT	NT	NT
	6	6.85	NT	1.69	NT	NT	NT
	7	10.0	NT	3.70	NT	NT	NT
	8	9.23	NT	2.16	NT	NT	NT
	9	6.85	38.3	1.86	4.69	6.38	>50
	10	11.4	NT	2.63	NT	NT	NT
	Bulk	2.63	17.3	0.56	2.86	1.44	37.2

	Virus ID	3BNC117	VRC01	10-1074	PGT121	PGDM1400	10E8
B155	1	5.02	4.75	0.10	0.86	>50	>50
	2	6.61	35.4	>50	>50	>50	27.5
	3	5.77	41.0	>50	>50	>50	35.6
	4	7.37	NT	>50	NT	NT	NT
	5	4.75	31.0	>50	>50	>50	47.2
	6	5.55	38.0	>50	14.5	>50	45.2
	7	6.20	45.5	>50	32.8	>50	>50
	8	5.30	39.2	>50	36.8	>50	43.8
	9	6.07	26.5	>50	>50	>50	36.4
	10	9.62	NT	>50	NT	NT	NT
	Bulk	7.56	21.3	>50	>50	>50	35.9
B199	1	32.0	>50	0.20	0.77	2.36	47.2
	2	2.99	31.0	6.67	>50	21.0	>50
	3	20.7	>50	8.63	>50	3.74	>50
	4	3.03	>50	0.76	2.75	0.76	>50
	5	6.55	>50	2.84	42.6	2.53	>50
	6	4.61	>50	0.59	1.15	0.59	>50
	7	1.88	36.0	1.88	12.2	0.56	>50
	8	34.6	>50	1.33	10.6	0.52	>50
	9	11.1	>50	1.60	3.34	0.50	>50
	Bulk	8.67	>50	1.39	37.8	2.99	>50

IC_{80}(ug/ml) 0-0.1 0.1-1.0 1.0-10.0 10.1-50.0 >50.0 NT

B

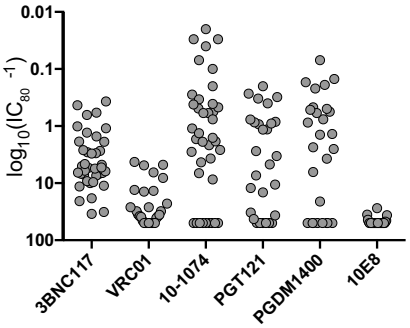

Fig. 6. Sensitivity of latent viruses to bNAbs. (*A*) Table shows antibody concentration that inhibits infection by 80% (IC_{80}) titers for selected outgrowth culture viruses (Fig. 2*A*) against a panel of bNAbs indicated at the top determined by TZM.bl assay. Viruses with low median tissue culture infectious dose ($TCID_{50}$) titers could not be tested against the entire bNAb panel (red, IC_{80} of 0–0.1 μg/mL; orange, IC_{80} of 0.1–1.0 μg/mL; yellow, IC_{80} of 1.0–10 μg/mL; green, IC_{80} of 10–50 μg/mL). NT, not tested. (*B*) Graph shows combined IC_{80} titers across all four patients for each bNAb. Each dot represents one virus.

$CD8^+$-depleted lymphoblasts from an HIV-negative donor were added to each well as target cells. At day 9, 1 mL of medium was removed and another 10^6 $CD8^+$-depleted $CD4^+$ T lymphoblasts were added to each well. At day 14, the supernatant of each well was tested for HIV-1 production using the Lenti-X p24 Rapid Titer Kit (Clontech) according to the manufacturer's instructions. Under these conditions, using 0.3×10^6 donor cells, less than 30% of all cultures were positive for all individuals.

Bulk Cultures. Bulk cultures were performed as previously described (45). Sequence analysis on bulk culture was also performed as previously described (28).

Q^2VOA Sequence Analysis. The supernatant from p24-positive cultures was extracted using the Qiagen MinElute Virus Spin kit QIAcube. The cDNA was produced using 1:10 diluted RNA with SuperScript III reverse transcriptase (Invitrogen Life Technologies) and the antisense primer env3out 5′–TTGCTACTTGTGATTGCTCCATGT–3′ followed by RNase H digestion (Invitrogen Life Technologies) for 20 min at 37 °C. The full gp160 *env* was amplified from 1:40 diluted cDNA using envB5out 5′–TAGAGCCCTGGAAGCATCCAGGAAG–3′ and envB3out 5′–TTGCTACTTGTGATTGCTCCATGT–3′ in the first round and the second-round nested primers envB5in 5′–TTAGGCATCTCCTATGGCAGGAAGAAG–3′ and envB3in 5′–GTCTCGAGATACTGCTCCCACCC–3′.

