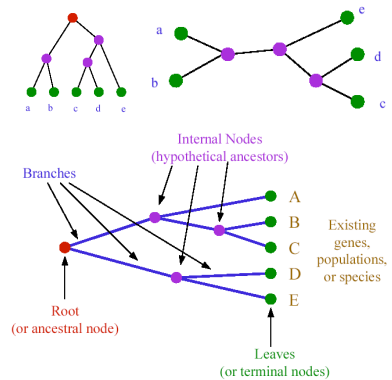


Phylogenetic trees

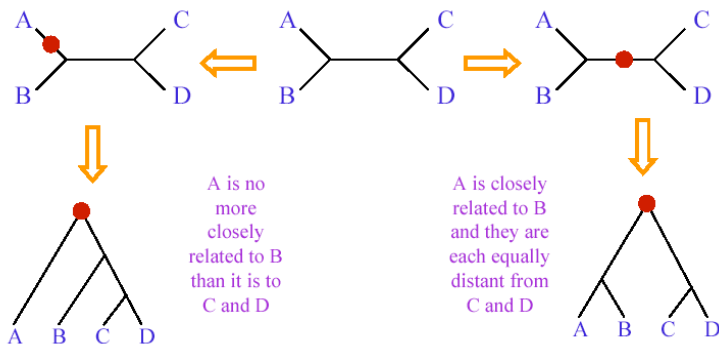
MEP: 20 Sept 2005

types of trees, terminology

- ♦ root: common ancestor of all taxa
- ♦ rooted, unrooted trees:
- ♦ topology: branching pattern of a tree
- ♦ branch (interior, exterior): defines the relationship between taxa in terms of descent and ancestry
- ♦ branchlength: represents the number of changes that have occurred in that branch
- ♦ node: represents a taxonomic unit; can be an existing species or ancestor
- ♦ clade: a group of two or more taxa including common ancestor and all descendants



where to root?



rooting can imply relationships

Types of trees

- ♦ Cladogram: only shows relationships
- ♦ Phylogram ("additive"): adds info about amount of change
- ♦ Phenogram ("ultrametric"): adds info about time since divergence

realized v. expect trees

- ◆ expected tree is made using infinitely long sequences (or expected number of substitutions for each branch)
- ◆ realized tree is made from the actual number of substitutions
- ◆ reconstructed (or inferred) tree is made from the observed sequences and given method

tree stats:

- ◆ rooted trees:
 - ◆ branches = $2m-2$
 - ◆ exterior branches = m
 - ◆ interior branches = $m-2$
 - ◆ interior nodes = $m-1$
- ◆ unrooted trees:
 - ◆ branches = $2m-3$
 - ◆ exterior branches = m
 - ◆ interior branches = $m-3$
 - ◆ interior nodes = $m-2$

number of possible topologies

- ◆ for rooted trees
 - ◆ for m taxa, # trees = $[(2m-3)!]/[2^{m-2}(m-2)!]$
 - ◆ for 10 taxa: 34,459,425
- ◆ for unrooted trees
 - ◆ for m taxa, # trees = $[(2(m-1)-3)!]/[2^{m-3}(m-3)!]$
 - ◆ for 10 taxa: 2,027,025
- ◆ only one tree is the TRUE TREE

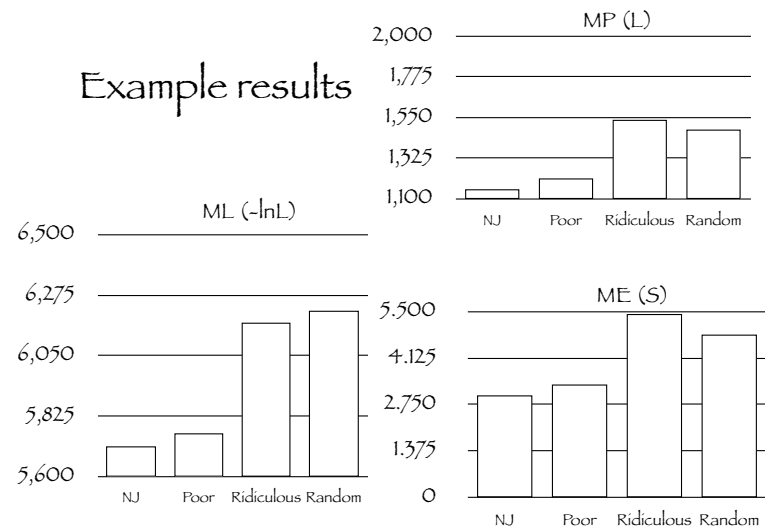
scoring trees

- ◆ the true tree is the one that:
 - ◆ has the smallest overall distance: "Minimum Evolution"; Rzhetsky & Nei (1993) provided a mathematical proof when unbiased estimates of evolutionary distance is used (i.e., the right model), irrespective of m
 - ◆ can be explained by the fewest number of nucleotide substitutions: "Maximum Parsimony"; works when there are no backwards, parallel, or multiple subst's and n is very large
 - ◆ has the highest likelihood given the sequence data: "Maximum Likelihood"; technically, it finds the set of branch lengths that maximize overall likelihood

scoring trees: example

- ◆ primate data with 3 trees:
 1. good, from NJ: (Lemur_catta, ((((((Homo_sapiens,Pan),Gorilla),Pongo),Hylobates), ((Macaca_fuscata,M._mulatta),M._fascicularis),M._sylvanus)),Saimiri_sciureus),Tarsius_syrichta)
 2. poor, gibbon & chimp are swapped: (Lemur_catta, ((((((Homo_sapiens,Hylobates),Gorilla),Pongo),Pan), ((Macaca_fuscata,M._mulatta),M._fascicularis),M._sylvanus)),Saimiri_sciureus),Tarsius_syrichta)
 3. ridiculous: (Hylobates, ((((((Homo_sapiens,Lemur_catta),Gorilla),Pongo),Macaca_fuscata), ((Pan,M._mulatta),M._fascicularis),M._sylvanus)),Saimiri_sciureus),Tarsius_syrichta)

Example results



searching treespace

- ◆ exhaustive search: evaluate all topologies and identify the one(s) that meet the optimality criterion; virtually impossible with more than 12-15 sequences; guaranteed to find the best tree(s); start with core tree of 3 taxa, each additional taxon is added one-at-a-time in the position that maximizes the optimality criterion (then restart with a different core tree)
- ◆ branch & bound search: used for 10-20 taxa; start with an arbitrary initial tree (often NJ) and compute a score; then we proceed with exhaustive search, quitting any avenue of trees that exceed the initial tree score; this greatly reduces the number of trees examined; guaranteed to find the best tree(s)
- ◆ heuristic search: used when $m > 20$ sequences; no guarantee to find the best tree(s); start with an initial tree and rearrange it to find similar trees with better scores; rearrangements through branch-swapping nearest neighbor interchanges (NNI), subtree pruning & regrafting (SPR), and tree bisection-reconnection (TBR)

search commands for PAUP:

exhaustive

- ◆ set criterion=parsimony|distance|likelihood
- ◆ AllTrees;
 - ◆ searches every last damn possible tree
 - ◆ obviously, it'll take PAUP a while

search commands for PAUP: branch-and-bound

- ◆ BandB;
 - ◆ faster than exhaustive, practical for $m < 20$

search commands for PAUP: heuristic

- ◆ Hsearch
 - ◆ start=stepwise|NJ|current
 - ◆ swap=none|NNI|SPR|TBR;
- ◆ faster