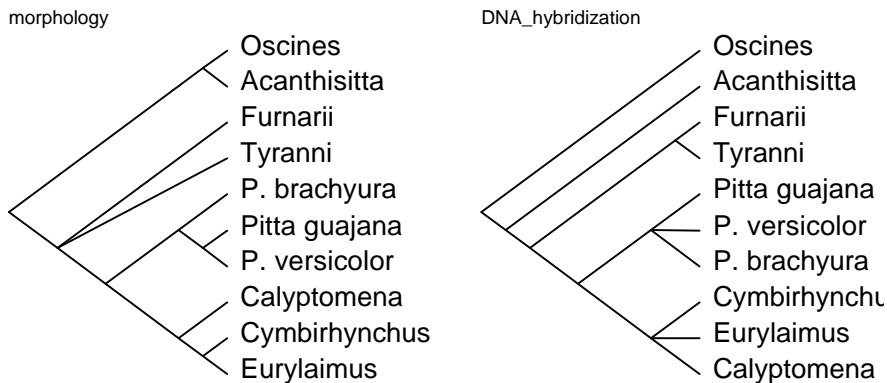


Exercises Using COMPONENT Lite

(These exercises are from the NERC/Glasgow Taxonomy Course 1997, and are © 1997, Roderic Page).

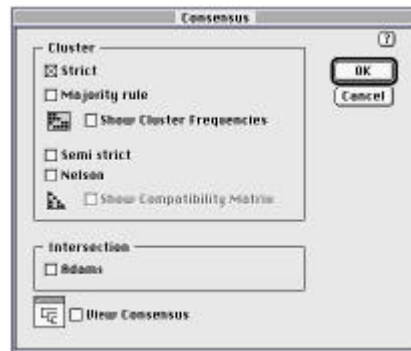
Exercise 1

On your machine you have a file **Passerines (DNA versus morphology)** which contains two trees for a group of passerine birds, one based on morphological data, one based on DNA hybridisation (Sheldon and Bledsoe, 1993). We will use these trees to explore strict and semi-strict consensus methods.



To do:

Use the **Consensus** command to compute a strict consensus tree for the two trees. When you chose the **Consensus** command COMPONENT will display this dialog box:



Check both the **Strict** and the **Show Cluster Frequencies** check boxes. You should see something like this in the output window:

Cluster table

	DRFTP	PPCEC	
	scuyi	..yua	
	carnt	mrl	
	innat	vbbyy	
	ntana	erilp	
1	●●●●●	50
2	●●●●●	50
3	●●●●●	100
4	●●●●●	100
5	●●●●●	50
6	●●●●●	50
7	●●●●●	100
8	50
9	100

The strict consensus tree comprises only those groups that are present in all the trees.

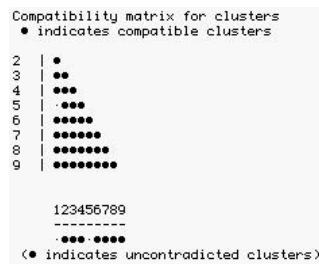
Q Find the clusters that are present in only one of the two trees. For each cluster unique to one tree, determine whether there is a cluster in the other tree that contradicts it.

Trees can differ because they are in conflict about relationships among taxa, or they can differ in degree of resolution, or both. If one data sets resolves a group that another data set fails to resolve, but does not contradict, then perhaps this is not genuine conflict.

Q Which clusters in the two trees genuinely conflict?

To do:

Now, compute a semi-strict consensus tree for the two trees. be sure to check the **Show Compatibility Matrix** option. For each pair of clusters this matrix contains a “●” if the pair is mutually compatible, and a ‘.’ if the two clusters contradict each other. At the bottom of the matrix is a line which tells you for each whether that cluster is contradicted by any other cluster.



Q How does the semi-strict consensus tree compare with the strict consensus tree? Can you explain why they differ?

Q Trees can differ not only in grouping and resolution, but also in how they are rooted. From the last analysis, why do the two trees disagree about the relationships of the Oscines and *Acanthisitta*? (Hint: look at the trees both rooted and unrooted).

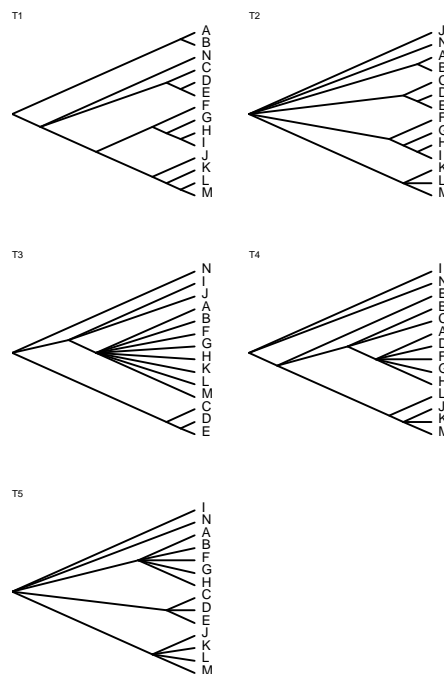
Q So, do the two trees disagree about bird relationships?

