Phylogenetic Investigator Version 2.0.1

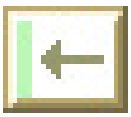
User's Manual

Steven D. Brewer University of Massachusetts
Robert Hafner Western Michigan University

A BioQUEST Library VII Online module published by the BioQUEST Curriculum Consortium

The BioQUEST Curriculum Consortium (1986) actively supports educators interested in the reform of undergraduate biology and engages in the collaborative development of curricula. We encourage the use of simulations, databases, and tools to construct learning environments where students are able to engage in activities like those of practicing scientists.

Email: bioquest@beloit.edu Website: http://bioquest.org



ABOUT PHYLOGENETIC INVESTIGATOR

Evolution, the central theme in biology, takes on added meaning for students when they can explore the construction and interpretation of evolutionary models. Phylogenetic Investigator (PI) facilitates creative problem-solving in phylogenetic inference for teaching and learning evolutionary biology. Users can identify characters and states, polarize characters, and engage in directed-search phylogenetic tree construction. PI also allows the user to (1) make inferences and represent them one step at a time, (2) vary representational features of their trees (such as angle of divergence and time between speciation events), (3) create reticulate tree patterns, and (4) view all of the character transformations at one time. In addition, PI can generate plausible data stochastically for modeling and practicing tree construction.

Phylogenetic Investigator was developed with support from the Department of Science Studies at Western Michigan University in Kalamazoo, Michigan . PI was created using SuperCard®.

TABLE OF CONTENTS

A PRIMER ON PHYLOGENETIC SYSTEMATICS

Introduction

What is phylogenetic systematics and why do people do it?

Each 'living thing' (or organism) is unique. Descended from some ancestor or ancestors and potential progenitors of offspring, organisms exist in populations of related organisms (species). Humans everywhere have named the species around them and evaluated the properties of each. Knowing whether a species was edible, medicinal, or poisonous could mean the difference between life and death. One of the fundamental aims of biology has been to create a nomenclature, or system of terms, that could systematically encompass the natural world.

It is axiomatic that species fall into natural kinds (See "A Quahog is a Quahog" in The Panda's Thumb Gould (1980)). Birds, although there are many different species, share features that appear to set them apart from all other kinds of living things. Similarly, these natural kinds seem to have some kind of hierarchical organization that can be represented by a taxonomy with species as the most basic taxon, or grouping, which can be placed within more and more inclusive taxa. A Red-winged Blackbird is one kind of blackbird which is one kind of perching bird which is one kind of bird which is one kind of the animals with backbones which is one kind of animal, and so on. Charles Darwin put forward a coherent explanation for this phenomenon that has come to be widely accepted. The theory of evolution proposes that living things are somehow related through ancestral/descendant relationships and that very similar things are more closely related than less similar things. Before a theory of evolution, taxa were usually based on the principle of overall similarity. The goal of phylogenetic systematics is the construction of a taxonomy based not on similarity, but on evolutionary relationship or genealogy.

The ability to describe how species are related has transformed how scientists understand evolution, systematics, and biogeography. Recently an issue of Bioscience was devoted to phylogenetic systematics (Simpson and Cracraft, 1995). Phylogenetic systematics, as a means to interpret the properties, activities, and distributions of species and groups of species, is illustrated from a variety of perspectives: biodiversity (Savage, 1995), agriculture (Miller and Rossman, 1995), ecology and behavior (Brooks et al, 1995), the study of organismal form and function (Lauder et al, 1995), and public health (Davis, 1995). In each of these examples, the ability to recognize the underlying relationships among species allows insight into the processes that have led to current conditions and makes it possible to predict future trends.

Phylogenetic Trees

What do they look like and what do all those things mean?

This section provides a brief description of phylogenetic trees, as they are conceptualized in Phylogenetic Investigator. Some of the concepts presented here are described at greater length elsewhere in the text.

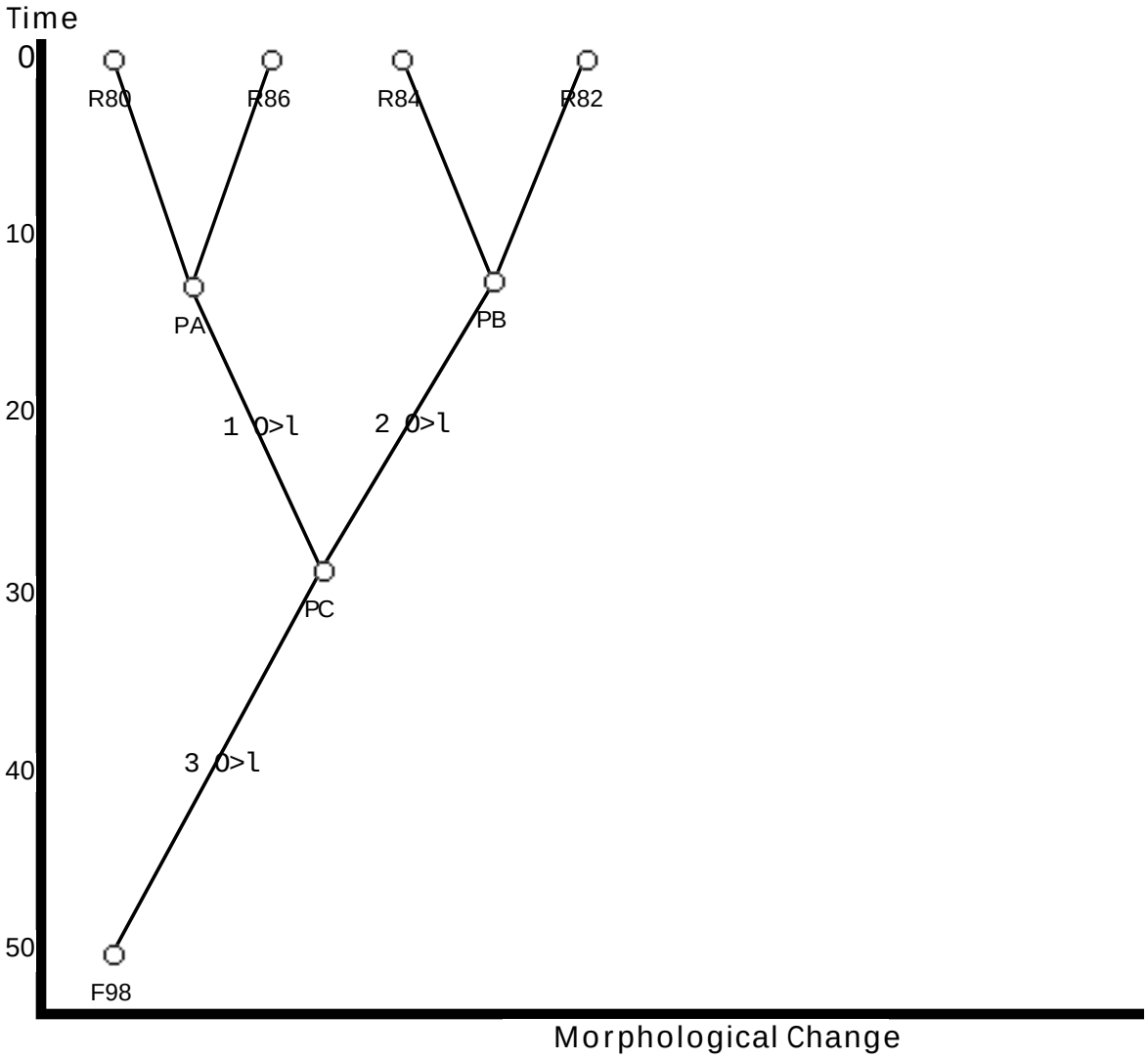
A phylogenetic tree is a diagram (Fig. 1) with time on the Y axis and evolutionary change (in PI this is assumed to be morphological change) on the X axis that illustrates a hypothesis of evolutionary relationships and the sequence of evolutionary events that gave rise to some group of taxa of interest (termed 'the ingroup'). In PI, phylogenetic trees are constructed of three kinds of pieces: nodes, links, and transitions.

Nodes represent taxa, for example species. Designations for nodes can have the prefix R, F, or P. Nodes that correspond to the observed taxa that are being studied, are numbered and have a letter prefix that is either R for Recent or F for Fossil. The ingroup in Figure 1 consists of R80, R86, R84 and R82. F98 is a fossil taxon from which the ingroup is believed to have descended. During tree construction, common ancestors of taxa are postulated to have existed in order to explain the data. Each of these nodes has a letter (e.g. A, B, C, etc.) with the prefix P (for Postulated).

Links connect nodes and represent hypothesized ancestor/descendant relationships between taxa. The slope of a link indicates the rate of morphological change: vertical lines indicate no change over time and the more a line tends to the horizontal, the more rapidly change is perceived as having taken place.

Transitions appear on links and represent the point at which evolutionary changes are believed to have occurred. Each transition represents some feature (character) of the taxa which has been numbered and described as having two conditions (states). One state is considered ancestral and is coded with a "0". The evolutionarily novel (or derived) state is coded with a "1". A transition shows the point where a character changes from "0" to "1" or from "1" to "0". Coded characters and states are organized by taxa in an associated data matrix.

Phylogenetic trees are just one type of a kind of branching diagram that appears often in biology. Other branching diagrams in biology include genealogies, that show relationships among individuals, and fate maps, that show how cells become canalized during the early stages of development. Both of these diagrams seeks to represent the systems of relationships that result from selective and reproductive processes at different hierarchical levels in biology (phylogeny at the level of species,



3 Steps

Characters

Taxa	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
R80	l	0	l																	
R82	0	l	l																	
R86	l	0	l																	
R84	0	l	l																	
F98	0	0	0																	

Problem: Synapomorphy 3

Figure 1. A phylogenetic tree as constructed using Phylogenetic Investigator.

genealogy at the level of individuals, and fate map at the level of cells). At the evolutionary level, these processes are microevolution (which causes lines to have a slope), speciation (which causes lines to branch), and extinction (which causes some taxa to leave no descendants).

Phylogenetic trees typically have dichotomous branching patterns, but trichotomies and even polytomies are possible. Each taxon is usually assumed to be derived from a single ancestral species, but using PI it is possible to create links to more than one ancestral species. These reticulating tree structures are occasionally used to illustrate hypotheses of interspecific genetic transfer (for example, hybridization).

A Brief History of Systematics

Traditional Linnaean classification still dominates how systematics is taught in most introductory biology texts. Linnaeus viewed species as unique and unchanging types or natural kinds. Each natural kind, according to Linnaeus, had particular morphological features that defined it. By describing those features systematically as taxonomic characters (a character being any attribute of an organism or group by which it may differ from another organism or group), each kind could distinguished from every other kind. Darwin's theory of evolution called for species to be historical entities which could change over time, produce new species, and go extinct. Systematics as a discipline has still not recovered from the impact of evolutionary theory and continues to be transformed today.

Systematics has become divided into two main schools of thought based primarily on different conceptions of the taxonomic goal (For a review see Ridley, 1986). Phenetic systematics seeks to represent a hierarchy based on the similarity of living things while phylogenetic systematics seeks to represent the hierarchy of evolutionary change. These forms of classification often result in similar, but different groupings. Phylogenetic inference seeks to define sets of species (taxa) which are all descended from one ancestral species (monophyletic). An incomplete set of descendant species is paraphyletic while a set which contains unrelated species is polyphyletic (Fig. 2). Phenetic classifications have been criticized because they sometimes group organisms that appear similar due to convergent evolution, but which are actually only very distantly related (resulting in polyphyletic groupings). They also sometimes fail to group things which are evolutionary related, but which have diverged greatly from one another (resulting in paraphyletic groupings) .

Although both phylogenetic and phenetic systematics seek to define groups based on shared similar characters, phylogenetic systematics makes a fundamentally different inference about the nature of some shared characters. Whereas phenetic classification treats all characters equally, phylogenetic classification is based solely on characters that are believed to demonstrate shared ancestry.

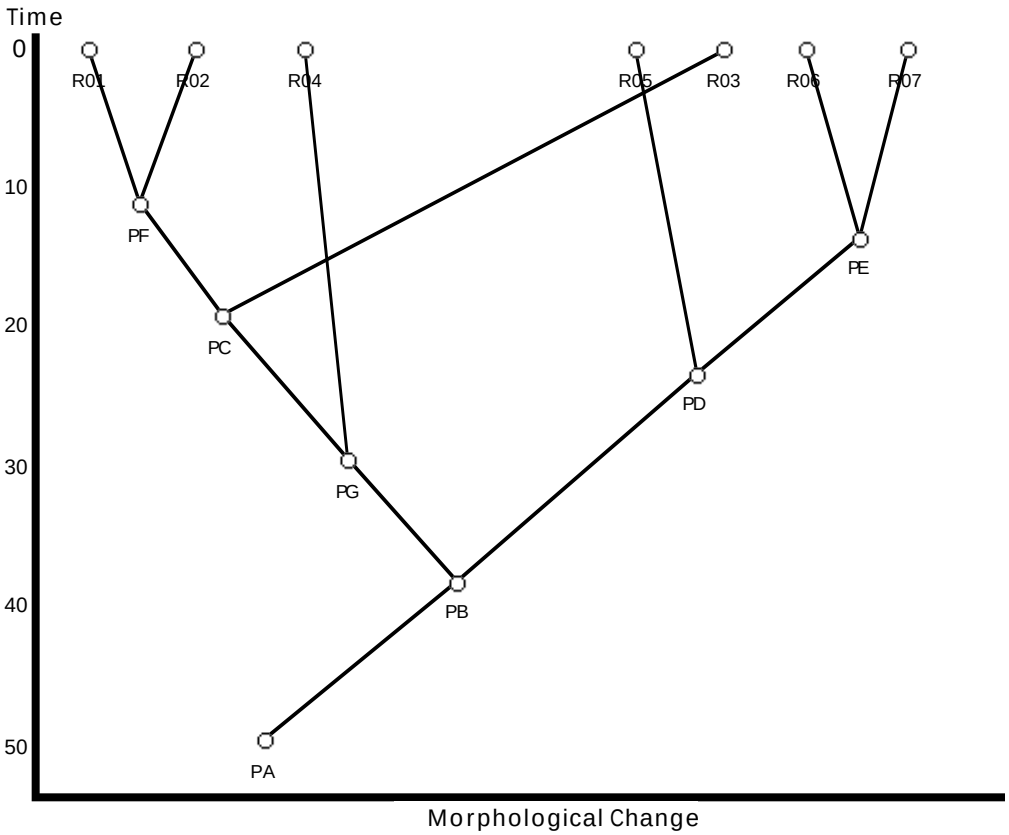


Figure 2. The placement of taxon R03 illustrates a paraphyletic grouping of {R01, R02 and R04} and a polyphyletic grouping {R05, R03, R06 and R07}. A group composed of all of the recent taxa {R01, R02, R03, R04, R05, R06, R07} is monophyletic.

