LAB 4 PHYLOGENETIC ANALYSIS – MORPHOLOGY

Modified from exercises described by Lemke and Jensen (2012), Gibson (2018), and others.

OBJECTIVES:

- Learn to read and interpret phylogenetic trees
- Infer character evolution from phylogenetic relationships
- Score morphological traits on organisms from biological collections
- Build a trait matrix for a group of organisms based on scored traits
- Use a computer program (Mesquite) to estimate phylogeny based on morphology

INTRODUCTION TO PHYLOGENETICS

Scientists make hypotheses about evolutionary relationships based on observations and gather evidence to test those hypotheses. We refer to this field of making testable hypotheses of evolutionary relationships as phylogenetic systematics.

Hypotheses about the evolutionary relationships among groups of organisms are based on the idea that they are arranged together by shared characters that have been inherited from a common ancestor.

These hypotheses are often represented by branching diagrams called phylogenetic trees.

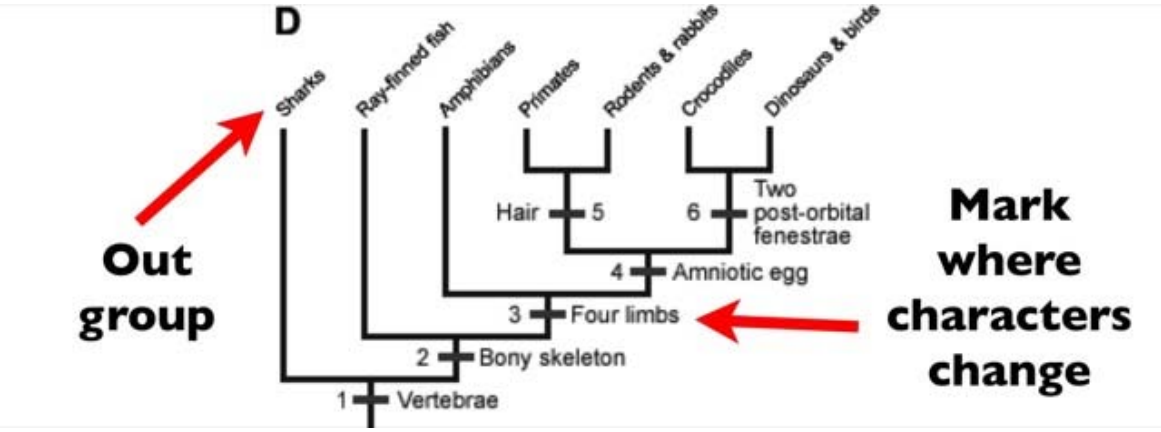
There are many types of data that can hypothesize evolutionary relationships and build phylogenetic trees. Morphological characteristics, DNA/molecular data, ecological information, behavioral characteristics and development can all be useful. The more information that can be collected about the group or **taxon, (plural taxa)** that we are interested in, the more likely we are to build an accurate representation of how these organisms are related.

This week, we will discuss the idea of shared characters, build a character matrix using morphological observations/measurement, and generate a phylogenetic tree using the software Mesquite. Next week, we will discuss the concept of parsimony as it applies to building and evaluating your phylogenetic hypothesis and look at DNA sequences as additional data for determining phylogenetic relationships.

The first step in building a phylogenetic tree of a group of organisms is to group the organisms by similarity (shared characters), and then evaluate which characteristics are shared or different. To organize the information, evolutionary biologists make a **character matrix**. This is a table of the characters used to categorize these organisms and the names of the organisms.

