



ASSOCIATION FOR BIOLOGY LABORATORY EDUCATION

**This article reprinted from:**

**Johnson, A. D. 2005. Morphological and molecular methods for creating phylogenetic trees. Pages 121-154, *in* Tested Studies for Laboratory Teaching, Volume 26 (M.A. O'Donnell, Editor). Proceedings of the 26th Workshop/Conference of the Association for Biology Laboratory Education (ABLE), 452 pages.**

---

Compilation copyright © 2005 by the Association for Biology Laboratory Education (ABLE)  
ISBN 1-890444-08-1

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the copyright owner. Use solely at one's own institution with no intent for profit is excluded from the preceding copyright restriction, unless otherwise noted on the copyright notice of the individual chapter in this volume. Proper credit to this publication must be included in your laboratory outline for each use; a sample citation is given above. Upon obtaining permission or with the "sole use at one's own institution" exclusion, ABLE strongly encourages individuals to use the exercises in this proceedings volume in their teaching program.

---

Although the laboratory exercises in this proceedings volume have been tested and due consideration has been given to safety, individuals performing these exercises must assume all responsibilities for risk. The Association for Biology Laboratory Education (ABLE) disclaims any liability with regards to safety in connection with the use of the exercises in this volume.

The focus of ABLE is to improve the undergraduate biology laboratory experience by promoting the development and dissemination of interesting, innovative, and reliable laboratory exercises.

Visit ABLE on the Web at:  
<http://www.ableweb.org>



# Chapter 7

## Morphological and Molecular Methods for Creating Phylogenetic Trees

*A. Daniel Johnson*

Dept. Biology  
Wake Forest University  
Rm. 226 Winston Hall  
Winston-Salem, NC 27109  
*johnsoad@wfu.edu*

**Dan Johnson** obtained his B.S. in Biology from University of North Carolina at Charlotte, and Ph.D. in cell biology from Wake Forest University School of Medicine in 1992. He was a post-doctoral fellow at Texas Heart Institute (Houston) and University of Virginia (Charlottesville). He has been a Lecturer in Biology and the Core Curriculum Coordinator since 1998, and teaches general biology for non-majors, as well as introductory cell biology and comparative physiology courses for pre-majors.

© 2005, A. Daniel Johnson

### Contents

Introduction	122
Student Outline: Systematics and Phylogenetics	122
Background	122
Pre-lab Homework	124
Week 1, Exercise #1: Creating a cladogram based on morphology	125
Week 1 Homework	130
Week 2 Exercise #2: Using sequence data to infer relationships	131
Week 2 Exercise #3: Aligning sequences using Workbench 3.2	132
Week 2 Homework	136
Materials	137
Notes for Instructor	137
Acknowledgements	140
Literature Cited	140
Appendix A: Handouts Used in Class	141
Appendix B: Handouts for Homework Assignments	142
Appendix C: Typical Results of Student Analyses	152

## Introduction

This exercise introduces students to the methods systematists use to uncover evolutionary relationships between organisms. We developed the lab for a freshman Ecology and Evolution course, but the general methods can be adapted and extended for a more advanced course. As written, the lab requires two, 2-hour blocks to complete. If you need a shorter lab, each week's exercise can be modified to stand alone. Photocopying the appropriate number of handouts each week is the only general pre-lab preparation required. In the second week, each group of 2–4 students will need access to a computer with a web browser (Macs and PCs work equally well.) It helps if the instructor has a computer with a projector so they can show the entire class simultaneously how to navigate the Workbench software interface.

The major concepts we stress with this lab are:

- There are many ways to arrange organisms into groups. However, only one reflects true phylogeny, that is, their actual evolutionary relationships.
- Neither morphological nor molecular data alone are infallible. Usually both are needed to uncover evolutionary histories.
- Synapomorphies are morphological traits that evolved recently and are shared between two organisms. They represent branch points in evolutionary history
- Differences in amino acids or nucleotides between two species can be treated as traits that have diverged from an ancestral sequence. Each position where a change has occurred is a distinct trait.
- Phylogenetic analyses rarely show a chain of species leading one to another. Usually 1 group splits into subgroups, which then diverge further. These separation/radiation events are due to many factors; an obvious one is geographic isolation.
- Different individuals can interpret the same dataset or evidence in different ways.

Students also learn these practical skills:

- How to locate & download DNA and protein sequence data from online databases.
- How to manipulate these sequences with bioinformatics software.
- How to determine phylogenetic relationships within a group of organisms.

## Student Outline: Systematics & Phylogeny

### Background

Diversity is a hallmark of life. Systematics is the branch of biology that tries to uncover the evolutionary history, or phylogeny, of groups of organisms. A diagram that summarizes such history is called a phylogenetic tree.

#### *The Origin and “Evolution” of Systematics*

Before the 1700's there was no universally accepted system of nomenclature, so scientists often had many short, descriptive names for one species. In 1735 Carolus Linnaeus published *Systema Naturae*, which outlined a classification system where different organisms were grouped because they shared morphological features (physical characteristics, development, behavior, or ecology). The traditional Linnaean system has 7 taxa: kingdom, phylum, class, order, family, genus, and species. General traits shared by a large number of organisms (such as single- or multi-celled,

photosynthetic or non-photosynthetic) are used to separate them into more inclusive groups such as kingdoms (plants, animals, fungi, etc.). More specific characteristics are used to assign organisms within a single kingdom to a particular phylum within that kingdom. Still more precise characteristics are used to further classify each organism into specific categories, down to a unique genus and species.

Linnaeus' scheme was not originally intended to describe evolutionary relationships. However, organisms in larger groups like Class or Order will share many general traits that appeared earlier in evolutionary history, while organisms within one genus usually share traits that evolved more recently. Modern systematists continue to revise and rearrange the Linnaean scheme so it reflects current evidence for the evolutionary history of life on Earth.

#### *Modern Systematics Uses Both Morphological and Molecular Data*

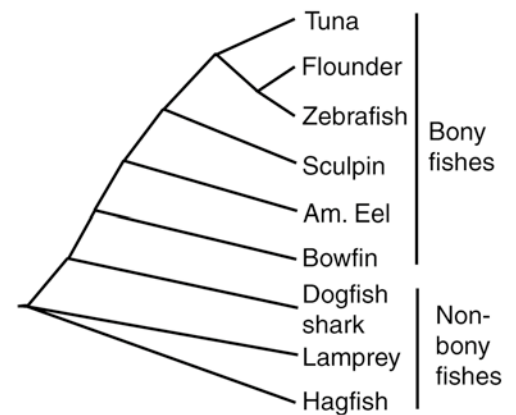
We have long used morphological characteristics to determine relationships between organisms, but they can be misleading. Complex eyes are an example. It seems logical that a simple eyespot arose once, then slowly changed into different, more complex eyes in response to various selective pressures. So it should be possible to use eye structure as a morphological feature to assign organisms to groups, right?

Wrong! Evidence suggests that primitive light gathering structures evolved as many as 60 different times! Complex eyes (like ours) have evolved independently at least three times, possibly more. So if we used eye structure alone to place organisms in various taxa, the groupings probably would not represent evolutionary history. This does not mean morphological features cannot be used to determine evolutionary relationships, but rather that a modern systematist must examine many different features in order to create a phylogenetic tree that reflects true evolutionary history.

Over the last 25 years or so, systematists have started using the tools of molecular biology. Inexpensive sequencing methods and powerful bioinformatics software allow them to compare long stretches of DNA or proteins from several organisms, calculate relative similarities between them, and create a phylogenetic tree that maps their genetic relatedness to one another. Figure 1 shows the possible evolutionary relationships between several bony and cartilaginous fishes, based on the sequence of amino acids in the insulin precursor protein.

Molecular sequence data have settled many arguments about who is related to whom, and overturned taxonomic relationships based on physical similarities alone. However, *molecular data are not infallible!* Like morphological features, DNA and protein analyses only indicate *likely* genetic relations between two or more organisms.

So, which type of data is most reliable in determining evolutionary histories? Most modern phylogenetic relationships are determined using mathematical algorithms that compile multiple individual trees (each built using morphological features or molecular sequence data) to create a single *consensus tree*. Figure 2 is an example of a consensus tree. It shows the relationships between Classes of insects, and is based on a combination of several sets of molecular sequence and morphological data.

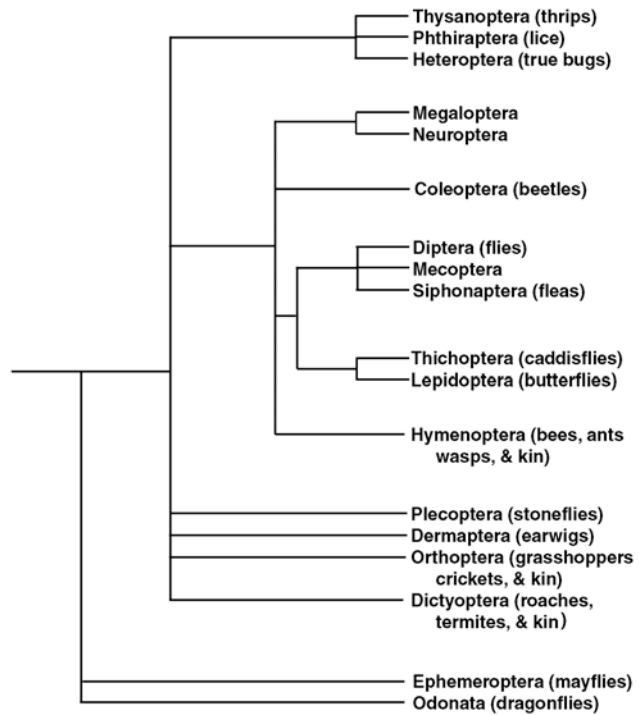


**Figure 1.** Phylogeny of fishes, based on sequence of insulin precursor protein.

**Figure 2.** Consensus phylogenetic tree for the major classes on insects.

*Goals For This Laboratory*

Even if the group of organisms is small, developing a consensus tree is a complex process. However it is fairly easy to learn to create the individual trees upon which a consensus tree is based. In the first week of this lab, you will create a phylogenetic tree based on morphology for a hypothetical group of organisms. For homework you will use the same methods to create a phylogenetic tree for a group of real organisms. In the second week of lab, you will learn to use bioinformatics software to download sequences, align them, and create a phylogenetic tree for several rodent species. The final homework assignment will be to use molecular sequences to create a phylogenetic tree for another group of real organisms that you will be assigned.



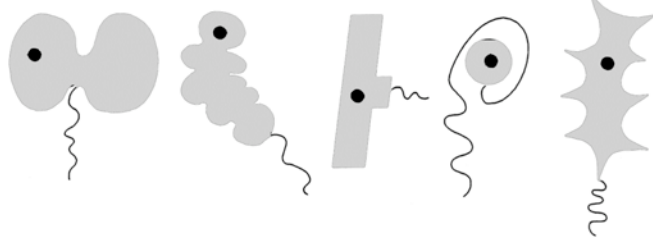
**Figure 3.** Schematic diagram of organisms within and outside the Mellinark clade.

**Pre-Lab Homework**

(Complete this exercise and bring it when you come to lab.)

Physical characteristics can be used to determine which organisms do or do not belong to a certain taxonomic group. Figure 3 illustrates this point. It is a schematic diagram of a series of organisms; some belong to a fictitious taxon called Mellinarks.

**All of these are Mellinarks**



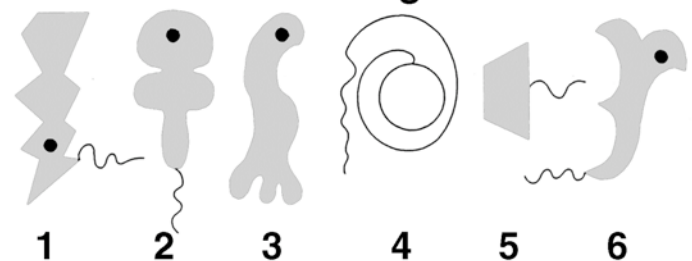
**None of these are Mellinarks**



*Questions:*

- Which organisms in the third row of Figure 3 are Mellinarks? What are the criteria for being a Mellinark, that is, for being a member of the **ingroup**?
- Say we wanted to determine the evolutionary history of Mellinarks.

**Which of the following is a Mellinark?**



You will learn this week that, in order to create a phylogenetic tree based on morphology, we must decide on an *outgroup*. Here the outgroup is simply the organism that is as close as possible to being a Mellinark without actually being one. Look back at all of the organisms you excluded in your answer to Question #1. Which would you choose as the representative of the outgroup, and why?

