

April 26, 2012: **Introduction to comparative methods for evolution, ecology, & behavior**

All we have time for today is a brisk look through the major categories of comparative methods. If you are going to do research in this half of phylogenetics, we strongly suggest continuing with IB 200B next spring. First, let's work through a little motivational exercise:

Assume the following data set of 12 taxonomic characters for eight tree species plus an inferred ancestor, and the distribution of three functionally important variables in these taxa.

(i) Construct a phylogenetic tree using the taxonomic characters. (ii) Point out which (if any) of the three following adaptive scenarios receives support from this tree.

taxa:	tax. characters:	fruit size:	fruit color:	fruit disperser:
one	1 0 1 1 0 1 1 1 1 1 0 0	L	RED	monkey
two	0 1 1 1 0 1 1 1 0 1 0 0	L	RED	monkey
three	0 0 1 0 1 1 1 1 0 1 0 0	L	GREEN	monkey
four	0 0 0 0 0 0 1 1 0 1 0 0	S	GREEN	monkey
five	1 0 0 0 0 0 1 0 0 1 0 0	S	GREEN	parrot
six	0 0 0 0 0 0 0 0 0 1 1 0	L	GREEN	monkey
seven	0 0 0 0 0 0 0 0 0 1 1 0	L	RED	monkey
eight	0 0 0 0 1 0 0 0 0 0 0 1	S	GREEN	parrot
Outgroup(s)	0 0 0 0 0 0 0 0 0 0 0 0	S	GREEN	parrot

A. Large fruits (L) evolved as an adaptive specialization (greater "reward") for monkey dispersal.

B. Monkeys evolved their preferences for fruits of certain trees because they are attracted to red colors.

C. Small fruits (S) evolved because parrots (given their particular food-size preferences) selected against trees with large fruit.

**A. Discrete-state characters on cladograms.** Sometimes we are interested in examining how discrete-state characters evolve on a tree individually and together. N.B. these are characters that met the 'discrete-state' criteria for taxonomic characters that we discussed earlier.

1. A review of character mapping, and its basis.

- Inferring history; Hypothetical Taxonomic Units (HTUs)
- Homology reviewed
- ACCTRAN vs. DELTRAN
- equivocal assignments to HTUs; ML approaches
- Possible errors in inference; what to do to correct or allow for these

2. Here are some types of hypotheses that can be tested:

- polarity of character changes in one character
- phyletic conservatism in one character (e.g. CI relative to randomization)
- helping to test adaptive hypotheses (the correct historical association of postulated cause and effect is part, but only part, of an adaptive explanation).

- association of state changes in two characters
  - undirected (Ridley's test)
  - directed (Maddison's test)

Maddison's (1990) concentrated changes test was introduced to test for correlations in the evolution of discrete traits, in a phylogenetic context. The test examines the question: are changes in character B concentrated on portions of the phylogeny where character A has a particular state, more than expected by chance? For example: does dioecy evolve from hermaphroditism more often than expected in lineages that have fleshy (vs. dry) fruits? Does larval gregariousness in butterflies evolve more often than expected in lineages with warning coloration?

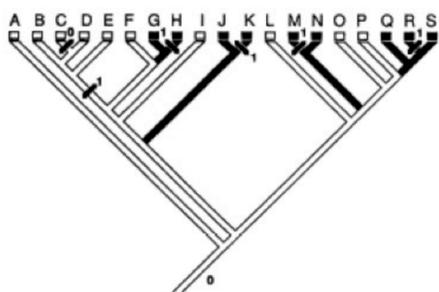


FIG. 1. Phylogenetic tree with the evolution of two characters traced on it. Changes in the first character are indicated by tick marks with the derived state listed ("0" or "1"). The states of the second character are indicated by the shading of the branch: white for one state and black for the other.

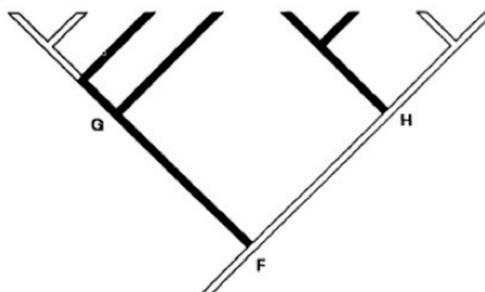


FIG. 2. Phylogenetic tree used in example calculations of Figure 3.

