Senegal HIV model

Fabricia F. Nascimento

2018-04-10

Introduction

- Phylogenetic trees were estimated for each HIV-1 subtype: B, C, and 02_AG;
- Using information on when sequences were collected, a phylogenetic tree in which branch length were in units of calendar times were also estimated;
- These 3 dated phylogenetic tree were merged into a single tree;
- Each tip of the pylogenetic tree was associated to a state (see section below);
- If a tip could not be associated to a state (because of missing information in the metadata), this particular tip was removed from the dated tree;
- The dated phylogenetic tree in which all tips could be assigned to a state were then used with phydynR to estimate the transmission rates and parameters of the model (see below for more information on which parameters we are estimating).

The Model

The model we fit is based on the structured coalescent models (Volz 2012). These models are used to estimate epidemiological parameters using a phylogenetic tree and information on states of each tip of the tree. These states are discrete-trait information representing each sequence.

In our mathematical model we have 4 different discrete-traits associated to each DNA sequence:

- gpf = infected heterosexual females from the general population;
- gpm = infected heterosexual males from the general population;
- msm = infected male that have sex with other men;
- src = source sample, which are infected individuals that are from other countries and not from Senegal.

Stage of infection

We fit the HIV epidemic in Senegal using ordinary differential equations (ODE) and only 1 stage of infection. This means that infected individuals would die and not recover from the infection. In our model we represented it as γ rate. We used 1 stage of infection, because the

metadata available for the Senegal sequences did not have information that we could use to determine the stage of HIV infection at the time the samples were collected.

How transmissions were modelled?

- An infected msm (I_{msm}) could transmite to another msm with probability $p_{msm2msm}$
- An infected msm (I_{msm}) could transmit to a gpf with probability $(1 - p_{msm2msm})$
- An infected gpf (I_{gpf}) could transmit to a gpm with probability $p_{gpf2gpm}$
- An infected gpf (I_{gpf}) could transmit to a msm with probability $(1 - p_{gpf2gpm})$
- An infected gpm (I_{gpm}) could also transmit to a gpf. For this event, we used the risk ratio of a male to transmite to a female, and fixed it to 1.02. This is the parameter $male_x$ of our model.

How about HIV incidence rate?

We also modelled the HIV incidence rate as a funtion of time (t) in msm and the gp (general population) as different spline functions (Eilers and Marx 1996), that in our ODEs are represented by $\lambda(t)$ and $\mu(t)$, respectively.

The *source* compartment

Finally, to model the HIV epidemic in Senegal, we also added an additional compartment named "source" (src), that represents the rate in which HIV lineages are imported to Senegal from other countries. We modelled this as a constant efective population size rate with two parameters to be estimeted – $srcNe$: the effective source population size; and the $import$ rate. Because the number of imported HIV balances the number of exported HIV, the infected src individuals along time are not represented in the ODEs.

The ODEs or mathematical model equations

$$\dot{I}_{gpf} = male_x \mu(t) I_{gpm} + (1 - p_{msm2msm})\lambda(t) I_{msm} - \gamma I_{gpf}$$

$$\dot{I}_{gpm} = p_{gpf2gpm} \mu(t) I_{gpf} - \gamma I_{gpm}$$

$$\dot{I}_{msm} = (1 - p_{gpf2gpm})\mu(t) I_{gpf} + p_{msm2msm} \lambda(t) I_{msm} - \gamma I_{msm}$$

Estimation of epidemiological parameters

For the Senegal HIV model, we are estimating the parameters using a Markov chain Monte Carlo (MCMC) as implemented in the R package BayesianTools.

Parameters to be estimated and priors

Parameters for estimating the spline function for the *gp*:

- *gpsp0*: prior chosen with mean around 1.1
- *gpsp1*: prior chosen with mean around 1.1
- *gpsp2*: prior chosen with mean around 1.1
- *gpsploc*

Parameters for estimating the spline function for the *msm*:

- *msmsp0*: prior chosen with mean around $R_0 = 1.1$
- *msmsp1*: prior chosen with mean around $R_0 = 1.1$
- *msmsp2*: prior chosen with mean around $R_0 = 1.1$
- *msmsploc*

Parameters that controls the *src*:

- *import*: prior chosen with mean around 0.03
- *srcNe*: prior chosen with mean around 0.05

Probability of certain events to occour:

- *pmsm2msm* : prior chosen with mean around 0.80
- *pgpf2gpm* : prior chosen with mean around 0.80

Initial population sizes:

- *initgp*: prior chosen with mean around 10
- *initmsm*: prior chosen with mean around 10

See Table 1 for a list of parameters that we are estimating and the priors used. Note that lower and upper bounds for the priors were used to keep the posterior distribution at sensible values when using the BayesianTools R package. If such bounds were not provided negative or very high values, when low values were expected, could be proposed during the MCMC.

Status of analysis

Part 1

I have started estimating the parameters of the model using phydynR. The first round of analysis, I did not estimate the initial population size of *gp* and *msm*, and I fixed maleX to 2.0. After meeting with Erik, he explained that these results are strange given that the estimated values for the spline function (the 3 shape parameters) should follow a trend of decreasing rather than increasing. For example: $gpsp0 > gpsp1 > gpsp2$

See attached pdf file (results_part1.pdf). In this pdf, the density plots represent several runs that were merged to achieve ESS for each parameter above 1,000. The black line represents

Table 1: Parameter definition, symbols used in the ODEs, priors with lower and upper bounds

Parameter	Symbol in R	Prior	Lower	Upper
Spline shape gp0	gpsp0	Gamma(3, 3/1.1)	0.01	5
Spline shape gp1	gpsp1	Gamma(3, 3/1.1)	0.01	5
Spline shape gp2	gpsp2	Gamma(3, 3/1.1)	0.01	5
Spline interval gp	gpsploc	U(1978, 2014)	1978	2014
Spline shape msm0	msmsp0	Gamma(3, 3/1.1)	0.01	5
Spline shape msm1	msmsp1	Gamma(3, 3/1.1)	0.01	5
Spline shape msm2	msmsp2	Gamma(3, 3/1.1)	0.01	5
Spline interval msm	msmsploc	U(1978, 2014)	1978	2014
Infectiouness ratio from male to female	maleX	Fixed(1.02)	Not applicable	Not applicable
Importation rate	import	Exp(30)	0	0.30
Effective population size of src	srcNe	Exp(20)	0.0001	0.30
Probability of infected msm to infect another msm	pmsm2msm	Beta(16, 4)	0.1	1
Probability of infected gpf to infect a gpm	pgpf2gpm	Beta(16, 4)	0.1	1
Initial number of infected msm	initmsm	Exp(1/10)	1	300
Initial number of infected gp	initgp	Exp(1/10)	1	300

11 merged independent MCMC that each were run for 18,000 iterations, while the blue line represents 10 merged independent MCMC that each were run for 21,000 iterations. The upper and lower bounds for the priors used to generate these runs were sligtly different for the one in Table 1.

Part 2

I am currently estimating the parameter values by additionally estimating the initial population sizes for *gp* and *msm*, and fixing maleX to 1.02.

I still don't have the results for these new analysis, and I am expecting to have them by next week.

References

Eilers, Paul H. C., and Brian D. Marx. 1996. "Flexible smoothing with B-splines and penalties." *Statistical Science* 11 (2):89–121. https://doi.org/10.1214/ss/1038425655.

Volz, Erik M. 2012. "Complex population dynamics and the coalescent under neutrality." *Genetics* 190 (1):187–201. https://doi.org/10.1534/genetics.111.134627.