AMERICAN MUSEUM ᵇ NATURAL HISTORY

Seminars on Science
Sharks and Rays: Myth and Reality

The Clompleat Cladist

A Primer of Phylogenetic Procedures

E.O. WILEY
D. SIEGEL-CAUSEY
D.R. BROOKS
V.A. FUNK

CHAPTER 1

INTRODUCTION, TERMS, AND CONCEPTS

The core concept of phylogenetic systematics is the use of derived or apomorphic characters to reconstruct common ancestry relationships and the grouping of taxa based on common ancestry. This concept, first formalized by Hennig (1950, 1966), has been slowly, and not so quietly, changing the nature of systematics. Why should we be interested in this approach? What about phylogenetic systematics is different from traditional systematics? The answer is simple: classifications that are not known to be phylogenetic are possibly artificial and are, therefore, useful only for identification and not for asking questions about evolution.

There are two other means of making statements of relationship: traditional systematics and phenetics. Traditional systematic methods employ intuition. In practical terms, intuition is character weighting. The scientist studies a group of organisms, selects the character(s) believed to be important (i.e., conservative), and delimits species and groups of species based on these characters. Disagreements usually arise when different scientists think different characters are important. It is difficult to evaluate the evolutionary significance of groups classified by intuition because we do not know why they were created or whether they represent anything real in nature. Because these groups may not be defined at all or may be defined by characters that have no evolutionary significance, such groups may be artificial.

Phenetics is an attempt to devise an empirical method for determining taxonomic relationships. In practice, phenetics is no better than traditional systematics in determining relationships because the various algorithms concentrate on reflecting the total similarity of the organisms in question. Organisms that appear to be more similar are grouped together, ignoring the results of parallel or convergent evolution and again creating possibly artificial groups.

Phylogeneticists differ from traditional systematists in that we employ empirical methods to reconstruct phylogenies and strictly evolutionary principles to form classifications rather than relying on intuition or authority. We differ from pheneticists in that our methods seek to find the genealogic relationships among the taxa we study rather than the phenetic or overall similarity relationships.

What all this means is that the groups we discover are thought to be natural, or monophyletic. Given any array of taxa, which two are more closely related to each other than either is to any other taxon? We attempt to discover the common ancestry relationships indirectly through finding evidence for common ancestry. This evidence comes in the form of shared derived characters (synapomorphies). For example, among Aves (birds), Crocodylia (alligators and crocodiles), and Squamata (lizards, snakes, and amphisbaenians), Aves and Crocodylia are thought to be more closely related because they share a number of synapomorphies thought to have originated in their common ancestor, which appeared after (later than) the common ancestor of all three taxa. This relationship is shown in the form of a phylogenetic tree, a reconstruction of the genealogic relationships.

In addition, phylogeneticists view the reconstructed tree (frequently termed a cladogram) as the classification, and when expressing it in a hierarchical scheme, we insist on maintaining monophyletic groups and sister-group relationships. The discovery of monophyletic groups is the basic quest of phylogenetics. Going to all the trouble of finding the groups and then throwing them away does not make sense to us.

Ever since the general theory of evolution gained acceptance, systematists have sought the one evolutionary history for organisms and have tried to fit that history into a hierarchical structure. We seek to reflect in our classifications the groups that we find in nature. Because phylogenetic reasoning delimits groups based on common ancestry, we can attempt to reconstruct evolutionary histories and from them develop a hierarchical ranking scheme. Phylogenetic groups are then a reflection

of the order in nature. Therefore, our classifications can be used for the study of other characters and for further investigations in biogeography, coevolution, molecular evolution, rates of evolution, ecology, etc. If you wish to use classifications to study evolution, they must reflect the genealogy of the taxa in question. Groups that are potentially artificial cannot be used in such investigations.

