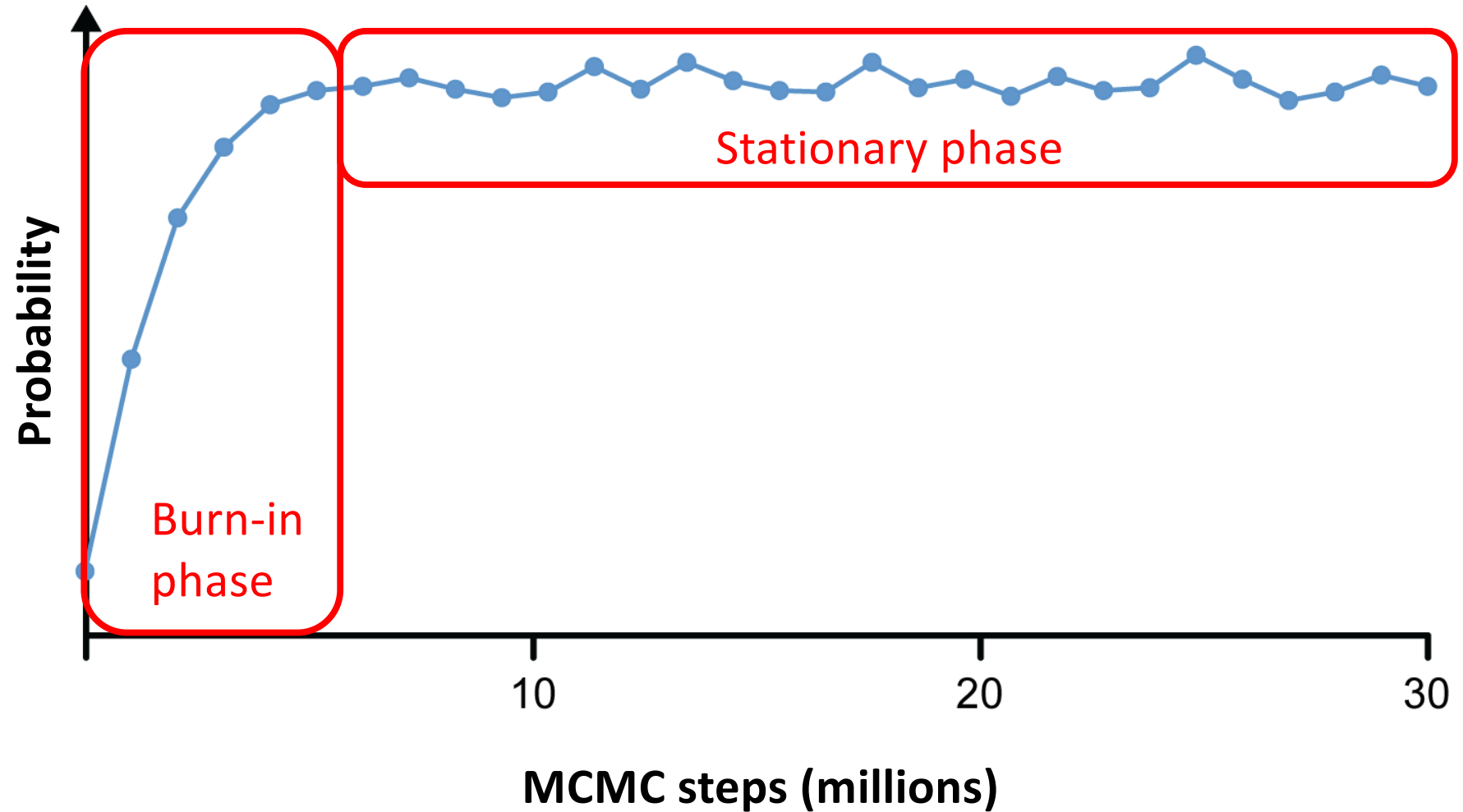

Lecture 2.3

Bayesian Phylogenetics II

MCMC Diagnostics

Samples from the MCMC



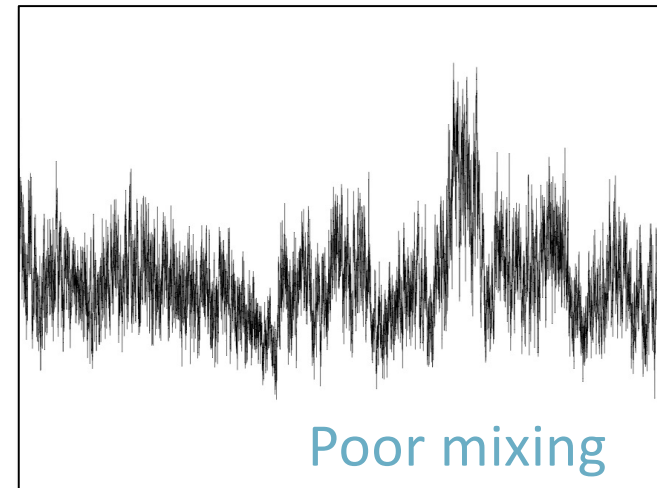
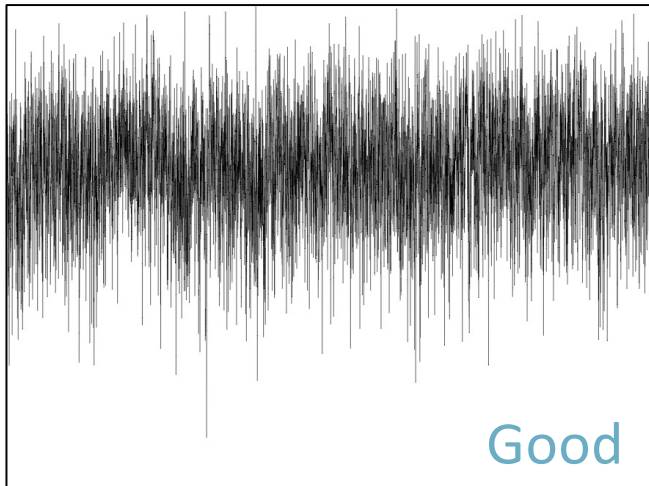
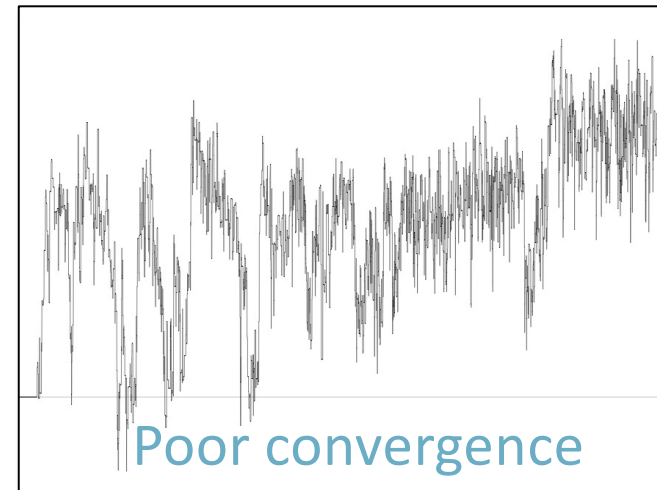
Diagnostics

1. Convergence

Are we drawing samples from the stationary distribution?

2. Sufficient sampling

Have we drawn enough samples to allow a reliable estimate of the posterior distribution?



Convergence

- Run at least 2 (preferably more) independent chains
- Likelihoods should be similar
- Estimates of model parameters should be similar

Sufficient sampling

- **Effective sample size (ESS)**

Have we drawn enough independent samples to produce a reliable estimate of the posterior distribution?

- ESS is preferably **>200** for each parameter
- ESS can be increased by:
 - Increasing the length of the MCMC
(and decreasing the frequency of sampling accordingly)
 - Modifying the MCMC proposals