3. One goal of phylogenetic analyses is to determine which organism in a group is most similar to that group's common ancestor. However, based on just its appearance, it is not always easy to choose one. Look again at the 5 Mellinarks in the first row. One of them is physically very similar to the common ancestor of all Mellinarks. So which of the 5 in the first row do you propose is the most primitive type of Mellinark, and why? (Hint: look for the one that does not have as many changes in its body structure than the others, or is less complex in one way or another.)
4. To create a cladogram that separates Mellinarks into ancestral and more recently evolved types, we need to identify some physical characteristics that are found in some Mellinarks but not others. What features OTHER than the ones you listed in Question #1 above could you use to further subdivide and categorize the Mellinarks?

### Week 1, Exercise #1: Creating a Cladogram Based on Morphology

In this exercise, you will look at physical characteristics of a theoretical group of organisms called **anchor arm creatures**. These are not real organisms, but we will discuss their morphological characteristics and evolutionary history as if they were. Figure 4 shows 7 different kinds of anchor arm creatures, plus a “sucker-arm creature,” a close relative that will be used to represent the outgroup. Anchor arm creatures are identified by a unique **internal spool**, which has **protein filaments** that extend out through its 1–3 arms and connect to **terminal anchors** that the creatures use to attach to food sources. The spools and terminal anchors have never been seen on any other organism (living or fossilized), so we are going to assume that they arose once during evolution, and that all organisms having them are related. Yet which anchor arm creature most closely resembles the ancestral type, and which ones evolved more recently?

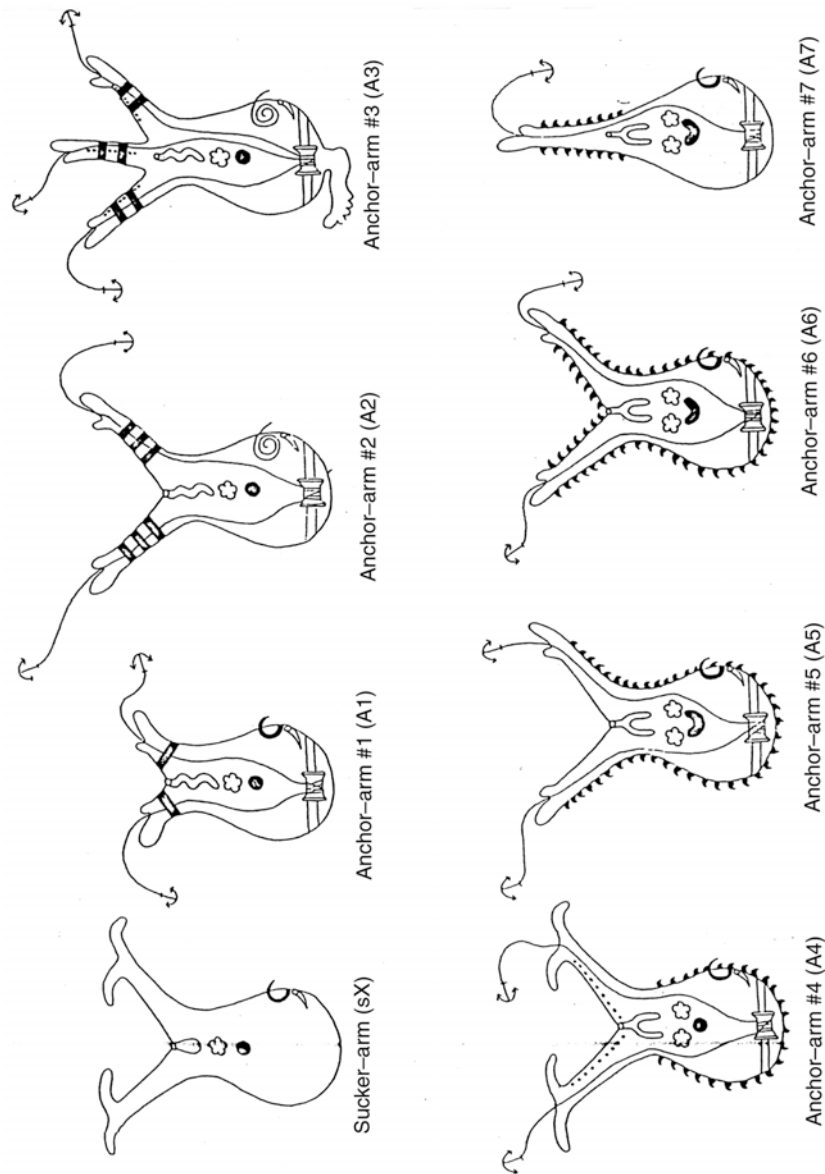
To answer these questions, you must determine the number of synapomorphies (shared, derived traits) for the group members, and use those values to construct a cladogram.

#### *Procedure*

The first step is to make a list of features that differ between members of the group. The traits need to be **binary**, meaning they can be scored as one of two possible **states**: present/absent, high/low, large/small, or some other pair of options. To make the process easier, Table 1 lists 12 characteristics that differ between anchor arm creatures. Selected examples of each trait are shown in Figure 5.

Next, you must **polarize characteristics**, relative to the outgroup. This means you must describe two states, named 0 and 1, where 0 represents how the characteristic looks in the outgroup. For example, look at the outgroup (sX) and first creature (A1) in Figure 4. A binary characteristic would be **length of the arms**, which can be in two states: **long** or **short**. The arms are long in the

outgroup (sX), so any creature (A1 through A7) that has long arms is given the value of 0 for that trait; and creature that has short arms has a state of “1” for that trait.



**Figure 4.** The 7 kinds of anchor arm creatures, plus a sucker creature.

In Table 1, assign states of **0 or 1** for each of the 12 different characteristics. If a characteristic has more than one state (the location of spines, for example), assign a 0 to the state as it appears in the outgroup (sX has no spines), then assign all of the other possible states a 1 (with spines). Later, you can create additional characteristics that can be used to further subdivide the groups.

Next, you mark the character states for each trait of each member of the in-group. Start with the first anchor arm creature, A1. For each of the 12 characteristics listed in Table 1, record whether A1 should be scored as a 0 or a 1. Repeat this procedure for the other creatures (A2 through A7).

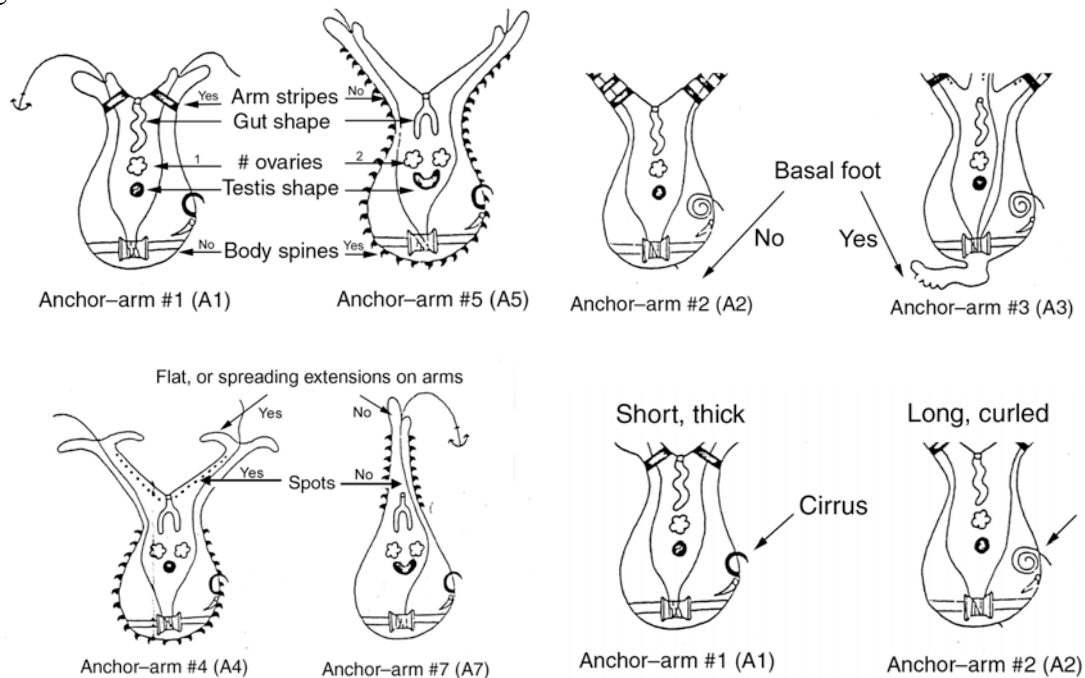
In Table 1, a score of “1” indicates a feature has changed in that anchor arm creature relative to its ancestors. These are **synapomorphies** (branch–points in evolution). If two creatures share several synapomorphies, they must be closely related. The fewer synapomorphies that are shared, the more distantly related the two organisms are.

Next, you must create a data matrix that shows how many synapomorphies each pair of organisms share. In Table 1, count the number of times that the same characteristic was scored “1” in **both** creatures A1 and A2. Enter that number in Table 2 Repeat this with each pairing (A1 and A3, A2 and A4, etc.).

Using Table 2 you can now construct a cladogram. Pick the two anchor arm creatures that have the **largest** number of synapomorphies. These will be the most closely related organisms in the group. On a separate sheet of paper, write their abbreviated names and connect them with a “y”-shaped line, preferably in pencil. (If you are unsure how to begin, refer back to Figure 1.)

Look for the creature that is next most closely related to the pair you just connected. Alternatively, there may be a 2nd pair that is more closely related to each other than to the two creatures you connected already. Keep trying to arrange and connect the 7 anchor arm creatures so that organisms sharing more synapomorphies (i.e., have a higher score in Table 2) are closer together than organisms that share only 1 or 2 synapomorphies. When you are done, you should have a cladogram similar to Figure 1.

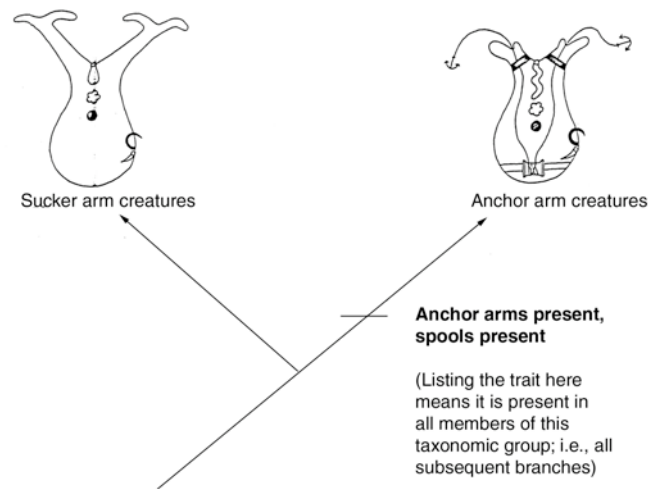
To check the accuracy of your cladogram, try to write a key morphological feature above each connecting branch; an example is shown in Figure 6. The morphological feature must apply to every organism on the particular branch of the tree, and must NOT be true for organisms that are not on that particular branch. If you cannot complete this checking procedure, your cladogram needs to be rearranged.



**Figure 5.** Basic anatomical features of anchor arm creatures. Arrows correspond to structures named in Table 1.



**Figure 6.** Example of how to annotate branches of the tree.



*Questions For Exercise #1*

1. What was the final arrangement of your cladogram for the anchor arm creatures? In your cladogram, which two creatures appear to be the most closely related, and which pair are the most distant, based on the evolutionary changes you charted?
2. Some morphological features of the anchor arm creatures had multiple states. For example, the gut (digestive tract) could be a round sac, a single long tube, or a forked tube (an inverted “Y” shape). Were you able to use any of the features that have multiple states to help determine the phylogeny? If so, how did you use them? If you had to ignore these traits, why?
3. (This is a thought question connecting this week’s results to our goals for next week. There is not one right answer. The point is to think about the problem.) Obviously, the physical structure of anchor arm creatures has changed dramatically as they evolved. Mutations in their DNA must have occurred, otherwise there would be no change in their body plan. Are these mutations going to be spread evenly throughout their entire genome, or will some alleles undergo more mutation than others?

For example, imagine that the anchor filament found in all 7 anchor arm creatures is composed of Filament Protein A. We have isolated and determined the DNA sequence that codes for Filament Protein A in all seven species.

- a. Compared to the average number of mutations in the ENTIRE genome, do you think the coding region for Filament Protein A contains a higher percentage of mutated bases, a lower percentage of mutated bases, or about the same percentage of mutated bases? What is your rationale?
- b. Imagine you are trying to determine the evolutionary relationships between the 7 anchor arm creatures using just DNA sequences. Think about your answer to Part A. Can you use the sequence of the coding region for Filament Protein A to determine the relationship between species? Why or why not?