One of the greatest strengths of the phylogenetic system is that the method and results are transparent, meaning that decisions, whether right or wrong, are based on data that can be examined by any and all persons willing to understand the nature of the data. The phylogenetic system does not depend on some special and mysterious knowledge about organisms that only the "expert" can understand. A critic cannot claim that your idea of the phylogenetic history of a group is wrong just because he has studied the groups longer than you have. Of course, there are valid disagreements, and there is room for change and improvement. But these disagreements are data based, not opinion based. Phylogenetics, to put it crudely, is a put-up-or-shut-up scientific discipline.

This workbook presents the basics of phylogenetic systematics as we use it today. We also cite references for those interested in following some of the debates currently underway among the proponents of phylogenetic systematics. We hope that this information will stimulate you and illustrate the importance of systematics as the basis of comparative biology. When you have finished this workbook, you should be able to reread this introduction and understand what we are trying to accomplish. As an acid test, go read Hennig (1966); it's the way we got started, and it remains the classic in the field.

All new scientific ideas and analytical methods are accompanied by new sets of terms and concepts, which can be unsettling to the tyro and even more unsettling to the experienced systematist who is called upon to abandon the "traditional" meanings of terms and embrace new meanings. The basic rationale for adopting the definitions and concepts presented in this workbook is twofold. First, it is vitally important for systematics and taxonomy to be integrated into in the field of evolutionary theory. Willi Hennig's major motivation for reforming systematics and taxonomy was to bring them in line with the Darwinian Revolution, making the results obtained through phylogenetic systematics directly relevant to studies in other fields of evolutionary research. Second, it is vitally important that the terms used in an empirical field be as unambiguous as possible so that hypotheses are as clear as possible. With these rationales in mind, we offer the following definitions for the basic terms in our field. They are largely taken from Hennig (1966) or Wiley (1980, 1981a). Other more specialized terms will be introduced in other chapters.

TERMS FOR GROUPS OF ORGANISMS

1. Taxon.-A taxon is a group of organisms that is given a name. The name is a proper name. The form of many of these proper names must follow the rules set forth in one of the codes that govern the use of names. The relative hierarchical position of a taxon in a classification can be indicated in many ways. In the Linnaean system, relative rank is denoted by the use of categories. You should not confuse the rank of a taxon with its reality as a group. Aves is a taxon that includes exactly the same organisms whether it is ranked as a class, an order, or a family.

2. Natural taxon.-A natural taxon is a group of organisms that exists in nature as a result of evolution. Although there are many possible groupings of organisms, only a few groupings comprise natural taxa. In the phylogenetic system, there are two basic kinds of natural taxa: species and monophyletic groups. A species is a lineage. It is a taxon that represents the largest unit of taxic evolution and is associated with an array of processes termed speciation. A monophyletic group is a group of species that includes an ancestral species and all of its descendants (Fig. 1.1a). Members of monophyletic groups share a set of common ancestry relationships not shared with any other species placed outside the group. In other terms, a

monophyletic group is a unit of evolutionary history. Examples include Mammalia and Angiospermae.

3. Clade.-A clade is a monophyletic group, i.e., a natural taxon.

4. Ancestral taxon.-An ancestral taxon   is a species that gave rise to at least one new daughter species during speciation, either through cladogenesis or reticulate speciation. By cladogenesis we mean speciation that results in two or more branches on the phylogenetic tree where there was only one branch before. By reticulate speciation we mean the establishment of a new species through a hybridization event involving two different species. A species that emerged from cladogenesis has one ancestral species but a species emerging from reticulate speciation has two ancestral species. In the phylogenetic system, only species can be ancestral taxa. Groups of species are specifically excluded from being ancestral to other groups of species or to single species. The biological rationale for this distinction is clear; there is an array of processes termed speciation that allow for one species to give rise to another (or two species to give rise to a species of hybrid origin), but there are no known processes that allow for a genus or a family to give rise to other taxa that contain two or more species ("genusation" and "familization" are biologically unknown). Thus, each monophyletic group begins as a single species. This species is the ancestor of all subsequent members of the monophyletic group.

