



ASSOCIATION FOR BIOLOGY LABORATORY EDUCATION

This article reprinted from:

Kosinski, R.J. 2006. An introduction to phylogenetic analysis. Pages 57-106, in *Tested Studies for Laboratory Teaching, Volume 27* (M.A. O'Donnell, Editor). *Proceedings of the 27th Workshop/Conference of the Association for Biology Laboratory Education (ABLE)*, 383 pages.

Compilation copyright © 2006 by the Association for Biology Laboratory Education (ABLE)
ISBN 1-890444-09-X

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>



An Introduction to Phylogenetic Analysis

Robert J. Kosinski

Department of Biological Sciences
132 Long Hall
Clemson University
Clemson, SC 29634-0314
Voice: (864) 656-3830 FAX: (864) 656-0435
rjksn@clemson.edu

Abstract: This exercise introduces students to analysis of evolutionary relationships both by analysis of molecular similarity and by cladistic techniques. The molecular analysis uses both paper-and-pencil and bioinformatics comparison of the amino acid sequence of the hemoglobin beta chain of eight vertebrates. The cladistic exercises introduce students to cladistic principles, and then allow students to solve hypothetical problems both with and without homoplasy. Students then test their cladistic ability by analyzing published data on the flightless birds. Finally, students solve an entertaining problem using “organisms” made out of nuts and bolts.

Keywords: cladistics, cladogram, phylogram, phylogeny, phylogenetic, evolution, taxonomy, bioinformatics

©2006 Robert J. Kosinski

Contents:

Student Outline	58
Classifying Based on Molecular Similarity	
Exercise A—Molecular Similarity (manual exercise)	58
Exercise B—Molecular Similarity (bioinformatics exercise)	61
Classifying Using Cladistic Analysis	65
Exercise C—Solving Cladistic Problems	69
Exercise D—The Flightless Birds	77
Exercise E—The “Boltidae”	78
Notes for the Instructor	79
Literature Cited	85
About the Author	85
Instructor Appendices	
Appendix A—Student Problem Handouts	86
Appendix B—Student Solution Handouts	94
Appendix C—Instructor Solution Pages	98
Appendix D—Flightless Birds Solution (transparency master)	106

Student Outline

Our system of taxonomy is based on phylogeny. That is, we classify organisms together because they have a common evolutionary ancestor. In most cases, we cannot determine ancestry directly because the fossil record is poor for most organisms. Instead we rely on shared, homologous features, and we say that organisms that share many features are closely related. Organisms that share many features probably had a relatively recent common ancestor.

For example, chimpanzees and humans share about 98% of their DNA because the common ancestor of chimps and humans lived only about 6 million years ago. In 6 million years, there has not been enough time for very much divergence to take place. However, the DNA of humans and yeast is more dissimilar because humans and yeast shared an early eukaryotic ancestor no more recently than about 1.2 billion years ago.

Evolutionary relatedness between two species is determined by the recency of their common ancestor. Recency of the ancestor can usually be determined by the two species' degree of similarity.

In this laboratory, we will practice two methods of determining the evolutionary relationships between organisms: molecular similarity, and cladistic analysis.

Classifying Based on Molecular Similarity

Comparison of the sequence differences in either nucleic acids or proteins can indicate relatedness. This method has the advantage that it is relatively easy to understand. Highly related organisms are expected to have many sequence similarities; distantly related ones are expected to have low similarity. There are two possible exercises here: Do Ex. A if you do not have computers in your lab, and Ex. B if you do have computers with Internet access. In both cases, we are going to draw conclusions about the relationships of humans (*Homo sapiens*), rhesus monkeys (*Macaca mulatta*), mice (*Mus musculus*), rats (*Rattus norvegicus*), domestic ducks (*Anas platyrhynchos*), Canada geese (*Branta canadensis*), Nile crocodiles (*Crocodylus niloticus*), and alligators (*Alligator mississippiensis*). Obviously, the eight species consist of four seemingly related pairs. What do you predict about the molecular similarity of organisms within each pair?

In both exercises A and B, we will use single-letter codes of the International Union of Pure and Applied Chemistry (IUPAC) to represent amino acids.

Table 1. Single-letter IUPAC codes for the 20 standard amino acids.

A alanine	G glycine	M methionine	S serine
C cysteine	H histidine	N asparagine	T threonine
D aspartic acid	I isoleucine	P proline	V valine
E glutamic acid	K lysine	Q glutamine	W tryptophan
F phenylalanine	L leucine	R arginine	Y tyrosine

Exercise A (if computers with Internet access are *not* available)

Table 2 on the next page shows the amino acid sequences for the first 60 amino acids of the beta chain of hemoglobin for the species above. Not all of the amino acids in the beta chain are shown because some of them are the same for all the species.

Table 2. The first 60 amino acids in the beta chain of hemoglobin in eight species.

Amino Acid	Human	Rhesus Monkey	Mouse	Rat	Domestic Duck	Canada Goose	Nile Crocodile	Alligator
1	V	V	V	V	V	V	A	A
2	H	H	H	H	H	H	S	S
3	L	L	L	L	W	W	F	F
4	T	T	T	T	T	T	D	D
5	P	P	D	D	A	A	P	A
6	E	E	A	A	E	E	H	H
8	K	K	K	K	K	K	K	R
9	S	N	A	A	Q	Q	Q	K
10	A	A	A	A	L	L	L	F
11	V	V	V	V	I	I	I	I
12	T	T	S	N	T	T	G	V
13	A	T	C	G	G	G	D	D
16	G	G	G	G	G	G	H	A
19	N	N	N	N	N	N	D	D
20	V	V	S	P	V	V	V	V
21	D	D	D	D	A	A	A	A
22	E	E	E	D	D	D	H	Q
23	V	V	V	V	C	C	C	C
25	G	G	G	G	A	A	G	A
26	E	E	E	E	E	E	E	D
29	G	G	G	G	A	A	S	S
31	L	L	L	L	L	L	M	M
33	V	L	V	V	I	I	I	I
38	T	T	T	T	T	T	K	K
39	Q	Q	Q	Q	Q	Q	R	R
41	F	F	Y	Y	F	F	Y	Y
43	E	E	D	D	A	S	E	E
44	S	S	S	S	S	S	N	H
47	D	D	D	D	N	N	D	K
48	L	L	L	L	L	L	I	M
49	S	S	S	S	S	S	S	C
50	T	S	S	S	S	S	N	N
51	P	P	A	A	P	P	A	A
52	D	D	S	S	T	T	Q	H
53	A	A	A	A	A	A	A	D
54	V	V	I	I	I	I	I	I
55	M	M	M	M	L	L	M	L
56	G	G	G	G	G	G	H	H
57	P	P	A	P	P	P	E	S
59	K	K	K	K	M	M	K	K

Procedure A

1. The first step is a laborious one: working in groups of four, you must find the number of *differences* in amino acid sequence between each pair of species. There are 28 pairs to be analyzed, as shown by the unshaded cells in Table 3 below:

Table 3. Differences in beta chain amino acid sequences between 28 pairs of species.

	Monkey	Mouse	Rat	Duck	Goose	Croc.	Allig.
Human							
Monkey							
Mouse							
Rat							
Duck							
Goose							
Croc							

You should divide the comparisons up among your group members so that each person is responsible for certain comparisons. To make sure you're doing the comparisons correctly, the human-rhesus monkey number should be 4 , and the crocodile-alligator number should be 17.

2. According to the data, which two species in Table 3 are most closely related? Why do we think these two lineages diverged the shortest time ago?
3. Which species is most closely related to humans? Which is most distantly related to humans?
4. To humans, alligators and crocodiles look very similar. However, which species difference in Table 3 comes closest to the difference between alligators and crocodiles? How can we explain this?

Exercise B (if computers with Internet access *are* available)

With modern bioinformatics tools, it is possible to do a more sophisticated analysis than is possible by hand. Also, protein databases have many organisms in them, so you can examine a great variety of evolutionary relationships. Let's attempt to look at the relationships of the beta chain of hemoglobin for our eight species using these tools.

Procedure B

1. "Swiss-Prot" is one of the world's leading protein databases. Go to the Swiss-Prot database at <http://us.expasy.org/sprot/>:

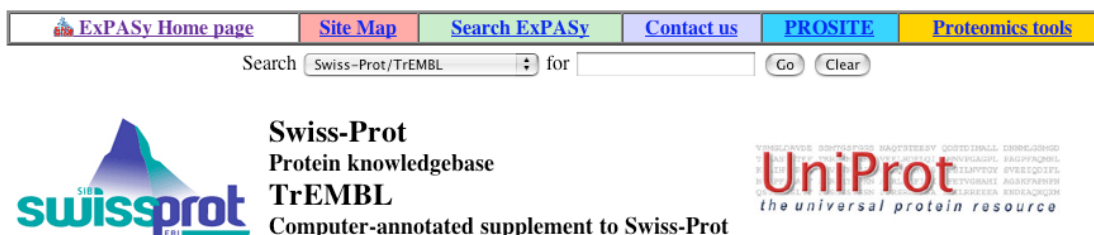


Figure 1. The Swiss-Prot protein database.

1. Using the text box at the upper right, enter the protein you want to find and the organism. For example, for human hemoglobin, we might put in "hemoglobin human" in the search box. Press the Go button. Do not use any ANDs or ORs in a Swiss-Prot search box.
1. You will get a page that lists several possible proteins, one of which is HBB_HUMAN, the beta chain of human hemoglobin. Click on that link and you'll be taken to a page that gives you the protein's official name, its gene name, many references on how it was sequenced, a summary of its function, and many other pieces of information.
1. However, now we just want to know the amino acid sequence. At the very bottom is information on the length (146 amino acids), molecular mass and amino acid sequence of the hemoglobin beta chain. The sequence is presented this way:

```

      10      20      30      40      50      60
      |      |      |      |      |      |
VHLTPEEKSA VTALWGKVNV DEVGGEALGR LLVYYPWTQR FFESFGDLST PDAVMGNPKV
      70      80      90     100     110     120
      |      |      |      |      |      |
KAHGKKVLGA FSDGLAHLDN LKGTFFATLSE LHCDKLHVDP ENFRLLGNVL VCVLAHHFGK
      130     140
      |      |
EFTPPVQAAV QKVVAGVANA LAHKYH

```

Figure 2. The hemoglobin beta chain sequence as presented by Swiss-Prot.

1. At the lower right is a link that refers to "FASTA Format." FASTA is a very common format for presenting either DNA or protein sequences to bioinformatics programs. If you click on the FASTA link, the sequence above changes to:

>sp|P02023|HBB_HUMAN Hemoglobin beta chain - Homo sapiens (Human), Pan troglodytes (Chimpanzee), and Pan paniscus (Pygmy chimpanzee) (Bonobo).

VHLTPEEKSAVTALWGKVVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK
VKAHGKKVLGAFSDGLAHLAHLNKGTFATLSELHCDKLHVDPENFRLGNVLCVLAHFF
GKEFTPPVQAAYQKVVAGVANALAHKYH

We could copy this amino acid sequence (starting with “VHLT...” onto our computer’s clipboard and then enter it into bioinformatics programs. We would then do a search for the hemoglobin beta chain of another organism, and so forth. For this exercise, we’ll use some hemoglobin sequences that have already been compiled, so don’t copy the human sequence above.

1. Using the Back function of your browser, go back to the main HBB_HUMAN page of Swiss-Prot, go to the top, and enter only “HBB” into the search box. Press Go. This asks the database to present all hemoglobin beta chains it has (about 500). This is an impressive diversity, representing organisms from stingrays to penguins to humans. This gives an idea of the variety of organisms that could be incorporated into your phylogenetic study using the hemoglobin beta chain. However, the sequences we need are already compiled on a Web site.
1. Go to <http://biology.clemson.edu/bpc/bp/Lab/111/hemo.htm> and click on the “Hemoglobin File” link. A Microsoft Word file will be downloaded to your desktop. It contains hemoglobin beta chain sequences for humans, rhesus monkeys, mice, rats, domestic ducks, Canada geese, Nile crocodiles, and alligators.
1. Open the hemoglobin file. Copy the human sequence (from “>human” to the end) onto your clipboard.
9. Go to a popular bioinformatics site: ClustalW at the European Bioinformatics Institute: <http://www.ebi.ac.uk/clustalw/index.html>. The “W” in this name stands for “Weights.”

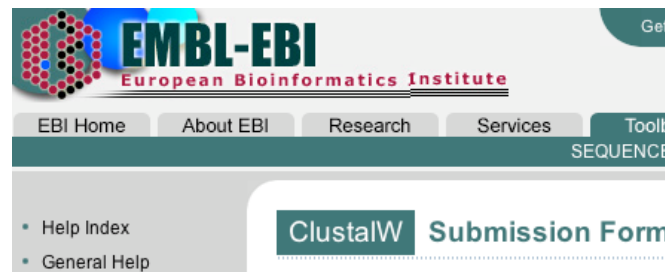


Figure 3. ClustalW at the European Bioinformatics Institute Web site.

