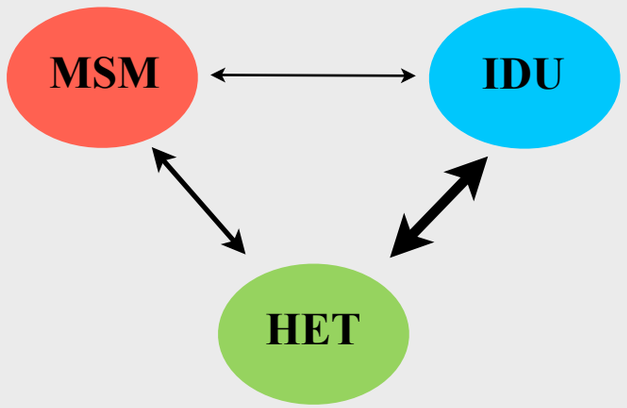
Recovering transmission structure and dynamics from viral sequence data

Tanja Stadler, Gabriel Leventhal, Sebastian Bonhoeffer
Institute of Integrative Biology, ETH Zurich

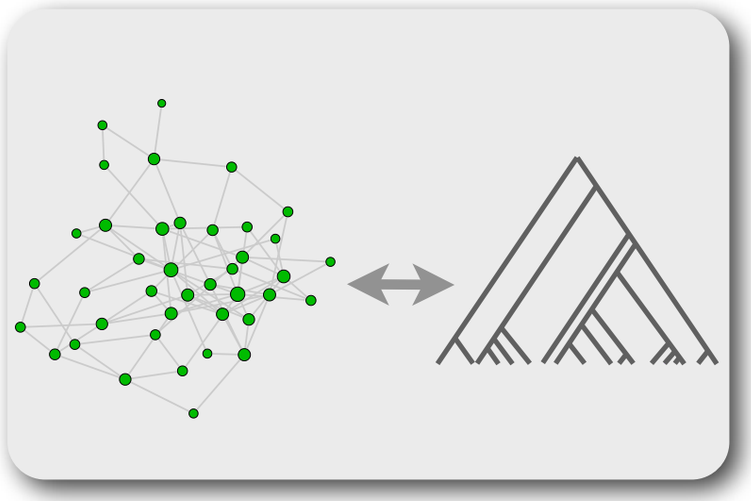
Roger Kouyos
University Hospital Zurich

Outline



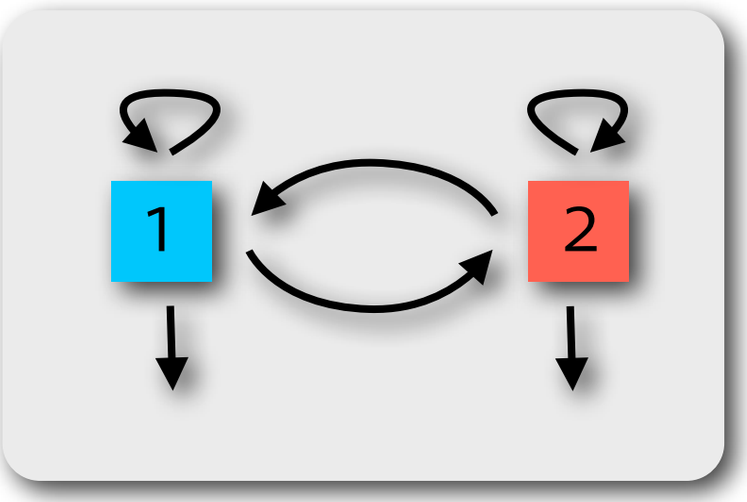
Introduction:
Risk group structure of HIV transmission in Switzerland

(Kouyos et al, JID, 2010)



Part I:
Inferring contact structure from phylogenetic trees

(Leventhal et al, PLoS Comp Biol, 2012)



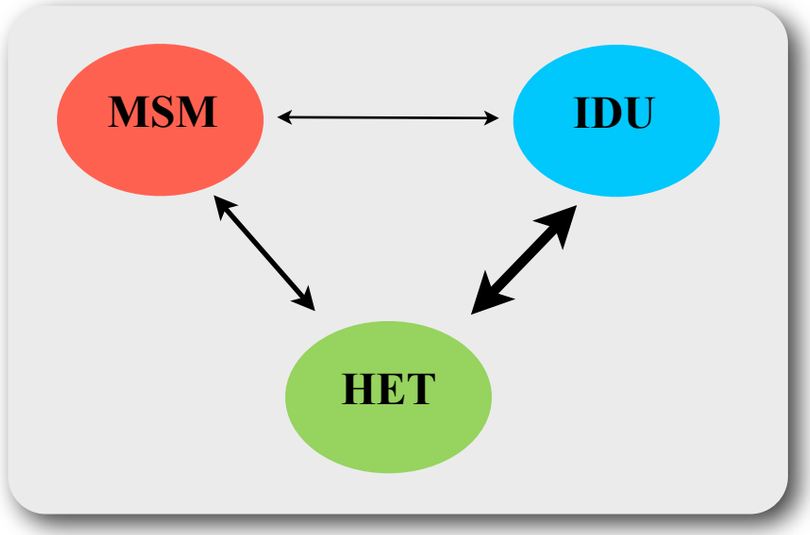
Part II:
Identifying and quantifying transmission structure

(Stadler & Bonhoeffer, Phil Trans 2013)

Introduction:
HIV transmission within and between risk groups in Switzerland

1. How much HIV transmission occurs within versus between risk groups?

2. Are there temporal trends in transmission between risk groups?



(Kouyos et al, JID, 2010)

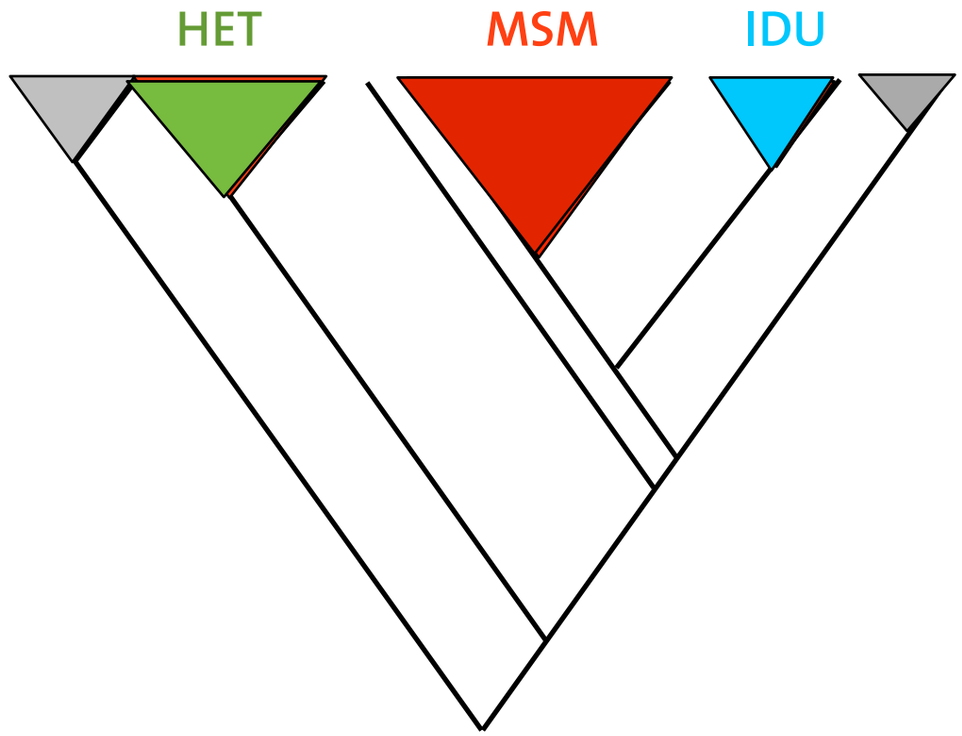
Data

- **HIV in Switzerland**
  - ~25000 HIV infections (mainly subtype B) / 7 Million inhabitants
  - Main transmission routes:
    - Men that have Sex with Men (MSM)
    - Intravenous Drug Users (IDU)
    - Heterosexual Contact (HET)

- **Swiss HIV Cohort Study (SHCS):**
  - Clinical and demographic data of >50% of HIV infections in Switzerland

Approach

I. **Construct tree with “Swiss” and “non-Swiss” sequences**
   - 5700 seqs from Switzerland (SHCS)
   - 5700 seqs from elsewhere (Los Alamos)



Key results

- **Clear evidence of transmission structure**
  - Two largely independent epidemics (HET/IDU and MSM)

- **No larger cluster is uniquely dominated by heterosexuals**
  - HET epidemics largely driven by IDUs

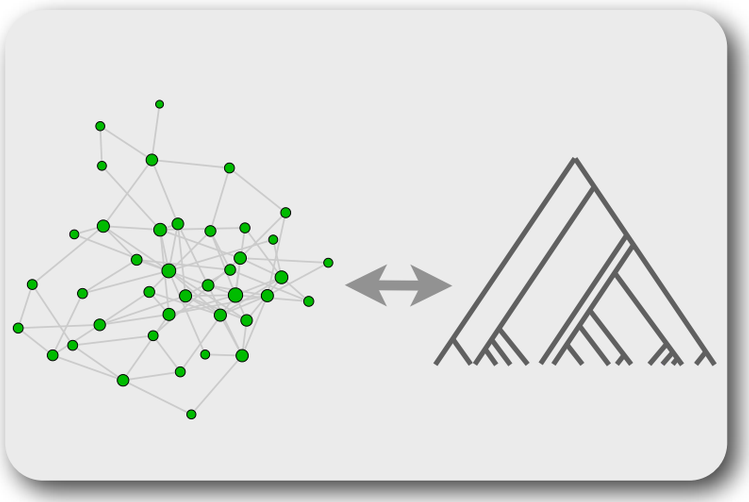
- **Strong reduction of impact of IDU on HET over time**
  - Heterosexuals have profited indirectly from prevention measures targeted at IDUs

Questions

- Can we infer contact structure from phylogenetic trees?
- Can we quantify transmission between risk groups using phylogenetic methods?

