HIV diversity and genetic compartmentalization in blood and testes during suppressive antiretroviral therapy

Rachel L. Miller[a], Rosalie Ponte[b,c], Bradley R. Jones[d], Natalie N. Kinloch[a], Fredrick H. Omondi[a], Mohammad-Ali Jenabian[e], Franck P. Dupuy[b,c], Remi Fromentin[f], Pierre Brassard[g], Vikram Mehraj[b,c,f], Nicolas Chomont[f,h], Art F. Y. Poon[i], Jeffrey B. Joy[d,j], Zabrina L. Brumme[a,d#], Jean-Pierre Routy[b,c] for the ORCHID study group

[a]Faculty of Health Sciences, Simon Fraser University, Burnaby, British Columbia (BC), Canada
[b]Research Institute, McGill University Health Centre, Montreal, Quebec (QC), Canada
[c]Chronic Viral Illness Service and Division of Hematology, McGill University Health Centre, Montreal, QC, Canada
[d]British Columbia Centre for Excellence in HIV/AIDS, Vancouver, BC, Canada
[e]Department of Biological Sciences, Université du Quebec a Montreal , Montreal, QC, Canada
[f]Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Montreal, QC, Canada
[g]The Metropolitan Centre for Plastic Surgery, Montreal, QC, Canada
[h]Department of Microbiology, Infection and Immunology, Université de Montréal, Montreal, QC, Canada;
[i]Department of Pathology & Laboratory Medicine, Western University, London, Ontario, Canada;
[j]Department of Medicine, University of British Columbia, Vancouver, BC, Canada

Running Head: HIV reservoirs in blood and testes

24

25 *# Address correspondence to*: Zabrina L. Brumme, email: zbrumme@sfu.ca

26

27 RLM and RP contributed equally to this work. ZLB and J-PR also contributed equally to this

28 work.

29

30 *Word Counts*: 250 (abstract); 5179 (text)

31

32 **Keywords:** HIV, reservoir, diversity, testes, genetic compartmentalization, clonal expansion

Abstract (250 words)

HIV's ability to persist during suppressive antiretroviral therapy is the main barrier to cure. Immune privileged tissues such as the testes may constitute distinctive sites of HIV persistence, but this has been challenging to study in humans. We analyzed proviral burden and genetics in blood and testes of 10 individuals on suppressive therapy who underwent elective gender-affirming surgery. HIV DNA levels in matched blood and testes were quantified by qPCR and subgenomic proviral sequences (*nef* region) were characterized from single templates. HIV diversity, compartmentalization and immune escape burden were assessed using genetic and phylogenetic approaches. Diverse proviruses were recovered from blood (396 sequences; 354 *nef*-intact) and testes (326 sequences; 309 *nef*-intact) of all participants. Notably, the frequency of identical HIV sequences varied markedly between and within individuals. Nevertheless, proviral loads, within-host unique HIV sequence diversity, and immune escape burden correlated positively between blood and testes. When all intact *nef* sequences were evaluated, 60% of participants exhibited significant blood-testes genetic compartmentalization, but none did so when restricting to unique sequences per site, suggesting that compartmentalization, when present, is attributable to clonal expansion of HIV-infected cells. Our observations confirm the testes as a site of HIV persistence and suggest that individuals with larger and more diverse blood reservoirs will have larger and more diverse testes reservoirs. Furthermore, while the testes microenvironment may not be sufficiently unique to facilitate the seeding of *unique* viral populations therein, differential clonal expansion dynamics may be at play, which may complicate HIV eradication.

57 **Importance** (150 words)

58

59 Two key questions in HIV reservoir biology are whether immune privileged tissues, such
60 as the testes, harbor distinctive proviral populations during suppressive therapy, and if so by
61 what mechanism. While our results indicated that blood-testes HIV genetic
62 compartmentalization was reasonably common (60%), it was always attributable to differential
63 frequencies of identical HIV sequences between sites. No blood/tissue dataset retained evidence
64 of compartmentalization when only unique HIV sequences per site were considered; moreover,
65 HIV immune escape mutation burdens were highly concordant between sites. We conclude that
66 the principal mechanism by which blood and testes reservoirs differ is *not* via seeding of
67 divergent HIV sequences therein, but rather due to differential clonal expansion of latently-
68 infected cells. Thus, while viral diversity and escape-related barriers to HIV eradication are of
69 broadly similar magnitude across blood and testes, clonal expansion represents a challenge.
70 Results support individualized analysis of within-host reservoir diversity to inform curative
71 approaches.

72 **Introduction**

73 The HIV reservoir, a small pool of primarily $CD4^+$ T-cells that harbor replication-
74 competent virus despite long-term suppressive combination antiretroviral therapy (cART) (1), is
75 the major barrier to cure. During untreated infection, HIV populations evolve within-host (2-5),
76 and individual viral sequences are continually archived into the reservoir as integrated proviruses
77 (6). There, they can persist long-term, either within the original infected cell or clonal
78 descendants thereof (7-10). Much of our understanding of HIV persistence is derived from HIV
79 sequences isolated from blood during long-term cART; these studies have revealed that the
80 within-host HIV reservoir is genetically diverse (11-13), that it frequently contains immune
81 escape mutations (14, 15), and that it commonly features clonally-expanded cell populations
82 harboring identical proviruses (10, 16-19).

83 The HIV reservoir in blood however, may not reflect that within tissues. HIV reservoir
84 composition in "immune privileged" sites, defined as those that are protected by physical barriers
85 that limit immune cell trafficking (20) and/or that are equipped with mechanisms to suppress
86 immune responses locally (21, 22), may be distinct from those elsewhere, but this has been
87 challenging to study in healthy humans. The testes represent such a site (21, 23): non-human
88 primate studies have demonstrated that cytokine responses of testicular T-cells to mitogen
89 stimulation are lower than those of blood T-cells (24, 25); moreover, human testes harbor higher
90 frequencies of $CD39^+$ T-regulatory cells (21), which are capable of suppressing HIV-specific
91 $CD8^+$ T-cell responses (26). Despite this however, leukocyte populations, including central and
92 effector memory CD4+ T-cells that represent potential viral targets, are nevertheless present in
93 the testes (24). Indeed, immunodeficiency viruses penetrate into the testes early during infection

(22, 27), and subsequently persist (22, 28-30). As such, the testes have been hypothesized to represent a distinctive HIV sanctuary site (21, 25).

A key outstanding question is whether proviruses persisting in testes are genetically distinct from those in blood, and if so, how. Blood-tissue genetic compartmentalization could arise a number of ways (31). It could happen if, during untreated HIV infection, tissue HIV populations replicated under reduced immune pressures and/or experienced limited subsequent mixing with blood: if this were the case, then provirus populations archived into blood and tissue HIV reservoirs would reflect their divergent evolution pre-cART. Genetically distinct HIV reservoirs could also arise as a result of restricted gene flow between compartments: for example if only a small number of HIV strains initially penetrated into tissue, descendant viral populations in tissue could form nested sub-clades within the more diverse populations in blood. Differential proviral distributions could also arise via unequal clonal expansion and/or contraction of latently HIV-infected cells in blood and tissue, either before or during cART. The latter would result in viral populations that are *not* different in terms of unique viral lineages, but rather different in terms of their frequencies of particular identical proviruses. Investigating blood-tissue genetic compartmentalization will shed light on the extent to which tissue microenvironments shape latent within-host HIV populations, and the extent to which the blood reservoir mirrors that elsewhere in the body, information that is relevant to the design of curative approaches guided by individualized analysis of within-host reservoir composition (32, 33).

While studies in non-human primates have compared genetic structures of within-host Simian Immunodeficiency Virus populations in male genital organs versus blood (28, 34), no studies to our knowledge have investigated this in healthy, HIV-infected humans. We thus investigate within-host proviral burden, genetic diversity and compartmentalization in 10 HIV-

infected, cART-suppressed individuals undergoing elective bilateral orchiectomy for gender affirmation surgery (21, 23).

Methods

Study participants and ethics statement

Blood and testicular tissues were donated by ten HIV-infected but otherwise healthy adults, aged 27 to 52, who underwent elective bilateral orchiectomy for gender affirmation (ORCHID cohort (21, 23)). At time of surgery, all participants had maintained viremia suppression on cART for at least 6 months, however full clinical histories (including infection date estimates) were unavailable. This study was approved by the Institutional Review Boards of the Research Institute of the McGill University Health Centre and Simon Fraser University. All participants provided written informed consent.

Blood and tissue collection

Peripheral Blood Mononuclear Cells (PBMC) were isolated from whole blood by Ficoll density gradient centrifugation and stored at -80°C until use. Testicular tissue was processed within one hour of surgery (21, 23). For genetic analysis, ~100µg tissue sections were snap-frozen upon reception in liquid nitrogen; for histology, $10mm^3$ sections were embedded in Cryomatrix (ThermoFisher) before snap-freezing.

Immunostaining

Frozen embedded tissues were cut into 5 µm-thick sections (Leica CM3050S cryostat), fixed in acetone/methanol for 5 minutes at 4°C, stained overnight at 4°C using anti-CD3 and anti-CD4 (DAKO-Agilent) and exposed to conjugated Alexa-Fluor secondary antibodies (ThermoFisher) for 30 minutes at room temperature. Slides were counter-stained using 4',6-

diamidino-2-phenylindole (DAPI) before mounting in Fluoromount-G (Southern Biotech). Images were acquired using a Nikon epi-fluorescent microscope with a Zyla sCMOS camera and analyzed using Image J 1.47g (National Institutes of Health).

HIV quantification and sequencing

DNA was extracted from $CD4^+$ T cells negatively isolated from PBMCs (StemCell Technologies) and directly from snap-frozen tissue (Qiagen). HIV DNA burden was measured using a quantitative PCR assay capable of detecting a single viral genome per PCR reaction, where primers are located in the HIV LTR-gag region and are optimized for detection of multiple HIV subtypes (21, 35). For each participant, a minimum of 5 million PBMCs plus at least four tissue sections, two each from right and left testes, were allocated for HIV sequencing, with the exception of participant 8, for whom only two right testis sections were available. Genomic DNA was extracted from PBMCs directly (Invitrogen), whereas tissues were further sectioned into ~50µg fragments prior to genomic DNA extraction. HIV amplification was performed using single-genome amplification of a subgenomic fragment (HIV *nef*). The *nef* gene was selected based on its relatively high within-host diversity, richness in phylogenetic signal, and its representativeness of within-host HIV evolution and diversity within the rest of the HIV genome (8). This gene also represents the one most likely to be intact within proviruses sampled during long-term cART (36). Single-genome amplification, performed using high fidelity enzymes (Roche Expand™ HiFi) and oligonucleotide primers designed to amplify HIV sequences from all major subtypes, was achieved by endpoint dilution such that ~25-30% of the resulting nested PCR reactions were successful (8). Negative PCR controls, included in every amplification, always remained negative. Amplicons were sequenced on an automated DNA sequencer using BigDye v3.1 chemistry (Applied Biosystems). Chromatograms were basecalled

using Sequencher v5.0 (GeneCodes). Sequences exhibiting nucleotide mixtures, defects, or hypermutations (identified using Hypermut (37)) were excluded, as were sequences exhibiting evidence of within-host recombination (identified using RDP4 (38)).

