

MAXIMUM LIKELIHOOD, ML (maximální věrohodnost)

15 coin tosses:

→ score TTHHHTHTTTHTHHT
tj. 7× head (H), 8× tail (T)



Likelihood = conditional probability of data (final score)
given the hypothesis:

$$L = \Pr(D \mid H) = \Pr(7 \times \text{head}, 8 \times \text{tail} \mid \text{hypothesis})$$

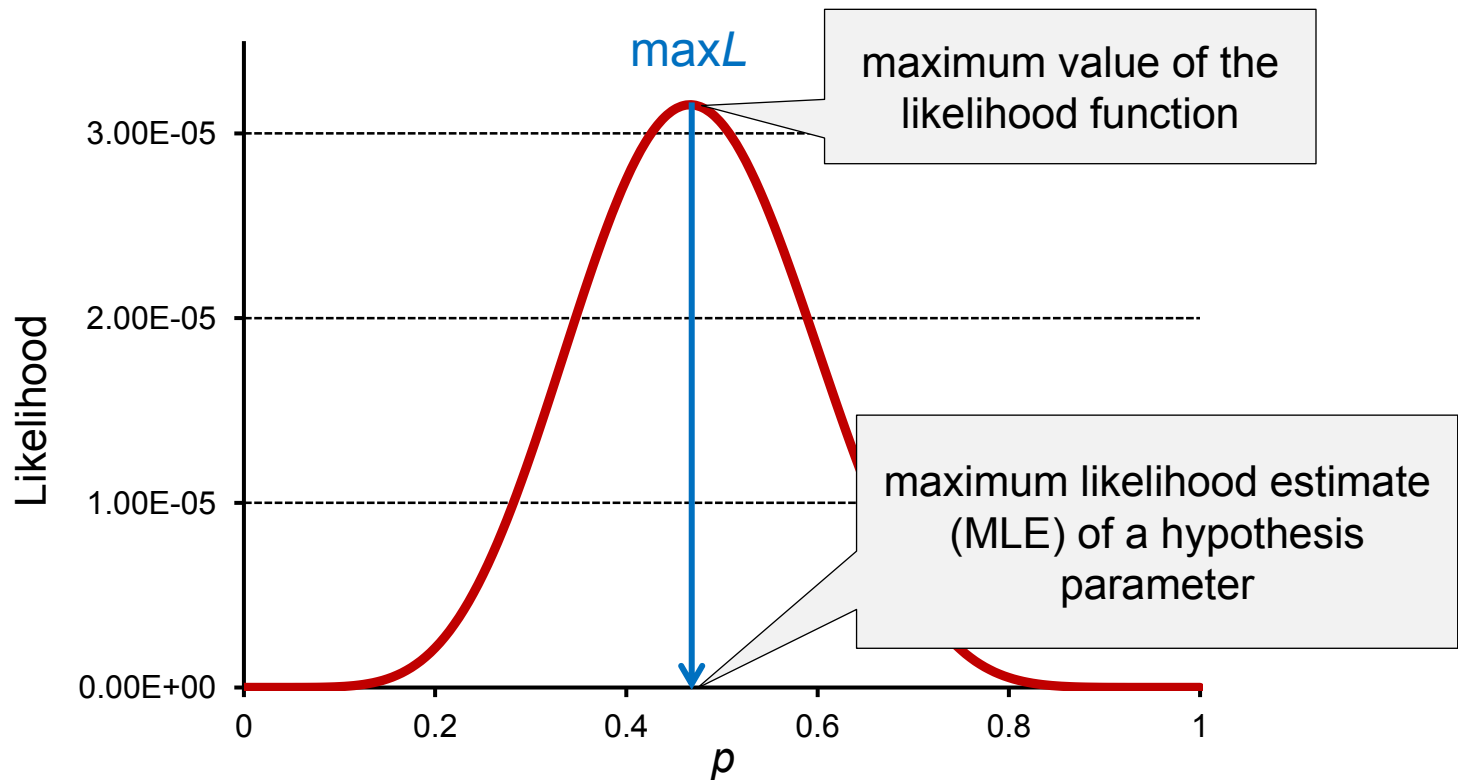
Probability of head = p , tail = $(1 - p)$

score TTHHHTHTTTHTHHT [7× head (H), 8× tail (T)]



Because tosses independent \Rightarrow probability of final score =
 $(1 - p) \times (1 - p) \times p \times p \times p \times (1 - p) \times p \times (1 - p) \times (1 - p) \times (1 - p) \times p \times (1 - p) \times p \times p \times (1 - p) =$
 $= p^7(1-p)^8$

maximum = $0,4666 \approx 7/15$



Hypothesis?

Eg. $H =$ coin is not „biased“, ie. $p = 1/2 \Rightarrow L = 3,0517 \cdot 10^{-5}$

If the coin is biased so that we get tail in $2/3$ cases:

$$p = 1/3 \Rightarrow L = 1,7841 \cdot 10^{-5}$$

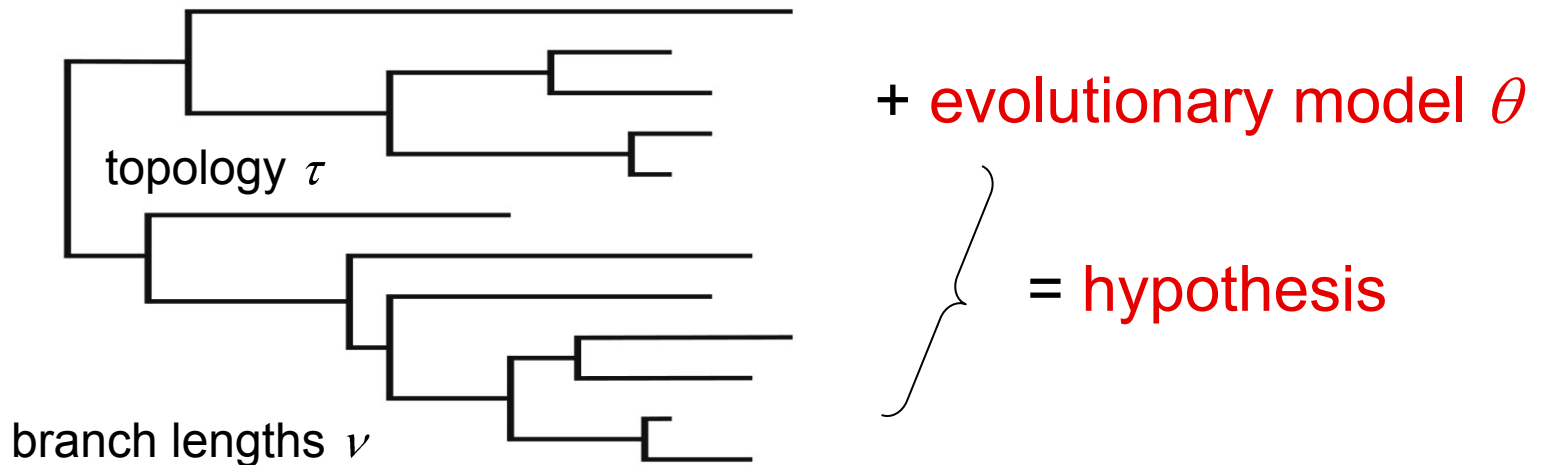
\Rightarrow result of tosses $1,7\times$ more probable with unbiased coin

Maximum likelihood in phylogenetic analysis

data:

```
1 TCAAAAATGGCTTTATTTCGCTTAATGCCGTTAACCTTGCGGGGGCCATG
2 TCCGTGATGGATTTATTTCCGCAATGCCTGTCATCTTATTCTCAAGTATC
3 TTCGTGATGGATTTATTGCAGGTATGCCAGTCATCCTTTTCTCATCTATC
4 TTCGTGACGGGTTTATCTCGGCAATGCCGGTCATCCTATTTTCGAGTATT
```

tree:



$$L = P(D \mid H): D = \text{sequence matrix (data)}, H = \tau (\text{topology}) + \nu (\text{branch lengths}) + \theta (\text{model})$$

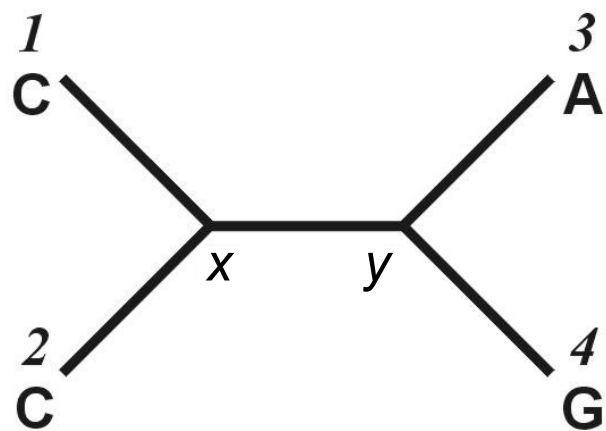
1 N

1 TCAAAAATGGCTTTATTTC**C**TTAATGCCGTTAACCTTGCGGGGGCCATG

2 TCCGTGATGGATTTATTT**C**GCAATGCCTGTCATCTTATTCTCAAGTATC

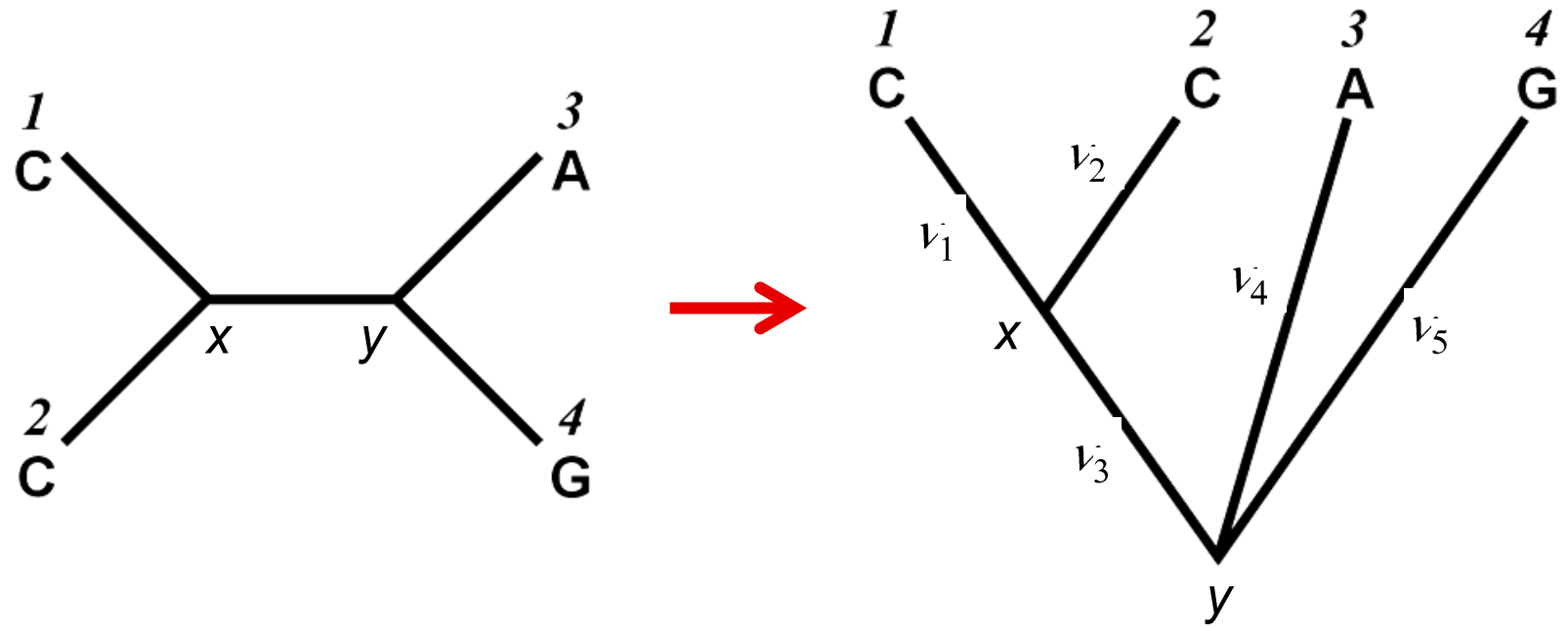
3 TTCGTGATGGATTTATTG**A**GGTATGCCAGTCATCCTTTTCTCATCTATC

