Bayroot: A Bayesian Phylogenetic Approach to Dating HIV Reservoir Sequences

Roux-Cil Ferreira[1] and Art F.Y. Poon[1]

[1]Department of Pathology and Laboratory Medicine, Western University, London, ON, Canada.

Western
UNIVERSITY · CANADA

Introduction

What is the latent viral reservoir?

- It is HIV-1 DNA (provirus) integrated into resting CD4+ T cells, *i.e.* a long-lived population of infected cells.

Why study the integration dates (provirus age) of proviruses in the reservoir?

- Homogeneous ages may indicate when during infection (early or late) the reservoir was established.
- Provirus age may influence susceptibility to immune-mediated or therapeutic eradication strategies.

Why change to Bayesian root-top-tip regression (RTTR) when dating HIV-1 reservoir sequences?

- Non-Bayesian RTTR makes limiting assumptions:
  - All sequence with age T carry exactly Y mutations from the root.
  - The phylogenetic tree is known without error.
  - The mutation rate continues to increase after treatment initiation, *i.e.* sequences can date to post-ART.
  - Sequences can also be assigned dates after their sampling date, *i.e.* map to future dates.
- We developed a Bayesian RTTR approach to address these problems.

Western
UNIVERSITY · CANADA

Method

- This study is based on dating 427 HIV-1 DNA sequences sampled from the reservoirs of 13 seroconvertes in the Zambia-Emory HIV Research Project.

- After screening for hypermutation and reconstructing phylogenies with RAxML the sequences were dated using non-Bayesian RRTR and our “new” Bayesian RTTR (bayroot).

- We compared estimated reservoir sequence ages between the two (RTTR and Bayesian RTTR) methods.

Western
UNIVERSITY · CANADA

Results

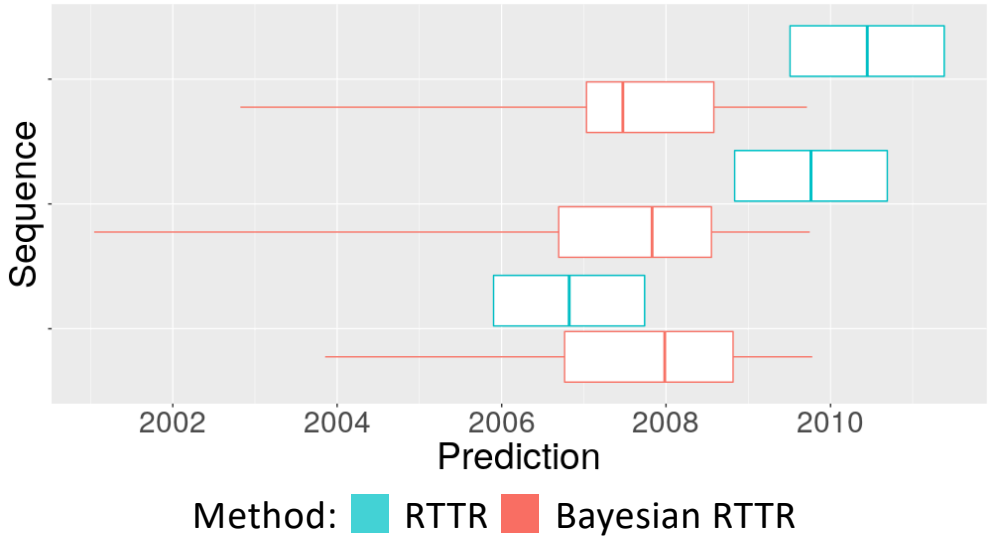

Figure 1: Example of three HIV-1 DNA sequence ages. Ages were estimated using RTTR (blue) and Bayesian RTTR (salmon). Confidence intervals (RTT regression, blue) or highest density intervals (bayroot, salmon) are shown depending on the method.

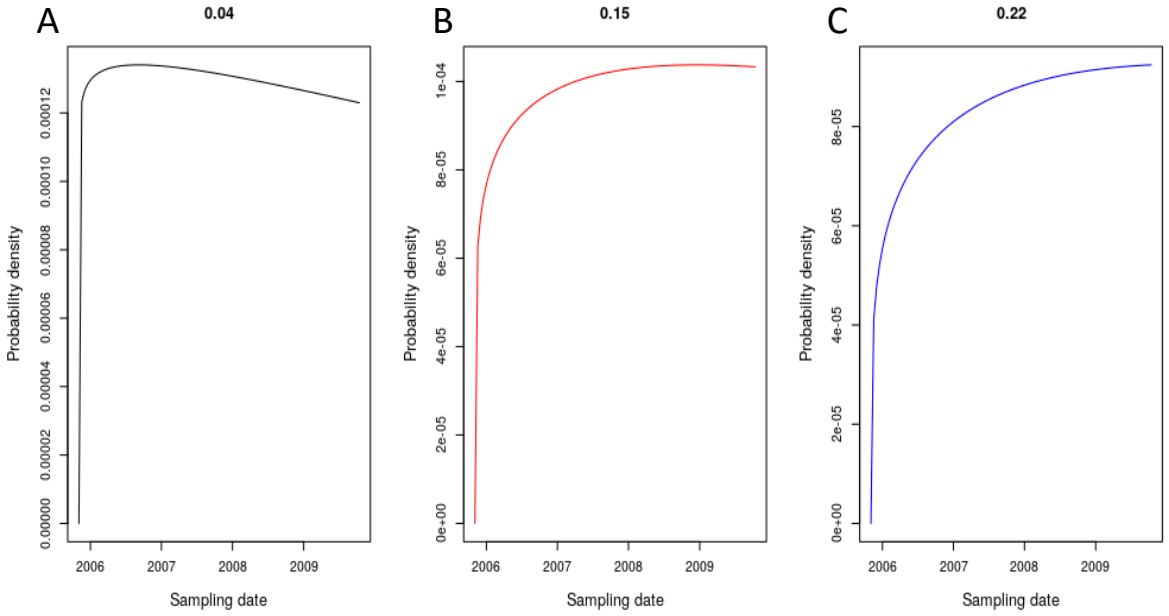

Figure 2: Examples of the "flat" posterior probability density of integration dates, $P(t|y)$, over all possible integration dates (infection - ART) for a sequence with (A) 0.04, (B) 0.15, and (C) 0.22 divergence respectively.

Western
UNIVERSITY · CANADA

Conclusion

- The **range of ages** supported by the data **were greater when the Bayesian RTTR method was used** compared to those supported by standard RTTR (Figure 1).
- Moreover, there is **little variation in the age ranges supported** for different HIV-1 DNA sequences when using Bayesian RTTR.
- We suspect that the wider range and less variation of ages supported by the Bayesian RTTR analysis is because the **posterior probability distribution of integration date is relatively flat** over the wider range of dates (Figure 2).
- Furthermore, the flat property of the posterior probability distribution of integration dates is present for various sequence divergences (Figure 2).
- By ignoring uncertainty in the location of the root and variation in the number of mutations **standard RTTR incorrectly narrows the range of probable HIV-1 DNA ages**.
- In summary, **it is difficult to estimate the ages of provirus sequences using sparse sequence data**, which may explain why different studies using different methods observe heterogeneous integration times.
- Further study using simulated data, where integration dates are known, may reveal how effectively sequence data can be use to estimate integration dates

Western
UNIVERSITY · CANADA