PCR assays were performed using High Fidelity Platinum Taq (Invitrogen) at 94 °C for 2 min; 94 °C for 15 s, 55 °C for 30 s, and 68 °C for 4 min × 35; and 68 °C for 15 min. Second-round nested PCR was performed using 1 μL of PCR1 product as a template and High Fidelity Platinum Taq at 94 °C for 2 min; 94 °C for 15 s, 57 °C for 30 s, and 68 °C for 4 min × 45; and 68 °C for 15 min. PCR2 products were checked using 1% 96-well E-Gels (Invitrogen). PCR bands with the expected HIV envelope size were quantified and subjected to library preparation using the Illumina Nextera DNA Sample Preparation Kit (Illumina) as described (46). Briefly, 10 ng of DNA per band was subjected to tagmentation, ligated to barcoded sequencing adapters using the Illumina Nextera Index Kit, and then purified using AmPure Beads XP (Agencourt). Ninety-six different purified samples were pooled into one library and then subjected to paired-end sequencing using Illumina MiSeq Nano 300 (Illumina) cycle kits at a concentration of 15 pM.

Sequence adapters were removed using Cutadapt v1.8.3. Read assembly for each virus was performed in three steps. First, de novo assembly was performed using Spades v3.6.1 to yield long contig files. Contigs longer than 255 bp were subsequently aligned to an HIV envelope reference sequence, and a consensus sequence was generated using Geneious 8. Finally, reads were realigned to the consensus sequence to close gaps, and a final read consensus was generated for each sequence. Sequences with double peaks (cutoff consensus identity for any residue <75%), with stop codons, or shorter than the expected envelope sizes were omitted from downstream analyses.

Proviral SGA. DNA from 1 to 10×10^6 $CD4^+$ cells from HIV-1–infected individuals was prepared as previously described (47). Briefly, cells were collected after magnetic isolation and lysed in Proteinase K buffer [100 mM Tris (pH 8), 0.2% SDS, 200 mM NaCl, 5 mM EDTA] and 20 mg/mL Proteinase K at 56 °C for 12 h. Genomic DNA was extracted by phenol chloroform precipitation. Aliquots of the resulting DNA were diluted and used as a template for full-length gp160 PCRs. All PCRs were performed using High Fidelity Platinum Taq (Invitrogen). The first PCR was performed at 94 °C for 2 min; 94 °C for 15 s, 58.5 °C for 30 s, and 68 °C for 4 min × 35; and 68 °C for 15 min. Second-round nested PCR was then performed using 1 μL of PCR1 product as a template and High Fidelity Platinum Taq at 94 °C for 2 min; 94 °C for 15 s, 61 °C for 30 s, and 68 °C for 4 min × 45; and 68 °C for 15 min. PCR2 products were checked using 1% 96-well E-Gels (Invitrogen). Bands with the expected size of the HIV-1 envelope obtained from diluted DNA samples that showed an amplification efficiency of less than 30% were subjected to library preparation and sequencing using the Illumina platform as described above. Hypermutants, sequences with double peaks (cutoff consensus identity for any residue <75%), sequences with stop codons, or sequences shorter than the expected envelope sizes were omitted from downstream analyses.

Viral Neutralization Testing by TZM.bl. Supernatants from p24-positive Q^2VOA wells were tested against a panel of bNabs using the TZM.bl neutralization assay, as previously described (48, 49). Neutralization assays were conducted in a laboratory meeting good clinical laboratory practice quality assurance criteria (laboratory of M.S.S.). Viruses with low median tissue culture infectious dose ($TCID_{50}$) titers were only tested against 3BNC117 and 10-1074.

GSI. The GSI (23) was used to calculate the degree of phylogenetic association of the *env* gene sequences. Phylogenies were inferred using PhyML version 3.0 (50). Multiple bifurcations with intervening zero-length branches were collapsed to polytomies using the di2multi method implemented in the ape package for R (51). These phylogram topologies were used to calculate the gsi values, and *P* values were determined using 10,000 replicate permutations (23).

Computational Analyses. Bayesian inference was implemented in Stan (52) to estimate jointly the frequency of integrated proviruses and their reactivation probability. Data from proviral DNA sequencing and viral outgrowth assays were combined across multiple visits, incorporating a prior probability for the gradual decay of the reservoir (3, 53). We assumed a weakly informative prior for the frequency of integrated provirus and a uniform prior for the probability of reactivation. Parameters were estimated individually for viruses with identical sequences observed at least three times across both visits (large clones), whereas sequences observed less frequently were grouped together (small clones). The observed relationship between clone frequency and reactivation probability also holds in simpler models incorporating data from single visits only, and when clones that were not observed in both assays during the same visit are excluded. Additional details of the inference methods used are provided in *SI Methods*.