Organisms share characters either because they are the result of shared ancestry (homology) or because they have evolved convergently in separate organisms (analogy). Only characters showing homology are useful for inferring phylogenetic relationships. In turn, homologous characters can be shared either because a character is generally ancestral or because it is modified from the ancestral. Ancestral characters may be retained by any combination of taxa regardless of phylogenetic relationship, but derived characters will be shared only by descendants of the ancestral species in which the character evolved. Therefore, only shared, homologous characters in the derived condition are useful for inferring phylogenetic relationships.

A METHODOLOGY OF PHYLOGENETIC INFERENCE

Should I draw phylogenetic trees and how do I do it?

Assumptions

Phylogenetic trees are hypotheses about how taxa are related to one another. Constructing phylogenetic trees requires a number of critical assumptions: (1) that all species in the ingroup, are descended from a single common ancestor, (2) that shared similarities among species are the result of sharing more recent common ancestors, (3) that ancestral and derived states of characters can be determined, and (4) that some form of character congruence indicates the most probable path of evolutionary relationship. Phylogenetic inference will yield accurate results to the extent that these assumptions are warranted. The reader should note that what is presented here is a general account of phylogenetic inference or what is sometimes termed Hennigian argumentation. Some recent forms of phylogenetic inference allow rejection or suspension of some of these assumptions.

The first assumption is an assumption of evolutionary process. Ancestral/descendant relationships, resulting from evolutionary processes, tie the diversity of living and fossil organisms together into a meaningful framework. Without this assumption, there would be no reason for supposing that there was any kind of underlying relationship among living things and phylogenetic inference would be meaningless. One could go through the mechanics of making groups based on shared derived characters, but there would be no coherent reason for doing so. (In fact, one school of systematics, which has come to be called transformed cladistics, has separated from the phylogenetic school arguing that the existence of patterns of character congruence, irrespective of models of evolutionary process, can serve as the raison d' etre for a systematic methodology. See Ridley, 1986 for a review.) On the other hand, the fact that phylogenetic inference appears to yield meaningful results is one of the pieces of evidence that has been used as support for the theory of evolution.

The second assumption deals with whether or not it is reasonable to postulate the links and common ancestors that will be used to construct a phylogenetic hypothesis. It is easy to imagine cases where this assumption would not be warranted and would result in a misleading analysis. Imagine the case of a species distributed over a continent which is subsequently inundated in a single event resulting in 5 islands with reproductively isolated populations. If the disjoint populations eventually evolved into 5 different species, one could deduce that any derived character states shared by these species could not be the result of recent common ancestors (Fig. 3). (Note: This issue is somewhat more complex than indicated here because, although there can be no common ancestors among populations after the inundation, some derived characters

may have had their origin prior to the separation of the populations and only been driven to fixation afterwards.) In this case the true phylogeny (Fig. 3) has only convergent characters. Every seemingly shared character must have

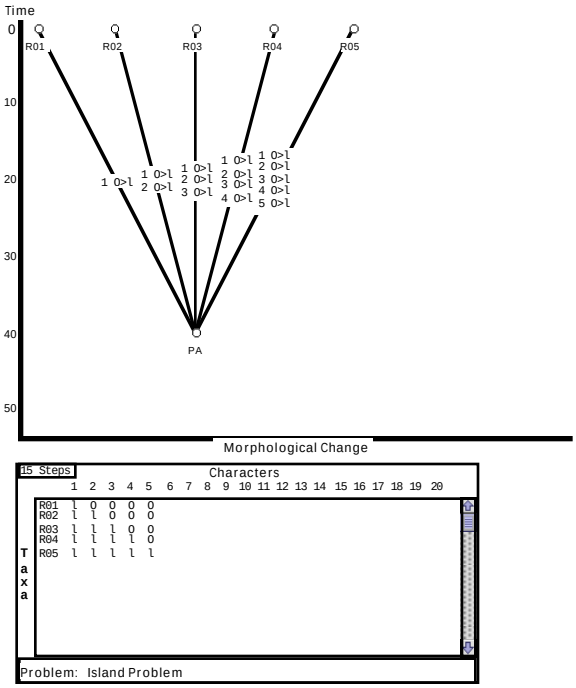


Figure 3. The "true" phylogeny in which 5 species are descended from a single common ancestor.

been independently acquired in each population because no more recent shared ancestors are possible. Methods of phylogenetic inference, however, would still yield a tree that explained all shared derived characters using shared common ancestors. In this case that assumption is unwarranted and the resulting phylogeny (Fig. 4) would be incorrect.

The third assumption deals with the determination of states of characters. If we cannot tell which characters are derived, then we cannot make groups on the basis of shared derived characters. Several techniques (e.g. outgroup, paleontological, and ingroup methods) are available for making determinations of states of characters and, although none are perfect, each can be evaluated to consider whether or not it can be counted on to provide meaningful results (Stuessy and Crisi, 1984). Futhermore, often several methods can be used and their results used to corroborate each other.

The last assumption deals with the issue of characters that suggest contradictory histories of descent. This can occur either because ancestral character states have been mistaken for derived states, or because of homoplasy (convergent evolution): either parallel appearance of a character in

the derived state or reversal of a character back to the ancestral state. In some cases, further study of the taxa themselves can illuminate the source of the conflict. A closer look at the taxa may show that two structures which appeared homologous are, in fact, substantively different. If further study does not diagnose the source of the conflict,

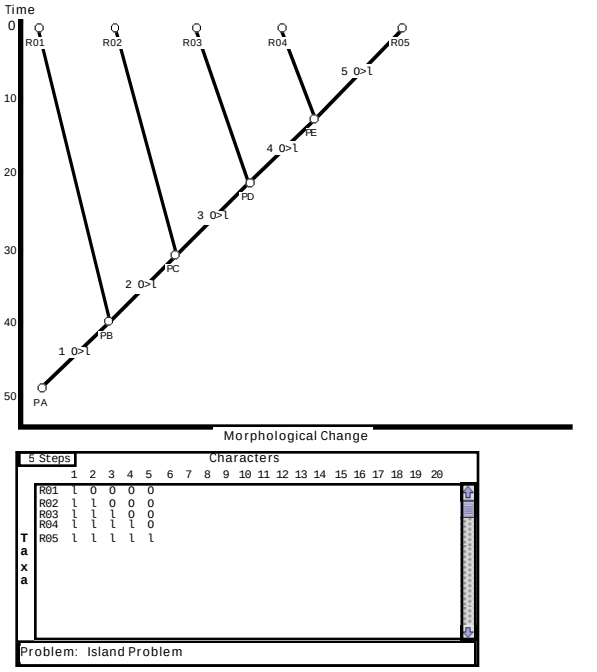


Figure 4. A phylogenetic tree representing 5 species descended from a single common ancestor through a nested series of more recent common ancestors.

statistical methods can be used to provide a basis for determining which of the possible trees should be preferred (Harvey and Pagel, 1991). The most common criterion has been termed parsimony and refers to selecting the tree that requires the fewest character state changes to explain the data. This criterion is based on the assumption that evolutionary events are rare and the hypothesis that invokes the fewest number of these rare events should be preferred. As long as the rate of evolutionary change is relatively low and can be assumed to be fairly equal among lineages, parsimony is probably a reasonable assumption (Felsenstein, 1983).

Other criteria for evaluating trees exist (See Harvey and Pagel, 1991). Compatibility analysis selects the tree or trees based on the largest possible set of non-homoplasious characters (Meacham and Estabrook, 1985). Compatibility analysis has been criticized for ignoring the potential that homoplasious characters may still carry some phylogenetic signal (i.e. some characters that could contribute meaningful information to the analysis would be ignored).

Maximum likelihood is another method that uses estimates of the probability for each possible evolutionary event to estimate the tree with the highest probability of having been produced. Maximum likelihood can be used where the assumptions required for parsimony are not valid.

Phases of Phylogenetic Inference

Phylogenetic inference can be divided into 4 phases: selection of ingroup and outgroup, identification of characters and states, assignment of polarity, and phylogenetic tree construction.

Selection of Ingroup and Outgroup

In scientific practice, the identification of the ingroup, or the group of taxa to be studied, is usually determined by a systematist who begins with a particular group of problem taxa in mind. Usually, it is assumed that the larger taxa are already monophyletic (Eldredge and Cracraft, 1980) and that the goal of analysis will be to establish the relationships of the ingroup. If these relationships are uncertain a lower-level study may be undertaken first to resolve uncertainty about the in-group. Lower-level studies often use large numbers of taxa to look for groups that appear to be monophyletic (Stevens, 1991)

The definition of the ingroup constrains selection of the outgroup. The outgroup" consists of taxa selected to determine which states of characters are ancestral or derived. The most desirable outgroup is the most closely related taxon to the ingroup, but in the event that this is unknown, any closely related species that are not within the ingroup can be selected (Stevens, 1991).

Determination of Characters and States

Any set of non-identical taxa can be divided by separating those that possess any feature "A", and those that do not. Any such feature can be used as a character for phylogenetic inference. For example, some plants contain enzyme A and some do not. "Enzyme A" would be the character and "present" and "absent" would be the two states of the character.

Some features do not seem to have just two states. For example, if we collected some evergreen branches, we might see that some have bundles of needles containing 1, 2, or 5 needles. This kind of multistate feature can be coded as a series of two binary characters in two different ways based on what is believed about the evolutionary sequence of events. If it is believed that 1 is ancestral to 2 and 2 is ancestral to 5 (1 -> 2 -> 5), then the first binary character will derived for those taxa with either 2 or 5 and the second binary character will be derived only for those with 5. If 1 is considered ancestral for both 2 and 5 (2 <- 1 -> 5) or if the sequence is unknown, then the first binary character will be derived only for those taxa with 2 and the second only for those with 5.

Assignment of Polarity

The assignment of character states as ancestral and derived, termed "polarity," is perhaps the most crucial step of phylogenetic inference. Phylogenetic methods require groupings based only on derived characters. Therefore, it is critical to be able to recognize them when they occur. Characters that have phylogenetic information will only contribute to the finished hypothesis if they are correctly polarized.

There are several methods for determining the polarity of characters. Three of the most important methods are outgroup, paleontological, and ingroup (Stuessy and Crisi, 1984). Each method has its strengths and weaknesses. Each can explain certain types of data and each has methods for explaining conflicting data. For all of the methods, conflicting data will be explained as homoplasy (convergent evolution) during tree construction.

Outgroup method

The outgroup method of determining polarity of character states is probably the most use commonly used. For each character, the state in which it exists in the outgroup is considered ancestral and the other state is derived. This method is based on the generalization that characters that have become derived for the ingroup will probably not be derived in a closely related group that diverged prior to the common ancestor of the taxa in the ingroup. The outgroup method can account for conflicting data by reevaluating whether some outgroups should be considered part of the ingroup or vice versa. The key to successful use of the outgroup method is to have well-resolved groups: knowledge about relationships among taxa in the outgroup improves the ability to estimate the ancestral state of characters for the ingroup. (See Maddison et al, 1984 for a more comprehensive description).

Paleontological method

The paleontological method uses fossil taxa for the outgroup. The state in which each character exists in the outgroup is considered ancestral and the other state is derived. Although one might think that fossil evidence could resolve all questions about the polarity of characters, there are two reasons why it does not: First, it is impossible to determine whether fossils represent taxa which are direct ancestors of living taxa or a taxon which diverged from the lineage leading to the present taxa. For this reasons, fossils should be treated essentially the same way as outgroups. Second, fossils often can not be accurately coded for many of the characters described from living taxa. Many features of organisms, like behavior, cannot be easily inferred from fossil evidence even under ideal conditions and fossils are often fragmentary and incomplete.

If the fossils are close in temporal position to ancestors of recent species and if a significant percentage of the characters can be unambiguously coded, then fossils can greatly improve the resolution of ancestral character states.

The paleontological method can account for conflicting data through appeals to the incompleteness of the fossil record.

In-group method

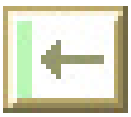
The in-group method is probably the weakest of the criteria described here. The most common form of a character among the ingroup is considered ancestral. For example, if 5 taxa have state A of character 1 and 3 taxa have state B, then state A is considered ancestral. This method is based on the generalization that the most common character states among the in-group represent the primitive condition. Older, larger, and diverse groups are less likely to preserve the primitive state as the most common character (Stuessy and Crisi, 1984). The in-group method is most useful as a form of corroboration or for use when other methods provide ambiguous results.

Tree Construction

Using parsimony, phylogenetic tree construction is a search among possible arrangements of relationships among taxa and characters that result in the fewest possible transitions of character states. For any data set, there are a finite number of possible arrangements of taxa and characters. For data sets with very few taxa, it is possible to construct all possible trees and see which require the fewest number of steps (transitions). The number of possible trees grows exponentially with the addition of taxa, however, and this method quickly becomes impractical to perform by hand. There are, however, strategies and heuristics which can allow the problem-solver to greatly limit the number of possibilities which must be considered. In most problems, only a few trees are actually supported by any of the data.

Each character in the data set, defines a group of taxa potentially descended from a postulated ancestor, and therefore can be seen as direct support for the existence of a postulated common ancestor or node. The real set of possible trees, then consists only of those trees which could be constructed from the available nodes.

Characters are inclusive/exclusive when they define identical, nested, or exclusive groups. For example, assume that character 1 defines a group of {R81, R82, and R83}. If another defines the same set of taxa, the characters are identical characters. If another character defines a subset or a superset of characters (e.g. {R81 and R82} or {R81, R82, R83, and R84}), the characters are nested with respect to each other. If another character defines completely different set of taxa (e.g. {R85 and R86}) the characters are exclusive with respect to one another. Characters conflict when they overlap incompletely. For example, assume that character 1 defines a group of {R81, R82, and R83} and character 4 defines a group of {R82, R83 and R84}. These two groups are contradictory because each character claims some, but not all of the taxa of the other. Character compatibility groups can be formed that place some or all of the characters into a hierarchical arrangement to evaluate how many of the



characters will support a particular hypothesis (arrangement of the taxa) and how many extra steps will be needed to account for incompatible characters.

