Phylodynamics of acute HCV infection in men having sex with men

Gonché Danesh[1], Victor Virlogueux[2], Christophe Ramiere[2], Caroline Charre[2], Laurent Cotte[2], Samuel Alizon[1]

[1] MIVEGEC, CNRS, IRD, University of Montpellier, France
[2] Hospice Civils de Lyon, France

Disease Modeling Symposium 2019

Two ongoing epidemics in Lyon

"Classical" hosts

- Injection drug users
- Detection during chronic infection
- HIV +

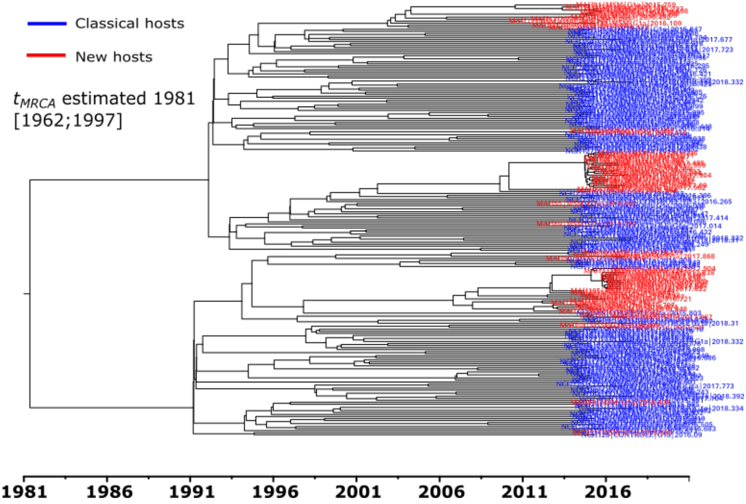
New hosts are being detected since 2000's

- Men who have sex with men (MSM)
- Detection during acute infection
- HIV - and HIV +

Ramiere *et al.* (2019, *Clinical Infectious Diseases*)

Dated viral phylogeny

213 sequences (NS5B gene, 322bp)
Known sampling dates from 2011 to 2018



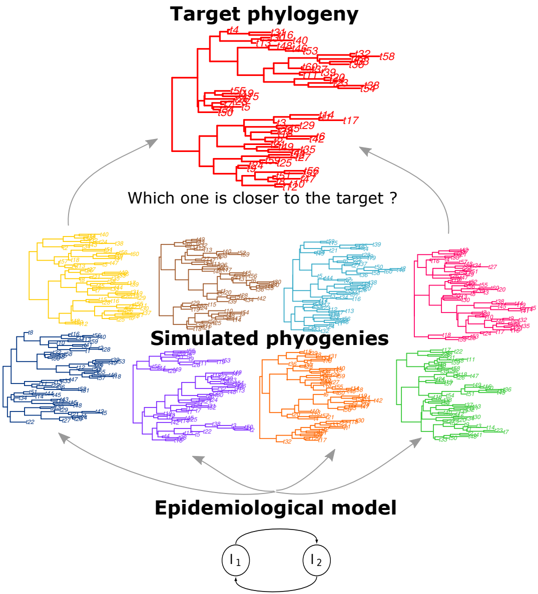
gonche.danesh@ird.fr

Our aim

→ Perform phylodynamics inference using two types of hosts.

→ Infer the basic reproduction number (R_0) for each epidemic simultaneously.

ABC Phylodynamics



Target phylogeny

Which one is closer to the target ?

Simulated phyogenies

Epidemiological model

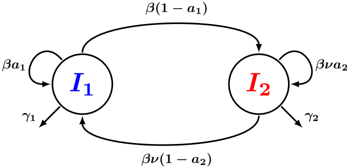
Epidemiological model

Host compartments

- I_1 : classical hosts
- I_2 : new hosts

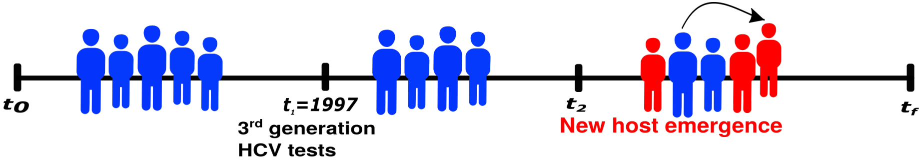
Parameters

- β : transmission rate
- ν : transmission differential
- a_i : assortativity
- $1/\gamma_i$: infectious period
- $R_0 = \frac{\beta}{\gamma}$: reproduction number



$$\frac{dI_1}{dt} = a_1\beta I_1 + (1-a_2)\nu\beta I_2 - \gamma_1 I_1$$
$$\frac{dI_2}{dt} = a_2\beta\nu I_2 + (1-a_1)\beta I_1 - \gamma_2 I_2$$

