Bayesian phylogenetic inference and time-calibrated trees

Heroen Verbruggen
School of BioSciences
heroen@unimelb.edu.au

Principle of Bayesian inference

Bayes theorem

Evidence or observations (D) are used to update the probability of a hypothesis (H).

likelihood

prior probability of hypothesis

$$P(\text{H}\,|\,\text{D}) = \frac{P(\text{D}\,|\,\text{H})\ P(\text{H})}{P(\text{D})}$$

posterior probability of hypothesis

marginal probability of data: constant

Bayesian inference in phylogenetics

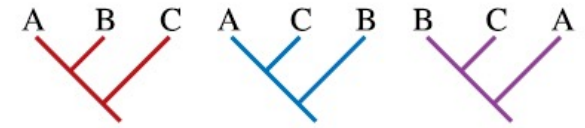$$P(H \mid D) = \frac{P(D \mid H) \; P(H)}{P(D)}$$

D = alignment
H = tree + branch lengths + other model parameter values

$$P(\text{Tree} \mid \text{Alignment}) = \frac{P(\text{Alignment} \mid \text{Tree}) \; P(\text{Tree})}{P(\text{Alignment})}$$

$P(\text{Tree})$: prior probability of tree
- if no prior knowledge → use vague prior
- posterior probability determined primarily by likelihood

Huelsenbeck et al. 2001 Science 294: 2310

Bayesian inference in phylogenetics

The **prior probability** of a tree represents the probability of the tree before the observations have been made. Typically, all trees are considered equally probable, a priori. However, other information can be used to give some trees more prior probability (e.g., the taxonomy of the group).
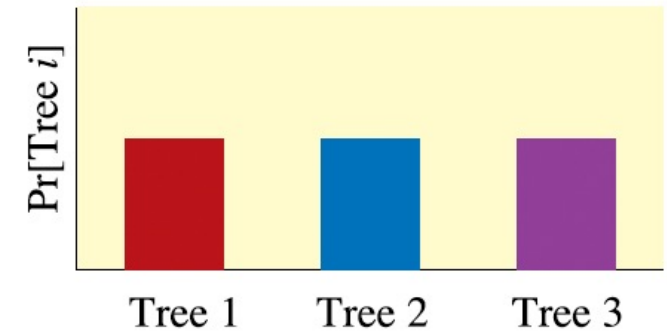
The **likelihood** is proportional to the probability of the observations (often an alignment of DNA sequences) conditional on the tree. This probability requires making specific assumptions about the processes generating the observations.
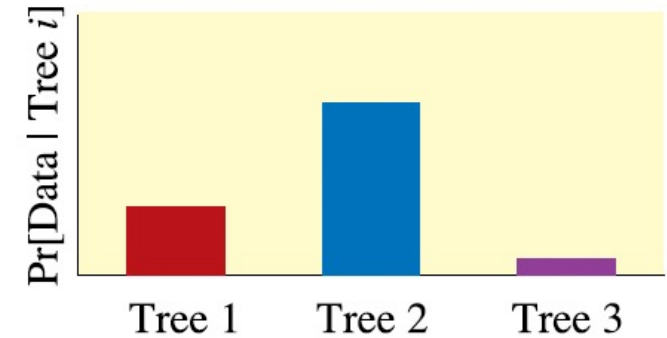
The **posterior probability** of a tree is the probability of the tree conditional on the observations. It is obtained by combining the prior and likelihood for each tree using Bayes' formula.

Huelsenbeck et al. 2001 Science 294: 2310
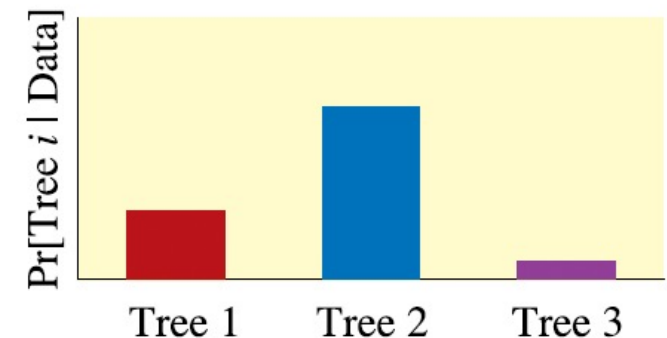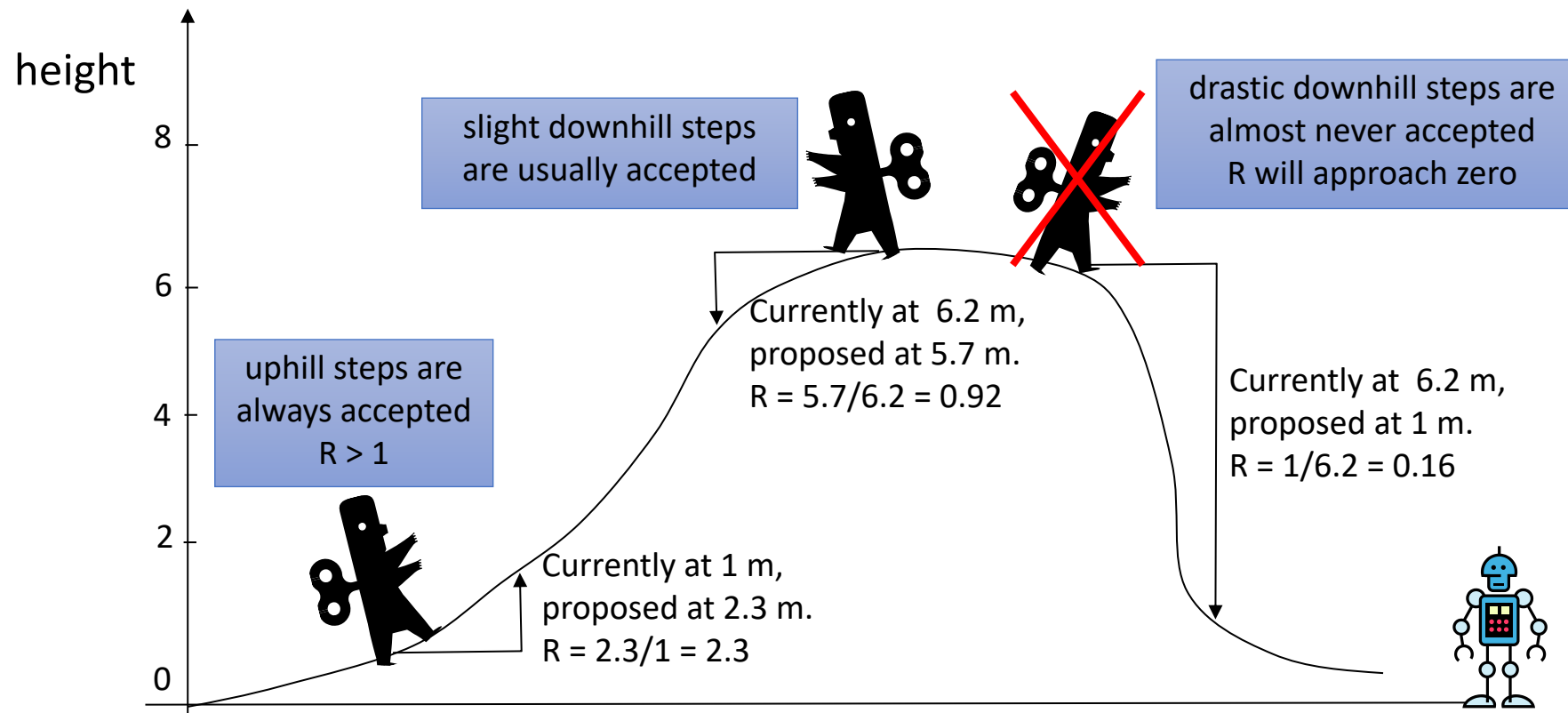
Markov chain Monte Carlo: MCMC

- Posterior probability distribution cannot be calculated analytically.
- Approximated using Markov **chain** Monte Carlo (**MCMC**) simulation.