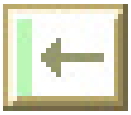
Ideally, all of the characters will agree in defining a single tree. In practice, some characters will define contradictory groups (groups that overlap incompletely). The largest possible group of inclusive/exclusive characters can serve as a working hypothesis from which to construct a phylogenetic tree. This tree can then be optimized for parsimony if so desired.

A phylogenetic tree is a branching path from a single point at which all of the character states are ancestral to several points where they are the same as the taxa in the ingroup. The lowest node, the node at the bottom of the tree, will be entirely ancestral, The postulated node above that will be linked to the lower node and will have a transition or transitions. Its states, then, are partially ancestral and partially derived. If it has the same states as any of the ingroup, they can be directly linked. The next postulated node has more derived states and may be linked to more recent taxa, until all of the taxa have been accounted for.

Constructing the phylogenetic tree involves adding postulated ancestors for each of the unique inclusive/exclusive characters, linking the ancestors together and to the taxa in the ingroup, adding the transitions for the characters which support the structure, and then distributing the homoplasious (conflicting) characters either as parallel gains or gains with subsequent reversals. (I suggest initially adding homoplasious characters as parallel gains, wherever possible. This makes it easy to spot duplicated characters each of which should be considered in order to evaluate alternate topologies and character optimizations.)

Once a tree has been constructed, it can be assessed and, if necessary, revised to ensure that it is a minimum length (most parsimonious) tree. Tree assessment should begin by examining each homoplasious character, beginning with the one that requires the most transitions, and considering (1) how many steps could be saved by "fixing" the character (rearranging the tree so that this character would have a single transition) and (2) how many more steps would be required in each other character that would be affected by those changes. If an arrangement is found that results in fewer steps, the tree should be restructured and then assessed again from the beginning. If an arrangement is discovered that results in an equal number of steps, assessment should continue until it is confirmed that no better tree is possible, and then all equally parsimoniously trees should be reported. The most difficult part of phylogenetic inference is assuring that all most parsimonious trees have been discovered. Rigorous assessment and systematic consideration of each homoplasious character provides the best probability of success.

For each most parsimonious tree, there should also be consideration of alternate character optimizations. Each homoplasious character should be considered for how it could be distributed on each most parsimonious tree. One



of the most important aspects of the interpretation of phylogenetic trees involves describing alternate hypotheses that could explain the data set and suggesting subsequent investigation that could provide insight into these uncertainties.

AN EXAMPLE PROBLEM USING PI

This example problem deals with a set of imaginary insect taxa among which several wing characteristics vary. Using diagrams of their wings as a data source, this guide will illustrate how to use PI to determine characters and states, assign polarity, and construct the most-parsimonious phylogenetic trees. This example is constructed to allow the reader to follow along using Phylogenetic Investigator by following the instructions given in italics. Program structures like windows, menus, and commands are printed in **boldface**.

For this example, I have selected only a subset of taxa (Fig. 5) from the data source (see Appendix B for the complete set of taxa). Taxa R04, R08, R11, R12, and R15 will be the ingroup. We will use R10 as an

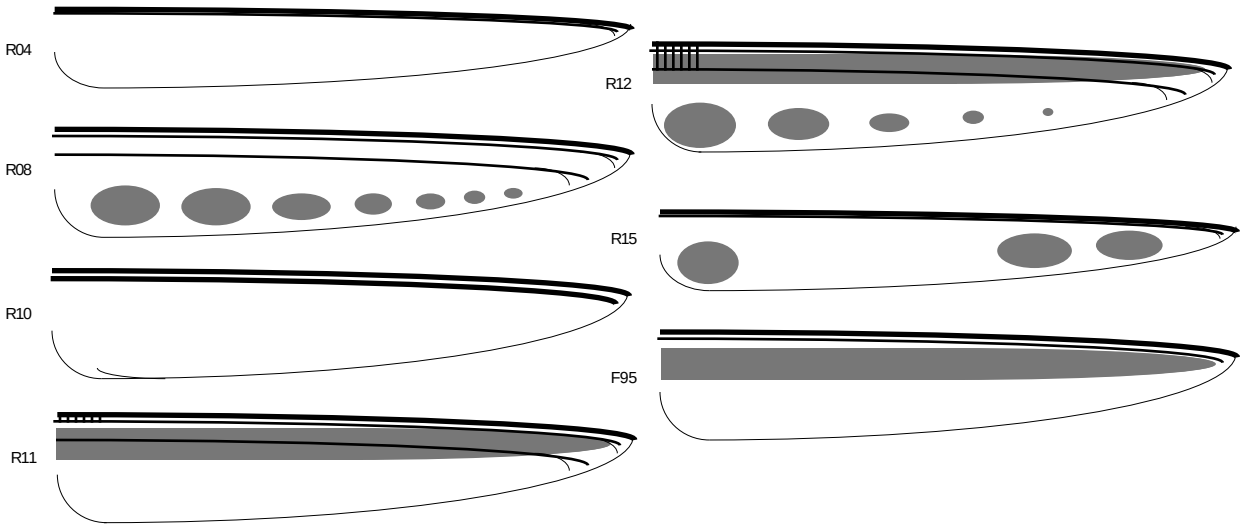


Figure 5. A set of taxa presented as an example problem of phylogenetic inference. R04, R08, R11, R12 and R15 are the ingroup, R10 and F95 are used to determine polarity by the outgroup and paleontological methods.

outgroup and F95 as a representative fossil. The decisions to use these particular taxa have been made more or less arbitrarily, in order to illustrate certain aspects of problem solving using PI. Ideally the ingroup will be composed of all of the taxa descended from some postulated ancestor and the outgroup will be the sister taxon, or the most closely related taxon not within the ingroup. In practice, one is constrained by current knowledge and the availability of study material. Our problem, then, is to define the system of evolutionary relationships among the ingroup. Having defined our problem, we are ready to start PI.

Double click on the program icon and, after the program finishes opening, select **Set-up Problem** from the **Problems** menu. This causes the **Set-up Problem** window to open which contains a scrolling list of taxa (Fig. 6).

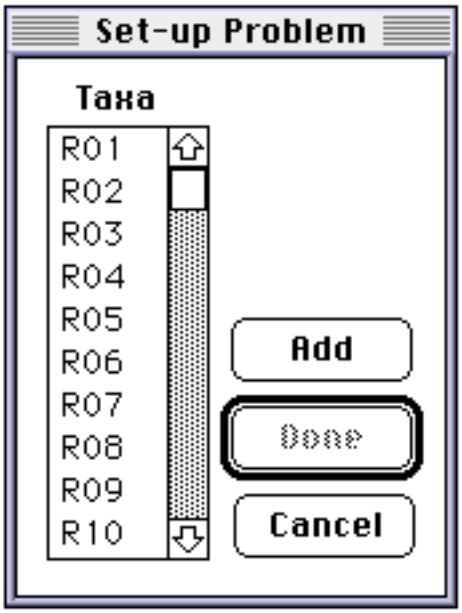


Figure 6. The Set-up Problem window. This window is opened by using the Set-up Problems item in the Problems menu.

Hold down Command key and select R04, R08, R10, R11, R12, R15, and F95. Click **Add** and then click **Done**. (Note that one could also select a single taxon, click **Add**, and repeat until all the desired taxa have been selected and then click **Done**). The recent and fossil nodes should appear in the drawing field and a new window, entitled **Chars & States** should open directly over them (Fig. 7).

At this point, we are ready to start identifying characters and states. We notice that some wings have spots and some don't. At this point, we need not be concerned which state is ancestral and which is derived. Simply enter the character and the two states. Click in the top field of the **Chars & States** window. Type "Spots" into top field and press tab -- this makes the Ancestral field active. Type "present" into active field and press tab -- this makes the Derived field active. Type "absent" into active field and press tab -- this moves the insertion point back up to the top field.

Click the zoom button at upper right hand corner of Chars & States window. This transforms the **Chars & States** window into a spreadsheet type format. The **Chars & States** window can be toggled between these two modes at any time and either window can be used for entering, modifying and deleting characters.

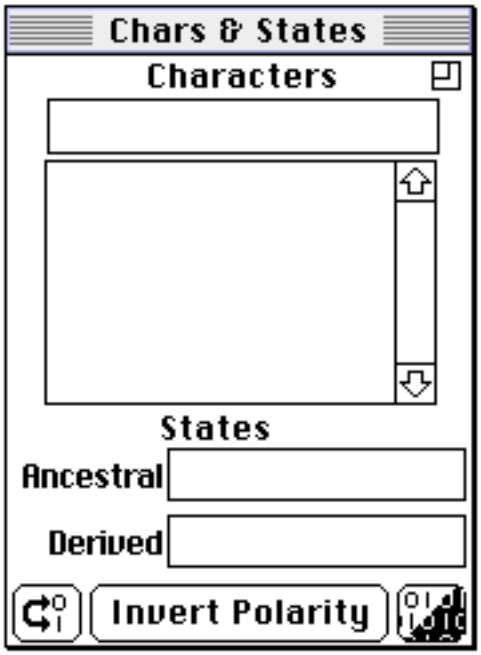


Figure 7. The compact version of the Chars & States window. Enter characters into the top field. Once entered they appear in the scrolling list. Enter states into the lower fields. The three buttons at the bottom allow exchanging character state names (left button), reversing polarity of data in the data matrix (right button), or both (middle button).

We notice that some wings have a little branch at the end of the veins and some don't. Click in the left most field of line 2 (the character field). Type "Vein branching" and press tab -- this moves the insertion point to the Ancestral field. You may notice that PI replaces any spaces within characters and states with underline characters. Type "present" and press tab -- this moves the insertion point to the Derived field. Type "Absent" and press tab -- this moves the insertion point to the next character field.

Enter the rest of the data as it appears in the Table 1. After entering all the data, click the zoom button at the upper right hand corner of the window. This will transform the **Chars & States** window back to the compact configuration in preparation for assigning polarity.

Table 1. Six characters and unpolarized states for the insect wing example.

	Characters	States: Ancestral	States: Derived
1	Spots	present	absent
2	Vein_branching	present	absent
3	Secondary_vein	absent	present
4	Stripe	absent	present
5	Vein_bars	absent	present
6	Wing_width	broad	narrow

Once all of the data has been entered, we're ready to start assigning polarity to the character states. Select the first line in the scrolling field in the middle of the small **Chars & States** window. This will bring up the two states assigned to it in the lower fields (Fig. 8).

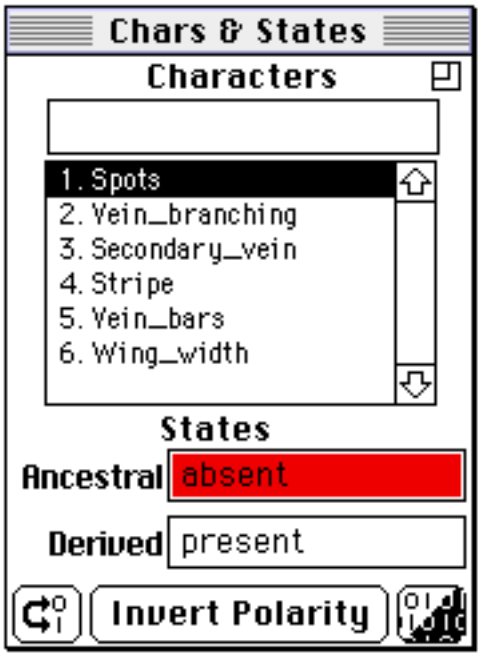
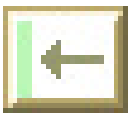


Figure 8. When a character is selected from the scrolling list, the states for that character can be modified or polarized. In this figure character 1 has already been polarized.

At the bottom of the **Chars & States** window are three buttons. The button on the left exchanges the words for the states in the **Chars & States** window. The button on the right inverts the coded data in the data matrix for a



character (exchanges 1's and 0's for a whole column). The button in the middle, labelled **Invert Polarity**, does both.

By looking at our data source, we see that spots are present neither in the outgroup (R10) nor in our fossil taxon (F95). Therefore, we will reverse the polarity of this character.

Press the left-hand button. This will exchange the two character state words -- after pressing the button your window should match Figure 7. As we look at the rest of the taxa we can see that some are already polarized correctly and others need to be exchanged.

When we get to character 4, we realize there is a problem. Character 4 is present in the fossil, but absent in the outgroup. In this case, we can use the ingroup method to evaluate which should be ancestral: it is present in 2 members of the ingroup, but absent in the other 3, therefore absent should be considered ancestral.

Polarize the rest of the characters. When you have polarized all your characters, they should match the table below.

Table 2. The characters and polarized states for the insect wing example.

	Characters	States: Ancestral	States: Derived
1	Spots	absent	present
2	Vein_branching	absent	present
3	Secondary_vein	absent	present
4	Stripe	absent	present
5	Vein_bars	absent	present
6	Wing_width	broad	narrow

Having finished polarity, we are ready to open the **Data Matrix** and code the data (Fig. 9). Select the **Data Matrix** item from the Windows menu. When the **Data Matrix** is initially opened, there should be a row for each taxon and a column for each character. These should all be 0's, unless the right hand **Invert Polarity** buttons have been used.

In the **Chars & States** window select character 1. Look at each taxon in turn, determine whether or not it possesses the ancestral or the derived condition for the character. If the taxon has the derived condition, click on the symbol where the row for that taxon and the column for character 1 intersect. This will cause the symbol to change from the ancestral "0" to the derived "1". A second click will cause it to toggle back. Code the rest of the data by selecting each character in turn and considering each taxon. At this point, we

are finished with and can close the **Chars & States** window. by selecting the **Chars & States** item from the Windows menu.

At this point we begin phylogenetic tree construction and begin to search for any patterns in the data matrix that indicate phylogenetic signal. In order to increase our ability to recognize patterns, we can organize the taxa more effectively and as we find patterns that appear to indicate phylogenetic signal, we can also restructure the matrix to aid recognition and memory. Organizing the taxa in the matrix as described here is not necessary for tree construction, but it can greatly aid finding patterns among the data.

