



National Institute for Public Health
and the Environment
Ministry of Health, Welfare and Sport

Phylodynamic analysis with BEAST

16th of June, 2023

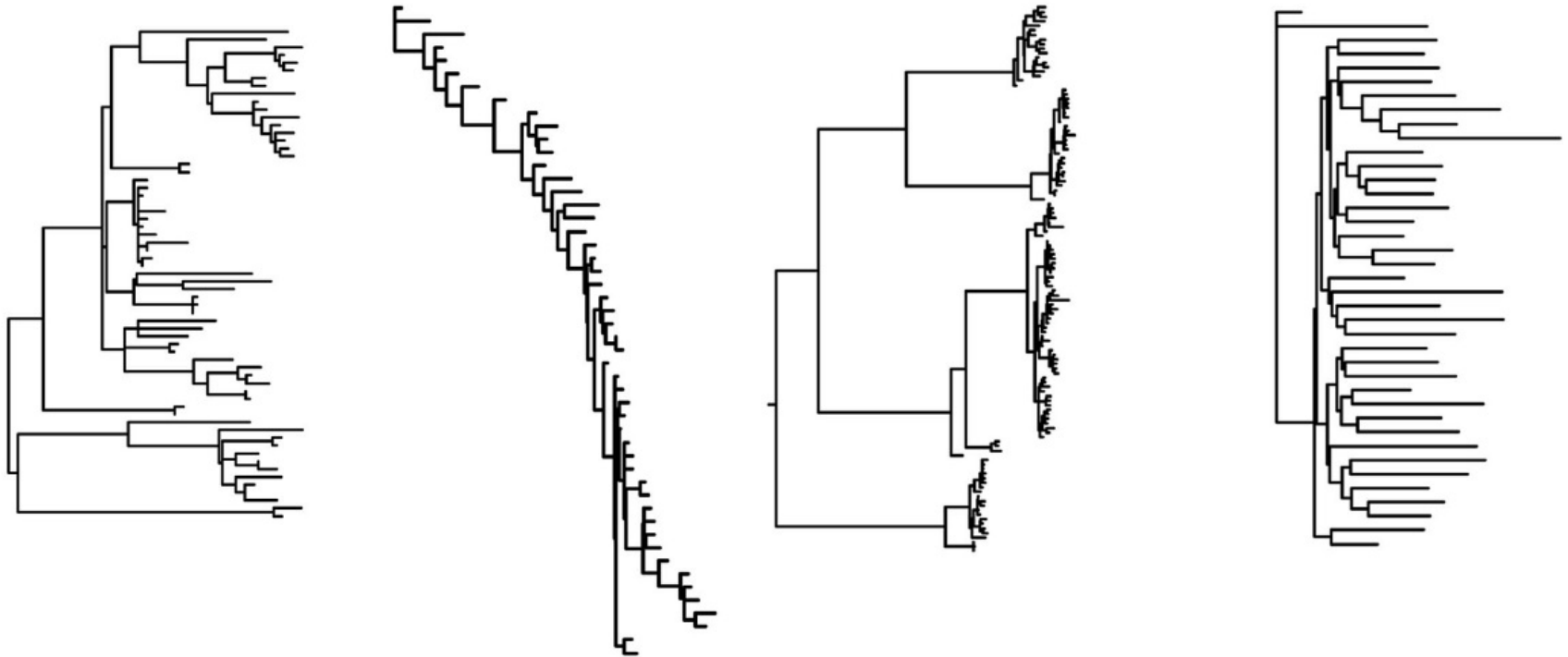
Schedule



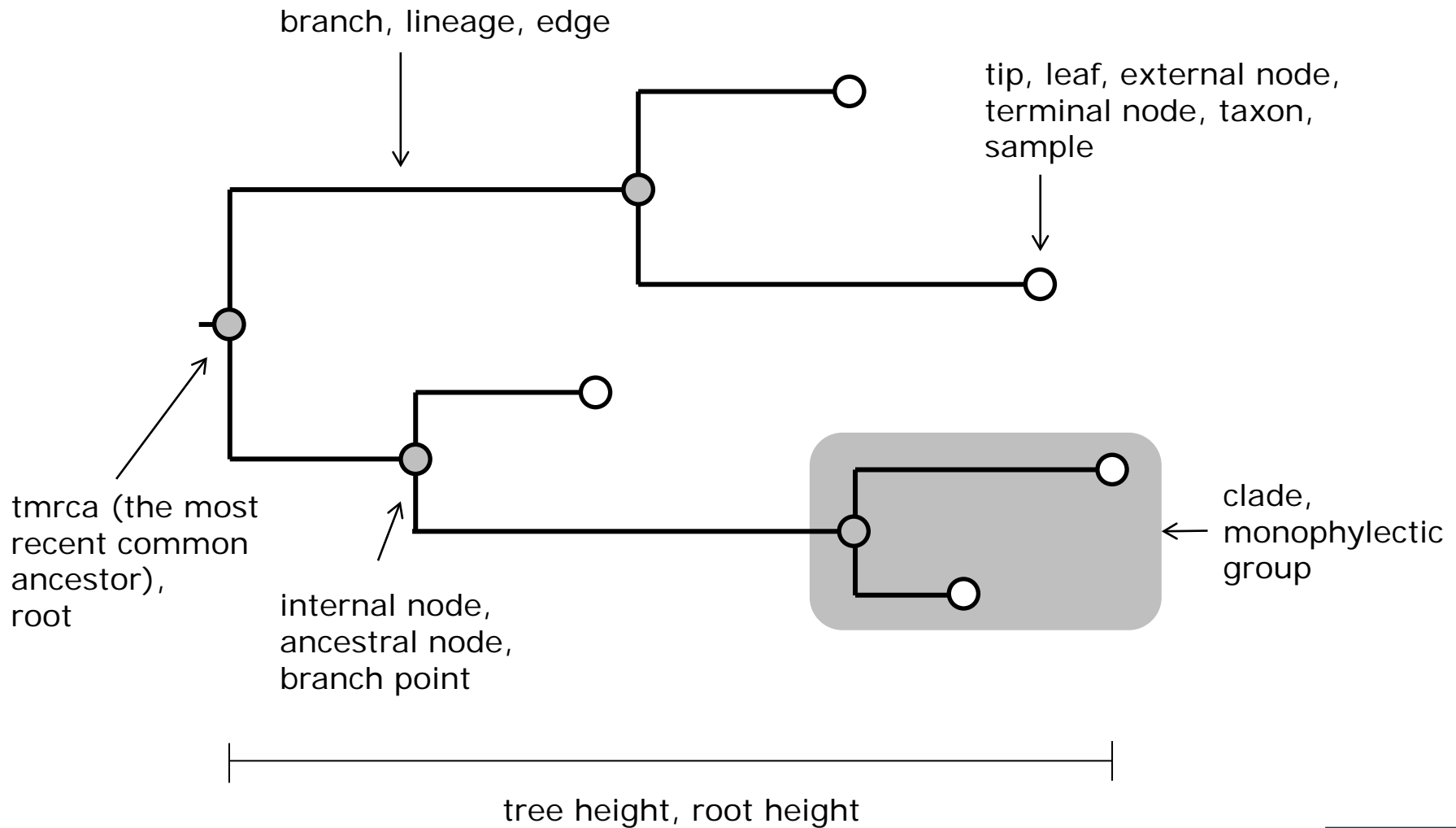
time	subject
9:00	introduction phylodynamics
9:30	tutorial Yellow fever virus <ul style="list-style-type: none">• substitution model• molecular clock model
10:30	break
11:00	tutorial Hepatitis B virus <ul style="list-style-type: none">• population model
12:00	DIY
12:45	wrap-up
13:15	end

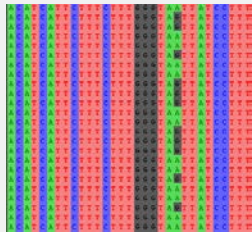


The study of how epidemiological, immunological, and evolutionary processes act and potentially interact to shape phylogenies

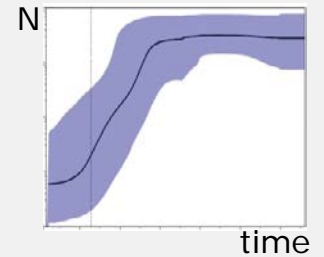


Phylogenies



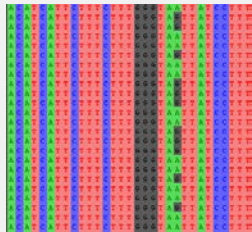


sequence
alignment

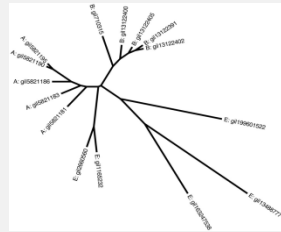


population
characteristics

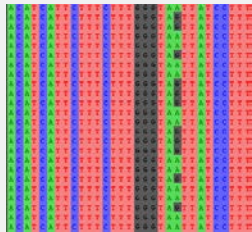
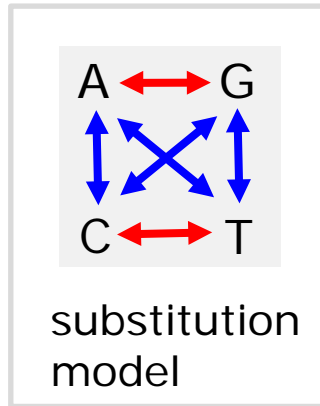
Phylodynamics



