CONFERENCE CAHR 2022

CONGRÈS DE L'ACRV 2022

Computational advances in molecular dating of within-host HIV systems

Bradley R. Jones[1,2], Kelsie Brooks[3], Eric Hunter[4], Zabrina L. Brumme[1,5], Jeffrey B. Joy[1,2,6]

[1]BC Centre for Excellence in HIV/AIDS, [2]Bioinformatics Program, University of British Columbia, [3]National Institute of Allergy and Infectious Diseases, National Institutes of Health, [4]Department of Pathology, Emory University, [5]Faculty of Health Sciences, Simon Fraser University, [6]Department of Medicine, University of British Columbia

Poster #238

Correspondence: bjones@bccfe.ca

Special thanks to the study participants without whom this research would not be possible

BRITISH COLUMBIA CENTRE *for* EXCELLENCE *in* HIV/AIDS

NIH National Institutes of Health

CIHR IRSC Canadian Institutes of Health Research / Instituts de recherche en santé du Canada

EMORY UNIVERSITY

SFU

UBC

Providence HEALTH CARE How you want to be treated.

BRITISH COLUMBIA Ministry of Health

BBD: BEAST2 package for extended date estimation

Phylogenetic dating is an important tool for estimating unknown tip dates and the time of most recent common ancestors. This enables us to estimate **integration dates of proviral sequences into the persistent reservoir** and **infection dates of persons living with HIV.**

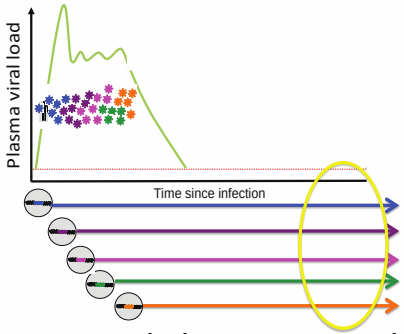


Figure 2.1: HIV replicates and evolves during active infection. Meanwhile, HIV proviruses are integrated into the persistent reservoir where evolution ceases. **By comparing the evolution of HIV sequences from active infection we can estimate the integration dates of the reservoir. Understanding HIV reservoir dynamics aids HIV cure research.**

Current phylogenetic methods of proviral integration dating either only recover dates sampled during active infection[1] or rely on an unrealistic strict molecular clock[2]. **The Bayesian method of tip date sampling has neither issue.**

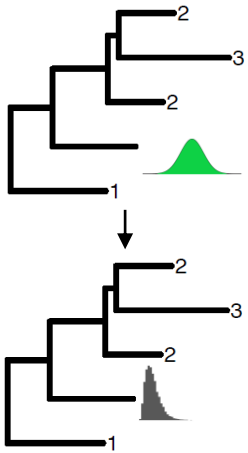


Figure 2.2: Diagram of tip date sampling. Sampling dates from a **prior distribution** subject to the tree likelihood gives a sampling of the **posterior distribution** of dates. Numbers represent collection dates of active sequences.

[1]Abrahams et al (2020) *Sci Transl Med*. doi:10.1126/scitranslmed.aaw5589.
[2]Jones et al (2018) *Proc Natl Acad Sci U S A.* doi:10.1073/pnas.1802028115.
[3]Bouckaert (2019) *PLoS Comput Biol*. doi:10.1371/journal.pcbi.1006650.

BBD is our new Java package for the software BEAST2[3] that includes additional priors for date estimation and tip date operators allowing for better date estimation.

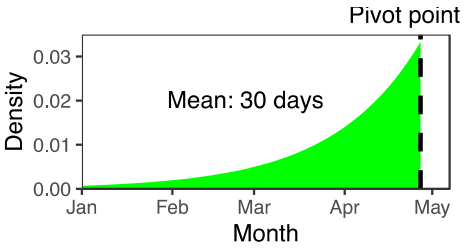


Figure 2.3: Reverse exponential distribution. **This distribution, included in BBD, allows us to use priors on dates acting backwards in time.**

BBD also includes an operator to re-root a tree without changing its topology.