Part I:
Inferring contact structure from phylogenetic trees

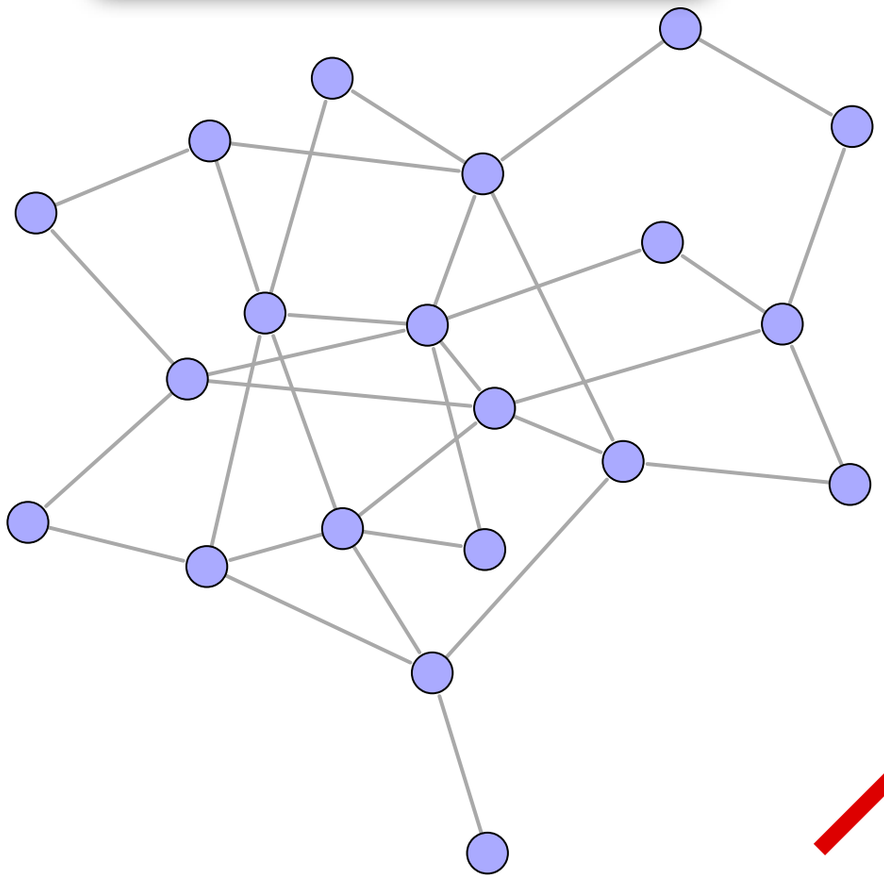
What does the phylogenetic tree tell us about host contact structure?



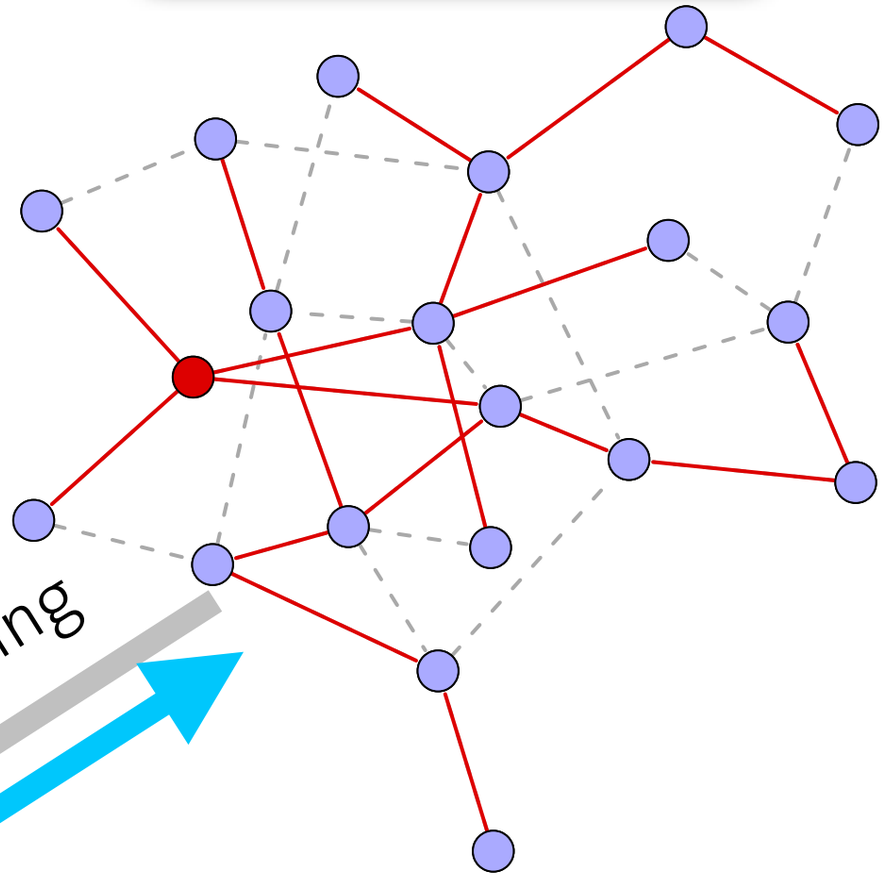
(Leventhal et al, PLoS Comp Biol, 2012)

