Phylodynamic analysis of rapidly evolving pathogens

Probability and Biological Evolution
CIRM, Marseille

Denise Kühnert

ETH Zürich

June 15, 2015

Outline

1. Why phylogenetics / phylodynamics?
2. How? (using branching processes)
3. Examples:
   - Globally circulating influenza virus
   - Risk group dynamics in HIV
   - Ebola virus in Sierra Leone 2014

1) Why phylogenetics / phylodynamics?

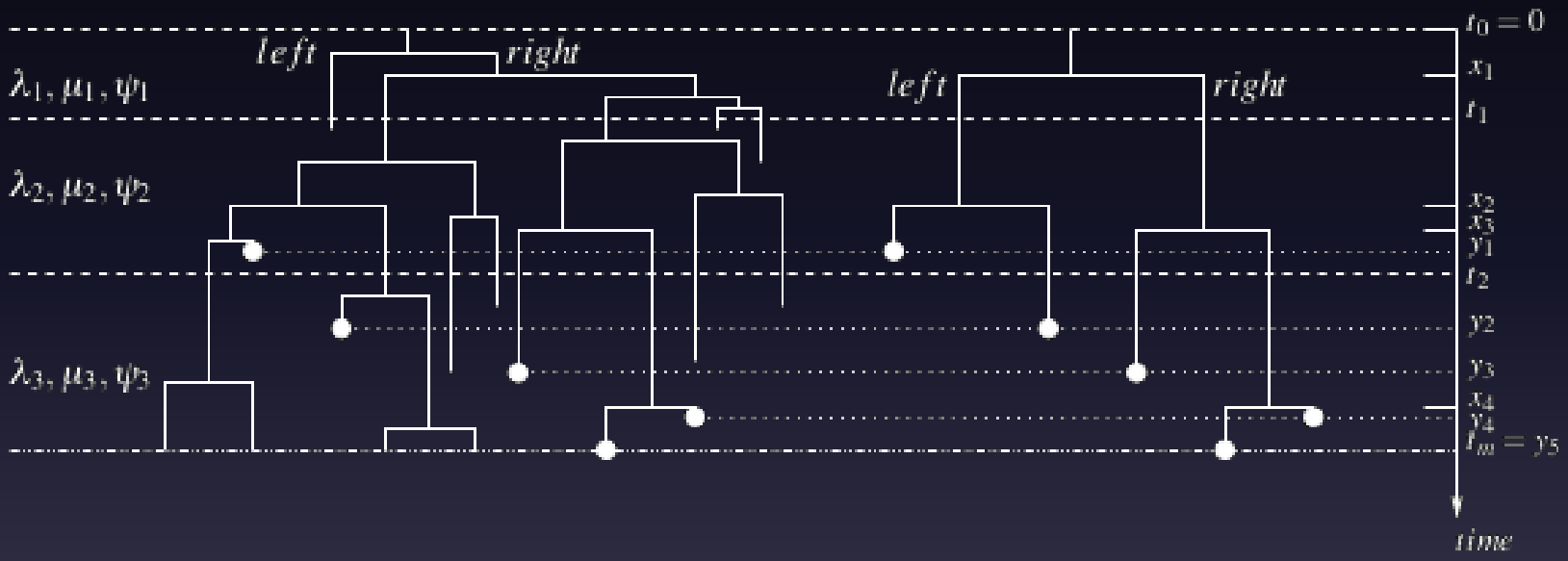
Virus phylogenetics / phylodynamics

- Estimate viral mutation rates
- Epidemic dynamics:
  - Basic/effective reproduction number R_0 / R_e
  - Infectious periods
  - Contact heterogeneity
  - ...

When incidence data is rare, variation in genomic data can provide valuable insight into epidemic dynamics.

Phylogenetic trees

Full tree (unknown) Reconstructed tree



Assume that reconstructed phylogeny is a good proxy for the observed transmission tree.

2) Branching processes in phylogenetics

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 $\mathcal{D}$:

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

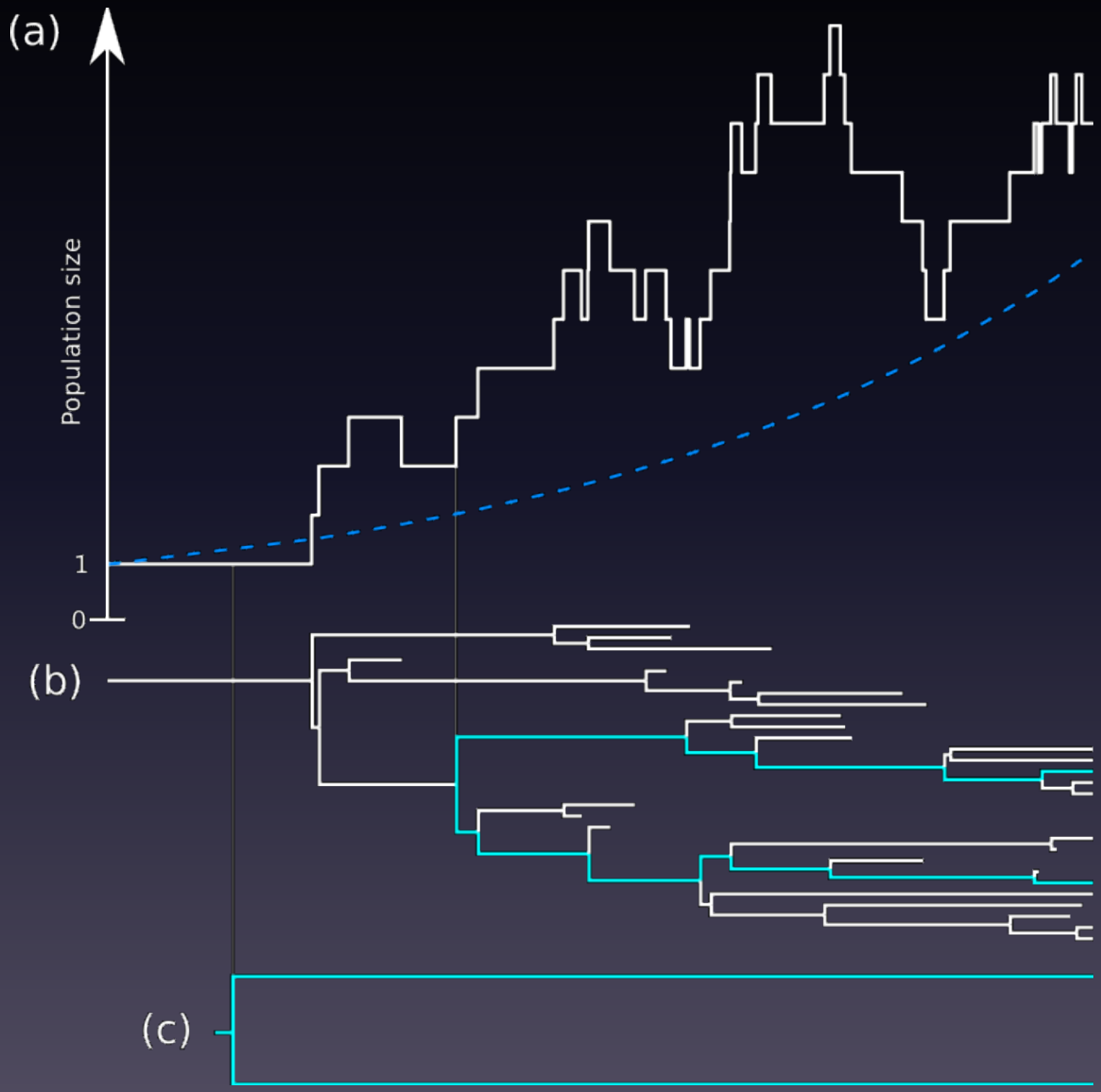
Analysis using Bayesian MCMC (in BEAST2).

