Phylogenetics in action

Detecting competition based on phylogenetic trees

joint work with Gabriel Leventhal & Rampal Etienne

Tanja Stadler
Theoretical Biology group — Department of Environmental Systems Science
tanja.stadler@env.ethz.ch

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

My main questions

Macro-evolution


http://conservingbiodiversity.yolasite.com/

- **What is the number of niches for which species compete?**

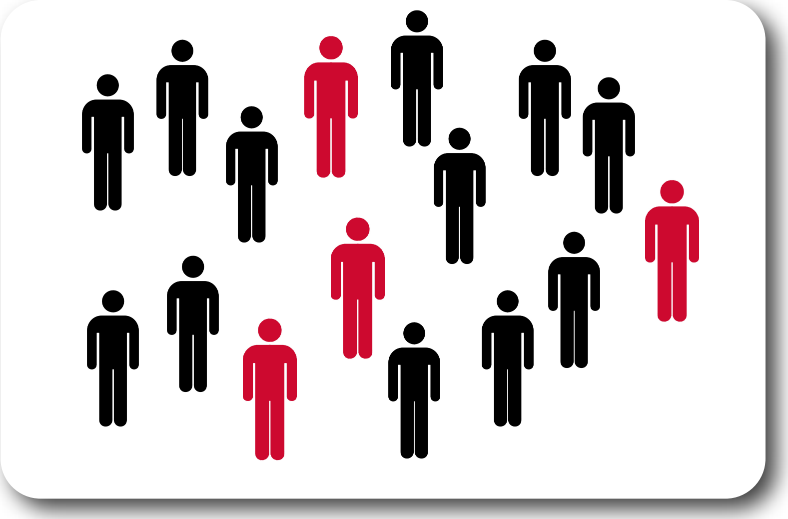
My main questions

Macro-evolution

http://conservingbiodiversity.yolasite.com/

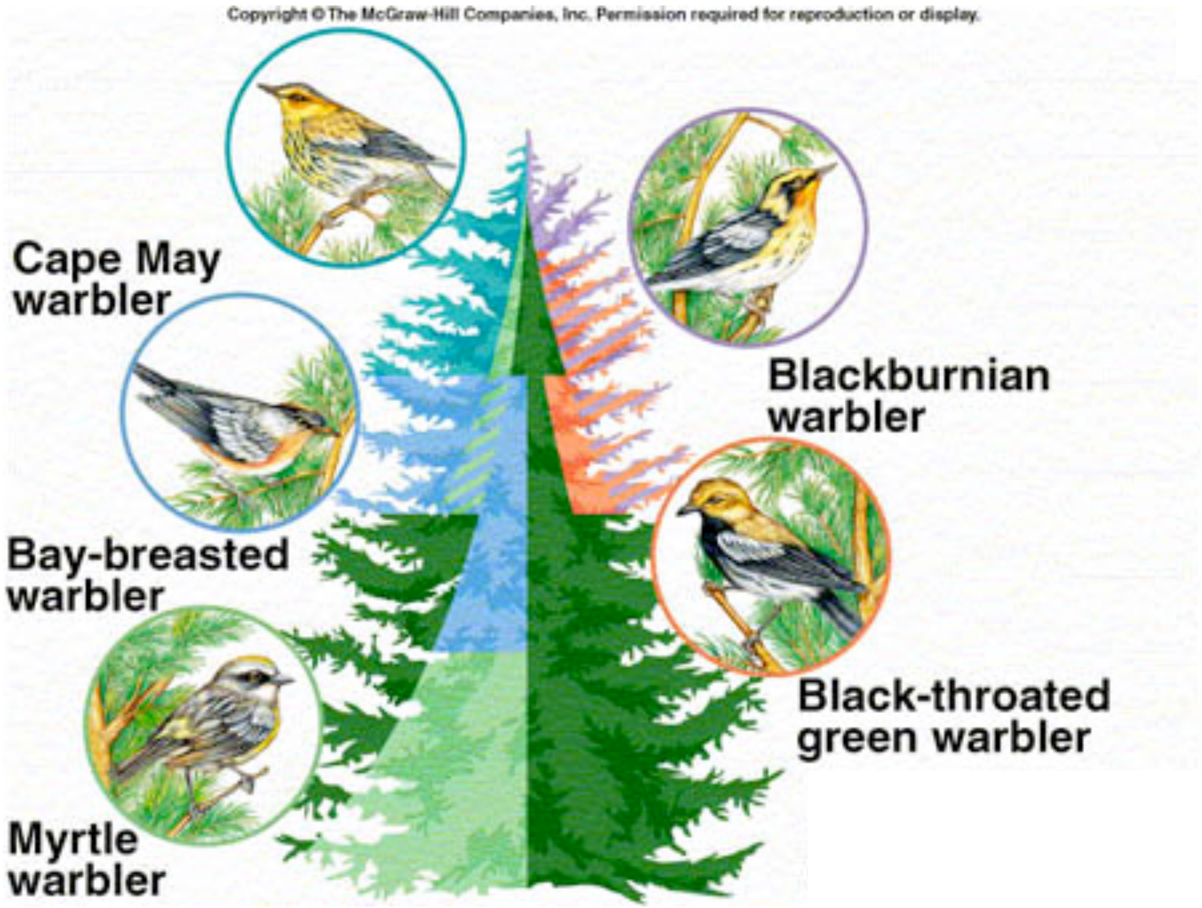
- **What is the number of niches for which species compete?**

