Maximum Parsimony Analysis

Three tree-building methods
- **Maximum Parsimony**
  - Not model based
  - Criterion based so includes near trees
- **Distance (or clustering) algorithms**
  - Explicit model of sequence evolution
  - Algorithmic so returns just one tree.
- **Maximum Likelihood**
  - Explicit model of sequence evolution
  - Criterion based so includes near trees
  - Slow search but statistically sound

Parsimony principle.
- Implicit assumption: evolutionary change is rare.
- Goal: find the tree(s) that require the fewest changes.
- Attempt to maximize homology and minimize homoplasy.
- Tree depends on “phylogenetically informative characters” only.
MacClade example

- Mac software for parsimony
  - Try Mesquite (for PCs and Macs, free)
- Example data set: primate.nex
  - 3 mtDNA protein coding genes
  - Previously downloaded and aligned.
  - Protein coding colors on nucleotides.
    - Note easily detected homologies.
    - Note “phylogenetically informative sites.

Examine amino acid sequence

Homology or homoplasy?

Homology or homoplasy?
3 phylogenetically informative sites

Singletons: parsimony uninformative.

Many constant sites with respect to codons

Nucleotide view
Parsimony informative.

Parsimony principle
- Find presumptive homologies.
- Phylogenetically informative sites:
  - At least two different characters and at least two taxa each.
- These are the characters used to infer the phylogeny - the evolutionary tree.

Parsimony uninformative.

Parsimony principle
- Find the tree that maximizes homology and minimizes homoplasy.
- Homology:
  - similar by common descent
  - corroborates the presumptive tree.
- Homoplasy:
  - recurring independent events.
  - e.g. convergence
  - contradicts the presumptive tree.
Random tree

Hominoid Primates with Capuchin outgroup
COL, ND2, Cyb mtDNA genes

Treelength: 3759 steps

We can improve the tree by hand

Treelength: 3759 steps
Shorter more parsimonious tree: 3716 steps

But we want the optimal tree.

Maximum parsimony tree: 3325 steps

MacClade parsimony search

- We ask MacClade to find the maximum parsimony (MP) tree
  - In MacClade we choose a search tool.
  - Run until dialog window says Close
  - New score is shorter by 434 steps.
- We use the rooting tool to change rooting.
  - What happens to the score? Why?

Trace the changes site by site.
Trace the changes.

Inconsistent with the MP tree  Consistent with the MP tree

Inconsistent with the MP tree  Consistent with the MP tree

Trace the changes.

Inconsistent with the MP tree  Consistent with the MP tree

Parsimony principle.

- Implicit assumption:
  - evolutionary change is rare.
  - Attempts to maximize homology, minimize homoplasy.
  - This is pleasingly intuitive.
  - We shall see later that the advantages and disadvantages of parsimony depend on the data set.
So let’s try phylip’s parsimony

- Works behind the scene in the way that MacClade works.
- Attempts to maximize homology, minimize homoplasy.
- Open the phylip folder.
  - Note that phylip has many modules.
  - Make a new folder called Analysis (say)
  - From the exe folder copy and move dnapars to the Analysis folder.
  - Also move the primates2seq.PHYLIP file into the Analysis folder.

Open dnapars

- Type in the name of the data file: primates2seq.PHYLIP
- Hit return
- Now examine the parsimony options.
- Toggle through the options by typing a letter.
- First we will use default settings.

Save tree.

- Rename output files to prevent overwriting.
- Tree1 (say) and Outfile1 (say).
- Now open dnapars again.
- Open the same file again.
- Now type J to start the tree with a random taxa.
- Input any odd number for a random seed.
- Number of times to jumble: type 10
- Type y (yes)

Save and examine the two trees.

- Rename output files to prevent overwriting.
- Tree2 (say) and Outfile2 (say).
- Now open the output files.
  - Outfile shows the tree and branch lengths.
  - Tree files are in the “newick” format
    - Recall: Felsenstein’s preferred format.
    - We mere mortals think this is difficult to interpret.
  - So open FigTree and then open both trees.
    - These two trees are the same
    - But different search strategies can give different trees.
Later we examine more search options

- These search the very large tree
  - space more or less quickly and…
  - more or less thoroughly.
- The j option (jumble) chooses to start from a random tree.
  - And you specify how many times to repeat the rearrange the initial tree.
  - This increases the probability of finding the truly most parsimonious tree(s).
- These options are very important for larger data sets.

Other software options for parsimony

- Mega for PCs
- Mesquite for PCs and for Macs
- PAUP (phylogeny analysis using parsimony) and MacClade ($100 each but very powerful).

Next week: Distance methods

And an introduction to models of sequence evolution