CAFCA

Chapter 2

Overview
of
Menus

February 1996

THIS PAGE INTENTIONALLY LEFT BLANK

II. OVERVIEW OF MENUS

OUTPUTFILE

```
┌─────────────────────────────────┐
│ OutputFile  Run   Print         │
├─────────────────────────────────┤
│ ▓ ► Read...            ⌘O   ▓    │
│ ▓   Save...            ⌘S   ▓    │
│     Delete...                   │
│     UnDelete...                 │
│     Trash...                    │
│     ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─      │
│     Quit...            ⌘Q        │
└─────────────────────────────────┘
```

Output from all types of CAFCA analyses, whether primary, secondary, biogeographical, or a user-tree evaluation, can either be written to or read from, as well as deleted, undeleted, or trashed from so-called APL.68000 component files.

The APL.68000 component file system as used by CAFCA has a fixed name, CAFCA.IO. At first use, CAFCA creates this file system in the CAFCA folder, i.e. the folder where the CAFCA program resides. You can have more than one outputfile system in use by CAFCA, although not in the same folder as all these file systems will carry the name CAFCA.IO.

READ...

This option enables you to read results from a former CAFCA run that has been saved as a component file int the CAFCA outputfile system, CAFCA.IO.

All available files are listed in a select box and labelled by the name of the data matrix.

Once selected and clicked for **OK** all results from the relevant CAFCA run will be read into memory and are then available for further action (e.g., printing, write to ASCII file, etc.).

SAVE...

CAFCA saves all items resulting from an analysis in separate components of a file in the CAFCA outputfile system (CAFCA.IO). The file containing the items carries the name of the data matrix and will be retrievable under that name by the **Read** option in the OutputFile Menu.

After all items have been saved the program returns control to you to resume any action you have in mind.

DELETE...

This option enables you to delete one or more files from the CAFCA outputfile system CAFCA.IO, containing results from CAFCA analyses.

All available files are listed in a select box and labelled by the name of the data matrix. After selection and click for **OK** the file is marked for deletion.

The space occupied by a file marked for deletion can be claimed by another file during a **Save** action.

The deletion mark can be undone by the **UnDelete** option in the same OutputFile menu, but only if the file space remained unclaimed.

UNDELETE...

This option enables you to un-delete one or more files from the CAFCA outputfile system CAFCA.IO, containing results from CAFCA analyses.

All available files presently marked for deletion will be listed in a select box and labelled by the name of the data matrix. After selection and click for **OK** the mark for deletion for the relevant file will be undone.

The deletion mark can be set (again) by the **Delete** option in the same OutputFile menu.

TRASH...

Trash removes a file from the list of OutputFiles in the CAFCA outputfile system CAFCA.IO.

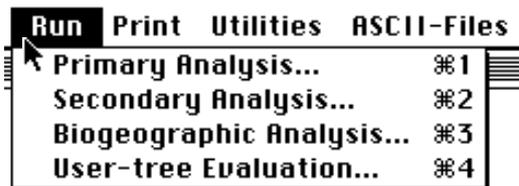Only files that are marked for deletion can be removed.

All files that are marked for deletion will be presented in a select box. When selected and clicked for **OK** the file will be removed.

Once removed, a file cannot be recovered.

QUIT

Quit quits CAFCA. You will be prompted for confirmation.

RUN

Run	Print	Utilities	ASCII-Files
Primary Analysis...			⌘1
Secondary Analysis...			⌘2
Biogeographic Analysis...			⌘3
User-tree Evaluation...			⌘4

CAFCA employs the methods of group-compatibility and component-compatibility to run a cladistic analysis of a data matrix. The data matrix may either be binary, multi-state, or mixed binary/multi-state. Missing values are allowed and should be indicated by a negative integer, or a question mark.