Priors



	t_0-t_1	t_1-t_2	t_2-t_f
γ_1	Unif$(0.1, 4)$	-	-
γ_2	0	0	Unif$(0.1, 3)$
ν	0	0	Unif$(1, 10)$
a_i	Unif$(0, 1)$	-	-
R_0^1	Unif$(0.9, 15)$	Unif$(0.1, 3)$	-

$$\beta_{t1} = R^1_{0,t1}\gamma_1$$

$$\beta_{t2} = R^1_{0,t2}\gamma_1$$

$$R^2_{0,t3} = \frac{\beta_{t2}\nu}{\gamma_2}$$

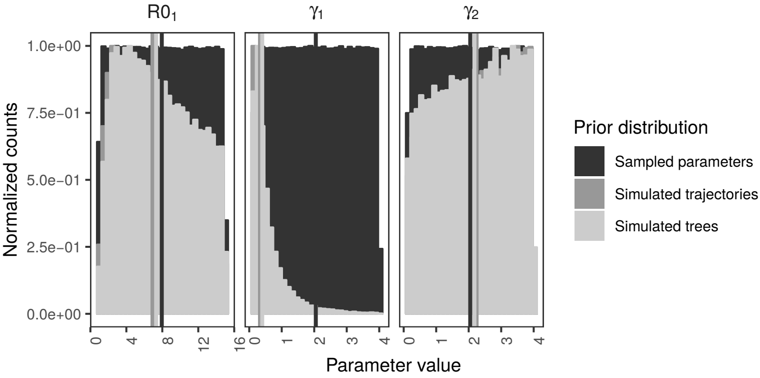
Simulations of phylogenies

- Simulation of dated phylogenies using a two-step procedure:
    1. Simulate epidemiological trajectories given our model using Gillespie Stochastic Simulations Algorithm (SSA)
    2. Simulate phylogenies using simulated trajectories and sampling dates, going backward-in-time
- Simulator EpidSim implemented in Rcpp
- Number of phylogenies simulated (training dataset) is key

Volz (2012, *Genetics*)

The feasibility of the simulations reduces the prior distribution

2.10^6 sets of parameters → 61.000 simulated phylogenies

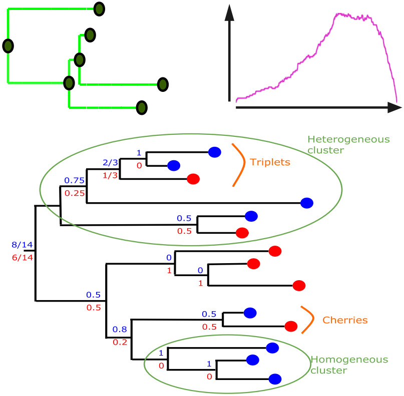


Summary statistics

- Branch lengths
- Tree topology
- LTT (lineage through-time) plot
- X-axis coordinates of the LTT plot
- Y-axis coordinates of the LTT plot

New summary statistics

- Cherries, triplets
- Branch lengths on homogeneous and heterogeneous clusters
- LTT 'label' plot



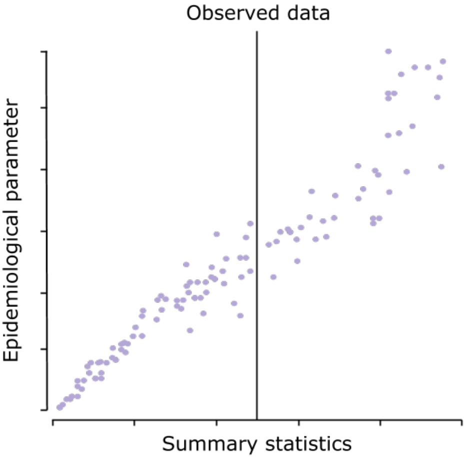
Saulnier *et al.* (2017, *PLOS Comput. Biol.*)