"The Coalescent"

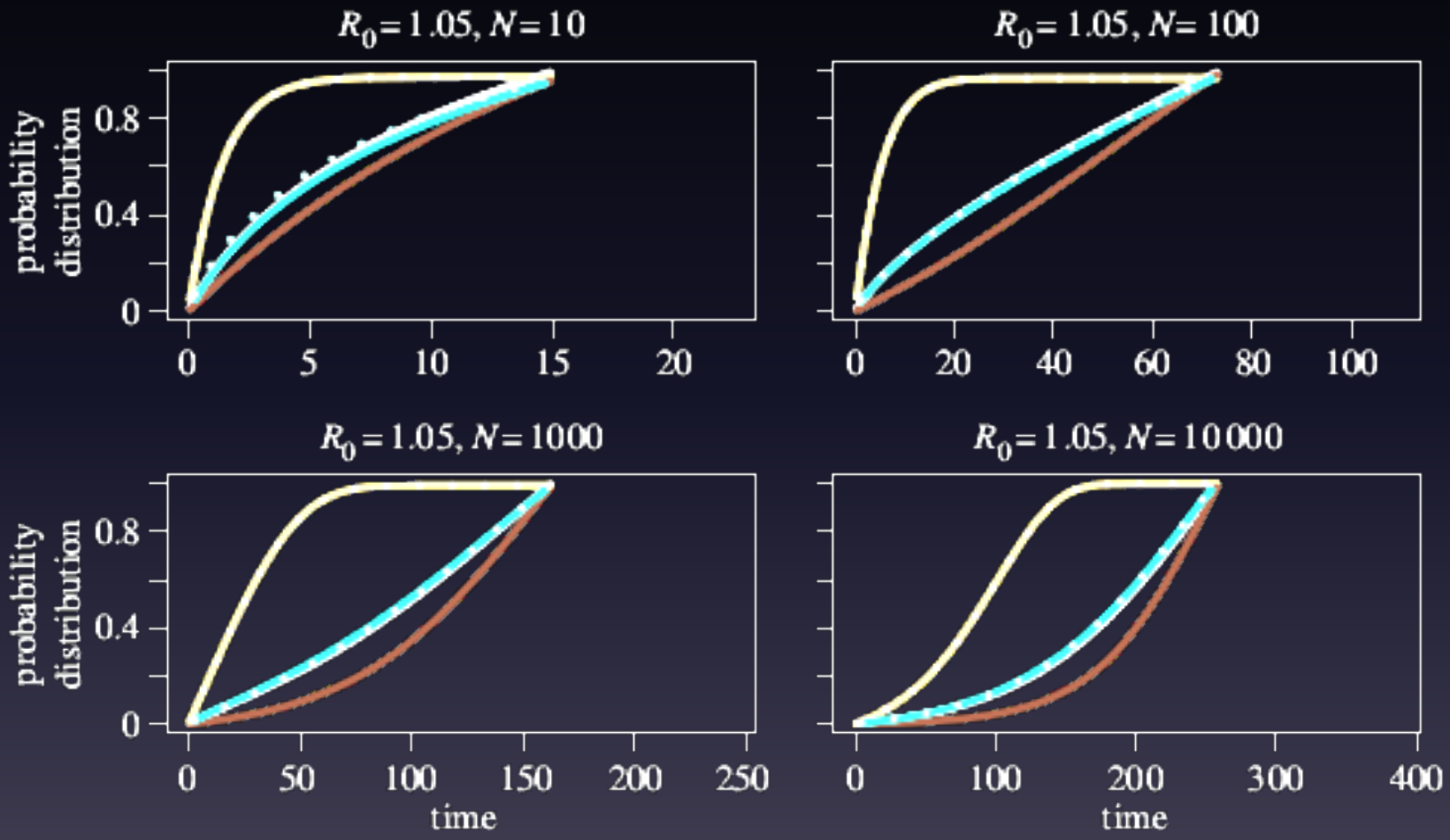
Most popular in phylogenetics:

- Neutral, deterministic population coalescent
  (Griffiths & Tavaré 1994)
- Simple & computationally efficient
- Exponential growth coalescent provides growth rate estimate
  $\rightarrow$ Estimate basic reproductive number R_0
  (Pybus et al. 2001)