Epidem-iology

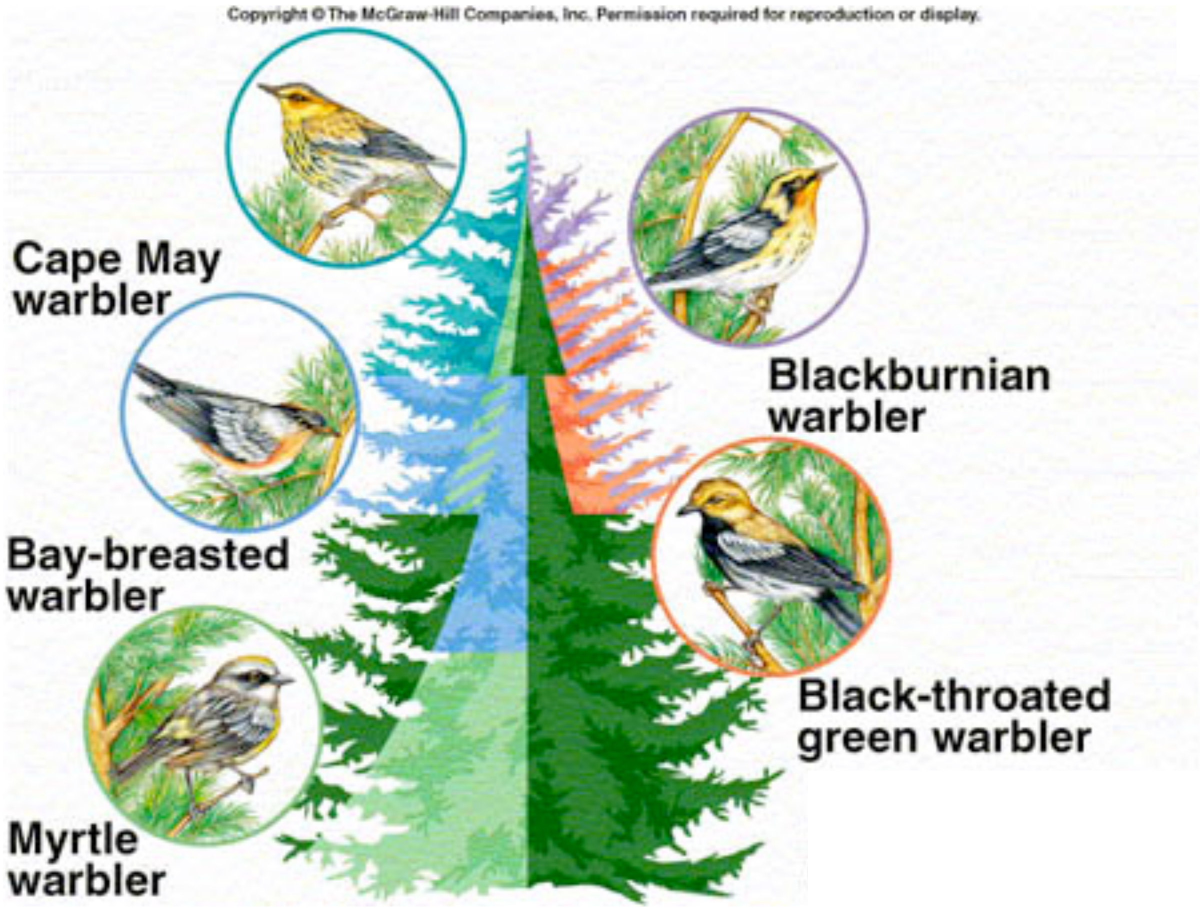


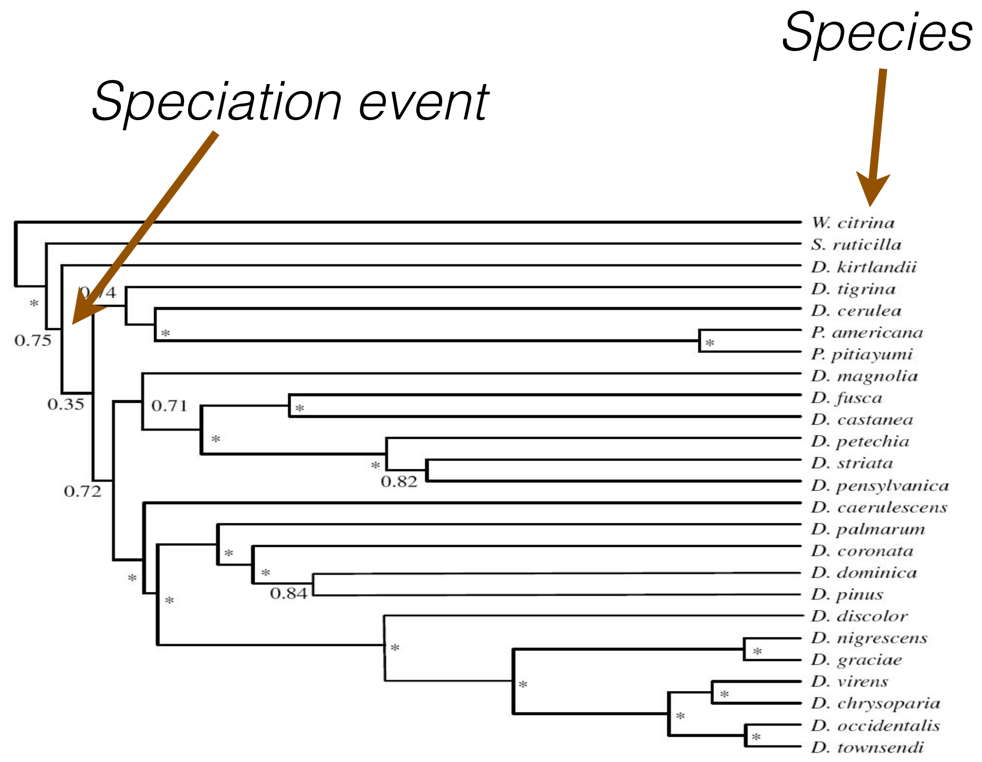
- **What is the number of hosts for which pathogens compete?**

Part A:
Competition between species



Part A:
Competition between species




Species

Speciation event

density-dependent speciation and extinction?

Likelihood approach for quantifying competition between species

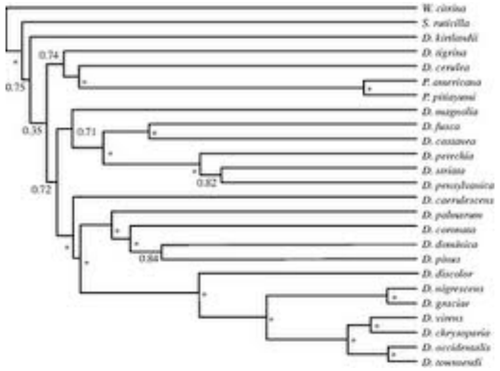
Mathematical

- Define a density-dependent macroevolutionary model
- Find efficient way to calculate likelihood function: *Probability of the phylogenetic tree given the parameters*

Computational

- Implement R packages TreePar & TreeSim

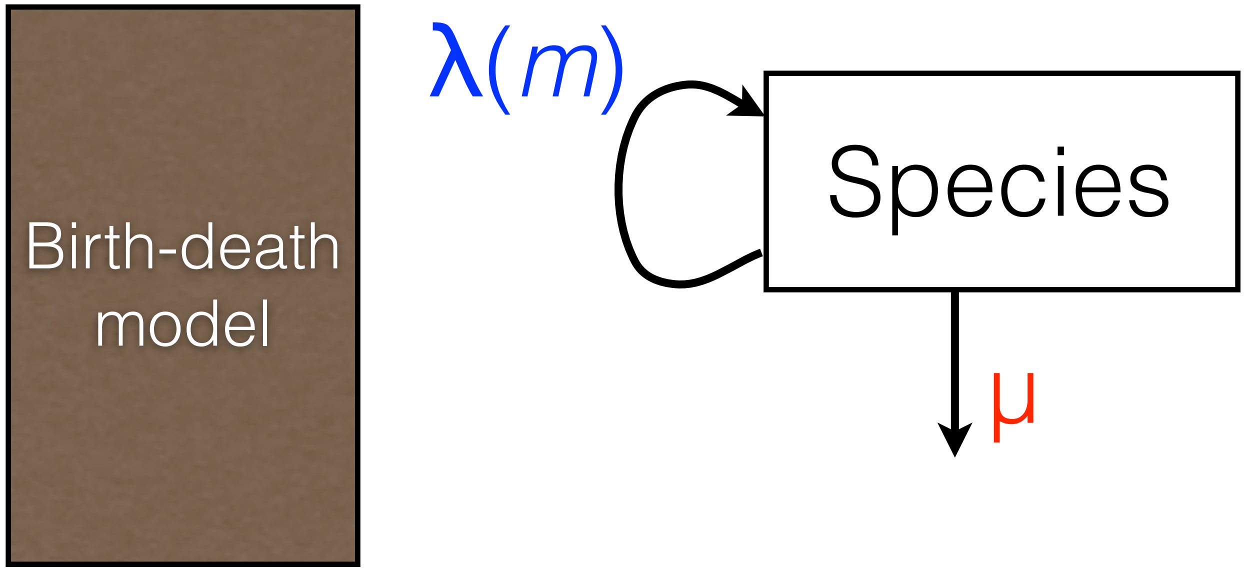
Empirical



Birds

Ants

Density-dependent speciation and extinction model



Birth-death model

$\lambda(m)$

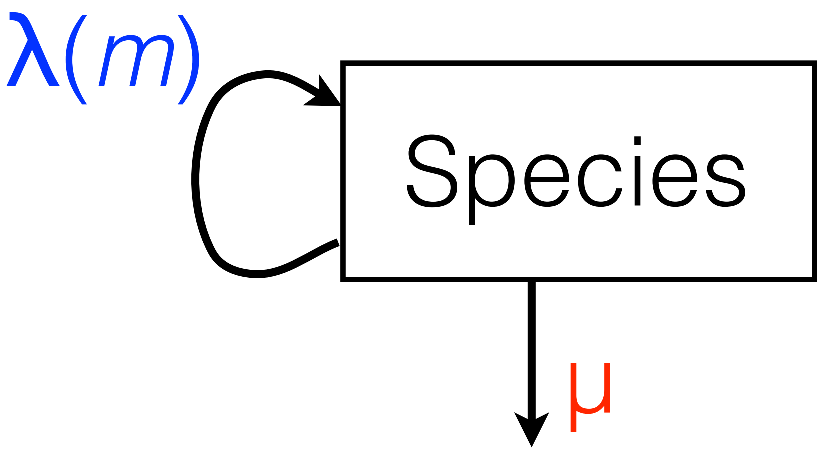
Species

μ

- stem age of a clade T
- speciation rate $\lambda(m)=\beta(1-m/K)$
  K is number of niches
- extinction rate μ
- sampling probability ρ

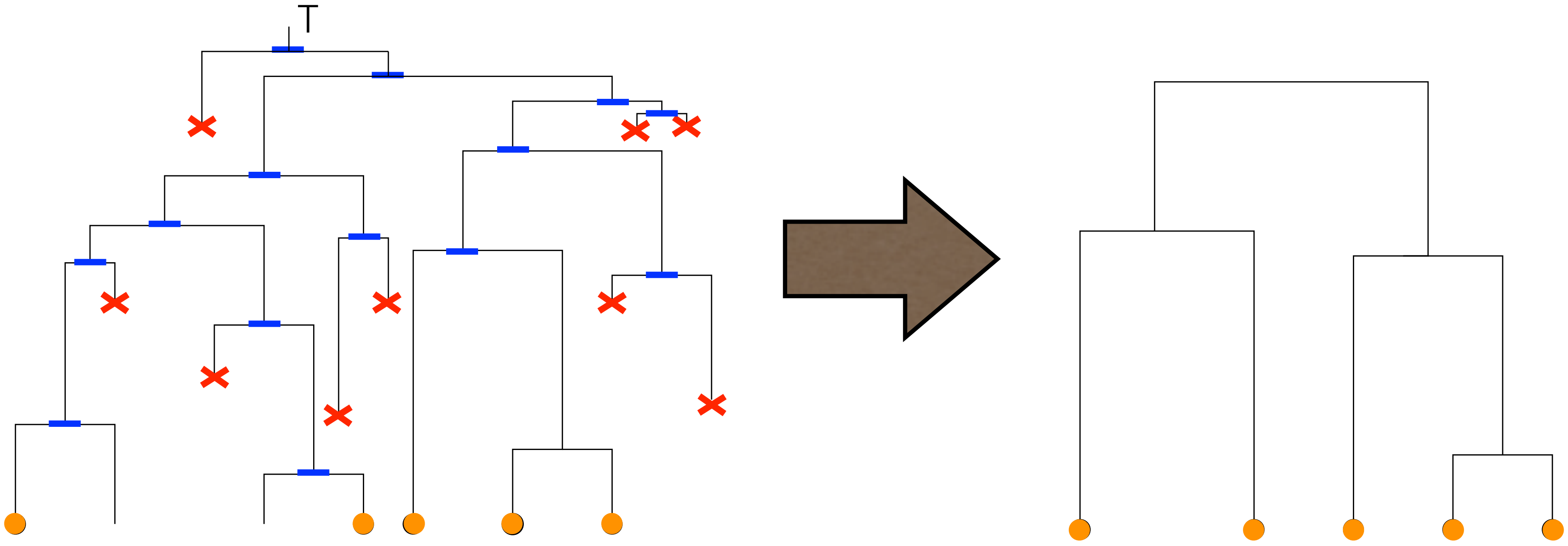
Density-dependent speciation and extinction model

