Phylogenetic tree IV- Tree selection, consensus, compromise

I. To choose or not to choose? In most cases an optimality criterion will result in a set of equally best trees. There are basically three views on the way to deal with this.

A. *All trees of the optimal set are equal. So no single tree should be selected.*

1. Alternative trees may be discussed and the strict consensus presented.
2. If a procedure requires a single, fully resolved tree then a random one can be used.
3. Using a secondary optimality criterion is not valid and (typically) equal weights were used.

B. *Optimal trees represent a select subset of all possible trees and implementing a secondary optimality criterion(a) that selects from those trees is legitimate.*

1. Internal evidence
   a. subjectively preferred character state transformation
   b. explicit (numerical) secondary optimality criteria methods that concentrate homoplasy and related tree optimality methods

"the trees themselves tell us how reliable the characters are" (Goloboff 1993).

SAW- Successive Approximations Character Weighting (Farris 1969)
- get starting MPTs
- use character fit to reweight (could be ci, ri or rc)
- search for MPTs with the new weights
- repeat until a stable set of trees is found.

IF this results in a subset of MPT from the original data those are preferred. However, often this results in a different set of trees. It was not originally introduced to be a secondary optimality criterion.

PIWE- Implied Weights (Goloboff 1993)
- weighting function is used to maximize weighted fit of characters to trees.
$fi=(k+1)/S_i+k+1-m_i$
k= constant (1...6); S_i = observed steps; m_i= minimum possible steps

e.g. For k=4 the cost of adding one step to a character with two extra steps is 54% of the cost to add a step to a "perfect" character.

- This kind of weighting function and SAW tend to push homoplasy into fewer characters and so the fittest tree(s) from the set of MPTs *could* be selected. But, Goloboff did not introduce this method a secondary optimality criterion.

AUCC- average unit character consistency, (Sang 1995). sum ci/number of characters

Tree	ci										AUCC
A	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	0.500
B	1	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/3	0.533
C	1	1	1	1/2	1/2	1/2	1/2	1/3	1/3	1/3	0.600
D	1	1	1	1/2	1/2	1/2	1/2	1/2	1/2	1/5	0.620
E	1	1	1	1	1/2	1/2	1/2	1/2	1/4	1/4	0.650
F	1	1	1	1	1	1	1	1	1	1/11	0.909

- Pack the most homoplasy in the fewest characters and thereby preserve the maximum number of initial hypotheses of homology. Choose the tree(s) with the highest AUCC.
- But….Why this measure? Others abound.... optimal character compatibility index (OCCI) (Rodrigo 1992); boil-down (Sharkey 1989) ....

2. External evidence
    a. correspondence to existing taxon hypotheses
    b. best fit to characteristics that are cannot be coded as characters
    c. preferred transformations and/or preferred character weights

C. *If secondary optimality criterion(a) can be justified it should be included in the initial search.*

a. upon recognition that many equal trees result from the first analysis a new analysis with differential weighting or explicit model is used

II. Consensus & Compromise: The representative summary of a set of source trees. Consensus trees can only be the most optimal tree when it is identical to one of the optimal source trees. Consensus trees have lost the information about what trees went into them, so reconstructing character evolution (mapping) and use of treelength on them should be avoided, or maybe done with extreme caution.

A. Strict consensus- Only monophyletic groups found in all source trees are found in the resultant tree. The tree excludes a subset of all possible trees and conversely includes a subset of possible trees, whether or not they are part of the source set. In some sense the most conservative consensus. However, consider the bush.

e.g. (A(B(CD))) + (A(C(BD))) = (A (BCD)) but this also implies (A(D(BC)))

NOTE: All trees below contain some resolution not supported in all source trees (Bryant 2003):

B. Semistrict (Bremer trees or combinable-components) - Only monophyletic groups found in at least **one** of the source trees and compatible (not in conflict) with all other source trees are found in the resultant tree, i.e. if a clade is never contradicted, but not always supported, then it is still included in this compromise tree.

C. Majority-rule - Shows groups that appear on pre-specified percentage of source trees, usually >50%. Used for summary of searches where plurality is important. Can result in a tree that contains two groups that are simultaneously found in **only one** of the source trees (minimum to make majority =0.5T+ 1).