Characters may either be polarised by indicating the (putative) ancestral state (zero), or polarised + ordered by means of (partial) additive binary coding, or kept neutral (no polarity, no order), or polarised and linearly ordered upon request (multi-state characters only). CAFCA has no options for step-matrices.

Taxa may either all belong to the ingroup, or the outgroup(s) may be included and (interactively) indicated by the user, or deduced from the data matrix (full zero row). The input data matrix must be available either as an ASCII (=TEXT) file (i.e., in e.g., TeachText™ format), or as part of an OutputFile, or you may use CAFCA's built-in editor to enter a data matrix.

The examples folder on the distribution disk contains examples of valid input files for CAFCA (see also chapter. 3, p 3.2)

PRIMARY ANALYSIS...

A primary analysis involves a cladistic character analysis; i.e., a data matrix is subjected to the group compatibility method to find cladograms for taxa.

This is a four-step procedure:

1. Recognition of groups of taxa (clada), to serve as building-blocks for cladograms. Clada are either partially or strictly monothetic sets, they can be based on all possible additive binary coding for each character, or are generated using three-taxon-statement permutations.

2. Search for sets of clada by a branch and bound algoritm, such that all clada in a set are mutually compatible (either include or exclude each other), and these sets are the largest possible (maximal cliques).

3. Each maximal clique corresponds to a cladogram. Given T taxa and cliques of size 2T-1, completely resolved (= fully dichotomous) cladograms are present.

4. All characters are put on the cladograms to evaluate the quality of each cladogram in terms of number of homoplasious events (= multiple origin of character states), support (= single origin of character states), total number of steps, the corrected extra length (CEL), redundancy quotient (RQ), consistency index (CI), the rescaled consistency index (RC), the average unit character consistency (AUCC), the homoplasy distribution ratio (HDR), and the compatible character state index (CCSI). The best cladograms for a chosen criterion enter the (CAFCA) selection.

SECONDARY ANALYSIS...

Secondary analyses serve to resolve polychotomies in cladograms resulting from a previous primary analysis where the set of clada contained insufficient information to achieve complete dichotomy.

Input files for a secondary analysis are present in CAFCA output of a primary analysis saved as an OutputFile and will be retrieved automatically when you click a name in a selection of OutputFiles presented after starting the secondary analysis. Input files for a secondary analysis are already present if you choose to run a secondary analysis immediately after a run of a primary analysis is finished.

In a secondary analysis of a cladogram the program will check all nodes for dichotomy. If a polytomous node is found all leaf side neighbouring nodes are isolated from the data matrix (in case of single terminal taxa) or from the internal node character state list (in case of groups of terminal taxa).

On this selection of clada a primary analysis is run using all characters. New branching patterns are kept in memory when found, while the other nodes are checked and treated in the same manner if polytomous. If for each of these nodes more than one better resolved solutions are found, this will result in multiple solutions for the complete cladogram.

BIOGEOGRAPHIC ANALYSIS...

Biogeographic analyses are run to explore by means of the component compatibility method the relations between the phylogeny of a group of taxa, or different phylogenies of different (unrelated) groups, and the geographical distribution over areas of the taxa involved, in order to resolve the historical relationships of these areas or biota's.

In fact, any co-evolutionary pattern can be studied for its historical implications by this type of analysis, be it parasites on hosts (hosts as areas for parasites)

or genes (characters) in taxa (taxa as areas for genes).

The method used is component compatibility analysis; it takes the same procedural steps as a primary character analysis.

Input consists of either one or more binary area-data matrices (areas x nodes). When an area-data matrix is not available the program can derive one from a binary distribution matrix for taxa (taxa x areas) and a cladogram matrix (taxa x nodes), either in parentheses' notation or as a binary matrix.

All input must be available as ASCII files (i.e., in e.g., TeachText™ format) or be present in an OutputFile.

In case of a generalised analysis several area-data matrices (one for every taxon cladogram) must be available, either as ASCII files or in an OutputFile. They cannot be derived from distribution and cladogram matrices in an intermediate step; i.e., standard analyses must precede a generalised one.