In this presentation, we show two examples of using the BBD package:

- Estimating HIV integration dates of the persistent reservoir using tip date sampling (slides 3–4)
- Estimating HIV infection dates applying time of HIV testing (slide 5)

Estimating HIV proviral integration dates with BBD

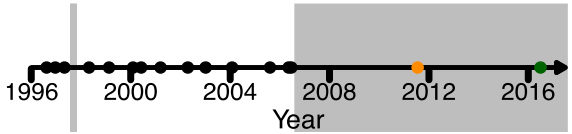


Figure 3.1 (left): Sampling scheme of Participant 1 from Jones et al[1]. Black points show plasma sampling and coloured points show latent sampling. Grey shading indicates periods of combined antiretroviral therapy (cART).

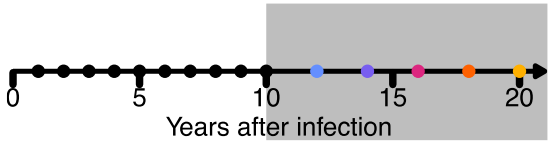


Figure 3.3 (above): Sampling scheme of simulation[2,3]. Black points show plasma sampling and coloured points show latent sampling. Grey shading indicates periods of cART.

Figure 3.2 (right): Maximum clade credibility tree of Participant 1 from Jones et al[1]. Closed black circles show plasma sequences. Open coloured circles show mean estimated proviral integration dates with coloured bars showing 95% highest posterior density interval. **Estimated dates are dispersed throughout time including two sequences whose estimates precede plasma sampling (circled in red).**

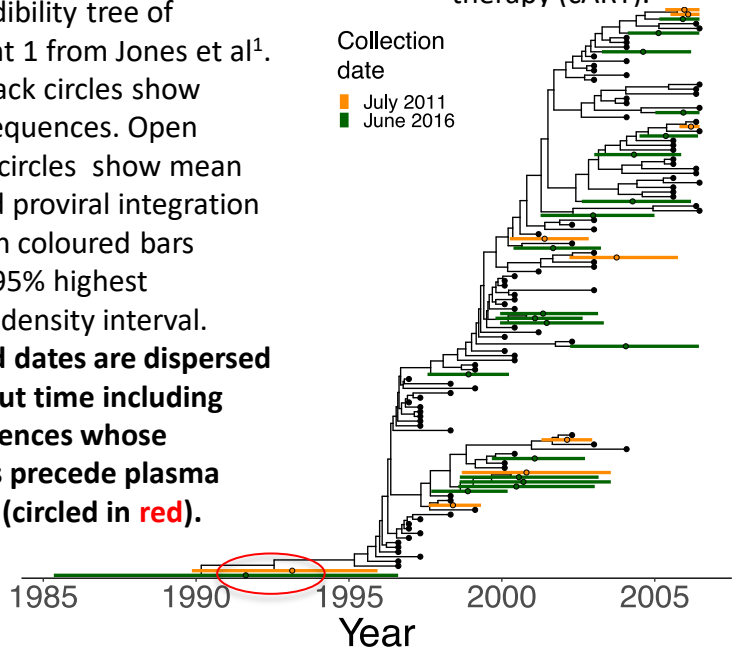


Figure 3.4 (right): Maximum clade credibility tree of simulated data. Closed black circles show plasma sequences. Open coloured circles show mean estimated proviral integration dates with coloured bars showing 95% highest posterior density (HPD) interval. Coloured crosses show actual integration dates. **All but two (circled in red) actual integration dates fell within the 95% HPD interval.**

[1]Jones et al (2018) *Proc Natl Acad Sci U S A.* doi:10.1073/pnas.1802028115.
[2]Jariani et al (2020) *Virus Evol*. doi:10.1093/ve/vez003.
[3]Jones et al (2020) *Virus Evol*. doi:10.1093/ve/veaa089.

BBD outperforms current date estimation methods on simulated data

Table 4.1: Integration date priors.

