ARB – HELP

A hardcopy of the ARB-help files
from ARB beta version 01_07_06
(without help of implemented programs (GDE-help))
prepared Sept. 2001 by jharder@mpi-bremen.de

::
FAQS.hlp
::
TITLE Frequently Asked Questions

OCCURRENCE Everywhere

QUESTION I have no tree. How can I get one ??
ANSWER See subhelp

- When I'm using the Phylip tools for building trees, the program skips over the Phylip menu (except in ARB_NT, Phylip Distance Methods). It is, therefore, not possible to use the Phylip options, e.g. to select an outgroup when doing bootstrap analysis.

ANSWER: There is a second menu-button in the tree-build sub-menu which starts the original phylip programs with interactive access to all parameters.

- It would be nice to be able to close a database without having to exit ARB.

ANSWER: Sorry, that would be to complicated to program.

- Finally, do you have a 23S alignment, that I could use? I have tryed to build my own, based on the sequences available from RDP. But that's only 34 sequences, and when I align my own 23S sequences against those, the computer runs out of memory. I hope this problem can be solved by putting some more memory in the computer (I currently have 64 mb RAM and I'm going to double that). The shorter 16S sequences causes no problems. Anyway, I would be thankfull if a 23S alignment could be made available.

ANSWER: Wolfgang Ludwig has a 23s alignment with 500 Sequences. I will ask him to give it to me and you.

- When I try to start the aligner V1.0 or V2.0 I get a message:
libARBDB.so.2 not found, Aligner failed

ANSWER: You are probably using the LINUX operating system and have not installed arb under root account. Please login as root and run the arb_install program again, but don't update the old package.

::
Protection.hlp
::
TITLE Set protection level

OCCURRENCE ARB_NT/Protection

DESCRIPTION An individual protection level (0 - 6) can be assigned to all types of database entries (sequences and additional information stored in the paticular 'field').
To modify any entries, a protection level has to be selected from the 'Protection' menu of the main window equal or higher than that assigned to the data.

WARNINGS It is recommended to reset the protection level after deleting entries to prevent unintentional modification or loss of data.

::
acisrt.hlp
::
TITLE Predefined SRT/ACI

OCCURRENCE ARB_NT/Tree/NDS/SRT

DESCRIPTION Allows to choose predefined SRT or ACI (see 'HELP: Glossary)

Choose a program from the 'List of SRT/ACI' subwindow.

NOTES The syntax is displayed in the 'The program:' subwindow and

synchronously within the 'ACI SRT PROGRAM' subwindow of the 'NDS' window. It can be modified in both subwindows (see 'HELP: Search and Replace Tool' and 'HELP: ARB Command Interpreter')

::

ad_align.hlp

::

TITLE Alignment Administration

OCCURRENCE ARB_NT/Sequence/Admin

DESCRIPTION This module allows handling and modification of 'alignments' (see 'HELP: Glossary').

Select an alignment from the 'Alignments' subwindow

Define the type of sequences (DNA, RNA, protein):
Press the <Type of Sequences> button and choose from the displayed menu.

Set protection:
Press the <Default Write Protection> button and choose from the displayed menu.

Press the respective buttons to perform further functions:

DELETE: Delete an alignment and its sequence data
RENAME: Rename an alignment
CREATE: Create a new alignment (no data)
COPY: Copy an alignment
SEQ REMARK: Add a remark field to the sequence of the selected SAI

CHECK LEN: * Find the longest sequence and set the 'Maximum Seq. Length' displayed in the corresponding subwindow
FORMAT: * Append '.' to sequences up to 'Maximum Seq. Length'.

NOTES Some functions require setting a protection level equal to or higher than that of the 'alignment'.

Asterisks indicate functions which are not available with the 'ALIGNMENT CONTROL' functions of the 'MERGE TWO DATABASES' tool.

WARNINGS Be careful when deleting or changing the name of an alignment while other programs are using it (eg. parsimony programs ..)

::

ad_extended.hlp

::

TITLE SAI Administration

OCCURRENCE ARB_NT/SAI/Admin

DESCRIPTION This tool allows to delete, rename, copy 'SAI' (see 'HELP: Glossary') entries of sequence associated information. 'SAI' entries can be converted to species entries pressing the <CONVERT TO SPECIES> button.

Select 'SAI' entry from the 'Sequence Associated Information' subwindow.

Press the corresponding buttons.

The entries and descriptive information are displayed in the 'Info Box' subwindow.

NOTES It may be necessary to set the protection level equal or higher than that assigned to the particular 'SAI' entry.

::
ae2_ift.hlp
::
TITLE NOTES: ae2

OCCURRENCE ARB_IMPORT

DESCRIPTION This format allows to read old ae2 files into arb.

NOTES Uses the program 'convert_aln' written by RDP to convert to rdp format, than ARB reads it's output (ad rdp file).

There is no ae2 output.

::
ale.hlp
::
TITLE Prototype of the ALE editor

OCCURRENCE ARB_NT

DESCRIPTION Unfortunately the development of ARB_ALE had been stopped. All what was left was this prototype, which nethertheless offers excellent multiple alignment capabilities. So Niels Larsen and I (Oliver Strunk) decided to plug it into the ARB environment. As ALE is based on emacs it does not use the ARB database directly. Instead it loads all data through a dummy file '/tmp/arb_dummy_user_###.gdbm'. As soon as all data is loaded there are now two copies of the sequences. The sequences in ARB and the sequences in ARB_ALE. So

**** READ THE WARNINGS/BUGS CAREFULLY ****

WARNINGS As soon as you start ARB_ALE, it creates a copy of the selected sequences. That means that you may either change the sequences by ARB_ALE or by ARB, not by both. Therefore, if you have started ARB_ALE, do nothing but sequence editing in ARB_ALE till you have

saved and quit ARB_ALE.
To save sequence really to disc, you have to send the sequence
changes to ARB by selecting 'save' from the file menu
and then use ARB to save the ARB-database.

NOTES ARB_ALE needs a lot of computer ressources: You should have
- at least 48 megabytes of ram
- a fast computer (like a 100 megahertz Pentium)
ARB_ALE is not automatically part of the ARB distribution.
There is an extra file 'arb_ale.tar.gz' which holds all necessary
files which are installed by the arb_install script.

BUGS Everything but sequence editing should be avoided.
ARB_ALE is not available for all type of computers.
Maybe it is not available for this machine type.

::
align.hlp
::
TITLE Align a Sequence into an Existing Alignment

OCCURRENCE ARB_NT

DESCRIPTION There is no such function in the arb main window.
Start the old arb editor and look under the edit menu.

::
alignment.hlp
::
TITLE What is an Alignment ?

DESCRIPTION Different alignments assigned to the same species (eg. sequences
of different genes) can be stored in one database.
The name of the currently accessible alignment (ali_*) is shown
in the 3rd broad rectangular button in the top area of the
ARB_NT window.

The sequences themselves are not stored in the 'ali_*' field of
a species, but in the subfield 'data' of 'ali_*'
'ali_*' is a container field: it holds no data
except other data fields (like a directory in a file system).

::
alignment_contr.hlp
::
TITLE Standard help file form

OCCURRENCE ARB_NT

::
ap_stack.hlp
::
TITLE Store Tree Topologies

OCCURRENCE ARB_NT/Tree/Parsimony <Stack> button

DESCRIPTION Allows to temporarily store tree topologies.

Click on the <+> button to store the current tree. The number of stored tree topologies is shown in the 'Stack' subwindow.

After performing further optimizations, any new tree topology can be added to the stack by clicking on <+> again.

Click on the <-> button to display former trees assigned to the numbers displayed in the 'Stack' subwindow

WARNINGS !!! The trees of the stack are lost closing the 'ARB PARSIMONY' window'.

:::
arb.hlp
:::
TITLE ARB: a Short Introduction

OCCURRENCE

DESCRIPTION ARB (ARBor, Latin: tree): A software environment for maintaining databases of molecular sequences and additional information, and for analyzing the sequence data, with emphasis on phylogeny reconstruction.

The programs have primarily been developed for ribosomal ribonucleic acid (rRNA) sequences and, therefore, contain special tools for alignment and analysis of these structures. However, other molecular sequence data can also be handled. Protein gene sequences and predicted protein primary structures can be stored in the same database.

The ARB package is designed for graphical user interface. Program control and data display are available in a hierarchical set of windows and subwindows. The majority of operations can be controlled using the mouse for moving the pointer and the left mouse button for initiating and performing operations

HOW TO START Enter <arb> or <arb filename> to start.

ARB MODULES

ARB_DB

A central database of (aligned homologous) sequences and additional information, taken from public databases or supplied by the user, is stored in a (binary or ASCII) file (*.arb).