4 TTCGTGACGGGTTTATCT**G**GCAATGCCGGTCATCCTATTTTCGAGTATT



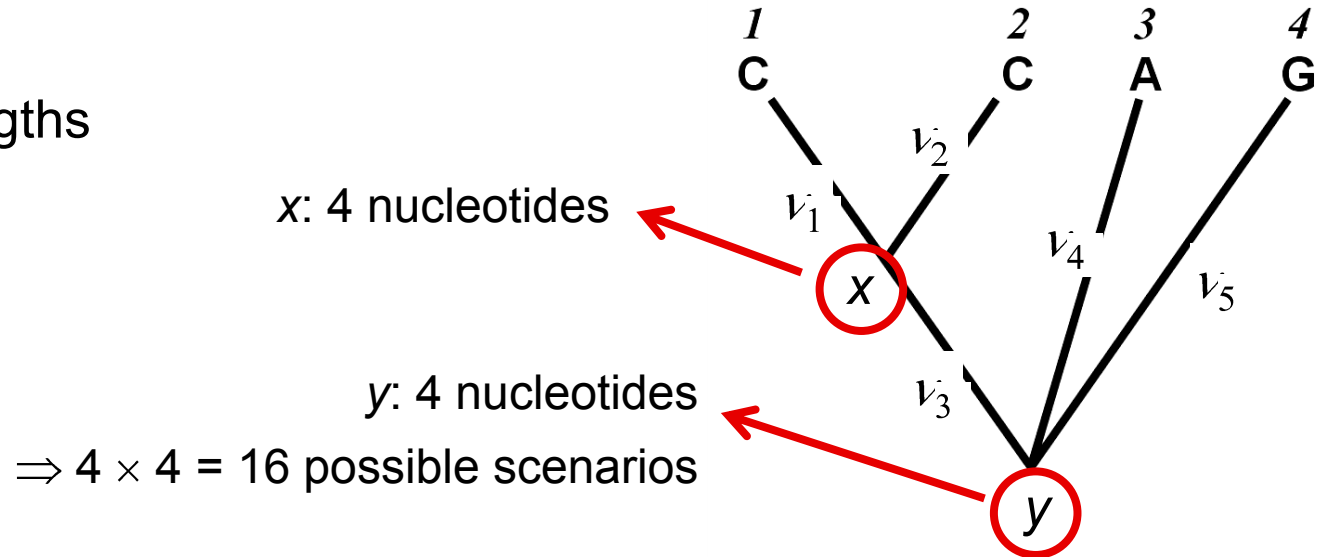
	1		j		N																																													
1	T	C	A	A	A	A	T	G	G	C	T	T	T	A	T	T	C	C	T	T	A	A	T	G	C	C	G	T	T	A	A	C	C	T	T	G	C	G	G	G	G	C	C	A	T	G				
2	T	C	C	G	T	G	A	T	G	G	A	T	T	T	A	T	T	T	C	G	C	A	A	T	G	C	C	T	G	T	C	A	T	C	T	T	A	T	T	C	T	C	A	A	G	T	A	T	C	
3	T	T	C	G	T	G	A	T	G	G	A	T	T	T	A	T	T	G	A	G	G	T	A	T	G	C	C	A	G	T	C	A	T	C	C	T	T	T	T	T	C	T	C	A	T	C	T	A	T	C
4	T	T	C	G	T	G	A	C	G	G	G	T	T	T	A	T	C	T	G	G	C	A	A	T	G	C	C	G	G	T	C	A	T	C	C	T	A	T	T	T	T	C	G	A	G	T	A	T	T	

$v_i =$ branch lengths



	1	<i>j</i>	<i>N</i>
1	TCAAAAATGGCTTTATTTC	C	TTAATGCCGTTAACCCCTTGCGGGGGCCATG
2	TCCGTGATGGATTTATTTTC	C	GCAATGCCTGTCATCTTATTCTCAAGTATC
3	TTCGTGATGGATTTATTG	A	GGTATGCCAGTCATCCTTTTTCTCATCTATC
4	TTCGTGACGGGTTTATCTC	G	GCAATGCCGGTTCATCCTATTTTCGAGTATT

v_i = branch lengths

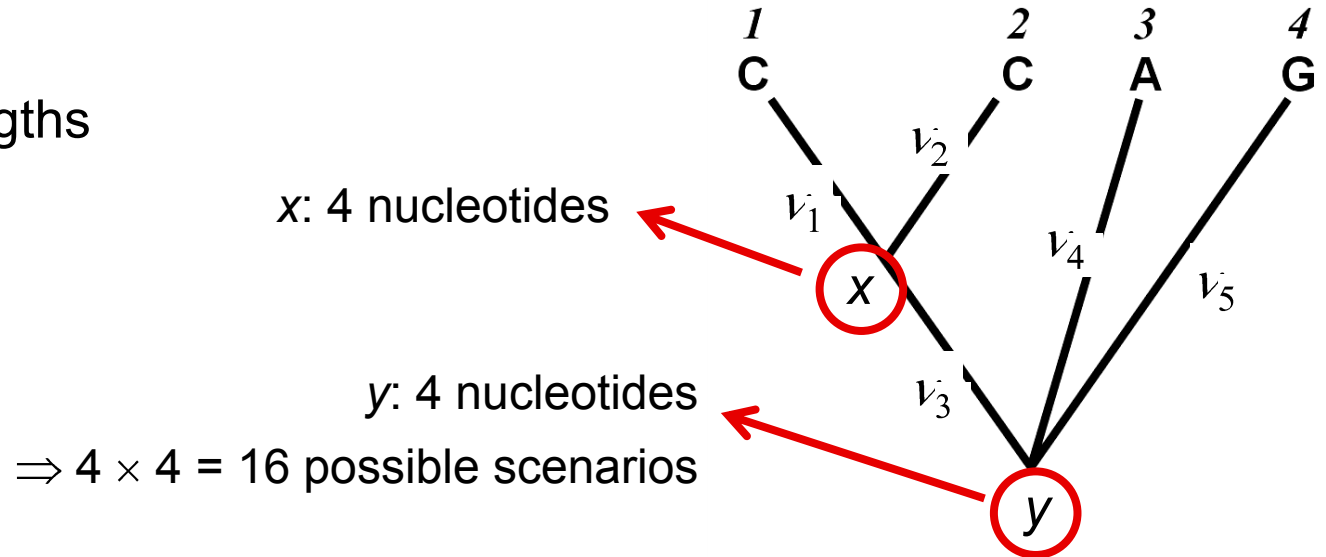


$$L(1) = P(y) \times P(y \rightarrow x) v_3 \times P(x \rightarrow C) v_1 \times P(x \rightarrow C) v_2 \times P(y \rightarrow A) v_4 \times P(y \rightarrow G) v_5$$

$$L(j) = P(\text{scenario 1}) + \dots + P(\text{scenario 16})$$

	1	<i>j</i>		<i>N</i>
1	TCAAAAATGGCTTTATTTC	C	TTAATGCCGTTAACCTTGC	GGGGGCCATG
2	TCCGTGATGGATTTATTT	C	GCAATGCCTGTCATCTT	ATTCTCAAGTATC
3	TTCGTGATGGATTTATTG	A	GGTATGCCAGTCATCCT	TTTTCTCATCTATC
4	TTCGTGACGGGTTTATCT	G	GCAATGCCGGTTCATCCT	ATTTTCGAGTATT

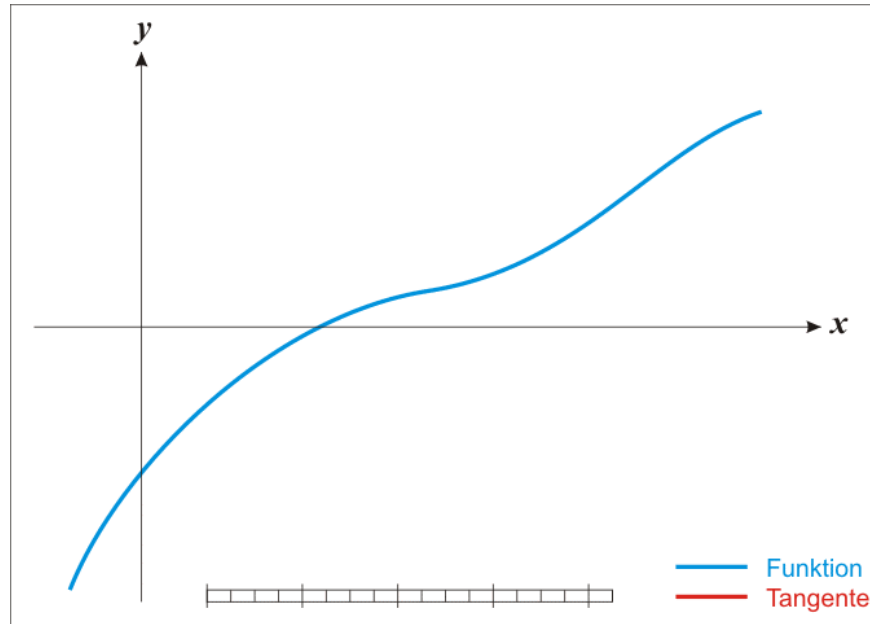
v_i = branch lengths



all sites: $L = L(1) \times L(2) \times \dots \times L(j) \times \dots \times L(N) = \prod_{j=1}^N L_j$

$$\ln L = \ln L(1) + \ln L(2) + \dots + \ln L(N) = \sum_{j=1}^N \ln L_j$$

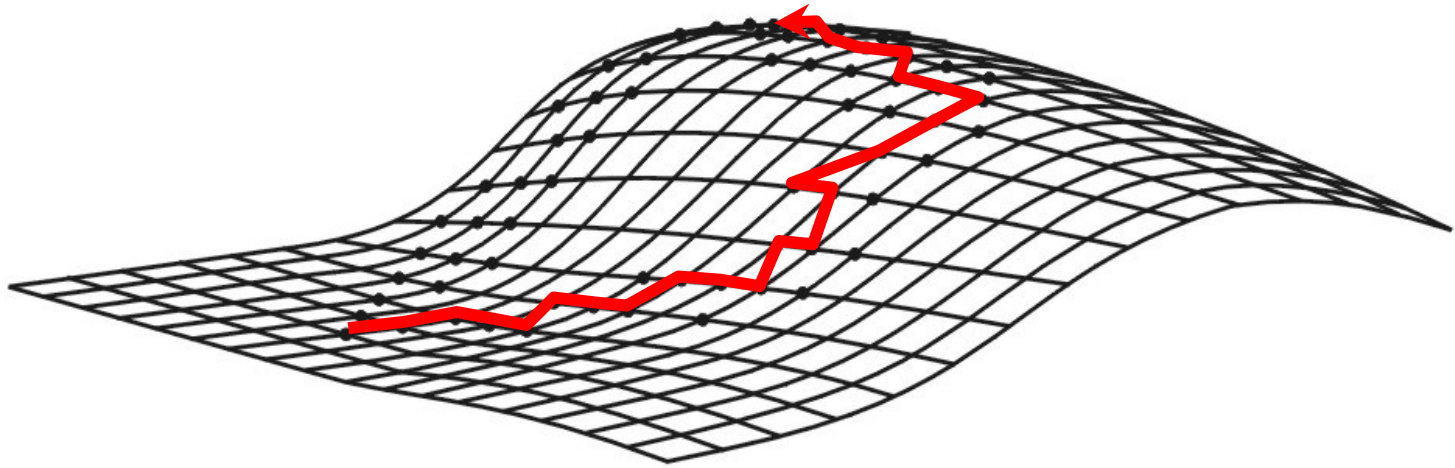
Search for maximum likelihood of the tree
→ eg. Newton (Newton-Raphson) method



https://upload.wikimedia.org/wikipedia/commons/e/e0/NewtonIteration_Ani.gif

Maximum likelihood tree search: heuristic search

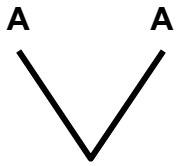
Heuristic search



stepwise addition ... eg. PHYLIP

star decomposition ... eg. MOLPHY; neighbor-joining tree

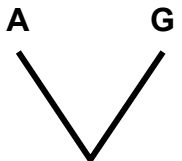
branch swapping



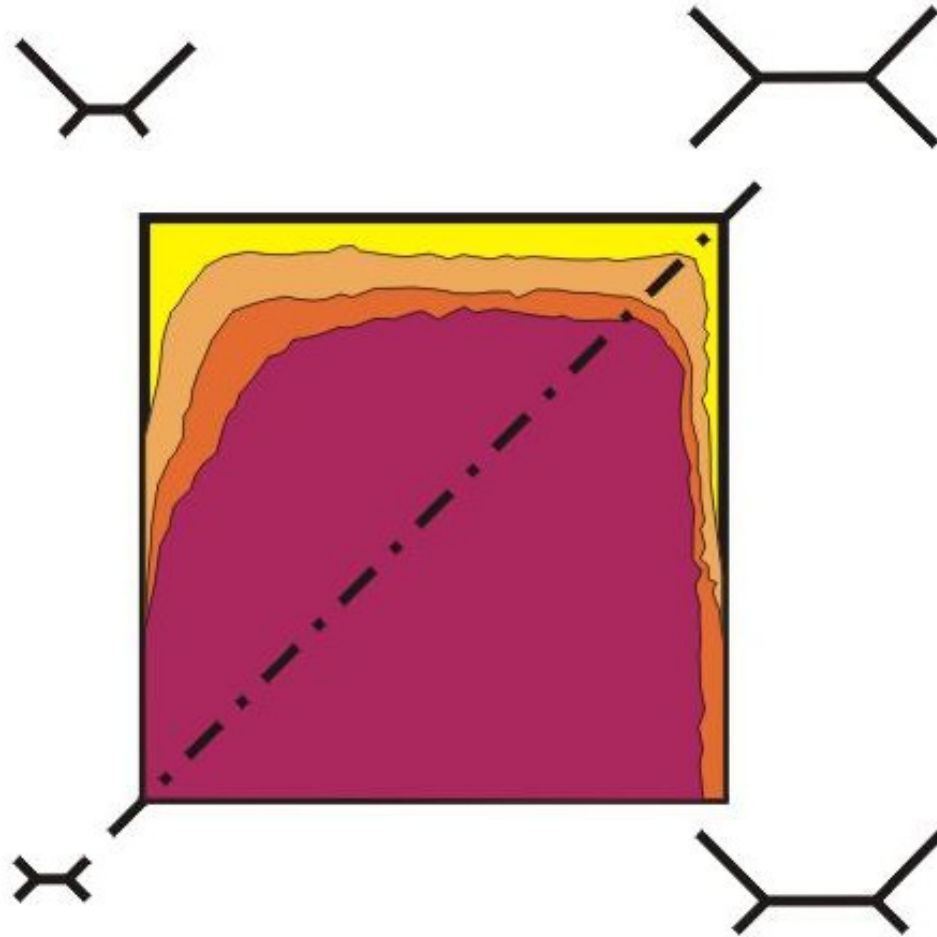
Likelihood (ML) and parsimony (MP)

