Heterogeneous age and genetic distribution of within-host HIV-1 reservoirs revealed by a novel phylogenetic dating strategy

Bradley R. Jones[1], Natalie N. Kinoch[2], Joshua Horascek[1], Philip Mwimanzi[2], Bemuluyigza Baraki[2], John Huang[3], Ronald Truong[3], Bruce Ganase[1], Marianne Harris[1], P. Richard Harrigan[1,4], R. Brad Jones[3], Mark A. Brockman[1,2], Jeffrey B. Joy[1,4], Art F. Y. Poon[5], Zabrina L. Brumme[2]

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

Background

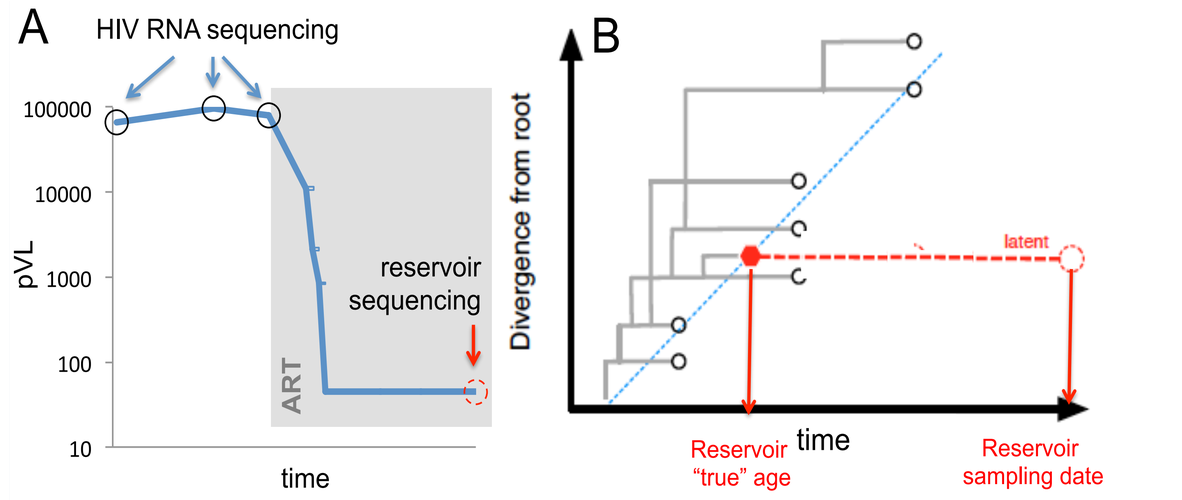
Untreated HIV infection is characterized by rapid and continual seeding of the reservoir and ongoing within-host viral evolution. Thus, by chronic infection, the HIV reservoir should constitute a heterogeneous population in terms of age, genetic divergence from the founder virus and immune escape mutation burden. If so, simultaneous reactivation of such a heterogenous viral pool, for example using latency reversal agents, could conceivably complicate immunotherapeutic cure approaches to reservoir elimination.

We apply a novel phylogenetic framework to characterize the age and diversity the HIV reservoir (including proviral, *in vivo* spontaneous reactivated and *in vitro* reactivated HIV sequences) in two individuals followed over 20+ years, including 10+ years on cART.