1. Start with a random tree and parameters
2. In each generation, randomly propose either
   - a new topology
   - a new value for a model parameter
3. If the proposed hypothesis (tree + model parameters) has
   - higher likelihood → add new hypothesis to posterior set
   - lower likelihood → probabilistic rule to add the old or new hypothesis to posterior set depending on how much lower the likelihood is
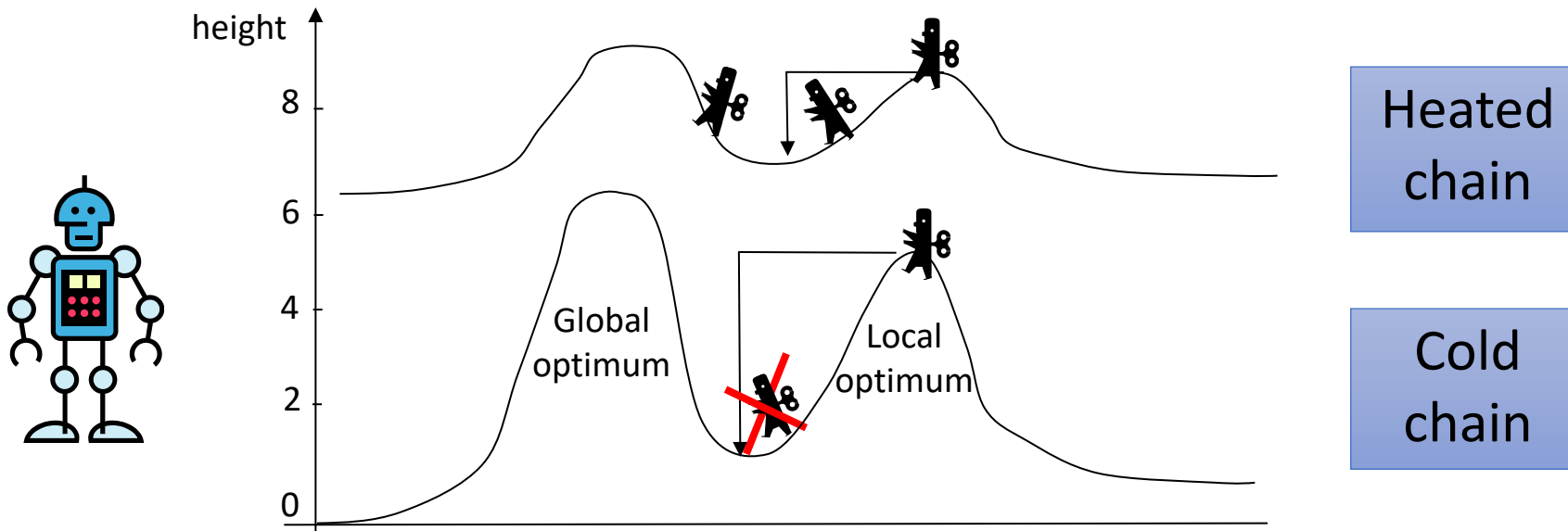4. Repeat steps 2 & 3 millions of times.
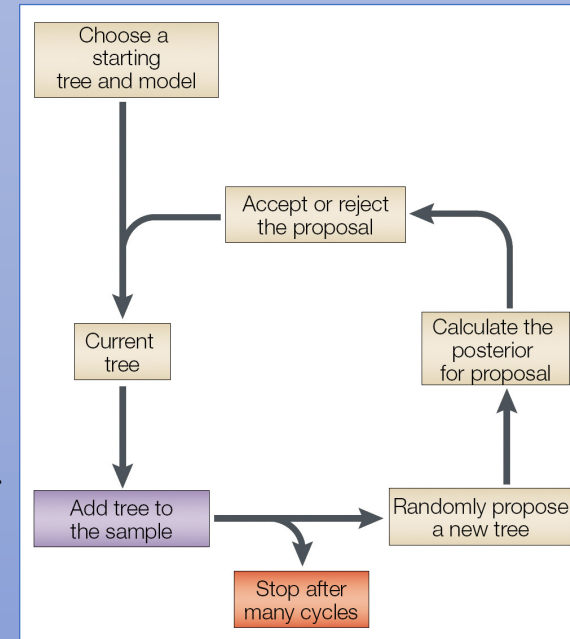
Markov chain Monte Carlo: MCMC

If a proposed step will take the robot downhill, it divides the elevation at the proposed location by its current elevation and only takes the step if it draws a random number (uniform on the open interval 0,1) that is smaller than this quotient.
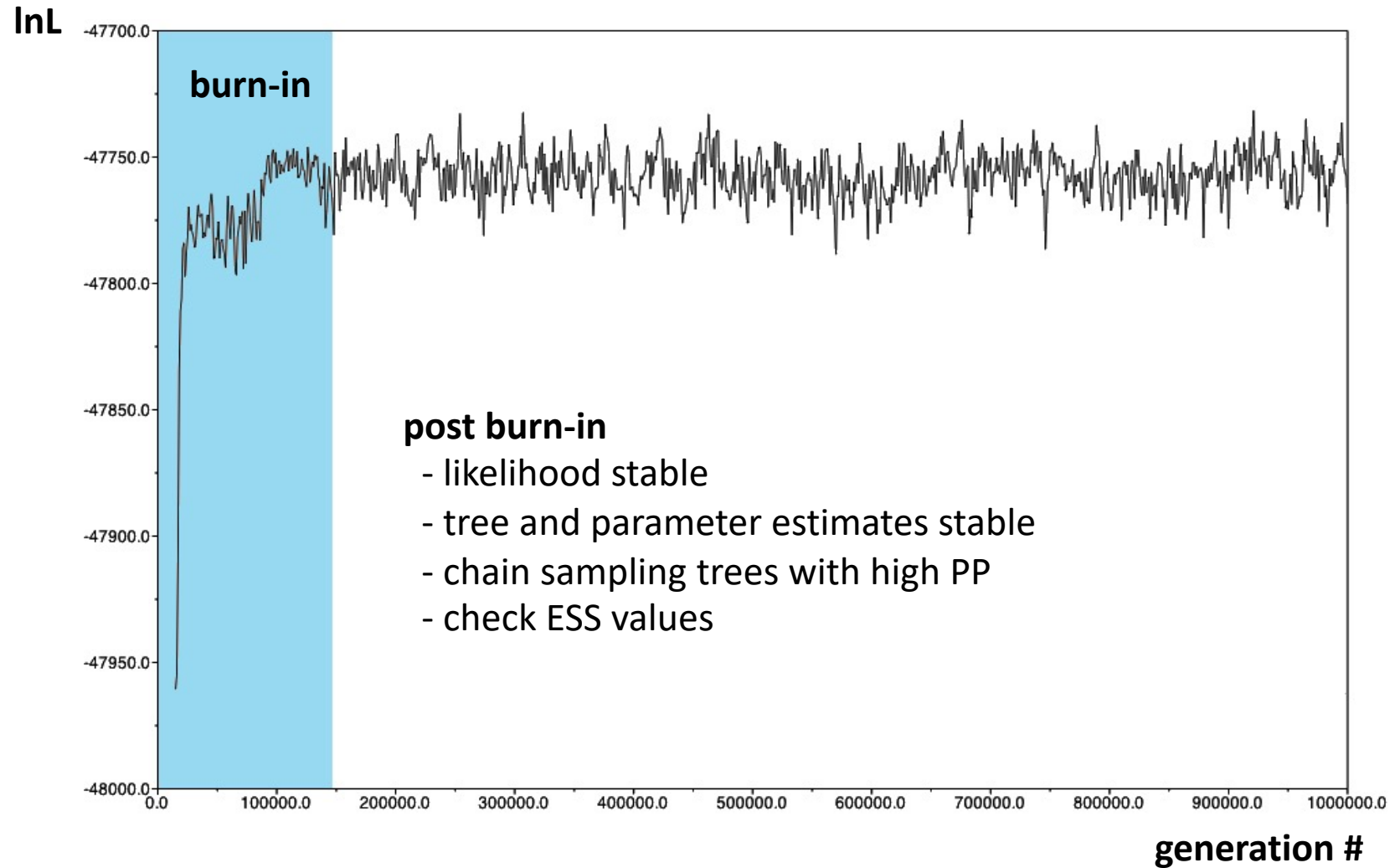