**Table 1.** Data Matrix of Character States of Anchor Arm Creatures

Character	States	A1	A2	A3	A4	A5	A6	A7
Presence of spools, filaments, and anchors	0= 1=							
Length of the arms	0= 1=							
Presence of a basal foot	0= 1=							
Shape of the cirrus	0= 1=							
Number of ovaries	0= 1=							
Shape of the testes	0= 1=							
Spines present	0= 1=							
Stripes on the arms	0= 1=							
Shape of the gut	0= 1=							
Spots on the arms	0= 1=							
Ends of arms are broadly spread or flattened	0= 1=							
Total number of arms	0= 1=							
Other Features (if needed)	0= 1=							
	0= 1=  0= 1=							

**Table 2.** Number of Synapomorphies (Shared, Derived Features) in Anchor Arm Creatures

Anchor Arm Creature	A1	A2	A3	A4	A5	A6	A7
A1							
A2							
A3							
A4							
A5							
A6							
A7							

**Homework for Week #1: Creating a Cladogram Based on Morphology**

You practiced constructing cladograms using theoretical organisms. For homework, you must construct a cladogram for a group of real organisms assigned by your instructor. You will receive an outline that describes several morphological characteristics of members of the ingroup, as well as a representative outgroup. Using the information in the outline, you must:

- Decide how to score the morphological characteristics you must use.
- Create the data matrix tables.
- Draw a cladogram and annotate it.

Your scoring matrices and final cladogram are due at the start of class next week.

## Week 2, Exercise #2: Using Sequence Data to Infer Evolutionary Relationships

As you learned last week, morphological features may not be homologous, but may develop independently in two organisms due to convergent evolution. Like physical traits, DNA and amino acid sequences in proteins change with time. So they can be used as synapomorphic traits to create a cladogram. Like morphological features, it is possible for a sequence change to appear more than once. However, if we generate cladograms using sequences of several different proteins, and always get the same phylogenetic relationships, we can be fairly confident that we have a good estimate of their true evolutionary history.

Suppose analyses of two proteins yield two different trees, or that the trees based on morphology and sequence data differ. In these cases, we use statistical methods to compile the trees into one that explains as much of the data as possible. These methods are beyond the scope of this lab, but are essential tools of systematists.

In this exercise you will use an existing cladogram to track the changes in an amino acid sequence as anchor arm creatures evolved into different species. In Exercise #3, you will learn how we align protein sequences and determine their relationships without using any morphological features.

### Background

For this exercise you will examine a variable region of Filament Protein A (FPA). This protein is the main component of the filaments that connect the anchors to the spools in anchor arm creatures. Figure 7 shows a 20 amino acid sequence of FPA for all 7 of the anchor arm creatures you saw last week (the ingroup), along with the closest matching protein sequence that can be found in the sucker arm creatures (the outgroup).

**Figure 7.** Aligned amino acid sequences for Filament Protein A

sX	VKAHGKKVLG	AFSDGLAHL
A1	VKVHGKKVIG	AFSDGLAHL
A2	VKVHGKKVIG	AFSDGLAHL
A3	VKVHGKKVIG	AFSDGLAHL
A4	VKVHGKKVAG	AFTDGLAHL
A5	VKVHGKKVAG	AFSEGLAHL
A6	VKVHGKKVAG	VFSEGLLHL
A7	VKVHGKKVAG	VFSEGLAHL
Pos	1 2 3 4 5 6 7 8 9 10	12 14 16 18 20

Changes in nucleotides or amino acids in a sequence can be used in much the same way that we used morphological features last week. The FPA sequence shown in Figure 7 has 20 different positions where the sequence could **potentially** differ

between the organisms. However, not every amino acid changes; in conserved regions (outlined in grey), the same amino acids are present. These conserved amino acids do not help us determine the relationships between anchor arm creatures. Only positions where changes have occurred can be used as characters to estimate evolutionary relationships. Out of the 20 possible characteristics (amino acids), 8 can be used to determine evolutionary relationships.

Now look at position #3 in the sequences. Sucker arm creatures have an alanine (A) at that position, while in all of the anchor arm creatures (A1–7), there is a valine (V). Put another way, a valine at position #3 is a characteristic shared by all anchor arm creatures.

*Procedure*

Your instructor will give you a cladogram that shows one *possible* tree for the anchor arm creatures from last week. The diagram has blanks next to each branch of the tree. On the diagram, write down the amino acid change(s) that distinguish each branch. To be valid, the amino acid change must occur in ALL of the anchor arm creatures on that branch. For help you get started, the first branch that leads away from sucker arm creatures to all of the anchor arm creatures has been annotated like this:

#3: A—> V

This means that in position #3 of the sequence of Filament Protein A, the amino acid has changed from A to V. Using the sequences in Figure 7, continue filling in the blanks in the diagram. If there is no amino acid change that is unique to a particular branch of the cladogram, simply write “none”.

To double-check your tree, write down the entire 20 amino acid sequence for sucker arm creatures. Pick any of the 7 anchor arm creatures, and then trace the tree from its base to that particular organism. Each time you pass over a sequence change, make that change in the sequence for the sucker arm creatures. When you reach the final branch, the edited sequence should be identical to the one listed in Figure 7.

*Questions For Exercise #2*

1. In Figure 7, why would some amino acids be conserved in all members of the ingroup? Why can other amino acids change more rapidly?
2. Did you find any branch points where there was not a difference in the amino acid sequence between anchor arm creatures? If so, how is this possible?
3. Suppose your cladogram from last week did NOT explain the differences in amino acid sequences of FPA. What are 2 possible explanations for why the morphological and molecular data do not correspond to each other?

**Exercise #3: Aligning Longer Sequences Using Bio Workbench 3.2**

In Exercise #2, you mapped changes in amino acid sequences by arranging them along a pre-existing tree. However, relationships between DNA or amino acid sequences also can be determined without a pre-existing cladogram. The process requires cumbersome calculations when there are more than a few short DNA or protein sequences to align, so we let computers do the heavy lifting for us. Today you will use Bio Workbench (a free Web-based bioinformatics package) to retrieve protein sequences from a database, align them, and construct a tree that shows the relationships between them.

*Background*

The Order Rodentia (rats, mice, and other gnawing mammals) is a widely distributed mammalian taxonomic group. All rodents have 4 prominent, open-rooted front incisor teeth that are adapted for gnawing and chiseling. A large gap separates their incisors and remaining teeth. Rodents are

extremely common worldwide, with 175 species found in North America alone. They live in a range of habitats, from tropical swamps and marshes to deserts and cold alpine regions.

In this exercise, you will use the amino acid sequences for the Type I insulin precursor protein to create a phylogenetic tree for several species of rodents. The species chosen come from different habitats in different parts of the world. You will try and determine the evolutionary relationships between these species, and at the end of the lab, see if the results of your molecular analyses match their geographic distribution, habitats, or both.

### *Getting Started*

**Bio Workbench 3.2** is supported by the San Diego Supercomputing Center. Their web address is: <http://workbench.sdsc.edu/>

You must set up an account to use the Bio Workbench online bioinformatics program. On the main page, you should see two options: entering the program, or signing up for a new account. Choose “**Set Up a Free Account.**” Fill out the form and submit it. To make it easy to remember, we suggest you use your e-mail log-on name and password to register for a Workbench account.

WorkBench stores your data in **Sessions**. You can think of them as file folders, where you store individual sequences, aligned sequences, or the results of your calculations. You can have more than one session, so most people store all the work related to each major project in a separate session. WorkBench always creates a dummy session called DEFAULT to hold a place for you in the computer. You cannot enter data to this session, so you must either create a new one for this laboratory exercise, or use a session you created earlier.

Go to the Bio Workbench main page, log on, and select **Session Tools**. From the list, select **Start New Session** then select **Run**. You will get a dialog box that lets you create a session. Give your session a name and click again on **Start New Session**. You will return to the main page.

### *Downloading the Sequence Data*

To begin, you need some sequence data. You can type sequences directly into a form, but millions of useful sequences have already been entered and stored in various databases. Today, you will retrieve amino acid sequences for insulin or insulin precursor proteins from a database called **SwissProt**.

Begin by clicking on **Protein Tools**. Scroll down the menu until you find the option **Ndjinn–Multiple Database Search**. Click on this option, then click on **Run**. ‘Ndjinn’ is similar to Google, Altavista, or any other web search engine, except it looks through bioinformatics databases. Like any search engine, the quality of the results you get out of Ndjinn depends upon how carefully you choose search terms.

When Ndjinn starts, you must type in your search terms, and choose the databases to search. Scroll down the list of available databases until you find SwissProt. Click on the check box beside it, then scroll back to the top of the window.

To find the insulin sequences you could simply type in **insulin**, and choose **Run**. However, this will return every last sequence in the entire database that has the word “insulin” anywhere in the file. This would include tens of thousands of sequences, including insulin receptors, insulin–like growth factor, insulin–related somatomedin, etc., for every organism in the database.

To narrow the search, you will use Boolean modifiers. These simply tell the computer search program to limit what it returns. Modifiers are AND, NOT, & OR, and are always typed in capital letters.

In the search box at the top of the Ndjinn page, type “insulin AND rat” (omit the quotation marks when you type), then choose **Run**. You still will get 50 or more sequences, ranked in order of how close they come to your search criteria. Look through them until you find one called “**Insulin 1 precursor, [Rat].**” This is the first amino acid sequence you need to import.

Click the check box next to the sequence, and scroll up or down until you see a button labeled **Import Sequence**. Select this, and the program should import the file and return you to the main page for Protein Tools.

Use this same method to locate and download all the sequences in the following list. All of them can be found using the search string “insulin AND           (common or scientific name)”:

- Mouse (*Mus musculus*; like rats, mice make 2 types of insulin so be sure to choose Type 1 insulin)
- Chinese hamster (*Cricetulus longicaudatus*)
- Israeli sand rat (*Psammomys obesus*)
- Nutria (*Myocastor coypus*)
- Guinea pig (*Cavia porcellus*)
- Chinchilla (*Chinchilla brevicaudata*)
- Degu (*Octodon degu*)
- Rabbit (*Oryctolagus cuniculus*); this will be your outgroup for the analysis.

#### *Setting the Outgroup*

After you have imported all 9 sequences, you must set one to be the outgroup for subsequent analyses. Unfortunately, Workbench does not let you choose the outlier group directly; you must modify the file’s name in a certain way.

Find the sequence file for rabbit insulin. Click on the file to check it. Then under **Protein Tools**, select the option **Edit Protein Sequence(s)**. Click **Run**.

A box will appear entitled “**Label:**” with the name of the file in it. Click in the box at the very beginning of the name of the file, and add “**zz**” to the name of the file. For example, if the file name was originally “**Rabbit insulin,**” the new name would be “**zzRabbit insulin.**” Scroll down the screen until you see the **Save** button, and click it.

You will return to the main page of protein editing tools. A copy of the newly renamed file will now be in the list of sequence files, along with the original file with the old name. Anytime you want to use that particular sequence as the outgroup, just choose the version which has “**zz**” added to the name.

#### *Aligning Similar Amino Acid Sequences*

Once you have set the outgroup, you can align the 9 imported sequences to find which amino acids are conserved and which vary between species. Begin by scrolling down through the list of options on the **Protein Tools** page. Select the option **CLUSTALW**, which is a sequence aligning algorithm.

Click on the check boxes next to each of the 9 sequences that you want to have included in the alignment. If you accidentally downloaded sequences you do not want to include, simply do not check them. Be sure that you do NOT check the original unedited file for rabbit insulin, but instead select the file for which you changed the name. Then select **Run**.

The program shows a list confirming the sequences to be aligned, and gives you the option of changing how it will perform the alignment. Change the first option from “**Aligned**” to “**Input**”

**Order**". This forces the file you renamed with the prefix "zz" to be used as the outgroup. You can leave the other remaining options at their default settings. Select **Submit**.

Check with your instructor about the report you receive. If the information in the sequence alignment is what you wanted to see, select **Import Alignments**; and you will go directly to the **Alignment Tools page**. If for some reason the CLUSTALW report does not show what you wanted, simply select **Return**. The program will delete the alignment without saving it, and you can do it again.

### *Generating a Parsimonious Phylogenetic Tree*

To reiterate what you have already accomplished, you downloaded protein sequences that were published by someone else, and aligned these sequences to show the conserved and non-conserved amino acids. These steps must always be done before creating any kind of tree from amino acid (or DNA) sequences. The next step is to see how the sequences are related to one another.