Birth-death model



- stem age of a clade T
- speciation rate $\lambda(m)=\beta(1-m/K)$
  K is number of niches
- extinction rate μ
- sampling probability ρ

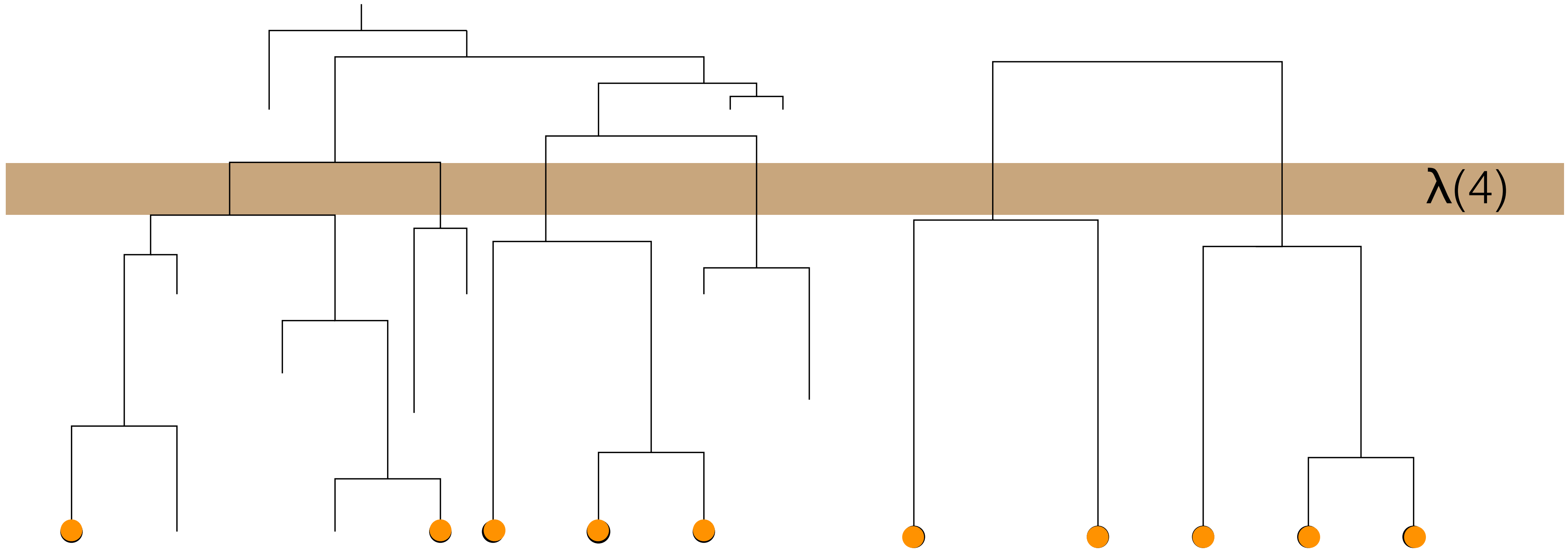
Phylo-genetic trees



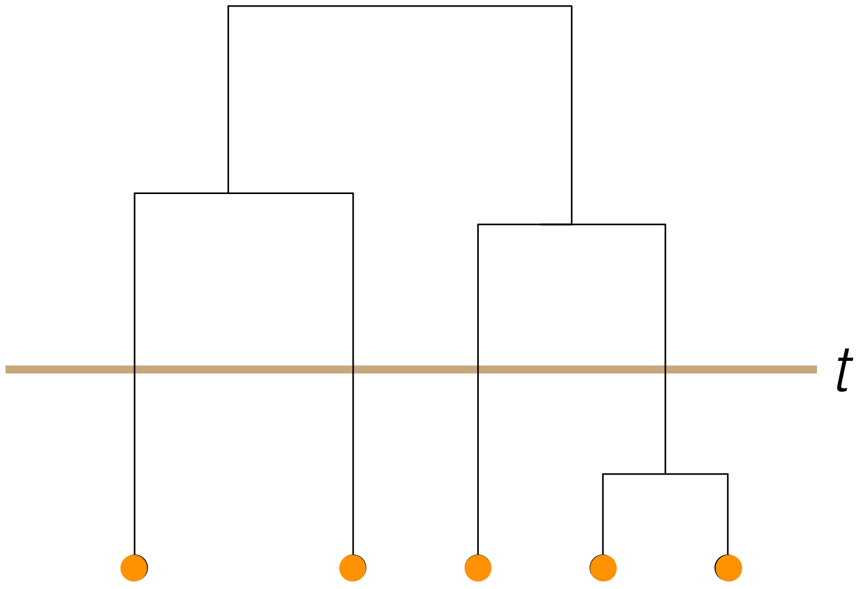
complete phylogeny

reconstructed phylogeny

Likelihood of a reconstructed tree

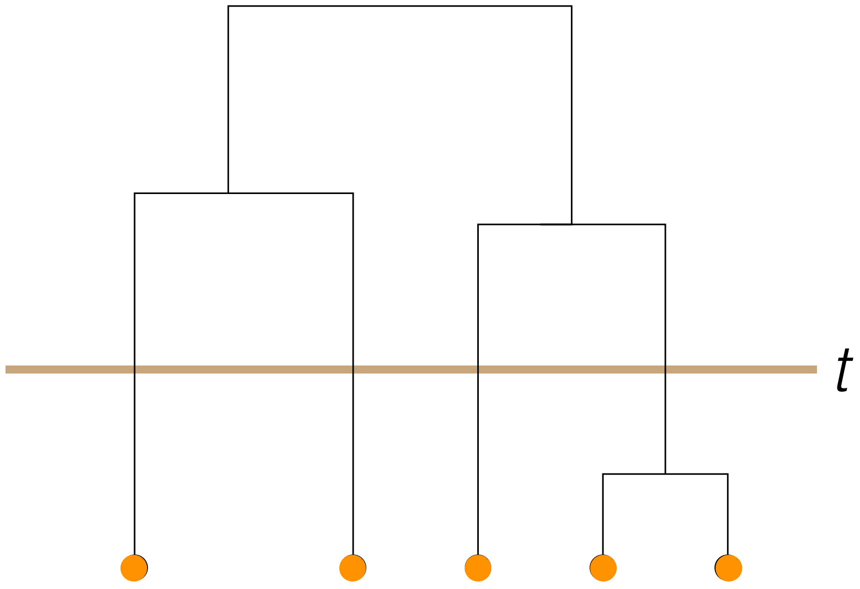


Likelihood of a reconstructed tree



- m: number of species at time t in complete tree
- n: number of species at time t in reconstructed tree
- $f(t|m)$: probability density of observing the reconstructed tree between t and present given m species at time t

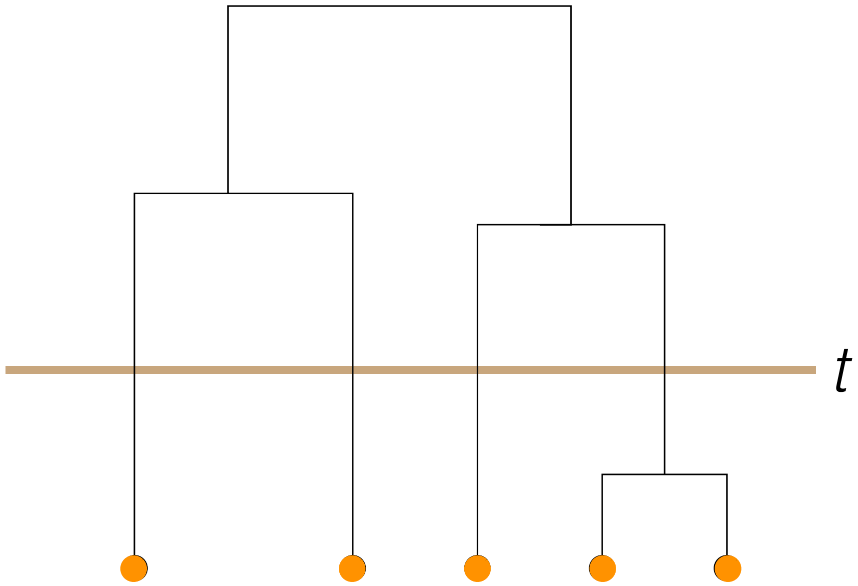
Likelihood of a reconstructed tree



- *m*: number of species at time *t* in complete tree
- *n*: number of species at time *t* in reconstructed tree
- $f(t|m)$: probability density of observing the reconstructed tree between *t* and present given *m* species at time *t*

$$\begin{aligned} f(t+\Delta t|m) \quad &= \quad [1-m(\lambda(m)+\mu)\Delta t]\, f(t|m) \\ &\quad +(k-n)\left[\lambda(m) f(t|m+1)+\mu f(t|m-1)\right]\Delta t \\ &\quad +2n\lambda(m)\Delta t f(t|m+1) \end{aligned}$$

