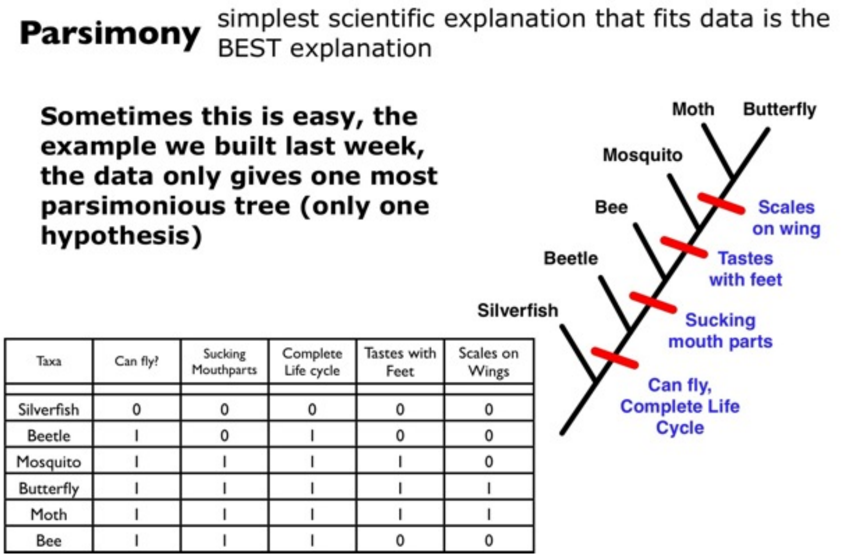
# LAB 5 PHYLOGENETIC ANALYSIS - DNA

## OBJECTIVES

- Learn about the use of parsimony in phylogenetic analysis
- Understand how gene sequence data are used for phylogenetic reconstruction
- Learn how data are used to distinguish between equally parsimonious trees and to resolve homoplasy
- Conduct sequence alignment and phylogenetic reconstruction with gene sequence data
- Interpret the evolution of morphological traits using gene sequence data

Last week we used morphological characters to make a phylogenetic tree of different taxa. At the end of lab, the program Mesquite used the character matrix you gave it to make a tree to explain the evolutionary relationships among the taxa you analyzed. You then marked your tree to show where the character states changed, either in the gain or loss of a character. You may have noticed that some characters conflicted. The program Mesquite looks at all the data you gave it and then determines the simplest phylogeny gven the shared characters you provided. This is called **parsimony**: the simplest explanation is the likeliest. In the case of phylogenetic analysis, the most parsimonious hypothesis of evolutionary relationships (or phylogenetic tree) is considered to have the fewest character changes. Sometimes the data suggest two phylogenetic trees are equally likely. Under these circumstances we say there are two MOST PARSIMONIOUS trees. Sometimes additional data or more traits will help resolve such conflicts. So Mesquite could provide you with more than one tree.

The example we did last week was relatively simple, To go over how we built that tree, see below.



**Parsimony** simplest scientific explanation that fits data is the BEST explanation

**Sometimes this is easy, the example we built last week, the data only gives one most parsimonious tree (only one hypothesis)**

| Taxa | Can fly? | Sucking Mouthparts | Complete Life cycle | Tastes with Feet | Scales on Wings |
|---|---|---|---|---|---|
| Silverfish | 0 | 0 | 0 | 0 | 0 |
| Beetle | 1 | 0 | 1 | 0 | 0 |
| Mosquito | 1 | 1 | 1 | 1 | 0 |
| Butterfly | 1 | 1 | 1 | 1 | 1 |
| Moth | 1 | 1 | 1 | 1 | 1 |
| Bee | 1 | 1 | 1 | 0 | 0 |

Because we were using parsimony to build the tree, this is the simplest explanation, or the tree with the fewest character state changes. When the data do not conflict, it's relatively easy to build a tree from a character matrix.