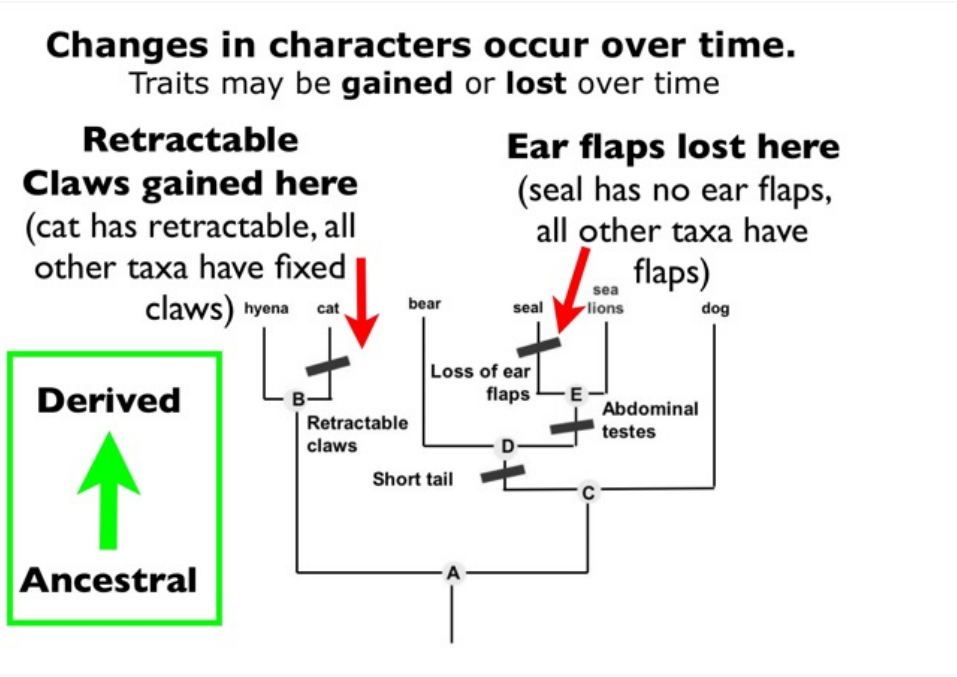
Outgroup	Vertebrae?	Bony skeleton?	Four limbs?	Amniotic egg?*	Hair?	Two post-orbital fenestrae?**
Sharks and relatives	YES	no	no	no	no	no
Ray-finned fishes	YES	YES	no	no	no	no
Amphibians	YES	YES	YES	no	no	no
Primates	YES	YES	YES	YES	YES	no
Rodents and rabbits	YES	YES	YES	YES	YES	no
Crocodiles and relatives	YES	YES	YES	YES	no	YES
Dinosaurs and birds	YES	YES	YES	YES	no	YES

In this example sharks are the most different from the rest of the organisms involved in the study. We call sharks the **outgroup:** a taxon that is closely related to the other organisms in the study, but less closely than any of the others are to one another. The outgroup serves as an important reference point, and it allows us to **root** the tree. As evolution progresses through time, accumulated mutations within lineages cause characters to change; new traits appear and other traits disappear. The tree allows us to see when characteristics developed during the process of diversification of the lineage and to group the organisms by SHARED characteristics, reflecting their shared ancestry.

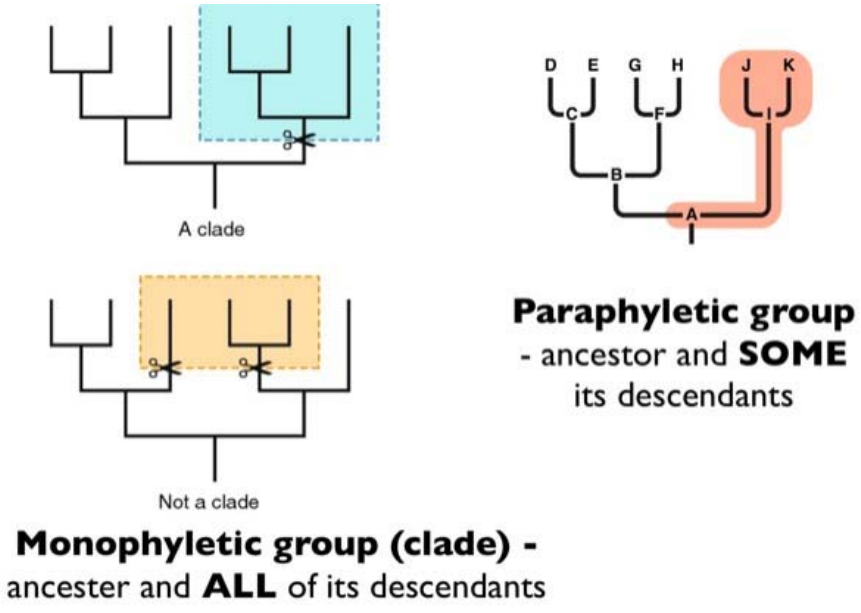


In this example, vertebrae evolved earliest and all organisms in the group share this. The number 1 "slash" shows when that occurred in time, and all the organisms above that slash (which is all of them) have that character. Bony Skeleton evolved a little later, and all organisms except sharks have a bony skeleton... so this trait first evolved in the ancestor to ray-finned fish and all the others. Four Limbs is shared by all organisms except Sharks and Ray finned fish, so it evolved at slash 3, in the ancestor to all the animals in the list except sharks and ray-finned fish. Using the groupings from the character matrix we can build a tree that depicts the historical relationships of these different organisms, AND the characters that define those relationships. (Note: there are MANY MANY characters that were scored to make this tree and that underpin these evolutionary relationships, shown are only six traits for simplification.)