Although taxa can be moved up and down in the data matrix at any time, characters can only be moved when no links are selected. Click on taxa to move them. This brings up a horizontal box which highlights the row to be moved and changes the cursor to a sideways arrow. Click between the two lines where the taxon is to be moved. To move a

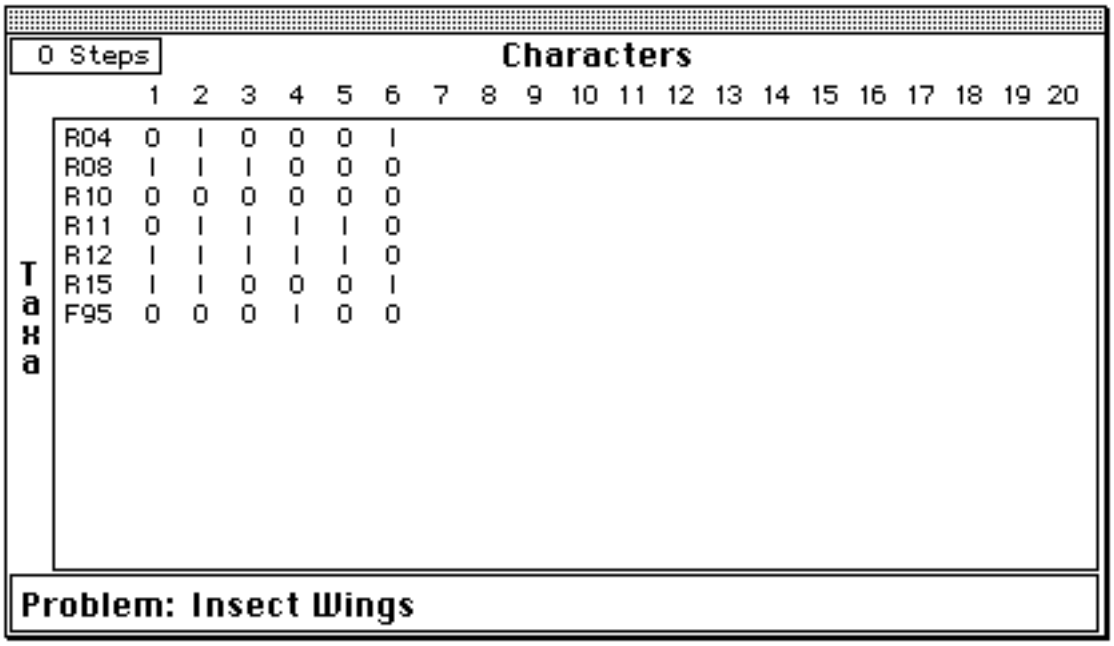


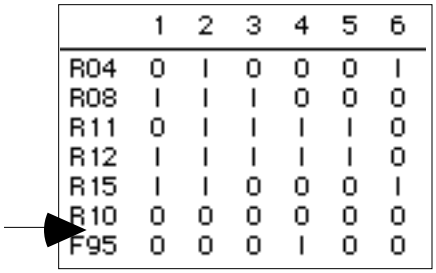
Figure 9. The Data Matrix window. Each row in the matrix represents the data for a taxon and each column represents a character. Characters are coded with symbols for ancestral (0) and derived (1) states.

character, click on the column heading when no link is selected and a vertical box which hilites the column is displayed. Click on a second column heading and the character is moved into that column.

	1	2	3	4	5	6
R04	0	1	0	0	0	1
R08	1	1	1	0	0	0
R10	0	0	0	0	0	0
R11	0	1	1	1	1	0
R12	1	1	1	1	1	0
R15	1	1	0	0	0	1
F95	0	0	0	1	0	0

Figure 10. The initial arrangement of the data matrix.

After initial inspecting the original data matrix (Fig. 10), we can make a change that will enhance our ability to recognize patterns: we can move the outgroup (R10) to the bottom of the matrix. This will separate the ingroup and outgroup taxa. Click on R10 and then, with the sideways arrow cursor, click between the rows where you want the taxon to appear -- in this case, just above F95 (Fig. 11).



	1	2	3	4	5	6
R04	0	1	0	0	0	1
R08	1	1	1	0	0	0
R11	0	1	1	1	1	0
R12	1	1	1	1	1	0
R15	1	1	0	0	0	1
R10	0	0	0	0	0	0
F95	0	0	0	1	0	0

Figure 11. R10 has been moved together with F95 separating the ingroup and outgroup taxa.

Now we can exclusively consider relationships within the ingroup. First, we notice that 6 and 3 have the opposite pattern. These characters are "inclusive/exclusive". If we put the 1's in character 6 together, we may be able to emphasize this pattern. Bring R15 up to just below R04 to put the 1's in character 6 together (Fig. 12).

	1	2	3	4	5	6
R04	0	1	0	0	0	1
R15	1	1	0	0	0	1
R08	1	1	1	0	0	0
R11	0	1	1	1	1	0
R12	1	1	1	1	1	0
R10	0	0	0	0	0	0
F95	0	0	0	1	0	0

Figure 12. R15 has been joined with R04 on the basis of character 6.

We can see now that 3 and 6 are exclusive from each other and can both be nested within 2. We can move 6 to the other side of 2 so as to emphasize

that pattern (Fig. 13). Click on the column heading for character 6. Once it is outlined, click on the column heading for character 2.

	1	6	2	3	4	5
R04	0	I	I	0	0	0
R15	I	I	I	0	0	0
R08	I	0	I	I	0	0
R11	0	0	I	I	I	I
R12	I	0	I	I	I	I
R10	0	0	0	0	0	0
F95	0	0	0	0	I	0

Figure 13. Character 6 has been moved to the other side of character 2 from character 3 to emphasize this division of the taxa.

Now we can see that 4 (disregarding the outgroup problems) and 5 nest nicely within 3. We can also see that 1 just doesn't fit at all. 1 conflicts with 6 and 4 and 5. Move 1 to outside the group of inclusive/exclusive characters to set it apart (Fig. 14).

	6	2	3	4	5	7	1
R04	I	I	0	0	0		0
R15	I	I	0	0	0		I
R08	0	I	I	0	0		I
R11	0	I	I	I	I		0
R12	0	I	I	I	I		I
R10	0	0	0	0	0		0
F95	0	0	0	I	0		0

Figure 14. Character 1 has been separated from the other taxa to separate homoplasious and non-homoplasious characters .

The organization of this matrix now represents an inclusion/exclusion hypothesis. It shows us that R04 and R15, based on sharing character 6, will be a group separate from R08, R11 and R12 (which share character 3). Also, we can see that the group of 3 taxa will contain a subgroup composed of R11 and R12 (because they share 4 (with homoplasy in F95) and 5. Now, with our completed inclusion/exclusion hypothesis, we're ready to draw some phylogenetic trees.

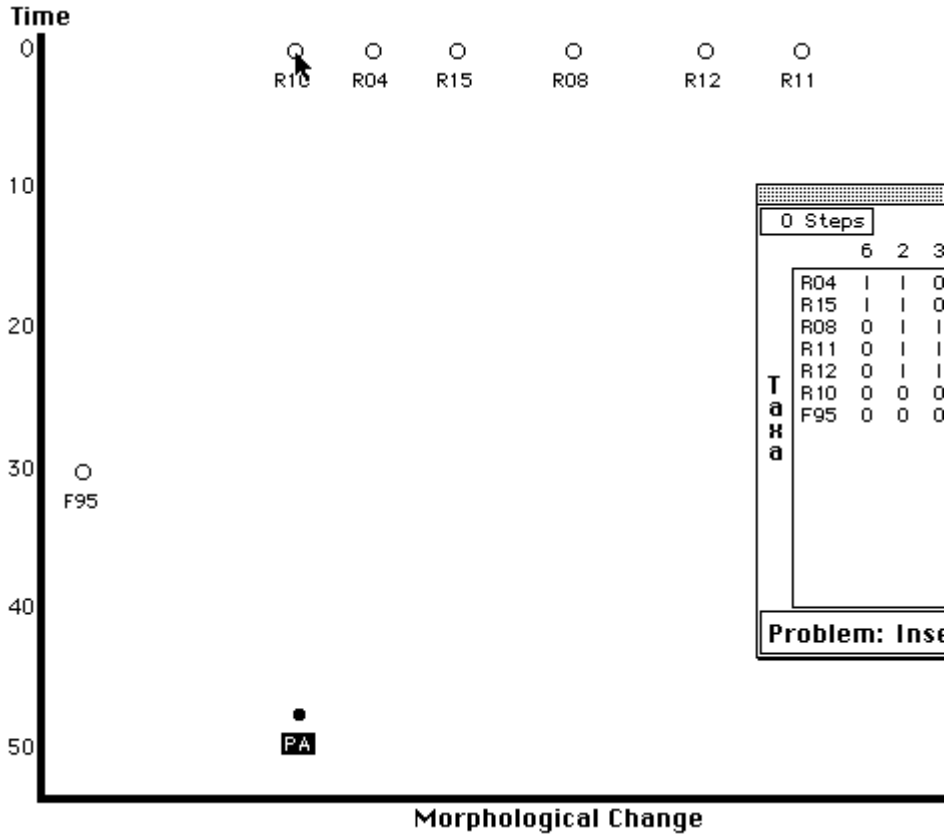


Figure 15. To make a link, select a second node while pressing the shift key. (Or press the shift key and select two nodes).

First, we can move the recent taxa at the top to represent the order described in the Data Matrix. F95 will at the extreme left and R10 at the left of the recent taxa. Then R04 and R15 will be together, then R08, and then R11 and R12. Within the two subgroups R04, R15 and R11, R12, order is not significant. This order will produce a diagram which appears to have a trend of increasing numbers of derived characters from left to right. This trend is actually an illusion: the branches could be arranged such that R04, R15 was on the right of R08, R11, R12. Nevertheless, it is often useful to use a consistent form of representation because it can facilitate both construction and interpretation.

Select **Add Node** from the Actions menu, and click near the bottom of the screen. This node will be our outgroup node. When the node appears it is selected. Because the outgroup node and the outgroup have the same distribution of characters states (all ancestral), they can be immediately linked. Holding the shift key down, we click on R10 (Fig. 15). This forms a link and unselects both nodes. Note that R10 is connected to PA with a vertical line. This indicates qualitatively, in addition to the fact that no transitions will appear on this line, that there are few or no differences between the ancestor and this descendant taxon.

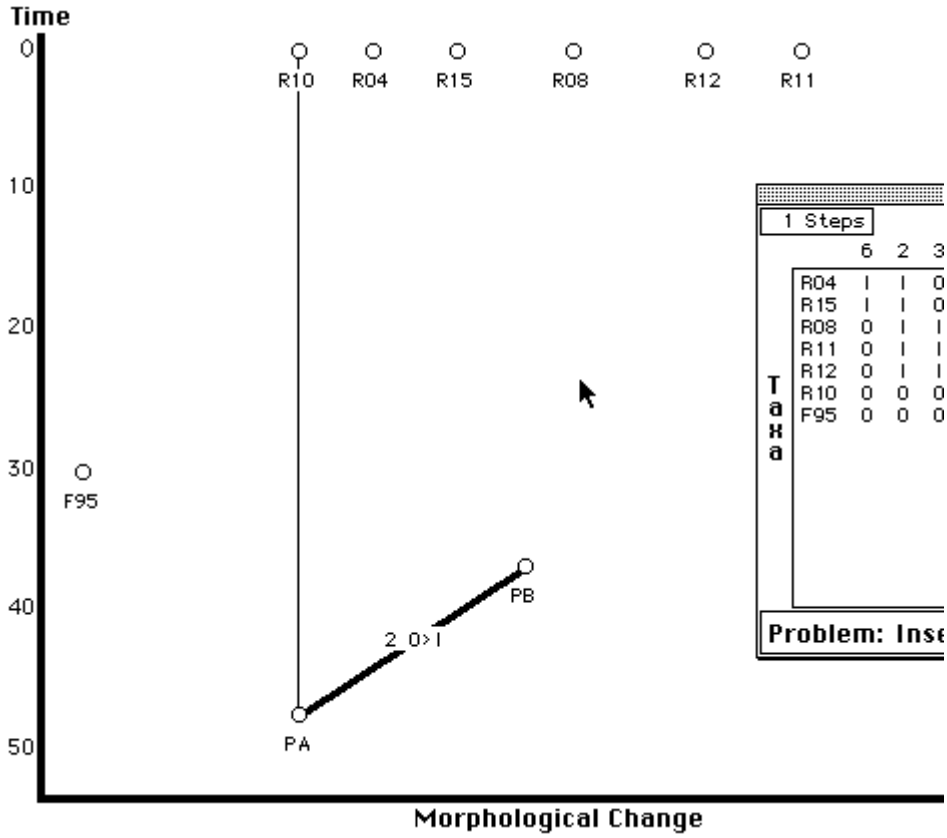


Figure 16. Click on a link to select it. Press the column heading in the data matrix while a link is selected to add a transition to a link.

We then create a second node. This node will be the ingroup node, from which all the taxa in the ingroup (all the taxa that share character 2) are descended. After linking this node to PA, click on the link, selecting it, and then click on the character 2 column heading in the **Data Matrix**. This will add a forward transition for character 2 to the selected link (Fig. 16). Note that there are no taxa which possess only character 2, so PB should not be linked directly to any taxa in the ingroup.

