MUST2000

The new version of MUST has one major advantage with respect to version 1.0: all the computer memory (or RAM) is now available and not only to the 640KB of DOS memory. This allows handling very large data sets (several millions of nucleotides). In addition, many new options have been added since version 1.0. For each program, a detailed description is provided below. Please note that MUST has been modified in order to work with the Unix program, alibaba, that allows the automatic retrieving of all the sequences detected by a blast search. Finally, the new version of MUST only concerns the management of aligned sequences (the FAST part). The database management (the USEB part) will no longer be improved except for the correction of the Y2K bug.

This MUST2000 version is still under testing and therefore likely contains some bugs. If you discover any bug, please inform Hervé Philippe (herve.philippe@umontreal.fr) while giving as much information as you can. If you have any ideas about new options, do not hesitate to ask Hervé Philippe about a possible implementation of them.

Download a version of MUST2000 for computer with few memory (40MB) (the maximum of usable species is 2000)

Download a version of MUST2000 for computer with the standard memory size (64MB) (the maximum of usable species is about 3000–5000 for ED)

Please note that the descriptions below are just for the main improvements. You should have a look at the Help files (F1 key) to see all the new options.

MAKETREE
This program must be executed right after you have downloaded MUST2000, because the way of recording the systematic frames has changed. Previously, a file SYSTEME.LST contained the numbers and the names of all the systematic frames that are contained in the files SYS0, SYS1 … SYSn. This resulted in problems when one would like to transfer a systematic frame from one computer to another. Now, each systematic frame is recorded in a separate file with the extension FRA in the directory MUSTEXE. This file can be easily copied from one computer to another. Another improvement concerns the number of groups per page, from 23 to 28.

ED
There are many new options, in particular the display of the species in color according to their classification in the systematic frame (these colors are conserved in NET and in TREEPLOT), an increased number of lines for the alignment (up to 50) and a semi-automatic alignment.

When the systematic frame is displayed:
F10: Selects a species by giving a part of the species name
SHIFT+F2: Deselects all the species of the current group
CTRL+F2: Cancels the choice of species in progress
SHIFT+F5 (instead of SHIFT+F3): Writes the selected sequences into a file usable by the program NET and activates this program automatically
SHIFT+F6: Writes the selected sequences into a file usable by the program NET

SHIFT+F10: Allows you to write the identifiers with their name's groups in a file .NOM usable by prep-hennig in order to apply the Slow-Fast method (Brinkmann & Philippe 1999)

F11, SHIFT+F11: Selects the color of the leaves in the systematic frame from the list of group's names (F11) or directly from the systematic frame(SHIFT+F11)

When the species of a group are displayed:

A + indicated that the species have already been selected in another ensemble.

F4 and F5: Looks for the best similarity between the first 40 characters of the current species and all the other sequences. The 15 best scores are displayed and you can automatically select the number of species of your choice by decreasing similarity values. This option allows quickly selecting the species that is more similar to a new sequence to be aligned.

F8: Automatically changes the group of the current species using the information contained in the file .INF (generated by alibaba)

F9: Changes the name of a species and thus moves it from its current group to any other with modification of the current file .NOM

F10: Moves a species from its current group to any other with modification of the current file .NOM

F12: Changes the identifier of the active sequence and saves the new name and the ancient group name in the current file .NOM

CTRL+DEL: Permanently removes a sequence from the current file .ALI and puts it in a file .NON [compliant with the program alibaba (Philippe Lopez unpublished)]. This option is particularly important to eliminate redundant, incorrect or partial sequences from your file . Such deletion is necessary because alibaba retrieves ALL the homologous sequences. The most interesting aspect of this option is that this sequence will not be added again by alibaba when you will update your alignment.

? : Indicates the full name of the species and, if possible, displays available information about this sequence in a file .INF (generated by the program alibaba)

When the alignment is displayed:

! : Looks for the first gap in the first species which is not shared by the current species, automatically acquires the first 20 characters at the right of this position, searches for similarities with the first displayed sequence , and automatically adds the required number of gaps if the first score is better than the second score by more than 5 identities. If not, it adds the number of gaps required by the best score.

a : Applies the option ! up to the end of the sequence. This option allows an automatic alignment between two closely related sequences.

: : Automatically acquires the first 20 characters at the right of the cursor and searches for similarities with the first displayed sequence (from -200 to +200 positions with respect to the cursor)

; : Automatically acquires the first 10 characters at the right of the cursor and searches for similarities with the first displayed sequence (from -100 to +100 positions with respect to the cursor)

, : Automatically acquires the first 20 characters at the right of the cursor and searches for similarities with the first displayed sequence , and, if possible, adds the required number of gaps. The cursor does not move after this operation.

