RECONSTRUCTING INTEGRATION DATES OF LATENT HIV SEQUENCES WITHIN-HOST

Bradley R. Jones[1], Natalie Kinloch[2], Joshua Horacsek[2], Bruce Ganase[1], Marianne Harris[1], P. Richard Harrigan[3], R. Brad Jones[4], Mark Brockman[1,2], Jeffrey Joy[1,4], Art Poon[5], Zabrina Brumme[1,2]

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

Abstract #372
CROI 2018, Boston, MA
March 4-7, 2018
Presenter: Bradley R. Jones
Email: bjones@cfenet.ubc.ca

Background

Given that HIV evolution and latent reservoir establishment occur continually within-host, and that latently HIV-infected cells can persist long-term, the HIV reservoir should comprise a genetically heterogeneous archive recapitulating within-host HIV evolutionary history. This has yet to be conclusively demonstrated however, in part due to the challenges of reconstructing within-host reservoir dynamics.

Towards this goal, we developed a phylogenetic framework to reconstruct the integration dates of individual latent HIV lineages within-host. Here, we apply the framework to characterize the age and diversity of latent reservoir sequences (including proviral and *in vivo* spontaneous reactivated HIV sequences) in two individuals followed over 20+ years, including 10+ years on cART.

Methods

Our method begins by inferring a phylogenetic tree relating longitudinally sampled HIV RNA sequences from plasma collected during active infection and HIV sequences collected from the latent reservoir. The tree is then optimally rooted using root-to-tip regression. Next we fit a linear model relating phylogenetic distances to time using the HIV RNA sequences from plasma collected during active infection. Finally, we use this linear model to reconstruct the integration dates of the HIV sequences collected from the latent reservoir using their phylogenetic distance from the root. The method was validated using simulated and published data sets. We applied the method to reconstruct latent HIV ages in two participants.

Results

The method (Figure 1) performed very well during validation on both simulated and published datasets (not shown) and on datasets from both participants (Figures 2, 3, 4, 5). **Importantly, for both participants the estimated integration dates of their latent reservoirs were interspersed throughout the period of active infection and some sequences dated to before the first sampling time, up to 20 years before they were collected.**

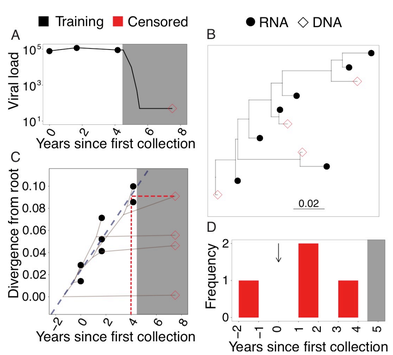

Figure 1: Method illustration. A) We begin with longitudinally sampled HIV RNA sequences from plasma and proviral DNA sequences from the latent reservoir. All sequences are isolated using single-genome amplification. B) We build a phylogenetic tree relating HIV RNA and DNA sequences. C) Using the RNA sequences to fit a linear model (blue dotted line), we relate divergence from the root to time and finally reconstruct the integration dates of the reservoir sequences using the linear model (red dotted lines). D) The distribution of estimated integration dates.