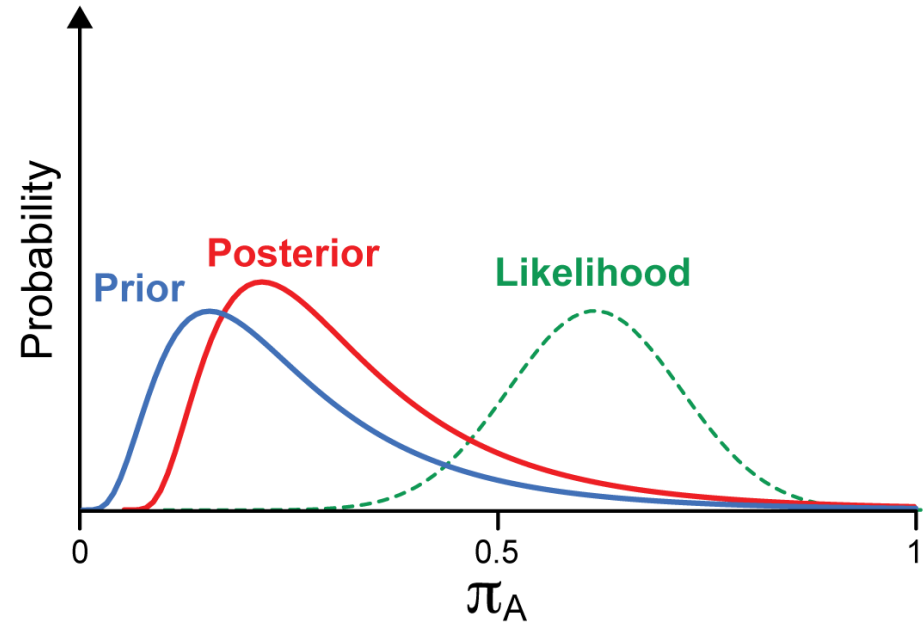
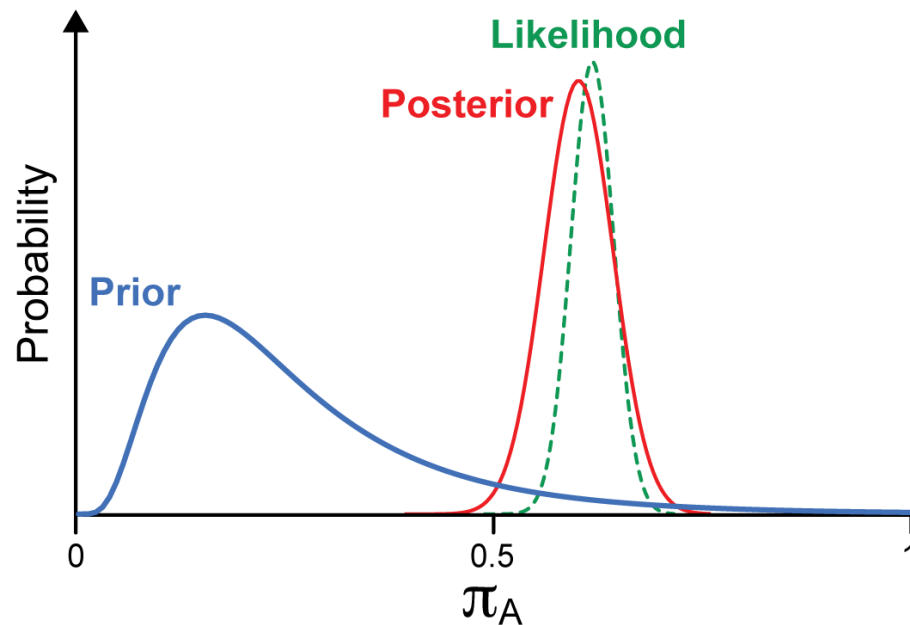
Advantages and Problems