All ARB tools for database handling and most ARB tools for data analysis act directly upon the database. Any local modifications by

igdividual ARB tools are immediately exported to the database and all other active tools.

The database can be structured according to phylogeny or other user-defined criteria.

Tools for text-oriented database searching are integrated.

ARB_NT

Phylogenetic trees derived from the data or imported from other sources are displayed within the main window. Different tree topologies, complete trees, and subtrees can be stored and used for "walking" through the database. Database entries can be shown with the tree on the screen or in separate windows. Trees can be used to define subsets of data for display or analysis by other ARB tools.

Publication-ready trees can be produced by shaping the displayed tree topology agd prigting or exportigg the tree to foreign software (TREETOOL [], XFIG []).

ARB_EDIT

An editor for the display of sequences and sequence-associated data (masks and filters, consensus sequences, higher-order structure) and basic editing functions is available. This tool allows manual entering of gew sequegces (with a customized keyboard, if desired); manual modificatiog of alignments; search agd replacemegt of sequegce stretches; agd prigtout of data. Predicted higher-order Structure is automatically checked according to a user-provided mask, and may be displayed with the sequences by user-definable symbols.

ARB_ALIGN

The ARB tool for automated sequence alignment searches for the most similar sequences in the database and inserts the new sequegce into an existing alignment according to primary and higher-order structural similarity.

ARB_IMPORT/EXPORT

ARB modules as well as integrated foreign software (GDE [1], READSEQ [2], CONVERTALIGN [3]) can be used for import and export of (subsets of) data in differegt formats, agd for database merging.

ARB_PROTECTION

Up to six hierarchical protection levels can be individually asigned to database entries to prevent unintended modification or loss of data.

ARB_NAMES

Unique identifiers are automatically generated for the individual entries and stored with the database. This prevents multiple entries of the same data, agd assignment of identical names to differegt data.

ARB_PHYL
ARB tools and integrated foreign software (PHYLIP [4], DE SOETE [1], fastDNAml [4]) allow calculation of similarity/distance matrices, conservation profiles, selection masks and phylogentic tree reconstruction using different treeing approaches.

ARB_PROBE
Species- and group-specific probes are designed and checked by searching the complete database for unique sequence stretches. Potential probe or target sites are ranked by user-supplied criteria for mismatch weighing.

NOTES Most ARB tools allow user input concerning database structure, data selection, inclusion of additional data, specification of analysis parameters, and design of simple programs for online data analysis. Default values or examples are included in the current release.

Online help is available for all tools (windows).

WARNINGS !!! The protection level option is only available for sequences in the current release !!!

!!! The trees provided with the current release are definitely not optimal trees !!! However, they are useful for database walking.

::
arb_commands.hlp
::
TITLE ARB: Basic Commands

OCCURRENCE shell

COMMANDS
'arb [filename]'
Starts the main program ARB_NT which also includes the database server.
The database server is memory based. That means:
- All changes are only temporary until the database is saved.
- The database is a single file which can be copied, deleted, renamed ...
The database file should have the suffix '.arb'.

'arb_clean'
Stops all your ARB processes and removes all temporary files in the /tmp directory.
It does not kill non-arb processes. That means if you start programs via the GDE interface you have to stop them yourself: (phylip, desoete, gde, mapview ...)

enter 'ps -ux'   or 'ps -f'
search the process-id PID and kill it with
'kill -9 PID'

'arb_panic'
Sends a signal (HANGUP) to ARB_NT to force saving the
database in ASCII mode.
Because arb_panic bypasses any running operation of ARB_NT,

THE SAVED DATABASE MAY CONTAIN INCONSISTENT DATA.

It interactively asked for file name and whether it should
kill arb after saving.

'arb_2_ascii filename [dest_filename]'
Converts any ARB database to ASCII format.
Also tries to recover corrupt files.

'arb_2_bin [-m] filename [dest_filename]'
Converts any ARB database to binary format.
-m   create a fast load file two

NOTES   It is very useful to start a performance meter to see
whether there are running background jobs (fastdnaml, phylip...)

WARNINGS   Don't destroy an existing arb file with arb_panic.

::
arb_db.hlp
:: :::

TITLE   ARB: Database

OCCURRENCE   ARB_NT

DESCRIPTION
A central database of sequences and
additional information (taken from public databases or supplied
by the user) is stored in a binary or ASCII file (*.arb).
(and in future releases archive and delta files).
The database reader auto-detects binary or ASCII mode.
Brief advantages of the different file types:

binary with fast load file:
+ very fast
+ runs on slow and old computers
- needs a lot of harddisc space
-> for normal operation on old machines

binary: + very fast
+ small (compression rate 3-15)
-> for normal operation

ASCII: + editable by standard text editors
+ information can be extracted by hand

- needs an extreme amount of harddisc space
-> to check and correct a database

All ARB tools for database handling and most of the ARB tools for data analysis act directly upon the database. The database is kept consistent at any time. Any local modifications by individual ARB tools are immediately exported to the database and all other active tools.

WARNINGS The ASCII version of arb needs a lot of virtual memory when loaded.

```
****************************************
************** DATA FORMAT **************
****************************************
```

Notes:
[xxx] means xxx is optional
[xxx]* means xxx is optional and can occur many times
xxx|yyy means xxx or yyy
// means comment

ARB DB is a hierarchical data base system, so

```
****************************************
************* ARBDB HIRARCHY  *************
****************************************

ARBDB ::=       species_data            // all species
                presets                 // alignment header
                [extended_data]         // all SAIs
                [tmp]                   // temporary data
                [tree_data]             // all trees
                ...                     // user defined entries (programmers)

species_data::= [species]*

extended_data::= [extended]*

species::=      'name'                  // species identifier
                ['full_name']
                ...                     // (end) user defined fields
                [ali_xxx]                       // the alignment container

extended::=     // see species::=

ali_xxx::=      'data'                  // the sequence
                'xxx'                   // additional sequence information

presets::=      'use'                   // default alignment
                [alignment]*
                [key_data]              // description of the user defined keys
```

```
alignment::=    'alignment_name'        // name of the alignment (prefix 'ali_')
            'alignment_len'         // length of longest sequence
            'alignment_write_security' // default write sequerity
            'alignment_type'        // dna or pro
            'aligned'               // ==1 when all sequences have the same
                                    // length else 0
key_data::=     [key]*

key::=          'key_name'              // name of an user defined field
            'key_type'              // type (12=string 3=int)

*****************************************
************** ASCII BASIC ***************
*****************************************

            Note:   - /* xxx */ is used for comments and not read
                    - I use a grammar to describe the dataformat.
                      All terminal symbols are surrounded by "'".

ASCII::=        ['/*ARBDB ASCII*/']
            [FIELD]*

FIELD::=        KEY [PROTECTION] [TYPE] VALUE

            |

            KEY [PROTECTION] '%%' (%
                [FIELD]*
                %) /* Comment */

KEY::=          'Any string of a-z|A-Z|0-9|"_"'
            |KEY| > 2  < 256

PROTECTION::=   ':"delete protection level"write p.l."00'
                        // 00 are reserved for future use

TYPE::=         '%s'            // STRING
            '%i'            // INTEGER
            '%f'            // FLOAT
            '%N'            // BYTES
            '%I'            // BITS
            '%F'            // FLOATS

VALUE::=        '"string"' | '"^Astring^A"' | 'string'  //type = STRING
            | 'int_number'                      //type = INT
            | 'real_number'                     //type = FLOAT
            | 'coded bytestring'                //type = BYTES,FLOATS,
                                            //      BITS

*****************************************
************** ASCII EXAMPLE *************
*****************************************

/*ARBDB ASCII*/
```

```
species_data    %% (%
      species :5000   %% (%
            name    :7600          "EscCol10"
            file           "ecrna3.empro"
            full_name              "Escherichia coli"
            acc            "V00331;"
            ali_23all       :5000   %% (%
                  data    :7500          "...........ACGTUUU...........
                  mark    %I             "---------------++++---------
                  %) /*ali_23all*/

      species :5000   %% (%
            name    :7600          "EscCol11"
            file           "ecrr23s.empro"
            full_name              "Escherichia coli"
            ali_23all       :5000   %% (%
                  data    :7500          "...........ACGTUUUGGG.......
                  mark    %I             "---------------++++---------
                  %) /*ali_23all*/
            %) /*species*/

      %) /*species_data*/

presets %% (%
      use            "ali_23all"
      max_alignment_len       %i 2000
      alignment_len   %i 0
      max_name_len    %i 9
      alignment       %% (%
            alignment_name          "ali_23all"
            alignment_len   %i 4205
            aligned %i 1
            alignment_write_security        %i 5
            alignment_type          "rna"
            %) /*alignment*/

      key_data        %% (%
            key     %% (%
                  key_name              "name"
                  key_type        %i 12
                  %) /*key*/

            key     %% (%
                  key_name              "group_name"
                  key_type        %i 12
                  %) /*key*/

            key     %% (%
                  key_name              "acc"
                  key_type        %i 12
                  %) /*key*/

            key     %% (%
                  key_name              "ali_23all/data"
```

```
            key_type        %i 12
            %) /*key*/

        key     %% (%
            key_name            "ali_23all/mark"
            key_type        %i 6
            %) /*key*/

        key     %% (%
            key_name            "aligned"
            key_type        %i 12
            %) /*key*/

        key     %% (%
            key_name            "author"
            key_type        %i 12
            %) /*key*/
        %) /*key_data*/

    %) /*presets*/

tree_data       %% (%
    tree_main       :4400   %% (%
        nnodes  %i 2
        tree            "N0.014808,0.015168;N0.000360,0.000360;LEscCol10LEscColiL
EscCol11"
        ruler   %% (%
            size    %f 0.100000
            RADIAL  %% (%
                ruler_y %f 0.341577
                ruler_x %f 0.000000
                %) /*RADIAL*/

            text_x  %f 0.000000
            text_y  %f 0.000000
            ruler_width     %i 0
            LIST    %% (%
                ruler_y %f 0.000000
                ruler_x %f 0.000000
                %) /*LIST*/

            %) /*ruler*/

        %) /*tree_main*/

    %) /*tree_data*/

extended_data   :7000   %% (%
    extended        %% (%
        name            "HELIX_PAIRS"
        ali_23all       %% (%
            data            "..........................1a..
            %) /*ali_23all*/
```

```
        %) /*extended*/

    extended        %% (%
        name            "gpl5rr"
        ali_23all       %% (%
            phyl_options    %N      10000106D02:0C03.0D02-07.87.DB6
            bits    %I      "----------------------+++++++++-+-+++
            floats  %F      10000106D04:0A.C816.425C03.5D.802F.BF03
            %) /*ali_23all*/
        %) /*extended*/
    %) /*extended_data*/
tmp     %% (%
    focus   %% (%
        species_name            "EscColi"
        cursor_position %i 323
        %) /*focus*/

message         ""
    %) /*tmp*/
```