Likelihood of a reconstructed tree



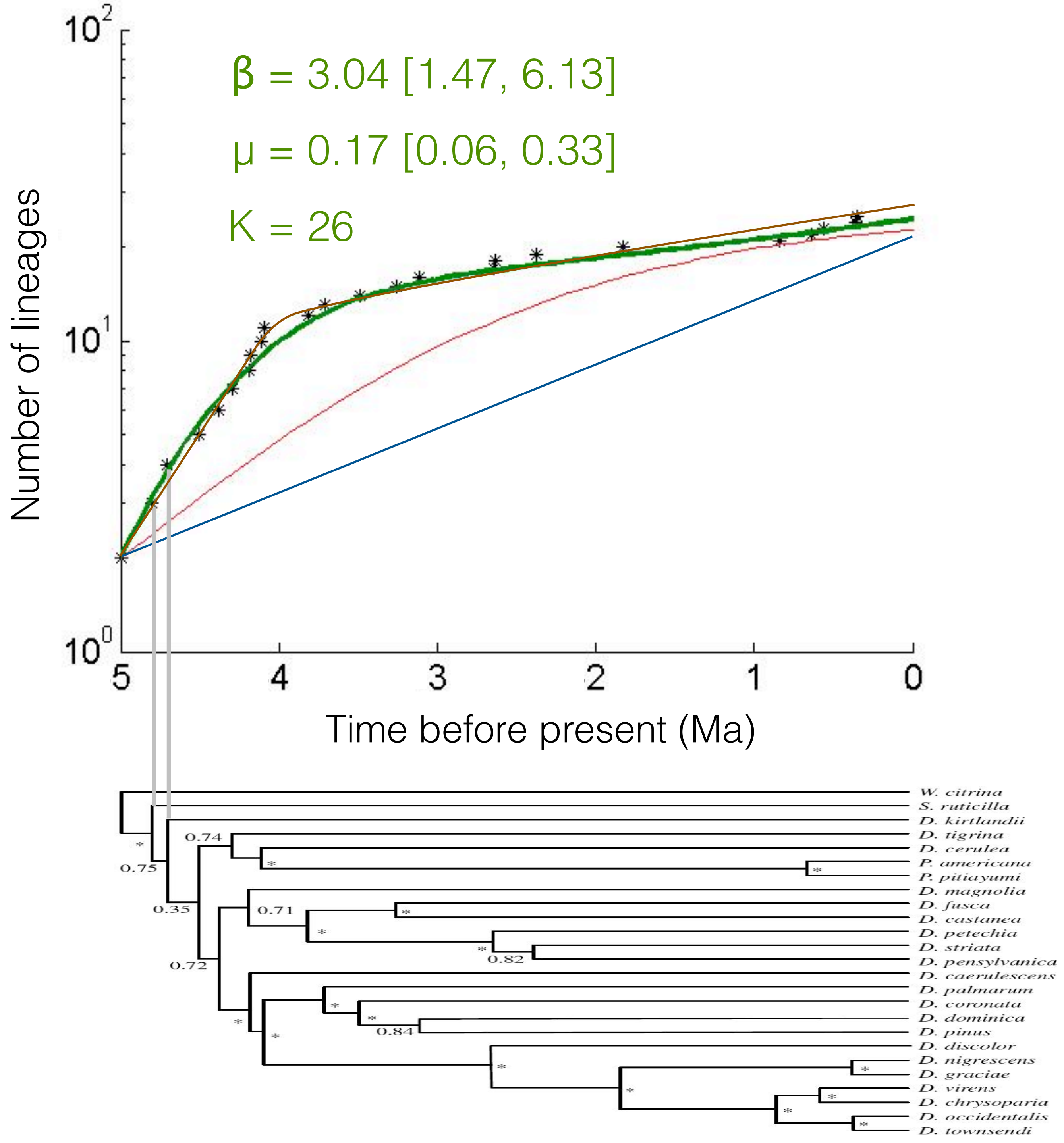
- *m*: number of species at time *t* in complete tree
- *n*: number of species at time *t* in reconstructed tree
- $f(t|m)$: probability density of observing the reconstructed tree between *t* and present given *m* species at time *t*

$$\begin{aligned} f(t+\Delta t|m) \quad &= \quad [1-m(\lambda(m)+\mu)\Delta t]\, f(t|m) \\ &\quad +(k-n)\left[\lambda(m) f(t|m+1)+\mu f(t|m-1)\right]\Delta t \\ &\quad +2n\lambda(m)\Delta t f(t|m+1) \end{aligned}$$

$\Delta t \rightarrow 0$

$$\frac{d}{dt} f(t|.) = M f(t|.)$$

Dendroica warbler phylogeny



✷ empirical data

our density-dependence model

environmental model (ΔAIC=5.93)
(Stadler, PNAS, 2011)

density-dependence w/o extinction (p=0.97)
(Rabosky et al., Proc. Roy. Soc. B, 2008)

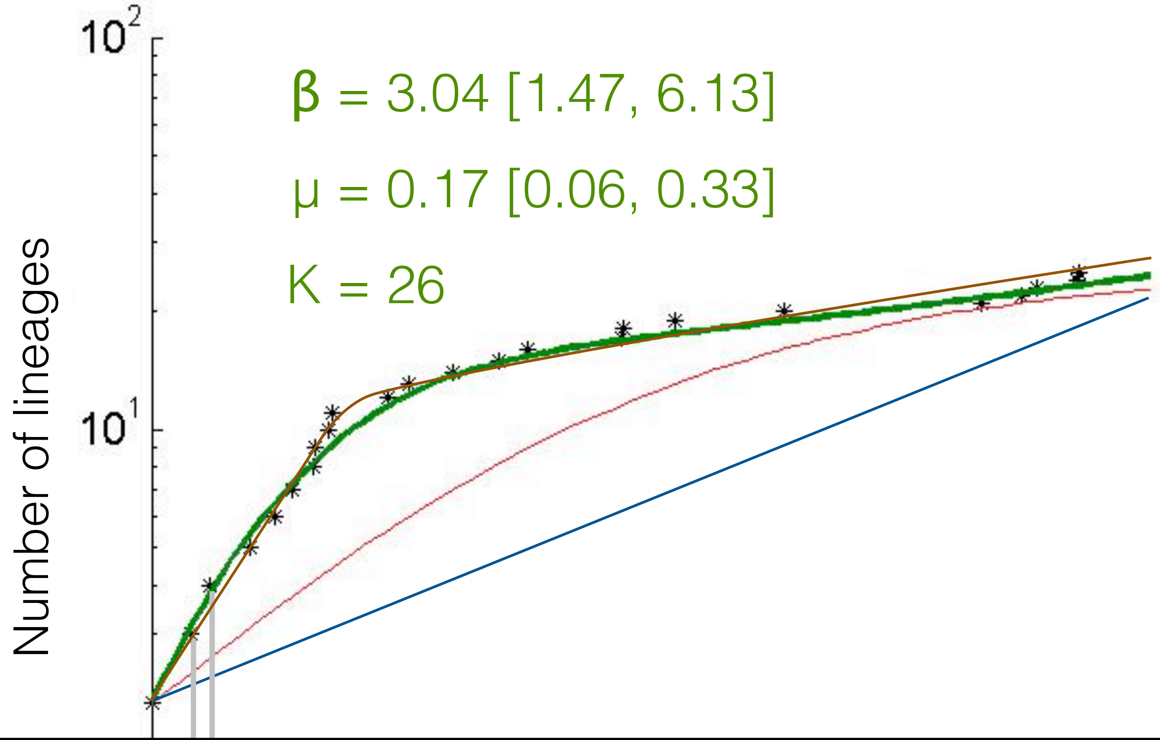
constant rates model (p=1.00)
(Nee et al., Phil. Trans. Roy. Soc. B, 1994)

Warbler phylogeny

(from Rabosky et al., Proc. Roy. Soc. B, 2008)

Etienne et al. (Proc. Roy. Soc. B, 2011)

Dendroica warbler phylogeny



β = 3.04 [1.47, 6.13]

μ = 0.17 [0.06, 0.33]