Methods

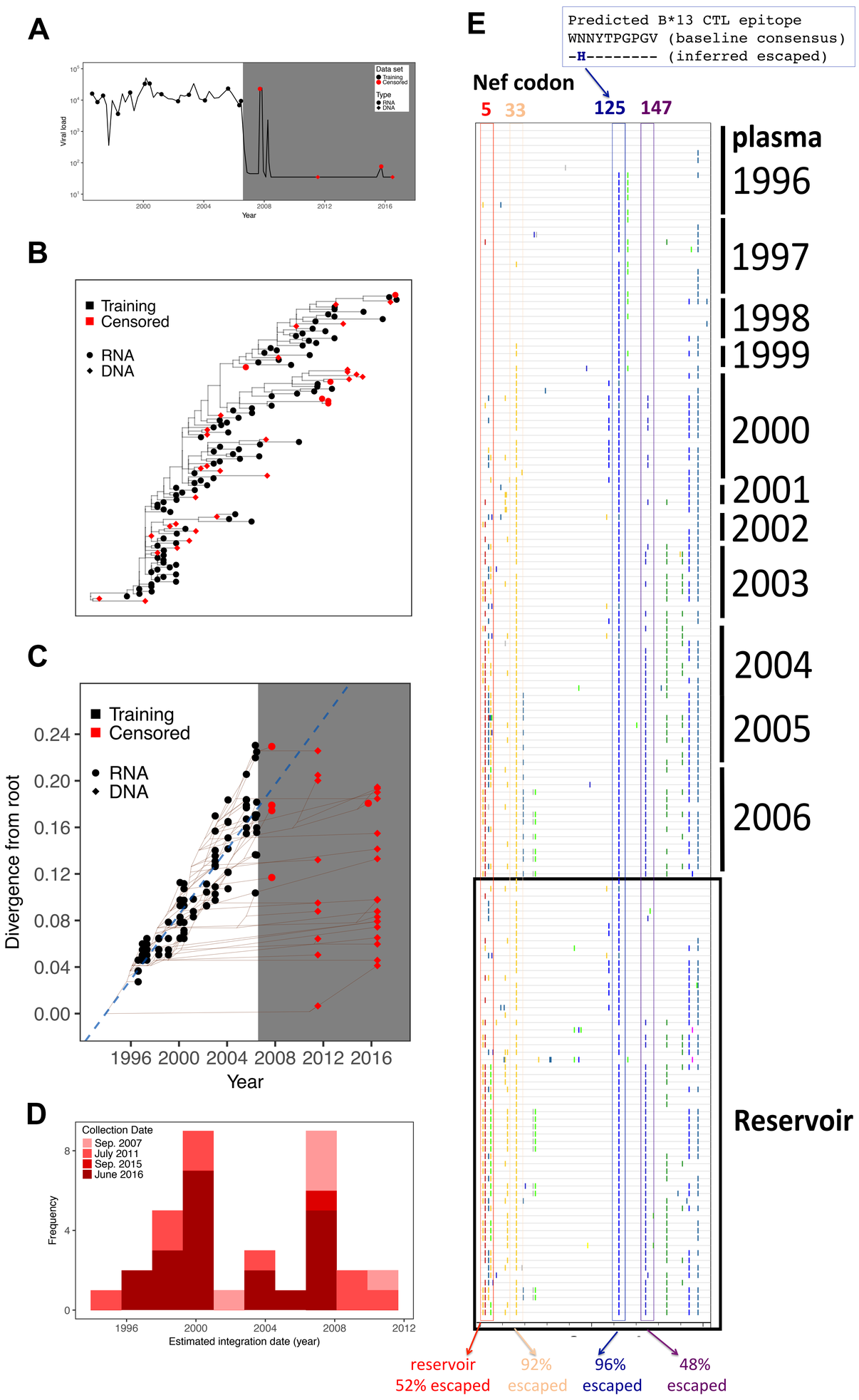
Host-specific molecular clocks for the two study patients were calibrated using HIV Nef sequences obtained via single genome amplification from longitudinal pre-cART plasma samples spanning ~10 years. These clocks were used to "date" proviral HIV DNA sequences isolated directly from, or following *in vitro* reactivation of, PBMCs collected up to 10 years following cART initiation; as well as HIV RNA from spontaneous in *vivo* viremia blips during cART (="*in vivo* spontaneous reactivated"). Sequences were HXB2-aligned with MUSCLE and trimmed manual)ly with AliView. Maximum likelihood phylogenetic trees were inferred from the sequences with RAxML and outgroup rooted using HXB2.

