Stockholm University

Inference on contact networks and sampling fractions from epidemic phylogenetic trees

Tom Britton

February, 2017

with Federica Giardina, Jan Albert, Ethan Romero-Severson and Thomas Leitner (PLoS Comp Bio, 2017 + ongoing work)

Stockholm University

Background and scientific questions

General aim: To study how reconstructed virus phylogenies of sampled cases can help inferring contact networks (published) and undetected fractions (ongoing)

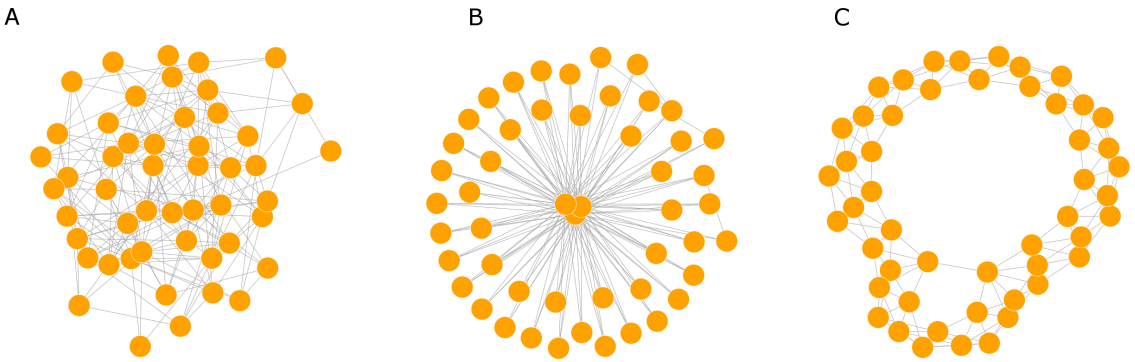
More specific assumptions

- Infectious disease. **Data**: time of diagnosis + virus sequence
- Simple contact structures: uniform mixing, Erdös-Renyi, Small-World, Preferential attachment (**unobserved!**)
- Assumption: evolution rate of pathogen comparable to disease spread
- Within host diversity acknowledged
- Application: HIV in Sweden

Stockholm University

Contact networks

- Completely random network: Erdös-Renyi (independent edges between all pairs)
- Preferential attachment (Barabasi-Albert): sequentially more connections to popular individuals
- Small-World (Watts and Strogatz): local/spatial connections + random connections

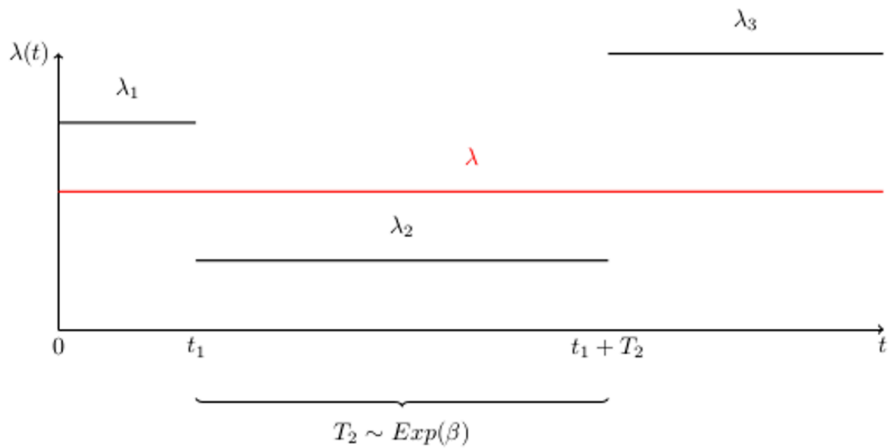


Stockholm University

Epidemic model

Epidemic model

- Susceptible-Infectious-Recovered (SIR) model
- Transmission to neighbours in network
- All individuals equal, or individual heterogeneity
- Infectivity profile: Constant infectivity, or three phases: high – low – very high (HIV)