ALT+F4: Acquires the first 20 characters at the right of the cursor, with possible modification of that sub-sequence and searching for similarities with all the other sequences (the 15 best scores are displayed)

ALT+F5: Acquires the first 20 characters at the right of the cursor, with possible modification of that sub-sequence and searching for similarities with all the other sequences that are selected (the 15 best scores are displayed)

SHIFT+F3: Remove from a sequence any number of the first blanks/stars at the left of the active position of the cursor

SHIFT+F4: Removes all the gaps at the right of the cursor

SHIFT+F8: Writes the consensus sequences in file .NET

SHIFT+F9: Writes all the selected sequences into a file usable for ALI (be careful with this option to not overwrite a correct file)

SHIFT+F10: Writes the identifiers with their name's groups in a file .NOM usable by prep-hennig

ALT+F12: Modification of the primary sequence of the current species (be careful with this option, there is no control at all and no backup. The only issue if you make a mistake is to quit the program without saving)

CTRL+DEL: Permanently removes a sequence from the current file .ALI and puts it in a file .NON [compliant with the program alibaba (Philippe Lopez unpublished)]

? : Indicates the full name of the species and, if possible, displays the available information about this sequence in a file .INF (generated by the program alibaba)

c : Looks for the first difference between the current species and the first one (option currently valid only if all the species are selected)

d : Discards the current species from the selection

f : Displays the current sequence in FASTA format

p : Replaces stars at the beginning and at the end of the current sequence by spaces (useful for partial sequence)

s : Removes all spaces and stars from the current sequence (useful to perform an alignment again)

x : Shortens the current sequence on the left or on the right of the cursor. The shortening is indicated by XXX.

+ : The first displayed sequence becomes the last one but the cursor does not move (i.e. remains on the same line)

- : The last displayed sequence becomes the first one but the cursor does not move (i.e. remains on the same line)

F9: Changes the identifier of the active sequence

F10: Change the group of the current species (and therefore its color)

F12: Change the identifier of the active sequence and saves the new name and the ancient group name in the current file .NOM

NET

Several new options have been implemented in NET, such as the computation of the NJ tree, the possibility to compute the number of differences (instead of the frequency, via the key F6 in the main menu), the use of colors for displaying the identifiers and the possibility to display 60 species on the alignment for selecting unambiguously aligned positions.

When the sequences are displayed:
On bottom left, two values appear: the number of the position in the file .ALI (as previously) and the absolute number of characters of the first sequence (i.e. without counting gaps)

CTRL+F1: Takes the complementary boundaries, i.e. the red selection becomes white, and the white becomes red.
CTRL+F2: Load the boundaries of the sequences that are retained as a function of the absolute numbering of the first species on the screen from a file .TXT
CTRL+F3: Save the boundaries of the sequences that are retained as a function of the absolute numbering of the first species on the screen in a file .TXT
CTRL+F4: Removes the sites selected by the Index (PRESENT1, PRESENT2, CHI2 and CODON), i.e. the "I" selection will become red.
ALT+F1: Switches the number of lines of the screen from 30 to 60 and vice versa

TREEPLOT

Several new options have been implemented in TREEPLOT, such as the use of colors for displaying the identifiers, the possibility to display 60 species per screen, the possibility to display 2 bootstrap values for each node from file .GRP (generated by various programs such as NJBOOT, PAUP and PROTML) and the writing of lists of species in the same order as on the tree (list that can be used in ED).

SHIFT+F4: Writes the list of species in the order displayed on the tree in a file .LIS in the MUST\DATA directory
SHIFT+F6: Writes the list of species in the order displayed on the tree in a file .LIS in the MUST\DATA directory, but only keeping species which are not closely related. A limit value is entered and only one species is retained within a sub-group defined by a distance lower than this limit
SHIFT+F7: Writes the list of species in the order displayed on the tree in a file .LIS in the MUST\DATA directory, but only keeping species which are closely related. A limit value is entered and only species which have a species related by a distance lower than this limit are retained.
CTRL+F1: Loads Bootstrap Proportions from a file .GRP and displays them on the tree
CTRL+F2: Computes the symmetry of the tree and of the current sub-tree
CTRL+F3: Saves the tree in its current form in a file .ARB
CTRL+F4: Saves the tree in its current form in a file .TRE that is usable by TREEVIEW (Roderic Page, 1998)
ALT+F1: Switches the number of lines of the screen from 30 to 60 and vice versa
ALT+F6: writes several matrices containing the patristic distances (as for F6). The only difference is that distances between couple of species are set to zero if the number of internal nodes on the path connecting them is below a threshold (1 to 10)
ALT+F7: writes a matrix in which the distance between two species is equal to the number of internal nodes on the path connecting them
? : Indicates the full name of the species and, if possible, displays available information about this sequence from a file .INF (generated by the program alibaba)

COMP_MAT

Few new options have been implemented in COMP_MAT, such as the display, in green, of the y=x line, an output for EXCEL and more importantly the possibility of sorting couples of species according to their distance to the x-axis, y-axis and the y=x line (instead of the regression line only).

When the diagram is displayed:
F10: Displays the pairs of taxa by increasing abscissa (X) or ordinate (Y) values, or by decreasing distance from the regression line. For each pair, the couple of (X,Y)

values is displayed together with the distance (in % of the greatest value) from either X-axis or Y-axis or regression line.

SHIFT+F3: Sets the origin of the X and Y axes at 0 and rounds the maximum values of both axes

SHIFT+F4: Performs SHIFT+F3 and F4 successfully

CTRL+F1: Writes the pairwise values from the two starting matrices into two columns with their corresponding taxa names. The file (*.XLS) is directly readable by EXCEL.

AFT_PAUP, AFT_PHYL and AFT_UNIX

These programs are designed to reformat the outputs of PAUP3, PHYLIP 3.57, MOLPHY and PUZZLE. The new option in AFT_PAUP and AFT_PHYL is the reformatting of the bootstrap values (contained in a file .OUT) to a file .GRP (readable by TREEPLOT, allowing the display of bootstrap values obtained by any method on tree obtained by any method). In AFT_UNIX, you can reformat a single tree (the best one).