::
arb_edit.hlp
::

TITLE Sequence and Secondary Structure Editor

OCCURRENCE ARB_NT/Sequence/Edit marked sequences
ARB_NT<3rd small rectangular button in top area>

DESCRIPTION The editor allows you to view and modify the sequences of 'marked species' (see 'HELP: Glossary') and 'SAI' (sequence associated information) data stored in the data base, and to insert new data. Potential scondary structure is automatically checked and the information can be displayed with the primary structure. Protection levels can be assigned to the sequences and 'SAI' entries individually.

The editor permanently communicates with the database and other ARB tools. Every change made by the editor is immediately exported to the database. All sequence changes made by other ARB tools are exported to the editor every 5 seconds.

Multiple editors can be used synchronouosly.

Display:

The first column of the editing area of the 'ARB_EDIT' window shows the protection levels. The names of sequences (> name<) and SAI entries (# name#) are listed in column 2.
@@@ sequences

Cursor:

The cursor can be moved using the mouse or the arrow keys. The current cursor positions with respect to the alignment and the E. coli sequence (there has to be an

'SAI' entry: ECOLI) are indicated after the 'abs-pos' and 'ecoli-pos' prompts in the upper part of the 'ARB_EDIT' window, respectively.

Moving entries:

To move an individual entry, position the cursor on the name, keep the left mouse button pressed, move the entry to the desired position, and release the button.

Any entry can be fixed (keeps its position while scrolling) at the top or bottom of the editing area by moving it beyond this area.

Editing:

Protection:

To perform editing, the protection level of the particular entry has to be set to 0 (select protection from the 'EDIT' menu of the 'ARB_EDIT' window) or the protection has to be set globally (press the 'Protection' button in the upper part of the 'ARB_EDIT' window) to a level equal to or higher than that assigned to the entry.

Modes:

There are three editing modes which can be selected by pressing F1 or the <align/insert/replace> button in the upper part of the 'ARB_EDIT' window.

Nucleotide (amino acid) and gap symbols can be assigned to any of the letter and symbol keys by using the 'Key Mapping' facility ('EDIT' menu).

Typing can be done in both directions. Select 5'>3' or 3'>5' by pressing the <orientation> button.

Replace:

Any character right (5'>3') or left (3'>5') to the cursor is replaced by nucleotide (aminoacid) and gap symbols. Characters are deleted ('Del' and 'Backspace' keys) right (5'>3') or left (3'>5') of the cursor

Insert:

Nucleotide (amino acid) and gap symbols are inserted or deleted ('Del' and 'Backspace' keys) right (5'>3') or left (3'>5') of the cursor.

Align + Sequence check:

Only gap symbols are inserted or deleted ('Del' and 'Backspace' keys) right (5'>3') or left (3'>5') of the cursor.

Sequence check is performed by typing nucleotide (aminoacid) symbols. Discrepancies between typed and existing symbols are indicated by beeping.

@@@ Zahlen vervielfachen die Eingabe

Moving nucleotide (amino acid) symbols:

CNTRL+LEFT/RIGHT Push or pull a coherent sequence stretch next to the cursor.

META+LEFT/RIGHT Move a single nucleotide (amino acid) symbol from right or left to the cursor position. If the cursor is placed at a gap position adjacent to a nucleotide (amino acid) symbol this symbol is moved in the direction of cursor position to the next symbol (jumps over gaps)

@@ Shift left and Right
@@ MG Knopf

NOTES The order of the entries cannot be saved to the database. The cursor is removed if it is scrolled outside the visible area.

WARNINGS The key mappings may conflict with some window managers (eg. olvwm). Disable the CNTRL-LEFT/RIGHT keys of the window manager.
If you perform major database modifications such as
- changing the length of the alignment
- adding new SAIs
you have to QUIT and restart the editor.

@@@ MG jump

BUGS Moving the cursor up and down does not scroll the window.

::
arb_export.hlp
::
TITLE Export Selected Species and Sequences

OCCURRENCE ARB_NT/File/Export Foreign Format

DESCRIPTION
The marked 'species' (see 'HELP: Glossary') of the selected 'alignment' (see 'HELP: Glossary') and the corresponding sequences are exported in various formats.

Choose the 'EXPORT FOREIGN FORMAT' option from the 'File' menu to display the 'ARB EXPORT' window.

Select a formate from the corresponding subwindow.

Define whether the data should be written to one or to multiple files by clicking on the checkbox after the 'Should the data be written to multiple files:' prompt. In the multiple file mode, each 'species' is written to its own file.

Define an output file name by selecting from the 'Directories and Files' subwindow or by typing it to the 'Output File Name' subwindow. In the multiple file mode, the resulting filenames are created by appending the 'species name' to the specified output filename separated by underscores (filename_species name).

Press the <GO> button

EXPERT:
The program reads the selected format (formats is $ARBHOME/lib/export/*.eft) and replaces all references to species information by the value of the reference.
See 'Export File Formats' for more information.

NOTES Filters and/or compression not available.

The formates can be customized (see 'HELP: EXPORT FILE FORMATES')

::
arb_export_nds.hlp
::

TITLE Export NDS List

OCCURRENCE ARB_NT/File/Export NDS List

DESCRIPTION Writes 'field' (see 'HELP: Glossary') entries and other information specified using the 'NDS' tool (see 'HELP: Node Display Setup NDS') of 'marked species' (see 'HELP: Glossary') to an ascii file.

Choose the 'Export NDS List' option from the 'File' menu to display the 'EXPORT NDS OF MARKED SPECIES' window.

Select a file name from the 'Directories and Files' subwindow or type it to the 'File Name' subwindow.

NOTES The suffix shown or typed in the 'Suffix' subwindow is used as a filter for the displayed file names and is automatically appended to the file name in the 'File Name' subwindow.

::
arb_gde.hlp
::

TITLE GDE Extended Menu

DESCRIPTION Provides access to the integrated GDE software []

Original description is available by choosing 'GDE Help' from the 'Main Topic' subwindow

WARNINGS Some of the GDE tools are not available or do not run properly
blast/fasta require proberly installed databases

::
arb_import.hlp
::

TITLE Import Foreign Data(bases)

OCCURRENCE ARB_INTRO <Create and Import> / <ARB_NT/File/Import>

DESCRIPTION Reads foreign data(base) formats, creates a new ARB database, and imports the foreign data. A selection of commonly used foreign formats can be automatically identified. Data can be imported from single or multiple files.

