**Laboratory IV**

**Phylogenetic Reconstruction**

**Objective:** In this week’s lab you will learn how to reconstruct evolutionary relationships. Biologists have experimented with a variety of methods for interpreting who is related to whom. However, many of these methods did not well reflect the processes that we believe produce diversity. The first goal of this lab is to master the vocabulary associated with cladistics. The point is not to memorize these terms and their definitions, but to understand how they are applied. Second, you will try your hand at reconstructing evolutionary relationships in two systems, one artificial (for which true evolutionary relationships are known) and one biological.

**Note that your report on toucan barbet phylogeny will be due in lecture on 18 February.**

**Reconstructing Evolutionary Relationships**

One of the most important aspects of paleontology is establishing relationships among the organisms that we encounter in the fossil record. Interpreting ecological or biological diversity depends on incorporating information about relationships. For example, biologists have noticed a striking similarity in the snail species found on the eastern (Caribbean) and western (Pacific) coasts of Panama. Some have suggested that the snails on either side of the isthmus represent lineages derived from a common ancestor that ranged freely between Caribbean and Pacific before land connected North and South America during the Pliocene. Alternatively, they could have recently dispersed, one species at a time, through the Panama Canal. Still another possibility is that these snails are not really closely related and just look similar because they occupy similar environments. Each of these hypotheses generates a different predicted pattern of relatedness that we can compare to a phylogeny of the snails themselves. Thus, we can actually test these hypotheses directly.

**Systematics** is the science of describing the relationships among organisms and the processes producing these relationships. It includes (but is sometimes incorrectly equated to) **taxonomy**, the rules for description and naming of groups of organisms, and **classification**, the organization of those named groups.

Since the ancient Greeks founded systematics\(^1\), overall or **phenetic** similarity has been the main basis for systems of classification of living things. Carl Linné’s (Linnaeus’) hierarchical system and binomial nomenclature was meant to reflect the natural order of God’s creation, as manifested in overall morphological similarity. Like most people of his time, Linnaeus believed that all species had been constant since their creation, although some might have become extinct. Even he recognized that the grouping criteria he used united many taxa that clearly did not belong together as well as split up morphologically coherent groups of species. Nonetheless, Linnaeus, and generations of biologists since, have accepted the system because it satisfies a practical, descriptive need.

After the acceptance of evolution through the work of natural scientists, systematists realized that they needed to incorporate knowledge about the evolutionary relationships among organisms into schemes of classification. For example, each distinguishable group of organisms

---

\(^1\) Actually, you might say taxonomy is the oldest profession. After all, in the Judeo–Christian tradition, God gave Adam the job of naming all the animals (Genesis 2:19–20).
should have a single common origin (be monophyletic). Under this principle, the classification doesn’t simply reflect overall similarity, but evolutionary history. Under such a system, we can use evolutionary groups to ask evolutionary questions.

**Evolutionary Systematics** Evolutionary systematics in the tradition of Ernst Mayr and George G. Simpson was practiced by most taxonomists of the 20th Century. In this school of thought, classification reflected relatedness as well as morphological similarity. The introduction to a lineage of a major new trait (apomorphy, e.g., flowers in the angiosperm lineage, or wings to birds) therefore results in the formation of a new so called "natural group" (a group that reflects relationships of descent). For example, the class "Reptilia" was thought to have evolved from the class "Amphibia" by the invention of the amniote egg. Unfortunately, this classification lead to the creation of a paraphyletic taxon (one that do not encompass all the descendants of their common ancestor), or evolutionary grades (reptiles) instead of truly monophyletic taxa.

This early version of evolutionary systematics was based on a very imprecise, subjective, and complicated set of rules that only scientists with lots of experience working with their organisms were able to use. The resulting phylogenies became impossible to reproduce other than by the specialists themselves. This practice of systematics was more art than science, and led to a call for more repeatable and objective methods.