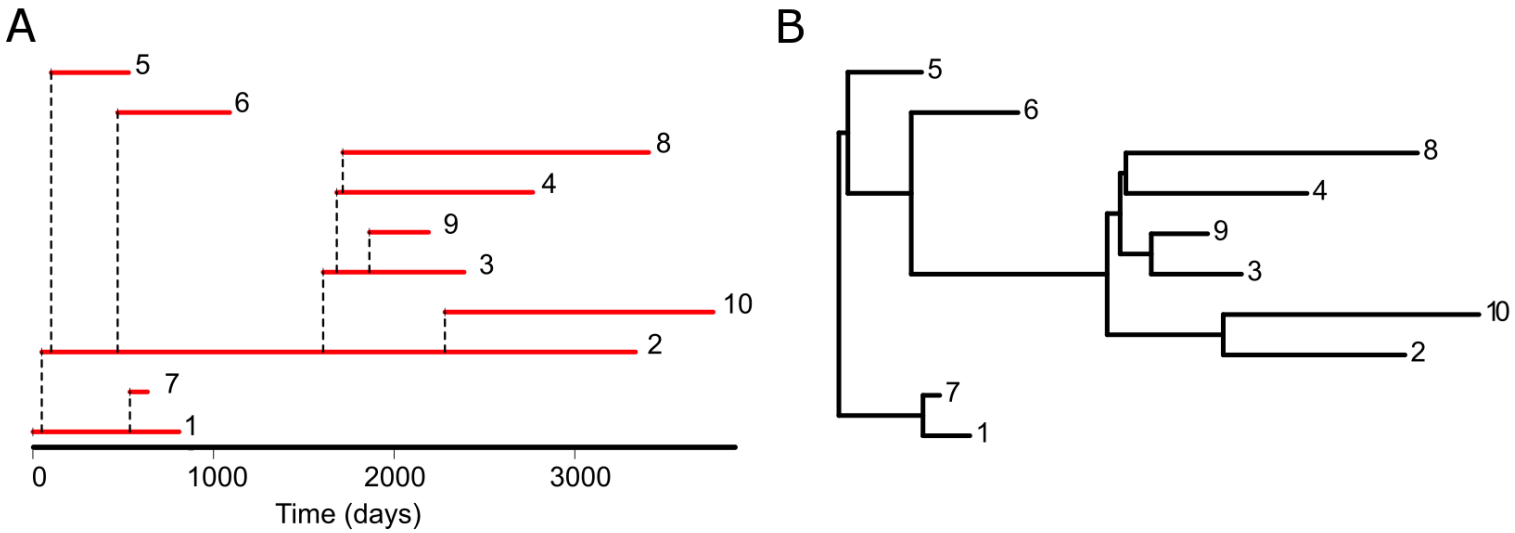
Stockholm University

Virus phylogeny

The network together with spreading model produce a *Transmission history*

Using a within-host model for virus evolution, this in turn gives a *virus phylogeny* which can be estimated

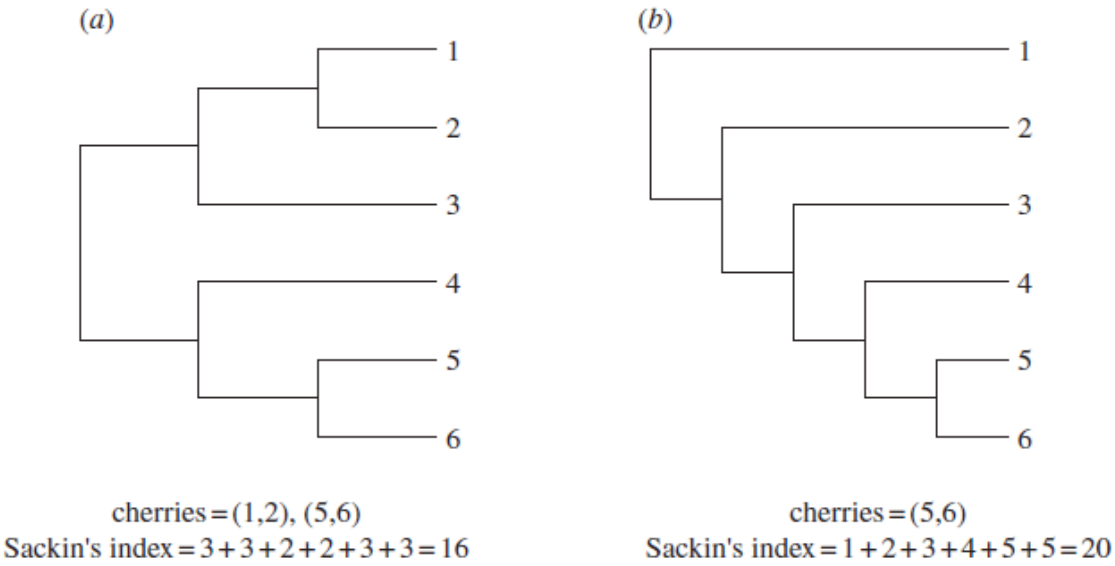
Who-infects-whom is lost. Topologies may also **differ** if within-host diversity! (Treated in model)