In general, when more characters are scored, the phylogeny can be better resolved, and the hypothesis about relationships among taxa is stronger. But if more than one tree is possible from the data, how do we know if they are equally parsimonious? The most parsimonious tree uses the FEWEST character changes. This includes both GAIN of a character and LOSS of a character, so we need to practice marking character changes in terms of both GAIN or LOSS. Here's an example:

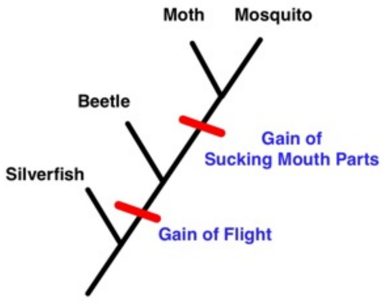
| Taxa | Can fly? | Sucking Mouthparts |
|---|---|---|
| Silverfish | 0 | 0 |
| Beetle | I | 0 |
| Mosquito | I | I |
| Moth | I | I |

This is a very basic matrix. Silverfish is the outgroup and Mosquito and Moth are grouped due to the shared character of "Sucking mouthparts". Given these data alone, the tree looks like this:

**Only one hypothesis based on the CURRENT data**

Remember to note either GAIN or LOSS of a character state

**Any tree must account for all the data**

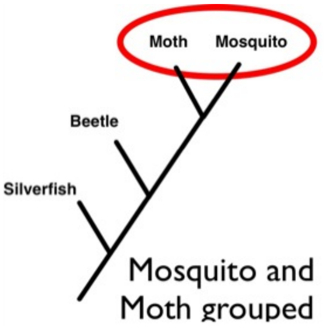


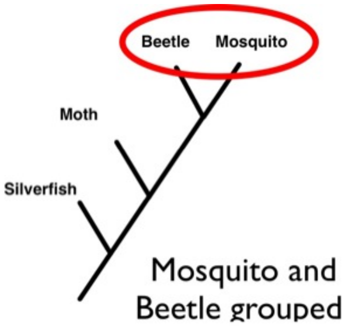
Notice we marked characters as a "GAIN" of the character. So far, this is easy. Let's score more characters:

| Taxa | Can fly? | Sucking Mouthparts | Chitinous Pupal Case |
|---|---|---|---|
| Silverfish | 0 | 0 | 0 |
| Beetle | I | 0 | I |
| Mosquito | I | I | I |
| Moth | I | I | 0 |

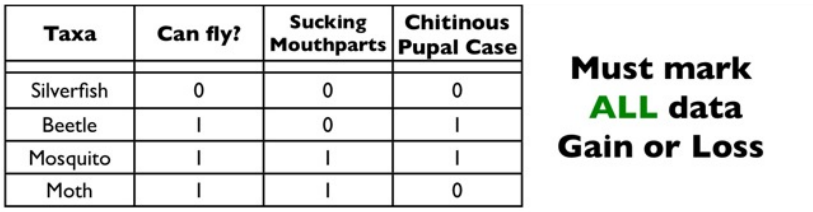
If you notice, these data have some conflicts. The character "Sucking Mouthparts" groups Mosquito and Moth together, but the character "Chitinous Pupal Case" groups together Beetle and Mosquito. Based on these data there are two possible hypotheses: (1) Moth and

Mosquito share a recent common ancestor or (2) Beetle and Mosquito share a recent common ancestor. The two alternative trees would look like this:





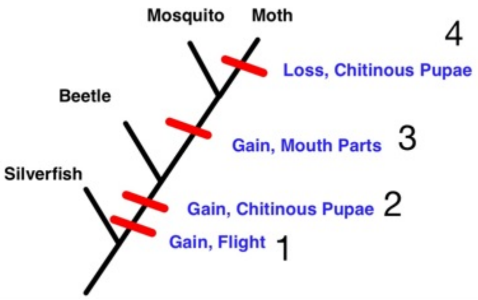
The question is, are these equally good? Are they equally parsimonious? To determine this, we must count the number of character state changes if Mosquito and Moth form a clade ('''Tree 1'''), and compare that to the number of character state changes if Mosquito and Beetle form a clade ('''Tree 2'''). First let's do''' Tree 1''':