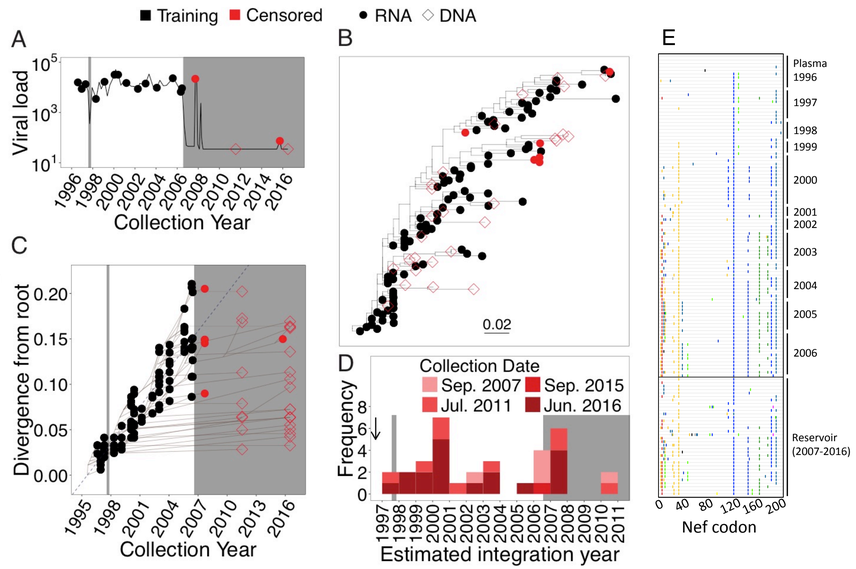

Figure 2: Participant 1. A) Clinical and sampling history. We collected 93 unique HIV RNA sequences (black dots) from plasma over 14 time points prior to suppressive cART spanning 10 years and 33 HIV sequences (red dots) from proviral DNA and plasma RNA from recrudescent viremia episodes on therapy. The gray shaded region indicates cART. B) The phylogenetic tree of participant 1's sequences. ***Note the interspersion of censored sequences (putative reservoir sequences) throughout the tree***. C) The linear model relating genetic divergence to time (shown as a dotted blue line). D) The distribution of estimated integration dates of the putative reservoir sequences. The downward arrow indicates the initial sampling date. ***Note the wide range of dates, and that there was a sequence collected in 2016 that dates to 1997, almost 20 years prior***. E) Highlighter plot depicting longitudinal within-host plasma HIV and reservoir diversity. ***Note how the reservoir sequences recapitulate the evolutionary history of the plasma HIV RNA***.