Tree diagnostics

Simulation of Transmission Histories (TH) $\rightarrow$ Virus Phylogenies (VP)

Comparison of trees using different tree diagnostics, e.g. *Sackin's index* and *Cherries* (Frost and Volz, 2013)

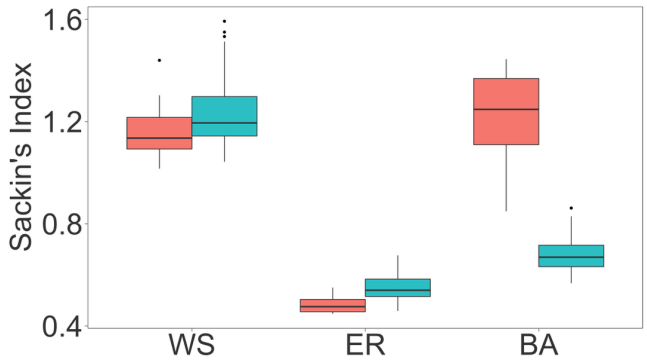


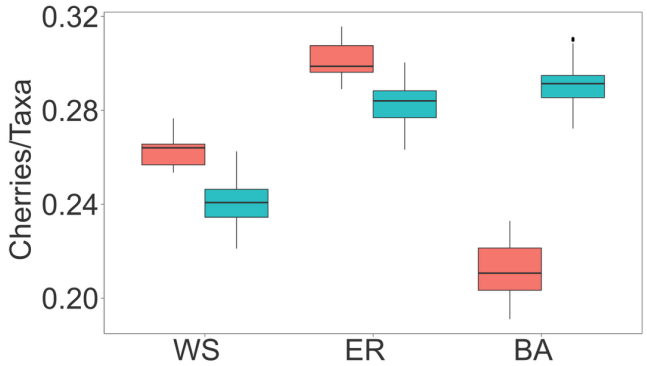
Stockholm University

Diagnostics from simulations: distinguishing networks

Aim: See effect of Network structures, Infectivity profiles, individual heterogeneity. Using various tree-diagnostics (cf. Leventhal et al., 2012)

Tree statistics from simulations in networks of size 1000: Transmission history (red) and Virus phylogeny (blue)



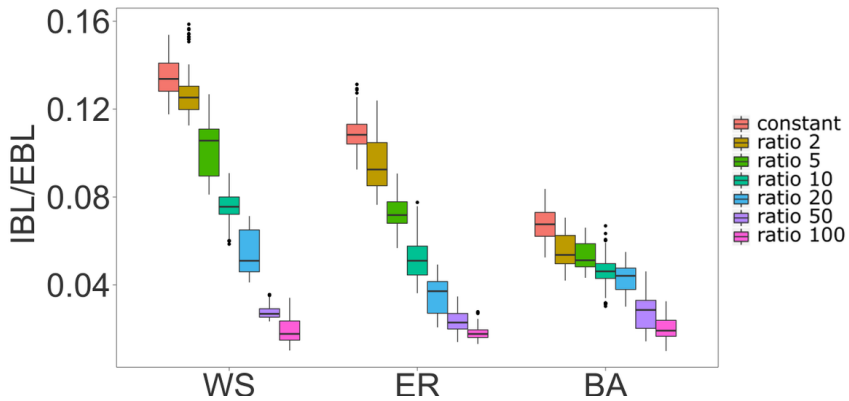


Conclusion: Type of network identifiable both for TH and VP

Stockholm University

Simulations diagnostics: distinguishing infectivity profile

Aim: See effect of time-varying infectivity on tree statistics of VP (higher ratio $\rightarrow$ bigger time-variation of infectivity)