Advantages

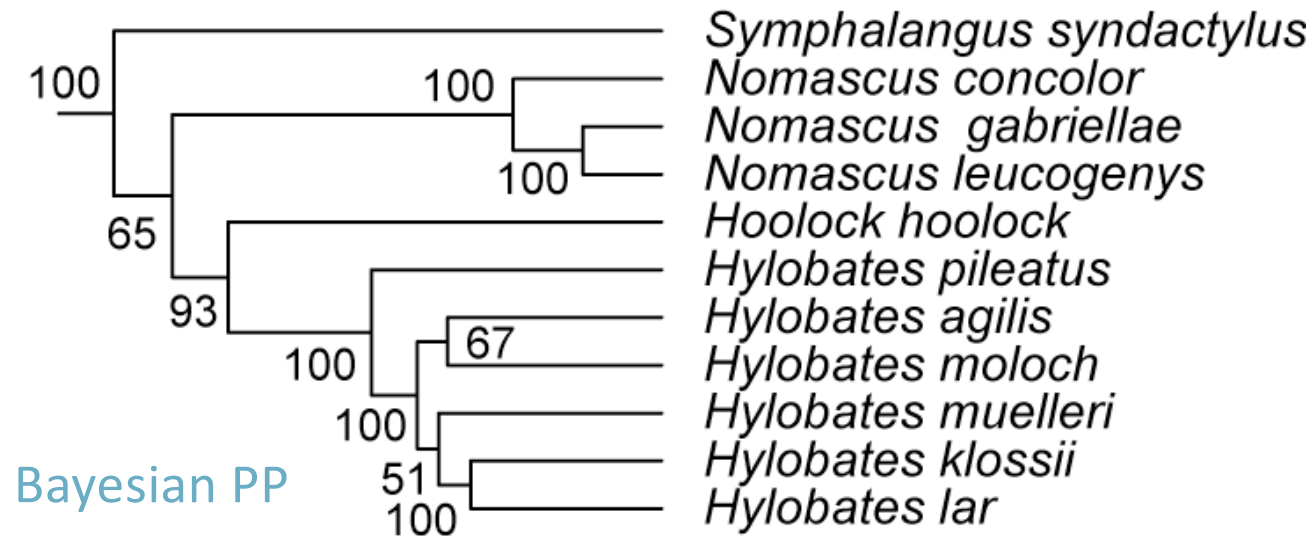
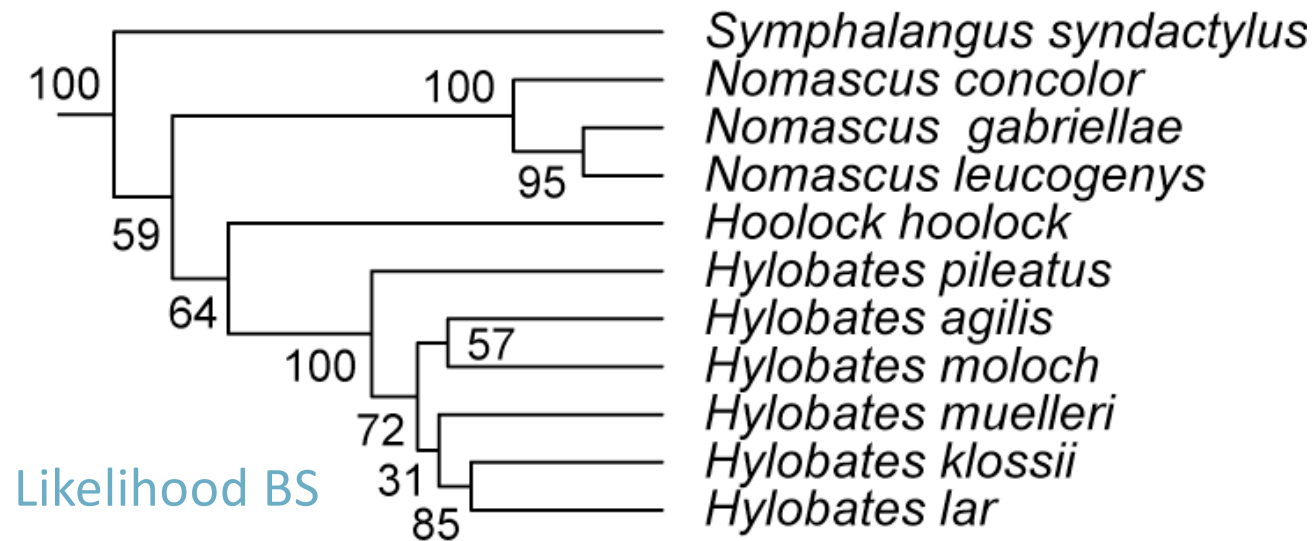
- Able to implement complex (highly parameterised) models
- Posterior probabilities have an intuitive interpretation
- Can incorporate independent information (in the prior)
- Integrate over 'nuisance' parameters

