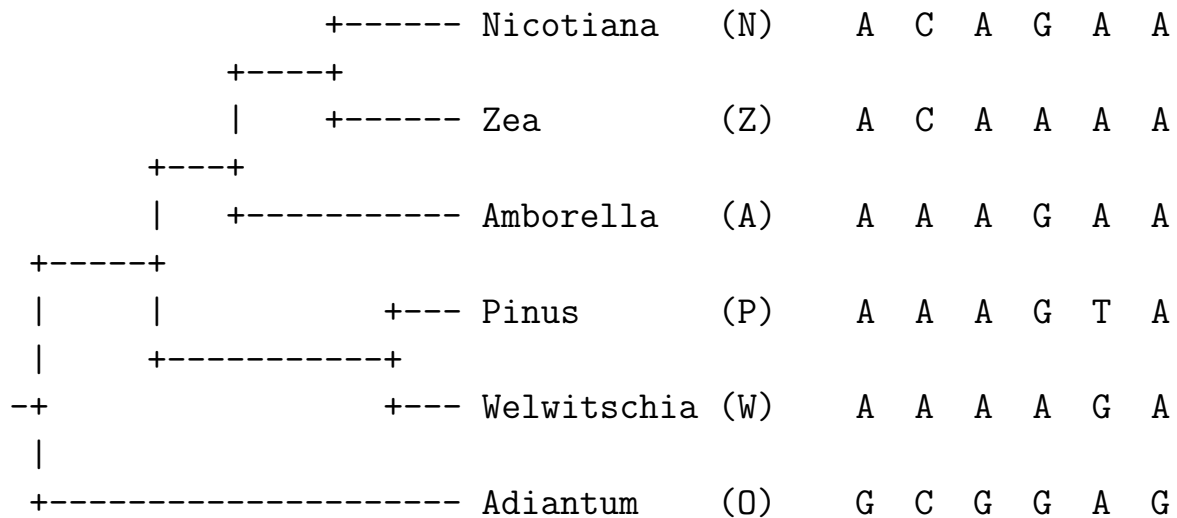


Name: _____



The first several problems refer to the phylogeny of several plant genera shown above. Letters in parentheses are one-letter abbreviations you may use below when asked to sketch trees (O is for outgroup). The data are the first six varied sites from the *rbcL* gene.

Recall that there are $1 \times 3 \times \dots \times (2n - 3) \equiv (2n - 3)!!$ rooted binary tree topologies with n leaves ($n \geq 2$).

- (3 points) Complete the following table showing the number of possible fully resolved rooted tree topologies for up to six taxa.

# of taxa	1	2	3	4	5	6
# of rooted trees	1	1				945

- (4 points) Sketch a tree with a different tree topology for these six taxa that has fewer labeled histories than the one shown.
- (4 points) Sketch a tree with the same tree topology as above that has a different labeled history than the one shown above.

-
4. (5 points) The tree has a clade comprising of *Nicotiana*, *Zea*, and *Amborella*. What proportion of all fully resolved rooted trees with these taxa contain the same clade?

 5. (5 points) What proportion of all fully resolved rooted trees with these taxa contain the N/Z/A clade and have *Adiantum* as an outgroup?

 6. (4 points) The figure shows an alignment of six varied sites. Which of these (from sites 1–6) are parsimony informative?

 7. (5 points) Find the parsimony score for the displayed tree for the second site.

 8. (4 points) Complete the following sentence: *If a site is not parsimony informative and it contains x distinct bases, then it adds _____ to the parsimony score for all possible trees.*

The following problems refer to this Q -matrix which parameterizes a continuous-time Markov chain for molecular evolution of a single site. The order of the states is A, C, G, T.

$$Q = \begin{pmatrix} -12.5 & 3.0 & 8.0 & 1.5 \\ 1.5 & -8.5 & 4.0 & 3.0 \\ 3.0 & 3.0 & -7.5 & 1.5 \\ 1.5 & 6.0 & 4.0 & -11.5 \end{pmatrix}$$

9. (4 points) If a base is currently a G and there is a substitution, what is the probability that the new base will be a C? Report your answer as a fraction.
10. (4 points) The stationary distribution of the Markov chain satisfies $\pi_A + \pi_C + \pi_G + \pi_T = 1$. Write down one more equation using values from Q that the stationary distribution must satisfy.
11. (4 points) The displayed Q matrix has not been scaled. If we choose an initial base from the stationary distribution and observe the process for one unit of time, would we expect to see more than or fewer than one substitution? Briefly explain (one sentence or phrase).
12. (4 points) The stationary distribution is $\pi = (\pi_A, \pi_C, \pi_G, \pi_T) = (0.15, 0.3, 0.4, 0.15)$. Write down a numerical estimate of the probability transition matrix $P(t) = e^{Qt}$ for a large time such as $t = 1000$.