Integrative Biology 200A "PRINCIPLES OF PHYLOGENETICS" Spring 2006
University of California, Berkeley Kipling Will- 11 Apr 2006

Phylogenetic tree IV- Data/Hypothesis Exploration and Support Measures

I. Overview.
-- The truest test involves empirical tests that examine all critical evidence. For phylogenetic hypotheses (tree, branching pattern, branch lengths, character state distribution), this involves the addition more characters and taxa. This is not always reasonable/feasible and when do we have enough anyway? This is an issue of philosophical or statistical confidence.
-- The simplest form of confidence, which is somewhat subjective, is to show character state changes on the cladogram. Groups supported by more, less homplastic and more complex character state changes are thought to be well supported. When more of our initial statements of homology survive and are compatible we have increased confidence in the hypothesis. Alas, this may be suitable for morphological data, however, it is very difficult to apply given the simplicity of DNA sequence.
-- A general or specific "fit" to external data (e.g. biogeographic patterns) also builds confidence. However, it is generally more narrative and subjective, making it hard to evaluate if you are not actively working within the system.
-- It is necessary to express some sort of confidence or make a statement of reliability in order to give *others* a sense of how well your data fit your hypothesis and to what degree the critical evidence refutes competing hypotheses even if we are confident in the result.
-- Many exploration methods seek some sort of statistical reliability or measure to give a notion of how bold or conservative we should be in regard to conclusions based on the phylogenetic pattern. The fact that a nearby sub-optimal solutions exist is not enough to cause us to move from one hypothesis to another.
-- Although there is a general notion that we are identifying well supported clades, exploration methods and support measures are really just as (more?) important for pointing to poorly supported parts of the tree. Poorly supported groups suggest where future efforts need to be applied.
-- Most statistical methods require some assumption of a universe from which the sample is drawn. Generally this is random sample of the universe of possible independent entities, i.e. they are independent and identically distributed (i.i.d).

II. Sensitivity and Resampling Analyses: Various heuristic explore how robust the hypothesis is likely to be if the underlying assumptions are wrong or expressed as some sort of "support".

A. Assumption sensitivity analyses:
HOW: Assumptions (= parameters) are varied in multiple analyses and the results compared in some way.

WHY: To look for (in)sensitivity to variation in model assumption (e.g. weights assigned to transitions/transversions changes topology). This has been used as an optimality criterion for deciding if a group should be accepted or rejected. Groups sensitive to variation are rejected. Also used as a means to select a set of alignment parameters.

WHAT IT TELLS US: Not truly a test of monophyly or support. Monophyly is tested in the corroboration of empirical evidence in light of some set of "valid" assumption.

It doesn't really test the support offered by the data. It does show which groups remain under a set of "reasonable" parameters and support is drawn from a variety of synapomorphy classes. Almost any topology can be supported under some set of parameters. Can't distinguish levels or different kinds of support and a group well supported under "mutation-parsimony" may be lacking under "in/del-parsimony"…….. So what?

((a,b) [10,10,10] -- ((c,d) [1,1,1] -- (e,f) [30,0,0]))

B. Bremer Support / Decay analyses or "index"[not really a mathematical index]
HOW: Record the number of extra steps required to loose a clade that is found in the most parsimonious tree. Any clade not found in the strict consensus of all MPTs has a Bremer support value of 0. Any clade not found in the strict

consensus of all trees one step longer than the MTPs has a Bremer support value of 1,2,3... until a shortest tree that does not contain any clade is found.

In reality this includes too many trees and a Bremer value is an *estimate* based on *heuristic* searches of suboptimal tree space.

WHY: To give a measure of decisiveness and indicate ambiguously supported nodes directly from the data.

WHAT IT TELLS US: An estimate of the degree to which the optimal solution is preferred to alternatives. As a heuristic it points to poorly supported groups that may have few synapomorphies or may be supported by conflicting characters. However, it does not discriminate between different types of support and does not have a clear statistical interpretation.

Matrix w/100 characters and MPTs of 200 steps, each character optimizes for 2 steps. Trees 2 steps longer (202 steps) could come by increasing one character to 4 steps [99*2 +1*4= 202] OR reducing 49 characters to 1 step and increasing 51 to 3 steps [49*1 + 51*3= 202].

