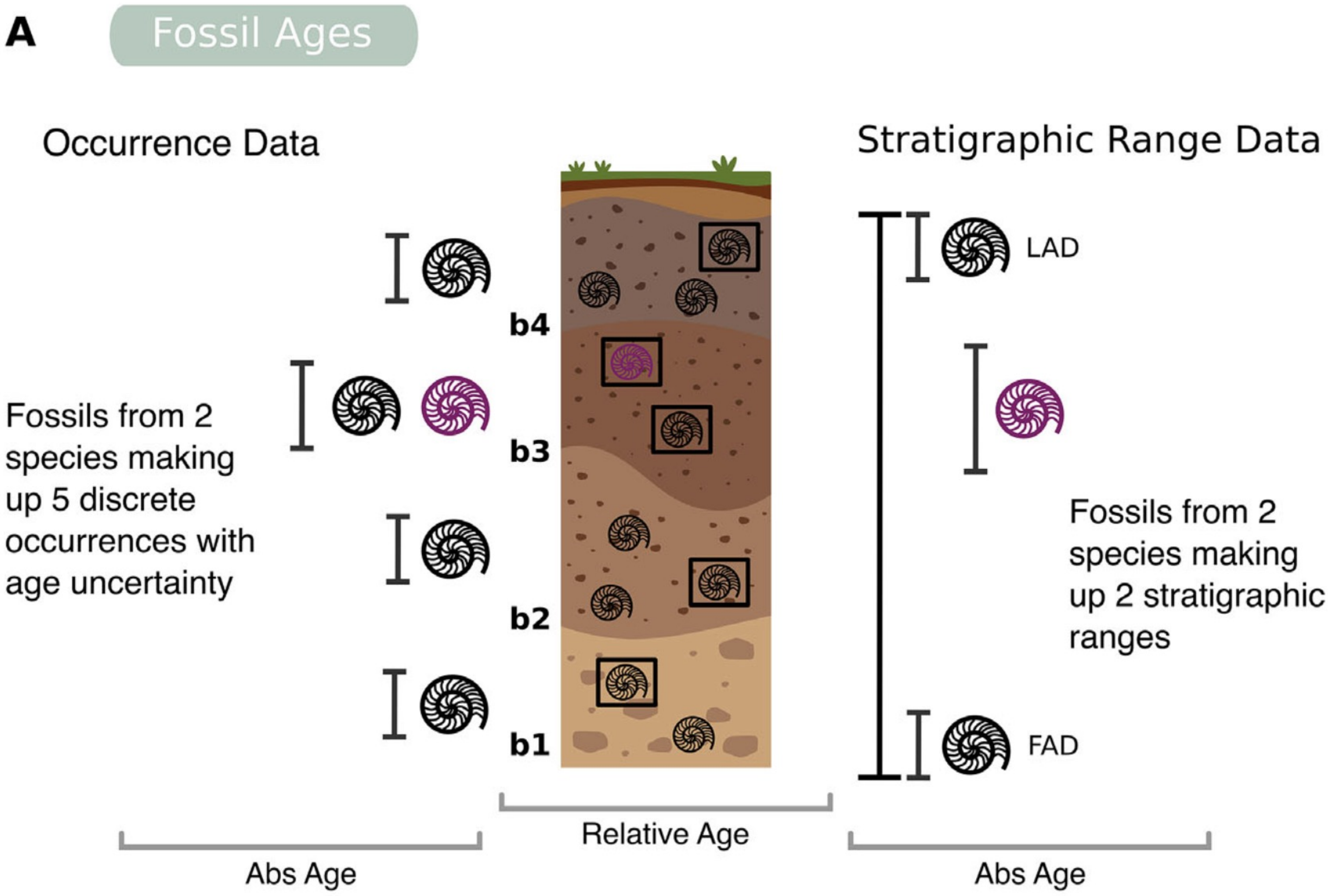


Mulvey, L. P. A., M. C. Nikolic, B. J. Allen, T. A. Heath, and R. C. M. Warnock (2025). From fossils to phylogenies: exploring the integration of paleontological data into Bayesian phylogenetic inference. Paleobiology 51, 214–236. https://doi.org/10.1017/pab.2024.47

Class focus area: Bayesian inference

EEB603: Brian O'Meara

All quotes and images from the above paper unless otherwise noted

Flip a coin three times:

$\boldsymbol{q}$ = probability of landing heads

What is a good estimate of q for this coin?

Likelihood approach:

$\boldsymbol{q}$ = probability of landing heads

- First choose a model
- In this case, we think that the coin has an innate probability of landing heads, regardless of whether the previous toss landed heads, whether it’s sunny outside, or whether it was tossed with the right or left hand.
- The simple model we will use is the binomial distribution.

Likelihood approach:

q = probability of landing heads

- First choose a model (binomial distribution)
- We now have our model and our data (2 heads out of 3 tosses).
- Definition:

Likelihood of the hypothesis given the data
is proportional to the
Probability of the data given the hypothesis

Likelihood approach:

q = probability of landing heads

Likelihood of the hypothesis given the data
is proportional to the
Probability of the data given the hypothesis

In our case, the hypothesis is just a given value for q, and the data are getting 2 heads out of 3 tosses.

Likelihood of the hypothesis given the data is proportional to the *Probability* of the data given the hypothesis

In our case, the hypothesis is just a given value for ***q***, and the data are getting 2 heads out of 3 tosses.

For example, if we set *q* to 0.4, the probability of seeing 2 heads out of 3 tosses is:

$$P(2\ H,\ 1\ T \mid q=0.4) = \frac{3!}{2!(3-2)!}(0.4)^2(1-0.4)^{3-2} = \mathbf{0.288}$$

Likelihood of the hypothesis given the data
is proportional to the
Probability of the data given the hypothesis

In our case, the hypothesis is just a given value for ***q***, and the data are getting 2 heads out of 3 tosses.