ClustalW performs multiple alignments (it aligns more than two sequences at the same time so corresponding sections are being compared), it determines the relationships between them, and it finally clusters them on a tree-like structure called a cladogram or phylogram. We hope this structure shows the degree of relatedness between the protein or nucleotide sequences.

9. Go to your downloaded “Hemo” file, select all the text, and paste it in the text box on the submission form:

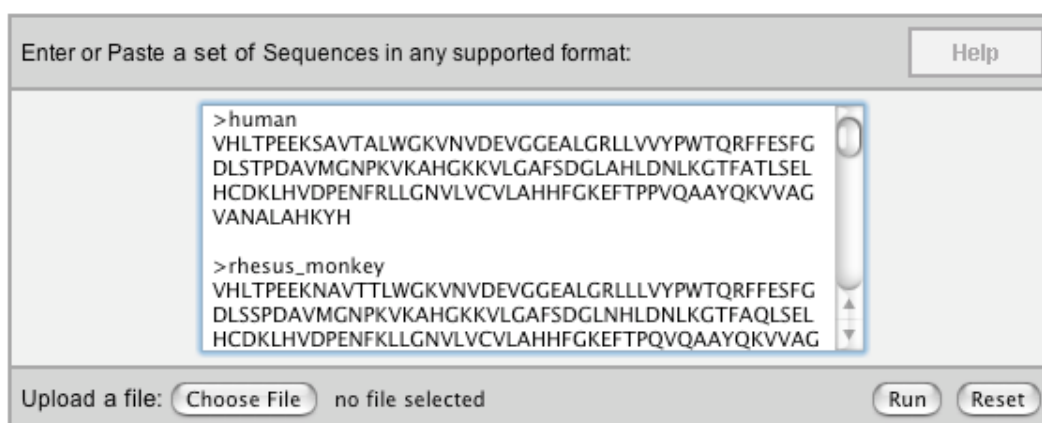


Figure 4. ClustalW ready to run with the hemoglobin sequences in the search box.

The only option to change is that “Output Order” (just above the text box) should be set to “Input” rather than to “Aligned.” Then press Run.

- After a short pause, you will get a screen that starts with a table of differences between the different organisms. For the hemoglobin beta chain, this table starts:

Scores Table

Sort by

SeqA Name	Len(aa)	SeqB Name	Len(aa)	Score
1 human	146	2 rhesus_monkey	146	94
1 human	146	3 rat	146	81
1 human	146	4 mouse	146	80
1 human	146	5 domestic_duck	146	69
1 human	146	6 Canada_goose	146	69
1 human	146	7 Nile_crocodile	146	54
1 human	146	8 alligator	146	48

Figure 5. The ClustalW hemoglobin results for comparison of the human with all the other species.

The scores in this table show that the hemoglobin beta chain sequences were 94% identical between humans and rhesus monkeys, 81% identical between humans and rats and so forth.

- At the bottom of the output, there is a cladogram, but this is something called a “guide tree” that has more to do with how ClustalW handled the problem than with evolution. So ignore that and go back up to the beginning of the multiple alignment, which starts as follows:

Alignment

[Show Colors](#)
[View Alignment File](#)

CLUSTAL W (1.82) multiple sequence alignment

```

human          VHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKV 60
rhesus_monkey  VHLTPEEKNAVTTLWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSSPDAVMGNPKV 60
rat            VHLTDAEKAAVNGLWGKVNPDVVGGEALGRLLVVYPWTQRYFDSFGDLSSASAIMGNPKV 60
mouse         VHLTDAEKAAVSCLWGKVNSDEVGGEALGRLLVVYPWTQRYFDSFGDLSSASAIMGNPKV 60
domestic_duck VHWTAEEKQLITGLWGKVNVDVCGAEALARLLIVYPWTQRYFDSFGDLSSPTAILGNPMV 60
Canada_goose  VHWTAEEKQLITGLWGKVNVDVCGAEALARLLIVYPWTQRYFDSFGDLSSPTAILGNPMV 60
Nile_crocodile ASFDPEKQLIGDLWHKVDVAHCGGEALSRLIVYPWKRRYFENFGDISNAQAIMHNEKV 60
alligator     ASFDAHERKFIVDLWAKVDVAQCADALSRMLIVYPWKRRYFEHFGMKCNHAHDILHNSKV 60
.             *:  :  ** **:  . *.:**.*:****.:** * *..  : * *
```

Figure 6. The beginning of the ClustalW alignment of the hemoglobin beta chains of the eight species.

This section aligns corresponding amino acids with one another so we may get an accurate estimate of how many amino acids have changed from one species to another. A “*” under a column means that all species had an identical amino acid in that spot, a “:” means all the amino acids were similar, and a “.” means they were less similar, and a “-” means one or more amino acids were markedly different. We’re now going to use these aligned sequences to do our evolutionary comparison.

9. Select all the text starting with “CLUSTAL W (1.82) multiple sequence alignment” and going down to the last line of the alignment that mentions amino acid 146.
9. Use the “back” key on your browser to go back to the ClustalW page where you entered your protein sequences. Clear all that text and paste in the contents of your clipboard. Then change the ClustalW parameters (last line above the text box) like this:

OUTPUT		PHYLOGENETIC TREE		
OUTPUT FORMAT	OUTPUT ORDER	TREE TYPE	CORRECT DIST.	IGNORE GAPS
aln w/numbers	input	nj	on	on

Figure 7. Settings for the second ClustalW run. “Nj” stands for “neighbor-joining,” a technique for doing trees. The distance and gaps setting may help to eliminate some inaccuracies.

9. Press “Run.” When your output is delivered, scroll down to the bottom and find the tree-like cladogram. Press the button labeled, “Show as a Phylogram Tree.” To understand the phylogram structure, refer to the phylogram below, which was done as an example using different organisms and triosephosphate isomerase, an enzyme in glycolysis:

Phylogram

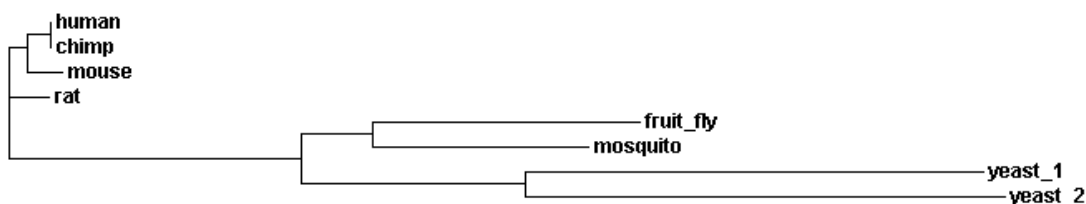


Figure 8. Phylogram showing ClustalW's comparison of an enzyme of several species. This is presented as a non-hemoglobin example of phylogram output.

The branches here group together related sequences, and the lengths of the branches indicate degree of difference between sequences. This phylogram indicates that the triosephosphate isomerase of humans, chimps, mice, and rats is rather similar, but this mammalian sequence is quite different from that of fruit flies, mosquitoes, and yeasts. Also, the two yeasts (with their long branches) differ from each other more than the fruit fly and mosquito differ from each other.

15. Now look at your hemoglobin output for our eight species. Which two species are most closely related? Why do we think these two lineages diverged the shortest time ago?

16. Which species is most closely related to humans?

16. To humans, alligators and crocodiles look very similar (much more similar than humans and rhesus monkeys, for example). However, according to the data, did the alligator and crocodile lineages separate before or after the human and rhesus monkey lineages separated? Why do we say this?

Classifying Using Cladistic Analysis

Cladistic analysis is a method of determining the relationships of groups by analyzing the features they share. If differences between lineages start to accumulate once lineages have separated, then many shared homologous features should indicate a recent common ancestor and a high degree of relatedness. On the other hand, few shared features indicate an ancient common ancestor and a low degree of relatedness.

Unfortunately, cladistics has a complicated terminology. You'll have to deal with the following terms:

Clade – a monophyletic lineage (a lineage that arose from one ancestor)

Character - some feature of an organism (e.g., either presence or absence of legs).

Ancestral (or plesiomorphic) character state - a character state that was seen in the early ancestors of an organism (e.g., for reptiles, having four legs is an ancestral state because the earliest reptiles had four legs).

Derived (or apomorphic) character state – an evolutionary innovation, or a character state that was *not* seen in the early ancestors of an organism (e.g., limblessness of snakes is a derived or apomorphic character state *for the reptiles*).

Synapomorphy - a *shared*, homologous, derived character state (e.g., all snakes are limbless, so limblessness is a synapomorphy for snakes).

Ingroup - the group (such as snakes) that is now being analyzed.

Outgroup - a group that is thought to be closely related to the ancestor of the ingroup, but is not part of the ingroup. Snakes evolved from lizards, so lizards would make a good outgroup for a phylogenetic analysis of snakes.

Parsimony – the scientific preference for the simplest possible explanation that will explain some data.

With this terminology, the first paragraph in this exercise might be rewritten as:

“Cladistics is the classification of an ingroup based on synapomorphic characters. If apomorphies start to accumulate once two lineages have separated, then numerous synapomorphies indicate a recent, common ancestor and a high degree of relatedness. On the other hand, few synapomorphies indicate an ancient common ancestor and a low degree of relatedness.”

Cladistic analysis starts with the thoughtful identification of characters. These must be relatively stable, heritable, must show both ancestral and derived states within the ingroup, and must be homologous and not be subject to convergent evolution. For example, number and structure of scales would make good characters for snakes because they do change within the snakes and reflect ancestry. Color would not be a good character because it's very variable and unstable, and totally unrelated snakes (and fish, birds, and insects too) might be the same color. The presence of a backbone would not be a good character either because it's *too* stable—*all* snakes have backbones (that is, this character is plesiomorphic in all snakes). Presence of a backbone indicates a common ancestor far older than the oldest snake, and so cannot be used to determine relatedness within the snakes.

Once the characters are chosen, the outgroup is used to determine the ancestral state of all characters. Then any difference from the outgroup in the chosen characters is scored as the appearance of a derived character. New lineages are “diagnosed” by the derived characters they have.

Imagine that species W, X, Y, and Z all evolved from a common ancestor after this ancestor split from a related outgroup (OG). This means that the WXYZ group is a clade or a monophyletic taxon (a taxon with one ancestor). In which order did W, X, Y, and Z evolve? This can be determined by considering the number of derived characters that these species share. Say that the evolution occurred as shown in the treelike figure below (known as a cladogram), where the appearance of derived characters are shown as ticmarks.

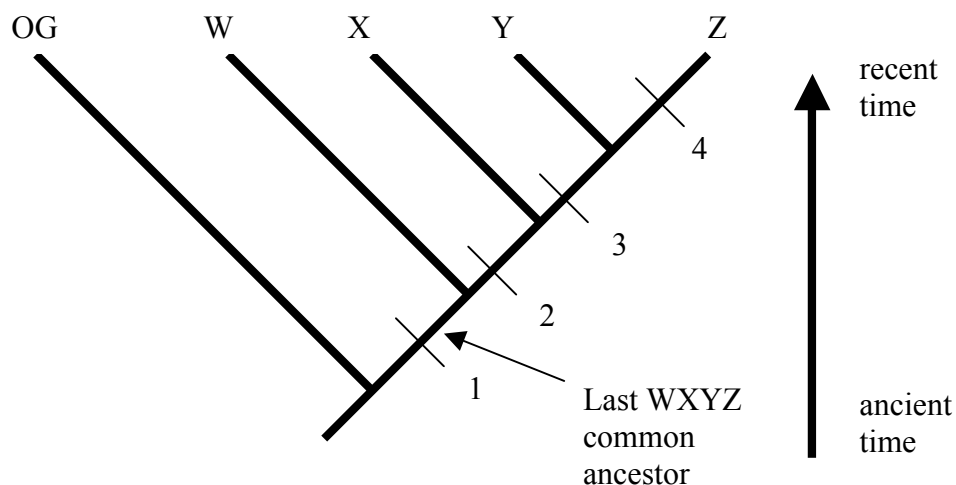


Figure 9. A cladogram with the appearance of derived characters shown by ticmarks.

At the lower left is the common ancestor of the outgroup and the WXYZ group. Then the outgroup split off and *after that* the WXYZ ancestor acquired the derived state of character 1. We know that this happened after the split because the outgroup still has the ancestral state of character 1, and W, X, Y and Z all have the derived state. The last common ancestor of the WXYZ clade is shown by the arrow just before the W lineage split from this group. After the W ancestor's divergence, the common ancestor of X, Y, and Z developed the derived state of character 2. Again, we know that this split happened before the derived state of character 2 originated because W doesn't have the derived state of character 2 but X, Y, and Z do. Then the X ancestor split off and the derived state of character 3 appeared in the Y-Z common ancestor. Finally, Y split from Z before the derived state of character 4 evolved in the Z ancestor.

