

Feb. 23, 2012. **Parsimony, Patterns & Processes**

1. Hennig and Parsimony: Hennig was not concerned with parsimony as an optimality criterion, but rather his general paradigm was consistent with parsimony as a guiding principle (e.g. Occam's Razor as a heuristic rule of thumb). The connection is in Hennig's Auxiliary Principle – *to assume homology if there is no evidence to suggest otherwise*. Hennig provided fundamental methods for the use of character data to form phylogenies and he made the relationship between character evidence and cladograms explicit in a way that had not been done previously, but he did not provide a clear method for choosing among competing alternatives.

Parsimony as used in phylogenetics is often defined as "*minimizing evolutionary changes*." In a sense this is correct, but it should not be construed to mean that one thinks evolution is parsimonious. If our character matrix consists of characters that have undergone rigorous character analysis to establish conjectural or primary homology, we then should seek hypotheses (trees) that maximizing our homologies. Conversely, we prefer trees that overturn as few as possible of our initial homologies, given that these initial hypotheses are the result of careful character analysis. The result is to *minimize ad hoc explanations* when we fail to get the primary homology right.

The two views that parsimony is "minimizing evolutionary changes" or "minimizing ad hoc explanations" is part of a larger tension between **pattern and process**.

Pattern cladists or transformed cladists are one extreme end of the spectrum. They put forward the idea that cladistic (in this case = strict parsimony) methods do not need, and in fact are better off without an evolutionary (process) justification. Three things are needed to justify building trees based on synapomorphies, 1. discoverability of characters, 2. hierarchy is the best representation of the natural world and 3. parsimony as an epistemological approach (Brower, A.V.Z. 2000. Evolution Is Not a Necessary Assumption of Cladistics. *Cladistics* 16, 143–154.). Also part of the pattern v. process debates was the accusation of circularity, e.g. Mitter (1981. "Cladistics" in botany. *Syst. Zool.*, 30:373–376.) "there is widespread (but not universal) agreement that ... systematic methods should be as free as possible from assumptions about how evolution works, because these assumptions are in general not testable without reference to systematic results." Much debate exists in the literature in regard to parsimony. Is it assumption free, assumption minimizing or just a case where assumptions are ignored?

The clear understanding of patterns we observe and summarize, as apposed to processes that explain such patterns is important (e.g. *multiple substitutions and insertions/deletions are inferred events, all processes, not observations*). However, a strict pattern view, which denies a role for evolution, does not provide a good explanation as to why any given character should or should not be included in an analysis.

For "reconstructionists" evolutionary models are definitely part of the character analysis and should be used (with caution) in tree building.

In the "estimation school" maximum parsimony (MP) is a just one model of evolution and many would say a rather poor one. Maximum likelihood (ML), and methods of that sort, benefit from an evolutionary model based justification. Exactly which model should be used, and where the values for parameters in the model should come from, are issues and this problem is

addressed through assessment of the impact of model violations and statistical support. More on this topic in the next couple of weeks.

Parsimony and likelihood are best viewed as belonging to a family of methods. They are character-matrix based, using information about individual hypotheses of homology, unlike the distance methods we talked about last time. The connection between parsimony and likelihood is shown clearly in the case of the “no common mechanism” model (Penny et. al. 1994, Tuffley and Steel 1997). This model loosens the assumptions of rate change so that there is potentially a different rate for every combination of branch and character across the tree, which comes back to the parsimony model. Essentially parsimony has both the property of being the simplest model (straightforward summing changes of observed states) and the most complex model (each character assumed to have its own rate), thus the relationship between MP and ML is more like a circle than a spectrum.

2. Consistency and long branches: Methods are said to be consistent if they converge on the true value as more data are used. Tree building methods can be inconsistent (“positively misleading”) – converging on the wrong tree as more data with the same properties are added. Parsimony methods can be inconsistent (“positively misleading”) in the “Felsenstein zone” as discussed last week. Other methods that account for unobserved changes (e.g. maximum likelihood) can be consistent in the Felsenstein zone IF the evolutionary process is adequately by the assumed model. If the model is wrong, ML methods can be inconsistent too.

How bad is this problem?

- All methods fail in some cases so the phenomenon is not exclusively a parsimony issue.
- If the rate of change is relatively slow as compared to the branching events in the phylogeny it isn't a problem.
- If the relative branch lengths are reasonably similar it isn't a problem.
- If the “true” relationships are that the long branches are in fact sister taxa, it isn't a problem for reconstructing the topology, but will artificially increase support for that relationship.

The “Zone” probably isn't that much of a problem for most carefully selected data. But when it is thought to be an issue you may...

- Add taxa that break up long branches
- Add *different* data with an appropriate rate of change for the problem being addressed

3. Parsimony as an optimality criterion.

Minimize the total number of steps over the tree where there is a set of parsimonious solutions.

How do we measure steps (length)? ---- A character has a length that is the number of independent origins of character states on any given cladogram. This is measured as steps or costs and is weighted depending on the model assumptions.

- We know that there is character conflict, i.e. character state distributions support groups that are not compatible.

Two kinds of equally parsimonious trees: 1. same topology but different character state distributions (optimization) 2. Different topologies.

Optimizations: ACCTRAN and DELTRAN. Two “extremes” of optimization that may alter the resolution of the tree and the implied transformational history of the character.

Character weighting vs. character-state weighting

The step matrix

A priori weight.

Equal weights is viewed by some as a minimal necessary assumption. For others differential weighting, or the *a priori* selection of a weight difference, introduces an unacceptable subjectivity. Exactly what the difference in weight should be cannot be unambiguously answered.

Others argue that equal weight for all characters and character state changes is a whopper of an assumption. Imagine 100 columns of DNA data 40 columns have only A and G, 40 columns have only C and T and 20 are mixed. The observation suggests that transitions occur more frequently than transversions. As such a 1:1 weighting is more of an assumption than 2:1 or 4:1. According to some, at least for molecular data, a slight weighting in the right direction is much better than none at all.

A posteriori weighting:

1. Character consistency weighting. (aka, successive approximations weighting, Farris 1969)
2. Implied weighting (Goloboff 1993)

4. Tree Searching

-We know that there are a huge number of possible cladograms for any modest number of OTUs.

Number of OTUs-	Number of rooted, resolved trees-
2	1
3	3
4	15
5	105
6	945
7	10395
10	34459455
20	$\sim 10^{21}$

This is a proven NP complete problem.

Strategies typically used to find most parsimonious trees (MPTs):

1. Enumeration- look at every possible cladogram and sum length of all characters. Of course, this is not practical for any significant matrix. *Thus we need heuristic methods and methods to escape local optima...*
2. Basic strategy of a search
 - a. Get a near optimal tree
 - i. This is often done by randomly putting three OTUs in a network and adding a forth to the edge that creates the shortest four OTUs network and so on until all OUTs are joined. This does not guarantee a shortest tree, but it usually is not too bad for a start.