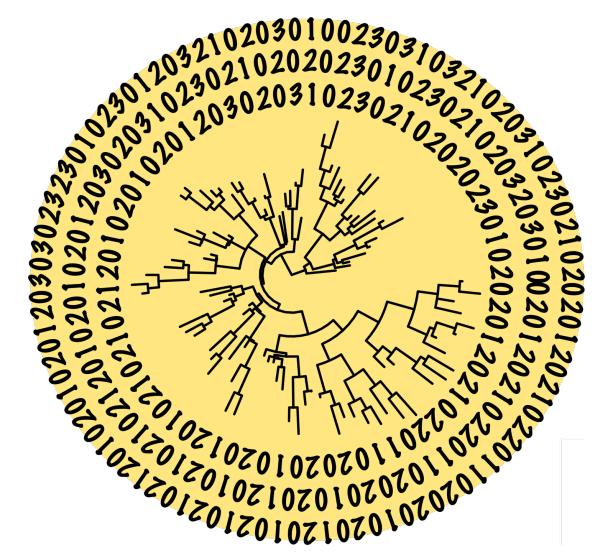


Morphological models and how to choose them

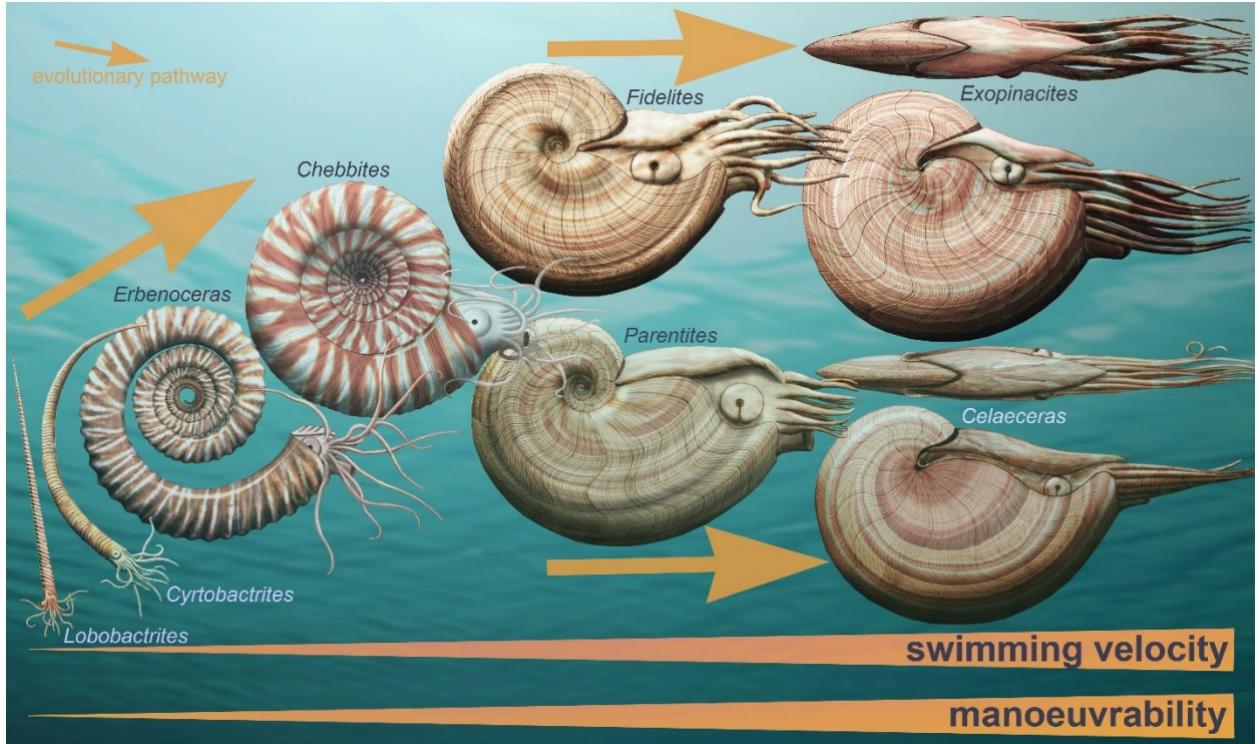
Laura Mulvey
Phylogenetics Workshop



Morphological data

Morphological data was the original type of information used in phylogenetic analysis

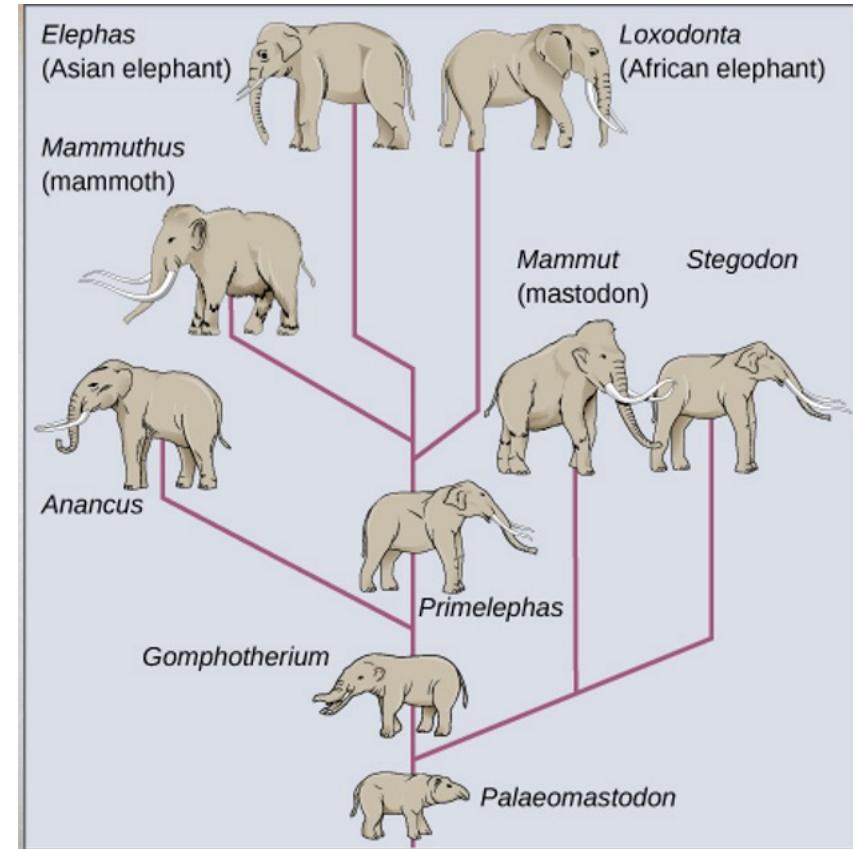
Fossils can be used to provide time calibrations, helps extant phylogeny, allows us to understand evolution through time



Morphological character data

Discrete Characters: Morphological data often consist of discrete characters, such as the presence or absence of certain traits, or more complex multistate traits (e.g., number of limbs, type of leaf, presence of a particular bone structure)

Continuous Characters: Some morphological data can be continuous, such as measurements of body size, length of bones, or other quantitative traits



Morphological character data

Discrete Characters: Morphological data often consist of discrete characters, such as the presence or absence of certain traits, or more complex multistate traits (e.g., number of limbs, type of leaf, presence of a particular bone structure)

Continuous Characters: Some morphological data can be continuous, such as measurements of body size, length of bones, or other quantitative traits

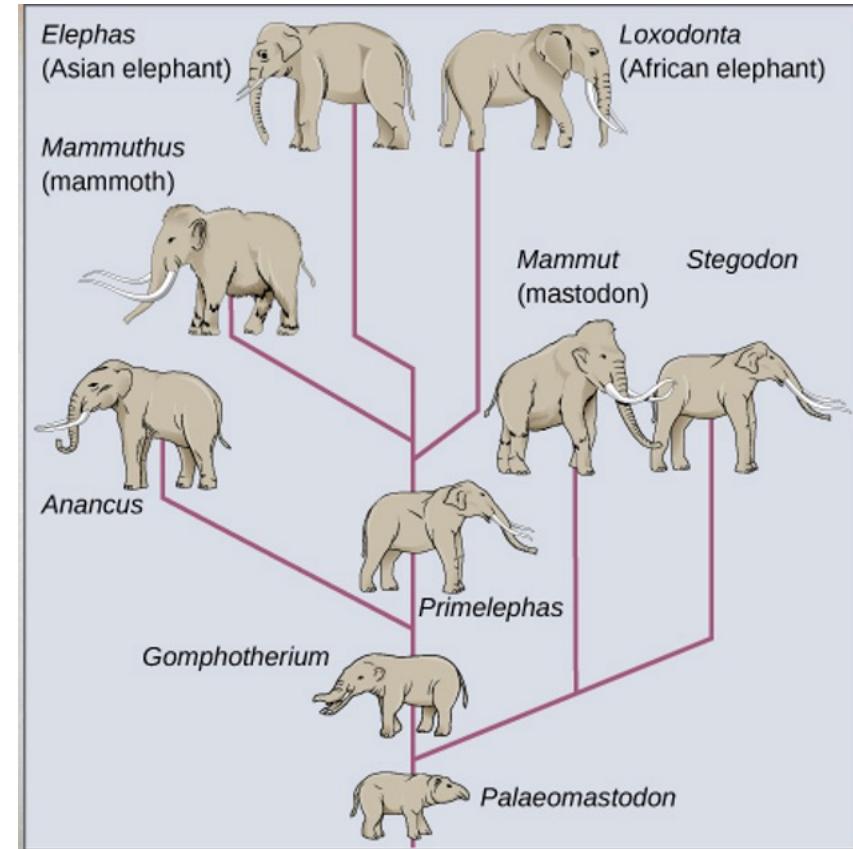
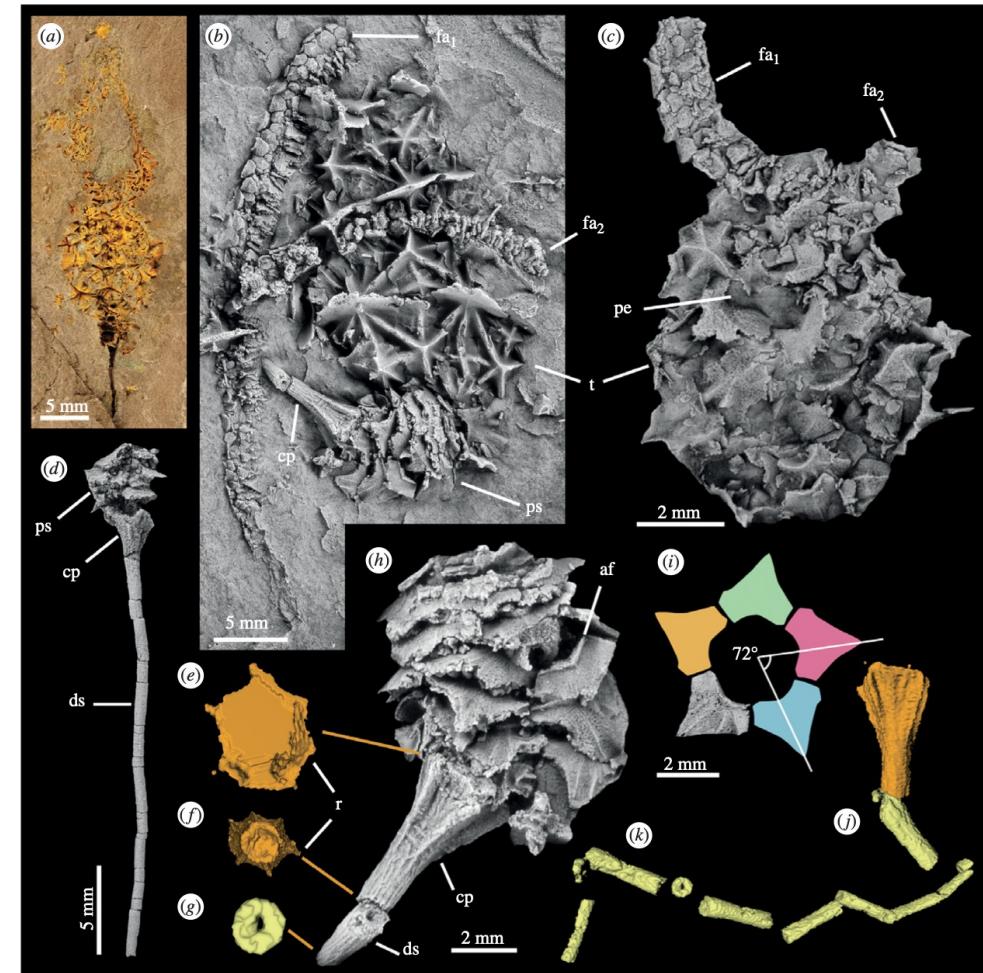