Stochasticity



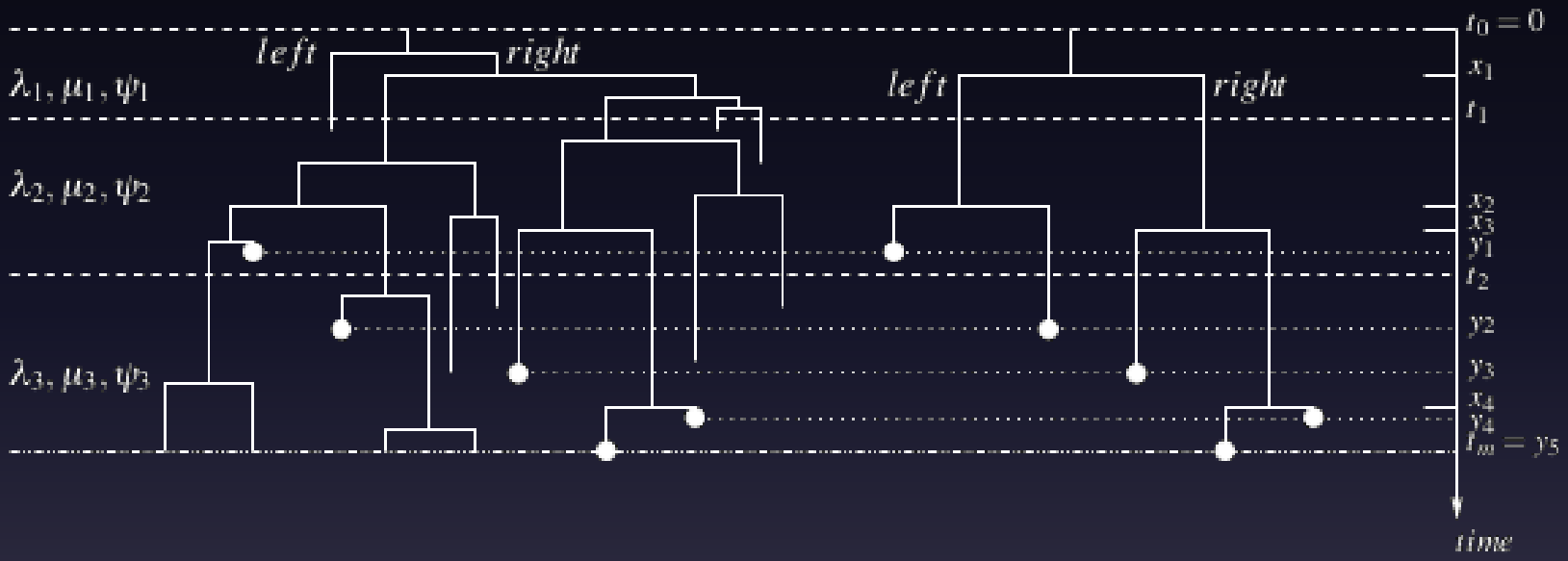
Coalescent times of 2-tip trees



Stadler et al. 2015, Proceedings B

Birth-death approaches

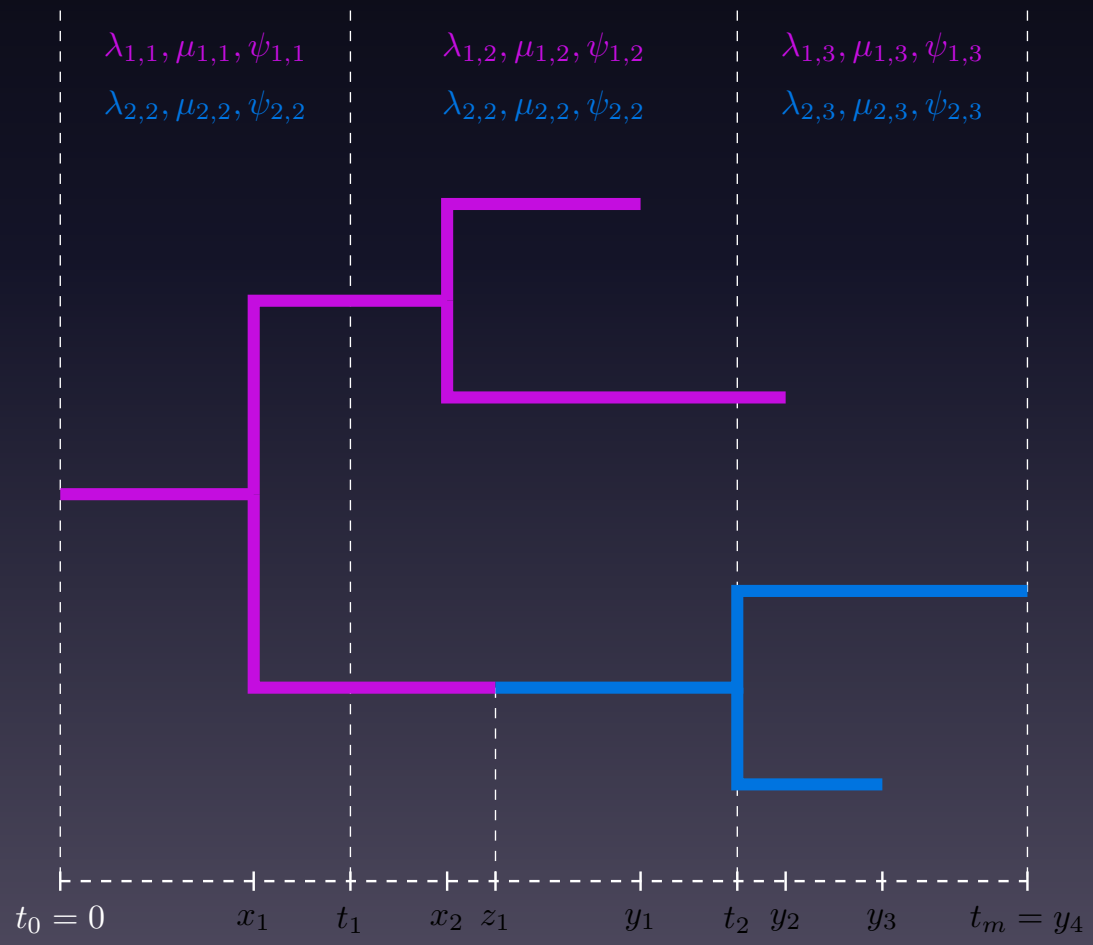
Full tree (unknown) Reconstructed tree



- Birth-death skyline model (Stadler, Kühnert et al. 2013)
- SIR-approximation "BDSIR" (Kühnert et al. 2014)

Multi-type birth-death tree prior

.. with $d \in \mathbb{N}$ demes and $m \in \mathbb{N}$ intervals



Calculating the tree likelihood

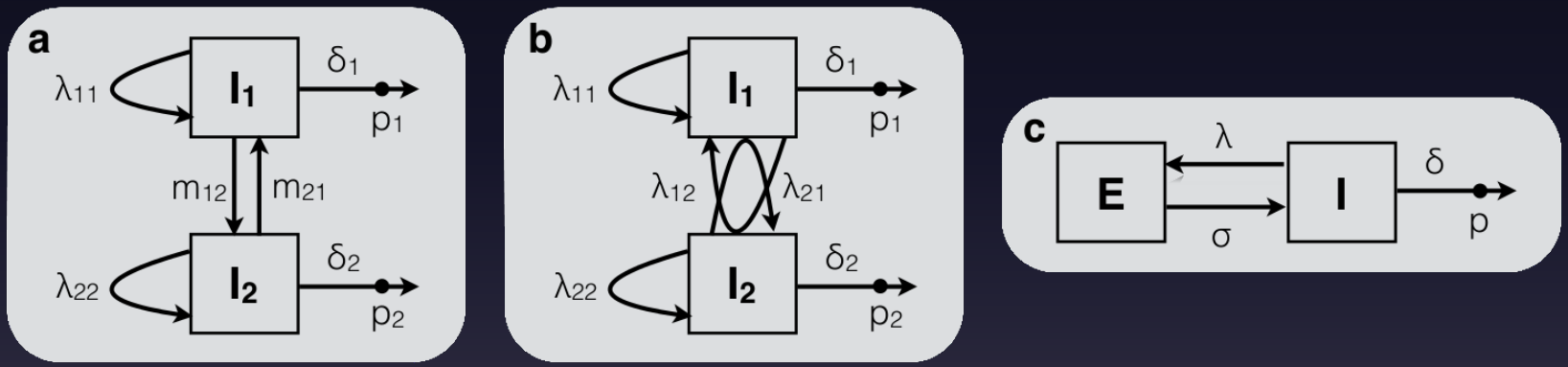
Probability of not having any sampled descendants:

$$\frac{d}{dt}p_i(t) = \mu_i - \left(\sum_{j=1}^{m}(\lambda_{i,j} + m_{i,j}) + \mu_i + \psi_i\right) p_i(t) + \sum_{j=1}^{m} m_{i,j}p_j(t) + \sum_{j=1}^{m} \lambda_{i,j}p_i(t)p_j(t)$$

Probability of having evolved as observed in the tree:

$$\frac{d}{dt}g_{Ni}(t) = -\left(\sum_{j=1}^{m}(\lambda_{i,j} + m_{i,j}) + \mu_i + \psi_i\right) g_{Ni}(t) + \sum_{j=1}^{m} m_{i,j}g_{Nj}(t) + \sum_{j=1}^{m} \lambda_{i,j}p_j(t)g_{Ni}(t) + \sum_{j=1}^{m} \lambda_{i,j}p_i(t)g_{Nj}(t)$$

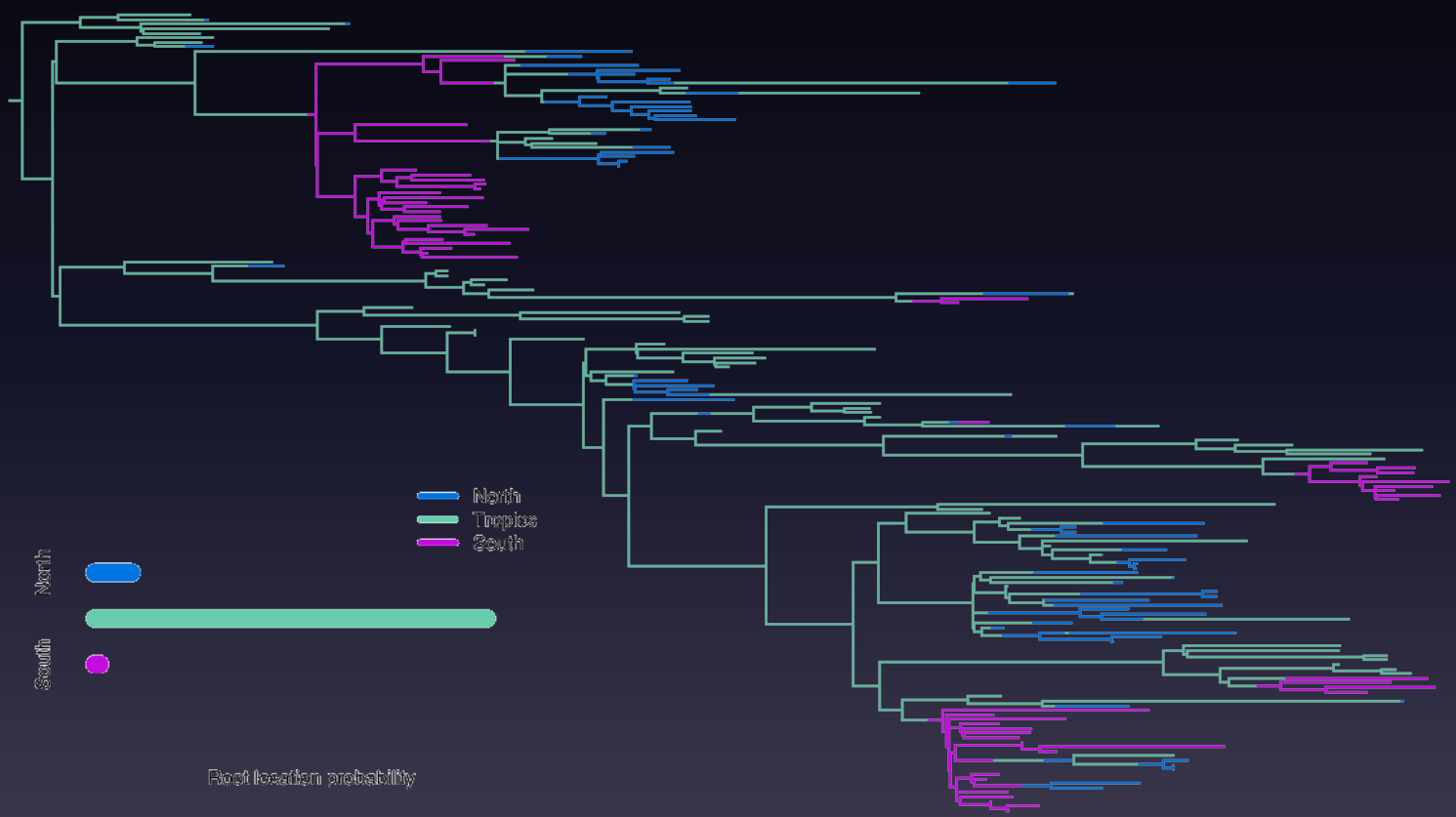
3) Examples


a
λ11
I1
δ1
p1
m12
m21
λ22
I2
δ2
p2
b
λ11
I1
δ1
p1
λ12
λ21
λ22
I2
δ2
p2
c
λ
E
I
δ
σ
p