The problem

Contact network

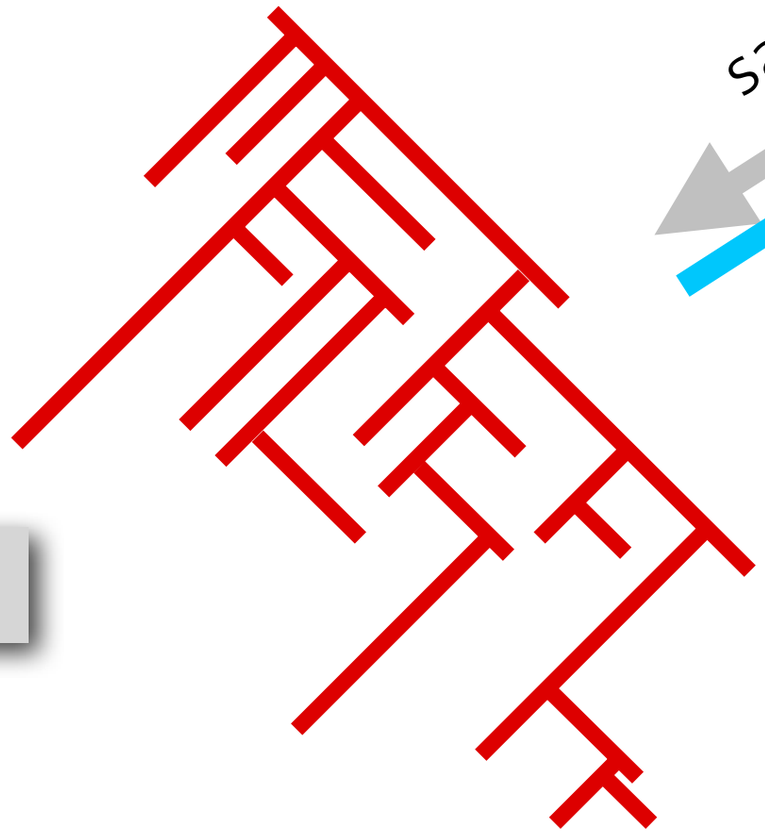


infection



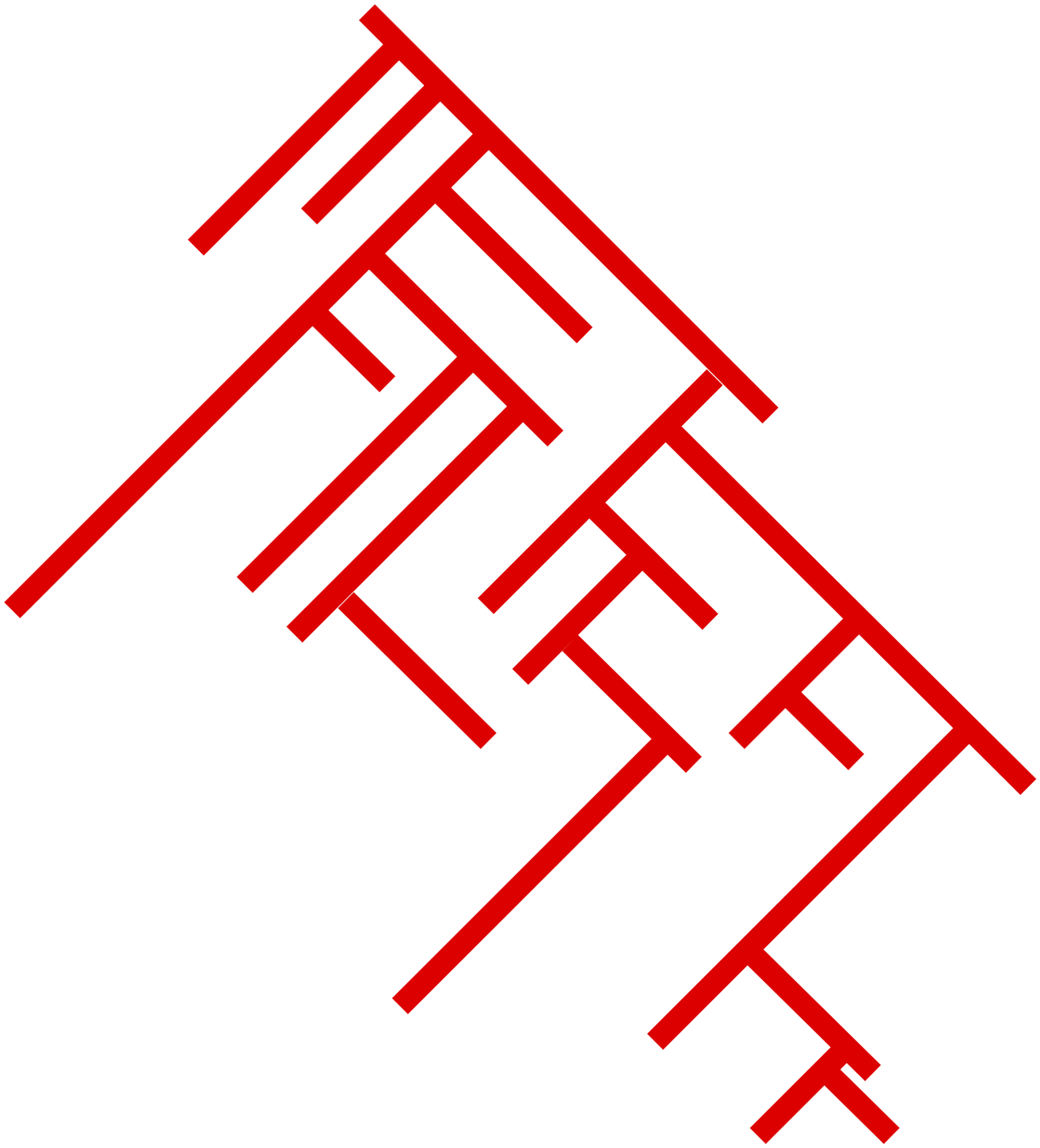
Transmission tree

sampling



Phylogenetic tree

Which property reveals contact structure?



Measures:

Branching topology?

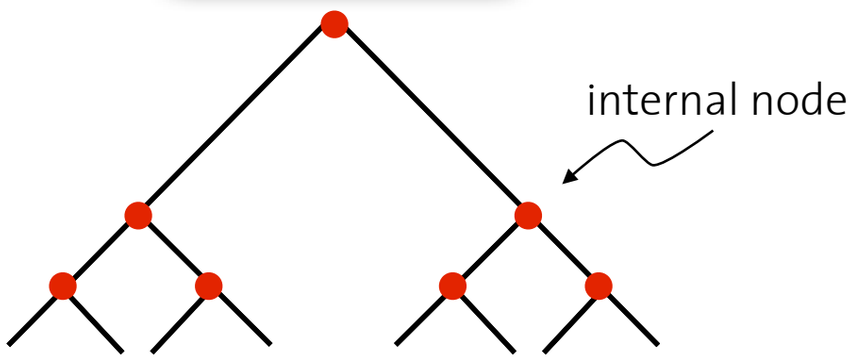
Branch length?

We use:

Balancedness

Sackin index

balanced

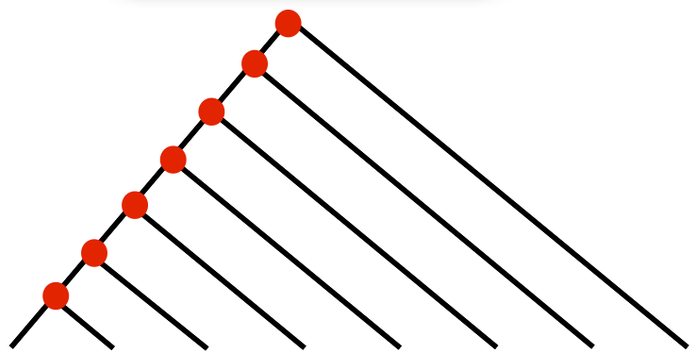


3 + 3 + 3 + 3 + 3 + 3 + 3 + 3 = 24

Sackin index:

Sum over all internal nodes from all tips to the root

unbalanced



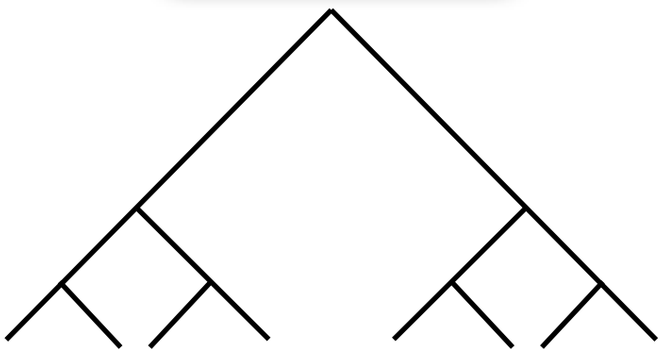
7 + 7 + 6 + 5 + 4 + 3 + 2 + 1 = 35

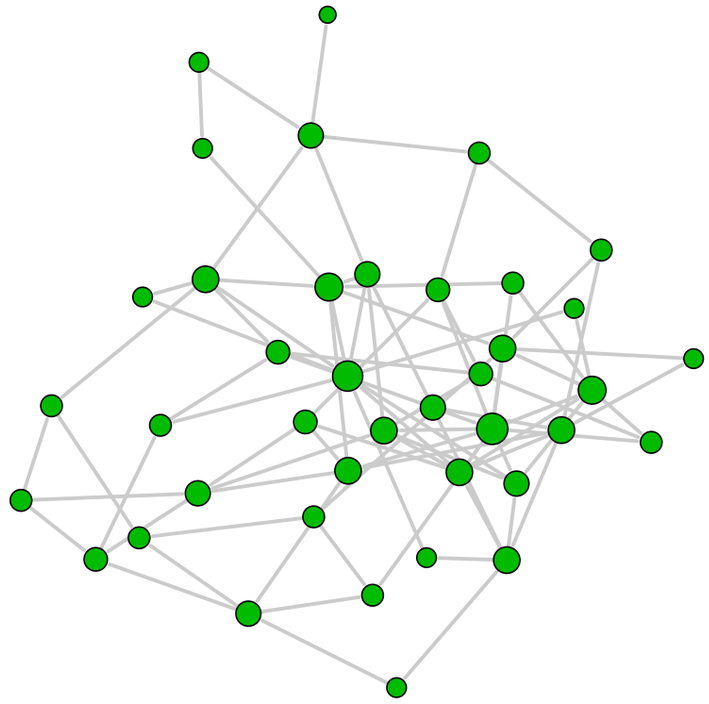
Normalized Sackin index:

- deviation of from expected Sackin index for epidemic with complete mixing
- normalized with regard to number of tips