	T1	T2	T3	T4	T5	T6	T7	TOT
AB								
CDE								
DE								
XCDE								
XDE								
XC								
XAB								
XB								
XE								
ABCDE								
XABCDE	1	1	1	1	1	1	1	7

D. Adams - Inconsistently placed taxa are moved to the first node that summarizes the possible topologies. Groups can appear in Adams consensus that are not found in **any** source tree. Adams trees have no biological or phylogenetic interpretation. They do point to "wildcard" taxa. Those taxa may be experimentally removed from the matrix and the resulting analysis compared to when they are included.

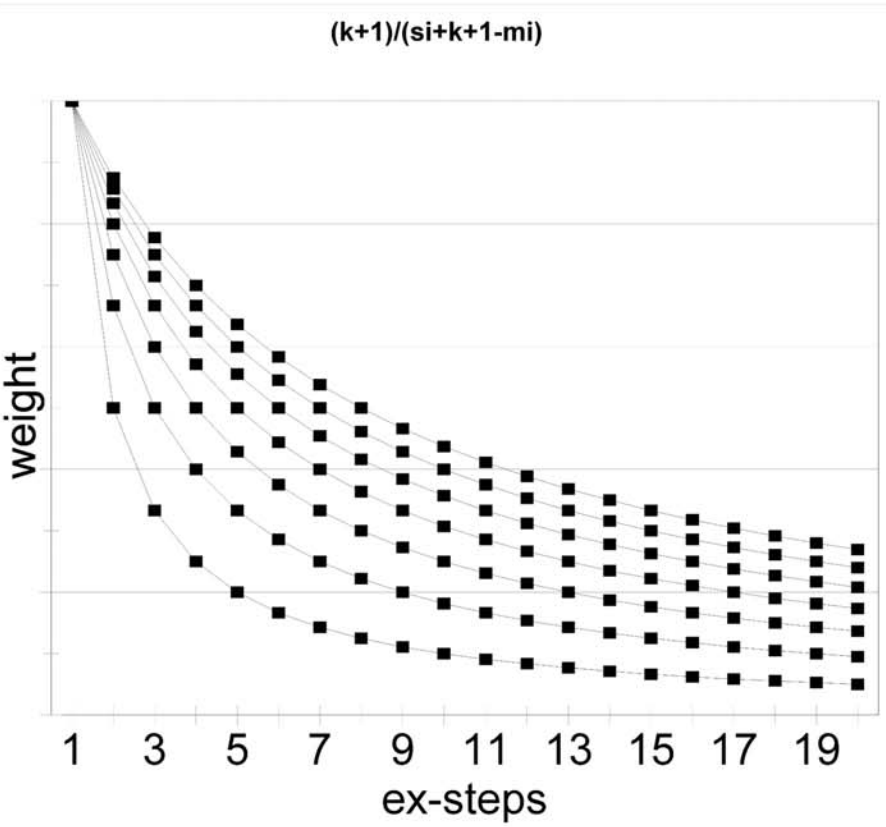
E. Greedy consensus. Groups ordered by frequency like in Majority-rule, then added in to the consensus tree as long as they are compatible. How will ties in frequency change the results?

F. Matrix representation with parsimony (MRP). A recoding consensus method that can be used for trees with different sets of taxa. Both topology and frequency are important.

```
Unrooted source trees
((a; b; c); (d; e; f)),   t1 t2 t3 t4 t5 t6 t7
((a; b; c); (d; e; f)), a 1  1  1  1  1  0  0
((a; b; c); (d; e; f)), b 1  1  1  1  0  1  0
((a; b; c); (d; e; f)), c 1  1  1  1  0  0  1
((a; d; e); (b; c; f)), d 0  0  0  0  1  1  1
((b; d; e); (a; c; f)), e 0  0  0  0  1  1  1
((c; d; e); (a; b; f)). f 0  0  0  0  0  0  0
```

((de) (abc)f)

Bryant, D. 2003. A classification of consensus methods for phylogenies. in Janowitz, M., Lapointe, F.-J., McMorris, F.R., Mirkin, B., Roberts, F.S. (eds) BioConsensus, DIMACS. AMS. 163–184.



