Biology 4415/5415
Evolution

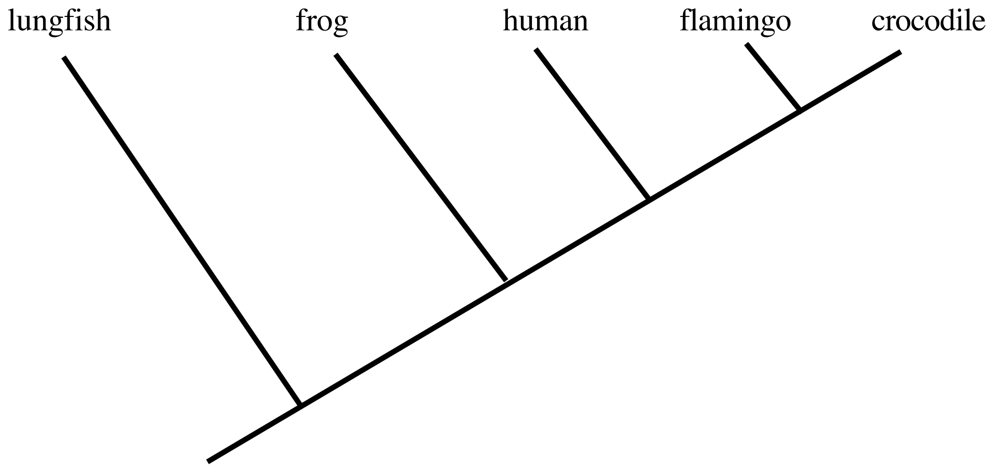
LABORATORY EXERCISE 6:

CLADISTICS I

Take a group of organisms. Let's use five: a **lungfish**, a **frog**, a **crocodile**, a **flamingo**, and a **human.** How to reconstruct their relationships? To use the method known as **cladistics**, we start by drawing up a list of **characters**, or features that may or may not be present in each organism. We then determine the **character states** for each organism, and draw up a matrix that looks like this:

	lung	vertebrae	developed limbs	amniotic membranes	scales or feathers	skull openings
lungfish	present	present	absent	absent	absent	none
frog	present	present	present	absent	absent	none
crocodile	present	present	present	present	present	two
flamingo	present	present	present	present	present	two
human	present	present	present	present	absent	one

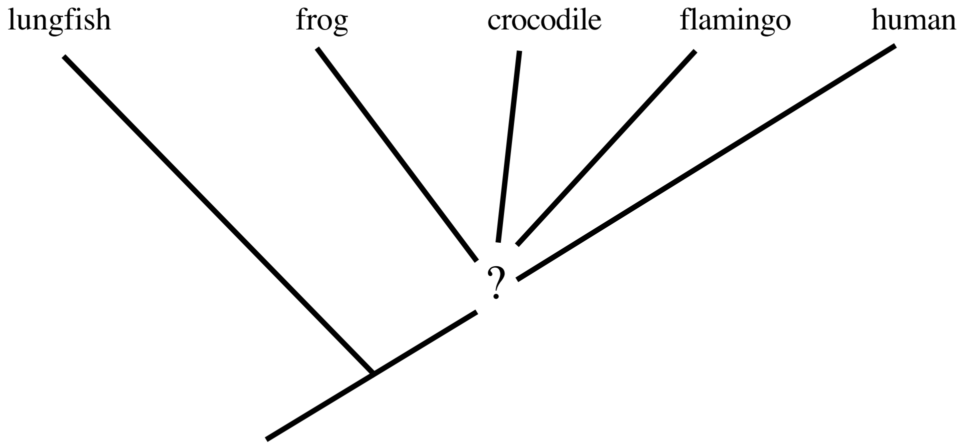
A **cladogram** is a branching, tree-like diagram that presents a hypothesis of common evolutionary ancestry. The taxa actually being studied in a cladistic analysis, the **terminal taxa**, lie at the tips of the branches. The points where the branches join (the **nodes**) represent **inferred common ancestors**. So a cladogram like this



simply means : "I hypothesize that flamingos and crocodiles, and their close relatives, had a common ancestor which was not an ancestor of humans, frogs, or lungfish. . . and flamingos, crocs and humans had a common ancestor, older than the croc-flamingo common ancestor, which was not an ancestor of frogs or lungfish. . ." and so on. This may or may not be true—but it's testable. But how? Most of the time you can't dig up The

