Phylogeographic analysis of infectious disease agents based on birth–death processes

Denise Kühnert

ETH Zürich

September 13, 2013

Overview

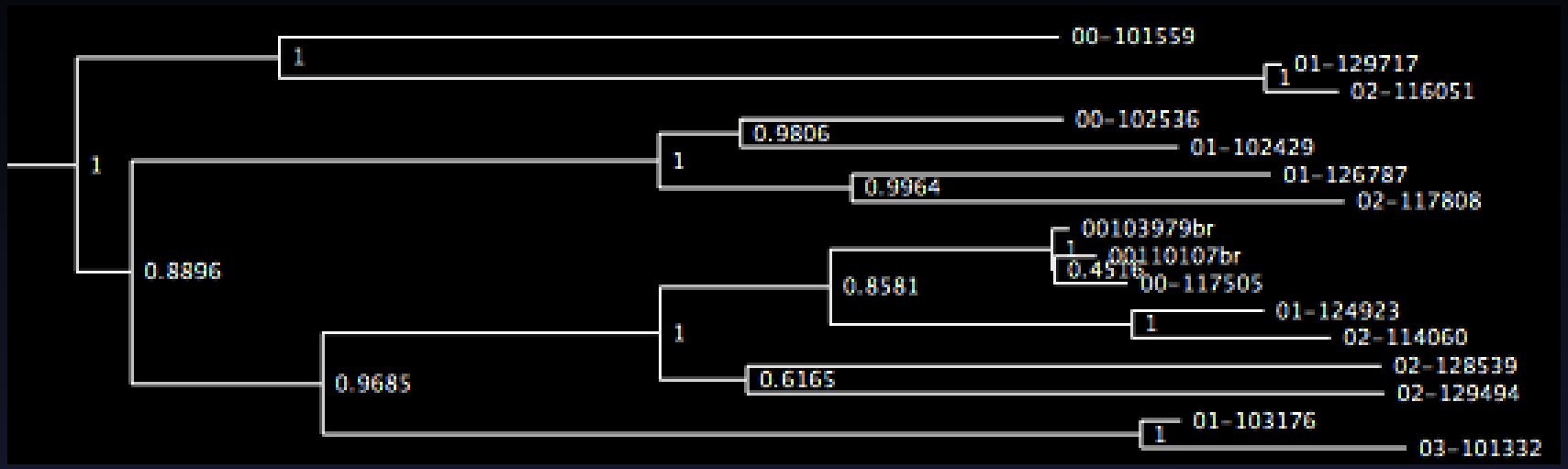
- Motivation
- What is a "tree prior"?
- The multi-type birth–death model
- Analysis of Latvian HIV data

Why phylogenetics?

- When and where did a virus originate in a certain host population (e.g. HIV in humans)?
- How fast does a pathogen evolve?
- What is the relationship among viral subtypes?

- How fast does a virus spread in a host population?
- Do intervention measures help to contain/stop an epidemic?
- How important are differences within the host population?

Bayesian inference of time-trees



Reconstruction of evolutionary history using Bayesian Markov Chain Monte Carlo methods:

- Approximate the likelihood of the model given the data
- Sample states (model parameter combinations) from prior distribution and then sample the desired stationary distribution
- Result: Posterior distribution of the model parameters of interest (e.g. phylogenetic tree, mutation rate)

Posterior distribution and "tree priors"