Sequences were codon-aligned using MAFFT (39) and edited in AliView v1.18 (40). For each within-host alignment, the best fitting substitution models for both genetic distance and phylogenetic-based tests were determined using jModelTest v2.1.10 (41). Phylogenies were inferred by approximate maximum likelihood using FastTree2 v2.1.10 (42) under the best available nucleotide substitution model (for participants 7 and 10, the generalized time reversible [GTR] model (43) with no rate heterogeneity among sites was the best fit; for all others a GTR model with rate heterogeneity using a gamma distribution was the best fit). Node support values were derived from 1000 bootstraps. Within-host phylogenies were mid-point rooted in FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) and visualized with ggtree (44).

Genetic compartmentalization tests and other statistical analyses

As there are different ways to define genetic compartmentalization (*e.g.* as genetic heterogeneity, or the presence of distinct lineages, between two viral subpopulations), compartmentalization tests occasionally yield discordant results (45). As recommended, we applied more than one test and classified a dataset as “compartmentalized” only when both tests agreed (45). We employed one genetic distance-based test (46) and one tree-based test (47), both implemented in HyPhy v2.22 (48). For the distance-based test, we employed Hudson, Boos and Kaplan's nonparametric test for population structure (46), a test that has been validated for within-host HIV datasets (49) including from the reservoir, where it is sometimes referred to as the "nonparametric test for panmixia" (49-51). This test compares the mean pairwise distances between sequences from different versus the same subpopulation (compartmentalization is

supported if mean pairwise distances of sequences from the same subpopulation are smaller than those from different subpopulations) and computes the K_{ST} statistic whose values range from 0 (denoting no compartmentalization) to 1 (denoting complete compartmentalization). Statistical significance, expressed as a p-value, is assessed via a population-structure randomization test. Using jModelTest v2.1.10 (41) we identified the Tamura-Nei 1993 (TN93) nucleotide substitution model as the best fitting distance-based model available in HyPhy v2.22 (48); this model is also is consistent with previous applications of this test to within-host HIV datasets (45, 52). For the tree-based test, we used the Slatkin-Maddison test (47). This test determines the minimum number of between-compartment migrations to explain the distribution of compartments on the tree tips: the smaller the number, the stronger the support for compartmentalization. Statistical significance is assessed via a lineage permutation test. Importantly, because identical HIV sequences (especially those present in abundance, and whose frequencies differ between sites) increase the likelihood of compartmentalization detection (53), we analyzed each dataset two ways: overall, and with identical sequences collapsed down to a single copy per compartment.

HLA class I sequence-based typing was performed to allele-level resolution (54). HLA-associated adapted (inferred HLA-escaped) and non-adapted (inferred HLA-susceptible) polymorphisms in HIV subtype B were defined using a published list derived from analysis of linked HIV/HLA genotypes from an independent cohort of 1888 subtype B-infected individuals (55). HLA-restricted, optimally-described CTL epitopes in participants' provirus sequences were defined using the Los Alamos HIV Molecular Immunology Database with current updates (56), where epitopes restricted by HLA alleles closely related to the participants' allele(s) were

identified as putative epitopes (55). Correlations were assessed using Spearman's correlation. P-values < 0.05 were considered statistically significant.

Data availability

Genbank accession numbers for intact sequences are MK301609 to MK302069 and MK849402 to MK849612.

Results

HIV characterization in blood and testes

At time of blood and tissue sampling, all participants had maintained plasma viremia suppression on cART for ≥6 months. At this time, blood HIV DNA levels in participants ranged from 11 to 4003 (median 512, Interquartile Range [IQR] 100-1111) HIV DNA copies/10^6 $CD4^+$ T cells. The presence of $CD4^+$ T-cells in testes was confirmed by immunostaining (**Figure 1A**). Testes HIV reservoir size, quantified in terms of viral copies *per total cell equivalent* ($CD4^+$ T-cells were not isolated from tissue), ranged from trace quantities to 198 copies/10^6 cells (median 3, IQR 0.4-10 copies/10^6 cells) in right testis and trace quantities to 39.2 copies/10^6 cells (median 1, IQR 0.4-15 copies/10^6 cells) in left testis (**Figure 1B**). HIV DNA loads in right and left testis correlated significantly between participants (Spearman's r=0.73, p=0.03, not shown). HIV DNA levels in blood and testes, calculated as the mean of right and left sections, also correlated marginally between participants (Spearman's r=0.54, p=0.1) (**Figure 1C**).

In total, 722 subgenomic HIV sequences (*nef*) were isolated by single-genome amplification from blood (N=396) and at least one tissue section (N=326) for all 10 participants, even those for whom testes HIV DNA loads were below the limit of quantification (participants 7 and 10). Even though different tissue sections were used for HIV DNA quantification and

231 sequence recovery, testes HIV DNA levels correlated with the total number of HIV sequences
232 retrieved from ~100μg tissue (Spearman's r=0.61; p=0.07, not shown), suggesting the absence of
233 major location-based sampling biases. After removal of 50 defective and/or hypermutated
234 sequences and 9 putative within-host recombinants, 663 intact *nef* sequences remained: 354 (26-
235 45 per participant, median 34, IQR 31-40) from blood and 309 (4-99 per participant, median 16,
236 IQR 10-35) from testes (**Table 1**). For eight participants (1-5, 7, 9 and 10), intact *nef* sequences
237 were isolated from both right and left testes, however for participant 6 only the right testis
238 yielded sequences and for participant 8, only right testis sections were available for analysis. Due
239 to the modest number of testes *nef* sequences recovered for some participants, our primary
240 analysis combined all within-host testes sequences together regardless of sampling location,
241 though we tested the validity of this assumption in participants for whom sufficient data were
242 available (see below).

243 All participants harbored both unique HIV *nef* sequences (observed only once) and
244 identical *nef* sequences (observed more than once), but at markedly different frequencies (**Table**
245 **1**). This can be visualized by plotting each participant's total number of unique sequences as a
246 function of their total number of sequences collected (**Figure 2**). While for some, notably
247 participant 6, unique sequences continued to be recovered during sampling, for many others (*e.g.*
248 participant 1), ongoing sampling largely yielded sequences that had been recovered previously.
249 Notably, there was no significant relationship between the frequency of unique sequences
250 recovered from blood and testes *within* a given participant (Wilcoxon signed rank test p=0.5;
251 **Figure 1D**). Participants 1 and 2 represented the most extreme examples of this: whereas 70% of
252 HIV *nef* sequences isolated from participant 1's blood were unique, the sequences isolated from
253 testes were almost all (92%) identical, whereas for participant 2 the opposite was true (**Table 1**).

This indicates that one cannot generalize, based on location alone, where identical HIV sequences are more likely to be observed. In blood, we observed a highly significant positive correlation between HIV DNA levels and unique HIV *nef* sequence frequency (Spearman's r=0.90, p=0.0008; **Figure 1E**). This supports the notion that larger blood reservoirs tend to be more genetically diverse (as opposed to more highly clonally expanded). In contrast, no significant correlation between HIV DNA levels and unique *nef* sequence frequency was observed in testes; in fact the relationship trended in the opposite direction (Spearman's r= -0.49, p=0.2; **Figure 1F**). This indicates that, in contrast to blood, the relationship between reservoir size and diversity in testes is less straightforward.

A phylogeny inferred from an alignment comprising all intact HIV *nef* sequences plus select subtype reference strains confirmed that each participant's sequences formed monophyletic clades with high (≥99.7%) bootstrap support, and revealed that participants 1-9 harbored HIV subtype B while participant 10 harbored subtype C (**Figure 3**). Though all within-host phylogenies featured both unique and identical HIV *nef* sequences, tree topology and diversity varied markedly between individuals (**Figure 4**). Participant 2's phylogeny was remarkable in that 25 of the 26 *nef* sequences isolated from blood were identical, but those testes were diverse (though the other lone unique blood *nef* sequence was also quite distant from the others) (**Figure 4A**). Of note, the abundant blood *nef* sequence was also recovered from both right and left testes; in addition there was a second *nef* sequence that was independently recovered from both testes. Participant 6 harbored diverse HIV *nef* sequences within both blood and right testis (no sequences were recovered from the left testis), and exhibited the highest within-host diversity overall (**Figure 4B**). Nevertheless, identical *nef* sequences, including three examples of identical sequences isolated from both blood and right testis, were identified. Participant 5 harbored

multiple closely-related monotypic clades, including one sequence that was abundant in both testes (and was detected in independent samplings thereof) and that was also present in blood, alongside other diverse sequences in blood and tissue (**Figure 4C**). Participants 1 and 9 were notable in that they both featured a large, genetically distinct monotypic clade detected solely in testes (in participant 1, this sequence was detected in both testes while in participant 9 it was detected in right testis only) yet both also featured other identical HIV *nef* sequences that were isolated from both blood and testes (**Figures 4D, 4E**). Participant 8 featured six clusters of identical sequences, four of which were detected solely in blood and two of which were detected in both blood and right testis (a left testis section was not available for this participant) (**Figure 4F**). Participant 3 harbored four clusters of identical sequences - one large one that was found predominantly in blood but that was also recovered from right testis, and three others that were exclusively observed in either blood or tissue (including one sequence observed in both right and left testis) (**Figure 4G**). Of all individuals studied, participants 4, 7 and 10 exhibited the most limited within-host diversity (participant 10's proviruses were the least diverse of all, suggestive of early cART initiation; of note this individual, the only one for whom some clinical information was available, had been diagnosed only 3 years prior) (**Figures 4H-J**). Nevertheless, like all others studied, participants 4,7, and 10 also featured at least one identical HIV *nef* sequence isolated from both blood and testis.

We next investigated the extent to which testes HIV DNA diversity correlated with that in blood. Because identical sequences will artificially lower average pairwise HIV diversity measures (and our data already indicate that identical sequence distributions varied markedly between compartments), we collapsed identical sequences down to a single copy per compartment and computed median within-host patristic (tip-to-tip phylogenetic) distances

between all pairs of sequences sampled from each site. Median within-host patristic distances were 2.25×10^{-2} (IQR 1.83×10^{-2}-3.70×10^{-2}) substitutions per nucleotide site in blood and 2.33×10^{-2} (IQR 1.79×10^{-2}-4.05×10^{-2}) in testes, values that correlated significantly with one another (Spearman's r=0.94, p=0.0002; **Figure 5**). These results indicate that the diversity of *unique* HIV sequences in testes generally reflects that in blood.

HIV genetic compartmentalization

Our major objective was to assess whether proviral sequences in blood and testes exhibit significant population structure or "genetic compartmentalization". As the presence of identical sequences influences compartmentalization detection (especially if their distribution varies markedly across sites), we analyzed the data two ways: by including all sequences ("overall") and by collapsing identical sequences down to a single copy per compartment ("unique").

When assessing within-host datasets overall, both distance- and tree-based tests agreed that 6 of 10 participants (1-5, 9) displayed significant blood-testes HIV genetic compartmentalization while participants 7 and 8 did not (**Figure 6A and Table 2**). For participants 6 and 10, the distance-based test did not detect compartmentalization but the tree-based test did; as per our predefined criteria these participants' overall datasets were deemed not compartmentalized. Notably, all six participants whose overall datasets were deemed compartmentalized harbored at least one abundant identical HIV *nef* sequence that predominated, or was exclusively found in, one compartment (**Figure 4**).

