

- Likelihood models take into account the probability that characters will change over time between nodes.
- The simplest model is picked from a variety of models available.
- Bayesian analysis asks the question: what is the probability of the model (including the tree topology) given the data?
- Bayesian analysis also includes both the tree and other parameters such as branch length.
- Bayesian analysis uses the formalities of likelihood calculations and explores posterior probability space using MCMC to find the model/tree topology with the highest posterior probability.
- For phylogenetic research, branch lengths are of concern only in so far as they affect our ability to estimate genealogical relationships, so the important part of the model in both kinds of analyses is the tree topology.

8

PHYLOGENETIC CLASSIFICATION

Academic taxonomy deals with classes; it merely arranges according to similarities; while natural taxonomy arranges according to kinships determined by generation.

—Kant, 1775; quoted in Dobzhansky (1962:93)

[A]ll true classification is genealogical.

—Charles Darwin (1859:420)

Classifications are systems of names organized to show relationships among the entities named. The names derive their meaning from the intent of the persons who are trying to communicate. Biological classifications are used to convey ideas of the relationships among organisms. Technically, biological classifications are not classifications at all because taxa are not classes. Thus, Griffiths (1974) prefers *systematization* and de Queiroz (1988) has suggested that confusion over classification as opposed to systematization inhibited the spread of phylogenetic classifications between the time of Darwin and Hennig. However correct *systematization* might be, the term has not caught on, and we may be stuck with an inappropriate term (*classification*) based on common usage. Still, we have no problem with the term *phylogenetic systematization*, although we will refer to it as *phylogenetic classification*.

Phylogenetic classifications are biological classifications that meet the minimum criteria of being a system of names that imply relationships that are logically consistent with the phylogenetic tree the classification references. Differences between phylogenetic classifications of the same organisms may come from two sources. First, phylogenetic classifications may differ because they adopt different conventions for showing relationships, e.g., a classification that names each branch as compared to

Phylogenetics: The Theory of Phylogenetic Systematics, Second Edition.
E. O. Wiley and Bruce S. Lieberman.
© 2011 Wiley-Blackwell. Published 2011 by John Wiley & Sons, Inc.

a classification that uses a listing convention, as described below. Or they may differ because one uses Linnean conventions and another uses numerical prefixes or PhyloCode conventions. Second, they may differ because the reference phylogeny is different. In the first case, we can think of the classifications as different ways of communicating about the same idea. In the second case, there are disagreements about the underlying relationships. Differences of the second kind are biologically important as they denote differences in the empirical data or the interpretation of the empirical data. Differences of the first kind are a matter of conventions adopted.

Classifications that include what we now recognize as monophyletic groups have been around long before Linnaeus, never mind Darwin. This indicates that the pattern of evolutionary descent in some groups is clear enough that their existence was recognized both by those who believed that natural order was divinely composed and by those who recognized that the order is the result of natural processes (e.g., Louis Agassiz and Charles Darwin, respectively). The distinctive nature of groups such as Vertebrata or Aves makes them stand out as natural individuals regardless of the methods of taxonomists (Patterson, 1977). However, it was left to Hennig (e.g., 1966) to codify the critical distinction between monophyletic and paraphyletic groups and thus clearly distinguish between artificial groups (those poly- and paraphyletic) and natural groups (those monophyletic).

In this chapter, we will discuss the general nature of classification, including some nonbiological examples. We will then discuss ways of making phylogenetic classifications and discuss the annotated Linnean Hierarchy (Wiley, 1979c, 1981a). The merits of this system are discussed and examples given. We will then discuss alternate ways of classifying phylogenetically, including such issues of classifications without rank, numerical prefix schemes, and the PhyloCode. We will end with our rationale for preferring phylogenetic classifications over classifications that include paraphyletic groups.

CLASSIFICATIONS: SOME GENERAL TYPES

The process of classifying is the activity of grouping entities or phenomena and giving names to the resulting groups. The placing of some things into one group to the exclusion of other things implies that the members or parts of the group share some relationships not shared with things outside the group. There are many ways to parse out the classification of classifications. One could recognize a basic dichotomy between hierarchical and nonhierarchical classifications. Or one could divide classifications into natural classifications and artificial classifications. For purposes of discussion, we will distinguish between three types of classifications: (1) those involving natural kinds, (2) those involving historical groups and individuals, and (3) artificial, convenience classifications. This discussion builds on the discussions in Chapter 3 on the nature of supraspecific taxa and in Chapter 5 where we discussed the relationship between characters and groups.

Classification of Natural Kinds

Natural kinds are formed when the entities classified have an indirect a-historical relationship. Entities are members of a kind by virtue of having the property of the

kind and further that property is a property that functions significantly in some theory of the world thought to be valid. The most obvious example is the semihierarchical classification of the Periodic Table. Each kind (Hydrogen, Helium, Lithium, etc.) has properties that are necessary and sufficient for kind membership such that entities can be placed in the kind. In this case, the property is atomic number. Certain kinds can be hierarchically grouped based on consequences of the property. For example, the group of inert kinds (Helium, Neon, etc.) have properties deriving from their atomic number. Entities that are members of these kinds have orbitals that are filled and thus do not participate in chemical reactions under normal, Earth-like, conditions.

Astronomers classify stars based on the size, luminosity, and temperature into such groupings as main sequence, blue giants, and red dwarfs. The pattern on which this classification is derived is the Hertzsprung-Russell (H-R) diagram shown in introductory astronomy classes (Fig. 8.1a). This classification is used for a variety of purposes as a predictive tool. The place of a star on the H-R diagram can be used to predict the age and future ontogeny of the star, within certain limits. Blue giants are all relatively young stars (100 million years old or less) because they are so massive that they burn fuel quickly. Red dwarfs have a life expectancy of many billions of years, because they burn their fuel slowly. The shape of the H-R diagram, when applied locally, has predictive properties. For example, the H-R diagram of stars in an old globular cluster looks much different from Fig. 8.1b due to an excess of giants and a low number of stars on the main sequence. The age of the cluster can be inferred by where along the main sequence the pattern is disrupted.

Classifications of natural kinds are rarely hierarchical and even the Periodic Table is only partly hierarchical, extending only one to a few levels. What all such classifications share in common is that the properties are not shared historically. Instead, they are properties gained by what an evolutionary biologist would term convergence. It is the power of such classifications that they organize and explain convergence in reference to laws and processes thought to be valid.

Historical Classifications (Systematizations)

Historical classifications are based upon inferred historical connections between the entities classified. The properties, as mentioned in Chapter 5, have an indirect historical relationship, but the entities have a direct historical relationship, that is, they form ancestor-descendant relationships. Historical classifications form deep hierarchies because the entities are replicators or they are made up of replicators; thus the entities formed have part-whole relationships. The replication need not be biological. For example, our present continents are "descended" from Pangea; North America is a historical part of Pangea, as is Africa. Phylogenetic classifications of organisms are a type of historical classification. *Homo* and *Rana* are parts of Tetrapoda, and Tetrapoda is part of Vertebrata. *Homo sapiens* is a replicator, whereas Tetrapoda is not.

The properties of entities in historical classifications are time-bound, a quality not found in the properties of natural kinds. The class of properties that systematists can study are what we term *homologies* (Chapter 5). Each grouping based on one to many synapomorphies is a hypothesis that the entities comprise a monophyletic group. A particular monophyletic group (Aves) is an entity (or more formally,

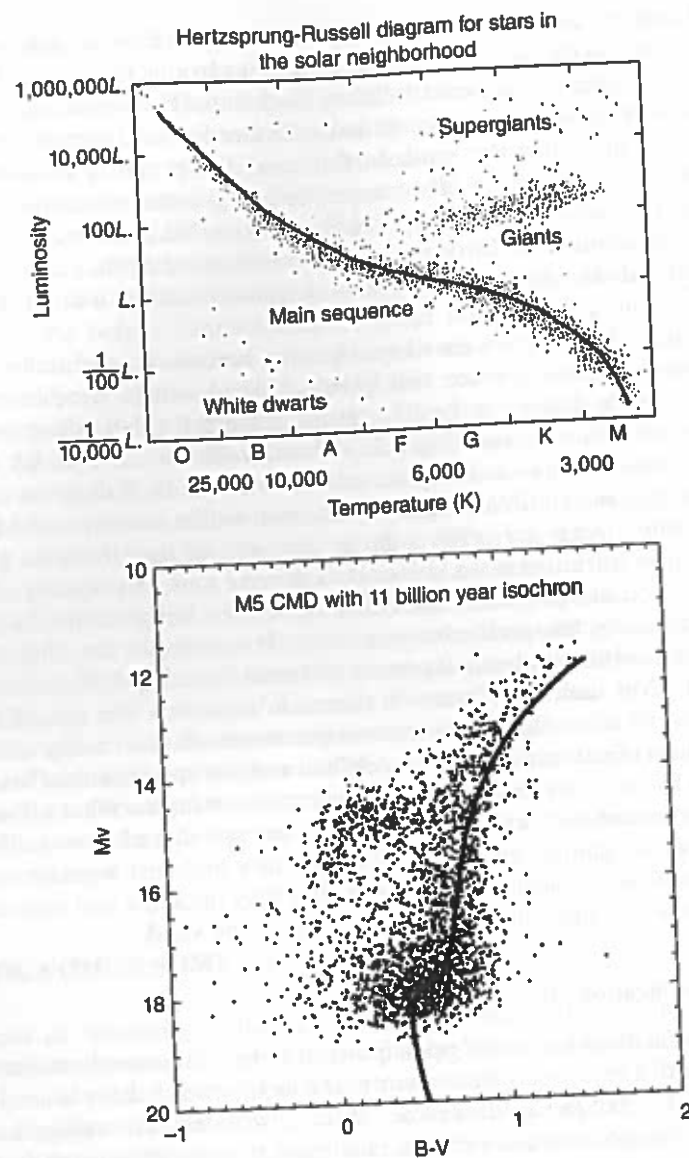


Figure 8.1. Examples of nonhierarchical classification, using Hertzsprung-Russell Diagrams to identify the nature of star clusters. (a) An H-R diagram of stars in the Pleiades open cluster (M45). (b) The H-R diagram of the globular cluster (or dwarf galaxy) 47 Tucanae. The age and classification of different clusters (open and globular) are determined by the pattern of distribution of their constituent stellar population. M45 is a relatively young open cluster with stellar distributions similar to a random selection of galactic stars and thus follows the "Main Sequence." The 47 Tucanae has an old stellar population that is atypical of the galactic population as a whole with many stars off the "main sequence." B-V is the difference between apparent blue (B) and visual (V) magnitude and cooler red stars are to the right of each graph. The M_v is the absolute visual magnitude of stars based on their apparent magnitude and the distance modulus of the cluster. Use of original data from the VizieR catalog service, Centre de Donnée astronomiques de Strasbourg, is gratefully acknowledged.

hypothesized to be an entity). It is important to recognize that although individual monophyletic groups are not natural kinds, the monophyletic group, in general, is a natural kind within evolutionary theory. Moreover, discovery of individual cases of monophyletic groups is confirmation that an important part of evolutionary theory (speciation) is valid.