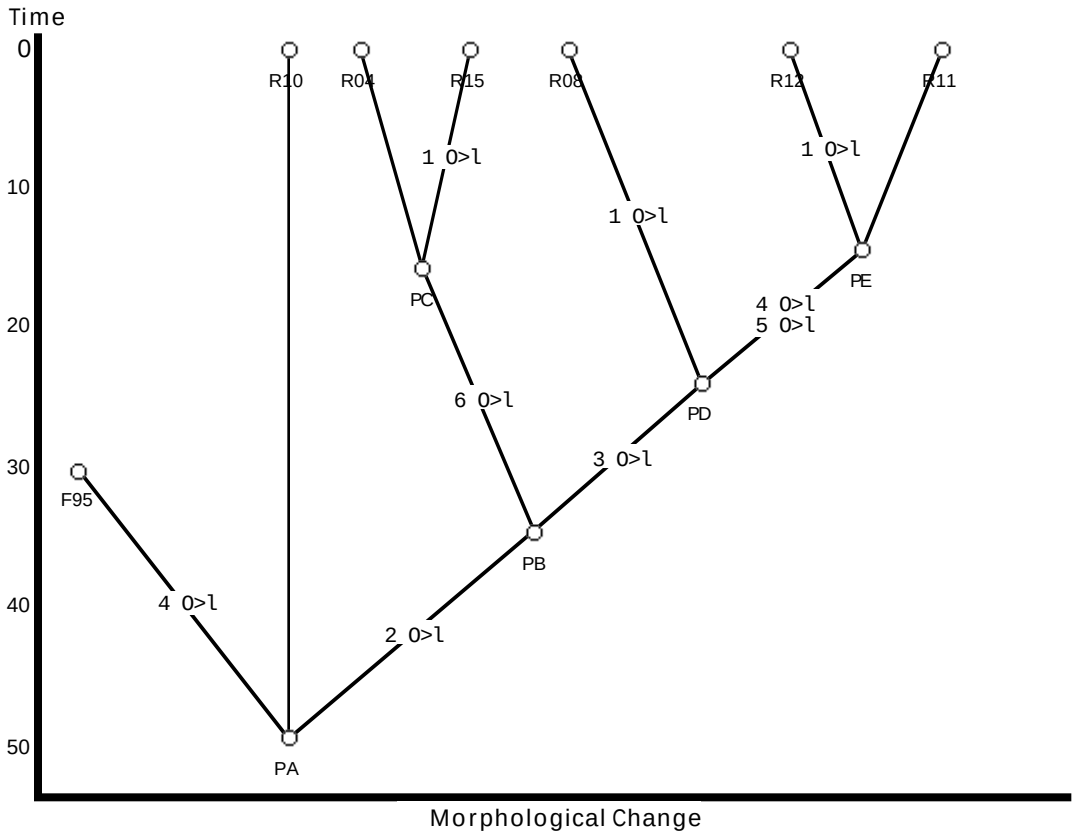


Figure 17. In this most parsimonious tree, character 1 is distributed as 3 convergent forward transitions (in R04, R08, and R12).

We can then add a node (PC) under R04, R15 for character 6, a node (PD) under R08, R11, R12 for character 3 and a node (PE) under R11, R12 for characters 4 and 5. We can then link up all the taxa (eventually linking F95 also to the outgroup node with a homoplasious gain for character 4). We are then left with character 1. Character 1 can be added as 3 separate gains in R08, R12, and R15. This implies that character 1 evolved separately three times (Fig. 17).

This optimization of character 1 provides an avenue of subsequent research. If character 1 evolved three separate times in recent history, perhaps some major climatic or environmental change occurred where these taxa occur. Perhaps a new predator appeared or arrived. Perhaps these taxa invaded new areas that placed similar constraints on evolutionary development. This optmization of character 1 predicts that if we discover fossil taxa closely related to PB, PC, PD and PE, none of them will have character 1 in the derived state. All of these are avenues for gaining further insight into character 1.

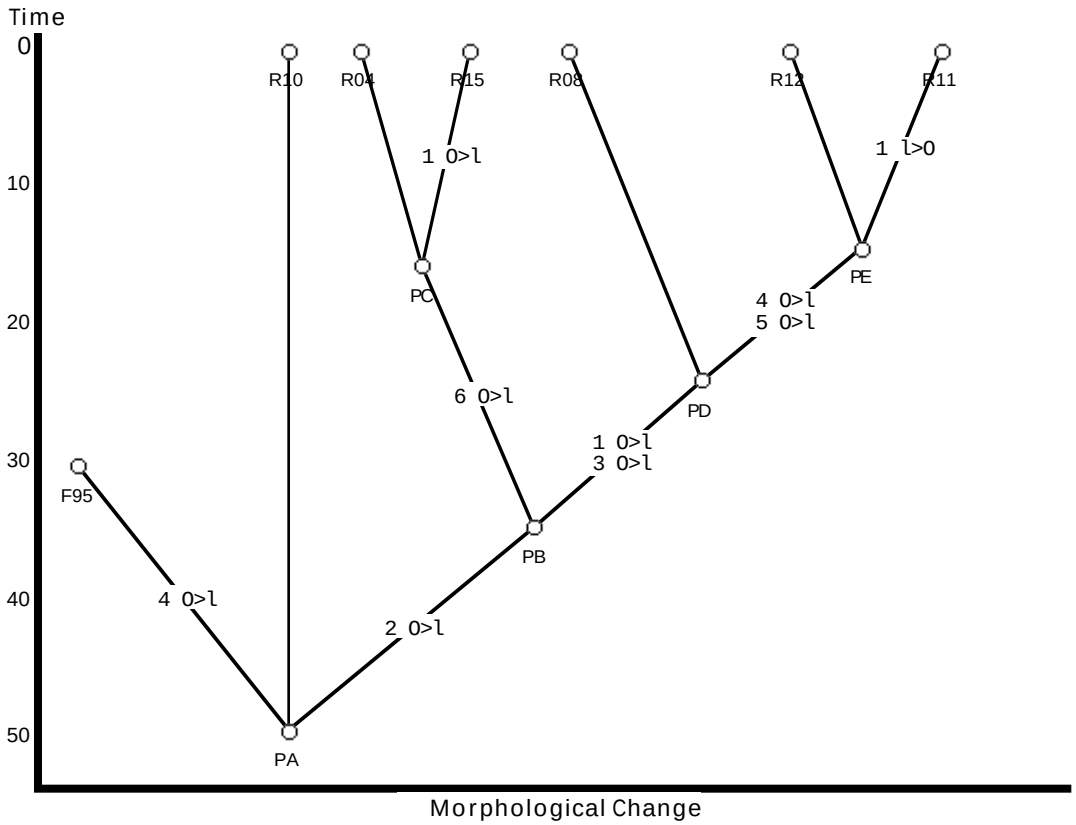


Figure 18. In this most parsimonious tree, character 1 is distributed as 2 convergent forward transitions (in PD and R15) and a reversal (in R11).

Character 1 can also be two gains (in R15 and PD) and a loss (in R11) (Fig 18). To generate this optimization from the previous arrangement, select link R12-PE and click the character 1 button twice. This causes the transition to change first to a reversal and then to be removed entirely. Do the same for link R08-PD. Then select link PB-PD and click (the character 1 button) once -- this adds the forward transition. -- and select link R11-PE and click (the character 1 button) twice. This adds a reversal for character 1.

This optmization of character 1 predicts that if we discover fossil taxa closely related to PD, it will have character 1 in the derived state, but that taxa closely related PB and PC will not. Biogeography might again offer insights into parallels between R15 and the other taxa.

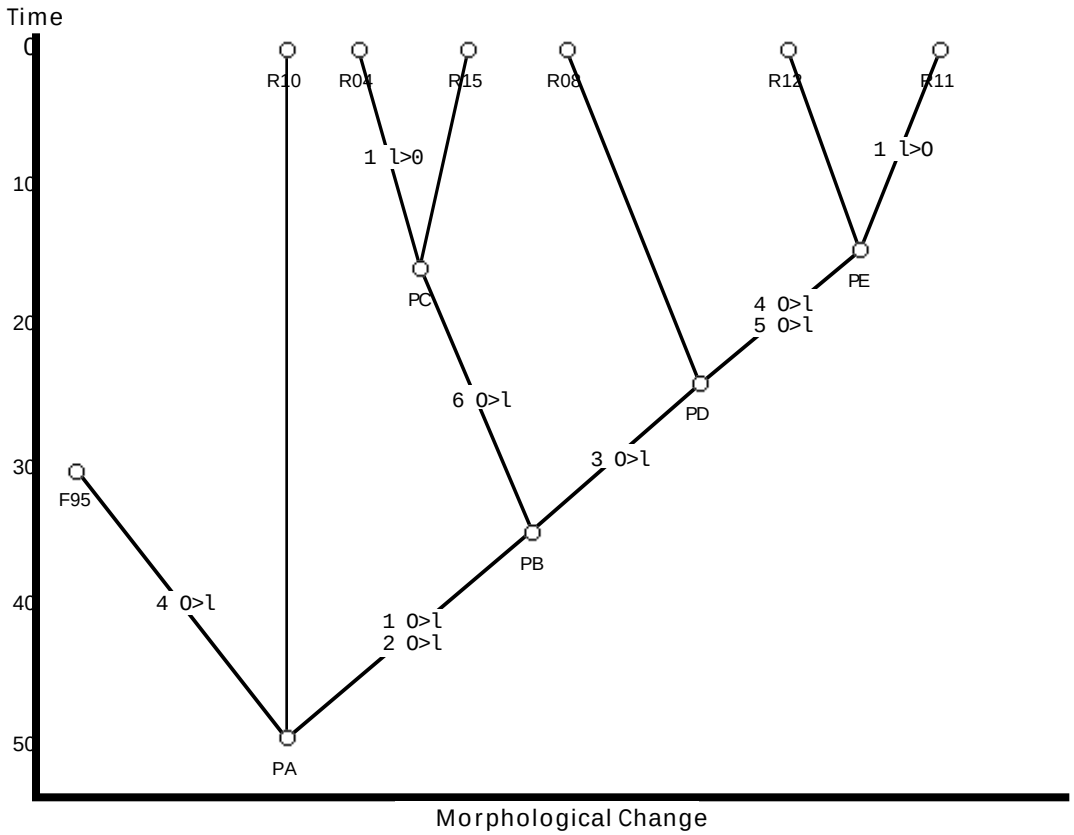


Figure 19. In this most parsimonious tree, character 1 is distributed as 1 convergent forward transitions (in PB) and two reversals (in R04 and R11).

Character 1 can also be 1 gain (in PB) and two losses (in R04 and R11) (Fig. 19). To generate this optimization from the previous arrangement, select link PC-PD and click the character 1 button twice. This causes the transition to change first to a reversal and then to be removed entirely. Do the same for link R15-PC. Then select link PA-PB and click once -- this adds the forward transition. -- and select link R04 PC and click twice. This adds a reversal for character 1.

This optmization of character 1 now focuses attention on the taxa which appear to have lost character 1. Is there some environmental or biogeographical factor that can be associated with the loss? Now, if we discover fossil taxa they should all have character 1 in the derived state.

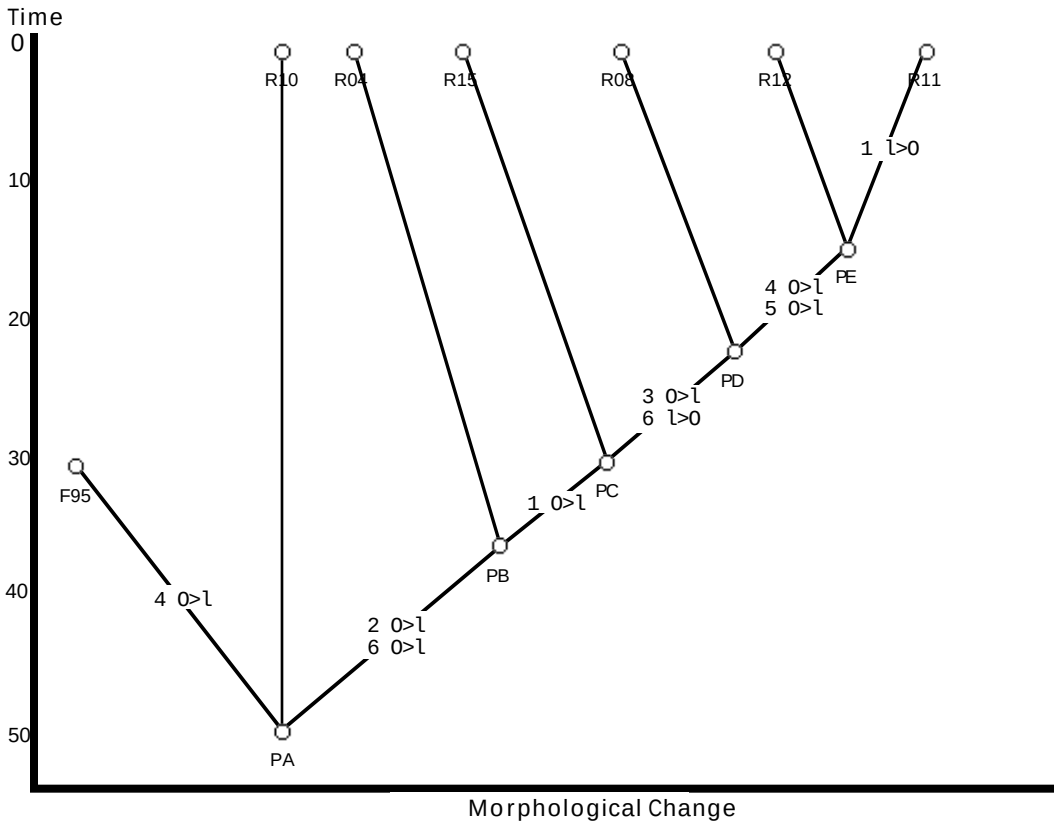


Figure 20. In this most parsimonious tree, character 1 is distributed as 1 forward transition (in PF) and 1 reversal (in R11). Saving a step in character 1 is achieved by explaining character 6 using 2 steps -- a forward transition (in PB) and a reversal (in PD).

It is also possible to construct a second topology which improves character 1 by a step, but adds a step to Character 6 (Fig. 20). Character 6 is then gained in PB and lost in PD and Character 1 is gained in PF and lost in R11. To construct this topology, select link PB-PD and select **Reassign Link** from the **Actions** menu. Use the pop-up menu PB to change the node assignment to PC. Then select link R04-PC and Reassign Link from PC to PB. Instead of using the menu command, it is also possible to select the link and hold down the shift key while selecting the node to be reassigned. This causes the pop-up menu to appear right on the drawing field.

Having constructed a phylogenetic tree or a series of phylogenetic trees, interpretation is necessary for them to become meaningful. Each speciation event and each character transition should be considered thoughtfully from a historical perspective: What was the environment? What other evidence (ecology, biogeography, etc.) might support or contradict the evidence used to construct the tree? The homoplasious characters are of particular interest: are these characters highly variable among other taxa? Is it possible to look at the

character more closely to investigate how it has been defined? Does the homoplasious character vary in function across groups?