[Technical note: Paup doesn't calculate Bremer support directly, so use MacClade to make a command file for Paup to read, this will generate Bremer numbers based on a set of constraint clade analyses (see MacClade manual). You can also use the program TreeRot with Paup. Bremer support can be directly calculated by Nona, however, it is VERY dependent on the search parameters and memory limitations. I suggest Paup for this one. As a rule of thumb, a score of 3 is good and 5 highly "supported". I have not done this with TNT. If you do let me know how it goes.]

C. Methodological concordance:

HOW: Multiple methods of phylogenetic analysis are used and the clades found in common are presumed well supported.

WHY: Controversy over methods and assumption can be avoided by a pluralistic approach that leads to reasonable results. Accurate methods will converge on the "truth" and a lack of agreement between methods indicates that none are recovering the true tree (Kim 1993).

WHAT IT TELLS US: Which clades are not affected by the assumptions and philosophical underpinnings of the methods that were used for analysis of the data. Since various methods address the problem from very different statistical and philosophical views, the fact that they converge may say something about the data or the methods but may have little to do with discovery of correct groups. There is no clear connection between convergence of methods and "accuracy."

D. Bootstrap/Jackknife

HOW: Resample with replacement (bootstrap) or without replacement (Jackknife) from your matrix. Essentially Bootstrap differentially weights some characters to build a matrix of equal size. Jackknife reduced some characters to weight of zero. For either method, the resulting matrix is used to build a set of trees. This is repeated many times to build a cloud of trees. A majority rule consensus tree for groups found in >50% of the trees is used to show well-supported groups.

WHY: To empirically estimate the variability. In phylogenetics used to assess uncertainty in the proposed phylogeny. These methods are usually applied to characters but also have been used to resample taxa.
The issue of independence of sampled elements is debatable, but generally in statistics these are only used for random samples that are independent. For taxa, most people agree that because they are more or less phylogenetically related they do not represent i.i.d. samples. However, some (e.g. Felsenstein) maintain that characters are less likely to be non-independent than taxa (an assumption made most of the time) or this can be corrected for. Some maintain that a sample of characters in a matrix is not drawn from an i.i.d. of all possible characters and this invalidates the method for phylogenetic characters. Other say that the sample need only be drawn independently from "some" universe of characters. But if the Bootstrap tree is different than the sample universe of empirical data, which we have, it must be a poor estimator. Characters that are not parsimony informative are potentially problematic. Bootstrap has been shown to be positively correlated to number of informative characters (parsimony); negatively correlated to number of taxa in analysis, number of taxa in a clade and tree asymmetry

(Siddall 2002). Also there is autocorrelation of nested clades (e.g. clade (D,E,) and supporting characters are not independent of (C(D,E)).)

WHAT IT TELLS US: Re-sampling biased data would only lead to an assessment of the accuracy of the bias. At best, as a heuristic it points to poorly supported groups that may have few synapomorphies or may be supported by conflicting characters.

E. PTP, etc.

HOW: Character state data is randomly and independently reshuffled among taxa, optimal trees are found for each permutation and compared to establish confidence limits, e.g. 95%. Either tree length (permutation tail probability- PTP) may be used or the clades (topological dependent permutation tail probability- T-PTP) may be compared.

WHY: To place confidence limits on the clades relative to Type-1 error (errors resulting from wrongly rejecting the null hypothesis = no structure.)

WHAT IT TELLS US: If the optimal score for the original data is far out in the distribution tail then significant, non-random structure is present in the data. However, PTP can show significant support for a group that has none in the original data. A single resolved node (either an internal polytomy or a pair of very close species) may give a significant result for an otherwise unstructured data set. Similarly the TPTP has a null hypothesis that there is no structure in the matrix anywhere, so it is likely to reject the null too easily.

III. Comparison/description

A. Skewness

HOW: Look at the number of changes on all possible topologies (actually a random sample). If there are a few trees of much lower score they will negatively skew the distribution.

WHY: Strongly skewed distribution suggest "strength" of the phylogenetic signal or decisiveness in the matrix.

WHAT IT TELLS US: Hard to tell. A number of published examples show it may fail to reflect phylogenetic structure and it is influenced by the central mass of the distribution more than the tail, influenced by character state distribution and requires arbitrarily resolved polytomies. Nevertheless, it is still being used in publications. Perhaps, because for no other reason than it is available in the Paup menu.

B. ILD (incongruence length difference), etc.
HOW: Compare the length of the most parsimonious trees for two or more data matrices or partitions to their length in the combined analysis and/or to randomly sampled partitions of equal size.