5. Artificial taxon.-An artificial taxon   is one that does not correspond to a unit involved in the evolutionary process or to a unit of evolutionary history. You will

Insert fig1_1

encounter two kinds of artificial groups. Paraphyletic   groups   are artificial because one or more descendants of an ancestor are excluded from the group (Fig. 1.1b). Examples include Dicotyledonae, Vermes, and Reptilia. Polyphyletic groups are artificial because the common ancestor is placed in another taxon (Fig. 1.1c). An example would be the Homeothermia, a group composed of birds and mammals. Note that the term "ancestor" is used in its logical sense, i.e., the ancestor is unknown but its inclusion or exclusion can be deduced as a logical consequence of the grouping. The important contrast is between monophyletic groups and nonmonophyletic groups. Paraphyletic groups are as artificial as polyphyletic groups. Further, it is not always possible to distinguish clearly the status of a group as either paraphyletic or polyphyletic.

6. Grade.-A grade is an artificial taxon. Grade taxa are frequently paraphyletic and sometimes polyphyletic but are supposed to represent some level of evolutionary progress, level of organization, or level of adaptation (e.g., Reptilia or Vermes).

7. Ingroup.-The ingroup   is the group actually studied by the investigator (Fig. 1.2a). That is, it is the group of interest.

8. Sister group.-A sister group   is the taxon that is genealogically most closely related to the ingroup (Fig. 1.2a). The ancestor of the ingroup cannot be its sister because the ancestor is a member of the group.

Insert fig1_2

9. Outgroup.-An outgroup   is any group used in an analysis that is not included in the taxon under study. It is used for comparative purposes, usually in arguments concerning the relative polarity of a pair (or series) of homologous characters. The most important outgroup is the sister group, and considerable phylogenetic research may be needed to find the sister group. Usually more than one outgroup is needed in an analysis. This will become apparent in Chapter 3.

Quick Quiz-Groups

Examine Fig. 1.1 and answer the following:
1.Why do we say that the group A+B+C and the group M+N are monophyletic?

2.Which taxa would have to be either included or excluded to change the paraphyletic groups into monophyletic groups?
3.Can polyphyletic groups ever contain monophyletic groups within them?
4.Where are the ancestors in these diagrams?

TERMS FOR THE RELATIONSHIPS OF TAXA

1. Relationship.-In the phylogenetic system, the term relationship refers to the genealogic or "blood" relationship that exists between parent and child or between sister and brother. In other systems, relationship can also refer to similarity, with the evolutionary implication that taxa that are more similar to each other are more closely related. This meaning is specifically excluded from the phylogenetic system.

2. Genealogy and genealogic descent.-A genealogy is a graphic representation of the descent of offspring from parents. Genealogic descent on the taxon level (i.e., between groups recognized as taxa) is based on the proposition that species give rise to daughter species through an array of mechanisms termed speciation.

3. Tree.-A tree is a branching structure and, in our sense, might contain reticulations as well as branches. A tree may be rooted (Fig. 1.2a) or unrooted (Fig. 1.2b) and is composed of several parts. A branch is a line connecting a branch point to a terminal taxon. A branch point, or node, represents a speciation event. This is true even if the taxa joined by the branch point are higher taxa such as families or phyla, because higher taxa originated as species. Branch points are sometimes represented by circles. An internode is a line connecting two speciation events and represents at least one ancestral species. (We say at least one because the statement is made relative to the species and groups we actually know about. It is always possible to find a new species or group of species that belongs to this part of the phylogeny. To make this addition, we would bisect the internode and create the possibility for an additional ancestral species.) The internode at the bottom of the tree is given the special term root. The term interval is a synonym of internode and is used in the Wagner algorithm (see Chapter 4). A neighborhood is an area of a tree relative to a particular taxon or taxa. In Fig. 1.2b, taxon B is the nearest neighbor of taxa A and C. Note that A may or may not be the sister of a monophyletic group B+C. This relationship cannot be established until the root is specified.