If we were dealing with plants, rather than insects, we might be asking whether some of the character incompatibility observed was due to the presence of hybrids. Hybridization is rare among animals, but often causes problems for phylogenetic inference with plants because hybrids may share characteristics of taxa from different lineages. Alternatively, derived characters are often recessive and some hybrids may have no derived characters at all. Hybrids can be dealt with in a variety of ways. One way is to simply remove them from the sample. Hybrids are not really taxa in that they often cannot themselves reproduce. Another way is to place them with links between them and the taxa from which they are derived.

PHYLOGENETIC INVESTIGATOR REFERENCE MANUAL

Phylogenetic Investigator (PI) is designed to facilitate modeling and practicing fundamental phylogenetic inference. We believe that beginning students of phylogenetic inference should be able to (1) inspect the data, make inferences, and build representations one step at a time, (2) vary representational features of their trees (such as angle of divergence and time between speciation events), (3) create reticulate tree patterns, and (4) view all of the character transformations at one time. No other available software package allows students to do any of these things. It was for these purposes that we created Phylogenetic Investigator.

PI provides tools for managing and manipulating up to 20 characters of binary phylogenetic data for 15 or fewer taxa. PI has been designed with 2 data sets in mind: the Caminalcules and the Dendrogrammaceae, but other data sets can be adapted for use (See the section on Set-up Problems below). With PI, students can wrestle with the assumptions, methods, goals, and limits of phylogenetic inference. Once students have become conversant with the concepts and functional relationships implied by phylogenetic inference other more research-oriented tools may be better suited. More advanced tools can allow students to use more complex transformation series, weight characters, and experiment with the effects of including and excluding characters and taxa.

The guide to PI below is organized systematically to facilitate finding information about particular features of the program. Windows are described first and then menus. Dialog boxes are described with the menu item that opens them.

Windows

PI uses two windows for data management (Chars & States and Data Matrix) and one for tree construction (Phylogenetic Tree). Most will open automatically when a problem is selected or set-up. None of these windows have close boxes and must be opened or closed using the Windows menu.

Chars & States

The Chars & States window will open automatically if Set-up Problem has been used to pose a problem. This window has two configurations and the user can move between them by clicking the zoom button at the upper right hand side of the window. Data can be entered using either configuration and the small configuration (Fig. 21) can be used for polarizing characters.

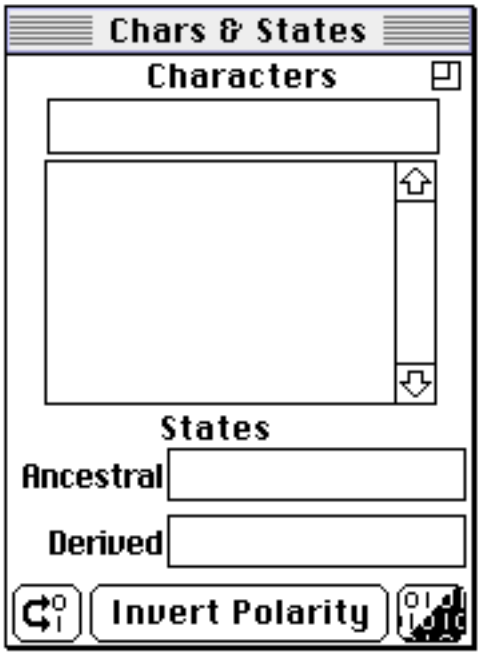


Figure 21. The compact version of the Chars & States window. Enter characters into the top field. Once entered they appear in the scrolling list. Enter states into the lower fields. The three buttons at the bottom allow exchanging character state names (left button), reversing polarity of data in the data matrix (right button), or both (middle button).

Small configuration

Upon opening, the upper left field should be active. The user enters Characters here, causing them to be entered into the list of characters below. As characters are entered here, a column is automatically created in the Data Matrix window for coding. A total of 20 characters can be defined. The user is automatically prompted to enter first the ancestral and then the derived state. All of these fields can contain only a single word and the program will automatically substitute underline characters for spaces, if entered.

Items in the list can be modified by shift clicking -- this will bring up a dialog that asks what the new item should be. Items can also be deleted by option clicking -- this will bring up a warning/confirmation dialog. By selecting different characters from the list, one can subsequently modify states for that character.

At the bottom of this window are three buttons. The middle button, labelled Invert Polarity, exchanges the terms entered for ancestral and derived characters and also exchanges 1's and 0's in the column for that character in the Data Matrix. The button to the left only exchanges ancestral and derived terms and the button on the right only inverts the polarity of the column in the data matrix.

Large configuration

In this mode, the window has a spreadsheet type format (not pictured). Characters and states can be entered, but only in order. A tab will move the insertion point to the next active field. A return will move the insertion point down one row (if that row is active). If a character is deleted, the user is asked to confirm deletion before the line of data from the data matrix is removed.

Data Matrix

The Data Matrix (Fig. 22) is a palette, meaning that this window will float over all the others. It is often useful to move this window to the right so that only that portion which contains data is visible. There are three fields in this window. The Problem field at the bottom shows the title of the problem that is currently being addressed. This field will be filled in automatically when a model or practice problem has been selected, but it is also user modifiable. The contents of this field is what is used as the default file name when a problem is saved for the first time. This field also communicates with the problem field in the expanded Chars & States window. The small field in the upper left shows the current tree length (in unweighted transitions). The large, central field contains the data matrix currently being used for problem-solving.

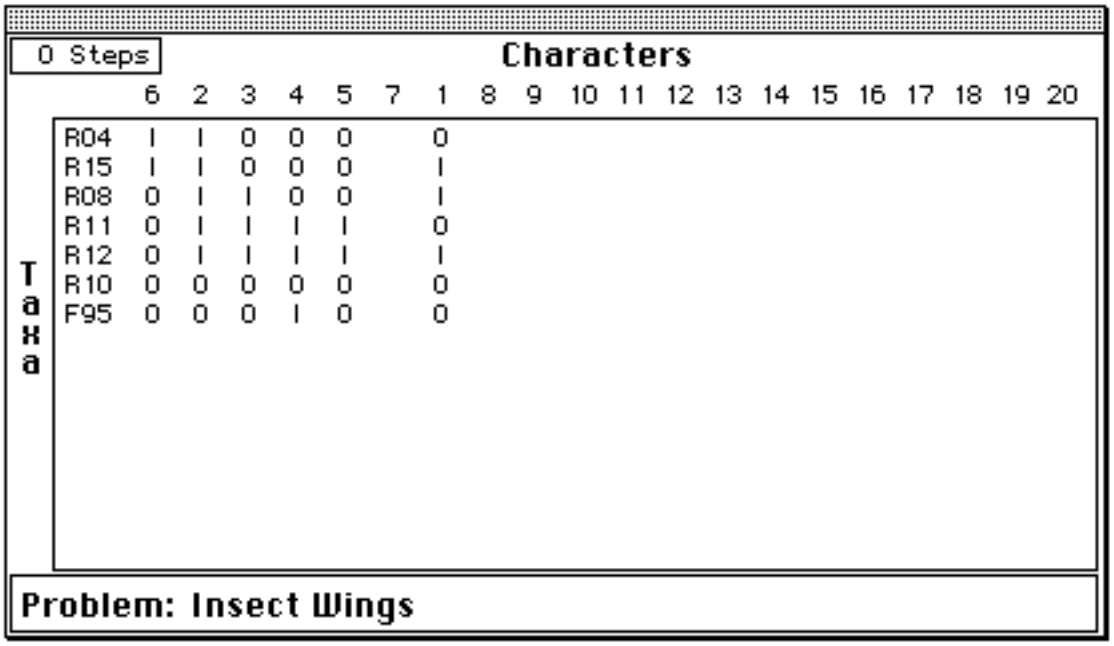
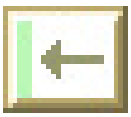


Figure 22. The Data Matrix. Data consists of 1's for ancestral and 0's for derived character states and is organized with taxa in rows and characters in columns. At the upper left is the number of unweighted transitions in the tree. The field at the bottom is user modifiable and contains the name of the problem.

In the data matrix, characters are in columns and taxa are in rows. When a link between nodes is selected in the tree construction window, a click on a



character button (in the row above the matrix) will add a transition for that character to the selected line. A second click will change the transition into a reversal and a third click will remove the
transition from the line. The tree length field is updated automatically. States for taxa can be modified by clicking on the state character for a taxon. This will toggle between the ancestral and derived characters. Holding down the option key and clicking allows one to change the character to X to indicate missing data. Rows can be moved by clicking on them, which will bring up a box outlining the row to be moved and different cursor. A second click, indicating where the row should be moved to (between rows or above or below another row) will move the row to this location. Columns can be moved by clicking on a character number above the data table while no line is selected on the phylogenetic tree. This will reveal a box outlining the column to be moved. Click on another character button to move the column into that space in the matrix.

Phylogenetic Tree

In Phylogenetic Investigator, trees are constructed from nodes, links, and transitions. Nodes and links can be selected by clicking on them. To de-select everything, click on the background. Nodes can be moved by dragging. To form a link, use the shift key to select two nodes. These nodes will be automatically linked and the link will subsequently follow the nodes if moved. Transitions are added to links by clicking on the character buttons in the Data Matrix window.

About Nodes

All organism designations (Nodes) begin with letters that indicate the organism's status R for recent, F for Fossil, and P for Postulated. Recent and Fossil organisms are numbered and can be constrained temporally (this property is controlled by the Time checkbox in the settings window). Postulated organisms have sequential letters are free to move in both axes. When nodes are selected, they can be deleted by using the Remove Nodes menu item. All associated links will also be removed (this is sometimes a fast way to reconstruct a tree for a revision). Holding the shift key down allows two nodes to be selected. Once a second node has been selected, a link is formed between them and both are de-selected. Holding the shift key down and selecting a node while a link is selected brings up a pop-up menu that allows reassigning the link from the selected node to any other node. If a node is selected and the Add Node command is executed while holding the shift key down, a new node will be added and linked to the previously selected node.

About Links

Links can be selected by clicking on them. Selected links can be removed or reassigned (by using menu items). Selected links can have transitions assigned to them by clicking on the character button in the Data Matrix. Holding the shift key down and selecting a node while a link is selected brings up a pop-up menu that allows reassigning the link from the selected node to any other node.

Settings

The settings window (Fig. 23) allows the user to modify the time scale on the phylogenetic tree, change the characters used for ancestral and derived characters, and to apply or remove temporal constraint from a problem. The temporal constraint is turned on by default. If turned off, it will remain off until turned on again (even between uses of the program).

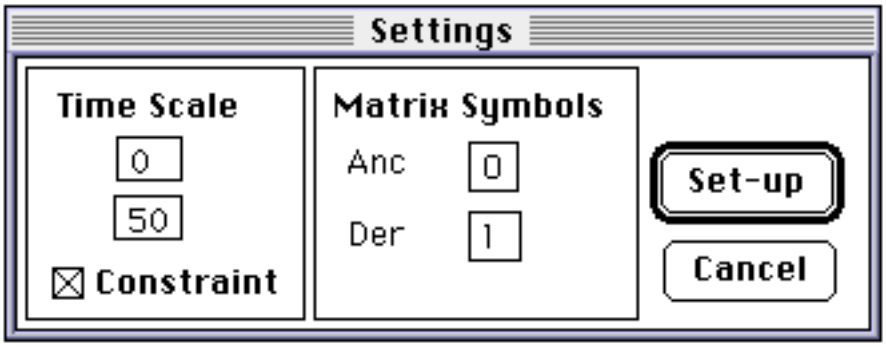


Figure 23. The Settings window. The time scale and constraint may be modified during problem-solving. Modifying the Matrix Symbols during problem-solving may result in erratic behavior.

The matrix symbols currently in use are the uppercase letter 'O' (as in Oliver) and lowercase letter 'l' (as in lollipop). These were what I thought looked the best after trying many other possibilities. (Real 1's and 0's don't line up right vertically as nicely as O's and l's.)

Note: Changing matrix symbols during problem-solving is probably a bad idea. It might not be fatal, but could cause some odd behavior with transitions.

Menus

Apple

The Apple Menu contains the About Phylogenetic Investigator item which opens the Phylogenetic Investigator splash screen.

File

New

This clears the drawing field, data matrix, and characters and states.

Open...

This item will open a PI Treefile

Save

Save As...

These items generate a PI Treefile. Treefiles contain a snapshot of the current state of the problem: Characters, states, coded data, nodes, links, locations, and transitions.

Open Nexus

This feature has not yet been implemented. Look for it in future versions of PI.

Save Nexus

This saves the current data in a form which can be read by PAUP and MacClade 3.x.

Save MacClade 2.1

This saves the current data in a form which can be read by the older version of MacClade.

Export Tree

This item creates a ClarisWorks PICT file with the current tree and Data Matrix.

Print...

This opens a dialog box (Fig. 24) with two radio buttons and three checkboxes. One can select to print the data as a practice problem or as a setup problem. As a practice problem, the data matrix and phylogenetic tree are put together on a single page and printed. As a setup problem, one can select phylogenetic tree, data matrix, and characters and states for printing. Each will appear on a separate page.

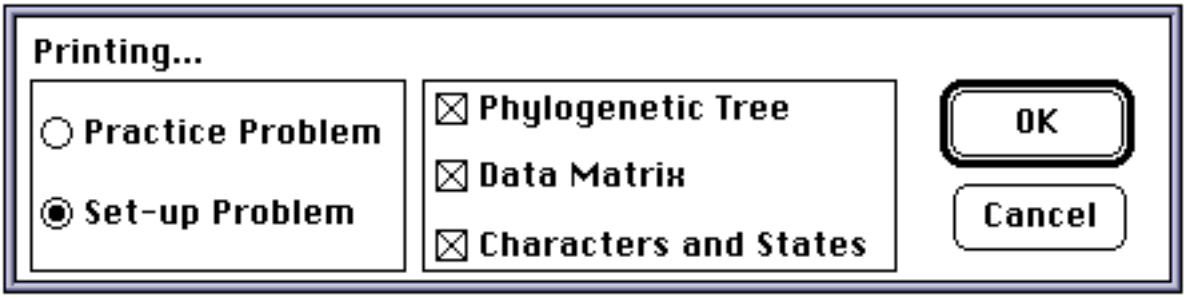


Figure 24. The printing dialog box.