Influence of priors

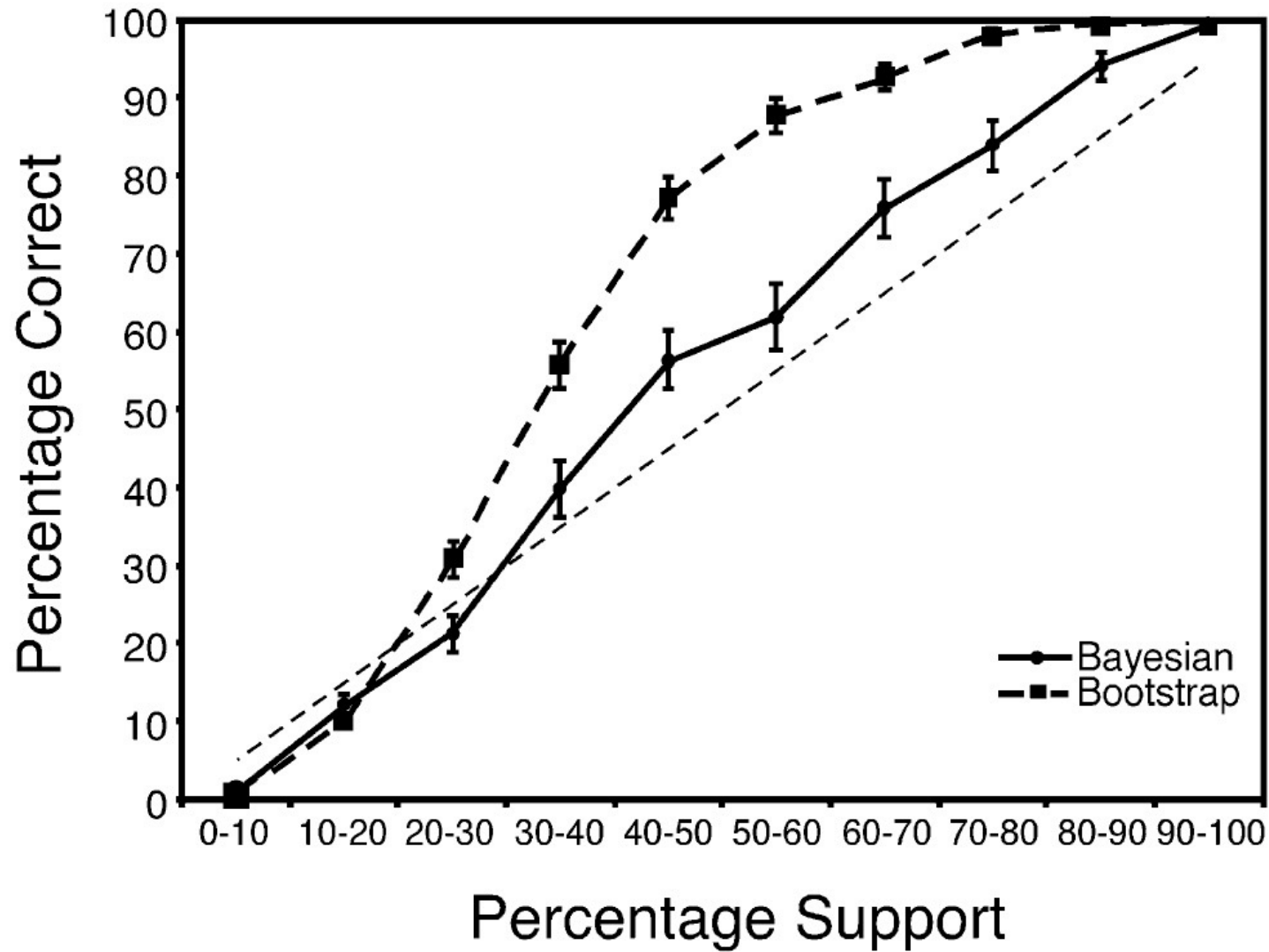
- Sensitivity of the posterior to the prior
- This problem can occur if the data are uninformative, the prior is strong, or both



Node support



Node support



Bayesian Model Selection

Bayesian model selection

- Bayesian model selection is usually based on the marginal probability of the data, conditioned on the model:

$$\mathbf{Pr(D | M)}$$

- This is a weighted average of the likelihood
- Weights are given by the prior distribution

Marginal likelihood of the model

Bayesian model selection

- Compare marginal likelihoods of competing models
- Ratio of marginal likelihoods is the **Bayes factor**

