

March 11, 2008: **Species Concepts**

I. Importance of the species problem

- a. nomenclatorial requirements (all organisms must belong to a species).
- b. practicality -- need to organize diversity, summarize information, communicate, give names to things.
- c. legal issues -- endangered species legislation; conservation
- c. connection to evolutionary theory -- desire to have species as units functioning in process theories (but which process theories? how to connect units and theories without circularity?)
- d. philosophical concerns -- need to be clear about properties of units: sets vs. individuals, universalism vs. pluralism.

II. Quasi-historical outline:

- 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)
- 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 (will return later)

III. 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)?

d. what are the spatio-temporal boundaries of species?

- monophyly?

- origin?

- extinction?

IV. The winning argument for phylogenetic classification (review)

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).

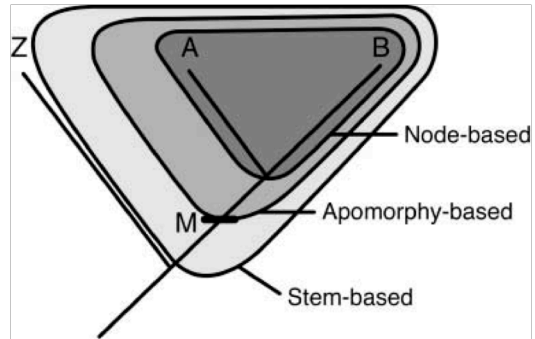
A summary of the arguments for why formal taxonomic names should be used solely to represent phylogenetic groups is as follows: evolution is the single most powerful and general process underlying biological diversity. The major outcome of the evolutionary process is the production of an ever-branching 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.

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 below 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.



- 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)

V. A phylogenetic solution under the current codes of nomenclature

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. The field of systematics in general has settled on restricting the use of formal taxonomic names to represent phylogenetically natural, monophyletic groups. So "species" should be the smallest one of those

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. How do we know when we have reticulation going on? We can only detect it by doing phylogenetics.

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. 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 (more on the possibility of getting rid of species in a couple weeks).

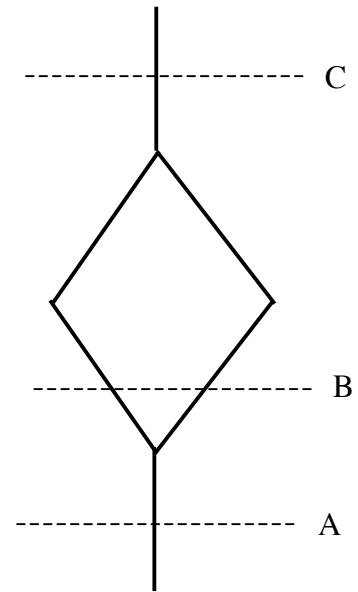
To summarize, a phylogenetic species concept can be defined based on these considerations (from B.D. Mishler and E. Theriot. 2000. In Q.D. Wheeler & R. Meier (eds.), *Species Concepts and Phylogenetic Theory: A Debate*. Columbia University Press). First, organisms should be grouped into species on the 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. The grouping concept (monophyly) in this view is monistic, while the ranking concept is pluralistic:

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.

[Does the ranking part of this sound arbitrary?? It is -- see below.]

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. Only "clades" can be monophyletic, not "lineages" (see diagram from last time). 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).

Can a group arising by hybridization be monophyletic? Given this synchronic definition it can, as long as it only arose once. In the diagram at right, in the first time-slice (A) there is one clade, (B) then two, then (C) one again.



VI. The winning argument for rank-free classification (review)

As discussed last week, 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. There are several problems with the current nomenclatorial system:

1. The current system, with its single type for a name, cannot be used to precisely name a clade. E.g., you may name a family based on a certain type specimen, and even if you were clear about what node you meant to name in your original publication, the exact phylogenetic application of your name would not be clear subsequently, after new clades are added to the tree.
2. There are not nearly enough ranks to name the thousands of levels of monophyletic groups in the tree of life. Therefore people are increasingly using informal rank-free names for higher-level nodes, but without any clear, formal specification of what clade is meant.
3. 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. E.g., when it was detected that the Cactaceae is nested inside of the Portulacaceae, one of these well-known family names has to be abandoned. Frivolous changes in names often occur when authors merely change the rank of a group without any change in postulated relationships.
4. 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 -- removing the ranks would serve the same purpose as child-proof door locks for the back seat of your car!

It is not completely clear at this point how exactly a new phylogenetic code of nomenclature should be written (see draft of the Phylocode and other materials 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, two major changes would be made:

1. The Linnaean ranks would be abandoned, Instead, names of clades would 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.
2. Two or more types (called "specifiers") would be used, for efficient and accurate representation of phylogenetic relationships. There are two types of specifiers, internal and external. Two or more of the former are used to name node-based groups. two or more of the former and one of the latter are used to name stem-based groups. Apomorphy-based names are controversial, and really shouldn't be used.

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

At the moment, the species rank is the most controversial topic among Phylocode advocates. Some (primarily zoologists) want to retain it as one fundamental rank in an otherwise rank-free system, others (primarily botanists) want to get rid of it.

1. Names of clades (including the terminal level), could be regarded be hierarchically-nested uninomials regarded as proper names (as at all levels in the Phylocode, current usage should be followed as much as possible to retain links to the literature, databases, and collections).
2. As at all taxonomic levels, recommended to use node-based names with two or more internal type specimens.
3. To deal with the rampant homonyms that will result from treating current species epithets as uninomials, several proposals have been made. My preference would be to regard all the higher clades to which a taxon belongs as part of its complete name. Thereby each clade would have a uninomial given name, but also a set of more and more inclusive "family names, " thus homonyms could be told apart by higher-level clade names.

Much of this taken from:

Mishler, B. D. 1999. Getting rid of species? Pp.307-315 in R. Wilson (ed.), Species: new interdisciplinary essays. MIT Press.

Mishler, B. D. and E. Theriot. 2000. The phylogenetic species concept sensu Mishler and Theriot: monophyly, apomorphy, and phylogenetic species concepts. Pp.44-54, 119-132, 179-184 in Q.D. Wheeler & R. Meier (eds.), Species concepts and phylogenetic theory: a debate. Columbia University Press, NY. [three chapters]

Three major logical phases of a phylogenetic analysis:

(1) *Character analysis 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 these methods).

→ 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.