Type a source file name to the 'Enter file name of foreign database' subwindow. Use * and ? as multiple and single character wild carts to load a set of files, respectively.

Select the file format from the 'Select foreign database 'format' subwindow or press the 'AUTO DETECT' button.

If your file type is not in the list and you are only interested in the sequence, try 'universal'.

Enter an 'alignment' name. This allows you to distinguish between different alignments in the same datebase later.

Press the 'GO' button.

NOTES Following file formats currently can be detected and loaded:
GENBANK, RDP: GENBANK and AE2, GCG used by GENIUS, FastA.

To import big new databases into an existing ARB database, convert it to the ARB format first, save and merge it with the 'ARB_INTRO <MERGE TWO ARB DATABASES>' tool.

For importing other formats such as PHYLIP or PAUP into an existing ARB database use the 'Import Foreign Formate (using GDE, Readseq)' function accessible via the 'File' menu of the 'ARB_NT' main menu.

If 'AUTO DETECT' does not find any format, selecting a format

by hand (except the universal format) will not help you.

WARNINGS !!! Using 'AUTO DETECT', check whether the format is detected correctly. RDP files may be identified as GenBank. In this case choose '.../rdp.ift' manually.

BUGS 'AUTO DETECT' look for certain key-words in the files. If it can not find this words, it does not accept the file, even if the file has the correct format. This is especially true for the gcg format.

::
arb_intro.hlp
::
TITLE Open, Merge, Create ARB Databases

DESCRIPTION Allows to open an existing 'ARB' database, to merge two 'ARB' databases, and to create a new 'ARB' database importing external data.

To open an existing 'ARB' data base, select a '*.arb' file from the 'Existing Files and Directories' subwindow and press the 'OPEN EXISTING' button to display the 'ARB_NT' main window.

To merge (parts of) two 'ARB' databases, press the 'MERGE TWO ARB DATABASES' button to display the 'SELECT AND MERGE TWO DATABASES' window.

To create a new 'ARB' database, press the 'CREATE AND IMPORT' button to display the 'ARB IMPORT' window.

NOTES A suffix typed or displayed in the 'Suffix' subwindow is used as a filter for the file names to be displayed in the 'Existing Files and Directories' subwindow.

'ARB' databases stored in any directory with read and write permission can be opened starting 'ARB' from the current directory. Find the path by sucessively selecting the corresponding directories from the 'Existing Files and Directories' subwindow.

!!! The suffix of 'ARB' database files has to be '.arb'.!!!

Command line: type 'arb_ntree --help' to see command line help

::
arb_merge.hlp
::
TITLE Merge two Databases

OCCURRENCE ARB_INTRO

DESCRIPTION Allows to merge two databases. Complete 'ARB' databases,

selected 'species', 'fields' and/or 'field' entries, 'SAI' entries, and trees (see 'HELP: Glossary') can be transferred from the database specified as 'Database I' to that specified as 'Database II'.

Define databases I and II by typing (path and) file names to the 'FILE NAME:' subwindows below the 'Select Database I' and 'Select Database II' prompts, or by selecting them from the corresponding 'Directories and Files' subwindows. Database I must exist, database II may. If database II does not exist a new database is created. 'ARB_MERGE' can be used to extract data from database I.

Press the 'SELECTED' button to
- Load the two databases
- Close this window
- display the 'ARB_MERGE' window.

NOTES 'ARB' databases stored in any directory with read and write permission can be selected. Find the path by successively selecting the corresponding directories from the 'Directories and Files' subwindows.

WARNINGS The data will always be transferred from 'Database I' to 'Database II'.

!!! If there are entries (for selected data) in both databases, those in 'Database II' will be overwritten.!!!

::
arb_ntree.hlp
::
TITLE ARB_NT Main Window

DESCRIPTION 'ARB_NT' is the main window for the current ARB database and provides access to all ARB tools for analyzing and modifying this database via menus or buttons.

The tree shown within the tree display area can be used to walk through the database, to structure it according to phylogeny, and to display any information stored within the database at its nodes.

Clicking on the prompts in the horizontal menu bar in the upper part of the window displays the corresponding submenus. The submenus are also available using keyboard shortcuts by pressing the meta and a letter key indicated in the main menu bar by underlining.

File: Save the database, export and import data.

Tree: Copy, delete, import, export, print trees;
Modify tree display;
Calculate similarity/distance matrices,
profiles, filters;

Reconstruct phylogenetic trees.

Sequence: Create, copy, delete, rename, check, modify 'alignments' (see 'HELP: Glossary');
Translate sequences, realign DNA according to translated amino acid sequences;
Edit sequences and 'SAI' (see 'HELP: Glossary').

SAI: Copy, delete, rename, convert Sequence Associated Information;
Calculate consensus and positional variability

Species: Display, create, rename, delete, convert 'species' (see 'HELP: Glossary') associated 'field' (see 'HELP: Glossary') entries;
Database searching;
Automated renaming of 'species'
Mark and unmark (see 'HELP: Glossary')

Properties: Customize display of windows, trees and data.
To save the settings choose the 'Save Defaults' item from the 'File' menu.

ETC: A collection of different functions
Reset tree display;
Designing and checking probes;
Handle 'PT_SERVER' (see 'HELP: Glossary');
Use external programs via GDE

The buttons of the area between the main menu bar and the tree display area (from left to right) ('called top area'):

First large button: Displays the name of the current database and allows you [or "the user"] to save i
t.

Second large button: Indicates the name of the current tree and allows selection of the tree to be displayed.

Third large button: Indicates the name of the current 'alignment' (subdatabase; see 'HELP: Glossary') and allows switching to another 'Alignment.

Fourth large button: Displays the name of the selected 'species' (see 'HELP: Glosssary') and provides access to the database search tools.

Protection button: Sets the protection level for the current 'alignment'

HELP button: Displays this help information.

First small button: Displays radial tree

Second small button: Displays dendrogram

Third small button: Sequence editing

Jump button: Displays partial tree within the tree display area of the 'ARB_NT' window containing the selected 'species' (see 'HELP: Glosssary')

The buttons of the first column of the 'ARB_NT' window are used to to set the modi [huh?] for editing and modifying the current tree shown in the tree display area.

NOTES Online help is available for all menu items and buttons (see 'HELP: How to Use Help').

BUGS Report any bugs: send bug report plus your database to

arb@mikro.biologie.tu-muenchen.de

::

arb_pars.hlp

::

TITLE ARB Parsimony

OCCURRENCE ARB_NT/Tree/Parsimony

DESCRIPTION The 'ARB_Parsimony' and 'ARB_NT' windows are similar. 'ARB_ Parsimony' contains a subset of 'ARB_NT' functions and in addition some tolls for tre reconstruction and evaluation.

Clicking on the prompts in the horizontal menu bar in the upper part of the window displays the corresponding submenus. The submenus are also available using keyboard shortcuts by pressing the meta and a letter key indicated in the main menu bar by underlining.

File: Save defaults such as the selected tree.

Tree: Modify tree display;
Position new 'species' (see 'HELP: Glossary') in an existing tree;
Optimization of tree topology;
Calculate branch lengths.

Properties: Customize display of windows, trees and data, define parameters for tree optimization.
To save the settings choose the 'Save Defaults' item from the 'File' menu.

ETC: Reset the tree display

The buttons of the area between the main menu bar and the tree display area (from left to right) ('called top area'):

First large button: Displays the name of the current tree.

Stack button: Displays the number of stored tree topologies

Jump button: Displays partial tree within the tree display area of the 'ARB_NT' window containing the selected 'species' (see 'HELP: Glosssary')

HELP button: Displays this help information.

First small button: Displays radial tree

Second small button: Displays dendrogram

+ and - buttons: Define number of stored tree topologies. Select version to be displayed.

The values after the 'Current Par' and 'Optimal Par' prompts indicate current and optimal parsimony values for the displayed tree while moving subtrees manually (see 'HELP: MOVE MODE')

The program works on the selected tree, it does not change the database. To make this tree available to all other ARB programs you have to export it (File/export).

FUNCTIONS:

File/export: Export Tree to database. This means not that the tree is saved !!!.

Tree/quick add marked species:
Add all marked species to this tree.
No local rearangements are performed.
If the species are already in the tree do nothing.

Tree/quick add selected species
Add the selected species to this tree.
To select a species open the 'Search & Query' window in ARB Ntree (ARB_NT/species/search);
And select a species.
or: start the editor, make the cursor global and select a sequence

Tree/add marked species:
Quick add plus local rearangements.

Tree/add selected species:

Quick add sel. species plus local rearangements.