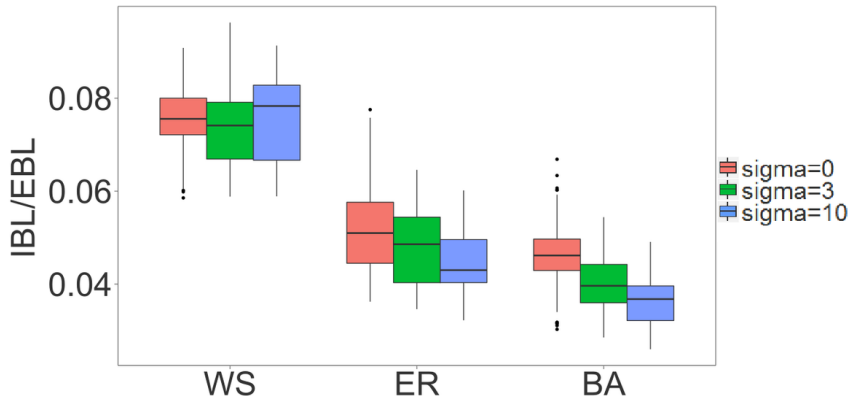


Conclusion: Possible to determine infectivity profile, but not type of network when infectivity profile highly variable

Stockholm University

Simulations diagnostics: distinguishing individual variability

Aim: See effect of variable infectivity on tree statistics of VP (larger $\sigma \rightarrow$ more variablity)



Conclusion: Individual variability not identifiable, but networks identifiable also with individual variability

Stockholm University

Inference from simulations of VP

Inference use ABC for model selection (type of network) and parameter estimation

Network size 1000, acute vs chronic 10:1, $\sigma = 3$ and sampling fraction =0.5

	BA	ER	WS
BA	**0.780**	0.210	0.010
ER	0.220	**0.765**	0.015
WS	0.003	0.006	**0.991**

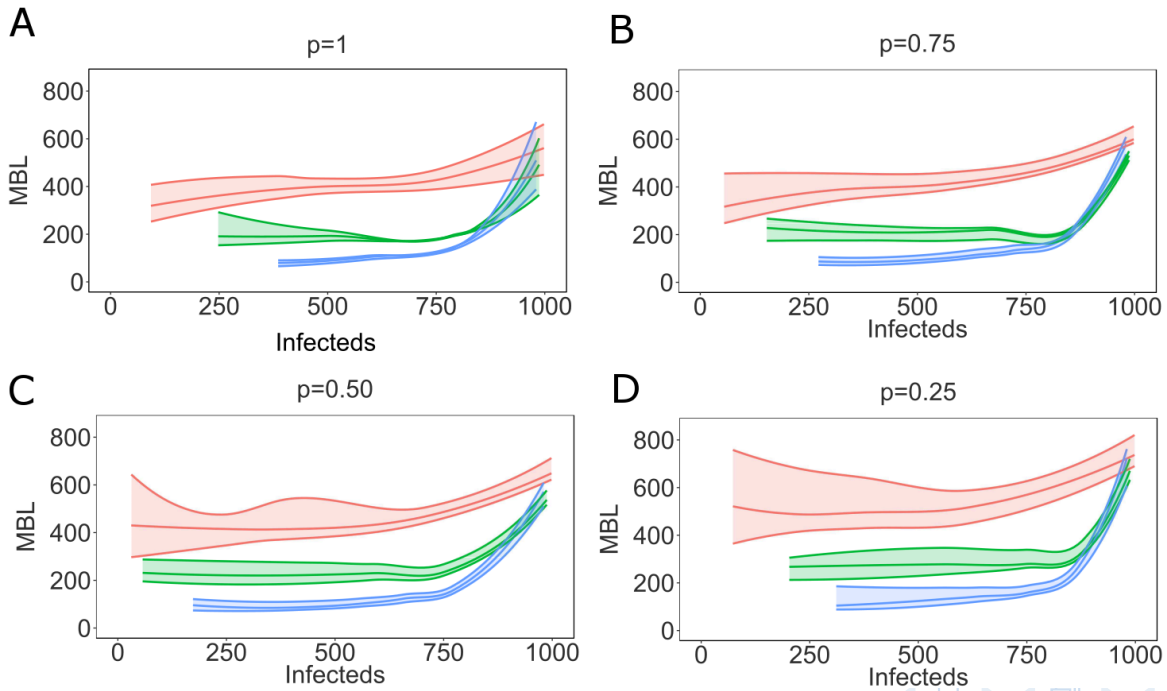
Parameter	Median	95% CrInt	True value
Mean NW degree	8.5	(7.8, 8.7)	8
Removal rate	0.25	(0.19, 0.37)	0.35
Acute st inf-rate	0.008	0.002, 0.010	0.005