For example, if we set *q* to 0.4, the probability of seeing 2 heads out of 3 tosses is:

$$P(2\text{ H}, 1\text{ T} \mid q=0.4) = \frac{3!}{2!(3-2)!}(0.4)^2(1-0.4)^{3-2} = \mathbf{0.288}$$

We can then say the **likelihood** of the hypothesis $q=0.4$ given the data of 2 heads out of 3 tosses is **0.288.**

Likelihood of the hypothesis given the data is proportional to the *Probability* of the data given the hypothesis

In our case, the hypothesis is just a given value for ***q***, and the data are getting 2 heads out of 3 tosses.

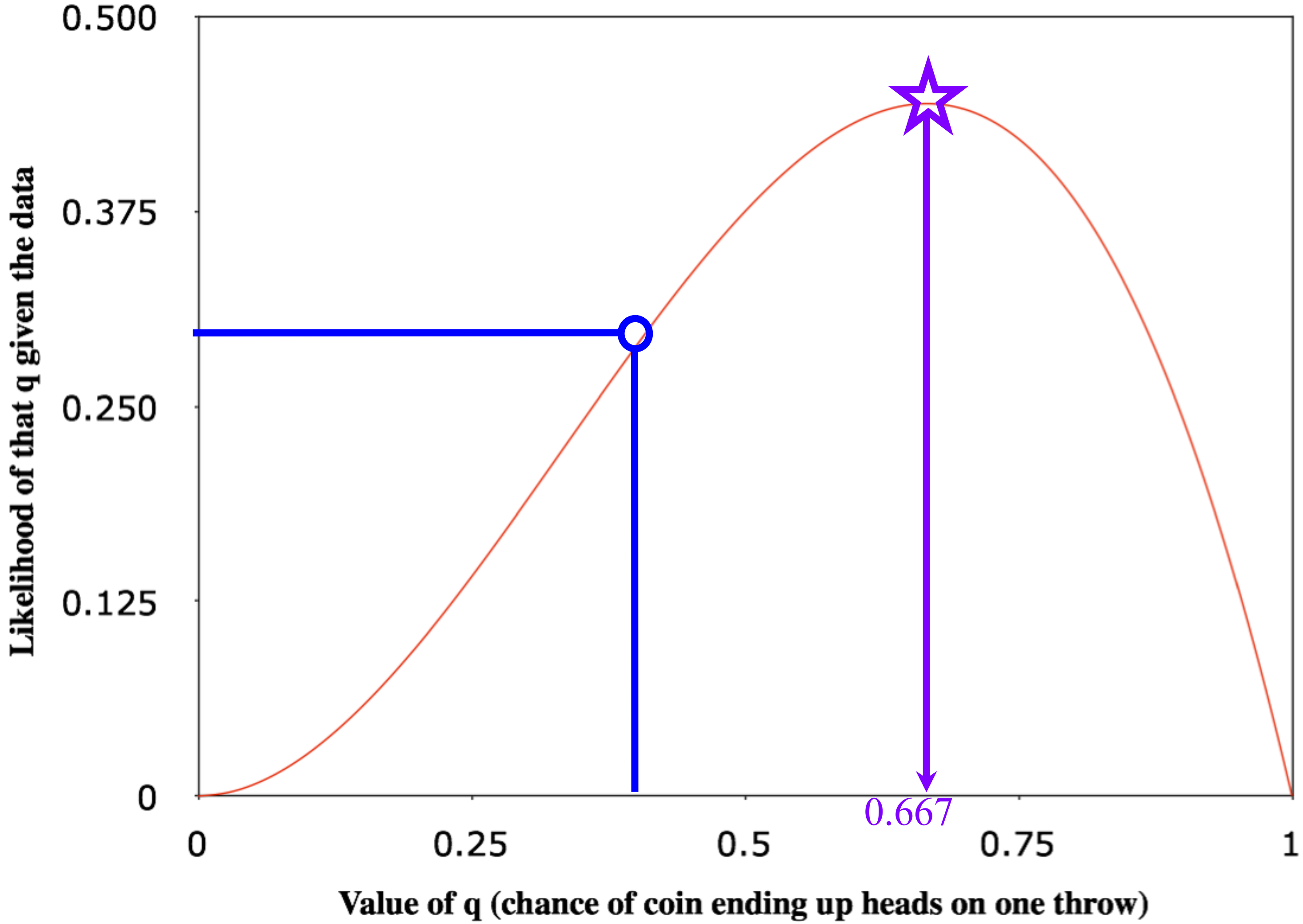
For example, if we set *q* to 0.4, the probability of seeing 2 heads out of 3 tosses is:

$$P(2\ H,\ 1\ T \mid q=0.4) = \frac{3!}{2!(3-2)!}(0.4)^2(1-0.4)^{3-2} = \mathbf{0.288}$$