Image from
<https://www.zoologytalks.com/>

Trait 1

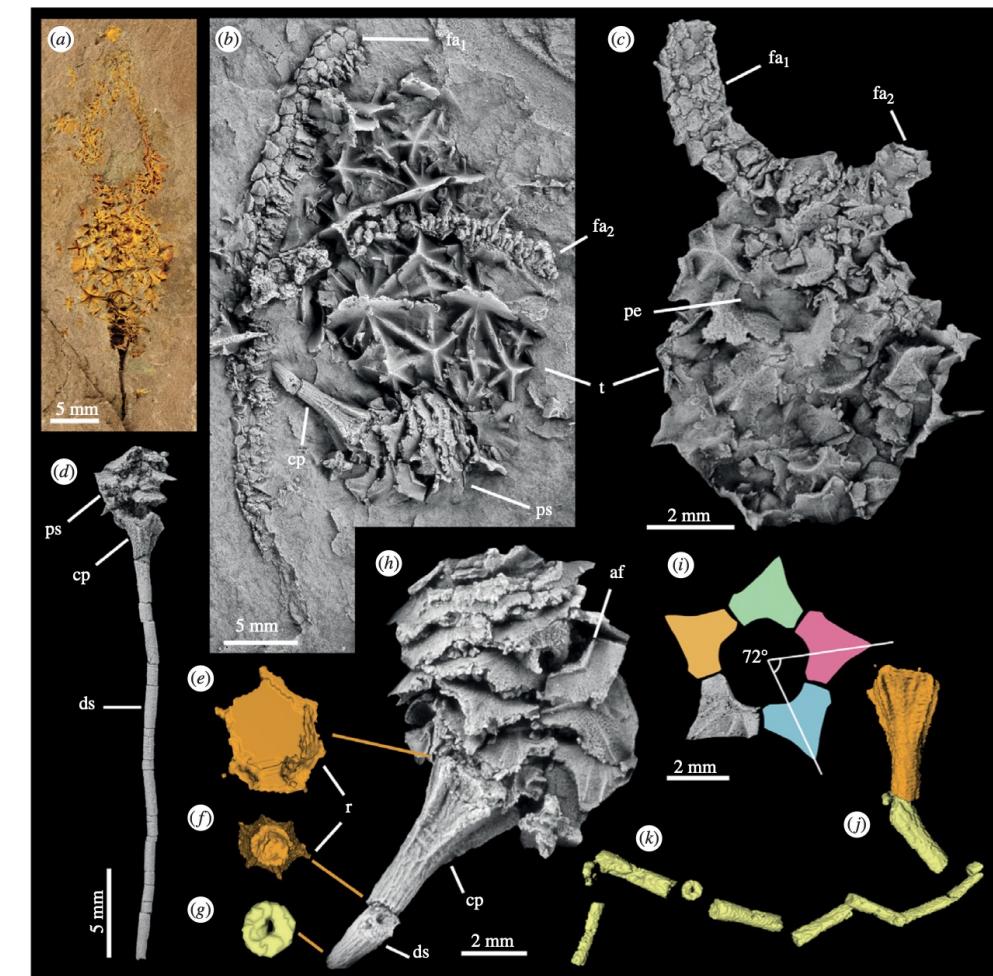
Taxa 1 001510010?00-100--000000000
000500010?200100--001001000
002500010?200100--0?1001000
00?5?0010?200100?-0???010110
0015000101201000430100011111
0015000101201010440111011111
??050?????201000440?11011111
01050?010-210000?501??010110
00020001002101003-1110010110
0002000100211001441121011111
000201111-210010?-??11011121
?103?0?11?1001104-0000010000
1005002110100010--0?00110?20
Taxa 14 1005002000101010540?00110020

Trait 28



Cambrian stalked echinoderms show
unexpected plasticity of arm construction
Zamora & Smith. 2012 Proc B

Appendage branching pattern	Cover plate arrangement	Presence	Absence
001510010?	00-100--	000000000000	
000500010?	200100--	0010010000	
002500010?	200100--	0?10010000	
00?5?0010?	200100?-0???	010110	
0015000101201000430100011111			
0015000101201010440111011111			
??050?????	201000440?	11011111	
01050?010-210000?	501??	010110	
00020001002101003-1110010110			
0002000100211001441121011111			
000201111-210010?-??	11011121		
?103?0?11?1001104-	0000010000		
1005002110100010--0?	00110?20		
1005002000101010540?	00110020		



Cambrian stalked echinoderms show
unexpected plasticity of arm construction
Zamora & Smith. 2012 Proc B

Discrete character data

Binary traits	0 1	Often describes the presence/absence of a trait

Discrete character data

Binary traits	0 1	Often describes the presence/absence of a trait
Multistate traits	0 1 2 3 4	Used to describe more complex traits and can capture greater variation between taxa

Discrete character data

Binary traits	0 1	Often describes the presence/absence of a trait
Multistate traits	0 1 2 3 4	Used to describe more complex traits and can capture greater variation between taxa
Missing characters	?	Used when the specimen is either too decayed to determine whether it has a certain character trait or not, or we are missing the relevant part of the body

Discrete character data

Binary traits	0 1	Often describes the presence/absence of a trait
Multistate traits	0 1 2 3 4	Used to describe more complex traits and can capture greater variation between taxa
Missing characters	?	Used when the specimen is either too decayed to determine whether it has a certain character trait or not, or we are missing the relevant part of the body
Non-applicable	-	Used when the trait is not associated with a taxon. They represent a type of nested coding where the presence of the trait is defined in a different trait

Discrete character data

Binary traits	0 1	Often describes the presence/absence of a trait
Multistate traits	0 1 2 3 4	Used to describe more complex traits and can capture greater variation between taxa
Missing characters	?	Used when the specimen is either too decayed to determine whether it has a certain character trait or not, or we are missing the relevant part of the body
Non-applicable	-	Used when the trait is not associated with a taxon. They represent a type of nested coding where the presence of the trait is defined in a different trait
Polymorphisms	0/1/2	Used when there are variations in a traits within species

Discrete character data

Binary traits	0 1	Often describes the presence/absence of a trait
Multistate traits	0 1 2 3 4	Used to describe more complex traits and can capture greater variation between taxa
Missing characters	?	Used when the specimen is either too decayed to determine whether it has a certain character trait or not, or we are missing the relevant part of the body
Non-applicable	-	Used when the trait is not associated with a taxon. They represent a type of nested coding where the presence of the trait is defined in a different trait
Polymorphisms	0/1/2	Used when there are variations in a traits within species
Uncertain	0/1/2	Used when it is not clear which character trait is present in the taxon

How do we model
morphological
evolution?

Mk Model

Assumes equal transition probabilities between states

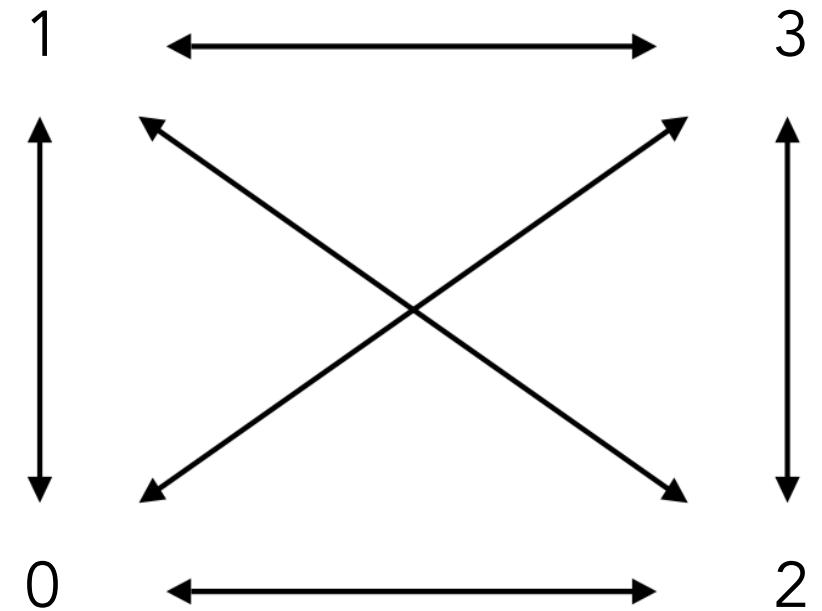


$$Q = \begin{bmatrix} -\mu_0 & \mu_{01} \\ \mu_{10} & -\mu_1 \end{bmatrix}$$

M_k Model

K can be any number of states

$$Q = \begin{pmatrix} -\mu_0 & \mu_{01} & \mu_{02} & \mu_{03} \\ \mu_{10} & -\mu_1 & \mu_{12} & \mu_{13} \\ \mu_{20} & \mu_{21} & -\mu_2 & \mu_{23} \\ \mu_{30} & \mu_{31} & \mu_{32} & -\mu_3 \end{pmatrix},$$



*4 state here as an example, can be any number from 2!

MkV model

What is one characteristic of morphological data that is extremely different to molecular though there are plenty.....

001510010?00-100--0000000000
000500010?200100--0010010000
002500010?200100--0?10010000
00?5?0010?200100?-0???010110
0015000101201000430100011111
0015000101201010440111011111
??050?????201000440?11011111
01050?010-210000?501??010110
00020001002101003-1110010110
0002000100211001441121011111
000201111-210010?-??11011121
?103?0?11?1001104-0000010000
1005002110100010--0?00110?20
1005002000101010540?00110020

MkV model

What is one characteristic of morphological data that is extremely different to molecular

though there are plenty.....