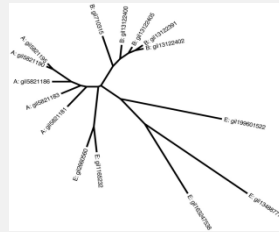
sequence
alignment



phylogeny
(branch lengths
in subst/site)



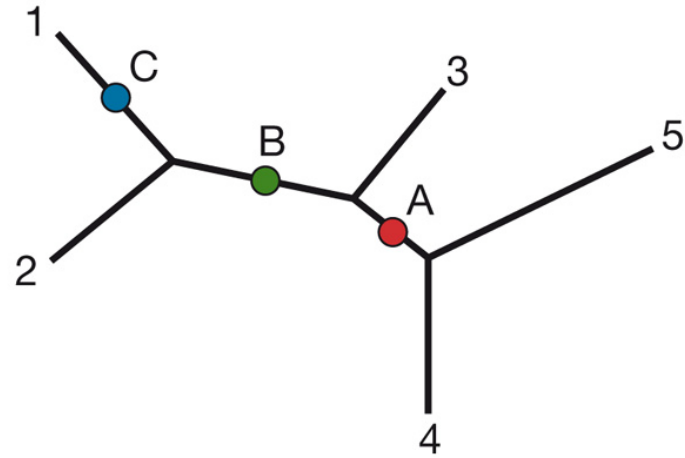
sequence
alignment



phylogeny
(branch lengths
in subst/site)



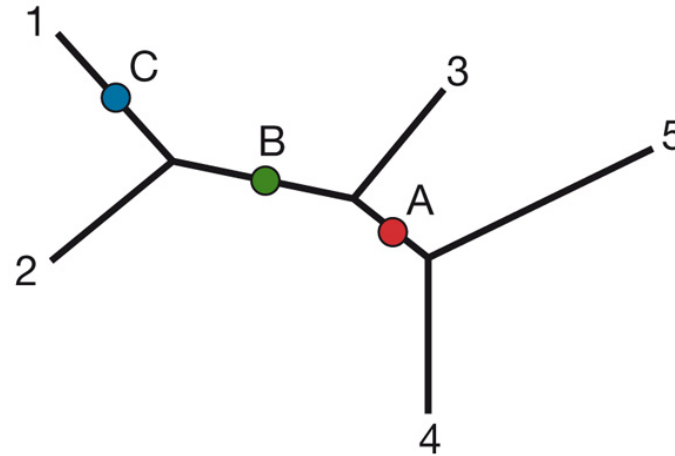
Unrooted tree



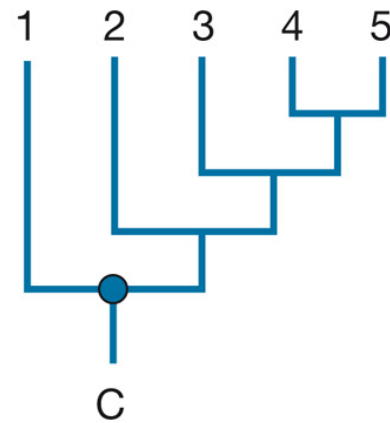
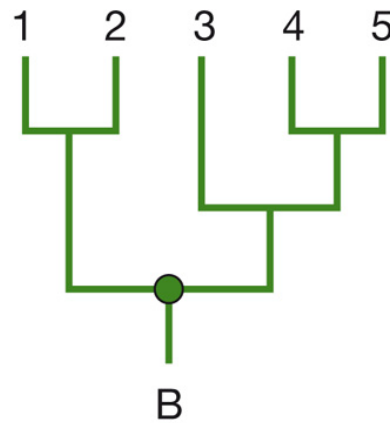
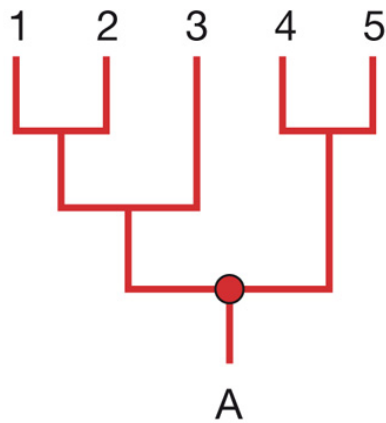
Phylogenies



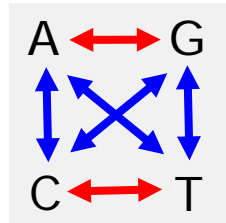
Unrooted tree



Rooted trees



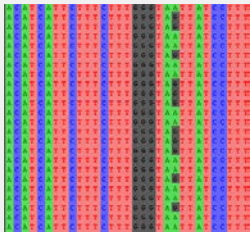
Phylodynamics



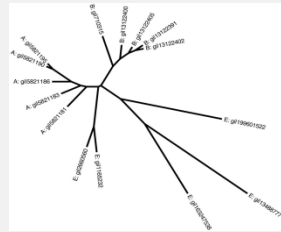
substitution
model



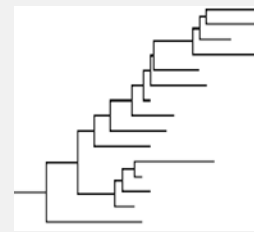
molecular
clock model



sequence
alignment

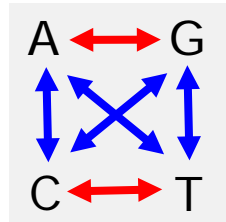


phylogeny
(branch lengths
in subst/site)



phylogeny
(branch lengths
in time units)

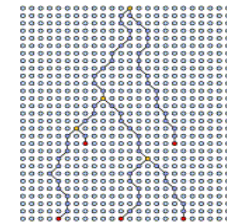
Phylodynamics



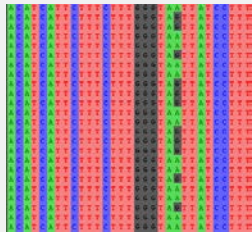
substitution
model



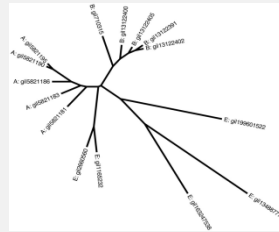
molecular
clock model



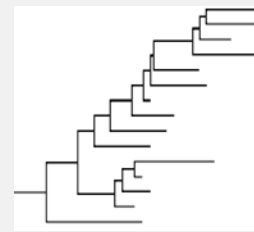
population
model



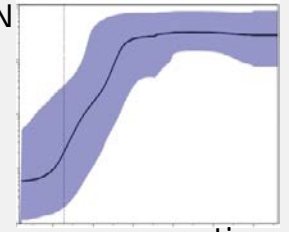
sequence
alignment



phylogeny
(branch lengths
in subst/site)

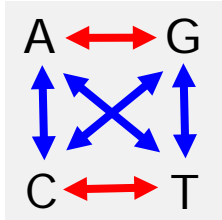


phylogeny
(branch lengths
in time units)



population
characteristics

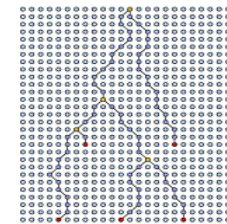
Phylodynamics



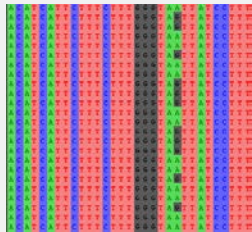
substitution
model



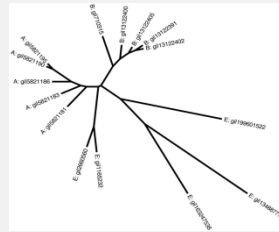
molecular
clock model



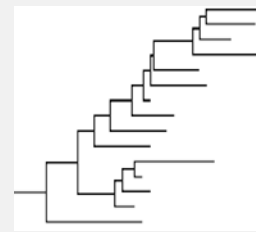
population
model



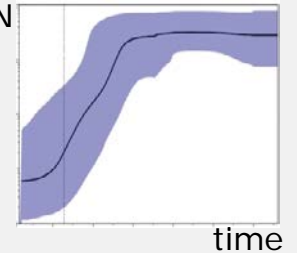
sequence
alignment



phylogeny
(branch lengths
in subst/site)



phylogeny
(branch lengths
in time units)