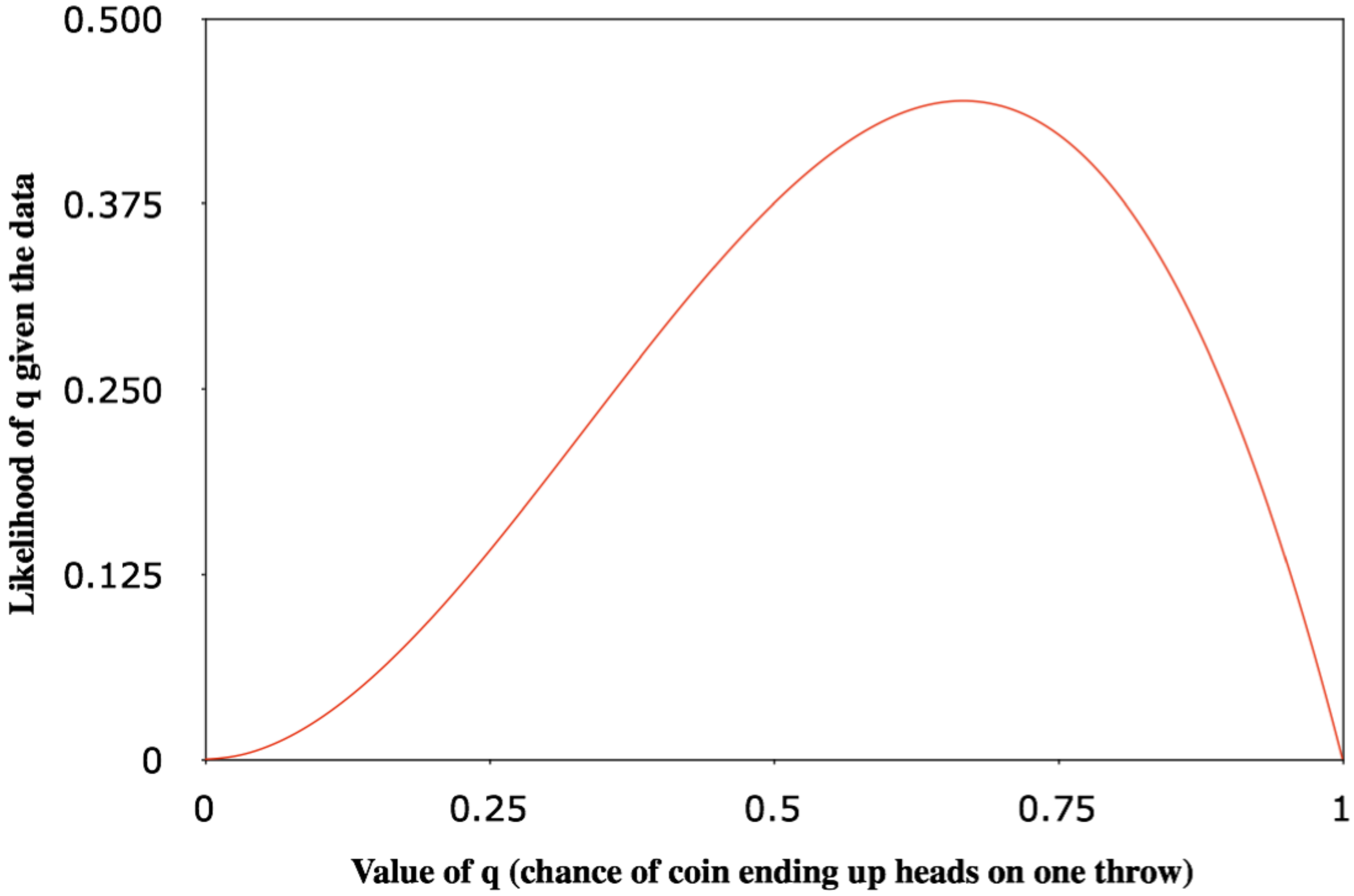
We can then say the **likelihood** of the hypothesis *q*=0.4 given the data of 2 heads out of 3 tosses is **0.288**.

Likelihood takes as the best estimate of *q* that value of *q* which maximizes the probability of the observed data.

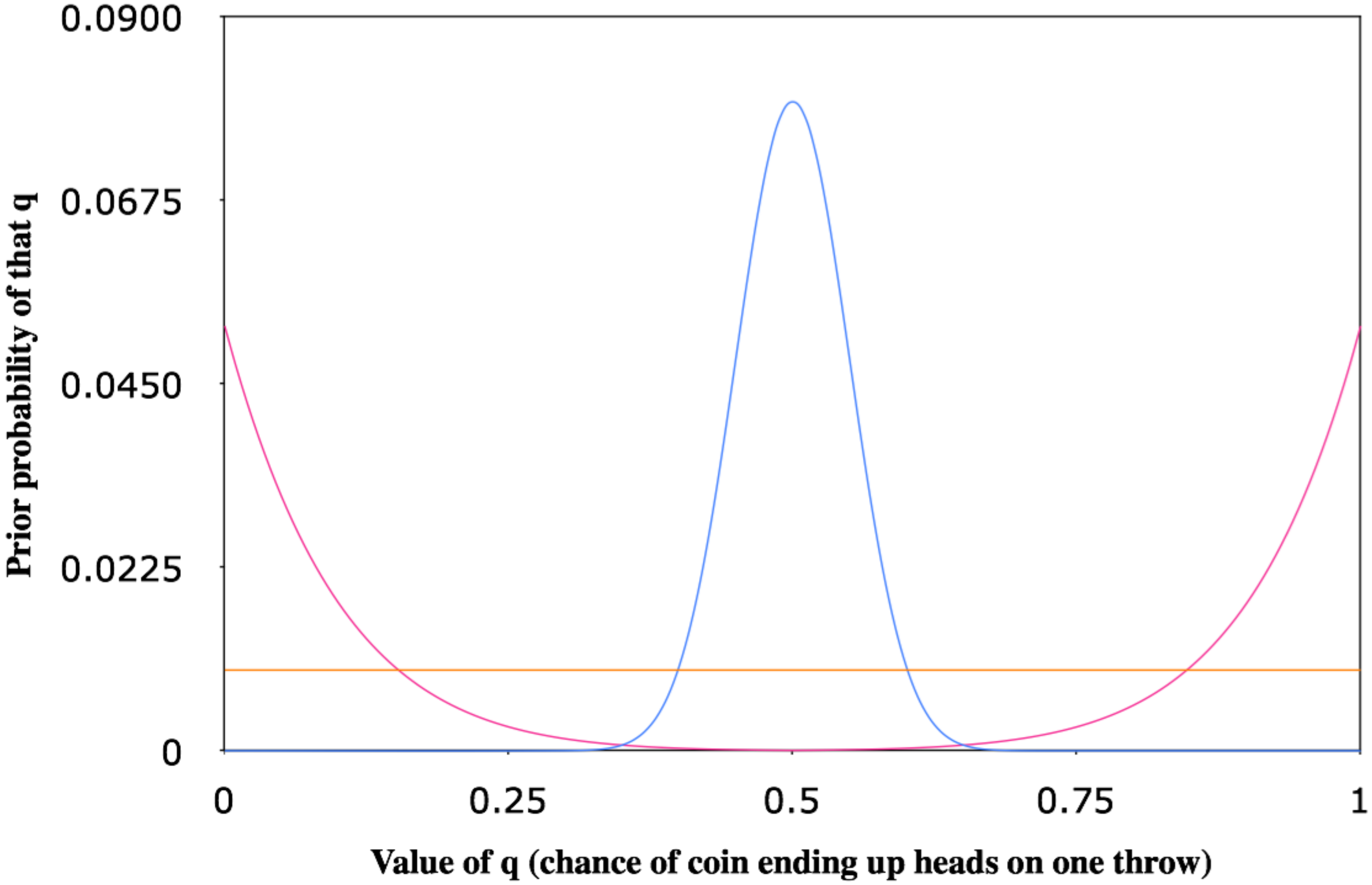
In our case, the hypothesis is just a given value for ***q***, and the data is getting 2 heads out of 3 tosses.



