A Phylogenetic Approach to Recover Integration Dates of Latent Human Immunodeficiency Virus Sequences Within-Host

Bradley R. Jones[1,2], Natalie N. Kinloch[3], Joshua Horacsek[1], Rachel Miller[3], Bemuluyigza Baraki[3], Hanwei Sudderuddin[3], Bruce Ganase[1], Marianne Harris[1], P. Richard Harrigan[2], R. Brad Jones[4], Mark A. Brockman[1,3], Art F. Poon[5], Jeffrey B. Joy[1,2], Zabrina L. Brumme[1,3]

1. BC Centre for Excellence in HIV/AIDS, 2. University of British Columbia, 3. Simon Fraser University, 4. The George Washington University, Washington, 5. University of Western Ontario

Poster BSP4.03
CAHR 2018
Bradley R. Jones
bjones@cfenet.ubc.ca

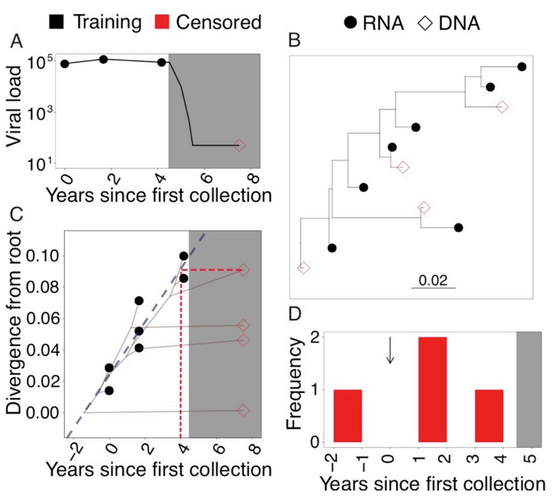
Background

- HIV evolution and latent reservoir establishment occur continually within-host, and latently HIV-infected cells persist long-term
- Thus, the within-host latent HIV reservoir should comprise a genetically heterogeneous archive recapitulating HIV's within-host evolutionary history.
- To demonstrate this, we developed a phylogenetic method to recover individual latent HIV lineage integration dates within-host.
- We apply this method to characterize the age and diversity of latent reservoir sequences (including proviral and *in vivo* spontaneous reactivated HIV sequences) in four individuals followed over 14+ years, including 10+ years on cART.

Methods

Data. For four participants, we collected:
- Longitudinal plasma HIV RNA Nef sequences from pre-ART timepoints ("training data"; used to calibrate host-specific HIV evolutionary rates)
- HIV DNA Nef sequences from PBMC collected while on long-term cART, as well as HIV RNA Nef sequences retrieved from spontaneous low-level viremia events on cART (representing putative release of HIV from the reservoir *in vivo)* ("censored data"; used for molecular dating)
- HIV sequencing was performed using single genome amplification (SGA)

Analysis.
- Inferred a distribution of 100 bootstrap maximum likelihood phylogenetic trees for each individual.
- Trees were rooted using root-to-tip regression using only the pre-ART plasma HIV RNA sequences.
- Host-specific linear models that related phylogenetic divergence from the root to pre-ART HIV RNA sampling times, were fit to each tree
- Integration dates of "censored" (putative reservoir) sequences were inferred by applying the linear model to their phylogenetic divergences from the root.
- We extensively validated our method on simulated, published and empirically collected HIV sequence data (results not shown).