USER-TREE EVALUATION...

User-tree evaluation can take place when you have entered a data matrix and one or more cladograms. The cladogram is, as a rule, not derived from the data matrix itself but usually comes from the literature, or is intuitive (see also below).

Characters from the data matrix will be put on the cladogram by parsimony mapping; i.e., such that for all characters taken together a minimum amount of character state transitions is sufficient to explain all character state distributions (= minimum step solution).

The input data matrix must be an ASCII file or be present in an OutputFile.

The user-tree (cladogram) must be present as an ASCII file of either one binary cladogram matrix, or one or more cladograms in parentheses' notation.

The user-tree need not be completely resolved. If it contains polychotomies it can be subjected to a secondary analysis after evaluation.

A possible use may come from a primary analysis on the best characters that,

however, did not result in a fully resolved cladogram. After saving, this cladogram can be entered as a user-tree to be evaluated against another data matrix containing the weaker characters, and subsequently subjected to a secondary analysis on the basis of the latter data matrix, etc...

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).

PRINT
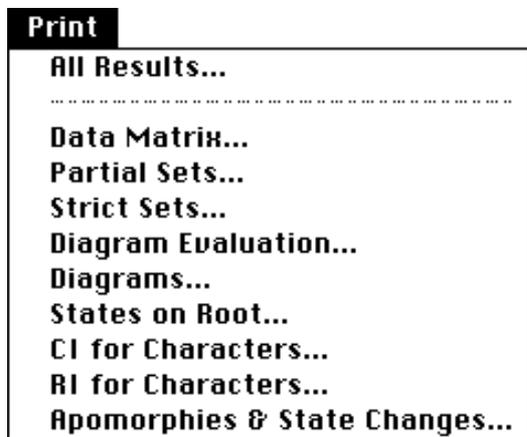
Output on the most relevant items generated by a CAFCA run can be printed to screen, or to printer (Any Appletalk™ connected printer, Apple ImageWriterII™, or HP's DeskJet™), or to (ASCII) file. These items can either be printed together in a coherent manner (**All Results** option) or separately.

Separate printing may come in handy especially for those items that do not pass the selection criterion, eventually (e.g., cladograms, apomorphies and state changes), and which are therefore not included in the **All Results** print.

Printing to a (laser)printer directly is reasonably up to Macintosh standards in this version of CAFCA. However, you may want to print to file first and use your favourite word processor when the output needs editing before printing.

ALL RESULTS...

This option enables printing of the most relevant results of a CAFCA run. Results can be printed either to screen, to printer, or to (ASCII) file. You can select the output device by way of a select box. The following items will be printed:
• Binary data matrix
• Column partitioning vector
• Multi-state data matrix (if present)
• Partial (or strict) monothetic sets of taxa (or areas)
• Partial (or strict) monothetic sets of character states
• Character states present at the root of the selected cladograms
• Cladogram evaluation data (# homoplasious events, support, CEL, steps, RQ, CI, RC, AUCC, HDR, CCSI)
• Topologies of selected cladograms, with labelled nodes
• Apomorphic character states per selected cladogram
• Compatible character states per selected cladogram
• Character state changes per selected cladogram

DATA MATRIX...

This option enables the printing of the data matrix that will be or has been subjected to a cladistic analysis.

If a multi-state data matrix is present, or deducible by means of the column partitioning vector, it will be printed together with its binary expression.

PARTIAL SETS...

This option enables the printing (to screen, printer, or file) of the partial monothetic sets of taxa (clada) or areas (components) as well as the corresponding partial monothetic sets of character states (or monophyletic groups).

Partial monothetic sets (of taxa or areas) are defined by sets of unique (separate) character states.

STRICT SETS...

This option enables the printing of the strict monothetic sets of taxa (clada) or areas (components) as well as the corresponding strict monothetic sets of character states (or monophyletic groups).