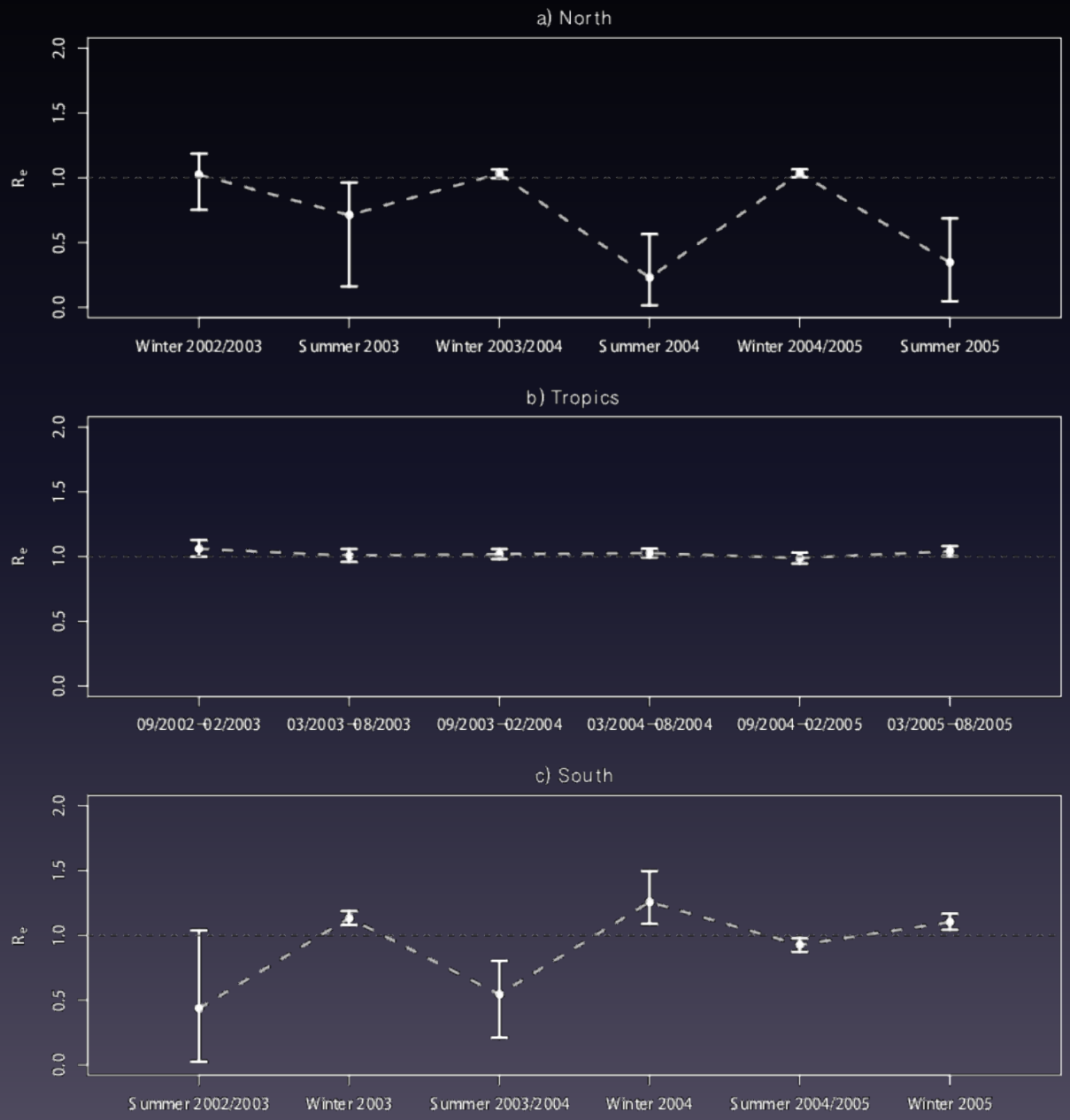
(i) Globally circulating influenza virus

- Global human influenza H3N2 virus data set
- 175 HA sequences from Northern, Southern hemisphere & Tropics
- Seasonal dynamics in temperate regions