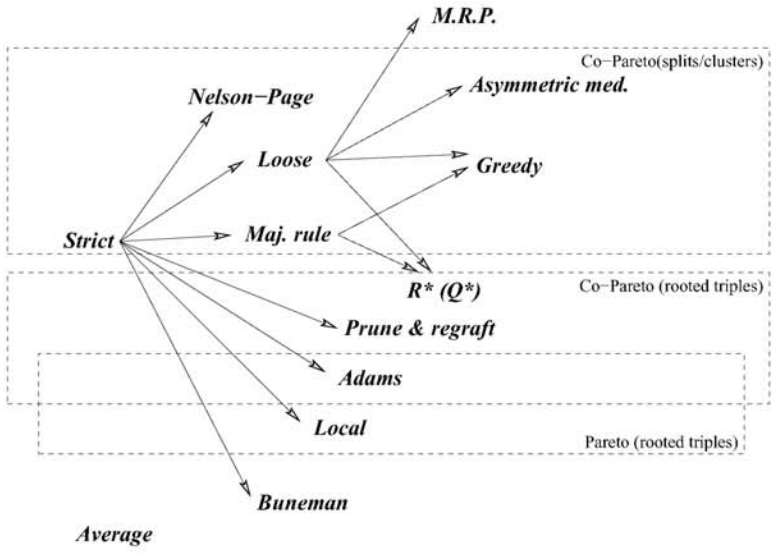
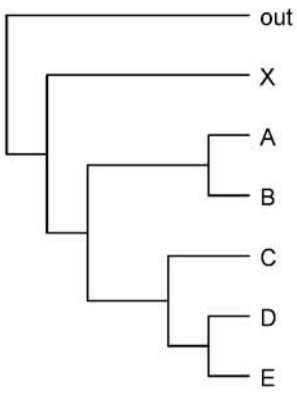
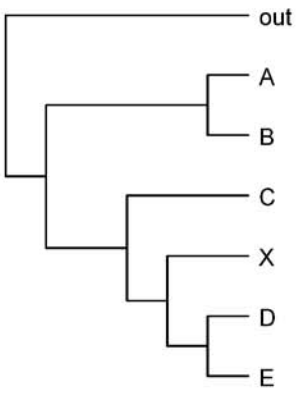


FIGURE 2. A classification of consensus methods. There is an arrow from one method to another if every split in the consensus tree produced by the first method is contained in every consensus tree produced by the second method.

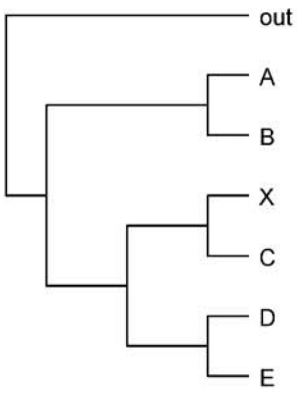
t1



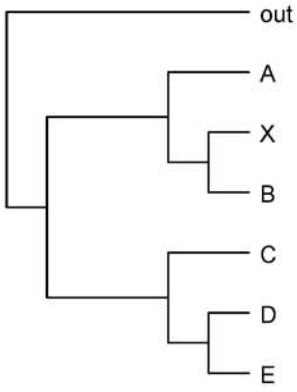
t2



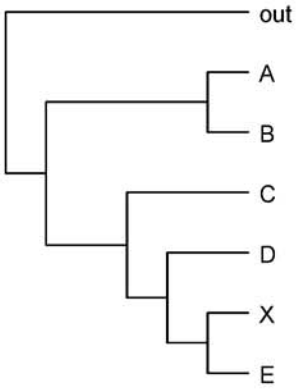
t3



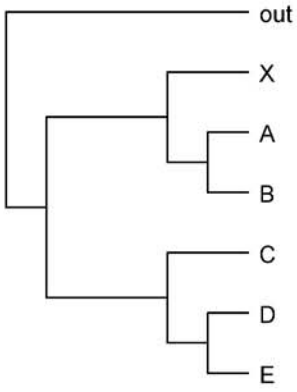
t4



t5



t6



t7