Which are the most related species on this cladogram? In other words, which pair of species has the most recent common ancestor? Y and Z share three derived characters, showing a relatively recent common ancestor. X, Y and Z are less related because they share two derived characters, and we'd have to go farther back in time to find their common ancestor. W, X, Y, and Z are only slightly related because they only share one derived character and their common ancestor is relatively ancient. Nevertheless, W, X, Y, and Z all belong to a monophyletic lineage or clade—they all came from common ancestor that acquired the derived state of character 1 and passed it on to all of them.

Most of the time the cladogram comes last and we start with a “character matrix” in which the derived characters are shown as a 1 and the ancestral characters are shown as a 0. The WXYZ data would produce a character matrix like this:

Table 4. Character states in the WXYZ group. 0 = ancestral and 1 = derived.

Species	Characters			
	1	2	3	4
OG	0	0	0	0
W	1	0	0	0
X	1	1	0	0
Y	1	1	1	0
Z	1	1	1	1

To construct the cladogram from this table, one would notice that character 1 divides the WXYZ ingroup from the outgroup (the outgroup has the ancestral state of character 1 and the whole ingroup has the derived state). Then character 2 separates the XYZ group from W. Character 3 separates Y and Z from X, and finally character 4 separates Z from Y.

Parsimonious Cladograms

“Parsimony” is the scientific preference for explanations that are as simple as possible. If we started with the character matrix in Table 4, consider these two possible cladograms that could explain the results:

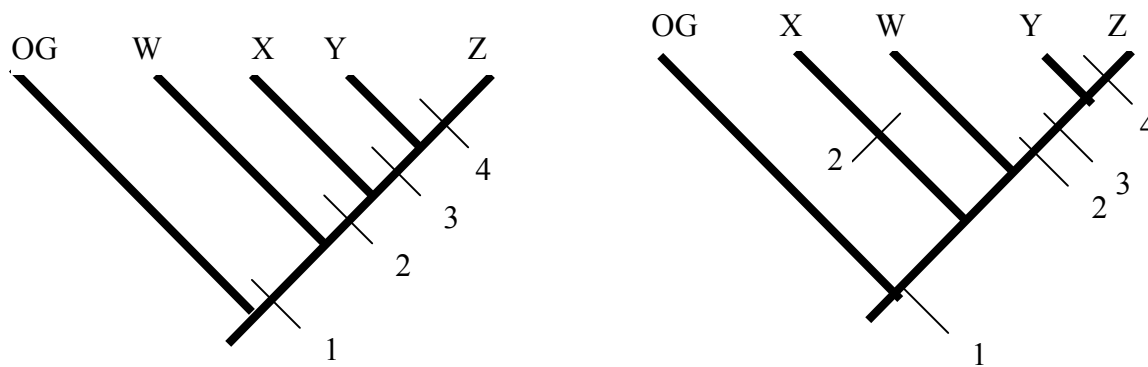


Figure 10. Parsimonious (left) and non-parsimonious (right) cladograms for W, X, Y, and Z.

We’ve already discussed the left one. In the right one, we hypothesize that X split first from the WXYZ lineage and that W split second, and that the derived state of character 2 appeared twice (independently in X and in the YZ ancestor). Both hypotheses could explain the data, but the left one is simpler, requiring 4 origins of derived characters rather than 5. Parsimony dictates that if we have no reason to suspect that a more complicated cladogram is accurate, we should always prefer a simpler cladogram. This is not just an arbitrary policy—independent origins of identical derived characters are rare in evolution, and we will be right the majority of the time if we favor a hypothesis that requires the fewest independent origins.

Remember that we should try to explain the data with the fewest possible steps.

The Heartbreak of Homoplasy

Homoplasy is the collective name for a series of conditions (such as convergences and reversals) that violate the assumptions of cladistic analysis.

Cladistic analysis makes the assumption that species share derived characters because they share a common ancestor that had that derived character. However, sometimes *convergences* occur and there are independent origins of a feature in unrelated lineages. For example, the fins of sharks, the flippers of dolphins, and the wings of penguins look alike but arose independently in the cartilaginous fishes, the mammals, and the birds. These structures are analogous, not homologous.

Second, cladistic analysis assumes that if a species does not have a derived character, it is because its ancestor diverged from the lineage that developed the derived character before the derived character originated. However, sometimes lineages have *reversals*—they have a derived character and it reverts to an ancestral type. For example, having hair is a derived character of mammals as compared with

reptiles. Dolphins have no hair because of a reversal—the mammalian ancestors of dolphins had hair and then the dolphin ancestor lost it. Hairlessness in dolphins and alligators (for example) is analogous, not homologous.

In this exercise, you will analyze simulated character data that both has homoplasy and lacks it. Homoplasy might lead you to draw the wrong conclusions even though you do everything right. For example, if species A has derived character 1 and species B has both derived characters 1 and 2, you'd be doing a correct analysis by parsimoniously assuming that A got 1, and then B split from A and developed 2 after the split. However, if homoplasy intervened, you may find that A got 1 and then B acquired both 1 and 2 independently.

Exercise C: Solving Cladistic Problems

In the following exercises, you will try to reconstruct the evolutionary history of a group of 6 hypothetical species (species A through species F). There are 20 possible characters, called 1-20. The data are hypothetical, and came from a computer simulation of random speciation. This has the advantage that the simulation keeps a record of which species evolved from which other species, so your conclusions can be checked.

There are two exercises:

- C1. The simulation only allows one origin of each derived character, and no derived characters can revert to the ancestral state. That is, there is no homoplasy.
- C2. Homoplasy (both multiple origins of the same derived character and loss of derived characters) may occur.

On both exercises, you will be shown a worked example for each type of exercise. Then pairs of students will do a different example that will be given to you by your lab instructor. After you write down your solution to each problem, you will be shown the true evolutionary history of your group of “species.”

One point about this exercise must be emphasized. The “answers” are the true evolutionary history of the group. Especially when homoplasy is present, doing a correct analysis will not necessarily produce the correct evolutionary history because homoplasy violates the assumptions of cladistic analysis. Therefore, your task is to *apply correct cladistic techniques* and *defend* your methods.

Exercise C1. No Homoplasy

Recall that in these exercises there are 20 possible characters, all of which started off in the ancestral state in the outgroup. The character matrix, showing an X where derived characters appear, is as follows:

Table 5. Character matrix for Species A-F. An X indicates the derived state of the character.

	3	4	5	6	8	11	12	15	16	17	18	20
A					X							
B	X	X		X	X	X						
C			X	X	X	X						X
D					X		X					
E			X	X	X	X		X	X		X	X
F			X	X	X	X				X		X

Your analysis will be based on determination of which derived characters are shared the most widely, then the next most widely, so so forth.

The character matrix shows that the most widely-shared derived character (most common synapomorphy) is 8 (shared by all species in the ingroup). The outgroup must have branched from the A-F group before the derived state of character 8 evolved. Next, derived characters 6 and 11 are shared by 4 of the 6 species, but not by species A and D. Why not? The A and D ancestors must have branched off from the other species before the derived states of 6 and 11 evolved. A diagram of the situation so far would look like this:

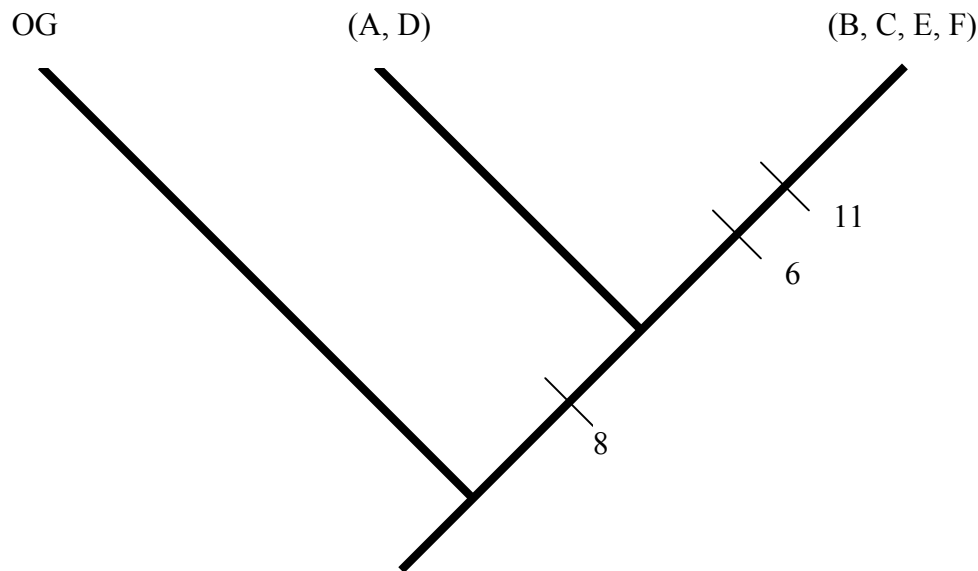


Figure 11. An incomplete cladogram for Example C1.

Of course, we still have to determine how A and D are related and how B, C, E, and F are related. Rather than continuing to draw diagrams, it will be faster to express our conclusions above by a “shared character” table. So far, the table looks like this:

Table 6. Initial shared character table for Example C1.

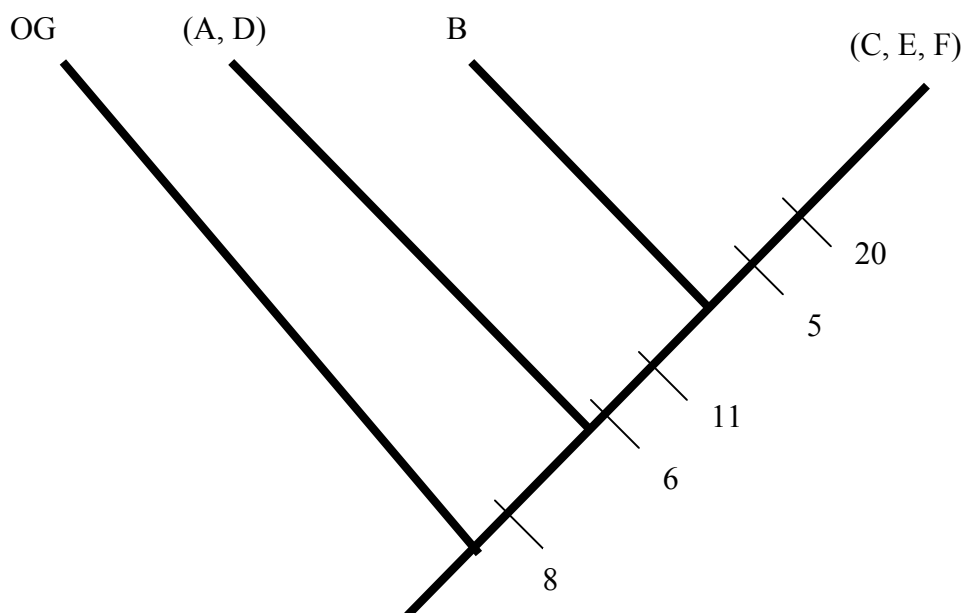
Shared Characters	A	B	C	D	E	F	Species Branched Off Before Shared Characters Evolved
8	X	X	X	X	X	X	Outgroup
6, 11		X	X		X	X	A, D

This table means that the derived form of character 8 is present in all species, that derived characters 6 and 11 are present in all except A and D. A and D have branched off after 8 evolved but before 6 and 11 evolved because A and D had an “X” on the “8” line but do not have it on the “6, 11” line. The next most common characters are 5 and 20, present in C, E, and F, but not in B, so B has branched off in the next step. Now our shared character table looks like this.

Table 7. More complete shared character table for Example C1.

Shared Characters	A	B	C	D	E	F	Species Branched Off Before Shared Characters Evolved
8	X	X	X	X	X	X	Outgroup
6, 11		X	X		X	X	A, D
5, 20			X		X	X	B
Remaining							C, E, F

After characters 5 and 20, we can't go any further with the Shared Character Table because there is no clear choice for the next most common shared character. Therefore, we'll list C, E, F (the remaining species) as the last branched group. Our cladogram now looks like this:

**Figure 12.** A more complete cladogram for Example C1.

Note that we do the cladogram by writing a branch for the “Branched Off” group and then putting the common character(s) from that line on the main diagonal *after* the branch. So the OG branch is followed by character 8, the (A, D) branch is followed by common characters 6 and 11, and the B branch is followed by common characters 5 and 20. These steps will allow you to complete the main outlines of the cladogram.

