

**Jan. 29, 2008.** The Hennig Principle: Homology, Synapomorphy, Rooting issues

The fundamental idea that has driven recent advances in phylogenetics is known as the Hennig Principle, and is as elegant and fundamental in its way as was Darwin's principle of natural selection. It is indeed simple, yet profound in its implications. It is based on the idea of homology, one of the most important concepts in systematics, but also one of the most controversial. What does it mean to say that two organisms share the same characteristic? The modern concept is based on evidence for historical continuity of information; homology would then be defined as a feature shared by two organisms because of descent from a common ancestor that had that feature.

*1. The Hennig Principle*

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 (fig. 1, below). Only the former are useful for reconstructing the relative order of branching events in phylogeny -- "special similarities" (synapomorphies) are the key to reconstructing truly natural relationships of organisms, rather than overall similarity (which is an incoherent mixture of synapomorphy, symplesiomorphy, and non-homology).

We must pay close attention to both ontology and epistemology, and the feedback relationship between the two: A given method makes sense only if the world really is a certain way, yet the view we have of how the world is organized is dependent on the methods we have used. For example, if species on earth are related genealogically and evolution is mainly by descent with modification (in a primarily diverging mode), then the Hennig Principle is the best method for reconstructing the history of life. Yet, the discovery of hierarchically nested characters is the best evidence we have on how evolution has occurred.

*2. Homology, or When are two things the same?*

These concerns are relevant to characters as well; the mere act of stating that two things are the same, or parts of two things are the same, is loaded with a (perhaps subconscious but nonetheless real) complex theoretical framework.

"Homology" -- One of the most important concepts in systematics, but also one of the most controversial.

- classes versus individuals (type/token).
- classical, pre-evolutionary views (Cuvier, Owen)
- nominalistic views (many botanists, pheneticists)
- the need for ontology.
- evolutionary views: historical connectedness.
- synapomorphy (Patterson, Stevens)
- historical continuity of information (Van Valen, Roth)\*\*

Our Ontology:

- Phylogenetic Homology (between organisms)
- Taxic (= synapomorphy)
- Transformational (plesiomorphy -> apomorphy)
- Iterative Homology (within one organism), e.g., Serial Homology or Paralogy (in molecular data)

Epistemology:

How do we recognize homology?

Best early codification of recognition criteria was that of Remane (See Wiley, 1981, pp. 130-158):

- position
- quality of resemblance
- continuance through intermediate forms

Also, an important contribution of the cladists has been the explicit formulation of a phylogenetic criterion:

**\*\* a hypothesis of taxic homology of necessity is also a hypothesis for the existence of a monophyletic group \*\***

--Therefore, congruence among all postulated homologies provides a test of any single character in question [the central epistemological advance].

--Is this circular? A digression into general concerns in the philosophy of science; reciprocal illumination.

Patterson formulated these epistemological concerns in his three tests of homology:

- similarity
- conjunction
- congruence

Alignment of DNA. A certain prominent paleontologist who was at Harvard (who shall remain nameless, but his initials are S.J.G.) once proclaimed that comparisons of DNA have "solved the problem of homology." Is this true? The very simplicity of molecular characters (i.e., no ontogeny, few possible character states) actually leads to special problems with determining homology! We have essentially a one-dimensional string, although we may also have some additional dimensions added by structural constraints.

Some epistemological concerns:

- mismatches vs. gaps -- gap penalty?

Example: TCAGACGATTG Which is the best alignment?

TCGGAGCTG [One approach:  $D = y + wz$  ]

( I ) TCAG-ACG-ATTG  
TC-GGA-GC-T-G

( II ) TCAGACGATTG  
TCGGAGCTG--

( III ) TCAG-ACGATTG  
TC-GGA-GCTG-

### 3. *Brief introduction to parsimony*

In the Hennigian system, individual hypotheses of putative homology are built up on a character-by-character basis, then a congruence test (using a parsimony principle) is applied to identify homoplasies (i.e., apparent homologies that are not congruent with the plurality of characters). The fundamental assumptions of a parsimony-based approach to phylogenetic reconstruction are shown in the figure on the last page, one that we will return to at several points later on in the course when we get into tree building in a serious way.