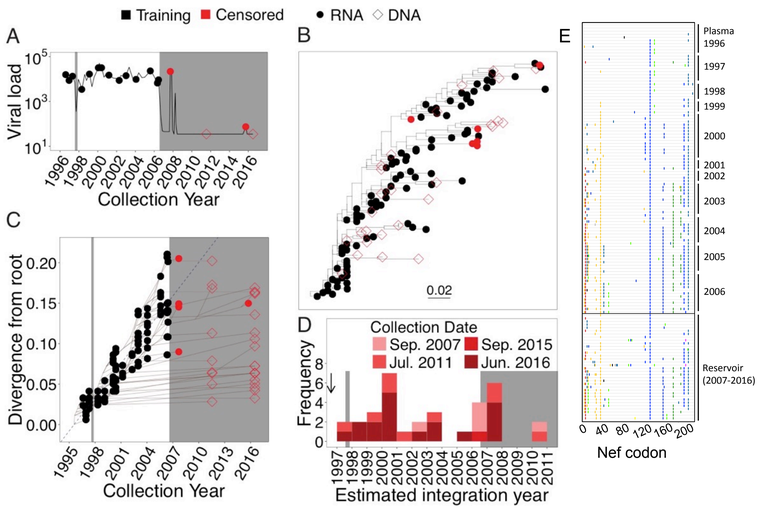
Method illustration.
A) We use SGA to obtain longitudinal pre-ART plasma HIV RNA sequences (used for model training) as well as latent HIV reservoir sequences during long-term cART (grey shaded area).
B) We infer a phylogenetic tree relating HIV RNA and DNA sequences.
C) We use the RNA sequences to fit a linear model (blue dotted line) that relates divergence from the root to time. This model is then used to recover integration dates of reservoir sequences (red dotted lines).
D) Distribution of estimated integration dates.

Conclusion

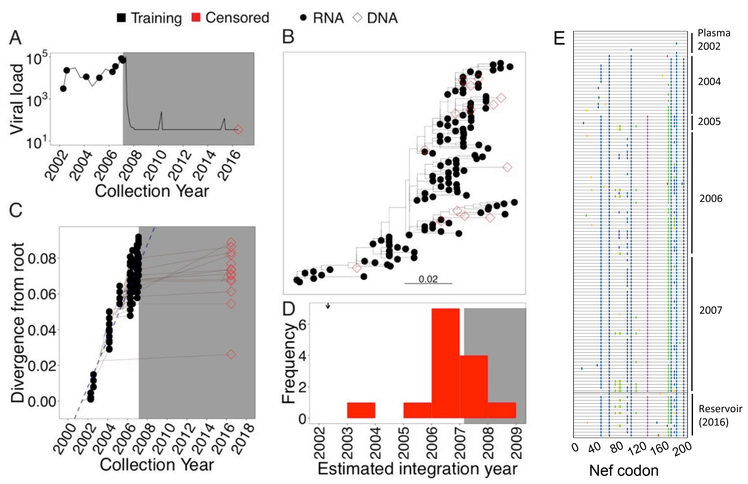
When applied to 4 HIV-infected individuals followed over 14+ years including 10+ years on cART, our method reveals a temporally and genetically diverse reservoir. Our results are consistent with persistence of reactivation-competent latent HIV lineages for >20 years. Our method for estimation of latent HIV ages may illuminate a variety of fundamental questions in HIV persistence.

Results

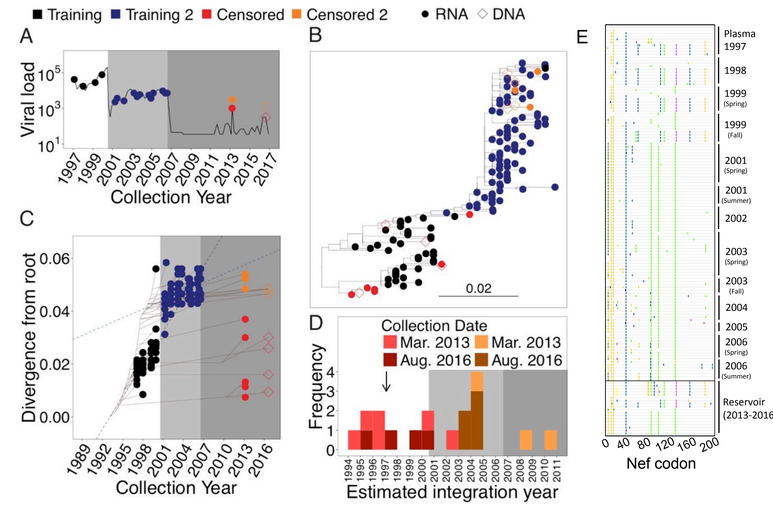
Overall, the linear models fit well to the participant data, substitution rate estimates ranging from 1.1×10^{-5} to 4.8×10^{-5} substitutions/base/day. All four participants displayed chronologically and genetically diverse latent HIV reservoirs. Some reservoir sequences dated prior to the first sampling time, up to 20 years before they were collected.