height

slight downhill steps are usually accepted

drastic downhill steps are almost never accepted
R will approach zero

Currently at 6.2 m, proposed at 5.7 m.
R = 5.7/6.2 = 0.92

uphill steps are always accepted
R > 1

Currently at 6.2 m, proposed at 1 m.
R = 1/6.2 = 0.16

Currently at 1 m, proposed at 2.3 m.
R = 2.3/1 = 2.3

Metropolis coupling: MCMCMC or MC^3

- Several chains are run in parallel: cold + incrementally heated

- Heating chains flattens out the posterior distribution, making it easier to hover through tree space

- Chain swapping: hot chains become colder when they reach peaks in the PP distribution

Chain convergence

burn-in

post burn-in
 - likelihood stable
 - tree and parameter estimates stable
 - chain sampling trees with high PP
 - check ESS values

InL axis values: -47700.0, -47750.0, -47800.0, -47850.0, -47900.0, -47950.0, -48000.0

generation # axis values: 0.0, 100000.0, 200000.0, 300000.0, 400000.0, 500000.0, 600000.0, 700000.0, 800000.0, 900000.0, 1000000.0

Traces:

Statistic	Mean	ESS
r(A<->T){2}	7.764E-2	866.246
r(C<->G){2}	0.266	755.498
r(C<->T){2}	0.207	765.019
r(G<->T){2}	7.455E-2	839.878
r(A<->C){3}	6.936E-2	134.86
r(A<->G){3}	0.183	13.92
r(A<->T){3}	1.266E-2	14.65
r(C<->G){3}	8.507E-2	105.464
r(C<->T){3}	0.574	15.014
r(G<->T){3}	7.658E-2	16.726
r(A<->C){4}	6.652E-2	701.327
r(A<->G){4}	0.242	546.43
r(A<->T){4}	0.12	555.595

BI tree?

- Result of MCMC run:
  posterior distribution of 1000s of trees

- Possibilities:
  - Optimality criterion to select one
 - Summarize information somehow