$\geq 4$  gains (out of 5 gains, 1 loss) on black branches:  $p = 0.0638$

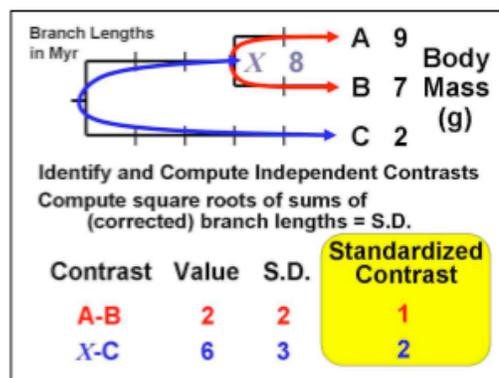
69 ways to arrange 2 gains and 0 losses 15 have 2 gains on black branches

## B. Evolution of quantitative traits on cladograms

1. The bad old way. Direct correlations of species values. Species as data points. Why not?
2. Advent of quantitative comparative approaches; importance of history (jobs!)

### Motives:

- A. "remove" influence of history
  - B. Describe variation due to phylogeny by tracing on phylogenetic tree
3. Taxic approaches (compare sister groups):
    - A. ANOVA and ANCOVA -- nested using taxonomy; partition variance, "subtract" phylogenetic effects.
    - B. Autocorrelation -- based on spatial statistics; time from divergence, i.e., distance on a phylogeny; may or may not weight by branch length.
    - C. Independent contrasts (Felsenstein) -- assume separate paths, based on Brownian model; compare differences in value of traits between sister groups (see fig at right)
    - D. Regression -- general linear model approach.



Courtesy Ted Garland

4. Transformational approaches (reconstruct values for nodes, look at direction of change):

A. How to map quantitative characters on a tree?

B. Minimum evolution models

1. Distance Wagner approach (median state rule)
2. Squared-change parsimony (minimize sum of squared changes; again assuming Brownian model)
3. Directional methods (Huey & Bennett; Losos)

C. Tempo and mode issues (combining character evolution with cladogram comparisons):

Gradual versus punctuational: is change along branches uniform?

Rates in morphology?

- standard deviation or *coefficient of variation* [= 100(s/y)]
- the *darwin*: change by a factor of  $e$  (=2.718) per million years; on a natural log scale that is a change in one unit per million years; millidarwin, etc.

How to test hypotheses about rates? (e.g., molecular clocks?) -- maximum likelihood tests (covered in earlier session)

How to test hypotheses about trends?

- An example using number of pollinators in orchids -- sign test

D. Comparing cladograms

-- There are many reasons why one would want to compare cladograms, falling into three basic categories:

-- *Within an analysis of one major clade, with the same OTUs*; e.g., equally or nearly equally parsimonious (or likely) trees, trees resulting from different character partitions, models of evolution, or methods of analysis, and comparisons with trees from the literature.

-- *Within an analysis of one major clade, with different OTUs*; trying to come up with a general tree for all OTUs, e.g. super trees, compartmentalization.

-- *Comparing analyses of different major clades*, e.g., gene family evolution, migration between populations, biogeography, host/ parasite relationships, symbiosis, community evolution, or any long-term ecological association

-- Methodology for comparing cladograms:

(1) consensus techniques (strict, semi-strict, majority rule, Adams) -- finding shared signal among trees (examples gone over before)

(2) tree-to-tree distance metrics (e.g., NNI interchange metric, partition metrics, agreement subtrees)

(3) component analysis (was covered in biogeography lecture) -- finding individual statements of relationship that are shared among trees, basically a node relating some taxa to the exclusion of others.



(4) parsimony approaches -- representing the grouping information in separate trees as characters in a matrix (Brooks' parsimony; was covered in biogeography lecture))

(5) maximum likelihood approaches (parametric bootstrapping) -- comparing alternative trees or alternative models of evolution for your data.

(6) supertrees -- combining different trees, e.g., using Brooks parsimony (called "matrix representation" in the literature).

## E. Comparing sister clades within one cladogram: The shape of evolution

In order to address various questions in both micro- and macro-evolution, we need to address issues such as the symmetry of balance of trees. What is the null expectation? Intuitively, we would expect balanced trees, perhaps, based on some sort of false analogy to coin flips. But, is this the right? How would you generate "random" trees?