Steps for the ABC Elastic Net regression

1. Test for multicolinearity using Variance Inflation Factors (VIF)
    - Stepwise variable selection
    - $VIF_j = \frac{1}{1-R_j^2}$
    - If $VIF_j > 10$, variable j is removed

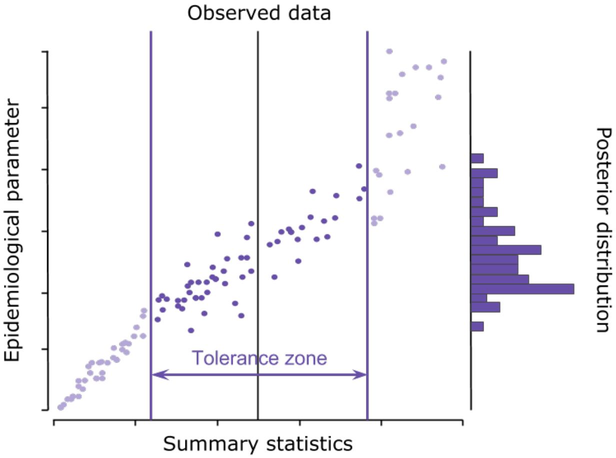
Csillery *et al.* (2010, *Trends Ecol Evol*)
Saulnier *et al.* (2017, *PLOS Comput. Biol.*)

Steps for the ABC Elastic Net regression



Csillery *et al.* (2010, *Trends Ecol Evol*)
Saulnier *et al.* (2017, *PLOS Comput. Biol.*)

Steps for the ABC Elastic Net regression



Csillery *et al.* (2010, *Trends Ecol Evol*)
Saulnier *et al.* (2017, *PLOS Comput. Biol.*)

Steps for the ABC Elastic Net regression

1. Test for multicolinearity using Variance Inflation Factors (VIF)
    - Stepwise variable selection
    - $VIF_j = \frac{1}{1-R_j^2}$
    - If $VIF_j > 10$, variable j is removed
2. ABC rejection step
    - Tolerance parameter $=$ 10 %
    - Results in a first posterior distribution

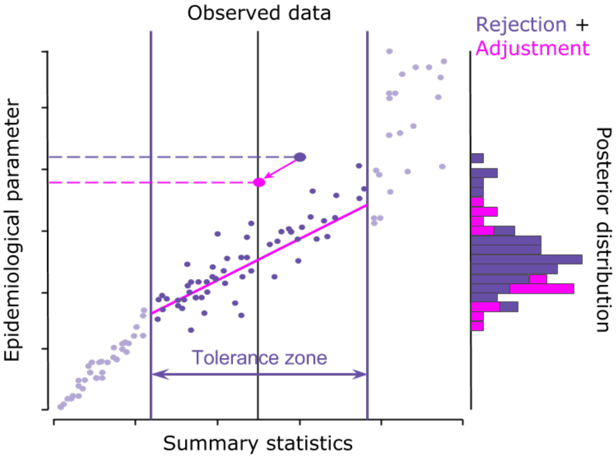
Csillery *et al.* (2010, *Trends Ecol Evol*)
Saulnier *et al.* (2017, *PLOS Comput. Biol.*)

Steps for the ABC Elastic Net regression

1. Test for multicolinearity using Variance Inflation Factors (VIF)
    - Stepwise variable selection
    - $VIF_j = \frac{1}{1-R_j^2}$
    - If $VIF_j > 10$, variable j is removed
2. ABC rejection step
    - Tolerance parameter = 10 %
    - Results in a first posterior distribution
3. Weighting of the non rejected summary statistics

Csillery *et al.* (2010, *Trends Ecol Evol*)
Saulnier *et al.* (2017, *PLOS Comput. Biol.*)

Steps for the ABC Elastic Net regression



Csillery *et al.* (2010, *Trends Ecol Evol*)
Saulnier *et al.* (2017, *PLOS Comput. Biol.*)