Characters can be categorized as ancestral or derived. All species, including our own, have a mixture of ancestral and derived characters. **Ancestral characters** are those that have been inherited from an ancestor and not evolved or changed – in other words, these traits are the same as they are in the ancestral taxon. **Derived characters** are those that have undergone more recent evolutionary change in a group. Species that share derived characters are grouped together because their most recent common ancestor also possessed those characters.

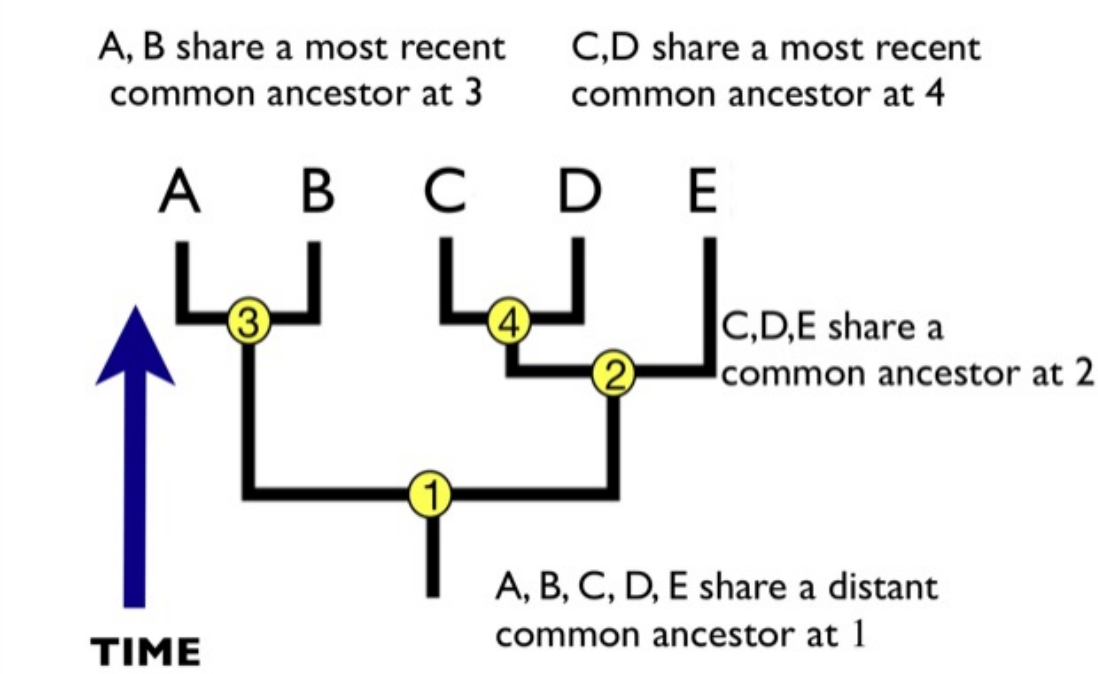


Groups of organisms that share a common ancestor are called **monophyletic groups.** More specifically a monophyletic group (also called a clade) is one ancestor and ALL its descendants. A paraphyletic group is an ancestor and only SOME of its descendants.



READING A PHYLOGENETIC TREE:

On the "tips" of phylogenetic trees are taxa. The nodes (labelled here as 1-4) indicate the hypothetical shared ancestors of the branches joined by that node. More closely related taxa are joined by more RECENT ancestors, which are closer to the tips of the tree. In this diagram for instance, D is more closely related to E than to B. This is because D and E share a more recent common ancestor at node 2, whereas the only common ancestor shared by D and B is represented at Node 1. Node 2 is closer to the tips of the tree, and so the ancestor represented by Node 2 lived at a more recent time than the one represented at Node 1.