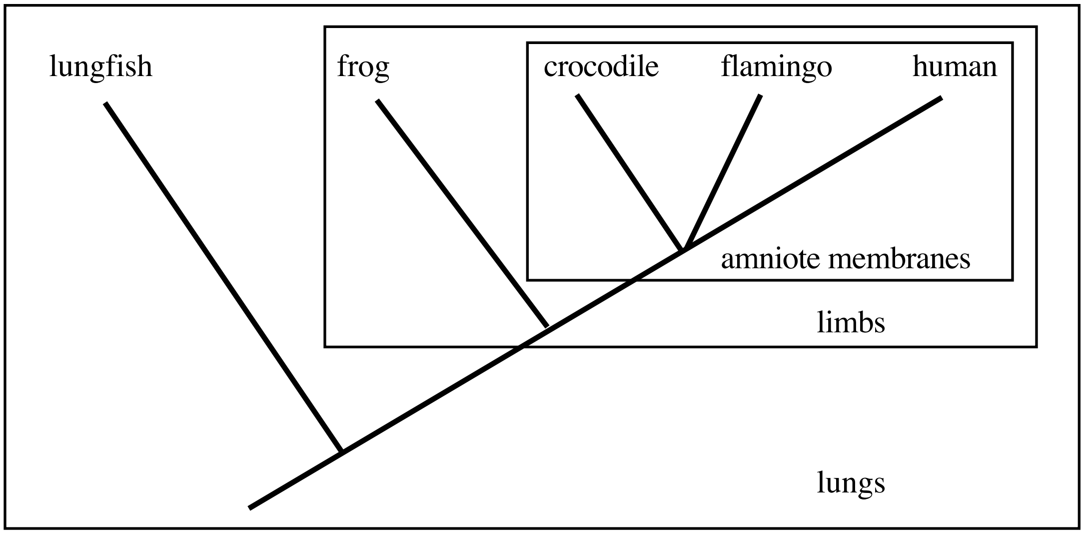
Common Ancestor as a fossil, and even if you did, you'd have a hard time making a convincing case that it was The Common Ancestor.

To draw a cladogram, you first have to make at least one taxon your **outgroup**, and the remaining taxa your **ingroup**—that is, you start with the assumption that all members of the ingroup are closer to each other than any is to the outgroup. Your outgroup need not necessarily be *the* ancestor, but it should reflect the presumed primitive character states. Here, the lungfish is a reasonable outgroup. It's a vertebrate, but it lacks characters that the other four taxa have, and we know from fossils that lungfish appeared on Earth before modern amphibians, reptiles, birds, and mammals. So at this stage, the tree looks like this:

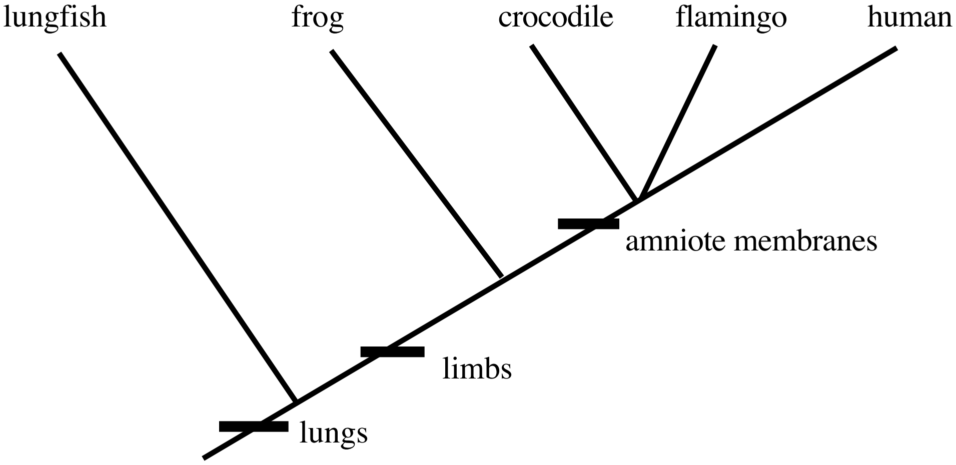


Notice something: All five of these taxa have lungs, so that character doesn't tell you anything about how they're related. Presumably, lungs were present in the common ancestor of all these taxa. The character "lungs," therefore, is a **plesiomorphy**. Other characters, however, are not shared by all taxa. They are potential **synapomorphies**—shared, derived characters. For instance, amniote membranes are found in three taxa—crocodile, flamingo, and human. We can hypothesize that amniote membranes evolved only once, in a common ancestor that was **not** ancestral to the frog. Similarly, legs—true legs, that is, not lobe-fins—are found in all taxa but the lungfish. We may reasonably assume they were inherited from a common ancestor that was not also ancestral to the lungfish. They could have evolved convergently—but the **principle of parsimony**, discussed further below, states that you should not assume convergent evolution without good reason. The limbs of land vertebrates are so anatomically similar in detail that we have no good reason to assume convergence.

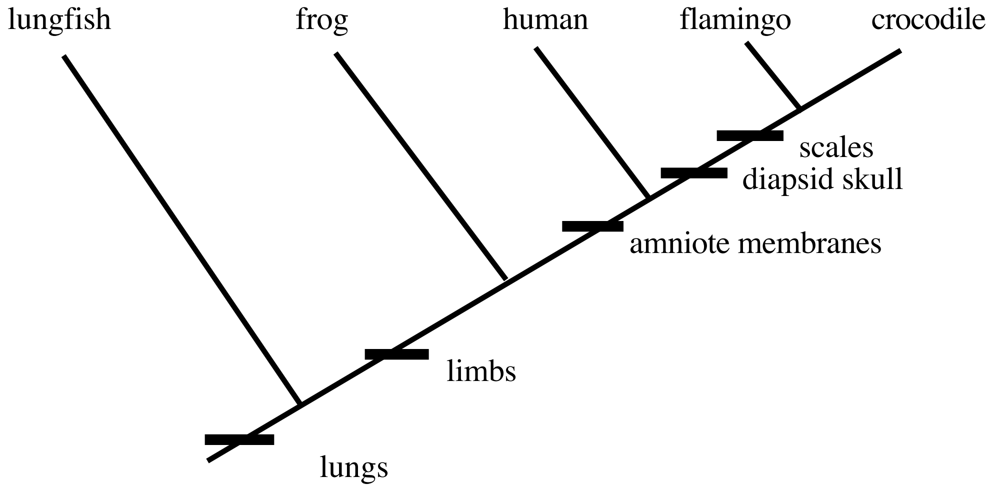
We could map the three characters mentioned so far onto our tree, which would give something like a Venn diagram:



But that’s a bit cumbersome to draw, so we usually just use tick marks to show that a character is shared by all taxa that lie above the level of the tick mark.