# Maximum A Posteriori (MAP) tree

MrBayes .trprobs file

```
tree tree_1  [p = 0.001, P = 0.001] = [&W 0.001304] (27,26,((4,(24,(23,25))),(22,(21,...
tree tree_2  [p = 0.001, P = 0.002] = [&W 0.001043] (27,26,((4,(24,(23,25))),(22,(21,...
tree tree_3  [p = 0.001, P = 0.003] = [&W 0.000782] (27,26,(((4,24),(23,25)),(22,(21,...
tree tree_4  [p = 0.001, P = 0.004] = [&W 0.000522] (27,26,((4,(24,(23,25))),(22,(21,...
tree tree_5  [p = 0.001, P = 0.004] = [&W 0.000522] (27,26,((4,(24,(23,25))),(22,(21,...
tree tree_6  [p = 0.001, P = 0.005] = [&W 0.000522] (27,26,((23,(25,(4,24))),(21,(22,...
tree tree_7  [p = 0.001, P = 0.005] = [&W 0.000522] (27,26,((4,(24,(23,25))),(21,(22,...
tree tree_8  [p = 0.001, P = 0.006] = [&W 0.000522] (27,26,((4,(24,(23,25))),((21,22)...
tree tree_9  [p = 0.001, P = 0.006] = [&W 0.000522] (27,26,((25,(23,(4,24))),(21,(22,...
tree tree_10 [p = 0.001, P = 0.007] = [&W 0.000522] (27,26,((4,(24,(23,25))),((21,22...
tree tree_11 [p = 0.001, P = 0.007] = [&W 0.000522] (27,26,((4,(24,(23,25))),(22,(21...
tree tree_12 [p = 0.001, P = 0.008] = [&W 0.000522] (27,26,((4,(24,(23,25))),(22,(21...
tree tree_13 [p = 0.001, P = 0.008] = [&W 0.000522] (27,26,((4,(24,(23,25))),((21,22...
tree tree_14 [p = 0.001, P = 0.009] = [&W 0.000522] (27,26,((4,(24,(23,25))),(22,(21...
tree tree_15 [p = 0.001, P = 0.009] = [&W 0.000522] (27,26,((4,(24,(23,25))),(22,(21...
tree tree_16 [p = 0.001, P = 0.010] = [&W 0.000522] (27,26,((4,(24,(23,25))),(22,(21...
...
...
```
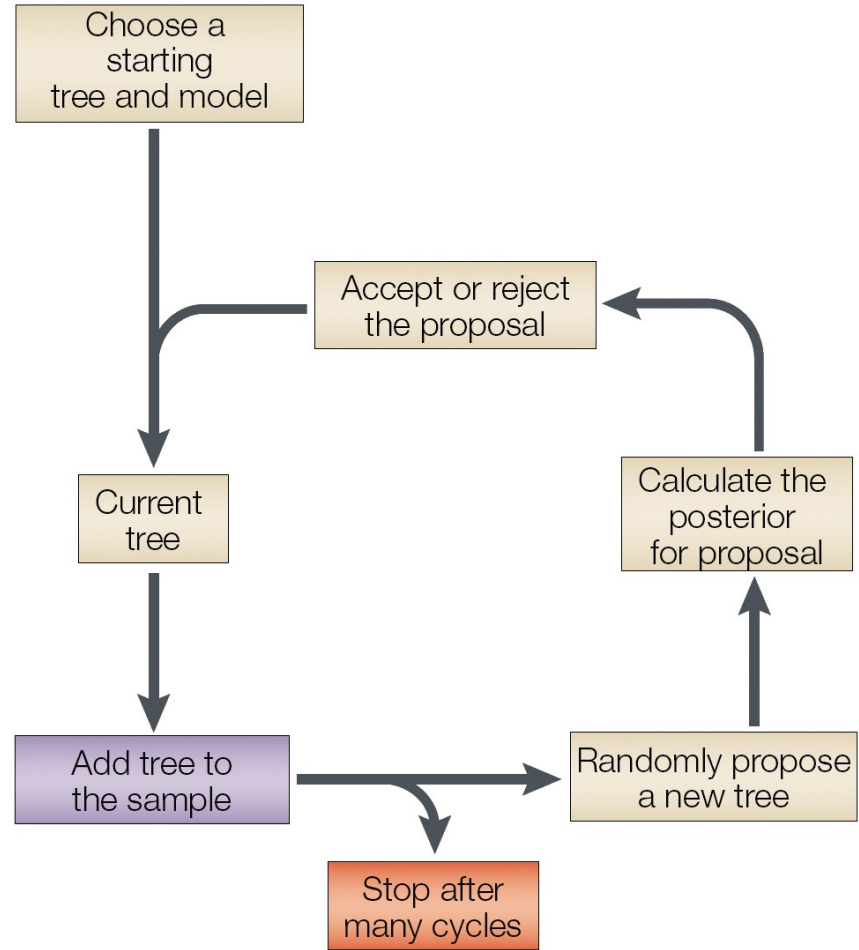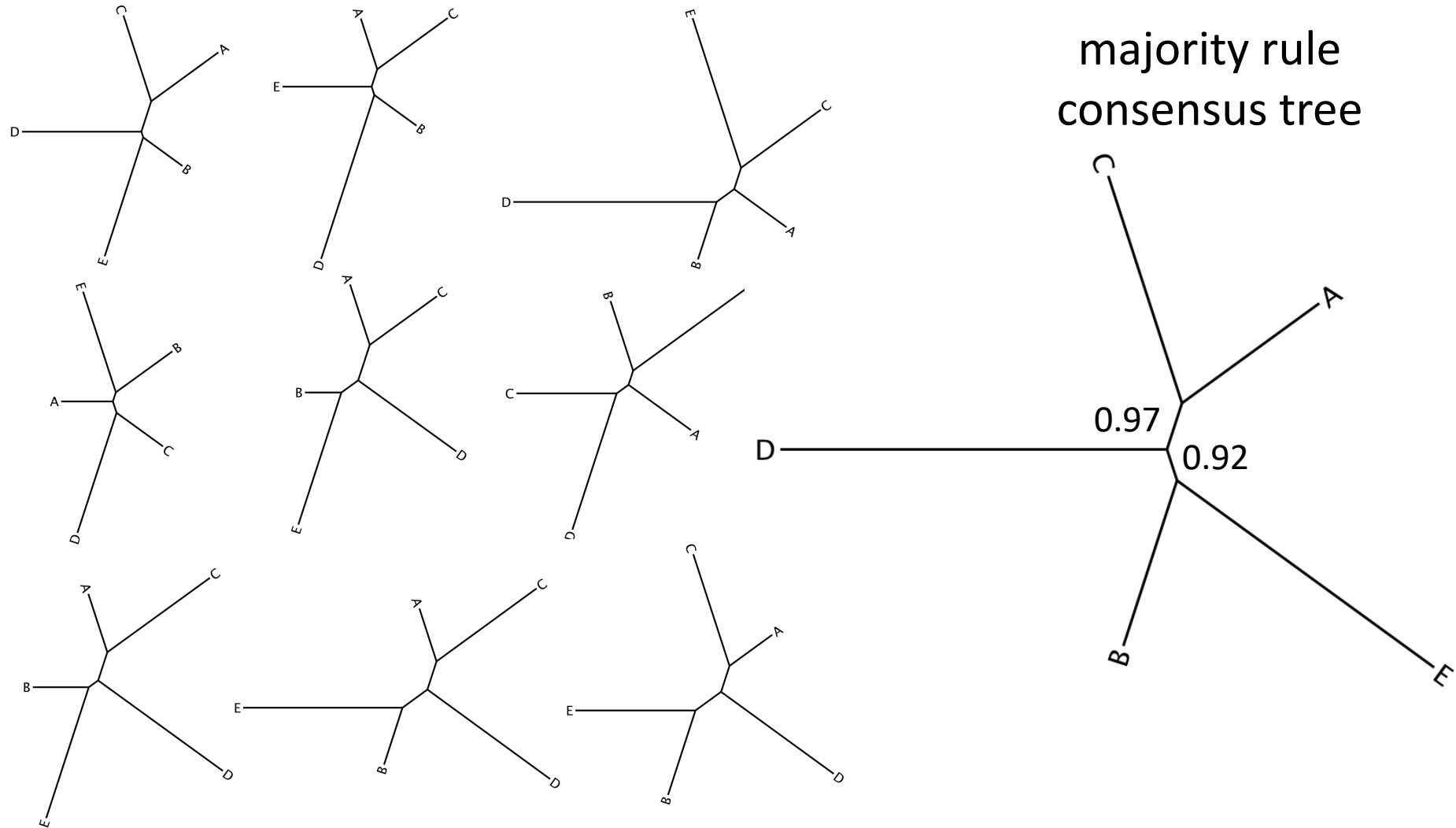
Consensus tree: summary