Stockholm University

Simulation diagnostics during an outbreak with unobserved

Previous plots were from final outbreak size and all cases sequenced

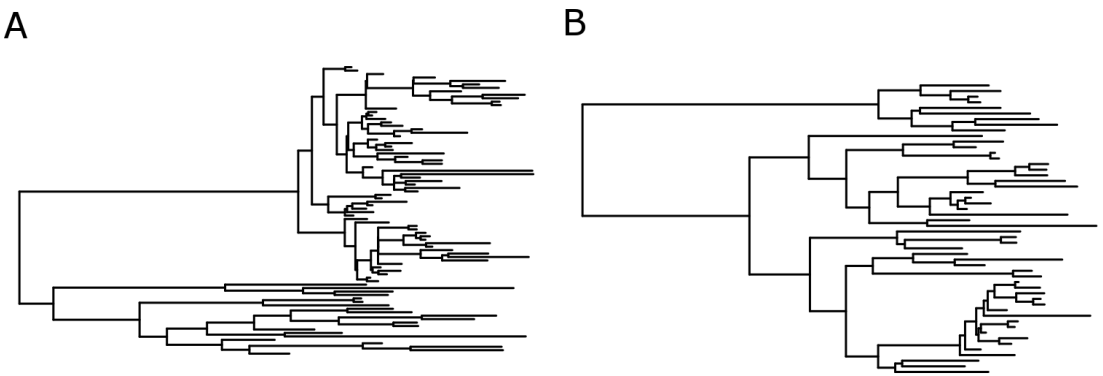
Similar analysis possible *during* an outbreak and when not all are diagnosed and/or sequenced (p = sequenced fraction): WS (red), E-R (green), B-A (blue)



Stockholm University

Two HIV-outbreaks in Sweden (IDU)

First outbreak is a rapid CRF01 outbreak (left) and slower subtype B outbreak (right)
Bayesian skyline coalescent model (Beast)



Result:

WS	ER	BA
0.16	0.39	0.45

WS	ER	BA
0.34	0.57	0.09

Stockholm University

Conclusions

- Virus phylogenies carry information on underlying network
- Inference from Virus phylogeny different from inference from Transmission history
- Time-varying infectivity identifiable but make network differences less pronounced
- Individual variability harder to identify

Stockholm University

Inference for N_U = number of unobserved cases

In many situations only a fraction of cases are observed (or sequenced): asymptomatic, not yet showing symptoms, not sequenced, ...

