

"PRINCIPLES OF PHYLOGENETICS: ECOLOGY AND EVOLUTION"

Integrative Biology 200B
University of California, Berkeley

Spring 2011
B.D. Mishler

Jan. 25, 2009. **What can we do with trees once we have them? Step 1: Classification, including species.**

Biological systematics encompasses three distinct activities: taxonomy, classification (which may or may not be a reasonable reflection of phylogeny) and nomenclature (Fig. 1). Although these three components are rigorously and distinctly practiced by systematists, they are often amalgamated under the term “taxonomy.” While the breadth of “taxonomy” is clearly understood among most practitioners, it can obscure the methodology and practices of modern systematics to others. On the other hand, not all systematists work across the full breadth of systematics. For example, they can be engaged in the study of molecular phylogenies without applying the results of their studies to the nomenclature of the group. Similarly, the resolution of nomenclatural issues can be carried out without a phylogenetic study of the species or the generation of a new classification, but usually not without extensive library resources.

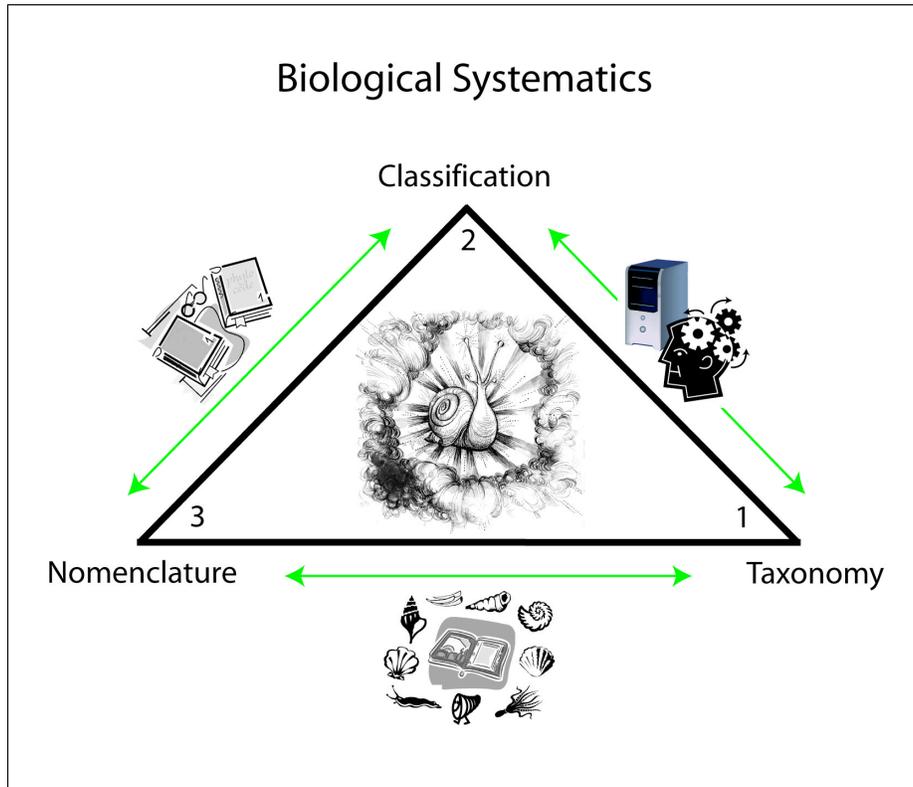


Figure 1. Biological Systematics a summary showing the relationship between taxonomy (1), classification (2), and nomenclature (3).

I. Purposes for classification:

- 1) Practicality
(Operationality, ease, stability)

- 2) Information Content
(Optimal summarization of what is known about entities)
- 3) Predictivity
(Of unknown features of entities)
- 4) Function In Theories
("Capture" entities acting in, or resulting from, natural processes)

II. History:

Systematics has always played a central role in the history of biology. The recognition of basic kinds of organisms, their properties and "relationships" in higher categories, was the earliest biological discipline. Developments in biology as a whole have interacted with systematics throughout. Detailed treatment of this history include Stevens, Hull, Dupuis, Donoghue & Kadereit, Mayr.

HISTORICAL PERIODS IN BIOLOGICAL SYSTEMATICS:

- 1) PRE-HISTORY -- FOLK CLASSIFICATIONS
- * 2) ANCIENT GREEKS THROUGH LINNAEUS--
ESSENTIALISM
- * 3) NATURAL SYSTEM -- OVERALL RESEMBLANCE;
"IMPORTANCE"
- 4) DARWIN -- EVOLUTIONARY LANGUAGE [only
a superficial effect]
- 5) NUMERICAL PHENETICS -- COMPUTERS [only
a superficial effect]
- * 6) PHYLOGENETIC SYSTEMATICS (CLADISTICS) --
SYNAPOMORPHIES, MONOPHYLY

* Argued to be the only true revolutions in the conceptual bases of systematics

III. Classification versus taxonomy versus systematics?

The majority view is that "systematics" is the more general term, being defined as the study of organismal diversity, whereas "taxonomy" is more narrowly the set of procedures and rules for naming entities (taxa) and producing "classifications" (ordering and/or nesting of taxa).

