**Phylogenetics Lab**

by

Tom Raffel (part 1)

and

Douglas Wendell (part 2)

**Lab Goals**

In this lab you will learn to use cladistics to determine evolutionary relationships between species. You will work with both morphological data and DNA sequence data.

**I. Phylogenetics:**

The field of phylogenetics is concerned with determining how different organisms are related to each other. Phylogenetic hypotheses can be displayed by the use of a phylogeny, a graphical model in which organisms are linked together by a branching pattern. A standard type of phylogenetic “tree” is shown below. By convention, extant (living) organisms are placed on either the top or right side of the phylogeny, with branches into the past extending down or to the left.

Species 1

Species 2

Species 3

Species 4

Species 4

Species 3

Species 2

Species 1

**II. Cladistics:**

Cladistics is a method for placing organisms in related groups, or clades, by minimizing the total number of evolutionary steps needed to explain the overall tree of relatedness. According to the principle of parsimony, the simplest explanation tends to be the correct one. Therefore the parsimonious tree is the one requiring the fewest evolutionary steps. For example, the phylogenetic tree presented in Figure A is simpler than the tree in Figure B, since Trait 1 only needs to evolve once.

Trait 1

**A.** Species 1 **B.** Species 1

Trait 1

Trait 2

Trait 2

Trait 1

Species 2 Species 2

Species 3 Species 3

Species 4 Species 4

The traditional method for constructing a phylogeny is to group organisms according to which ones share the greatest numbers of derived characteristics. A derived characteristic is one that has changed from the ancestral state. For example, humans habitually walk upright on two legs (bipedalism), but our primate ancestors generally used all four limbs for locomotion. Therefore bipedalism is a “derived” characteristic of humans, relative to other primates.

If two organisms share a characteristic in common, the simplest explanation is often that their common ancestor had that characteristic. This type of similarity is called “homology”, and we would refer to the shared characteristic as a “homologous trait”. Alternatively, the two organisms might have evolved the same characteristic independently, possibly for similar reasons. This type of similarity is called “homoplasy” or “analogy”, and we refer to the shared characteristic as an “analogous trait”. For example, ostriches and humans both have four limbs (2 wings/arms and 2 legs) because our last common reptilian ancestor had four limbs, making this character homologous. However, ostriches and humans evolved bipedal walking independently of each other, making this character analogous. When we create a cladogram, we attempt to maximize the number of apparent homologies and minimize the number of apparent homoplasies needed to explain the overall tree of relatedness.

Robust cladistics analysis requires selection of an appropriate outgroup species, which is used as a control. The outgroup should be related to, but outside of, the group of organisms being analyzed. An outgroup is useful for determining whether characteristics of the focal species are “ancestral” (present in the common ancestor) or “derived” (recently evolved in most of the organisms). Characteristics found in both the focal group and the outgroup are probably ancestral, whereas characteristics found only in the focal group are probably derived. Generally, characteristics should only be considered derived if they are absent in the outgroup.

**Rules for creating a phylogeny using cladistics analysis:**

* Organisms should be listed on the right with evolutionary relationships extending back in time to the left.
* The goal of Cladistics is to minimize the total number of transitions to create parsimonious phylogeny. The simplest strategy to finding the parsimonious phylogeny is to work backwards, starting with species pairs that share more characters with each other (and presumably diverged more recently).
* As you add each species to the phylogeny, note which derived characters are possessed by that organism. This will help you keep track of the transitions needed to achieve that organism’s set of characteristics.
* Label all evolutionary transitions with letters from Table 1. Transitions can include either evolution of a new characteristic (e.g., “+A”) or the loss of a previously evolved characteristic (e.g., “-A”).

**WEEK 1:. Phylogenetic Analysis Using Morphological Data**

#1. Examine a flatworm specimen, which we will use as the outgroup for this analysis. **Does it possess any of the characteristics listed in Table 1 (on the following page)?**

**#2.** Examine specimens of the other 12 species, and **indicate whether each possesses any of the derived characteristics listed in Table 1.** Use a “+” for the presence of a character and a “-“ for its absence. Ask your instructor if you are uncertain whether an organism possesses a particular characteristic.

**#3.** Count the number of DERIVED characteristics (both “+”) shared by each pair of species, and **record the number in the corresponding cell of Table 2.**

**#4.** In the space below, list pairs of species that share at least 5 derived characteristics.

**#5.** In the space below, list pairs of species that share at least 4 derived characteristics.

**#6.** In the space below, list any species not included in the species pairs for #1 or #2. Do any of these species form obvious groupings (i.e., share 2-3 characteristics only with each other)?

