Senegal HIV model

Fabricia F. Nascimento

2018-06-27

Introduction

- Phylogenetic trees were estimated using the RaxML for each HIV-1 subtype: 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 using treedater;
- These dated phylogenetic trees were analysed in separate;
- 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 for each subtype 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. This is the risk ratio of a male (gpm) to transmite to a female (gpf). This is the parameter $male_x$ of our model.

See Figure 1 for a partil schematic representation of the transmission model for HIV in Senegal. In this figure gpf, gpm and msm represent the infected individuals.

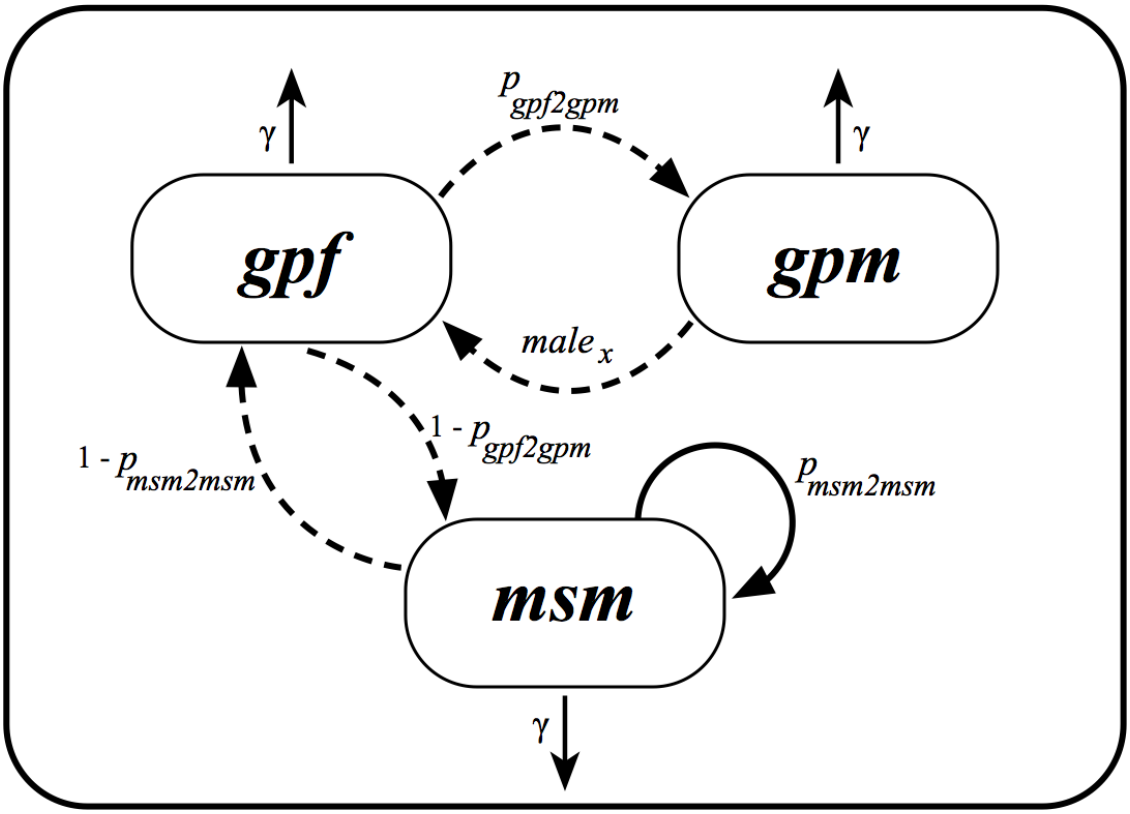

Figure 1: Transmission model for HIV in Senegal. gpf, gpm and msm represent infected individuals.

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 in order to have the mean of $R_0 = 1.1$
- *gpsp1*: prior chosen in order to have the mean of $R_0 = 1.1$
- *gpsp2*: prior chosen in order to have the mean of $R_0 = 1.1$
- *gpsploc*

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

- *msmsp0*: prior chosen in order to have the mean of $R_0 = 1.1$
- *msmsp1*: prior chosen in order to have the mean of $R_0 = 1.1$
- *msmsp2*: prior chosen in order to have the mean of $R_0 = 1.1$
- *msmsploc*

Note that the spline shape parameters (*gpsp0*, *gpsp1*, *gpsp2*, *msmsp0*, *msmsp1*, *msmsp2*) represent the number of transmissions per infected individual.

