Feb 28th**.  Classification 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)

**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).

**Natural classification**
     All three schools of systematics want to produce "natural" classifications, but disagree on what counts as "natural":
1. *Pheneticists* view natural groups as those taxa linked by the greatest similarity to each other; characters not weighted.  Note that "phenetic" vs. "phylogenetic" is not the same distinction as "phenotypic" vs. "genetic" (an unfortunate similarity in sound).
2. To *evolutionary systematists*, "natural" groups are defined by gaps between taxa in characters for which an evolutionary scenario can be argued.  Characters are weighted by adaptive significance.
3. *Cladists* consider natural groups to be monophyletic, and thus "natural" classifications reflect the tree of life.  Note that "cladist" does not equal "parsimony advocate" -- the former term reflects a philosophy of classification, the latter term reflects a preference for one of several related character-based approaches to tree-building (to be discussed later), all of which yield cladograms.

**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.

As discussed earlier, the German entomologist Willi Hennig codified the meaning of these evolutionary outcomes for systematics, in what has been called the Hennig Principle (Hennig, 1965, 1966).  Hennig's seminal contribution was to note that in a system evolving via descent with modification and splitting of lineages, characters that changed state along a particular lineage can serve to indicate the prior existence of that lineage, even after further splitting occurs.  The "Hennig Principle" follows from this: homologous similarities among organisms come in two basic kinds, synapomorphies due to immediate shared ancestry (i.e., a common ancestor at a specific phylogenetic level), and symplesiomorphies due to more distant ancestry.  Only the former are useful for reconstructing the relative order of branching events in phylogeny.   A corollary of the Hennig Principle is that classification should reflect reconstructed branching order; only monophyletic groups should be formally named. 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.


**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?
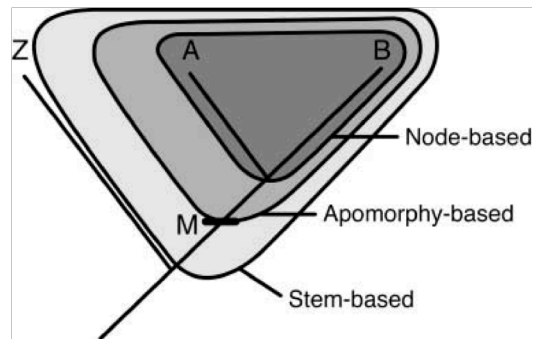
3. Clade versus Lineage (see next page for illustration).  They are not the same thing -- "clade" is a synchronic concept, a snapshot of a lineage -- while a lineage is a diachronic concept, a series of replicators.

- <u>Clade</u>: a synchronic, monophyletic set of lineage-representatives (i.e., Hennig's semaphoronts)

- <u>Monophyletic</u>: all and only descendants of a common ancestor

- <u>Lineage</u>: a diachronic ancestor-descendant connection ("species" in the de Queiroz sense)
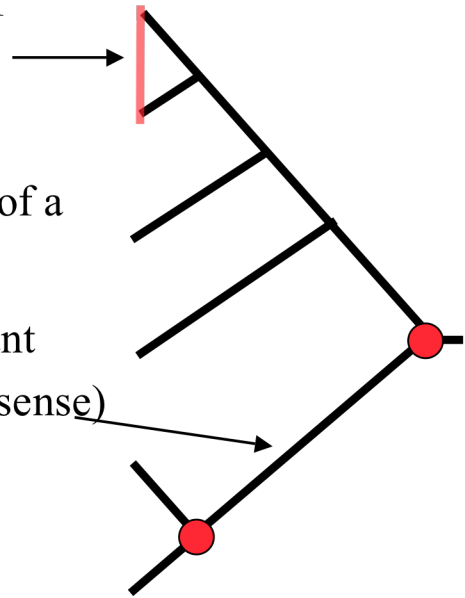
**Three major logical phases of a phylogenetic analysis:**

(1) *Pre-cladistic phase* (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) *Cladistic analysis* (analytic, deductive)
➔ the data matrix is translated into a cladogram using a parsimony/ ML model (more in later lectures on unweighted or weighted parsimony in relationship to maximum likelihood).
➔ 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 studies*
➔ 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.