CAMINALCULES ACTIVITY

To practice collecting data on shared characters for a character matrix, and to practice using a character matrix to build a phylogenetic tree in the software Mesquite, we will be using a classical teaching activity in systematics known as “Caminalcules.” The Caminalcules are imaginary animals for explaining phylogenetic systematics. They are named after Dr. Joseph Camin, a biologist from the University of Kansas who developed the exercise to teach evolution. Camin invented both "living" (extant) Caminalcules, and fossil versions, and designed them to reflect evolutionary patterns that biologists also observe in real species. We will work with 13 “extant” taxa, and attempt to reconstruct the phylogenetic tree that Camin designed for his creations.

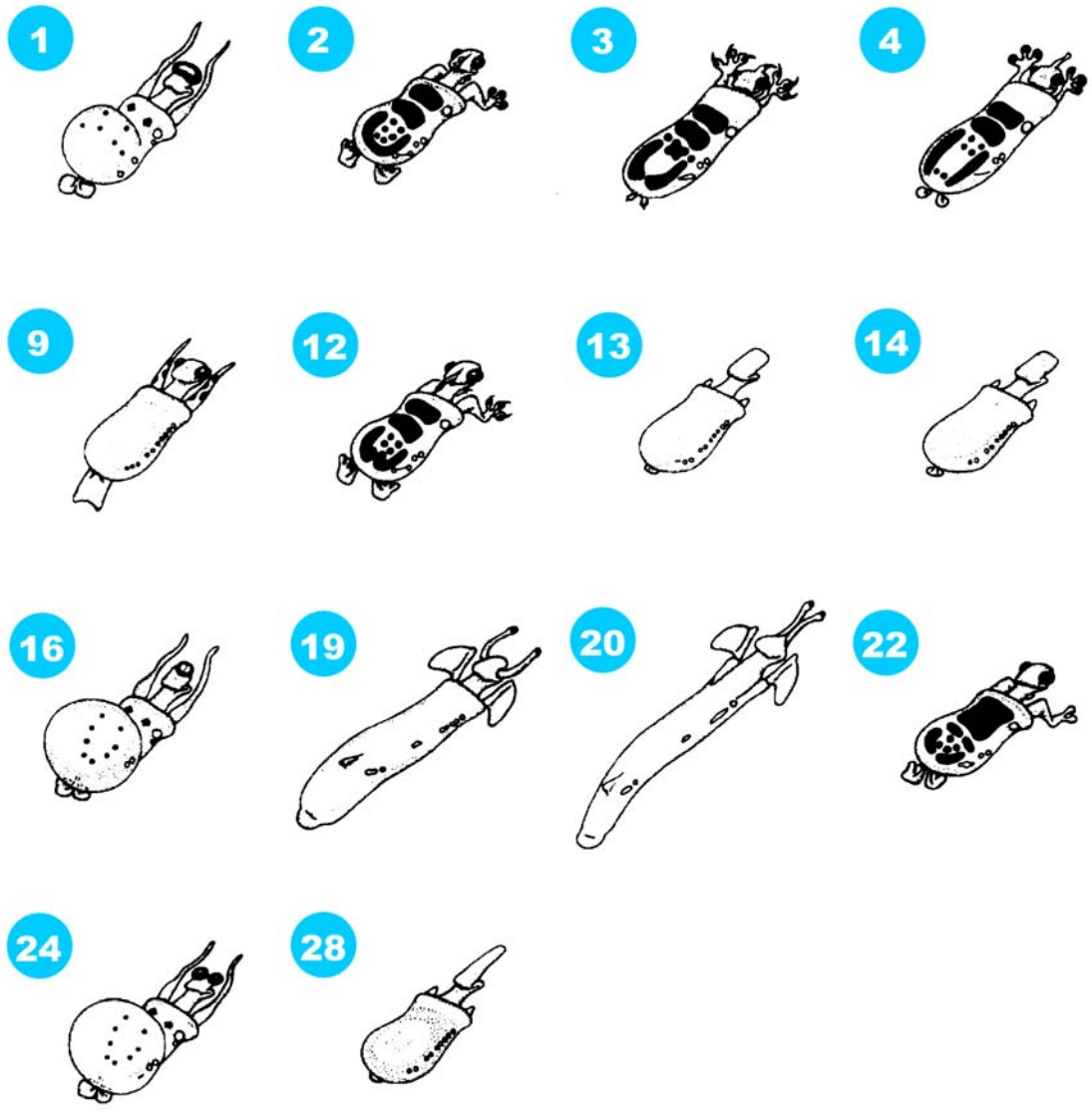
First group them by similarity:

1. Open the Mesquite software program

2. File > Open File… > browse for Caminacules_empty.nex

3. Click on Show Matrix. Characters should be scored as present or absent using the number one if present and zero if absent. Taxon 20 has been scored as the outgroup. That is, taxon 20 should be similar to the others, but not contain any characters in common with the other taxa. In contrast to the outgroup, all the other taxa in the study are referred to as the ingroup."'