You will be creating a phylogenetic tree based on the **principle of parsimony**. To understand this principle, remember that: 1) mutations are relatively uncommon; and 2) the only way an amino acid sequence can change in an organism is if there has been a non-lethal mutation. Since changes in amino acid sequences are rare events, it is extremely unlikely that they will occur at random. Instead, changes in amino acid sequence tend to accumulate in a logical order. The principle of parsimony builds on this thinking; it states that the phylogenetic tree that requires the fewest evolutionary changes (mutations) to arrange the sequences in a logical order is the best explanation of the diversity in the group.

In Bio WorkBench, the **ProtPars** (short for **Protein Parsimony**) alignment program draws trees based on this principle of parsimony. How does it actually find the most parsimonious tree? The program starts with the amino acid alignment created by CLUSTALW and repeatedly rearranges the sequences into various trees, where similar sequences are on nearby branches. After each tree is constructed, the program counts how many times an amino acid would have to change in order to make the arrangement of sequences on that tree. The program selects the arrangement of amino acid sequences that requires the fewest number of changes, and displays it as the most parsimonious tree. If ProtPars cannot find just one best arrangement, it will display more than one tree, each of which is equally likely to have occurred.

If you are not already there, go to the page for **Alignment Tools**. Scroll down through the list of options. You will see several options for drawing trees; each uses a different algorithm (calculation method) to determine relationships between protein or DNA sequences. Look for the option **ProtPars**.

Select ProtPars, check the box next to your group of aligned insulin precursor sequences, then select **Run**. Again you will have the option of setting several parameters. Make sure that the first one "**Randomize order of sequences,**" is set to **No**. Leave the rest at their default settings, and click **Submit**.

The program will return in a few moments with one or more **unrooted parsimonious trees**. Each arm of the tree will have a number on it. These indicate the number of changes that were necessary to arrive at the amino acid sequences represented on that particular branch point on the tree.

Copy the parsimonious tree(s) from your screen onto paper. Alternatively, you can use the combination of **Alt + PrntScrn** to put a copy of the window in the computer's clipboard. You can then paste the image directly into a Word or PowerPoint document.



Once you have completed the online part of the exercise, your instructor will give you a handout that summarizes the geographic distribution and natural history of each of the species you analyzed. You must use both your parsimonious tree and the printed handout to answer the questions at the end of this exercise.

### *Questions For Exercise #3*

When you turn in your answers, include a printed copy of the tree that you created.

1. Are rodents that have similar geographic distribution (i.e., natives of the same continent) close to each other on your tree? What do you think this means?
2. Are rodents that live in similar habitats, but on different continents, close to one another on your tree? How could you interpret this result? Do you think you would see the same thing if you looked at another protein?
3. Many changes in sequences are **semi-conserved changes**, where the new amino acid has the same general chemical properties as the old one. For example S (serine) often is changed T (threonine), and vice versa. Both of these hydrophilic amino acids have an –OH side group. In contrast, a **non-conserved change** means the new amino acid has different properties. An example would be if S (which is hydrophilic) is mutated to give V (valine), a hydrophobic amino acid.

Both types of changes occur in proteins, but one will be more common than the other. Which would be more common, and why?

## **Homework for Week #2: Find & Align Your Own Molecular Sequence Data**

You have been working with a pre-determined dataset to construct phylogenetic trees. For homework, you will construct a tree using sequence data for an assigned group of organisms. You will receive an outline listing several members of the group by both common and scientific name. The outline will also suggest proteins you might use to determine plausible evolutionary relationships.

Using Bio Workbench, you must download the necessary sequences, align them, and create a phylogenetic tree for 4–5 of the representatives in the group. Next week, you will need to turn in a copy of the sequence alignment, and your phylogenetic tree(s).

### *Helpful Hints*

- When you retrieve nucleotide or amino acid sequences, the file for one species may contain a much longer sequence than another; alternatively, one file may contain only a partial sequence, and another, the complete sequence. Do not try to align sequences that are very different lengths. Instead, trim off the longer sequence file so that the same regions are being compared.
- Phylogenetic analyses take time. Do not be discouraged if you do not get a perfect alignment the first time. You may have to try one time, find the errors in how you made the alignment, edit the files, and try again. This is normal.

## Materials

The handouts are reproduced in Appendices A and B. For homework after the first lab, each student needs 1 of the 4 handouts summarizing morphological features of a clade of plants or animals. Assuming we have 200 students, we make 50–55 copies of each of the 4 handouts then randomly assign a clade to each student. For second week of lab, each student needs their own copy of the pre-drawn tree for anchor-arm creatures, and the handout describing the natural history of rodents. For homework after Week 2, they need a copy of 1 of the 4 handouts describing clades to be used in molecular analyses.

## Notes for Instructors

### General Comments

The students' instructions for this lab are quite detailed, but it is still essential that the instructor work through the entire exercise before the first lab meeting so they can see how the individual components are linked together.

More assertive students will try to rush through this exercise to get answers without understanding the underlying concepts. Make sure these students demonstrate conceptual understanding before you let them work independently. Other students may panic at the apparent complexity of the lab (it's a variation of math anxiety!) and claim they cannot do or understand it at all. To lessen their anxiety we remind students that they will complete the exercises in small, manageable steps. Also, we do not emphasize formal definitions and terms, but instead help them develop an intuitive understanding of the goals and rationale behind phylogenetic analyses. Finally, we always have 2–4 students work together, rather than alone.

To make the second part run more smoothly we use a computer projector to show the entire class how to perform each step in the molecular analyses. Then students can help each other with the mechanics of data retrieval and alignment, while the instructor focuses on linking the concepts together. If you do not have a projector and need a version of the exercise that includes figures showing each step, contact the author.

Finally, it is essential that the students understand there may be more than one way to interpret the morphological features of the anchor arm creatures. Otherwise they become focused on finding the "right" answer. An essential element of Exercise #1 is that there is not one correct answer. We need the sequence data in Exercise #2 to help resolve certain questions. Some students will (hopefully) ask: "why do we not use just sequence data to start with, rather than morphology and sequence data?" The sequences in Exercise #2 were designed so that, if they are analyzed with Workbench, there is more than one possible phylogenetic tree. So if we just used sequences, we still would need the morphological data to clarify the phylogenetic relationships of anchor arm creatures.

### Comments, Tips, Hints, etc. for Specific Exercises

1. In Exercise #1, traits with more than one state can create problems. For example, some of the anchor arm creatures have spines and others do not; stated that way, the spines trait has 2 states. However, there are several different positions for the spines (body only, arms and body, arms only). It is difficult to score traits when they have multiple states. The solution is to split each multi-state trait into two or more traits, each having just 2 states. In the case of the spines, students might use spines on arms as one 2-state trait, and spines on other body parts besides the arms as a second 2-state trait.

In phylogenetic analysis there is no single correct way to assign binary states. As long as students have a rational explanation for their divisions, they can try to use them to create a phylogenetic tree.

How can students tell if they have split a multi-state trait like this one in a way that reflects evolutionary history? One way is to create several variations of a tree, and compare them. One of our instructors has different groups of students split up the multi-state trait in different ways, then generate a tree for anchor–arm creatures. Whoever gets the simplest tree (i.e., the fewest branches) PROBABLY divided the multi-state trait in a way that reflects the actual evolutionary history of anchor-arm creatures. Another way to deal with multi-state traits is simply ignore them and look for other binary morphological traits. In this example, we do not have any more morphological traits to use. So we use molecular analyses instead; this leads into the exercises in the second week.

2. There is a hidden bomb in the data for Exercise #1. Most students try to use number of arms as a binary trait, and score it as “0=2 arms, 1=more or fewer than 2 arms.” This would make Creature #3 (with 3 arms) related to Creature #7 (with 1 arm). They are NOT closely related if you eliminate arm number and look at other traits.

We left this bomb in place because it teaches a valuable lesson about evolutionary mechanisms. When students lump Creature #3 and Creature #7 together, they are saying a SINGLE evolutionary change occurred that SIMULTANEOUSLY resulted in some creatures having 3 arms, and others having 1 arm. This is possible, but it makes less sense than two distinct evolutionary changes, one that caused loss of an arm, and a second that added an arm.

Since it only sidetracks students for a few minutes, we let them make the mistake of lumping 1 and 3 arms together. As a result, there will be too few matches in the data matrix (Table 2) to create a tree. We tell them to go back and look for traits that might be due to multiple evolutionary events. Invariably someone questions the validity of arm number, giving us a reason to explain the concepts described above.

3. In Week #2, Exercise #2, students mark amino acid changes in the filament protein sequence onto a photocopied tree for the anchor arm creatures. DO NOT give them this diagram until the second week. If they see the diagram the first week they assume they must match it. Also, as they annotate the tree on the handout, they will find some branches do not have corresponding changes in amino acid sequence of FPA. Students see these blanks and assume they did something wrong. This is how we lead them into discussing the mechanisms by which sequences change. We explain that speciation does not mean every protein or DNA region has to change, and talk about why particular sequences do or do not change easily.
4. In Week #2, Exercise #3, students use parsimonious sequence alignment to determine relatedness of 8 species of rodents. They are asked to determine which are more closely related: rodents that have different lifestyles but are geographically close, or rodents with similar lifestyles that are geographically separated. For this particular clade, geography rather than lifestyle reflects evolutionary relatedness. Despite having a range of lifestyles, rodents in South America are closer to each other than to the species from other continents, because they radiated out from just a few colonizing species, and evolved in isolation. This pattern is common, but is not always true. In some other clades, geographically separated species that share the same lifestyle may be

descendants of a common ancestor. Shared ancestry in geographically distant groups is more common with migratory or highly mobile species like birds or fish.

### Questions From Participants at the ABLE Conference

1. Is there a way to perform the morphological analyses on the computer as well?

I have not found a way to enter data matrices into Workbench so it can perform morphological analyses. Standalone programs are available but require students to learn a second computer program. The most widely used programs are PAUP and PHYLIP (PC, Mac), and MacClade (Mac only). Information about these and many other programs can be found at:

<http://evolution.genetics.washington.edu/phylip/software.html>

2. Aside from those provided at the conference (see Appendix B), what are some other good clades for students to explore?

We have not tested these with students, but for morphological trees, try using marsupials of South America and Australia to demonstrate geographic isolation and radiation. Many instructors already use skull and bone morphology of mammals as a general demonstration of phylogenetic change. If you already have such a lab, it would be simple to add the data analyses used in Exercise #1 to it. To help engage pre-medical students, try using molecular analysis to determine relationships between strains of HIV, different strains of influenza, or viruses within a single group such as the Poxviridae (the pox viruses). There is still controversy about whether photosynthesis originated more than once; have students align the sequences for RUBISCO or other photosynthetic enzymes from various plants and prokaryotes.

3. Are there papers highlighting phylogenetic analyses that students could understand?

A Nature paper (D'Erchia, 1996) had the blunt title, "The guinea pig is not a rodent." It immediately created controversy, which was summarized by Cao (1997). Both are short papers, and they demonstrate the ongoing discussion that accompanies most from phylogenetic studies.

Another recent analysis suggests the traditional invertebrate groups should be rearranged. Protostomes should be divided into Ecdysozoa, Lophotrochozoa, and 1 or more clades containing flatworms, rotifers, and several others. A brief, informative paper by Haase et al. (2001) summarizes and references some of the supporting data. The paper itself describes an enzyme marker (i.e., a morphological feature) that appears to be specific to Ecdysozoans.

4. What are some useful general references and web sites?

For morphological methods, there is a good textbook by Judd, et al. (2002). It deals with classifying plants, but most of the general discussion and many of the methods apply to animals as well. Two standard references for molecular methods are by Avise (1994), and Hillis, et al. (1996).

Online, an excellent starting point is the Tree of Life (<http://tolweb.org/tree/>). It is a collaborative effort to summarize data about phylogeny and diversity, and has dozens of annotated and referenced trees. The site also has an extensive set of links to other sites, including lab groups working on specific clades.

### Guide to the Appendices

Appendix A contains the handouts for the second week of the lab. Give students Part A (figure page) when they begin Exercise #2. Give them Part B (the natural history summary) once they have completed their sequence alignments for the 8 rodent species.

Appendix B contains the handouts for the 2 homework assignments. Sections A–D describe morphological features of 4 different clades. Each student needs a copy of 1 of the 4 handouts at the end of the first lab. Sections E–H describe clades that are to be analyzed using sequence data; again, each student needs only 1 of the 4 handouts.

Appendix C is for the instructor only. It is a key for the lab and homework assignments. The most commonly derived trees are shown.