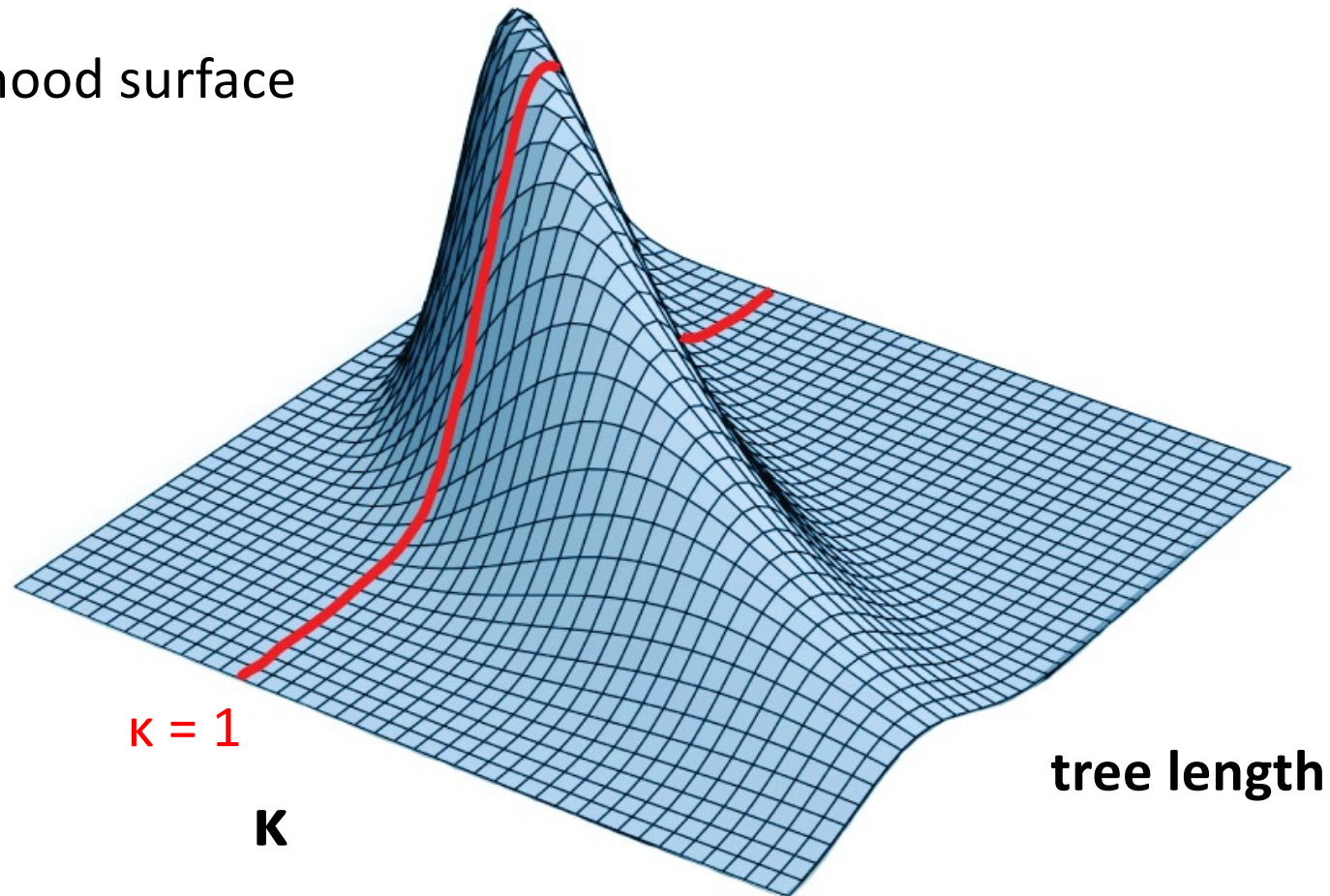
$$\mathbf{BF} = \frac{\mathbf{Pr(D | M_1)}}{\mathbf{Pr(D | M_2)}}$$

$$\mathbf{logBF = logPr(D | M_1) - logPr(D | M_2)}$$

- Models do not need to be nested
- Do not need to correct for number of parameters

Bayesian model selection

Likelihood surface



Bayesian model selection

- Interpreting the Bayes factor

BF	$\log BF$	Evidence against M_2
1 – 3	0 – 1	Not worth mentioning
3 – 20	1 – 3	Positive
20 – 150	3 – 5	Strong
> 150	> 5	Very strong