Approach

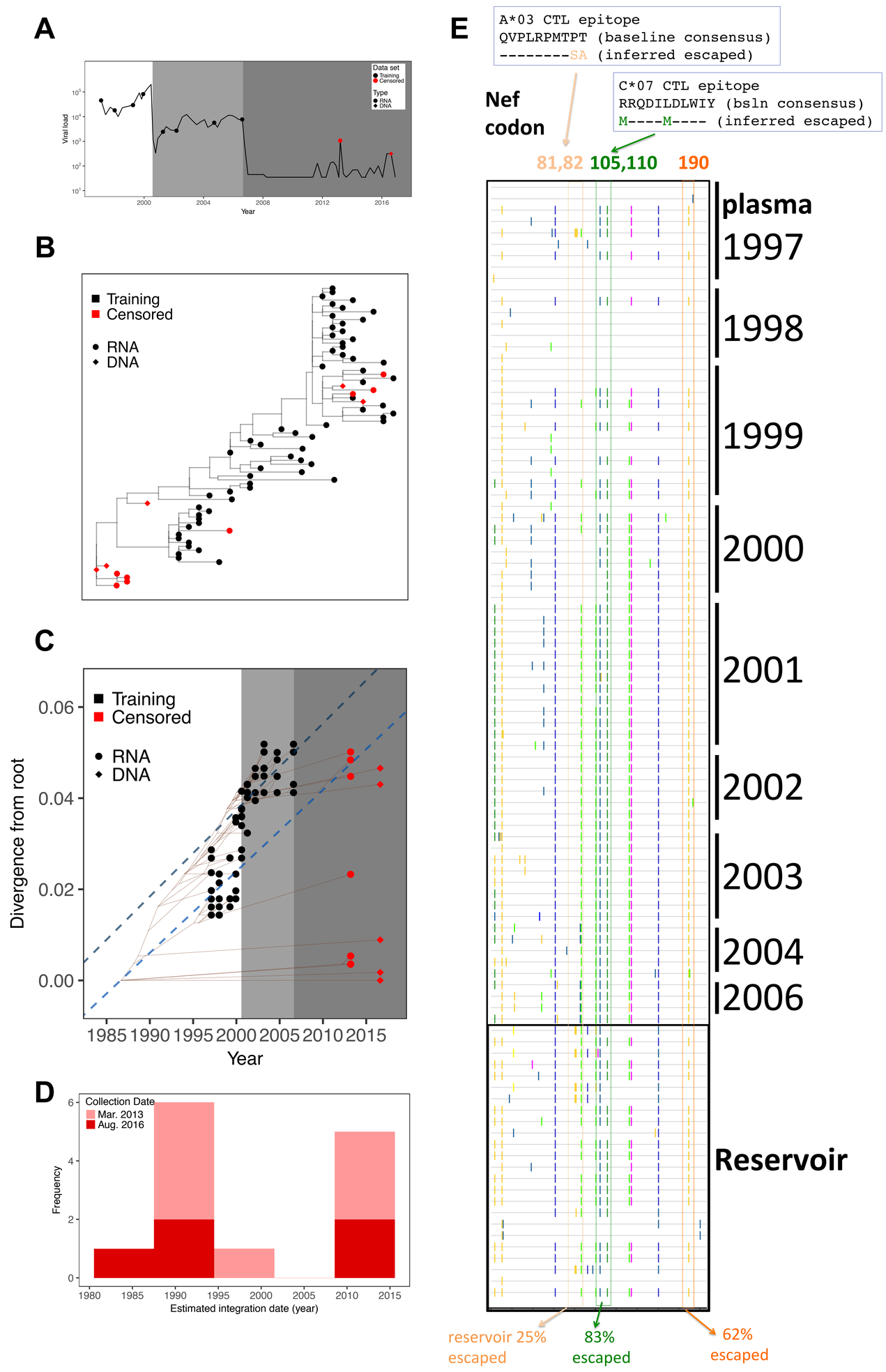


Longitudinal, single-template amplified, within-host HIV sequences sampled pre-cART are used to fit a linear model relating phylogenetic distance from the inferred root (founder virus) to time (phylogenies rooted using the HIV-1 HXB2 reference sequence). This host-specific rate of HIV evolution (denoted by the dashed diagonal in **B**) is used to infer the "true age" (i.e. integration date) of individual HIV reservoirs from their sequence.