4. Phylogenetic tree.-A phylogenetic tree is a graphic representation of the genealogic relationships between taxa as these relationships are understood by a particular investigator. In other words, a phylogenetic tree is a hypothesis of genealogic relationships on the taxon level. Although it is possible for an investigator to actually name ancestors and associate them with specific internodes, most phylogenetic trees are common ancestry trees. Further, phylogenetic trees are hypotheses, not facts. Our ideas about the relationships among organisms change with increasing understanding.

5. Cladogram.-Cladograms are phylogenetic trees. They have specific connotations of implied ancestry and a relative time axis. Thus, a cladogram is one kind of phylogenetic tree, a common ancestry tree. In some modifications of the phylogenetic system, specifically what some have termed Transformed Cladistics, the cladogram is the basic unit of analysis and is held to be fundamentally different from a phylogenetic tree. Specifically, it is purely a depiction of the derived characters shared by taxa with no necessary connotation of common ancestry or relative time axis.

6. Venn diagram.-A Venn diagram is a graphic representation of the relationships among taxa using internested circles and ellipses. The ellipses take the place of internode connections. A typical Venn diagram is contrasted with a phylogenetic tree in Fig. 1.3.

Insert fig1_3

Quick Quiz-Relationships

Examine Fig. 1.2a and answer the following:

I . What is the sister group of the clade N?
2. What is the sister group of the clade M?
3. What is the sister group of a group composed of M+N?
4. Where is the hypothetical ancestor of the ingroup on the tree?
5. How many ancestors can a group have?
6. Draw a Venn diagram of Fig. 1.2a.

TERMS FOR CLASSIFICATIONS

1. Natural classification.-A classification containing only monophyletic groups and/or species is a natural classification. A natural classification is logically consistent with the phylogenetic relationships of the organisms classified as they are understood by the investigator constructing the classification. That is, the knowledge claims inherent in a natural classification do not conflict with any of the knowledge claims inherent in the phylogenetic tree. The protocols for determining if a classification is logically consistent with a phylogenetic tree are given in Chapter 6.

2. Artificial classification.-An artificial classification is a classification containing one or more artificial groups (i.e., one or more paraphyletic or polyphyletic groups). An artificial classification is logically inconsistent with the phylogenetic relationships of the organisms as they are understood by the investigator making the classification. That is, some of the knowledge claims inherent in the classification conflict with knowledge claims in the phylogenetic tree.

3. Arrangement.-An arrangement is a classification of a group whose phylogenetic relationships are not known because no investigator has ever attempted to reconstruct the evolutionary history of the group. The vast majority of current classifications are arrangements. A particular arrangement may turn out to be either a natural or an artificial classification. Arrangements serve as interim and completely necessary vehicles for classifying organisms until the phylogenetic relationships of these organisms can be worked out.

4. Category.-The category   of a taxon indicates its relative place in the hierarchy of the classification. The Linnaean hierarchy is the most common taxonomic hierarchy and its categories include class, order, family, genus, and species. The formation of the names of taxa that occupy certain places in the hierarchy are governed by rules contained in various codes of nomenclature. For example, animal taxa ranked at the level of the category family have names that end in -idae, whereas plant taxa ranked at this level have names that end in -aceae. It is important to remember that the rank of a taxon does not affect its status in the phylogenetic system. To the phylogeneticist, all monophyletic taxa are equally important and all paraphyletic and polyphyletic taxa are equally misleading.

Classifications and arrangements are usually presented as hierarchies of names, with relative position in the hierarchy (rank) noted by categories. However, these classifications can be portrayed as tree diagrams and as Venn diagrams. The use of these methods of presenting classifications is discussed in Chapter 6.

Quick Quiz-Classification

1. In the phylogenetic system, must the taxa be clades?
2. In the phylogenetic system, must categories be clades?
3. Which is more important, a phylum or a genus?