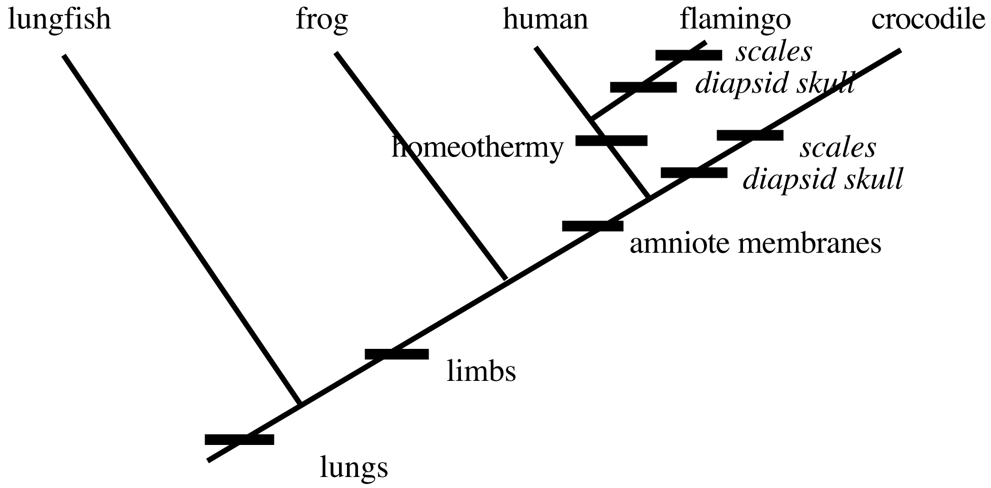


Let's look more closely at the croc-bird-human group. In our original matrix, there were two characters that supported grouping crocs and birds: scales / feathers (both are derived from common embryonic structures) and two openings in the roof of the skull through which muscles pass (a **diapsid** skull). That would make the tree look like this:

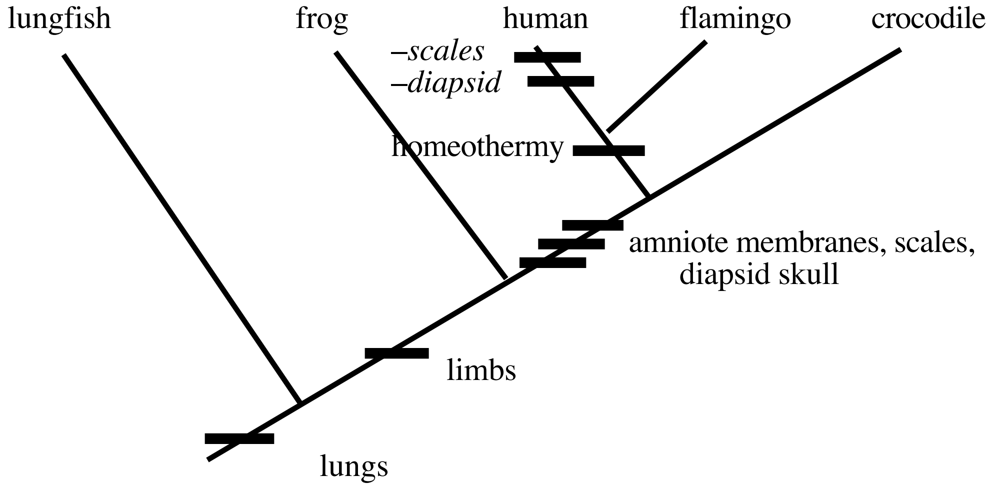


But you know that there are other characters that birds and mammals have in common. Let's add one to the list: endothermic homeothermy ("warm-bloodedness"). All of a sudden, the picture is less clear: there is no way that these characters could have the distribution that they do without *either* secondary loss *or* convergent evolution *somewhere* on the tree. (Both are technically called **homoplasy** by cladists; they'll be indicated by italics on these trees.)