Patient 1: A*26:01/30:01, B*13:02/14:01, C*06:02/08:02



Patient 2: A*03:01/29:02, B*44:03/49:01, C*07:01/16:01



Patient 1 reservoir dynamics/diversity. A. Clinical history and sampling timeline. Grey shading denotes cART. Each point represents a sampling timepoint, and whether isolated HIV sequences were used to calibrate the molecular clock ("training"; black) or were retrieved from putative reservoirs ("censored"; red). **B.** Maximum likelihood tree inferred from the sequences, rooted using HXB2 (not shown), coloured as in A. *Note the reservoir sequences intersperse throughout the phylogeny.* **C.** Linear model relating within-host genetic divergence from the root versus collection time. Dashed blue diagonal is the linear regression (shaded area denotes cART). Grey lines indicate the phylogenetic relationship of the sequences. **D.** Distribution of reservoir integration dates. **E.** Highlighter plot depicting longitudinal within-host plasma HIV and reservoir diversity, with example putative escape sites highlighted. *Note the reservoir recapitulates plasma HIV RNA evolution.*

Patient 2 reservoir dynamics/diversity. A. Clinical history and sampling timeline. Light grey shading denotes dual ART; darker grey shading denotes cART. Sampling timepoints colour coded as for patient 1. **B.** Maximum likelihood tree inferred from patient 2's HIV sequences, rooted using HXB2 (root not shown), coloured as in A. *Note that clade of reservoir sequences that branch closer to the root than the earliest plasma samples, suggesting that they pre-date earliest plasma sampling.* **C.** Linear models relating within-host genetic divergence from the root versus collection time (one each for pre-ART [light blue] and dual-ART [dark blue] periods). **D.** Distribution of reservoir integration dates. **E.** Highlighter plot depicting longitudinal within-host plasma HIV and reservoir diversity, with example putative escape sites highlighted. *Note that again, the reservoir reflects longitudinal within-host plasma HIV RNA evolution and includes sequence heterogeneity suggestive of immune escape within published HLA-matched CTL epitopes.*

Summary and Implications

Analyses of within-host HIV dynamics over a 20-year timeframe reveal a genetically complex and heterogeneous reservoir that recapitulates HIV's within-host evolutionary history. Results further suggest that the reactivation-competent portion of the reservoir may fully reflect this diversity. Specifically, HIV evolution in both patients featured genetic bottlenecks characteristic of host-driven selection, including in published/predicted HLA-restricted CTL epitopes, with the oldest reservoirs, including reactivated sequences, exhibiting ancestral (pre-bottleneck) forms. Simultaneous reactivation of such a heterogenous viral pool, for example using latency reversal agents, could conceivably produce an HIV sequence pool whose diversity vastly exceeds that ever "seen" by the immune response at any given point during infection. This could complicate immunotherapeutic strategies for reservoir elimination. In particular, the co-existence of ancestral (*i.e.* presumed immunologically susceptible) and adapted (*i.e.* presumed escaped) forms of the same HLA-restricted CD8+ CTL epitope within the reservoir may have important implications for immunotherapeutic approaches aimed at reservoir elimination. This finding highlights the need to consider immunotherapeutic approaches focused exclusively on conserved, immunologically susceptible regions and/or approaches aimed at targeting a range of viral epitopes and their within-host variants. Results also support the potential utility of HIV reservoir genetic characterization approaches to inform the design of 'personalized' immunogens.

Project funded by NIAID award AI126617 (co-funded by NIDA, NIMH, NINDS), the Canadian HIV Cure Enterprise Team Grant HIG-133050 (co-funded by CIHR, CANFAR, IAS), and CIHR project grant PJT-376711 (funded by CIHR)