

CAFCA

Chapter 5

User-Tree Evaluation

© M. Zandee 1989,1996 All Rights Reserved.

February 1996

THIS PAGE INTENTIONALLY LEFT BLANK



V. USER-TREE EVALUATION

INTRODUCTION

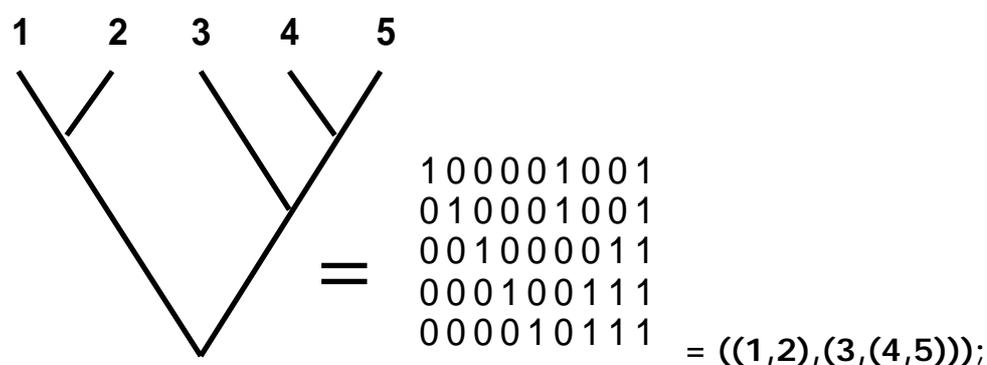
In a user-tree evaluation we have one or more cladograms available for which we want to know how well they fit a particular data matrix. These cladograms may, for instance, represent our intuitive ideas about the historical relations for the taxa or areas or hosts in our study group. A user-tree evaluation, then, provides a tool for comparing, in terms of selection criteria, our intuitive insights with those produced by, for instance, a group compatibility analysis of the same data.

The cladogram can also come from the literature. They are then usually based on a different set of characters than we ourselves have available. In that case we apply a user-tree evaluation as a tool for comparing the effect of different data sets.

Another possible use of user-tree evaluation may result from running a primary analysis on a data matrix containing the 'better' characters that, however, do not give a completely resolved cladogram. After saving, this cladogram can be entered as a user-tree and evaluated against another data matrix containing the 'weaker' characters, and consequently subjected to a secondary analysis on the basis of the 'weaker' characters.

User-tree evaluation can also be applied in the study of coevolution for those cases where independent cladograms are available for hosts and parasites (or genes and taxa: molecular vs morphological data).

User-trees must be available as ASCII (TEXT) files, either in parentheses format (in CAFCA spaces instead of commas are allowed, but the closing semi-colon must be present), or as a binary matrix with the tree topology presented in additive binary coding (see example below).



We can also copy a cladogram matrix from a CAFCA OutputFile to use as a user-tree. Such a cladogram already has the format of a binary matrix.

On the other hand, it may also be very likely that you use PAUP 3.0 for the Macintosh to run your phylogenetic analyses, but that occasionally you get more than one most parsimonious tree and you want CAFCA to calculate the Redundancy Quotients or Compatible Character State Indices for these trees. In that case you **export** your data matrix to **file** in **simple text** format by means of the PAUP **File** menu and you **save** your **trees to file** (include a translation table, but leave out any comments) by means of the PAUP **Trees** menu. Table 5.1 shows an example of the *Second* data matrix, and the tree file as exported by PAUP, ready to be used in a user-tree evaluation by CAFCA.

```

Aus 10000110011
Bus 10000110110
Cus 01000001100
Dus 01001001100
Eus 00100000100
Fus 00010100111
Gus 00010100001
Hus 00010110010
Ius 00101000110
Anc 00000000000

#NEXUS

begin trees; [Treefile saved Monday, June 8, 1992 6:04 PM]
[!>Heuristic search settings:
> Addition sequence: simple (reference taxon = Aus)
> 1 tree(s) held at each step during stepwise addition
> Tree-bisection-reconnection (TBR) branch-swapping performed
> MULPARS option in effect
> Steepest descent option not in effect
> Initial MAXTREES setting = 100
> Branches having maximum length zero collapsed to yield polytomies
> Topological constraints not enforced
> Trees are rooted
> Total number of rearrangements tried = 968
> Length of shortest tree found = 18
> Number of trees retained = 3
> Time used = 1.12 sec
]

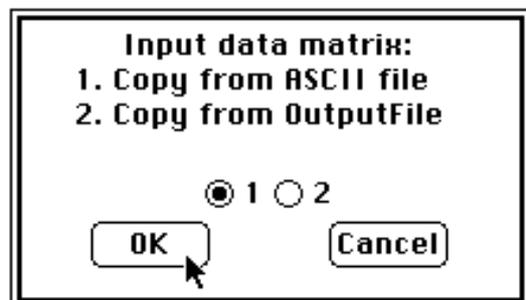
    translate
      1 Aus,
      2 Bus,
      3 Cus,
      4 Dus,
      5 Eus,
      6 Fus,
      7 Gus,
      8 Hus,
      9 Ius,
     10 Anc
    ;
tree PAUP_1 = (((1,2),((6,7),8)),((3,4),(5,9))),10);
tree PAUP_2 = (((((1,2),8),(6,7)),((3,4),(5,9))),10);
tree PAUP_3 = ((((((1,2),8),6),7),((3,4),(5,9))),10);
endblock;