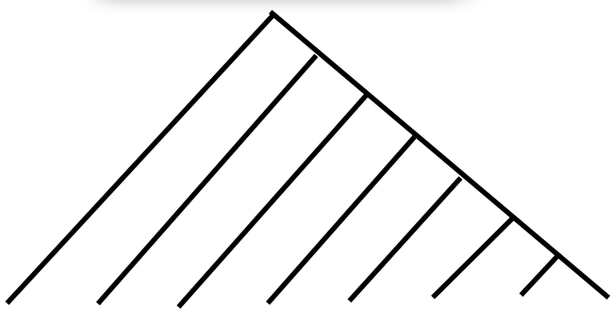
Contact structure and tree balance

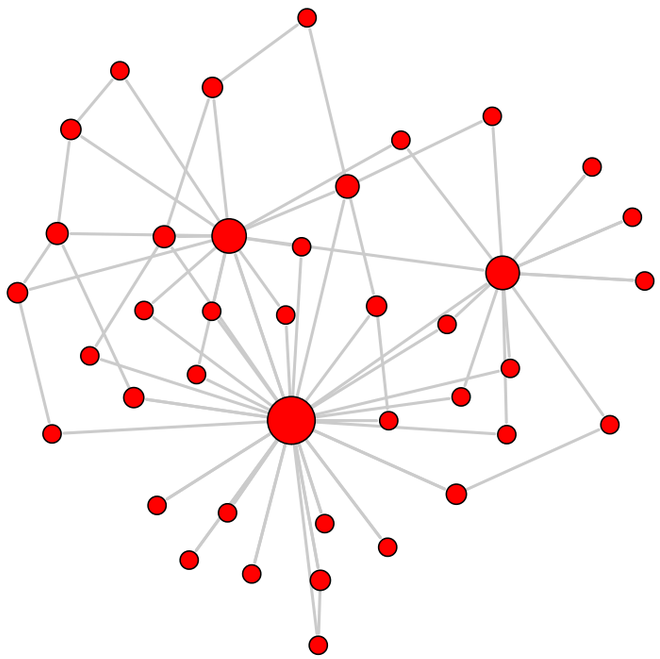
balanced





unbalanced

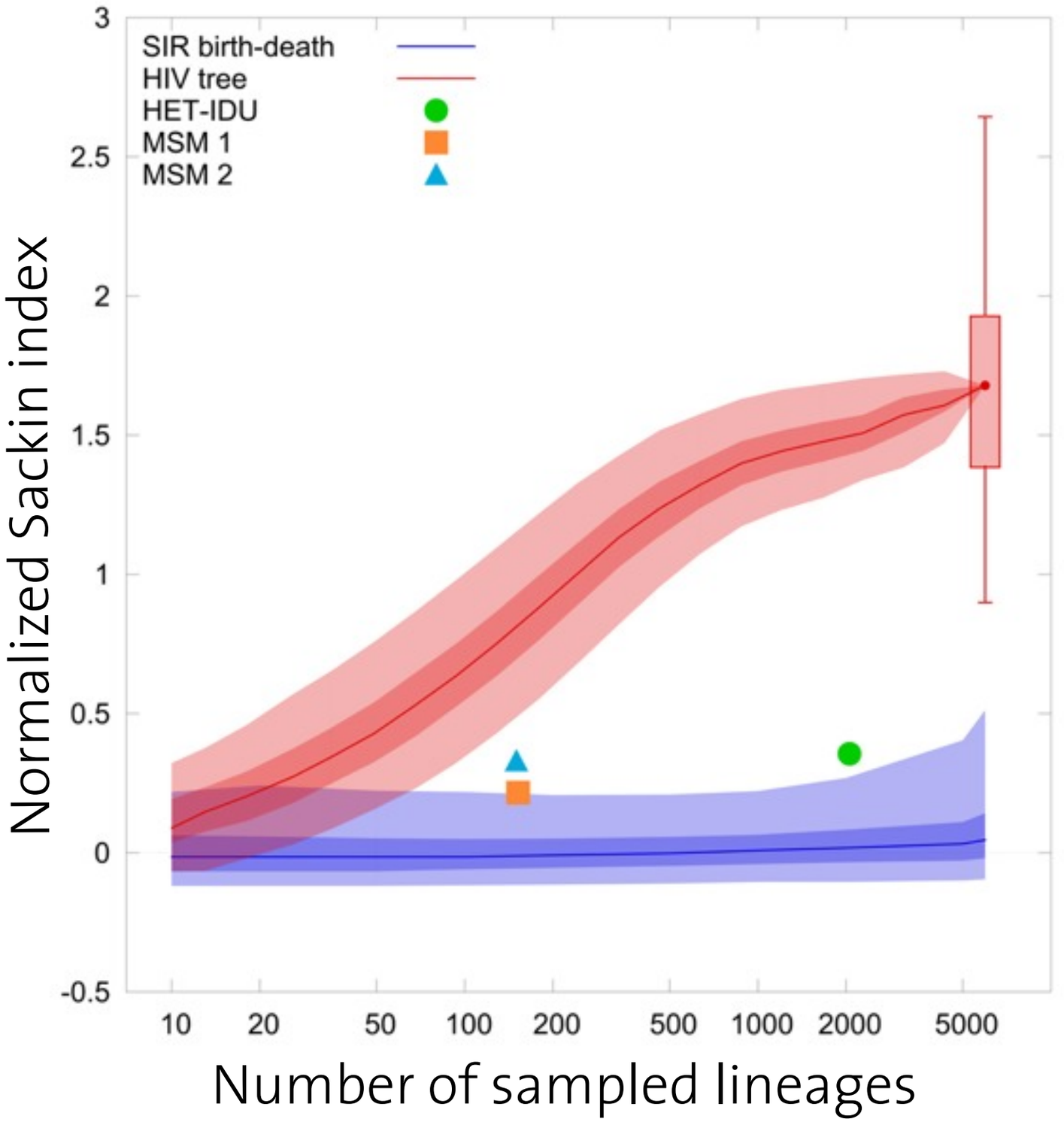




Swiss HIV epidemic

- HIV tree is much more unbalanced then expected under random mixing

- Transmission group structure does not account for all of the unbalancedness



(Leventhal et al, PLoS Comp. Biol.. 2012)

Summary

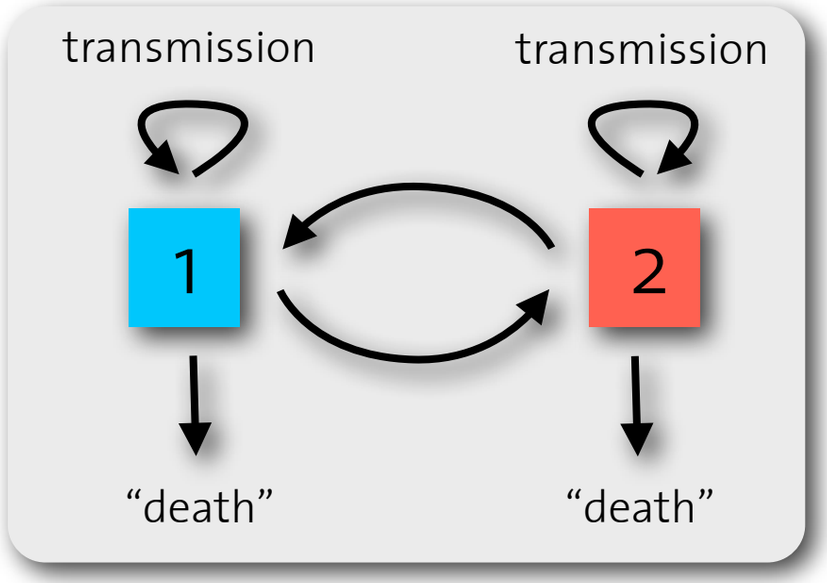
- **Key results**
  - Tree-balancedness can be used to distinguish between alternative models of contact structure
    - Refute random mixing model for HIV transmission
    - Evidence for non-random contact structure even within HIV transmission groups
  - We cannot directly infer the contact structure, but we test whether null models of contact structure can be rejected statistically

- **Outlook**
  - Use branch length in addition to tree topology to distinguish between models of contact structure (=> Gabriel Leventhal)

Part II:
Identification and quantification of transmission structure

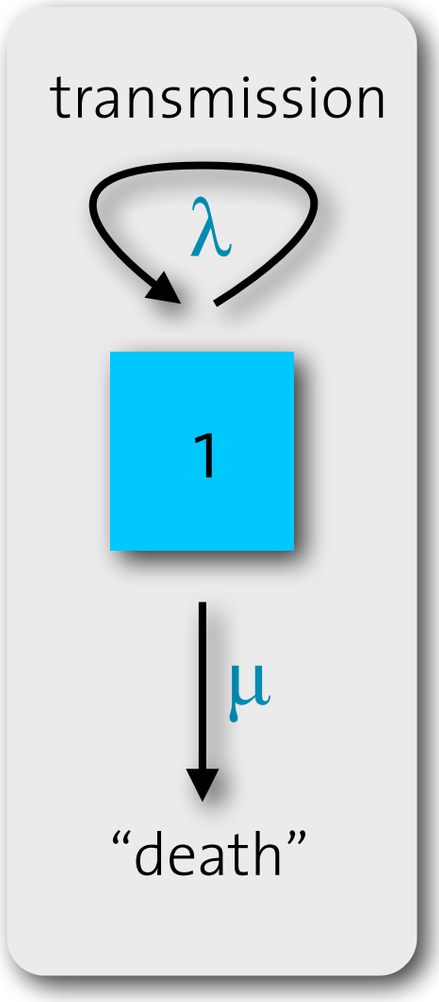
Can we quantify transmission between risk groups ?

Can we identify and quantify super-spreading?



(Stadler & Bonhoeffer, Phil Trans 2013)

Birth-death model



- “Death” <=> Any process of becoming non-infectious such as death or treatment (sampled or not sampled)
- “Birth” <=> Transmission

- Birth-death model / BEAST
  - Inference of epidemiological parameters such as R_0

