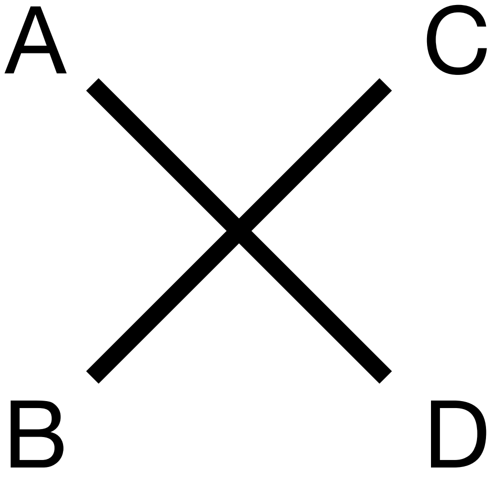
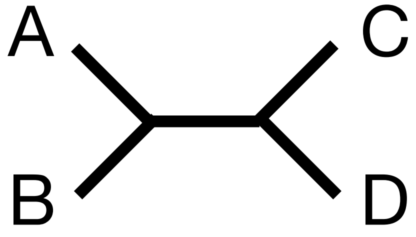
Polytomies and rjMCMC

Simulate data on the star tree. Is marginal posterior on topology uniform?

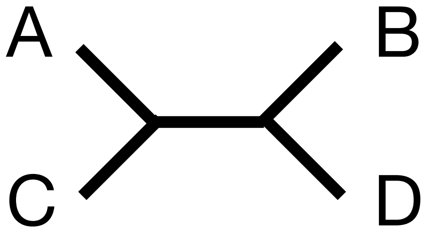


Data simulated from this tree

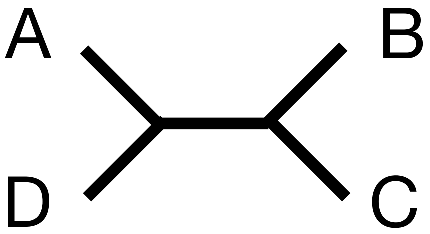
$\frac{1}{3}$



$\frac{1}{3}$



$\frac{1}{3}$



Marginal posterior uniform?

1 site sampled

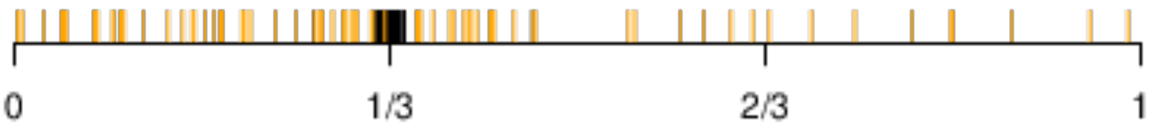
0.3405	0.3242	0.3353
0.3410	0.3279	0.3311
0.3308	0.3327	**0.3365**
0.3318	0.3329	**0.3353**
0.3302	**0.3367**	0.3331
0.3380	0.3343	0.3277
0.3324	**0.3357**	0.3319
0.3348	0.3293	**0.3359**
0.3344	0.3291	**0.3365**
0.3367	0.3320	0.3313
0.3355	0.3262	**0.3383**
0.3367	0.3318	0.3315
0.3358	0.3310	0.3332
0.3466	0.3211	0.3323
0.3429	0.3295	0.3276
0.3216	0.3390	**0.3394**
0.3318	0.3321	**0.3361**
0.3369	**0.3374**	0.3257
0.3348	0.3332	0.3320
0.3284	**0.3407**	0.3309
0.3288	**0.3370**	0.3342
0.2660	**0.4605**	0.2735
0.3345	**0.3358**	0.3297
0.3355	0.3267	**0.3378**
0.3430	0.3339	0.3231