The basic assumptions for this method are five: (1) Reproduction (replication in the sense of Hull, 1980) must be occurring, to form lineages (the diachronic ancestor-descendent relationship). (2) Heritability (in the population genetic sense) must obtain, wherein particular features to be used as historical markers (characters) have discrete variants (character states empirically, transformational homologs ontologically) that show a strong correlation between parent and offspring. (3) Divergence (branching of lineages) must occur predominantly, as compared to reticulation, giving rise to patterns of taxic homologs shared among sister groups (the synchronic monophyly relationship). (4) Independence must occur among different characters; that is, no process (e.g., natural selection, gene conversion, developmental constraints) is operating to produce character correlations except for common history. (5) Transformation in particular characters must occur at a relatively low rate ( $\lambda$  = the expected number of characters changes on a given branch, more in a later lecture).

These five basic criteria amount to a joint assumption that an apparent homology [N.B., this a feature that has already passed strict observational and experimental tests of detailed similarity, heritability, and independence to be discussed next week] is more likely to be due to true taxic homology than to homoplasy, unless evidence to the contrary exists, i.e., a majority of apparent taxic homologies showing a different pattern. This is Hennig's auxiliary principle.

Why is there a scientific preference for simplicity? Consider Ockham's Razor (meant to be an epistemological tool, not ontological!), a general concern in all systematic schools of thought and indeed in all scientific inference (see Sober, 1988). There is a universal need in scientific inference to minimize ad hoc hypotheses (e.g., lost car keys; solution of a murder mystery).

#### *Example showing the different between parsimony and distance*

Examine the data matrix on the next page (Box 3.4, from Lundberg & McDade, 1990), and be sure you can see why those data support the cladogram shown. To see the effect of homoplasy, consider a new character 13, with the distribution 10010. Note that the pattern of overall similarity (a "phenogram") would give a different result, and group E with A and B rather than with C and D (this will be demonstrated on the board). Why would using overall similarity give you the wrong tree, even when the data fit the tree perfectly? That is the big question, and one of the main bones of contention in the "systematics wars" that Hull (1988) chronicles. If all methods gave the same result, then there would be no point to a controversy over methods. But, they don't...

### 4. *Polarization and rooting*

In order to apply this distinction between an apomorphic state and a plesiomorphic state, it is necessary to polarize a character. Three primary criteria have been used to do this before an analysis:

- (1) Stratigraphy (the state seen earliest in the fossil record is plesiomorphic).
- (2) Ontogeny (the state seen first in development is plesiomorphic).

(3) Outgroup comparison (the state seen outside the study group is plesiomorphic).

An alternative, commonly applied approach to polarizing the characters before an analysis is to first construct the topology of the tree as an unrooted network, and then "pull it down" into a tree shape in one of two ways:

- (1) By bending at the point where the outgroup joins the ingroup: Outgroup rooting.
- (2) By seeing where an ancestral vector of hypothesized character states would attach: Lundberg rooting.

The correct way to use these approaches will be briefly discussed here, along with some cautions, but we'll need to return to this important issue later in the semester.

### *5. Classification*

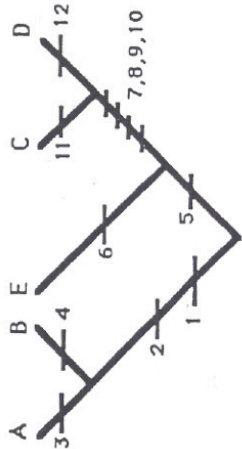
Finally in the Hennigian system, classifications are applied to the resulting branching diagram (cladogram). A corollary of the Hennig Principle is that classification should reflect reconstructed branching order; only monophyletic groups should be formally named. A strictly monophyletic group is one that contains all and only descendents of a common ancestor. A paraphyletic group is one that excludes some of the descendents of the common ancestor. See figure 1 for the distinction between these two types of groups. Again, we will return to deal with the ramifications of this approach to classification later in the course.