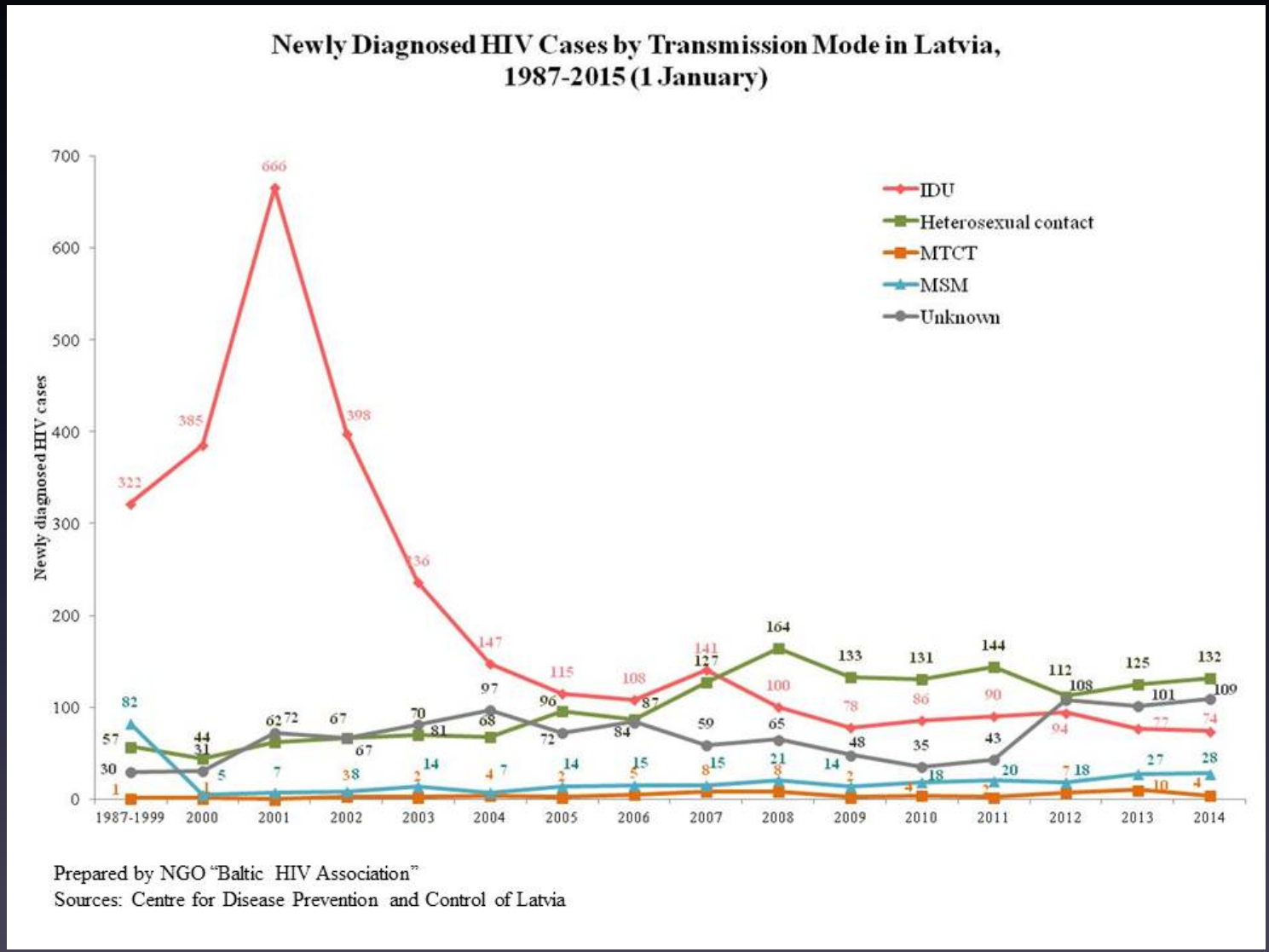
H3N2 maximum posterior tree



H3N2 seasonal transmission dynamics



(ii) HIV epidemic in Latvia



The data set

- 196 (p17) / 199 (V3) sequences of HIV subtype A from Latvia
- Sampled between 1998-2005, published by Balode et al. 2004, Graw et al. 2012
- Risk groups: heterosexuals (HET) and injecting drug users (IDU)

Epidemiological dynamics within and between HET & IDU

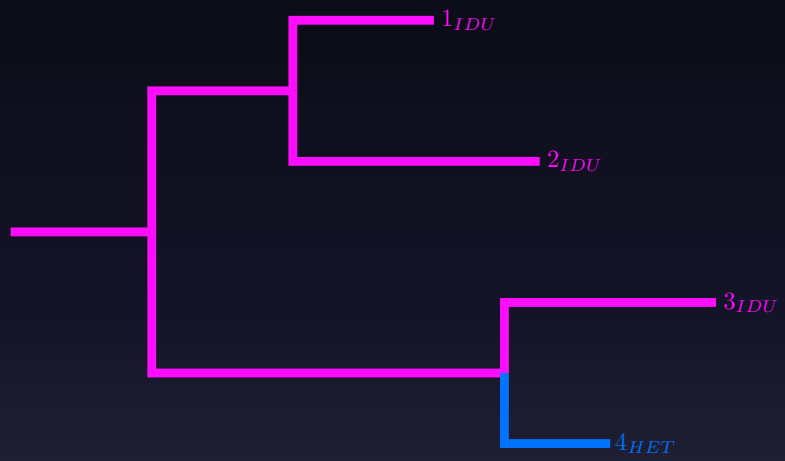
- *Is the Latvian HIV epidemic mainly driven by IDU's?*
- *Can the HET sub-epidemic persist on its own, or does it depend on IDU's?*

A tree prior for multi-type trees - Maximum Likelihood

1. Reconstruct phylogeny
2. Estimate epidemiological parameters

$$R_d = \frac{\lambda_d}{\psi_d + \mu_d}$$

$d \in$ {HET, IDU, HET>IDU, IDU>HET}

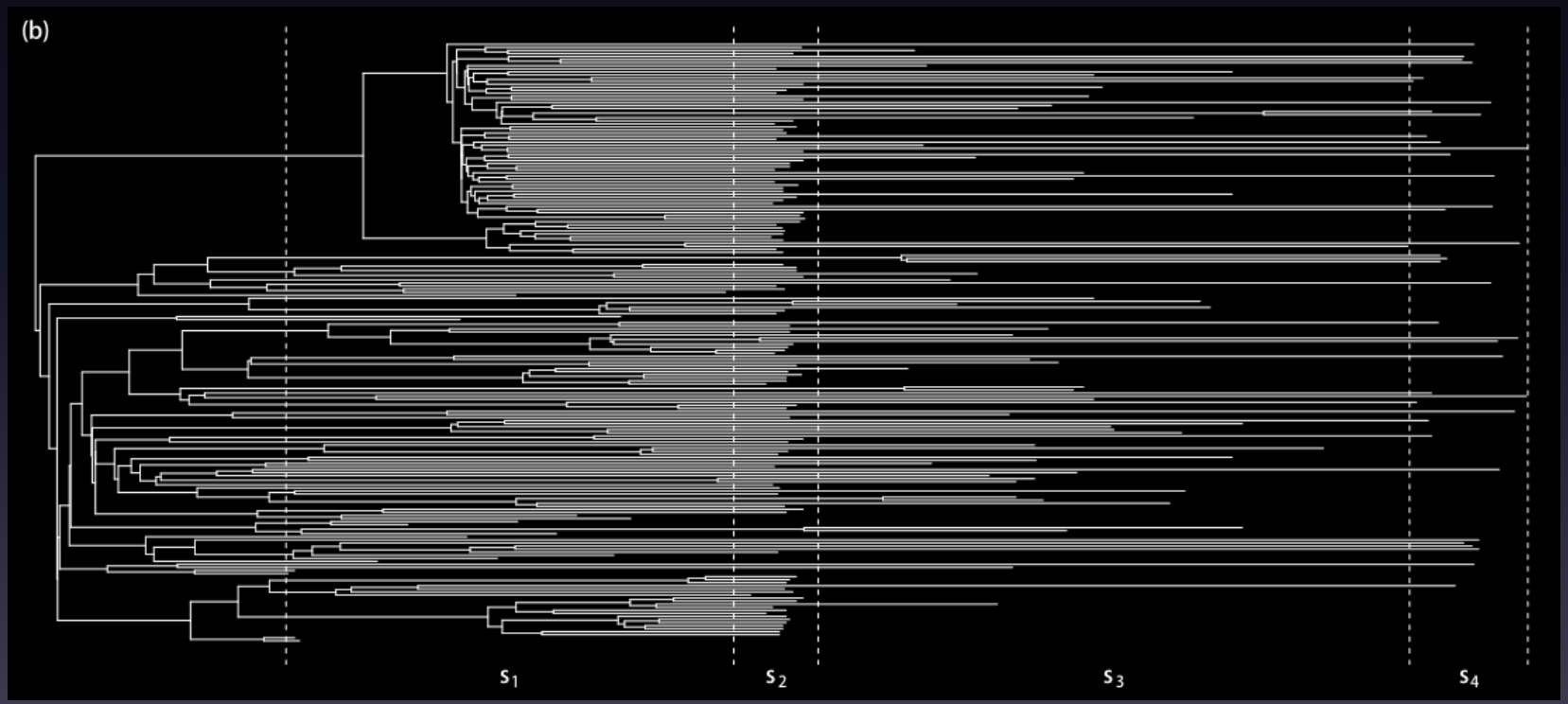


Stadler, Bonhoeffer 2013 PhilTrans, median ML estimates:

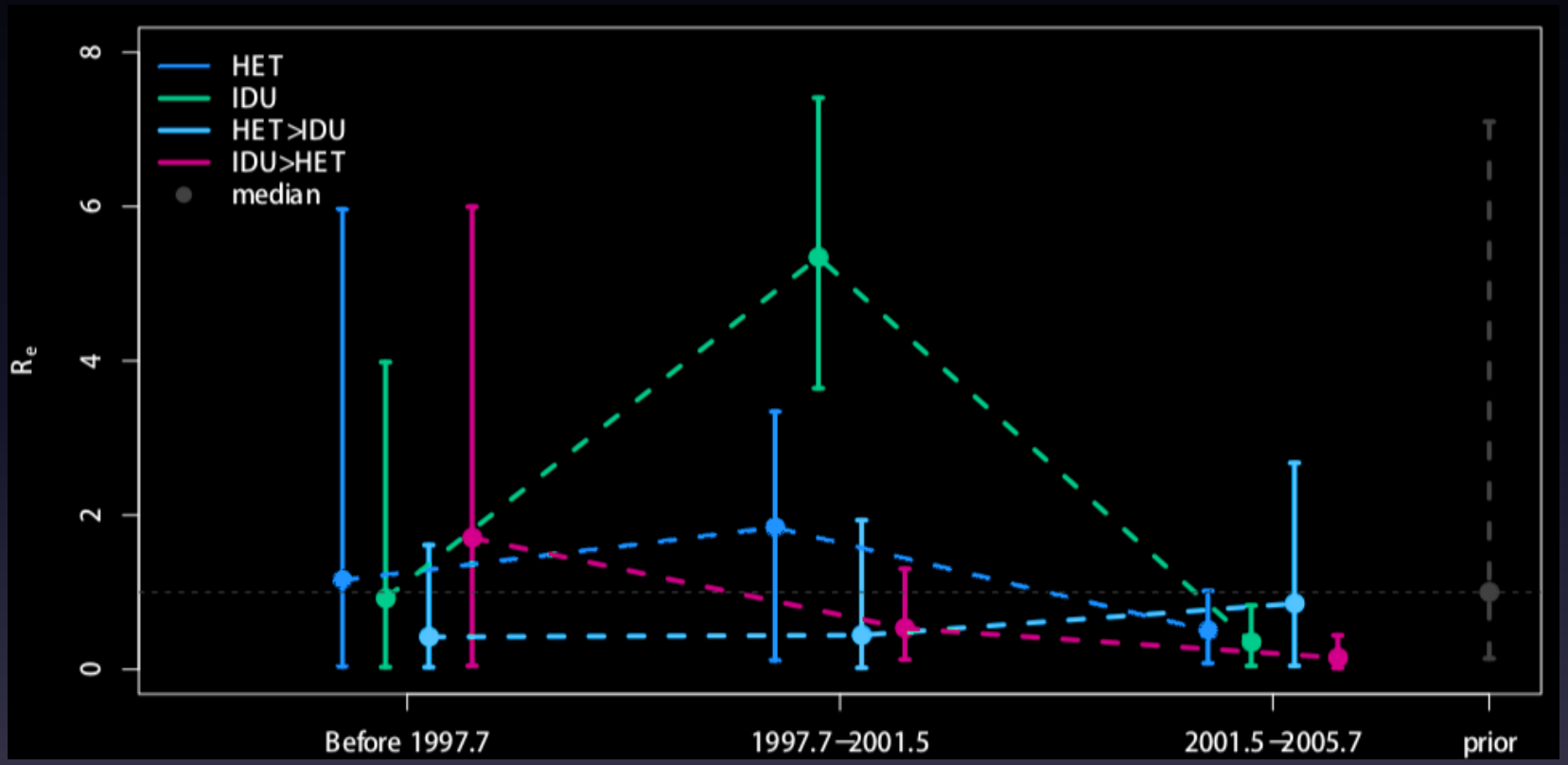
R_{HET}	R_{IDU}	$R_{H \to I}$	$R_{I \to H}$	recovery-rate
0.38	1.13	7x10-7	0.04	4.36

Sampling through time

Allow four sampling periods $s_1, \ldots, s4$ to account for changes in sampling effort through time and between risk groups:



Reproduction number through time



- median recovery rate : 0.46

Conclusion - HIV risk group analysis

- Sampling process can bias estimates of epidemiological parameters
- V3 and p17 results agree with each other
- Epidemic mainly driven by IDU risk group
- Significant transmission from HET to IDU can not be ruled out

Phylodynamics of the ebola virus outbreak in Sierra Leone, 2014

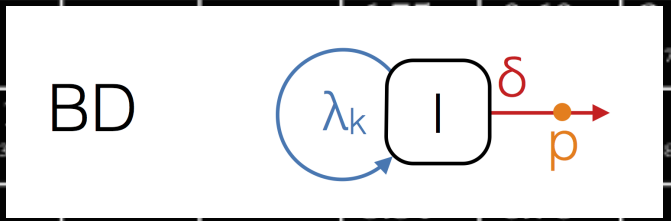
Sciencexpress

Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak

Stephen K. Gire,[1,2*] Augustine Goba,[3*†] Kristian G. Andersen,[1,2*†] Rachel S. G. Sealfon,[2,4*] Daniel J. Park,[2*] Lansana Kanneh,[3] Simbirie Jalloh,[3] Mambu Momoh,[3,5] Mohamed Fullah,[3,5‡] Gytis Dudas,[6] Shirlee Wohl,[1,2,7] Lina M. Moses,[8] Nathan L. Yozwiak,[1,2] Sarah Winnicki,[1,2] Christian B. Matranga,[2] Christine M. Malboeuf,[2] James Qu,[2] Adrianne D. Gladden,[2] Stephen F. Schaffner,[1,2] Xiao Yang,[2] Pan-Pan Jiang,[1,2] Mahan Nekoui,[1,2] Andres Colubri,[1] Moinya Ruth Coomber,[3] Mbalu Fonnie,[3‡] Alex Moigboi,[3‡] Michael Gbakie,[3] Fatima K. Kamara,[3] Veronica Tucker,[3] Edwin Konuwa,[3] Sidiki Saffa,[3] Josephine Sellu,[3] Abdul Azziz Jalloh,[3] Alice Kovoma,[3‡] James Koninga,[3] Ibrahim Mustapha,[3] Kanc
Momoh Foday,[3] Mohamed Yillah,[3] Franklyn Kanneh,[3] Willie F
James L. B. Massally,[3] Sinéad B. Chapman,[2] James Bochicc
Murphy,[2] Chad Nusbaum,[2] Sarah Young,[2] Bruce W. Birren,[2]
Grant,[3] John S. Scheiffelin,[8] Eric S. Lander,[2,7,9] Christian Ha
M. Gevao,[11] Andreas Gnirke,[2§] Andrew Rambaut,[6,12,13§] Rob
Garry,[8§] S. Humarr Khan,[3‡§] Pardis C. Sabeti[1,2†§]