Strict monothetic sets are defined by unique combinations of character states (neither one of the separate states needs to be unique).

DIAGRAM EVALUATION...

This option enables the printing of the evaluation data available for cladograms, area-cladograms or user-trees. These evaluation data concern the following items:

1. The number of homoplasious events (= characters requiring extra steps in surplus of their theoretical minimum to explain the distribution of their states on the diagram).
2. The number of single origins (= support = character states requiring but a single origin to explain their distribution on the diagram).
3. The Corrected Extra Length (CEL; Turner & Zandee), i.e., the number of extra steps [in the cladogram as compared with the theoretical minimum] plus 1 minus the average unit retention index (ri: Farris, 1989) for the n characters in the cladogram.
4. The total amount of character state changes (= steps) on the diagram.
5. The Redundancy Quotient (RQ; Zandee & Geesink). The degree in which a diagram represents an optimum distribution of character states in the context of hierarchical information theory.
6. The Consistency Index. The ratio of the minimum number of steps required to explain all character states as single origin events (**M**), and the actual number of steps needed in the most parsimonious explanation of character states distributions on the diagram (**S**).

7. The Rescaled Consistency Index (RC: Farris, 1990). The ratio of the difference between the number of steps in a completely unresolved diagram (**G**) and the number of steps needed in the most parsimonious explanation of character states distributions on the actual diagram (**S**), and the difference between **G** and the minimum number of steps required to explain all character states as single origin events (**M**), times the ratio of **M** and **S**.
8. The avergage unit character consistency (AUCC; Sang, 1995), calculated as the average of total unit character consistencies:

$$AUCC = \left[\Sigma\ c(i)\right] / n$$

where c(i) is the unit character consistency (UCC) of character i (Kluge & Farris, 1969). AUCC is maximized when homoplasy is distributed most asymmetrically, i.e. all the homoplasy occurs in one character. AUCC actually varies in the interval [CI, 1] (Sang, 1995).
9. The homoplasy distribution ratio (HDR; Sang, 1995) is calculated as the ratio of the homoplasy distribution index (HDI; Sang, 1995)) to the homoplasy index (HI; Sang, 1995)

HDR = HDI / HI

where

HDI = AUCC - CI, and

HI = 1 - CI

Since, whenever homoplasy occurs, the AUCC is smaller than 1, AUCC-CI is smaller than 1-CI. Thus, the HDR exists in the interval [0,1] (Sang, 1995). According to Sang (1995), HDR measures level of homoplasy and its distribution and can be a relalively accurate indicator of reliability of parsimonious cladograms. Although a cladogram may have a relatively low CI, it still can be considered reliable if its HDR is high, because in such a case the homoplasy is concentrated in a small group of cladistically unreliable characters.
10. The compatible character state index, **CCSI**, is calculated as the ratio

of the number of compatible character *states*, i.e., the character *states* that are identical with components of the cladogram, and the total number of character *states*. Autapomorphies are not excluded, although always consistent and thereby inflating the value of CCSI. This also applies to uninformative character states, i.e. states that are present in all the taxa concerned. For polarized multi state characters, the state that is assumed to be the start of the transformation series is also considered uninformative.

CCSI varies in the interval [0,1]. It is zero for a bush, and reaches its maximum value when all character states are consistent with the cladogram. CCSI is a measure of "goodness of fit" for primary homologies.

CCSI is related to OCCI (Rodrigo, 1991). OCCI counts the number of fully compatible characters for a cladogram. OCCI is meant to be used as a discriminator for MPT's.

CCSI also measures taxonomic efficiency in the sense of Rodrigo (1991); "the ease of which we may identify taxon membership in practice". Character states consistent with the cladogram have a single origin and as such represent unique identifiers for clades.

11. The no-order lower limit for 4, 5, and 6; i.e., their value in a completely unresolved diagram (one polytomy).

DIAGRAMS...