100,000 sites sampled

0.2990	0.3288	**0.3722**
0.3172	0.0464	**0.6364**
0.1584	**0.7969**	0.0447
0.4625	0.3600	0.1775
0.7077	0.0881	0.2042
0.0884	0.0262	**0.8854**
0.9551	0.0422	0.0027
0.1826	**0.5511**	0.2663
0.3043	**0.4224**	0.2733
0.6559	0.0707	0.2734
0.0073	**0.9892**	0.0035
0.3029	0.2922	**0.4049**
0.4607	0.1362	0.4031
0.6704	0.0975	0.2321
0.6120	0.1852	0.2028
0.3605	0.3570	0.2825
0.5455	0.2505	0.2040
0.4253	**0.4254**	0.1493
0.1595	**0.7465**	0.0940
0.4436	0.1697	0.3867
0.3994	0.3904	0.2102
0.1151	**0.5912**	0.2937
0.8333	0.0951	0.0716
0.8317	0.0736	0.0947
0.2703	**0.4112**	0.3185

Most extreme posteriors in red for each case

1 site sampled in **black**
100K sites sampled in **orange**



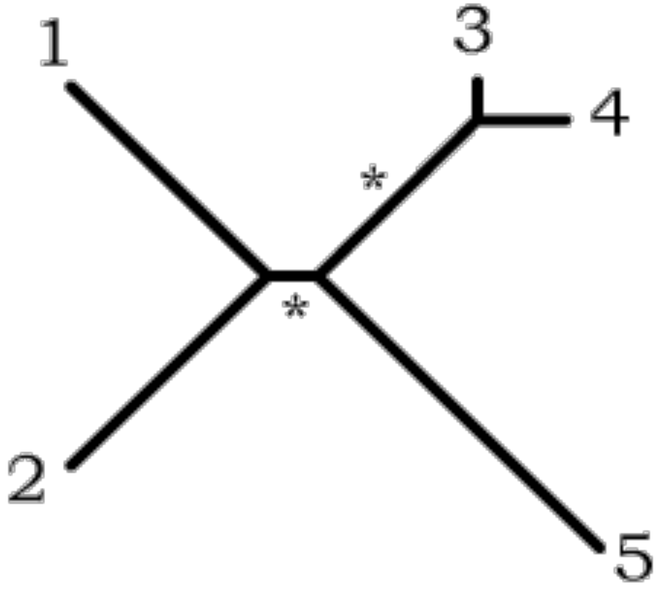
Reversible-jump MCMC

- Used when the **complexity** of the model changes during an MCMC analysis
- Requires inventing proposal mechanisms ("moves") that change the **dimension** of the model

more **complex** models have higher **dimension**
(i.e. more parameters)

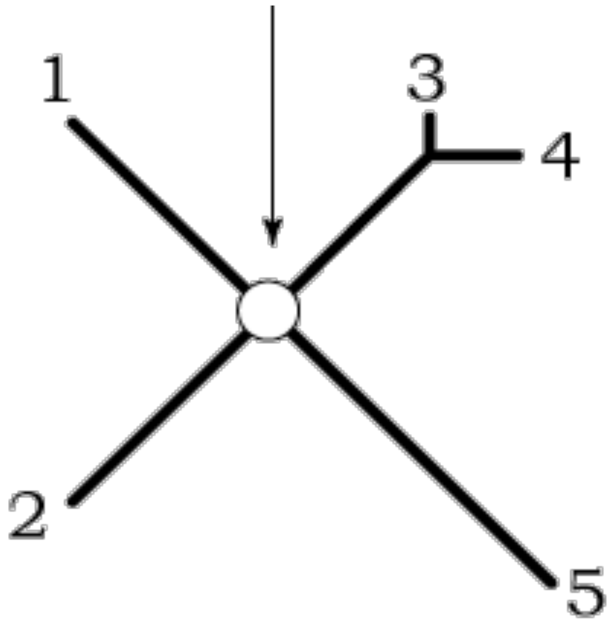
Green 1995

Delete-edge Proposal



Asterisks (*) indicate possible internal edges to delete; one of these is chosen at random