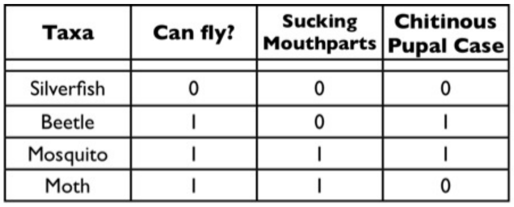
| Taxa | Can fly? | Sucking Mouthparts | Chitinous Pupal Case |
|---|---|---|---|
| Silverfish | 0 | 0 | 0 |
| Beetle | 1 | 0 | 1 |
| Mosquito | 1 | 1 | 1 |
| Moth | 1 | 1 | 0 |

**Must mark ALL data Gain or Loss**

For this tree to explain the data, **Pupal Case has to be gained before Beetle and lost before Moth**



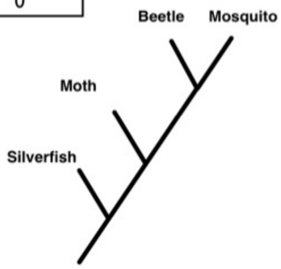
Ok now you do '''Tree 2''' (seriously...do it):



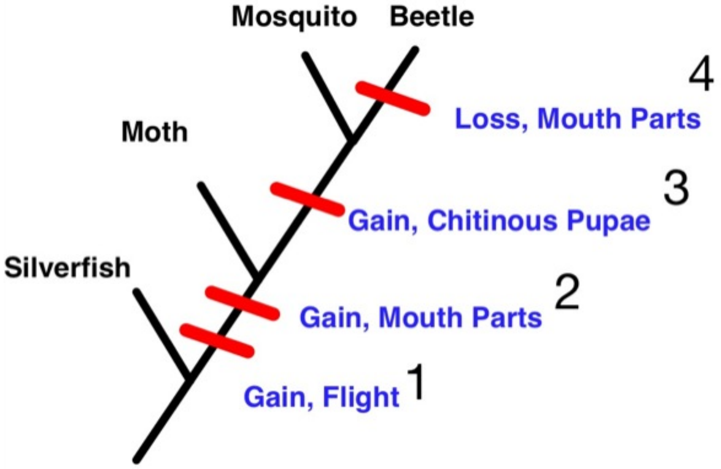
| Taxa | Can fly? | Sucking Mouthparts | Chitinous Pupal Case |
|---|---|---|---|
| Silverfish | 0 | 0 | 0 |
| Beetle | 1 | 0 | 1 |
| Mosquito | 1 | 1 | 1 |
| Moth | 1 | 1 | 0 |

**Now you mark the character changes on the second tree.**

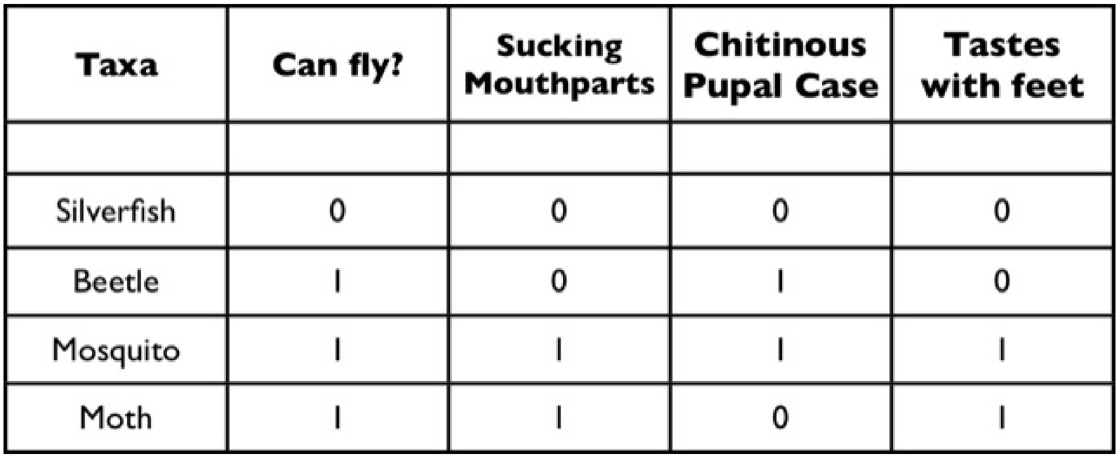
**How many changes?**



If you account for all GAINS and LOSSES if Mosquito and Beetle form a clade, you should have the following: 4 character changes:



As you see, this is the same number of character state changes as was hypothesized with Tree 1 above, in which Mosquito & Moth formed a clade. Therefore we conclude that the two trees (1 and 2) are equally parsimonious, or stated another way, both are equally feasible representations of the relationships among these taxa, given the data available. With these data, we cannot determine which hypothesis has the greatest support. To try to resolve this conflict, let's collect more data:



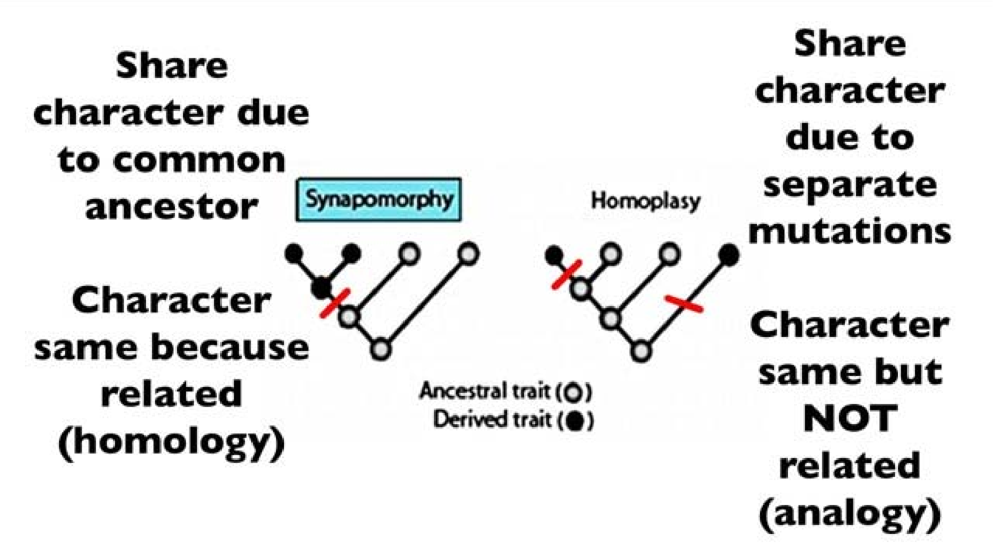
| Taxa | Can fly? | Sucking Mouthparts | Chitinous Pupal Case | Tastes with feet |
|---|---|---|---|---|
| | | | | |
| Silverfish | 0 | 0 | 0 | 0 |
| Beetle | 1 | 0 | 1 | 0 |
| Mosquito | 1 | 1 | 1 | 1 |
| Moth | 1 | 1 | 0 | 1 |

Go back and add the GAINS and LOSS that would be required to explain the new data (Tastes with Feet) in Tree 1 and then in Tree 2. Based on number of overall character changes for each tree, which tree is now most parsimonious (least number of changes)? Include this answer in your lab notebook, along with a sketch of the most parsimonious tree, along with all character state changes.

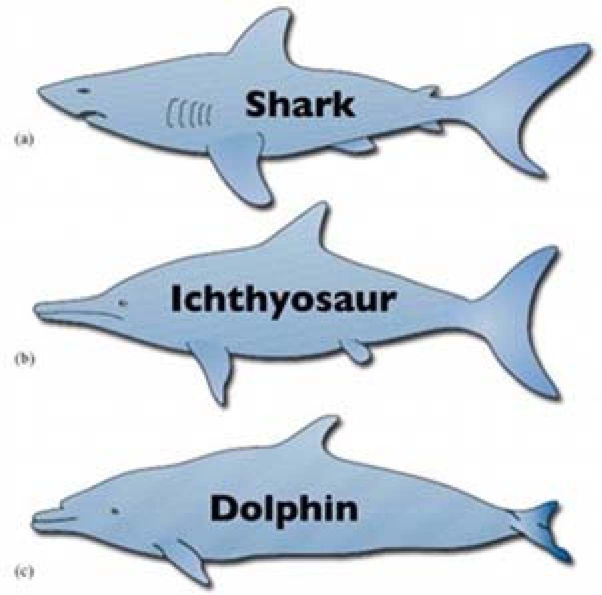
When you use morphology and shared characters to measure evolutionary relationships, there are a few challenges. First you need to score A LOT of characters which can be time consuming, and you have to measure those morphological characters very carefully which can be hard because the differences may be extremely small, but still important. In addition, looking similar doesn't always mean you share a recent common ancestor. There are two reasons two taxa can share a common character:

1. The two share a character because they both inherited that character from a recent common ancestor.
2. The two share a character, not because they share a recent common ancestor, but because that character evolved independently in two different branches of the tree.

For a visual example:


Share character due to common ancestor
Character same because related (homology)
Synapomorphy
Homoplasy
Share character due to separate mutations
Character same but NOT related (analogy)
Ancestral trait (○)
Derived trait (●)

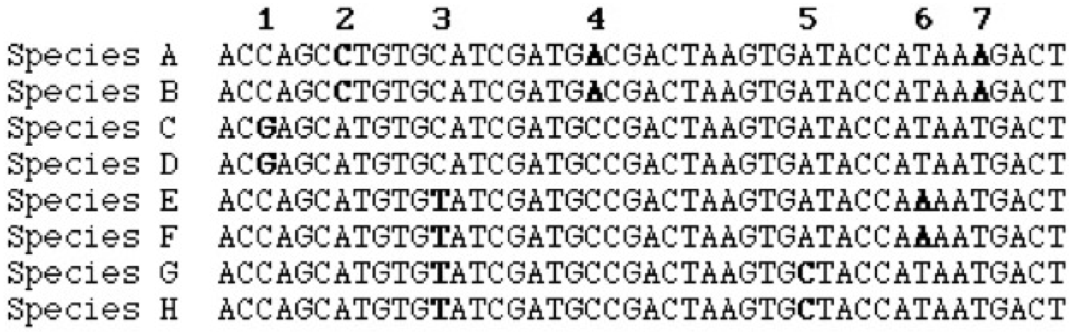
When two lineages (branches of the tree) develop the same character but do not share a recent common ancestor, it’s usually because this morphology was adaptive in a particular environment. The character evolves in multiple lineages because when mutations associated with the morphology arise in the population, it helps those individuals survive or reproduce – in other words, the morphology improves fitness. A good example of this is the comparison of dolphin, shark and the extinct icthyosaur.



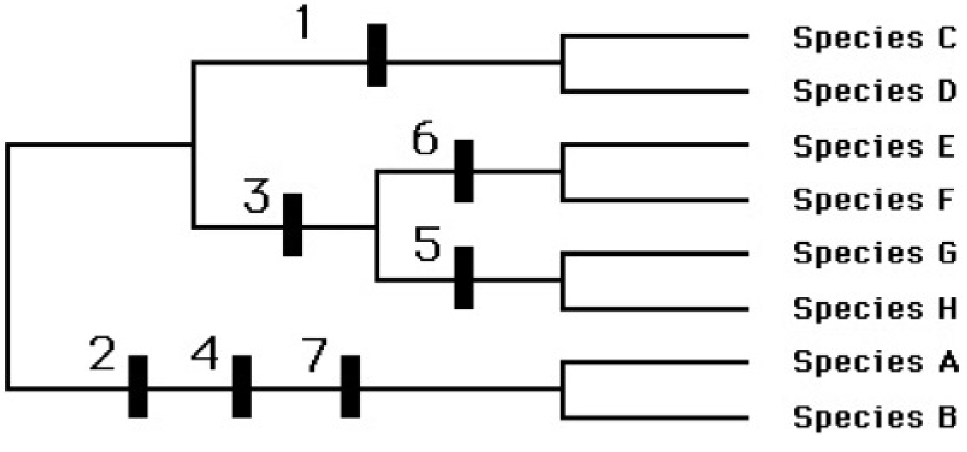
All three of these taxa have a similar, streamlined body shape, and if we were scoring them solely on their morphology, we might conclude they share a recent common ancestor. They all have flippers, a dorsal fin, a tail, and a streamlined body. Based on many other anatomical traits, however, it is most likely that their last common ancestor was over 400 million years ago. This is an example of **convergent evolution:** when similar adaptive traits evolve independently in different lineages. The three different lineages above each evolved these similar morphologies independently. This most likely occurred because they are all water-dwelling, fish-eating predators, and their morphology is adaptive in such environments.