Importantly however, when analysis was restricted to unique *nef* sequences per site, no participant retained consistent evidence of compartmentalization between blood and testes (**Figure 6B and Table 2**). These observations are notable for two reasons. First, the observation that 40% of within-host datasets lacked any evidence of significant HIV population structure

between blood and testes overall, indicates that HIV proviral composition in testes is not always distinct from that in blood. Secondly, the observation that no participants exhibited evidence of blood-testes HIV compartmentalization when analysis was restricted to unique *nef* sequences per compartment strongly suggests that HIV population structure, when present, is driven by the differential distribution of identical sequences across blood and tissue, and not by the presence of distinct HIV lineages within these sites.

Our primary analyses combined all sequences from a given site together. However, given that only a limited amount of biological material was analyzed, sampling biases are a potential concern. For participants 5, 6 and 7, independent PBMC aliquots were shipped and analyzed at different times, allowing us to investigate whether independent sampling of HIV *nef* sequences from blood yielded consistent distributions. It did in all cases: proviruses isolated from independent PBMC aliquots exhibited no evidence of compartmentalization by any method (**Table 3**), supporting their sampling from the same overall gene pool. Moreover, for seven participants (1, 3-5, 7-9) sufficient HIV sequences were isolated from different sections of the same testis (see tip label annotations in **Figure 4**), yielding eight opportunities to investigate whether independent samplings of the same site yielded consistent HIV sequence distributions (for participant 5, independent samplings of both right and left testes yielded sufficient sequences for evaluation). Of these eight within-testes datasets, consistent evidence of overall compartmentalization was only detected in one: the right testis of participant 7 (**Table 3**). Note that consistent support for within-tissue compartmentalization in participant 7's right testis did *not* remain after restricting to unique sequences (not shown), though limited statistical power is acknowledged. Phylogenetic segregation of the HIV sequences retrieved from these two independent samplings of the right testis is indeed apparent (**Figure 4J**), however further

sampling of additional tissue sections yielded HIV sequences that intermixed with these, and it is important to underscore that there was no support for compartmentalization *between* blood and testes within this participant (**Figure 6**). Together, these observations suggest that HIV sampling from blood and testes was not majorly biased by the specific aliquot or tissue section analyzed.

For seven participants (1-5, 7, 9) we had sufficient HIV *nef* sequences to investigate compartmentalization *between* right and left testes. When analyzing *overall* within-host datasets, only participant 9 consistently exhibited between-testes compartmentalization (**Table 2**). However, after restricting analysis to only unique *nef* sequences per site, no consistent evidence of compartmentalization remained. Consistent with the results of our analyses of HIV sequences in blood and tissue, these results suggest that in the rare cases where population structure is observed *between* proviral populations in right and left testes, that this is driven by the differential distribution of identical sequences between sites.

HIV immune escape in blood and testes

Given that cellular immune responses are suppressed in the testes (21, 23), we wished to investigate whether blood and testes HIV sequences differed with respect to immune escape mutation burden. We investigated this two ways. Total escape burden was estimated by identifying all HIV codons associated with one or more host HLA allele(s) and classifying the HIV residue at that site as adapted (inferred escaped) or susceptible based on published lists of HLA-associated polymorphisms defined at the population level in HIV subtype B (55). As such, analysis was restricted to the 9 subtype B-infected participants. For each sequence, we calculated the % HLA-associated sites exhibiting an adapted (or possibly adapted) form, and computed the mean for each within-host dataset (*e.g.* participant 7's dataset is 43% adapted to host HLA; **Figure 7A**). Next, we estimated within-host escape complexity by quantifying the percentage of

published or predicted optimally-described HLA-restricted CTL epitopes exhibiting within-host amino acid variation (*e.g.* 5/5 [100%] for participant 7). Both metrics correlated positively in blood and testes (Spearman's r=0.98, p=0.0001 for escape burden, **Figure 7B**; and Spearman's r=0.93, p=0.001 for escape complexity, **Figure 7C**). In fact, these values were highly concordant (Lin's concordance coefficient for the data presented in Figure 7B was 0.97; 95% CI 0.90-0.99 while that for Figure 7C is 0.95; 95% CI 0.80-0.99). Examples of the generally concordant nature of within-host HIV variation in key HLA-restricted epitopes in blood and testes are shown in **Figure 7D**. Consistent with previous reports (15), these examples also reveal that HLA-susceptible and adapted forms of the same CTL epitope commonly co-exist in the reservoir: 6 of the 9 subtype B-infected individuals harbored at least one epitope where this occurred.

Discussion

The availability of matched blood and tissue from our unique cohort (21, 23) allowed the first genetic exploration of the hypothesis that the testes constitute a distinctive HIV sanctuary site in healthy HIV-infected humans on suppressive cART (21, 25). Our study confirmed that CD4+ T-cells are present in testicular tissue and that HIV DNA could be detected in the testes all participants, though in two cases these were below quantification limits. Overall, testes proviral burdens measured in the present study were lower than those previously reported in an autopsy study of cART-treated individuals, however in the latter study some individuals had discontinued cART near the end of life, which could have increased proviral levels in tissue (30). Proviral burden in blood correlated with that in testes (**Figure 1C)**, suggesting that individuals with larger blood reservoirs will tend to have larger tissue reservoirs (57). These observations confirm the testes as a site of HIV persistence during cART.

392 Our results also yielded insights into within-host proviral diversity. First, identical HIV
393 *nef* sequences, which are suggestive of clonal expansion, were observed at markedly different
394 frequencies within and between participants. While in some (*e.g.* participants 1 and 5) a
395 particular *nef* variant dominated within testes, in others (*e.g.* participant 2) a particular variant
396 dominated in blood, while in yet others (*e.g.* participant 6) identical *nef* sequences were
397 relatively rare. These observations are consistent with clonal expansion of latently-infected cells
398 as a driver of HIV persistence during cART in many persons (10, 16-18, 58, 59). However, the
399 frequencies of identical *nef* sequences followed no consistent pattern between blood and testes,
400 suggesting that one cannot generalize, based on location alone, where clonally-expanded
401 latently-infected cell populations are more likely to reside. Disregarding clonal expansion
402 however, within-host blood and testes HIV diversity correlated significantly (**Figure 5**),
403 suggesting that individuals with diverse blood reservoirs will also harbor diverse HIV sequences
404 in testes.
405 Our study also sheds some light on two outstanding questions: first, whether the within-
406 host testicular microenvironment may be distinct enough to influence proviral composition
407 therein, and second, in the cases where blood-testes HIV genetic compartmentalization is
408 detected, how this arises (31). Our observations reveal that, while blood-testes genetic
409 compartmentalization was reasonably common if one analyzed within-host HIV *nef* sequences as
410 a whole (two established tests consistently deemed 60% of overall within-host datasets to be
411 compartmentalized), no participant retained consistent evidence of blood-testes
412 compartmentalization when only unique HIV *nef* sequences per compartment were analyzed.
413 Strictly speaking therefore, testes proviral populations in the majority of cART-suppressed
414 individuals do differ from those in blood, but only because of their unbalanced distributions of

identical sequences. This in turn suggests that the principal mechanism in which blood and testes HIV reservoirs differ from one another is *not* via seeding of divergent viral sequences therein pre-cART, nor due to restricted trafficking of viral lineages between blood and testes. Rather, our observations suggest that the extent and/or dynamics of proliferation of latently HIV-infected cells (and/or subsequent waning of descendant populations (10)) can differ substantially between blood and testes, and that this in turn drives the differences in HIV population structure between these sites. While it is tempting to speculate that the site harboring the sequence in abundance is where the clonal expansion occurred, and the other is where clonal descendants subsequently migrated, we cannot rule out that cells harboring identical HIV proviruses pre-existed in both sites as a result of a previous expansion and migration event, and that these underwent another expansion, in parallel, more recently. Our observation that HIV population structure, when present, is due to the differential distribution of identical sequences across anatomical sites, is consistent with studies of viral diversity within the female genital tract (53, 60). More broadly, our observations are consistent with the notion that major blood-tissue compartmentalization during cART, at least in terms of unique HIV lineages, may not be the norm (61-63).

At first glance, our finding that the composition of *unique* proviral lineages in testes does not differ significantly from that in blood may appear to contradict the former's status as an HIV sanctuary site (21, 64). The recovery of proviruses from testicular tissue of all participants however confirms the testes as a site of HIV persistence during cART. Rather, our results suggest that the testes microenvironment may not be sufficiently unique to persistently maintain distinct HIV lineages therein, at least not over the timecourse of cART suppression of the participants studied. Similarly, while reduced anti-HIV immune responses likely facilitate HIV persistence in the testes in general, our observation that HIV immune escape burden correlated

strongly between blood and testes (**Figure 5**) suggests that the extent of local immunosuppression may not be sufficient to markedly influence the immune escape landscape therein. However, there is an alternative explanation for the lack of genetic compartmentalization of unique HIV lineages between blood and testes: we cannot exclude the possibility that low-level HIV replication may still occur in certain tissues, including testes, during cART, and that this could lead to the ongoing re-shaping of HIV populations persisting in blood.

Other limitations of our study and data interpretation should also be acknowledged. Though proviral sequences were obtained via single-template amplification, only a subgenomic HIV region (*nef*) was analyzed due to the challenges of recovering HIV from testes (including from two participants for whom testes proviral loads were below quantification limits). As such, we cannot rule out that sequences that are intact within *nef* did not harbor defects elsewhere in the genome, nor can we definitively classify identical *nef* sequences as being derived from clonally expanded cell populations. As HIV sequences were isolated from PBMC or whole tissue, the specific immune cells harboring them are unknown. This may be particularly important for testes, which are rich in macrophages, which may be contributing to HIV persistence there (30). Limited sampling must be acknowledged, particularly from testes, which could compromise our power to detect compartmentalization. However, our independent recovery of the same sequence from separate testis sections, our lack of observation of compartmentalization within and between testes in 21/22 comparisons performed (**Tables 2 and 3**), our recovery of at least one identical HIV sequence from both blood and testes in each participant, and the observation that, for many participants, increased proviral sampling largely yielded sequences that had been recovered previously, suggest that our results are not simply attributable to biased and/or limited sampling. Furthermore, our qualitative observation that

blood and testes sequences were reasonably well-intermixed in all phylogenies (that is, that no dataset exhibited overt segregation by site) further supports the notion that these two sites do not differ substantively in terms of unique HIV lineage distribution. We must also acknowledge that limited within-host diversity can contribute to uncertainty in phylogenetic reconstruction (as evidenced by low bootstrap values for some nodes in the within-host trees), which could in turn influence the results of tree-based compartmentalization tests, but the relatively high concordance between tree- and distance-based tests (**Figure 6, Tables 2 and 3**) suggests this is not a major concern. As *nef* was sequenced, the possibility of differential distribution of antiretroviral resistance mutations across sites (as a result of reduced penetration of some antiretrovirals into testes (64)) could not be investigated. As clinical histories were unknown, we could not investigate correlates of blood-tissue HIV diversity and compartmentalization, nor can we rule out the possibility that evidence of genetic compartmentalization between blood and tissues may only become apparent after much more prolonged durations of cART suppression (*e.g.* as a result of latently-infected cells having different half-lives in blood vs. testes as a result of reduced immune surveillance in the latter). Finally, proviral sequences could not be studied in context of pre-cART viral populations, as these samples were not available.

In conclusion, and again noting the caveats associated with subgenomic HIV sequencing, our observation that *unique* proviral sequence distribution and immune escape mutation burden in testicular tissue is not unrepresentative of that within blood suggests that diversity and escape-related barriers to HIV eradication are likely to be of broadly similar magnitude across these sites. However, the marked differences in the quantity and distribution of identical HIV sequences between hosts and between anatomical sites within-host, underscore the potential

challenges of differentially-distributed clonally-expanded latently-infected cell populations to HIV eradication.