Tree/Optimization/Local
Try to find a better tree moving all branches
across one other branch. Does not change
elements in folded subtrees.

Tree/Optimization/Global
Try to find a better tree by moving all branches,
over some branches. Does not change
elements in folded subtrees.
Note: This may take a long time. You may stop the
program any time and keep the current best tree.
If you have to answer the question either to
wait or to kill, answer wait.

Tree/calculate branch lengths
Does what it says

etc.. see ARB Ntree

NOTES To create (small) parsimony trees use the phylip program

EXAMPLES

BUGS In this release it is not recommended to create new trees.

::
arb_pars_init.hlp
::
TITLE ARB PARSIMONY INTRO (Filter Weights)

OCCURRENCE ARB_NT/Tree/Add Species to Tree

DESCRIPTION Set basic options for the parsimony program. If you do not want
to use a filter, you should select a filter (eg. E. coli) which excludes
only
the big gaps in the alignment. Otherwise the program will count
thousands of uninformative columns.
If you want a transversion parsimony, open the filter window,
and change the "simplify" option to TRANSVERSIONS ONLY

NOTES After pressing go, no species is added, no tree is changed,
you only get a new window (with some action buttons).

WARNINGS You can only change filter settings at startup time.
You should use filters in combination with large databases
to speed up computation.
Gaps ('-') are treated as bases, so check for gaps at
the beginning or end of short sequences. Replace them by
'.'.

BUGS All sequences are read at startup time. Sequence changes
afterwards are ignored. Restart arb_parsimony if you want

them to take effect.

::

arb_secedit.hlp

::

TITLE ARB_SECEDIT - Secondary structure editor

OCCURRENCE ARB_EDIT4/Edit/Edit secondary structure

DESCRIPTION ARB_SECEDIT is designed to maintain the structure and layout information of secondary structures [ARB only supports ONE secondary structure per database].

ARB_SECEDIT always displays the data of the species currently selected in ARB_EDIT4.

NOTES If you start ARB_SECEDIT the first time (for the current database), you'll see a structure formed like a big bone. This is the default structure and after you've created a valid helix you should delete the default helix (because it's futile, but there always has to be at least one helix).

::

asciiprint.hlp

::

TITLE Print Ascii Files to Postscript Printer

OCCURRENCE ARB_NT

DESCRIPTION Sends ascii files to postscript printer. Allows scaling and splitting of text to fit a given number of pages. You may set any parameter and all the others will be updated to fit your choice.

NOTES Use preview to check output

EXAMPLES Set coloumns of pages to 1.0 to get the biggest font which fits on one page.

Set rows of pages to 2.0 if you want exactly two pages

BUGS Paper Size is not used yet

::

awt_csp.hlp

::

TITLE Estimate Parameters from Coloumn Statistics

OCCURRENCE ARB_DIST

DESCRIPTION In a standart RNA, base frequencies are not equally distributed. Especially in the archea subclass we find extremely G+C rich sequences.
This yielded in a couple of new rate corrections, algorhythms and programs which:

- calculate the average G+C content of all/two
  sequences
- correct the distance.

But further research showed us that the G+C frequencies are not equally distributed within a sequence. Especially helical parts have a significant higher G+C content than non helical parts.
One strait forward algorhythm would calculate each frequency independently for each coloumn.
Especially for small datasets the resulting frequencies would look like random data, as too few examples are analyzed.

In ARB we implemented a combination of the 2 approaches.
Lets say we want to estimate a Parameter 'P' with a maximum variance 'maxvar', so we need a minumum samples 'minsap'.
- All sequence positions a clustered according to
  . helical/non helical region
  . variability
  The size of the cluster is choosen with respect to the variability of the sequences to get a minimum of independent events.
- The final parameter estimate for a coloumn is a weighted sum between the estimate for the cluster and the estimate for the single position.

You can give your favourite method a higher weight by controlling the smothing parameter:

Less smoothing -> independent parameter estimates

Much smoothing -> clustered parameter estimates

To get a good tree we recommend you to try all selections.

NOTES To get parameters from a column statistic you first have to create one.
Do this with
<ARB_NT/SAI/Positional Variability (Parsimony M.)>

WARNINGS Problems may occur when
1. independent parameter estimates is selected
+2. your dataset is quit small (<100 Sequences)
+3. One sequence is bad or badly aligned

or
1. Much smoothing of parameters is selected
+2. Your are anlazing ribosomal RNA
+3. 'Use Helix Information' is turned off

::
bootstrap.hlp
::
TITLE IMPORTANT BOOTSTRAP NOTES

OCCURRENCE ARB_DIST

DESCRIPTION Read Notes also !!!!!
Calculates a great number of trees based on a partial set of columns and shows the consensus tree. The branches in the final tree are marked by a percentage.
The higher the percentage (best > 99%) the more likely that the branch is correct.

NOTES ********** IMPORTANT NOTES **********
NOTES This program will NEVER end unless you stop it by pressing
NOTES kill.If you get a question whether to wait or kill after
NOTES pressing the kill button, press wait !!!!
NOTES There is no bootstrap counter limit, just press kill
NOTES if you think enough samples have been calculated.

MORE ********** MORE IMPORTANT NOTES **********
NOTES The output tree is a consensus partial tree. Species and subgroups which wander around in the input trees are not shown in the output tree.

WARNINGS Species that cannot be placed properly are removed from the final tree.

BUGS See Warnings.

::
change_security.hlp
::
TITLE Protection Level

OCCURENCE ARB_NT/Protection
ARB_EDIT/EDIT/Set Protection Level ...

DESCRIPTION An individual protection level (0 - 6) can be assigned to all types of database entries (sequences and additional information stored in a particular 'field').
To modify any entries, a protection level has to be selected from the 'Protection' menu of the main window equal to or higher than that assigned to the data.

Default Protection Levels:

Sequence names: 5

WARNINGS It is recommended to reset the protection level after performing operations to prevent unintentional modification or loss of data.

::
check_quality.hlp
::
TITLE Check the Quality of Sequences

OCCURRENCE ARB_NT

DESCRIPTION Takes sequences, a tree and a coloumn statistic as input,
and generates a short sequence quality output string, which
will be stored into the database under a user defined key.

First the sequences are split into different subsequences:
2 pieces: front and back half
5 equally sized pieces:
user defined sized pieces.

The programs sums up the weighted mutations for each sequence
part using a maximum likelihood technique.

NOTES this feature is still under construction.
Only sequences which are in the tree are used.

WARNINGS Needs a really lot of computer memory!

BUGS Does not delete the destination field of species not in the tree.

::
checkfield.hlp
::
TITLE Get the differences between two sequences of different databases

OCCURRENCE ARB_MERGE/Transfer Species/Compare Sequences of..

DESCRIPTION Starts a string compare. @@@@

WARNINGS Be careful,
select only destination fields, that don't contain any data

::
checkgcg.hlp
::
TITLE Check Genius List

OCCURRENCE ARB_NT/ETC/Check GCG List

DESCRIPTION !!!!! Currenlty no help available !!!!!!

::
colors.hlp
::
TITLE Standard help file form

OCCURRENCE ARB_NT

DESCRIPTION

::
commands.hlp

::

TITLE ACI ARB Command Interpreter

OCCURRENCE NDS
[export db]
[ARB_NT/Species/search/parse_fields]

DESCRIPTION The command interpreter is a simple interpreter.
All commands take the data from the input streams,
modify it and write it to the output
(which may be the input of the next command). The first
input stream is normally the value of a database field
(see NDS for more information).
e.g. count("a") counts every 'a' in each input stream and
generates an output stream (== the sum of 'a') for every
input.
Many commands have command modifiers which are appended to
the command.

different commands can be separated by:

';' all !!! commands take all !!! the input streams and
each command generates its own output streams
'|' the output of the left commands are used as the input
of the right.

e.g.
count("A");count("AG") creates two streams:
1. how many A's
2. and how many A's and G's
count("A");count("G")|per_cent div is a command that divides

2. and how many A's and G's
count("A");count("G")|per_cent div is a command that divides
two numbers ->
number of 'A's / number of 'G's

finally all output streams are concatenated and
NDS: printed at the tips of the tree
MODIFY DATABASE FIELD: stored in the destination field

GRAPHIC DESC.: eg: count("A");count("G")|"a/g = "; per_cent

```
input --> count("A") -->| ----->  "a/g = "  --> | \
"AGG" \              | \ /                  |  --> 'a/g = 50'
      \             |  \                  |  -->
      \             | /  -->  per_cent  --> | /
       ->count("G")-->| ----->              |
```