K = 26

✷ empirical data

our density-dependence model

environmental model (ΔAIC=5.93)
(Stadler, PNAS, 2011)

density-dependence w/o extinction (p=0.97)
(Rabosky et al., Proc. Roy. Soc. B, 2008)

constant rates model (p=1.00)

Limitation

Method only works for trees up to 50 species

Warbler phylogeny

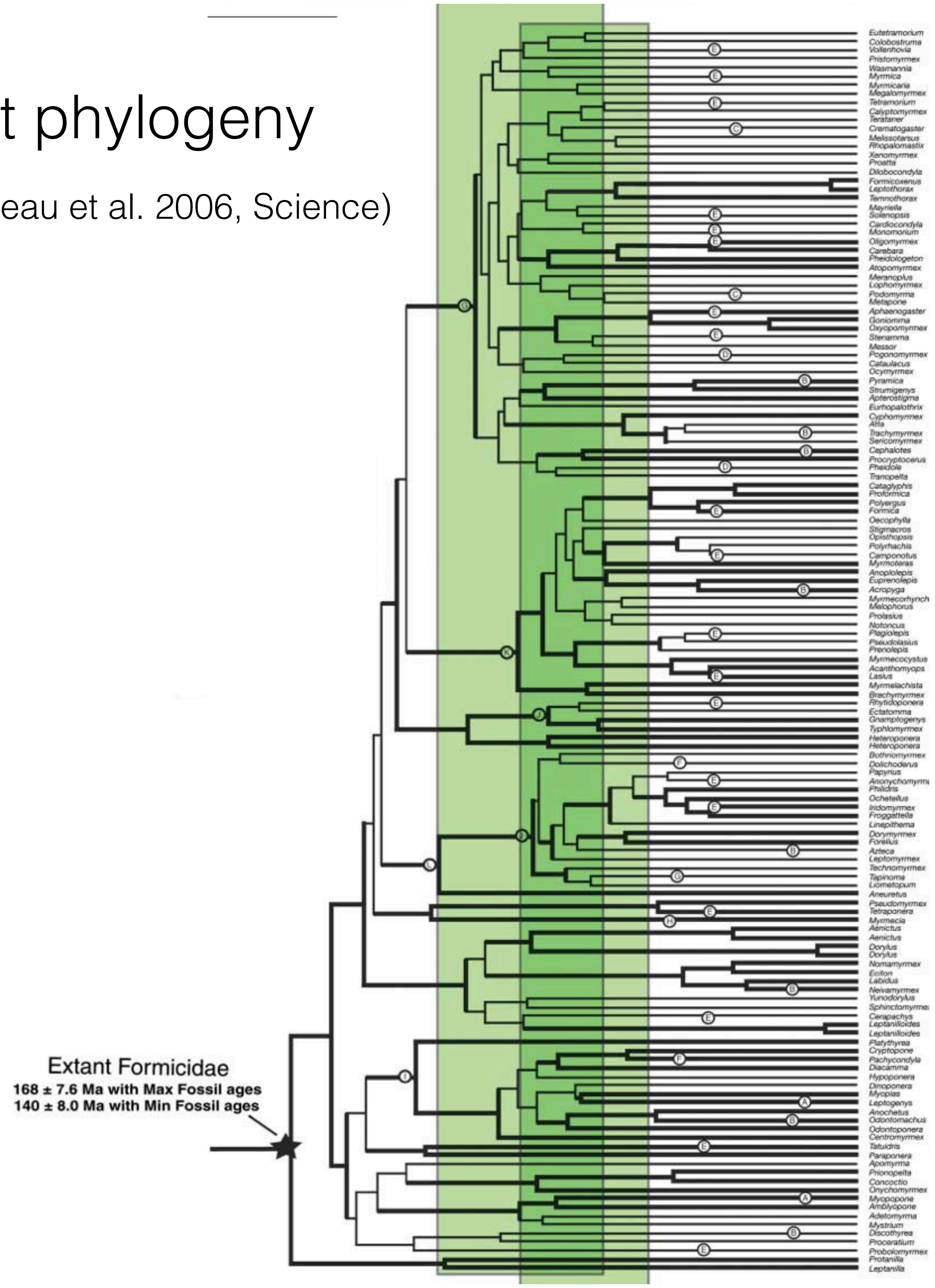
(from Rabosky et al., Proc. Roy. Soc. B, 2008)

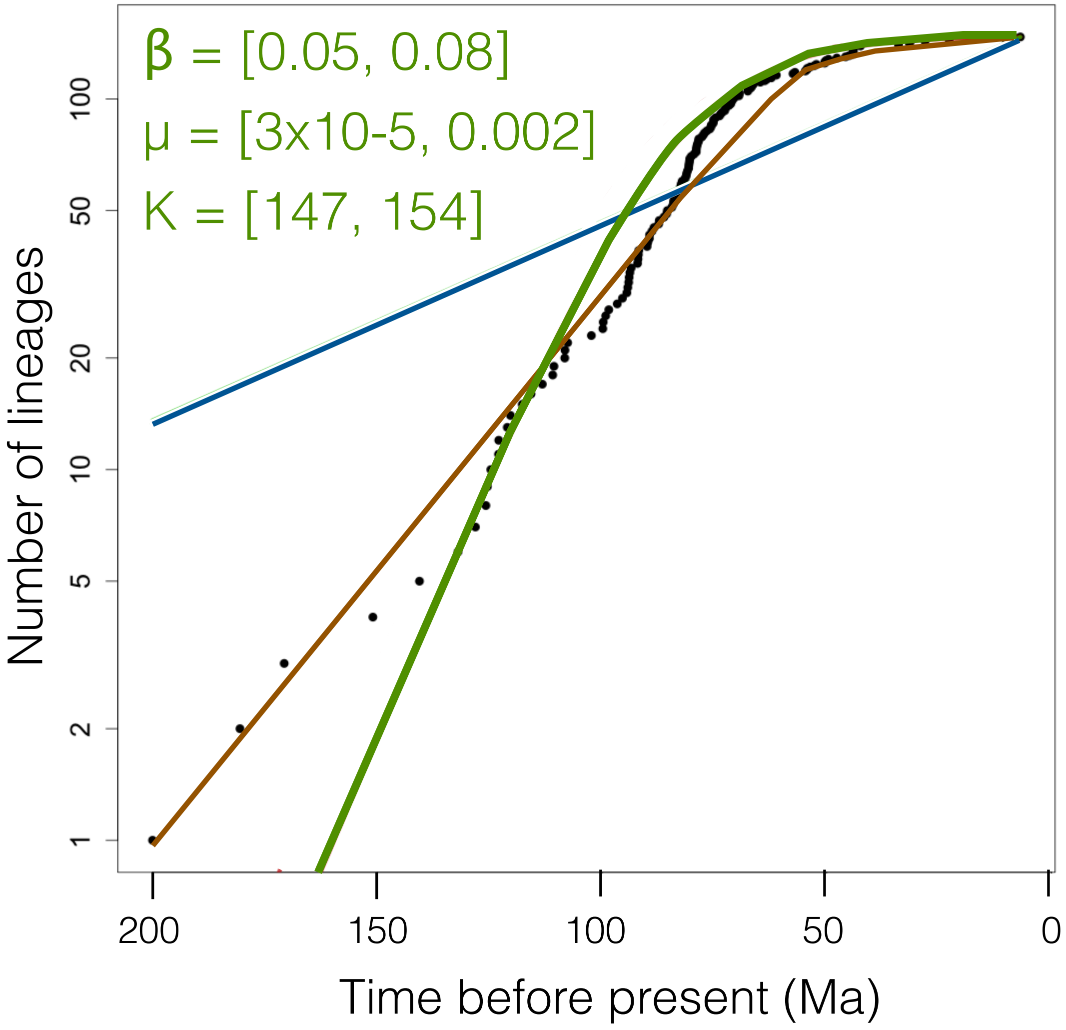
Etienne et al. (Proc. Roy. Soc. B, 2011)

Formicidae ant phylogeny

Ant phylogeny

(from Moreau et al. 2006, Science)





- empirical data

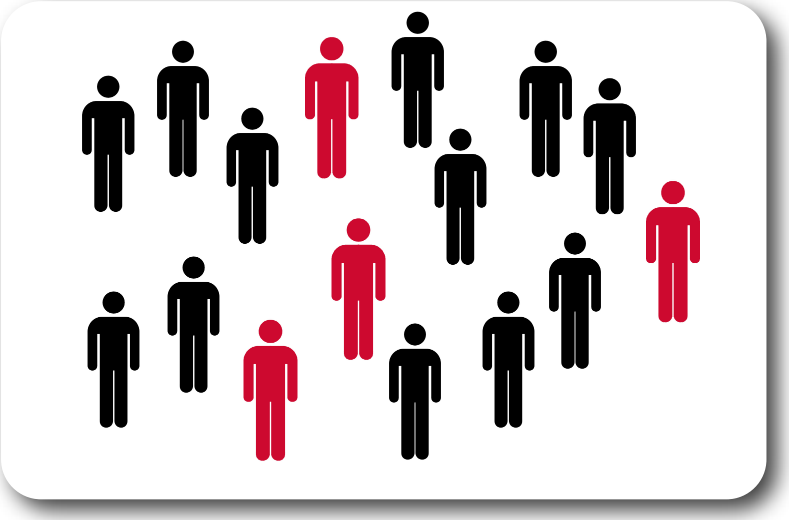
our density-dependence model

environmental model (ΔAIC=36.43)

constant rates model (p=1.00)