**Phenetics** Numerical taxonomy or phenetic systematics sought a more empirical and objective way of generating and choosing among phylogenies. This method was made possible by the advent of computers in the 1960’s because phenetics relies on extensive mathematical calculations. Organisms are grouped by a mathematical analysis of a large set of characters (preferably all possible characters!) according to overall similarity. The results are certainly reproducible, but fail to create groups that reflect evolutionary relationships. Phenetics failed to create truly evolutionary groups because organisms can be similar because they live in similar environments or because they make their living in similar ways, not just because they are descended from a common ancestor. The main advocates of phenetic systematics are James Rohlf, Robert Sokal, and Peter Sneath.
Phenetics broke the log–jam of systematic methodology, opening biologists to the idea of considering characters in an objective and reproducible way. However, the overall–similarity methods failed to reconstruct evolutionary groups, so new methods were sought.

Despite the failure of phenetics to reconstruct evolutionary relationships, the numerical methods themselves have applications to a variety of other problems, such as delimiting variation within and among populations of individuals. This approach can be particularly useful to paleontologists as we try to circumscribe species with their associated ranges of variation from collections of individual specimens. Often observation alone leads the paleontologist to focus on just one or two characters because it is difficult for our brains to juggle all possible characters at once. The numerical or morphometric methods use the power of the computer to consider all characters together in recognizing groups within a collection or a gradation between dissimilar end members.

**Cladistics** Willi Hennig developed a systematic methodology that emphasized:

- objectiveness and reproducibility (to be more objective than the opinions of individual taxonomists), and

- accordance with evolutionary patterns.

Hennig’s system, phylogenetic systematics or cladistics, is now the standard method of phylogenetic inference among evolutionary biologists and paleobiologists. Just like phenetics, cladistics takes advantage of the availability of electronic computing power to analyze large data sets. However, the cladistic algorithms think about characters in a way that is more consistent with our ideas about how evolution works.

The most important insight of cladistics is that if you consider all character states shared by a
number of organisms (i.e., if you look at overall similarity), you will not necessarily get a
classification that reflects actual evolutionary relationships. Instead you need to concentrate on
only certain characters—those that provide evolutionary information.

Hennig defined a few terms to describe the distinction between his approach and others. The
term **apomorphy** means a specialized or derived character state; **plesiomorphy** refers to a
primitive or ancestral trait. An **autapomorphy** is a derived trait that is unique to one group,
while a **synapomorphy** is a derived trait shared by two or more groups. A **sympleiomorphy** is
similarly a shared primitive trait. These terms are defined relative to a particular node (e.g.,
representing a taxonomic level) on the cladogram. This means that a trait can be a
synapomorphy and a sympleiomorphy if different nodes are considered. For example: The
multicellular sporophyte (in plants, having a multicellular diploid body) is a synapomorphy of
the land plants (*clade b*) but a sympleiomorphy for the vascular plants (*clade c*).

![Cladogram](image)

Cladistic vocabulary. At node A, “multicellular sporophyte” is an autapomorphy for clade b (land
plants). At node B (liverworts) and clade c (vascular plants), “multicellular sporophyte” is a
synapomorphy. At node C, this trait is a sympleiomorphy for conifers and angiosperms.

Since current evolutionary theory says that traits arise (are derived) in lineages through
evolution, only synapomorphies give us information about evolutionary relationships.
Autapomorphies contain no information about relationships (because they don’t group organisms
together); sympleiomorphies should not be used to unite taxa because organisms share them.
because of who their ancestors are. For example, having five digits on each hand and foot is a symplesiomorphy for primates; all primates have this trait because the common ancestor of all mammals had this trait. Thus, five digits won’t help you group primates.

Organisms that are united by one or more synapomorphies share a common ancestor which possessed these derived traits. They belong to a **monophyletic** group in which all descendants of the common ancestor have to be included. This is also referred to as a "natural" or "evolutionary" group or as a **lineage**. In modern evolutionary biology, we work hard to recognize only monophyletic groups. If a group does not include all the descendants of a common ancestor, the group is termed **paraphyletic**, or a grade. An example of this is the Reptilia (discussed above), which includes crocodiles, lizards and dinosaurs, but not birds. If the group includes some or all of the descendants, but not the common ancestor, it is called **polyphyletic**. For example, a group of all flying vertebrates, regardless of their ancestry, would be extremely polyphyletic. A **sister group** (or sister taxon) is defined as the closest relative to a monophyletic group, as determined by one or more synapomorphies uniting the groups.

![Cladistic Diagram](image)

More cladistic vocabulary. (A) Monophyletic clades; (B) paraphyletic clade; (C) a polyphyletic clade.