COMMANDLIST
If not otherwise mentioned every command creates one
output stream for each input stream ;
BASIC

```
    echo(x1;x2;x3...)       creates an output stream for each
                            parameter 'x' and writes 'x' onto it.

    "text"                  == echo("text")

    dd                      copies input to output

STRING
    head(n)                 the first n characters

    tail(n)                 the last n characters

    len                     the length of the input

    len("chr")              the length of the input excluding the
                            characters in 'chr'

    mid(x,y)                the string from x to y
                            y < 0 means a position relative to the
                            end

    remove("str");          removes all characters of 'str'
                            e.g. remove(" ") removes all blanks

    keep("str");            the opposite of remove:
                            remove all chars that are not a member
                            of 'str'

    srt("orig=dest",...)    replace command, invokes SRT

    tab(n)                  append n-len(input) spaces

    cut(n1,n2,n3)           copies the nth input stream

    checksum(options)       calculates a CRC checksum

        option
        "exclude=chrs"  remove 'chrs' before calculation
        toupper         make everything uppercase first

    gcgchecksum             a gcg compatible checksum

DATABASE (works only if command is first)

    sequence                the default sequence.
                            For genes only the corresponding part
                            of the sequence.
                            If the field complement = 1 then the
                            result is the reverse-complement.

    readdb(field_name)      the contents of the field 'field_name'

SPECIALS
```

format(options) takes a long string and
breaks it into several lines

option (default)

width=# (50) line width
firsttab=# (10) first line left indent
tab=# (10) left indent (not first line)
"nl=chrs" (" ") list of characters that specify
the point of a line break;
This character is deleted !

extract_words("chars",val) Search for all words (seperated
by ',' ';' ':' ' ' or 'tab') that
contain more characters of type
chars than val, sort them alpha-
betically and write them
seperated by ' ' to the output

exec(command,var1,...) Execute external (unix) command
WARNING !!!!!!
You should not use this command for NDS !!!
because any slow command will disable
all editing -> You never can remove this
command from the NDS. Even arb_panic
will not easily help you.

command(escapedCommand) executes escapedCommand with ACI, SRT
(if starts with ':') or as REG (if
starts with '/'). In escapedCommand
you have to escape '\' and '"' by
preceeding a '\'. If you nest calls
you have to use multiple escapes.

CALCULATOR

plus takes exactly the first two input
numbers and
creates one output with the sum
minus " difference
mult " product
div " quotient
per_cent " quotient * 100

SEQUENCE SPECIALS

sequence the default sequence
(in case of genes this is only the
corresponding part of the sequence)
sequence_type the default sequence's type(rna/dna..)
sequence_name the default sequence name(ali_16s,..)

format_sequence(options) takes a long string (sequence) and
breaks it into several lines

option (default)

```
        width=#    (50) sequence line width
        firsttab=# (10) first line left indent
        tab=#      (10) left indent (not first line)
        numleft    (NO) numbers on the left side
        gap=#      (10) insert a gap every # seq. characters.

    statistic              creates a character statistic of the
                           sequence
                           (not implemented yet)

    filter(options)
        exclude=#       characters in filter sequence which
                        exclude the coloumn
    or   include=#       vice versa
        SAI=#           SAI of filter sequence
    or   species=        species of filter sequence

    diff(options)          show differences to default alignment
                           of another species
        SAI=#
        species=#
        [equal=.]

    extract_sequence("chars",rel_len)
                           like extract_words, but do not sort
                           words, but rel_len is the minimum
                           percantage of characters of a word
                           that mach a character in 'chars'
                           before word is taken. All words will be
                           seperated by white space.

EXAMPLES       sequence|format_sequence(firsttab=0;tab=10)|"SEQUENCE  ";dd

                  fetches the default sequence, formats it,
                  and prepends 'SEQUENCE  '.

    sequence|remove(".-")|format_sequence
                  get the default sequence, remove all '.-' and
                  format it

    sequence|remove(".-")|len

                  the number of non '.-' symbols (
                        sequence length )

    sequence|statistic

:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
configuration.hlp
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
TITLE          Selection of Species ( == Configurations )

OCCURRENCE     ARB_EDIT4
```

DESCRIPTION Each species is either marked or unmarked. All species that are marked are called marked species. In order to handle multiple sets of marked species we invented configurations (== a selection of species). You may create a configuration from all marked species by pressing:
<ARB_NT/Species/Create Selection ...>
delete and extract configurations.

Each configuration has a unique name and is part of the database.
Also the current tree displayed within the ARB_NT main window is stored into the configuration.

The ARB_EDIT4 sequence may load all marked species or all species stored in a configuration.

NOTES Saving the database means also saving all configurations.

::
cons_params.hlp
::
TITLE Consensus Parameters

OCCURRENCE ARB_NT/SAI/Consensus/expert

DESCRIPTION Allows to define the paramters for consensus calculation.

The left part of the 'Expert Window' is similar to the 'Consensus' window the right part contains descriptions of the parameters as well as the buttons and subwindows to adjust them.

After setting the parameters, press the 'Go' button.

NOTES The consensus sequence is stored within the database as an 'SAI' entry.

::
consensus.hlp
::
TITLE Consensus

OCCURRENCE ARB_NT/SAI/Consensus

DESCRIPTION Calculates consensus for marked or all sequences.

1. Choose an 'alignment' (see 'HELP: Glossary) from the 'Alignment' subwindow.

2. Define whether 'marked' (see 'HELP: Glossary) or all sequences should be used by pressing the respective checkbox after the 'species' prompt.

3. Define whether to use or to ignore gaps entirely. If you count gaps and the gap frequency exceeds 'threshold for gaps', the

result will be shown as '-'. If the switch is 'off',
the algorythm will virtually remove all gaps.
That means if you have a coloumn with two 'A's
and 500 gaps the program thinks
of 100% 'A'. If the switch would be 'on', the relative
number of 'A's would be 2%.

4. Define whether the most frequent base or the IUPAC code
(more then one base) should be shown at the particular
position by pressing the respective checkbox after the
'simplify' prompt.
The IUPAC codes are displayed after pressing the <show
IUPAC> button.
Characters are used for IUPAC encoding only if their
frequency exceeds 'threshold for character'.
Example:
If you have 40% 'A', 10% 'C', 40% 'G' and 10% 'T' and
'threshold for character' is set to 20%, the
program looks for a iupac code only for 'A' and 'G'.

5. Define a name for the consensus sequence by selecting one
from the 'SAI' subwindow or by typing it to the
'Name of New SAI' field.

NOTES The consensus sequence is stored within the database as an 'SAI'
entry and automatically edited with 'ARB_EDIT'.

Display detailed help in postscript format by choosing
'Postscript: Consensus.ps' from the 'Main Topic' subwindow.

If a column contains only gaps the result will be shown as '='.

BUGS Our postscript viewer cannot show 'Postscript: Consensus.ps'
You may print it on your printer: $ARBHOME/lib/ps/Consensus.ps

BUGS IUPAC symbols in source sequences are ignored.

::
copyright.hlp
::
TITLE Some selected copyright notices

********** ARB **********
The ARB software and documentation are not in the public domain.
External functions used by ARB are the property of their
authors. This release of the ARB program and documentation may not
be sold, or incorporated into a commercial product, in whole or in
part, without the expressed written consent of the Technical
University of Munich and of its supervisors
Oliver Strunk & Wolfgang Ludwig.

All interested parties may redistribute the ARB as
long as all copies are accompanied by this
documentation, and all copyright notices remain

intact. Parties interested in redistribution must do so on a non-profit basis, charging only for cost of media.

********** GDE **********
The Genetic Data Environment (GDE) software and documentation are not in the public domain. Portions of this code are owned and copyrighted by the The Board of Trustees of the University of Illinois and by Steven Smith. External functions used by GDE are the property of their authors. This release of the GDE program and documentation may not be sold, or incorporated into a commercial product, in whole or in part without the expressed written consent of the University of Illinois and of its author, Steven Smith.

All interested parties may redistribute the GDE as long as all copies are accompanied by this documentation, and all copyright notices remain intact. Parties interested in redistribution must do so on a non-profit basis, charging only for cost of media. Modifications to the GDE core editor should be forwarded to the author Steven Smith. External programs used by the GDE are copyrighted by, and are the property of their respective authors unless otherwise stated.

********** PHYLIP **********
(c) Copyright 1986-1993 by Joseph Felsenstein and the University of Washington.
Permission is granted to copy this document provided that no fee is charged for it and that this copyright notice is not removed.

