Chapter 6
Random trees

COMPONENT can generate a variety of trees at "random," and for small numbers of leaves it can generate all possible trees. These trees have several uses:

■ as user trees for phylogeny programs that lack the facility to examine all trees or sets of random trees (e.g., Hennig86 and PHYLIP).

■ to estimate distributions of tree comparison measures.

■ to generate all possible tree shapes.

■ as the basis of statistical tests.

Generating all possible trees

COMPONENT can exhaustively generate all possible trees with a given number of leaves. The trees can be either rooted or unrooted, and labelled or unlabelled. Unlabelled trees are sometimes called "topologies" or "shapes," terms which are used interchangeably in this manual. The following table lists the number of each kind of binary tree for 1-10 leaves.

Table 6.1
Numbers of binary rooted and unrooted, labelled and unlabelled trees for 1-10 leaves

| Leaves	Rooted		Unrooted	
	Labelled	Unlabelled	Labelled	Unlabelled
1	1	1	1	1
2	1	1	1	1
3	3	1	1	1
4	15	2	3	1
5	105	3	15	1
6	945	6	105	2
7	10,395	11	945	2
8	135,135	23	10,395	4
9	2,027,025	46	135,135	6
10	34,459,425	98	2,027,025	11

For example, for six taxa there are 945 possible binary rooted trees, spread over six different tree shapes, whereas the are only 105 unrooted binary trees with two different shapes.

You can generate all possible trees either by using the ALL command in an input file (see Chapter 3) or from the **Generate** menu using the **All trees** command.

Generating all possible trees from the menu

■ Choose **All trees** from the **Generate** menu. The following dialog box will appear:

Figure 6.1
The Generate All
Trees dialog box
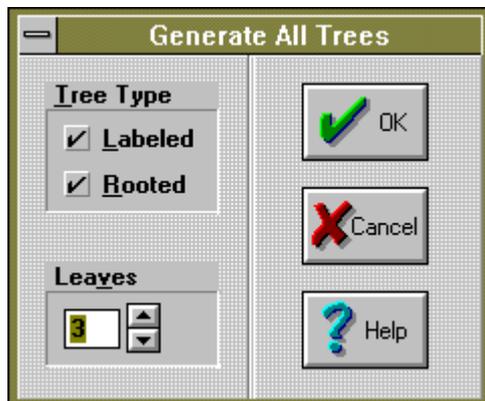
■ Use the controls in the **Tree Type** group to specify what kind of trees you want generated.

When the **Labelled** check box is checked all terminally labelled binary trees are generated (the corresponding option in the ALL command is MODEL = EQUIPROBABLE). Otherwise all unlabelled binary trees are generated (this corresponds to MODEL=UNLABELED).

When the **Rooted** check box is checked then all the rooted binary trees (either labelled or unlabelled) are generated, otherwise unrooted trees are generated.

■ Using either the input box or the spin button you can specify the size of the trees generated. The permissible values are:

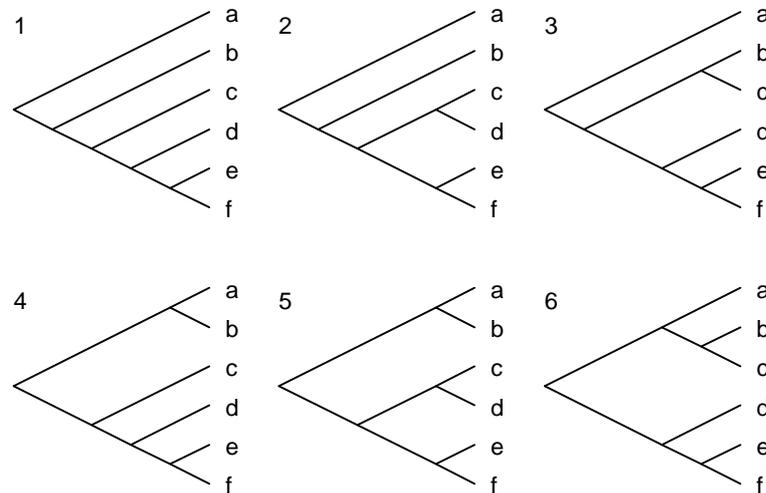|	Labelled	Unlabelled
Rooted	3-7	3-15
Unrooted	4-8	4-19

The size of the labelled trees is limited by tree buffer storage space, the size of unlabelled trees by the limitations of integer arithmetic on a PC.

■ COMPONENT will now generate the requested trees and place them in an untitled Tree window. If you attempt to close this Tree window COMPONENT will first ask you if you want to save the trees to a disk file.

An example

This is the set of trees generated by the **All trees** command when you select unlabelled, rooted trees with six leaves:

*Figure 6.2
Unlabelled six taxon
trees generated by the
All trees command*

The trees above are the six possible rooted shapes for a tree with six leaves. Tree 1 has shape 6_1, tree 2 has shape 6_2, and so on (see Chapter 2, section 2.9).

☞ *Even though the trees are "unlabelled" COMPONENT still labels the leaves: these labels are arbitrary and are a programming convenience.*

If you had chosen to generate labelled trees COMPONENT would have generated all the 945 possible six-leaf trees.

Algorithms

To generate all unlabelled binary trees COMPONENT uses the numbering scheme developed by Furnas (1984). This scheme establishes a one-to-one correspondence between the integers 1 to k, where k is the number of unlabelled trees for n taxa. Each number can be "unpacked" to give the corresponding tree (see Appendix B).

To generate all labelled trees COMPONENT uses a simple recursive algorithm.