$ILD = L_{AB} - (L_A + L_B)/L_{AB}$

WHY: Should some data be excluded or reweighted rather than direct and equal combination? Are two data "partitions" combinable? What model should be preferred (for a recent interesting paper in this see Aagesen et al. 2005)?

WHAT IT TELLS US: Significant incongruence suggests that partitions may have a different evolutionary history. Some people would not combine data that showed significant incongruence. However, without evidence that there is some process that would cause the incongruence, down-weighting or eliminating character data simply because it is incongruent is really not scientific. Others would combine the data but consider that the result heuristically points to a need for more study. Examining RI and CI (see below) of partitions would be as informative and would allow for all data partitions to be examined in light of all critical evidence.

C. Basic Descriptive Indices:

Consistency Index (CI & ci)
Measure of how data fits the tree topology. Give the amount of homoplasy in a character or matrix for a give tree.

ci = m/s

where m = minimum number of steps in a character (number of states -1)
s = steps actually realized on a given tree

e.g. binary character m=1 actually has 1step on the tree then ci=1.0 if it has 2steps on the tree then ci=0.5

This index falls between 0 and 1.0 but is usually reported as scaled between 0-100

Ensemble CI (for the whole matrix) is the sum of all m/ total length of the tree (CI=M/S). In general, a high CI indicates that the data matrix "fits" the tree well (i.e., contains little homoplasy for the particular tree topology), whereas a low CI does not.

Characters with the same ci may not be contributing to the tree topology equally (e.g., autapomorphies ci=1.0 and perfect synapomorphies ci=1.0), so CI may be an overestimate if these are included. CI is NOT comparable between different sets of taxa as more taxa decreases CI.

Retention Index (RI & ri)
Measure grouping in formation in the data.

ri = (g – s)/(g – m)

where g= minimum steps on the worst tree (=bush)

Ensemble RI (for the whole matrix) like CI is based on sums RI=(G-S)/(G-M)

These problems for CI noted above may be overcome by excluding autapomorphies OR calculating a Rescaled Consistency Index.

RC = RI*CI

This removes the impact of any characters that do not contribute to the "fit" of the data to the tree (e.g., autapomorphies ci=1.0 and ri=0.0)

WHAT THESE TELLS US: These describe aspects of the tree and matrix or partitions of the matrix (e.g. 3rd position might have a lower CI and/or RI than 1st) or a particular sequence may contribute more to the resolution that another.

>>>>>>>>
Aagesen, L., Petersen, G. and Seberg, O. 2005. Sequence length variation, indel costs, and congruence in sensitivity analysis. Cladistics. 21(1):15-30 .

Grant, T. and Kluge, A. 2003. Data exploration in phylogenetic inference: scientific, heuristic, or neither. Cladistics 19:379–418.

Kim, J., 1993. Improving the accuracy of phylogenetic estimation by combining different methods. Syst. Biol. 42, 331–340.

Siddall, M.E., 2002. Measures of support. In: DeSalle, R., Giribet, G. and Wheeler, W.C. (Eds.), Techniques in Molecular Systematics and Evolution. Birkhaauser Verlag, Basel, Switzerland, pp. 80–101.

Table 1
List of data exploration methods assessed in this paper. Approaches that involve especially diverse methods are divided accordingly. See text for details.

Kind of method	Data exploration method(s)	Page(s)
Sensitivity analysis	Wheeler's sensitivity analysis	384, 388
	Decisiveness/ambiguity	388
	Bremer support	
	Double decay	
	Total support	
	Clade stability index	389
	Transformation series additivity	390
	Methodological concordance	391
	Sensitivity to prior probabilities (Bayesian phylogenetic inference)	393
	Skewness test	394
	Computer-intensive sampling	395
	Bootstrap	
	Jackknife	
	PTP	
	T-PTP	
	RT-PTP	
	HER	
	Long-branch attraction	398
	Likelihood ratio test for model selection	398
	Amount of evidence (missing data)	400
	Safe taxonomic reduction	
	Phylogenetic trunk	
	RILD test	
	Multiple regression analysis	
	Polymorphism	402
	Clade concordance index	403
Quality analysis	Relative rate comparison (saturation analysis)	386, 388
	Character compatibility	403
	Spectral analysis	404
	Relative apparent synapomorphy analysis (RASA)	405
	Data partition methods (taxonomic congruence)	406
	Topological incongruence test	
	Global congruence	
	χ^2 test	
	Mickevich–Farris incongruence index	
	Miyamoto incongruence index	
	ILD test	
	Partitioned Bremer support	
	Congruence with an empirically "known" phylogeny	410