Convenience Classifications

Convenience classifications are similar to classification of natural kinds in that a hierarchy, if it exists, does not extend to all entities covered by the classification. For example, the kind "felon" might include the kinds "murderer" and "arsonist" but not the kinds "speeder" or "adulterer." Convenience groups have properties that are indirectly a-historical, and their justification does not depend on invoking natural processes. There are many useful convenience classifications, including such classifications as the Dewey Decimal and Library of Congress systems of classifying books.

BIOLOGICAL CLASSIFICATIONS

Biological classifications may fall into any of the three kinds listed above and may be hierarchical or nonhierarchical, and they may group entities into kinds or historical groups (or even convenience groups). An example of a largely nonhierarchical classification of kinds would be energy-flow classifications with kinds such as primary producers that might comprise both photosynthetic and chemosynthetic organisms and with primary and secondary consumers. These kinds relate directly to process theories about movement of energy through ecosystems. We do not expect the entities that are members of each kind to form monophyletic groups that are associated with synapomorphies; both pitcher plants and lions are secondary consumers, and these character properties did not arise via common descent of an ancestor that was a secondary consumer. As mentioned above, evolutionary theory also has its natural kinds. For example, monophyletic group or species, but also Mendelian population. However, systematists are usually interested in fully nested historical classifications, specifically classifications of the natural hierarchy that comprises the tree of life.

Constituents and Grouping in Phylogenetic Classifications

The constituents of phylogenetic classifications are taxa: species and monophyletic taxa. A taxon is a group of organisms, and taxon names are proper names. As there are different views on proper names, we will discuss their nature in a section below. For now, we are worried about taxa, not their names. There are many possible groupings of organisms, but only those grouping of species that result in monophyletic taxa *sensu* Hennig (1966) are recognized as natural in the phylogenetic system. We claim an even deeper significance: monophyletic groups are the only natural taxonomic groups of species in evolutionary biology. The relationship of a constituent to the group is a part-whole relationship. Obvious nonmonophyletic groups of species are dismembered and allocated to monophyletic groups if a phylogeny is available. Many taxa are not associated with a phylogeny at all, and these groups

serve as placeholders in general classifications until such time as they are subject to phylogenetic investigation.

As Hennig (1966) stated, the actual classification of a particular phylogenetic hypothesis is a relatively straightforward procedure accomplished by applying whatever conventions the investigator wishes to adopt. Any classification that is logically consistent with the hypothesized phylogeny is a phylogenetic classification (Wiley, 1981b). The extent to which the classification accurately reflects the topology of the phylogeny is the extent to which the classification informs the community as to the evolutionary/genealogical relationships of the organisms classified. If two phylogeneticists construct different classifications of the same organisms and agree upon the phylogenetic hypothesis, then the difference lies in the conventions adopted. For example, one investigator might construct a completely subordinated classification (McKenna, 1975) while another might name only terminal taxa and use a listing convention to show relationships (Nelson, 1974a). Both classifications are logically consistent and fully informative of the tree, and there are only two rules for classification to be termed *phylogenetic*:

1. Taxa classified without qualification are monophyletic groups or species (Hennig's Criterion; Hennig, 1966).
2. The classification must be logically consistent with the phylogeny, and the conventions adopted must reveal the genealogical relationships among the groups and species classified (Hull's Criterion; Hull, 1964).

The traditional conventions for classifying taxa are embodied in the Linnean system of nomenclature, and we shall use Linnean nomenclature. The advantages and disadvantages of alternative systems are discussed after introducing the Annotated Linnean Hierarchy.

THE LINNEAN HIERARCHY

The Linnean Hierarchy is one of several conventions for classifying phylogenies, and scientists who study each major group of organisms (plants, animals, prokaryotes) have developed rules for naming and the use of names for their organisms. (We will review these rules in Chapter 11.) Species names are formed in two parts, using a genus name and a species epitaph. Taxa of higher rank (genus and above) receive a single name. (More elaborate names are available for taxa recognized below the species level.) The Linnean system expresses the relative position of a taxon within the hierarchy by using a set of tags, categories, that denote relative subordination of taxa relative to other taxa and amends the root of the name with a suffix that is particular to the rank assigned for certain categorical rank levels. Within a single clade, taxa of high rank are hypothesized to have originated earlier than taxa of low rank. Occasionally a taxon of relatively high rank will contain a single species. In this case, the higher taxon is redundant or monotypic (containing the same species as the taxon included within it; Buck and Hull, 1966). Such a taxon functions purely to denote relative age ("age of origin," Hennig, 1966:162) and position of the species or clade in the hierarchy. Two features of the Linnean system must be understood by systematists:

1. Higher rank categories are not comparable between clades. Put simply, a family of frogs does not have the same biological characteristics (e.g., time of origin, degree of distinctness, etc.) as a family of tulips. Rank is relative within clades not absolute between clades. Thus, rank has no particular biological meaning. See Forey et al. (2004) for a particularly good discussion of this point.
2. Named species are potential units of process and as potential units of process they may be compared across clades to study the general characteristics of speciation.

The Linnean system has three major disadvantages:

1. Rank categories and the relative position of each rank to the others must be memorized.
2. Shifting ranks cause changes in the suffix of certain group names that are formed with roots and suffixes.
3. A great number of categorical ranks and name endings would be needed in order to completely name and rank every clade of organisms.

These disadvantages may be a major motivation for seeking different ways of expressing hierarchical position within the Linnean system, and some have been cited as a reason for abandoning the system altogether, as we shall see.

Definition of Linnean Higher Categories

Five higher category ranks are commonly used in the various Codes of Nomenclature: genus, family, order, class, and phylum/division. We shall briefly touch on the fine points of nomenclature in Chapter 11. For now, we wish to draw distinctions between old and new concepts of the categories themselves.

Phylogeneticists take a different attitude toward higher categories than such workers as Mayr (1969:92) who defined the genus as a taxonomic category separated from other genera by a decided "gap." Wiley (1979c, 1981a) noted that phylogeneticists rejected such gaps, and by 1991 Mayr and Ashlock (1991:135) agreed, using a completely comparative and pragmatic definition based on a deeper understanding of the difference between the nature of a taxon and its rank. We take the opportunity to further develop definitions based on Wiley (1979c, 1981a):

1. *Category*. A tag of convenience that denotes relative subordination of a taxon within a particular clade.
2. *Species Category*. A category below genus. Names of species are formed by the name of the taxon ranked at the level of genus plus a species epitaph and name formation follows the appropriate code.
3. *Genus Category*. A category between species and family. Name formation follows the appropriate code.
4. *Family Category*. A category between genus and order. Names of taxa assigned at this level in the hierarchy are uninominal and have endings that are set by the various codes.

5. *Order and Class Categories.* Categories between family and phylum/division. Names formed for taxa at this level in the hierarchy are uninominal, and the codes differ on how such names are treated (e.g., whether they are subject to priority or whether endings are uniform).
6. *Phylum/Division Category.* A category between class and kingdom. Names formed at this level in the hierarchy are uninominal, and the codes differ on how names are treated.
7. *Kingdom Category.* The highest category normally treated by the codes. Again, names formed at this level in the hierarchy differ among the codes.

There are additional categories formally recognized by various codes (e.g., tribe) and subdivisions of categories (e.g., subfamily), and there are also informal categories not recognized by the codes. For example, the informal category "species group" is frequently encountered as an informal category to group species within genera without introducing a formal category such as subgenus.

Conventions for Annotated Linnean Classifications

Regardless of the eventual fate of the Linnean system of nomenclature, international agreements in place at the present govern the names of taxa that use this system. Further, major repositories for genetic information, such as Genbank, use Linnean names, and there is no reason to think that Linnean nomenclature will disappear any time soon. Thus, it is worth dealing with. We will first discuss a series of conventions designed to minimize the use of rank categories. We will then suggest some modern ways to integrate Linnean nomenclature with Web-based technology to further extend its utility. Various authors have found parts of the conventions listed below useful (e.g., Judd et al., 2008).

Convention 1. The Linnean Hierarchy will be used, with certain other conventions, to classify organisms.

Convention 2. Minimum taxonomic decisions will be made whenever possible to construct a classification or to modify an existing classification. This will be accomplished in two ways. First, no empty or redundant categorical ranks and associated taxon names will be used unless they are needed to show the sister group relationships of a small clade or single species relative to its sister. Second, the ranks of well-known clades will be retained whenever possible.