Acknowledgements

The research reported in this publication was funded in part by the Canadian HIV Cure Enterprise Team Grants HIG-133050 (to J-P.R./Z.L.B./A.F.Y.P.) and HB2-164064 (to J-P.R.) from the Canadian Institutes for Health Research (CIHR) in partnership with CANFAR and the International AIDS Society (IAS), by CIHR team grant HB1-164063 (to Z.L.B), by CIHR project grants PJT-159625 (to J.B.J./Z.L.B) and PJT-155990 (to A.F.Y.P.), by the National Institutes of Health under award number NIHR21A127029 (to J.B.J/A.F.Y.P./Z.L.B.) by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health with co-funding support from the National Institute on Drug Abuse, the National Institute of Mental Health, and the National Institute of Neurological Disorders and Stroke under award number UM1AI126617 (to Z.L.B.) and a Simon Fraser University Next Big Question fund award (to Z.L.B.). This study was also supported by the Fonds de recherche du Québec- Santé (FRQ- S): Réseau SIDA/Maladies infectieuses and Thérapie cellulaire (to J-P.R.). N.N.K. was supported by a CIHR Frederick Banting and Charles Best M.Sc. award. F.H.O. was supported by the Canadian Queen Elizabeth II Diamond Jubilee Scholarships program, a joint initiative of the Rideau Hall Foundation, Community Foundations of Canada and the Association of Universities and Colleges of Canada; and also received a fellowship from the Sub-Saharan African Network for TB/HIV Research Excellence (SANTHE), a DELTAS Africa Initiative [grant # DEL-15-006]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant # 107752/Z/15/Z] and the UK government. M.A.J. holds the CIHR Canada Research Chair tier 2 in Immuno-virology. A.F.Y.P.

508 is supported by a CIHR New Investigator Award (FRN-130609). N.C. is supported by a
509 Research Scholar Career Award of the Quebec Health Research Fund (FRQ-S, #253292). Z.L.B.
510 is supported by a Michael Smith Foundation for Health Research (MSFHR) Scholar Award. J-
511 P.R. is the holder of the Louis Lowenstein Chair in Hematology and Oncology at McGill
512 University. The content is solely the responsibility of the authors and does not necessarily
513 represent the official views of the National Institutes of Health, AAS, NEPAD Agency,
514 Wellcome Trust, UK government or other funding agencies listed.

515 We thank Gabriel Stang and Jeff Knaggs for software programming and bioinformatics
516 assistance, and Mark Brockman for helpful discussions. We are grateful to the participants who
517 made this study possible by their donation of blood and tissue.

References

1. **Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, Chaisson RE, Quinn TC, Chadwick K, Margolick J, Brookmeyer R, Gallant J, Markowitz M, Ho DD, Richman DD, Siliciano RF.** 1997. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science **278:**1295-1300.
2. **Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, Gupta P, Rinaldo CR, Learn GH, He X, Huang XL, Mullins JI.** 1999. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J Virol **73:**10489-10502.
3. **Herbeck JT, Rolland M, Liu Y, McLaughlin S, McNevin J, Zhao H, Wong K, Stoddard JN, Raugi D, Sorensen S, Genowati I, Birditt B, McKay A, Diem K, Maust BS, Deng W, Collier AC, Stekler JD, McElrath MJ, Mullins JI.** 2011. Demographic processes affect HIV-1 evolution in primary infection before the onset of selective processes. J Virol **85:**7523-7534.
4. **Dapp MJ, Kober KM, Chen L, Westfall DH, Wong K, Zhao H, Hall BM, Deng W, Sibley T, Ghorai S, Kim K, Chen N, McHugh S, Au L, Cohen M, Anastos K, Mullins JI.** 2017. Patterns and rates of viral evolution in HIV-1 subtype B infected females and males. PLoS One **12:**e0182443.
5. **Zanini F, Brodin J, Thebo L, Lanz C, Bratt G, Albert J, Neher RA.** 2015. Population genomics of intrapatient HIV-1 evolution. Elife **4**.
6. **Whitney JB, Hill AL, Sanisetty S, Penaloza-MacMaster P, Liu J, Shetty M, Parenteau L, Cabral C, Shields J, Blackmore S, Smith JY, Brinkman AL, Peter LE, Mathew SI, Smith KM, Borducchi EN, Rosenbloom DI, Lewis MG, Hattersley J, Li B, Hesselgesser J, Geleziunas R, Robb ML, Kim JH, Michael NL, Barouch DH.** 2014. Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. Nature **512:**74-77.
7. **Brodin J, Zanini F, Thebo L, Lanz C, Bratt G, Neher RA, Albert J.** 2016. Establishment and stability of the latent HIV-1 DNA reservoir. Elife **5**.
8. **Jones BR, Kinloch NN, Horacsek J, Ganase B, Harris M, Harrigan PR, Jones RB, Brockman MA, Joy JB, Poon AFY, Brumme ZL.** 2018. Phylogenetic approach to recover integration dates of latent HIV sequences within-host. Proc Natl Acad Sci U S A **115:**E8958-E8967.
9. **Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, Margolick JB, Kovacs C, Gange SJ, Siliciano RF.** 2003. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. Nat Med **9:**727-728.
10. **Wang Z, Gurule EE, Brennan TP, Gerold JM, Kwon KJ, Hosmane NN, Kumar MR, Beg SA, Capoferri AA, Ray SC, Ho YC, Hill AL, Siliciano JD, Siliciano RF.** 2018. Expanded cellular clones carrying replication-competent HIV-1 persist, wax, and wane. Proc Natl Acad Sci U S A **115:**E2575-E2584.
11. **Hiener B, Horsburgh BA, Eden JS, Barton K, Schlub TE, Lee E, von Stockenstrom S, Odevall L, Milush JM, Liegler T, Sinclair E, Hoh R, Boritz EA, Douek D, Fromentin R, Chomont N, Deeks SG, Hecht FM, Palmer S.** 2017. Identification of Genetically Intact HIV-1 Proviruses in Specific CD4+ T Cells from Effectively Treated Participants. Cell Rep **21:**813-822.

12. **Kearney M, Spindler J, Shao W, Maldarelli F, Palmer S, Hu SL, Lifson JD, KewalRamani VN, Mellors JW, Coffin JM, Ambrose Z.** 2011. Genetic diversity of simian immunodeficiency virus encoding HIV-1 reverse transcriptase persists in macaques despite antiretroviral therapy. J Virol **85:**1067-1076.
13. **Josefsson L, von Stockenstrom S, Faria NR, Sinclair E, Bacchetti P, Killian M, Epling L, Tan A, Ho T, Lemey P, Shao W, Hunt PW, Somsouk M, Wylie W, Douek DC, Loeb L, Custer J, Hoh R, Poole L, Deeks SG, Hecht F, Palmer S.** 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 U S A **110:**E4987-4996.
14. **Deng K, Pertea M, Rongvaux A, Wang L, Durand CM, Ghiaur G, Lai J, McHugh HL, Hao H, Zhang H, Margolick JB, Gurer C, Murphy AJ, Valenzuela DM, Yancopoulos GD, Deeks SG, Strowig T, Kumar P, Siliciano JD, Salzberg SL, Flavell RA, Shan L, Siliciano RF.** 2015. Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations. Nature **517:**381-385.
15. **Brumme ZL, Sudderuddin H, Ziemniak C, Luzuriaga K, Jones BR, Joy JB, Cunningham CK, Greenough T, Persaud D.** 2019. Genetic complexity in the replication-competent latent HIV reservoir increases with untreated infection duration in infected youth. AIDS **33:**211-218.
16. **Bui JK, Sobolewski MD, Keele BF, Spindler J, Musick A, Wiegand A, Luke BT, Shao W, Hughes SH, Coffin JM, Kearney MF, Mellors JW.** 2017. Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. PLoS Pathog **13:**e1006283.
17. **Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, Spindler J, Ferris AL, Mellors JW, Kearney MF, Coffin JM, Hughes SH.** 2014. HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. Science **345:**179-183.
18. **Wagner TA, McLaughlin S, Garg K, Cheung CY, Larsen BB, Styrchak S, Huang HC, Edlefsen PT, Mullins JI, Frenkel LM.** 2014. HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. Science **345:**570-573.
19. **Lee GQ, Orlova-Fink N, Einkauf K, Chowdhury FZ, Sun X, Harrington S, Kuo HH, Hua S, Chen HR, Ouyang Z, Reddy K, Dong K, Ndung'u T, Walker BD, Rosenberg ES, Yu XG, Lichterfeld M.** 2017. Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4+ T cells. J Clin Invest **127:**2689-2696.
20. **Mital P, Hinton BT, Dufour JM.** 2011. The blood-testis and blood-epididymis barriers are more than just their tight junctions. Biol Reprod **84:**851-858.
21. **Jenabian MA, Costiniuk CT, Mehraj V, Ghazawi FM, Fromentin R, Brousseau J, Brassard P, Belanger M, Ancuta P, Bendayan R, Chomont N, Routy JP, Orchid study g.** 2016. Immune tolerance properties of the testicular tissue as a viral sanctuary site in ART-treated HIV-infected adults. AIDS **30:**2777-2786.
22. **Le Tortorec A, Le Grand R, Denis H, Satie AP, Mannioui K, Roques P, Maillard A, Daniels S, Jegou B, Dejucq-Rainsford N.** 2008. Infection of semen-producing organs by SIV during the acute and chronic stages of the disease. PLoS One **3:**e1792.
23. **Ponte R, Dupuy FP, Brimo F, Mehraj V, Brassard P, Belanger M, Yurchenko E, Jenabian MA, Bernard NF, Routy JP, group Os.** 2018. Characterization of myeloid

cell populations in human testes collected after sex reassignment surgery. J Reprod Immunol **125:**16-24.

24. **De Rose R, Fernandez CS, Hedger MP, Kent SJ, Winnall WR.** 2013. Characterisation of macaque testicular leucocyte populations and T-lymphocyte immunity. J Reprod Immunol **100:**146-156.

25. **Winnall WR, Lloyd SB, De Rose R, Alcantara S, Amarasena TH, Hedger MP, Girling JE, Kent SJ.** 2015. Simian immunodeficiency virus infection and immune responses in the pig-tailed macaque testis. J Leukoc Biol **97:**599-609.

26. **Nikolova M, Carriere M, Jenabian MA, Limou S, Younas M, Kok A, Hue S, Seddiki N, Hulin A, Delaneau O, Schuitemaker H, Herbeck JT, Mullins JI, Muhtarova M, Bensussan A, Zagury JF, Lelievre JD, Levy Y.** 2011. CD39/adenosine pathway is involved in AIDS progression. PLoS Pathog **7:**e1002110.

27. **Shehu-Xhilaga M, Kent S, Batten J, Ellis S, Van der Meulen J, O'Bryan M, Cameron PU, Lewin SR, Hedger MP.** 2007. The testis and epididymis are productively infected by SIV and SHIV in juvenile macaques during the post-acute stage of infection. Retrovirology **4:**7.

28. **Matusali G, Dereuddre-Bosquet N, Le Tortorec A, Moreau M, Satie AP, Mahe D, Roumaud P, Bourry O, Sylla N, Bernard-Stoecklin S, Pruvost A, Le Grand R, Dejucq-Rainsford N.** 2015. Detection of Simian Immunodeficiency Virus in Semen, Urethra, and Male Reproductive Organs during Efficient Highly Active Antiretroviral Therapy. J Virol **89:**5772-5787.