4. Score the matrix from the images in the figure below. Once the matrix is complete, you can make a tree.



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

6. In the popup menu choose Treelength (then click OK). Then choose SPR rearranger (click OK).

7. Keep the number of Maxtrees at the default of 100, and on the next window that comes up, click Not Separate Threads.

8. The tree search will begin – when finished, a window will open, click Yes. There are two remaining steps to arrange the tree.

9. Choose the re-root tool:

and click anywhere on the branch leading to taxon 20 (the outgroup).

10. Next choose the ladderize clade tool

and click on the branch representing the root or common ancestor to the whole tree.

11. Click File > Save Tree as PDF. Save the file to your USB drive, print it out and place in your lab manual, along with an appropriate figure caption.

12. Does your tree seem to group similar Caminalcules together? Now, open the file for Caminalcules.nex in Mesquite. This file has a simplified character matrix scored to reflect their intended phylogenetic relationship (Sokal 1983). Use the steps above to have Mesquite make a phylogenetic tree from this matrix, then compare it to the one you made from your scoring of the characters of the Caminalcules. Were you correct??

USING COLLECTIONS TO BUILD A MORPHOLOGICAL PHYLOGENY

Biological research museums have a mission similar to that of libraries, but instead of preserving books, biological collections preserve individual organisms. It might seem that museums are static places, but in reality, collections grow continuously through activities of faculty, students, and other researchers, as well as by exchange with other museums. As they grow, collections increase in value, preserving samples of natural variation, documenting the occurrence of species in space and time, and providing the ultimate basis for our understanding of species identity. Biological collections are also the source of information regarding phylogenetic (evolutionary) relationships, whether morphological or genetic. Without this phylogenetic context, comparative biology would be impossible.

Chico State has a rich history of biological research collections, which are actively used by students, faculty, and visiting scholars for conducting investigations. We have collections of both living and preserved organisms including plants, insects, birds, mammals, and reptiles, as well as other organisms. These collections represent a great resource for biological inquiry. As you continue in your study of organismal biology, you will have opportunities to use and study these collections. For today, we have provided samples from these collections for you to examine and use to build a character matrix and phylogenetic tree.

1. Open the file [Taxon]Morph.nex (where [Taxon] is the group you are studying), and Click on Show Matrix.

The taxa available for today's lab may have already been entered. It is up to you to design characters and score them on your taxa.

2. First group your organisms by similarity, and choose an outgroup that, while related, is most dissimilar from the rest of your taxa. If you think about the characters you scored on the Caminalcules, you'll realize that some characters are shared by related taxa, and some are unique to a taxon. The best dataset from which to construct a phylogeny includes combinations of characters, some of which are similar to most groups of taxa, others of which are similar to a subset of taxa, and others that are unique. Some characters ultimately will be more valuable than others, so the more characters that are scored, the stronger the inference in phylogeny.

3. You should try to come up with about 15 characters to start with. To add more characters to the matrix file click on the Add Characters tool:

Click anywhere on the character matrix and a popup box will prompt you to enter the number of characters you wish to add. Double click on the slanted columns to add text describing each character.

Try to find a mix of characters that are (a) shared by all species, (b) shared by several species, and (c) shared by only two species. The characters you choose must be scored in a binary fashion (e.g., 1 or 0 for present or absent, respectively). Some non-binary traits can be re-scored as binary traits by defining some threshold value (e.g., greater than the threshold = 1; lower than the threshold = 0), but true binary traits are less ambiguous.

4. Once you have scored your characters, save the character matrix under a new name. Transcribe your character matrix into your lab notebook, or include a copy in your notebook, along with an appropriate table heading. **NOTE**: Save this matrix because you will need it next week! Then, using the procedure described above, make and save a tree. Remember to re-root your tree using the outgroup you've defined, as described above.