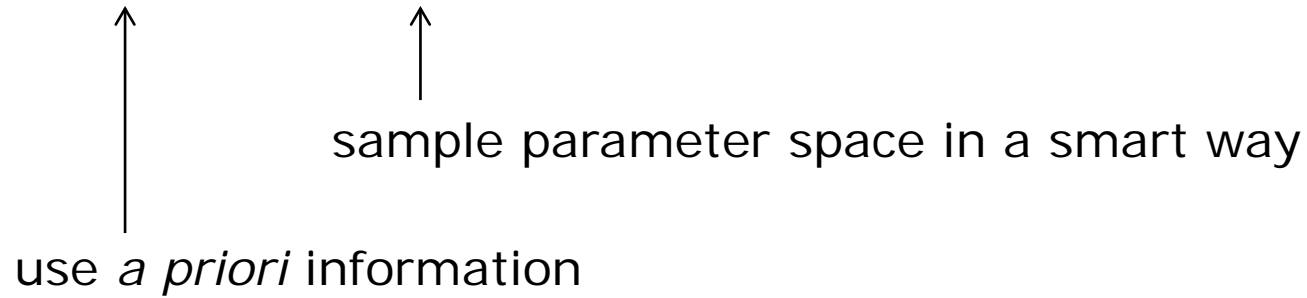
population
characteristics

Bayesian MCMC



sample parameter space in a smart way

Bayesian MCMC



Bayesian MCMC

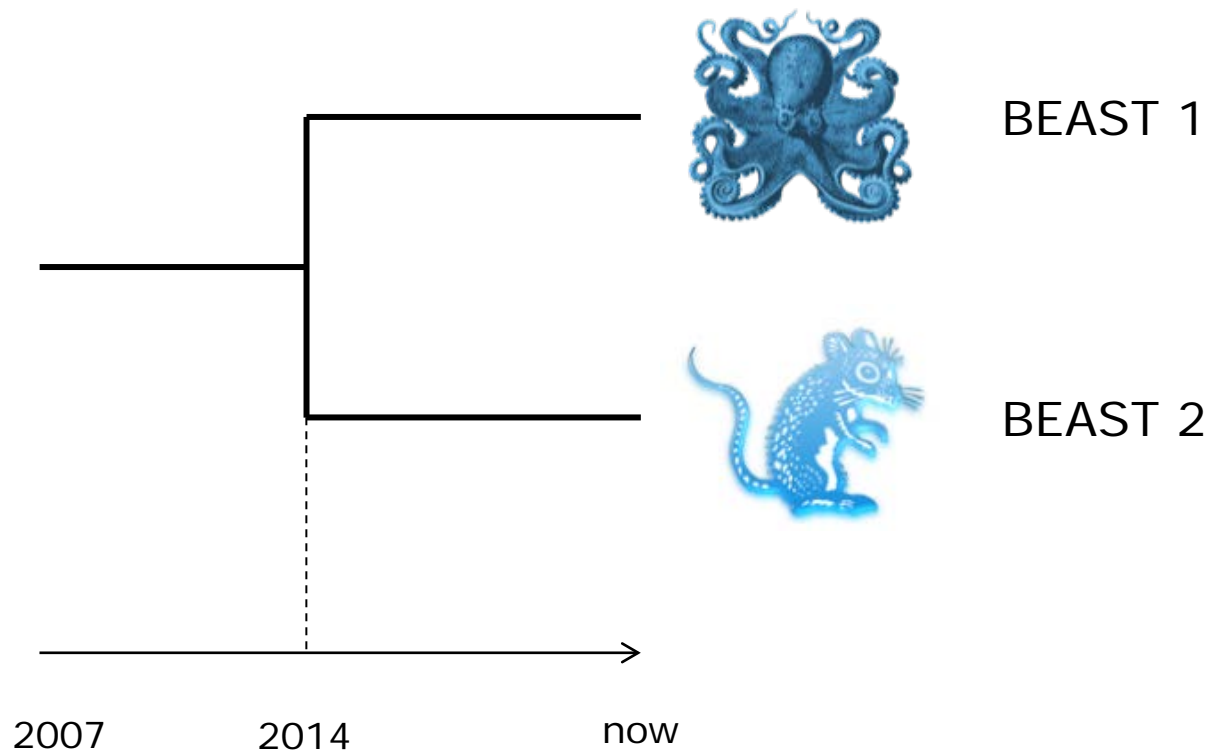


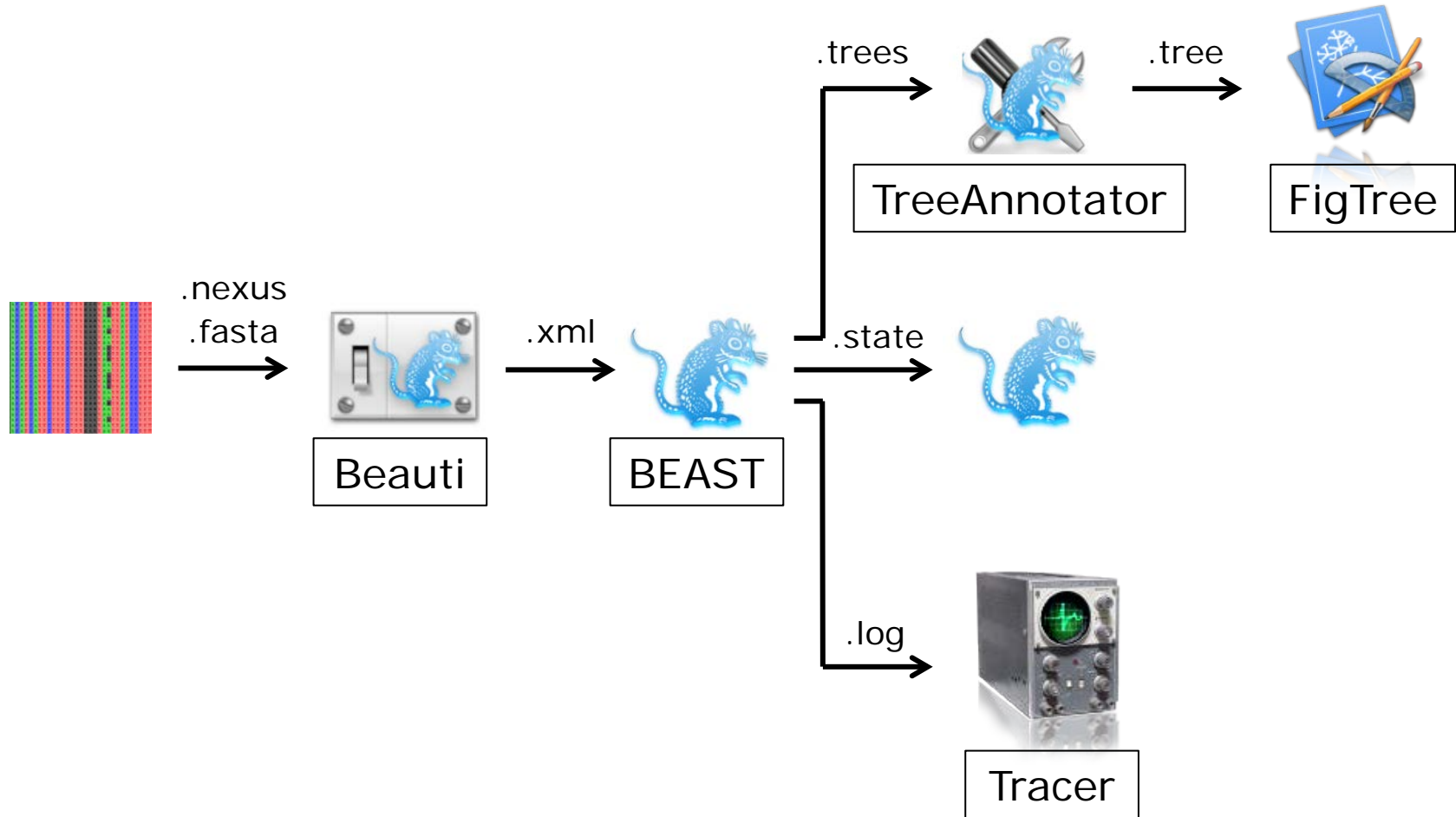
↑
use *a priori* information

↑
sample parameter space in a smart way

$$\overbrace{p(\text{par}|\text{data})}^{\text{posterior}} \propto \underbrace{p(\text{data}|\text{par})}_{\text{likelihood}} \underbrace{p(\text{par})}_{\text{prior}}$$

Bayesian Evolutionary Analysis Sampling Trees



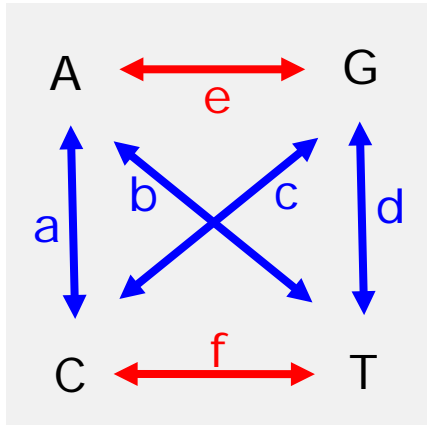




Tutorial 1

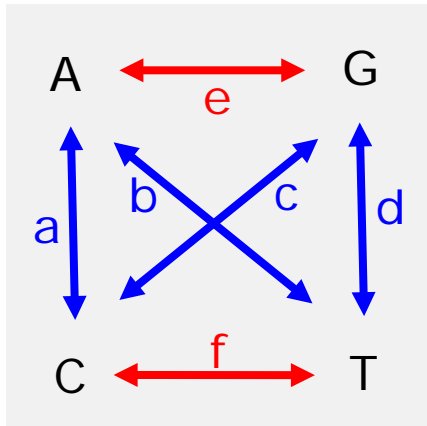
Yellow fever virus

Substitution models



model	relative rates	fractions	#par
JC69	$a=b=c=d=e=f$	$\pi_A=\pi_C=\pi_G=\pi_T$	1
⋮	⋮	⋮	⋮
HKY85	$a=b=c=d \quad e=f$	$\pi_A \neq \pi_C \neq \pi_G \neq \pi_T$	5
⋮	⋮	⋮	⋮
GTR	$a \neq b \neq c \neq d \neq e \neq f$	$\pi_A \neq \pi_C \neq \pi_G \neq \pi_T$	9