These conventions simply declare that we shall use Linnean nomenclature and that the classification will be minimally redundant and maximally informative (following Farris, 1976). Note that the wording of Convention 2 differs from Wiley (1979c, 1981a) who advocated that redundant names be restricted to "mandatory" categories. No such restriction need be placed on the Annotated System because there are no mandatory categories. (Not even genus and species are mandatory as a fragment of a fossil can be classified to family without having to be classified to genus or species.) An example of the conventions is provided by gar classification (Wiley, 1976).

Division Ginglymodi
Family Lepisosteidae
Genus *Atractosteus*
Genus *Lepisosteus*

The family contains only two genera. There is no need to assign each genus to a separate subfamily as the included taxon and diagnosis of each subfamily are redundant. Likewise, assigning both genera to a single subfamily would render the subfamily redundant relative to the family. However, the classification does contain one redundancy as Lepisosteidae is redundant relative to Ginglymodi. The redundant rank and name does serve a purpose: the sister group of gars (as presently understood) is the taxon division Halecomorphi, a clade that includes one living and many extinct species of fishes. Use of the redundant clade named Ginglymodi at the rank division is purely a device used to permit taxa to be ranked below the rank of division within the sister group of gars. Thus, although it is redundant relative to gars, it is useful, needed, and not redundant relative to bony fish classification writ large.

Convention 3. Asymmetric trees containing terminal taxa may be placed at the same hierarchical rank and listed in order of their branching sequence (Nelson, 1972a, 1974a).

This is the sequencing convention of Nelson. It is used to preserve the phylogenetic information concerning sister group relationship without the need to introduce additional hierarchical levels in the classification. For example, Schuh's (1976) classification of the hemipteran family Miridae exactly reflects the branching order of his phylogenetic hypothesis (Fig. 8.2).

Family Miridae
Subfamily Isometopinae
Subfamily Psallopinae
Subfamily Phyllinae
Subfamily Cylapinae
Subfamily Mirinae
Subfamily Bryocorinae

Without this convention, Schuh (1976) would have required 14 taxon names and appropriate categorical ranks rather than the seven names and two categorical ranks used.

Convention 4. Entirely fossil clades should be noted as such.

This is a revision of the original Convention 3 of Wiley (1979c), which called for the use of the rankless category "plesion" and the sequencing of extinct clades. We discuss the reasons for and the history behind the revision more fully below, but briefly here it derives from the fact that it was thought that fossil taxa were inherently less informative, when it came to phylogenetic matters, than extant taxa. Moreover, others were concerned that newly discovered fossil taxa might overturn or change existing classifications, necessitating the establishment of several new

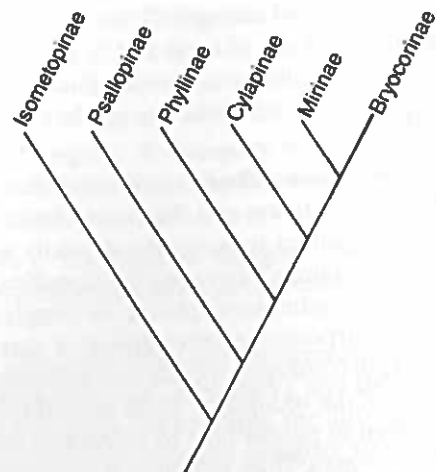


Figure 8.2. Schuh's (1976) hypothesis of relationships among subfamilies of the hemipteran family Miridae (after Schuh, 1976, from Wiley, 1979c).

taxonomic ranks. (This begs the question of whether something similar could happen with the discovery of a new extant organism.) The plesion convention can certainly still be used, but it is potentially problematic because it implies that an organism should be classified differently simply because it went extinct. Given that fossil taxa are just as much a part of the tree of life as extant taxa, and further, they are also natural, historical entities, we would argue that they should not be treated any differently. Although the plesion concept still is occasionally used, it is more common to rank fossil clades part and parcel with extant clades, although they can be set apart with the use of a symbol such as the dagger (†). A reasonable way to exactly reflect the relationships of a fossil clade to its recent relatives would be to use this convention in concert with the listing convention to preserve as many hierarchical ranks as possible.

As mentioned above, the placement of fossil groups vexed some early phylogeneticists. Convention 4 acknowledges how the community has dealt with the problem. It emerged from consideration of three proposals debated between 1966 and around 1980.

Proposal 1. Fossils should be classified separately from clades with recent organisms (Crowson, 1970) and divided along time lines (e.g., Hennig, 1966). This proposal creates a tidy classification of clades with living constituents and takes care of the placement of fossil ancestors, but there are problems. Patterson and Rosen (1977) point out that (1) there is a decrease in information content compared to a combined fossil-recent classification and (2) there would be a needless proliferation of names to accommodate relatively few fossil taxa. In addition, they pointed to some illogical features of such a scheme when the same taxon's range transcends the time boundaries assigned to different classifications. Wiley (1979c) pointed to other problems: (1) horizontal divisions in time are inherently arbitrary and likely to be clade specific, (2) the system produced paraphyletic grades, and (3) the same higher taxon

would occupy one hierarchical rank at one time interval and another rank at a later time period. Such objections also argue against Lovtrup's (1977) axiomizations concerning fossil classification, which were based on accepting Crowson's (1970) proposal.

Proposal 2. Fossil and recent organisms should be classified together and treated the same. This was McKenna's (1975) proposal (within the phylogenetic discussions of the time). It seems to be the proposal that has won out, but it does require that the ranks to which certain taxa are assigned in the hierarchy be radically adjusted from time to time. For example, McKenna (1975) demonstrated the utility of his scheme with a classification of Mammalia, but assigned the rank of class to the clade. In contrast, Nelson (1969) assigned ranks of suborder to Aves and Mammalia, which is more in line with his overall classification of Vertebrata, but this proposal was not well received among ornithologists nor mammalogists.

Proposal 3. Fossil and recent taxa should be classified together, but fossils should be treated differently (Hennig, 1966; Nelson, 1972a, 1974a; Griffiths, 1974; Patterson and Rosen, 1977). There were several proposals for how to accomplish this kind of classification. Hennig (1966) suggested the concept of the stem group (*stammgruppe*). *Archaeopteryx* and other fossil birds basal to recent ratites would be allocated to the avian stem group. The problem with this solution is that it encourages paraphyletic groups, just the kind of groups we wish to avoid. It is also not very useful for clades where fossil members of the clade are more diverse than living members (Patterson and Rosen, 1977). Nelson (1972a) simply suggested that fossils be tagged with a dagger and inserted within the classification at the appropriate level after categorical ranks are assigned to the living clades. Patterson and Rosen (1977) suggested that the rankless category "plesion" be assigned to fossil clades and that they be sequenced with their living relatives. The plesion concept can certainly be used, but seems out of favor at this time relative to the dagger convention of Nelson (1972a), which has long been used in traditional classifications to denote entirely fossil groups.

Convention 5. Monophyletic groups that form polytomies are given appropriate equivalent rank and placed *sedis mutabilis* at the level of the hierarchy at which their relationships to other taxa are known (Wiley, 1979c).

The *sedis mutabilis* convention is necessary in order to set apart lists of taxa that form ascending dichotomies from list of taxa that form polytomies. For example, the tree in Fig. 8.3 would be classified in the following manner using both the sequencing convention and the *sedis mutabilis* convention.

Family XYZidae
 Genus X, *sedis mutabilis*
 Genus Y, *sedis mutabilis*
 Genus Z, *sedis mutabilis*
 Z aus
 Z bus
 Z cus
 Z dus

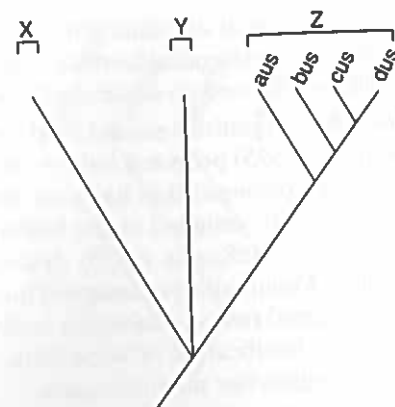


Figure 8.3. Phylogenetic relationships among some members of the hypothetical family XYZ (from Wiley, 1979c).

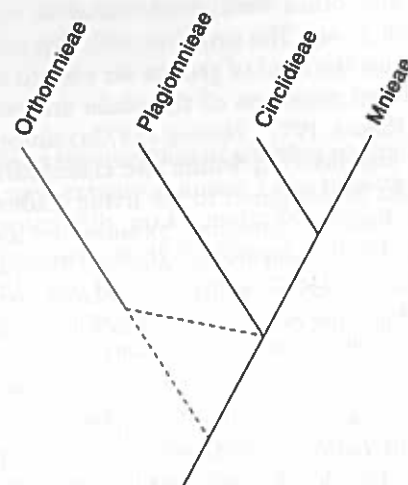


Figure 8.4. Koponen's (1968) hypothesis of relationships of mosses of the family Mniaceae.

Convention 6. Monophyletic taxa of uncertain relationships will be placed in the hierarchy *incertae sedis* at the level and ranks at which their relationships are best understood.

Although some workers have restricted *incertae sedis* only to fossil taxa (e.g., Nelson, 1972a, 1973a; Patterson and Rosen, 1977), this restricted use seems arbitrary and we prefer to follow McKenna (1975). The convention is used in both phylogenetic and traditional classification to denote ambivalence relative to the classification of a taxon of low rank relative to one or more taxa of higher ranks. For example, Koponen (1968) analyzed the bryophyte family Mniaceae and found that three of the four traditional tribes could be parsed phylogenetically into an ascending hierarchy (Fig. 8.4). One tribe, however, could not be placed as it lacked the synapomorphies uniting it with the other tribes. Using this convention, Wiley (1979c)

suggested the following classification, which uses both the sequencing convention and the *incertae sedis* convention.