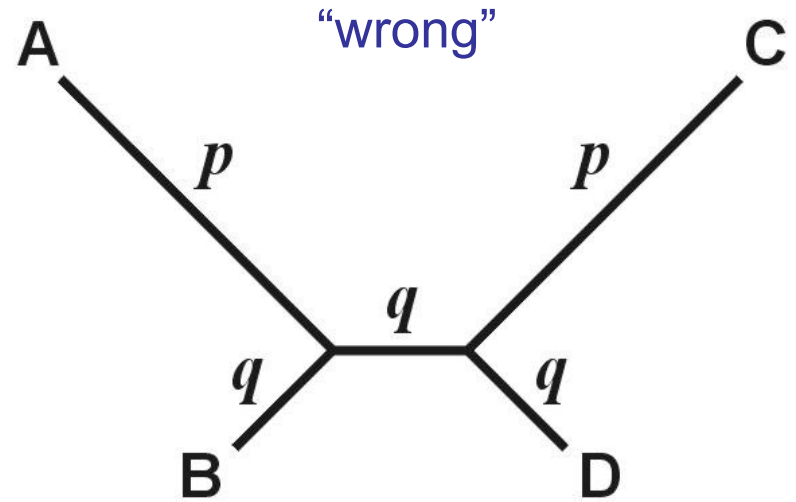
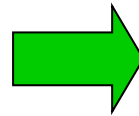
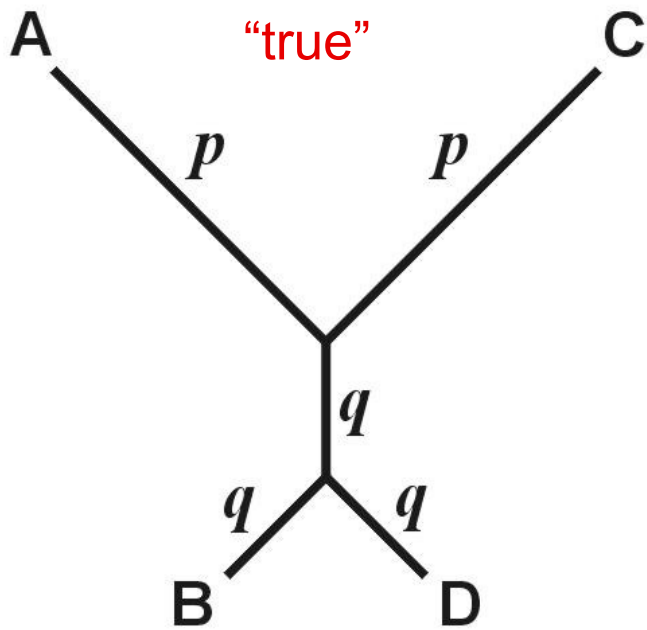
No. changes	Parsimony	$\nu = 0,01$	$\nu = 0,10$	$\nu = 0,20$	$\nu = 1,00$
		(0,2475)	(0,2266)	(0,20611)	(0,11192)
0	100	99,99	99,83	99,31	82,17
1	0	0,00	0,00	0,00	0,00
2	0	0,0011	0,11	0,44	9,13
3	0			0,034	3,55
4	0				0,0027

No. changes	Parsimony	$\nu = 0,01$	$\nu = 0,10$	$\nu = 0,20$	$\nu = 1,00$
		(0,00083)	(0,00786)	(0,01462)	(0,04602)
0	0	0,00	0,00	0,00	0,00
1	100	99,66	96,64	92,36	66,54
2	0	0,33	3,22	6,22	21,19
3	0		0,12	0,48	8,61
4	0		0,003	0,023	2,05
5	0			0,0037	0,42



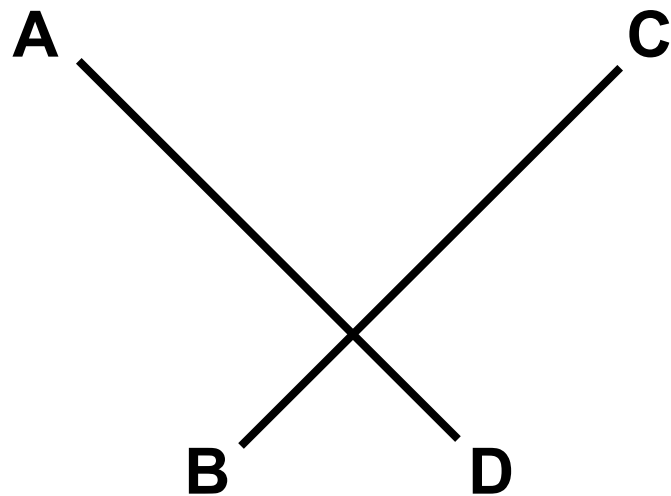
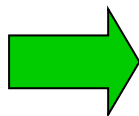
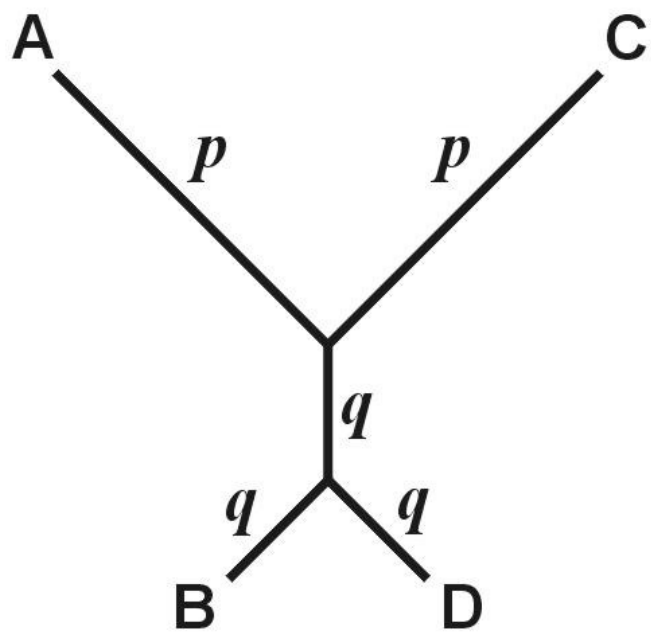
Likelihood and consistency





Farris
(anti-Felsenstein,
inverse Felsenstein)
zone

“long-branch repulsion”



BAYESIAN ANALYSIS (Bayesovská analýza)



ML: Probability of data given hypothesis

Bayesian approach:

Conditional probability of hypothesis given data

$$P(H | D)$$







Example.: set of 100 dice, from which we choose one
we know that of 100 dice, 80 are 'fair' and 20 biased for 6

2 throws: 1. throw =  2. throw = 

What is the probability our dice is biased?

probability of individual results:

all the same in unbiased dice, varied in biased dice:

	unbiased	biased
	$1/6$	$1/21$
	$1/6$	$3/21$
	$1/6$	$3/21$
	$1/6$	$4/21$
	$1/6$	$4/21$
	$1/6$	$6/21$

$P(H | D)$ is called **posterior probability** (aposteriorní pravděpodobnost)

posterior probability is a function of likelihood $L = P(D | H)$

and **prior probability** (apriorní pravděpodobnost) reflecting our a priori expectation or knowledge

Posterior probability that the coin is biased is given by the Bayes equation:

$$P(H | D) = \frac{P(D | H) \times P(H)}{\sum [P(D | H_i) \times P(H_i)]}$$

likelihood

prior probability



sum of numerators across all alternative hypotheses











Thomas Bayes

For our example of 2 dice throws:

prior probability (biased) = 0,2
(20/100 biased dice in the set)

Pr. of getting   with unbiased dice:
 $P = 1/6 \times 1/6 = 1/36$

Pr. of getting   with biased dice:
 $P = 3/21 \times 6/21 = 18/441$

	unbiased	biased
	1/6	1/21
	1/6	3/21
	1/6	3/21
	1/6	4/21
	1/6	4/21
	1/6	6/21

$$\begin{aligned}
 P(\text{biased} | \text{2 dots, 6 dots}) &= \frac{P(\text{2 dots, 6 dots} | \text{biased}) \times P(\text{biased})}{P(\text{2 dots, 6 dots} | \text{biased}) \times P(\text{biased}) + P(\text{2 dots, 6 dots} | \text{fair}) \times P(\text{fair})} \\
 &= \frac{18/441 \times 2/10}{18/441 \times 2/10 + 1/36 \times 8/10} = \underline{0,269}
 \end{aligned}$$

Bayesian method in phylogenetic analysis:

posterior
probability

likelihood

prior probability

$$P(\boldsymbol{\tau}, \boldsymbol{\nu}, \boldsymbol{\theta} | \mathbf{X}) = \frac{P(\mathbf{X} | \boldsymbol{\tau}, \boldsymbol{\nu}, \boldsymbol{\theta}) P(\boldsymbol{\tau}, \boldsymbol{\nu}, \boldsymbol{\theta})}{\sum_{i=1}^{B(s)} [P(\mathbf{X} | \boldsymbol{\tau}, \boldsymbol{\nu}, \boldsymbol{\theta}) P(\boldsymbol{\tau}, \boldsymbol{\nu}, \boldsymbol{\theta})]}$$

sum across all hypotheses
(= marginal likelihood)

Parameters of Bayesian analysis mostly continuous \Rightarrow
 $P \rightarrow$ probability density functions

$$P(\mathbf{X}|\boldsymbol{\tau}, \boldsymbol{\nu}, \boldsymbol{\theta}) = \int P(\mathbf{X}|\boldsymbol{\tau}, \boldsymbol{\nu}, \boldsymbol{\theta}) dF(\boldsymbol{\nu}, \boldsymbol{\theta})$$

likelihood

prior distribution

posterior distribution

$$f(\boldsymbol{\theta}|D) = \frac{\Pr(D|\boldsymbol{\theta})f(\boldsymbol{\theta})}{\Pr(D)}$$

marginal likelihood

$\boldsymbol{\theta}$ = set of (continuous) parameters in the model, including the tree, substitution model parameters, clock rates, etc.

either ML estimates \rightarrow **empirical BA**

or all combinations \rightarrow **hierarchical BA**

outcome = probability distribution

Tracer

Trace Files:

Trace File	States	Burn-In
batRABV_short.log	114820000	11482000

Reload

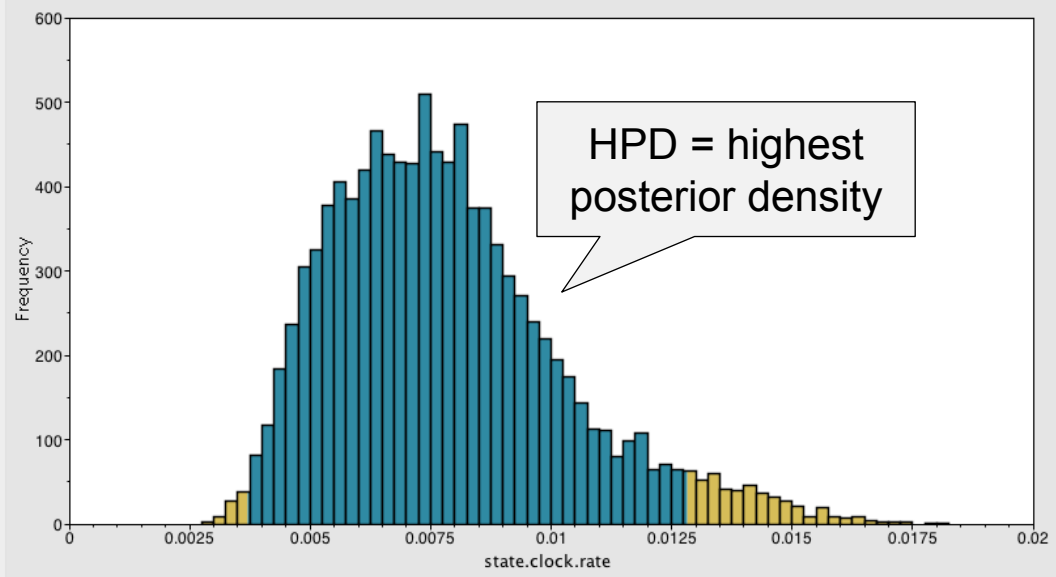
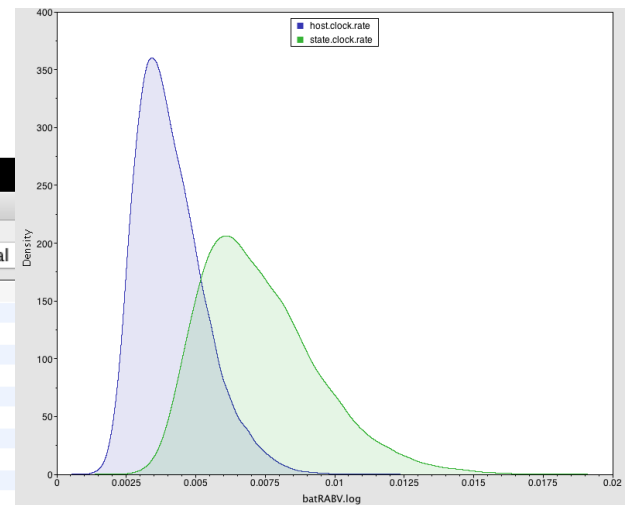
Traces:

Statistic	Mean	ESS	...
skygrid.logPopSize42	5.667	42	R
skygrid.logPopSize43	5.547	44	R
skygrid.logPopSize44	5.4	47	R
skygrid.logPopSize45	5.266	50	R
skygrid.logPopSize46	5.108	54	R
skygrid.logPopSize47	4.917	58	R
skygrid.logPopSize48	4.698	63	R
skygrid.logPopSize49	4.464	70	R
skygrid.logPopSize50	4.198	82	R
skygrid.cutOff	250	1	R
kappa	11.1	7622	R
frequencies1	0.292	4174	R
frequencies2	0.217	4908	R
frequencies3	0.212	4392	R
frequencies4	0.279	4428	R
alpha	0.214	8285	R
batRABV.clock.rate	2.48E-4	24	R
host.clock.rate	4.339E-3	28	R
state.clock.rate	7.824E-3	27	R
host.nonZeroRates	21.985	1926	I
state.nonZeroRates	19.285	114	I
root.host	n/a	-	C
root.state	n/a	-	C
host.rates.Ap.Ef	0.945	310	R
host.rates.Ap.Lb	0.954	467	R
host.rates.Ap.Lbl	0.974	457	R
host.rates.Ap.Lc	0.759	824	R
host.rates.Ap.Li	1.003	464	R
host.rates.Ap.Ln	0.975	552	R
host.rates.Ap.Ls	1.081	510	R
host.rates.Ap.Lx	1.041	492	R
host.rates.Ap.Ma	0.929	494	R
host.rates.Ap.Mc	1	513	R
host.rates.Ap.Ml	1.012	520	R
host.rates.Ap.My	1.083	265	R

Estimates Marginal Density Joint-Marginal

Summary Statistic

	state.clock.rate
mean	7.824E-3
stderr of mean	4.6621E-4
stddev	2.414E-3
variance	5.8276E-6
median	7.4921E-3
value range	[2.864E-3, 0.0181]
geometric mean	7.4749E-3
95% HPD interval	[3.7032E-3, 0.0128]
auto-correlation time (ACT)	3.855E6
effective sample size (ESS)	26.8
number of samples	10335



Setup...