Random trees

Since the number of trees rapidly increases with size it quickly becomes impractical to generate all possible trees. COMPONENT can generate a random set of trees drawn from various uniform distributions of trees.

What is a random tree?

A random tree is a tree drawn at random from a set of possible trees. In this context "at random" means that each tree in the set of trees has an equal chance of being sampled. Another way of saying this is that the distribution of trees is "uniform". COMPONENT allows you to sample from three uniform distributions:

■   unlabelled binary trees (rooted and unrooted)

■   labelled binary trees (rooted and unrooted)

■   dendrograms (rooted by definition)

The relationship between these distributions and the distributions described by Simberloff et al. (1981) is discussed below .

You can generate random trees either by using the RANDOM command in an input file (see Chapter 3) or from the menu using the **Random trees** command from the **Generate** menu.

Generating random trees from the menu

■   Choose **Random trees** from the **Generate** menu. The following dialog box will appear:

Figure 6.3
The Random Trees
dialog box
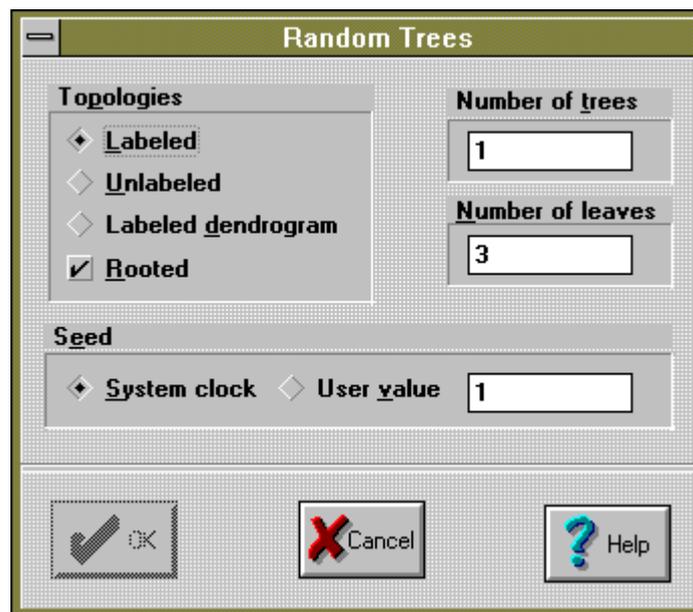
■   Use the controls in the **Topologies** group to specify what kind of trees you want generated.

When **Labelled** is selected the trees generated are drawn from the uniform distribution of labelled binary trees. This corresponds to MODEL = EQUIPROBABLE in the RANDOM command (see Chapter 3).

When **Unlabelled** is selected the trees generated are drawn from the uniform distribution of unlabelled binary trees (this corresponds to MODEL=UNLABELED in the RANDOM command). Note that the trees generated are labelled trees, but the underlying distribution of unlabelled trees is uniform (this corresponds to Simberloff et al.'s [1981] topologies equiprobable model, see below).

When **Labelled dendrogram** is selected the trees generated are drawn from the uniform distribution of labelled binary dendrograms (this corresponds to MODEL=MARKOVIAN in the RANDOM command).

Check **Rooted** to generate rooted trees (the default); otherwise the trees are unrooted. Because dendrograms are by definition rooted trees this option does not apply to dendrograms. The **Labelled dendrogram** check box is disabled if the **Rooted** check box is not checked.

■   Specify how many trees you want (in the range 1-10,000) and how many leaves they have (in the range 3-100 for rooted trees, 4-100 for unrooted trees).

■   By default the seed for COMPONENT's random number generator is taken from the system clock. By selecting the **User value** button you can enter your own seed, between 1 and 2,147,483,647. A set of random trees generated from the same distribution using the same seed will be identical.

■   COMPONENT will generate the random trees and place them in an untitled Tree window. The value of the seed used (whether taken from the system clock or supplied by the user) will be shown in the display buffer.

Simberloff et al.'s (1981) distributions

Using various options you can generate random binary trees corresponding to most of the distributions found in the literature. For example, the three distributions discussed by Simberloff et al. (1981) can be generated as follows:

H_D All labelled binary rooted trees are equiprobable. Select **Labelled** topologies and check the **Rooted** check box.

H_E Labelled cladograms with unlabelled topologies equiprobable. Select **Unlabelled** topologies and check the **Rooted** check box.

H_M Trees generated by a simple dichotomous branching process ("Markovian" model). Select **Labelled dendrograms** (see Page, 1991).

Algorithms

Unlabelled trees

COMPONENT uses Furnas' (1984) numbering scheme which establishes a one-to-one correspondence between the integers 1 to k, where k is the number of unlabelled trees for n taxa. Each number can be "unpacked" to give the corresponding tree (see Appendix B). Using a uniform random number generator (see below) a random number between 1 and k is obtained. This number is then unpacked to obtain a random tree.

Labelled trees

COMPONENT uses the algorithm described by Furnas (1984:203-207).

Dendrograms

Starting with a two leaf tree the program randomly chooses which leaf to add a new leaf. This is repeated until the desired number of leaves have been added. The resulting tree is then labelled by randomly assigning a label to each leaf.

The random number generator

COMPONENT uses a uniform random number generator based on Schrage's (1979) FORTRAN implementation of the linear congruential method. The sequence of random numbers, X, is generated using the recursion:

$$X_{n+1} = aX_n \bmod p$$

where $p = 2^{31} - 1 = 2147483647$ and $a = 7^5 = 16807$. The seed X_0 is supplied either by the user or by the system clock.