This option enables the printing of cladograms or area-cladograms resulting from a cladistic analysis. The index numbers of the available diagrams will be presented in a select box in which you can click one or more numbers and **OK** to select one or more diagrams to be printed.

Printing will be either to screen, printer, or file, as indicated by you.

STATES ON ROOT...

This option enables the printing of the character states present at the root of each cladogram or area-cladogram.

These states can be inferred to represent (one of) the most parsimonious estimates of character states present in the hypothetical ancestral group at the rootnode.

C.I. FOR CHARACTERS...

For each diagram resulting from the present analysis a list of the consistency indices for all characters present in the data matrix will be given, as well as the average CI for characters over all cladograms.

The consistency index is the ratio of the minimum number of steps required to explain all character states as single origin events, and the actual number of steps needed in the most parsimonious explanation of the distribution of character states on the diagram.

R.I. FOR CHARACTERS...

For each cladogram resulting from the present analysis a list of the retention indices for all characters present in the data matrix will be given, as well as the average RI for characters over all cladograms.

The retention index for a character is expressed by

$$\mathbf{RI} = \mathbf{g} \text{ - } \mathbf{s} \text{ / } \mathbf{g} \text{ - } \mathbf{m}$$

where **m** represents the minimum number of steps required to explain all character states as single origin events, **s** represents the actual number of steps needed in the most parsimonious explanation of the distribution of character states on the cladogram, and **g** represents the number of steps for the character on an unresolved cladogram.

APOMORPHIES & STATE CHANGES...

This option allows you to select one or more diagrams and subsequently list the apomorphies, character state compatibility's, and character state changes present in each of the selected diagrams.

The character state changes represent the most parsimonious explanation of the distribution of character states over the terminal taxa in the diagram . When more alternative equally parsimonious explanations are possible only one is listed, i.e., the one favouring reversals over parallelisms (= accelerated transformation). In all other cases the choice is most parsimonious but arbitrary.

State changes that can be considered evolutionary novelties for a group of taxa enter the list of apomorphies. Note that apparently identical state changes can acquire apomorphic status more then once, even in the same lineage (= contiguous linear sequence of nodes of the diagram), although in that event such state changes can no longer be considered to represent homologies.

Character states that are fully compatible with the cladogram are also listed. These compatibility's are not identical with steps on the cladogram (no single origins, nor apomorphies), just congruencies between groups in the cladogram and character states.
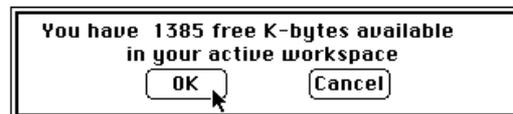
UTILITIES

The options in this menu are a mixed collection of utilities that may serve you in several ways to extend its analyses beyond CAFCA or facilitate inspection of some limits imposed by either hard- or software.
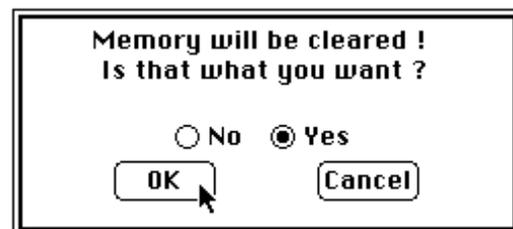
SHOW FREE MEMORY

This option shows the available free memory in the active workspace.

CLEAR MEMORY...