```

Table 5.1 Example of a data matrix as **exported** by PAUP (top) and a tree file (bottom) resulting from a **save trees to file** operation in PAUP, both based on the data matrix as used in chapter 4 on secondary analyses in CAFCA.

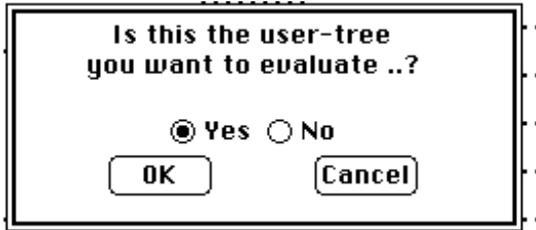
TUTORIAL

1. Select **User-tree Evaluation** from the **Run** menu.
2. We will assume that there is no data matrix present in your workspace. That's why the following dialog appears. If, however, there is a data matrix present because you did first run another analysis, either primary, secondary or biogeographic, CAFCA continues with step 6.

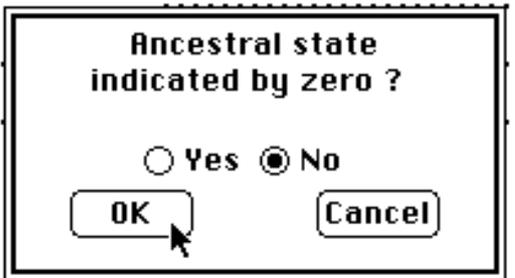


Click **OK** for the default value 1 (Copy from ASCII file).

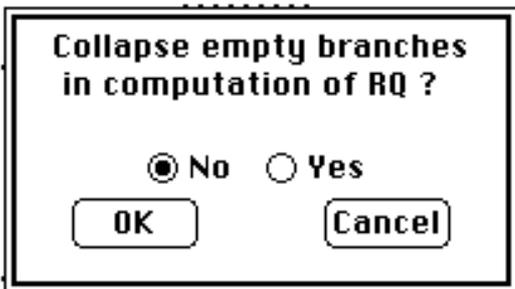
of the user-tree on display. Click **OK** for the default **Yes**.



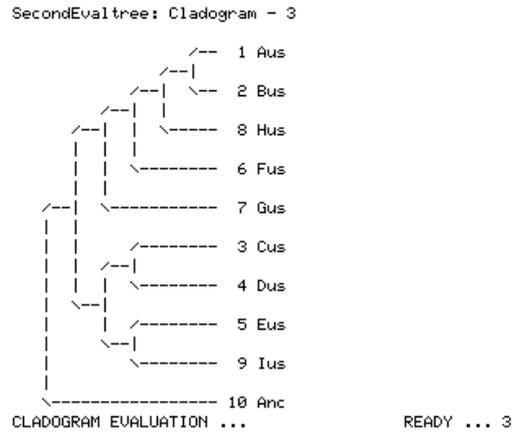
- You will not see the standard CAFCA parameter dialog as most entries are irrelevant now. However, you must indicate whether zero's in the data matrix are to be interpreted as ancestral or not. Click **OK** for the **NO** in the dialog, as we did not interpret zero's to be forced on the root in the primary and secondary analysis either.



- You must also indicate whether empty branches must be collapsed in the computation of the Redundancy Quotient. We take the default **No** here, as in the standard primary or secondary analysis empty branches were not collapsed either.



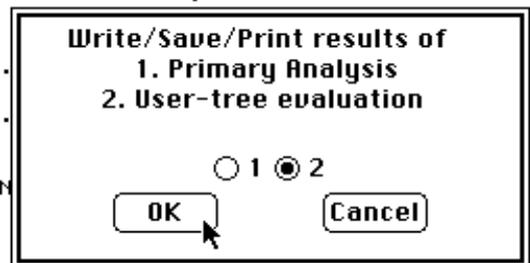
- Evaluation of the user-tree will now start. Its progress is indicated on the screen.