Stadler et al, Mol. Biol. Evol. 2012

- Assumption: Parameters constant over time

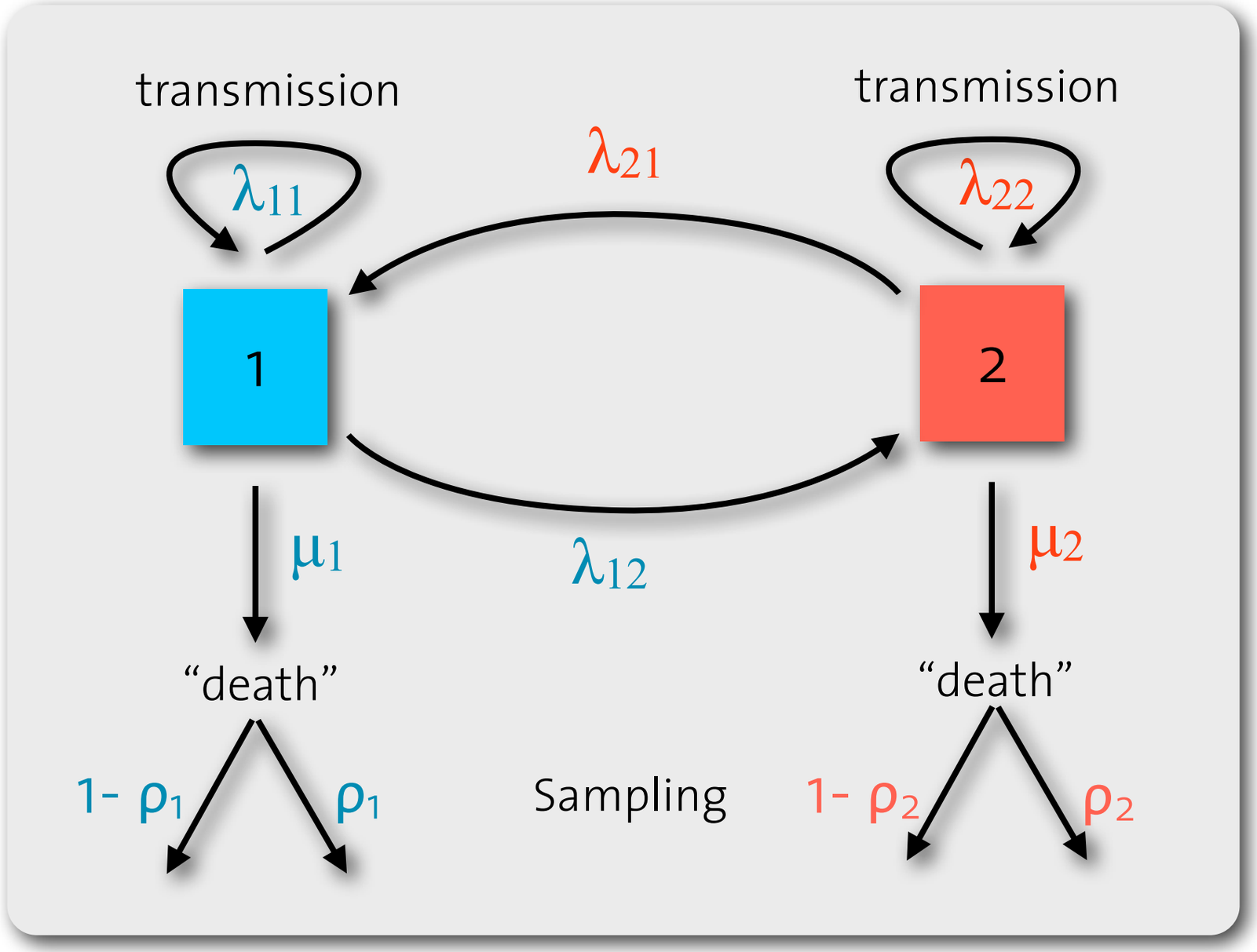
➡ Birth-death skyline plot

Stadler, Kuehnert, Bonhoeffer, Drummond, PNAS 2013

➡ Extend to SIR models

Leventhal et al, submitted

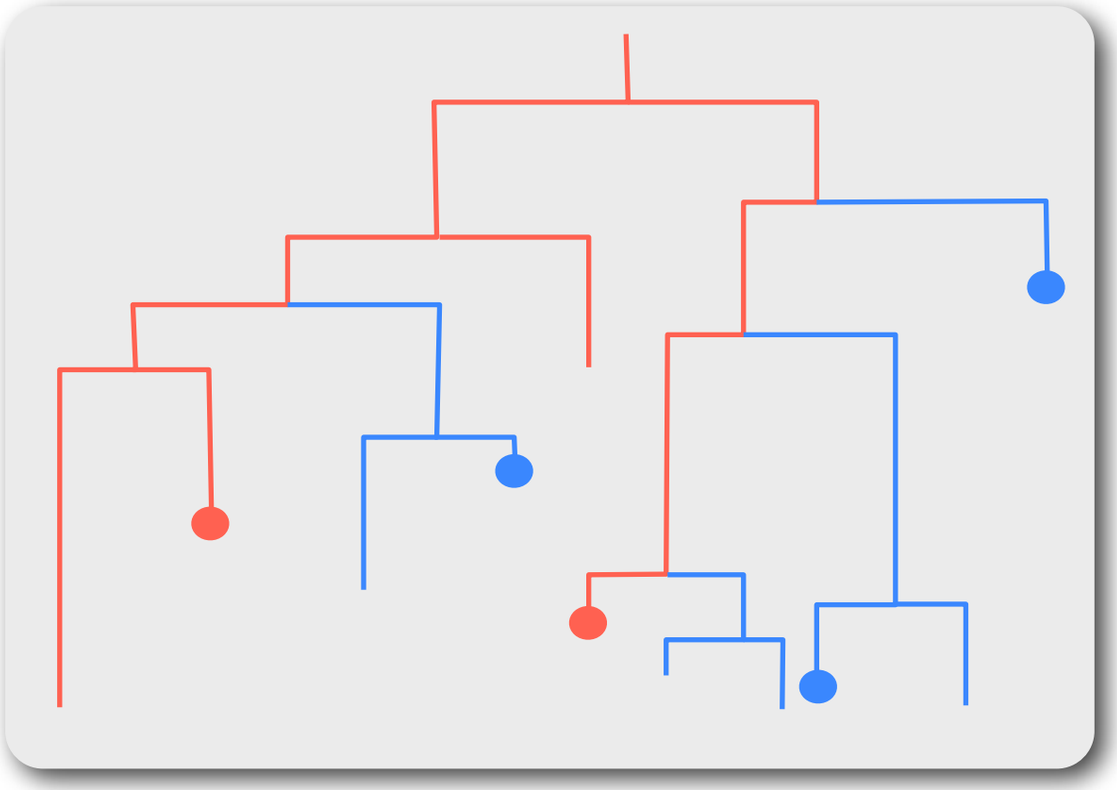
Structured birth-death model



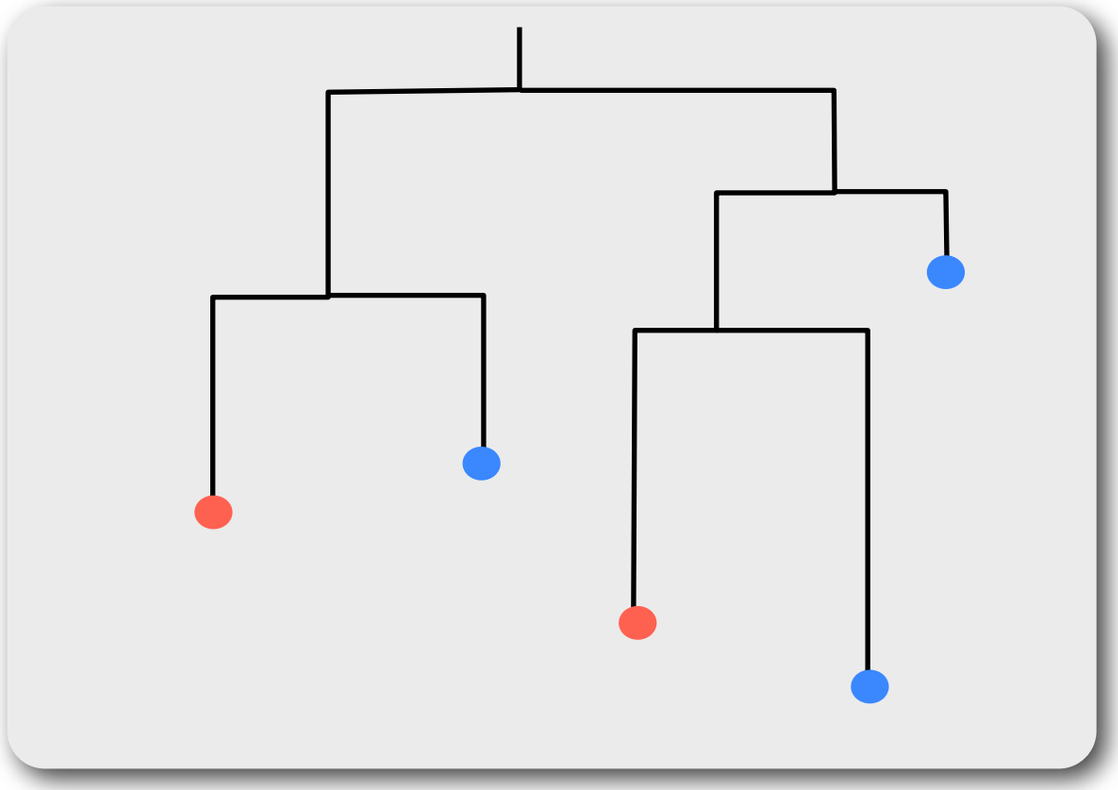
- Multitype branching process:
- Can be extended to more than 2 populations

Infection tree and reconstructed tree

Infection tree



Reconstructed tree



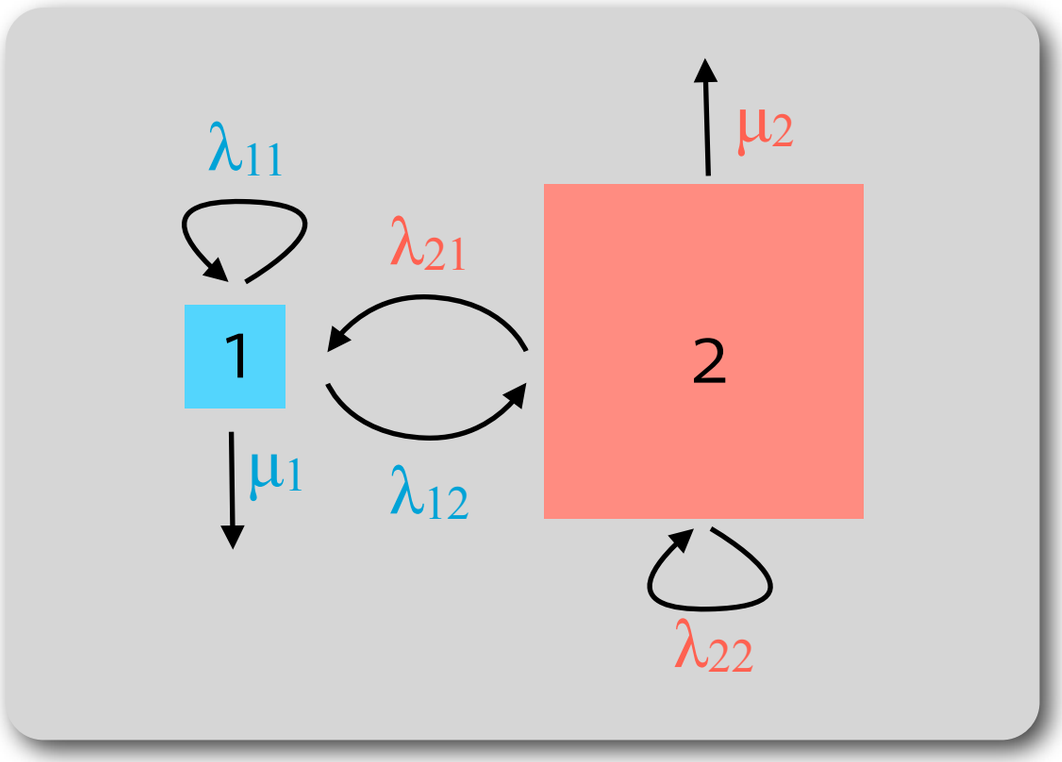
Maximum likelihood approach:

- Determine probability of observing exactly the reconstructed tree given parameters for the structured birth death model
- Determine parameters maximizing this probability

Simulated data

Simulation:

Two populations with different size and parameters



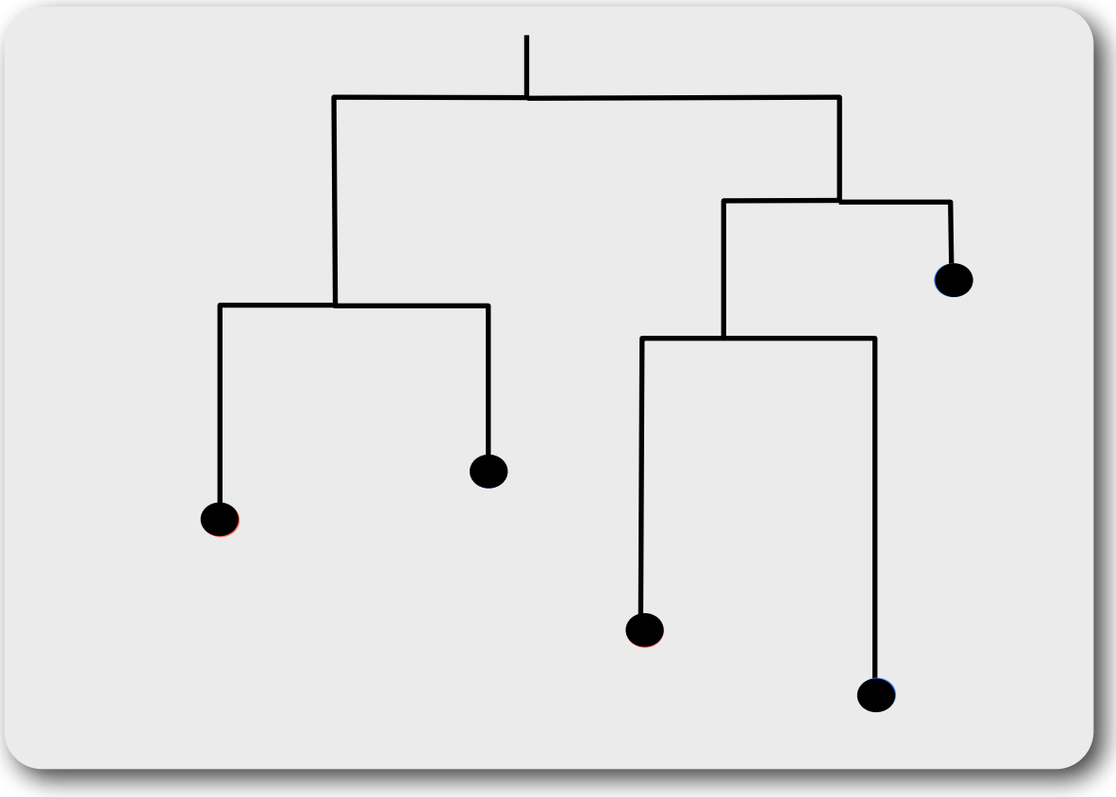
	true	median	2.5%	97.5%
λ_{11}	15.00	13.65	0.00	17.01
λ_{12}	3.00	3.04	0.00	17.09
λ_{21}	5.00	6.03	0.01	15.27
λ_{22}	7.00	7.56	0.00	12.04
μ_1	6.00	6.70	4.31	20.39
μ_2	2.00	1.51	0.62	8.64

Fixed $\rho = 0.25$

Results:

- Method infers correct parameters
- If data comes from a **single homogeneous** population, then the method **accepts** a **homogeneous** model (90 %)
- If data comes from a **heterogenous** population, then the method **rejects** a **homogeneous** model (52 %)

Simulated data: Unknown patient status



	true	median	2.5	97.5
λ_{11}	2.00	2.30	1.64	4.72
λ_{12}	20.00	19.13	8.00	25.43
λ_{21}	0.10	0.12	0.05	1.01
λ_{22}	1.00	1.03	0.30	2.04
μ	2.00	1.88	1.00	2.33

Fixed $\rho = 0.25$

Results:

- Method can detect population heterogeneity
- If data comes from a **single homogeneous** population, then the method **accepts** a **homogeneous** model (70 %)
- If data comes from a **heterogenous** population, then the method **rejects** a **homogeneous** model (99 %)
- acceptance/rejection depends on parameter choice

Data: HIV epidemic in Latvia

Data from *Balode et al, AIDS Res Hum Retrovir 2012*

Latvia

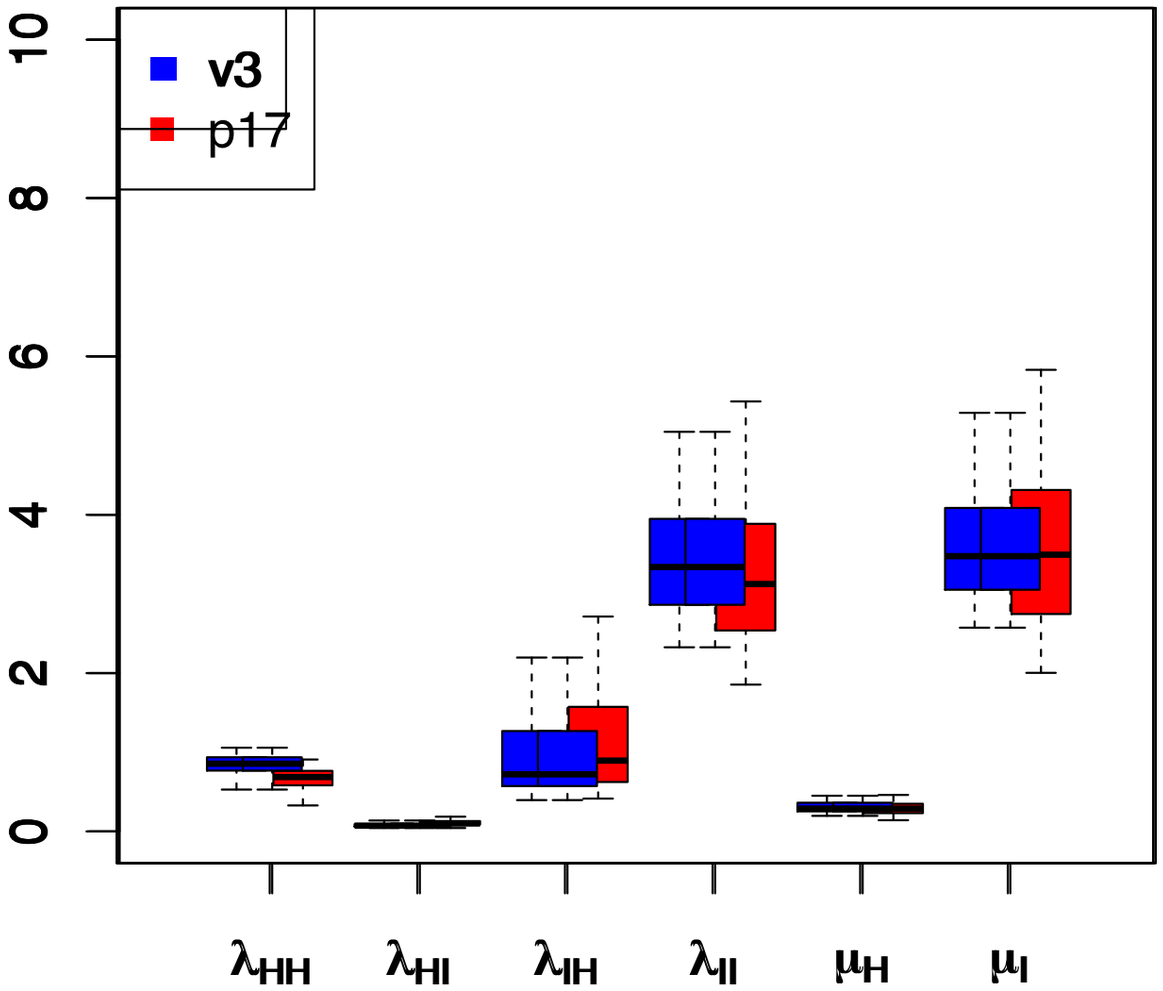
- Estimated total number of HIV infecteds:
  - 4'600 (Balode)
  - 8'000 - 9'000 (CIA Fact Sheet ~ prev 0.7%)

- Patient sequences
  - 65 subtype B (mostly MSM)
    - we take 45 seqs of Clade 1, all MSM
  - 230 subtype A (mostly HET and IDU)
    - we take 220 seqs all HET/IDU
  - *gag* p17 and *env* V3