29. **North TW, Higgins J, Deere JD, Hayes TL, Villalobos A, Adamson L, Shacklett BL, Schinazi RF, Luciw PA.** 2010. Viral sanctuaries during highly active antiretroviral therapy in a nonhuman primate model for AIDS. J Virol **84:**2913-2922.

30. **Lamers SL, Rose R, Maidji E, Agsalda-Garcia M, Nolan DJ, Fogel GB, Salemi M, Garcia DL, Bracci P, Yong W, Commins D, Said J, Khanlou N, Hinkin CH, Sueiras MV, Mathisen G, Donovan S, Shiramizu B, Stoddart CA, McGrath MS, Singer EJ.** 2016. HIV DNA Is Frequently Present within Pathologic Tissues Evaluated at Autopsy from Combined Antiretroviral Therapy-Treated Patients with Undetectable Viral Loads. J Virol **90:**8968-8983.

31. **Nickle DC, Jensen MA, Shriner D, Brodie SJ, Frenkel LM, Mittler JE, Mullins JI.** 2003. Evolutionary indicators of human immunodeficiency virus type 1 reservoirs and compartments. J Virol **77:**5540-5546.

32. **Theiler J, Yoon H, Yusim K, Picker LJ, Fruh K, Korber B.** 2016. Epigraph: A Vaccine Design Tool Applied to an HIV Therapeutic Vaccine and a Pan-Filovirus Vaccine. Sci Rep **6:**33987.

33. **Tumiotto C, Riviere L, Bellecave P, Recordon-Pinson P, Vilain-Parce A, Guidicelli GL, Fleury H, Provir/Latitude 45 collaborating g.** 2017. Sanger and Next-Generation Sequencing data for characterization of CTL epitopes in archived HIV-1 proviral DNA. PLoS One **12:**e0185211.

34. **Houzet L, Perez-Losada M, Matusali G, Deleage C, Dereuddre-Bosquet N, Satie AP, Aubry F, Becker E, Jegou B, Le Grand R, Keele BF, Crandall KA, Dejucq-Rainsford N.** 2018. Seminal Simian Immunodeficiency Virus in Chronically Infected Cynomolgus Macaques Is Dominated by Virus Originating from Multiple Genital Organs. J Virol **92**.

35. **Vandergeeten C, Fromentin R, Merlini E, Lawani MB, DaFonseca S, Bakeman W, McNulty A, Ramgopal M, Michael N, Kim JH, Ananworanich J, Chomont N.** 2014. Cross-clade ultrasensitive PCR-based assays to measure HIV persistence in large-cohort studies. J Virol **88:**12385-12396.
36. **Bruner KM, Wang Z, Simonetti FR, Bender AM, Kwon KJ, Sengupta S, Fray EJ, Beg SA, Antar AAR, Jenike KM, Bertagnolli LN, Capoferri AA, Kufera JT, Timmons A, Nobles C, Gregg J, Wada N, Ho YC, Zhang H, Margolick JB, Blankson JN, Deeks SG, Bushman FD, Siliciano JD, Laird GM, Siliciano RF.** 2019. A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. Nature **566:**120-125.
37. **Rose PP, Korber BT.** 2000. Detecting hypermutations in viral sequences with an emphasis on G --> A hypermutation. Bioinformatics **16:**400-401.
38. **Martin DP, Murrell B, Golden M, Khoosal A, Muhire B.** 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol **1:**vev003.
39. **Katoh K, Misawa K, Kuma K, Miyata T.** 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res **30:**3059-3066.
40. **Larsson A.** 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics **30:**3276-3278.
41. **Darriba D, Taboada GL, Doallo R, Posada D.** 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods **9:**772.
42. **Price MN, Dehal PS, Arkin AP.** 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One **5:**e9490.
43. **Tavare S.** Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences, p 57–86. *In* (ed),
44. **Yu G, Smith D, Zhu H, Guan Y, Lam TTY.** 2017. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution **8:**28-36.
45. **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.
46. **Hudson RR, Boos DD, Kaplan NL.** 1992. A statistical test for detecting geographic subdivision. Mol Biol Evol **9:**138-151.
47. **Slatkin M, Maddison WP.** 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics **123:**603-613.
48. **Pond SL, Frost SD, Muse SV.** 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics **21:**676-679.
49. **Achaz G, Palmer S, Kearney M, Maldarelli F, Mellors JW, Coffin JM, Wakeley J.** 2004. A robust measure of HIV-1 population turnover within chronically infected individuals. Mol Biol Evol **21:**1902-1912.
50. **Chaillon A, Gianella S, Lada SM, Perez-Santiago J, Jordan P, Ignacio C, Karris M, Richman DD, Mehta SR, Little SJ, Wertheim JO, Smith DM.** 2018. Size, Composition, and Evolution of HIV DNA Populations during Early Antiretroviral Therapy and Intensification with Maraviroc. J Virol **92**.
51. **Kearney MF, Spindler J, Shao W, Yu S, Anderson EM, O'Shea A, Rehm C, Poethke C, Kovacs N, Mellors JW, Coffin JM, Maldarelli F.** 2014. Lack of detectable HIV-1

molecular evolution during suppressive antiretroviral therapy. PLoS Pathog **10:**e1004010.

52. **Gianella S, Kosakovsky Pond SL, Oliveira MF, Scheffler K, Strain MC, De la Torre A, Letendre S, Smith DM, Ellis RJ.** 2016. Compartmentalized HIV rebound in the central nervous system after interruption of antiretroviral therapy. Virus Evol **2:**vew020.

53. **Bull M, Learn G, Genowati I, McKernan J, Hitti J, Lockhart D, Tapia K, Holte S, Dragavon J, Coombs R, Mullins J, Frenkel L.** 2009. Compartmentalization of HIV-1 within the female genital tract is due to monotypic and low-diversity variants not distinct viral populations. PLoS One **4:**e7122.

54. **Cotton LA, Rahman MA, Ng C, Le AQ, Milloy MJ, Mo T, Brumme ZL.** 2012. HLA class I sequence-based typing using DNA recovered from frozen plasma. J Immunol Methods **382:**40-47.

55. **Carlson JM, Brumme CJ, Martin E, Listgarten J, Brockman MA, Le AQ, Chui CK, Cotton LA, Knapp DJ, Riddler SA, Haubrich R, Nelson G, Pfeifer N, Deziel CE, Heckerman D, Apps R, Carrington M, Mallal S, Harrigan PR, John M, Brumme ZL.** 2012. Correlates of protective cellular immunity revealed by analysis of population-level immune escape pathways in HIV-1. J Virol **86:**13202-13216.

56. **Frahm N, Baker B, Brander C.** 2008. Identification and Optimal Definition of HIV-Derived Cytotoxic T Lymphocyte (CTL) Epitopes for the Study of CTL Escape, Functional Avidity and Viral Evolution, p 3-24. *In* Korber BTM, Brander C, Haynes BF, Koup R, Moore JP, Walker BD, Watkins DI (ed), HIV Molecular Immunology 2008. Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico.

57. **Chomont N, El-Far M, Ancuta P, Trautmann L, Procopio FA, Yassine-Diab B, Boucher G, Boulassel MR, Ghattas G, Brenchley JM, Schacker TW, Hill BJ, Douek DC, Routy JP, Haddad EK, Sekaly RP.** 2009. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. Nat Med **15:**893-900.

58. **Mullins JI, Frenkel LM.** 2017. Clonal Expansion of Human Immunodeficiency Virus-Infected Cells and Human Immunodeficiency Virus Persistence During Antiretroviral Therapy. J Infect Dis **215:**S119-S127.

59. **Murray AJ, Kwon KJ, Farber DL, Siliciano RF.** 2016. The Latent Reservoir for HIV-1: How Immunologic Memory and Clonal Expansion Contribute to HIV-1 Persistence. J Immunol **197:**407-417.

60. **Bull ME, Learn GH, McElhone S, Hitti J, Lockhart D, Holte S, Dragavon J, Coombs RW, Mullins JI, Frenkel LM.** 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.

61. **Dahl V, Gisslen M, Hagberg L, Peterson J, Shao W, Spudich S, Price RW, Palmer S.** 2014. An example of genetically distinct HIV type 1 variants in cerebrospinal fluid and plasma during suppressive therapy. J Infect Dis **209:**1618-1622.

62. **Chun TW, Nickle DC, Justement JS, Meyers JH, Roby G, Hallahan CW, Kottilil S, Moir S, Mican JM, Mullins JI, Ward DJ, Kovacs JA, Mannon PJ, Fauci AS.** 2008. Persistence of HIV in Gut-Associated Lymphoid Tissue despite Long-Term Antiretroviral Therapy. J Infect Dis doi:10.1086/527324.

63. **Kearney MF, Anderson EM, Coomer C, Smith L, Shao W, Johnson N, Kline C, Spindler J, Mellors JW, Coffin JM, Ambrose Z.** 2015. Well-mixed plasma and tissue

viral populations in RT-SHIV-infected macaques implies a lack of viral replication in the tissues during antiretroviral therapy. Retrovirology **12:**93.

64. **Huang Y, Hoque MT, Jenabian MA, Vyboh K, Whyte SK, Sheehan NL, Brassard P, Belanger M, Chomont N, Fletcher CV, Routy JP, Bendayan R.** 2016. Antiretroviral drug transporters and metabolic enzymes in human testicular tissue: potential contribution to HIV-1 sanctuary site. J Antimicrob Chemother **71:**1954-1965.

751 **Figure Legends**

752 **Figure 1. Quantifying and characterizing the HIV reservoir in blood and testes.**

753 *Panel A*: Representative immunostaining of a frozen testis section (participant 5) showing cells

754 expressing CD3 (red) or CD4 (green). Nuclei were counterstained with DAPI (blue). CD4 T cells

755 ($CD3^+$ $CD4^+$) appear in yellow in the merged image. *Panel B:* HIV DNA loads, expressed in

756 copies per million cells in whole tissue lysate, in left and right testes. ND=not determined.

757 Trace=HIV DNA was detected but not quantifiable. *Panel C:* Marginal positive correlation

758 between HIV DNA loads in blood and testes (where the latter is expressed as average of left and

759 right measurements). *Panel D:* Unique HIV sequence distribution in blood and testes, matched

760 by participant. *Panel E:* Significant correlation between HIV DNA load and unique HIV

761 sequence distribution in blood. *Panel F:* Lack of significant correlation between HIV DNA load

762 and unique HIV sequence distribution in testes.

763

764 **Figure 2. Relationship between the total number of unique HIV sequences collected per**

765 **participant as a function of their total number of sequences collected.** The dotted horizontal

766 line denotes a hypothetical dataset where every sampled sequence is unique.

767

768 **Figure 3. Between-host HIV phylogeny.** Numbers on branches indicate bootstrap values

769 supporting within-host monophyletic clades. Scale in estimated substitutions per nucleotide site.