50% majority rule (halfcompat)
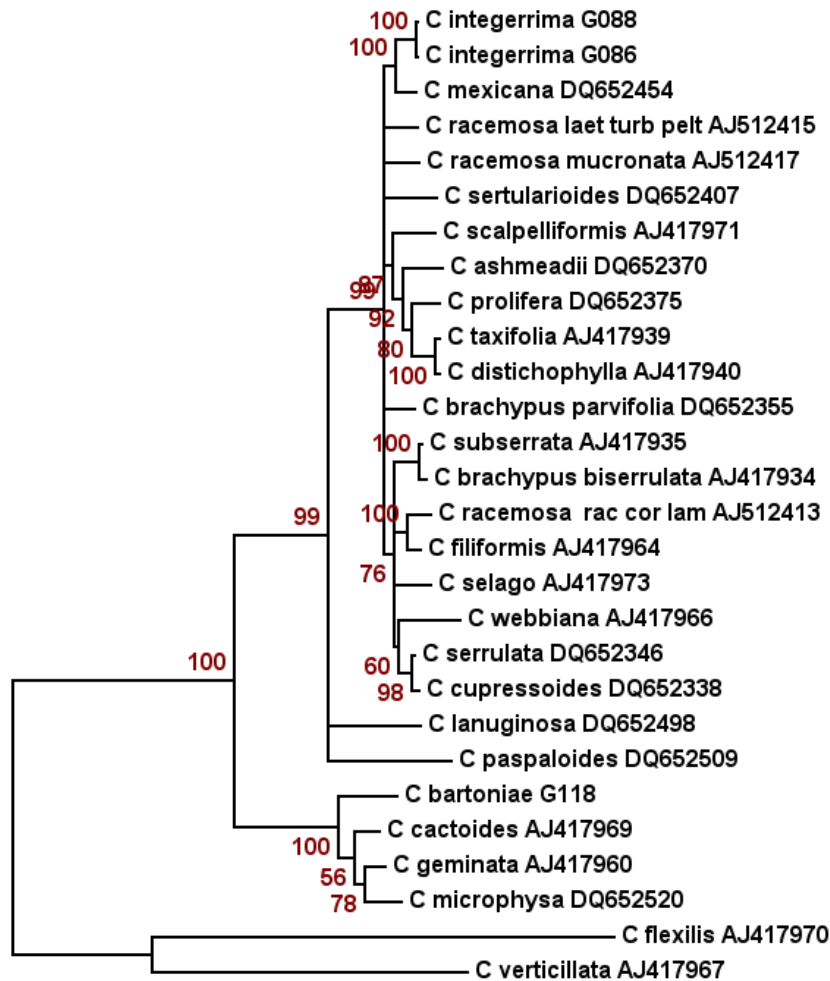
Majority rule with all compatible groups added (allcompat)
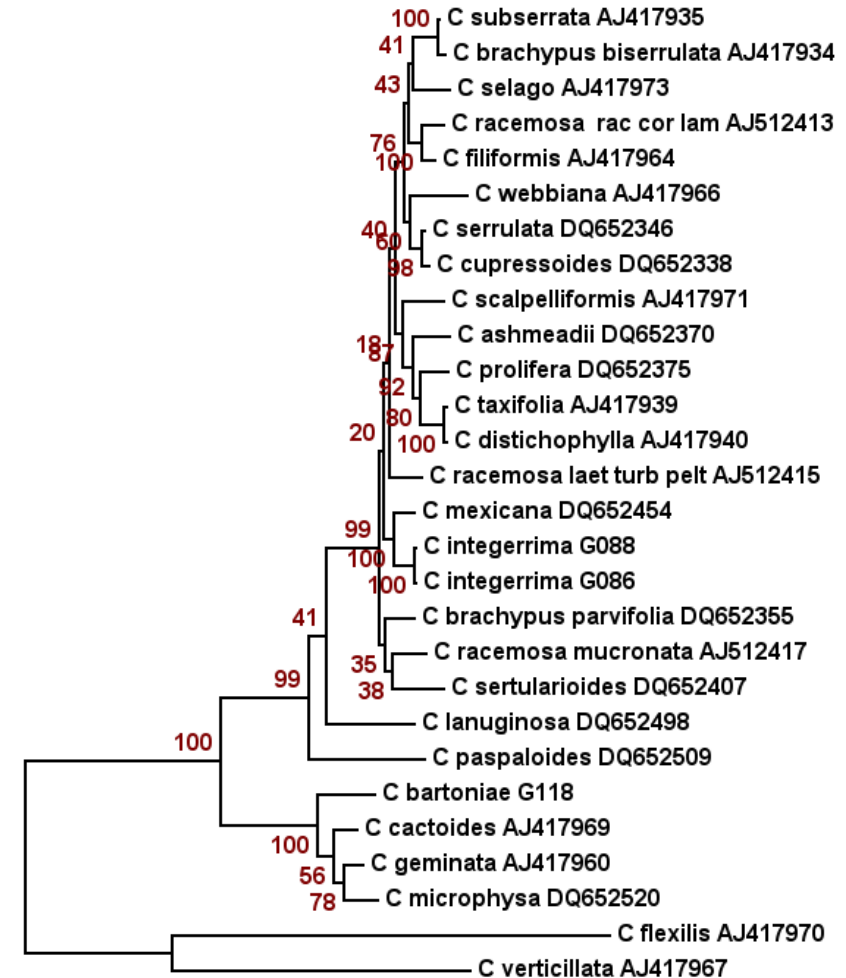
Posterior probabilities for phylogenetic uncertainty

Posterior probabilities for phylogenetic uncertainty

majority rule consensus tree

0.97

0.92

Posterior probabilities vs. bootstrap values

- Observation: **PP >> BV**

- Statistical interpretation of PP and BV

    - PP: probability that the clade is correct given the data, model and prior

    - BV: no simple statistical interpretation; could be seen as robustness to changing character weights

- Interpretation in empirical work:

PP	BV	Interpretation
> 0.90	> 50	indication that branch may be real
> 0.95	> 70	good support
= 1.00	> 80	very good support

- Should not be used without consideration of the conditions under which values were determined