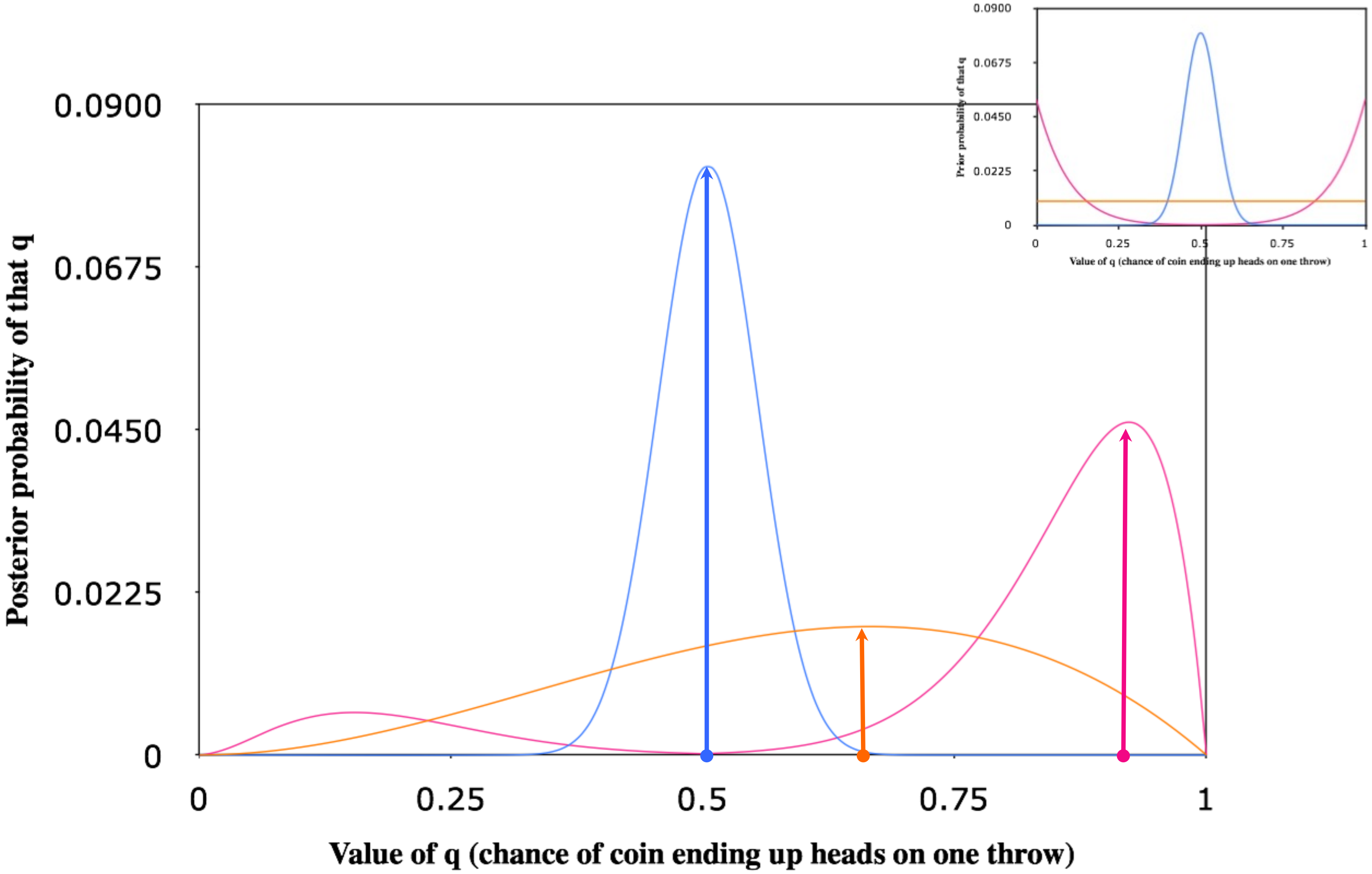
Likelihood

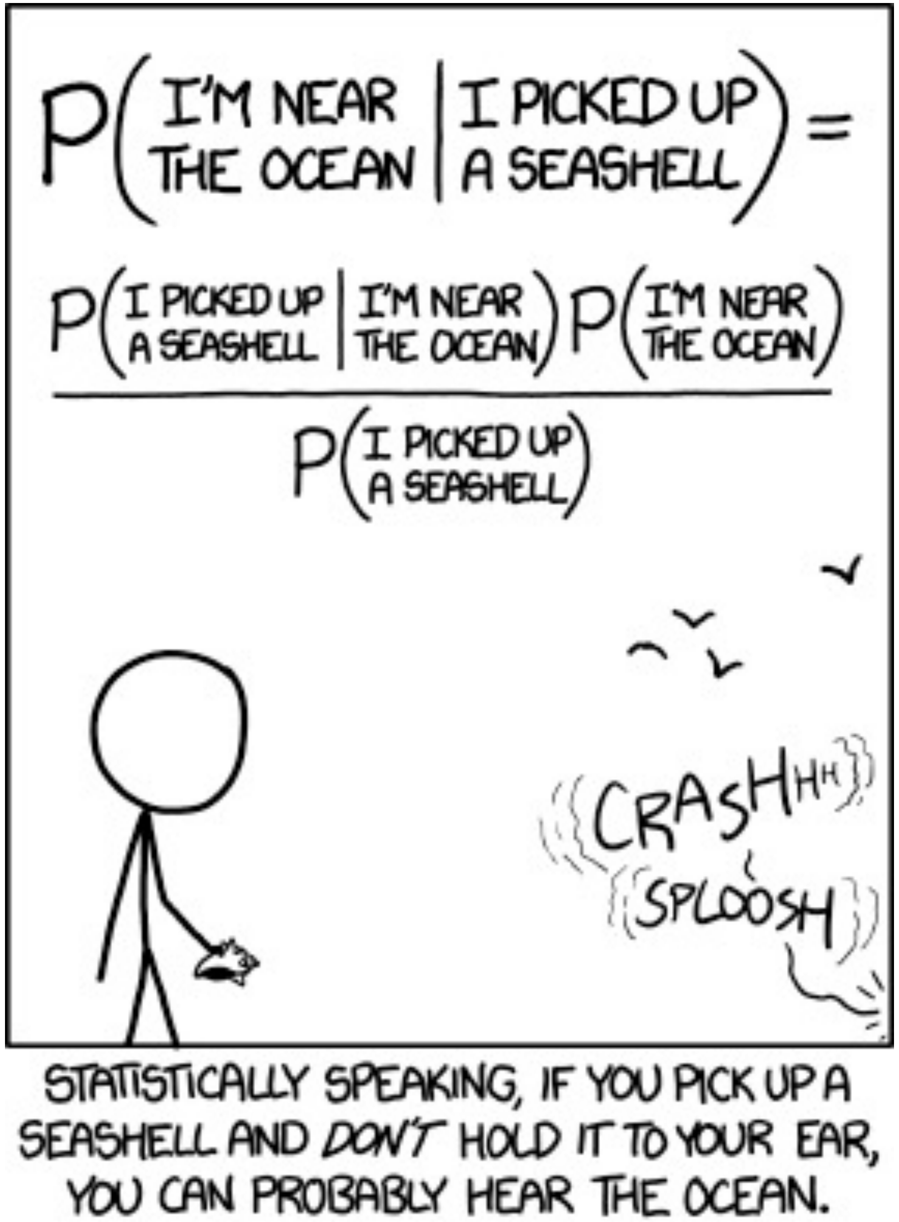

0.500
0.375
0.250
0.125
0
Likelihood of q given the data
0
0.25
0.5
0.75
1
Value of q (chance of coin ending up heads on one throw)