### Acknowledgments

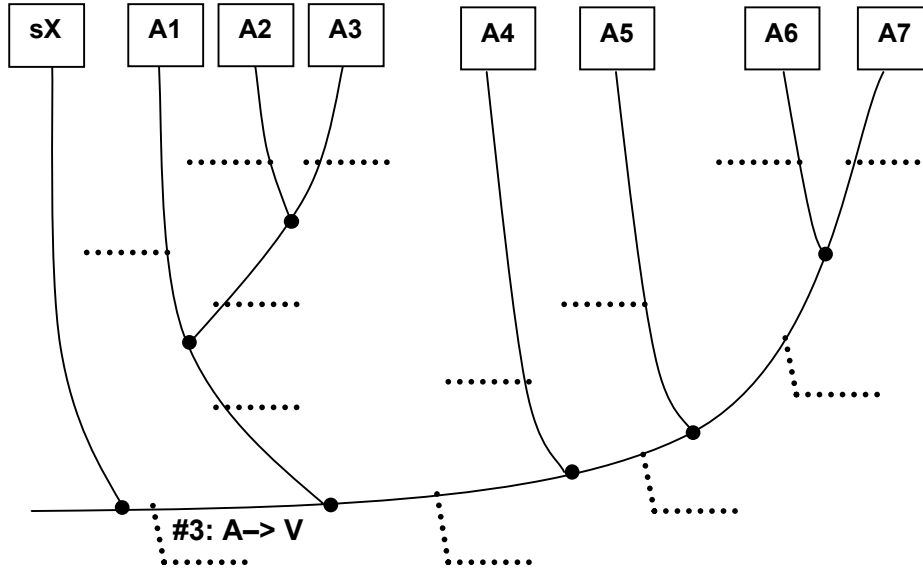
The author would like to thank Dr. Kathy Kron, Wake Forest University, for contributing the original outline of the exercise on morphology of anchor arm creatures, and for numerous discussions and constructive comments as these exercises were developed. Thanks also to the graduate teaching assistants, Jill Awkerman, Joel Fellis, Craig Moretz, Katherine Thorington, and Julie Wyatt, for their many suggestions for improving the exercises, and for sharing their teaching experiences and tips.

### Literature Cited

- Avise, J. C. 1994. *Molecular markers, natural history, and evolution*. Chapman & Hall, New York, 511 pages.
- Cao, Y., N. Okada, and M. Hasegawa. 1997. Phylogenetic position of guinea pigs revisited. *Molecular and Biological Evolution*, 14:461–464.
- D'Erchia, A. M., C. Gissi, G. Pesole, C. Saccone, and U. Arnason. 1996. The guinea pig is not a rodent. *Nature*, 381:597–600.
- Haase, A., M. Stern, K. Wachtler, and G. Bicker. 2001. A tissue-specific marker of Ecdysozoa. *Development, Genes & Evolution*, 211:428–433.
- Hillis, D. M., C. Moritz, and B. K. Mable. 1996. *Molecular systematics*. Second edition. Sinauer Associates, Sunderland, Massachusetts, 655 pages.
- Judd, W. S., C. S. Campbell, E. A. Kellogg, P. F. Stevens, and M. J. Donoghue. 2002. *Plant systematics: a phylogenetic approach*. Sinauer Associates, Sunderland, Massachusetts, 576 pages.

## Appendix A: Handouts Used in Class During Week #2

### A. One Possible Phylogenetic Tree for Anchor Arm Creatures



### B. Habitats and Natural Histories of Selected Rodent Species

The species of rodents you used for the molecular analysis are very different from one another. Use this summary information to help you answer the questions at the end of Exercise #3.

- Norway rat (*Rattus norvegicus*)
- House mouse (*Mus musculus*)

These are the common species found in houses. They originated in the temperate zone of Europe, but have been inadvertently introduced worldwide by human colonists. In many areas, house mice, Norway rats, and black rats (*Rattus rattus*) have completely displaced the native rodents. Although they are found from the Arctic to the tropics, and from deserts to mountain tops, Old World rodents can only live with or near humans.

- Chinese hamster (*Cricetulus longicaudatus*)
- Israeli sand rat (*Psammomys obesus*)

Hamsters and sand rats are found in different areas of the world, but in similar habitats. Chinese hamsters (which are paler colored versions of pet hamsters) are native to the semi-arid areas of central China and Mongolia. Israeli sand rats are found throughout Palestine, the Arabian peninsula, and over into Northern Africa. These and other sand rats (which actually look like very fat mice) often live in true deserts.

- Nutria (*Myocaster coypus*)
- Guinea pig (*Cavia porcellus*)
- Chinchilla (*Chinchilla brevicaudata*)
- Degu (*Octodon degu*)

South America are considered a “hot-spot” for rodent diversity, because there is an incredible variety of them. One reason is the range of available habitats. South America is best known for its extensive Amazonian rainforests and coastal swamplands. However, wet temperate and tropical woodlands are only one of many habitats on the continent. South America also has the Andes mountains, which rise to over 20,000 feet and have permanent glacier fields. There are also large areas of semi-desert, and high grassy plains.

Apparently, the first rodents that colonized South America were isolated by the narrow strip of land that forms the Isthmus of Panama. As a result, they evolved in relative isolation from rodent populations in North America and the rest of the world. Isolation allowed them to diversify and exploit all the various habitats available to them. You have

analyzed sequences from 4 South American rodents: nutria, chinchillas, degus, and guinea pigs. Each species represents a different habitat within South America.

Nutria are natives of the lowlands and coastal areas of the South American continent, but have been accidentally or intentionally introduced several times in Texas and Louisiana, where they have become pests. Their habits are very similar to a beaver or muskrat. They are active in the twilight and at night, and live in family groups or hierarchically stratified colonies. They move and feed mainly in the water where they build floating platforms they use as dining areas and resting sites. Their diet is mostly aquatic or shoreline plants.

Chinchillas are originally from Southern Peru, Bolivia, northern Argentina and Chile. They live in mountainous regions at altitudes of from three thousand feet to fifteen thousand feet above sea level, where the climate is cold and dry. Their habits are very similar to the marmots (a relative of groundhogs) that live in the boulder fields of the U.S. Rocky Mountains. They live among the rocks and boulders, where they feed on bark, branches, and leaves of low shrubs, as well as coarse mountain grasses.

Degus originated in the Andes mountains in Chile, and like chinchillas and guinea pigs, were domesticated for research, commercial ranching, or as pets. The best way to describe one is to imagine a pet gerbil that has grown to the size of a common gray squirrel, and has longish ears. Degus are not widely distributed across the South American continent. Their range is restricted to the west coast of Chile, from the plains up into the Andes mountains around 5000 ft up. Within this restricted area though they can be very abundant. They live in large social colonies, similar to prairie dogs in central North America. Again, like prairie dogs, they are considered to be agricultural pests.

Wild guinea pigs are native to Venezuela, Columbia, Brazil and Argentina, where extended families or social groups of up to 10 individuals inhabit burrows in rocky areas, savannas (a type of grassland), forest edges, and swamps. They are most active at night, when they forage for a wide variety of plant materials. Like European rats, guinea pigs are generalists that can adapt to a great range of climates, including the low, hot Amazon basin, the high Andes, and the arid western coastal plain.

## Appendix B: Handouts for Homework Assignments

### A. Morphological Phylogeny: the Fagales

This group includes oaks, birches, myrtles, beeches and walnuts. Several members of this familiar group are economically important for various reasons. Oddly enough, the most similar related group are the squashes, which you will use as the outgroup.

**Myricaceae:** bayberry, wax myrtle. These small trees or shrubs produce a fragrant resin that was the original source for the scent of bayberry soaps and candles. Wax myrtles are salt tolerant trees planted in coastal areas as windbreaks or hedges.

**Fagaceae:** beeches and oaks. Beeches are medium-sized trees, with species distributed nearly worldwide. In the US, the beeches are often found as climax species in hardwood forests. The fruit is 2 or 3 edible nuts within a small bur with soft spines. The oaks are also found worldwide. Oaks are slow-growing under normal conditions, making the wood exceptionally dense and hard. Like beeches, oaks are commonly the dominant climax species in a hardwood forest. The nuts or acorns of both types of trees are a staple in food caches of squirrels and other foraging mammals.

**Juglandaceae:** walnuts, hickories, and pecans. In addition to providing nuts, walnut wood is often used for fine furniture. Because it is so dense, hickory wood is often used for tool handles. Walnut trees produce several allelochemicals (the most common is juglone) that act as herbicides and kill all competing ground cover beneath the trees.

**Betulaceae:** birches. These slender, often multi-trunked trees have relatively thin, flexible bark that often peels from the trees in sheets. Many species are adapted to riverbanks and other moist habitats.

**Casuarinaceae:** she-oaks and tropical ironwoods. Rapidly growing tropical trees that reach heights of 40 m. They form dense stands under which little else grows, which suggests they release a plant toxin (similar to the juglone released from walnuts) that kills competitors. They are beneficial as windbreaks, erosion control, and for nitrogen fixation.

**Cucurbitaceae:** squashes. This large and diverse group contains pumpkins, winter squashes, and several other vegetables that form on large mounding vines.

Some anatomical features of plants have unfamiliar names. Those used in the table are described below. Do not focus on learning plant anatomy; instead, focus on determining who is related to whom.

**Cupule:** a cup-shaped anatomical structure, like that holding an acorn.

**Hairs:** often leaves, stems, or other structures are covered with coarse or fine hairs.

**Nectaries:** the nectar-producing glands at the base of stamens.

**Pollen types:** pollen grains have many different features that are visible under a microscope. Trisulcate grains have 3 rows or furrows in them. Triporate grains have three rounded indentations.

**Sepal:** the (usually) green leaf-like cup structure surrounding the base of a flower.

**Gametophyte:** the embryonic plant precursor within a seed.

**Ovule:** the structure that develops into a seed.

#### Morphological Features of the Fagales

	Cucurbitaceae (outgroup)	Myricaceae	Casuarinaceae	Fagaceae	Betulaceae	Juglandaceae
Nectaries	Present	Absent	Absent	Absent	Absent	Absent
Gametophytes/ ovule	One	Many	Many	One	Many	Many
Aromatic glands	No	Yes	No	No	No	Yes
Pollen surface texture	Smooth	Tiny spines in rows	Tiny spines in rows	Smooth	Tiny spines in rows	Tiny spines in rows
Sepals and petals	Reduced	Normal	Normal	Normal	Normal	Normal
Cupule	Smooth	Smooth	Smooth	Scaly	Smooth	Smooth
Hair type	Nonglandular, unbranched	Glandular, branched	Glandular, branched	Glandular , branched	Glandular, branched	Glandular, branched
Pollen type	Trisulcate	Trisulcate	Triporate	Trisulcate	Triporate	Trisulcate

#### B. Morphological Phylogeny: The Scaraboid Beetles

Beetles (Order Coleoptera) are among the most numerous organisms on Earth, and scarabs and their kin are the most common beetles. Scarabs include several species you already know, such as June beetles, Japanese beetles, and golden beetles (they frequently come to outdoor lights in summer). Scarabs are represented by several families:

**Glaresidae:** the burrowing fungus scarabs. This beetle family has one genus, *Glaresis*, that is found in arid and sandy regions worldwide. Members of the family are small (2-6mm long) and light to dark brown. Adults are collected at lights, and **stridulate** (squeal or squeak) weakly when handled. It has been hypothesized from gut contents that they feed on subterranean fungi, but they have not been successfully reared in lab conditions. Based on several features, Glaresids are considered the most primitive living Scaraboid beetle. They will be the outgroup for this phylogenetic analysis.

**Passalidae:** bessbugs. These are large beetles (20–70mm long), elongate or cylindrical, usually black; ventral surface with or without erect, moderately dense yellow hairs. Adults and larvae live in subsocial groups in rotting logs. All stages are found in galleries excavated by the adults. Eggs are usually placed together in a primitive “nest”. Adults and larvae communicate by stridulating and can produce 14 different calls. Adults care for larvae and prepare food by chewing it and presumably mixing it with saliva. Both adults and larvae need to feed on adult feces to obtain nutrients released by gut bacteria.



**Lucanidae:** stag beetles. Range in size from 1-9cm. Color is amber or reddish brown to black, sometime with yellow markings. Most adult males possess greatly enlarged, curving mandibles that are sometimes used in combat with male opponents during displays to establish dominance. Females lack enlarged mandibles. Lucanids are usually associated with decaying wood and logs in coniferous and deciduous forest habitats. Adults of some species are attracted to lights at night and some feed at sap flowing from tree wounds. Adults of smaller species may feed on flowers. The eggs are customarily laid in crevices in bark or logs, and the larvae feed on decaying wood.