Given the equation for R_0, we have:

$R_0 = \beta/\gamma$

In our model $\gamma = 0.1$, and β will be represented by each of the spline shape parameters. We are aiming to have a curve representing the number of tansmissions per infected individuals

Parameters that controls the *src*:

- *import*: prior chosen with mean around 0.03
- *srcNe*: prior chosen with mean around 100

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 3
- *initmsm*: prior chosen with mean around 3

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.

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, 3/0.1)	0.05	1
Spline shape gp1	gpsp1	Gamma(3, 3/0.1)	0.05	1
Spline shape gp2	gpsp2	Gamma(3, 3/0.1)	0.05	1
Spline interval gp	gpsploc	U(1978, 2014)	1978	2014
Spline shape msm0	msmsp0	Gamma(3, 3/0.1)	0.05	1
Spline shape msm1	msmsp1	Gamma(3, 3/0.1)	0.05	1
Spline shape msm2	msmsp2	Gamma(3, 3/0.1)	0.05	1
Spline interval msm	msmsploc	U(1978, 2014)	1978	2014
Infectiouness ratio from male to female	maleX	U(0.5, 2)	0.5	10
Importation rate	import	Exp(30)	0	0.30
Effective population size of src	srcNe	Exp(1/100)	1	5000
Probability of infected msm to infect another msm	pmsm2msm	Beta(16, 4)	0	1
Probability of infected gpf to infect a gpm	pgpf2gpm	Beta(16, 4)	0	1
Initial number of infected msm	initmsm	Exp(1/3)	1	300
Initial number of infected gp	initgp	Exp(1/3)	1	300

Partial results

- A total of 116 and 355 sequences were analysed for subtype C and 02_AG, respectively (including sequences that were not from Senegal, to represent the source compartment). From those, 100 and 302 are sequences from Senegal, respectively.
- After fitting the mathematical model to the phylogenetic tree, we can calculate the effective number of infections, the number of new cases, and PAF (proportion attributable fraction of transmissions), as exemplified in the plots for subtypes C and 02_AG. Note that in each plot the solid line is the median and the dashed line is the MAP (maximum a posteriori).

Parameter of interest

Below is a table showing the estimated value for a parameter of interest: *pmsm2msm*

Table 2: Median, MAP, and credible interval for parameter *pmsm2msm* as estimated by our model

Parameter	Median	MAP	2.5%	97.5%
C	0.842	0.869	0.721	0.923
02_AG	0.848	0.899	0.706	0.949

Some thoughts about the results

Results observed by subtype are very different, and we think this is because the way samples were sampled were not random, and the distribution of subtypes C and 02_AG differs between the heterosexual general population and msm. For example: 40% of msm in Senegal are infected with subtype C, while only 4-10% infects general population and FSW (female sex workers) (Ndiaye et al. 2009, 2013).

In the general population, the picture is different, and subtype 02_AG infects 64% of the general population (Ndiaye et al. 2013).

These observations could reflect the different results observed by the different HIV subtypes.

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.

Ndiaye, Halimatou Diop, Edmond Tchiakpe, Nicole Vidal, Ousseynou Ndiaye, Abdou Khoudia Diop, Martine Peeters, Souleymane Mboup, and Coumba Toure-Kane. 2013. "HIV type 1

subtype C remains the predominant subtype in men having sex with men in Senegal." *AIDS Research and Human Retroviruses*. https://doi.org/10.1089/aid.2013.0140.

Ndiaye, Halimatou Diop, Coumba Toure-Kane, Nicole Vidal, Fabien Roch Niama, Pape Amadou Niang-Diallo, Tandakha Dièye, Aissatou Gaye-Diallo, Abdoulaye Sidibe Wade, Martine Peeters, and Souleymane Mboup. 2009. "Surprisingly high prevalence of subtype C and specific HIV-1 subtype/CRF distribution in men having sex with men in Senegal." *Journal of Acquired Immune Deficiency Syndromes*. https://doi.org/10.1097/QAI.0b013e3181af70a4.

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.