Exam II: Macroevolution & Phylogenetics (100 pts)

Name: _____

1. Question 1
   a. What's (potentially) wrong with running a phylogenetic analysis with concatenated gene sequences (assume hybridization is not a major factor)? Identify the phenomenon that leads to a problem. (5 pts)

   b. Specify the conditions when this problem is expected to be most severe. (5 pts)

2.

a. Fill out a 5x5 variance-covariance matrix (**V**) for Brownian Motion given the phylogeny and assuming a value of $\sigma^2 = 1$. (5 pts)

A   B   C   D   E

0.5   0.5

1   1

2

1

2.5

1
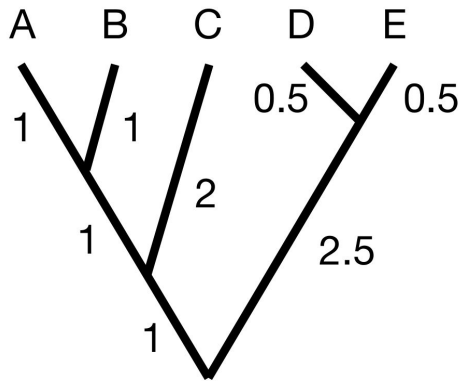
V =

b. Provide R code for how you would simulate continuous trait data for a single trait using this VCV matrix *without using a phylogenetics package*. (5 pts)

3. Compare and contrast traditional methods for time-scaling a phylogenetic tree estimated from molecular data using fossil calibrations against tip-dating methods using the fossilized birth death process. Include two figures (hand-drawn is fine) , one that explains how the fossil is incorporated into the model for node calibration and one that shows how it is incorporated using the FBD process.
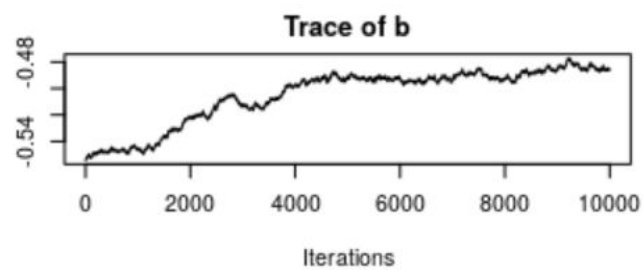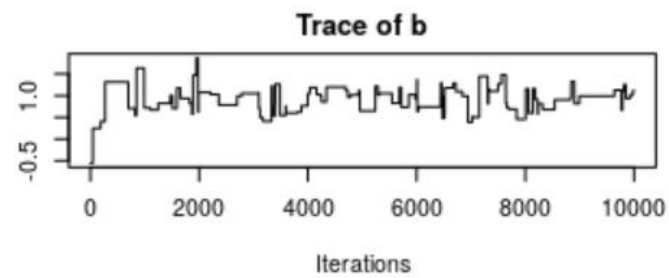(10 pts)

4. Consider the following 3 MCMC traces (4 pts):

a.

Trace of b

Iterations

b.

Trace of b

Iterations

c.

Trace of b

Iterations
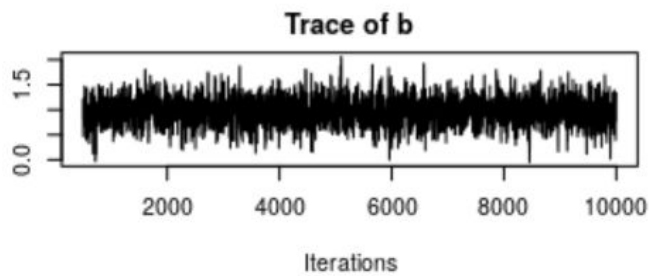
d. None of the above

Which of these traces is an MCMC that has too small a proposal distribution? ___
Which of these traces is an MCMC that has too large a proposal distribution? ___
Which of these traces are we sure has converged? _____
Which of these MCMC chains would you use as your best estimate of the parameter *b*? _____

5. Define phylogenetic signal. Explain how single-optimum OU models can be used to measure phylogenetic signal, and how this relates to Brownian Motion. (8 pts)

6.

a. What is the difference between *model adequacy* and *model selection*? Explain by providing the questions that each approach answers. (5 pts)

b. You are studying a continuous trait evolving on a phylogeny. You fit the following three models: Brownian Motion (BM), Early burst (EB) and Ornstein-Uhlenbeck (OU). You get the following model fits by AIC. What would you conclude about the pattern of evolution of this trait from this information? (3 pts)