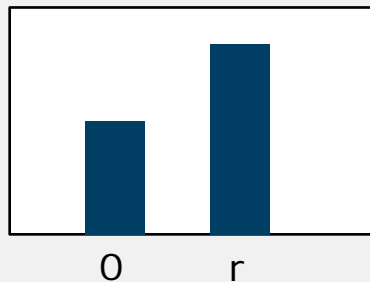
Substitution models



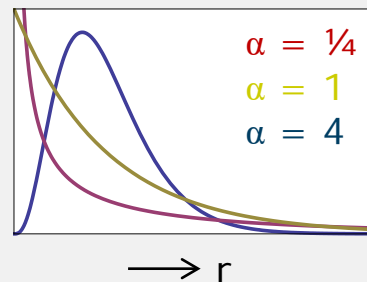
model	relative rates	fractions	#par
JC69	$a=b=c=d=e=f$	$\pi_A=\pi_C=\pi_G=\pi_T$	1
⋮	⋮	⋮	⋮
HKY85	$a=b=c=d \quad e=f$	$\pi_A \neq \pi_C \neq \pi_G \neq \pi_T$	5
⋮	⋮	⋮	⋮
GTR	$a \neq b \neq c \neq d \neq e \neq f$	$\pi_A \neq \pi_C \neq \pi_G \neq \pi_T$	9

take heterogeneity between sites into account:

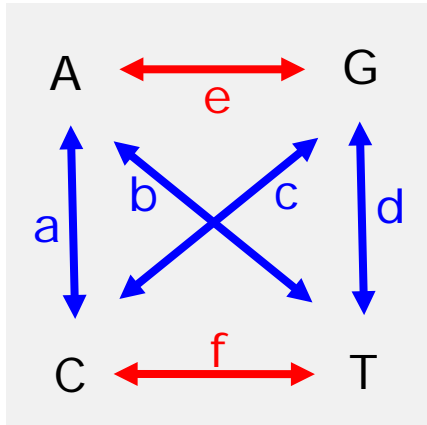
proportion invariant



gamma distributed



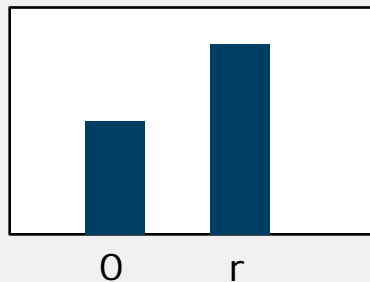
Substitution models



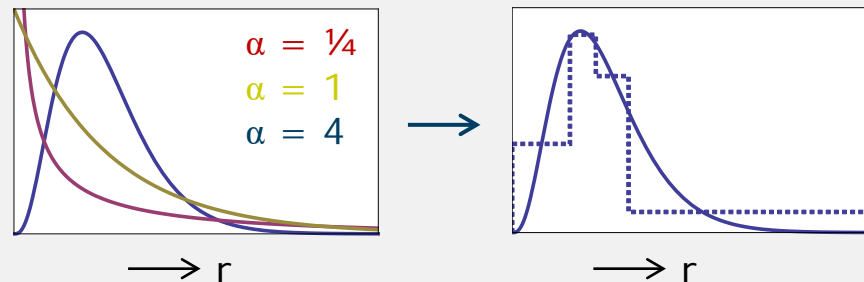
model	relative rates	fractions	#par
JC69	$a=b=c=d=e=f$	$\pi_A=\pi_C=\pi_G=\pi_T$	1
⋮	⋮	⋮	⋮
HKY85	$a=b=c=d \quad e=f$	$\pi_A \neq \pi_C \neq \pi_G \neq \pi_T$	5
⋮	⋮	⋮	⋮
GTR	$a \neq b \neq c \neq d \neq e \neq f$	$\pi_A \neq \pi_C \neq \pi_G \neq \pi_T$	9

take heterogeneity between sites into account:

proportion invariant



gamma distributed, discretized in categories:



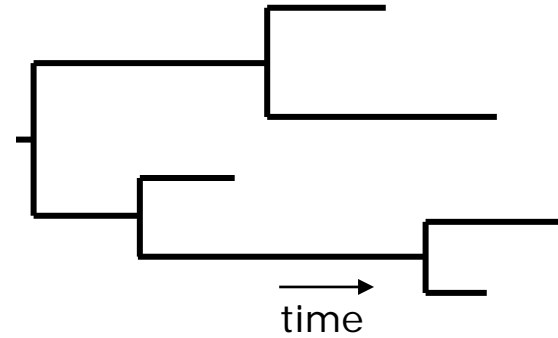
Molecular clock model



serially sampled tips +



=



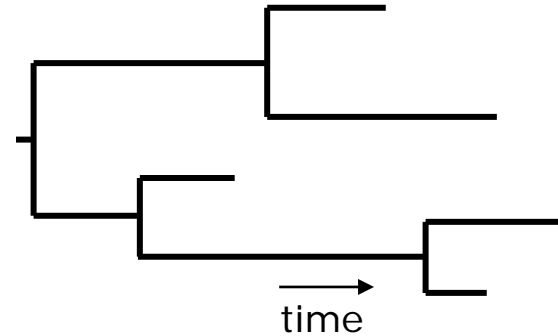
Molecular clock model



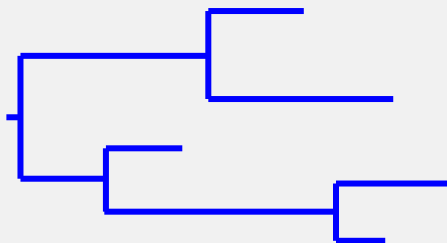
serially sampled tips +



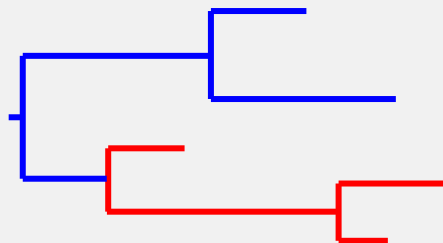
=



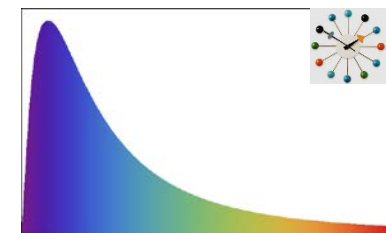
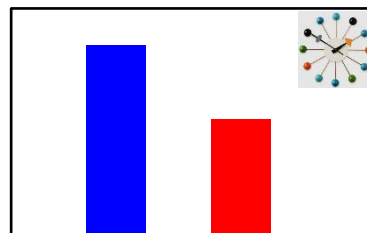
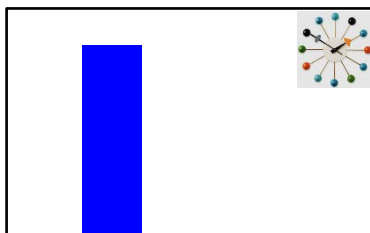
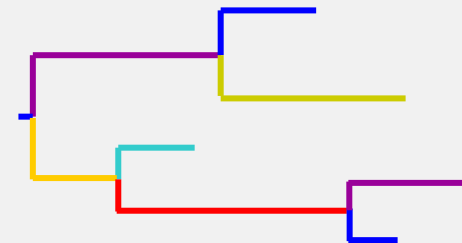
strict



random local



relaxed



Yellow fever virus



Data:

- 36 taxa of YFV
- serially sampled over period 1940 – 2009

In BEAUti:

- Use alignment: YFV.nex
- Substitution model: HKY with rate heterogeneity
- Molecular clock: Strict
- Population model: Coalescent constant population
- 2 mln MCMC steps
- Log and trees every 1000 steps
- Screenlog every 10000 steps



Tutorial 2

Hepatitis B virus

Hepatitis B virus



Data:

- 42 taxa of MSM, diagnosed with acute HBV
- serially sampled over period 1992 – 2006
- S-gene sequenced (661 nucleotides)

In BEAUti:

- Use alignment: HBV.fasta
- Substitution model: HKY with rate heterogeneity
- Molecular clock: strict
- Population model: ...

