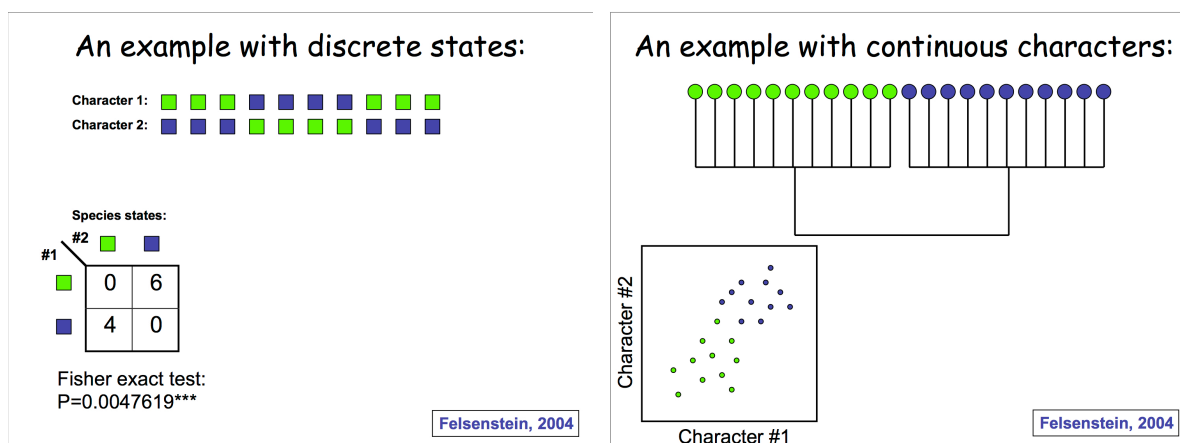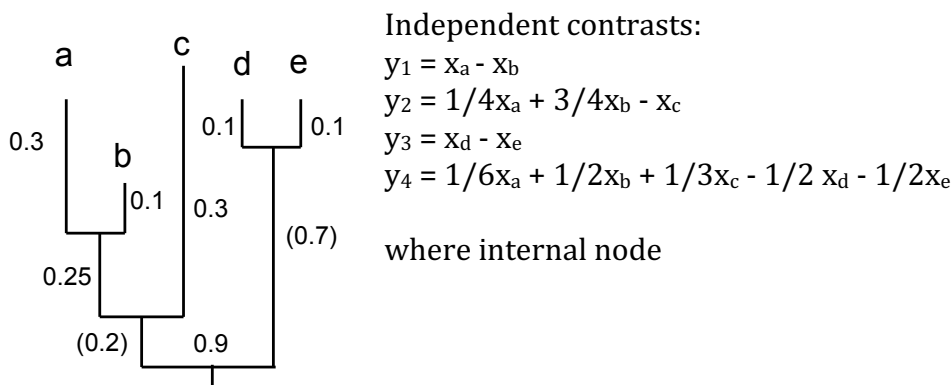# Phylogenetic comparative methods

In Biology we are often interested in co-evolution of various traits (e.g., population density and body size; IQ and reproductive success; genome size and cell size; etc.). Conventional statistical methods assume that samples are independent. Such an assumption is often inappropriate in Biology, where all species share some common history. Phylogenetic comparative methods (PCMs) aim to incorporate information on the evolutionary relationships of organisms (phylogenetic trees) to compare species.

Two examples:

An example with discrete states:

Character 1:
Character 2:

Species states:

	#2	
#1	0	6
	4	0

Fisher exact test:
P=0.0047619***

Felsenstein, 2004

An example with continuous characters:

Character #2

Character #1

Felsenstein, 2004

Continuous characters: the independent contrasts method

Independent contrasts:
$y_1 = x_a - x_b$
$y_2 = 1/4 x_a + 3/4 x_b - x_c$
$y_3 = x_d - x_e$
$y_4 = 1/6 x_a + 1/2 x_b + 1/3 x_c - 1/2 x_d - 1/2 x_e$

where internal node

- Expectations: 0
- Variances: proportional to the sum of branch lengths
- Can be standardized by dividing contrast score by its standard deviation

1) Find two tips on the phylogeny that are adjacent (nodes i and j) and have a common ancestor, say node k.
2) Compute the contrast Xi - Xj. This has expectation 0 and variance proportional to vi + vj,
3) Remove the two tips from the tree; the ancestor k becomes a tip. Assign it the character value (I) the weighted average of Xi and Xj.
4) Lengthen the branch below node k from vk to vk + vivj/(vi + vj).
5) Repeat 1-4.

Finding the most parsimonious ancestral states

Information from above and below the node is distilled and combined to yield the estimate. Thus the state estimated for a node will depend on the information distilled (1) down through the clade of the node's left descendant, (2) down through the right descendant's clade, and (3) up from the part of the tree below and beside the node. These three sources of information converge at the node and indicate its state, as shown below.