**Ochodaeidae:** sand scarabs. A widely distributed family of small, mostly brown non-metallic beetles ranging in length from 3-10mm. Predominantly active at night, they are most often collected at lights, sometimes in large numbers. Adults of a few species are active during the day. Most species prefer sandy areas, and many stridulate. Adults may spend the daylight hours in subterranean burrows, and may feed on fungi. Little else is known about the habits of either adults or immature stages.

**Trogidae:** skin and hide beetles. A small family of 300 species that are found on all continents. Members of the family are 2-20mm long, oval or oblong, and easily recognized by their overall warty, brown to grey or black, dirt-encrusted appearance, and their flat abdomen. Adults and larvae are found in the dry remains of dead animals; they are among the last of the succession of insects that invade and degrade carcasses. They are also found in the nests of birds or mammals where they feed on loose hair, feathers, or skin. When disturbed or frightened, adults feign death and remain motionless. This, in combination with their dirt-encrusted appearance, enables them to evade predators that might be scavenging in a carcass. Adults stridulate.

**Scarabaeidae:** the true or shining scarab beetles. Adults of many scarab beetles are noticeable due to their large size, bright colors, elaborate ornamentation, and interesting life histories. The family includes the giant Goliath Beetles from Africa, which are the heaviest insects (up to 100 g). The group includes nearly 28,000 species worldwide, an intriguing array of life histories, and many interesting adaptations. Lengths range from 2-180mm. Shape and colors vary. They are incredibly diverse; adults feed on dung, carrion, fungi, vegetation, pollen, fruit, compost, or roots. Some live in the nests of ants, termites, rodents, or birds. Species of dung beetles care for their larvae or the larval brood ball. Adults of some scarabs are diurnal and are seen on flowers or vegetation. Many other species are nocturnal and commonly come to lights at night. Adults or larvae of a few species are economically important because they cause considerable damage to turf or ornamental plants as they feed. Other scarabs are beneficial because they pollinate plants, recycle plant materials, and are major dung recyclers.

Many anatomical features of insects have unfamiliar names. Some that are in the table on the next page are described below. Do not focus on learning beetle anatomy; instead, focus on determining who is related to whom.

**Antennae, antennal club:** in all scarabs, the antennae end in 3-7 oval or flattened plates. The plates can be several different shapes or lengths, and the base of the antennae can be long and thick, short, thin, etc. To protect the plates or leaves of the antennal club, some species can close them into a single ball.

**Apical veins:** wings of all insects have “veins”, but they do not carry blood. Instead they are air-filled tubes used to inflate the wings when the insects mature. The pattern of veins often changes as insects evolve, so the presence or absence of certain veins is an important diagnostic trait. Apical veins are simply the veins near the wing tips.

**Labrum:** the tough chitinous “upper lip” that covers all of the mouthparts.

**Mandibles:** the jaws. In insects, these move side to side rather than up and down.

**Mentum:** a chitinous plate that connects to the labrum. On a human face, the mentum would be about where your nose is, but would be stiffer.

**Sternites:** the plates of the exoskeleton that cover the abdomen.

**Tibial spurs or spines:** on insects, the lower leg is called the tibia. Each tibia often has specialized spurs, spines, combs, or other appendages growing out of it. These specializations are used in grooming, mating, etc.

## Morphological Features of the Scarab-like Beetles

	Glaresidae (outgroup)	Passalidae	Lucanidae	Trogidae	Ochodaeidae	Scarabeidae
Front tibia flattened for digging	Yes	Yes	Yes	Yes	Yes	Yes
Antennal club closes	Yes	No	No	Yes	Yes	Yes
Mentum has central horn	No	Yes	No	No	No	No
Exoskeletal extension divides eye	Yes	No	No	No	No	No
# of ventral sternites	5	5	5	5	6	6
Middle tibial spurs are:	Smooth	Smooth	Smooth	Smooth	Comb-like	Smooth
Length of mandibles, compared to labrum	Shorter	Longer	Longer	Shorter	Shorter	Shorter
Basal segment of antennae	Thin	Robust	Robust	Robust	Thin	Thin
# of apical veins in wing	0	2	2	2	1 to 2	1
Shape of genitals in males	3 lobes	3 lobes	3 lobes	3 lobes	2 lobes	2 lobes
Mandible extensions are toothed or brushlike	Toothed	Brush-like	Brush-like	Brush-like	Toothed	Toothed

## C. Morphological Phylogeny: The Rodents

## Morphological Features of the Rodents

	L'morpha (o'grp)	Caviidae	Geomyidae	Sciuridae	Muridae	Heteromyidae	Castoridae	Cricetidae
Diastema present?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hard palate end	Above molars	Above molars	Above molars	Above molars	Beyond molars	Above molars	Above molars	Above molars
Fibula contacts calcaneum	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Masseter m. extends forward on jaw.	Deep head only	Deep head only	Both heads	Lateral only	Both heads	Both heads	Both heads	Both heads

Structure of forelimbs	Generalized	Generalized	Fossorial, clawed	Generalized	Generalized	Fossorial, clawed	Generalized	Generalized
Masseter m. passes thru infraorbital canal	Yes	Yes	No	No	No	No	Yes	No
Number of incisors	4	2	2	2	2	2	2	2
External, fur-lined cheek pouches	None	None	Present	None	None	Present	None	None
Size of tympanic bullae	Normal	Normal	Normal	Normal	Normal	Enlarged	Normal	Normal
Orientation of masseter fibers to infraorbital canal	Parallel	Parallel	Parallel	Parallel	Transverse	Parallel	Parallel	Transverse
Thickness of cranial bones	Thick	Thick	Thick	Thick	Thick	Thin	Thick	Thick

Among mammals, the rodents (Order Rodentia) are among the most successful in terms of sheer numbers. There are approximately 34 living families, and are distributed worldwide. One or more species of rodents are major elements within nearly all ecosystems. However, it is difficult to sort out their phylogenetic relationships, because there are numerous examples of convergent evolution in rodents. One major feature that all rodents share is their chisel-like front incisor teeth. These teeth grow throughout life, and must be sharpened continuously by gnawing or the animal will die. The second major feature is a diastema, a large gap or space that separates their incisors from the other teeth. The musculature and bones of rodent skulls are also extensively modified to increase the cutting force generated. This adaptations partly why rodents are so notorious for their destructive chewing habits.

Rodents have been subdivided many different ways by various authors. For the purposes of this phylogenetic analysis, we will look at 7 families, plus 1 outgroup:

**Castoridae:** beavers. Once a much larger group (which included some giant forms), there is a single living genus of beavers. They are highly adapted for their aquatic habitat, with a flattened tail and webbed digits. Their tail is naked and scaly. Mostly nocturnal, beavers feed mainly on aspen, poplar, birch, maple, willow and alder trees growing near sluggish bodies of water. They fell smaller saplings first, but will readily cut down a larger tree to obtain the tender small branches at the tips of limbs. They are loosely colonial, and several adults may share in damming, maintaining, and defending a single beaver pond.

**Geomyidae:** pocket gophers. These are small to medium rodents (120–230mm) with fur-lined external cheek pouches. They are burrowers, and solitary. Feed mostly below ground, so are rarely seen. Tunnels are commonly seen in southern and middle Rocky mountains. Considered pests in agricultural areas, where they destroy root crops, cut roots of alfalfa, and pile up dirt which prevents harvesting.

**Heteromyidae:** pocket mice, kangaroo rats. Small rodents with fur-lined cheek pouches. Hind feet are strongly developed and elongated like a rabbit's. Tail is long and well-developed, usually furred. Front feet are reduced, so they travel mainly by hopping. Highly adapted to arid areas, they burrow in the ground, and many species are able to survive without ever drinking from open water sources.

**Cricetidae:** hamsters, field mice, lemmings, voles. Small to medium sized (50–250mm). In voles, hamsters, and lemmings, tails are typically short or absent, with short hair; eyes are small. Field mice have large ears and eyes and long tails. Mostly are ground or burrow-dwellers. Mostly foragers. Many species are major prey items in ecosystems. Pet hamsters also belong to this group.

**Muridae:** Old World rats, mice. Originally from Europe, these rodents have been accidentally introduced nearly worldwide. Tail is invariably long, naked and scaled. They are the group from which albino laboratory rats and mice originated.

**Sciuridae:** squirrels, chipmunks, marmots, prairie dogs. All have hair-covered, often bushy tails. Almost all are active in the day, and all but the tree squirrels are ground or burrow nesters. Most store caches of food, and many hibernate in areas with harsh winter conditions. Ground-dwelling species have internal cheek pouches used to carry food to caches or a burrow before eating. In some areas, various species become localized pests. Several are also common enough to be major prey items for larger carnivores.

**Caviidae:** guinea pigs and Patagonian hares. While they are chunky with moderately short limbs, Caviids often have vestiges of musculature and bone structures that are commonly found in jumping animals like rabbits. One species, the Patagonian hare, even leaps like a rabbit. They are social herbivores, and often live in large colonies.

**Lagomorpha:** the rabbits, hares, and pikas. Rabbits are not rodents, but share several features with them. Lagomorphs are considered the closest group to the rodents, so they will be the outgroup for your phylogenetic analysis.

Some of the anatomical terms that may be unfamiliar are described below. Do not focus on learning the anatomy; instead, focus on determining who is related to whom.

**Calcaneum:** a key support bone in the ankle.

**Diastema:** a space separating incisor teeth from the remaining premolar teeth.

**Fibula:** one of the two bones that comprise the lower or hind limb in mammals.

**Fossorial:** adapted to living underground and digging. Fossorial limbs typically have thick bones with enlarged ridges for insertion of muscles used in digging. Claws are often thick and large as well.

**Infra-orbital canal:** a space or channel running along the lower surface of the bony socket surrounding the eye (the orbit).

**Masseter muscle:** the primary muscle used for chewing or movements of the lower jaw. The muscle can be divided into multiple sections, called heads, that can be enlarged or reduced singly or collectively.

**Tympanic bullae:** a small rounded region encasing the bones of the middle ear. It is located just behind the entrance to the auditory canal on the skull.

#### D. Morphological Phylogeny: the Caryophyllales

This group seems an odd mix at first glance, because it includes weeds, plants that are used for cut flowers or for landscaping, and several cacti! Yet this is not a random grouping; morphological features clearly indicate that they are a single clade. Families represented in the Caryophyllales are outlined below:

**Phytolacaceae:** pokeweeds. These are the large purple-stemmed perennial plants that are common in disturbed pastures, along fences, and in overgrown areas. Edible when they first emerge, pokeweed leaves become mildly poisonous once they mature. The purple berries are also poisonous, and readily stain clothing or other surfaces when they begin to fall in late summer or early fall.

**Cactaceae:** cactuses and related succulents. The green, fleshy portion of a cactus is not a leaf. It is a stem that has been extensively modified to act as a water storage organ. The spines on the cacti are the actual leaves.

**Caryophyllaceae:** carnations, pinks, dianthus, chickweed, andampions. You are probably very familiar with members of this family used by florists. However, the family also contains the pinks, which are ornamental flowers that are commonly planted as annuals in southern gardens. Champions are common, but not well known by most people. Chickweed is a very common broadleaf weed in lawns, and is often difficult to eliminate.

**Portulacaceae:** the true purslane family. Another succulent, low-growing herb that is commonly found in sandy soils or waste places. Occasionally, purslane will become a pest in lawns, especially in areas with sandy soil.

**Aizoaceae:** ice plants and sea purslanes. These are low-growing annual or perennial herbs with thick, fleshy or succulent leaves. Sea purslanes are fairly common wildflowers, particularly in sandy areas or beaches. The ice plants have been cultivated extensively, and several varieties are commercially available as groundcover plants for landscaping. Most of the cultivars tolerate poor, arid soil, and are common in rock gardens.

**Simmondsiaceae:** the jojoba family. These evergreen shrubs are common in many desert or arid areas such as the American Southwest. The fleshy leaves have relatively thick blades with smooth margins. Distal limbs are often fleshy rather than woody. Superficially, jojoba shrubs resemble the large jade trees that are commonly grown as houseplants.

**Polygonum:** knotweeds. Several species in this family are common broadleaf weeds in lawns and golf courses.

Some anatomical features of plants have unfamiliar names. Those used in the table are described below. Do not focus on learning plant anatomy; instead, focus on determining who is related to whom.