**Cladistic Method**

**Choosing taxa** The first step in a phylogenetic analysis is choosing the taxa (called operational taxonomic units or OTUs) you want to study. For the purposes of this lab, you will be given a set of OTUs with which to work. In original research, you might want each OTU to be a species. In this case, you have to decide what constitutes a species and which specimens belong in which species. Confronted with a very diverse group, you might need to decide which species to include and which to leave out. Most of these are practical decisions, but they may ultimately influence your results.

**Choosing characters and character states** Next, you need to choose a set of evolutionarily meaningful characters. So how do you know what characters are synapomorphies and "good" characters for phylogenetic analysis? First, a good character shows greater variation among the taxa you are interested in than within each OTU. Second, this variation must be heritable and independent of other characters. For example, the character "has a chitinous exoskeleton" and "grows by molting" are not independent because the only way an organism with a chitinous
exoskeleton can grow is by molting. Third, you must make sure that the characters and character states that you are examining are truly comparable, that is, homologous. The term homology was first introduced by Sir Richard Owen in 1843. The word is derived from "homologia" in Greek which means "agreement". Homology denotes structures and organs that have evolutionary correspondence, regardless of their current function. It can be recognized based on three criteria of homology:

– correspondence in position and details in structure
– correspondence in developmental origin
– an evolutionary series of character states with no significant breaks (from the plesiomorphy in the ancestor to the apomorphy in the descendant), so called transformational homology

Cladistic assumptions:

(A) replication (organisms reproduce);
(B) heritability (variation correlated between parent and offspring), giving rise to transformational homology;
(C) evolution occurs by splitting or divergence (at least as a predominating pattern), giving rise to taxic homology or homology between sister groups.

Correspondence in structure between sister groups is referred to as taxic homology. The wings of birds, forelimbs of a lizard and human arms are (taxic) homologies, because they are all derived from the same primitive structure in the common ancestor of these groups. Similarly, the fronds of ferns, the leaves of flowering plants and the needles of pines are also taxic homologies. Synapomorphies must always be taxic homologies.

The term homoplasy was coined by Lankester in 1870. It refers to analogous structures, structures that show similarity and may perform the same function, but that are not derived from a structure found in a common ancestor. The wings of bats and insects are analogous (homoplasic) because they both function for flight, but evolved from different primitive structures. Homoplasy is due to convergent evolution, parallel evolution or character reversal.

A character is any recognizable attribute of an organism, such as "eye color" or "presence of leaves". A character state is the value of the character, for example "blue" and "yes", respectively. Note that characters that are related to some function are more likely to be homoplastic or convergent (e.g., size and shape of leaf) because natural selection may exert substantial influence toward some really functional design.

The character states are entered into a matrix as for example "1" for "present" and "0" for "absent", or "1" for "green", "2" for "blue" and "3" for "red". For many characters it is possible
to hypothesize an evolutionary order— which character state must follow another in an evolutionary character transformation series. However, it is more difficult to decide on the polarity: the actual direction of evolutionary change. If you are able to do this, it becomes much easier to decide what organisms are basal (that is, near the root of the cladogram). The polarity of characters can be deduced by looking at a variety of lines of evidence: ontogenetic precedence (a character state present early in development but subsequently lost), stratigraphic character precedence (a character state present in organisms of the same lineage earlier in Earth’s history), and outgroup comparison (comparison with proposed sister taxa).

**Assembling the data matrix** Once taxa and characters are selected, you assemble your matrix for analysis. Each OTU occupies a ROW in your data matrix; each character is given a COLUMN. The character state for each character and for each OTU is then scored to fill up your data matrix.

**Analyzing the data matrix** The data matrix can then be analyzed by any of several computer programs that use an algorithm that joins OTUs based on the greatest number of shared derived characters (synapomorphies). The algorithm that is most appropriate for morphological data and that you are going to experiment with uses as a grouping principle maximum parsimony. Parsimony ("Okkham’s razor") is based on three assumptions about evolution (see also figure above):

1. Organisms reproduce and transmit characters to their offspring
2. Character states and character state changes are inherited
3. Branching occurs, and is the predominant pattern (i.e. it happens at a relatively high rate as compared to the rate of changes in character states in a single lineage)

Consequently, similarity in a derived character state is more likely to be due to a common ancestry (homology) than through random or non–random assimilation of the same feature in two different lineages (homoplasy). In the program you will use (PAUP = Phylogenetic Analysis Using Parsimony), the algorithm Maximum Parsimony chooses the phylogeny that gives the minimum amount of "steps" or character state changes. This follows Okkham’s razor in saying that the simplest solution is the most probable.