In practice this distilling process is done in two or three passes up and down the tree. The details of the process may differ with the different assumptions used. We will consider the algorithm for unordered characters.

Combining sets in downpass   Combining sets in uppass   Combining sets to yield MPR set
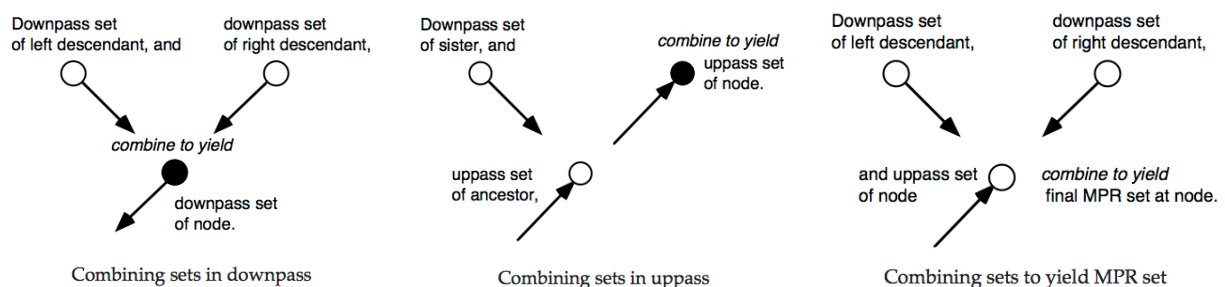
We created three sets of character states. The first set is called the downpass set of the node, because this is the set of states preferred by that part of the tree above the node. The states assigned on the downpass are most parsimonious according to the clade above it, but may not be most parsimonious according to the whole tree because the state calculations have not yet considered nodes below the visited node.

The set of states assigned on the uppass will be called the uppass set of the node. This is the set of states preferred by that part of the tree below the node. The uppass state set of a node is calculated from the uppass set of the node below it (its ancestor) with the downpass set of the node beside it (its sister node), using the same cal-culation for combining state sets as for the downpass.

The final set combines the first two and is called the MPR set of the node.

Example from the class:
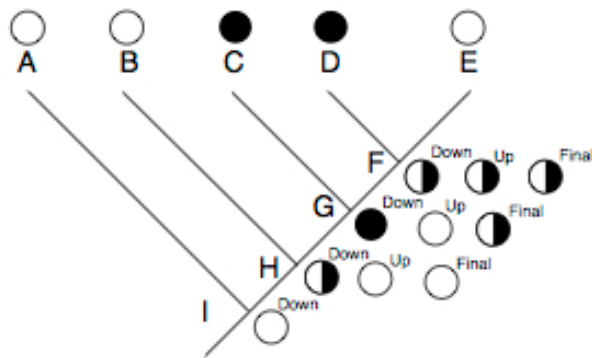
Reconstruction uncertainty.

"Character state reconstructions can provide a powerful mechanism for studying many facets of the evolutionary process. However, the zeal with which these techniques are sometimes advocated belies the complexity of the problem"

The first expression of uncertainty is ambiguity in the reconstruction or the existence of multiple equally parsimonious reconstructions. Ideally, when faced with multiple reconstructions, we should examine all of them; methods that examine only a subset of the reconstructions should be avoided, if possible. In reality, the number of possible reconstruction is often prohibitory large and only a few of them are examined/presented.

Two popular methods of choice are ACCTRAN and DELTRAN, which yield extremes of reversals versus parallelisms in the reconstruction.

ACCTRAN and DELTRAN
For characters of unordered and ordered type, ambiguities in character tracings can be resolved so as to choose the assignments that delay or accelerate transformations (look at the figure above and think what will happen to our reconstruction if we assign white/black color to node G). The DELTRAN option prefers most parsimonious assignments that delay changes away from the root; this maximizes parallel changes. The ACCTRAN tracing shows those assignments that accelerate changes toward the root; this procedure maximizes early gains and thus forces subsequent reversals. Again, ACCTRAN and DELTRAN are but two of various methods to select from among the most parsimonious reconstructions. It is best to examine all reconstructions, if this is possible, especially if you are not certain about what reconstructions are most critically related to your biological question.

Uncertainty does not end with ambiguity in the reconstruction. Even unequivocal assignment may be incorrect. Reconstructions of ancestral states are subject to error, as are all estimates of history. Studies of the reliability of ancestral state reconstructions

have yielded mixed results, although it is clear that when rates of evolution are high over the time scale of the tree, error rates can be high. Also, the rate of change from one character state to another character state can be different (e.g., losses can be much easier than gains).

Probabilistic methods (ML and Bayesian)

Hence the use of probabilistic methods (ML) provides several advantages:
1) Use an explicit model of character evolution
2) Consider branch length
3) Can estimate the relative probability of each character state at every node.

Bayesian methods provide additional advantages
1) avoid potential errors in fixed parameter estimates
2) address phylogenetic uncertainty