Problem: calculations too complex \Rightarrow impossible to solve analytically,
only numerically

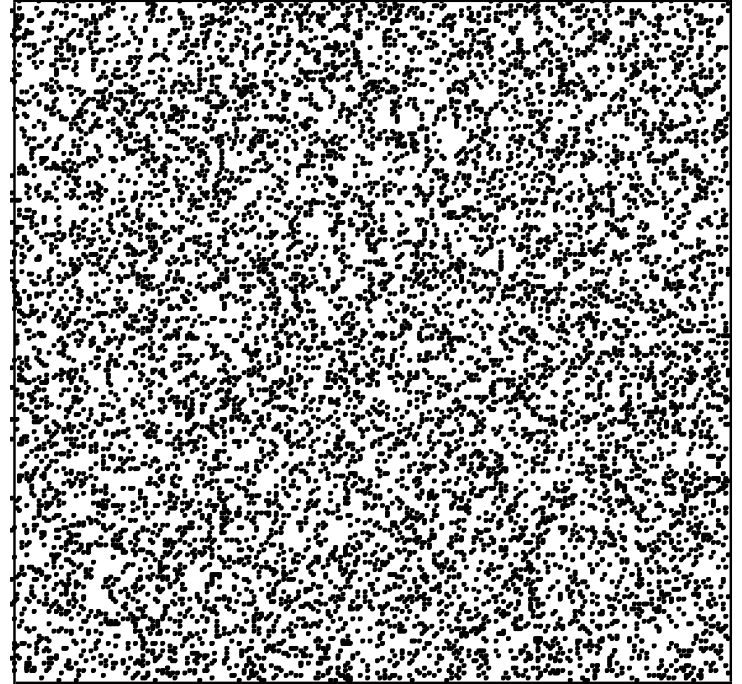
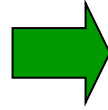
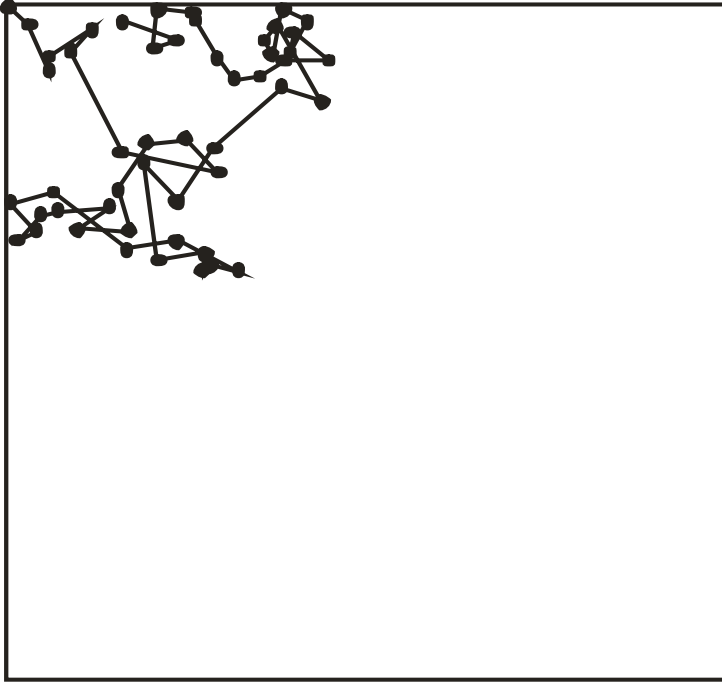
solution: **Monte Carlo methods**

random sampling, approximation of reality when sample size high

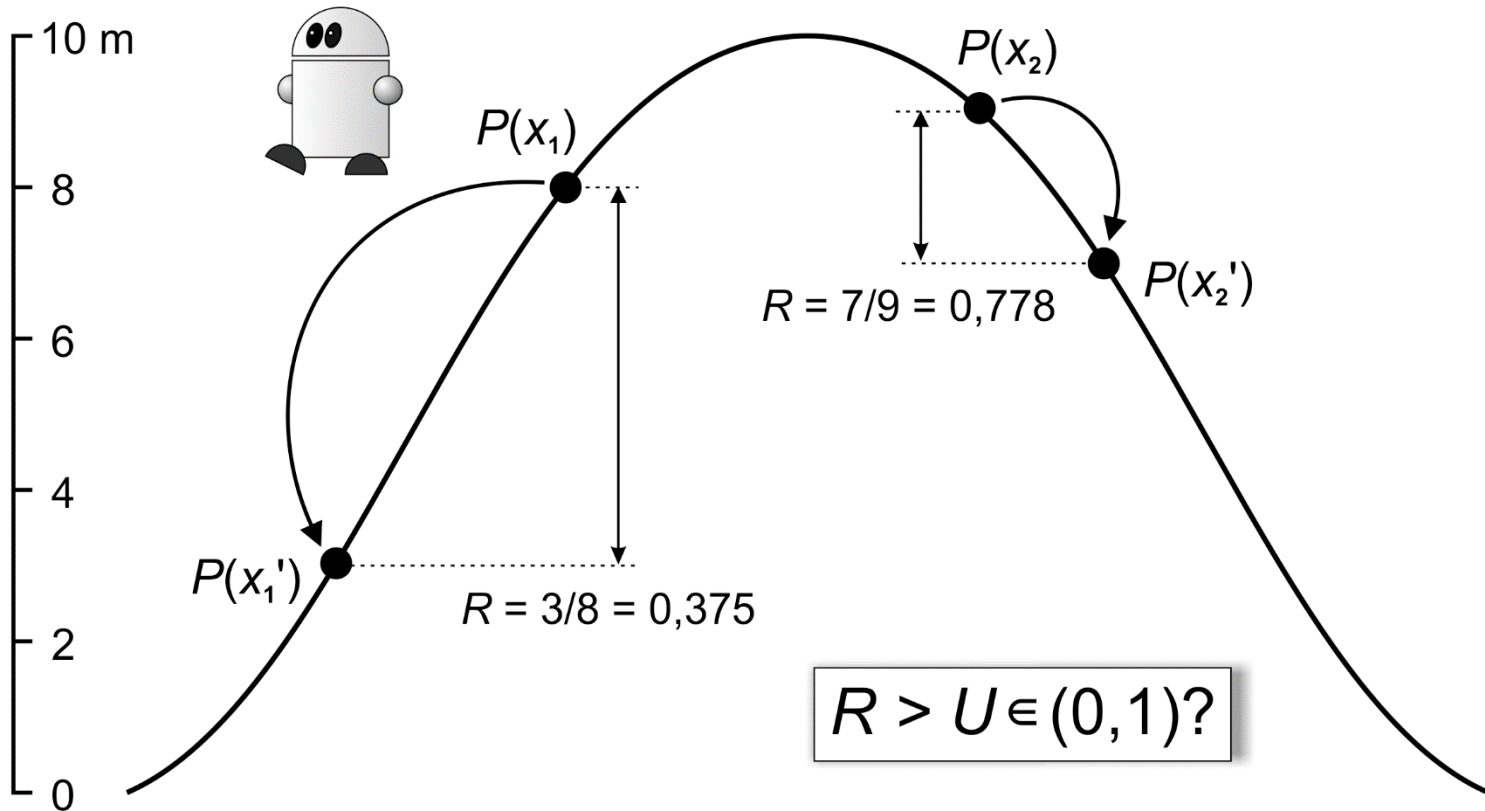
Markov chain Monte Carlo (MCMC)

Markov process: $t_{-1}: A \rightarrow t_0: C \rightarrow t_{+1}: G$

... P same across the whole phylogeny = **homogenous Markov process**



Metropolis-Hastings algorithm:

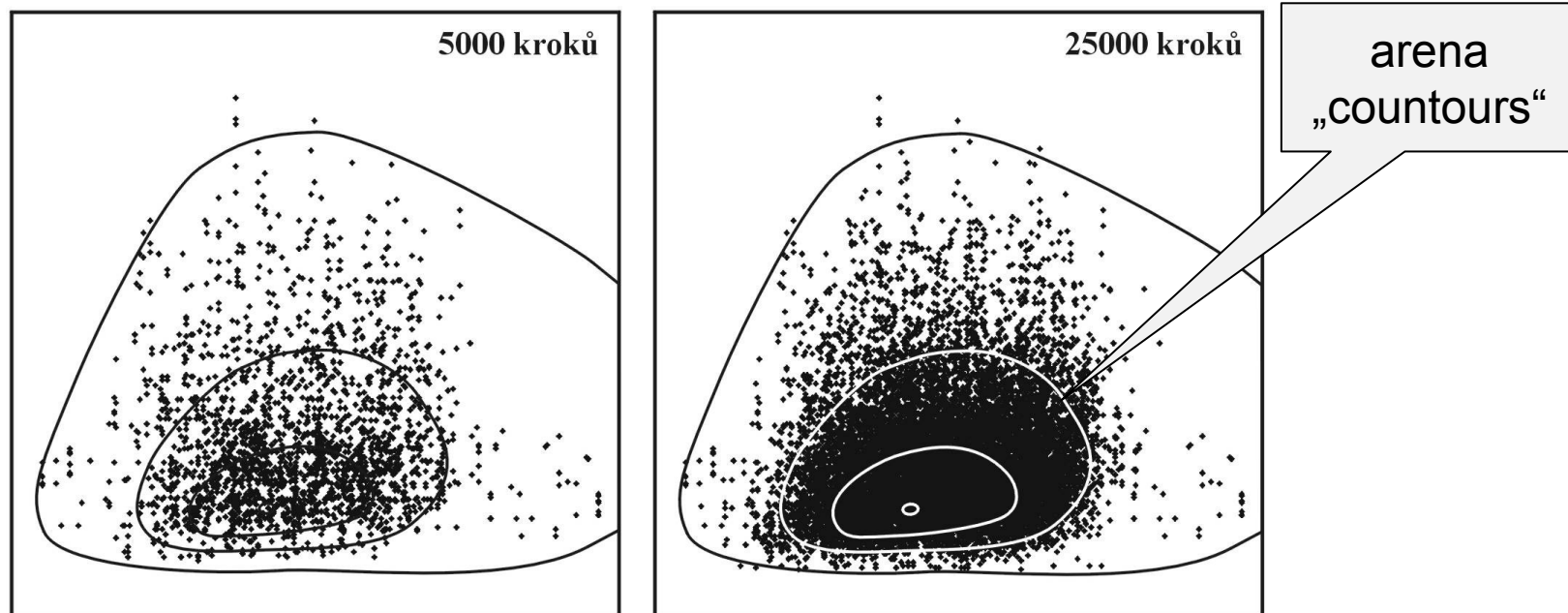


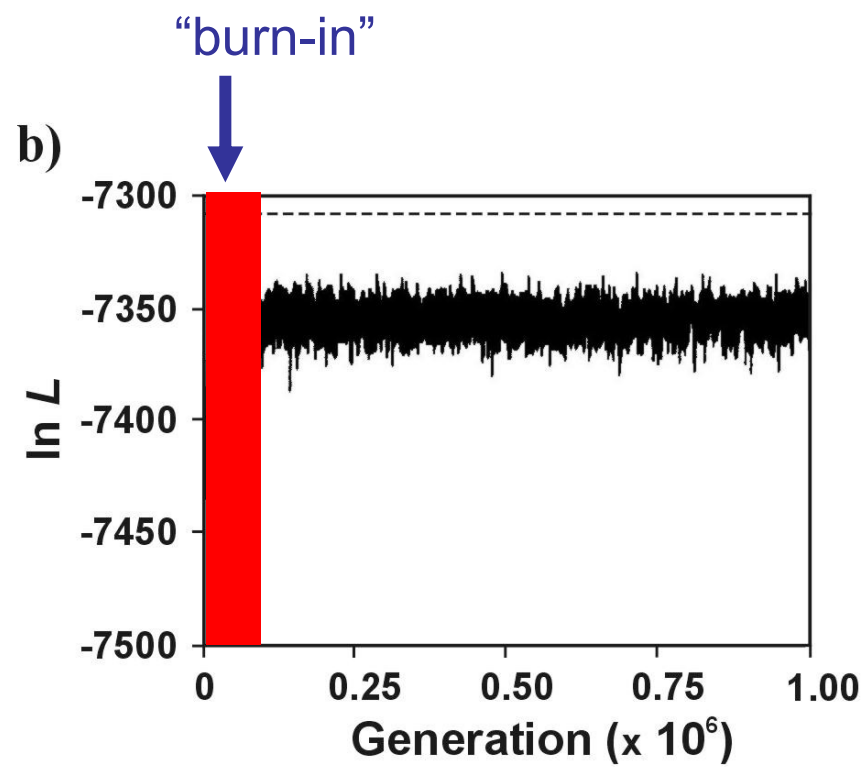
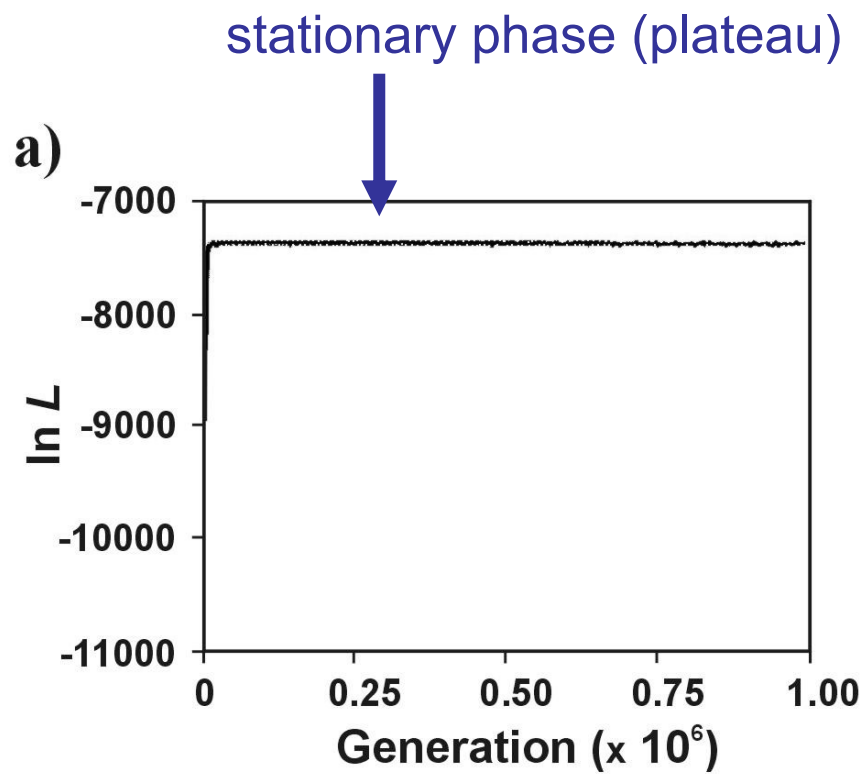
Metropolis-Hastings algorithm:

Change of parameter $x \rightarrow x'$

1. if $P(x') > P(x)$, accept x'
2. if $P(x') \leq P(x)$, calculate $R = P(x')/P(x)$
since $P(x') \leq P(x)$, R must be ≤ 1
3. generate random number U from uniform distribution from interval $(0, 1)$
4. if $R \geq U$, accept x' , if not, retain x

directed movement of robot across arena:





Reversible jump MCMC:

allows changing number of parameters in each MC step

we can use eg. for modelling variation of evolution between sites in sequences, for choosing models or for making non-homogenous substitution models (eg. different base composition along branches)

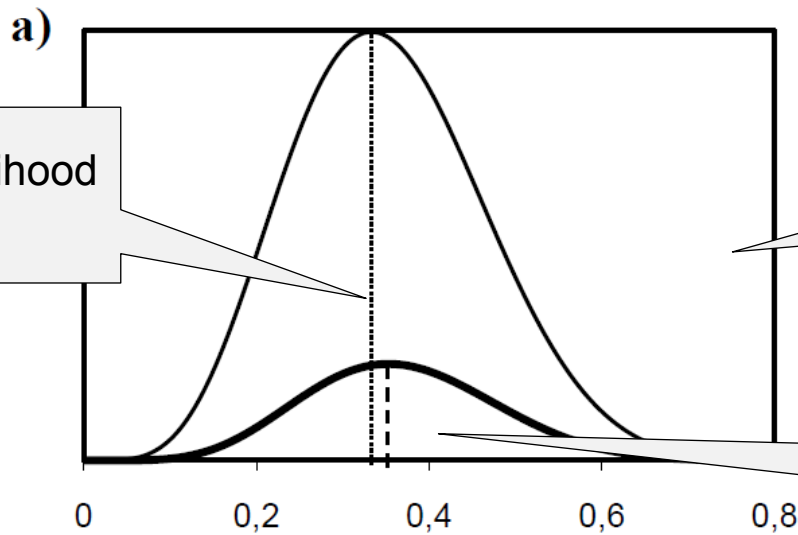
Metropolis coupled MCMC (MCMCMC, MC³):

1 „cold“ chain, 3 „heated“ chains

same starting point, due to stochasticity rapid divergence of „robots“

MrBayes: <http://morphbank.ebc.uu.se/mrbayes/>

Problem with priors

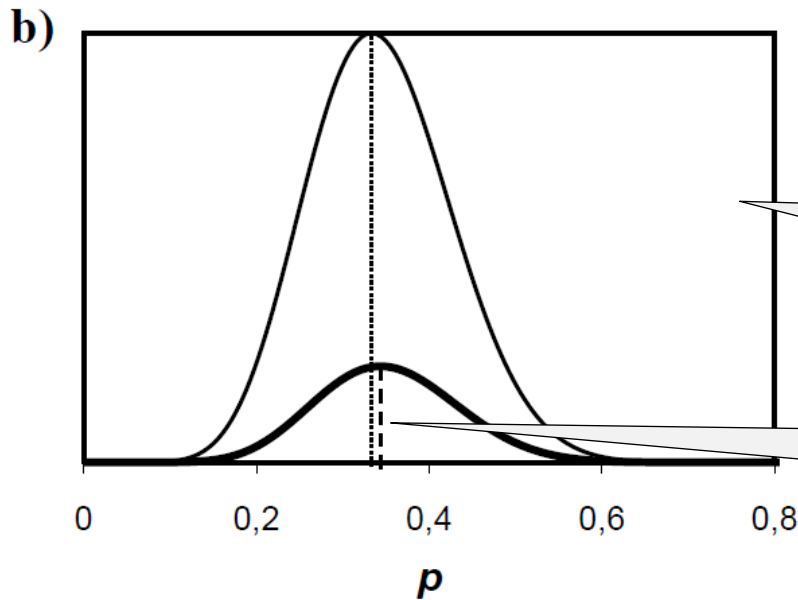


maximum likelihood
= 0,333

prior = 0,5

15 coin tosses
score 5 H : 10 O

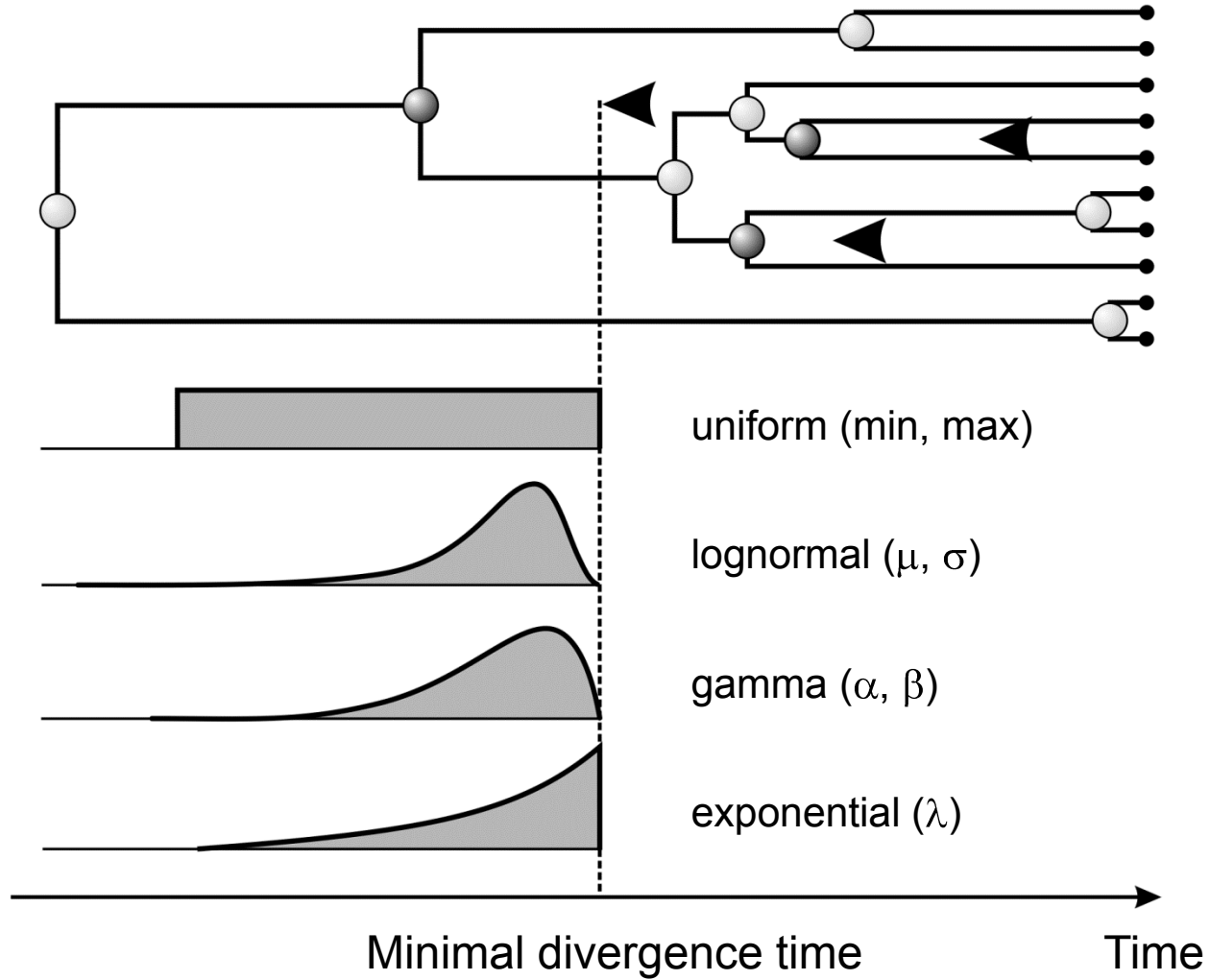
due to prior,
posterior pr.
shifted to the right



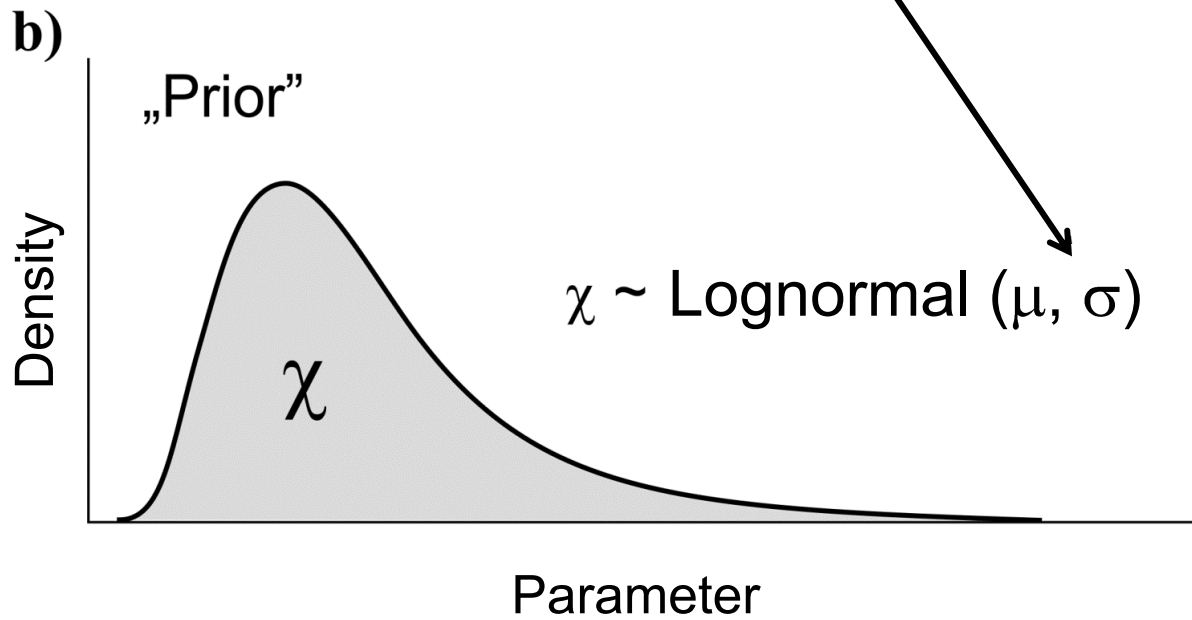
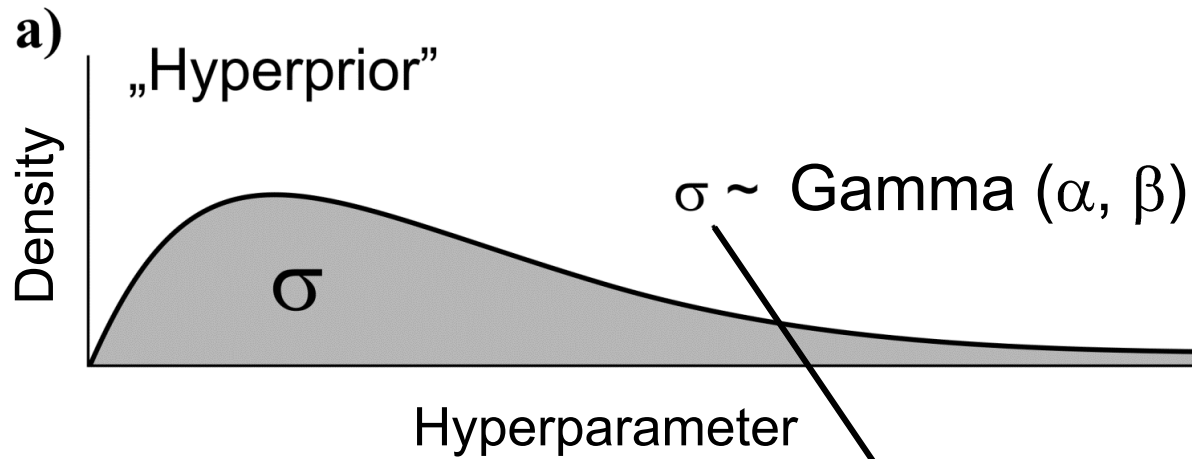
30 coin tosses
score 10 H : 20 O

difference from
ML smaller

Which priors to choose?