IV. Natural classification:

All three schools of systematics wanted to produce "natural" classifications:

1. Pheneticists view natural groups as those taxa linked by the greatest similarity to each other.
2. To evolutionary systematists, "natural" groups are defined by gaps between taxa in characters for which an evolutionary scenario can be argued.
3. Cladists consider natural groups to be monophyletic, and thus "natural" classifications reflect the tree of life.

V. The winning argument for phylogenetic classifications:

The debate over classification has a long and checkered history (see Hull 1988; Stevens, 1994). A conceptual upheaval in the 1970's and 80's resulted in a true scientific revolution --Hennigian Phylogenetic Systematics. Many issues were at stake in that era, foremost of which was the nature of taxa. Are they just convenient groupings of organisms with similar features, or are they lineages, marked by homologies? A general, if not completely universal consensus has been reached, that taxa are (or at least should be) the latter (Hennig, 1966; Nelson, 1973; Farris, 1983; Sober, 1988).

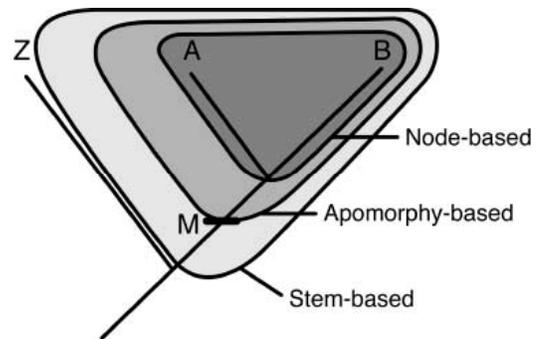
A summary of the arguments for why formal taxonomic names should be used solely to

represent phylogenetic groups is as follows: evolution is the single most powerful and general process underlying biological diversity. The major outcome of the evolutionary process is the production of an ever-branching phylogenetic tree, through descent with modification along the branches. This results in life being organized as a hierarchy of nested monophyletic groups. Since the most effective and natural classification systems are those that "capture" entities resulting from processes generating the things being classified, the general biological classification system should be used to reflect the tree of life. Phylogenetic taxa will thus be "natural" in the sense of being the result of the evolutionary process.

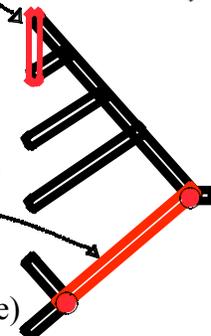
This isn't to say that phylogeny is the only important organizing principle in biology, There are many ways of classifying organisms into a hierarchy, because of the many biological processes impinging on organisms. Many kinds of non-phylogenetic biological groupings are unquestionably useful for special purposes (e.g., "producers," "rain forests," "hummingbird pollinated plants," "bacteria"). However, it is generally agreed that there should be one consistent, general-purpose, reference system, for which the Linnaean hierarchy should be reserved. Phylogeny is the best criterion for the general purpose classification, both theoretically (the tree of life is the single universal outcome of the evolutionary process) and practically (phylogenetic relationship is the best criterion for summarizing known data about attributes of organisms and predicting unknown attributes). The other possible ways to classify can of course be used simultaneously, but should be regarded as special purpose classifications and clearly distinguished from phylogenetic formal taxa.

VI. Phylogenetic classification in detail:

1. Three ways of defining a name: Node-, Stem-, or Apomorphy-based.
2. Different ways of defining monophyly: *synchronic* (i.e., "all and only descendants of a common ancestor") or *diachronic* (i.e., "an ancestor and all of its descendants"). Which is better? Should the word "species" appear in the definition of monophyly? Does it matter?



- Clade: a synchronic, monophyletic set of lineage-representatives (i.e., Hennig's semaphoronts)
- Monophyletic: all and only descendants of a common ancestor
- Lineage: a diachronic ancestor-descendant connection (1 species in the de Queiroz sense)



3. Clade versus Lineage (next page)

Three major logical phases of phylogenetics:

(1) *Character Analysis* (synthetic, inductive)

the elements of a cladistic data matrix (i.e., OTUs, characters, and character-states) are assembled.

this complex process involves considerable reciprocal illumination (since developing hypotheses of distinct, independent characters with discrete states goes hand in

hand with developing hypotheses of homogeneous OTUs).