- Compare support under different models

# Time-calibrated trees

Chronogram
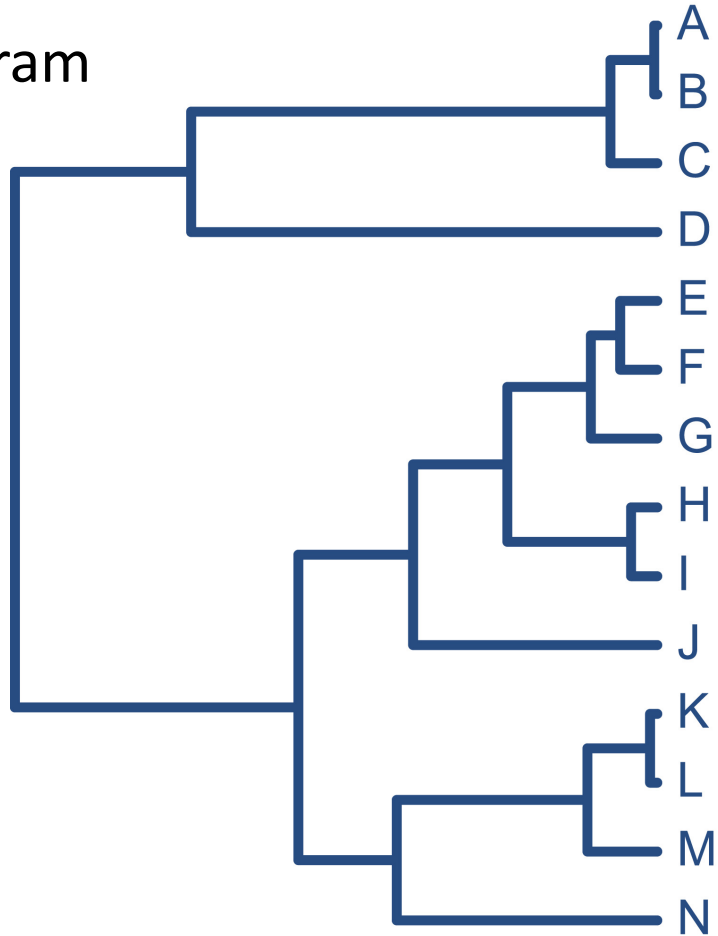
Phylogram

branch lengths ∝ time
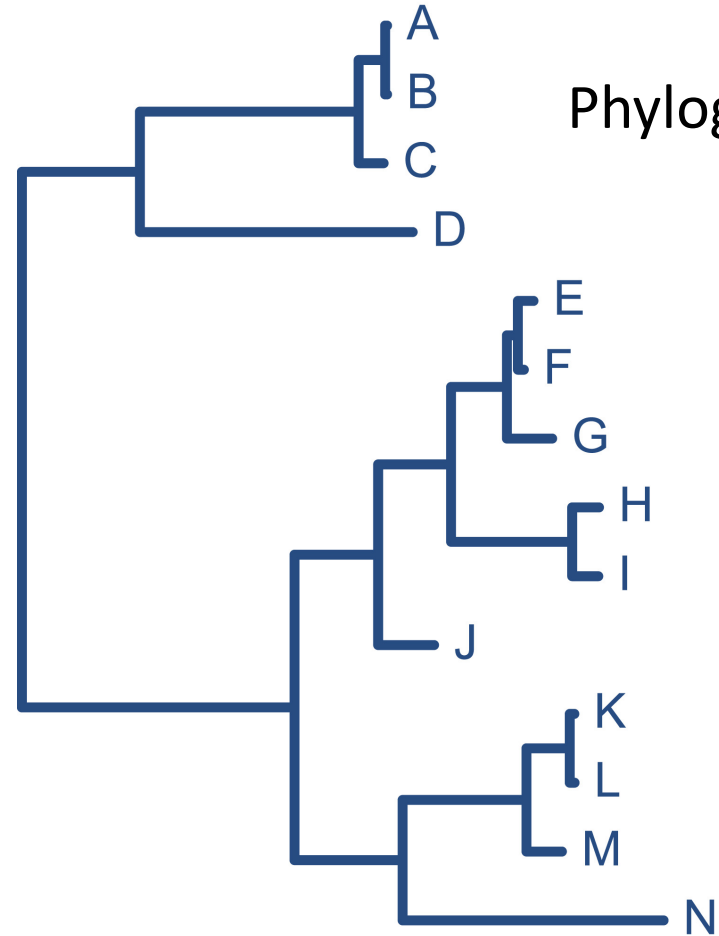
branch lengths ∝ amount of molecular evolution

Dating trees for richer interpretaion

- Interpret branching times with earth history

- Localise key events in time (e.g. genome duplication, radiation of gene family)

- Interpret biogeographic scenarios

- When did characters originate?
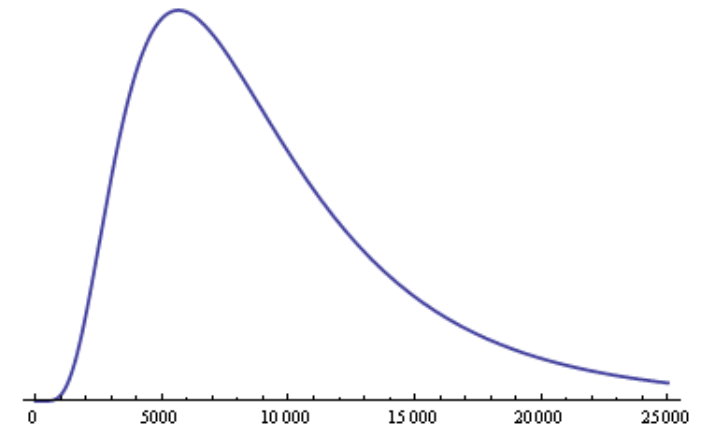
- Calculate rates of speciation