Model	AIC
BM	382.3
EB	384.3
OU	384.6

c. You are unconvinced that any of these models is actually a good description of the patterns in your data. Explain the procedure for a *parametric bootstrap*. (1 paragraph, 5 pts)

d. You use the R package Arbutus to conduct a parametric bootstrap for 1000 simulations using the Maximum Likelihood estimates of parameters under a Brownian Motion model and find the following results for six test statistics.
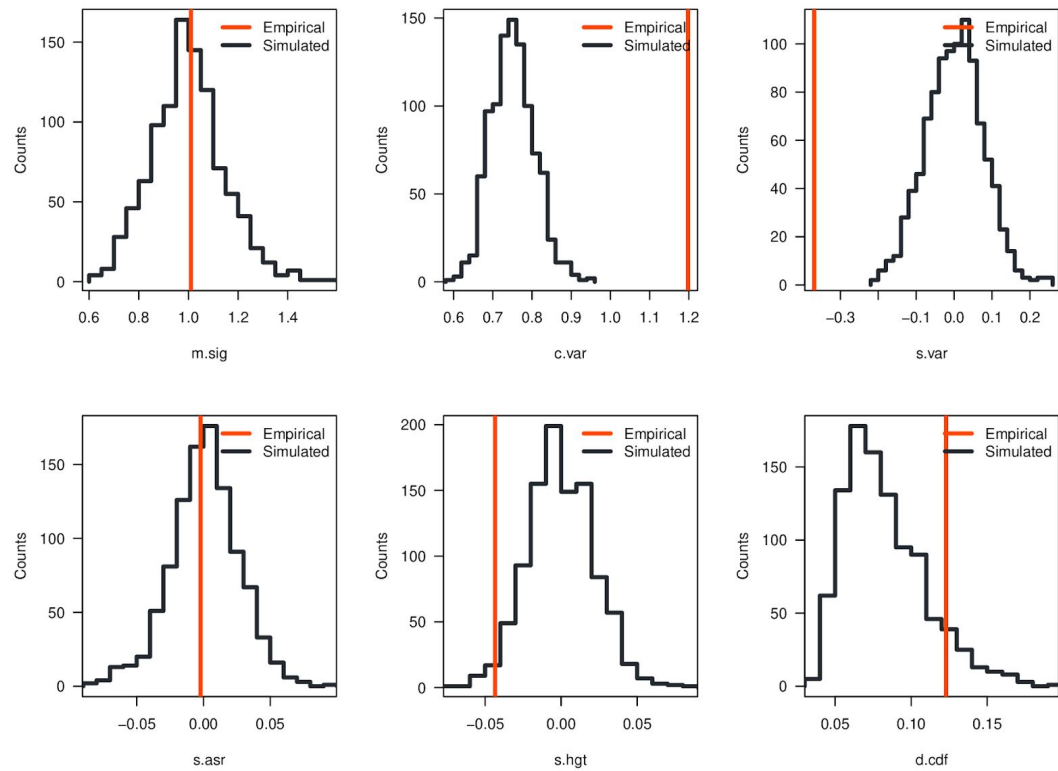
The interpretation of these test statistics is taken from this table (next page) in the Arbutus manuscript (Pennell et al. 2015). Is the model a good model? If not, what do you suspect may be wrong? (1 paragraph, 5 pts)

Table 1: Description of test statistics used to assess model adequacy

Test statistic	Description
M_{SIG}	The mean of the squared contrasts. This is equivalent to the restricted maximum likelihood estimator of the Brownian motion rate parameter σ^2 (Garland et al. 1992; Rohlf 2001). M_{SIG} is a metric of overall rate. Violations detected by M_{SIG} indicate whether the overall rate of trait evolution is over- or underestimated.
C_{VAR}	The coefficient of variation (standard deviation/mean) of the absolute value of the contrasts. If C_{VAR} calculated from the observed contrasts is greater than that calculated from the simulated contrasts, it suggests that we are not properly accounting for rate heterogeneity across the phylogeny. If C_{VAR} from the observed is smaller, it suggests that contrasts are even more than the model assumes. We use the coefficient of variation rather than the variance because the mean and variance of contrasts can be highly correlated.
S_{VAR}	The slope of a linear model fitted to the absolute value of the contrasts against their expected variances (following Garland et al. 1992). Each (standardized) contrast has an expected variance proportional to the sum of the branch lengths connecting the node at which it is computed to its daughter lineages (Felsenstein 1985). Under a model of Brownian motion, we expect no relationship between the contrasts and their variances. We use it to test whether contrasts are larger or smaller than we expect based on their branch lengths. If, for example, more evolution occurred per unit time on short branches than long branches, we would observe a negative slope. If S_{VAR} calculated from the observed data deviates substantially from the expectations, a likely explanation is branch length error in the phylogenetic tree.
S_{ASR}	The slope of a linear model fitted to the absolute value of the contrasts against the ancestral state inferred at the corresponding node. We estimated the ancestral state using the least squares method suggested by Felsenstein (1985) for the calculation of contrasts. (We note that this is not technically an ancestral state reconstruction [see Felsenstein 1985]; it is more properly thought of as a weighted average value for each node.) We used this statistic to evaluate whether there is variation in rates relative to the trait value. For example, do larger organisms evolve proportionally faster than smaller ones?
S_{HGT}	The slope of a linear model fitted to the absolute value of the contrasts against node depth (after Purvis and Rambaut 1995). This is used to capture variation relative to time. It is alternatively known as the "node-height test" and has been used to detect early bursts of trait evolution during adaptive radiations (for uses and modifications of this test see Freckleton and Harvey 2006; Slater and Pennell 2014).
D_{CDF}	The D statistic obtained from a Kolmogorov-Smirnov test from comparing the distribution of contrasts to that of a normal distribution with mean 0 and standard deviation equal to the root of the mean of squared contrasts (the expected distribution of the contrasts under Brownian motion; see Felsenstein 1985; Rohlf 2001). We chose this to capture deviations from normality. For example, if traits evolved via a "jump-diffusion"-type process (Landis et al. 2013) in which there were occasional bursts of rapid phenotypic evolution (Pennell et al. 2013), the tip data would no longer be multivariate normal owing to a few contrasts throughout the tree being much larger than the rest (i.e., the distribution of contrasts would have heavy tails).