Setting priors:



Time-trees and phylodynamics

Phylodynamics = synthesis of mathematical epidemiology and statistical phylogenetics

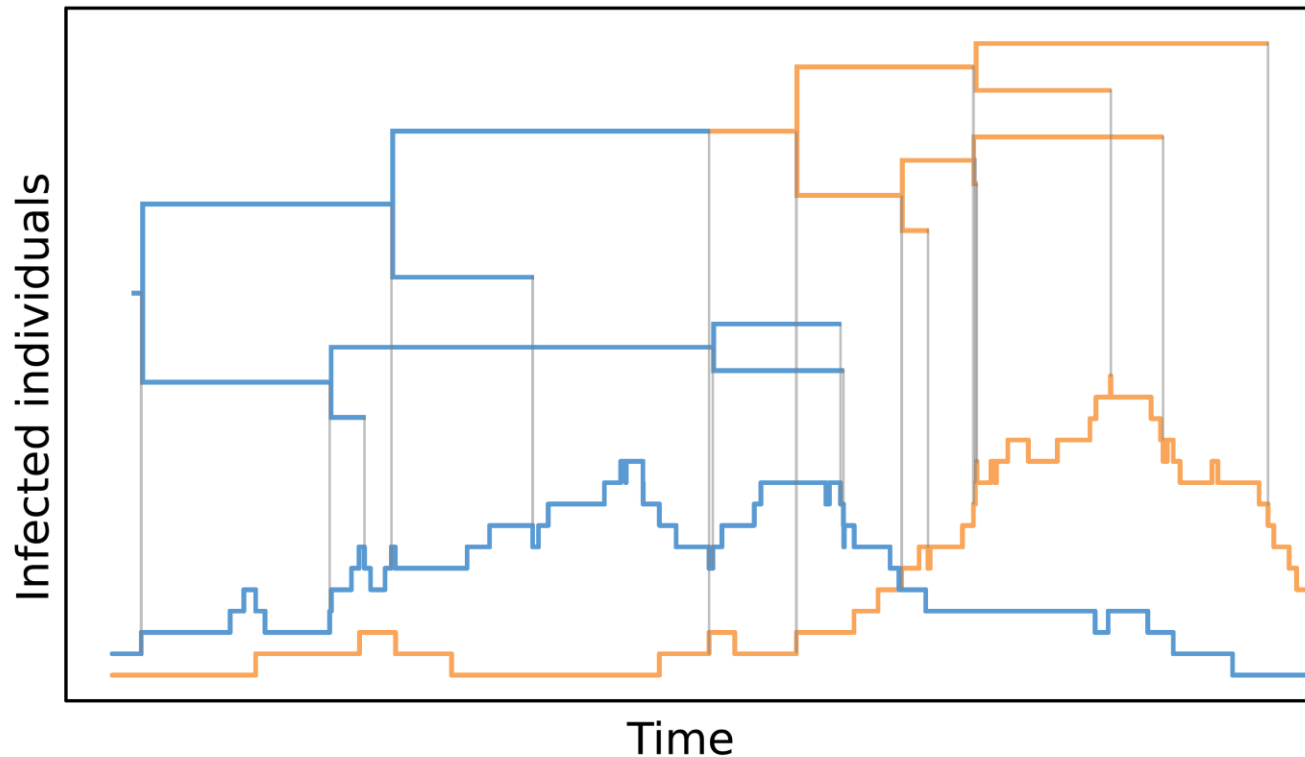
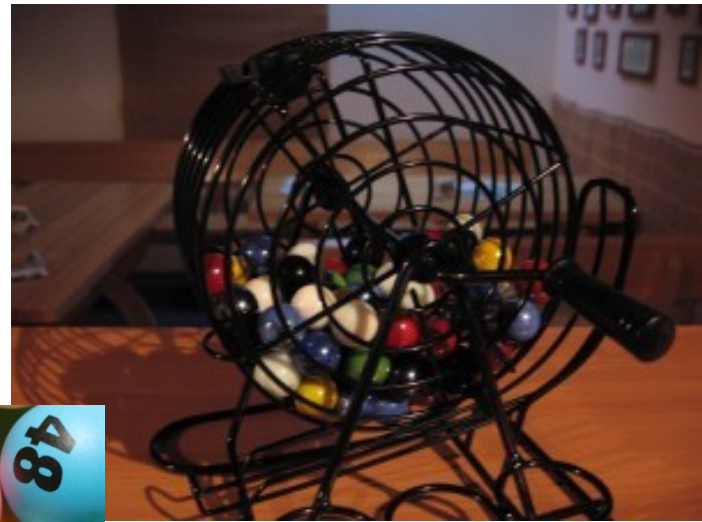


Figure 5.4 A two-deme phylodynamic time-tree with associated stochastic dynamics of infected compartments. (With thanks to Tim Vaughan for producing this figure).

Measuring tree reliability

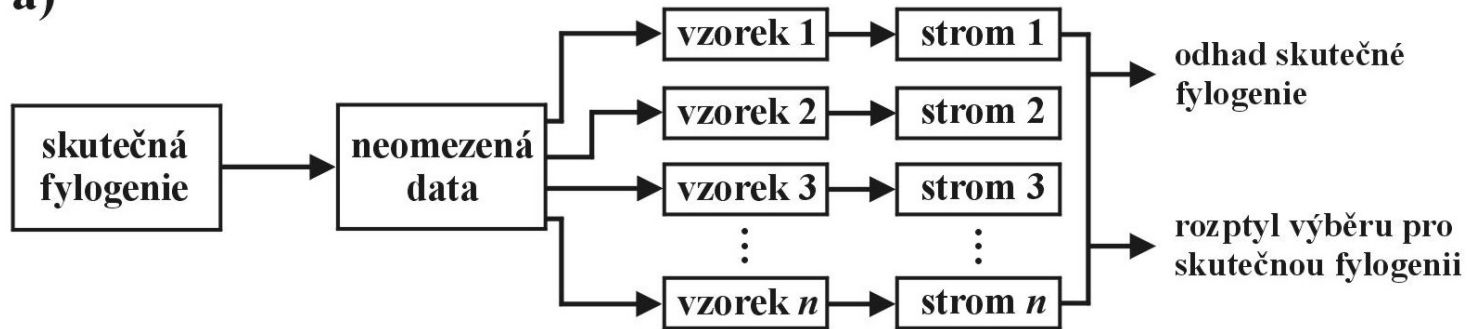
Resampling methods

without replacement = **jackknife**
with replacement = **bootstrap**



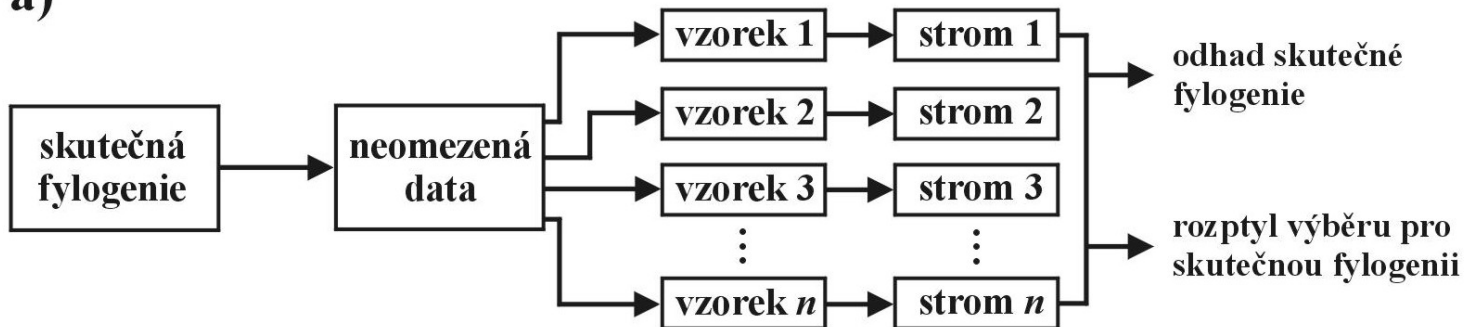
bootstrap:

a)

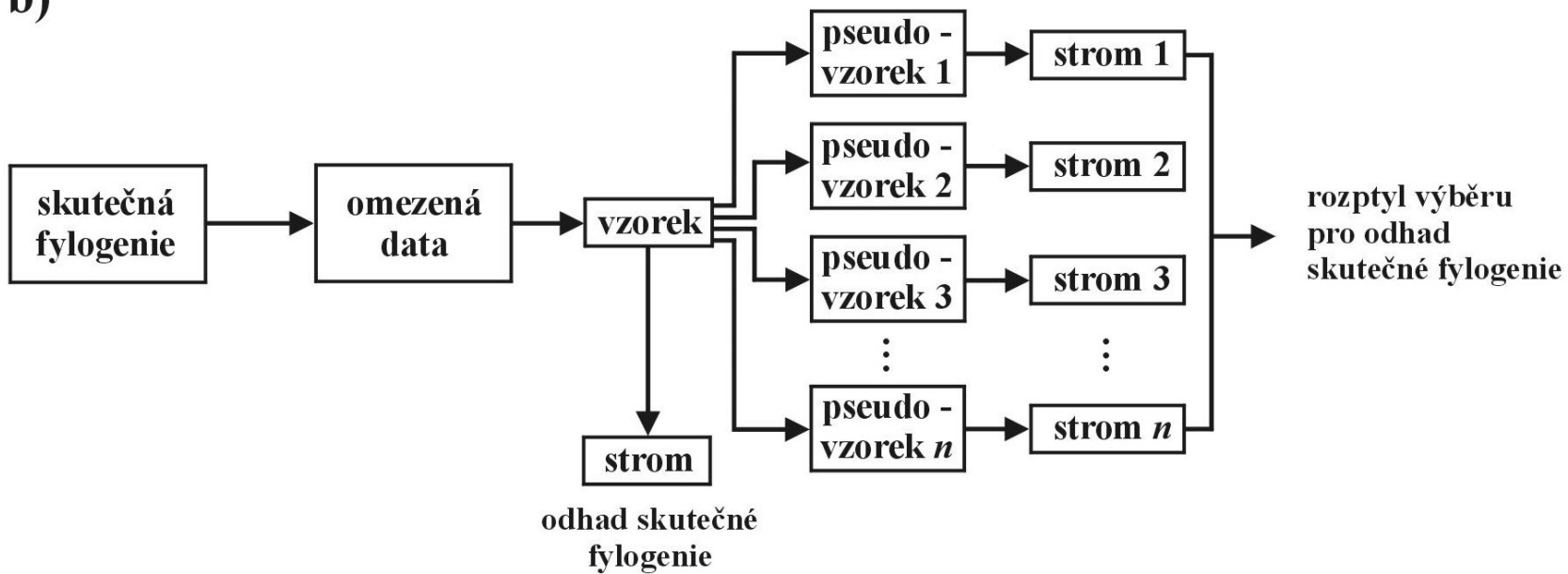


bootstrap:

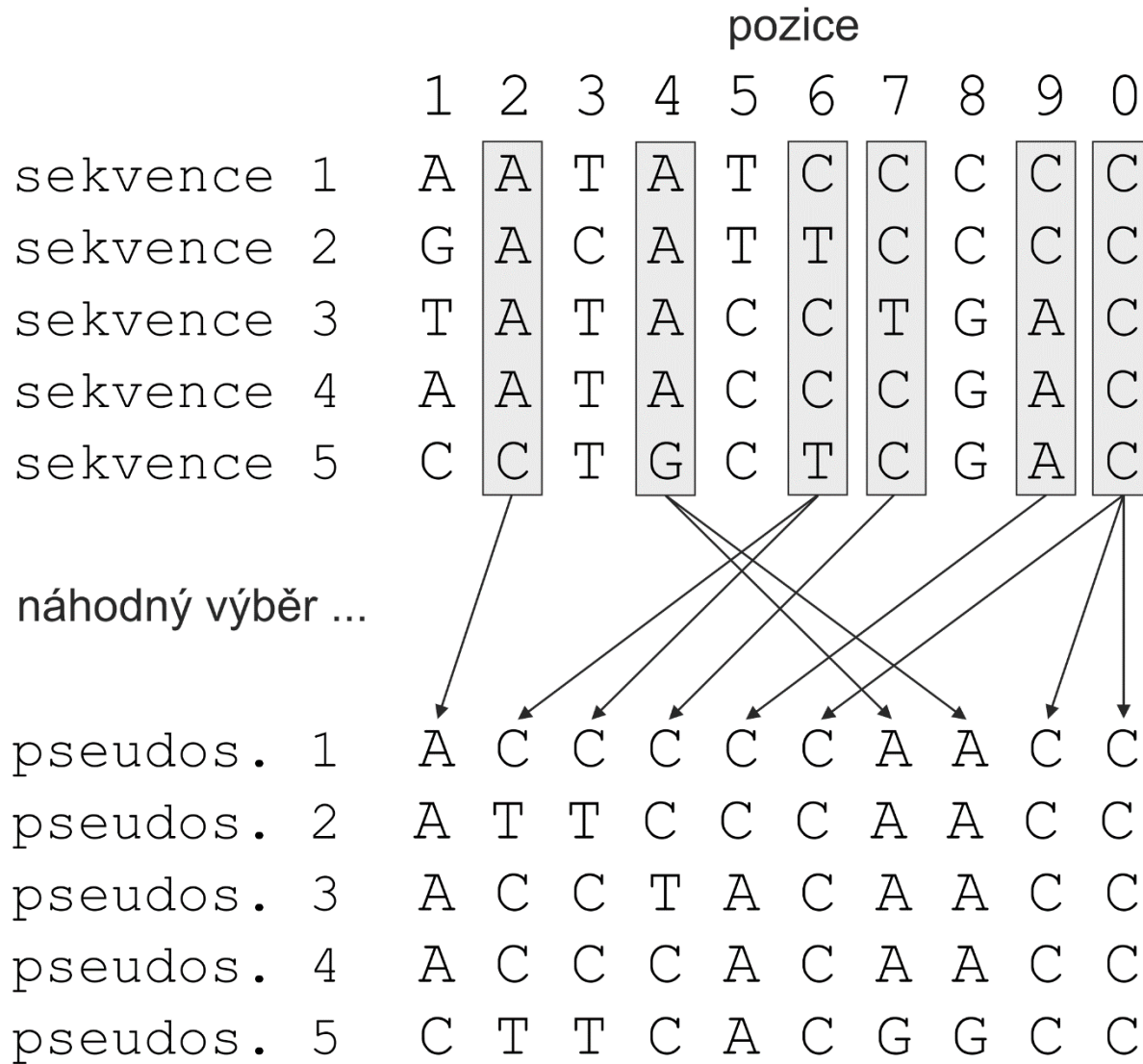
a)



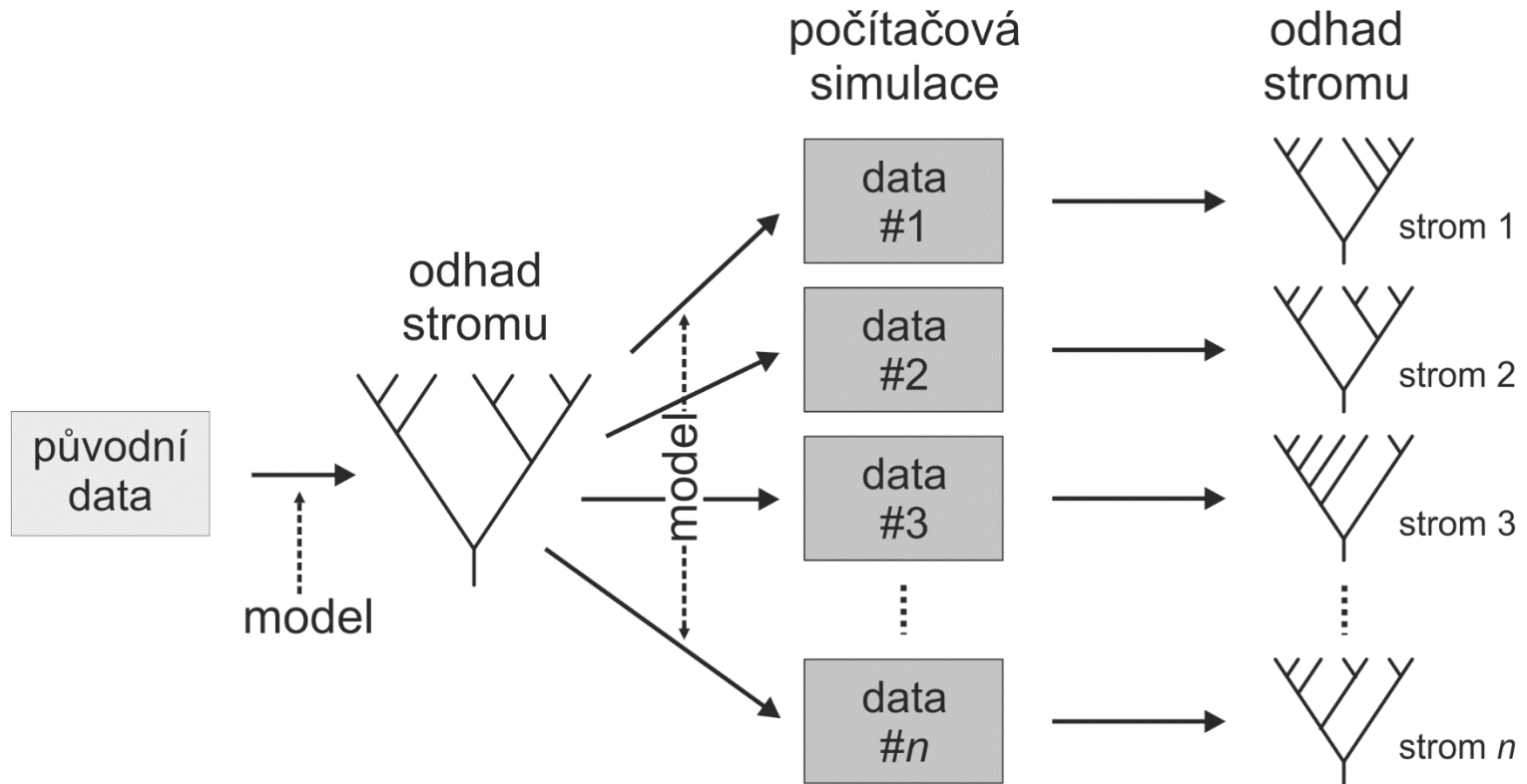
b)



bootstrap:



parametric bootstrap: evolutionary model



Bayesian analysis: posterior probabilities

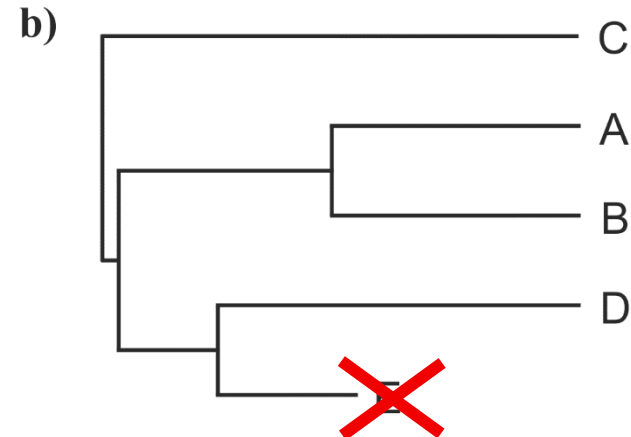
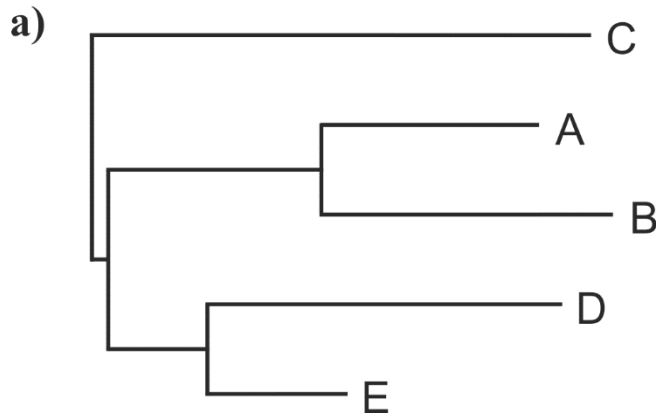
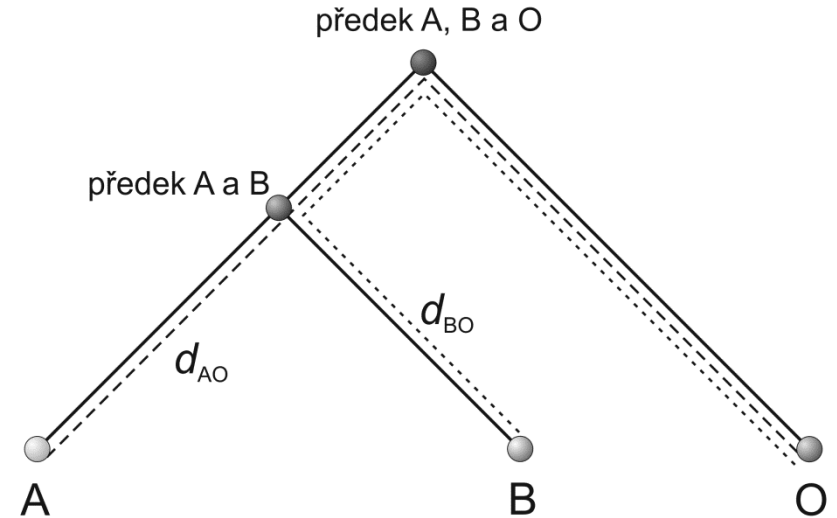
Hypothesis testing

Test of molecular clock:

Relative rate test (RRT): $AC=BC$?

Linearized trees

removing significantly different taxa

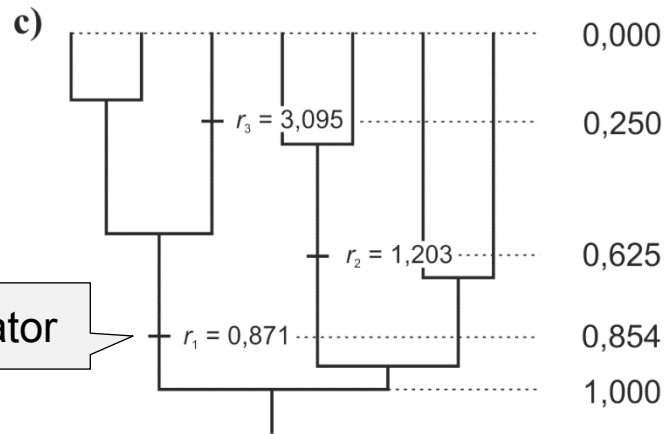
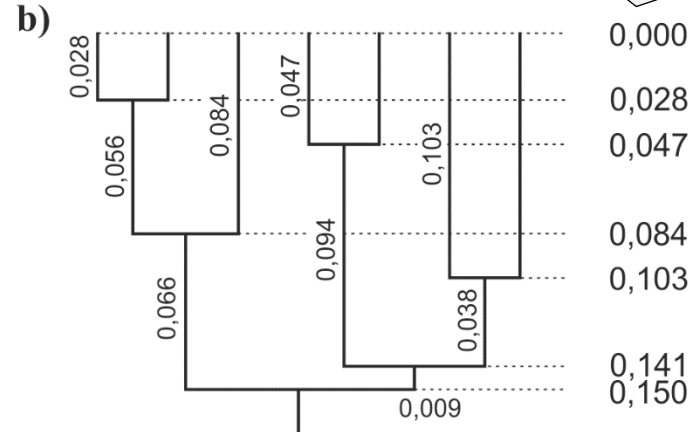
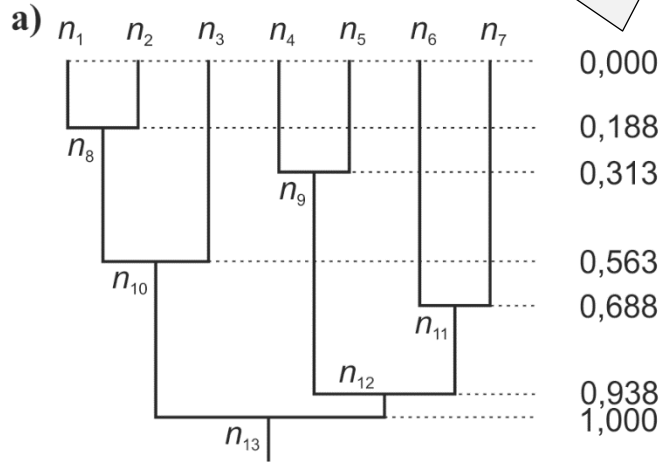


Relaxed molecular clock

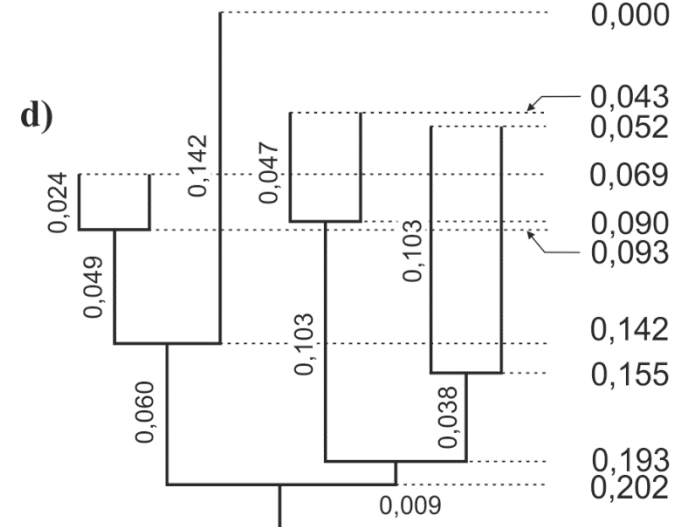
enable changing rates along branches

unscaled time

scaled time
(expected no. substitutions/site)



multiplier



Tree comparison

Are two trees significantly different?