This option will delete all objects (variables) from the active workspace (RAM). You will be prompted for confirmation.
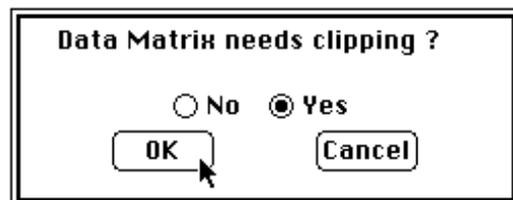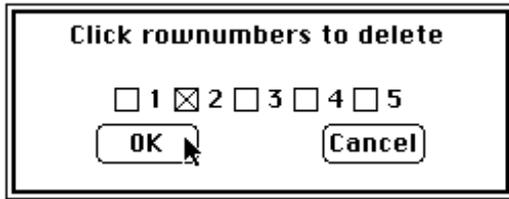

CLEAR SCREEN

This option will clear the display of its current contents.
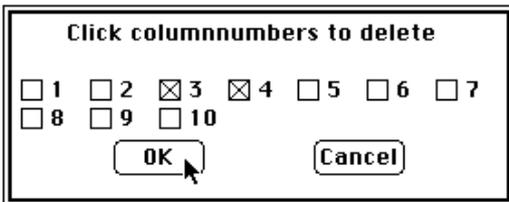
CLIP DATA MATRIX...

This option enables you to clip (= delete) rows and/or columns from the data matrix.

Row and column numbers will be presented in a select box from which you can click the numbers to be deleted.

Click rownumbers to delete

☐ 1 ☒ 2 ☐ 3 ☐ 4 ☐ 5

[OK]   [Cancel]

N.B. Once removed, these rows and columns can not be recovered.

Click columnnumbers to delete

☐ 1 ☐ 2 ☒ 3 ☒ 4 ☐ 5 ☐ 6 ☐ 7
☐ 8 ☐ 9 ☐ 10

[OK]   [Cancel]

The multi-state data matrix will be used for clipping if the analysis started from a multi-state character matrix, or if the construction of such a matrix is possible using the column partitioning vector.

OPTIONS AVAILABLE IN THE EXPORT DATA MATRIX ITEM

Export data matrix in format:

○ PAUP-(PC)   ⦿ NEXUS
○ McClade-2.1   ○ NTSYS
○ Hennig86

[OK]   [Cancel]
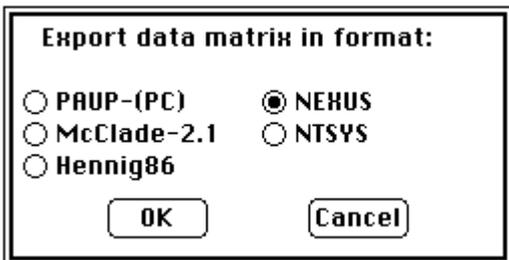
PAUP FORMAT...

The program will use the namelist of the taxa (or areas) and join it with either the binary or the multi-state data matrix to build a file that can be used as input for the PAUP program, PC version 2.4.1.

A PAUP parameter and data statement are added to this new file. If an ancestor is indicated in the active workspace this indication is also present in the file (*).

NEXUS FORMAT...

This option generates a file compatible with the NEXUS file format. It contains a data- and assumptions-block with the data matrix and the names for the taxa.

You are prompted whether the NEXUS input file should be based on the binary or the multi-state expression of the data matrix. You are also prompted whether the file should contain a TREES block including the cladograms as selected by CAFCA. You must also provide a name for the NEXUS file when asked.

This file-type can be used by PAUP vs 3.x and MacClade 3.0.x

SPECTRUM FORMAT...

This option generates an inputfile in NEXUX format for Michael Charleston's program SPECTRUM (http://www.taxonomy.zoology.gla.ac.uk/mike/spectrum/spectrum.html). Normally, sequence data are used as input for this program. Here, instead of the data matrix the list of clada (components) is used. This enables you to make use of CAFCA's different options to generate these building blocks for cladograms and then use the Hadamard transform approach of SPECTRUM to find the closest tree for these clada.

MACCLADE FORMAT...

The program will use the namelist of the taxa (or areas) and join it with either the binary or the multi-state data matrix, as well as the parentheses' notation of the selected cladograms (up to a maximum of 25), to build a file that can be used as a data file for the MacClade program, vs 2.1. You are prompted whether the MacClade input file should be based on the binary or the multi-state expression of the data matrix. You must also provide a name for the MacClade input file when asked.