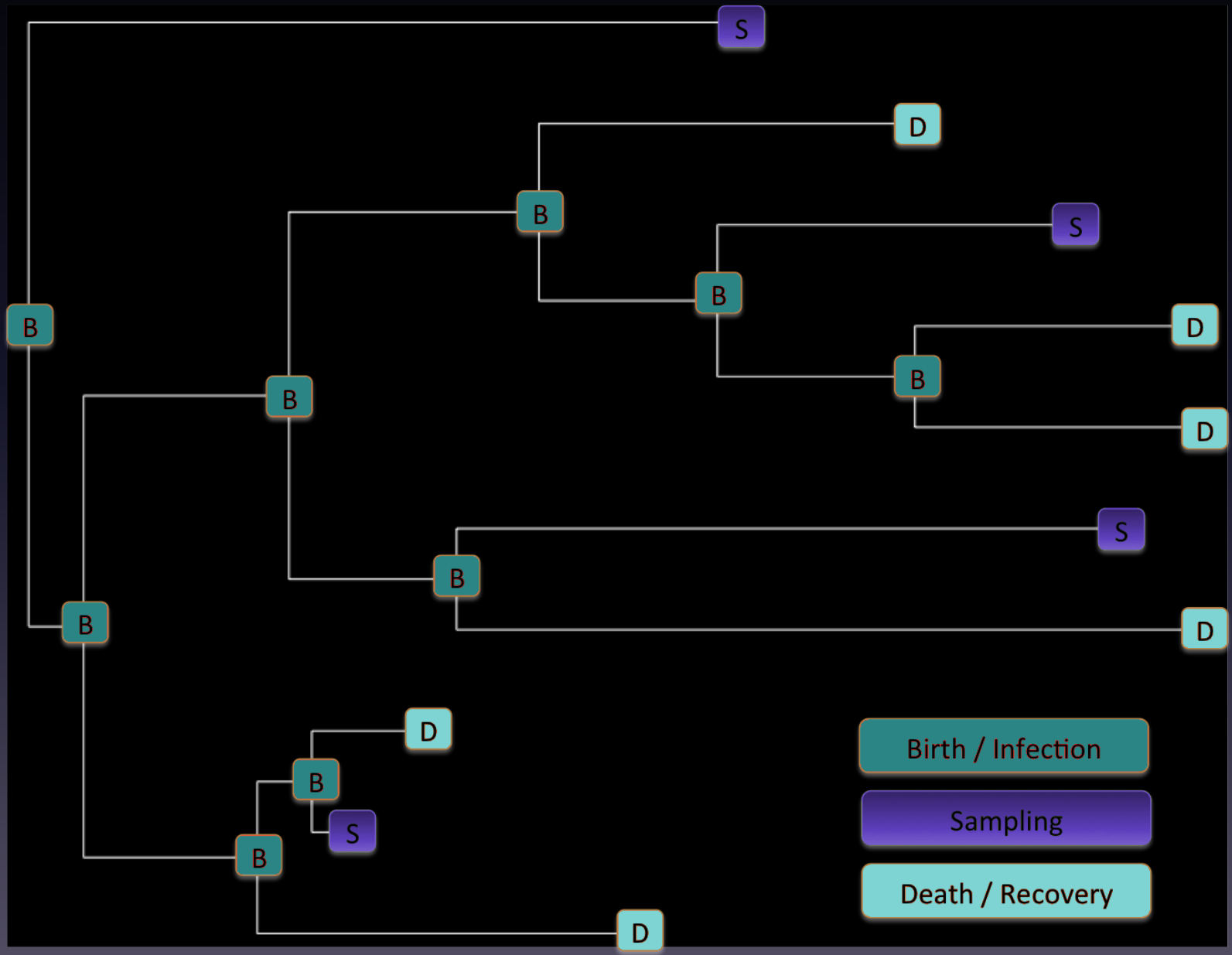
Posterior distribution of the time tree $\mathcal{T}$, the tree generating rates r and the model parameters θ given the data $\mathcal{D}$:

$$f[\mathcal{T}, r, \theta|\mathcal{D}] \propto f[\mathcal{D}|\mathcal{T}, \theta] \underbrace{f[\mathcal{T}|r]}_{\text{"tree prior"}} f[r]f[\theta]$$

Examples of tree generating processes:

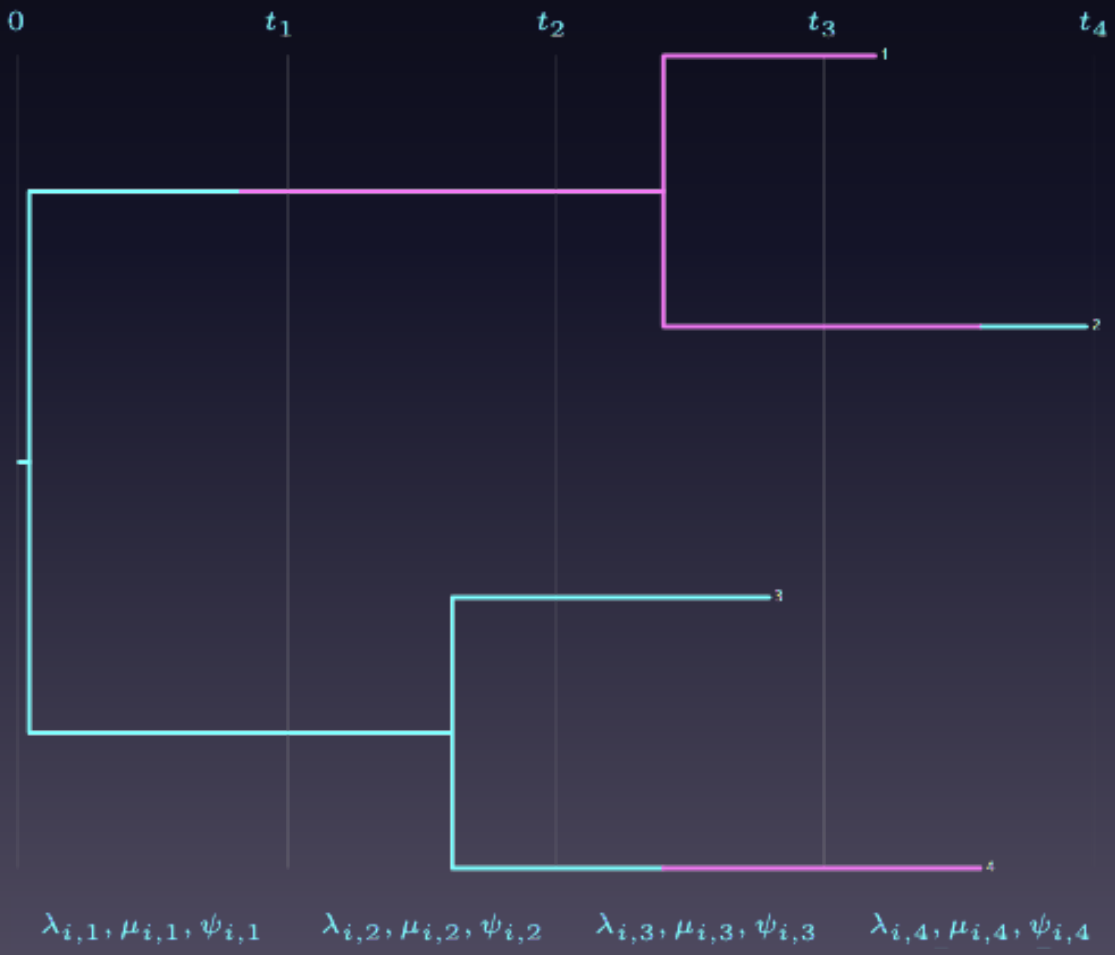
Single population	**Single + rate change**	**Population structure**
Birth–Death–Sampling model (Stadler 2010)	BDSKY (Stadler, K et al. 2013)	Multi-type Birth–death + rate change
Kingman's coalescent (Kingman 1982)	BSP (Drummond et al. 2005)	Structured coalescent (Notohara 1990)

The Birth–Death–Sampling tree prior



A tree prior for *'coloured'* trees

Birth–death–sampling process $\mathcal{N} = [\mathcal{N}_1, \ldots, \mathcal{N}_m]$, where $\mathcal{N}_i$ is the number of infected individuals in population $i \in \{1, \ldots, m\}$ and individuals can migrate among the discrete locations.



Model parameters

Birth, death and sampling rates can differ between the n intervals and m states/locations:

$$\begin{aligned}\boldsymbol{\lambda} &= ((\lambda_{1,1},..,\lambda_{1,m}),(\lambda_{2,1},..,\lambda_{2,m}),\cdots,(\lambda_{n,1},..,\lambda_{n,m}))\\ \boldsymbol{\mu} &= ((\mu_{1,1},..,\mu_{1,m}),(\mu_{2,1},..,\mu_{2,m}),\cdots,(\mu_{n,1},..,\mu_{n,m}))\\ \boldsymbol{\psi} &= ((\psi_{1,1},..,\psi_{1,m}),(\psi_{2,1},..,\psi_{2,m}),\cdots,(\psi_{n,1},..,\psi_{n,m}))\end{aligned}$$