Tests of paired positions:

winning sites test

Felsenstein's z test

Templeton's test

Kishino-Hasegawa test (KHT, REL)

a)

$$d_i^* = \ln L_{T1}^* - \ln L_{T2}^*,$$

kde i je bootstrapový replikát

$$d_1^* = \ln L_{T1}^* - \ln L_{T2}^*$$

$$d_2^* = \ln L_{T1}^* - \ln L_{T2}^*$$

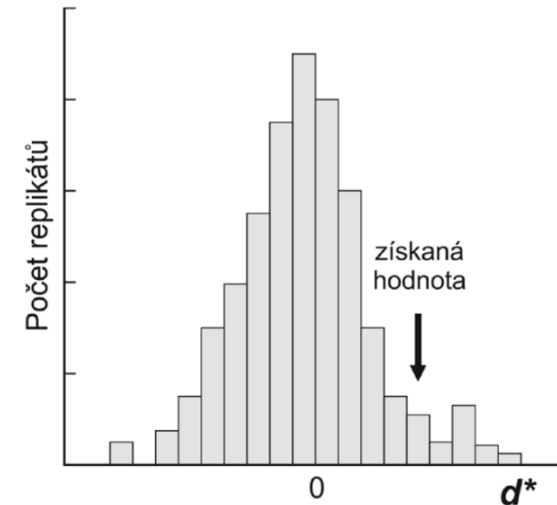
$$d_3^* = \ln L_{T1}^* - \ln L_{T2}^*$$

...

$$d_n^* = \ln L_{T1}^* - \ln L_{T2}^*$$

For more than two trees:

Shimodaira-Hasegawa (SH) test



Tree comparison

To what degree are two trees different?

Tree distances:

partition metric

quartet metric

path difference metric

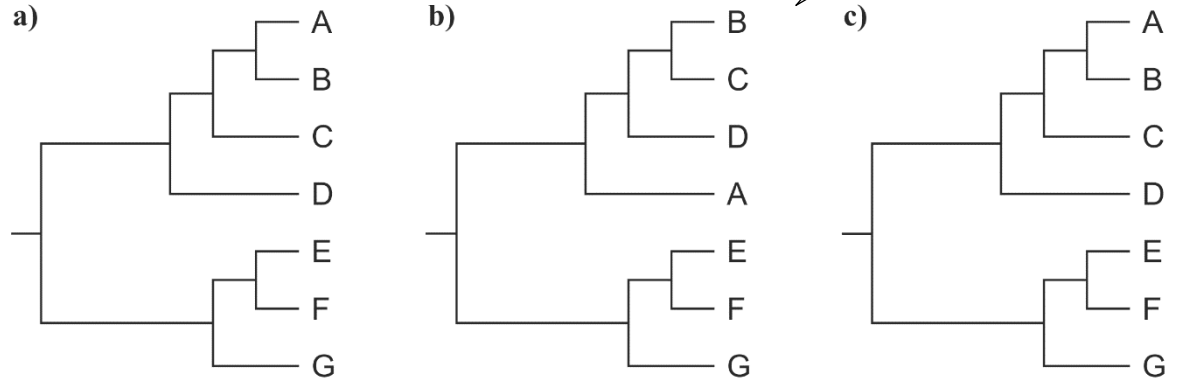
methods incorporating branch lengths

Problems with tree distances

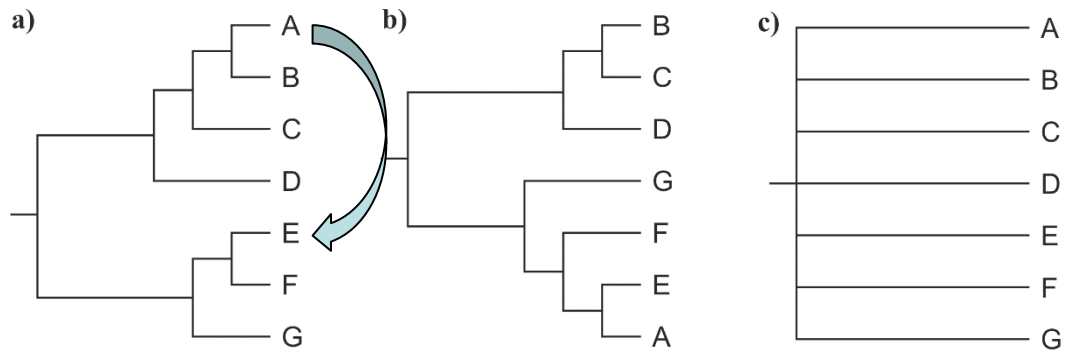
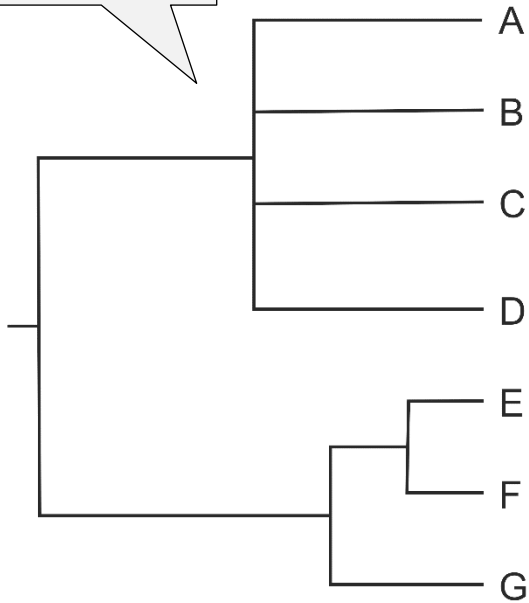
Consensus trees

strict consensus

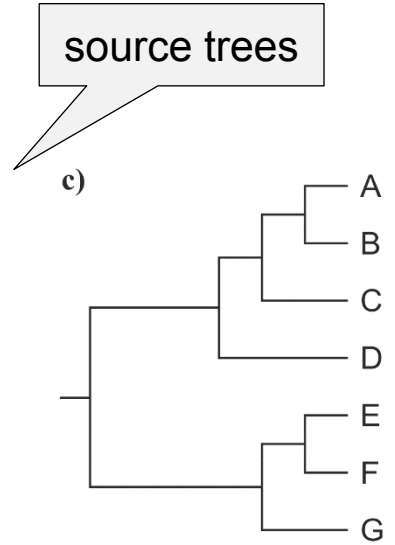
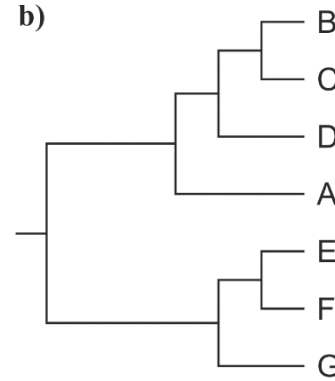
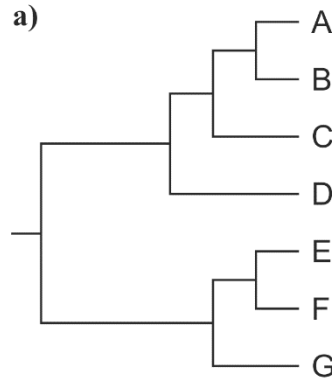
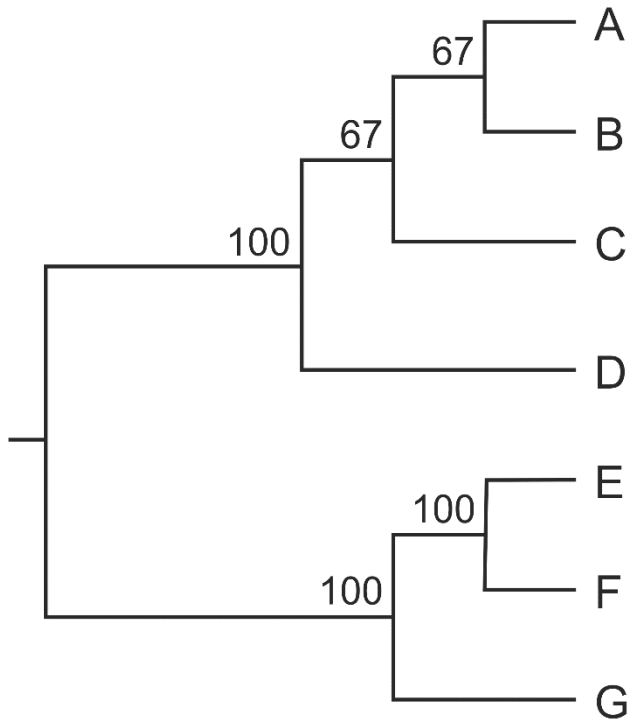
source trees



strict consensus tree



majority-rule



source trees

majority-rule
consensus

Consensus trees

problem with consensus trees – combined vs.
separate analysis, supermatrix vs. supertree

consensus trees in resampling methods

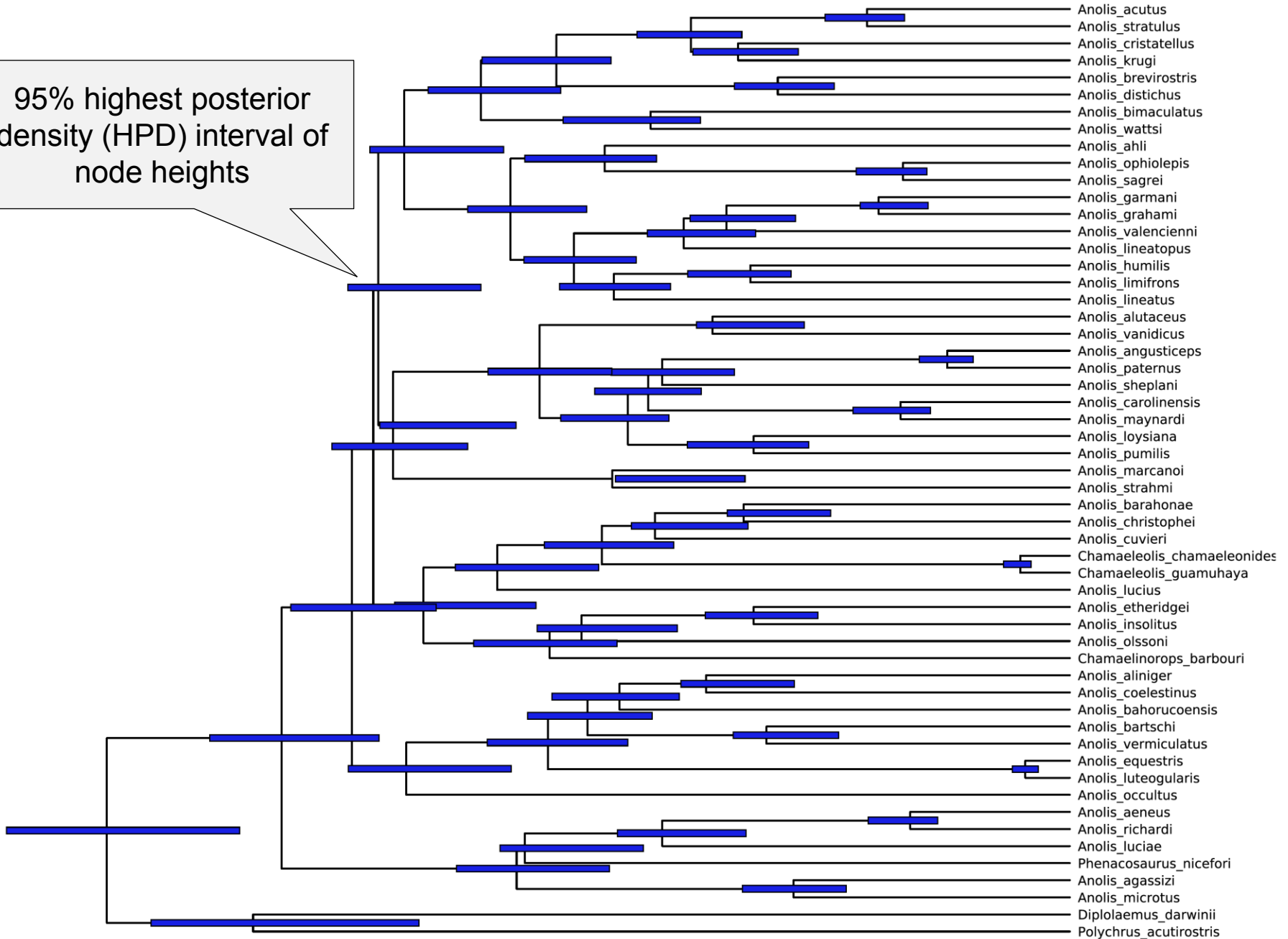
Bayesian analysis:

consensus tree

maximum *a posteriori* tree = tree with greatest posterior probability
(i.e. was sampled most often in the MCMC)

maximum credibility tree = tree with the maximum *product* of the posterior
clade probabilities (BEAST, TreeAnnotator)

95% highest posterior density (HPD) interval of node heights



Phylogenetic programs

phylogeny inference:

<http://evolution.gs.washington.edu/phylip/software.html>

PAUP*

PHYLIP

MOLPHY, PHYML, MEGA ... ML

MrBayes, BEAST ... BA

managing trees:

TreeView

FigTree