PROCESS TERMS

Three process terms are of particular importance in the phylogenetic system. Speciation   results in an increase in the number of species in a group. Speciation is

not a single process but an array of processes. Cladogenesis   is branching or divergent evolution and is caused by speciation. Anagenesis   is change within a species that does not involve branching. The extent to which anagenesis and cladogenesis are coupled is an interesting evolutionary question but not a question that must be settled to understand the phylogenetic system.

TERMS FOR THE   ATTRIBUTES OF SPECIMENS

1. Character.-A character   is a feature, that is, an observable part of, or attribute of, an organism.

2. Evolutionary novelty.-An evolutionary novelty   is an inherited change from a previously existing character. The novelty is the homologue of the previously existing character in an ancestor/descendant relationship. As we shall see below, novelties are apomorphies at the time they originate.

3. Homologue.-Two characters in two taxa are homologues if one of the following two conditions are met: 1) they are the same as the character that is found in the ancestor of the two taxa or 2) they are different characters that have an ancestor/descendant relationship described as preexisting/novel. The ancestral character is termed the plesiomorphic character,   and the descendant character is termed the apomorphic character.   The process of determining which of two homologues is plesiomorphic or apomorphic lies at the heart of the phylogenetic method and is termed character polarization or character argumentation.
Three (or more) characters are homologues if they meet condition 2.

4. Homoplasy.-A homoplasy is a similar character that is shared by two taxa but does not meet the criteria of homology. Every statement of homology is a hypothesis subject to testing. What you thought were homologues at the beginning of an analysis may end up to be homoplasies.

5. Transformation series.-A transformation   series (abbreviated TS in some tables and exercises) is a group of homologous characters. If the transformation series is ordered, a particular path of possible evolution is specified but not necessarily the direction that path might take. All transformation series containing only two homologous characters (the binary condition) are automatically ordered but not necessarily polarized (contrast Fig. 1.4a and Fig. 1.4b). Transformation series having more than two characters are termed multicharacter (or multistate) transformation series. If a multistate transformation series is unordered (Fig. 1.4c), several paths might be possible. Ordered transformation series are not the same as polarized transformation series (compare Figs. 1.4d and 1.4e). An unpolarized transformation series   is one in which the direction of character evolution has not been specified (Figs. 1.4a, c, d). A polarized transformation series   is one in which the relative apomorphy and plesiomorphy of characters has been determined by an appropriate criterion (Figs. 1.4b, e). It is possible for a transformation series to be both unordered and polarized. For example, we might know from outgroup comparison that 0 is the plesiomorphic state, but we might not know whether 1 gave rise to 2, or vice versa, or whether 1 and 2 arose independently from 0. Ordering and polarization of multicharacter transformation series can become very complicated, as we shall see in Chapter 3. Our use of the convention "transformation series/character" differs from that of many authors who use "character" as a synonym for "transformation series" and "character state" as a synonym for "character." We use "transformation series/character" instead of "character/character state" in our research and in this workbook for philosophical reasons. The "character/character state" convention reduces "character" to a term that does not refer to the attributes of organisms but instead to a class construct that contains the attributes of organisms,

Insert fig1_4

homologues or not. For example, dandelions do not have "color of flower" as an attribute; they have "yellow flowers." We adopt "transformation series/character"

because it explicitly avoids the construction of character classes and implicitly encourages the investigator to use characters hypothesized to be homologues of each other.

6. Character argumentation.-Character argumentation is the logical process of determining which characters in a transformation series are plesiomorphic and which are apomorphic. Character argumentation is based on a priori arguments of an "if, then" deductive nature and is based on outgroup comparison. This process is frequently termed "polarizing the characters." Polarity refers to which of the characters is plesiomorphic or apomorphic. Character argumentation will be covered in detail in Chapter 3.

7. Character optimization.-Character optimization consists of a posteriori arguments as to how particular characters should be polarized given a particular tree topology. Character optimization might seem a priori when used in a computer program, but it is not.