Molecular clocks

- The "molecular clock" ticks at a constant rate
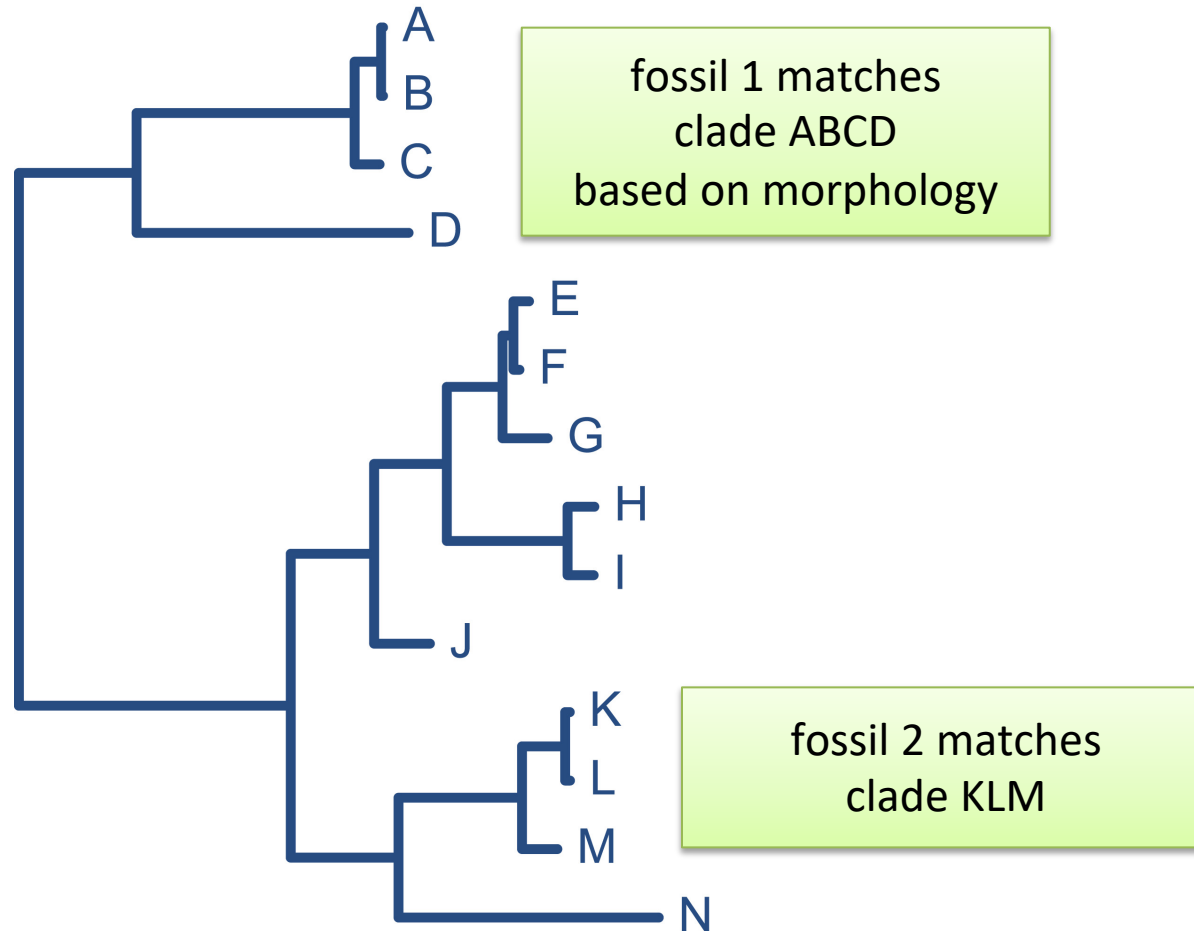- Seldom true → relax that assumption by modelling rate variation
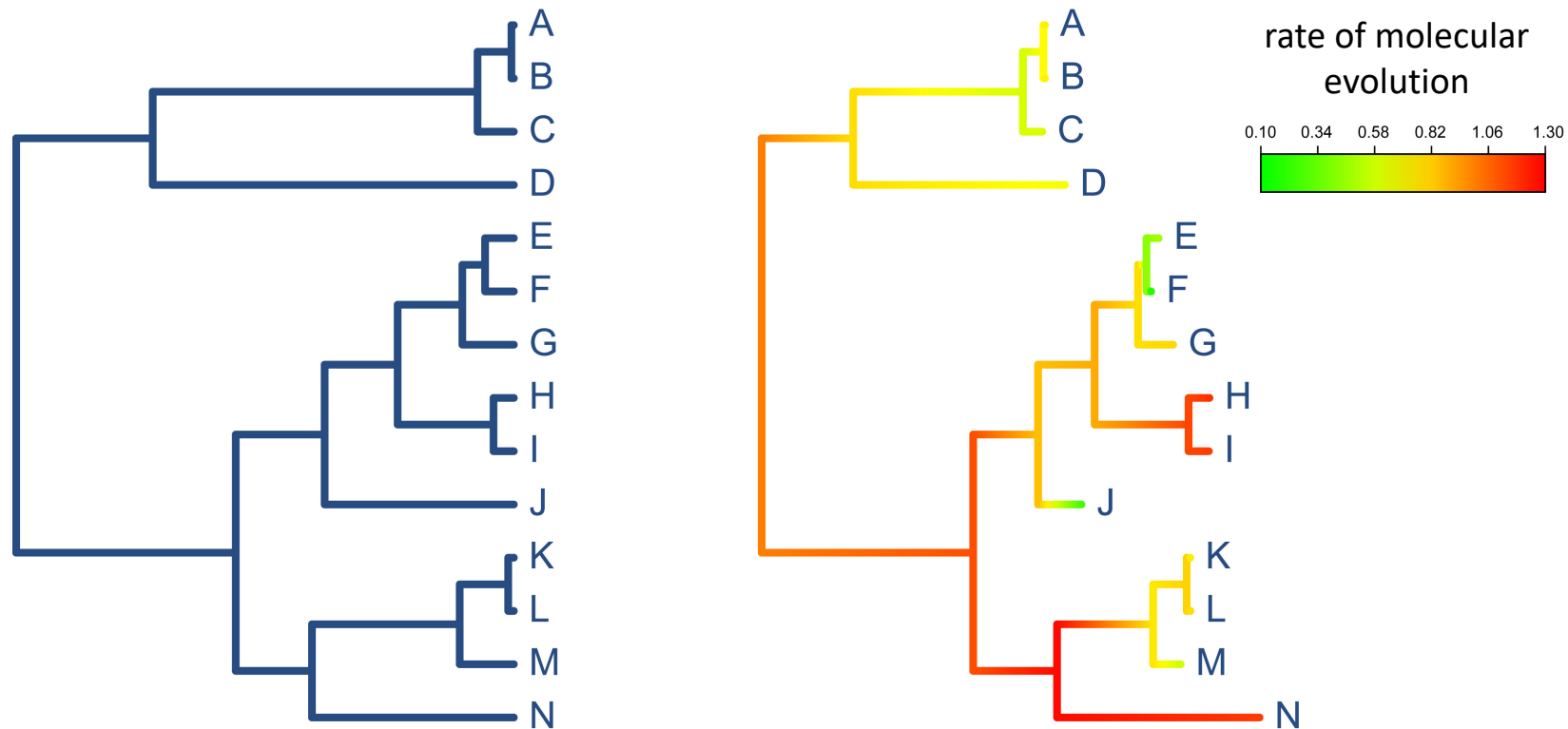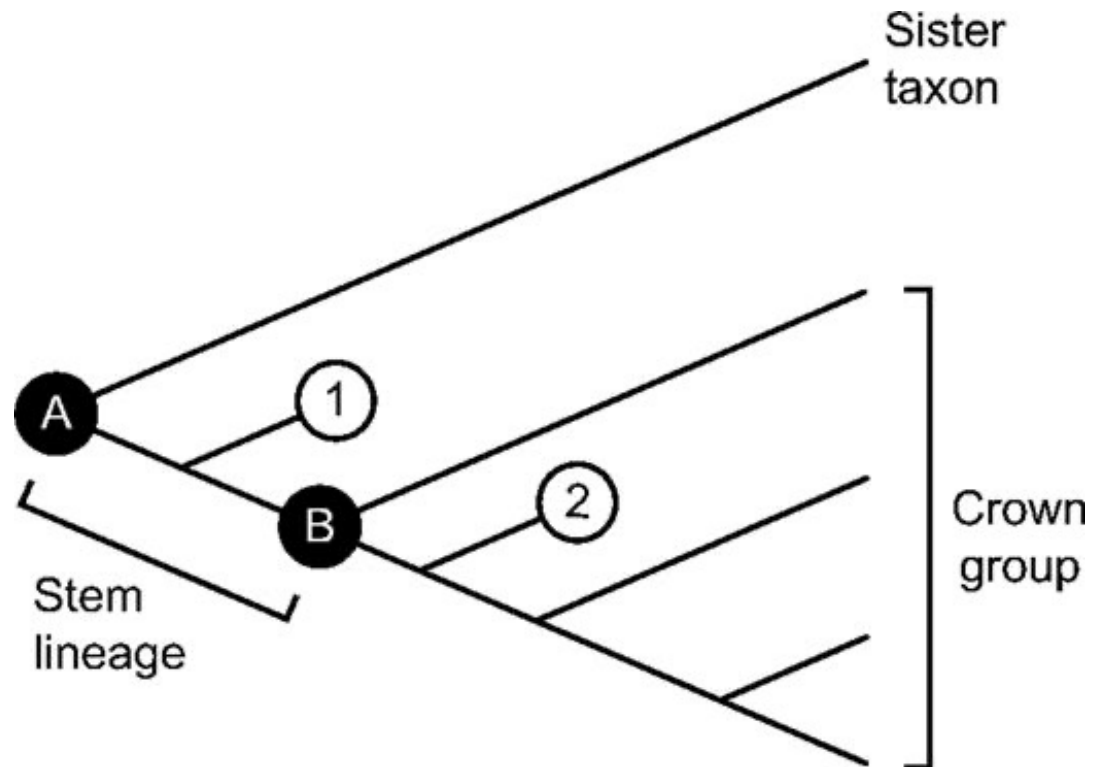
Relaxing the molecular clock

- Rate autocorrelation: rates are hereditary
  - Generation times
  - DNA proofreading mechanisms
  - Habitat conservatism
  - Small rate changes are more likely than large changes