**#7.** Examine the example phylogeny on page 4, to familiarize yourself with the format of the phylogeny. Next, create a hypothesized phylogeny for five invertebrate species **(Crayfish, Crab, Scorpion, Spider, and Flatworm) based on the characteristics listed in Table 1.** **WAIT TO CREATE YOUR PHYLOGENY UNTIL YOUR INSTRUCTOR HAS REVIEWED THE CHARACTERISTICS IN TABLE 1.**

**#8.** On page 5, draw a cladogram showing hypothesized evolutionary relationships among all 13 invertebrates, using the characteristics provided in Table 1 and following the guidelines below:

Table 1. Derived characteristics:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Derived trait:** | **Polychaete** | **Crayfish** | **Clam** | **Grasshopper** | **Sea cucumber** | **Scorpion** | **Chiton** | **Cockroach** | **Crab** | **Millipede** | **Spider** | **Starfish** | **Flatworm** |
| A | Protein (chitin) exoskeleton |  |  |  |  |  |  |  |  |  |  |  |  |  |
| B | Calcium carbonate shell |  |  |  |  |  |  |  |  |  |  |  |  |  |
| C | Waterproof skin (fully terrestrial) |  |  |  |  |  |  |  |  |  |  |  |  |  |
| D | Segmentation (repeated parts) |  |  |  |  |  |  |  |  |  |  |  |  |  |
| E | Jointed legs |  |  |  |  |  |  |  |  |  |  |  |  |  |
| F\* | 3 pairs of jointed legs |  |  |  |  |  |  |  |  |  |  |  |  |  |
| G\* | 4 pairs of jointed legs (not claws) |  |  |  |  |  |  |  |  |  |  |  |  |  |
| H\* | 20+ pairs of jointed legs |  |  |  |  |  |  |  |  |  |  |  |  |  |
| I | Radial symmetry |  |  |  |  |  |  |  |  |  |  |  |  |  |
| J | Grasping claw(s) |  |  |  |  |  |  |  |  |  |  |  |  |  |
| K | Single muscular hydrostatic "foot" |  |  |  |  |  |  |  |  |  |  |  |  |  |
| L | Multiple hydrostatic “tube-feet” |  |  |  |  |  |  |  |  |  |  |  |  |  |
| M | Jointed antennae |  |  |  |  |  |  |  |  |  |  |  |  |  |
| \*Only select up to one of these three (F, G, H) per organism | | | | | | | | | | | | | | |

Table 2. Tally of **SHARED DERIVED** characteristics:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Crayfish** | **Clam** | **Grasshopper** | **Sea cucumber** | **Scorpion** | **Chiton** | **Cockroach** | **Crab** | **Millipede** | **Spider** | **Starfish** |
| Polychaete |  |  |  |  |  |  |  |  |  |  |  |
| Crayfish |  |  |  |  |  |  |  |  |  |  |  |
| Clam |  |  |  |  |  |  |  |  |  |  |  |
| Grasshopper |  |  |  |  |  |  |  |  |  |  |  |
| Sea cucumber |  |  |  |  |  |  |  |  |  |  |  |
| Scorpion |  |  |  |  |  |  |  |  |  |  |  |
| Chiton |  |  |  |  |  |  |  |  |  |  |  |
| Cockroach |  |  |  |  |  |  |  |  |  |  |  |
| Crab |  |  |  |  |  |  |  |  |  |  |  |
| Millipede |  |  |  |  |  |  |  |  |  |  |  |
| Spider |  |  |  |  |  |  |  |  |  |  |  |
| Starfish |  |  |  |  |  |  |  |  |  |  |  |

Example (6 Characteristics; 7 Transitions):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Derived trait:** | **Species 1** | **Species 2** | **Species 3** | **Species 4** |
| A | + | + | + |  |
| B | + |  | + |  |
| C | + | + | + |  |
| D | + | + |  |  |
| E | + | + |  | + |
| F |  |  |  | + |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Species 1** | **Species 2** | **Species 3** | **Species 4** |
| Species 1 |  | 4 | 3 | 1 |
| Species 2 |  |  | 2 | 1 |
| Species 3 |  |  |  | 0 |
| Species 4 |  |  |  |  |



**#8.** In the space below, create a hypothesized phylogeny for the following **FIVE** organisms: Crayfish, Crab, Scorpion, Spider, and Flatworm. **WAIT TO CREATE YOUR PHYLOGENY UNTIL YOUR INSTRUCTOR HAS REVIEWED THE CHARACTERISTICS IN TABLE 1.**

**STEPS:**

1. Identify pairs of organisms that share the most characteristics and connect these together first.

2. Note the characteristics of each organism to the right (e.g., “ACDE”)

4. Determine what characteristics needed to change following divergence of each species pair, and add these transitions to their branches of the phylogeny.