We'll have to resolve the (A, D) and (C, E, F) groups by some parsimonious guessing. We'll rewrite the character matrix with A and D listed together and C, E, and F listed together. We won't include species B because it is not involved in a group with undetermined relationships.

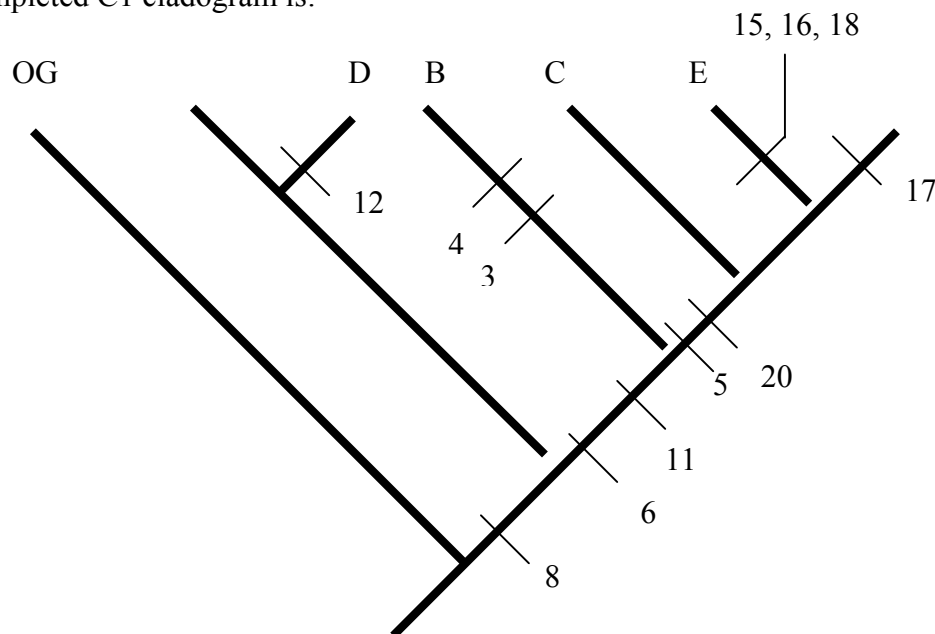
Table 8. Reordered character matrix for Example C1.

	3	4	5	6	8	11	12	15	16	17	18	20
A					X							
D					X		X					
C			X	X	X	X						X
E			X	X	X	X		X	X		X	X
F			X	X	X	X				X		X

In the (A, D) group, which species came first? It seems clear that the A ancestor evolved first and then the D ancestor came from the A ancestor by adding character 12. Likewise, for the (C, E, F) group, E is C with the addition of derived characters 15, 16, and 18, and F is C with the addition of derived character 17. Therefore, it is parsimonious to conclude that the E and F ancestor evolved from the C ancestor.

Finally, in addition to its other derived characters, B has 3 and 4, which no other species has. We express this by putting the ticmarks for derived characters 3 and 4 on the “B” branch. Derived characters on the main diagonal are passed along to species that branch off the diagonal after that point, but derived characters on the branches are only passed to the species that come from that branch.

The completed C1 cladogram is:

**Figure 13.** The complete cladogram for Example C1.

Can you follow the cladogram and see that species B has derived characters 8, 6, 11, 3, and 4, and that E has derived characters 8, 6, 11, 5, 20, 15, 16, and 18?

The cladogram above not only can explain all the data with a minimum number of steps, but it has accurately determined the way the computer simulation generated species A-F. This gives us confidence that cladistic methods can be used to infer evolutionary history.

This cladogram illustrates two more considerations. First, note that D branches from A above. This is a valid way of showing the relationship of A and D, but it would have been just as valid to present it as it is shown in the cladogram fragment below:

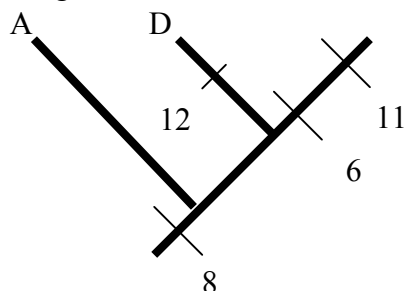


Figure 14. An alternate way of presenting the relationship between A and D in Example C1.

As long as there are no tickmarks on the main diagonal between the branching of A and the branching of D, D is considered to be branching from A. Also, there are aesthetic considerations in the drawing of cladograms. Cluttered cladograms with branches coming from branches are not preferred. For example, C, E, and F all branch from B in Figure 13. However, it would have been inelegant to draw three branches from the B branch in the cladogram in Figure 13. The larger the group, the more you should try to present it as branches from the main diagonal. Try to minimize branches coming from branches.

To Draw a Cladogram

1. Look at the Character Matrix and determine which derived character or group of characters is the most common, the second most common, etc.
2. Fill in a Shared Character Table. Determine the most common character and the species that share it, the second most common character and the species that share it, etc. At the right end of each line of the table, list the species that lack an X on that line of the table *but have an X on the line above*. If there is no clear choice of the next most common shared character, list all remaining species as branched off and go to step 3.
3. Draw a rough cladogram with only the main diagonal shown. Use the Shared Character Table to put in the main branches and the character tickmarks along the main diagonal. Remember that each species listed as branched off on each line branches *before* the shared characters on the left end of the line.
4. Take each branched-off group of two or more species and look at its shared derived characters (synapomorphies). Using parsimony, determine how the branching within each group probably happened.
5. Finally, fill in the details of branching on the cladogram for each group that branched off. You will probably have to do a rough draft of this first. Try to minimize branches coming from branches by putting the largest unresolved group along the main diagonal.

The cladogram at this point looks like this:

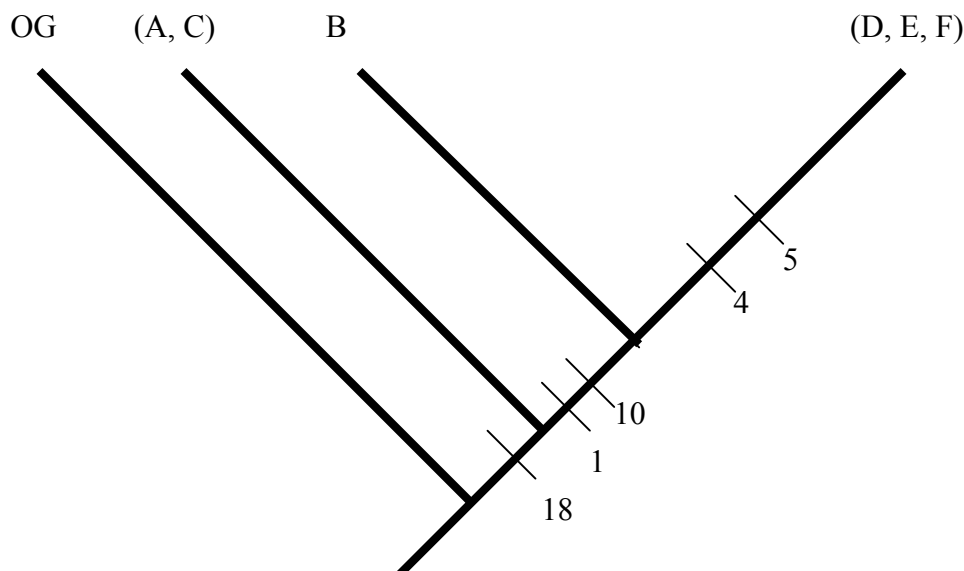


Figure 15. Partially complete cladogram for Example C2.

When we group (A, C) and (D, E, F) together in the character matrix, and we find:

Table 11. Reordered character matrix for Example C2. B has been omitted because it is not a member of an unresolved group.

	1	2	3	4	5	6	8	9	10	12	13	18
A			X									X
C							X	X		X		X
D	X			X	X				X	X		X
E	X	X		X	X				X			X
F	X			X	X				X	X	X	X

Looking at (A, C) first, it seems clear that the A ancestor arose first and then the C ancestor split from the A lineage by adding 8, 9, and 12. After the A-C split, A added 3. Notice that 3 is found only in A, so the 3 tickmark will be on the A branch after A split from C. And now, there's a problem. C branched off before B, D, and F ever evolved, so they should share only derived character 18 with B, D, and F. Instead, C shares 12 with those species. Look at Figure 15 again, and ask yourself: How could C, B, D, and F share derived character 12 without E having 12 as well??

There are three fairly parsimonious answers to this question:

- 12 originated on the main diagonal after A evolved but before C evolved, and then was lost from E (two steps).
- 12 originated in C and then again along with 1 and 10, and was lost from E (three steps). Because this explanation uses more steps than the first one, it is less parsimonious and not preferred.
- 12 originated independently in C, in B, and either D or F (three steps)...also not preferred.

The first explanation requires the fewest steps, so we select it. According to this idea, 12 would have to be added on the main diagonal (because it is passed onto all ancestors of C), but 8 and 9 would be added on the C branch because only C has them. Likewise, only A has 3, so the “3” ticmark is written on the A branch. Looking at the (D, E, F) group, there’s no clear ancestor. Both D and E have 6 derived characters. If D gave rise to E, E would have had to gain 2 and lose 12 (two steps). If E gave rise to D, D would have to lose 2 and gain 12 (two steps). It’s unclear which one originated first. We’ll say D originated first, and that E and F came from D. An origin of F from D is more parsimonious than an origin of F from E because F and D both have 12. Therefore, the final cladogram would look like this:

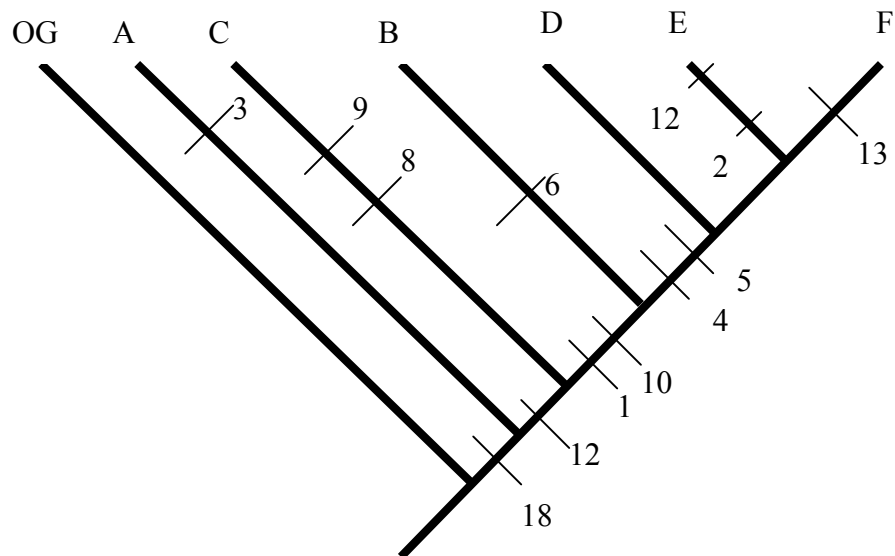


Figure 16. The completed cladogram for Example C2.

Figure 16 shows only one homoplastic event--the loss of 12 from E. This is a very parsimonious solution, and explains all the data.

However, the simulation *did not* happen as depicted above. Instead of originating after 18 and then being lost from E, 12 originated twice (in C after its split from the main diagonal and again with 1 and 10), and then was lost from E. We explained the data with two steps, but in reality, there were three steps. So we made a reasonable but unlucky guess about the origin of 12. On the other hand, we did guess correctly about the evolution of D, E, and F. What matters is that despite the ambiguity caused by homoplasy, our reasoning was *parsimonious* and *defensible*.

Dealing with Homoplasy

1. Use the most common synapomorphies to construct a Shared Character Table.
2. Construct a partial cladogram using the corrected table.
3. Using parsimony, try to sort out the unresolved groups of species. Multiple origins of derived characters and losses of derived characters might occur, but try to explain the data with the minimum number of steps.
4. Complete the cladogram.

Figure 17. Your cladogram for the flightless bird data.

Exercise E. Cladistic Analysis of the “Boltidae”

After your laborious solution to the Exercise C and D problems, we thought you might like a more intuitive challenge involving “organisms” composed of nails, bolts, nuts, and washers. In this exercise, you are not told what is ancestral and what is derived, or which “species” is a member of the outgroup. You must determine that. While almost everyone will agree on some aspects of the classification, there is room for legitimate disagreement on other aspects.

Exercise E1. This is an easy warmup exercise. You will be given five “organisms,” identified as A through E. The tape labels on them are just labels, not characters. Only pay attention to the metal parts of the “organisms.” Examine their characteristics, determine which one will serve as the outgroup, lay them out on your lab bench in the most parsimonious order in which they evolved. There is no need to draw a cladogram in this exercise.