All varying characters

001510010?00-100--0000000000
000500010?200100--0010010000
002500010?200100--0?10010000
00?5?0010?200100?-0???010110
0015000101201000430100011111
0015000101201010440111011111
??050?????201000440?11011111
01050?010-210000?501??010110
00020001002101003-1110010110
0002000100211001441121011111
000201111-210010?-??11011121
?103?0?11?1001104-0000010000
1005002110100010--0?00110?20
1005002000101010540?00110020

MkV model



Corrects for ascertainment bias

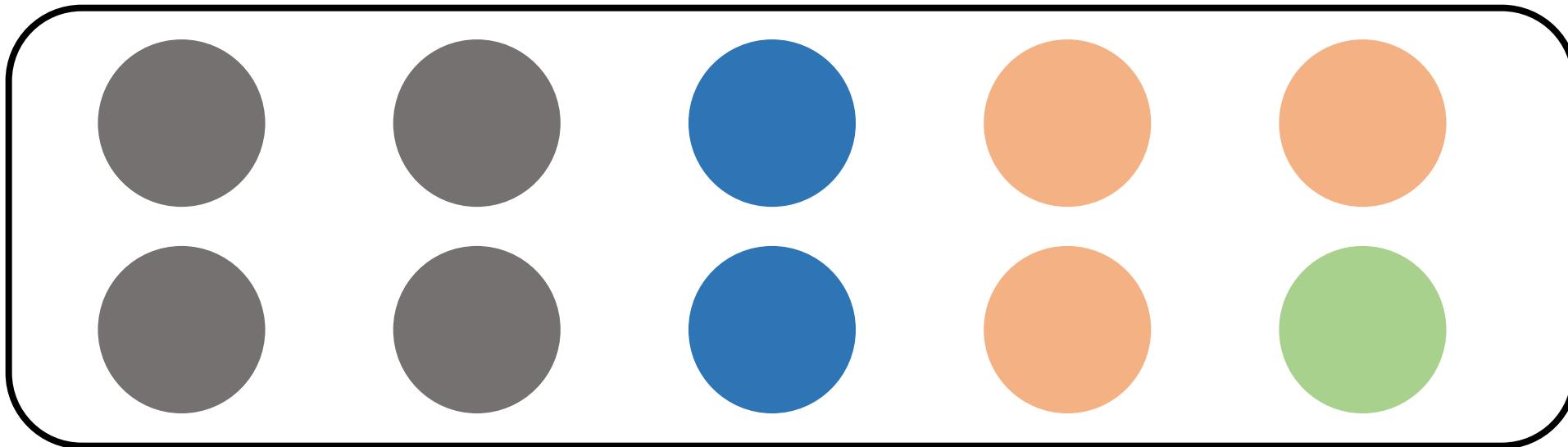
Failing to account for this can lead to **overestimations in branch lengths** and which can further lead to errors in topology!

Condition the likelihood
on there only being
varying site

$$\Pr(D | V) = \frac{\Pr(D, V)}{\Pr(V)}$$

MkV model

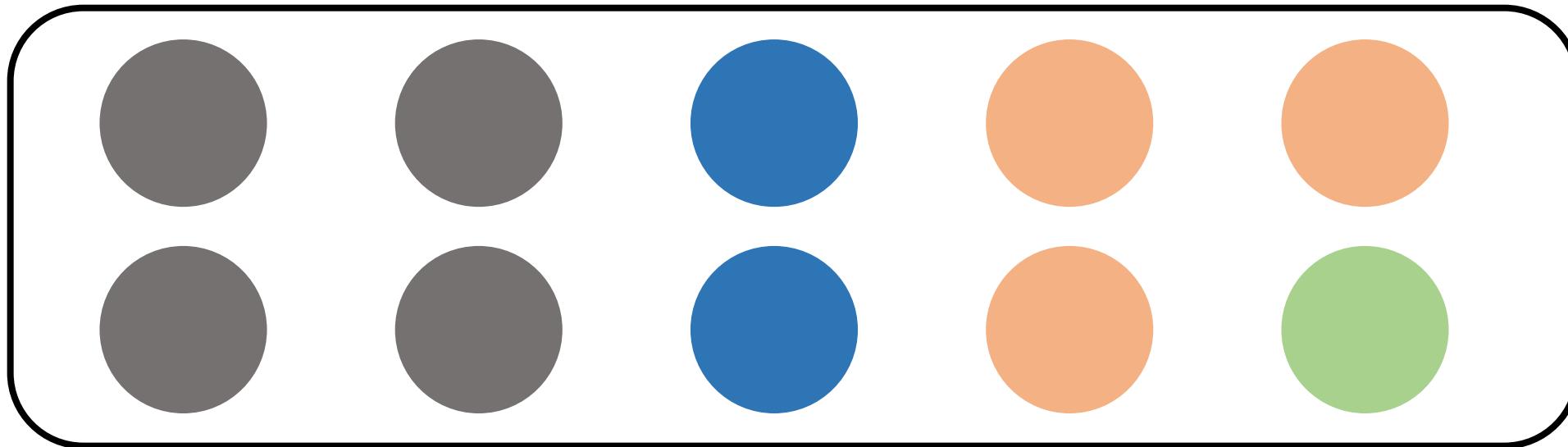
What is the probability for picking a certain colour ball?



$$1 = 0.4 + 0.2 + 0.3 + 0.1$$

MkV model

What is the probability for picking a certain colour ball?

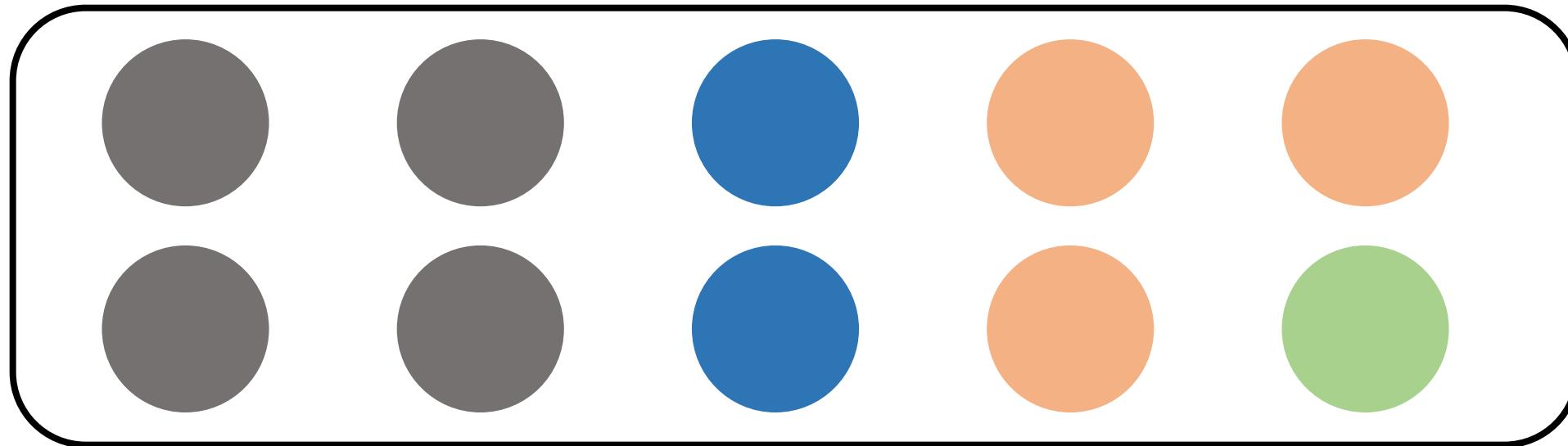


$$1 = 0.4 + 0.2 + 0.3 + 0.1$$

Probability of choosing an orange ball = 0.3

MkV model

What is the probability for picking a certain colour ball?



$$1 = 0.4 + 0.2 + 0.3 + 0.1$$

Probability of choosing an orange ball = 0.3

Probability of choosing an orange ball given it is not grey = $0.3/0.6 = 0.5$

Adapted from Paul Lewis PhyloSeminar

MkV model



Corrects for ascertainment bias

Failing to account for this can lead to overestimations in branch lengths and which can further lead to errors in topology!

Probability of the data
given character is
variable



$$\Pr(D | V) = \frac{\Pr(D, V)}{\Pr(V)}$$

MkV model



Corrects for ascertainment bias

Failing to account for this can lead to overestimations in branch lengths and which can further lead to errors in topology!

Probability of the data
given character is
variable

Probability of the
data and character
is variable

$$\Pr(D | V) = \frac{\Pr(D, V)}{\Pr(V)}$$

MkV model

Corrects for ascertainment bias

Failing to account for this can lead to overestimations in branch lengths and which can further lead to errors in topology!



Probability of the data
given character is
variable

Probability of the
data and character
is variable

$$\Pr(D | V) = \frac{\Pr(D, V)}{\Pr(V)}$$

Probability that
character is variable

MkV model



Corrects for ascertainment bias

Failing to account for this can lead to overestimations in branch lengths and which can further lead to errors in topology!

$$\Pr(D | V) = \frac{\Pr(D, V)}{\Pr(V)}$$

$$\Pr(V)$$

Probability that
character is variable



$$1 - \Pr(\text{character is constant})$$

This value, $\Pr(C)$ can be obtained using a **dummy character** having the same state for all internal nodes

MkV model

In RevBayes this is done internally and all non varying characters will be removed before the inference

In Beast you will see the dummy characters in the xml file produced from beauti

```
<data
id="Zamora_Smith_part"
spec="Alignment"
dataType="standard">
<sequence id="seq_Kinzerocystis" spec="Sequence" taxon="Kinzerocystis" totalcount="6"
value="001510010?00-100--00000000012345"/>
<sequence id="seq_Gogia" spec="Sequence" taxon="Gogia" totalcount="6"
value="000500010?200100--00100100012345"/>
<sequence id="seq_Sinoeocrinus" spec="Sequence" taxon="Sinoeocrinus" totalcount="6"
value="002500010?200100--0?100100012345"/>
<sequence id="seq_Akadocrinus" spec="Sequence" taxon="Akadocrinus" totalcount="6"
value="00?5?0010?200100?-0???01011012345"/>
<sequence id="seq_Ridersia" spec="Sequence" taxon="Ridersia" totalcount="6"
value="001500010120100043010001111012345"/>
```

MkV model

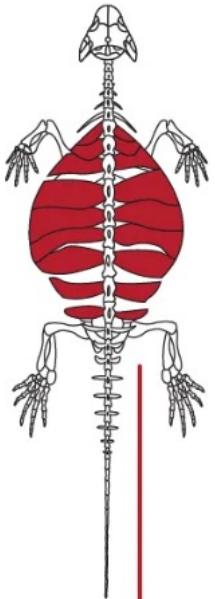


	True Branch Length	Mk	Mkv
Percent correct	-	74.0	99.8
Branch A	0.2	241,750 (±349,100)	0.206 (±0.060)
Branch B	0.05	0.43210 (±0.13756)	0.050 (± 0.018)
Branch X	0.05	54.646 (±1,725.3)	0.052 (± 0.023)
Branch C	0.2	143,950 (±228,910)	0.206 (± 0.059)
Branch D	0.05	0.022 (± 0.054)	0.051 (±0.019)

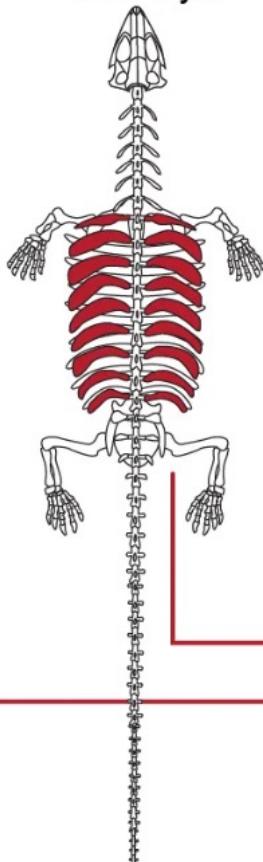
Among-character rate variation

Turtle shell evolution

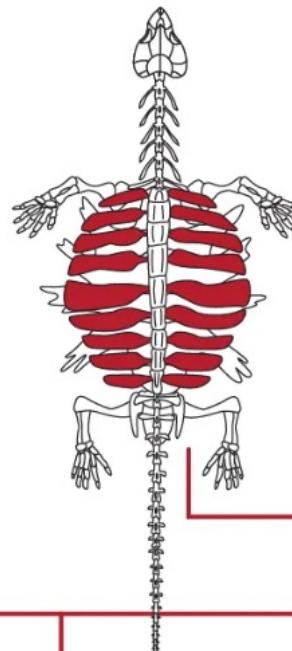
Eunotosaurus
~260 mya



Pappochelys
~240 mya



Odontochelys
~220 mya



Proganochelys
~210 mya



Image [source](#)

Among-character rate variation

	T1	T2
Taxa A	0	0
Taxa B	0	1
Taxa C	1	2

The transition rate
will impact branch
lengths

Slow rate of evolution



Fast rate of evolution

Relative to each other!