**Crystals:** some plants secrete sharp calcium or other crystals into the leaves to deter herbivores.

**Physiology:** in case your plant biology is rusty (or absent), all plants use photosynthesis to produce sugars. Some plants absorb the necessary carbon dioxide immediately before they use it to produce glucose. These are referred to as C3 plants. Other plants absorb carbon dioxide and temporarily store it bound to an intracellular molecule; these are referred to as CAM plants.

**Traces:** when buds or scales form at the nodes (leaf junctions) of a plant, they create scars on the stem called traces. The arrangement of traces can be seen under a microscope.

**Xylem & Phloem:** the xylem carries water and soil minerals from roots to shoots. Phloem carries sugars and water from leaves to storage areas, often in the roots. The arrangement of these tubes is very consistent within a related group of plants.

Morphological Features of the Caryophyllales

	Polygonum (out-group)	Phytolacaceae	Cactaceae	Caryophyllaceae	Portulacaceae	Aizoaceae	Simmondsiaceae
# traces/node	3	1	1	1	1	1	1
Crystals in leaves	Absent	Present	Absent	Absent	Absent	Present	Absent
Embryo shape	Straight	Curved	Curved	Curved	Curved	Curved	Straight
Flower pigment	Anthocyanins	Betalains	Betalains	Anthocyanins	Betalains	Betalains	Anthocyanins
Physiology	C3	C3	CAM	C3	CAM	C3	C3
Xylem & phloem arrangement	Single ring of each type	Concentric rings	Concentric rings	Concentric rings	Concentric rings	Concentric rings	Concentric rings
Stamen #	<10	>20	>20	<10	>20	>20	<10

### E. Molecular Phylogeny: the Cetaceans

This group includes whales, dolphins, and porpoises. They can be subdivided into two major groups: Odontoceti, or toothed whales, and Mysticeti, or baleen whales. In baleen whales, teeth have been replaced by plates that act as strainers, allowing baleen whales to feed on plankton and small marine invertebrates. Toothed whales are active predators that hunt a variety of foods including squid, fish, or other marine mammals.

The question you must answer is, are all baleen whales derived from one ancestor? In other words, did their filter-feeding lifestyle evolve once, or more than once?

**Some Common Species of Cetaceans:**

Blue whale (*Balaenoptera musculus*), a baleen whale  
 Rorqual or Sei Whale (*Balaenoptera borealis*), a baleen whale  
 Finback whale (*Balaenoptera physalus*), a baleen whale  
 Humpback whale (*Megaptera novaeangliae*), a baleen whale  
 Sperm whale (*Physeter macrocephalus*), a toothed whale  
 Beluga whale (*Delphinapterus leucas*), a toothed whale  
 Killer whale (*Orcinus orca*), a toothed whale  
 False killer whale (*Pseudorca crassidens*), a toothed whale  
 \*\*Atlantic bottle-nosed dolphin (*Tursiops truncatus*), the outlier group

**Possible Sequences to Use:**

Cytochrome B (a protein used by mitochondria)  
 Insulin  
 Hemoglobin A or B chain  
 Myoglobin

**Hints:** assign dolphins as the outlier group. You may not be able to find the same sequence data for every possible species. Your overall goal is to determine if all the whales that have baleen are descended from one common ancestor, or have evolved more than once from the toothed whales. Start by searching the word “cetacean” in the protein database, and see which sequences from the list are present for the most species. From there you can choose which species and protein to use.

**F. Molecular Phylogeny: the Caryophyllales**

This is the same group for which you may already have a parsimonious tree based on morphology. If you did not work with them previously, this group includes several plants that, at first glance, do not seem related. Morphological features though suggest they are close kin. Molecular data are available for:

**Cactaceae:** cactuses and related succulents. The green, fleshy portion of a cactus is not a leaf. It is a stem that has been extensively modified to act as a water storage organ. The spines on the cacti are the actual leaves.

**Caryophyllaceae:** carnations, pinks, dianthus, chickweed, andampions. You are probably very familiar with members of this family used by florists. However, the family also contains the pinks, which are ornamental flowers that are commonly planted as annuals in southern gardens. Campions are common, but not well known by most people. Chickweed is a very common broadleaf weed in lawns, and is often difficult to eliminate.

**Portulacaceae:** the true purslane family. Another succulent, low-growing herb that is commonly found in sandy soils or waste places. Occasionally, purslane will become a pest in lawns, especially in areas with sandy soil.

**Aizoaceae:** ice plants and sea purslanes. These are low-growing annual or perennial herbs with thick, fleshy or succulent leaves. Sea purslanes are fairly common wildflowers, particularly in sandy areas or beaches. The ice plants have been cultivated extensively, and several varieties are commercially available as groundcover plants for landscaping. Most of the cultivars tolerate poor, arid soil, and are common in rock gardens.

**Simmondsiaceae:** the jojoba family. These evergreen shrubs are common in many desert or arid areas such as the American Southwest. The fleshy leaves have relatively thick blades with smooth margins. Distal limbs are often fleshy rather than woody. Superficially, jojoba shrubs resemble the large jade trees that are commonly grown as houseplants.

**Polygonum:** knotweeds. Several species in this family are common broadleaf weeds in lawns and golf courses.

If you did not create a phylogenetic tree for this group previously, ask a classmate to borrow theirs before you begin this exercise. The question you must answer is: do molecular data support the phylogenetic tree created using morphological features?

**Some Common Species of Caryophyllales:**

Golden ball cactus (*Parodia leninghausii*)  
 Other miscellaneous cacti (*Neoporteria*, *Eriosyce* sp.)  
 Pinks and carnations (*Dianthus* sp.)  
 Portulacaceans (examples include *Calandrinia* sp., *Claytonia* sp., *Lewisia* sp.)  
 Common ice plant (*Mesembryanthemum crystallinum*)  
 Jojoba (*Simmondsia chinensis*)  
 \*\*Common knotweed (*Polygonum* sp.)—assign this species as the outlier

You are free to substitute other species in the same genus as the representatives for a particular family. For example, you could use any species of *Parodia*, *Neoporteria*, or *Eriosyce* to represent the family Cactaceae.

**Note:** when you search for plant sequences, you may not find enough proteins in Swiss-Prot's database. If this happens, switch to using DNA sequences. First, go back to the main page and switch to **Nucleic Acid Tools**. In the NDJINN search, use **GBPLN**, which is a subset of the Genbank database that contains DNA sequences from plants, fungi and algae.

**Some sequence options:**

You are most likely to find the sequences for the large or small subunits of ribulose biphosphate carboxylase (abbreviated rbcL and rbcS, respectively). These 2 proteins are essential for photosynthesis, and have been used to establish many phylogenetic relationships between photosynthetic organisms.

Internal transcribed spacers 1 or 2 (ITS-1, -2). These DNA sequences are found in nearly all organisms, and have been used many times for molecular cladistic analyses. However, they may not be available for all plants.

18S ribosomal RNA. Ribosomes are another universal feature of all organisms. Like the ITS sequences though, the sequence may not be available for all organisms.

### G. Molecular Phylogeny: the Fagales

This is the same group for which you may already have a parsimonious tree based on morphology. It includes oaks, birches, myrtles, beeches and walnuts. Several members of this group are economically important for various reasons. Fagales include these families:

**Myricaceae:** bayberry, wax myrtle. These small trees or shrubs produce a fragrant resin that was the original source for the scent of bayberry soaps and candles. Wax myrtles are salt tolerant, and often planted in coastal areas as windbreaks or ornamental hedges.

**Fagaceae:** beeches and oaks. Beeches are medium-sized trees, with species distributed nearly worldwide. In the US, the beeches are often found as climax species in hardwood forests. The fruit is 2 or 3 edible nuts within a small bur with soft spines. The oaks are also found worldwide. Oaks are slow-growing under normal conditions, making the wood exceptionally dense and hard. Like beeches, oaks are commonly the dominant climax species in a hardwood forest. The nuts or acorns of both types of trees are a staple in food caches of squirrels and other foraging mammals.

**Juglandaceae:** walnuts, hickories, and pecans. In addition to providing nuts, walnut wood is often used for fine furniture. Because it is so dense, hickory wood is often used for tool handles. Walnut trees produce several allelochemicals (the most common is juglone) that act as herbicides and kill all competing ground cover beneath the trees.

**Betulaceae:** birches. These slender, often multi-trunked trees have relatively thin, flexible bark that often peels from the trees in sheets. Many species are adapted to riverbanks and other moist habitats.

**Cucurbitaceae:** squashes. This large and diverse group contains pumpkins, winter squashes, and several other vegetables that form on large mounding vines.

If you did not create a phylogenetic tree based on morphology for this group previously, ask a classmate to borrow theirs. The question you must answer is: do molecular characteristics support the cladogram created using morphological features?

**Some Common Species of Fagales:**

Beechnut (*Fagus sylvatica*)

River birch (*Betula nigra*)

Wax myrtle (*Myrica cerifera*)

Red oak (*Quercus rubra*)

Black walnut (*Juglans nigra*)

\*\*Squashes (*Cucurbita pepo*, *C. maxima*, *Ovifera* sp.)—these are your outgroup

You are free to substitute other common species (such as pecan for walnut, white oak for red oak, etc.) as representative of the various families in the Fagales.

**Note:** when you search for plant sequences, you may not find enough proteins in Swiss-Prot's database. If this happens, switch to using DNA sequences. First, go back to the main page and switch to **Nucleic Acid Tools**. In the NDJINN search, use GBPLN: the subset of the Genbank database that contains DNA sequences from plants, fungi and algae. Then proceed as usual.

**Possible Sequences to Use:**

18S ribosomal RNA. Ribosomes are a universal feature of all organisms. The sequences may not yet be available for all organisms, but are among the first sequences chosen by molecular phylogenists.

You are also likely to find several sequences for the large or small subunits of ribulose biphosphate carboxylase (abbreviated *rbcL* and *rbcS*, respectively). These 2 proteins are essential for photosynthesis, and have been used to establish many phylogenetic relationships between photosynthetic organisms.

Internal transcribed spacers 1 or 2 (ITS-1, -2). These DNA sequences are found in nearly all organisms, and have been used many times for molecular cladistic analyses. However, they may not be available for all plants.

**Hints:** you may not be able to find the same sequence data for every possible species. Your overall goal is to determine if morphological features and molecular evidence match up. Start by searching the word "Fagales" in the database, and see which sequences are present for the most species. Alternatively, enter a single genus (say, "Quercus") and see which sequences are available to you. From there you can choose which species and protein to use.

## H. Molecular Phylogeny: the Primate group Haplorhini

Primates are divided into two major suborders: Strepsirhini, and Haplorhini. The Strepsirhini, or prosimians, include animals such as lorises, lemurs, and pottos. The Haplorhini, or simian (ape-like) primates, includes tarsiers (a small nocturnal primate the size of a kitten), Old and New World Monkeys, gibbons, and the great apes. The press has given a lot of coverage to the fact that humans are closely related to chimpanzees. But how close are we? Furthermore, what are our relationships with the other members of the Haplorhini?

Once you work through this phylogenetic analysis, answer these questions: 1) are we really closer to chimps than any other great ape? 2) how are the other apes related to one another?

**Representative Species of Haplorhini:**

Unlike great apes, there are dozens of species of gibbons, Old World monkeys, and New World monkeys. The genera listed below have been studied extensively, so they should be well-represented in the molecular sequence databases.