Exercise E2. This is more difficult. You will be given another set of “organisms” to *add* to your first set. Here, relationships are not so simple, and homoplasy may be present. There is more than one parsimonious solution. Add them to the previous group of organisms on your lab bench in the pattern in which you think they evolved. Again, only pay attention to the metal parts as you make your decisions. You should be able to defend your solution as parsimonious.

When finished, put the first set (marked with white labels) in one Ziplock bag, and the other set (marked with pink labels) in the second Ziplock bag.

Literature Cited

Cracraft, J. 1974. Phylogeny and evolution of the ratite birds. *Ibis*, 116:494-521.

Phylogenetic Analysis--Notes for the Instructor

The attached laboratory presents five exercises intended to teach students the fundamentals of phylogenetic analysis, especially by cladistic methods. It has been successfully used as a 3-hour lab at Clemson University since 2001.

The Exercises

- A. Explains how to classify organisms by molecular similarity. This exercise attempts to determine relatedness of eight vertebrate species (humans, rhesus monkeys, mice, rats, Canada geese, domestic ducks, alligators, and crocodiles) based on differences in the beta chain of their hemoglobin. Ex. A uses a paper-and-pencil solution only, and should be used if computers with Internet access are not available. Exercise A is based on an exercise in Nickels and Nelson (2001).
- B. Same as Ex. A, but this exercise uses a more sophisticated approach involving bioinformatics databases and tools. Ex. B takes less time to complete, and would be used if computers with Internet access are available. However, both exercises A and B reach the same conclusion about the eight species.
- C. This largest exercise explains how to take a table of derived characters for a clade and produce a parsimonious cladogram. The data are derived from a computer simulation of evolution, so the accuracy of the students' conclusions can easily be assessed. The students solve a problem without and with homoplasy (reversal of derived characters and/or independent origin of derived characters). While there is only one instance of homoplasy per problem, Ex. C2 requires judgment and parsimonious guessing rather than just a problem-solving algorithm. There are four problems without homoplasy and four with homoplasy.
- D. The students demonstrate their skills by producing a parsimonious cladogram from published data on eight families of flightless birds. After exercise C, they regard this exercise as relatively easy.
- E. The students produce a cladogram involving "organisms" made out of nuts and bolts. Although this exercise could have multiple valid solutions, it is a fun end to the lab. This exercise is based on Boese (2002).

The salient characteristic of this lab is that (aside from Ex. B), it is completely low-tech. Rather than focusing on teaching the students to use phylogenetic analysis software, it teaches them to solve small cladistic problems by observation and thought alone. Thus they learn how cladistic analysis works rather than just how to operate software. Also, because the "evolutionary history" that gave rise to the characters is known from the computer simulation, the students can be told whether they inferred the correct evolutionary relationships or not. At Clemson, we then have them write a paper that evaluates the success of cladistic methods in determining the true pattern of evolutionary history. The usual conclusion is that cladistic methods are very successful if there is no homoplasy present, but can easily produce incorrect conclusions where homoplasy is a problem.

Materials Supplied

Aside from computers with Internet access, this manuscript contains all necessary materials for the use of the lab. These materials include these Instructor's Notes, the Student Exercises, and a series of Instructor Appendices. The Appendices include four different sets of student problems for Ex.C, four

sets of student answers to the problems (handed out in class after the students have finished their solutions), and a set of instructor answers to the problems. The student answers give the true evolutionary history of the simulated clade so the students can evaluate their success, but they do not show a cladogram. The instructor answers show this true history presented as a cladogram as well. This should be the cladogram the students produce in problems without homoplasy. In problems with homoplasy, the students may follow good procedure and not reproduce the true history. Finally, the Instructor Appendices include a transparency master showing a parsimonious solution to the flightless bird problem in Ex. D.

Advice on Implementing the Laboratory

Introduction. First, the instructor must explain the idea that differences between lineages start to accumulate once interbreeding stops. We use the following parable:

Say that you are a fourth grade teacher, you're giving an arithmetic test, and halfway through the test, you see that Billy is cheating from Bobby. You move them to opposite sides of the room. When you correct the papers, you may see that both Billy and Bobby have lots of mistakes, but up until the point where you separated them, the mistakes on both papers were _____ (the same or different). On the other hand, after you'd separated them, the mistakes on the two papers were _____ (the same or different). This principle is the basis of molecular similarity techniques of judging evolutionary distance.

If two groups of organisms are interbreeding or have just recently stopped interbreeding, we would expect that their mutations at a locus would be _____ (the same or different). If two groups of organisms have not interbred for a long time, we would expect that their mutations at a locus would be _____ (the same or different).

Exercise A. The students should do Ex. A if no computers are available, and Ex. B if computers with Internet access are available. In Exercise A, the students have to manually count up differences in the hemoglobin beta chain between all possible pairs of eight species (28 pairs). If the instructor organizes groups of four, in a class of 24, each group will have to do about five comparisons, a reasonable workload. The directions are complete, and no explanations are usually necessary.

The correct numbers of differences are as follows:

Table 15. Correct number of amino acid differences between all possible pairs of species in Ex. A.

	Monkey	Mouse	Rat	Duck	Goose	Croc.	Allig.
Human	4	13	13	19	19	29	37
Monkey		13	13	18	18	29	37
Mouse			5	23	23	29	36
Rat				20	20	29	36
Duck					1	28	32
Goose						28	31
Croc							17

It is obvious that humans and rhesus monkeys, mice and rats, ducks and geese, and crocodiles and alligators are related pairs (because they had relatively recent common ancestors). Mammals are also more related to each other than to birds or reptiles (again, mammals have a more recent common

ancestor). What might be a surprise is that domestic ducks and Canada geese are so closely related, and that mice and rats are more related *to humans* than crocodiles are to alligators. In fact, question 4 on p. 10 points out that the species difference that is closest to the alligator-crocodile difference (17 amino acids) is the difference between rhesus monkeys and either ducks or geese (18 amino acids). Alligators and crocodiles seem to have been separate lineages for a long time.

Exercise B. There are three reasons to do this exercise rather than A if computers with Internet access are available. First, the exercise goes faster than Exercise A. Second, the students will be introduced to online bioinformatics tools. These include the Swiss-Prot protein database, an impressive source of data on over 195,000 protein sequences (as of October 2005) that can be used in other taxonomic studies, and ClustalW, a popular tool for doing multiple sequence alignments and taxonomic comparisons. Third, these tools can handle more difficult comparisons, such as amino acid sequences that are not aligned correctly. The reason *not* to do Exercise B is that the students may never actually look at the amino acid sequences, and instead spend their time merely operating software.

The ClustalW results, depicted in a phylogram below, make the same points about relationships as were made in Exercise A:

Phylogram

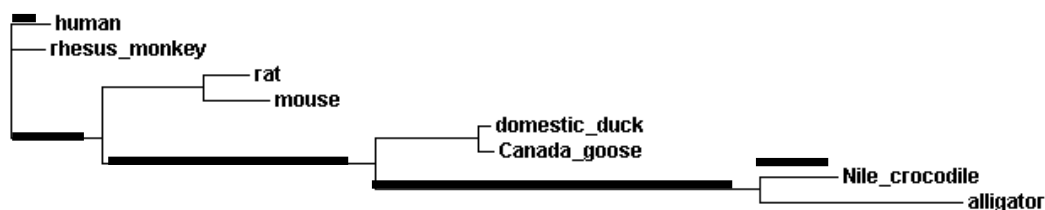


Figure 18. ClustalW's phylogram for the eight species in Ex. B. The heavy line segments indicate the segments whose lengths should be added to determine the evolutionary distance between humans and crocodiles (see text).

Related pairs are grouped together, and the horizontal length of their branches indicates the degree of difference from the other member of the pair. It is clear that the most similar pair is duck-geese, and the least similar is crocodile-alligator. In fact, while someone who needed a field guide to tell the difference between a crocodile and an alligator could easily distinguish between a rhesus monkey and a human, the molecular differences between crocodiles and alligators are three-four times as large as between rhesus monkeys and humans.

If we think of the lines connecting any two species as roads, the degree of dissimilarity between organisms that are *not* in pairs can be determined from the *horizontal* "driving distance" between any two species. For example, between humans and crocodiles, the degree of dissimilarity is proportional to the sum of the heavy line segments in Figure 18.

ClustalW calculates dissimilarities based on the number of amino acid differences alone. Therefore, it is similar to the manual Exercise A in its results.

Exercise C. This longest exercise includes an explanation of cladistics and cladograms. The best approach is to work through the C1 and C2 examples in the student handout with the students to make sure they understand the principles. At Clemson, we make sure we pass the exercise out to the students in advance of the lab so they can read it over first.

Then teams of two students should get either problems 1 and 2, or 3 and 4, or 5 and 6, or 7 and 8. The reason for this is that the student answer sheets have all these sets of problems on them. The first problem in each pair does not have homoplasy, but the second one does. If an instructor is concerned about the ability of the students to solve difficult cladistic problems, or if time is limited, Ex. C2 and the even-numbered problems above should be omitted.

After the student pairs work through the problems, the instructor should give out the student answer sheets that give the true evolutionary history of the clade, but do not show the corresponding cladogram. Instructor answer sheets (that *do* show the cladogram) are provided for each of the eight problems. Both of these answer sheets are found in the Instructor Appendices.

Our experience is that Ex. C1 is not too difficult for the students. Ex. C2 can be difficult because each problem includes one instance of homoplasy (either duplicate origin of a derived character or reversal of a derived character). Severe homoplasy can make a cladistic problem impossible to solve, but experience shows that students can solve these mildly homoplastic problems. Even if they have difficulty, the uncertainty of cladistic methods when homoplasy occurs is one of the main lessons of the lab.

Exercise D. Analysis of the flightless birds in Table 13 should not cause the students any problem. The correct cladogram appears below:

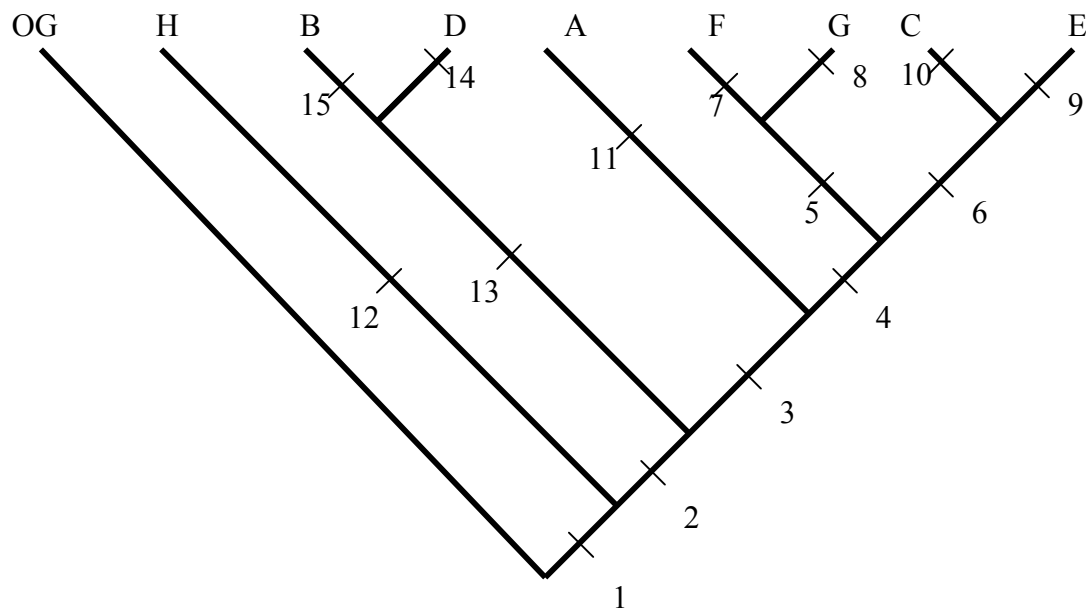


Figure 19. A parsimonious solution to the flightless birds data (Ex. D).

This cladogram accounts for all the derived characters in the clade with only one origin for each one. It can easily be derived using the techniques in the student exercise. We've found that the students are proud of the fact that they can solve a moderately complex problem like this relatively quickly.

Exercise E. The “Boltidae” exercise is more like recreation than a serious problem. We divide our nuts and bolts “organisms” into two groups in two Ziplock bags per pair of students. Each organism consists of the parts below, with either a white tape label giving its code letter or a pink tape label for the code letter:

Bag 1--White Labels (Ex. E1)

E nail (the outgroup)

A machine screw with threads and slotted head

D machine screw with threads, slotted head, and hex nut

C machine screw with threads, slotted head, hex nut, and flat washer

B machine screw with threads, slotted head, hex nut and flat washer and lock washer



Figure 20. The white label “organisms” arranged in the order in which they “evolved.” Each “organism” adds one more derived character to the organism before it in the series. The nut in each case is glued onto the bolt so students don’t remove them.

