"Introduction to Phylogenetic Classifications"

Classifications are required to provide a context for biodiversity. Thus given the importance of taxonomic classification, the criteria for these classifications should rest on the highest quality of data available, and if these data change the classification should be adjusted accordingly. We expect nothing less of the models used in quantum physics, molecular biology, health sciences, or engineering, and we should expect nothing less for biodiversity estimates and comparisons.

There are multiple ways in which a group of organisms can be classified, and some are better than others are. Color is usually available and could produce obvious and repeatable groupings. However, it might be a little dangerous to make conservation decisions for cave faunas based on such a classification. The best criteria by which to classify organisms are the principles of the process that produces biodiversity in the first place – classifications based on the phylogenetic relationships of the organisms. Haven't we been doing this? No. "Name-ably distinct", criterion of "so different", overall similarity, etc.; all of these proxies.

In what other scientific or technological undertaking do we make decisions based on data categorized by proxies when superior data is readily attainable, testable, and is predictive? Analogous approximations are not acceptable in the launching of the space shuttle, PCR, the value of π, quantum physics, or even the preparation of a soufflé.

Classification versus taxonomy versus systematics?

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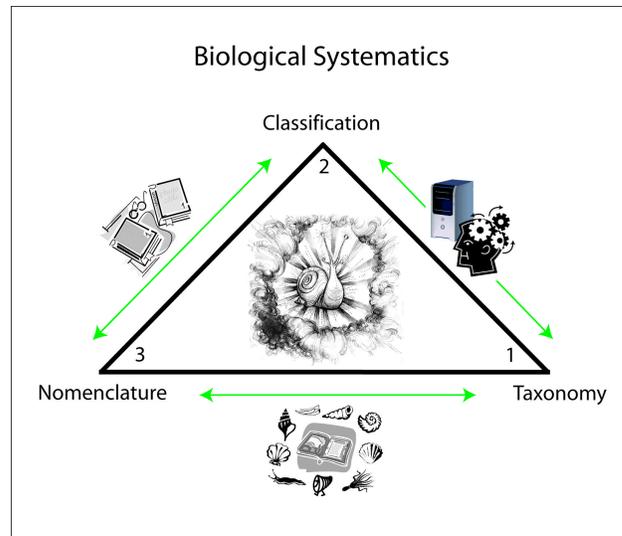
Figure 1. Biological Systematics a summary showing the relationship between taxonomy (1), classification (2), and nomenclature (3).

"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.

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 scientic 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.

Obviously, there are alternative ways to classify 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, relationships are 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.

Purposes for classification:

1) Practicality - operationality, ease, stability.

2) Information Content - Optimal summarization of what is known about entities.

3) Predictivity - unknown features of other entities

4) Function in Theories - "Capture" entities acting in, or resulting from, natural processes

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?

3. Clade versus Lineage
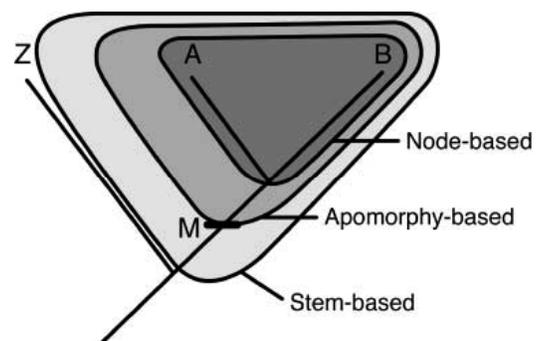
- Clade: a synchronic, monophyletic set of lineage-representatives (i.e., Hennig's semaphoronts)
- Monophyletic: all and only descendants of a common ancestor
- Lineage: a diachronic ancestor-descendant connection ("Species" in the de Queiroz sense)
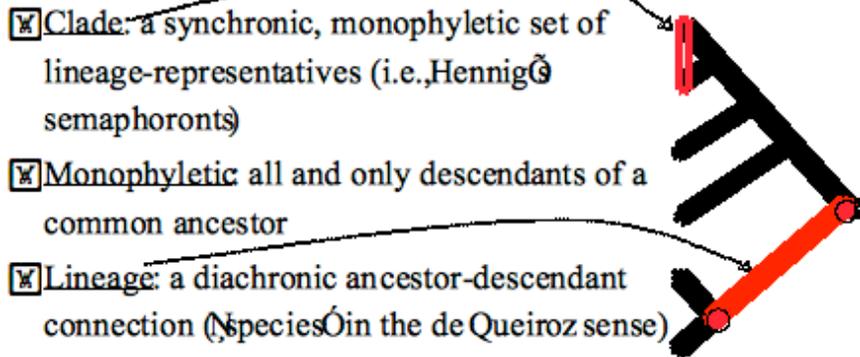
Four logical phases of phylogenetics:

(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.
➔ comparative methods applied

(4) *Nomenclature*

➔ 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.