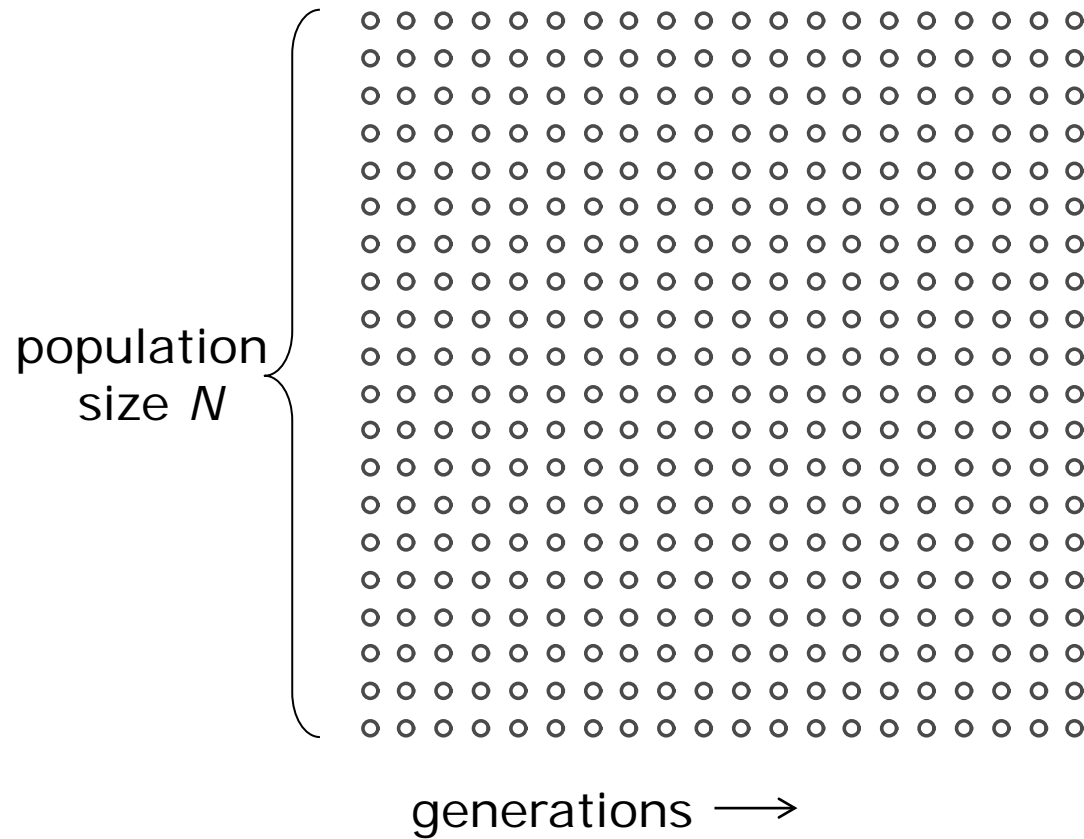
Coalescent theory



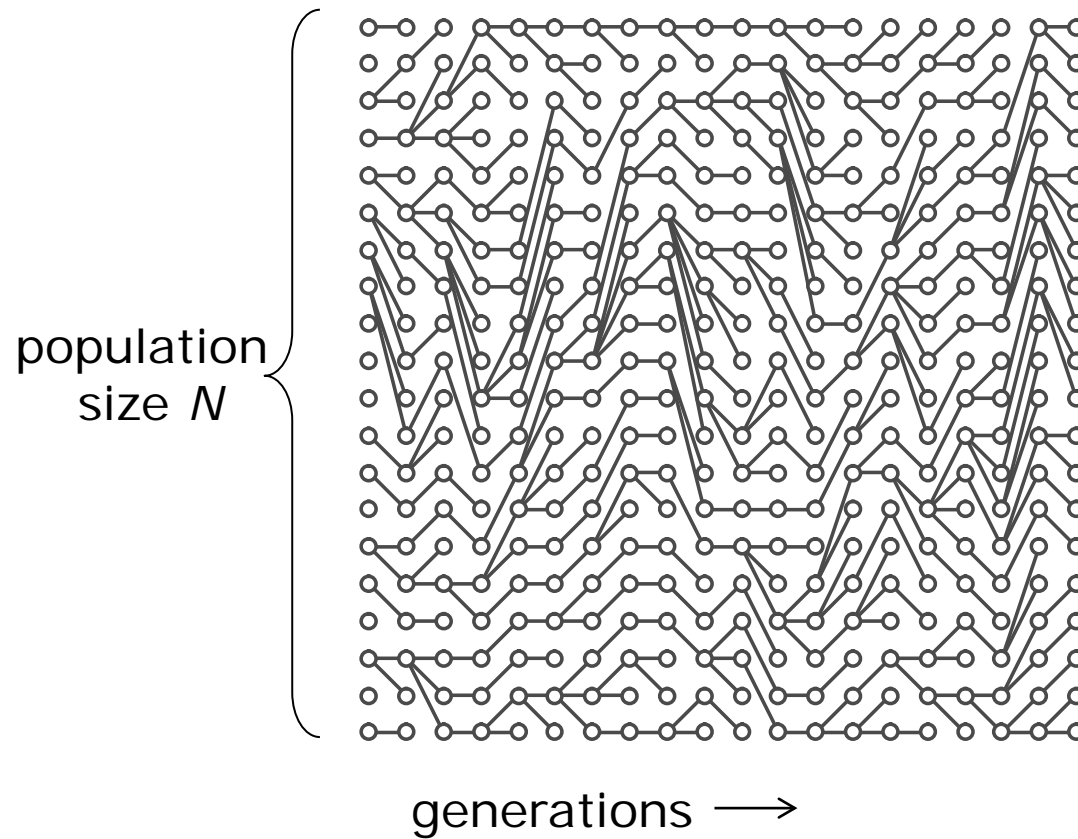
Coalescent theory



Coalescent theory

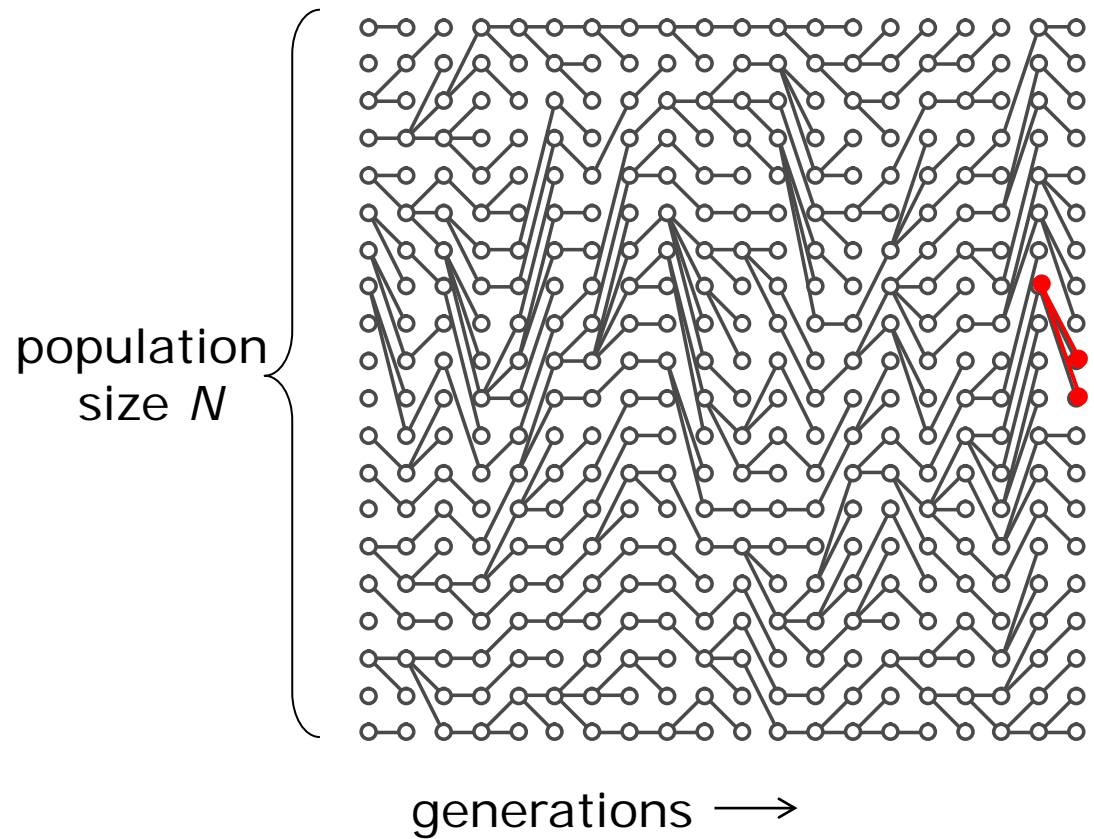


Coalescent theory



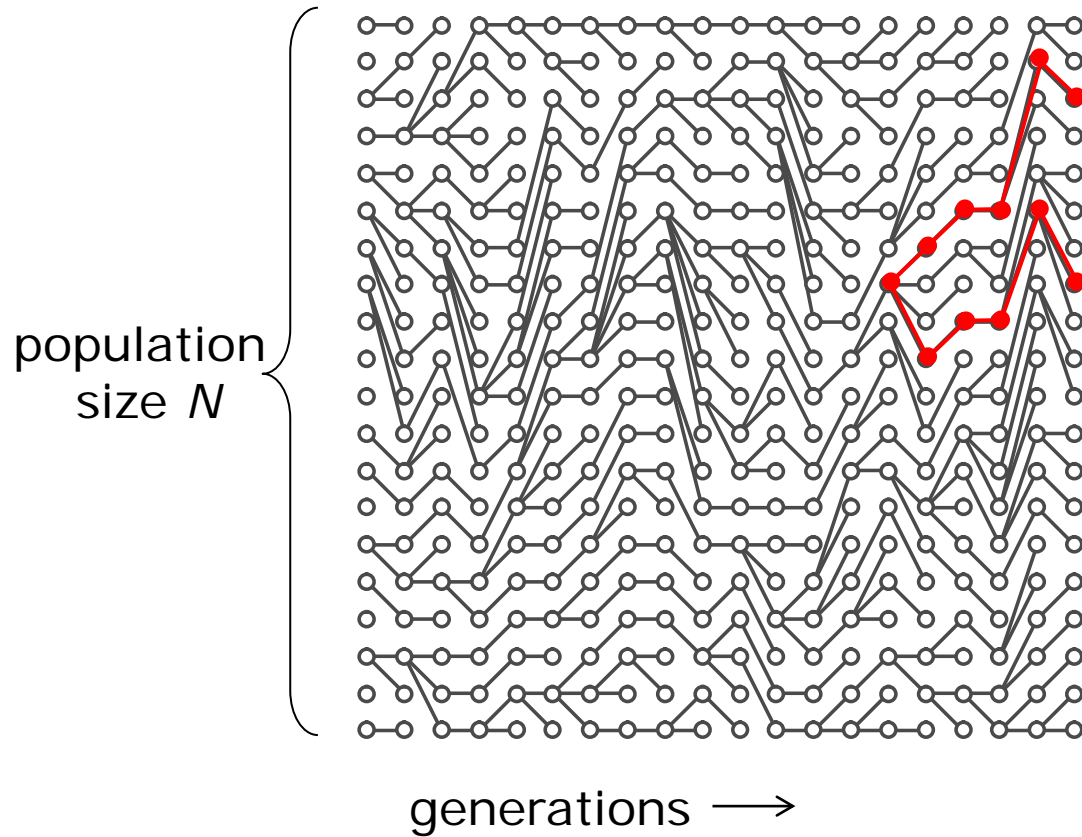
Wright-Fisher population (Wright S (1931) Evolution in Mendelian populations, *Genetics*)

Coalescent theory



probability of 2 individuals
coalescing in 1 generation:
 $\frac{1}{N}$

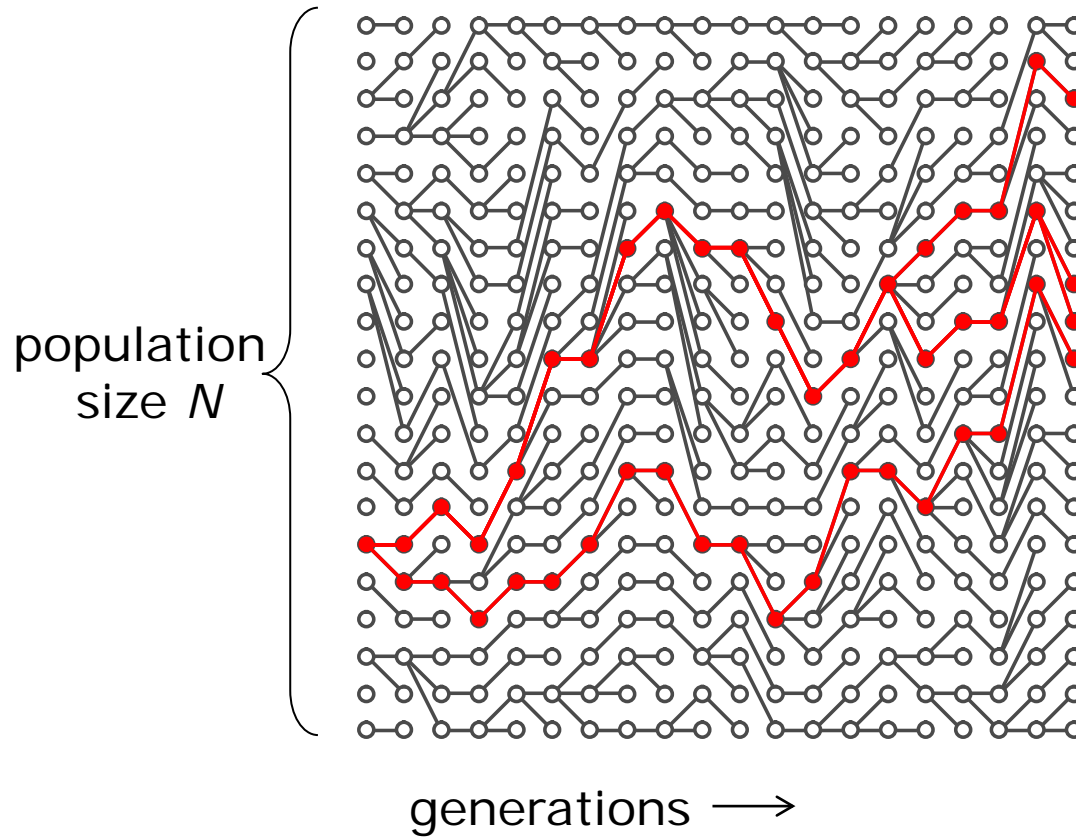