Bag 2--Pink Labels (Ex. E2, to be added to the E1 group)

G machine screw with hex nut and two lock washers, but no flat washer

J machine screw with wing nut and flat washer

H hex-headed bolt

F hex-headed bolt with hex nut

I hex-headed bolt with hex nut and flange below head

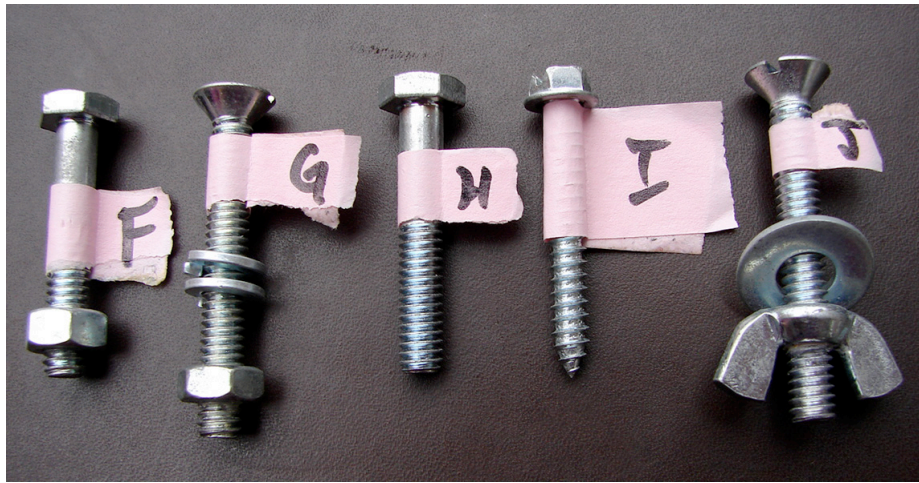


Figure 21. The pink label “organisms” arranged alphabetical order. There could be several different proposals as to how these organisms “evolved.” The bottom nut is glued onto the bolt to prevent students from removing it.

The parsimonious solution to Exercise E1 is obvious because the cladogram has no homoplasy. Each organism adds one more derived character to the organism before it in the sequence, as can be seen from the descriptions above. Therefore, the cladogram is simple:

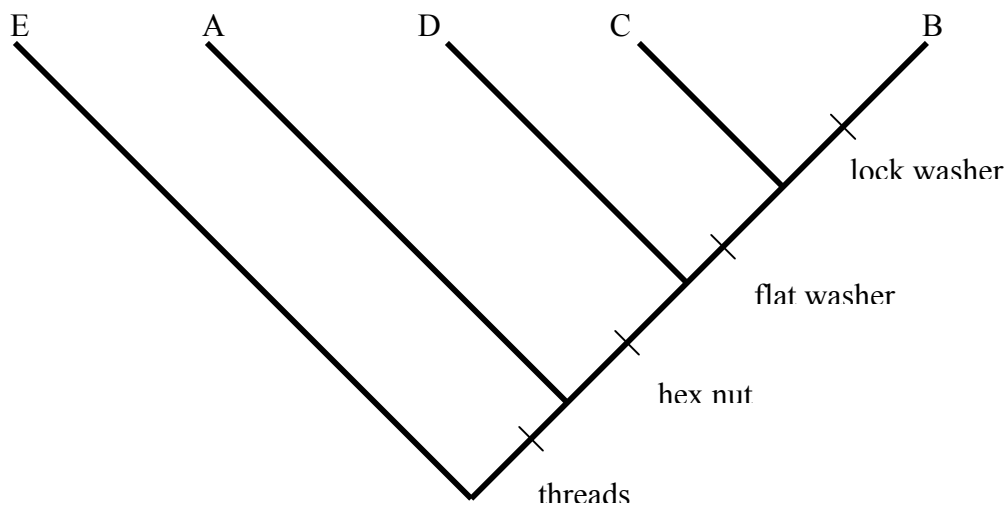


Figure 22. A parsimonious solution to Ex. E1.

There are several possible solutions when the pink label “organisms” (F-J) are added. One solution involves two origins of the hex head (for H and I as a pair, and also for F), and loss of a flat washer from G.

Literature Cited

- Boese, S. 2002. Classification and field guides to the fictitious genus Boltadae: An introduction to phylogeny and biodiversity. Page 391, *in* Tested Studies for Laboratory Teaching, Volume 23 (M.A. O'Donnell, editor). Proceedings of the 23rd Workshop/Conference of the Association for Biology Laboratory Education (ABLE), 392 pages.
- Cracraft, J. 1974. Phylogeny and evolution of the ratite birds. *Ibis*, 116:494-521.
- Nelson, C. E. and M. E. Nickels. 2001. Using humans as a central example in teaching undergraduate biology labs. Pages 332-365, *in* Tested Studies for Laboratory Teaching, Volume 22 (S.J. Karcher, editor). Proceedings of the 22nd Workshop/Conference of the Association for Biology Laboratory Education (ABLE), 489 pages.

Basic Explanations of Cladistics on the Web

- Clos, L. M. *What is Cladistics?* Web site at <http://www.fossilnews.com/1996/cladistics.html>. This is rather simple and brief. It attempts to introduce cladistics to paleontologists who don't know anything about it.
- Collins, A. G., R. Guralnick, and B. R. Speer. *Cladistics in Brief*. Web site at <http://www.ucmp.berkeley.edu/clad/clad1.html>. Very understandable introduction from Berkeley.
- Arens, N. C., C. Stromberg, and A. Thompson. *Laboratory II. Phylogenetics*. Web site at <http://www.ucmp.berkeley.edu/IB181/VPL/Phylo/PhyloTitle.html>. Another Berkeley Web site with more detailed explanations and plant examples.

About the Author

Bob Kosinski is a professor of Biology at Clemson University, where he is the sole lecturer in the Introductory Biology course for majors and also the coordinator of the labs for that course. He received his BS degree from Seton Hall University and his Ph.D. in Ecology from Rutgers University. His interests include laboratory development, investigative laboratories, and the educational use of computer simulations, all in introductory biology. He has attended every ABLE meeting since 1989, has presented at 10 of those meetings, and acted as the chair of the host committee for the 2000 ABLE meeting at Clemson University.

Appendix B: Student Answer Sheets

This appendix includes true evolutionary histories of the simulated taxa in the problems. Each page contains a pair of problems that should be given to one student team. Therefore, the four student answer pages cover:

- Problems 1 and 2
- Problems 3 and 4
- Problems 5 and 6
- Problems 7 and 8

Solution

Problem 1

Character Matrix

Species	4	6	7	10	11	13	14	17	18	20
A	X				X		X	X	X	X
B					X		X			X
C					X	X				X
D										X
E					X		X	X	X	X
F		X	X	X	X		X	X	X	X

True Evolutionary History

- A came from E by adding 4.
- B came from C by adding 14.
- C came from D by adding 11 and later 13.
- D came from OG by adding 20.
- E came from B by adding 17 and 18.
- F came from E by adding 6, 7, and 10.

Solution

Problem 2

Character Matrix

Species	2	3	4	5	7	9	11	12	19
A	X	X	X	X	X			X	
B	X	X		X	X			X	
C	X	X		X	X	X		X	
D	X	X		X	X		X	X	
E	X			X				X	
F	X	X		X	X				X

True Evolutionary History

- A came from B by adding 4.
- B came from E by adding 3 and 7.
- C came from B by adding 9.
- D came from B by adding 11.
- E came from OG by adding 2, 5, and 12.
- F came from B by adding 19 and losing 12.

Instances of homoplasy are underlined.

Solution

Problem 3

Species	1	2	3	4	5	7	8	9	15	16
A	X						X		X	X
B	X			X			X	X		
C	X						X			
D	X	X	X	X			X			
E	X			X	X	X	X			
F	X			X			X			

True Evolutionary History

A came from C by adding 15 and 16.

B came from F by adding 9.

C came from OG by adding 1 and 8.

D came from F by adding 2 and 3.

E came from F by adding 5 and 7.

F came from C by adding 4.

Solution

Problem 4

Character Matrix

Species	2	4	5	7	8	12	13
A	X			X	X	X	X
B					X		X
C				X	X		X
D							X
E	X	X	X				X
F	X	X		X	X	X	X

True Evolutionary History

A came from C by adding 2 and 12.

B came from D by adding 8.

C came from B by adding 7.

D came from OG by adding 13.

E came from D by adding 2, 4, and 5.

F came from A by adding 4.

Instances of
homoplasy are
underlined.

Solution

Problem 5

Character Matrix

Species	2	4	5	6	7	9	10	11	20
A	X	X	X	X	X		X		
B			X	X			X		
C			X	X	X		X		
D			X	X	X		X		X
E	X	X	X	X	X	X	X		
F							X	X	

True Evolutionary History

A came from C by adding 2 and 4.

B came from F by adding 5 and 6.

C came from B by adding 7.

D came from C by adding 20.

E came from A by adding 9.

F came from OG by adding 10 and later 11.

Solution

Problem 6

Character Matrix

Species	5	8	10	11	14	15	17	20
A						X		
B		X	X		X	X		
C	X			X		X		
D		X	X		X	X		X
E		X	X	X	X	X		X
F		X	X		X	X	X	

True Evolutionary History

A came from OG by adding 15.

B came from A by adding 8, 10 and 14.

C came from A by adding 5 and 11.

D came from B by adding 20.

E came from D by adding 11.

F came from B by adding 17.

Instances of
homoplasy are
underlined.

Solution

Problem 7

Character Matrix

Species	1	3	4	8	12	18	19	20
A				X	X			
B	X		X	X	X			
C				X				X
D	X		X	X	X	X		
E	X	X	X	X	X	X		
F				X			X	X

True Evolutionary History

A came from C by adding 12 before C had added 20.

B came from A by adding 1 and 4.

C came from OG by adding 8 and later 20.

D came from B by adding 18.

E came from D by adding 3.

F came from C by adding 19 after C had added 20.

Solution

Problem 8

Character Matrix

Species	1	2	4	5	6	8	11	20
A	X			X	X			
B	X			X	X		X	X
C	X		X	X	X			
D	X					X		
E	X	X		X	X			X
F	X						X	

True Evolutionary History

A came from F by adding 5 and 6 before F had added 11.

B came from A by adding 20 and later 11.

C came from A by adding 4.

D came from F by adding 8 before F had added 11.

E came from B by adding 2 before B had added 11.

F came from OG by adding 1 and later 11.

Instances of
homoplasy are
underlined.

Appendix C: Instructor Answer Sheets

This appendix includes instructor answers (including cladograms) for all 8 exercise C problems. The cladogram in each case represents the true evolutionary history of the group. If students produce a more parsimonious solution that achieves the same result, they have guessed incorrectly about an ambiguous situation, but they have performed a correct analysis.

Solution

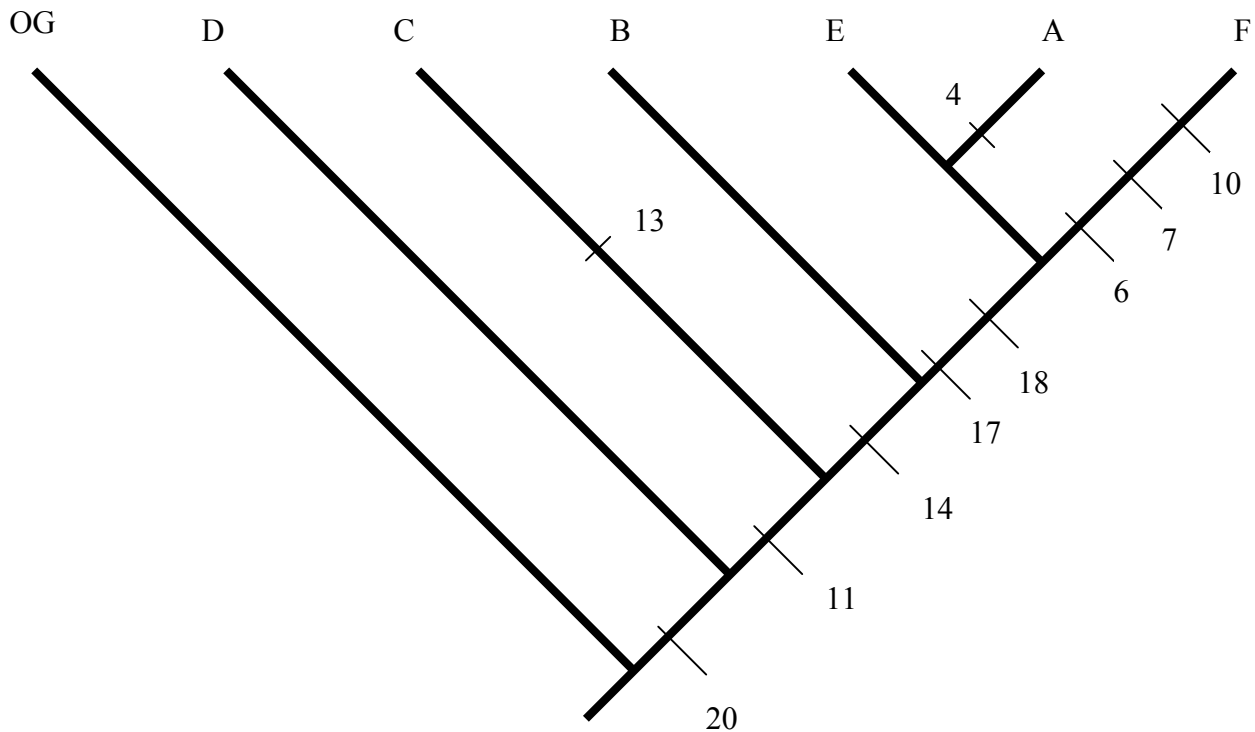
Problem 1

Character Matrix

Species	4	6	7	10	11	13	14	17	18	20
A	X				X		X	X	X	X
B					X		X			X
C					X	X				X
D										X
E					X		X	X	X	X
F		X	X	X	X		X	X	X	X