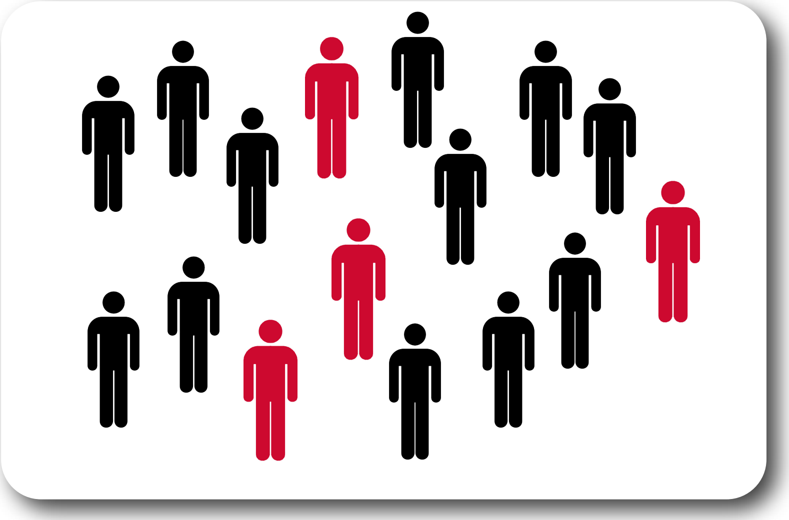
Leventhal & Stadler (unpublished)

Part B:
Competition for susceptible host

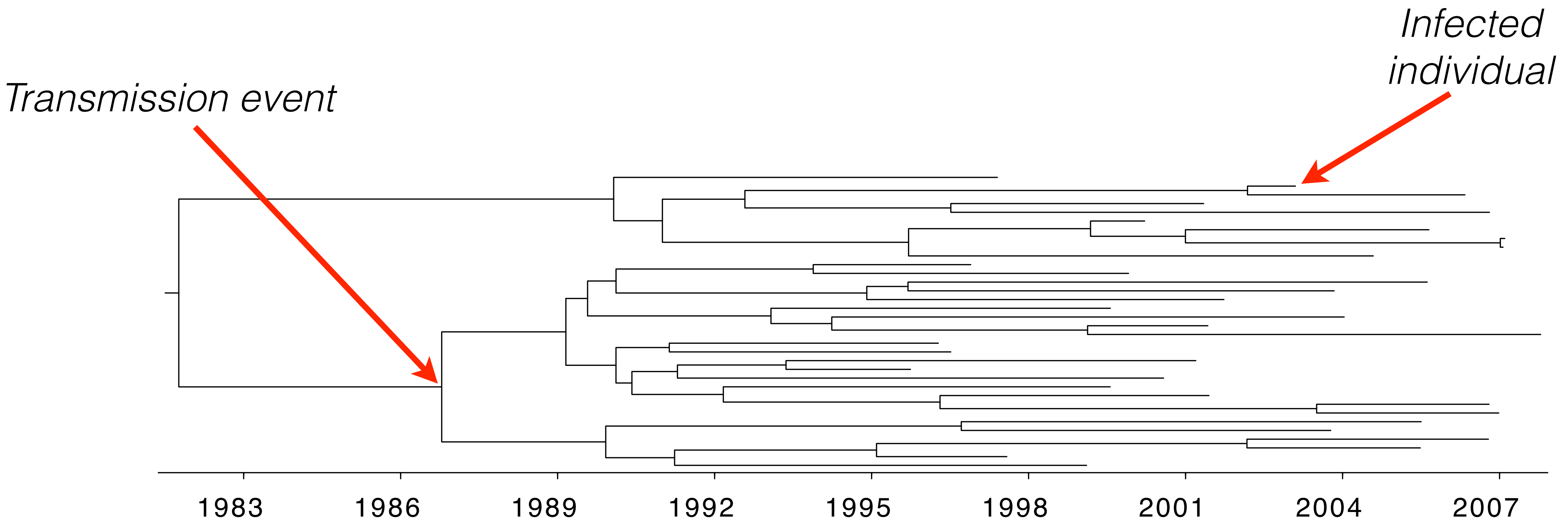


What is the number of hosts for which pathogens compete?

Part B:
Competition for susceptible host



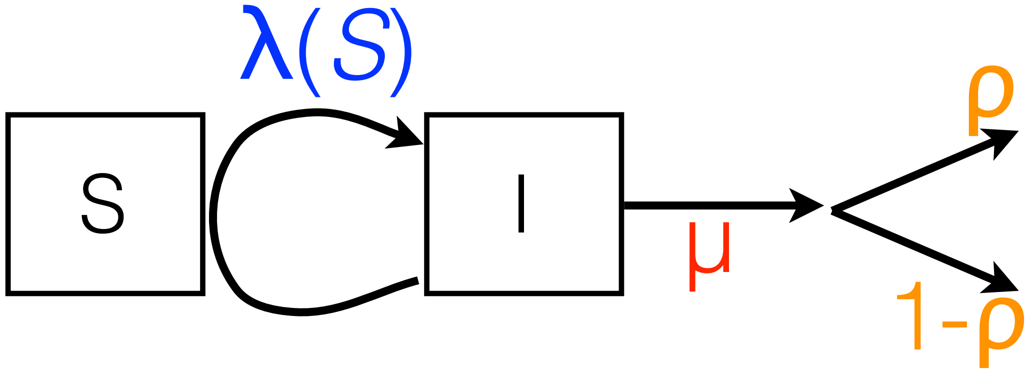
What is the number of hosts for which pathogens compete?



Swiss HIV transmission chain

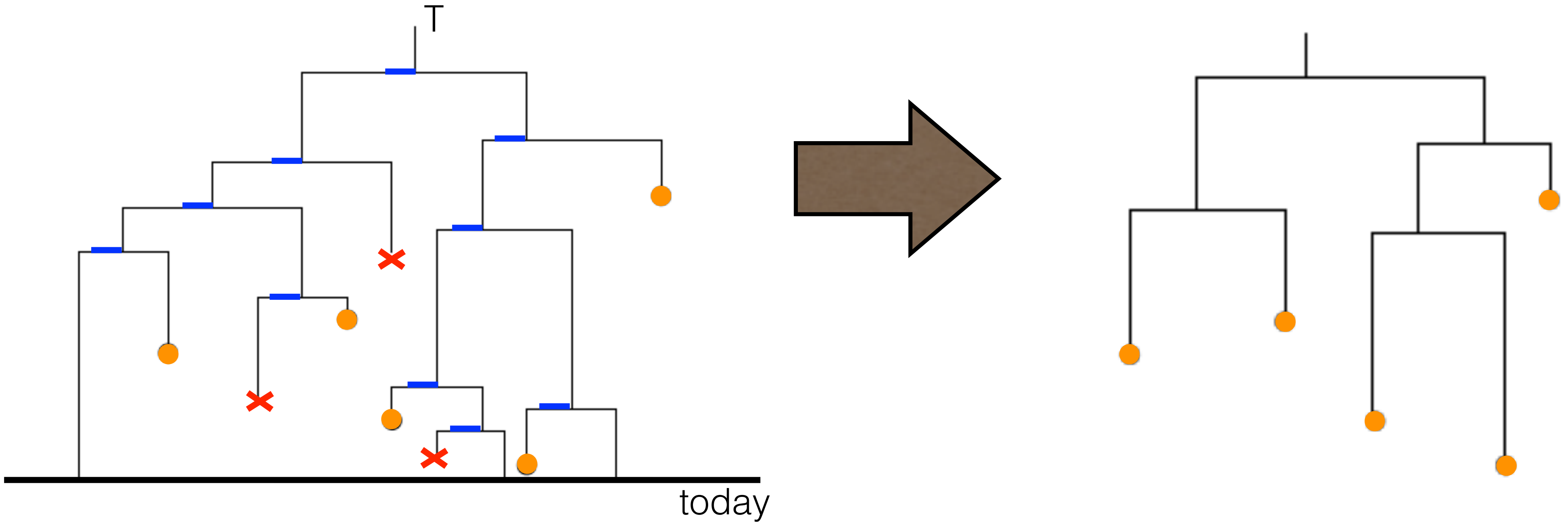
Model for epidemic spread (SI model)