********** LSADT **********
LEAST SQUARES ALGORITHM FOR FITTING ADDITIVE TREES TO PROXIMITY DATA

GEERT DE SOETE -- VERSION 1.01 - FEB. 1983
VERSION 1.02 - JUNE 1983
VERSION 1.03 - JULY 1983

'C' version by Michael Macuikenas, University of Illinois

REFERENCE: DE SOETE, G. A LEAST SQUARES ALGORITHM FOR FITTING ADDITIVE TREES TO PROXIMITY DATA. PSYCHOMETRIKA, 1983, 48, 621-626.
DE SOETE, G. ADDITIVE TREE REPRESENTATIONS OF INCOMPLETE DISSIMILARITY DATA. QUALITY AND QUANTITY, 1984, 18, 387-393.

REMARKS
---

1) THE PROGRAM USES SUBROUTINES FROM THE PORT LIBRARY FOR

ERROR HANDLING, DYNAMIC STORAGE ALLOCATION AND SPECIFICA-
TION OF MACHINE-DEPENDENT CONSTANTS.
CF. FOX, P.A., HALL, A.D., & SCHRYER, N.L.
- THE PORT MATHEMATICAL SUBROUTINE LIBRAY. ACM TRANS.
ON MATH. SOFTW., 1978, 4, 104-126.
- ALGORITHM 528. FRAMEWORK FOR A PORTABLE LIBRARY.
ACM TRANS. ON MATH. SOFTW., 1978, 4, 177-188.
2) UNIFORMLY DISTRIBUTED RANDOM NUMBERS ARE GENERATED BY A
PROCEDURE DUE TO SCHRAGE. CF.
SCHRAGE, L. A MORE PORTABLE FORTRAN RANDOM NUMBER GENERATOR.
ACM TRANS. ON MATH. SOFTW., 1979, 5, 132-138.
3) SUBROUTINES VA14AD AND VA14AC ARE ADAPTED FROM THE
HARWELL SUBROUTINE LIBRARY (1979 EDITION).
4) ALTHOUGH THIS PROGRAM HAS BEEN CAREFULLY TESTED, THE
AUTHOR DISCLAIMS ANY RESPONSABILITY FOR POSSIBLE
ERRORS.

********** BLAST **********

/*
==========
==
*
* PUBLIC DOMAIN NOTICE
* National Center for Biotechnology Information
*
* This software/database is a "United States Government Work" under the
* terms of the United States Copyright Act. It was written as part of
* the author's official duties as a United States Government employee and
* thus cannot be copyrighted. This software/database is freely available
* to the public for use. The National Library of Medicine and the U.S.
* Government have not placed any restriction on its use or reproduction.
*
* Although all reasonable efforts have been taken to ensure the accuracy
* and reliability of the software and data, the NLM and the U.S.
* Government do not and cannot warrant the performance or results that
* may be obtained by using this software or data. The NLM and the U.S.
* Government disclaim all warranties, express or implied, including
* warranties of performance, merchantability or fitness for any particular
* purpose.
*
* Please cite the author in any work or product based on this material.
*
*
==========
==*/
Warren Gish
NCBI/NLM

********** CONVERT_ALN **********
convert_aln -- an alignment(or sequence) converter written by Wen-Min Kuan
for the Ribsomal Database Project(RDP), April 28, 1992.

********** TREETOOL **********

Written by Mike Maciukenas, at the RDP, with design and implementation guidance by Gary Olsen, Niels Larsen, Carl Woese.

********** GHOSTVIEW **********

********** XFIG **********

********** fastdnaml **********

```
/*  Copyright notice from dnaml:
 *
 *  version 3.3. (c) Copyright 1986, 1990 by the University of Washington
 *  and Joseph Felsenstein.  Written by Joseph Felsenstein.  Permission is
 *  granted to copy and use this program provided no fee is charged for it
 *  and provided that this copyright notice is not removed.
 *
 *  Conversion to C and changes in sequential code by Gary Olsen, 1991-1992
 *
 *  p4 version by Hideo Matsuda and Ross Overbeek, 1991-1992
 */

/********** Disclaimer **********/
      THE AUTHORS GIVE NO WARRANTIES,
      EXPRESSED OR IMPLIED FOR THE SOFTWARE AND
      DOCUMENTATION PROVIDED, INCLUDING,
      BUT NOT LIMITED TO WARRANTY OF
      MERCHANTABILITY AND WARRANTY OF
      FITNESS FOR A PARTICULAR PURPOSE.
```

:::::::: ::: :: ::::::::::::::::::
del_list.hlp
::
TITLE DELETE LISTED SPECIES/GENES

OCCURRENCE ARB_NT/Species/Search: DELETE LISTED
ARB_NT/Genes/Search: DELETE LISTED

DESCRIPTION Deletes all database entries (sequence and additional information stored within fields) assigned to the listed species/genes.

NOTES A protection level has to be selected from the Protection menu of the main window (ARB_NT/Protection) equal to or higher than that assigned to the selected alignment (ARB_NT/ali_*; third broad rectangular button in the upper part of the main window).

WARNINGS It is recommended to reset the protection level after deleting entries to prevent unintentional loss of data.

::
dewachter_ift.hlp
::
TITLE NOTES: dewachter

OCCURRENCE ARB_IMPORT

DESCRIPTION Tries to read files produced by the deWachter group. Preserves as much additional data as possible.

NOTES Nearly all output fields are tagged with [DEW]
::
dist.hlp
::

TITLE Neighbour joining

OCCURRENCE ARB_NT/Tree/Neighbour joining

DESCRIPTION Reconstructs a tree for all or marked species by first calculating binary distances and subsequently applying the neighbour joining method.

The tree topology is stored in the database and can be displayed within the tree display area of the 'ARB_NT' window.

[1.] Mark all interesting species.

2. Select all or marked species from the 'Select Species' menu

of the 'NEIGHBOUR JOINING' window.

3. Select Alignment from the 'Select Alignment' subwindow of the 'NEIGHBOUR JOINING' window.

[3.] Display the 'Select Filter' window by pressing the button after the 'Filter' prompt and define an alignment-associated mask which defines aligment positions to include for treeing.

[4.] Define Weights: !!! not implemented !!!

[5.] Select rate matrix: !!! not implemented !!!

6. Type characters for the exclusion of aligment postions to the 'Exclude Column' subwindow. The positions are excluded from the calculation of binary distance values if one of the specified characters is present in one or both sequences. The described function acts as a second filter and affects only the particular sequence pairs, not the whole alignment.

7. Select the type of distance correction from the 'Distance Correction' submenu. You can use the program to detect the best correction for you by pressing the AUTODETECT button.

none:
Differences/Sequence length. May be a good choice for short sequences (length < 300)

similarity:
1.0 - Differences/Sequence_Length

jukes-cantor:
Accounts for multiple base changes, assumes equal base frequencies.
Good choice for medium sized sequences
(300 - 1000/2000 sequence length)

felsenstein:
Similar to jukes-cantor transformation. Allows unequal base frequencies.
(length > 1000/2000)

olsen:
As Felsenstein, except the base frequencies are calculated for each pair of sequences.

!!! The other correcting functions are in an experimental state.
Wait for new release.!!!

8. Select a name for the tree from the 'Trees in Database' subwindow or type a new tree name.

The tree name has to be 'tree_*'.
An existing tree with that name will be deleted.

9. Press the 'CALCULATE TREE' button

11. Now you may display the new tree in the ARB_NT main window
by selecting its name from the <Tree/Select> subwindow.
If its name is already selected, you will not need to
reselect it.

The distance matrix can be written to an ascii file:
Press the <SAVE MATRIX> button to display the 'SAVE
MATRIX' window. Select a file from the 'Directories
and Files' subwindow or type a file name to the 'FILE
NAME' subwindow. Press the <SAVE> button.
The suffix displayed in the 'SUFFIX' subwindow is added
to the typed file name and defines the selection of
files listed in the 'Directories and Files' subwindow.

@@@ Calculate Compressed Matrix

NOTE Read the NDS help text

NOTE Computing time can be estimated using the following
formula:

time = (Sequence_Length * Nr.of.Spec * Nr.of.Spec)/
Computer Power

Examples: Sparc 10,
74 Sequences, length 8000 characters

-> 10 Seconds

WARNINGS Don't try to build a tree with the 'similarity' distance
correction selected.

::
e4.hlp
::

TITLE Prototype of The New Sequence and Secondary Structure Editor

OCCURRENCE ARB_NT/Sequence/Edit marked sequences

DESCRIPTION The editor allows you to view and modify the sequences of 'marked
species' or 'a selection of species' and 'SAI' (sequence associated
information) stored in the data base.
Potential secondary structure is automatically checked
and the information can be displayed with the primary structure.
In addition, an online column statistic may help you find
sequence and alignment errors.

The editor permanently communicates with the database and other ARB tools. Every change made by the editor is immediately exported to the database. All sequence changes made by other ARB tools are exported to the editor every 5 seconds.