- Family Mniaceae
 - Mniaceae *incertae sedis*: tribe Orthomniaceae
 - Tribe Plagiomniaceae
 - Tribe Cinclidiaceae
 - Tribe Mniaceae

Convention 7. A group whose status as monophyletic is unknown or suspect may be included in a phylogenetic classification if its status is clearly indicated by placing the name in shutter quotes to indicate that all included taxa are actually *incertae sedis* at the level of the hierarchy at which the taxon is classified. Such a group will not be accorded a formal rank.

This convention was first used by Patterson and Rosen (1977) to indicate the status of certain Mesozoic fish groups was either unknown (monophyly not demonstrated) or perhaps para- or polyphyletic. Specifically, the "Semionotidae" was considered a collection of diverse taxa that fits somewhere between the gars and the teleost fishes in actinopterygian evolution. Some might be more closely related to bowfins, others to teleosts and others spread out in the phylogeny between gars and teleosts. Part of Patterson and Rosen (1977) illustrates this use.

- Infraclass Neopterygii (higher bony fishes)
 - Division Ginglymodi (garfishes)
 - Division Halecostomi
 - Halecostomi *incertae sedis*: "Semionotidae"
 - Subdivision Halecomorphi (bowfins and relatives)
 - Subdivision Teleostei (teleosts)

Ancestors in Phylogenetic Classification The conventions presented above all deal with the classification of recent and fossil taxa while minimizing the use of rank categories. Exceptions include the placement of hybrid species and the placement of ancestral species (if known). Because part of the solution to hybrid species deals with how to place ancestral species, we shall take up ancestors first.

As mentioned in Chapter 4, ancestral species, if present in the analysis, should form polytomies with two or more descendant clades because ancestors have the synapomorphies of the group but none of the synapomorphies of descendant clades or descendant species. Thus the issue of whether one has an ancestor that needs classification depends on accepting a polytomy as a hard or true polytomy and not simply the result of missing data, a soft polytomy. A hard polytomy might be definitively recognized if repeated character analysis fails to resolve the polytomy. We would also assume that other conditions are met, such as the fact that we have biogeographic and stratigraphic information amenable to the hypothesis that the species involved has characteristics we might expect from an ancestral species. Given that there must be many ancestors awaiting discovery (and perhaps many

have been discovered but were not called ancestors), no phylogenetic classification philosophy would be complete if phylogenetic classifications were incapable of correctly classifying ancestral species, regardless of the difficulties of that enterprise. Naturally, the issue of classifying "ancestral groups" does not occur in phylogenetics because such "groups" would necessarily be paraphyletic and thus discarded in favor of monophyletic groups. Whether or not we will ever have enough information to actually identify an ancestral species is another point of debate (Hennig, 1966; Brundin, 1966; Crowson, 1970; Griffiths, 1974); traditionally phylogeneticists have rejected the idea that an actual ancestral species could be identified as a stem species using the tools now at our disposal and this concern was clearly articulated by Hennig (1966:72):

Naturally, in practice this [ancestor recognition] meets with basically insurmountable difficulties because it is scarcely ever possible to determine with certainty whether one (and in this case which) of the known species of *Archaeopteryx* (to continue our example) is the stem species of all other known species of Aves.

Patterson and Rosen (1977) suggested that ancestors be treated as terminal taxa. Given the peculiar topologies that are predicted to result when an ancestral species is included in a phylogenetic analysis, this proposal would invoke the *sedis mutabilis* convention. Consider, for example, that the ancestral species of all higher bony fishes was discovered. The classification would appear as below (based on Wiley and Johnson, 2010).

Subclass Neopterygii (higher bony fishes)
Neopterygius primus, *sedis mutabilis*
 Infraclass Holostei, *sedis mutabilis*
 Infraclass Teleostei, *sedis mutabilis*

Such a classification bypasses controversies of whether or not *Neopterygius primus* is the ancestor of all descendants classified as neopterygian fishes. If one was bold enough to actually propose that *N. primus* was the ancestor, then some other convention would have to be applied. Indeed, in a truly general system of classification, we should anticipate that future investigators might be able to reliably identify stem species and take the view that, in general, phylogenetic classification must be able to accommodate all species, not simply descendant species (Wiley, 1979c, 1981a). Hennig (1966:71–72) gives a clue used by Wiley (1979c) for solving the "ancestor classification question."

From the fact that ... the boundaries of a "stem species" coincide with the boundaries of the taxon that includes all of its successor species, it follows that the "stem species" itself belongs in this taxon. But since, so to speak, it is identical with all the species that have arisen from it, the "stem species" occupies a special position in the taxon. If, for example, we knew with certainty the stem species of the birds (and it is only from such a premise that we can start in theoretical considerations), then we would no doubt have to include it in the group "Aves." But it could not be placed in any of the subgroups of the Aves. Rather, we would have to express unmistakably the fact that in the phylogenetic system it is equivalent to the totality of all species in the group.

Convention 8. A stem species (ancestral lineage/ancestral species) of a supraspecific taxon will be classified in a monotypic genus and placed in the hierarchy in parentheses at the side of the supraspecific taxon of which its descendants are parts.

Given *N. primus* as the ancestor of all other neopterygian fishes, it would be classified in the following manner:

Subclass Neopterygii (*Neopterygius primus*)
 Infraclass Holostei, *sedis mutabilis*
 Infraclass Teleostei, *sedis mutabilis*

This convention treats stem species as biologically relevant, yet preserves the relationship between the phylogeny and the classification. Consider that we discover the monophyletic nature of Aves through phylogenetic analysis. Although a monophyletic group now, Aves arose as a single ancestral species. It is perhaps irrelevant that if we were alive during the Jurassic and collecting specimens of this species we would never expect it to give rise to all birds. We would likely have been more concerned that one or several large and carnivorous similar "dinosaurian" forms were apt to eat us. It is only looking back through history and fitting Aves into the larger tree of life that we arrive at the place in the hierarchy that causes us to assign the rank superorder to Aves. If, sitting in the Jurassic, we could have predicted what would unfold in the future, we might have placed *Aves ancestroriscus* in its own monotypic superorder, but without knowledge of the next 150 million years, we very much doubt it.

This convention has additional benefits (Wiley, 1981a). (1) If we can discover stem species, their incorporation into existing phylogenies will have minimal impact. (2) Because the classification of ancestral species will have minimal impact, bold hypotheses can be proposed without having to dramatically change classifications to accommodate the hypothesis. (3) Only the phylogenetic system of classifying clades provides a ready backbone of classification that can accommodate the placement of all ancestral species while preserving both the logical correspondence between the phylogenetic tree and the classification and the biological significance of the ancestral species themselves.

There is an additional benefit that ties ancestors to the clade containing their descendants in an empirical manner. Synapomorphies are the evidence we use to circumscribe monophyletic groups. They are the historical effects of common ancestry. That is, at least some of the descendants of an ancestral species have the property we characterize as a synapomorphy, because they are descended from a common ancestral species that had the property by the time it speciated to leave descendants. Each synapomorphy "points to" at least one ancestral species where the synapomorphy was originally fixed as an autapomorphy and several may point to the same ancestor when its history is understood in total. Theoretically, that apomorphy that diagnoses the ancestral species is the same synapomorphy that groups the descendants in the part-whole relationship. This is not a typological or essentialistic position. In saying that all subsequent species are descended from an ancestor that has a particular synapomorphy, we are not saying that all descendants must have that synapomorphy. All tetrapods *sensu* Gaffney (1979) are descended from an ancestral species that had the tetrapod limb, but not all descendants of that ancestor need have a tetrapod limb (consider extant snakes for example).

Species and Higher Taxa of Hybrid Origin Species of hybrid origin are relatively rare among animals (White, 1978) but more common among plants (Grant, 1981). Botanical classification recognizes such taxa as nothotaxa and furnishes various rules for naming them (*International Code of Botanical Nomenclature*, 2000). Naming is one task, but phylogenetic classifications must also be capable of showing the relationships of taxa of hybrid origin relative to their parental species (Wiley, 1981a). The solution we adopt here is similar to the solution adopted in annotated classifications for placing stem species.

Convention 9. Taxa of hybrid origin will be classified with one or both parental species and its hybrid nature (apart from any nomenclatural rules applied) will be indicated by placing the names of the parental species, if known, beside the hybrid's name in parentheses. The sequence of the hybrid in a list carries no connotation of branching relative to nonhybrid taxa in a sequenced list of taxa.

Wiley (1981a) used an example of phylogenetic relationships of some members of the composite genus *Anacyclus* L. published by Humphries (1979) to illustrate use of this convention. Of the 12 species of *Anacyclus*, 3 were hypothesized to be of hybrid origin (Fig. 8.5). The classification below uses the sequencing convention and the hybrid convention to illustrate both the origin of hybrid species and the origin of species from species further removed on the tree. Note that two of the sections in the listing convention were not named by Humphries (1979).

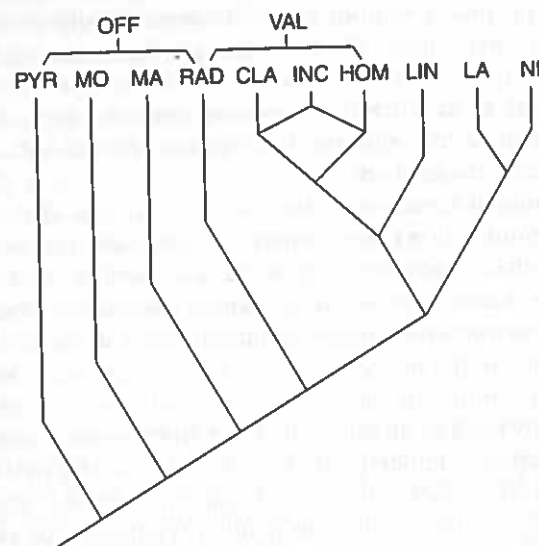


Figure 8.5. Humphries' (1979) hypothesis of relationships of the composite genus *Anacyclus* L. Abbreviations: CLA, *A. clavatus*; HOM, *A. homogamos*; INC, *A. x inconstans*; LA, *A. latealatus*; LIN, *A. linearilobus*; MA, *A. maroccanus*; MO, *A. monanthos*; NI, *A. nigellifolius*; OFF, *A. officinarum*; PYR, *A. pyrethrum*; RAD, *A. radiatus*; VAL, *A. x valentinus*.