NTSYS-PC FORMAT...

This option writes the namelist for taxa (or areas) plus the data matrix and an appropriate parameter statement to disc, as an ASCII file in NTSYS format.
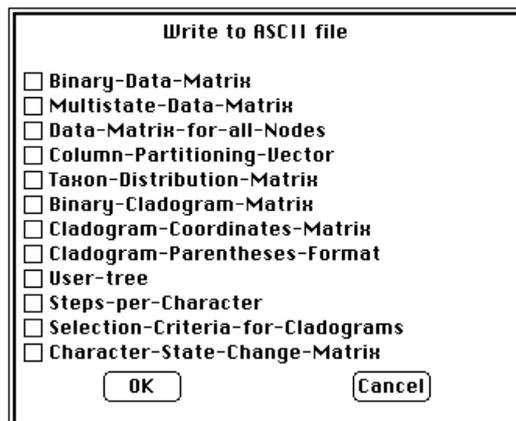
You must provide a name for this file when prompted to do so. This file can be used directly as input for the NTSYS program package.

HENNIG86 FORMAT...

The program will use the namelist of the taxa (or areas) and join it with either the binary or the multi-state data matrix to build a file that can be used as a procedure file for the HENNIG86 program.

A 'ccode -.;' statement is included to let the characters be treated as unordered.

You are prompted whether the HENNIG86 input file should be based on the binary or the multi-state expression of the data matrix. You must also provide a name for the HENNIG86 input file when asked.
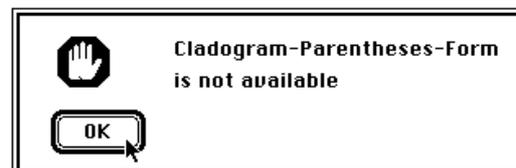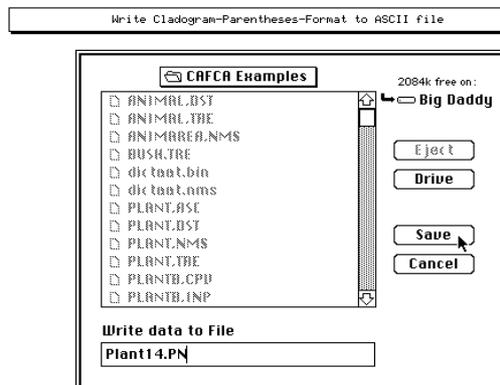
ASCII

This option enables you to deal with ASCII (= TEXT) files in several ways. You can either choose to **Write** items generated by a CAFCA run to disk, or **View** ASCII files present on the default drive for inspection (e.g., before using them as input files for a CAFCA run), or **Delete** ASCII files that are present on the default drive.

WRITE FILE...

This option enables you to write several CAFCA files from the active workspace (RAM) to disk (ASCII files) for future use, e.g., as input for further analyses.

The items that can be written to an ASCII file will be presented in a select box in which you can click an item and **OK** to select it and write it to file. The items that can be written to file are:

Write to ASCII file

☐ Binary-Data-Matrix
☐ Multistate-Data-Matrix
☐ Data-Matrix-for-all-Nodes
☐ Column-Partitioning-Vector
☐ Taxon-Distribution-Matrix
☐ Binary-Cladogram-Matrix
☐ Cladogram-Coordinates-Matrix
☐ Cladogram-Parentheses-Format
☐ User-tree
☐ Steps-per-Character
☐ Selection-Criteria-for-Cladograms
☐ Character-State-Change-Matrix

[OK]   [Cancel]

If for one reason or another the type of data you selected is not available you will be alerted.

Cladogram-Parentheses-Form is not available

[OK]

In other cases you will get a file selector box where you can enter a name for the file that will contain the data previously selected.