(2) *Phylogenetic or cladistic analysis* (analytic, deductive)

☒ the data matrix is translated into a cladogram using a parsimony/ ML model.

☒ reciprocal illumination is often involved here as well, since incongruence between characters or odd behavior of particular OTUs may lead to a return to phase 1 (a reexamination of OTUs, characters, and models) primarily to check for fit to assumptions of the cladistic method:

☒ OTUs should be homogeneous for the characters employed, and be the result of a diverging process rather than a reticulating process; characters should be discrete, heritable, and independent.

(3) *Classification and evolutionary analysis*

☒ the cladogram is translated into a classification, based on an assessment of the relative support for different clades.

☒ formal taxa (including species) are named here, on the basis of clear support for their existence as monophyletic cross-sections of a lineage, and for their utility in developing and discussing process theories.

☒ comparative methods applied (the remainder of this class!)

VII. Species in detail:

A. Different approaches:

- a. typological or essentialist approach (i.e., systematics through Linnaeus)
- b. phenetic, morphological, or "natural" approach
 - older botanists (Gray, Bentham, Hooker) plus many recent botanists (Cronquist, Levin, Sokal & Crovello)
 - some recent cladists (!) (Nelson & Platnick, Cracraft, Nixon & Wheeler)
- c. "biological" species concept: interbreeding groups
 - classic isolation approach (most zoologists, e.g., Mayr, Dobzhansky)
 - newer recognition approach (some zoologists, see Paterson)
- d. "evolutionary" species concept: lineages (Simpson, Wiley, De Quieroz)
- e. "ecological" species concept: niches (Van Valen)
- f. "species as individual": integrated, cohesive units with spatio-temporal boundaries (Ghiselin, Hull)
- g. "phylogenetic" species concept: basal monophyletic groups (more below)

B. Reason for the existence of a species problem:

- a. most of the above concepts and criteria conflict in most real cases -- different concepts (and processes) "pick out" different groups in each particular case, thus the implied correspondence between different criteria relied on by the BSC (and phenetic concepts) is abundantly falsified.
- b. operationality -- how to apply various concepts in a practical sense.
- c. what causes integration/cohesion of species? -- concerns:
 - breeding relationships are often clinal and/or non-transitive (what does "potential" interbreeding mean?)
 - gene flow is often very limited or lacking (what causes the evident distinctness of many asexual species?)
 - what is a niche?
 - developmental constraints (phylogenetic inertia)?

C. A Phylogenetic Solution:

Recognize that there is no species problem per se in systematics. Rather, there is a taxon

problem. Once one has decided what taxon names are to represent in general, then species taxa should be the same kind of things -- just the least inclusive. There is an element of arbitrariness to the formal Linnaean nomenclatorial system. Evolution is real, as are organisms (physiological units), lineages (phylogenetic units), and demes (interbreeding units) for example. On the other hand, our classification systems are obviously human constructs, meant to serve certain purposes of our own: communication, data storage and retrieval, predictivity. These purposes are best served by classification systems that reflect our best understanding of natural processes of evolution, and the field of systematics in general has settled on restricting the use of formal taxonomic names to represent phylogenetically natural, monophyletic groups.

Grouping vs. ranking. There are two necessary parts to any species definition. The criteria by which organisms are grouped into taxa must be specified, as well as the criteria by which a taxon is ranked as a species rather than some other hierarchical level. Following the arguments given previously supporting a Hennigian phylogenetic system of classification, the **grouping** criterion that should be used is monophyly. Under this view, apomorphies are considered to be the necessary empirical evidence for unambiguous phylogenetic species, as for phylogenetic taxa at all levels.

There are difficulties applying the concept of monophyly at this level. As you consider less inclusive levels in the genealogical hierarchy there is an increasing probability that reticulating ("hybridizing") events will occur, rather than the diverging phylogenetic relationships assumed by the cladistic approach. However, the problem of reticulation is not specific to the species level; indeed reticulation can occur throughout the hierarchy of life, and so is one of more general difficulty, and one that is receiving a lot of attention in the professional literature. It is becoming clear that while a certain amount of reticulation does not preclude cladistic reconstructions of phylogeny, extensive reticulation can cause major problems (more in a later lecture).

Note in passing that reproductive criteria cannot be used to **group** organisms into phylogenetic species. The fundamental inappropriateness of using breeding compatibility in cladistic analysis is because the ability to interbreed (potential or actual), is a plesiomorphy by definition, thus not a phylogenetically valid grouping criterion.