- Uncorrelated model:
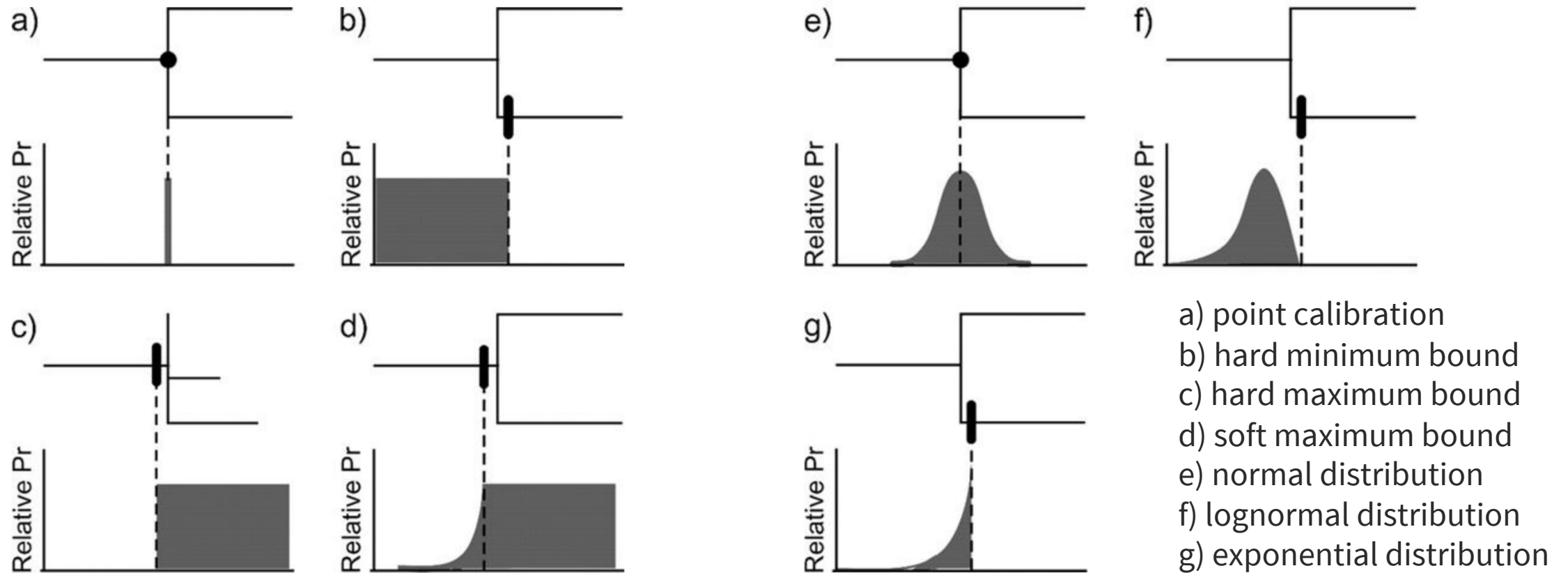 - Branch lengths sampled from a distribution

Using fossils to constrain node ages
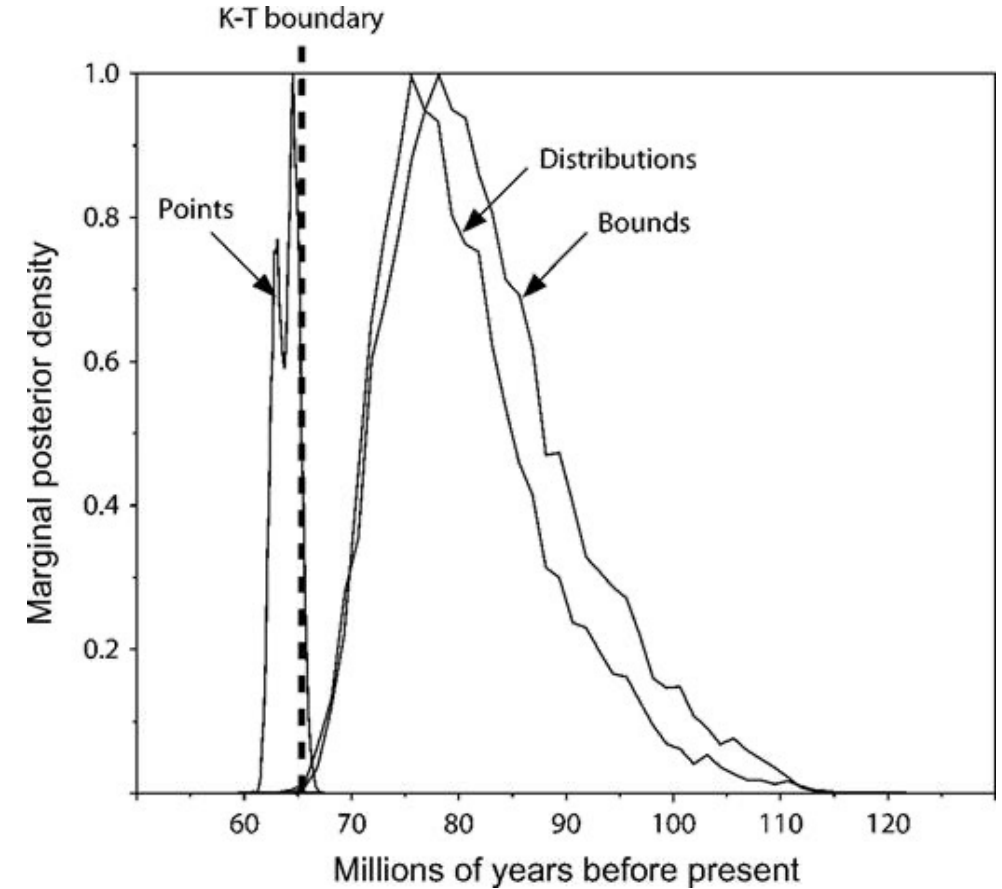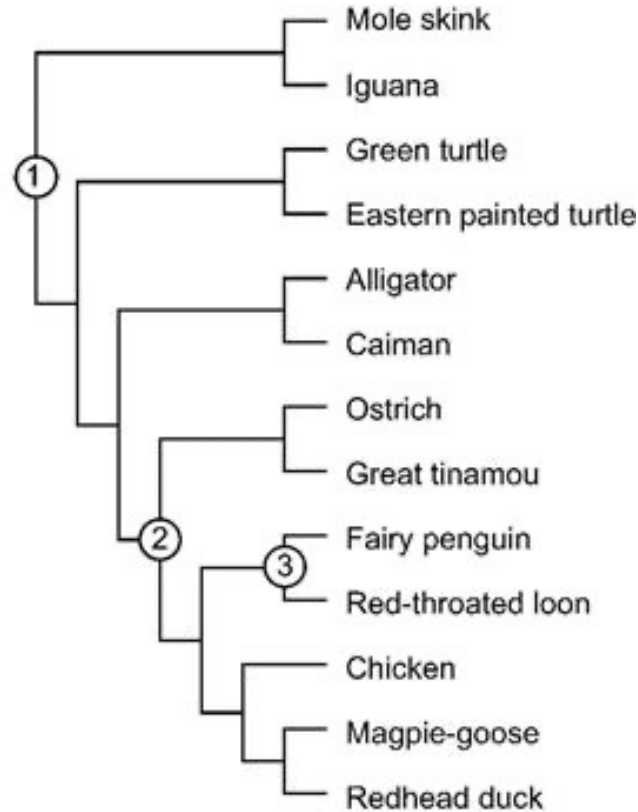
Using fossils to constrain node ages

Using fossils to constrain node ages: node priors

a) point calibration
b) hard minimum bound
c) hard maximum bound
d) soft maximum bound
e) normal distribution
f) lognormal distribution
g) exponential distribution

Using fossils to constrain node ages: node priors

Ho & Phillips 2009 Syst Biol 58: 367

Result: chronogram with node ages and uncertainty bars