Shared Character Table

Shared Characters	A	B	C	D	E	F	Species Branched Off Before Shared Characters Evolved
20	X	X	X	X	X	X	OG
11	X	X	X		X	X	D
14	X	X			X	X	C
17,18	X				X	X	B
Remaining							A, E, F



Solution

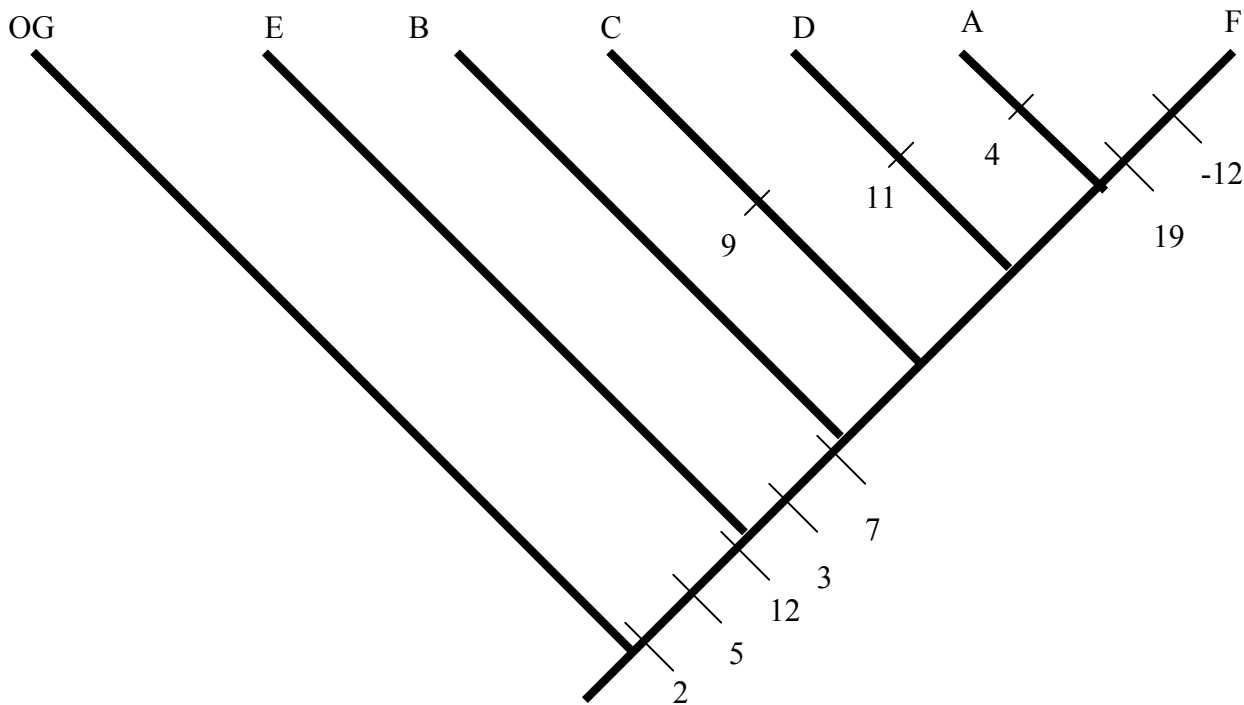
Problem 2

Character Matrix

Species	2	3	4	5	7	9	11	12	19
A	X	X	X	X	X			X	
B	X	X		X	X			X	
C	X	X		X	X	X		X	
D	X	X		X	X		X	X	
E	X			X				X	
F	X	X		X	X				X

Shared Character Table

Shared Characters	A	B	C	D	E	F	Species Branched Off Before Shared Characters Evolved
2, 5	X	X	X	X	X	X	OG
3, 7	X	X	X	X		X	E
Remaining							A, B, C, D, F



Solution

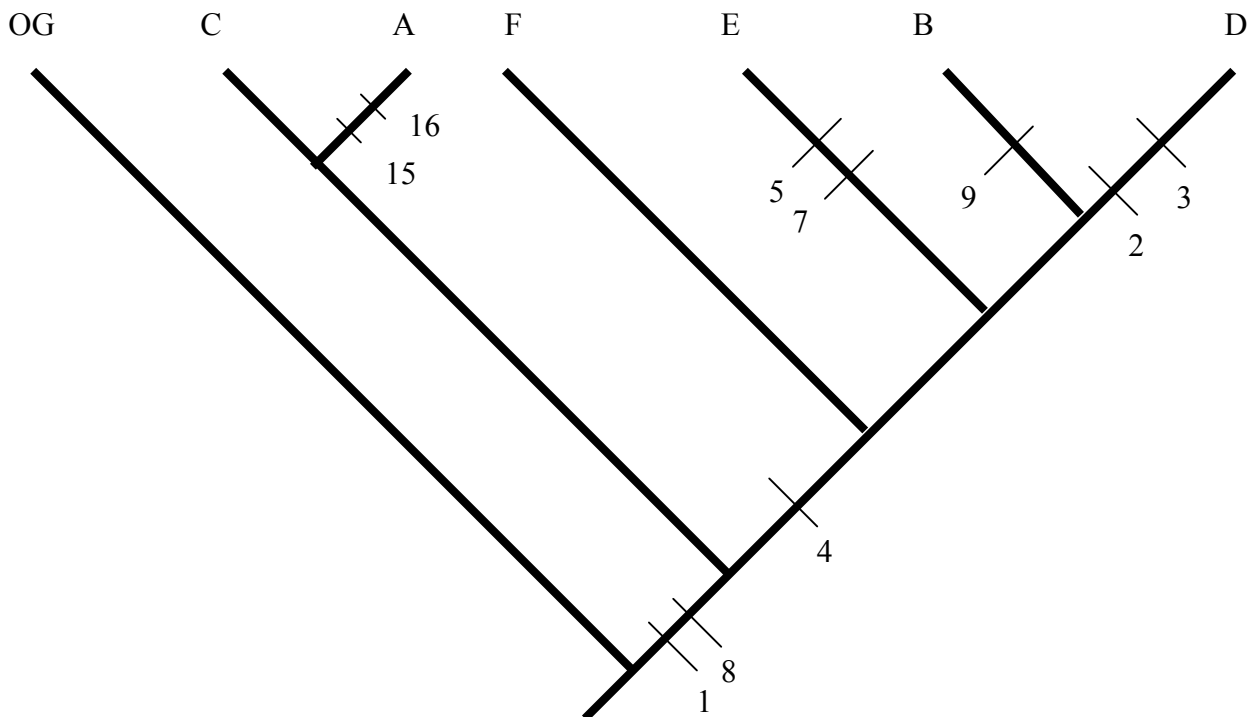
Problem 3

Character Matrix

Species	1	2	3	4	5	7	8	9	15	16
A	X						X		X	X
B	X			X			X	X		
C	X						X			
D	X	X	X	X			X			
E	X			X	X	X	X			
F	X			X			X			

Shared Character Table

Shared Characters	A	B	C	D	E	F	Species Branched Off Before Shared Characters Evolved
1, 8	X	X	X	X	X	X	OG
4		X		X	X	X	A, C
Remaining							B, D, E, F



Solution

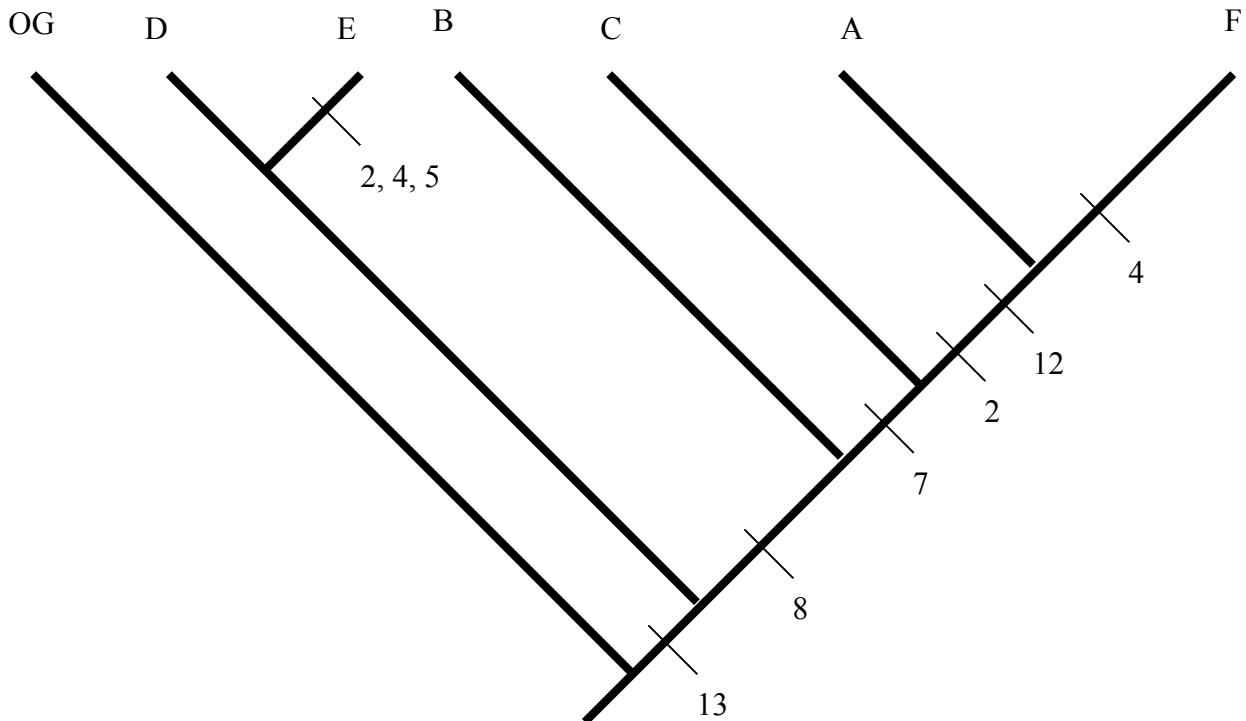
Problem 4

Character Matrix

Species	2	4	5	7	8	12	13
A	X			X	X	X	X
B					X		X
C				X	X		X
D							X
E	X	X	X				X
F	X	X		X	X	X	X

Shared Character Table

Shared Characters	A	B	C	D	E	F	Species Branched Off Before Shared Characters Evolved
13	X	X	X	X	X	X	OG
8	X	X	X			X	D, E
Remaining							A, B, C, F