770 NL4-3 and MJ4 are subtype B and C reference sequences, respectively.

771

772 **Figure 4. Within-host HIV reservoir diversity in blood and testes.** Phylogenies, inferred from

773 nucleotide sequence alignments (HIV *nef*), are midpoint rooted, with scales denoting estimated

substitutions per nucleotide site. Matched highlighter plots, made from amino acid alignments, show substitutions relative to the master sequence (the top sequence in the phylogeny). Symbols denote sampling location: blood (filled circle), right testis (open square), left testis (open diamond). Red "A" and black "B" markers in participants 1, 3, 4, 5, 7, 8 and 9 identify sequences collected from independent sections of the same testis (HIV genetic compartmentalization results for these within-tissue comparisons are summarized in Table 3). Bootstrap values between 70-90% are reported to the left of their respective nodes; those > 90% are marked with asterisks.

Figure 5. Significant positive correlation between blood and testes proviral diversity. Values represent average within-host patristic (tip-to-tip phylogenetic) distances between all pairs of sequences sampled from each site, after collapsing identical *nef* sequences down to a single copy per compartment.

Figure 6. Genetic compartmentalization results for each participant. “Distance” denotes the results of Hudson Boos and Kaplan's nonparametric test for population structure, "Tree" denotes the results of the Slatkin-Maddison test, and “Consensus” denotes the final result (compartmentalization was declared only if both tests agreed). Blue denotes compartmentalization, grey denotes no compartmentalization. *Panel A* ("overall") shows results when comparing all *nef* sequences from blood to all those from testes; *Panel B* ("unique") shows results with identical *nef* sequences collapsed down to a single copy per compartment.

Figure 7. Immune escape landscape in blood and testes.

796 *Panel A*: Example of escape burden analysis using participant 7's HIV Nef amino acid
797 alignments in blood and testes. Red, orange and blue represent HLA-adapted, possibly adapted
798 and susceptible forms respectively. For each sequence, the sum of HLA-adapted and possibly
799 adapted sequences are divided by the total number of HLA-associated sites to yield the overall %
800 adapted sites (values shown at the end of each sequence). These values are then averaged to yield
801 site- and participant-specific averages. For viral sites associated with more than one HLA allele
802 (*e.g.* Nef codon 105), the inferred escape profile is shown for the underlined allele. *Panel B:*
803 Correlation between escape burden in blood and testes. *Panel C:* Correlation between epitope
804 complexity, defined here as the % of known or predicted HLA-restricted CTL epitopes
805 exhibiting amino acid variation, between blood and testes. *Panel D:* Examples of escaped and
806 susceptible forms coexisting within known or predicted CTL epitopes within participants'
807 reservoirs, where the height of the residue represents its frequency.

Table 1. Participant HLA and HIV sequence characteristics

ID	HLA-A	HLA-B	HLA-C	HIV Subtype	Intact Blood Sequences Total N (N; % unique)	Intact Testes Sequences Total N (N; % unique)
1	A*32:01 - A*34:02	B*15:18 - B*52:01	C*07:04 - C*16:01	B	30 (21; 70%)	99 (8; 8%)
2	A*23:01 - A*30:01	B*07:02 - B*42:01	C*07:02 - C*17:01	B	26 (2; 8%)	9 (7; 78%)
3	A*02:01 - A*25:01	B*08:01 - B*18:01	C*07:01 - C*12:03	B	36 (10; 28%)	40 (24; 60%)
4	A*11:01 - A*24:02	B*15:02 - B*35:03	C*08:01 - C*12:02	B	31 (20; 65%)	15 (12; 80%)
5	A*02:01 - A*26:01	B*15:01 - B*38:01	C*03:04 - C*12:03	B	31 (20; 65%)	84 (21; 25%)
6	A*33:01 - A*68:01	B*14:02 - B*44:02	C*07:04 - C*08:02	B	45 (40; 89%)	16 (15; 94%)
7	A*24:02 - A*29:02	B*18:25 - B*27:05	C*01:02 - C*12:03	B	37 (16; 43%)	13 (9; 69%)
8	A*02:01 - A*02:01	B*08:01 - B*44:02	C*05:01 - C*07:01	B	41 (23; 56%)	8 (5; 63%)
9	A*02:01 - A*11:01	B*18:01 - B*27:04	C*07:04 - C*12:02	B	45 (25; 56%)	21 (8; 38%)
10	A*26:01 - A*26:01	B*38:01 - B*51:01	C*12:03 - C*15:02	C	32 (7; 22%)	4 (3; 75%)
TOTAL					**354 (184; 52%)**	**309 (112; 36%)**

808

Table 2. Blood-testes and right-left testis HIV genetic compartmentalization results

Comparison	Test	Result	Participant 1	2	3	4	5	6	7	8	9	10
Blood vs. Testes Overall	Distance	K_{ST}[a]	0.61	0.53	0.09	0.06	0.17	0.01	0.002	0.02	0.26	0.11
		p value	**	**	**	*	**	NS	NS	NS	**	NS
	Tree	#Migrations[b]	7	4	16	9	18	10	12	7	7	3
		p value	**	**	**	*	**	**	NS	NS	**	*
Blood vs. Testes Unique	Distance	K_{ST}	0.10	0	0.04	0.01	0.004	0.01	0	0.02	0.10	0
		p value	NS	NS	NS	NS	NS	NS	NS	NS	*	NS
	Tree	#Migrations	3	2	8	9	15	10	9	4	6	2
		p value	**	NS	NS	NS	NS	*	NS	NS	NS	NS
Right Testis vs. Left Testis Overall	Distance	K_{ST}	0.14	0.02	0.03	0.05	0		0		0.65	
		p value	*	NS	NS	NS	NS		NS		**	
	Tree	#Migrations	38	3	9	1	30		4		2	
		p value	NS	NS	NS	*	NS		NS		**	
Right Testis vs. Left Testis Unique	Distance	K_{ST}	0	0.02	0	0.07	0		0		0.21	
		p value	NS	NS	NS	NS	NS		NS		*	
	Tree	#Migrations	3	3	9	1	9		4		2	
		p value	NS	NS	NS	*	NS		NS		NS	

[a] Statistic produced by Hudson, Boos and Kaplan's nonparametric test for population structure, where 0 indicates no compartmentalization and 1 indicates complete compartmentalization.

[b] The number of migrations between compartments required to justify the input phylogeny, as defined by the Slatkin-Maddison test. The smaller the number, the stronger the support for compartmentalization.

* denotes $0.01 < p < 0.05$, ** denotes $p < 0.01$

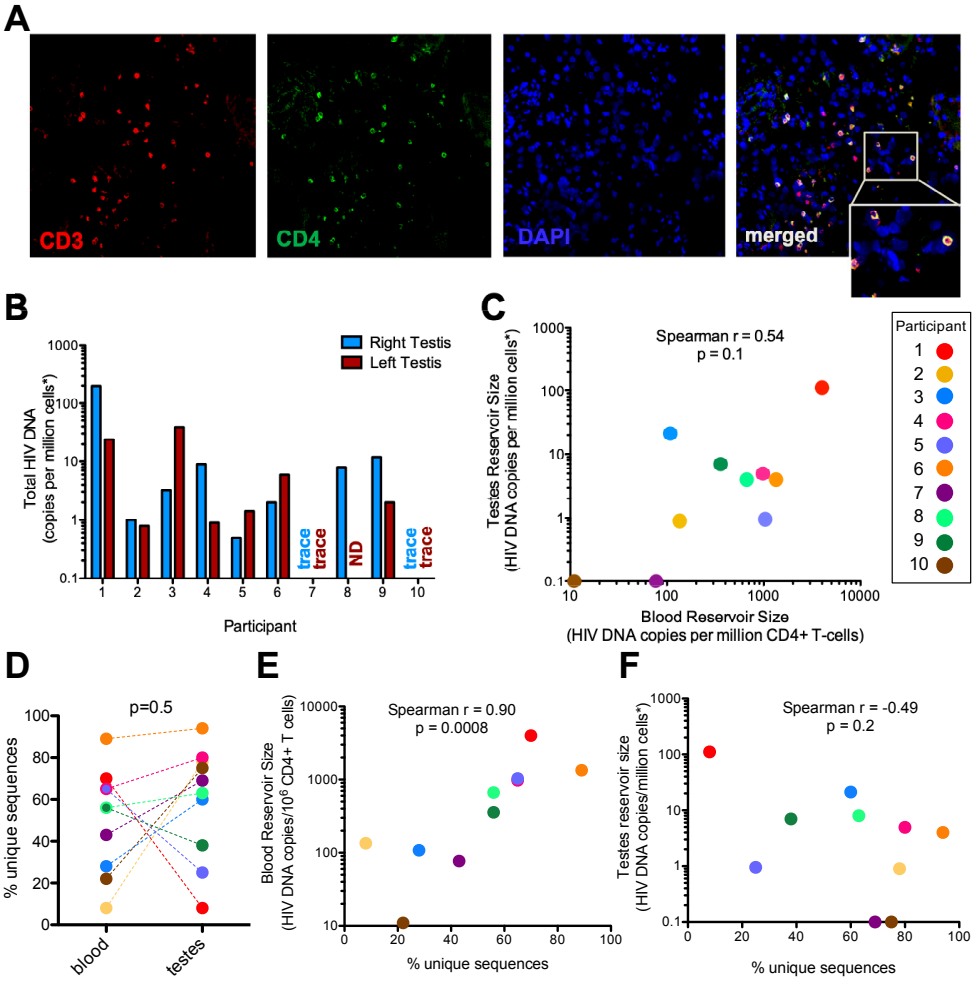
Table 3. Genetic compartmentalization when sampling HIV sequences from the same site or tissue

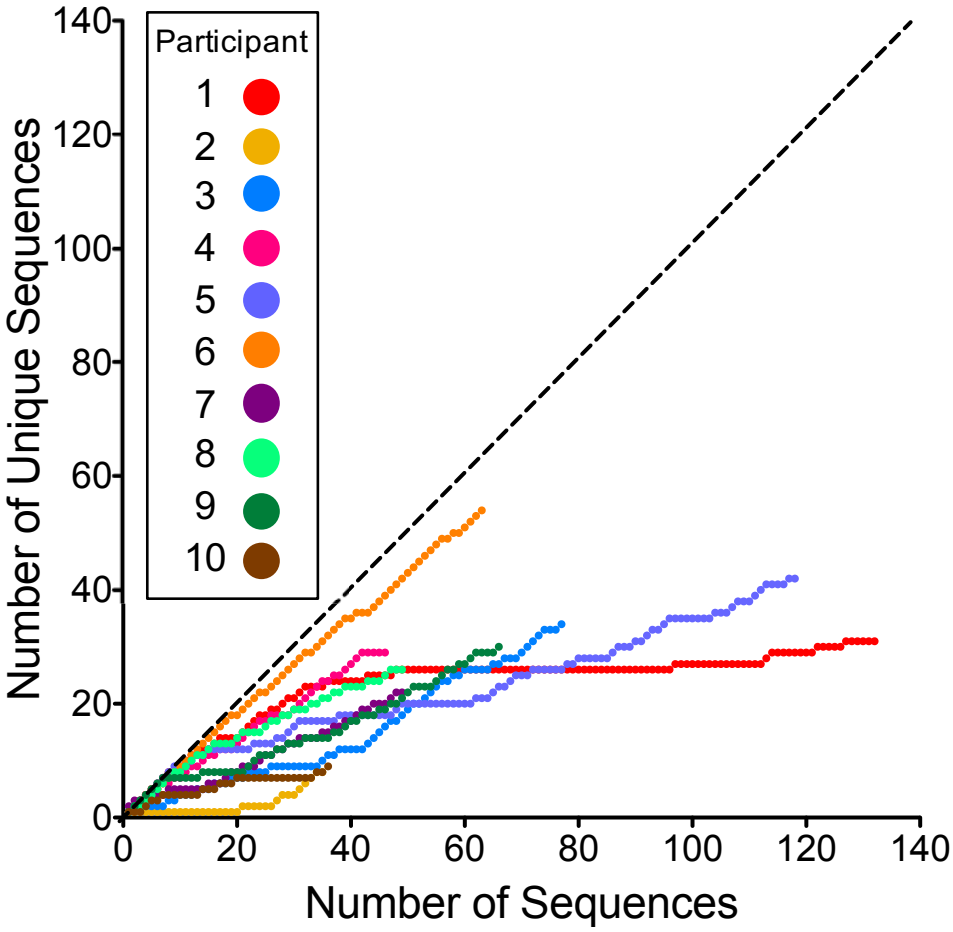
Comparison	Test	Result	Participant 1	2	3	4	5	6	7	8	9	10
PBMC Aliquot A vs. Aliquot B Overall	Distance	K_{ST}[a]					0.02	0	0			
		p value					NS	NS	NS			
	Tree	# Migrations[b]					9	11	13			
		p value					NS	NS	NS			
Right Testis Location A vs. Location B Overall	Distance	K_{ST}			0.01		0		0.51	0.03	0.273	
		p value			NS		NS		**	NS	NS	
	Tree	# Migrations			12		5		1	3	6	
		p value			NS		NS		*	NS	NS	
Left Testis Location A vs. Location B Overall	Distance	K_{ST}	0			0.09	0					
		p value	NS			NS	NS					
	Tree	# Migrations	17			3	4					
		p value	NS			NS	NS					