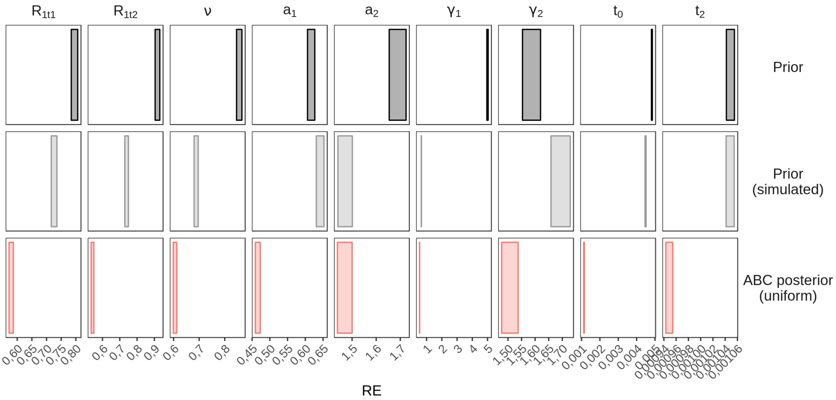
Steps for the ABC Elastic Net regression

1. Test for multicolinearity using Variance Inflation Factors (VIF)
    - Stepwise variable selection
    - $VIF_j = \frac{1}{1-R_j^2}$
    - If $VIF_j > 10$, variable j is removed
2. ABC rejection step
    - Tolerance parameter = 10 %
    - Results in a first posterior distribution
3. Weighting of the non rejected summary statistics
4. Adjust the posterior distribution using Elastic Net linear regression model
    - Combines LASSO regression penalty and Ridge regression penalty
    - LASSO: tends to set coefficients to zero
    - Ridge: shrinks coefficients
    - Results in a final posterior distribution

Csillery *et al.* (2010, *Trends Ecol Evol*)
Saulnier *et al.* (2017, *PLOS Comput. Biol.*)

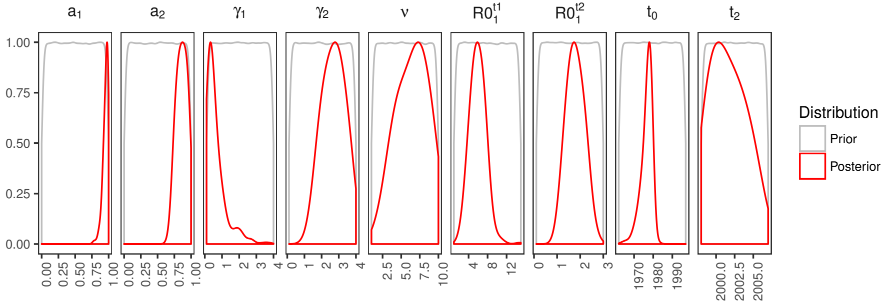
Cross-validation

- 100 replicates: 1 test data x 60.000 training data
- Inference error measurement using relative error (RE)



Given the phylogeny structure and the low number of sequences, our ABC regression method has the power to infer most parameters.

Posterior distributions



a_1	a_2	γ_1	γ_2	ν	$R_0^{1,t1}$	$R_0^{1,t2}$	$R_0^{2,t3}$	t_0	t_2
0.96	0.86	0.6	2.7	6.5	5.9	1.79	2.35	1977	2002

$1/\gamma_1 = 20$ months , $1/\gamma_2 = 4$ months

Conclusions

- New MSM hosts transmit more than classical IDU hosts.
- The R_0 for the "classical" and the "new" epidemics are respectively 1.80 and 2.35.
- The new epidemic is characterized by a shorter infectious period duration (4-5 months) than the classical epidemic (18 months).

Conclusions

- We developed new summary statistics to apply regression-ABC to a multi-hosts typed data to infer different R_0.
- We inferred epidemiological parameters despite the low number of sequences.
- The cross-validation results show how robust our results are for most parameters.
- Perspectives: model comparison, correlations between parameters.

Thank you for your attention !

- The R_0 for the "classical" and the "new" epidemics are respectively 1.80 and 2.35.

- The new epidemic is characterized by a shorter infectious period duration (4-5 months) than the classical epidemic (18 months).

- We developed new summary statistics to apply regression-ABC to a multi-hosts typed data to infer different R_0 simultaneously.

This work was supported by the Fondation pour la Recherche Médicale (FRM grant number ECO20170637560 to GD)