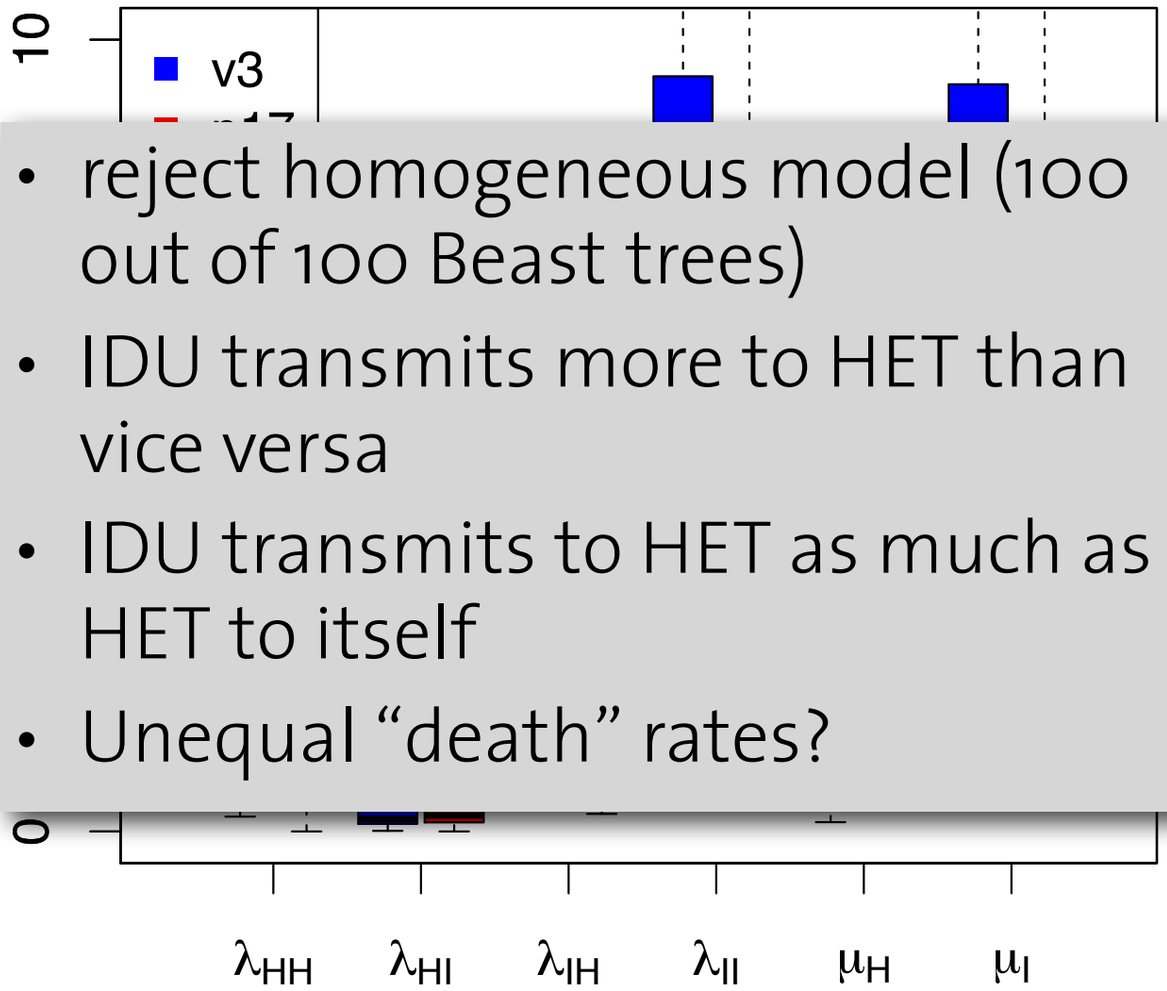
Inference on Latvian data

H=HET; I = IDU

Fraction sampled: 10%



Fraction sampled: 1%

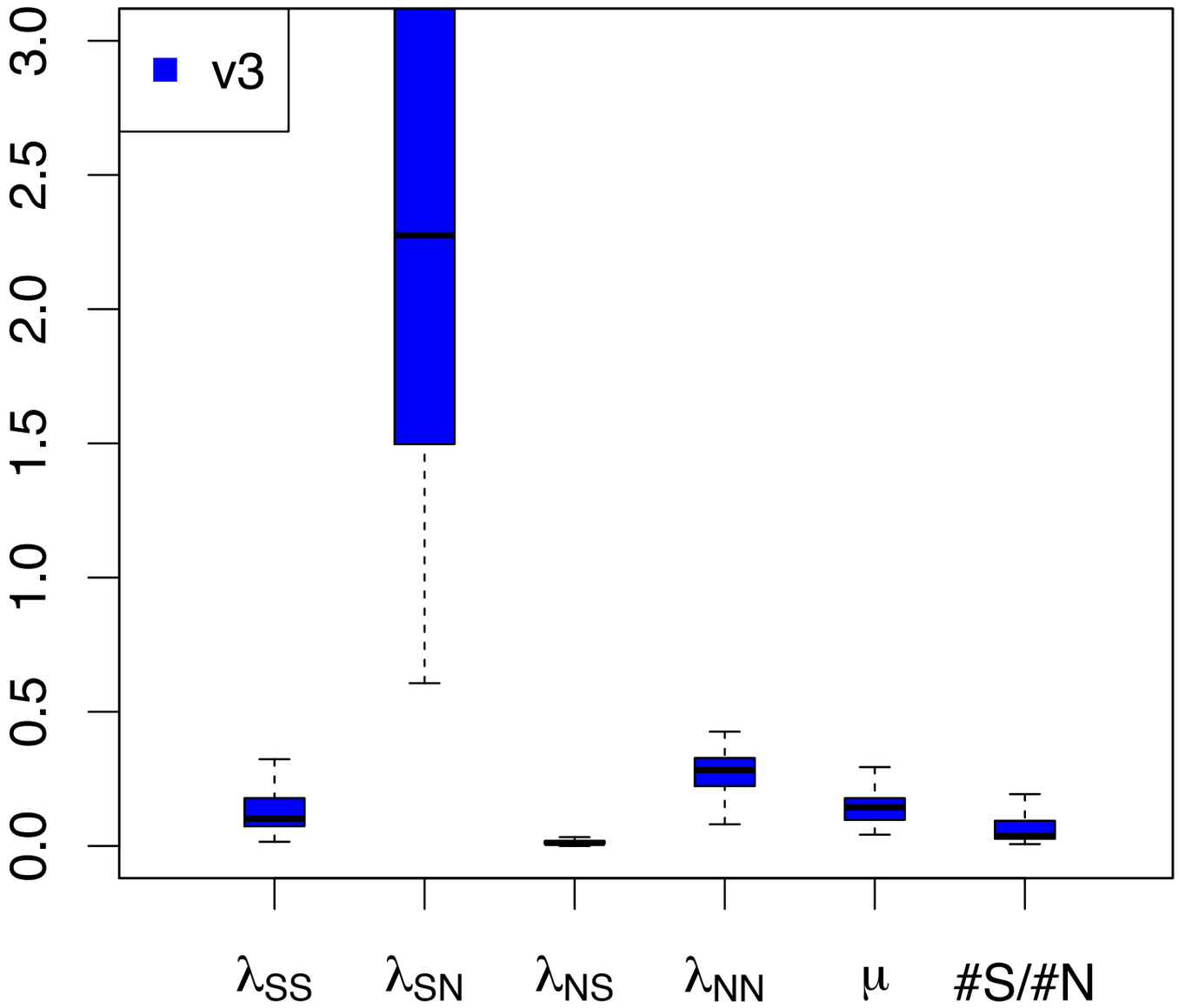


- reject homogeneous model (100 out of 100 Beast trees)
- IDU transmits more to HET than vice versa
- IDU transmits to HET as much as HET to itself
- Unequal "death" rates?

- Qualitatively similar results for
  - fraction sampled 10% or 1%
  - V3 and p17

Superspreaders

- **All MSM**
- **S=Superspreader; N = Normal**
- **Fraction sampled: 10%**



- reject homogeneous model (70 out of 100 Beast trees)
- preliminary evidence for superspreaders in MSM
- Superspreader about 20-fold less frequent than normal spreaders ($\lambda_{SS}/\lambda_{SN}$)
- Superspreaders transmit about 10-fold more than normal spreader ($\lambda_{SS}/\lambda_{NS}$)

Summary

- **Approach**
  - Novel method to quantify transmission in structured populations

- **HET/IDU**
  - Evidence for transmission structure in HET/IDU community
  - IDUs transmit more to HETs than vice versa

- **MSM:**
  - Method detects superspreaders amongst MSMs
  - about 20-fold less frequent than normal spreaders
  - about 10-fold higher transmission than normal spreaders

Acknowledgements

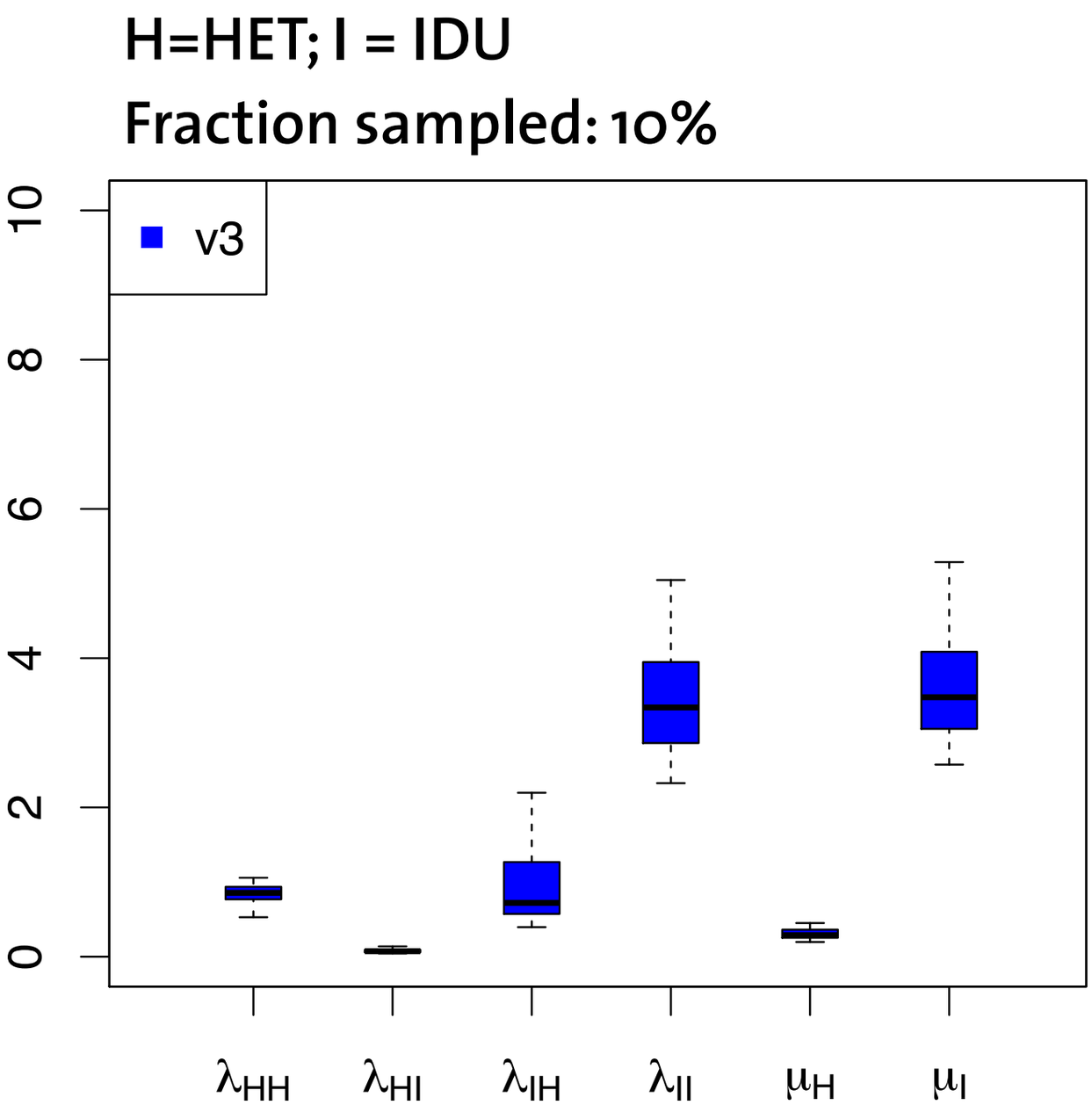
Tanja Stadler, Gabriel Leventhal
Institute of Integrative Biology, ETH Zurich