Thus, some similar characters may not reflect common ancestry and may instead reflect convergent evolution. Characters that are the same due to convergent evolution are called **homoplasious** characters. Additional data can often help to identify homoplasy.

One useful source of additional data comes from gene or genomic sequences. The DNA sequence (ATCG bases) of a gene provides many, many characters, because each nucleotide position can potentially evolve independently (or acquire mutations) as a different character. Moreover, there are potentially four character states (A, C, T, or G) for each position. By lining up the sequence of a gene from different taxa, and using the sequence as a character matrix, a phylogenetic tree can be constructed. Consider, for example, the matrix below:



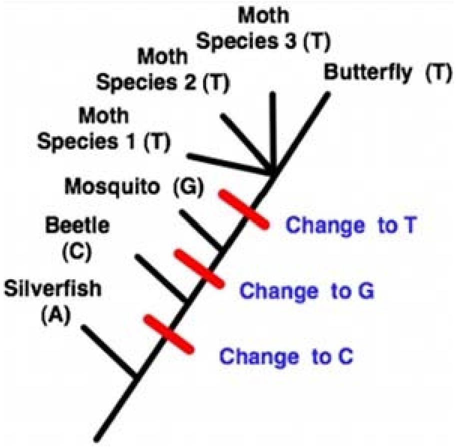
```
             1   2    3       4          5     6  7
Species A  ACCAGCCTGTGCATCGATGACGACTAAGTGATACCATAAAGACT
Species B  ACCAGCCTGTGCATCGATGACGACTAAGTGATACCATAAAGACT
Species C  ACGAGCATGTGCATCGATGCCGACTAAGTGATACCATAATGACT
Species D  ACGAGCATGTGCATCGATGCCGACTAAGTGATACCATAATGACT
Species E  ACCAGCATGTGTATCGATGCCGACTAAGTGATACCAAAATGACT
Species F  ACCAGCATGTGTATCGATGCCGACTAAGTGATACCAAAATGACT
Species G  ACCAGCATGTGTATCGATGCCGACTAAGTGCTACCATAATGACT
Species H  ACCAGCATGTGTATCGATGCCGACTAAGTGCTACCATAATGACT
```



Note that most of the base positions are identical across all taxa for this sequence, reflecting conserved regions of the gene. Regions with variations are highlighted in bold and given a number at the top of the matrix, which is used to indicate the character changes in the phylogeny above.

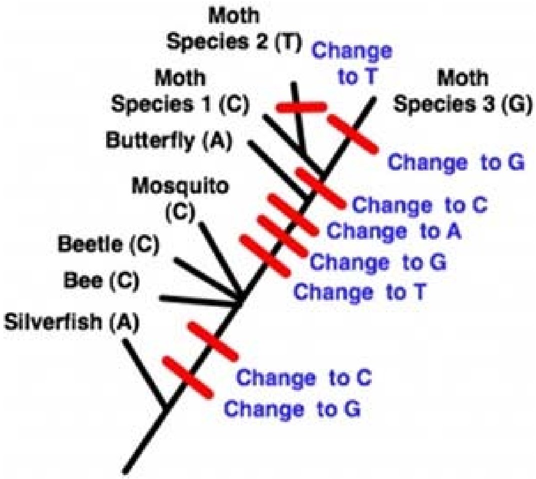
The choice of gene to examine is critical. Some genes evolve at a faster rate over time than others. Depending on the group of taxa studied, a faster evolving gene or a slowly evolving gene may be more useful.