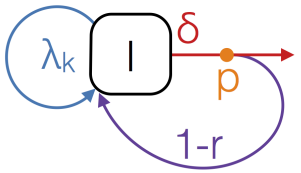
Pawel Gaul / iStockPhoto

	R_0/R_e initial	R_e middle	R_e recent	Days exposed	Days infected	Sampling	Sampling Ancestors	r	origin	MRCA
BD 1	1.65 (1.02-2.70)	-	-	-	6.09 (2.84-18.84)	0.65 (0.20-1.00)	-	-	7.5. (7.4.-22.5.)	15.5. (3.5.-22.5.)
BD 3	0.95 (0.22-2.56)	1.57 (0.73-2.91)	1.81 (1.07-3.03)	-	6.15 (3.22-	0.70 (0.27-1.00)	-	-	8.4. (30.12.-21.5.)	12.5. (24.4.-23.5.)
BDsa 1	1.75 (1.04-2.95)	-						0.93 (0.71-1)	8.5. (10.4-22.5.)	15.5. (3.5.-23.5.)
BDsa 3	0.96 (0.20-2.65)	1.6 (0.74-						0.93 (0.70-1)	9.4. (31.12.-20.5.)	12.5. (24.4-23.5.)
BDSIR	1.81 (1.12-2.84)	-			(3.61-18.78)	(0.24-1.00)		-	4.5. (11.4-19.5.)	15.5. (3.5-22.5.)
BDEI 1*	**2.18 (1.24-3.55)**	-	-	**4.92 (2.11-23.20)**	**2.58 (1.24-6.98)**	**0.71 (0.62-0.80)**	-	-	**8.5. (10.4-21.5.)**	**14.5. (3.5-22.5.)**
BDEI 3*	2.00 (0.66-5.46)	1.85 (0.57-3.71)	3.15 (1.43-6.09)	5.92 (2.49-24.92)	2.71 (1.28-9.22)	0.71 (0.63-0.80)	-	-	5.5. (3.4.-21.5.)	13.5. (30.4.-22.5.)



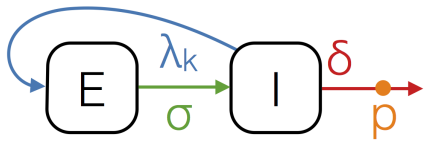
	R_0/R_e initial	R_e middle	R_e recent	Days exposed	Days infected	Sampling	Sampling Ancestors	r	origin	MRCA
BD 1	1.65 (1.02-2.70)	-	-	-	6.09 (2.84-18.84)	0.65 (0.20-1.00)	-	-	7.5. (7.4.-22.5.)	15.5. (3.5.-22.5.)
BD 3	0.95 (0.22-2.56)	1.57 (0.73-2.91)	1.81 (1.07-3.03)	-	6.15 (3.22-17.04)	0.70 (0.27-1.00)	-	-	8.4. (30.12.-21.5.)	12.5. (24.4.-23.5.)
BDsa 1	1.75 (1.04-2.95)	-	-	-	6.75 (3.14-24.10)	0.60 (0.17-1.00)	2 (0-7)	0.93 (0.71-1)	8.5. (10.4-22.5.)	15.5. (3.5.-23.5.)
BDsa 3	0.96 (0.20-2.65)	1.61 (0.74-3.00)	1.88 (1.09-3.23)	-	6.54 (3.24-22.14)	0.65 (0.19-1.00)	2 (0-8)	0.93 (0.70-1)	9.4. (31.12.-20.5.)	12.5. (24.4-23.5.)
BDSIR	1.81 (1.12-2.84)	-	-	-	6.64 (3.61-[illegible])	0.70 (0.24-1.00)	-	-	4.5. (11.4-19.5.)	15.5. (3.5-22.5.)
BDEI 1*	**2.18 (1.24-3.55)**	-						-	**8.5. (10.4-21.5.)**	**14.5. (3.5-22.5.)**
BDEI 3*	2.00 (0.66-5.46)	1.8[illegible] (0.57-3.71)	(1.43-6.09)	(2.49-24.92)	(1.28-9.22)	(0.63-0.80)		-	5.5. (3.4.-21.5.)	13.5. (30.4.-22.5.)

BDsa



	R_0/R_e initial	R_e middle	R_e recent	Days exposed	Days infected	Sampling	Sampling Ancestors	r	origin	MRCA
BD 1	1.65 (1.02-2.70)	-	-	-	6.09 (2.84-	0.65 (0.20-1.00)	-	-	7.5. (7.4.-22.5.)	15.5. (3.5.-22.5.)
BD 3	0.95 (0.22-2.56)	1.5 (0.73-						-	8.4. (30.12.-21.5.)	12.5. (24.4.-23.5.)
BDsa 1	1.75 (1.04-2.95)	-						0.93 (0.71-1)	8.5. (10.4-22.5.)	15.5. (3.5.-23.5.)
BDsa 3	0.96 (0.20-2.65)	1.6 (0.74-						0.93 (0.70-1)	9.4. (31.12.-20.5.)	12.5. (24.4.-23.5.)
BDSIR	1.81 (1.12-2.84)	-			18.78)			-	4.5. (11.4-19.5.)	15.5. (3.5-22.5.)
BDEI 1*	**2.18 (1.24-3.55)**	-	-	**4.92 (2.11-23.20)**	**2.58 (1.24-6.98)**	**0.71 (0.62-0.80)**	-	-	**8.5. (10.4-21.5.)**	**14.5. (3.5-22.5.)**
BDEI 3*	2.00 (0.66-5.46)	1.85 (0.57-3.71)	3.15 (1.43-6.09)	5.92 (2.49-	2.71 (1.28-9.22)	0.71 (0.63-0.80)	-	-	5.5. (3.4.-21.5.)	13.5. (30.4.-22.5.)

BDEI



- λ — infection rate
- σ — incubation rate
- δ — becoming-noninfectious rate
- p — sampling probability

Summary phylodynamics of ebola virus

- Estimation of sampling proportion possible
  $\Rightarrow$ Estimate unobserved cases
- Average infectious period is short (2-3 days)
- Duration of incubation period difficult to estimate
- Weak signal for superspreading

Thank you for your attention

ETH Zurich:
Tanja Stadler
Sebastian Bonhoeffer
David Rasmussen
Louis du Plessis

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

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