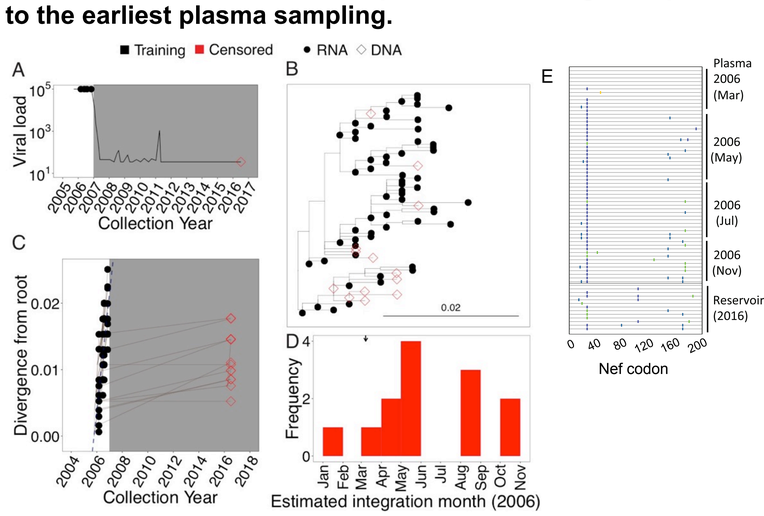
Participant 1.
A) pVL history and collection times. Grey shading indicates cART.
B) Phylogenetic tree relating HIV RNA sequences to reservoir sequences, rooted using root-to-tip regression.
C) Divergence versus time plots with linear regression (dotted blue line). Ancestral traces are shown in brown.
D) Histogram of estimated integration dates. The arrow indicates time that first HIV RNA sample was taken.
E) Highlighter plots of codon substitutions in Nef.
We collected 93 unique plasma HIV RNA sequences (black dots) over 14 pre-cART time points spanning a 10 year period, and 33 unique HIV sequences (red symbols) from proviral DNA and plasma RNA from recrudescent viremia episodes on therapy. **The ladder-like structure of the phylogenetic tree is typical of within-host HIV evolution. Reservoir sequences sampled after long-term cART were integrated throughout the tree and recapitulate HIV's within-host evolutionary history as evident in the highlighter plot in panel E.**



Participant 3.
We collected 113 unique plasma HIV RNA sequences (black dots) over 8 pre-cART time points spanning a 5 year period, and 14 HIV sequences (red diamonds) from proviral DNA on cART. **Note how most of the reservoir sequences in this participant date to just prior to cART initiation. Nevertheless, genetic diversity in the reservoir is relatively high, with one sequence estimated to have integrated 14 years prior to sampling.**



Participant 2.
We collected 39 unique plasma HIV RNA sequences (black dots) over 4 pre-ART time points spanning 3 years, 80 plasma HIV RNA sequences (blue dots) over 12 time points spanning 5 years during dual therapy, and 18 HIV sequences (red and orange symbols) from proviral DNA and plasma RNA from spontaneous viral rebounds while on cART. Because of this participant's period of incompletely-suppressive dual therapy, we used two linear regressions to estimate the reservoir integration dates. The blue training sequences were used to fit the second regression (shown as a dotted lighter blue line in Figure 3C). The dates of the orange reservoir sequences were estimated from this regression (chosen based on their position in the tree). **Note the two clades of HIV circulating in plasma during the pre-ART period, one of which disappears from circulation after the initiation of dual ART. However the clade that disappeared from circulation is clearly preserved in the latent reservoir. There is also a clade of latent HIV sequences that integrated prior to the earliest plasma sampling.**



Participant 4.
We collected 59 unique plasma HIV RNA sequences (black dots) over 4 pre-cART time points spanning an 8 month period, and 13 HIV sequences (red diamonds) from proviral DNA on cART. Participant 4 differs from the others in that he initiated cART comparatively earlier in infection. **Nevertheless, the reservoir is genetically diverse, with individual latent HIV sequences dating to the time between infection and cART initiation.**