The phylogenetic tree and data matrix are printed exactly as they appear on the screen. The Characters and States are automatically transferred to a form for printing.

Quit

This item retains the current problem and quits the application

Edit

Cut, Copy, Paste, and Clear are implemented.

Actions

Add Node

When this item is selected, the cursor changes to appear like a postulated node and when the mouse is clicked, a new postulated node is placed at that point and selected.

Remove Link

If a line is selected, this command will remove it and updates tree length if transformations were present on the link removed. Links can also be deleted by pressing the delete key.

Remove Node

If a node is selected, the program confirms and then removes the selected node and attached links. Nodes can also be removed by pressing the delete key.

Reassign Link...

If a line is selected, this command will open a dialog box (Fig. 25) with a line and two pop-up menus. Select the pop-up menu for the end of the line to be moved and select the node it is to be reassigned to. Selecting either of the nodes that already terminate the line, or clicking the cancel button, will cancel this command and close the window.

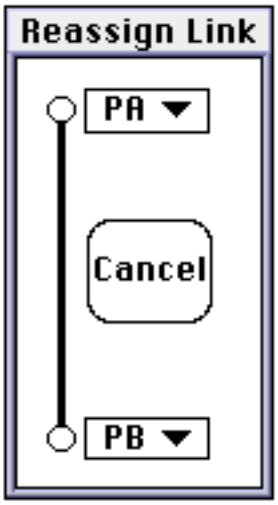


Figure 25. The Reassign Link dialog box.

Links can also be reassigned by selecting a line, holding down the shift key and selecting one of the nodes at either end of the line. This will cause a pop-up menu to appear at that node. Selecting one of the nodes

from the menu will cause that end of the link to be reassigned to the selected node.

Problems

There are three types of problems that can be selected under the problem menu. At the top of the menu is the **Set-Up Problem...** command which opens a dialog box and allows the user to define a set of organisms for a problem. The lower two sections of this menu provide tree construction problems for students which are useful for learning the mechanics of tree construction prior to addressing determining characters and states and assigning polarity. The second area of the menu contains Model problems. These problems always display particular characteristics, but the specific taxa and the arrangement of the characters will vary each time. The lowest area on the menu contains 5 problems of generally increasing complexity. Each of these problems will display similar characteristics each time it is selected, but may produce substantially different results.

Set-Up Problem

Opens the Set-up Problem dialog box (Fig. 26). Select the taxa from the scrolling list and Add them to the problem set. When complete, select Done and the selected taxa will be placed in the tree

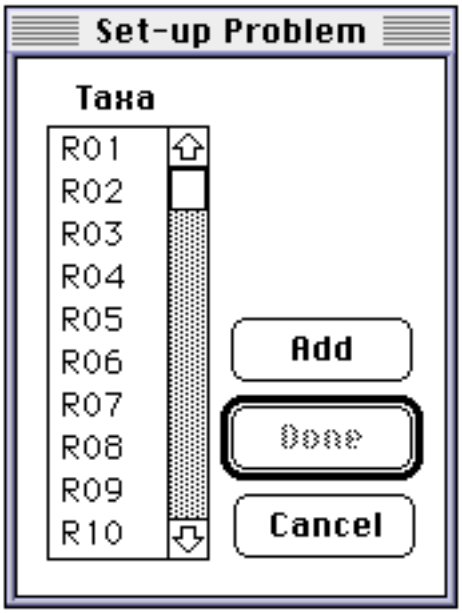
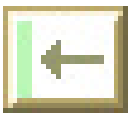


Figure 26. The Set-up Problem dialog box.

construction window. Non-contiguous selections can be made by using the Command (cloverleaf) key.

The taxa listed here represent the Caminalcules (R1-29, F1-77) the Dendrogrammaceae (R1-18), and the model problem taxa (R80-R89, F90-



F99). Other sets of taxa can be adapted for use within PI by assigning a label for each one. For recent taxa either R1-29 or R80-89 can be used. For fossil taxa less than 50 million years old, F90-F99 can be adapted (they appear in pairs at 10 million year intervals). Future versions of PI may permit modification of the taxon data base.

Each taxon that is added here will be given a line in the data matrix for coding character and state data. The software can accommodate up to 15 taxa in a problem set. It is not recommended to construct problems with more than this number of taxa.

Taxa can be added at any time during the problem solving process. Taxa added after characters have been defined will be coded with an "X" for each character. Note: It is nonsensical and not-advised to add taxa to a model or practice problem.

Model Problems

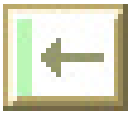
The second area defined in the Problems menu contains a list of predefined problems: Autapomorphy; Synapomorphy 1, 2, and 3; and Homoplasy 1 & 2, 3, and 4. Each of these problems, when selected, will produce a data matrix and add several taxa to the drawing field. In every case, the taxa selected and the order of the taxa and characters in the matrix will be randomized, but the form of the resultant phylogenetic tree will be the same each time. An example problem with solution and comments is provided for each model problem in Appendix B.

Practice Problems

Like the model problems, the practice problems randomly select and arrange a group of taxa and characters each time they are selected. These problems show much greater variability than the model problem. Problem 5 has two parts. After solving the first part, select Problem 5b and an additional taxon with data is added to the problem.

Windows

Each menu item simply opens the window named (or brings it to the front, if hidden or closed).



OTHER SOFTWARE FOR PHYLOGENETIC ANALYSIS

There are many sets of software tools for phylogenetic research. Three of the most important are MacClade, PHYlogenetic Inference Package (PHYLIP), and Phylogenetic Analysis Using Parsimony (PAUP). Most packages now allow some form of automated searching for trees that meet various criteria (tree length, etc.).

MacClade (Maddison & Maddison, 1989) is a well designed Macintosh software package which allows the user to evaluate the effects of swapping branches on the tree length. This is particularly useful for evaluating a series of closely related hypotheses. An early version of MacClade (2.1) appears on the BioQUEST CD-ROM and is freely distributable. More recent versions are available for purchase. All distribution is by Sinauer Associates, Sunderland, Massachusetts 01375, USA. Their phone number is: (413) 665 3722, FAX: (413) 665 7292.

PHYLIP (Felsenstein, 1993) is a large set of free programs which appear to have designed for the UNIX environment, but which have been ported to Macintosh and DOS platforms. PHYLIP's interface is not very Macintosh-like (for lack of a better term). PHYLIP is available by "anonymous ftp" over electronic networks (including the PCDOS, 386 PCDOS, 386 Windows, and Macintosh executables) from evolution.genetics.washington.edu (128.95.12.41). Contact Joe Felsenstein <joe@genetics.washington.edu> for details or start by fetching file pub/phylip/Read.Me.

PAUP (Swofford, 1991) is probably the single most widely used package by researchers. It provides a fairly Macintosh-like interface and allows a wide variety of options for searching for phylogenetic trees. Previous versions have been available from the Center for Biodiversity, Illinois Natural History Survey, 607 East Peabody Drive, Champaign, Illinois 61820, U.S.A.

SUGGESTED READINGS

For a highly readable treatment of the evolutionary issues relevant to cladistics, read Stephen Jay Gould's (1989) Wonderful Life: The Burgess Shale and the nature of history.

For a general account of cladistics, try Mark Ridley's (1986) Evolution and Classification: The reformation of cladism.

For a thorough and readable introduction to cladistic applications, read Daniel Brooks and Deborah McClennan's (1991) Phylogeny, Ecology, and Behavior.

For an in-depth treatment of the scientific revolution in cladistics try David Hull's (1988) Science as a Process.

For a thorough background on phylogenetic diagrams, try Niles Eldredge and Joel Cracraft's (1980) Phylogenetic Patterns and the Evolutionary Process: Method and theory in comparative biology.

For a thorough treatment on the philosophy of phylogenetic inference try Elliott Sober's (1988) Reconstructing the past: Parsimony, evolution and inference..

The English version of the book that started it all is Wili Hennig's (1966) Phylogenetic Systematics.

BIBLIOGRAPHY

Brooks, D. R., & McClennan, D. A. (1991). Phylogeny, Ecology, and Behavior. Chicago: University of Chicago Press.

Brooks, D. R., McLennan, D. A., Carpenter, J. M., Weller, S. G., & Coddington, J. A. (1995). Systematics, ecology and behavior. Bioscience, 45(10), 687-695.

Davis, G. M. (1995). Systematics and public health. Bioscience, 45(10), 705-714.

de Queiroz, K. (1985). The ontogenetic method for determining character polarity and its relevance to phylogenetic systematics. Systematic Zoology, 34(3), 280-299.

Duncan, T., Phillips, R. B., & W.H. Wagner, J. (1980). A comparison of branching diagrams derived by various phenetic and cladistic methods. Systematic Botany, 5(3), 264-293.

Eldredge, N., & Cracraft, J. (1980). Phylogenetic Patterns and the Evolutionary Process: Method and theory in comparative biology. New York: Columbia University Press.

Felsenstein, J. (1983) Parsimony in systematics: biological and statistical issues. Annual Review of Ecology and Systematics, 14, 313-333.

Felsenstein, J. (1993). PHYLIP: Phylogeny inference package. Distributed by the author. University of Washington.

Gould, S. J. (1980). The Panda's Thumb. New York: W. W. Norton & Company.

Gould, S. J. (1989). Wonderful Life: The Burgess Shale and the nature of history. New York: W.W. Norton and Company.

Harvey, P. H., & Pagel, M. D. (1991). The comparative method in evolutionary biology. New York: Oxford University Press.

Hennig, W. (1966). Phylogenetic Systematics. Chicago: University of Illinois Press.

Hull, D. L. (1988). Science as a Process. Chicago: University of Chicago Press.

Lauder, G. V., Huey, R. B., Monson, R. K., & Jensen, R. J. (1995). Systematics and the study of organismal form and function. Bioscience, 45(10), 696-704.

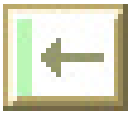
Maddison, W., & Maddison, D. (1989). MacClade: Software for cladistic analysis.

Maddison, W. P., Donoghue, M. J., & Maddison, D. R. (1984). Outgroup analysis and parsimony. Systematic Zoology, 33(1), 83-103.

Meacham, C. A., & Estabrook, G. F. (1985). Compatibility methods in systematics. Annual Review of Ecological Systematics, 16, 431-446.

Miller, D. R., & Rossman, A. Y. (1995). Systematics, biodiversity, and agriculture. Bioscience, 45(10), 680-686.

Ridley, M. (1986). Evolution and Classification: The reformation of cladism. New York: Longman Group Limited.
Savage, J. M. (1995). Systematics and the biodiversity crisis. Bioscience, 45(10), 673-679.
Simpson, B. B., & Cracraft, J. (1995). Systematics: The science of biodiversity. Bioscience, 45(10), 670-672.
Sober, E. (1988). Reconstructing the past: Parsimony, evolution and inference. Cambridge: MIT Press.
Sokal, R. R. (1983a). A phylogenetic analysis of the Caminalcules: I. The data base. Systematic Zoology, 32(2), 159-184.
Sokal, R. R. (1983b). A phylogenetic analysis of the Caminalcules: II. Estimating the true cladogram. Systematic Zoology, 32(2), 185-201.
Stevens, P. F. (1991). Character states, morphological variation, and phylogenetic analysis: A review. Systematic Botany, 16, 553-583.
Stuessy, T. F., & Crisi, J. V. (1984). Problems in the determination of evolutionary directionality of character-state change for phylogenetic reconstruction. In T. Duncan & T. F. Stuessy (Eds.), Cladistics: Perspectives on the reconstruction of evolutionary history (pp. 71-87). New York: Columbia University Press.
Swofford, D. L. (1991). PAUP: Phylogenetic Analysis Using Parsimony. Champaign: Illinois Natural History Survey



APPENDIX A -- MODEL PROBLEMS

Model Problems

The model problems were created to demonstrate fundamental concepts in phylogenetic biology. In teaching, these can be useful both for modeling problem-solving techniques and allowing students to practice recognizing these patterns in the data. Each time a problem is selected, the taxa and characters are randomly arranged, but the form of the solution will remain constant.

A solved example is provided for each of the model problems below with a description of the number of taxa, characters, solutions (both topologies (arrangements of taxa) and optimizations (character interpretations), and steps (number of unweighted character transitions). For some problems, there are also comments indicating particular features of interest.

These problems are also available separately in the "Model Problems" document.

Autapomorphy

2 taxa, 1 character, 1 solution with 1 step

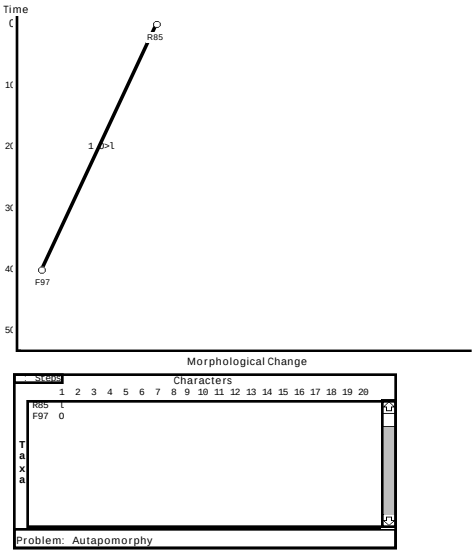


Figure 27. Autapomorphy: a phylogenetic tree representing an autapomorphy.

This problem demonstrates the essence of the phylogenetic problem: A taxon at one point in time is ancestral (F99) with respect to a character of interest (1) while a recent taxon (R84) has the character in the derived state. The problem can be resolved by establishing a link of ancestral-descendant relationship and placing a transition for the character on the link.