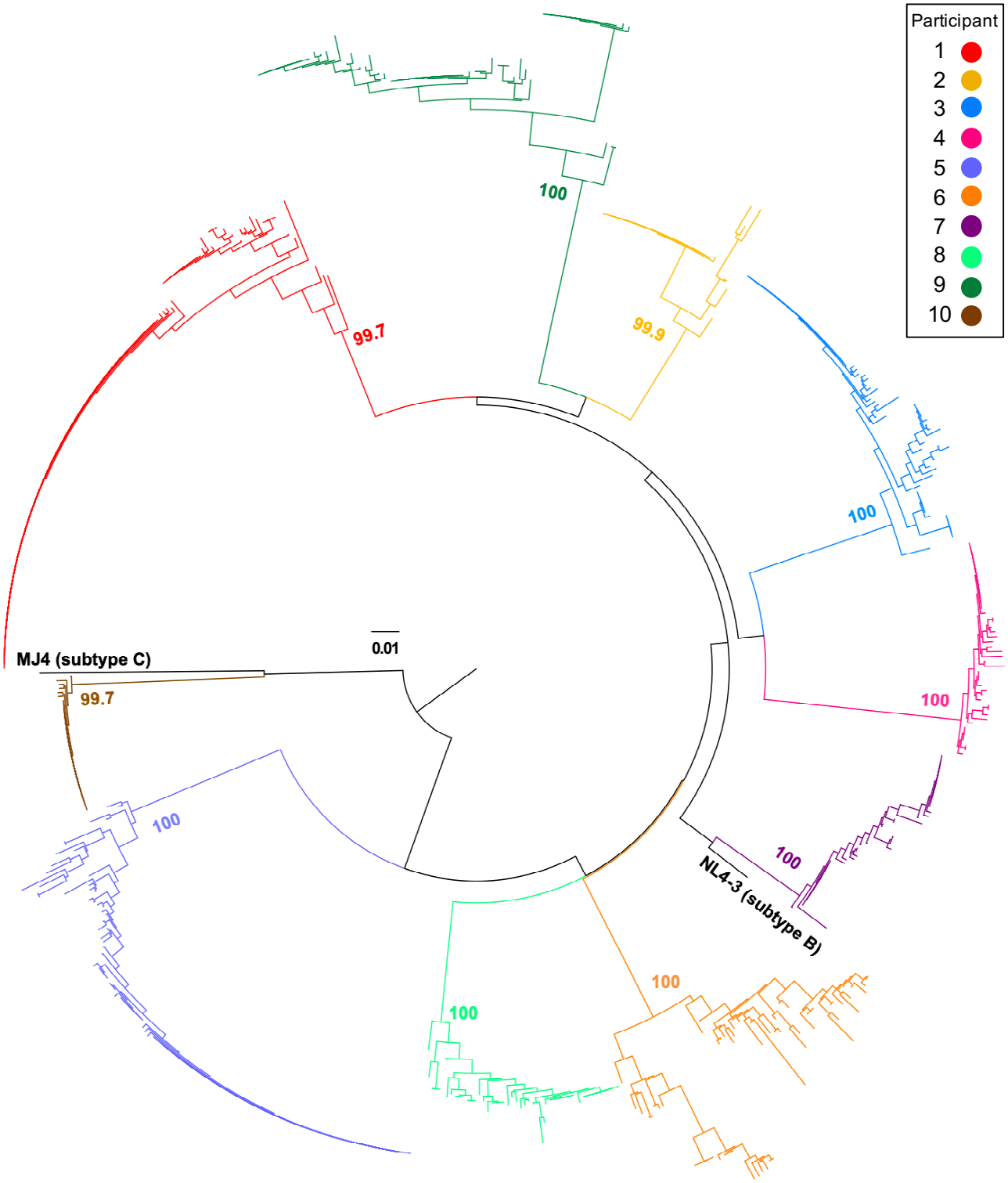
[a] Statistic produced by Hudson, Boos and Kaplan's nonparametric test for population structure, where 0 indicates no compartmentalization and 1 indicates complete compartmentalization.

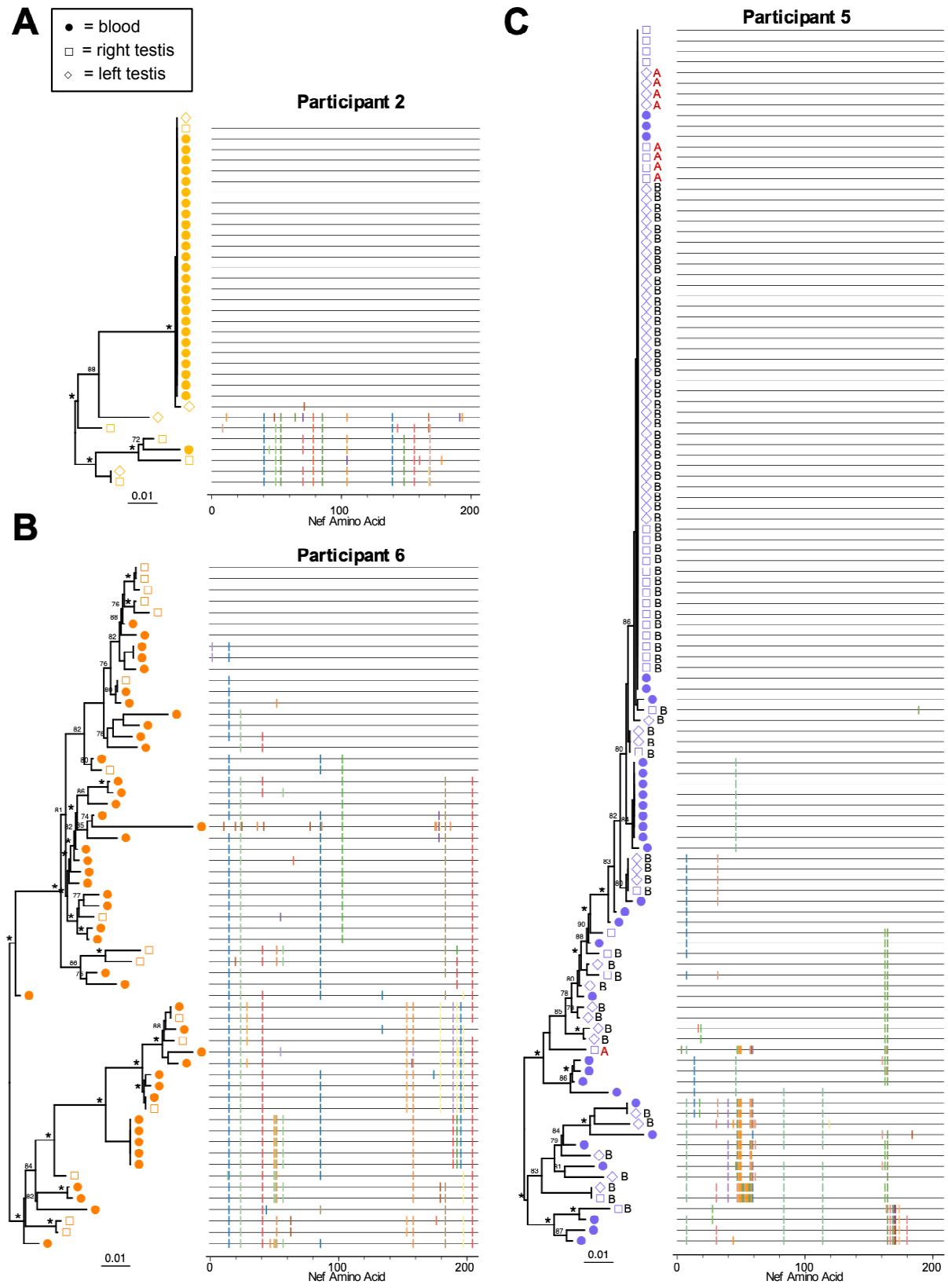
[b] The number of migrations between compartments required to justify the input phylogeny, as defined by the Slatkin-Maddison test. The smaller the number, the stronger the support for compartmentalization.

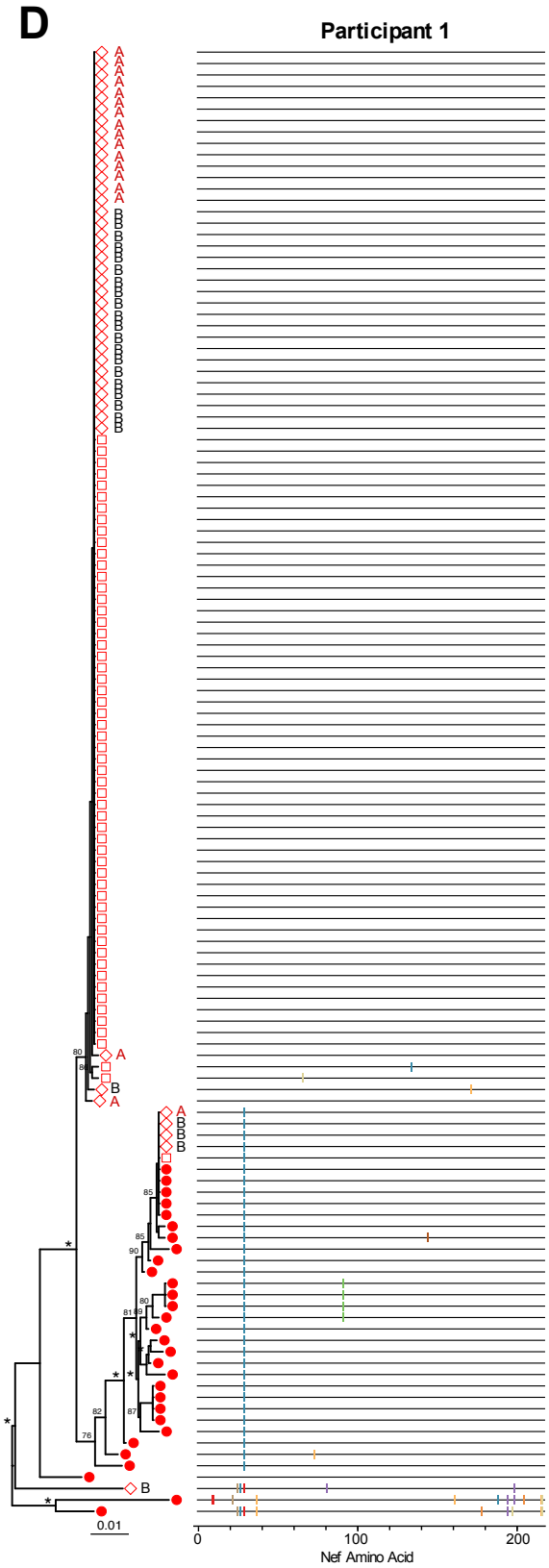
* denotes $0.01 < p < 0.05$, ** denotes $p < 0.01$