The **ranking** decision should involve practical criteria such as the amount of character support for a group and may also involve biological criteria in better known organisms, including reproductive criteria, e.g., the origin of a distinctive mating system at a particular node or the acquisition of **exclusivity** (a condition in which each allele in a lineage is more closely related to another allele in the lineage than it is to an allele *outside* the lineage). This ranking decision is forced because systematists have legislatively constrained themselves to use a ranked Linnaean hierarchy. A larger issue are recent calls for reforming the Linnaean system to remove the concept of ranks. This move would keep the hierarchy of named phylogenetic groups, but remove the ranks (including species) associated with the names (more in a later lecture). This move would decrease the arbitrariness of ranking decisions at the "species level," and will probably happen some day, but for now we assume that the current Linnaean system of ranked classifications is to remain in place.

To summarize, a phylogenetic species concept can be defined based on these considerations. First, organisms should be grouped into species on the basis of evidence for monophyly, as at all taxonomic levels; breeding criteria in particular have no business being used for grouping purposes. Second, ranking criteria used to assign species rank to certain monophyletic groups must vary among different organisms, but might well include ecological criteria or presence of breeding barriers in particular cases. See Mishler & Brandon (1987).

The Phylogenetic Species Concept:

A species is the least inclusive taxon recognized in a formal phylogenetic classification. As with all hierarchical levels of taxa in such a classification, organisms are grouped into species because of evidence of monophyly. Taxa are ranked as species rather than at some higher level because they are the smallest monophyletic groups deemed worthy of formal recognition, because of the

amount of support for their monophyly and/or their importance in biological processes operating on the lineage in question.

Some elaboration of the term monophyly from this definition is needed. Monophyly is here defined synchronically to be: *all and only descendants of a common ancestor, existing in any one slice in time*. This ancestor was not an ancestral species, but rather a less inclusive entity such as an organism, kin group or population. The synchronic approach is necessary to avoid the time paradoxes that arise when classifying ancestors with descendants [i.e., questions like: Was your grandmother your grandmother before your parents were born?]. The evidence required for a hypothesis of monophyly is primarily corroborated patterns of synapomorphy (but possibly also including other factors, such as geography).

VIII. Wave of the future: phylogenetic taxonomy

A number of recent calls have been made for the reformation of the Linnaean hierarchy (e.g., De Queiroz & Gauthier, 1992). These authors have emphasized that the existing system is based in a non-evolutionary world-view; the roots of the Linnaean hierarchy are a specially-created world-view. Perhaps the idea of fixed ranks made some sense under that view, but under an evolutionary world view they don't make sense. Most aspects of the current code, including priority, revolve around the ranks, which leads to instability of usage. For example, when a change in relationships is discovered, several names often need to be changed to adjust, including those of groups whose circumscription has not changed. Frivolous changes in names often occur when authors merely change the rank of a group without any change in postulated relationships. While practicing systematists know that groups given the same rank across biology are not comparable in any way (i.e., in age, size, amount of divergence, diversity within, etc.), many users do not know this. For example, ecologists and macroevolutionists often count numbers of taxa at a particular rank as an erroneous measure of "biodiversity." The non-equivalence of ranks means that at best (to those who are knowledgeable) they are a meaningless formality and perhaps not more than a hindrance. At worst, in the hands of a user of classifications who naively assumes groups at the same rank are comparable in some way, formal ranks lead to bad science.

It is not completely clear at this point how exactly a new code of nomenclature should be written (see draft at: <http://www.ohiou.edu/phylocode/index.html>), but the basics are clear. Such a new code should maintain the principle of priority (the first name for a lineage should be followed) and other aspects of the current code that promote effective communication of new names to the community. However, the major change would be that the Linnaean ranks should be abandoned, for efficient and accurate representation of phylogenetic relationships. Instead, names of clades should be hierarchically nested uninomials regarded as proper names. A clade would retain its name regardless of where new knowledge might change its phylogenetic position, thus increasing nomenclatorial stability. Furthermore, since clade names would be presented to the community without attached ranks, users would be encouraged to look at the actual attributes of the clades they compare, thus improving research in comparative biology.

What about the species rank? The most controversial level of all, but the above arguments can be applied to the species rank just as well as to higher levels. In my opinion, this rank need to be eliminated as well. Below is how it could be done in practice; we will keep returning to this subject later in the semester, in terms of its implications.

How could rank-free classification be applied to terminal taxa?

- Names of clades (including the terminal level), should be hierarchically nested uninomials regarded as proper names (current usage should be followed as much as possible to retain links to the literature).
- As at all taxonomic levels, use node-based names with two or more internal type specimens.
- All the higher clades to which a taxon belongs should be regarded as part of its complete name.
- Therefore each clade has a uninomial given name, but also a set of more and more inclusive "family names". Homonyms can thus be told apart by higher level clade names.

IX. Classification Do's and Don'ts

- Don't use taxonomic ranks, including species, for comparative purposes or ecological/evolutionary inferences.
- Do use trees, branching events (represented by monophyletic taxa), and branch lengths for comparative purposes and ecological/evolutionary inferences.