Write Cladogram-Parentheses-Format to ASCII file

🖿 CAFCA Examples   2084k free on:
                          ↳ ⬜ Big Daddy
▢ ANIMAL.DST
▢ ANIMAL.TRE
▢ ANIMAREA.NMS   [Eject]
▢ BUSH.TRE
▢ dictaat.bin   [Drive]
▢ dictaat.nms
▢ PLANT.ASE
▢ PLANT.DST
▢ PLANT.NMS   [Save]
▢ PLANT.TRE   [Cancel]
▢ PLANTB.CPU
▢ PLANTB.INP

Write data to File
Plant14.PN

VIEW FILE...

This option enables you to view the contents of ASCII files present on disk in the current default drive.

The names of all ASCII files will be presented in a select box in which you can click a name to select a file to get a view of it on the screen in a separate window.
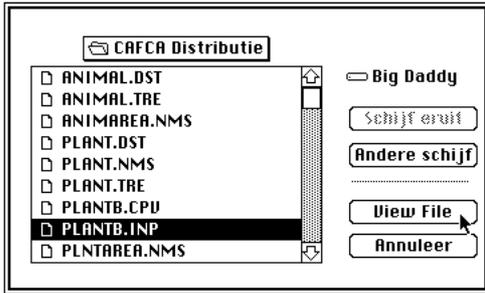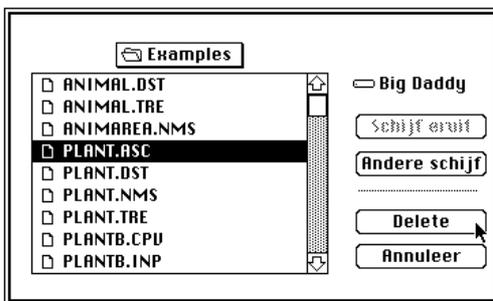
This option thus facilitates the inspection (but not the editing) of ASCII files before they are used as input for an analysis, eventually.
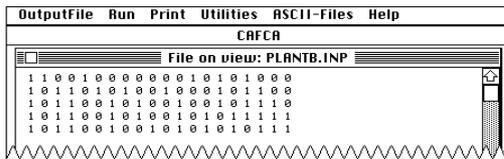
DELETE FILE...

This option enables the deletion of (ASCII) files.

The names of all (ASCII) files in the default volume will be presented in a select box from which you can click a name to select a file to delete it.

You will be prompted for a confirmation.

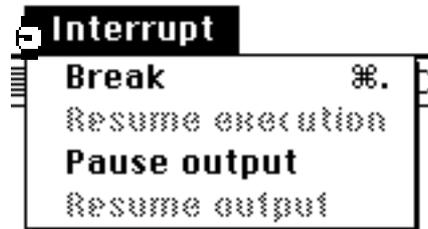## HELP

Speaks for itself.

ABOUT CAFCA...

Displays a copyright message and your registration (unless you downloaded your copy from the web) .

INTERRUPT

This menu will appear as the only available item during the execution of all options from the **Run** and **Print** menus.

BREAK

Clicking this option during any **Run** or **Print** session (fig 2.13) will halt execution as soon as the APL attention check routine can intervene in any running process. All processes will be halted (i.e., not kept pending!) except the main event loop; thus you will be returned to the standard CAFCA screen. Resuming execution of current processes will not be possible, but you may try to print (again) any results available until the break, or

you may start anew with another selection from any menu.
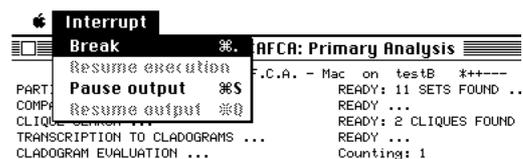
PAUSE OUTPUT

This option will pause output appearing on the screen during any **Print** option.

RESUME OUTPUT

This option will resume output appearing on the screen during any **Print** option.

THIS PAGE INTENTIONALLY LEFT BLANK