8. Character code and data matrix.-Phylogenetic systematists are quickly converting to computer-assisted analysis of their data. When using a computer, the investigator produces a data matrix. Usually, the columns of the matrix are transformation series and the rows are taxa. A code is the numerical name of a particular character. By convention, the code " 1 " is usually assigned to the apomorphic character and "0" to the plesiomorphic character of a transformation series if the polarity of that series is determined (=hypothesized) by an appropriate method of polarization. If the transformation series consists of more than two characters, additional numerical codes are assigned. Alternatively, the matrix might be coded using binary coding as discussed in Chapter 3. There are many ways of reflecting the code of a character when that character is placed on a tree. We will use the following convention: characters are denoted by their transformation series and their code. The designation 1-1 means "transformation series 1, character coded 1." Some basic ways of coding characters are discussed in Chapter 3.

9. Tree length.-The length of a tree is usually considered the number of evolutionary transformations needed to explain the data given a particular tree topology.

You will probably need some time to assimilate all of the definitions presented. A good strategy is to review this chapter and then go to Chapter 2, working your way through the examples. We have found that deeper understanding comes from actual work. Although we cannot pick a real group for you to work on, we have attempted the next best thing, a series of exercises designed to teach basic phylogenetic techniques that we hope will mimic real situations as closely as possible.

Quick Quiz-Characters

1. How would the transformation series in Fig. 1.4c look if it were polarized and unordered?
2. Is character "1" in Fig. 1.4e apomorphic or plesiomorphic?

CHAPTER NOTES AND REFERENCES

1. There is no substitute for reading Hennig (1966). We suggest, however, that you become familiar with most of the basics before attempting to read the 1966 text. Hennig (1965) is the most accessible original Hennig. Other classics include Brundin (1966) and Crowson (1970). An interesting analysis of Hennig's impact on systematics can be found in Dupuis (1984). A considerable portion of the history of phylogenetic thought (and indeed post-1950 systematics) can be followed in a single journal, Systematic Zoology. We highly recommend that students examine this journal.

2. Post-Hennig texts that are suitable for beginners are Eldredge and Cracraft (1980), Wiley (1981a), Ridley (1985), Schoch (1986), Ax (1987), and Sober (1988a). A more difficult text written from the point of view of the transformed cladists is Nelson and Platnick (1981).

3. A very readable review of the entire field of systematics is Ridley (1985), whose defense of phylogenetics and criticisms of traditional (evolutionary) taxonomy, phenetics, and transformed cladistics are generally on the mark.

QUICK QUIZ ANSWERS
Groups

1. They are monophyletic because no descendant of their respective common ancestor is left out of the group.
2. To make the group O+A+B monophyletic, you would have to include C. To make the group N+O+A+B+C monophyletic, you could either include M or exclude N.
3. Yes; e.g., N+A+B+C contains the monophyletic group ABC.
4. You were pretty clever if you answered this one because we haven't covered it yet. The ancestors are represented by internodes between branches. Obviously they are hypothetical because none of them are named.

Relationships

1. The ingroup, (A+B+C) is the sister group of N.
2. N plus the ingroup is the sister group of M.
3. A group composed of M and N is paraphyletic. Paraphyletic groups are artificial and thus cannot have sister groups.
4. The internode labeled "Internode."
5. A bunch, stretching back to the origin of life. But we usually refer only to the immediate common ancestor.
6. INSERT mnabc

Classification

1. Only clades (monophyletic groups and species) are permitted in the phylogenetic system. Grades are specifically rejected. As you will see in Chapter 6, this is because classifications that contain even a single grade are logically inconsistent with the phylogeny of the group containing the grade.
2. Categories are not taxa. They are designations of relative rank in a classification. As such, categories are neither clades nor grades.
3. All monophyletic taxa are equally important and interesting to the phylogeneticist.

Characters

1. INSERT 012triangle
2. Character 1 is both apomorphic and plesiomorphic. It is apomorphic relative to 0 and plesiomorphic relative to 2.