Genus *Anacyclus* L.

Section *Pyretharia* DC

Anacyclus pyrethrum (L.) Link

Anacyclus officinarum Hayne (*A. pyrethrum* x *A. radiatus*)

Section *Anacyclus* L.

Anacyclus monanthos (L.) Thell.

Anacyclus maroccanus (Ball) Ball

Anacyclus radiatus Loisel

Anacyclus x valentinus L. (*A. radiatus* x *A. homogamos*)

Section (Unnamed 1)

Anacyclus linearilobus Boiss. & Reuter

Anacyclus homogamos (Maire) Humphries

Anacyclus clavatus (Desf.) Pers.

Anacyclus x inconstans Pomel (*A. homogamos* x *A. clavatus*)

Section (Unnamed 2)

Anacyclus latealatus Hub.-Mor.

Anacyclus nigellifolius Boiss.

ALTERNATIVE METHODS OF CLASSIFYING IN THE PHYLOGENETICS COMMUNITY

The Linnean system is only one of several ways of classifying organisms. Wiley (1981a) examined two others, numerical prefix schemes and rankless indentation. We shall examine both briefly in this section and then move on to the most recent proposal, the PhyloCode.

Numerical prefix schemes denote hierarchical rank with a prefix that is unique to each taxon and fixes the hierarchical level of each taxon relative to others. Early discussions of this approach are provided by Hull (1966), Hennig (1969, 1981), and Griffiths (1974). Hennig (1969, 1981) used numerical prefixes in conjunction with traditional Linnean names (i.e., the suffixes applied to names with roots and suffixes that conformed to Linnean nomenclature or to traditional practice within entomology) to classify insects in his landmark book *Insect Phylogeny*. An example of the system is shown in Fig. 8.6. Hennig (1969, 1981), began his classification with the prefix 1.0 to denote Entognatha and 2.0 Ectognatha. Each subordinate level added to the prefix designation (1.1, 1.1.1, 1.1.2, etc.). This produced an internally consistent classification and has some decided advantages. For example, fossil species can be inserted anywhere in the system without changing the hierarchical level of recent taxa (Griffiths, 1974). It is also preadapted to work well with computer languages, and the hierarchical levels are self-sustaining in that there is no need to create new rank categories as these are made "on the fly" as a consequence of adding taxa. There are disadvantages, however (Wiley, 1981a). Numerical prefixes are not the language of humans and are foreign to our efforts to communicate. The prefixes are unique, and thus, there will be as many prefixes as there are branches of the tree. It is easy to begin with 1.0 at

2.2.2.2.4.6 Mecopteroidea
 2.2.2.2.4.6.1 Amphiesmenophora
 2.2.2.2.4.6.1.1 Trichoptera
 2.2.2.2.4.6.1.2 Lepidoptera
 2.2.2.2.4.6.2 Antilophora
 2.2.2.2.4.6.2.1 Mecoptera
 2.2.2.2.4.6.2.2 Diptera

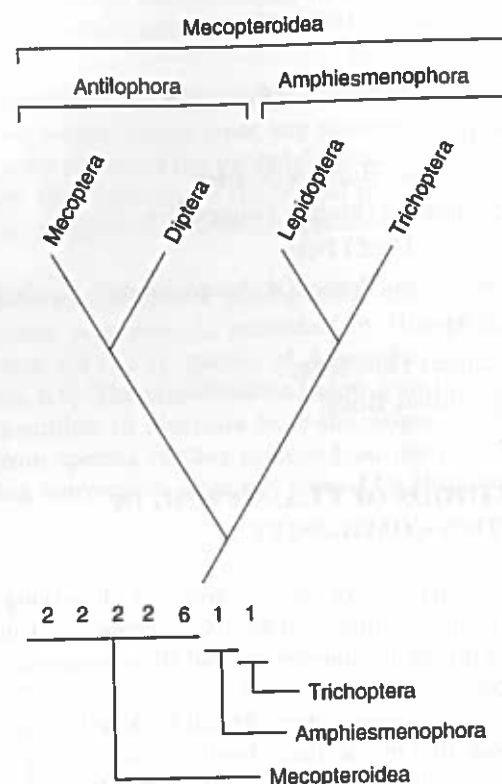


Figure 8.6. A hypothesis of the relationships among mecopteroid insects and the classification of Trichoptera using numerical prefixes (after Griffiths, 1974; from Wiley, 1979c).

Entognatha, but what if one began with prokaryotes? What would be the length of the prefix for Entognatha? Linnean ranks at least have the possibility of being reused (although they have the disadvantage of being interpreted by the unwary or unschooled of being biologically comparable between clades). Although Lovtrup's (1977) proposal that binary coding could be shortened by concatenation would certainly cut down on the length of the prefix, but this does not guarantee that the concatenation would yield unique prefixes (i.e., 1.1.1.1 = 4.0, but so does 2.2).

De Queiroz (1997:132) defends prefixes, stating that they are simple devices for representing hierarchical relationships (see also de Queiroz and Gauthier, 1992).

To de Queiroz, those who have criticized prefixes as "cumbersome and difficult to use in verbal communication" (e.g., Wiley, 1979c, 1981a; Eldredge and Cracraft, 1980; Ax, 1987) have misunderstood their use, as they are really not substitutes for Linnean ranks but "simple devices for representing hierarchical relationships." We make the following observations. First, if Linnean ranks serve to place groups of organisms such that their hierarchical relationships are shown, we fail to see why numerical prefixes, which serve the same purpose, are not "substitutes for Linnean ranks." Second, we leave to the reader to decide if 2.2.2.1.1.0.1 is simple and not cumbersome for humans. Third, we note that while Hennig used numerical prefixes, he also used formal Linnean name endings. In effect, Hennig (1969, 1981) hedged his bets.

Subordination by indentation is another alternative. In such schemes, one may use either ranked taxa or unranked taxa. If the taxa are ranked, the rank does not denote relative position in the hierarchy (Farris, 1976) but the indentation does serve this function. So, it is possible for one family to be included in another family by indenting. This clearly differentiates rank indentation schemes from indenting a traditional Linnean Hierarchy. In the latter, indentation is a way of presenting a visual clue as to the relative position of taxa. Pure indentation schemes dispense with ranks entirely and use only indentation to denote subordination. For example, the classifications below represent a traditional Linnean approach (above) and a pure indentation scheme (below).

Class Vertebrata
 Subclass Myxini (hagfishes)
 Subclass Petromyzontia (lampreys)
 Subclass Gnathostomata (jawed vertebrates)
 Infraclass Chondrichthyes (sharks, etc.)
 Class Teleostomi (bony fishes and tetrapods)

Vertebrata
 Myxini
 Petromyzontia
 Gnathostomata
 Chondrichthyes
 Teleostomi

Both classifications exactly reflect the underlying phylogeny generally accepted by many vertebrate systematists. The pure indentation scheme has an advantage in that if the phylogenetic position of hagfishes and lampreys changes relative to jawed vertebrates they can be easily moved. A disadvantage of the Linnean classification is that one has to memorize the hierarchical ranks order. However, rankless classifications have one practical difficulty (Wiley, 1981a): one must be able to line up coordinate taxa to confirm their sister group relationships. This presents no particular difficulty if the classification is relatively small and confined to a single page, but it does present difficulties when the classification is long and sister taxa appear on different pages. One might tend to lose one's place unless some standard was

introduced to measure the amount of indentation (Wiley, 1981a, suggested a marginal blue line for registration). In all fairness, we point out that this might also be a potential problem for the listing convention outlined above.

Our criticisms of both numerical prefixes and pure (or mixed) indentation schemes are practical, not theoretical. The fact is, all methods of classifying that are logically consistent with the underlying phylogenetic hypothesis and which are informative of that hypothesis serve the same basic purpose: to create a series of names of clades and their relationships that can be discussed by those interested in the history of the clades.

THE PHYLOCODE

A relatively recent development in classification is the proposal for an alternative formal system of nomenclature, the PhyloCode (available online: www.ohiou.edu/phylocode/index.html). The aim of the PhyloCode is the same as that of phylogenetic taxonomy in general (including phylogenetic taxonomy using Linnean nomenclature): to produce classifications that are logically consistent and fully informative concerning relationships among organisms. As of January 2011, this alternative code had not been implemented, but apparently it will be with the establishment of the *First Book of Phylogenetically Defined Names: A Companion to the PhyloCode*. This book will be the equivalent of the *Systema Naturae* for those who follow the PhyloCode and will contain names of clades that are approved by a PhyloCode nomenclature committee. Unlike the three major codes, the PhyloCode is not sanctioned by the International Union of Biological Sciences, and its rules of nomenclature will not carry the force of international "sanction" until such time as it is so sanctioned. This, of course, may be a matter of time or politics depending on the reception the PhyloCode receives when its governing body implements its rules. And besides, taxonomists have never let such formalities get in the way if they think the formalities are not useful.

The PhyloCode is designed to ensure the stability of names by defining the names of taxa through the use of specifiers. Specifiers are existing taxa or character homologies referenced to define the name relative to taxa included within the clade. There are three ways of defining the name relative to specifiers. The first is an inclusion statement: "Xinae is the name that refers to the clade stemming from the common ancestor of (the taxa named) *Xus* and *Yus*," where *Xus* and *Yus* are included taxa. The second is an inclusion/exclusion statement: "Xinae is the name of the clade that consists of all species that share a common ancestor with the taxon named *Yus* but not with that named *Zus*." The third is a synapomorphy statement: "Xinae is the name of the clade stemming from the first species to have the character 'hole on top of the head' that is homologous with the hole on top of the head found in the taxon named *Yus*." At least two specified things must be referenced in each case, either a combination of names of taxa or a combination of characters and taxa. There are alternative ways of specifying that are suitable depending on the intention of the author. Specifically, some forms of the definition are more suited for "crown clade" definitions (those designed to circumscribe only extant clades) as compared to "total clade" definitions where the taxon is circumscribed to include a number of more basal fossil species or clades. The PhyloCode differs in some important aspects

from most classification schemes in several ways. These should be understood by those wishing to employ it.