7. Unaccounted for measurement error will increase contrasts (across/at the tips/at the base)_____ of the phylogeny. Resulting in an increase in support for (EB/BM/OU) _____ models (4 pts).

8. Explain the relationship between Brownian Motion and Ornstein-Uhlenbeck models with:
   a. Genetic drift and natural selection at the microevolutionary scale (1 paragraph, 5 pts)

   b. Macroevolutionary models of trait evolution (1 paragraph, 5 pts)

9. You are teaching a senior undergraduate student about the comparative method. In less than 2 paragraphs (< 2 minutes, you can include a quick sketch or two as well), explain the problem and solution proposed by Felsenstein 1985 in an easily understood way with minimal jargon and equations. (10 pts)

10. In his guest lecture, Dr. Diego Porto presented his work on morphological evolution and the "tweaks" to models that we can use to better reflect evolutionary reality. He presented an example from his own work using the following figure. At each of the focal nodes, there are 7 circles corresponding to 7 anatomical partitions. Filled in circles indicate that that anatomical partition analyzed individually had > 0.90 posterior probability in support of a given node. Explain why some partitions are not filled and do not have strong support for the same relationships, and possibly even conflict with it. (1 paragraph, 8 pts)

species tree
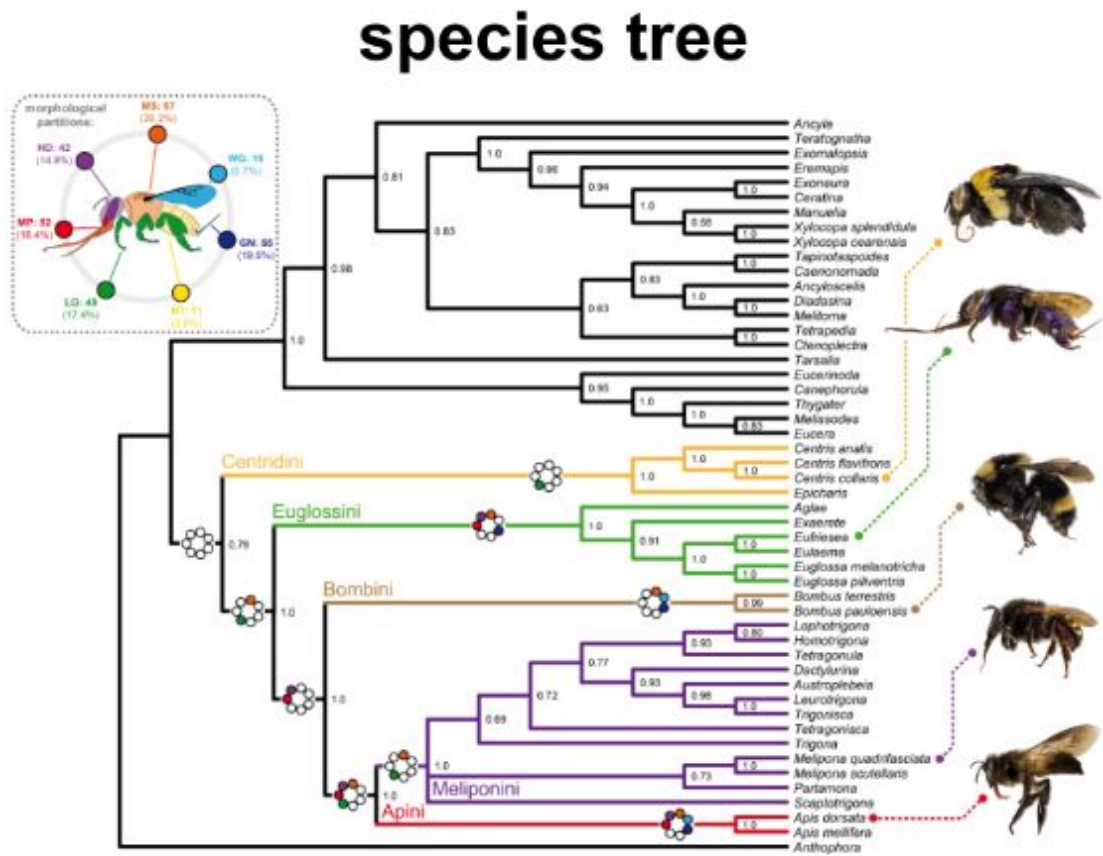
11. Reflect on Prof. Rosana Zenil-Ferguson's visit, what did you take away most from our discussion?
    (8 pts, 1-2 paragraphs; open-ended)