Synapomorphy 1

2 taxa, 1 character, 1 solution with 1 step

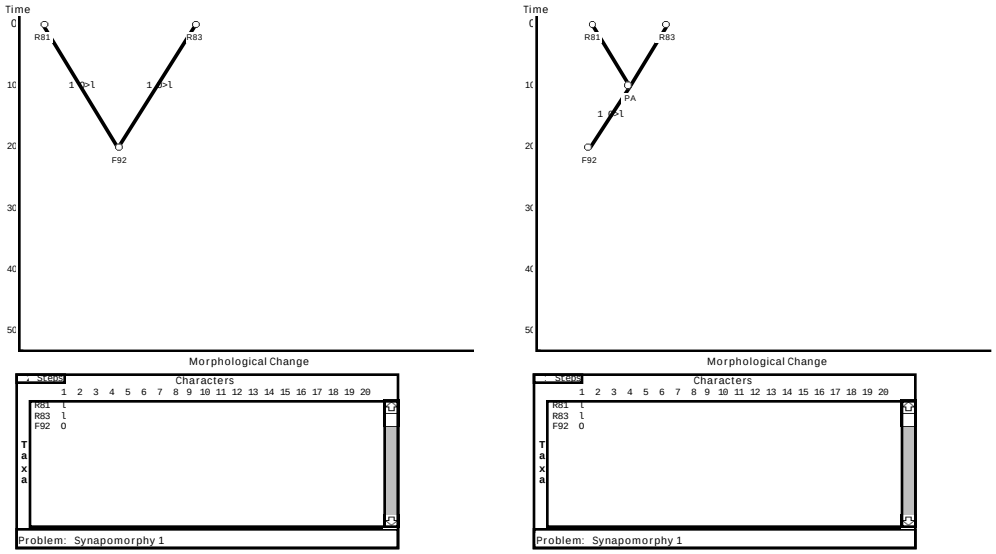


Figure 28. Synapomorphy 1: The data matrix contains a single character shared in the derived state by two recent taxa. On the right this data is represented as two autapomorphies (2 steps). More parsimonious is the phylogenetic tree on the left representing a 2 taxon synapomorphy (1 step).

This problem demonstrates the fundamental assumption of modern phylogenetic tree construction (What is sometimes called 'the auxiliary rule'). The two recent taxa share a derived character which is ancestral in the root. A common ancestor can be postulated, linked to the recent taxa and to the oldest taxon, and the transition for the character can be placed prior to the common ancestor. For classroom modeling, it is often useful to initially construct this problem as a convergence (both taxa linked directed to the ancestral taxon with the transition occurring on each link) and then to reconstruct the problem (using reassign links) to show synapomorphy. The principle of parsimony can be introduced at this point.

Synapomorphy 2

3 taxa, 2 characters, 1 solution with 2 steps

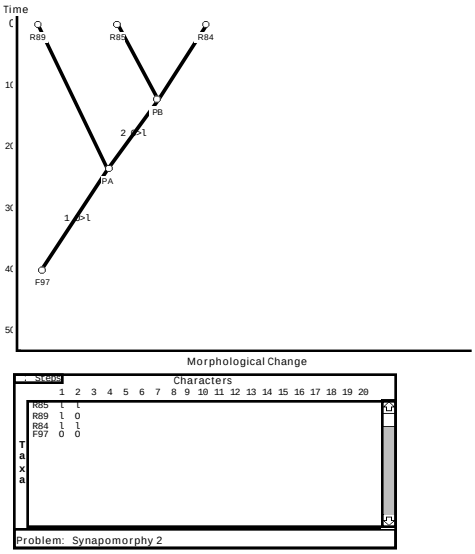


Figure 29. Synapomorphy 2: A phylogenetic tree representing a 3 taxon nested synapomorphy.

This problem illustrates nested characters (What is sometimes called 'the inclusion rule'). Characters 1 and 2 are nested because character 1's distribution is included entirely within character 2's distribution. Being nested is one way that characters can be 'consistent' or 'compatible.' Nested characters represent a stronger hypothesis than exclusive characters.

Synapomorphy 3

4 taxa, 3 characters, 1 solution with 3 steps

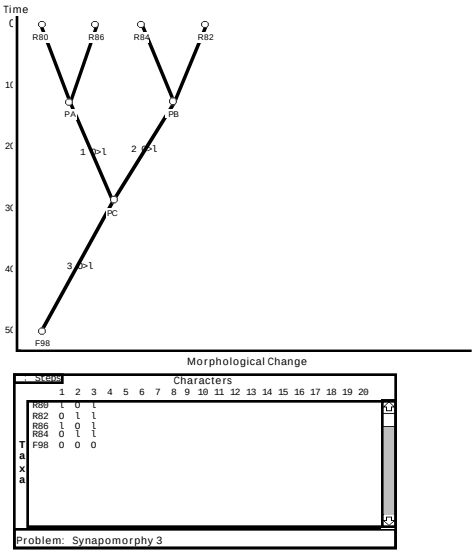


Figure 30. Synapomorphy 3: two 2-taxon synapomorphies joined by a whole ingroup synapomorphy.

This problem illustrates mutually exclusive characters (What is sometimes called 'the exclusion rule'). Characters 2 and 3 are exclusive because their distributions do not overlap. Exclusive characters, like nested characters, are 'compatible' or 'consistent' with one another.

Homoplasy 1 & 2

3 taxa, 4 characters, 1 topology and 2 optimizations, 4 steps

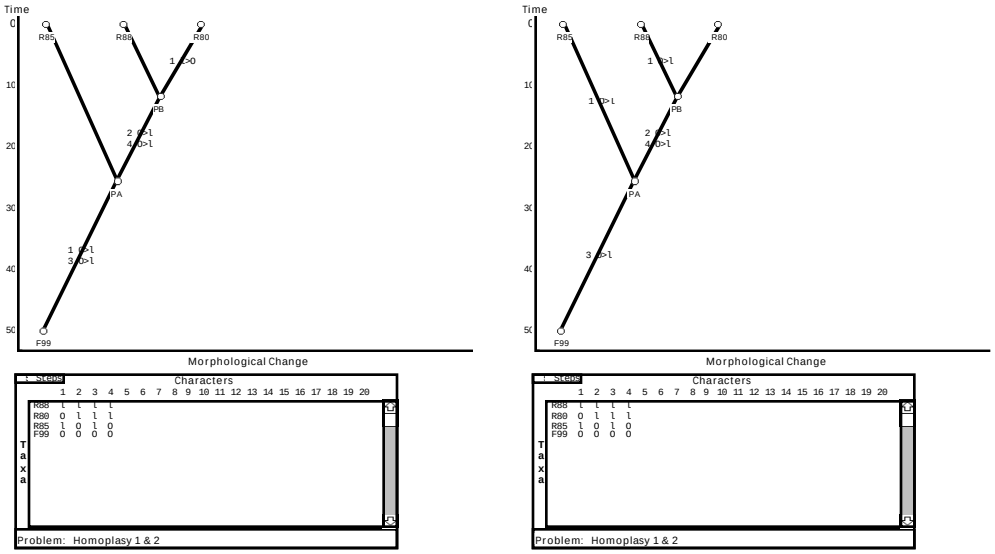


Figure 31. Homoplasy 1 &2: A nested synapomorphy problem with 2 equally parsimonious character optimizations..

This problem illustrates multiple character optimizations (convergence or reversal). One character (3) conflicts with two other characters (1,2) resulting in two different interpretations of the conflicting character. In one interpretation, the conflicting character is gained twice. In the other, it is gained once (before PA) and lost once (in R86)

Homoplasy 3

4 taxa, 4 characters, 1 solution 5 steps

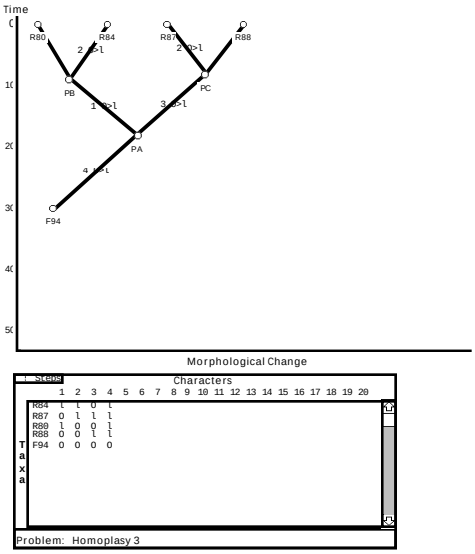


Figure 32. Homoplasy 3: A convergence problem.

This problem illustrates homoplasy with a single resolution (convergence). Two characters (1 and 3) are compatible and exclusive and nested within character 2. Character 4 conflicts with characters 1 and 3, but only one interpretation is possible in this case. Constructing this solution as a reversal, would require two reversals and, therefore, not be most parsimonious.

Homoplasy 4

3 taxa, 3 characters, 2 topologies each with 2 optimizations, 4 steps

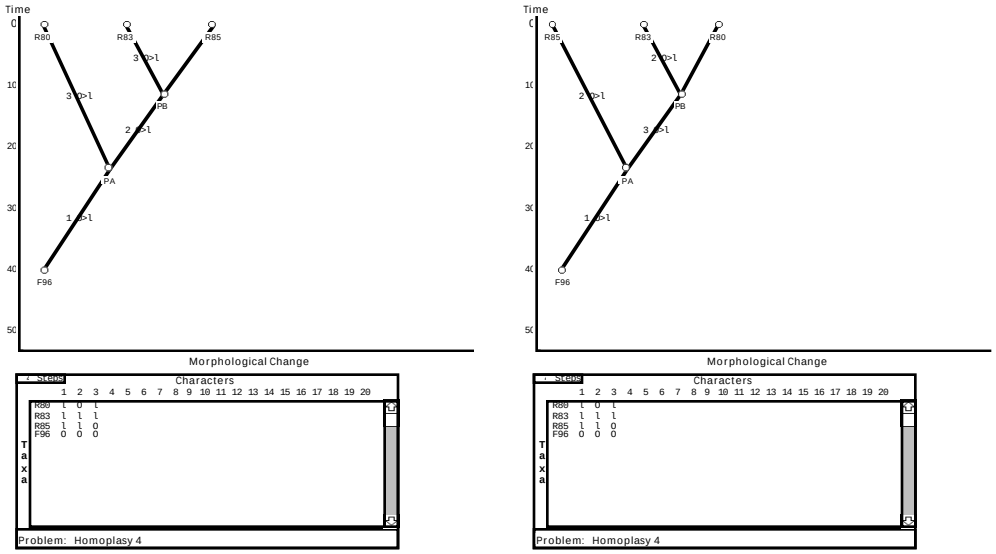


Figure 33. Homoplasy 4: A nested synapomorphy problem with 2 equally parsimonious topologies each with 2 character optimizations (not shown).

This problem illustrates multiple topologies. It is similar to Homoplasy 1 & 2, except that there are now only two characters (1 and 2) that conflict with each other. In the Homoplasy 1 & 2, having two identical characters unambiguously defines the tree's structure. In this case, either character is equally believable resulting in two arrangements of the taxa each with two character interpretations (only the convergence optimization is shown for each topology).

APPENDIX B -- INSECT WING DATA SOURCE

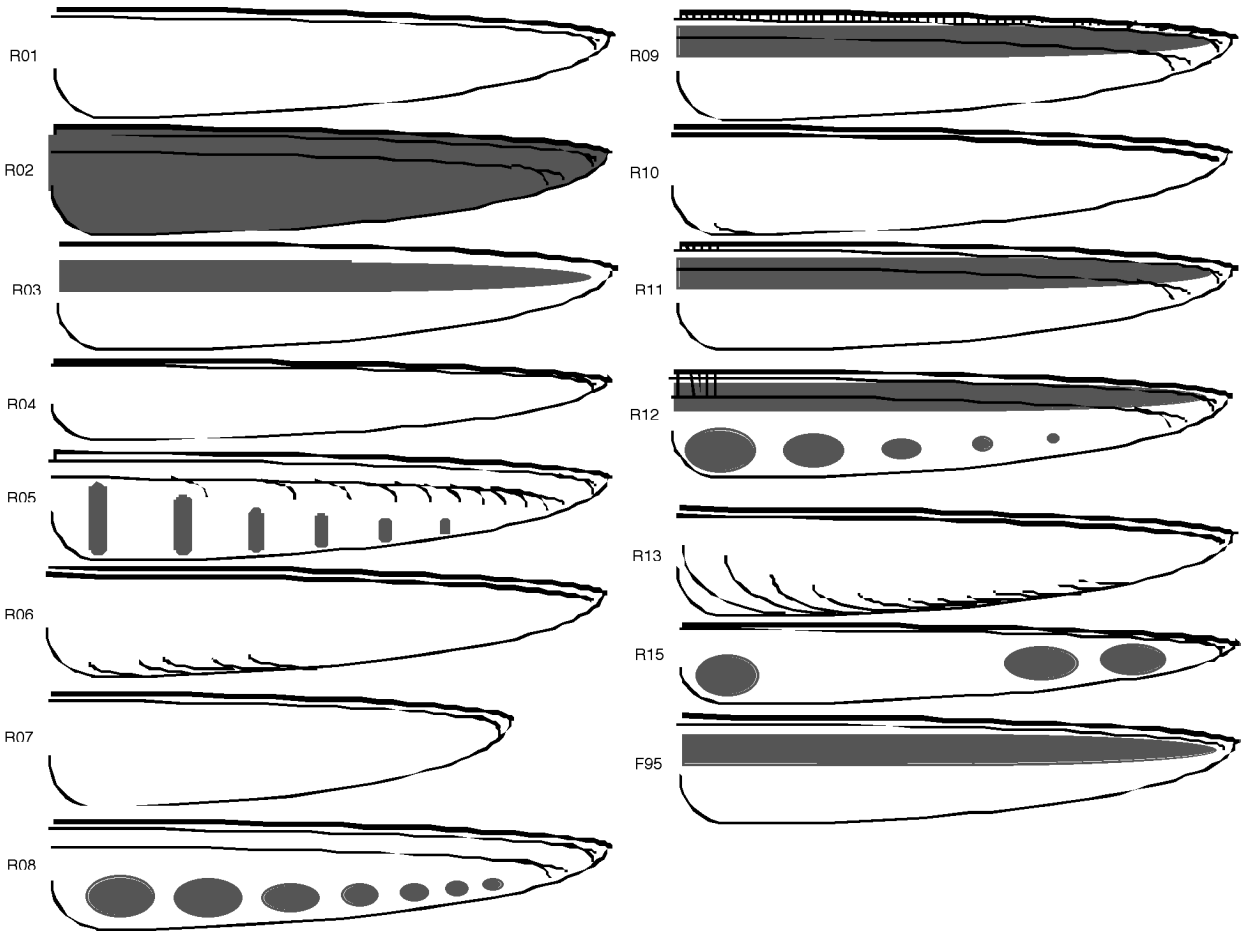


Figure 34. The complete data source from which the insect wing example is selected.

The insect wings are also available as a PICT file, "Insect Wings.pict".