Edge deleted, creating a new polytomy

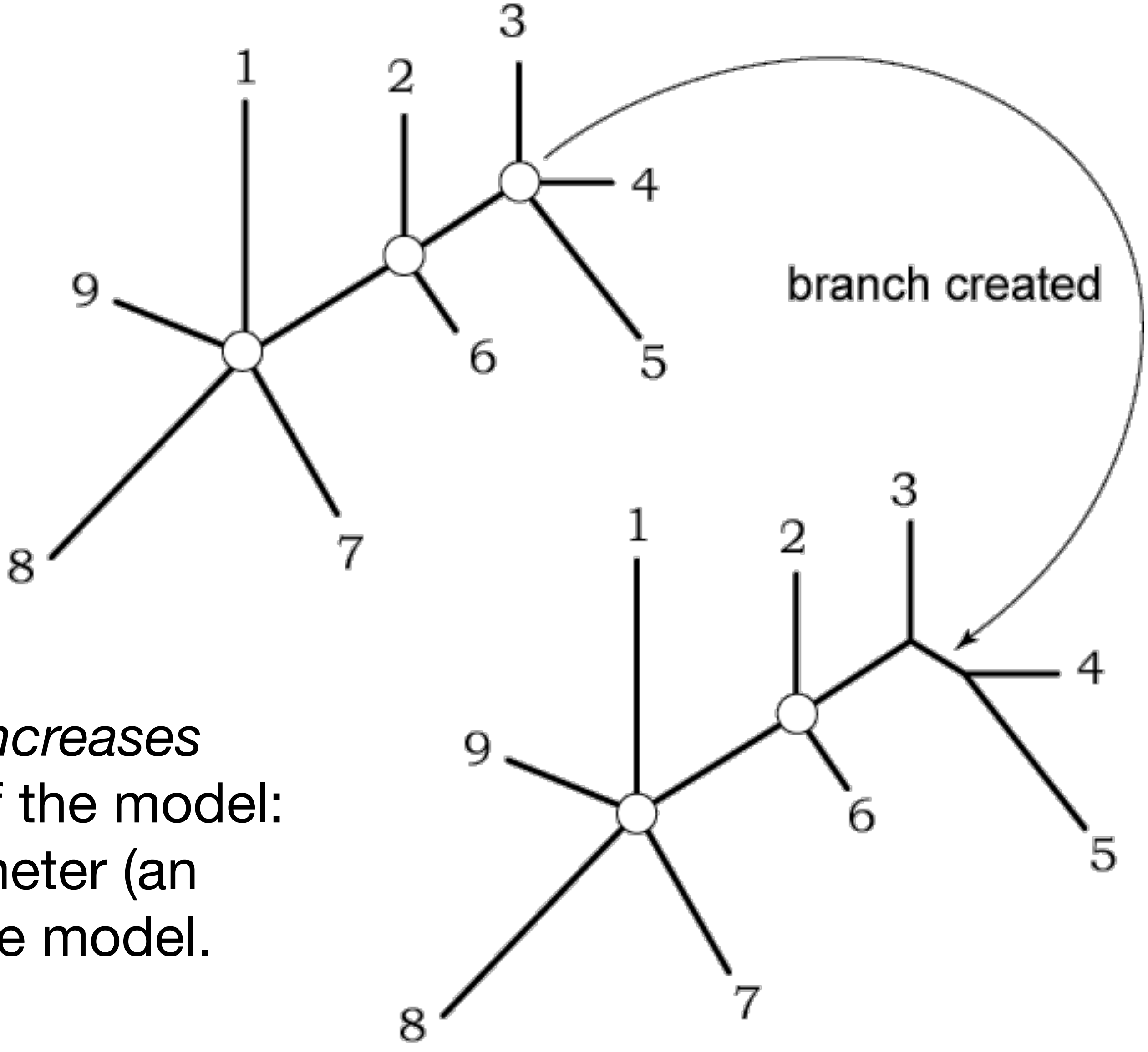


Deleting an edge *reduces* the dimension of the model because it removes one parameter (an edge length) from the model.