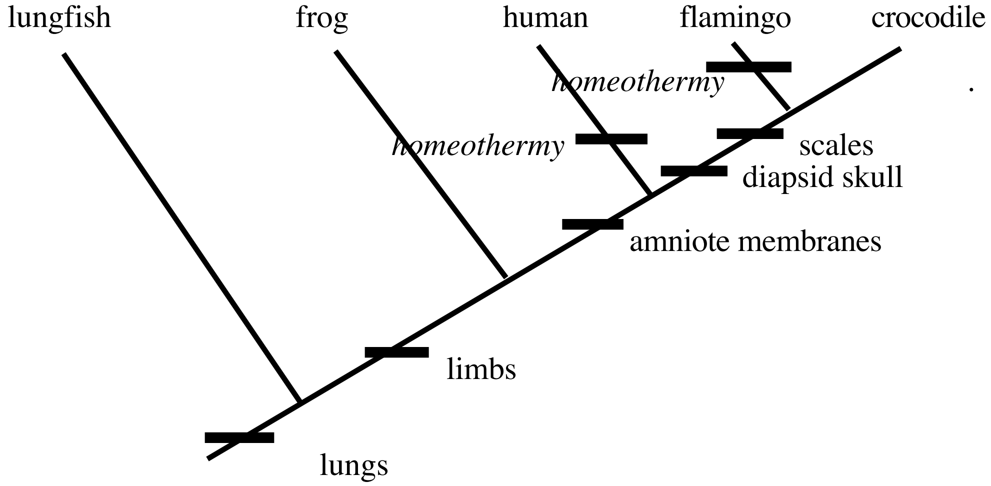
There are several possiblities. Either humans and flamingos shared a common ancestor that was homeothermic, and flamingos and crocs convergently evolved the diapsid skull and scales:



A second alternative is that the common ancestor of humans, crocs, and flamingos had homeothermy, scales and a diapsid skull, and the last two character states were secondarily lost in humans (shown using the minus sign):



A third alternative is that humans and flamingos—well, mammals and birds—evolved homeothermy independently:



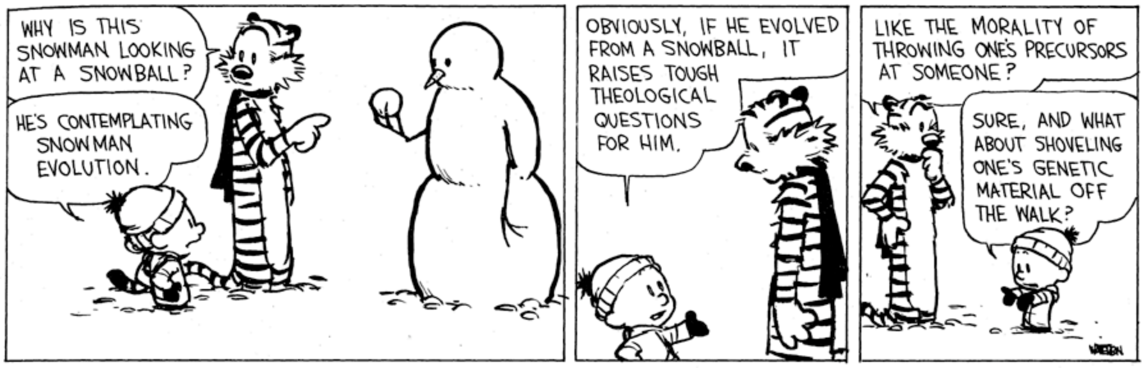
Which is correct? Basic to cladistics is the **principle of parsimony.** All that this states is that, given a choice of evolutionary hypotheses, the simplest one is preferable—until and unless there is good reason to pick a more complex hypothesis. The "complexity" of a cladogram, in this sense, is its **length**—the number of character state changes on the tree. Count the number of tick marks: the first and second trees have eight character state changes, including two cases of convergence. The third tree, however, has only seven changes, including only one case of convergence. We say it's the **shortest tree.** A cladist would prefer the third, shortest tree—until and unless some new data forced her to pick another.

That's how a cladogram is a testable hypothesis. You can always expand the data you use, whether with new characters or with new taxa or both, to keep testing and refining your evolutionary hypothesis. For example: crocs and birds have other characters in common, such as an open space in the skull in front of the eye (the **antorbital fenestra**). That gives further support to the crocodile-bird grouping. If a hypothesized pattern of relationships just doesn't fit your data, it's falsified.

By the way: A character state that is found in only one taxon isn't helpful in reconstructing ancestry. The fact that humans have hair, or that frogs have highly modified pelvises and spinal columns, tells us nothing about their relationship to crocs and flamingos. Such a character is called an **autapomorphy**, and would not usually be included in an analysis.

