Inference of transmission network structure from HIV phylogenetic trees

Federica Giardina¹,²*, Ethan Obie Romero-Severson², Jan Albert³,⁴, Tom Britton¹, Thomas Leitner²

¹ Department of Mathematics, Stockholm University, Stockholm, Sweden
² Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM
³ Department of Microbiology, Tumor and Cell Biology, Karolinska Institute, Stockholm, Sweden
⁴ Department of Clinical Microbiology, Karolinska University Hospital, Stockholm, Sweden

* E-mail: federica@math.su.se

Supplementary information

Probability to escape infection

We calculated the probability to escape infection from a neighbour, \( \pi \), as the exponential of the total infections pressure. Let \( X \) denote the time to diagnosis (or death from AIDS if never diagnosed) and assume that \( X \) follows an exponential distribution with rate parameter \( \gamma \), i.e. \( X \sim \text{Exp}(\gamma) \). In the first model specification (constant infectivity) \( \pi \) can be written as \( \pi = \mathbb{E}[\text{Exp}(-\lambda X)] \) where \( \lambda \) is the constant transmission rate. Therefore, \( \pi = \mathbb{E}[\text{Exp}(-\lambda X)] = \int_0^{\infty} \gamma e^{-\gamma x} (e^{-\lambda x}) dx = \frac{\gamma}{\lambda + \gamma} \).

In the second model specification, which included stage-varying infectivity, let \( t_1 \) denote the deterministic time spent in the first stage (acute phase), \( T_2 \) the random time spent in the second stage (chronic phase), represented by an exponential random variable with rate parameter \( \beta \), i.e. \( T_2 \sim \text{Exp}(\beta) \), and \( t_3 \) the time spent in the pre-AIDS stage. Here, the transmission rates in each of the three infection stages are \( \lambda_1, \lambda_2 \) and \( \lambda_3 \), respectively. Thus, the infectious pressure has a different expression depending on when the diagnosis occurs (in the acute, chronic or pre-AIDS stage). The probability \( \pi \) is the mean of the infectious pressure calculated over \( X \) and \( T_2 \). We have:

\[
\pi = \mathbb{E}[\text{Exp}(-\lambda_1 \mathbb{I}_{X<t_1} - (\lambda_1 t_1 + \lambda_2 (X-t_1)) \mathbb{I}_{t_1<X<t_1+T_2} - (\lambda_1 t_1 + \lambda_2 T_2 + \lambda_3 (X-t_1-T_2)) \mathbb{I}_{X>t_1+T_2}]
\]

\[
= \int_0^{\infty} \gamma e^{-\gamma x} (e^{-\lambda_1 x}) dx + \int_0^{\infty} \int_{t_1}^{t_1+T_2} \beta e^{-\beta t_2} e^{-\gamma x} e^{-(\lambda_1 t_1 + \lambda_2 (X-t_1-T_2))} dt_2 dx + \int_0^{\infty} \int_{t_1+T_2}^{\infty} \beta e^{-\beta t_2} e^{-\gamma x} e^{-(\lambda_1 t_1 + \lambda_2 t_2 + \lambda_3 (X-t_1-T_2))} dt_2 dx
\]

\[
= -\frac{\beta}{\beta + \gamma + \lambda_1} e^{-t_1(\gamma + \lambda_1)} - \frac{\beta}{\beta + \gamma + \lambda_2} e^{-t_1(\gamma + \lambda_2)} + \frac{\beta}{\beta + \gamma + \lambda_3} e^{-t_1(\gamma + \lambda_3)}
\]

Equating the two expressions of \( \pi \) under the two different model specifications, we can calculate the three stage dependent transmission rates \( \lambda_1, \lambda_2 \) and \( \lambda_3 \) corresponding to a given \( \lambda \). For example, let us assume:

\( t_1 = 30, \beta = 1/(365 \times 8), \gamma = 1/(2.8 \times 365), \lambda_1 = 100\lambda_2, \lambda_3 = 100\lambda_2, \lambda = 0.001 \). We obtain: \( \lambda_2 = 0.000121 \).