For genes that evolve slowly over time (i.e., nucleotide substitutions are very infrequent), changes in that sequence occur only over extremely long periods of time (called deep time). For example



This kind of gene is useful for resolving relationships between taxa that diverged over long periods of time, like silverfish, beetle, mosquito and moth. But it doesn't change fast enough to help us resolve relationships among species of moths or butterflies that have diverged over a short period of time. When more than one taxon emerges from a single node (as in moth species 1, moth species 2, moth species 3 above) we call that a **polytomy**. A polytomy means that the data are not informative for resolving relationships among those taxa. Using this gene gives us a polytomy for moths.

For understanding relationships among recently diverged moths, a faster evolving gene is needed – one that changes frequently over relatively short periods of time. For example:



This kind of gene does a better job of resolving relationships among the moths. It cannot, however, resolve relationships among the beetle, bee, and mosquito.... that's because those taxa diverged over long periods of time, and mutations to this gene occur so frequently that any given base position may rotate through A,T,C,G multiple times over that long period – i.e., the character state changes are too fluid to allow resolution in deep time. As a consequence, beetle, bee and mosquito may end up with the <u>same</u> nucleotide which prevents us from determining how they are related.

For our purposes, we will simplify this by providing you with a "slowly evolving gene," but scientists who use DNA to study relationships among a group of organisms must carefully select the genes with appropriate substitution rates to get the most useful information.

In class you will be provided with files of the 12S ribosomal RNA sequences and cytochrome b amino acid sequences of your taxa.

Open your file in Mesquite, and choose show matrix. To properly compare nucleotides across different taxa, you must align the sequences. The alignment software MUSCLE is linked to Mesquite. To align sequences:

1. Matrix > Align multiple sequences > Muscle align

2. Run on separate thread, click NO

3. Type Muscle into path, click OK

Once you have the sequences aligned you should see a character matrix that Mesquite can use to construct a phylogenetic tree, using the same procedure as last week:

1. Click on Taxa & Trees > Make New Trees Block from > Tree Search > Mesquite Heuristic

2. In the popup menu, choose Treelength (click OK) > SPR rearranger (click OK)

3. Keep the number of Maxtrees at the default of 100, and on the next window click Not Separate Threads

4. Remember to ROOT the tree with your outgroup, as before, by first choosing the re-root tool and clicking on the branch leading to the outgroup. Second, choose the ladderize clade tool, and click on the branch leading to the root, or common ancestor to the whole tree.

5. Click File > Save Tree as PDF, and save the tree to your USB drive. Be sure that the file extension (e.g., Name_Of_File**<u>.PDF</u>**) is included in the file name, or it will not open. Include a copy in your lab notebook, either by transcribing it from the computer screen, or printing it out, along with an appropriate figure caption.

6. Compare the tree you generated last week, using morphological data, with the tree you generated this week using gene sequence data.

7. Add the morphological traits to your gene sequence tree generated this week. **NOTE**: These are the traits you scored from LAST WEEK's lab – refer to the matrix you saved from last week to determine the morphological traits scored. Indicate all GAINS <u>and</u> LOSSES of morphological traits that are implied by the gene sequence tree. You may find that the implied evolution of the morphological traits, and the number of character state changes, is different than the tree you generated last week.

8. Write a paragraph in your lab manual, using correct phylogenetic terminology, that describes the patterns of morphological evolution implied by the gene tree, and contrast it with the patterns observed last week.

**Questions**

1. Which morphological traits showed a different pattern of evolution in the gene tree vs. the morphology tree?

2. Which of these traits appear to have evolved convergently, based on the gene tree?

3. What do the above observations suggest about the evolution of the morphological traits?

4. Which of your morphological traits showed the least changed (or unchanged) pattern of evolution in the gene tree vs. the morphology tree?

5. Do any of these traits appear to have evolved through convergent evolution?

6. Spend some time on the website tolweb.org finding the taxa that you've studied (or their relatives) and make a sketch in your lab notebook, including an appropriate figure caption, of the currently accepted phylogeny of the relationships among your taxa.

7. Which of your phylogenies (gene tree or morphology tree) appears to be most consistent with the phylogeny on tolweb.org? What does this imply about the evolution of the characters (base positions or morphological traits) that you studied?