Birth-death model



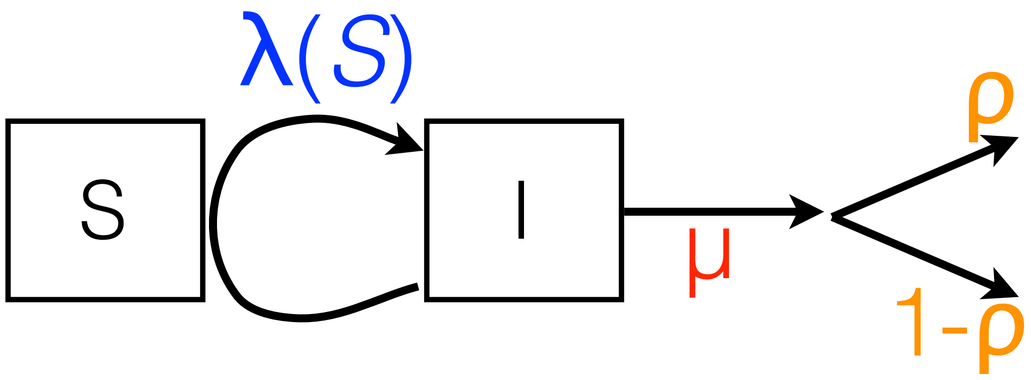
- time of epidemic outbreak T
- transmission rate $\lambda(S)=\beta S/K$
- “becoming-non-infectious” rate μ
- sampling probability ρ

Phylo-genetic trees

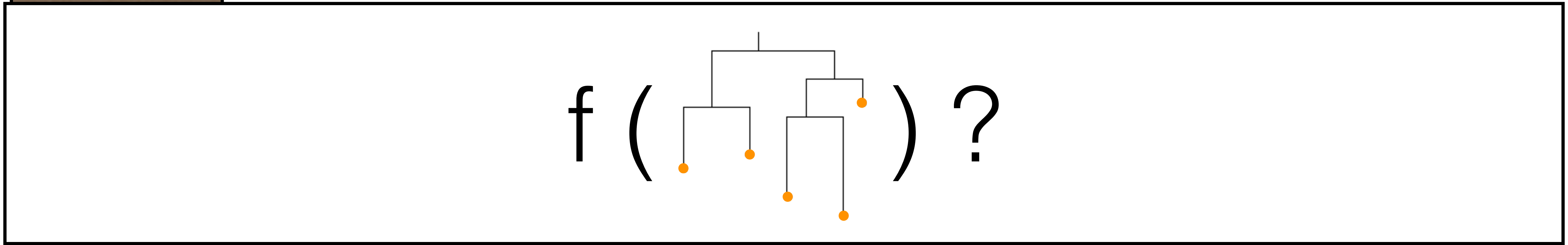


Model for epidemic spread (SI model)

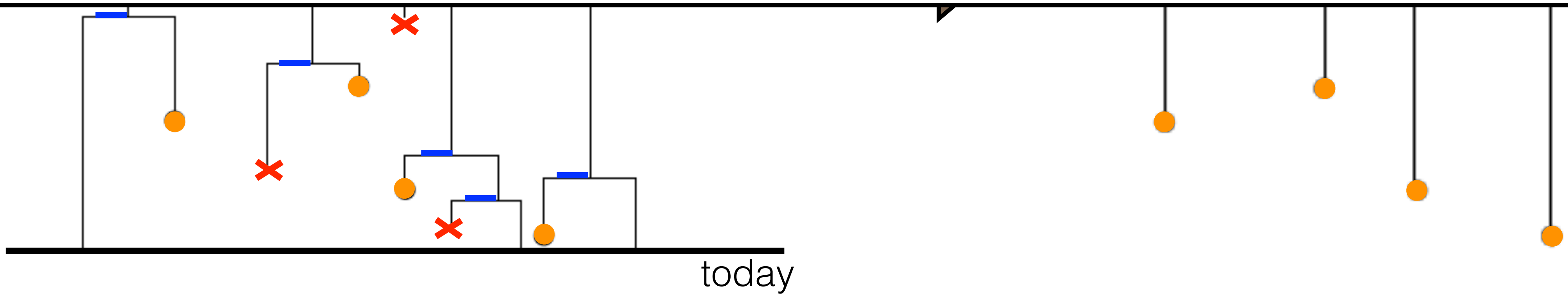
Birth-death model



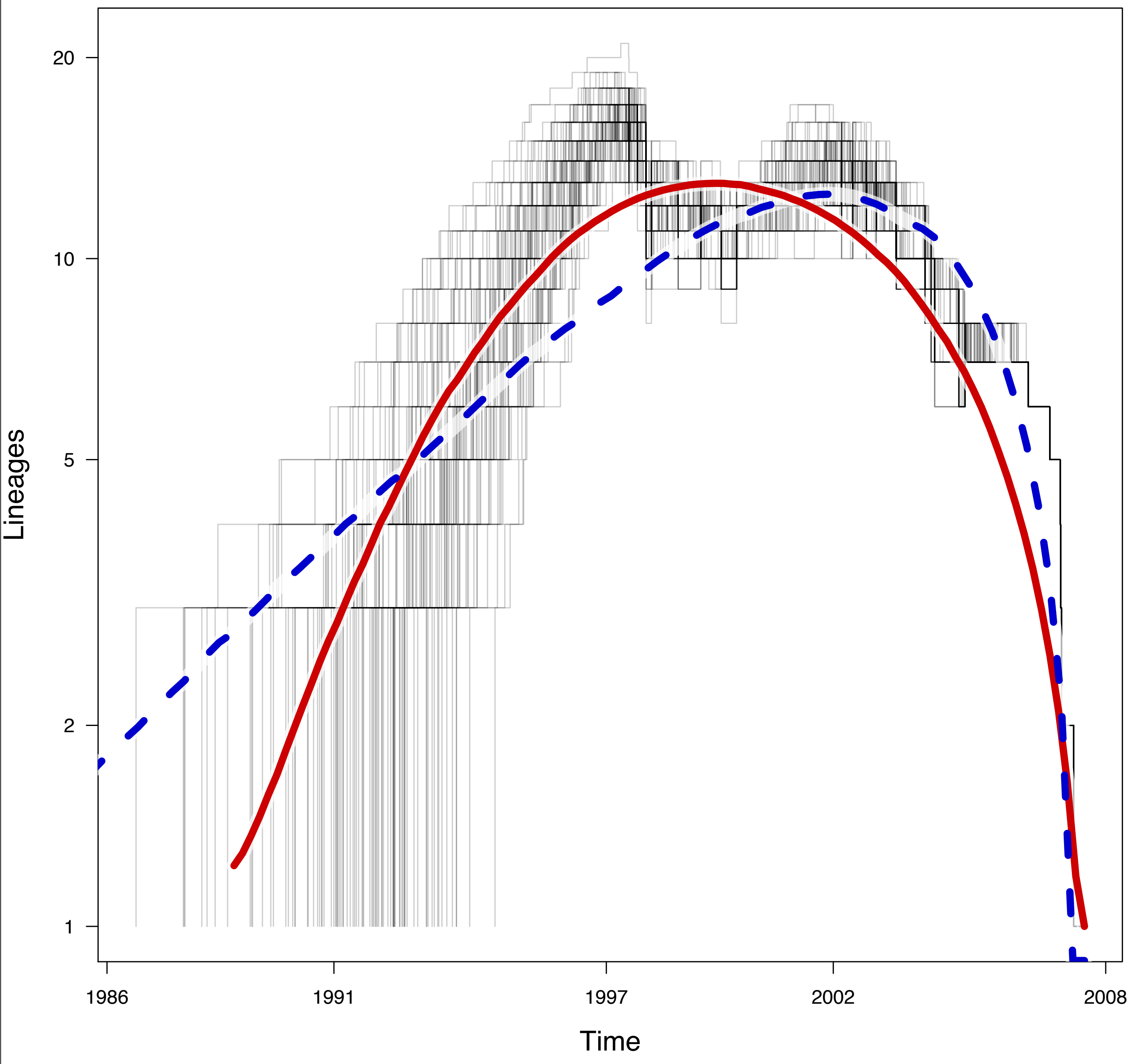
- time of epidemic outbreak T
- transmission rate $\lambda(S)=\beta S/K$
- “becoming-non-infectious” rate μ
- sampling probability ρ