5. Connect up the rest of the phylogeny and insert any necessary character transitions.

**No. Characters:**  **No. Total Transitions: \_\_\_\_\_**

**#9. Parsimonious Cladogram for all 13 organisms: No. Characters:**  13 **No. Total Transitions: \_\_\_\_\_**

**WAIT TO CREATE YOUR PHYLOGENY UNTIL YOUR INSTRUCTOR HAS REVIEWED THE CHARACTERISTICS IN TABLE 1.**

**#9.** Homoplasies (analogous characteristics) can be identified by finding where the same evolutionary transition occurs twice in the same phylogeny.Were there any examples of apparent homoplasy in your phylogeny? Did any organisms need to lose derived characteristics (e.g., -A) to make your phylogeny work?

**WEEK 2: Phylogenetic Analysis of Amino Acid Sequence**

**Introduction**

In this second part of the Phylogenetics Lab, you will learn about molecular phylogenetics. A molecular phylogeny is one that is made by comparing the degree of shared gene sequences. Gene sequences can be evaluated by analysis of the genomic DNA sequence or the amino acid sequence of the proteins encoded by those genes. Although the use of DNA and amino acid sequences in phylogenetic analysis is newer than the use of morphological features, molecular phylogeny does not replace morphological phylogeny. When evolutionary biologists develop phylogenies, they use a combination of morphology, molecular sequences, and fossil record evidence as their data.

In this lab we will focus on the gene encoding enolase. You heard of enolase previously in the Metabolism Lab. Enolase catalyzes a key step in glycolysis and is the enzyme that is inhibited by sodium fluoride (NaF). This enzyme, and thus a gene encoding it, is found in most known living things, so it is a good target for a phylogenetic analysis. However, it is important to note that in actual practice, molecular phylogeny makes use of multiple genes or even large portions of genome sequence.

**Lab outline**

A. Relationship between amino acid sequence identity and phylogenetic relationships.

B. Using amino acid sequence data to build a phylogeny.

C. Demonstration of amino acid sequence evolution and protein structure.

**A. Relationship between amino acid sequence identity and phylogenetic relationships.**

In this part of the lab, you will analyze the amino acid sequence of part of the enolase protein from some of the species whose evolutionary relationships you already established by analysis of numerous morphological characters. We will use the enolase protein because all of these species have an enolase gene that they inherited from a common ancestor. In other words their enolase genes are homologous. When two or more species have a homologous protein, there will be shared portions of sequence which have persisted from the original protein found in their common ancestor, and there will be changes where parts of the amino acid sequence have diverged since the species diverged from each other. The more closely related two species are, the fewer changes will have happened and the more similar their genes will be. Thus there will be more similarity between the amino acid sequences of the proteins encoded by these genes.

Sequence alignment is the method to identify these shared, or **conserved**, portions of the sequence. Finding an optimal alignment is a complex process that is done with computer programs. We will use the web-based tool **Clustal Omega** to do this. Below is the output from this program showing the alignment of a part of the enolase protein amino acid sequence for chiton, crab, crayfish, starfish, and flatworm. Amino acids are represented by single capital letters (for example M = methionine, C=cysteine). Positions where all five species have the same amino acid, *i.e.* a conserved amino acid, are indicated by an asterisk (\*) below the alignment. If there is no asterisk at a given position, it is because there is some difference between some of the species. This latter type is what is useful for phylogenetic analysis because these represent changes that have happened since the species have diverged from their common ancestor.



**KEY**

* Capital letters are standard single letter abbreviations for amino acids.
* A dash – in the sequence indicates that there is a deletion in that sequence relative to the others.
* An asterisk **\*** below the alignment indicates that all sequences compared have the same amino acid at that position.
* Any position where there is not an asterisk, there is some difference between two or more of the sequences compared. Whether there is a blank , period . , or colon : , indicates the degree of chemical difference between the divergent amino acids. However, in this lab we will be counting all changes equally.