Among-character rate variation

What do we do?

	T1	T2
Taxa A	0	0
Taxa B	0	1
Taxa C	1	2

Allow these traits to evolve at different rates:

- Specify which traits evolve fast
- Use a gamma model to account for rate heterogeneity

Among-character rate variation

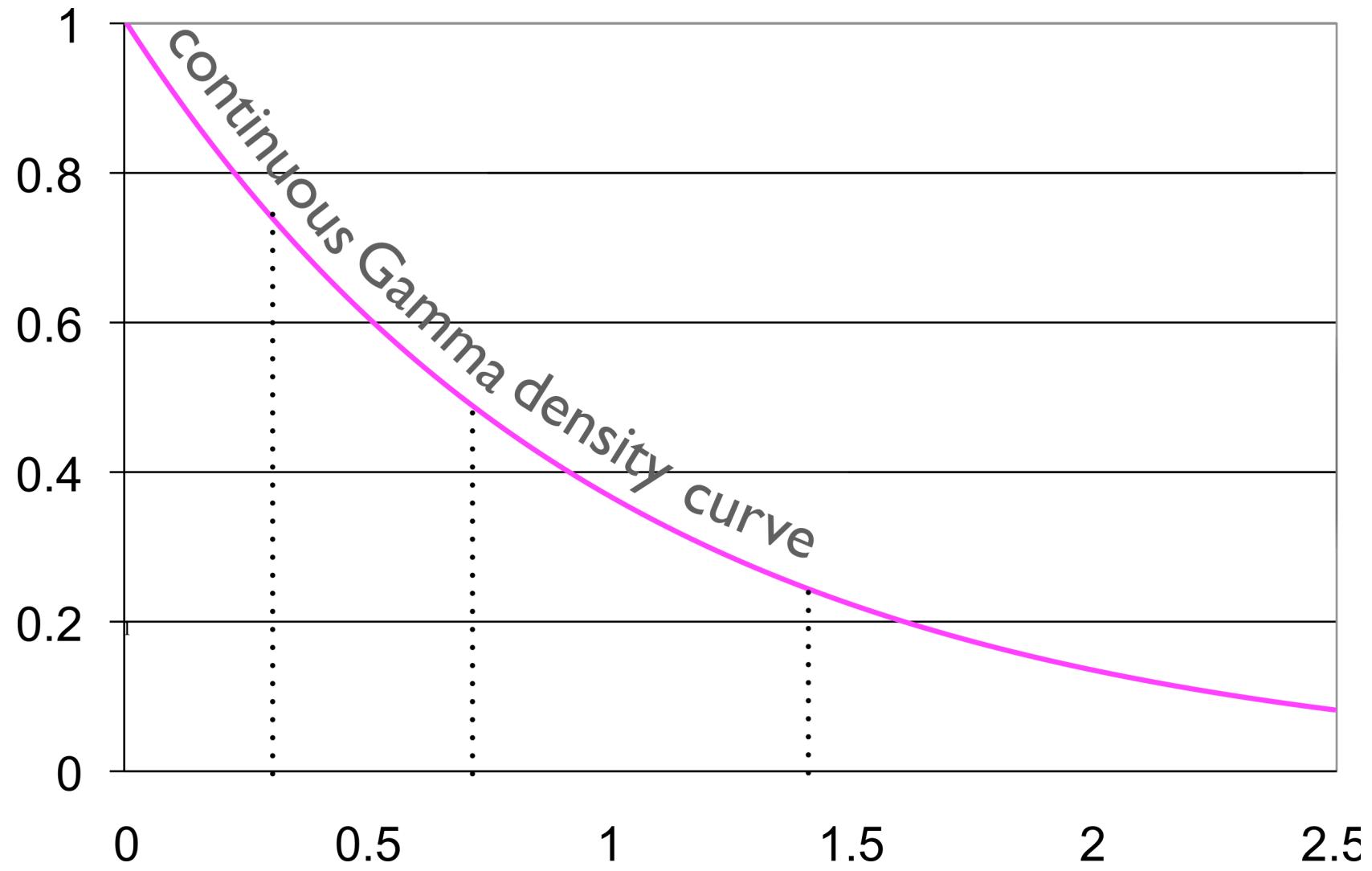
What do we do?

	T1	T2
Taxa A	0	0
Taxa B	0	1
Taxa C	1	2

Allow these traits to evolve at different rates:

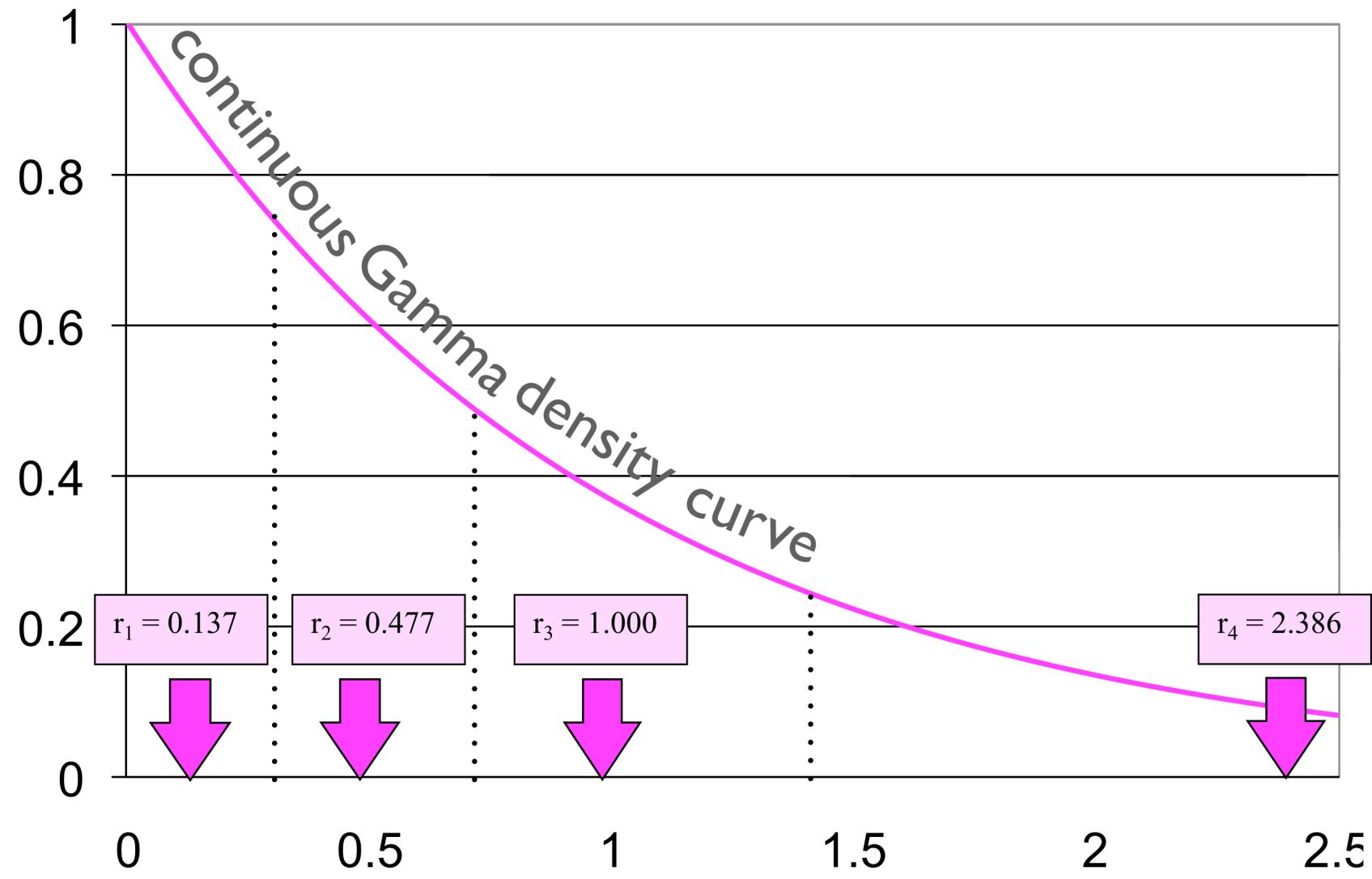
- Specify which traits evolve fast
- Use a gamma model to account for rate heterogeneity