Exercise 2

In many cases different trees will contain conflicting groups. Rather than simply keep those that occur in all trees (**strict**), or are uncontradicted by any group (**semi-strict**), we may ask “what are the groups most commonly agreed upon?” One way of answering this is to construct a **majority-rule** consensus tree, which contains only the groups occurring in a majority (i.e., more than one half of the trees).

As an example the file **Nelson (1979)** contains five trees used by Nelson (1979) to illustrate his consensus method. (As an aside, the method described by Nelson himself, and the “Nelson consensus tree” that appears in many papers are quite different — some authors prefer to call strict consensus trees “Nelson trees.” Indeed, the **nelsen** command in Hennig86 produces strict consensus trees. You have been warned!)

The five trees are shown below:



To do:

Compute strict and semi-strict consensus trees for these trees. When you do, display the cluster table as well.

Q Are there any uncontradicted clusters among these five trees?

To do:

Now, construct a majority-rule consensus tree.

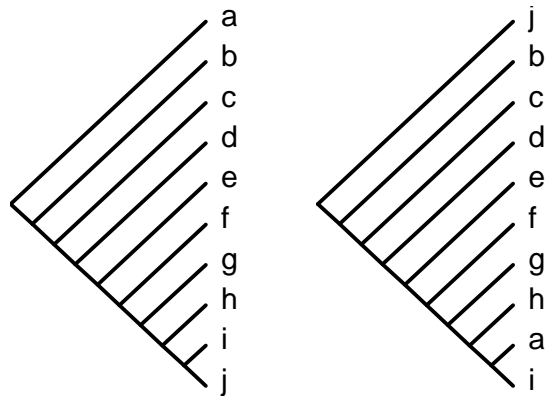
Q “Any two clusters that occur in more than one half of a set of trees must be mutually compatible.” Why is this true? (Don’t spend too much time on this!).

Exercise 3

In choosing a consensus measure we have seen that we need to decide what frequency of occurrence of a group we desire before including a group into a tree; strict consensus requires that the group appear in every tree, other methods may be less strict. Another decision to consider is what kind of information we are looking for in our trees. So far we have restricted our attention to clusters — monophyletic groups. However, is that the only information present in a tree?

To do:

Consider the two trees in the file **Adams consensus example**:



Compute a strict consensus tree for these two trees.

Q What does the strict consensus tree look like? Do you think it reflects the actual similarity between these two trees?

To do:

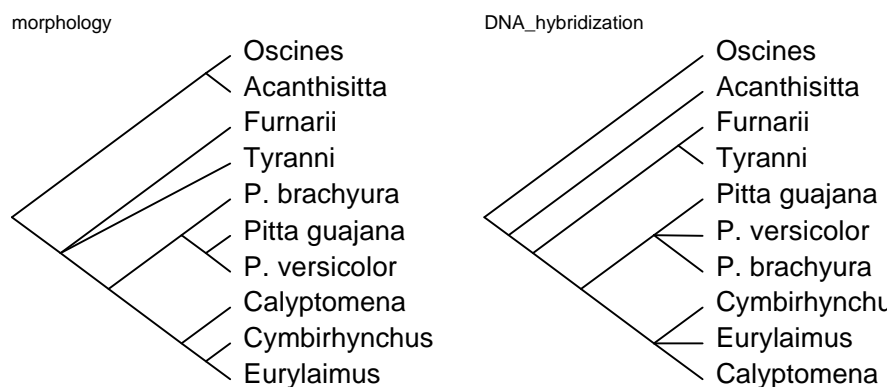
Now, compute an Adams consensus tree for these two trees. Draw it below:

Q How does this consensus relate to the two original trees? Why are taxa *a* and *j* basal on the tree?

Q When do you think Adams consensus might be useful?

Exercise 4

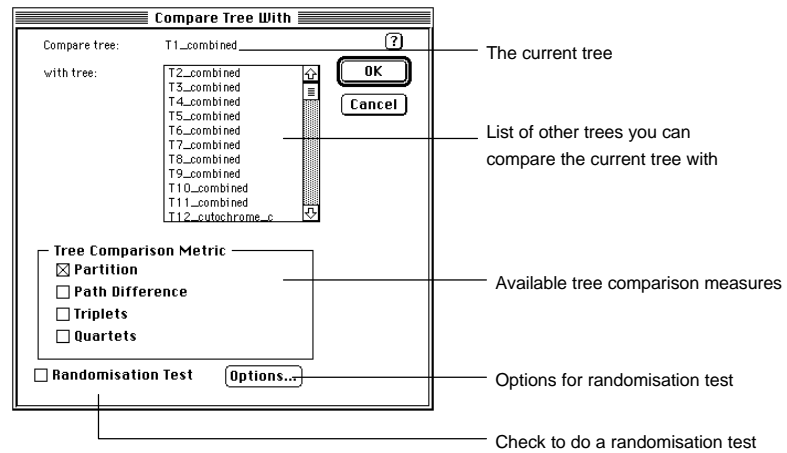
Now we turn to comparing trees. There are a range of tree comparison measures designed to quantify the similarity between two or more trees. This exercise will introduce you to some of these measures. Subsequent exercises will apply these measures to specific problems.