- Humans (*Homo sapiens*)
- Chimpanzee (*Pan troglodytes*)
- Gorilla (*Gorilla gorilla*)
- Orangutan (*Pongo pygmaeus*)
- Gibbons (*Hylobates lar*, *H. agilis*, *H. concolor*, *Hylobates* sp.)
- Old world monkeys (*Macaca* sp., *Cercopithecus* sp., *Papio* sp.)
- Tarsier (*Tarsius bancanus*, *T. syrichta*, *T. pumilus*, *T. spectrum*)
- \*\*New world monkeys (*Cebus* sp., *Aotus* sp., *Callicebus* sp.): your outgroup

**Some potential sequences to use:**

- Hemoglobin B chain
- Hemoglobin A chain
- Insulin
- Myoglobin (an oxygen storage pigment in muscle)
- Cytochromes (used for mitochondrial energy production)
- 18S ribosomal RNA

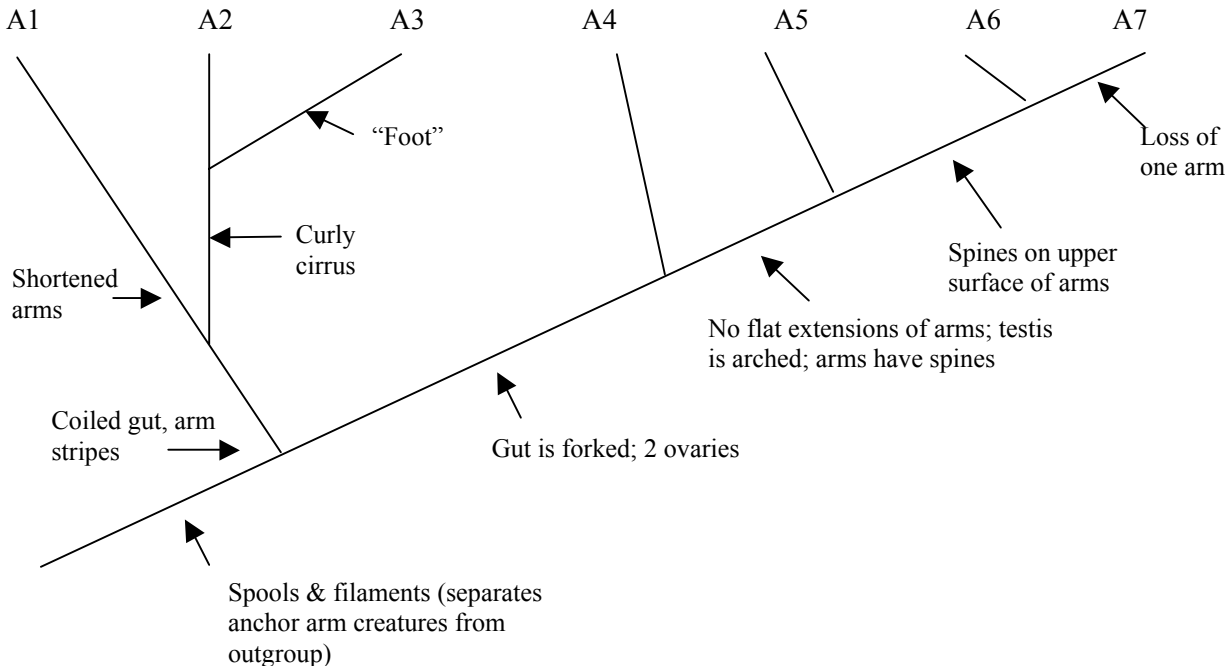
**Note:** as you search for sequences, you are not restricted to using just SwissProt's database. A good alternative is the Genbank database for DNA sequences from primates. You can even search both simultaneously; when you start an NDJINN search, just check the box next to SwissProt, and also the box next to the Genbank Primate Database.

**Hint:** First, do not try to download and align multiple sequences for each species, or find every possible member of large groups like the Old World monkeys. Instead, pick one representative from each major group, and one protein or DNA sequence to use. Also, if you have trouble finding sequences, search the databases using the genus names. You will get more sequences than if you used a species name, but fewer sequences than if you search with a generalized term like "primates" or "apes".



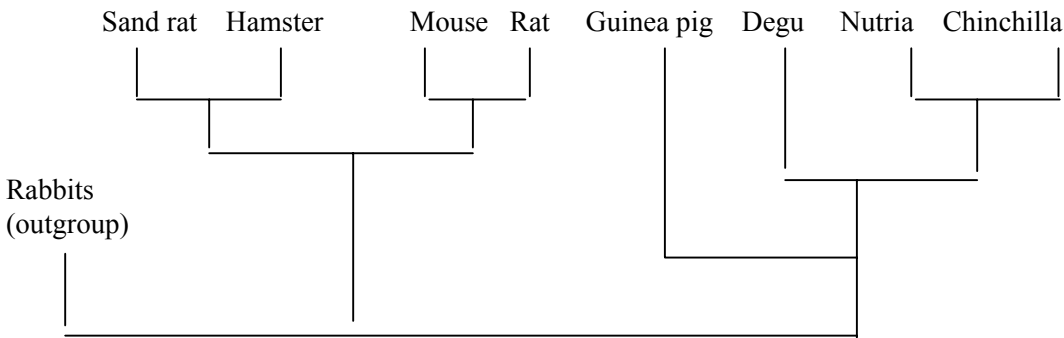
**Appendix C: Typical Results of Students' Analyses**

**Week 1: Phylogeny of Anchor Arm Creatures (based on morphology)**



The figure shows the most commonly derived arrangement, with key features next to each branch. However it is NOT the only arrangement students generate. Different views of the same features are routine in systematics. The lab exercise tells students that disagreements are precisely why systematists use many different datasets to generate a consensus tree. Use any disagreement between students' trees to reinforce this idea.

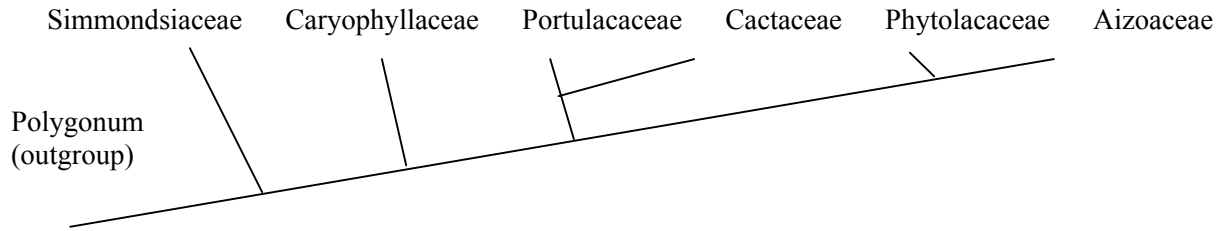
**Week 2: Phylogeny of Rodents (based on insulin protein sequences)**



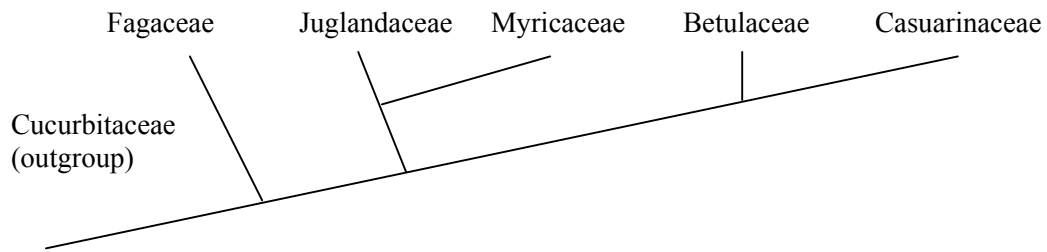
This is the tree usually created by ProtPars in Exercise #3. The 4 South American rodents form a single cluster. Eurasian sand rats and hamsters, and European rats and mice, each form 2 more obvious subgroups. In this example, geography rather than lifestyle is the better predictor of which species are related.

**Key to Homework Assignments**

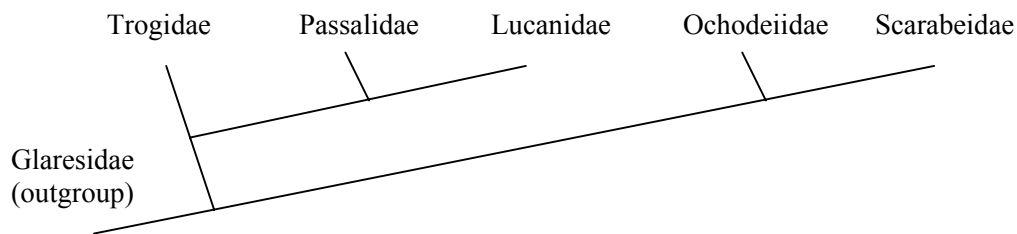
**Typical Phylogeny of Caryophyllales, based on morphology**



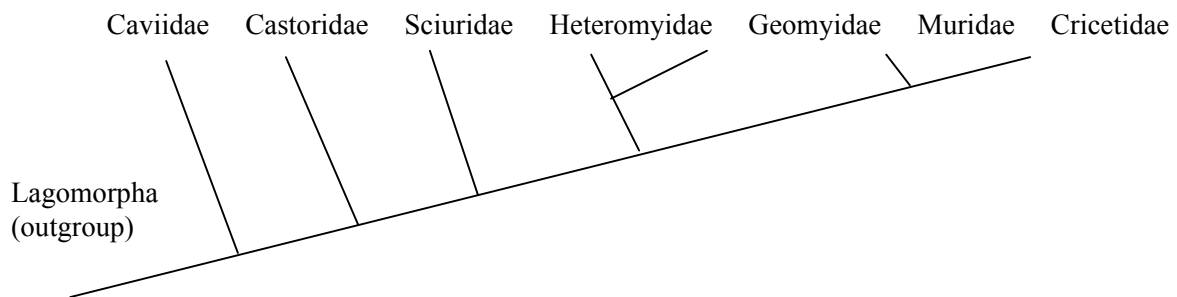
**Typical Phylogeny of Fagales, based on morphology**



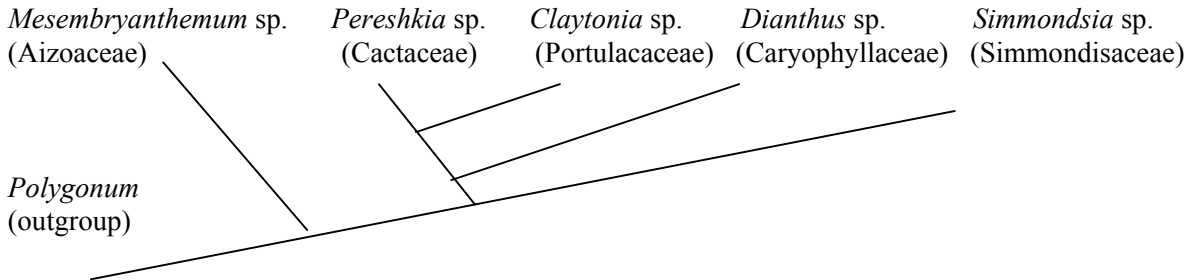
**Typical Phylogeny of Scarab beetles, based on morphology**



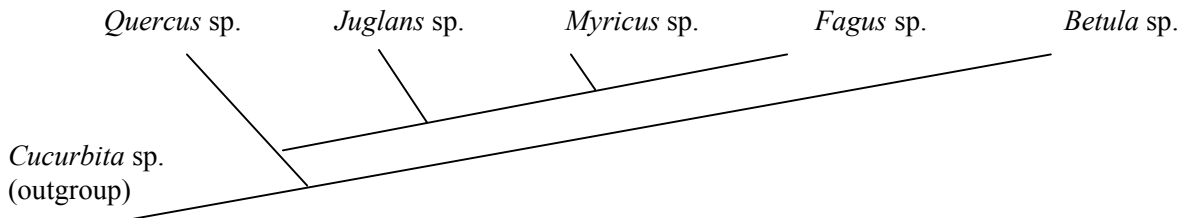
**Typical Phylogeny of Rodents, based on morphology**



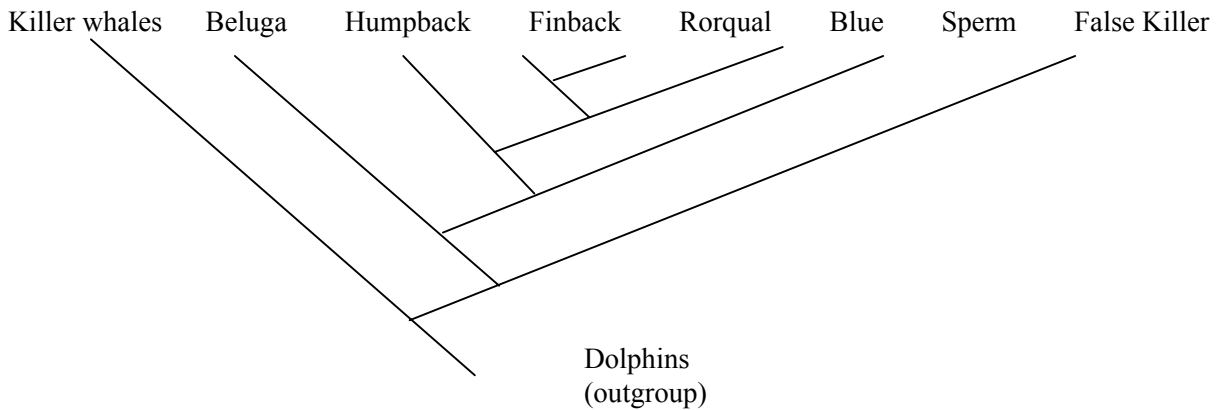
**Phylogeny of Caryophyllales, using rbcL DNA sequence**



**Phylogeny of Fagales, using 18S ribosomal DNA sequence**



**Phylogeny of Cetaceans, using Cytochrome B protein sequence**



**Phylogeny of Haplorhini, using hemoglobin B chain sequence**