Lewis, Holder, and Holsinger 2005

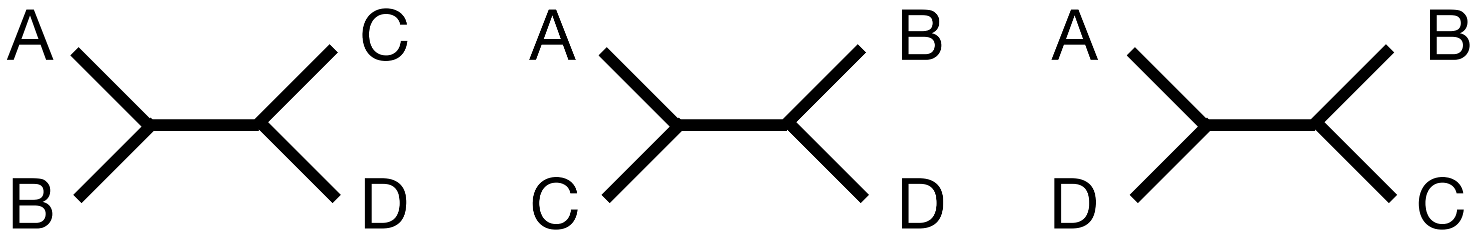
Add-edge Proposal

Open circles indicate existing polytomies that could be reduced in size by adding an internal edge; one of these polytomies is chosen at random, and an internal edge added that sequesters a random subset of the taxa involved in the polytomy



Adding an edge *increases* the complexity of the model: it adds one parameter (an edge length) to the model.

Priors: tree topologies



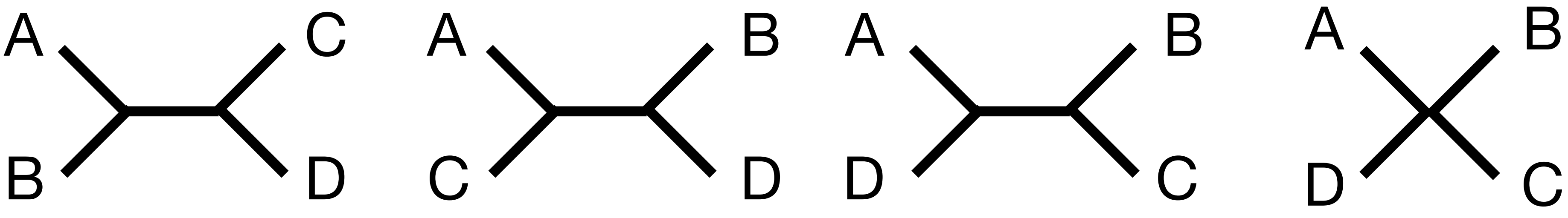
$$\frac{1}{3} \qquad \frac{1}{3} \qquad \frac{1}{3}$$

Usually, every possible tree topology is given equal prior probability in a Bayesian analysis

Priors: tree topologies

Standard prior is not really flat if polytomies are considered.