1. The PhyloCode assumes that taxa governed by the PhyloCode have certain biological characteristics; names must refer to clades (Preamble, 2). (Traditional codes have no such assumptions; they are "biology free.")
2. As it purports to govern only the names of clades, it does not serve as a general system for classification because explicit phylogenetic hypotheses are not available for many groups of organisms.
3. The PhyloCode extends the type concept and priority (through the use of specifiers and a registry of names, to be published) to all levels of the hierarchy in zoology, botany, and prokaryote classification.
4. Name endings do not change with changes in hierarchical rank. More to the point: the formation of names is independent of categorical rank.
5. Priority for the use of names is not determined by first use of the name in the literature but by first registration of the name in a list maintained by the governing body of the PhyloCode. Preexisting names can be "converted" by applying the appropriate definitional phrases and specifiers, but priority for the use of that name rests with the date the name was formally registered, not the date they were originally published.
6. Stability of names is achieved by fixing the name to a specific context determined by the specification as recorded. Once used in this context, the name is not available for use in other contexts.
7. The meaning of *clade* is not synonymous with the meaning of *monophyletic group*. A clade can be a monophyletic group of species or a clone or even individual organisms.
8. Clades can be overlapping (restricted to the case of taxa of hybrid origin).
9. There are several ways of recognizing clades (see above) but a single concept of monophyly. This can lead to difficulties depending on what kind of graph is referenced. For example, if one applies the "node-based" concept to a phylogenetic tree rather than to a Hennig tree, the ancestral species of the group would seem to be excluded from the group. This is problematic in a phylogenetic perspective because it is contra the entire phylogenetic enterprise, for instance, see Hennig (1966:71–72). To quote the PhyloCode (Article 2, Note 2.1.4 as of April 2007): "A node-based clade is a clade originating from a particular node on a phylogenetic tree, where the node represents a lineage at the instant of a splitting event." And from Article 9, Note 9.4.1: "A node-based definition may take the form 'the clade stemming from the most recent common ancestor of A and B. ...'"

PhyloCode names differ from Linnean names in a number of other respects. Available names are placed in an approved database of clade names maintained by the PhyloCode commission. *Xinae* will forever mean "the name of the clade stemming from the common ancestor of *Xus* and *Yus*." If one wishes to extend the name *Xinae* to include more basal taxa, then one would form a panclade name, *Pan-Xinae*, and define it appropriately.

The PhyloCode also differs from Linnean codes in its treatment of ranks. Ranks might be used (they are completely optional), but name endings do not change with a change in rank. For example, if we find that Agamidae is a clade that includes Chamaeleonidae, then Chamaeleonidae would be included within Agamidae without changing the suffix of the root name. This treatment of names is one of the sticking points for those who use Linnean nomenclature. In particular, name endings mean something in Linnean nomenclature at certain levels of the hierarchy, where they serve as exclusion devices (a member of Agamidae cannot also be a member of Chamaeleonidae if both are monophyletic). However, name endings are meaningless in PhyloCode nomenclature at all levels of the hierarchy. It should be noted that name changes in, for example, the Zoological Code, are only affected if the name is referred to a clade of the rank family or below, no name changes are governed for the names of taxa ranked higher than the family group. This does not mean that name endings will not change, but the changes are not governed by the code.

The PhyloCode also differs from the Linnean Codes in its view of the meaning of the taxon names. The Linnean Codes do not purport to give biological meaning to the names of taxa. In this regard, they are Millian, that is, they treat names as mere labels in accordance with the philosophy of proper names espoused by John Stewart Mill (1872). In contrast, PhyloCode names are Russellian, that is, they treat names as synonyms of their definitions in accordance with the philosophy of proper names espoused by Bertrand Russell (1919). For a discussion on these points, see Härlin (1998, 2003a) and Härlin and Sundberg (1998). We pursue this topic more fully below in the section on the meaning of proper names.

PhyloCode Controversies

A vigorous debate has ensued over the PhyloCode since it was first proposed by de Queiroz and Gauthier (1992). Part of this debate has resulted from a seeming misunderstanding, on the part of PhyloCode proponents, about the Linnean Codes. A number of such misunderstandings are significant.

Early advocates of what became the PhyloCode claimed that a new code was needed because Linnean classifications are essentialistic (de Queiroz and Gauthier, 1990), classification has not caught up with the Darwinian revolution, and Linnean taxonomic practices have inhibited the modernization of classification. However, these charges seem to be misplaced. In particular, claims that Linnaeus was an essentialist apparently trace back no farther than Cain (1958), and although repeated by such workers as Mayr (1959, 1963, 1968, 1976, 1982), they have been demonstrated to be false (Winsor, 2006). According to Winsor (2006) it is not clear that Linnaeus studied logic at all, much less Aristotelean logic. In fact, Linnean classifications made by Linnaeus and his followers "seemed to involve an active neglect of the classic rules of logical definition" (Whewell, 1847, cited in Winsor, 2003:3). Apparently the "essential characters" of Linnaeus were "key characters," those used in "keys" for identification; essential characters are characters of convenience not essential characters in the sense of Aristotelian logic. As evidence, one only need to consider the writings of Linnaeus:

Anyone who thinks that he can understand botany from the essential character and disregards the natural one is therefore deceiving and deceived; for the essential

character cannot fail to be deceptive in quite a number of cases. The natural character is the foundation of the genera of plants, and no one has ever made a proper judgment about a genus without its help; and it is and always will be the absolute foundation of the understanding of plants (Linnaeus, 1751:143, translation in Winsor, 2006).

Further demonstration that Linnaeus was not following Aristotle is the fact that he used genus and species as fixed hierarchical terms; their use in Aristotelian logic is relative (e.g., bird is a genus containing the species swan; bird is a species contained in the genus animal; Winsor, 2006). This "nonessentialistic" concept of the Linnean system was in fact recognized by Darwin (1859:413–414):

Such expression as that famous one of Linnaeus, and which we meet with in a more or less concealed form, the characters do not make the genus, but that the genus gives the characters, seem to imply that something more is included in our classification, than mere resemblance. I believe that something more is included, and that is propinquity of descent,—the only known causes of the similarity of organic beings,—is the bond, hidden as it is by various degrees of modification, which is partially revealed to us by our classifications.

Apparently, de Queiroz (1997:132) agrees. However, just because Linnean nomenclature is not inherently essentialistic does not mean it is to be preferred over a proposition such as the PhyloCode. It only means that rejecting the Linnean system because it is essentialistic is misplaced.

De Queiroz (1997) observes that changing paradigms from a creationist perspective to an evolutionary perspective did little to change Linnean classification or call into question the efficacy of the Linnean system. However, this change, according to de Queiroz (1997:128), "contradicted the Aristotelian context within which the Linnean Hierarchy was originally developed." Setting aside the question of whether Linnaeus was an Aristotelian (addressed above), it is certainly true that the basis for perceived hierarchy changed, whatever Linnaeus might have thought. But does this matter? The lag between the general acceptance of descent with modification and its explanatory power vis-à-vis groups-within-groups/part-whole hierarchies seems more a matter of two later developments: (1) the concept that genealogy alone should be the primary basis of relationships expressed in hierarchies and (2) the realization that many groups previously thought to be monophyletic were actually paraphyletic. These are conceptual issues that underlie the Hennigian Revolution and were not, so far as we can determine, inhibited (or promoted) by the use of the Linnean system of nomenclature. Thus, while it might be true that descent with modification played a rather superficial role in classifications between 1859 and today (a matter for historians of science to examine and which we doubt), it has not been established that the formalities of Linnean nomenclature caused this inhibition.

De Queiroz (1997) suggests that the realization that species are lineages effectively "redefined the Linnean category Species" and decoupled it from the rest of the Linnean Hierarchy. We certainly agree that species are lineages (Wiley, 1978, 1981a; Lieberman, 1992; Wiley and Mayden, 2000a; Wiley, 2002, 2007). Further, that species-as-taxa are different from clades-as-taxa is not doubted, nor do we doubt that real

species and real clades have objective reality apart from Linnean nomenclature. But we fail to understand why this would lead to rejection of Linnean nomenclature. The various international codes disavow placing biological meaning on names; biological meaning is left to the biologist to interpret (bringing us back to the distinction between Millian and Russellian meanings of words discussed previously). Systematists are free to consider a particular species such as *Fundulus nottii* as a lineage or a phenetic cluster. (Those who wish for species to function as comparative tools in evolutionary biology will gravitate to one concept, and those who see species names as only taxonomic devices may not.) As for higher taxa, we suspect that many are not clades, but there is little we can do about this until a phylogenetic analysis is performed on such a group's members and presumed relatives. In fact, the codes for the Linnean category "species" simply outline the rules for naming entities thought to be species and govern the use of such names when conflicts occur. It is up to the biological community to examine whether these names apply to lineages (or whatever your favorite species concept might be). It is true that there is nothing in the Linnean Codes that prohibits the naming of para- or even polyphyletic taxa. But it is also true that there is nothing in the Linnean system per se that prohibits a purely phylogenetic system of classification (Barkley et al., 2004a, b).

Two additional misconceptions need to be addressed. First, claims that the Linnean system is typological because names are defined by characters (de Queiroz and Gauthier, 1992) are false. In Linnean nomenclatural systems the *taxa* are "defined" or diagnosed by characters, not names of taxa. Names are formed using certain rules at certain levels. The form of the name may be dictated by the form of the name of a type, but if anything actually defines the name, it is the type (where typification applies), not the characters of the type. Second, in spite of repeated claims that there are five mandatory taxonomic categories (e.g., Laurin, 2005), such mandatory categories are nonexistent.

