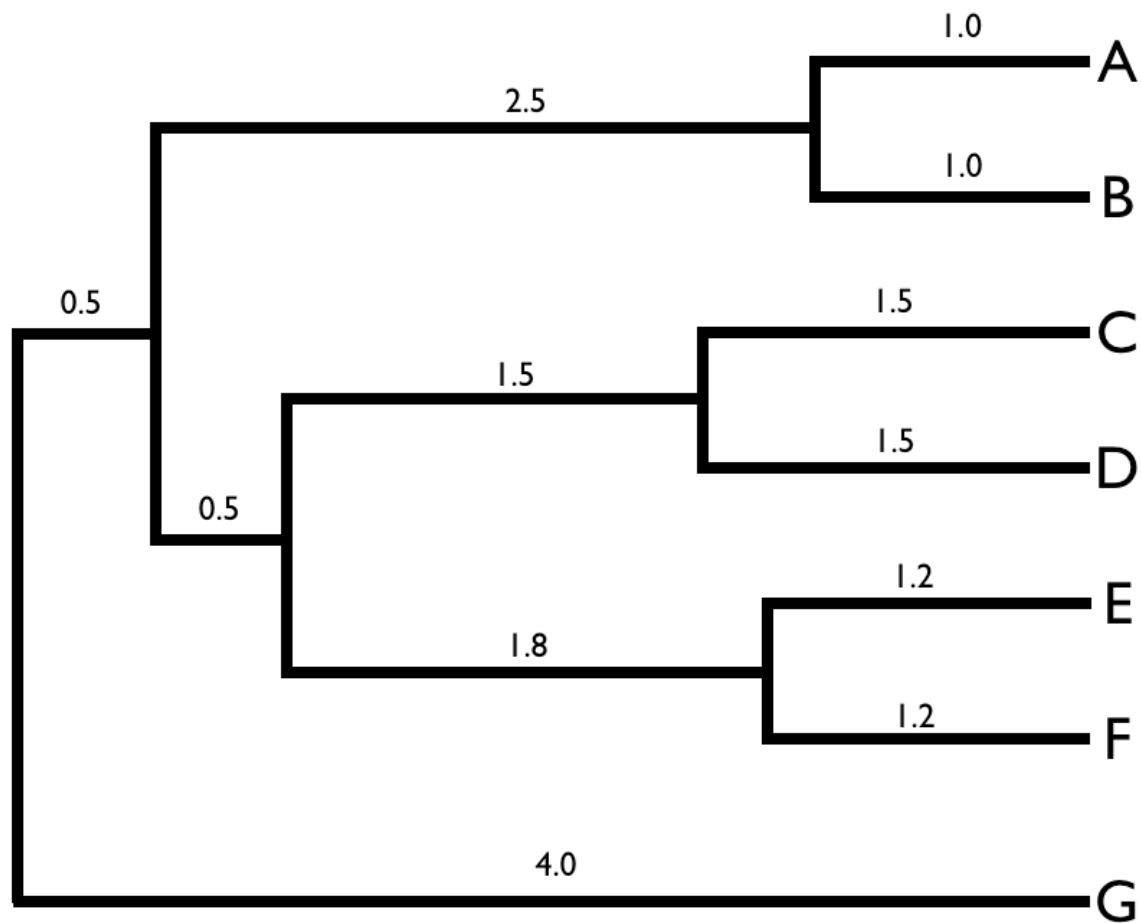


1. There is a seven-species phylogeny with branch lengths below.



a. What is the probability of obtaining the (1,6) split at the base of this tree under an ERM model?

$$\Pr[(k,n-k)] = \begin{cases} \frac{2}{n-1} & k \neq n-k \\ \frac{1}{n-1} & k = n-k \end{cases}$$

$k = 1$   
 $n = 7$   
 $\Pr[(1,6)] = 2/(7-1) = 1/3$

b. What are the other possible splits one could have at the base of a seven-taxon tree? Which of these is the most likely?

**(2, 5) and (3, 4). All three of these possibilities are equally likely ( $Pr = 1/3$ ).**

2. What is an AIC score? How are these scores interpreted? What can we use them for in comparative biology?

**AIC stands for Akaike Information Criterion. It is a measure of goodness of fit of a statistical model fit to a given data set. In general, the formula for an AIC score is  $AIC = 2k - 2 \ln L$ , where  $\ln L$  is the log-likelihood of that model given the data, and  $k$  is the number of estimated parameters. There is also a version of this formula that corrects for small sample sizes (AICc); this is the preferred formula for all real applications of AIC.**

**To interpret these scores, one compares the AIC values for a set of models fit to the same data set. The model with the lowest AIC score is the preferred model. Differences in AIC scores among competing models can be used to evaluate the strength of support for the preferred model. In general, AIC differences of about 4 mean that the model with the lower AIC score is substantially better than the competing model.**

**AIC scores can be used to select from among a set of competing models. The model with the lowest AIC score is the best explanation, among those considered, for the data; models with AIC scores that are much higher than the best model are not supported at all. There were a number of examples of this in class. I would expect you to list off a couple of examples here.**

3. We discussed a common model for growing phylogenetic trees, the birth-death model.

a. How many parameters does the birth-death model have? What are these parameters?

**2: b, the birth rate, and d, the death rate**

b. If you are growing trees under a birth-death model, from what distribution are the waiting times drawn?

**Exponential distribution with a parameter that depends on b, d, and n**

c. If  $b > d$ , how will the number of species change through time under a birth-death model?

The number of species will increase exponentially through time, with rate parameter equal to  $b-d$ .

$$E[N_t] = N_0 e^{(b-d)t}$$

4. What is “phylogenetic signal?” What can we learn by measuring it?

**See Revell, Harmon, and Collar 2008 to read exactly what I think about this. In short, it measures whether close relatives resemble each other; and you can’t learn very much biology by measuring it, because any number of processes can cause signal to be high or low.**

5. You collect data on two discrete characters for the species you study: whether or not they have hairy ears (0=no, 1=yes), and whether or not they yodel (0=no, 1=yes). You suspect that there is a relationship across species between having hairy ears and yodeling, because the hair protects females from having to listen to that nonsense all day long.

To address this problem, you fit three models to the data. For model 1, both characters evolve independently under an Mk model. For model 2, both characters evolve independently but each character can have a separate forward ( $q_{01}$ ) and backward ( $q_{10}$ ) rate. Model 3 is a dependent model where both forward and backward rates can depend on the state of the other character.

Model	lnL
1	-120.4
2	-114.2
3	-70.3

Carry out a model selection procedure on these results. What do you conclude?

**I will do AIC; you could also have done likelihood ratio tests.**

**Model 1:  $k = 2$**

$$\text{AIC} = 2k - 2\ln L = 4 - 2(-120.4) = 236.8$$

**Model 2:  $k = 4$**

$$\text{AIC} = 2k - 2\ln L = 8 - 2(-114.2) = 220.4$$

**Model 3:  $k = 8$**

$$\text{AIC} = 2k - 2\ln L = 16 - 2(-70.3) = 124.6$$

**We conclude that there is very strong support for model three over the other two. This means that the characters do seem to be evolving in a correlated fashion.**