&

Roger Kouyos
University Hospital Zurich

Thank you for your attention

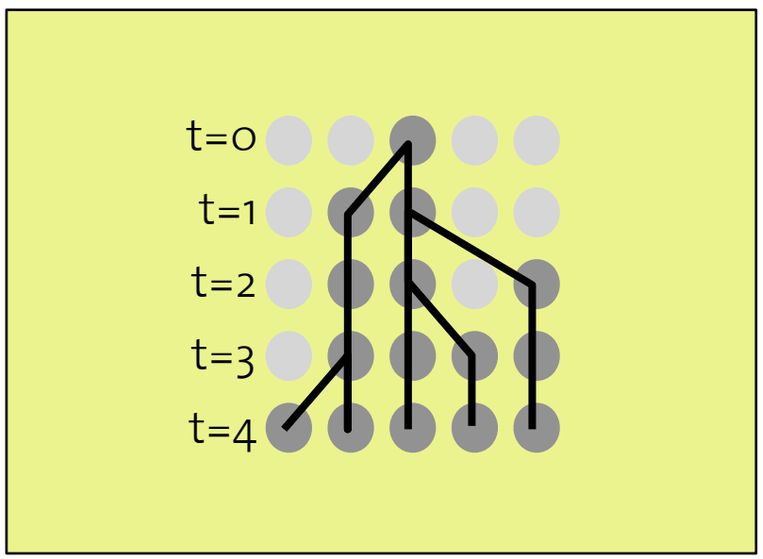
Transmission between HET and IDU



- Homogenous model rejected in 100 of 100 BEAST trees
- Higher transmission from IDUs to HETs than vice versa
  - see also Kouyos, JID, 2010
- HET to HET is much lower than IDU to IDU transmission
- Different “death” rates?

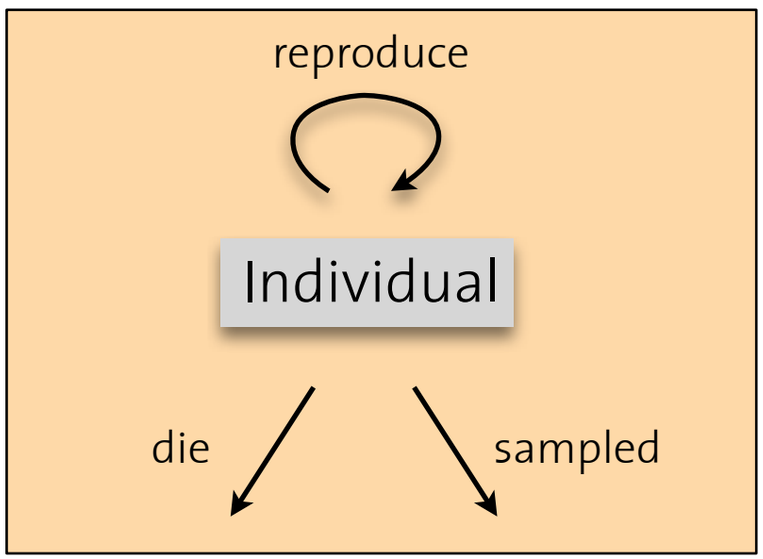
Linking phylogenetics and epidemiology

Coalescent



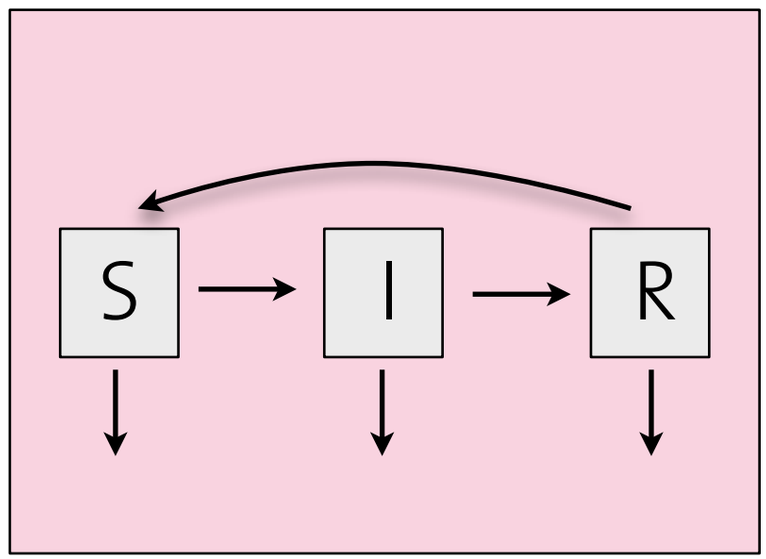
- **standard model:**
  - population genetics
- **analytical understanding:**
  - very good
- **Biological realism:**
  - poor
  - constant pop size
  - discrete time
  - sparse sampling
  - birth equals death

Birth/Death Model



- **standard model:**
  - species trees
- **analytical understanding:**
  - good
- **Biological realism:**
  - better
  - inc/decreasing pop size
  - continuous time
  - sparse or dense sampling
  - birth/death independent

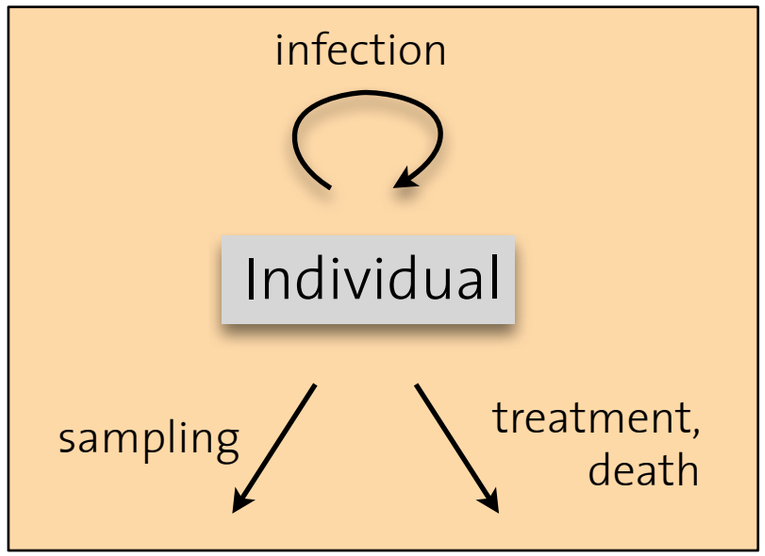
SIR Model



- **standard model:**
  - epidemiology
- **analytical understanding:**
  - poor
- **Biological realism:**
  - good
  - changing pop size
  - continuous time
  - accounts for temporal changes of susceptibles/infecteds

Birth/Death Model and R_o

Birth/Death Model



- We implemented birth/death model in BEAST to infer rates of:
  - infection (“birth”)
  - becoming non-infectious(treatment, death, ...)
  - sampling

Basic reproductive number: $R_o = \dfrac{\text{infection rate}}{\text{“death” rate}}$

Average number of secondary infections generated by an infected individual

R_0 estimates for Swiss HIV data

- Swiss HIV Cohort Data
- 10 largest clusters with no migration
- Size: 12 - 34 individuals sampled between 1995 and 2008

		R_0				
		mean	SD	median	2.5%	97.5%
Case (i)		2.17	0.37	2.14	1.57	3.00
Case (ii)	1	2.86	3.51	1.90	1.07	10.34
	2	1.59	0.62	1.40	1.02	3.23
	3	1.06	0.19	1.03	0.72	1.53
	4	1.43	0.50	1.27	0.99	2.68
	5	4.01	17.14	1.90	1.05	16.92
	6	1.79	1.17	1.47	1.02	4.44
	7	2.81	3.43	1.92	1.08	9.84
	8	1.25	0.55	1.13	0.70	2.45
	9	3.51	7.74	2.03	1.06	14.14
	10	1.66	0.86	1.40	0.92	4.00

Part III: Summary

- R_0 or other epidemiological / evolutionary parameters can be inferred directly from genetic data

- $R_0 \sim 2$ in Switzerland
  - transmission group independent
  - little changes over time observed

Take home

- **Part I: HIV transmission structure**
  - HIV transmission among heterosexuals in Switzerland appears to be driven by intravenous drug users
  - The impact of IDUs on the HET epidemic decreased over time, which is likely due to public health interventions targeted at IDUs

- **Part II: Inferring contact structure**
  - Tree-balancedness can be used to distinguish between alternative models for host contact structure
  - HIV transmission in Switzerland reveals high levels of non-random contact structure

- **Part III: Quantifying epidemiological parameters**
  - Using birth-death models rather than the coalescent in the phylogenetic analysis of viral transmission trees allows to estimate R_0
  - $R_0 \sim 2$ in Switzerland

Outlook

- **Part I: HIV transmission structure**
  - Quantify transmission between risk groups
  - Modify birth-death model to allow for multiple groups (=> Tanja Stadler)

- **Part II: Inferring contact structure**
  - Use branch length in addition to tree topology to distinguish between models of contact structure (=> Gabriel Leventhal)

- **Part III: Quantifying epidemiological parameters**
  - Using SIR rather than birth death models for phylogenetic analysis (=> Gabriel Leventhal and Tanja Stadler)