Mk(V) + Gamma



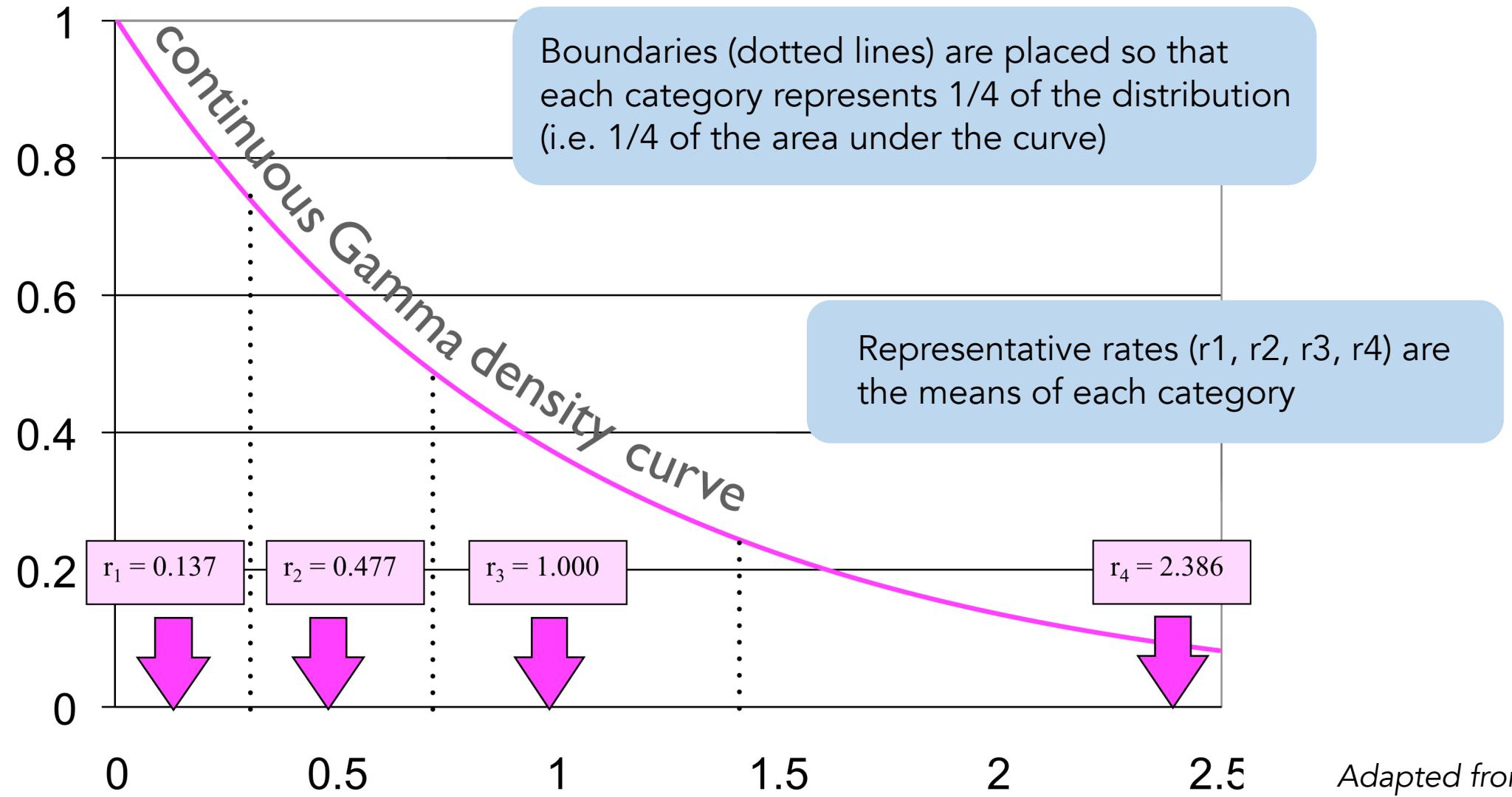
Adapted from Paul
Lewis PhyloSeminar

Mk(V) + Gamma



Adapted from Paul
Lewis PhyloSeminar

Mk(V) + Gamma



Mk(V) + Gamma

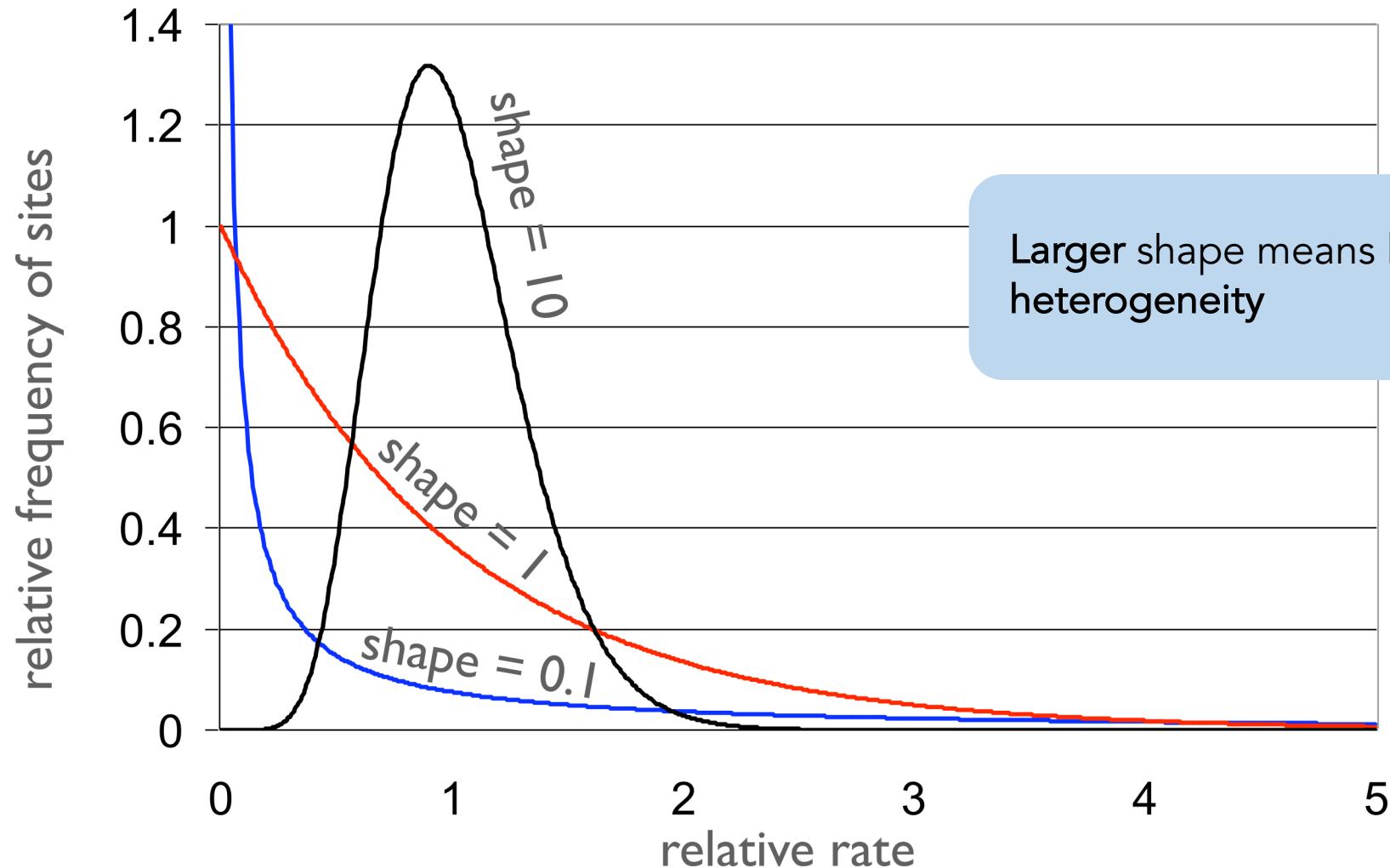
What do we do?

	T1	T2	
Taxa A	0	0	
Taxa B	0	1	
Taxa C	1	2	Faster R (R4)
Slower R (R1,2)			

Allow each trait to evolve according to the rates drawn from the gamma distribution

One rate will fit the best and be the most influential for the likelihood calculation

Mk(V) + Gamma



Adapted from Paul
Lewis PhyloSeminar

Partitioning Data

Grouping together parts of the alignment that have similar characteristics and or may have **evolved together** due to evolutionary pressures

The **defaults** in many phylogenetic software is to group by maximum observed state size

$$Q = \begin{pmatrix} -\mu_0 & \mu_{01} & \mu_{02} & \mu_{03} \\ \mu_{10} & -\mu_1 & \mu_{12} & \mu_{13} \\ \mu_{20} & \mu_{21} & -\mu_2 & \mu_{23} \\ \mu_{30} & \mu_{31} & \mu_{32} & -\mu_3 \end{pmatrix},$$

$$\begin{bmatrix} -\mu_0 & \mu_{01} \\ \mu_{10} & -\mu_1 \end{bmatrix}$$

$$\begin{bmatrix} -\mu_0 & \mu_{01} & \mu_{02} \\ \mu_{10} & -\mu_1 & \mu_{12} \\ \mu_{20} & \mu_{21} & -\mu_2 \end{bmatrix}$$

Partitioning Data

When should we partition our data?

Partitioning Data

When should we partition our data?

If we have presence (1) absence (0) traits partitioning will always be a logical approach: what would transitioning to state 2 in this scenario even mean?

Partitioning Data

When should we partition our data?

If we have presence (1) absence (0) traits partitioning will always be a logical approach: what would transitioning to state 2 in this scenario even mean?

We should be cautious for traits describing a trait – just because we do not observe a state 2 can we be absolutely certain there never was one?

Justifying partitioning schemes is very important as they have a major impact on inference results

Other morphological models

Ordered Characters

Ordered characters can be placed in an order so that transitions only occur between adjacent states.



For example, “intermediate” species that are somewhere in between limbed and limbless – for example, the “mermaid skinks” (*Sirenoscincus*) from Madagascar, so called because they lack hind limbs. An ordered model might only allow transitions between limbless and intermediate, and intermediate and limbed; it would be impossible under such a model to go directly from limbed to limbless without first becoming intermediate.

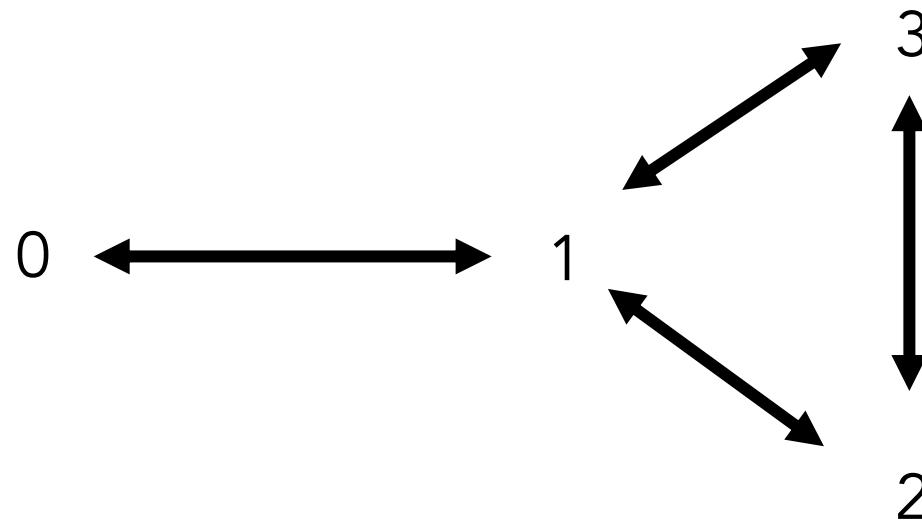
For unordered characters, any state can change into any other state.

Ordered Characters

All characters ordered:



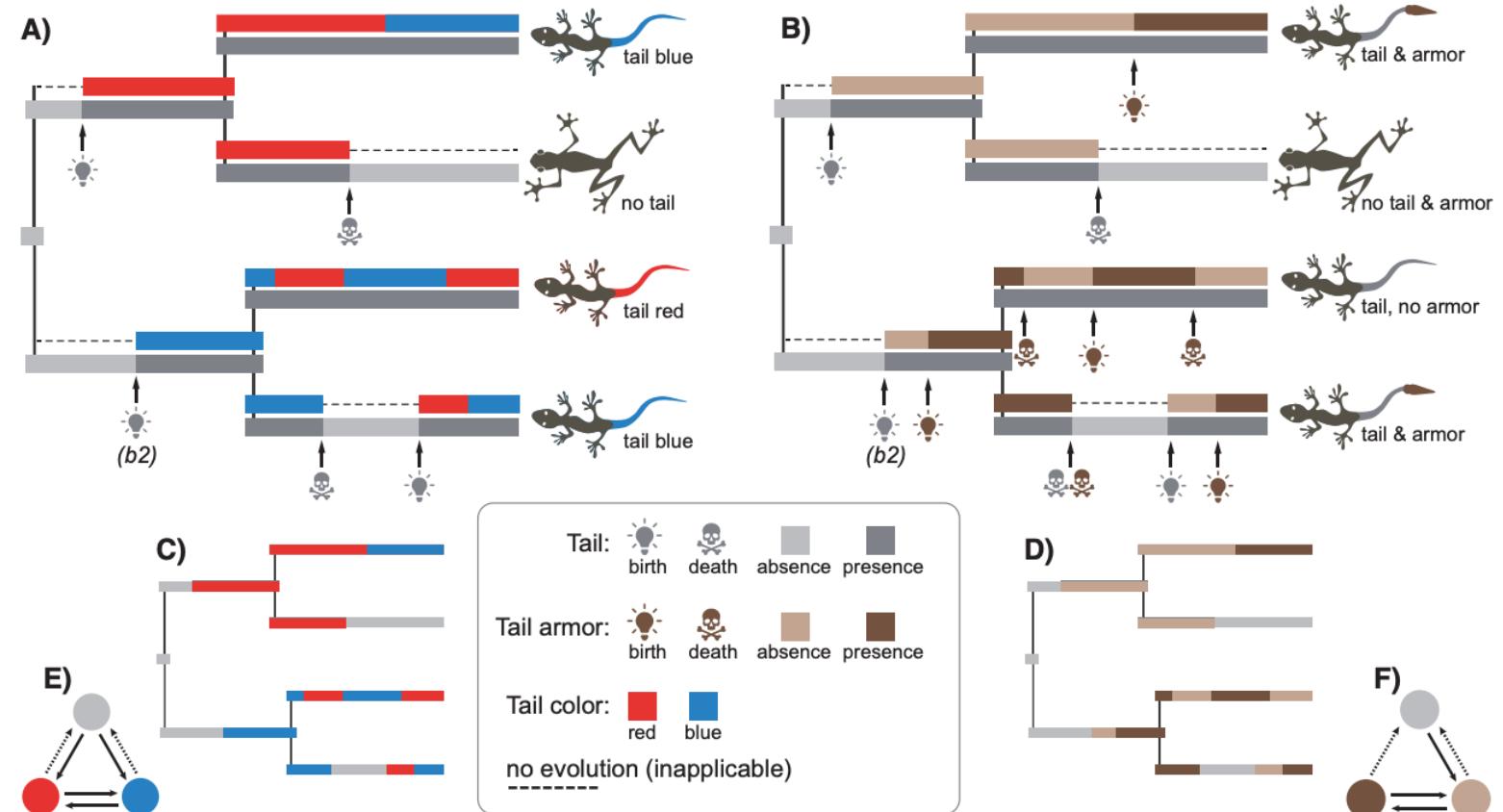
Specific characters
ordered:



Embedded dependency model

Markov models for phylogenetic inference with anatomically dependent (inapplicable) morphological characters

Non-applicable characters only considered when they are present (1)

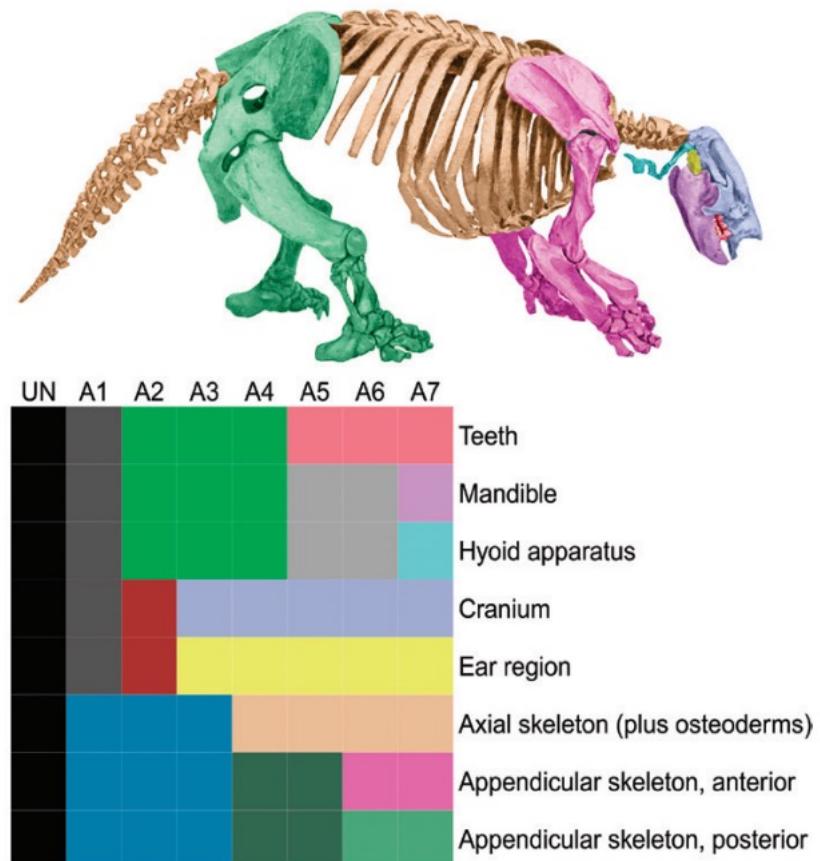


Alternative Partitioning schemes

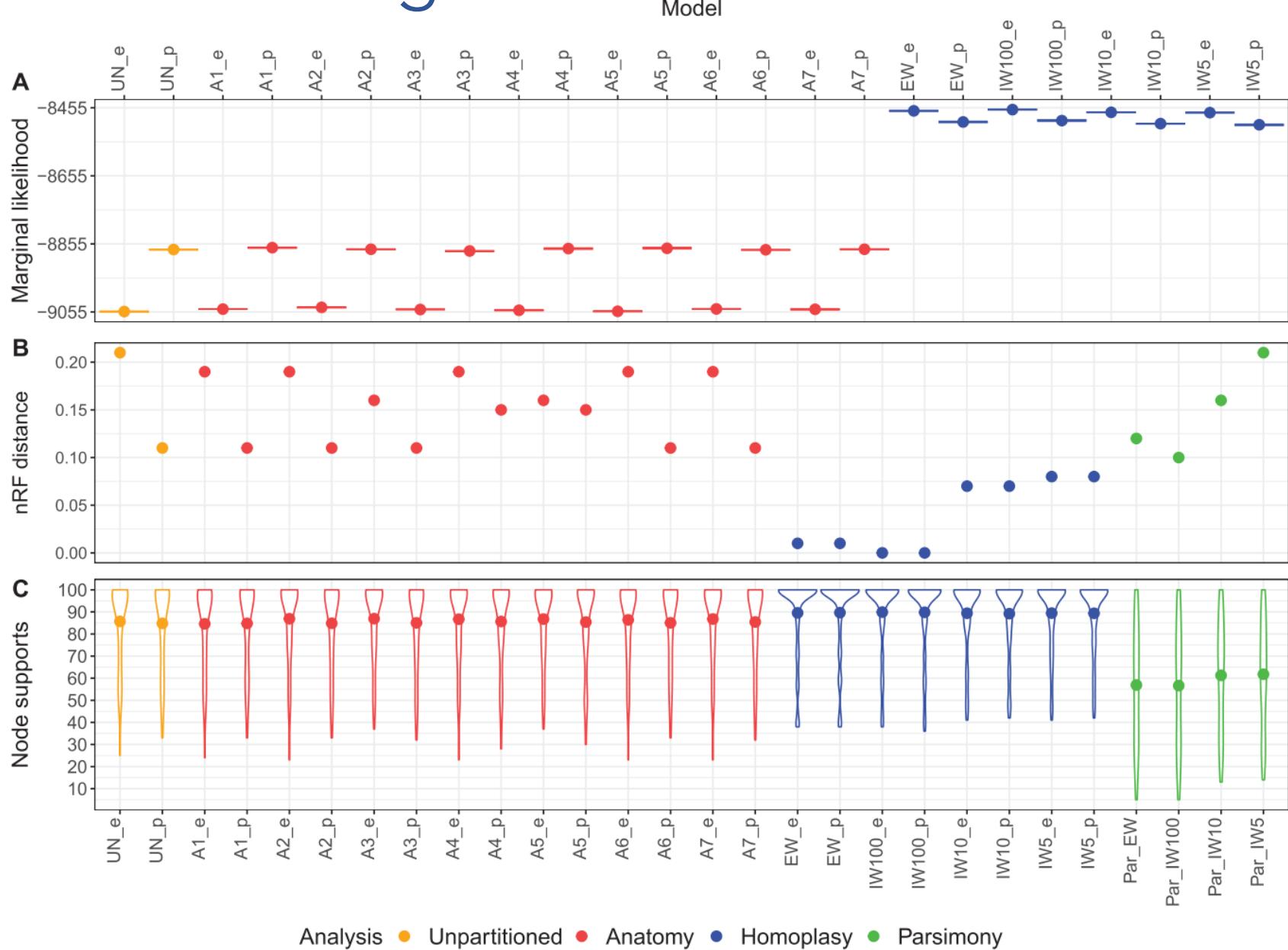
Reassessing the phylogeny and divergence times of sloths (Mammalia: Pilosa: Folivora)

Characters can be groups based on anatomical region

Other criteria such as the degree of homoplasy present in a character was explored in this study – and found to be a better fit using Bayes factors



Alternative Partitioning schemes



Challenges with morphological data

Generalising assumptions across different traits is often
not possible

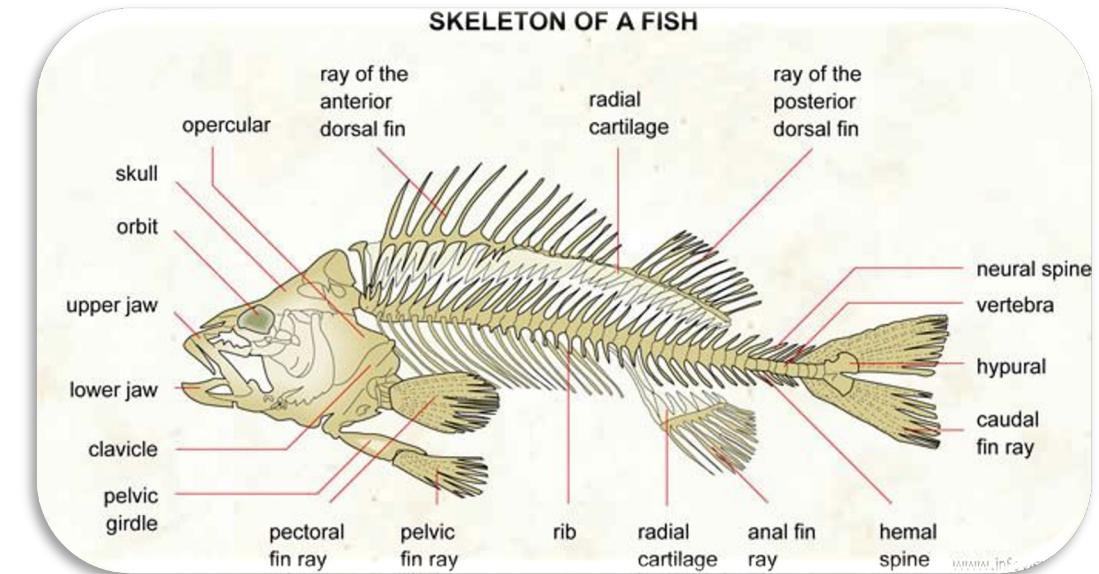
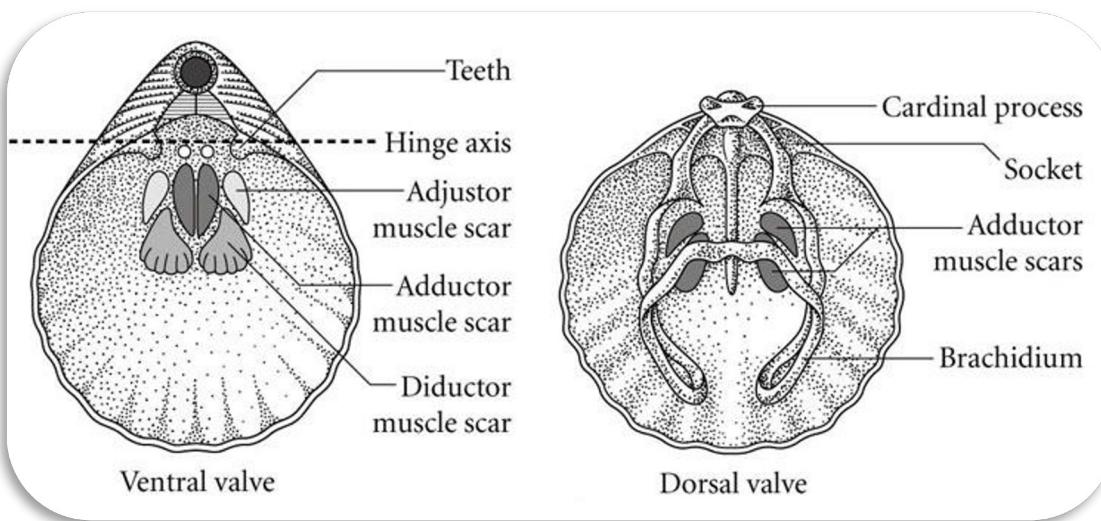
Modelling special characters in matrices

```
001510010?00-100--0000000000
000500010?200100--0010010000
002500010?200100--0?10010000
00?5?0010?200100?-0??010110
0015000101201000430100011111
0015000101201010440111011111
??050?????201000440?11011111
```

Challenges with morphological data

Morphological matrices are often quite small:

- Collection is very time consuming
- Number of characters available can be very small depending on the group