*Coda 1:* This elegant correspondence between synapomorphy, homology, and monophyly is the basis of the cladistic revolution in systematics.

*Coda 2:* The process of systematics can be defined most simply in three words: "Interpret homologies parsimoniously" (Patterson)

**Box 3.4 Character State Distribution**

The use of ancestral character states (symplesiomorphies) and uniquely derived character states (synapomorphies) to recognize monophyletic groups is illustrated by the phylogeny of species A-E, here diagrammed with 12 numbered character changes from ancestral to derived states.



This pattern of character evolution results in the taxonomic distribution of states given in the following character-by-taxon matrix; character states are ancestral (0) or derived (1).

Species	1	2	3	4	5	6	7	8	9	10	11	12
A	1	1	1	0	0	0	0	0	0	0	0	0
B	1	1	0	1	0	0	0	0	0	0	0	0
C	0	0	0	0	1	0	1	1	1	1	1	0
D	0	0	0	0	1	0	1	1	1	1	0	1
E	0	0	0	0	1	1	0	0	0	0	0	0

Together, the derived (1) states (synapomorphies) are hierarchically nested and characterize the nested monophyletic groups on the phylogenetic tree.

Characters	Groups sharing "1", state
1,2	A+B
3	A
4	B
5	C+D+E
6	E
7,8,9,10	C+D
11	C
12	D

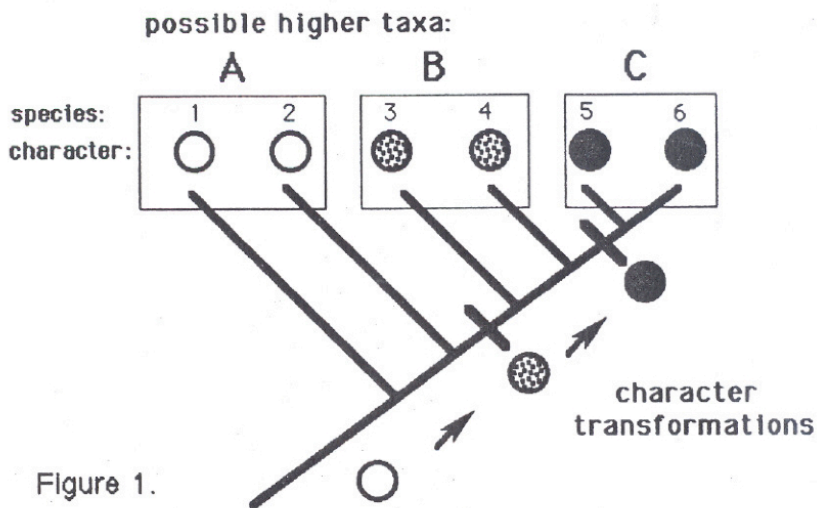


Figure 1.

## Phylogenetic Inference via Parsimony

### Assumptions:

- 1) Reproduction (replication)



-----> Genealogy

- 2) Heritability (variation, correlated between parent and offspring)

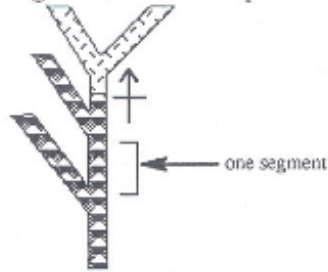


posterior condition

prior condition

-----> Homology  
[transformational]

- 3) Divergence (at least as a predominating pattern)



-----> Sister Groups  
[taxic]

- 4) Transformation to a particular posterior state occurs at a relatively low rate, as compared to Divergence.

Simulation studies would place the "safe" rate below  $\lambda = 0.1$  [i.e.,  $\approx$  one out of ten characters changing state on a particular segment], unless  $\lambda$ 's for different segments are similar (over).

This amounts to an assumption that apparent taxic homology is more likely to be due to true taxic homology than to homoplasy, unless evidence exist to the contrary (i.e. other apparent taxic homologies).