Software for Bayesian Phylogenetics



BEAST 1

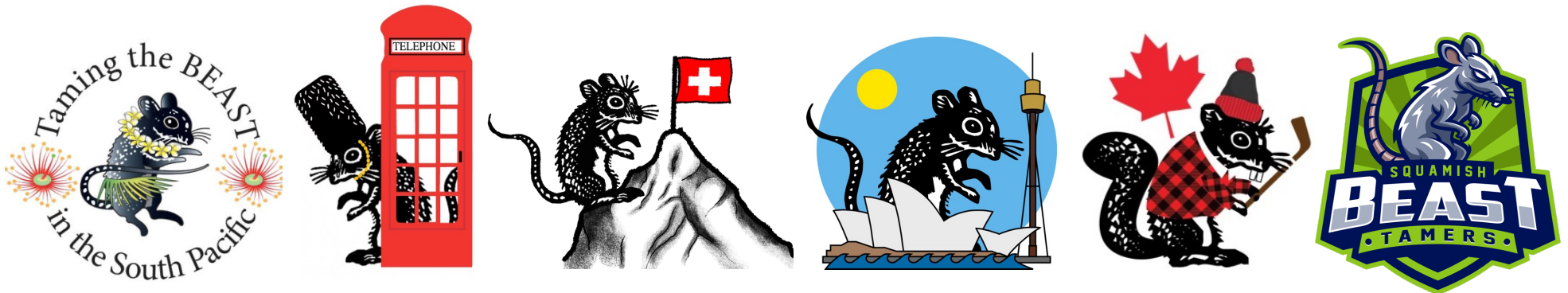


- **B**ayesian **E**volutionary **A**nalysis by **S**ampling **T**rees
- Analyse population- or species-level data
- Simultaneous estimation of tree and node times
- Range of clock models
- Range of tree priors and demographic models



- Re-write of *BEAST* to increase modularity
- Users can extend *BEAST* by adding packages
- Additional tree priors not available in *BEAST* 1
- Capacity to perform simulations

For a comparison of *BEAST* 1 and 2:
www.beast2.org/beast-features



MrBayes

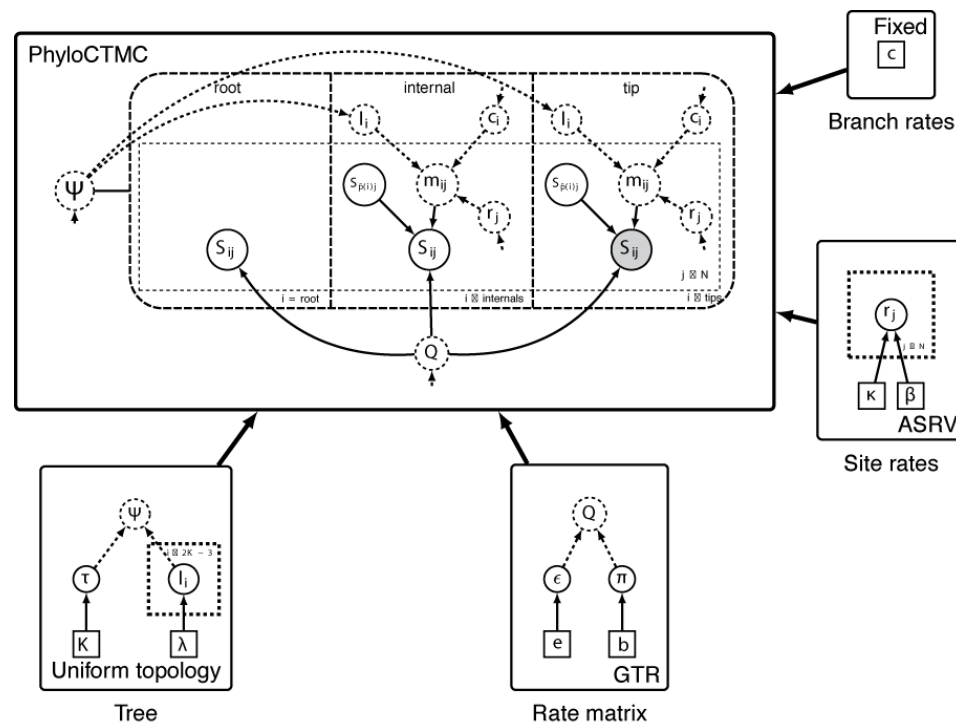
- Primarily designed for species-level data
- Simultaneous estimation of tree and node times
- Range of clock models
- Range of tree priors
- Multiple chains and MCMC diagnostics



RevBayes



- Uses its own R-like language, Rev
- Interactive construction of graphical model
- Flexible and can be used for simulation and inference



Useful references