$$\frac{1}{3} \qquad \frac{1}{3} \qquad \frac{1}{3} \qquad 0$$

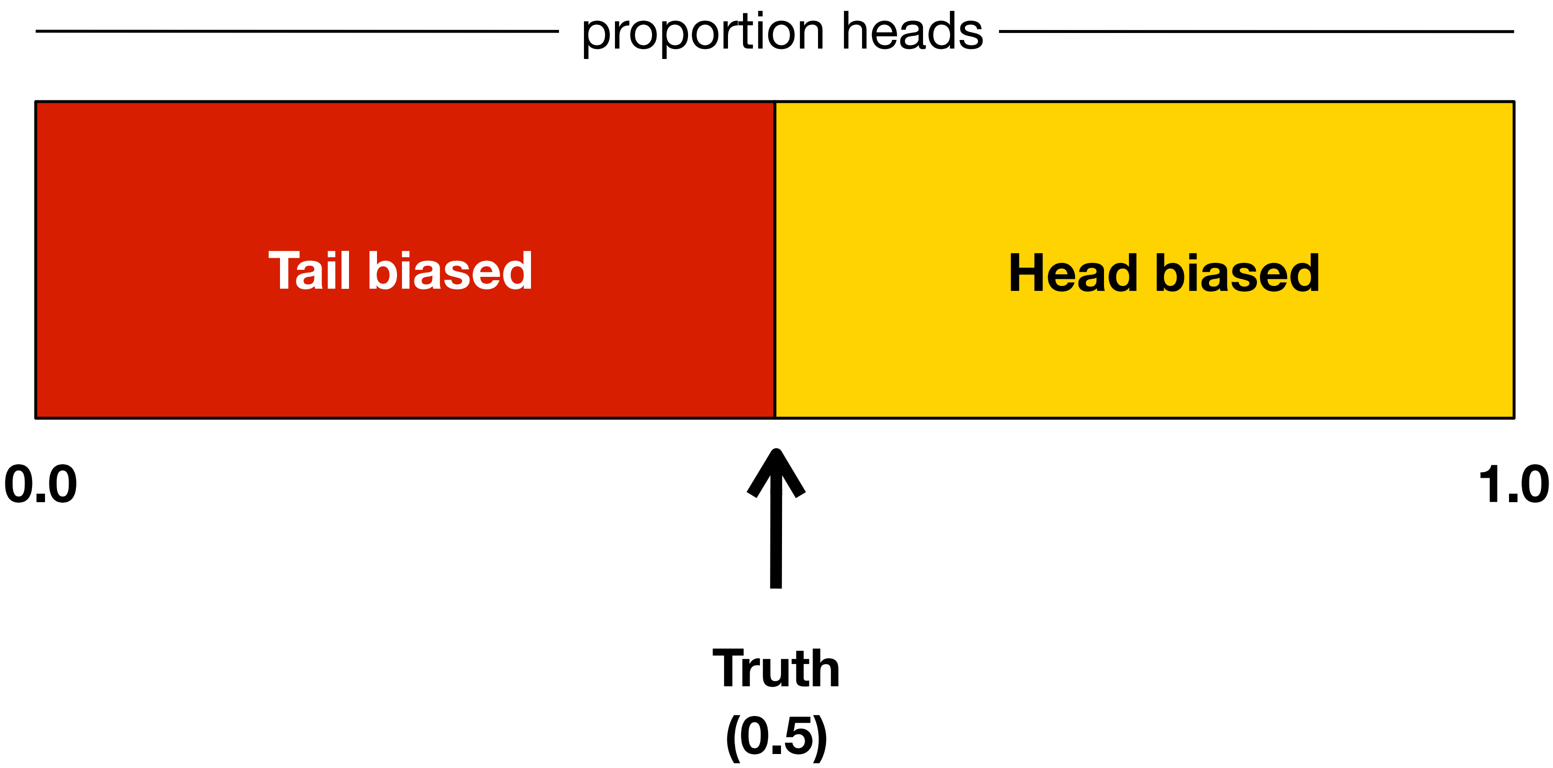


$$\frac{1}{4} \qquad \frac{1}{4} \qquad \frac{1}{4} \qquad \frac{1}{4}$$

A “polytomy prior” allows trees containing polytomies to receive some posterior support.

Fair Coin Analogy

The phylogenetic problem is analogous to a coin flipping problem in which the only hypotheses considered are *tail biased* or *head biased*, and the hypothesis *fair coin* is given zero prior probability