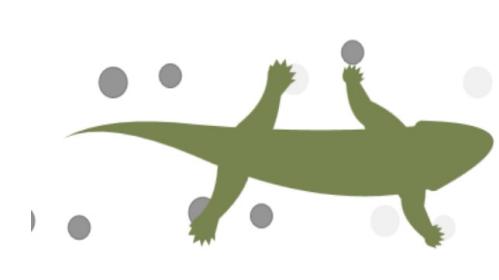


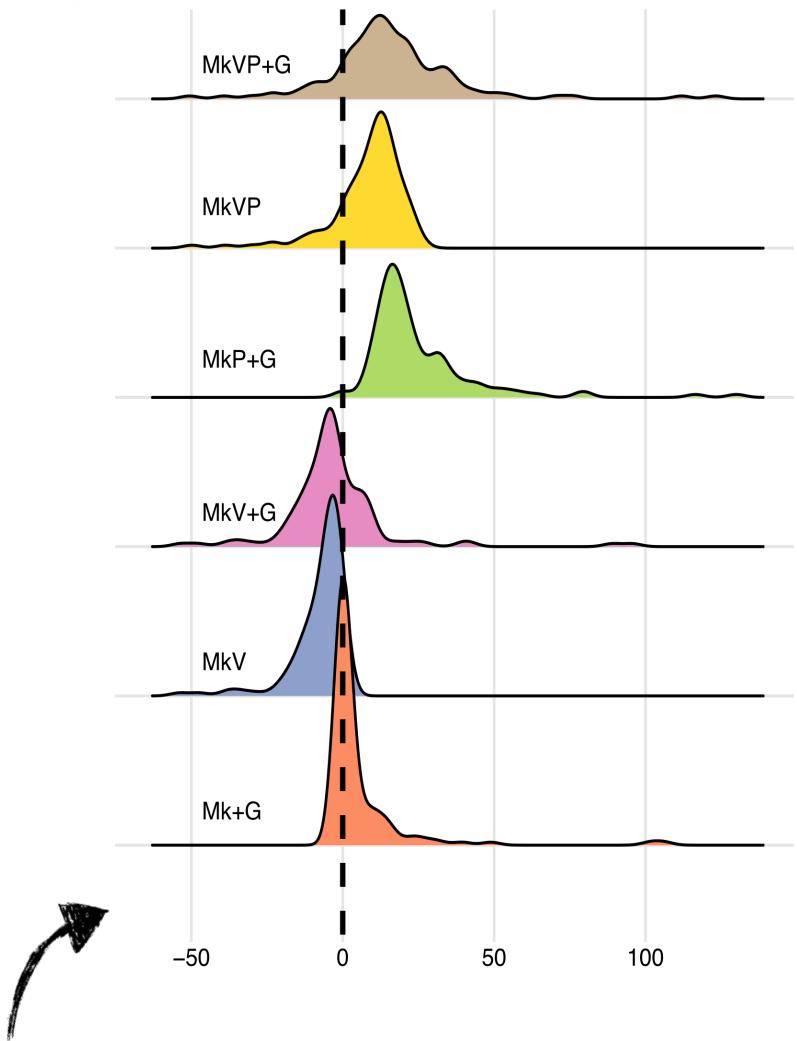
Model impact on key parameter estimates

Example of 114 empirical tetrapod matrices

Looked at the impact on:

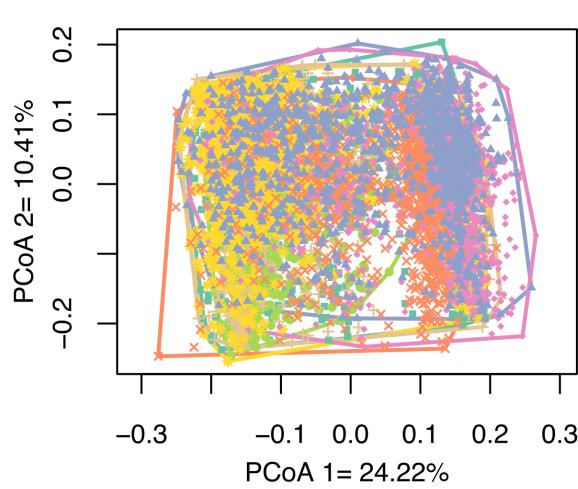
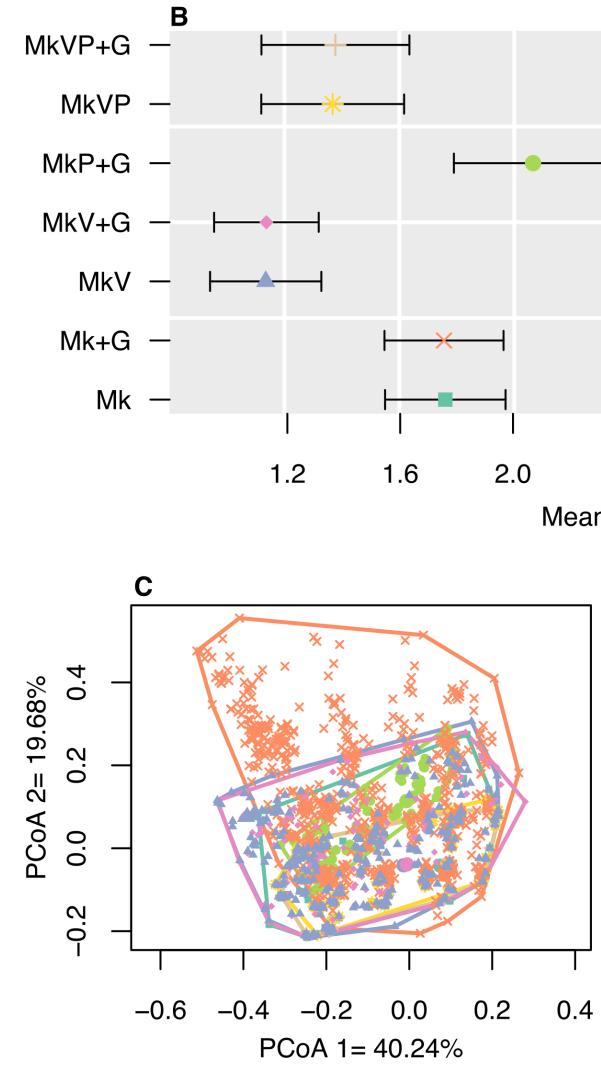
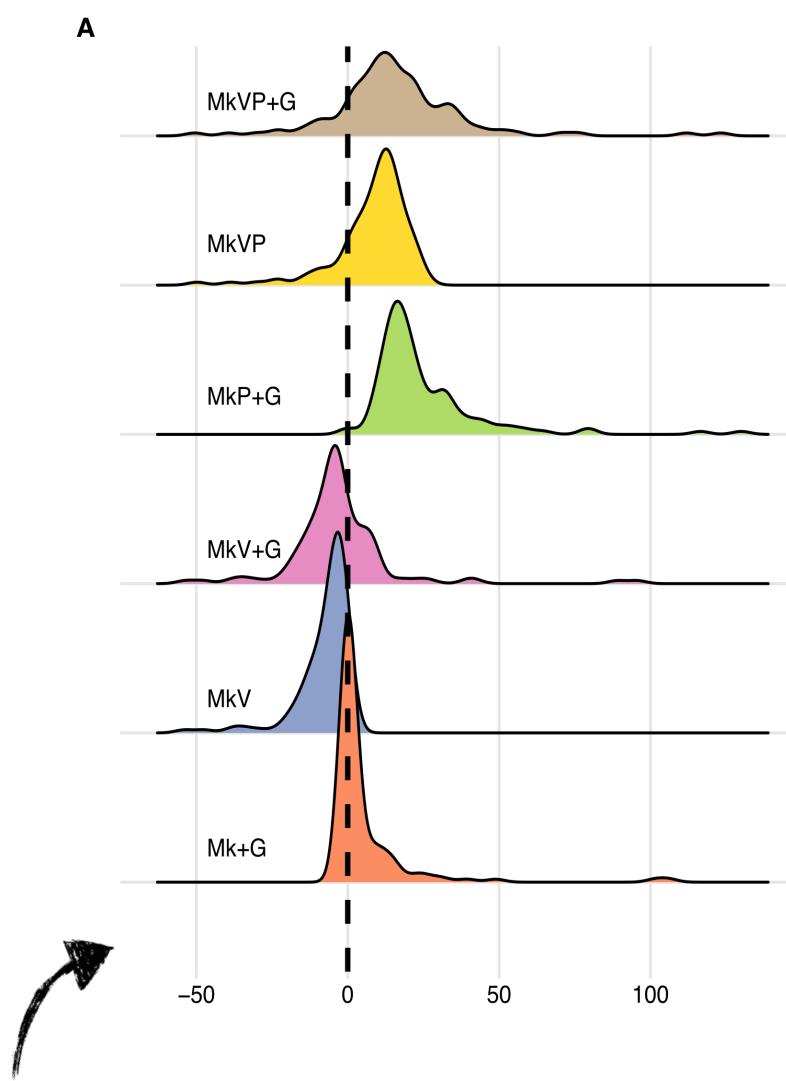
- branch lengths (**evolutionary distances**)
- Tree topology (**species relationships**)



A

Percentage difference in
tree length relative to Mk
model

- Mk
- Mk+G
- MkV
- MkV+G
- MkP+G
- MkVP
- MkVP+G



Tree length of two different data sets

Rf distances of two data sets

■ Mk ✕ Mk+G ▲ MkV ♦ MkV+G
● MkP+G * MkVP + MkVP+G

How do we choose a
model?

Model selection

Bayes factors are commonly used to determine the **relative fit** between model.

It relies on comparing the marginal likelihoods approximated from different models.

The ML measures the average fit of a model to our data.

We use MCMC to avoid calculating this number as it is computationally expensive and often not directly possible.

Model selection

$$P(\text{model} \mid \text{data}) = \frac{P(\text{data} \mid \text{model}) P(\text{model})}{P(\text{data} \mid \text{model})}$$

Posterior

Likelihood

Priors

Marginal likelihood

The diagram illustrates the formula for Bayesian model selection. The posterior probability $P(\text{model} \mid \text{data})$ is calculated by dividing the likelihood $P(\text{data} \mid \text{model})$ by the marginal likelihood $P(\text{data} \mid \text{model})$. The prior probability $P(\text{model})$ is also included in the numerator. Orange arrows connect the labels 'Posterior', 'Likelihood', 'Priors', and 'Marginal likelihood' to their corresponding components in the equation.

Marginal likelihood

Marginal probability of the data (denominator in Bayes' rule) is the expected value of the likelihood with respect to the prior distribution.