Reparametrization into *reproduction ratio* R, *become non-infectious rate* δ and *sampling proportion* s

$$R_{i,j} = \frac{\lambda_{i,j}}{\mu_{i,j}+\psi_{i,j}},\ \delta_{i,j} = \mu_{i,j}+\psi_{i,j},\ s_{i,j} = \frac{\psi_{i,j}}{\delta_{i,j}}$$

Migration matrix $\boldsymbol{M}$ is constant over time:

$$\boldsymbol{M} = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{bmatrix}$$

The likelihood of a 'coloured' tree

The probability density f of a tree is

$$f(\mathcal{T}|\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\psi}, \boldsymbol{M}, t) = \sum_{k=1}^{m} s_k \cdot g^e_{l(t),k}(t)$$

with $l(t) = i$ iff $t_{i-1} \leq t < t_i$ and s_k denoting the probability that an individual is in state k, and the probability density $g^e_{i,k}(t)$ at time $t_{k-1} \leq t < t_k$ that an individual in state $i \in \{1, .., n\}$ evolved as observed in the tree.

Application to Latvian HIV data

Data set:

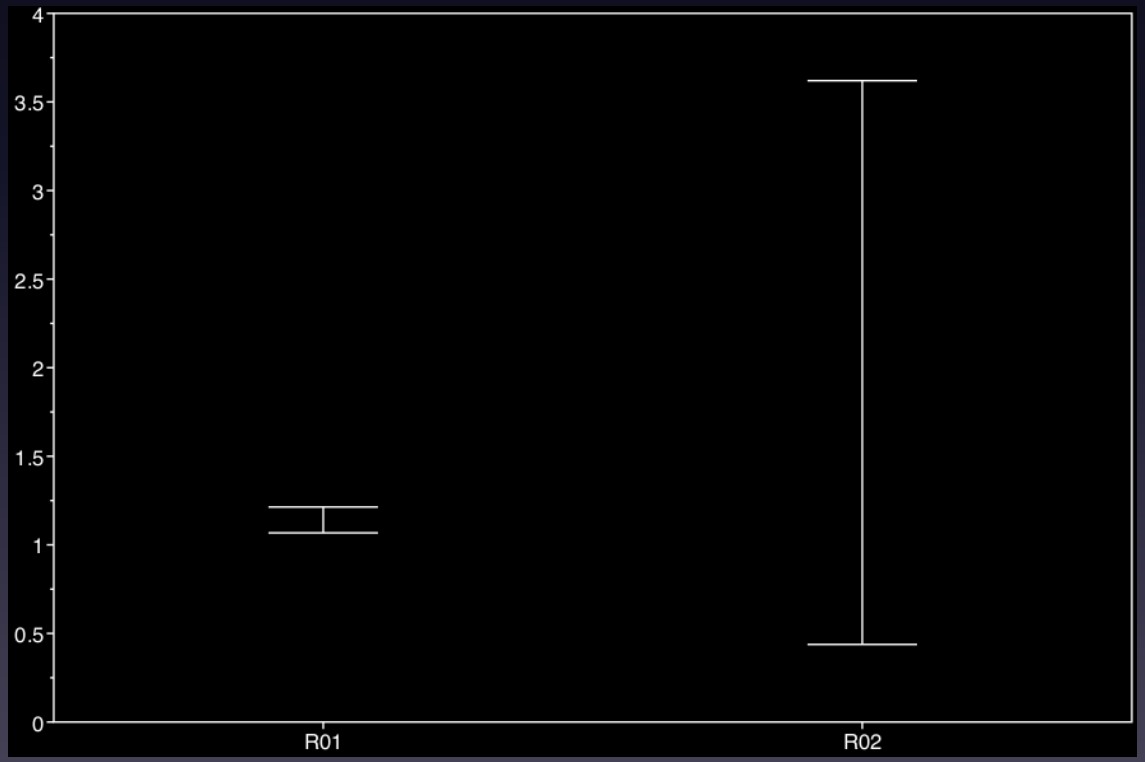
- 130 sequences of HIV subtype A from Latvia
- Matrix protein p17
- 65 sequences from HET's, 65 from IDU's
- Subsample of previously published data by Balode et al. 2004, 2012

Questions:

- What are the epidemiological characteristics of both risk groups?
- Is the IDU sub epidemic 'stronger' than the HET subepidemic?
- Does the HET subepidemic depend on IDU's?
  cf. Stadler, Bonhoeffer 2013

Preliminary results - R

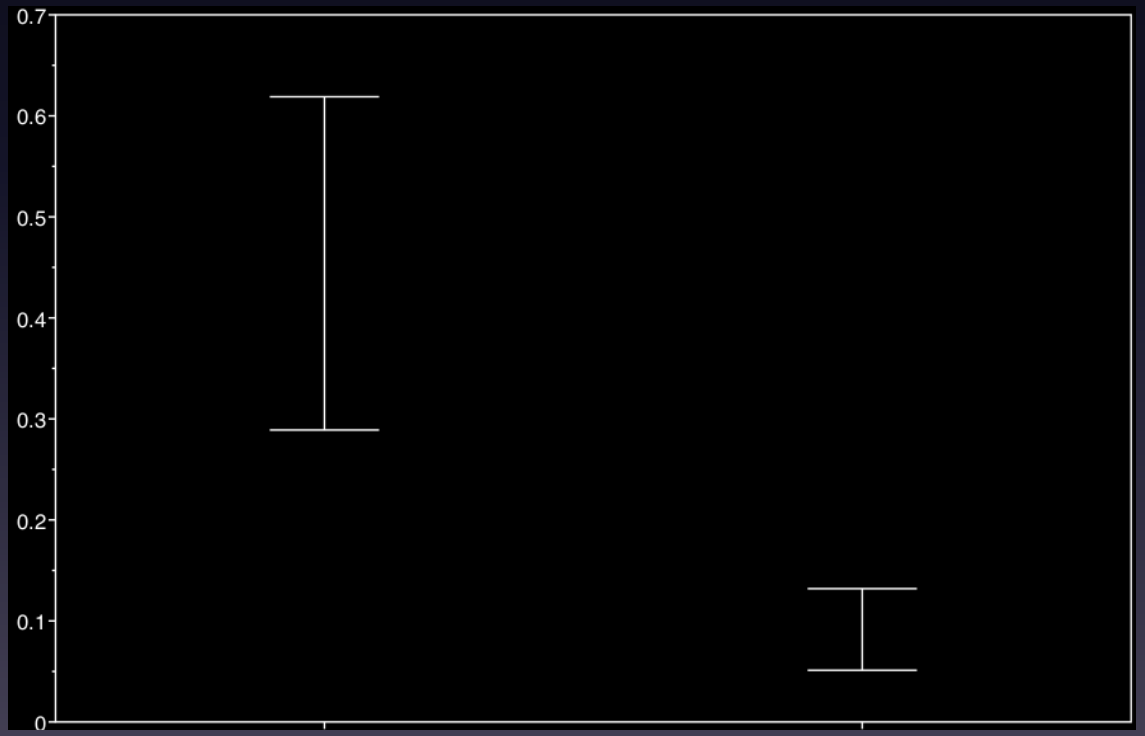
	R_{HET} (median)	R_{IDU} (median)
MCMC	1.14	1.63
ML*	0.38	1.13



*Stadler & Bonhoeffer 2013, PhilTrans

Preliminary results - migration rates

HET $\rightarrow$ IDU	IDU $\rightarrow$ HET
0.43	0.086



Summary

- Multi-type birth–death model provides a tree prior for phylogeographic analysis, including migration rate estimation
- Enables simulation of 'coloured' phylogenies
- Epidemiological rates can differ among locations
- Simultaneous reconstruction of the phylogeny and epidemiological rates
- Latvian HIV requires further analysis:
  $\rightarrow$ Use exactly the same data set as Stadler, Bonhoeffer 2013
  $\rightarrow$ Allow infections between compartments instead of migration
  $\rightarrow$ Compare with V3 region
  $\rightarrow$ Investigate large become non-infectious rate (med $\delta = 2.54$)

Thank you for your attention

Acknowledgements

ETH Zurich:
Dr Tanja Stadler
Prof Sebastian Bonhoeffer

University of Auckland:
Prof Alexei J Drummond
Dr Tim G Vaughan
Dr Remco Bouckaert

The Marsden Fund - Te Pūtea Rangahau a Marsden
The Allan Wilson Centre