Solution

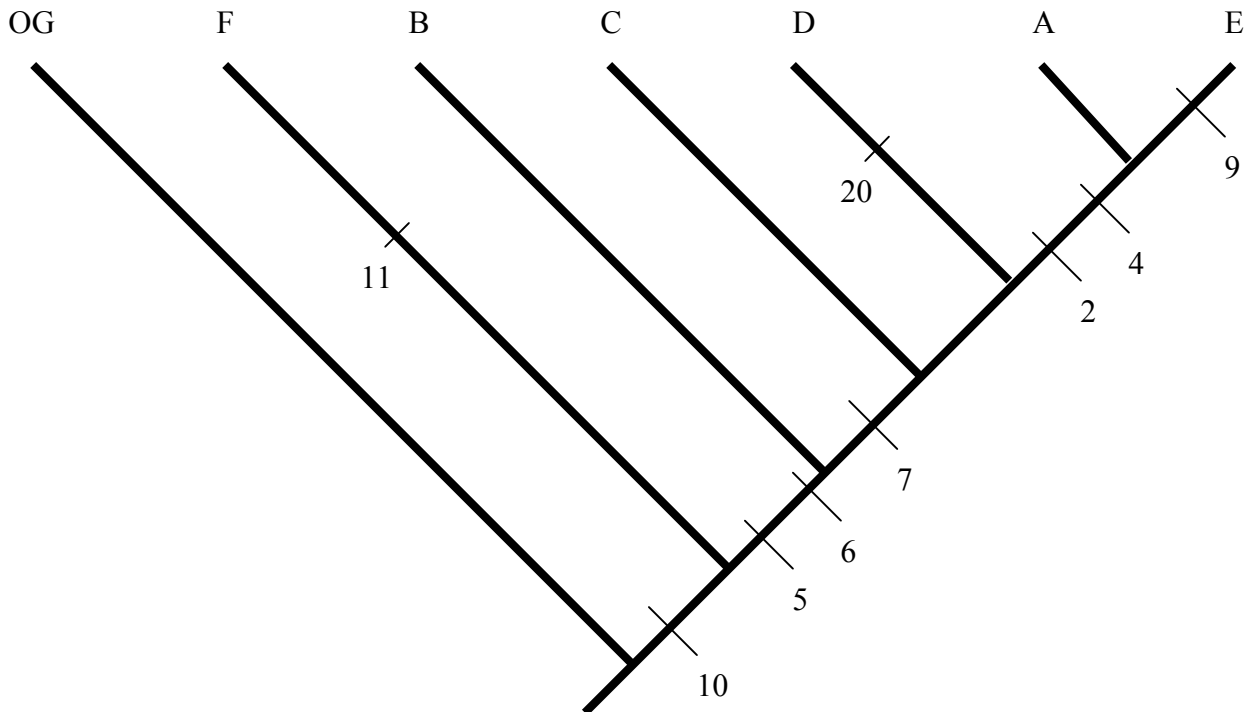
Problem 5

Character Matrix

Species	2	4	5	6	7	9	10	11	20
A	X	X	X	X	X		X		
B			X	X			X		
C			X	X	X		X		
D			X	X	X		X		X
E	X	X	X	X	X	X	X		
F							X	X	

Shared Character Table

Shared Characters	A	B	C	D	E	F	Species Branched Off Before Shared Characters Evolved
10	X	X	X	X	X	X	OG
5, 6	X	X	X	X	X		F
7	X		X	X	X		B
2, 4	X				X		C, D
Remaining							A, E



Solution

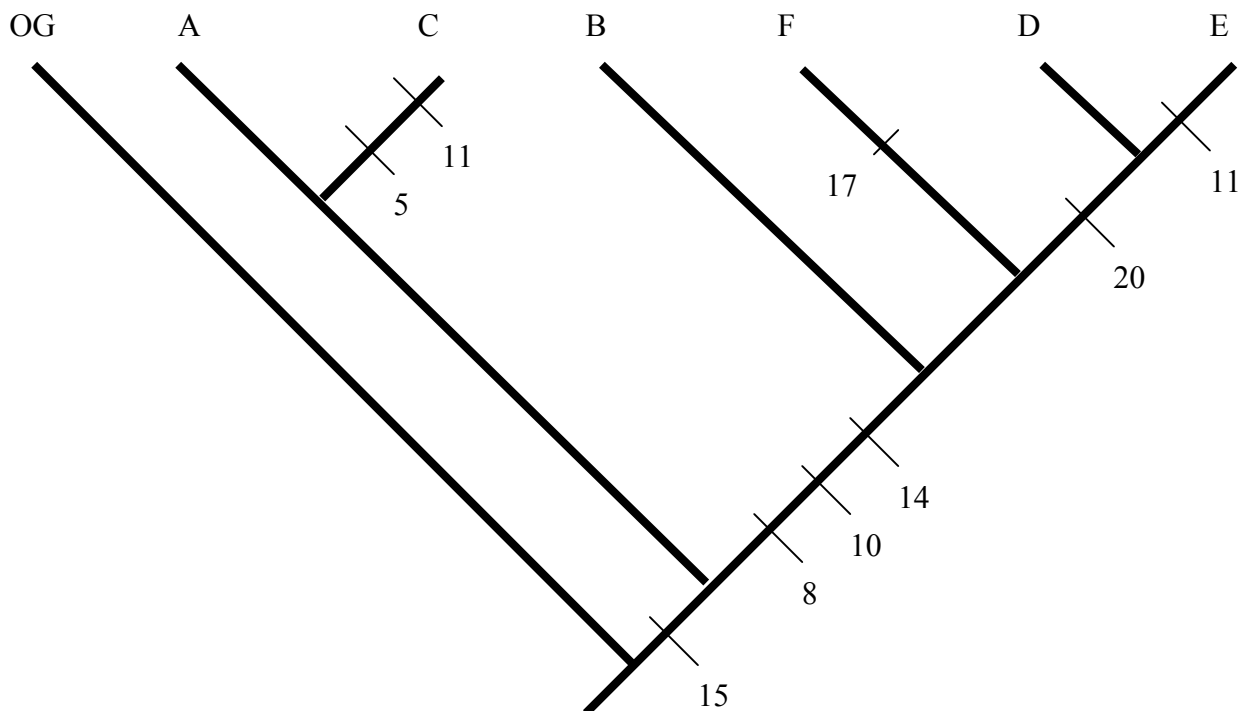
Problem 6

Character Matrix

Species	5	8	10	11	14	15	17	20
A						X		
B		X	X		X	X		
C	X			X		X		
D		X	X		X	X		X
E		X	X	X	X	X		X
F		X	X		X	X	X	

Shared Character Table

Shared Characters	A	B	C	D	E	F	Species Branched Off Before Shared Characters Evolved
15	X	X	X	X	X	X	OG
10, 14		X		X	X	X	A, C
8		X		X	X		F
Remaining							B, D, E



Solution

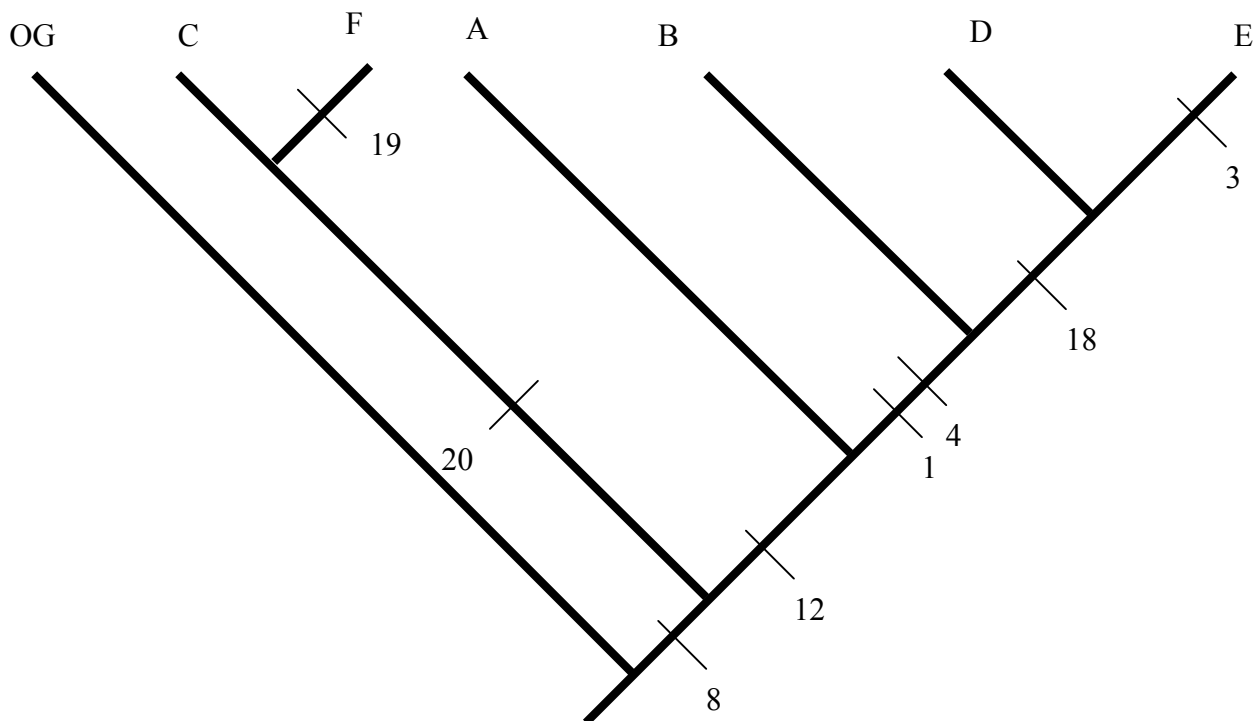
Problem 7

Character Matrix

Species	1	3	4	8	12	18	19	20
A				X	X			
B	X		X	X	X			
C				X				X
D	X		X	X	X	X		
E	X	X	X	X	X	X		
F				X			X	X

Shared Character Table

Shared Characters	A	B	C	D	E	F	Species Branched Off Before Shared Characters Evolved
8	X	X	X	X	X	X	OG
12	X	X		X	X		C, F
1, 4		X		X	X		A
18				X	X		B
Remaining							D, E



Solution

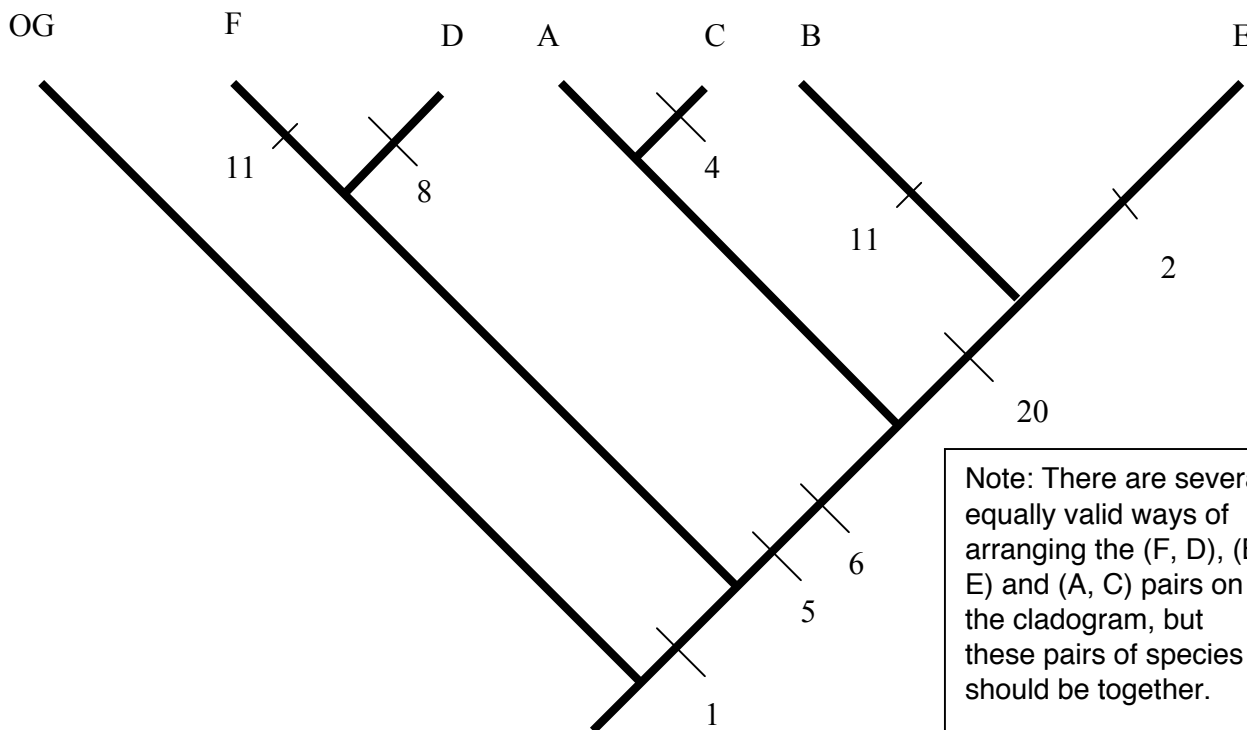
Problem 8

Character Matrix

Species	1	2	4	5	6	8	11	20
A	X			X	X			
B	X			X	X		X	X
C	X		X	X	X			
D	X					X		
E	X	X		X	X			X
F	X						X	

Shared Character Table

Shared Characters	A	B	C	D	E	F	Species Branched Off Before Shared Characters Evolved
1	X	X	X	X	X	X	OG
5, 6	X	X	X		X		D, F
Remaining							A, B, C, E



Appendix D: Transparency Master for the Flightless Birds (Ex. D)

The following transparency master shows the parsimonious solution to the flightless bird data published by Cracraft (1974). Cracraft came to the same conclusions, although these conclusions were formatted somewhat differently from the cladogram shown here.

Phylogeny of the Flightless Birds

Cracraft (1974)