Prior probability



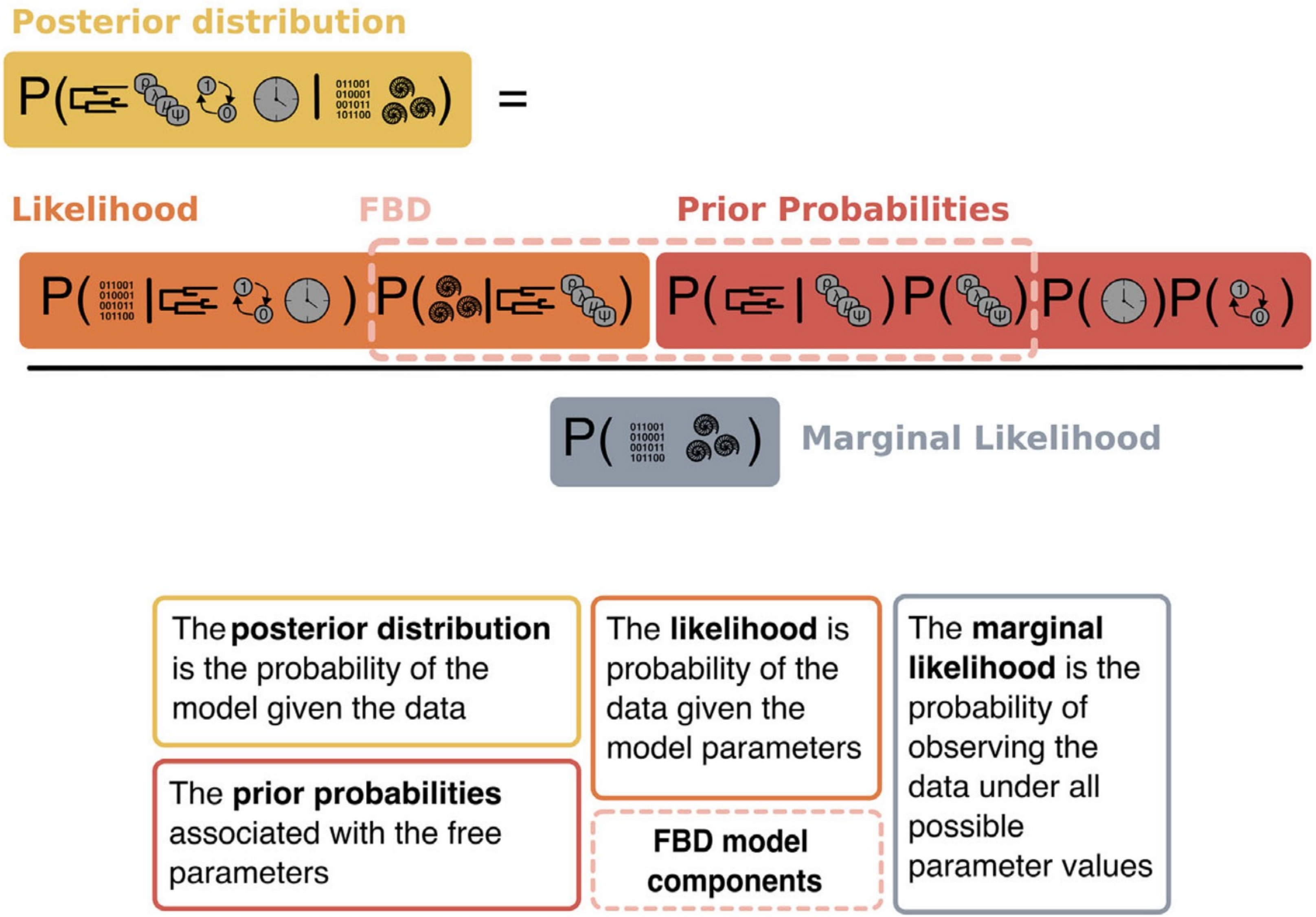
Posterior probability





https://xkcd.com/1236/

$$P(\mathrm{Model}|\mathrm{Data}) = \frac{P(\mathrm{Data}|\mathrm{Model})\,P(\mathrm{Model})}{P(\mathrm{Data})}$$



The **posterior distribution** is the probability of the model given the data

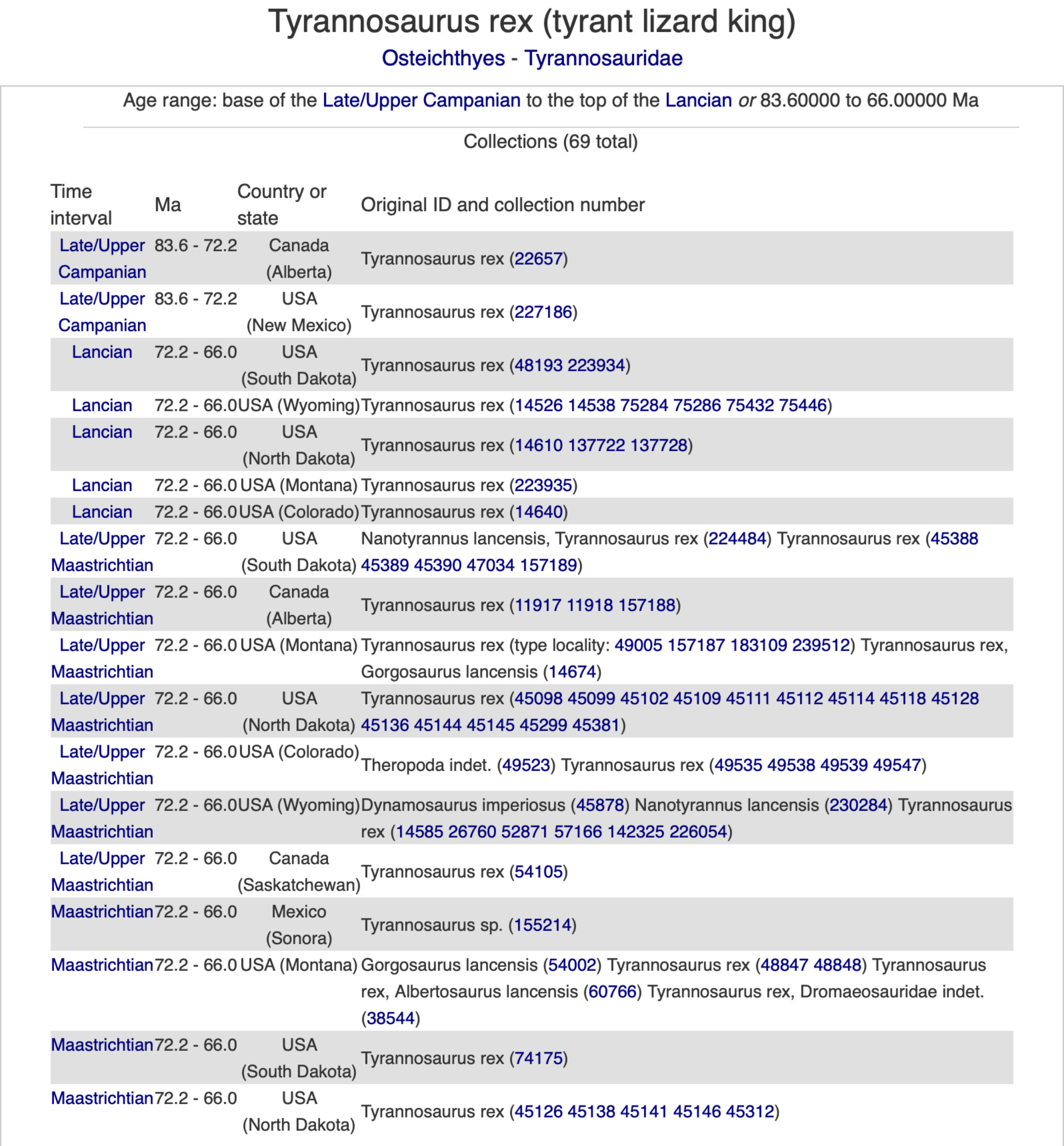
The **likelihood** is probability of the data given the model parameters

The **marginal likelihood** is the probability of observing the data under all possible parameter values

The **prior probabilities** associated with the free parameters

FBD model components

Figure 3. Symbolic representation of Bayes theorem (eq. 1) for a phylogenetic analysis of fossil ages and morphological characters. The data components include a morphological character matrix and fossil ages. The model parameters are a phylogeny with branch times, the diversification and sampling parameters of the fossilized birth–death (FBD) model, the lineage-specific branch rates of the clock model, and the parameters of the morphological substitution model (Mk model; Lewis 2001). The probabilities are delineated to highlight the joint posterior distribution, likelihood, and prior probability distributions. The FBD probability density includes some components for which we calculate prior probabilities (the tree topology, branch times, and diversification and sampling parameters) and some that are observations in the likelihood (fossil ages). Thus, these are separated to clarify the contributions to the posterior density coming from the prior and those coming from the data.