ACKNOWLEDGMENTS. We thank all study participants who devoted time to our research. We also thank members of the M.C.N. laboratory for helpful discussions. J.C.C.L. is supported by Award 248676/2013-0 from the Brazil Conselho Nacional de Pesquisas "Ciencia sem Fronteiras." Y.Z.C. is supported

by Grant KL2TR001865, National Center for Advancing Translational Sciences (NCATS), and Grant UL1 TR000043, National Institutes of Health (NIH) Clinical and Translational Science Award (CTSA) program. J.C.C.L. and Y.Z.C. are additionally supported by NIH/National Institute of Allergy and Infectious Diseases Grant U01 AI118536 (to M.C.). The contributions of J.P.B. and A.K.C. were supported by the Ragon Institute of MGH, MIT, and Harvard. E.F.K., G.H.L., and B.H.H. were supported by BEAT-HIV Delaney Collaboratory Grant UM1 AI126620. M.C.N. is a Howard Hughes Medical Investigator.

1. Chun TW, et al. (1997) Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc Natl Acad Sci USA* 94(24):13193–13197.
2. Finzi D, 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(5): 512–517.
3. Crooks AM, et al. (2015) Precise quantitation of the latent HIV-1 reservoir: Implications for eradication strategies. *J Infect Dis* 212(9):1361–1365.
4. Hughes SH, Coffin JM (2016) What integration sites tell us about HIV Persistence. *Cell Host Microbe* 19(5):588–598.
5. Churchill MJ, Deeks SG, Margolis DM, Siliciano RF, Swanstrom R (2016) HIV reservoirs: What, where and how to target them. *Nat Rev Microbiol* 14(1):55–60.
6. Eriksson S, et al. (2013) Comparative analysis of measures of viral reservoirs in HIV-1 eradication studies. *PLoS Pathog* 9(2):e1003174.
7. Ho YC, et al. (2013) Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* 155(3):540–551.
8. Cohn LB, et al. (2015) HIV-1 integration landscape during latent and active infection. *Cell* 160(3):420–432.
9. Bruner KM, et al. (2016) Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med* 22(9):1043–1049.
10. Chun TW, et al. (1997) Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* 387(6629):183–188.
11. Chun TW, et al. (1995) In vivo fate of HIV-1-infected T cells: Quantitative analysis of the transition to stable latency. *Nat Med* 1(12):1284–1290.
12. Evering TH, et al. (2012) Absence of HIV-1 evolution in the gut-associated lymphoid tissue from patients on combination antiviral therapy initiated during primary infection. *PLoS Pathog* 8(2):e1002506.
13. Josefsson L, et al. (2013) The HIV-1 reservoir in eight patients on long-term suppressive antiretroviral therapy is stable with few genetic changes over time. *Proc Natl Acad Sci USA* 110(51):E4987–E4996.
14. von Stockenstrom S, et al. (2015) Longitudinal genetic characterization reveals that cell proliferation maintains a persistent HIV type 1 DNA pool during effective HIV therapy. *J Infect Dis* 212(4):596–607.
15. Wagner TA, et al. (2013) An increasing proportion of monotypic HIV-1 DNA sequences during antiretroviral treatment suggests proliferation of HIV-infected cells. *J Virol* 87(3):1770–1778.
16. Barton K, et al. (2016) Broad activation of latent HIV-1 in vivo. *Nat Commun* 7:12731.
17. Günthard HF, 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(11):9404–9412.
18. Anderson JA, et al. (2011) Clonal sequences recovered from plasma from patients with residual HIV-1 viremia and on intensified antiretroviral therapy are identical to replicating viral RNAs recovered from circulating resting CD4+ T cells. *J Virol* 85(10): 5220–5223.
19. Bui JK, Mellors JW, Cillo AR (2015) HIV-1 virion production from single inducible proviruses following T-cell activation ex vivo. *J Virol* 90(3):1673–1676.
20. Lee SK, et al. (September 27, 2016) Quantification of the latent HIV-1 reservoir using ultra deep sequencing and primer ID in a viral outgrowth assay. *J Acquir Immune Defic Syndr*, 10.1097/QAI.0000000000001187.
21. Laskey SB, Pohlmeyer CW, Bruner KM, Siliciano RF (2016) Evaluating clonal expansion of HIV-infected cells: Optimization of PCR strategies to predict clonality. *PLoS Pathog* 12(8):e1005689.
22. Palmer S, Josefsson L, Coffin JM (2011) HIV reservoirs and the possibility of a cure for HIV infection. *J Intern Med* 270(6):550–560.
23. Cummings MP, Neel MC, Shaw KL (2008) A genealogical approach to quantifying lineage divergence. *Evolution* 62(9):2411–2422.