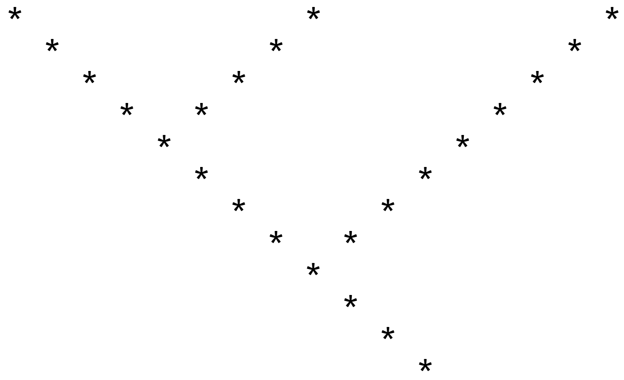
In the short example I've given, finding the **shortest** tree—the one with the fewest cases of reversal or convergence—is straightforward. But the number of possible trees that have to be sifted through rises rapidly with the number of taxa: in fact, it's a function of the factorial of the number of taxa. If you had ten taxa, there would be about two million possible trees that you'd have to search to be sure of finding the shortest. If you had eleven, there would be over twenty million. Now you see why working cladists use computers extensively to find the shortest trees. Even then, a large data set may be impossible to handle directly (I've had more software crashes than I like to think about!). There are several computational "shortcuts" for finding the shortest trees, but we need not go into those here. In this lab, the datasets will be small enough that you should be able to find the shortest trees more or less by trial and error without too much difficulty.

IMPORTANT: The shortest tree or trees—the **most parsimonious** trees, as cladists say—are not necessarily the *correct* ones, the ones that shows the true evolutionary history of the taxa. In fact, it often happens that there is more than one most parsimonious tree; obviously, only one can be correct. But you have to have some way of whittling down the huge number of possible hypotheses to a manageable set of plausible ones; otherwise, you'd never get anywhere. Parsimony is not perfect, and in fact there are other ways to draw up cladograms which sometimes perform better. But parsimony is a reasonable criterion in most cases.



MATERIALS
set of real organisms
notepads and pens

1. Stand in a “tree” pattern similar to the one shown here, as seen from above.



Space yourselves out so that you are about arm’s length from each other. Your instructor will tell you exactlly how to arrange yourselves, and may also assign you different arrangements to stand in.

2. The instructor will whisper a sentence to the person standing at the base of the “tree.” Each student then repeats it to the next one down the “tree.” A student standing at a branching point must pass the phrase on to both of her neighbors down the tree.

Please do not willfully garble or “improve on” the phrases you hear; repeat what you hear as honestly and clearly as possible. Don’t be shy or overly conscientious; you *will* make errors in passing the sentences on, so don’t be embarrassed—errors in transmission are what this lab is all about!

3. As the sentences reach the tips of the tree, the students at the tips should write their versions down on the notepads provided.

4. We will repeat this exercise several times, rearranging the students each time. At the end, we will discuss how the sentences you were repeating have changed over time. Can we identify synapomorphies, autapomorphies, and plesiomorphies? Does the pattern of similarities match the real “evolutionary” history of these sentences?

5. Once you’ve practiced with this method, you will be given a set of five or six real organisms (marine snail shells). Observe them carefully, look for characters they have in common, and draw up a character matrix. Use that matrix to create a cladogram, and then identify characters as being synapomorphies, autapomorphies, plesiomorphies, or homoplasies.

6. In the full lab write-up that you turn in, remember to include:

1) your cladogram of the snail shells, each with the complete data matrix attached, and identification of which characters are synapomorphies, plesiomorphies, autapomorphies, or convergences;

2) a discussion of the cladogram, stating your conclusions, any problems you had, ways the cladograms could be tested further, and so on;

3) a discussion of what the first part of the exercise implies about the assumptions of cladistics—do they work out in practice?

Embarrassing moments at gene parties