How to randomize a labeled tree (n.b., this is different than randomizing tree shapes if tips are considered unlabeled!)

equiprobable trees (picking out of a set of all possible trees -- bias towards asymmetry)

random joining (models a random speciation process -- intermediate bias)

random partition (bias towards symmetry)

How to grow a tree:

the most common approach is the Yule "pure birth" model (Markovian)

birth + death models (N.B., it is hard to tell higher birth rate from lower death rate!)

What do random trees look like? Slowinski & Guyer (1989) showed a non-intuitive result: the probability of generating a  $1 + (n-1)$  tree is  $2/(n-1)$ , which is equal to the probability of **any** division of species into lineages of unequal size (the probability is  $1/(n-1)$  when the species are evenly divided). Thus, even a tree in which one species is the sister taxon to 39 other species is not significantly non-random at the  $P = .05$  level ( $P > .051$ ).

What do real trees look like? Often are quite asymmetrical; could this be a methodological bias? Even if real, how do we judge whether it is significantly asymmetric? Furthermore, even if it is significantly asymmetric, how do we associate that with some specific factor postulated to be the cause of that asymmetry? That leads to the topics of "key innovations" and "adaptive radiations"

Adaptive radiations -- how best to define?

\*\*\* "Diversity" (number of subclades) vs. "disparity" (ecology, morphology, genetics)

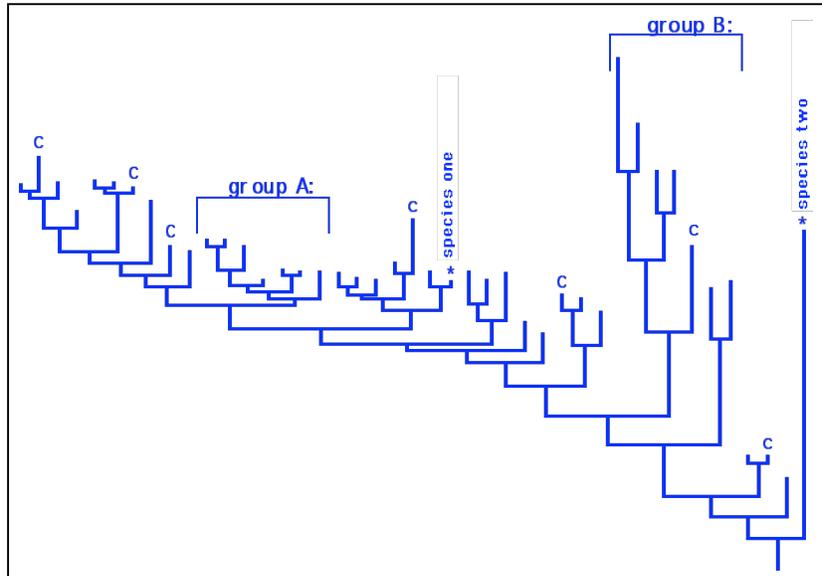
Levels of selection problem -- "key innovation" as analog of "adaptation" at organism level -- how do we assess whether selection among clades is actually occurring?

**F. Comparing branch lengths within and between cladograms.**

In order to address diverse types of questions ranging from rates of character change in adaptive radiations, to molecular dating of phylogenies, to conservation biology, we often need to study and compare branch lengths within and between cladograms.

What are the best units to measure branch lengths in? DNA, morphology, or both? Should we convert to *time*, and if so, how?

To the right is a hypothetical cladogram illustrating the potential role of phylogeny in setting conservation priorities. Shown is a phylogeny of 43 "species"; the branch lengths are proportional in the vertical direction to the number of evolutionary character changes along that branch. If you were offered three alternative conservation scenarios that would save either group A or B or C, how would you rank them? How would you explain your ranking to a biodiversity manager who is not familiar with phylogenetic concepts?



*Phylogenetic diversity:*  
 PD = sum of branch lengths of subtree incorporating set of selected taxa.

Biodiversity assessment: to the right is a map of Australia showing grids with significantly high PD of *Acacia* in green, and significantly low PD in red, based on a randomization test (Mishler et al., in prep).