Tyrannosaurus rex (tyrant lizard king)

Osteichthyes - Tyrannosauridae

Age range: base of the Late/Upper Campanian to the top of the Lancian *or* 83.60000 to 66.00000 Ma

Collections (69 total)

Time interval	Ma	Country or state	Original ID and collection number
Late/Upper Campanian	83.6 - 72.2	Canada (Alberta)	Tyrannosaurus rex (22657)
Late/Upper Campanian	83.6 - 72.2	USA (New Mexico)	Tyrannosaurus rex (227186)
Lancian	72.2 - 66.0	USA (South Dakota)	Tyrannosaurus rex (48193 223934)
Lancian	72.2 - 66.0	USA (Wyoming)	Tyrannosaurus rex (14526 14538 75284 75286 75432 75446)
Lancian	72.2 - 66.0	USA (North Dakota)	Tyrannosaurus rex (14610 137722 137728)
Lancian	72.2 - 66.0	USA (Montana)	Tyrannosaurus rex (223935)
Lancian	72.2 - 66.0	USA (Colorado)	Tyrannosaurus rex (14640)
Late/Upper Maastrichtian	72.2 - 66.0	USA (South Dakota)	Nanotyrannus lancensis, Tyrannosaurus rex (224484) Tyrannosaurus rex (45388 45389 45390 47034 157189)
Late/Upper Maastrichtian	72.2 - 66.0	Canada (Alberta)	Tyrannosaurus rex (11917 11918 157188)
Late/Upper Maastrichtian	72.2 - 66.0	USA (Montana)	Tyrannosaurus rex (type locality: 49005 157187 183109 239512) Tyrannosaurus rex, Gorgosaurus lancensis (14674)
Late/Upper Maastrichtian	72.2 - 66.0	USA (North Dakota)	Tyrannosaurus rex (45098 45099 45102 45109 45111 45112 45114 45118 45128 45136 45144 45145 45299 45381)
Late/Upper Maastrichtian	72.2 - 66.0	USA (Colorado)	Theropoda indet. (49523) Tyrannosaurus rex (49535 49538 49539 49547)
Late/Upper Maastrichtian	72.2 - 66.0	USA (Wyoming)	Dynamosaurus imperiosus (45878) Nanotyrannus lancensis (230284) Tyrannosaurus rex (14585 26760 52871 57166 142325 226054)
Late/Upper Maastrichtian	72.2 - 66.0	Canada (Saskatchewan)	Tyrannosaurus rex (54105)
Maastrichtian	72.2 - 66.0	Mexico (Sonora)	Tyrannosaurus sp. (155214)
Maastrichtian	72.2 - 66.0	USA (Montana)	Gorgosaurus lancensis (54002) Tyrannosaurus rex (48847 48848) Tyrannosaurus rex, Albertosaurus lancensis (60766) Tyrannosaurus rex, Dromaeosauridae indet. (38544)
Maastrichtian	72.2 - 66.0	USA (South Dakota)	Tyrannosaurus rex (74175)
Maastrichtian	72.2 - 66.0	USA (North Dakota)	Tyrannosaurus rex (45126 45138 45141 45146 45312)