**Instructions**

You will be given a handout with an alignment of amino acid sequences of crayfish, crab, chiton, and starfish. You will individually analyze a portion of this data and then combine your results with those of the rest of the class for a phylogenetic comparison between these species.

Your instructor will assign you two species to compare. For the pair of species you are assigned, count how many positions meet both of the following criteria:

1. There is **no** asterisk

**AND**

2. The two species you are comparing have the **same** amino acid.

Record your count here \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Once you are done, provide your count to your instructor. Record the results shared by the whole class in the following table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | chiton | crab | crayfish | starfish |
| chiton |  |  |  |  |
| crab |  |  |  |  |
| crayfish |  |  |  |  |
| starfish |  |  |  |  |

**Discussion Questions:**

**1.** How do the degrees of shared amino acids correspond with the phylogenetic relationships that you previously proposed from morphology?

**2.** How is this analysis similar to the analysis of shared derived morphological characters that you did last week?

**B. Molecular Phylogeny**

In this part of the lab you will construct a phylogeny of a large number of species using the enolase amino acid sequence. You will learn how to retrieve sequences from public databases and how to align and construct a phylogenetic tree from amino acid sequences. You will work in pairs for this part.

You will be provided with the enolase amino acid sequences of many of the invertebrate species in the morphology portion of this lab. These are in the document named “Enolase Sequences for Phylogeny”. Not every species that you examined last week is there because for some of them, the full amino acid sequence for enolase has not yet been determined. You will collect enolase amino acid sequence on additional species as directed below and then use all of these to construct a phylogeny.

**Instructions for obtaining additional sequences**

The first thing you will do is obtain enolase amino acid sequences for additional species to allow use to build a broader phylogenetic tree. You will obtain sequences for the following:

|  |  |
| --- | --- |
| **genus and species** | **common name or description** |
| *Escherichia coli* | enteric gut bacterium |
| *Homo sapiens* | human |
| *Mus musculus* | house mouse |
| *Saccharomyces cerevisiae* | baker’s yeast |
| *Salmonella enterica* | enteric gut bacterium |

**1.** Go National Center for Biotechnology Information (NCBI) protein data web site <http://www.ncbi.nlm.nih.gov/protein>

**2.** To search for the amino acid sequence of enolase in a given species enter

enolase AND genus species

where genus species is replaced by that of the organism whose data you seek. For example

enolase AND Salmonella enterica

Then click on the **Search** button in the upper right corner of the page.

**3.** You will get a huge number of records, but don’t worry. Among those records that appear on the first page, select the one with the largest number of amino acids (aa) because this will be the most complete sequence. If they are all the same length, select the first one.

Click on the link of the chosen record.

**4.** Once you get to the page for the specific record, click on the link labeled FASTA that is under the title of the protein.

**5.**  The FASTA format amino acid sequence will have a > symbol followed by a string of letters and numbers. The amino acid sequence will start on a new line and consist of capital letters, with a single letter for each amino acid. It will start with M for methionine because methionine is encoded by the universal start codon. Copy only the amino acid sequence.

**6.** Paste the amino acids sequence in the Word document in place of the [paste sequence here] under the appropriate sequence.

**7.** Save the sequence data with a new file name.

**Instructions for Making a Phylogeny**

You will use the program Clustal Omega to make a phylogeny with this data.

**1,** Go to the following web site <http://www.ebi.ac.uk/Tools/msa/clustalo/>

Go to your Word document and select and copy ALL of the amino acid sequences. Paste them all into the dialog box on the Clustal Omega start page.

**2.** Click the **Submit** button near the bottom of the page. Wait for the program to perform the alignments. When the alignment is finished, a new page will appear showing the alignment of the sequences from all of the submitted species. As with the previous alignment program, bases shared in all sequences are indicated by an asterisk (\*). You will notice that there are far fewer asterisks than in your previous analysis. Why do you suppose that is?

**3.** Click on the **Phylogenetic Tree** button near the top. After the tree is formed and a new page appears, scroll to the bottom to a cladogram based on the amino acid sequence you gave. Make a sketch of this cladogram in the space below.

**Discussion:** For the species that you examined by morphology, how do the relationships compare between the phylogeny based on morphology and the phylogeny based on enolase sequence?

For the newly added species, reconcile their positions in the enolase sequence phylogeny with what you know about them?