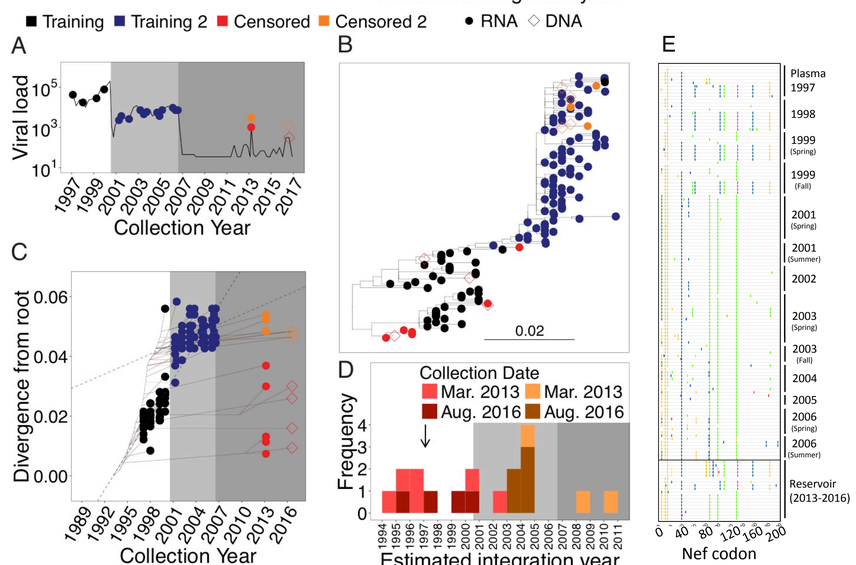

Figure 3: Participant 2. A) Clinical and sampling history. We collected 39 unique HIV RNA sequences (black dots) from plasma over 4 time points prior to therapy spanning 3 years, 80 HIV RNA sequences (blue dots) over 12 time points spanning 5 years during dual therapy, and 18 HIV sequences (red and orange dots) from proviral DNA and plasma RNA from spontaneous viral rebounds while on cART. Lighter gray shading indicates dual ART and darker gray shading indicates cART. B) The phylogenetic tree of participant 2's sequences. ***Note the clade of reservoir sequences close to the root of the phylogeny***. C) Because of the influence of therapy on within-host dynamics we applied two separate linear models, one for the pre-therapy period (dotted blue line) and the other for the dual therapy period (lighter dotted blue line). The model to estimate the integration date of each reservoir sequence depended on its location in the tree (red, first linear model; orange, second linear model). D) The distribution of estimated integration dates of the putative reservoir sequences. The downward arrow indicates the initial sampling date. ***Note that some of the reservoir sequences date to before the first sampling***. E) Highlighter plot depicting longitudinal within-host plasma HIV and reservoir diversity. ***Note the lineage that disappears in plasma following dual ART, but stayed archived in the reservoir 20 years later***.

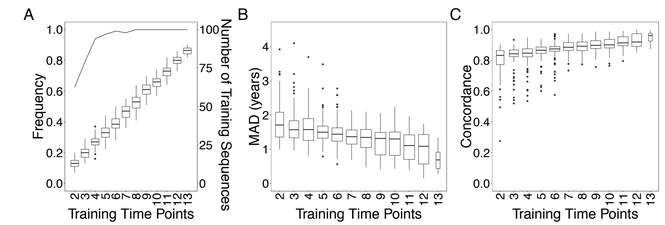

Figure 4: Method is robust to limited sampling. We subsampled the training (HIV RNA) sequences from participant 1 (1096 times) and reconstructed the integration dates of the reservoir sequences using each of these subsampled data sets. A) The proportion of subsampled data sets which have adequately fitting linear models (ΔAIC > 10 and estimated root date < first sampling date) for each number of total training time points (shown as a line). The number of training sequences in the data set (shown as boxplots). ***Note that when more than 8 training time points are used, all the models result in an adequate fit; however even when using 3 training time points 75% of the models have an adequate fit.*** B) The median absolute difference (MAD) between the estimated integration dates estimated from each subsampled data set with a fitting linear model and those derived from the full data set (shown in Figure 2). C) The concordance between the estimated integration dates between each subsampled data set with a fitting linear model and the full data set shown in Figure 2.

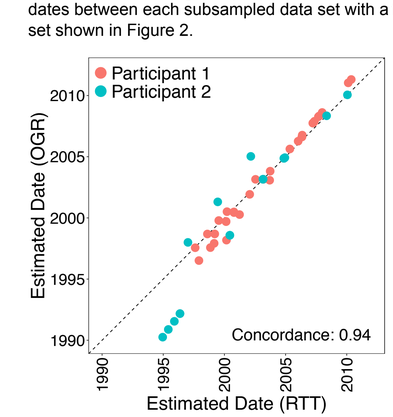

Figure 5: Method is robust to rooting method. Our primary rooting strategy was root-to-tip regression (RTT), which optimizes the fit of the linear model. Another commonly used method is outgroup rooting (OGR), which roots the tree according to the intersection of the sequences to one or more outgroup sequences. We compared inferred reservoir integration dates using RTT to those derived using OGR (using HXB2 as the outgroup sequence). The estimated integration dates using the two methods were highly concordant, with Lin's concordance coefficients of 0.99 and 0.90 respectively for participants 1 and 2 and 0.94 overall.

Conclusions

When applied to HIV-infected individuals followed over 20+ years including 10+ years on cART, **our method reveals a diverse reservoir in terms of age and genetics. Our results are consistent with persistence of reactivation-competent latent HIV lineages for at least 20 years.** Sensitivity analysis shows that our model is robust to rooting method and accurate results can obtained with as few as 3 training time points. Our method for estimation of latent HIV ages may help shed light on a variety of fundamental questions in HIV persistence.

CROI Conference on Retroviruses and Opportunistic Infections
CIHR IRSC Canadian Institutes of Health Research / Instituts de recherche en santé du Canada
National Institutes of Health
BELIEVE Martin Delaney Collaboratory
CanCURE Together, we Can CURE HIV
The George Washington University Washington, DC
Western
SFU
British Columbia Centre for Excellence in HIV/AIDS
UBC
Providence Health Care How you want to be treated.