proportion heads

Tail biased

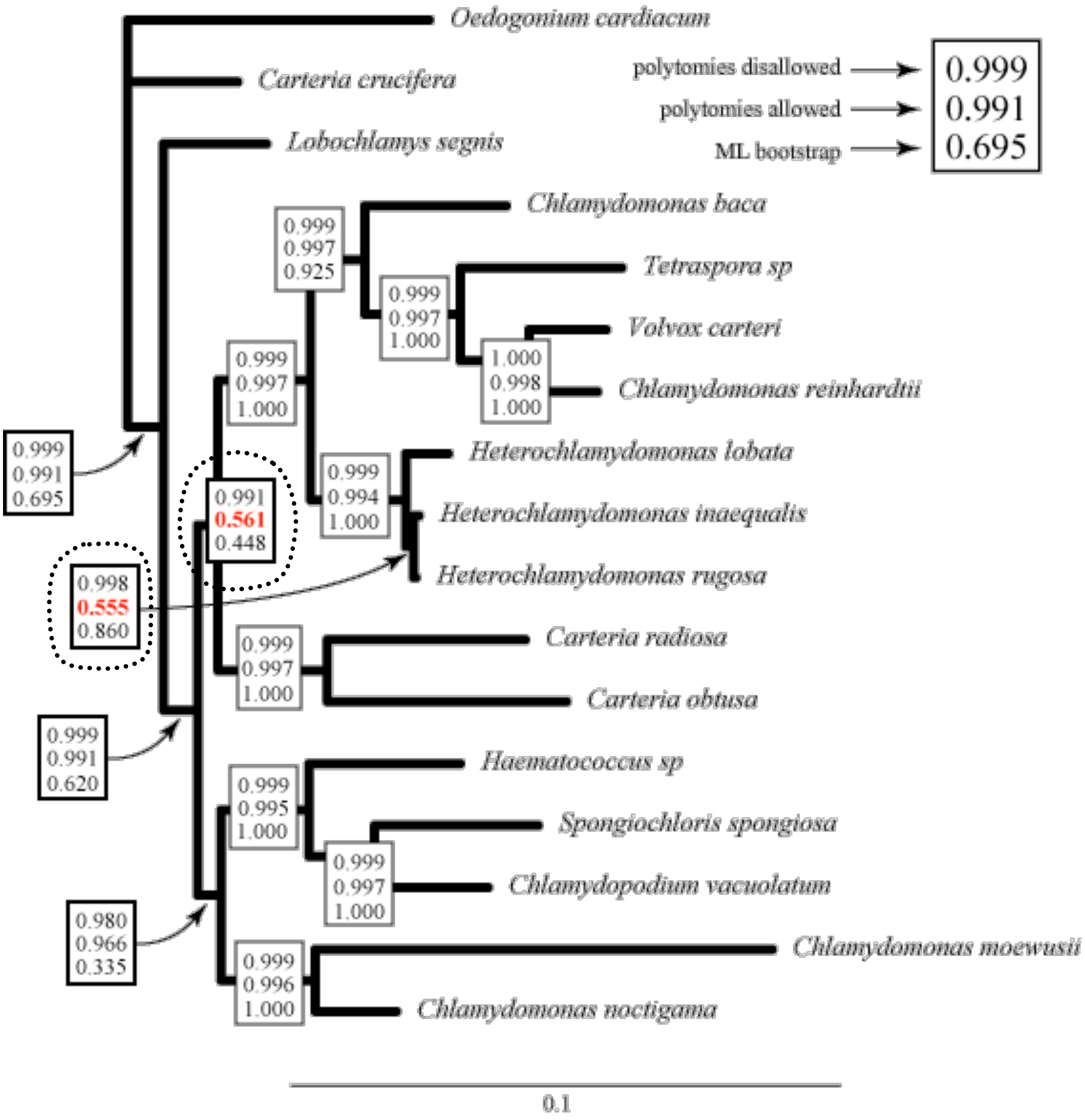
Head biased

0.0

1.0

Truth
(0.5)

Two cases where the posterior split probabilities did not agree with ML bootstrap proportions. In both of these cases, short edge lengths made high split posteriors suspicious.



Lewis, Holder, and Holsinger 2005

Bayesian vs. Bootstrap

What if only **one site out of 1000** supported a particular split? Further, imagine this site is **very reliable** (low rate but nevertheless variable).

The **nonparametric** b**ootstrap value** would be **low** because this site is missing from many bootstrap replicates, leaving no sites supporting X.

Number of sites = n

Pr(site chosen) = $\frac{1}{n}$

Pr(site not chosen) = $1 - \frac{1}{n}$

Pr(site not chosen *n* times) = $\left(1 - \frac{1}{n}\right)^n \xrightarrow{n \to \infty} \frac{1}{e} \approx 0.368$