If we set aside their misconceptions surrounding Linnean Codes, PhyloCode proponents do have some valid points. For example, there are simply not enough categorical levels, especially within the genus group, to serve the needs of some practicing phylogeneticists (e.g., Hillis, 2006). In fact, this lack of hierarchical ranks is exactly what precipitated the many proposals incorporated into the Annotated Linnean Hierarchy and is behind other alternative systems discussed above. Theoretically, there is no reason to restrict the number of hierarchical ranks (Farris, 1976). However, coming up with unique suffixes is problematic, and thus, the inclusion/exclusion function of name endings might only serve for as many ranks as can be provided reasonable and grammatical endings to be added to the root. (Note that this is a problem not articulated by PhyloCode proponents.)

Second, it is true that changing a phylogeny may precipitate many name changes, with the same taxon changing name endings as it is pushed up or down the phylogeny. Of course, phylogeneticists who adhere to Linnean nomenclature might welcome such changes as signals of a paradigm shift in the classification of their groups (Härlin, 2003a). So, changing the meaning of names by changing the content of clades or recognizing that a name did not belong to a clade at all is not necessarily bad.

If we set aside adherence to pure tradition of Linnean nomenclature, there are other, additional criticisms of the PhyloCode. At one level, PhyloCode circumscription of taxon names is unproblematic. Indeed, the same strategy could be used

within the Linnean system, although one of the specifiers would have to be the type where typification applies. However, there are problems. These problems exist on two levels. The first problem is empirical: the stability of names relative to specifiers is given precedence to the stability of the total content of the clades named (point made by Härlin, 2003b). The second problem is ontological. To achieve stability of names relative to specifiers, taxa must be treated as kinds and not as individuals (Härlin, 1998, 1999, 2003b). This seems paradoxical, especially given that one of the original reasons for developing the PhyloCode was to combat essentialistic thinking, but we will show why this is the case.

Stability of Names Relative to Clade Content

The PhyloCode purports to define names through the use of specifiers. As pointed out by Härlin (1999) and Forey (2001), future research might have unexpected nomenclatural results relative to the contents of clades, if the relationships among the specifiers changes. The node relative to the specifiers might not have a stable place in the larger phylogeny, and thus, the ancestor involved might actually end up having different relationships and a different group of descendants than originally intended. One is left wondering if this is the same ancestor as that intended by the person who formed the proper name originally (it is not, as a matter of contingency; Härlin, 2003b). Traditional classifications have the same problem, shifting ideas of relationship results in changing content and meaning of clades. The difference is that PhyloCode names must follow the specifiers even with content change, while traditional (Linnean) names do not. Linnean names only have to follow the rules of priority if they fall within the scope of the particular code and community consensus if they do not.

Part of the problem with the reception of the PhyloCode by its critics is what we might term the "part-whole" instability problem (e.g., Härlin, 1999; Nixon and Carpenter, 2000; Forey, 2001, 2002; Carpenter, 2003). While striving for the stability of names, it seems to create instability in the content of clades relative to names. Since no one must pay attention to the historical seniority of names as an arbitrator, and since PhyloCode extends priority to all levels, the "neo-seniority" of PhyloCode names by the registration process may cause unintended problems. For example, the phylogeny and names of three clades of apical bony fishes as of 2009 is shown in Fig. 8.7a. We form and register a node-based name, *Halecomorphi*, with specifiers *Amia calva* (AC, the living bowfin) and *Elops saurus* (ES, a teleost). Later we form the node-based name *Neopterygii* with specifiers *Lepisosteus osseus* (LS, a gar) and *E. saurus*. However, Grande (2010) demonstrates that the bowfin is actually more closely related to the gar than to the teleost (Fig. 8.7b). The name used since the 1920s to designate the group that includes gars, bowfins, and teleosts must now be replaced by the name adopted in the 1970s meant to exclude gars.

Stem-based names also create problems with unintended consequences. Consider *Halecostomi* as that clade defined as the clade that sprang from the common ancestor of *A. calva* but not from the common ancestor of *L. osseus* and *A. calva* (Fig. 8.7c) and the definition that *Neopterygii* is that clade that sprang from the common ancestor of *L. osseus* but not from the common ancestor of *L. osseus* and the paddlefish, *Polyodon spatula* (Fig. 8.7c). Grande's (2010) phylogeny would dictate that

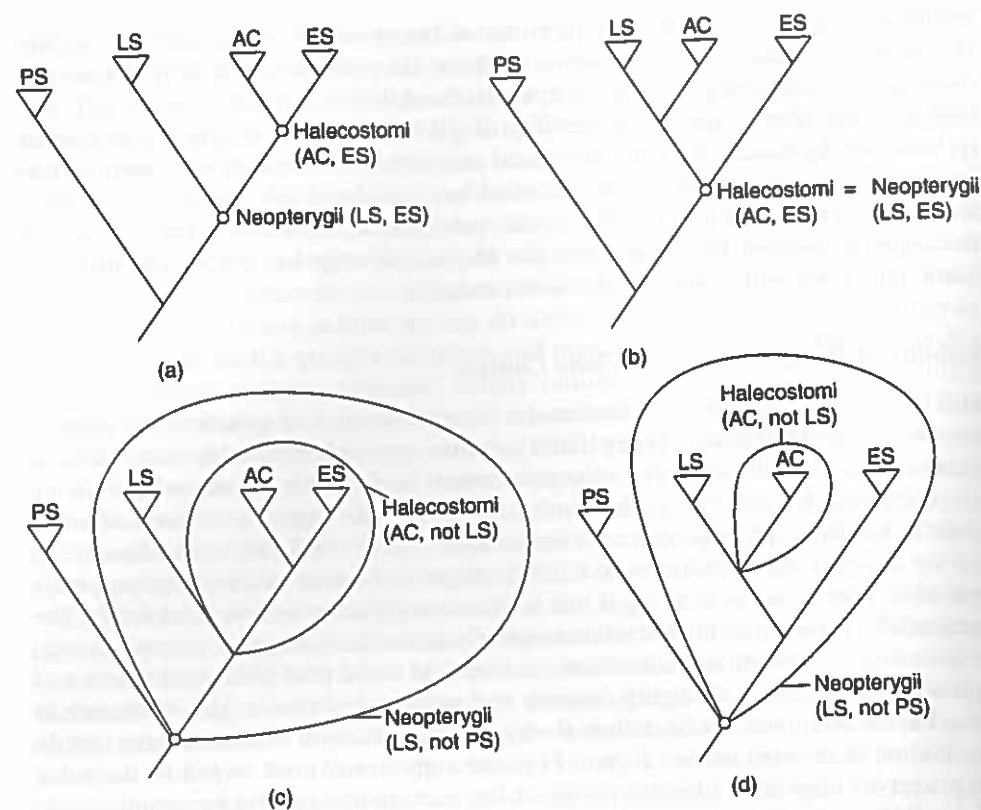


Figure 8.7. Name and content changes following the PhyloCode. (a–b) Alternative hypotheses of the relationships of basal actinopterygian fishes. If the 2009 community consensus hypothesis (a) is replaced by the alternative (b), then the names Halecostomi and Neopterygii become synonyms and Neopterygii is discarded as the younger name. (c–d) Node-based names fare no better. The tree in (c) is the traditional tree with groups circumscribed by ovals. If we accept tree (b), then Halecostomi is restricted only to *Amia calva* and the synapomorphy of this group must be changed from having an interoperculum to, for example, having double vertebral centra.

Halecostomi includes *A. calva* but excludes *E. saurus*, and this was not the intent of previous workers such as Rosen and Patterson (1977). Of course, we might be fortunate to have registered Neopterygii before Halecostomi, but since there are no rules about the historical priority of names relative to PhyloCode registration, there is no guard against such a scenario.

Now consider what might happen to the characters associated with the names. The presence of an interoperculum has been traditionally associated with the name Halecostomi as a synapomorphy uniting bowfins with teleosts. Grande (2010) showed that some fossil gars, basal on the gar phylogeny, have interoperculars. In traditional nomenclature the presence of interoperculars would simply be included in the diagnosis of Neopterygii (Halecostomi being discarded), and this would cause no problems with a node-based PhyloCode Halecostomi (=Neopterygii) so long as

the community was willing to accept Halecostomi as a replacement for Neopterygii. If, however, we opted for stem-based names, Halecostomi would need a replacement character diagnosis since the interoperculum no longer diagnoses Halecostomi but the larger group Neopterygii. Finally, there is the problem of homoplasy; characters that are used as specifiers for clades that are later found to be homoplasious will cause needed name changes (Forey, 2001).

We have no doubt that the PhyloCode can work as its adherents intend. In the end, after all, any system of nomenclature will work if we adhere to the rules of the system and this system contains no internal conflicts. And we do not question the good intent of PhyloCode adherents to wishing to make nomenclature totally phylogenetic. However, we wonder, along with Forey (2001, 2002), whether the benefits are worth the costs. Is the stability of names rather than the community meaning of those names what we are really striving to achieve?

We have primarily been focused on perceived epistemological problems of PhyloCode names. There is an additional ontological problem and we now deal with this problem because it is a larger issue of more general interest to our understanding of names in taxonomy.

PROPER NAMES OF TAXA

Given that taxa are individuals (Ghiselin, 1966; Hennig, 1966), the names of taxa are proper names. There are two basic philosophies of the nature of proper names that are related (broadly) to this discussion. These two philosophies cause controversy among phylogeneticists about the nature of proper names in taxonomy. The controversy is not easy to comprehend, because it underlies different philosophies of meaning and language. John Stewart Mill held that “a proper name is but an unmeaning mark which we connect in our minds with the idea of an object, in order that whenever the mark meets our eyes or occurs to our thoughts, we may think of that individual object” (Mill, 1872:22). Proper names are attached to an object (or entity) and are not dependent on any properties of the object. As Wettstein (1999:124) concludes: “A proper name, once attached, becomes a socially available device for making the relevant name bearer a subject of discourse.”