Coalescent theory



probability of 2 individuals
coalescing in j generations:

$$\frac{1}{N} \left(1 - \frac{1}{N}\right)^{j-1}$$

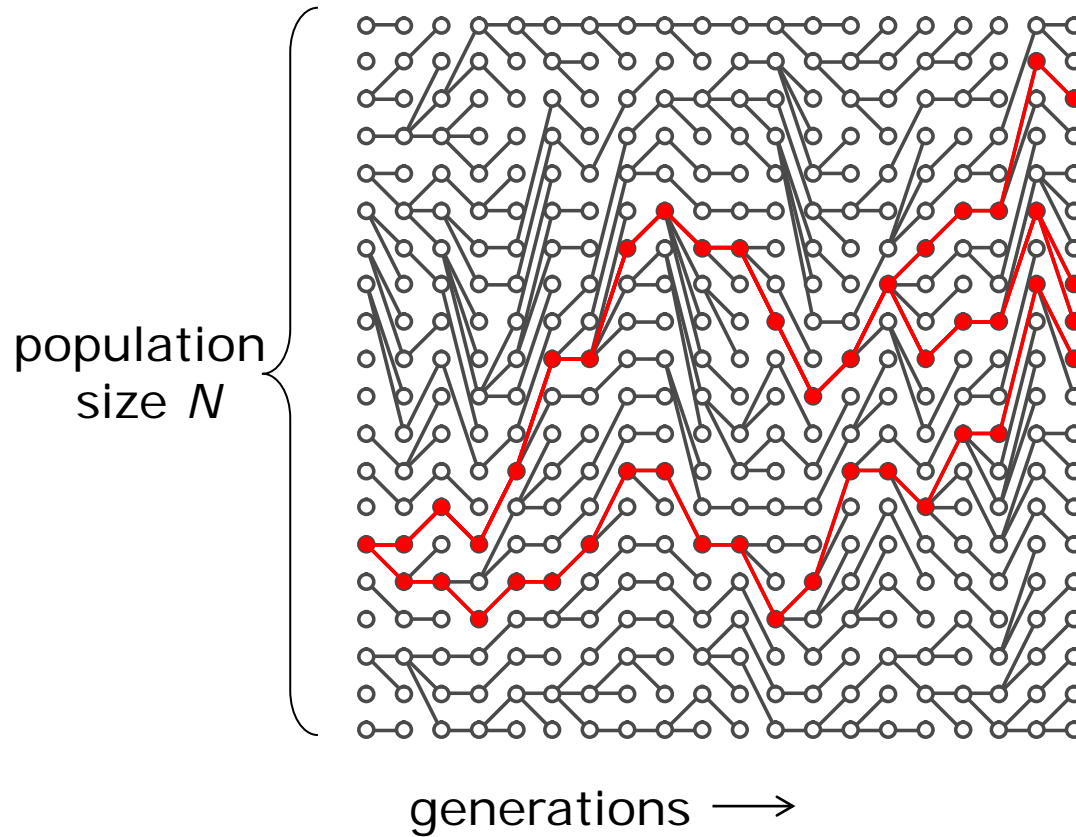
Coalescent theory



probability of 2 of k individuals
coalescing in j generations:

$$\binom{k}{2} \frac{1}{N} \left(1 - \binom{k}{2} \frac{1}{N} \right)^{j-1}$$

Coalescent theory



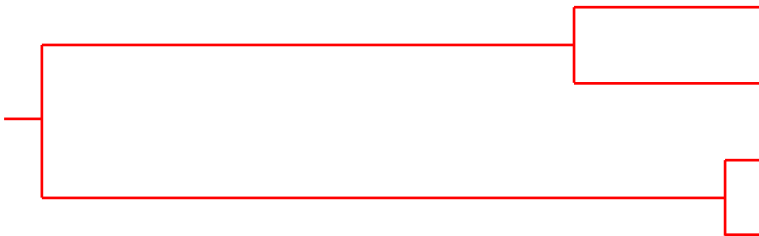
probability of 2 of k individuals
coalescing in j generations:

$$\binom{k}{2} \frac{1}{N} \left(1 - \binom{k}{2} \frac{1}{N} \right)^{j-1}$$

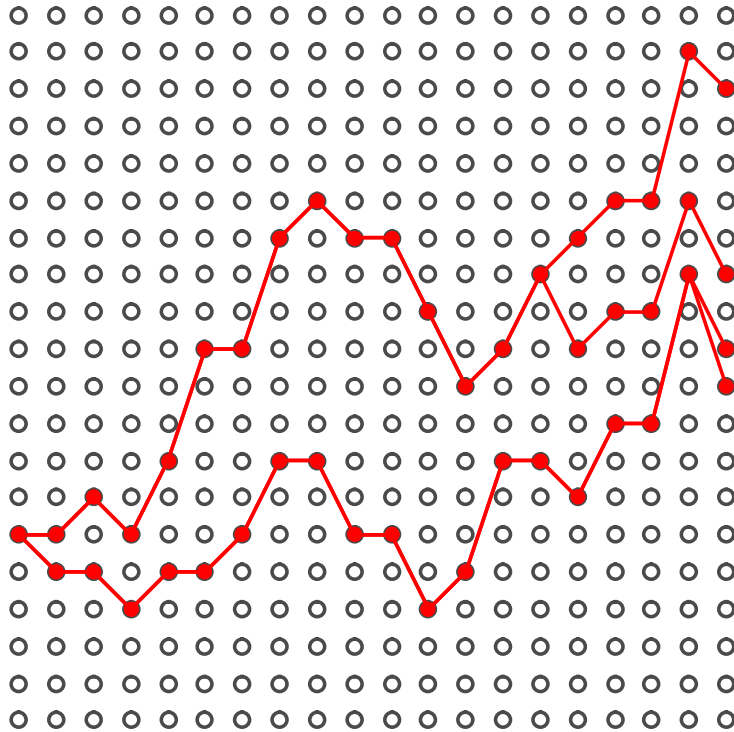
in continuous time and
large population:

$$1 - e^{-\binom{k}{2} t}$$

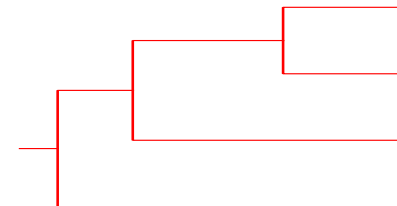
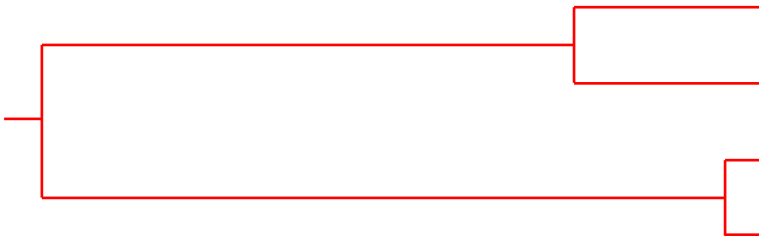
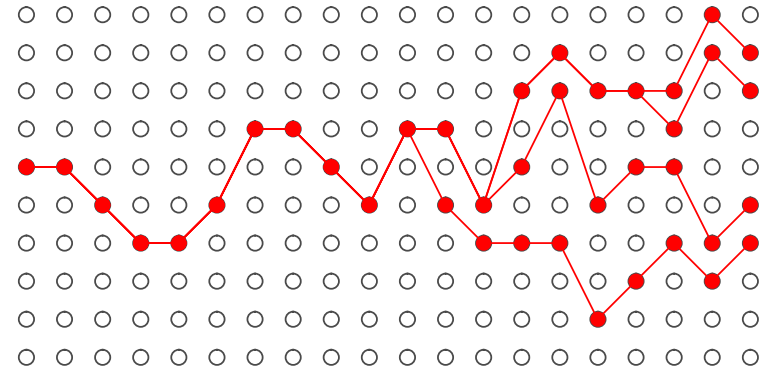
Coalescent theory



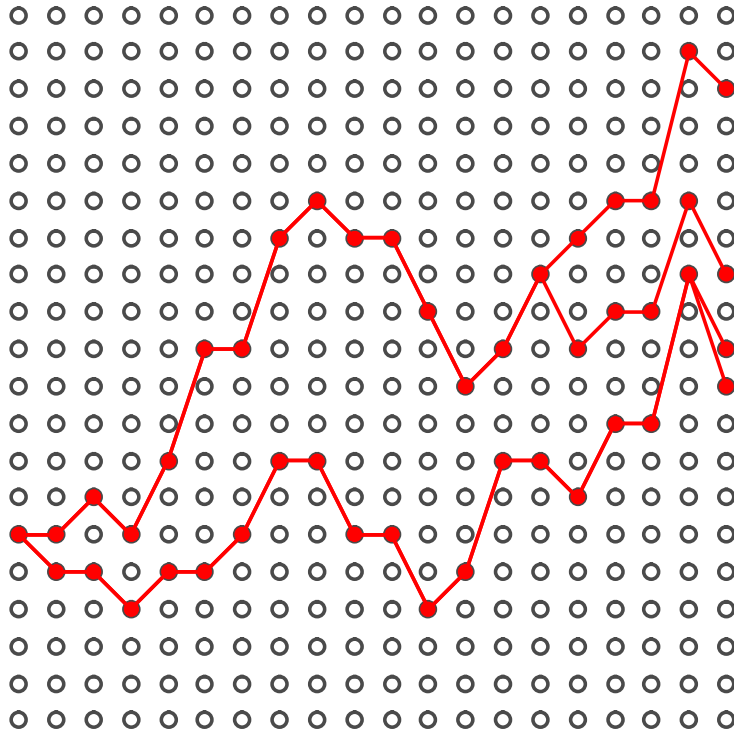
Coalescent theory



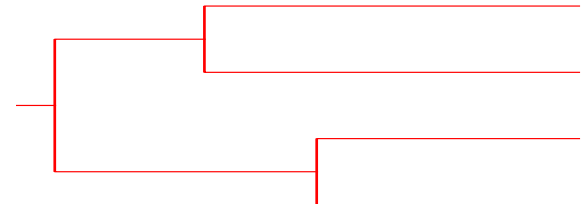
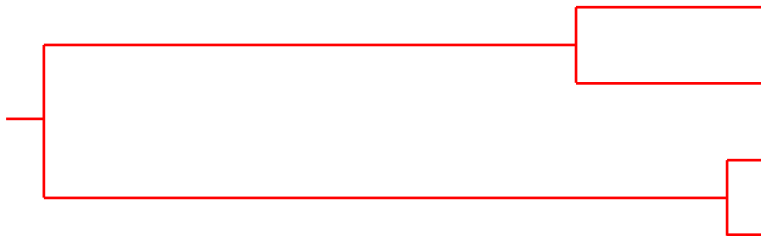
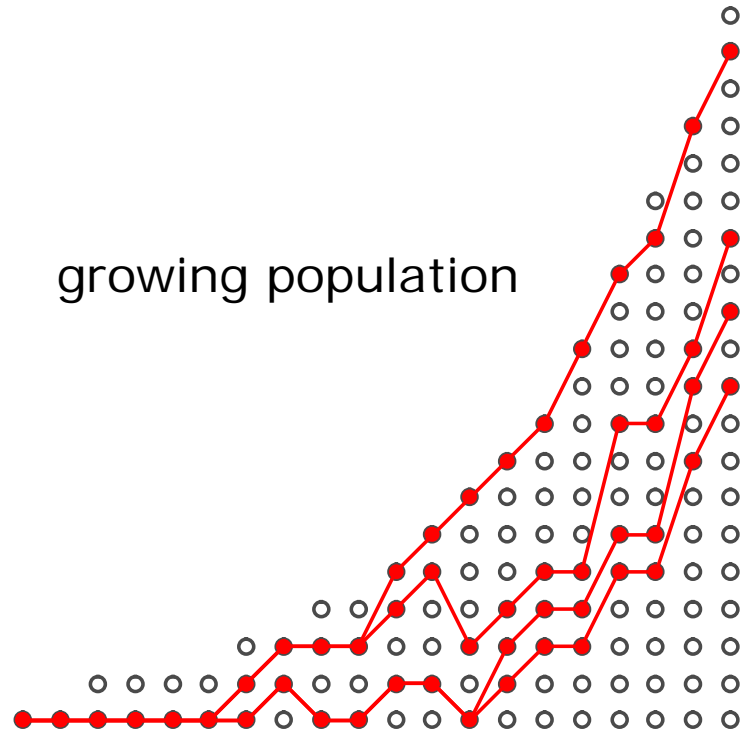
smaller population