Name	Prior	Variable	Parameters
Unif1	Uniform	Date	Bounds given by plasma sampling range
Unif2	Uniform	Date	Lower bound: earliest plasma sampling date minus plasma sampling range Upper bound: sampling date
Norm	Normal	Date	Mean: plasma sampling midpoint Sigma: ¼ plasma sampling range
Lnorm1	Reverse Log normal	Years/days before latest plasma sampling date	Real mean: plasma sampling midpoint Sigma: chosen to make real standard deviation ¼ plasma sampling range
Lnorm2	Reverse Log normal	Years/days before sequence sampling date	Real mean: plasma sampling midpoint Sigma: chosen to make real standard deviation ¼ plasma sampling range
Exp1a	Reverse Exponential	Years/days before latest plasma sampling date	Mean: plasma sampling midpoint
Exp1b	Reverse Exponential	Years/days before latest plasma sampling date	Mean: estimated with an exponential prior whose mean is the plasma sampling midpoint
Exp2a	Reverse Exponential	Years/days before sequence sampling date	Mean: plasma sampling midpoint
Exp2b	Reverse Exponential	Years/days before sequence sampling date	Mean: estimated with an exponential prior whose mean is the plasma sampling midpoint

[1]Abrahams et al (2020) *Sci Transl Med*. doi:10.1126/scitranslmed.aaw5589.
[2]Jones et al (2018) *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.1802028115.
[3]To et al (2016) *Syst Biol*. doi:10.1093/sysbio/syv068.

Figure 4.1 (right): Classical multidimensional scaling comparing the root mean squared deviation between the estimated dates between different date estimation methods and priors. Axes units are in years. **The BBD estimates with Exp1a, Exp1b and Unif1 priors clustered with the actual dates. Estimates using alternative methods gave similar results when using different tree building software.**

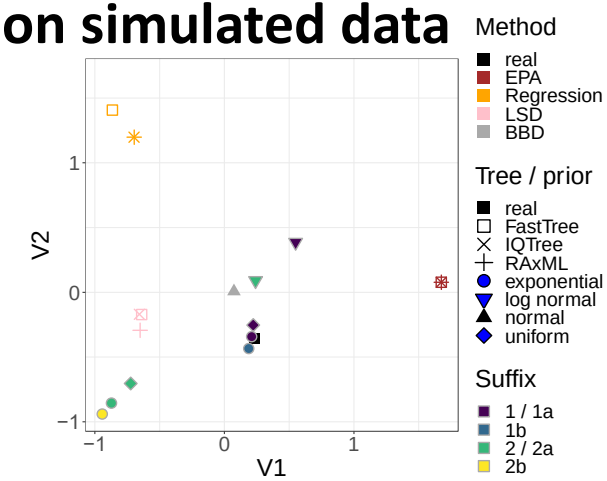


Figure 4.2 (below): Histogram of actual and estimated integration dates. Dashed and dotted lines show the first and last plasma sampling time respectively and the green line shows the best fitting exponential (Exp1a) prior of the dates. **BBD recaptures the actual distribution of integration dates.**

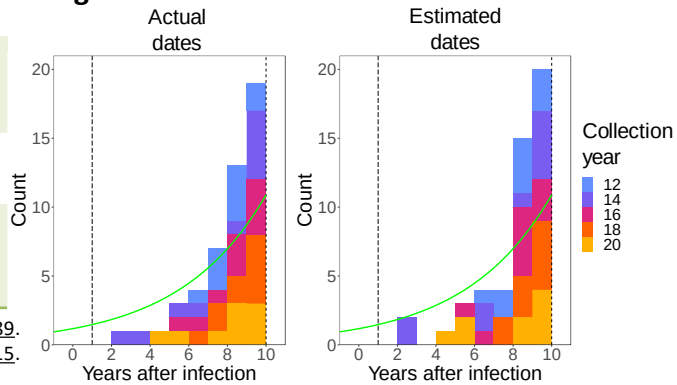


Table 4.2: Comparison of date estimation software. **Our BBD method is more accurate with a lower root mean squared error (RMSE) and higher concordance.**
*EPA values were calculated over sequences where estimation was possible.

Method	RMSE (years)	Concordance
BBD	0.62	0.937
EPA[1]	1.64*	0.453*
Regression[2]	1.89	0.572
LSD[3]	1.23	0.805

Informative priors improve HIV infection date estimates

The time of the most recent common ancestor (tMRCA) can be used as a proxy for infection date.