We will return to the example we used to illustrate consensus methods (i.e., the file **Passerines (DNA versus morphology)**). Given these two trees for the same group of birds we can use a variety of methods to quantify their similarity.

To do:

From the **Trees** menu choose the **Compare Tree With** command. You will see a dialog box like this:



This dialog box lists the trees which you can compare. One tree is always the tree currently displayed in the active tree window. The other tree or trees in the file are listed in the list box. You can select any of these trees (for most examples here there will be only one other tree). Below the list box are the various methods available.

For the first exercise, compare the two trees using the **Partition** metric.

Q What value of this metric did you obtain? Can you work out how COMPONENT got this result?

Q How does the value of the partition metric compare relate to the strict consensus tree you got for these trees in Exercise 1?

To do:

Now, repeat the comparison but use the triplet and quartet measures.

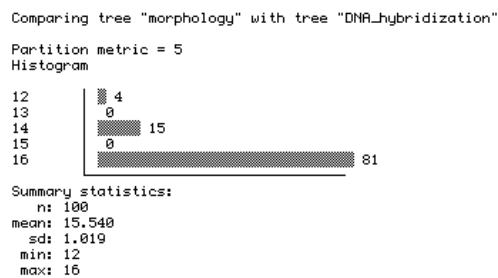
Q The triplet and quartet measures seem to give different answers, why is this? (Hint: think about rooted versus unrooted trees).

Exercise 5

A measure of tree similarity by itself doesn't tell us a great deal unless we know something about the distribution of that measure. In other words, is a value of 10 for the partition metric a little or a lot?

To do:

Still using the file **Passerines (DNA versus morphology)**, repeat the tree comparisons above (Exercise 4) but this time check the **Randomisation test** check box. This time COMPONENT will perform a simple randomisation test to generate the expected distribution of the tree comparison metric under the null hypothesis that the two trees being compared were drawn from a random distribution of trees. You should see something like this:



This histogram shows the expected distribution of the partition metric for random trees with the same number of taxa as those you are comparing. Below this chart is a statement of the probability of getting trees with the observed similarity due to chance alone. In this case it is < 0.01 .

Q What does this p value mean: are you trees similar or different?

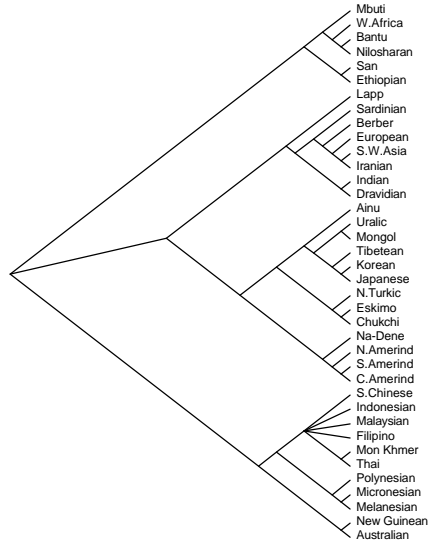
Q What would a p value of 0.50 mean?

Exercise 6

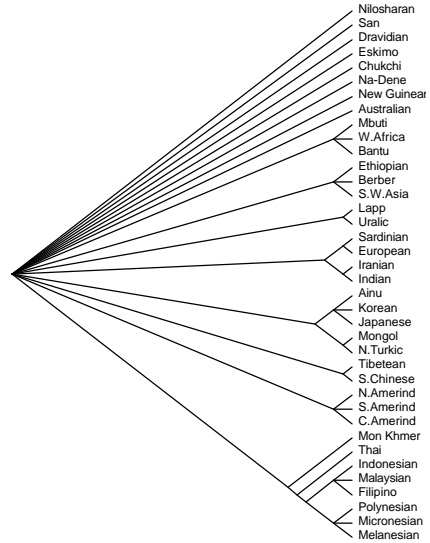
The file **Genes and language** contains two trees for modern human populations, one based on gene frequencies, the other based on languages (Penny *et al.*, 1993). The question naturally arises, "how similar are these two trees?" More precisely, one could ask whether both genetic and language data can tell us something about our recent evolutionary past.

Q Would you expect genetic and linguistic history to be similar or different? Give some reasons for your answer.

PHYLIP_1



PHYLIP_2



To do:

Suppose our null hypothesis is that the gene and language trees are no more similar than we could expect due to chance. We can test this using the **Compare Tree With** command in COMPONENT. Select your preferred measure (or measures) and do the test.

Q Are the gene and language trees more similar than you would expect due to chance alone?

REFERENCES

Nelson, G. (1979). Cladistic analysis and synthesis: Principles and definitions, with a historical note on Adanson's *Famille des Plantes* (1763-1764). *Syst. Zool.* **28**: 1-21.
 Penny, D., Watson, E. E. and Steel, M. A. (1993). Trees from languages and genes are very similar. *Syst. Biol.* **42**: 382-384.
 Sheldon, F. H. and Bledsoe, A. H. (1993). Avian molecular systematics, 1970s to 1990s. *Annu. Rev. Ecol. Syst.* **24**: 243-278.