Coalescent theory



growing population





Effective population size N governs coalescence rate

Size and history of the population shape phylogeny

Coalescent theory

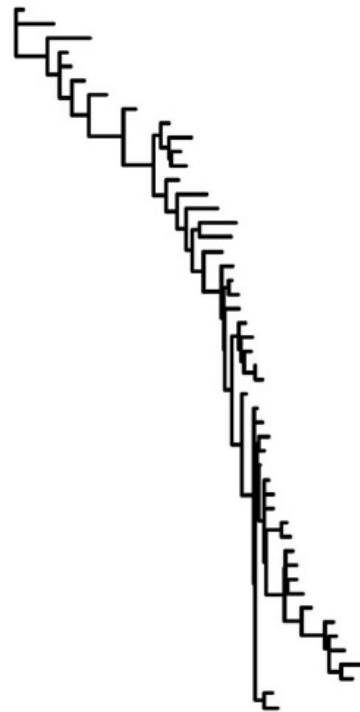


Effective population size N governs coalescence rate

Size and history of the population shape phylogeny



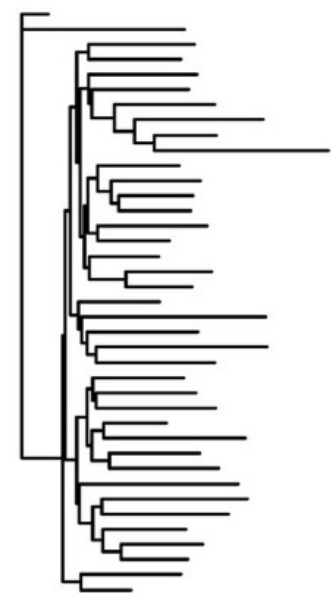
measles



influenza A



dengue

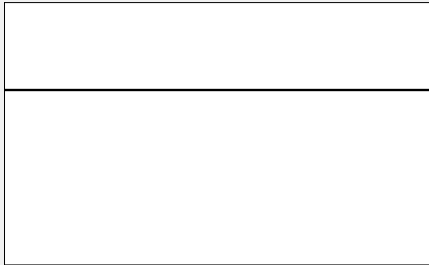


HIV

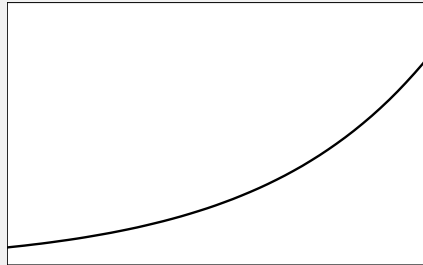
Coalescent models



constant population



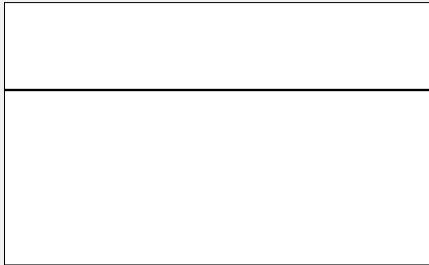
exponential growth



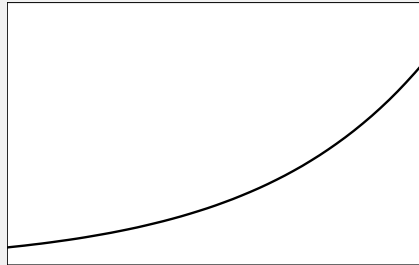
Coalescent models



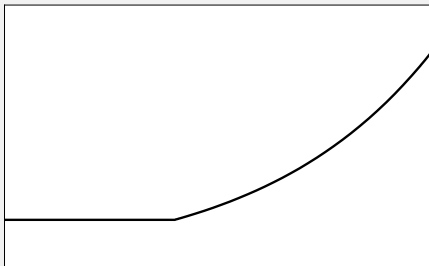
constant population



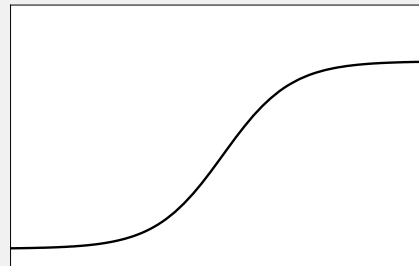
exponential growth



expansion growth*



logistic growth*

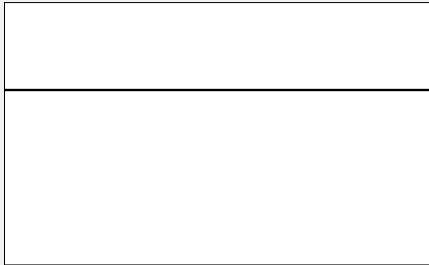


*only in BEAST 1

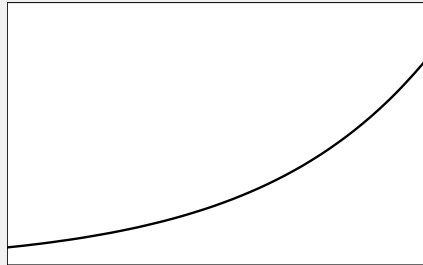
Coalescent models



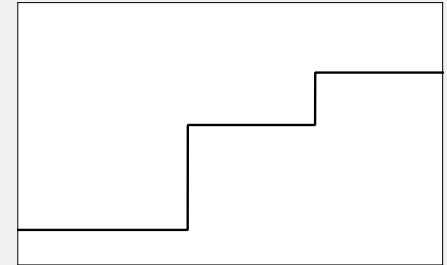
constant population



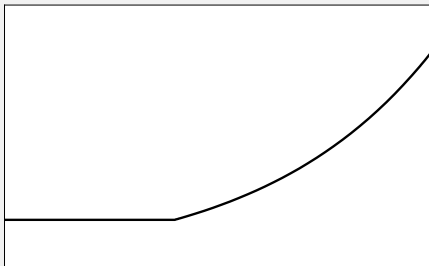
exponential growth



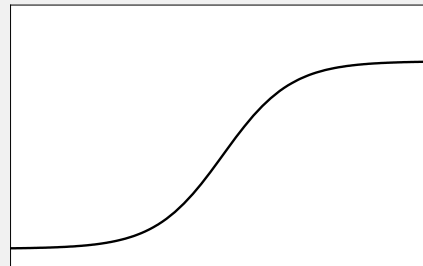
Bayesian Skyline



expansion growth*



logistic growth*

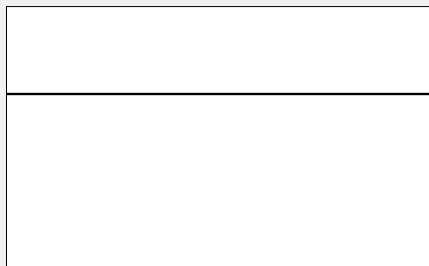


*only in BEAST 1

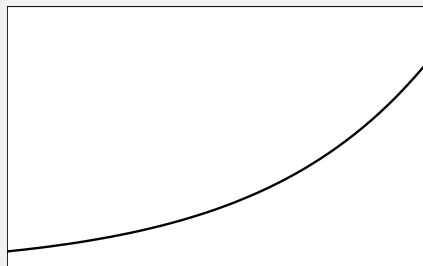
Coalescent models



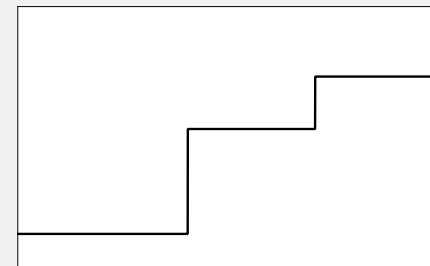
constant population



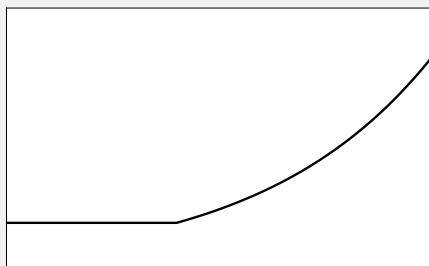
exponential growth



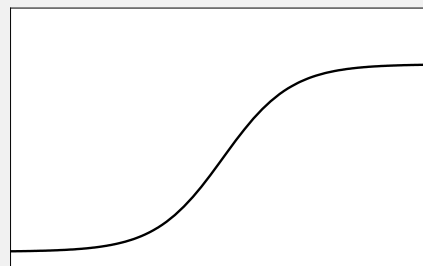
Bayesian Skyline



expansion growth*



logistic growth*



Extended Bayesian Skyline

Bayesian Skyride*

Bayesian Skygrid*

*only in BEAST 1

Hepatitis B virus



Data:

- 42 taxa of MSM, diagnosed with acute HBV
- serially sampled over period 1992 – 2006
- S-gene sequenced (661 nucleotides)

In BEAUti:

- Use alignment: HBV.fasta
- Substitution model: HKY with rate heterogeneity
- Molecular clock: strict
- Population model: Coalescent Bayesian Skyline
- 10 mln MCMC steps
- Log and trees every 5000 steps
- Screenlog every 10000 steps



DIY

© ERIC SCHICKLER PHOTOGRAPHY



Data Influenza A H1N1:

- 50 taxa of Influenza A H1N1 (pandemic strain)
- serially sampled in 2009

Data SARS-CoV-2:

- 47 taxa of 2019-nCoV (original name of SARS-CoV-2)
- serially sampled in December 2019 – January 2020

Questions:

- Which population model is appropriate?
- What is the timing of the MRCA (estimate + uncertainty)?



Reviews on viral phylodynamics:

- Grenfell, Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 2004
- Pybus & Rambaut, Evolutionary analysis of the dynamics of viral infectious disease. *Nature* 2009
- Volz, Koelle & Bedford, Viral phylodynamics. *PLoS Comp Biol* 2013
- Bromham et al., Bayesian molecular dating: opening up the black box. *Biol. Rev.* 2017

Tutorials BEAST(2)

- rega.kuleuven.be/cev/ecv/tutorials (BEAST1)
- beast.community (BEAST1)
- beast2.org/tutorials (BEAST2)
- taming-the-beast.github.io/tutorials/ (BEAST2)

Forum BEAST:

- Google group [beast-users](https://groups.google.com/forum/#!forum/beast-users)