We used BEAST2 to estimate the tMRCA of 11 participants with HIV *env* sequences from plasma at diagnosis, 1 year after diagnosis and cART initiation[1]. We estimated the tMRCA without a prior and with one of 4 priors informed by their last negative test and first positive test.

Prior	Parameters	BBD?
Normal	Mean: HIV test date mid point Sigma: ¼ HIV test date range	No
Uniform	Lower: Last negative test date Upper: First positive test date	No
Reverse Exponential	Mean: HIV test date mid point	Yes
Reverse Log Normal	Real mean: HIV test date mid point Sigma: $\sqrt{\log 4/5}$	Yes

Table 5.1: Priors on tMRCA (infection date). BBD priors act on the number of days before the first positive test.

Figure 5.1 (right): Density of infection date estimates for participant N133M. Histogram shows the density of the estimated dates using different priors (data restricted to the 95% highest posterior density). Solid grey line shows the mean estimated infection date. Red line shows the prior distribution. Dotted lines show HIV test dates. **Informative priors bring the estimated infection date within the testing range.**

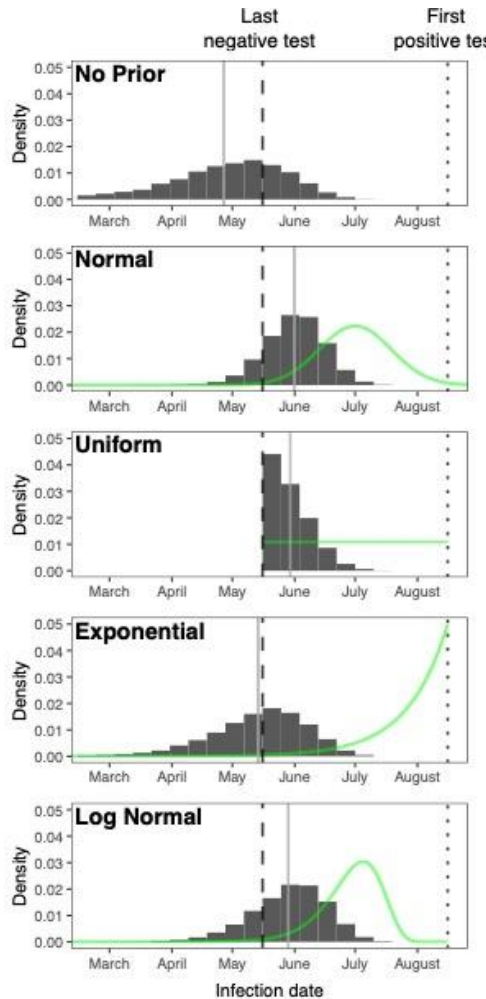
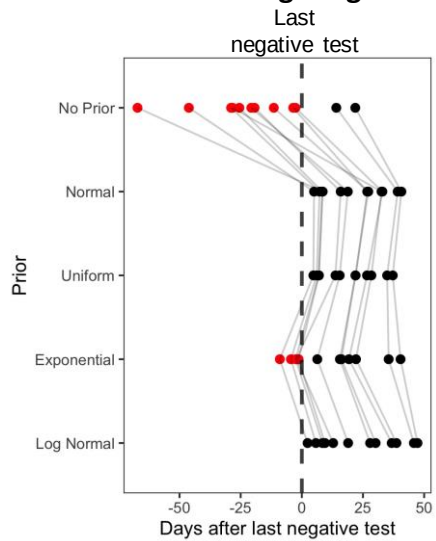


Figure 5.2: Mean estimated infection date from all participants. Points from the same participant are joined. Red points indicate estimates that are earlier than the last negative test. **When an exponential prior or no prior are used the estimated infection date may precede the last negative test date.**



BBD gives more accurate date estimates of the reservoir and infection that will propel our understanding of within host HIV dynamics leading to an HIV cure.

[1]Brooks et al (2020) *PLoS Pathog*. doi:10.1371/journal.ppat.1008378.