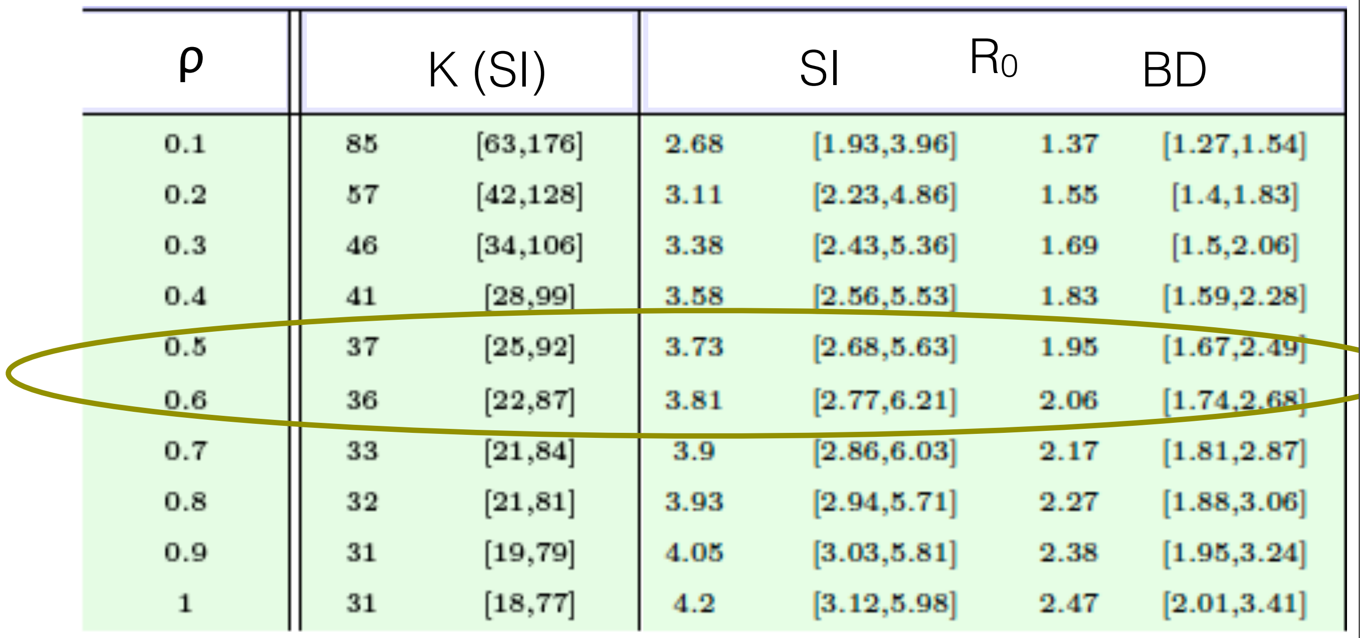


genetic trees



Swiss HIV-1 cluster (29 patients)





ρ	K (SI)		R0			
			SI		BD	
0.1	85	[63,176]	2.68	[1.93,3.96]	1.37	[1.27,1.54]
0.2	57	[42,128]	3.11	[2.23,4.86]	1.55	[1.4,1.83]
0.3	46	[34,106]	3.38	[2.43,5.36]	1.69	[1.5,2.06]
0.4	41	[28,99]	3.58	[2.56,5.53]	1.83	[1.59,2.28]
0.5	37	[25,92]	3.73	[2.68,5.63]	1.95	[1.67,2.49]
0.6	36	[22,87]	3.81	[2.77,6.21]	2.06	[1.74,2.68]
0.7	33	[21,84]	3.9	[2.86,6.03]	2.17	[1.81,2.87]
0.8	32	[21,81]	3.93	[2.94,5.71]	2.27	[1.88,3.06]
0.9	31	[19,79]	4.05	[3.03,5.81]	2.38	[1.95,3.24]
1	31	[18,77]	4.2	[3.12,5.98]	2.47	[2.01,3.41]

- - - - exponential growth model

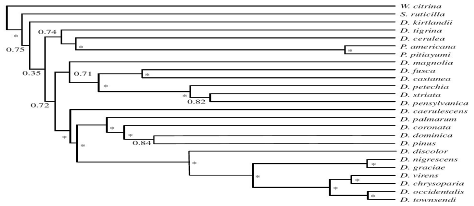
—— SI model

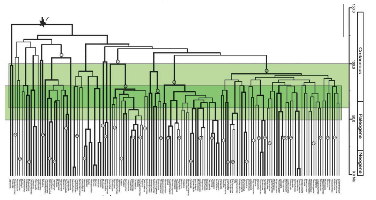
Likelihood ratio test rejects exponential growth model at 2.5% level

Leventhal & Stadler (unpublished)

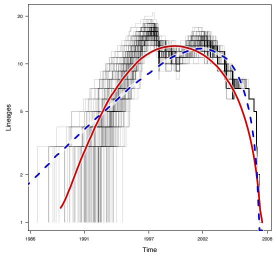
Phylogenetic methods...

- ...for determining the number of ecological niches





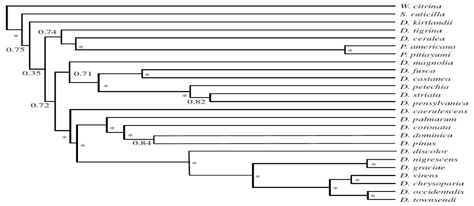
- ...for determining the host population size *(SI model)*

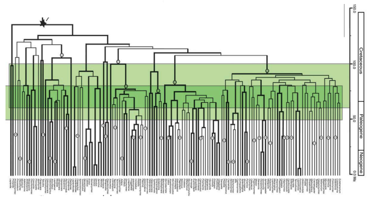


- ...available in the R packages TreeSim and TreePar

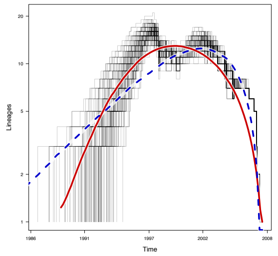
Phylogenetic methods...

- ...for determining the number of ecological niches





- ...for determining the host population size *(SI model)*



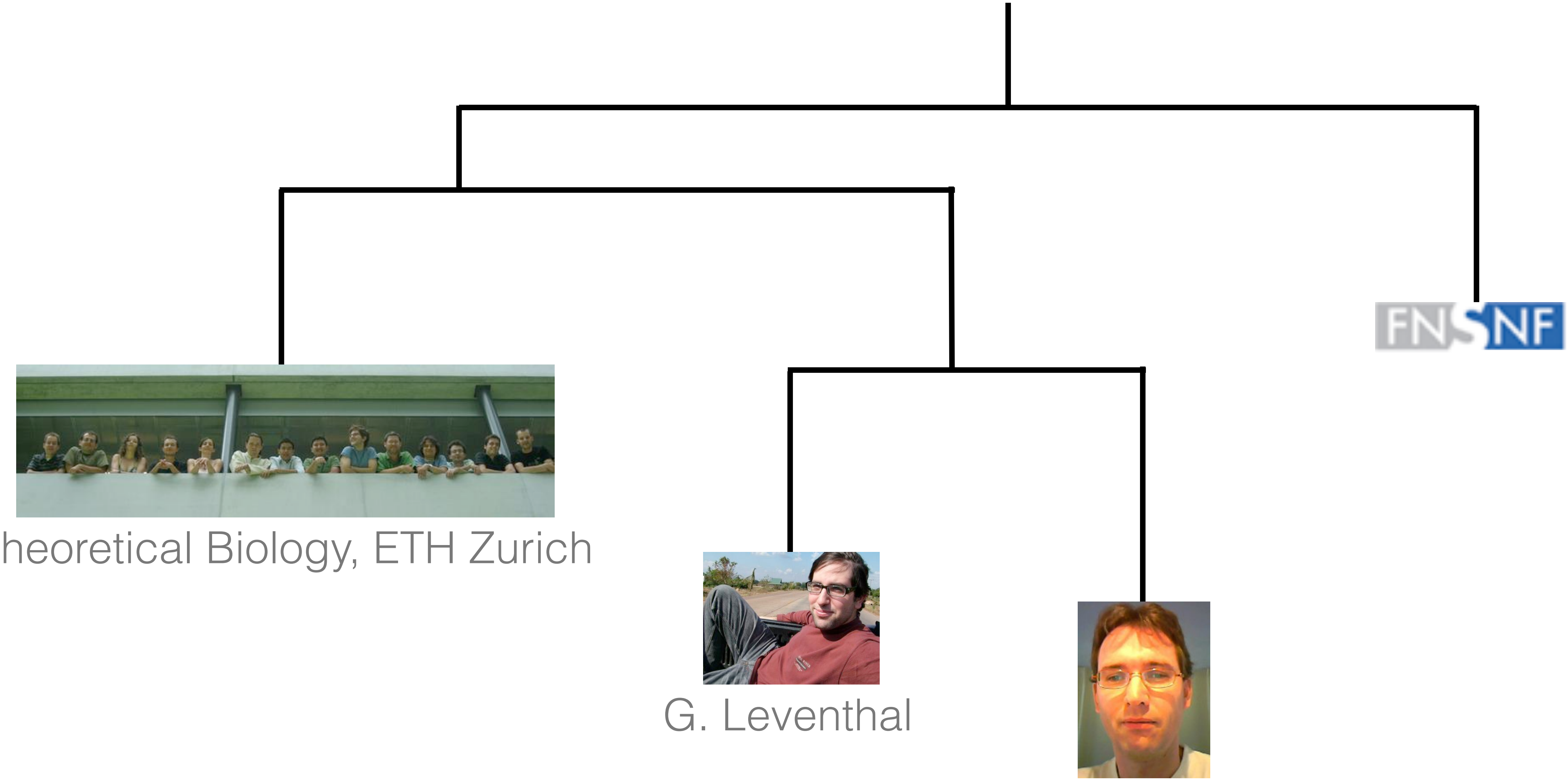
- ...available in the R packages TreeSim and TreePar

Can we calculate the likelihood faster?

At which taxonomic level do species compete?

Can we allow for recovered & immune hosts (SIR)?

Phylogeny of Acknowledgements



Theoretical Biology, ETH Zurich

G. Leventhal

R. Etienne