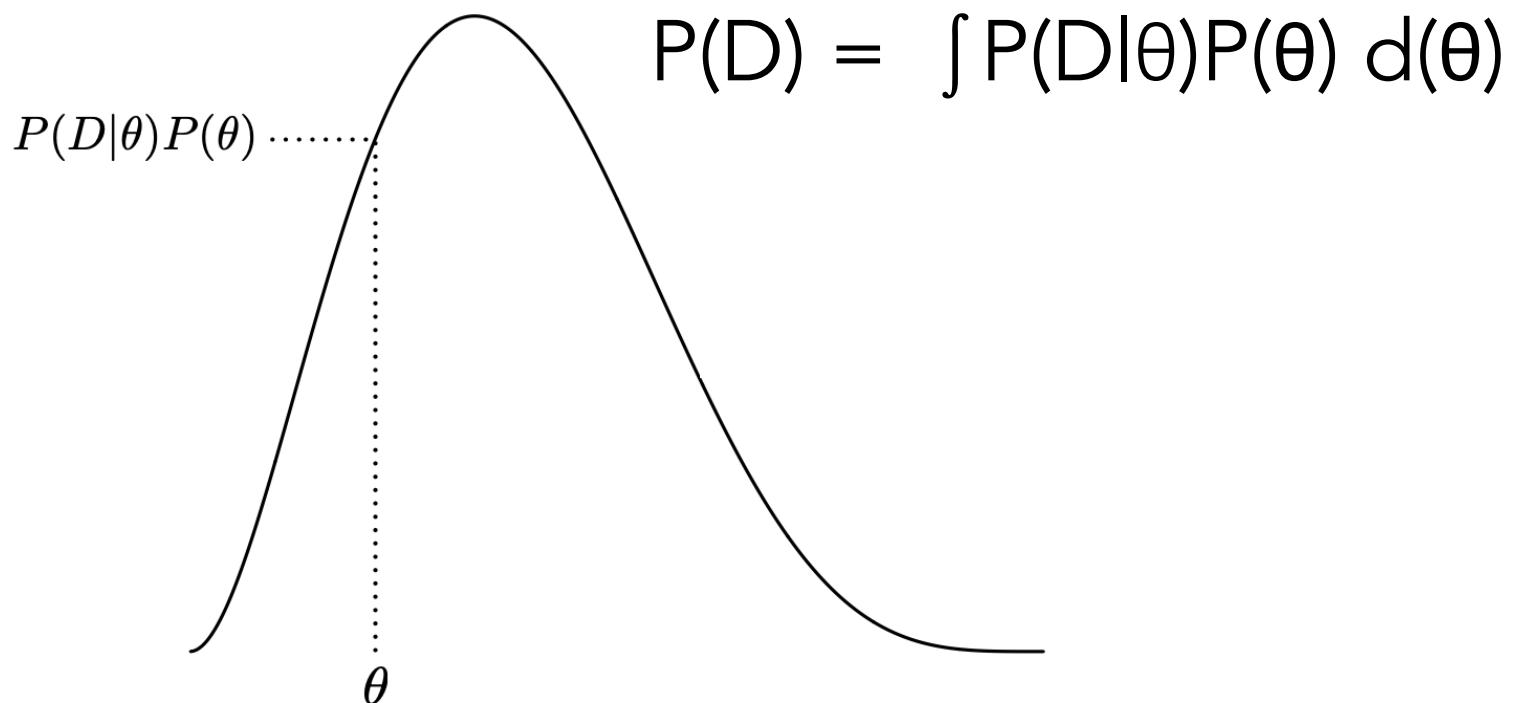
If likelihood measures model fit, then the marginal likelihood measures the average fit of the model to the data over all parameter values.

What is the expected value?

Marginal likelihood

$P(\text{data} | \text{model})$

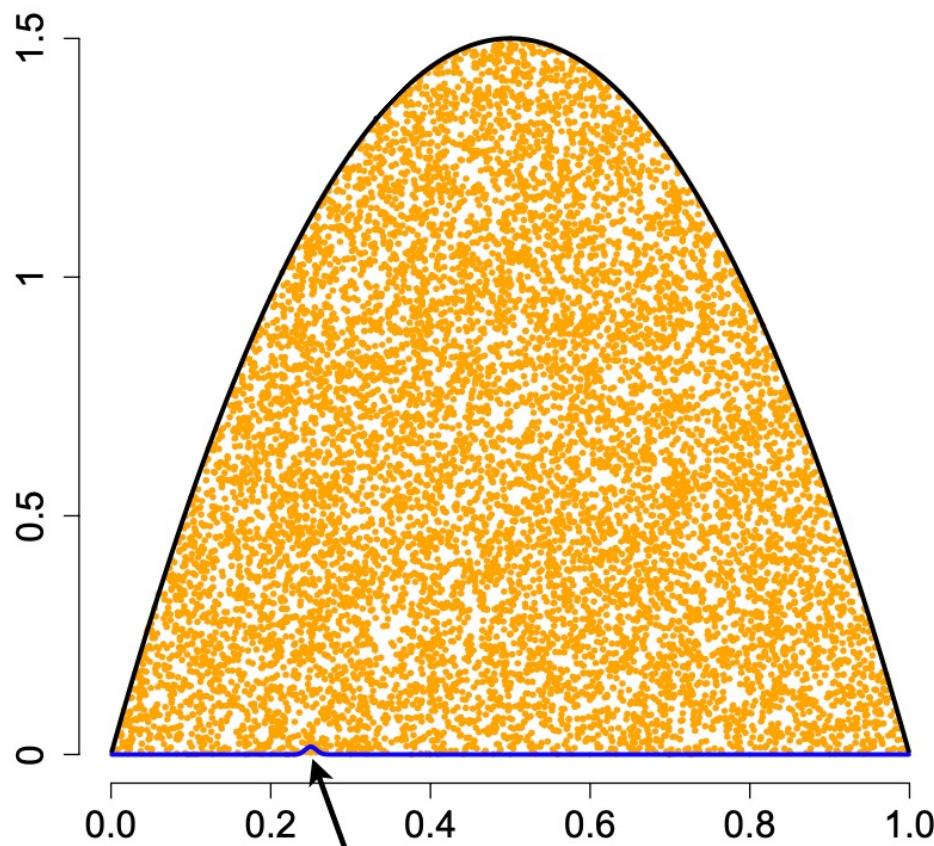
The marginal likelihood is used to evaluate the overall fit of the model to the data, integrating over all parameter values.



Marginal likelihood

$$P(\text{data} | \text{model})$$

Very small, single number between the posterior distribution and the prior



Approximating the marginal likelihood

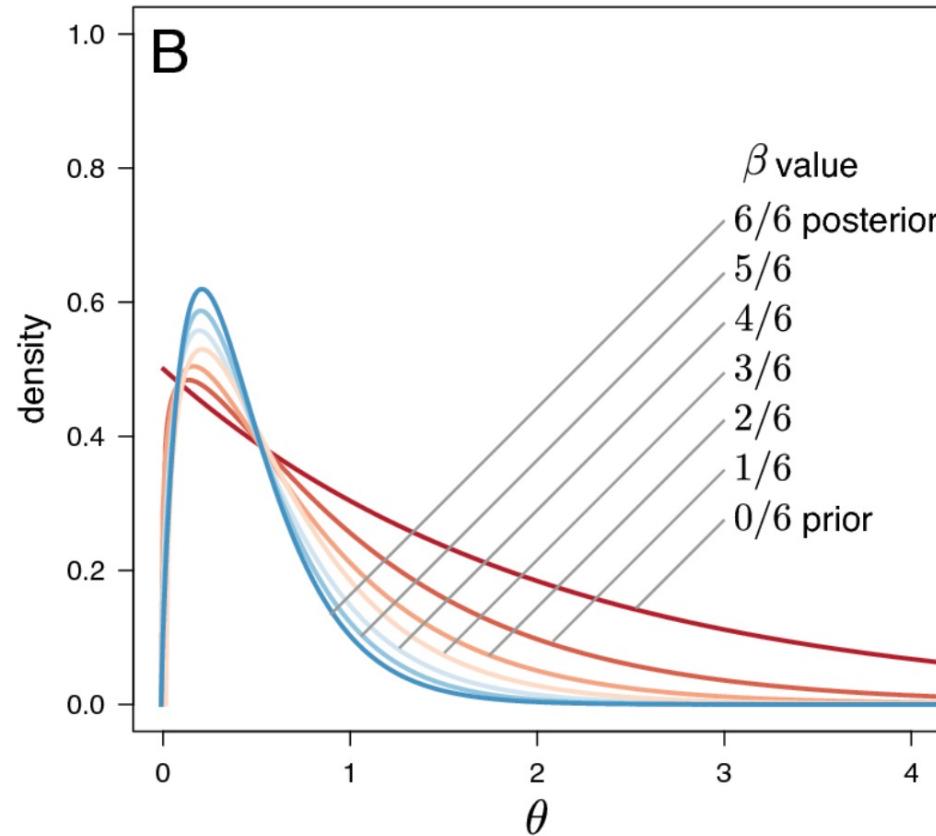
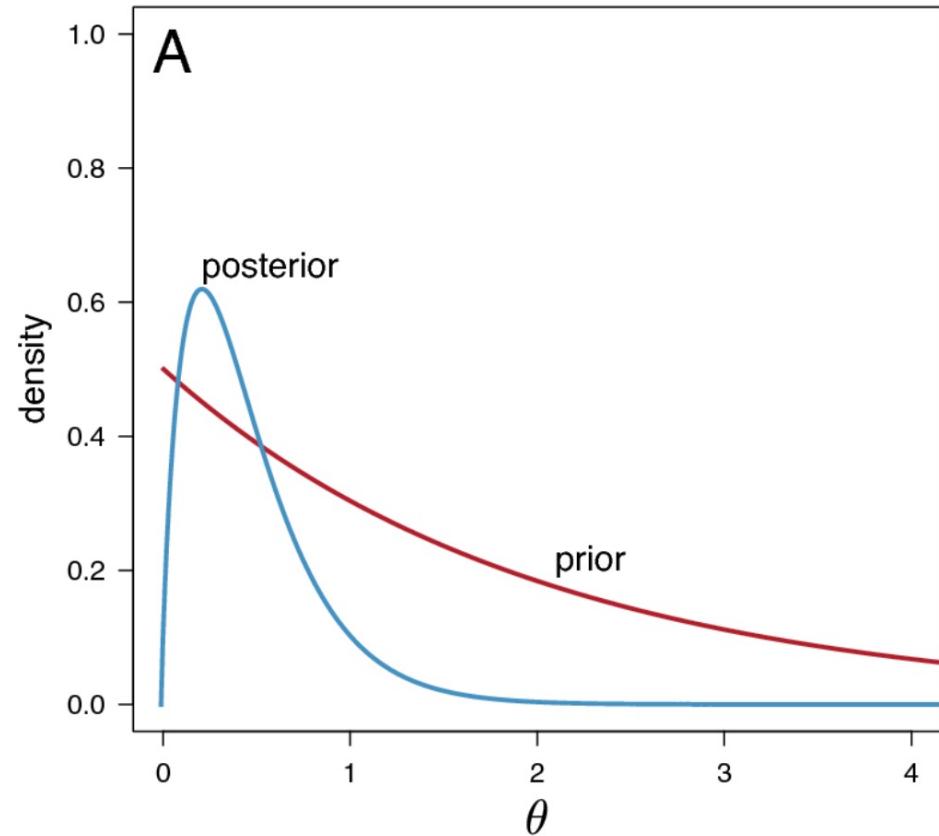
There are two common algorithms to do this:

- Stepping stone
- Path sampling

Both of these approaches are computationally expensive

Stepping-stone algorithms are like a series of MCMC simulations that iteratively sample from a specified number of distributions that are discrete steps between the posterior and the prior probability distributions.

Stepping stone algorithm



$\beta = 1$ this is the posterior distribution
 $\beta = 0$ this is the prior distribution

Bayes factors

$$B_{01} = \frac{P(D | M_0)}{P(D | M_1)} = \frac{\text{Marginal likelihood for model } M_0}{\text{Marginal likelihood for model } M_1}$$

Bayes factors

$$B_{01} = \frac{P(D | M_0)}{P(D | M_1)} = \frac{\text{Marginal likelihood for model } M_0}{\text{Marginal likelihood for model } M_1}$$

Marginal likelihoods are often on the log scale so the Bayes factor can be calculated as:

$$\log B_{01} = \log P(D | M_0) - \log P(D | M_1)$$

Interpreting Bayes factors

Strength of evidence	$BF(M_0, M_1)$	$\log(BF(M_0, M_1))$
Negative (supports M_1)	<1	<0
Barely worth mentioning	1 to 3.2	0 to 1.16
Substantial	3.2 to 10	1.16 to 2.3
Strong	10 to 100	2.3 to 4.6
Decisive	>100	>4.6

Exercise:
Use stepping stone to
determine the fit
between models