Important to estimate how many additional people are infected

Ongoing work (feedback welcome!): trying to estimate N_U in some toy examples (cf. Gamado et al., 2016)

Epidemic model: Simplest SIR epidemic model

Population structures: homogeneous mixing, 2-type population, household model

Data types: **a)** Diagnosis time only, **b)** diagnosis time and virus phylogeny, **c)** diagnosis time and transmission history

Stockholm University

Preliminary findings

Homogeneous mixing:
– N_U can be estimated from data c
– Virus phylogeny/transmission history only helps by partly determining infection times

Structured models:

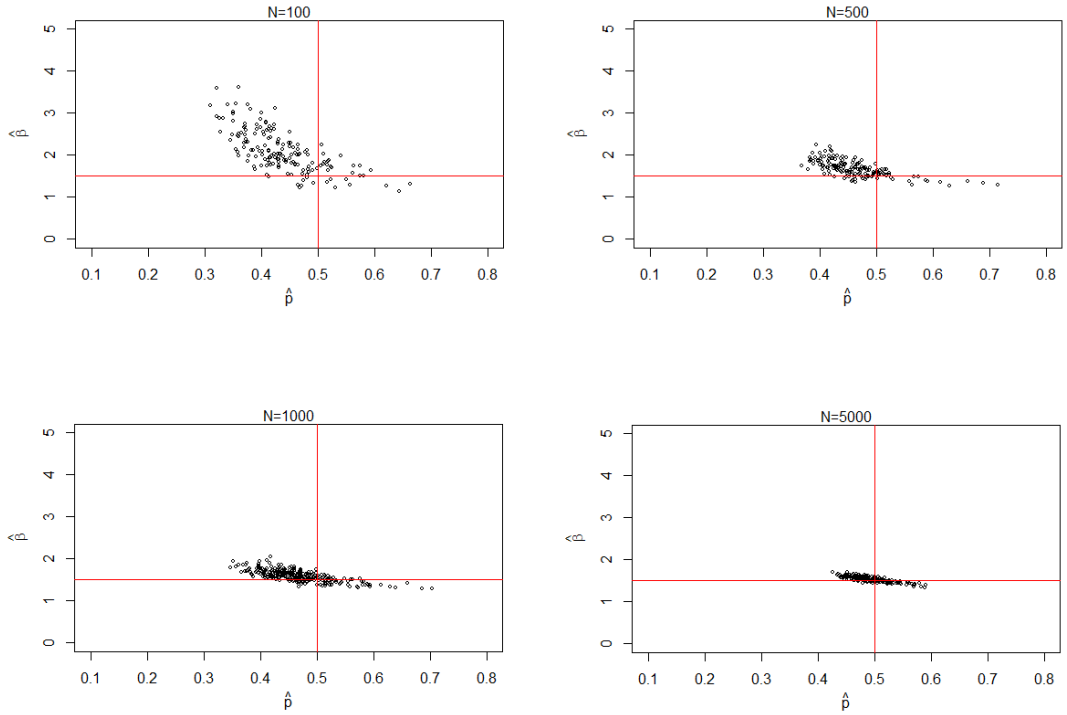
– Virus phylogeny/transmission history more important (helps estimating inter- and cross-transmission rates)

Stockholm University

Preliminary findings: result for homogeneous mixing

Estimation of transmission rate β and fraction sampled p (true values: 1.5, 0.5). (Recovery rate γ easily estimated)

Data: transmission history and recovery times



Stockholm University

Preliminary findings: : result for homogeneous mixing

Estimated precision

N	$\hat{\beta}$	$\hat{p}$	$Var(\hat{\beta})$	$Var(\hat{p})$	$corr(\hat{\beta}, \hat{p})$
50	1.53(0.82,3.02)	0.50(0.31,0.99)	0.35	0.06	-0.59
100	1.53(0.88,2.75)	0.48(0.32,0.98)	0.24	0.05	-0.70
500	1.55(1.16,2.02)	0.50(0.36,0.97)	0.05	0.027	-0.76
1000	1.19(1.46,1.75)	0.48(0.40,0.75)	0.02	0.007	-0.77

Stockholm University

Questions of general interest?

Can we estimate the fraction of unreported cases?

Can we estimate how many asymptomatics there are and how much they contribute to an outbreak?

More generally:

Can we from an estimated phylogeny infer how many samples are missing, and if and how this makes our estimated phylogeny biased?

Can we identify locations in phylogeny that miss data?