As you are no doubt realizing, the cladistic method is not entirely objective. The systematist still has to make interpretations about homology, choose characters and make decisions about weighting and ordering characters for analysis. However, the cladistic paradigm does force the systematist to state explicitly why he or she believes characters are homologous or should be ordered in a certain way. This provides bases for falsifying hypotheses and promotes testing of alternatives. This transforms the subjectivity into science.

**Interpreting results** The result of a cladistic analysis is a cladogram, a branching diagram of nested synapomorphies that defines relationships in a relative way. These synapomorphies are used to recognize monophyletic clades (monophyletic groups of organisms of any taxonomic rank), arranged in a hierarchical manner. Note that a cladogram does not have a time axis, and does not make any statements as to the mechanism of evolution other than it occurs by splitting of ancestral lineages. It is merely a hypothesis of character change and relationships. As such it
can be carefully compared—character by character—to competing hypotheses, tested, and refuted with additional data.

What if I get more than one most parsimonious tree? If you have a lot of characters and OTUs, there are often a number of ways in which the algorithm can organize a most parsimonious tree. As a result, you might end up with more than one most parsimonious hypothesis of relationships. In order to choose the most likely cladogram, you might try adding more characters. You might use character weighting, letting some characters have a larger influence on how the phylogeny comes out. Many scientists are conservative and only report those relationships that show up in all the most parsimonious solutions. This is called strict consensus.

**Phylogenetic Exercise – Comparing Cladograms to Evolutionary Trees**

As noted above, cladograms aren’t evolutionary trees. They do not show ancestor–descendent relationships. Rather, they present a graphical hypothesis of branching relationships based on the distribution of synapomorphies. To help you better understand this point, we will be looking at a make–believe example. The advantage of this example is that the evolutionary history will be known. Therefore, after you have completed the analysis, you can compare your cladogram to the actual evolutionary tree.

Work in groups of two or three for this exercise.

**Method**

1. **Meet the Ovids.** Ovids are mythical creatures. You don’t need to know too much about their ecology and life history because your phylogeny will focus only on morphological characters. Ovids have evolved by branching of lineages, with new traits acquired at the branching events. Thus, they are good candidates for cladistic analysis. I have already selected OTUs and examples of each are presented on your handout.

2. **Identify characters.** First you must identify morphological characters. You must have more characters than OTUs and try to identify at least two characters for each ingroup OTU. If you can think of more characters, include them. That will make your analysis stronger. Consider the rules for identifying good characters discussed above. For each character identify two character states. One of the states will be coded "0" in the data matrix. By convention, the zero state is the state possessed by the outgroup. The alternative (derived) state will be coded "1". Describe the characters and the character states for each on your Ovids worksheet.

3. **Complete the data matrix.** For each character, note the character state possessed by each OTU, thus completing the data matrix. Now enter your data into MacClade. Start MacClade by double–clicking on the MacClade icon. Select "new" when asked what file you want to open. Use your cursor to drag the number of columns to match the characters you have, and the rows to seven. Enter your data. Save your matrix with a unique file name and place it in the class folder.

4. **Analyze your data.** Start the application PAUP by double–clicking on PAUP icon. When asked which file to execute, choose you named file and click "execute". In the "analysis" menu, choose "exhaustive search". Click "search". The search should only take a moment.
Now, root your tree by telling PAUP which taxon is the outgroup. In the "tree" menu, select "root trees". Click the "rooting options" in the dialogue box that appears. Click the "define outgroup" button in the dialogue box that appears. Select your outgroup from the ingroup box and move it to the outgroup box. Click "ok"; click "ok" again; click "root".

Save your tree. In the "tree" menu, select "save trees to file?" Give the tree file a unique name and place it in the class folder.