In contrast, Russell (1919) held that proper names were substitutes for a description or set of properties. Russell’s approach is associated with Frege in that both held that proper names are abbreviated descriptions. This is sometimes called the Russellian or descriptivist philosophy of simple proper names.

Mill’s approach (the Millian philosophy) was considered by Kripke (1980), who argued that simple proper names (the names given to individuals) were not determined by a descriptive condition but by a causal chain that links name to references. Härlin (1998) called attention to these different approaches and asserted that only the Millian approach is consistent with the view that taxa are ontological individuals, as asserted by Ghiselin (e.g., 1966, 1995, 1997, 2007).

Both views have their philosophical problems, and two examples will show this. The Millian account of names has problems with statements that refer to nonexistent entities like “Santa Claus lives at the North Pole” or “Reptilia is paraphyletic.” Millians counter that while neither Santa Claus nor Reptilia are real, thoughts are nevertheless communicated, and thus the proper names of unreal entities serve a

useful purpose in that communication is achieved. Kripke's "causal chain that links name to reference" can be invoked to ensure that communication is meaningful. "Reptilia is paraphyletic" can be meaningfully linked: "Reptilia *sensu* Romer (1966) is paraphyletic." Indeed, by linking name to reference, we can always know the meaning of a Millian proper name, so long as we know the reference. Patterson (1977) meant Teleostei to be that clade of bony fishes found in the phylogeny above the branch leading to *Amia calva* (a more inclusive clade). Arratia (1999) meant Teleostei to be that clade found in the phylogeny above *Proleptolepis* (a less inclusive clade). We completely understand how the same proper name is used in different ways if we understand the reference, or chain of inference. If one particular use of the proper name is dominant or its use is unproblematic, we dispense with the referent. For example, among neonotologists, uttering "Teleostei" conjures up the same concept of a particular clade of living fishes, but among paleontologists who are worried about whether a particular fossil is or is not a teleost, a finer distinction might be needed to tie down exactly what we are talking about. Is *Pachycormiformes* a teleost *sensu* Patterson and *sensu* Arratia? (Answer: it is to Patterson but not to Arratia).

The descriptivist account has problems of rigidity and descriptive adequacy; while purporting to provide a description, the description when applied to evolving systems is frequently inadequate. For example, while the clade Tetrapoda *sensu* Gaffney (1979) is diagnosed with the synapomorphy of the tetrapod limb, not all tetrapods have this limb. (Indeed, no tetrapod has this limb during all phases of its life cycle.) Strict application of the definition would leave out snakes and other limbless tetrapods. Descriptivist accounts of natural kind names are quite another matter. Given that candidates for natural kinds have necessary and sufficient definitions, one could argue that helium, for example, is the kind name for those atoms with the property of having two protons. However, kind names are not proper names and philosophers of both camps generally agree that kinds do not have proper names.

Running through this controversy is the problem of what proper names should be referring to in the first place. The usual referent is a species or a clade. When Arratia or Patterson refers to Teleostei, they are referring to an entire clade (a whole), not just to some exemplars (only parts of the whole). These are contingent propositions that reference particular phylogenies, not necessary truths that refer to all possible phylogenies (Ghiselin, 1995; Härlin and Sundberg, 1998). They are not associated with necessary properties, but only contingent properties—those properties that are true given that the phylogeny in question is true. As such, they can hardly be mistaken for descriptivist proper names. This is so even within the Linnean system that uses type species to form names. The type species of *Homo* (*H. sapiens*) does not define the genus *Homo* in a descriptivist manner. Instead, it limits the content to the proper names of that clade, ranked as a genus, to those other species that share a common ancestor with *Homo sapiens* but not with, for example, *Pan tryglodytes* (because *Pan* is the name of another clade ranked as a genus). *Homo* is a Millian name, and if in doubt, a referent can be cited to tie down the meaning in a particular context.

Kripke (1980) coined the term *rigid designator* and claimed that proper names can only be used rigidly. This is a Millian concept, given that no definite description gives meaning to a proper name. But Kripke argues that one might "fix the reference" of a proper name. One way of fixing the reference is to point at the individual.

This amounts to fixing the reference by ostention. In phylogenetic classifications, the reference for clade names are to all the parts of the clade. The parts of the clade are "pointed at" through the inclusion of subclades within the named clade by classifying the parts. One always "knows" what a clade name means as a contingent proposition because the people who use the name signaled their intention for the meaning of the name by pointing directly at or alluding to the subclades within the clade so named. If we follow Kripke, this name would be rigidly designated and would apply to the clade in all possible worlds where the clade exists as a clade. (Of course, in worlds where the clade did not exist as a clade, then the name would not apply to anything.)

Another way to ostensibly fix the reference for a name is to make a declarative statement of the sort advocated by advocates of the Phylocode. They convey the intention by pointing at two parts of the clade and making a declarative statement. For example: "Mammalia might be defined as the clade stemming from the most recent common ancestor of horses and echnidas" (de Queiroz, 1995:224). Fixing the reference through such ostensive "definitions" is said to lead to taxonomic stability. Unfortunately, this is not true, as the names are proper names. Consider Smith and Jones. Let's imagine the phylogeny of Smith has hippos and kangaroos joining the phylogenetic tree between echnidas and horses, but the phylogeny of Jones has kangaroos branching before echnidas. *Mammalia sensu* Jones is not the same clade as *Mammalia sensu* Smith because the whole of *Mammalia sensu* Jones is not the same whole as that of Smith.

The simple solution to this problem is to adopt the concept that taxa are not individuals, but rather, some sort of kind or class (Härlin, 1998). This seems the solution advocated by de Queiroz (1995). This would lead to exactly the same sort of stability as the stability of other kinds. Helium is always the kind of atom that has two protons. Members of the kind vary from day to day, depending on, among other things, the number of hydrogen fusions and uranium decays in the universe. *Mammalia* is always the kind that includes the common ancestor of echnidas and humans, and whether hippos or horses or kangaroos happen to be members of the kind is quite irrelevant.

But there is a problem with this solution. *Mammalia*, as a kind, can hardly be a natural kind. *Mammalia* is not predicted by any general theory of natural processes. Clades, in general, are predicted to exist due to speciation, but particular clades are not predicted; they are a matter of historical contingency. If *Mammalia* is a kind, it must be a nominal kind, not a natural kind. But if taxa are nominal kinds, then why are names of monophyletic taxa any better than names of polyphyletic taxa?

We conclude that the benefit of treating the names of taxa as kind names with necessary and sufficient properties in order to achieve name "stability" is far outweighed by the cost, both philosophical and phylogenetic. Stability in classification, as such, is not a particular goal of the phylogenetic system. Rather, consilience of the phylogeny with the classification is what we seek. And with that consilience the names will take care of themselves.

THE FUTURE OF LINNEAN NOMENCLATURE

The future is a matter of historical contingency and thus impossible to predict even if we know the constraints of history. However, there are several possibilities

for the future of biological nomenclature. With the rise of computers it is conceivable that classification per se might not be needed at all, simply some rules of priority. For example, we could hyperlink every name to a tree graph and place the names in any order we wish. Select a name from the list and up pops a tree detailing its closest relatives. We suspect that nomenclature could evolve into even more hybrid systems than now provided by systems like the Annotated Linnean Hierarchy. For example, Linnean ranks could be used to denote some number of levels of hierarchy with appropriate rules (priority, naming, etc.) and within these levels another system could be used (e.g., indentation, numerical prefixes, or simply lists hyperlinked to trees). How phylogenetic classification will evolve is up to the community and is unlikely to be well served solely by committees, however well meaning.

ALTERNATIVE "SCHOOLS" AND LOGICAL CONSISTENCY

In Chapter 4 we used the tool of logical consistency to examine the nature of paraphyletic groups relative to monophyletic groups. We demonstrated that there were two kinds of taxon groupings relative to phylogeny: monophyletic and nonmonophyletic. This was based on the observation that paraphyletic groups, like polyphyletic groups, were not logically consistent relative to a phylogeny that contains the groups. Wiley (1981a) used a variety of arguments to counter the claims by evolutionary taxonomists such as Mayr (1969, 1974), Ashlock (1971, 1972), Simpson (1961, 1975), and Bock (1974) that their classifications containing paraphyletic groups were superior to those containing only monophyletic groups. Wiley (1981a) refuted the notion that the school of evolutionary taxonomy was superior to the school of phylogenetic systematics. However, Hull (1964) had provided a more succinct argument earlier. He showed that the claims by Simpson (1961) that classification should be logically consistent with the underlying phylogeny were true, but Simpson's (and other evolutionary taxonomists') recognition of paraphyletic groups renders classifications containing such groups logically inconsistent with the underlying phylogeny. Again, this undermines the entire program of evolutionary taxonomy. This fact, and the fact that phenetics has largely disappeared, negates the need for a separate chapter contrasting the phylogenetic system with alternatives, because there are no other.

CHAPTER SUMMARY

- Phylogenetic classifications are systematizations.
- Phylogenetic classifications are logically consistent with the phylogeny that they purport to summarize.
- There are a variety of ways of classifying that are both natural and useful but have different knowledge goals.
- Classifications of natural kinds tend to be nonhierarchical or only partly hierarchical.

- Phylogenetic classifications are part-whole hypotheses and fully hierarchical.
- Linnean ranks do not rank comparable groups between clades.
- Classifications using the Linnean system are but one of several ways of achieving the goals phylogenetic classification
- An annotated Linnean system is summarized.
- The PhyloCode is critiqued, and the philosophy of proper names is discussed.