paleobiodb.org

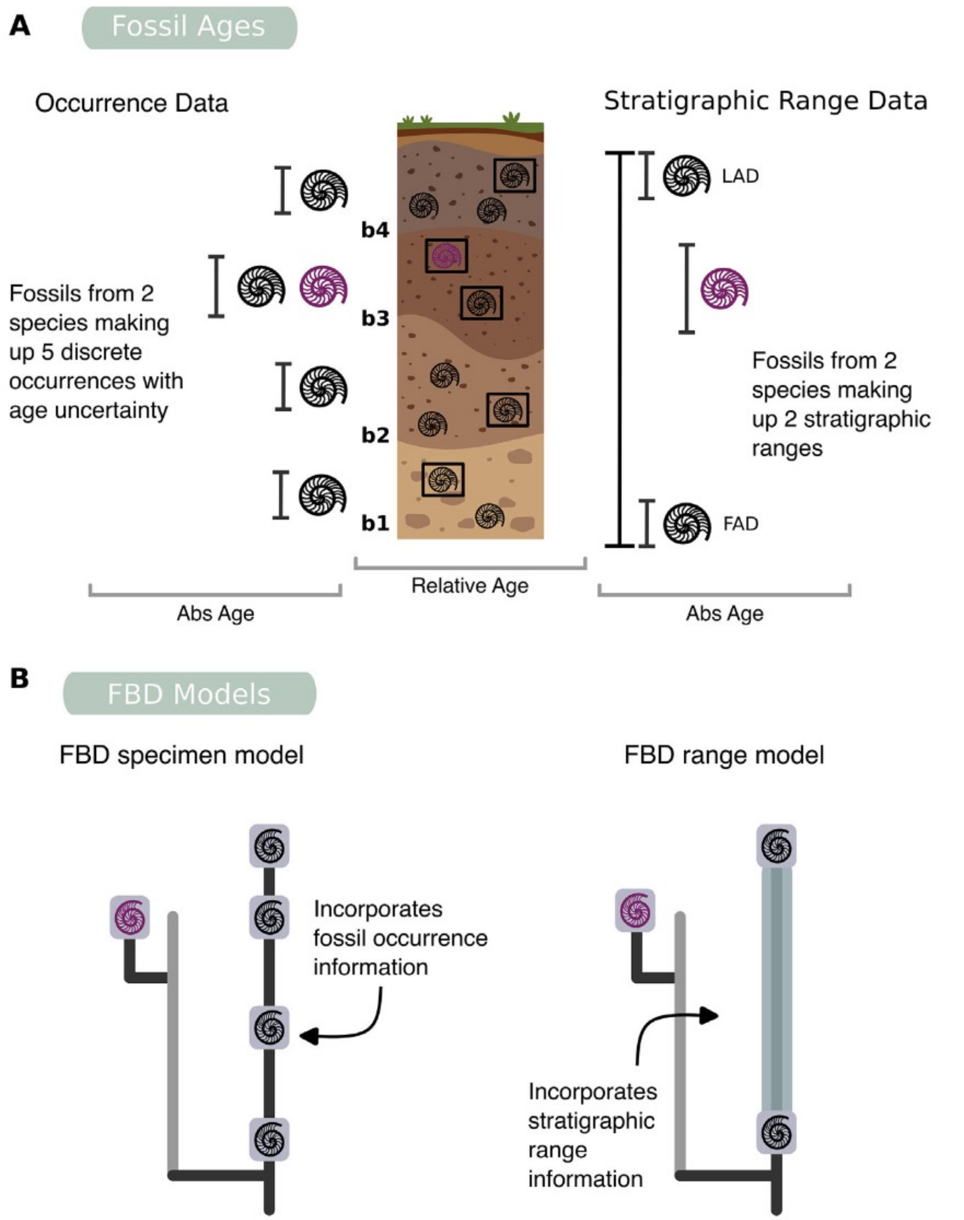


Figure 4 The temporal information available from fossils and how it can be incorporated into fossilized birth–death (FBD) models. **A,** Section with four fossil beds, b1–b4. Within each bed, there are fossils that can be used to provide temporal information for an FBD analysis. In this section, there are two different fossil taxa depicted as purple and black ammonites. Fossil age information can be taken as either occurrence data or stratigraphic range data. Occurrence data describe the age uncertainty associated with an individual sample or a discrete interval (shown to the left of the section). Stratigraphic range data describe the age around multiple fossils of the same taxon. The lower and upper bounds of the range (i.e., the first and last appearances) will also have a degree of age uncertainty around each of them (shown to the right of the section). Different FBD models are available to incorporate these are fundamentally different way of using fossil age information. **B,** How these different models incorporate the temporal information. The FBD specimen model uses occurrence information. Note that multiple fossil specimens from the same bed that are associated with the same age uncertainty should only be incorporated into the analysis once. FBD models do not currently have a way to account for abundance information. The FBD range model uses stratigraphic range information. In this case, it uses the first and last appearance fossil ages. Note, for the taxon in purple, there is only one fossil (i.e., a singleton); therefore, the occurrence and range information are the same. The gray branches on the tree represent unsampled lineages or taxa.

Box 6. Choosing priors.

Under the standard fossilized birth–death (FBD) parameterization, we place prior distributions directly on the FBD model rate parameters, namely origination (λ), extinction (μ), and fossil sampling (ψ). Often, we use exponential priors for these rates—this distribution places a high probability on values close to zero, which is typical of estimates obtained from fossil data, but does not preclude larger values.

The rates λ, μ, and ψ are always used to calculate the probability under the FBD model, but we also have the option to parameterize the model using different combinations of transformed parameters, enabling us to instead set priors on these values. For instance, we can place priors on diversification (d), turnover (r), and the probability of sampling before extinction (s), which can be transformed during inference using the formulas shown below to recover λ, μ, and ψ. One advantage of parameterizing the model using d, r, and s is that r and s can be bounded within the range [0, 1], if we assume $\lambda > \mu$, that is, that net diversification is positive, in contrast to λ, μ, and ψ, which are all in the range $(0, \infty)$ (Heath et al. 2014). It will not always be appropriate to assume $\lambda > \mu$ (see Marshall 2017), but users can and should make the choice based on their specific datasets.

Many possibilities for constraining and transforming model parameters exist within a flexible Bayesian framework. This is particularly true of more parameter-rich models, such as skylines (see Table 2, Box 5), which can allow for a more complex set of priors. In an analysis of Cambrian echinoderms, Wright et al. (2021) used the FBD skyline model, which permits variation in evolutionary rates between time intervals. They used an exponential prior for speciation, and constrained extinction such that turnover was within the range (0.90, 1.05), reflecting the observation that λ and μ tend to be correlated. Rates can also be linked across adjacent intervals (for details, see Zhang et al. 2023). Finally, it is possible to constrain or even fix the FBD model parameters based on independent estimates, which is often done for extant sampling probability, ρ (see Section The FBD Process), but can easily be done for other parameters, such as fossil sampling (e.g., O'Reilly and Donoghue 2021). Within paleobiology, we often work with per-interval sampling probabilities (Foote and Sepkoski 1999; Alroy 2008), which can also be transformed to recover ψ (e.g., Warnock et al. 2020), allowing us to take advantage of previous work in paleobiology.

Net diversification	$d = \lambda - \mu$
Turnover	$r = \lambda/\mu$
Probability of sampling before extinction	$s = \psi/(\mu + \psi)$
Origination/speciation	$\lambda = d/(1 - r)$
Extinction	$\mu = rd/(1 - r)$
Fossil sampling	$\psi = (s/1 - s)\ (rd/1 - r)$
Per-interval sampling probability	$P_t = 1 - e^{-\psi_t \triangle t}$
Per-interval rate	$\psi = -\ln(1 - P_t)/\triangle t$