24. Maldarelli F, et al. (2014) HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* 345(6193):179–183.
25. Wagner TA, et al. (2014) HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* 345(6196):570–573.
26. Mouquet H, et al. (2012) Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. *Proc Natl Acad Sci USA* 109(47):E3268–E3277.
27. Bar K, et al. (November 9, 2016) Effect of HIV antibody VRC01 on viral rebound after treatment interruption. *N Engl J Med*, 10.1056/NEJMoa1608243.
28. Scheid JF, et al. (2016) HIV-1 antibody 3BNC117 suppresses viral rebound in humans during treatment interruption. *Nature* 535(7613):556–560.
29. Siliciano JM, Siliciano RF (2015) The remarkable stability of the latent reservoir for HIV-1 in resting memory CD4+ T cells. *J Infect Dis* 212(9):1345–1347.
30. Chun TW, Moir S, Fauci AS (2015) HIV reservoirs as obstacles and opportunities for an HIV cure. *Nat Immunol* 16(6):584–589.
31. Chun TW, et al. (1998) Early establishment of a pool of latently infected, resting CD4(+) T cells during primary HIV-1 infection. *Proc Natl Acad Sci USA* 95(15):8869–8873.
32. Siliciano RF, Greene WC (2011) HIV latency. *Cold Spring Harb Perspect Med* 1(1): a007096.
33. Barton K, Winckelmann A, Palmer S (2016) HIV-1 reservoirs during suppressive therapy. *Trends Microbiol* 24(5):345–355.
34. Vallejo A, et al. (2012) The effect of intensification with raltegravir on the HIV-1 reservoir of latently infected memory CD4 T cells in suppressed patients. *AIDS* 26(15): 1885–1894.
35. Dinoso JB, et al. (2009) Treatment intensification does not reduce residual HIV-1 viremia in patients on highly active antiretroviral therapy. *Proc Natl Acad Sci USA* 106(23):9403–9408.
36. Joos B, et al.; Swiss HIV Cohort Study (2008) HIV rebounds from latently infected cells, rather than from continuing low-level replication. *Proc Natl Acad Sci USA* 105(43): 16725–16730.
37. Simonetti FR, et al. (2016) Clonally expanded CD4+ T cells can produce infectious HIV-1 in vivo. *Proc Natl Acad Sci USA* 113(7):1883–1888.
38. Persaud D, et al. (2007) Slow human immunodeficiency virus type 1 evolution in viral reservoirs in infants treated with effective antiretroviral therapy. *AIDS Res Hum Retroviruses* 23(3):381–390.
39. Halper-Stromberg A, et al. (2014) Broadly neutralizing antibodies and viral inducers decrease rebound from HIV-1 latent reservoirs in humanized mice. *Cell* 158(5): 989–999.
40. Bruel T, et al. (2016) Elimination of HIV-1-infected cells by broadly neutralizing antibodies. *Nat Commun* 7:10844.
41. Halper-Stromberg A, Nussenzweig MC (2016) Towards HIV-1 remission: Potential roles for broadly neutralizing antibodies. *J Clin Invest* 126(2):415–423.
42. Lu CL, et al. (2016) Enhanced clearance of HIV-1-infected cells by broadly neutralizing antibodies against HIV-1 in vivo. *Science* 352(6288):1001–1004.
43. van 't Wout AB, Schuitemaker H, Kootstra NA (2008) Isolation and propagation of HIV-1 on peripheral blood mononuclear cells. *Nat Protoc* 3(3):363–370.
44. Laird GM, Rosenbloom DI, Lai J, Siliciano RF, Siliciano JD (2016) Measuring the frequency of latent HIV-1 in resting CD4$^+$ T cells using a limiting dilution coculture assay. *Methods Mol Biol* 1354:239–253.
45. Caskey M, et al. (2015) Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117. *Nature* 522(7557):487–491.
46. Kryazhimskiy S, Rice DP, Jerison ER, Desai MM (2014) Microbial evolution. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* 344(6191):1519–1522.
47. Klein IA, et al. (2011) Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* 147(1):95–106.
48. Li M, et al. (2005) Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J Virol* 79(16):10108–10125.
49. Montefiori DC (2005) Evaluating neutralizing antibodies against HIV, SIV, and SHIV in luciferase reporter gene assays. *Curr Protoc Immunol* Chapter 12:Unit 12 11.
50. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307–321.
51. Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20(2):289–290.
52. Stan Development Team (2016) PyStan: The Python Interface to Stan, Version 2.9.0. Available at mc-stan.org.
53. Siliciano JD, et al. (2003) Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat Med* 9(6):727–728.
54. Gelman A, et al. (2008) A weakly informative default prior distribution for logistic and other regression models. *Ann Appl Stat* 2(4):1360–1383.
55. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–2690.