5. **Study your trees.** Return to MacClade. Open your data file again. In the "display" menu, select "go to tree window". Select "open tree file" and choose your saved tree. Click "open"; click "get tree". Your tree should now be displayed. Copy the tree topology (branching pattern) into your notebook and then onto the blackboard. When everyone is finished, we’ll have a look at the real evolutionary relationships among the *Ovids* and compare that to our phylogenies.

**Phylogenetic Exercise – Toucan Barbet Phylogeny**

In this exercise we will be working with a real example, where the true relationships are not known. For this analysis, your characters will focus on the color and markings of 13 species of tropical birds, the toucan barbets. Plumage color in birds evolves very easily and can become very complex. Because birds have excellent color vision, color can be very important to them. Thus, color and color patterns make good evolutionary characters.

Work in a different group of two or three for this exercise.

**Method**

1. **Meet the toucan barbets.** Toucan barbets are a group of Neotropical birds. They live in forested habitats in both lowland and montane environments. They eat a variety of fruit, seeds and insects. Your OTUs will be the established species of toucan barbets shown on the sheet.

2. **Identify characters.** Identify at least two morphological characters for each OTU. Of course, your analysis will be stronger if you have more characters. Consider the rules for identifying good characters discussed above. For each character identify two character states (0, 1). Describe the characters and the character states for each on your toucan barbet worksheet.

   The pictures from which you will be developing and scoring your characters are cartoons. They are designed to steer you to three types of characters: beak shape, color and markings. Consider only these types of characters in your analysis.

3. **Complete the data matrix.** For each character, note the character state possessed by each OTU, thus completing the data matrix. Now enter your data into MacClade as before. Select "new" when asked what file you want to open. Use your cursor to drag the number of columns to match the characters you have, and the rows to seven. Type your data. Save your matrix with a unique file name and place it in the appropriate class folder.

4. **Analyze your data.** Start the application PAUP by double-clicking on PAUP icon. When asked which file to execute, choose you named file and click "execute". In the "analysis" menu, choose "heuristic search". Click "search". The search should not take too long, however, the
amount of time required for the search will depend on the number of characters you have generated. We will not be rooting trees in this exercise.

Save your tree. In the "tree" menu, select "save trees to file?" Give the tree file a unique name and place it in the class folder.

5. **Study your trees.** Return to MacClade. Open your data file again. In the "display" menu, select "go to tree window". Select "open tree file" and choose your saved tree. Click "open"; click "get tree". Your tree should now be displayed. Copy the tree topology (branching pattern) into your notebook, note the file name of your tree on the "stickie" on the computer desktop. Prof. Arens will collect these and print them out on Monday morning so that you may include the tree in your report.

6. **Write up your results.** A short report on your analysis will be **due in lecture** on Monday 18 February. Your report should be co–authored by the members of the group, typed, and contain the following elements.

   – A list of the taxa you studied
   – A list and description of the characters and their character states
   – Your character matrix
   – A copy of your tree
   – A brief discussion interpreting your tree. Consider the following questions in your discussion:

     1. What monophyletic clades did you identify?
     2. Are named genera monophyletic?
     3. How might you change the classification (taxonomy) of the toucan barbets to reflect natural groups?

---

2 Write up original descriptions of your characters and a new copy of your data matrix. Your character and matrix worksheet will not be accepted for this character description or data matrix.

3 A printed copy of your tree will be available before lecture on Monday. You may include this in your report.
Suggested Reading

If you are interested in learning more about how phylogenies are constructed or how they can be used to answer evolutionary questions, some of these resources may help.


Questions for Further Thought

1. Do you see any homoplasy in the true evolutionary relationships of the *Ovids*? If so, what characters evolved more than once? How did your phylogenetic analysis handle this problem? What might have helped your cladistic analysis more accurately replicate the true evolutionary relationships of the *Ovids*?

2. What was the most difficult aspect of developing the toucan barbet cladogram?

3. Cladograms are only hypotheses about relationships. Based on your experience, where do you see the most potential for error (producing cladograms that do not reflect the true evolutionary relationships) in these analyses?

4. Cladistics was originally touted as helping eliminate the subjectivity from the reconstruction of phylogenetic relationships. Does the method succeed completely? If not, where do you see the most subjectivity creeping in, based on your experience with this exercise?