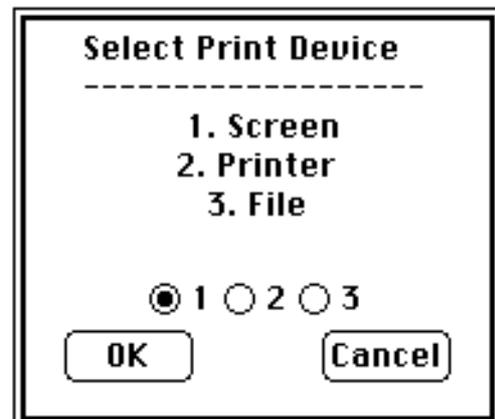
When the evaluation of the user-tree is finished, **READY** will appear on the screen. You can now print the results.

- Select **Diagram Evaluation...** from the **Print** menu.

As a data matrix etc.. is present in the workspace, CAFCA can not know whether you want to print the results of a primary analysis you could have been running, or the results of a user-tree evaluation. Click **OK** for **2 (User-tree evaluation)** in the next dialog.



- Click **1 (Screen)** or **3 (File)** in the **Select Print Device** dialog box.



DISCUSSION OF RESULTS

There is not much to discuss as the output of the cladogram evaluation looks the same as in a primary analysis. Note, however, that you can recognize by the name of the data matrix that these results are from a user-tree evaluation. This name is identical with the name you entered in the dialog (tutorial step 4), except for the addendum **tree**.

```

Selection criteria for cladograms of: Secondtree
Column numbers refer to numbers of cladograms
-----
Row 1 : Total number of homoplasous events
Row 2 : Total number of single origins (Support)
Row 3 : Corrected Extra Length (x1000; CEL: Turner + Zandee)
Row 4 : Total number of state changes (S: Steps)
Row 5 : Redundancy Quotient (x1000; RQ: Zandee + Geesink)
Row 6 : Rescaled Redundancy Quotient (x1000; RQc)
Row 7 : Consistency Index (x1000; CI), with autapomorphy correction
Row 8 : Rescaled Consistency Index (x1000; RC: Farris)
Row 9 : Average Unit Character Consistency (x1000; AUCC: Sang)
Row 10: Homoplasly Distribution Ratio (x1000; HDR: Sang)
Row 11: Compatible Character State Index (x1000; CCSI: Zandee)

      1      2      3
-----
1 |      6      6      6
2 |      6      6      6
3 |    7197    7197    7220
4 |      18     18     18
5 |     521     521     512
6 |     147     147     133
7 |     611     611     611
8 |     417     417     417
9 |     742     742     742
10|     338     338     338
11|     273     273     273

No-Order Limit for Steps, Extra Steps, RQ, and CI:

      S      ES      RQ      CI
-----
    33     22    438    333
    
```

Table 5.2 Result of the evaluation of the user-trees from the file **Second.pau.trees**

Looking at the result of the cladogram evaluation listed in table 5.2 we notice that cladograms number 1 and 2 (fig 5.1) have the highest value of RQ and the lowest for CEL. Comparing this result with that obtained in the secondary analyses we also notice that all cladograms from the PAUP analysis are among the cladograms generated by CAFCA in its primary and secondary analyses on the same data matrix, as shown in chapter 5. PAUP, however, finds these different most parsimonious solutions very quickly and direct (contrary to the indirect nature of the secondary analysis by CAFCA) with either the branch & bound or the exhaustive search option.

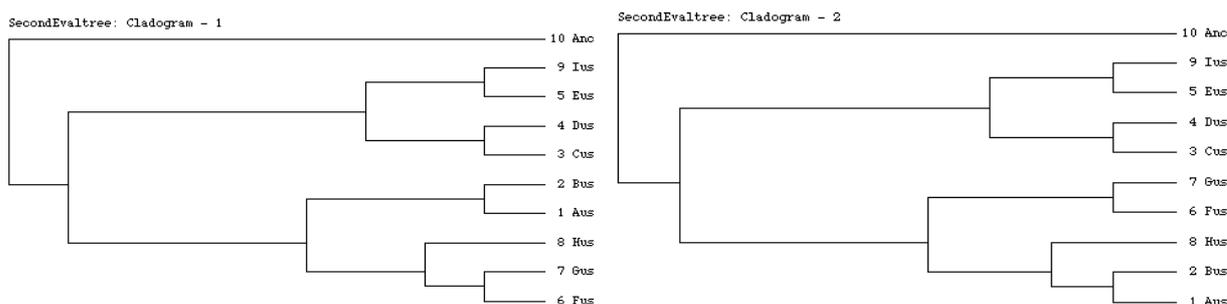


Figure 5.1. Two cladograms from PAUP analysis, used as user-trees, with highest RQ.

THIS PAGE INTENTIONALLY LEFT BLANK