Key for BioConcept Questions, Chapter 7

1. If all of the rooted trees in Figure 7.2B are equivalent to the unrooted tree in Figure 7.2A, why is it so important to develop an algorithm for choosing among them? Describe in evolutionary terms in what important ways these trees are different.

   Although the branch lengths are the same, the evolutionary pathways in the rooted trees are different. For example, in the tree on the top left in Figure 2B, species *a* and *c* descend from a common ancestor that is *not* also an ancestor of species *b* and *d*. In the tree to its right, the most recent common ancestor of *a* and *c* is also a distant ancestor of *b* and *d*. If the tree can be rooted, then how it is rooted is critical to understanding *how* we got to the observed distances.

2. We explored a number of distance metrics in Chapter 6 that attempted to model what happens biologically as DNA mutates over evolutionary time. Yet, many researchers choose to use character-based tree-building methods that essentially ignore any calculation of distance. What limitations do you see in the distance metrics that might keep us from accepting distance-based methods as the single best approach?

   The distance metrics only work for nucleotide sequences, not protein sequences. Their results depend heavily on the assumptions made by their developers, such as the frequency of transitions and transversions, the likelihood of multiple mutations, etc. Erroneous assumptions can reduce the accuracy of a tree based on these metrics. Character-based methods attempt to gather information from the sequences themselves without depending on an explicit set of assumptions.

3. The distance metrics used in Chapter 6 apply specifically to nucleotide sequences. In this chapter's exercises, we'll use amino-acid sequence alignments as the basis for tree building, and you may notice that we won't explicitly discuss distance metrics. In what way is a distance metric implicit in the alignment of protein sequences?

   Because protein alignments use substitution matrices that account for the likelihood of any given substitution, the alignment builds-in a model of protein evolution. This doesn't account for possibilities like multiple mutations affecting the same amino-acid position, but it does take into account factors such as whether a given substitution is likely or unlikely.

4. Suppose you are studying a group of organisms that are genuinely descended from a common ancestor and have many orthologous genes. Given a relatively constant rate of mutation and a relatively even distribution of mutations across the genome, we would expect that *any* of the orthologous genes could be used to construct a phylogenetic tree and that whatever gene we picked would give essentially the same results. It turns out, however, that not all genes are equal in terms of phylogenetic analysis. What factors can you think of that might account for differences between genes?

   Some genes are under strong selective pressure, such as genes that have alleles producing strongly negative phenotypes. Fewer changes in such genes would be observed. Others might encode many amino acids that are not critical to protein function, which would reduce the impact of mutation and allow more mutations to persist.