5. Print your tree, and then annotate the image with the character states that define each branch. Put a copy of this in your lab notebook, along with an appropriate figure caption that includes the "tree length" or total number of character state changes.

6. To have Mesquite indicate which characters defined the branching pattern:

Analysis:Tree >Trace Character History > Parsimony Ancestral States

Then toggle through the characters and on the tree indicate the nodes that each character supports.

7. Include a summary in your lab notebook – using correct phylogenetic terminology - that describes the inferred character evolution from your morphological phylogeny. Indicate which of the characters show homoplasy, and which appear to have a single origin.

HOMEWORK FOR NEXT WEEK:

Play CSI! (5 pts)

Download the file HIVSequenceFile.nex from Blackboard. In this file are the gene sequences of the V3 region of the HIV gp120 gene from a series of individuals. These data

are from GenBank and were used in a real court case in which an HIV positive dentist was accused of infecting his patients with HIV through improper hygienic practices. Eight patients of the dental practice (Patients A-H) tested positive for HIV during the time of receiving dental treatment.

Some of the patients, however, had additional risk factors, suggesting they may have acquired HIV outside of the dental practice.

To determine the relatedness of the strains of HIV in the dentist and in his patients, you will do a phylogenetic analysis on the following:
1) The DNA sequence of several taxa of HIV from the dentist
2) The DNA sequence of several taxa of HIV from Patients A,B,C,D,E,F,G,H
3) The DNA sequence of people in the local population with HIV (Labelled LC for "local control")
4) The DNA sequence of an African strain of HIV to use as an out group

Question: What is meant by “several taxa of HIV”?

Answer: HIV evolves at an unusually rapid rate. It evolves so fast that even within a single infected patient, new mutations can diverge into new strains of the virus. Thus, when the CDC collected these data, some individuals had more than one distinct HIV DNA sequence. In these cases, the initial strain of HIV the person was infected with had likely experienced several novel mutations. Thus, for a thorough analysis, we need to include multiple, genetically distinct strains of HIV collected from single patients. We've labeled those samples as, for example, Patient B1, Patient B2 etc...all collected from the same person (“B”), but different sequences (e.g., “1”, “2”, etc) for the HIV gene that was studied from the same sample.

You will need to do a **sequence alignment** on the data first, and then conduct a tree analysis in Mesquite. We've provided you with an outgroup in the form of an African strain of HIV. This African strain should be used to use to properly root your tree.

The sequence alignment will be conducted using a free software program called “Muscle.” To do this at home, you will need to separately download both Muscle and Mesquite (see below). These programs will work together in an integrated fashion.

When Mesquite asks for the path to Muscle (i.e., asking for the location of the program), click "BROWSE" and navigate to where Muscle has been installed. At that point, it should be linked to Mesquite. It's super easy (see next week’s lab for additional hints). Then you will have research tools for phylogenetics on your computer.

IMPORTANT NOTE: To make the analysis in Mesquite less time intensive (this is a bigger dataset than we were running in class), **instead of setting Max Trees to 100, set it to 1 max tree instead**. If you set Mesquite to the same selections as we've been using, and only 1 max tree, the analysis should take 10 to 15 minutes.

From this tree, determine which patients (if any)have HIV sequences that are closely related to the dentist's sequences, suggesting infection from the dentist, and which patients (if any)

appear to have unrelated sequences that suggest they acquired HIV from a different source potentially people in the local population.

Hand in at the beginning of lab next week:
1. The HIV phylogenetic tree from this data (1 pt)
2. A list of which patients (if any) were infected by the dentist, and which were not (1 pt)
3. A brief description of how you know that from the tree (using appropriate phylogenetic terminology) (3 pts)

You can do this assignment at home:
Download Mesquite here (free!): http://mesquiteproject.org/mesquite/mesquite.html
Download Muscle Alignment program (free!) here: http://www.drive5.com/muscle/

LITERATURE CITED

Gibson, J. P. 2018. Morphological and molecular analysis of plant diversity. Botanical Society of America, QUBES. doi: 10.25334/Q4ZH6G.

Lemke, H. and J. Jensen. 2012. Exploring systematics and phylogenetic reconstruction using biological models. Proceedings of the Association for Biology Laboratory Education 33: 123-144.