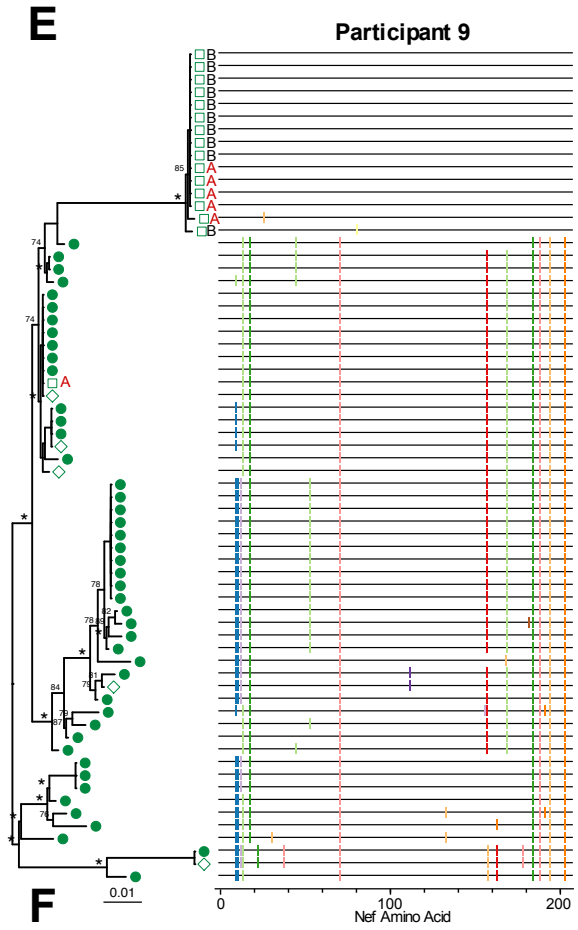


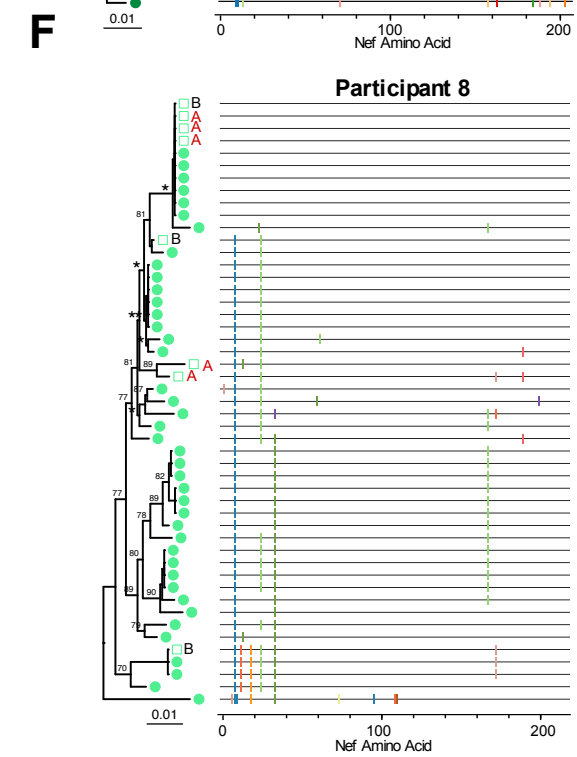


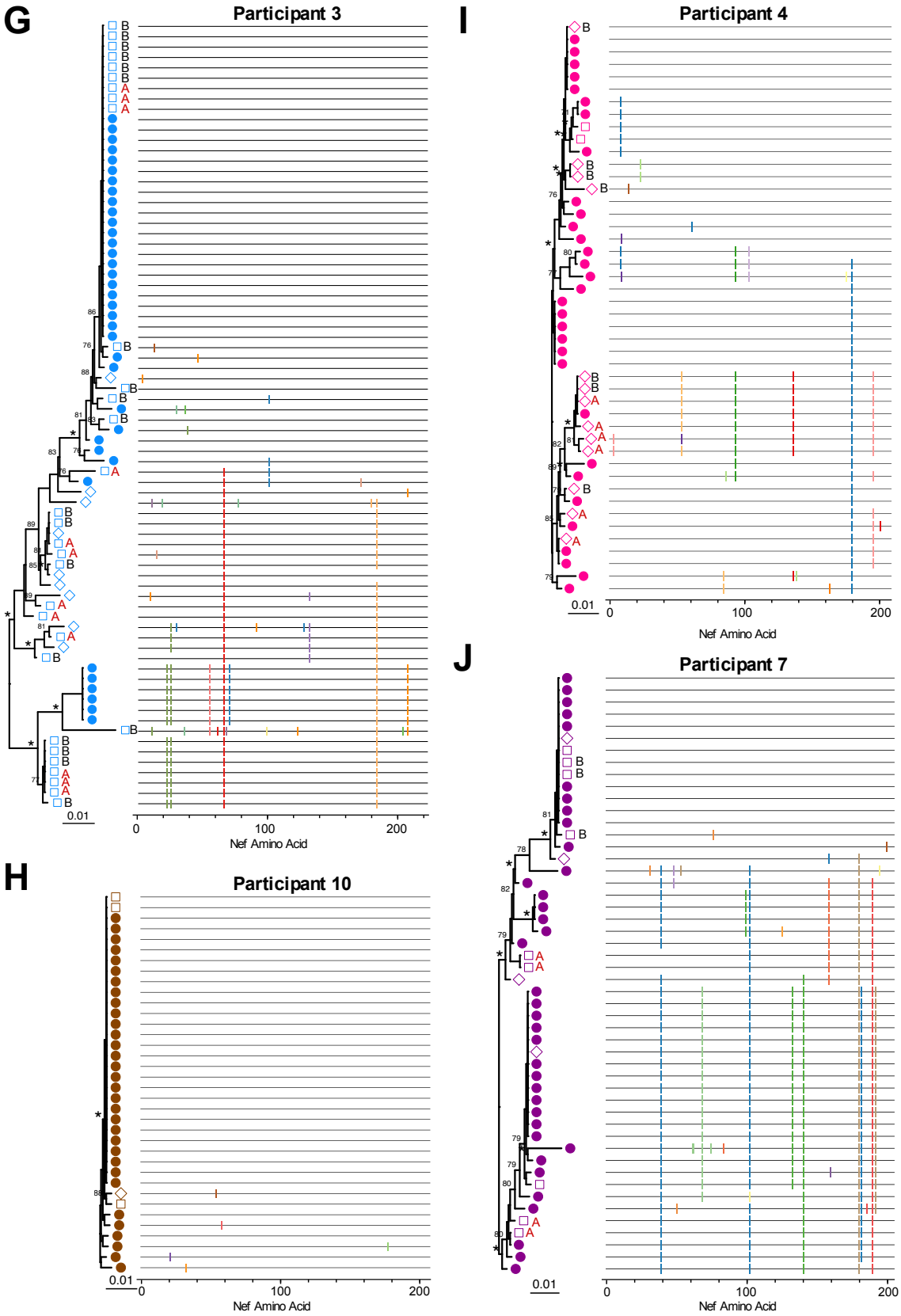


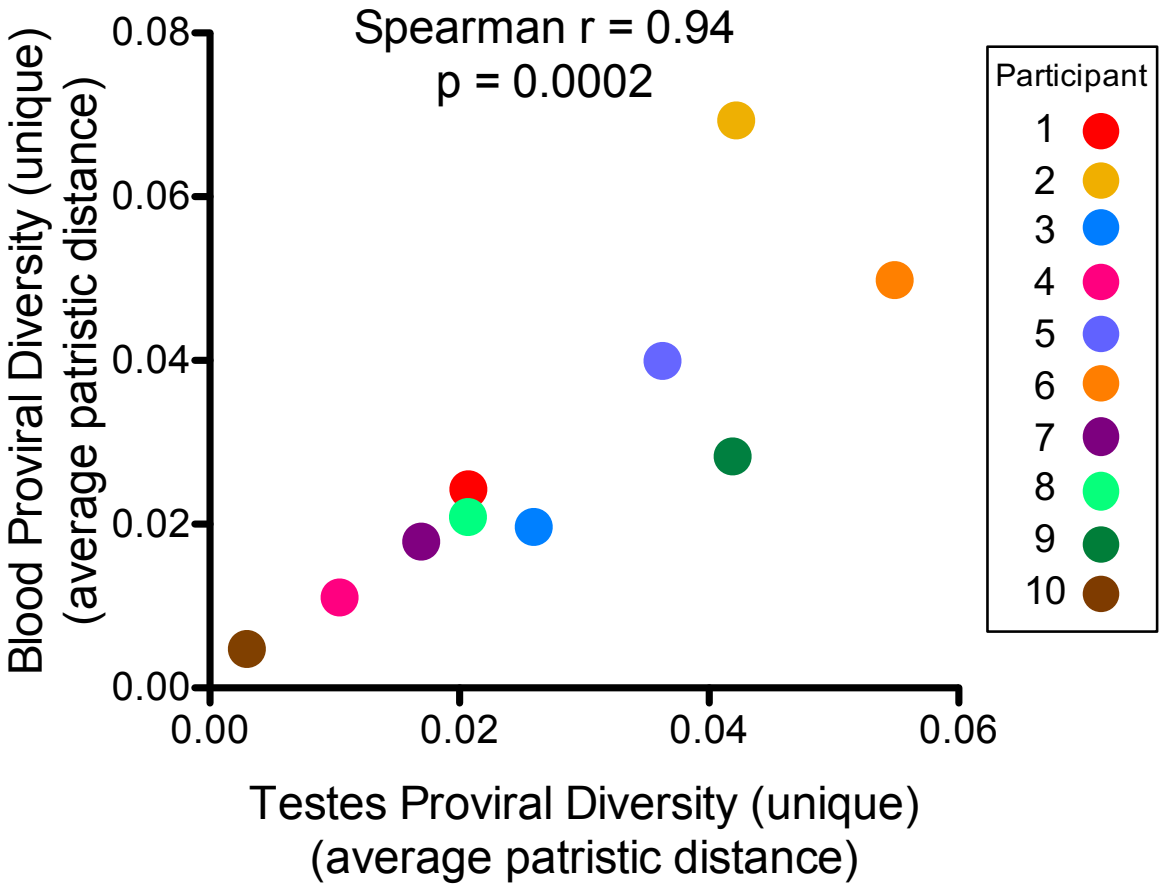












Journal of Virology