Multiple editors can be used synchronuously.

Display:

First Column: Name of sequences or name of groups.
Second C. Protection level and subtype of sequence.
Last c. Sequence and secondary structure and more.

Cursor:

The cursor can be moved using the mouse or the arrow keys. The current cursor positions with respect to the alignment and the E. coli sequence (there has to be an 'SAI' entry: ECOLI) are indicated after the 'Position' and 'E.coli' prompts in the upper part of the 'ARB_EDIT4' window.

These 3 positions and the IUPAC-display refer to the position RIGHT of the cursor.

Position counting now starts with 1 (too many people were confused about position counting starting at position 0).

Moving entries:

To move an individual entry, position the cursor on the name, keep the left mouse button pressed, move the entry to the desired position, and release the button.

Any entry can be fixed (keeps its position while scrolling vertically) at the top of the editing area by moving it somewhere above the double line (=top area).

Undo/Redo:

This undoes/redoes everything you did.

Jump:

If you selected a species in any other ARB component (i.e. in the Tree or in the Search Hitlist), you can jump to the selected species using this button.

In general the editor will automatically jump to the selected species, unless the species is in a folded group. In this case all necessary groups will be unfolded.

Get:

If the species is already in the editor 'Get' does the same as 'Jump'. If the species is NOT loaded in the editor it will be inserted into the group 'More Sequences'.

If you like to load several species into the editor, mark those species and use 'ARB_EDIT4/Edit/Load marked species'

Editing:

Protection:

To perform editing, the protection level of the particular entry ('EDIT/Set protection of selected species') has to be set below or equal the global protection level (use the 'Protect' button on the menuboard).

Note: There are two global protection levels: one in edit-mode and another one in align-mode.

Modes:

ARB_EDIT4 supports 2 modes: Align-mode and Edit-mode. To toggle between these two modes use the 'Align/Edit' button on the menuboard or press CTRL-E.

Align-mode:

Only gap symbols can be inserted ('-', '.' or 'Spc' key) or deleted ('Del' or 'Backspace' key)

Sequence data cannot be modified - it only can be checked in this mode. Sequence check is performed by typing nucleotide (amino acid) symbols. Discrepancies between typed and existing symbols are indicated by beeping.

Edit-mode:

Edit-mode is devided into two submodes: Insert-mode and Replace-mode. Toggle between these submodes with the 'Insert/Replace' button on the menuboard or with CTRL-I.

In Replace-mode inserted Nucleotides/gaps overwrite existing Nucleotides/gaps.
In Insert-mode inserted Nucleotides/gaps do not overwrite, instead the whole sequence is shifted.

Direction:

ARB_EDIT4 should perform any editing function into both directions - forward and backward. To toggle the editing direction use the "5'->3'" button on the menuboard.

Repeat editing functions:

Nearly every editing function can be performed repeatedly by typing some digits before you perform the editing function.

Example: '9-' inserts nine '-'

If you need to insert digits, use the toggle in 'Properties/Options'.

Key mapping:

Nucleotide (amino acid) and gap symbols can be assigned to any of the letter and symbol keys by using the 'Key Mapping' facility ('Properties' menu).

Moving nucleotide (amino acid) symbols:

SHIFT+LEFT/RIGHT
Push or pull a coherent sequence stretch next to the cursor.

META+LEFT/RIGHT (or CTRL+O/P)
Move a single nucleotide (amino acid) symbol.

If you move your cursor towards a gap, the next nucleotide (amino acid) symbol in movement direction is fetched and moved to the actual position.

If you move your cursor towards a nucleotide (amino acid) symbol, it will jump aside the next nucleotide (amino acid) symbol.

Other keys:

CTRL+LEFT/RIGHT
Jumps to the start of the next gap.
HOME/END
Jumps to the start/end of sequence.

Messages:
Many (less serious) errors will not be announced by a popup window. Instead they appear at the small text window in the upper-right corner of the editor window.
You can press one of the small buttons at the left side of this text window:
- press the small lens to see more errors
- press the small X to get rid of all errors

WARNINGS The key mappings may conflict with some window managers (eg. olvwm). Disable the CNTRL-LEFT/RIGHT keys of the window manager.

If you perform major database modifications such as
- adding new SAIs
you have to QUIT and restart the editor.

BUGS - sometimes the editor crashes after aligning
- some update problems (please report system and circumstances)
(workaround: resize the editor window)

::
TITLE ARB_EDIT4 Block Operation

OCCURRENCE ARB_EDIT4/Block

DESCRIPTION This menu allows you to perform different block operations.

All block operations work on what we call 'Selected Species'.
To select some species use the right mouse button on the
name or sequence data of the species/groups.

More selection functions are availiable in the first section
of the Block menu:

- 'Deselect all' de-selects all species
- 'Select all' selects all species (in middle area)
- 'Invert all' inverts selection of all species (in middle area)
- 'Invert group' inverts selection of all species in group
- 'Line block <-> column block' toggles blocktype (see below)

Use the second section of the block menu to copy selection to
marks or vv.

[Please consider the difference between the following terms:
- Marked species: species marked in ARB_NT
- Selected species: species marked in ARB_EDIT4
- Current species: species under cursor]

Several actions (third section of Block menu) can be
performed on the selected block. These actions can be performed
repeatedly by typing the number of repeats into the editor window
(check 'Properties/Options/Use digits to repeat edit commands?')

There are 2 types of selected blocks:

- Line blocks include the whole sequence. To activate a line
  block use 'ARB_EDIT4/Block/Deselect all' and right-click
  on the species name.
- Column blocks include only a part of the sequence. To
  activate a column block right-click on the sequence data.

Add/remove species to/from selected:
[All clicks below are right clicks!]

- You can add or remove species by clicking on their name.
- The first click on a group name will select the whole group
  (group name is colored), the second click will select all
  species contained in the group and the third click will
  de-select all.
- click on group bracket will deselect all in group

::
e4_consensus.hlp

::

TITLE Standard help file form

OCCURRENCE ARB_NT

DESCRIPTION

::
e4_options.hlp
::

TITLE ARB_EDIT4 Options

OCCURRENCE ARB_EDIT4/Properties/Options...

DESCRIPTION This window allows you to adjust some display parameters

Online Sequence Compression

Online Sequence Compression (OSC) is a way to hide column positions (normally: column positions containing only or many gaps) in order to simplify editing of aligments with wide gaps.

OSC affects only the manner how sequences are DISPLAYED in the editor and does not affect the sequences themselves.

All compression modes are named by the following naming convention:

GAP: 0 Hide every column consisting only of gaps

GAP: SOME Dynamically show some columns for each block of columns consisting only of gaps.

HIDE: 0 Do NOT hide any column containing bases.

HIDE: xx% Hide all columns containing less than xx percent bases.

Those above mentioned column values are calculated on the basis of all sequences currently loaded in the editor.

SPEC. Special edit mode to edit over gaps
EDIT (not working yet)

COM Deactivate OSC
OFF

Editing

::
e4_replace.hlp
::
TITLE ARB_EDIT4 Replace

OCCURRENCE ARB_EDIT4/Block/Replace

DESCRIPTION This menu allows you to replace parts in the sequence data.

The selected range (see subtopics) will be searched for occurances of the string given in the first input field. If found it will be replaced by the string given in the second input field.

You can use '?' as a joker for any character in the search pattern.

WARNINGS There are no further questions if you press GO.
If you pressed GO accidently, press UNDO to clean up the mess.

BUGS None

::

e4_search.hlp

::

TITLE Search

OCCURRENCE ARB_EDIT4/Edit/Search

DESCRIPTION Search

Search patterns

In the text field you can enter multiple search patterns. Different patterns are seperated by newlines or commas.

'?' is treated as single letter wildcard

'#' is an end-of-line comment

Text written behind # will not be used for search. Instead this text will be displayed in the message window when you position your cursor on a found pattern.

Last/Next

Jumps to the last/next occurance of any of the given patterns. You can repeat your last search by pressing CTRL-S.

Show ?

If checked, the found parts are shown in different background colors (defined at Properties/Data Search)

If the different search patterns overlap, they are shown in the following order:

User (shown above all others)
Probe
Primer (shown below all other)

Open folded?

If checked, the Last/Next-Button will open folded groups, to jump to the next occurance. Otherwise search will jump over folded groups.

Auto jump?

If checked, the cursor will automatically jump to the nearest occurance, if you change the search pattern or other search parameters.

Ignore gaps in sequence?

If checked, gaps in sequence will be ignored.
(ACGU will find A-CG-U, AC---GU, ...)

Ignore gaps in pattern?

If checked, gaps in the search pattern will be ignored.
(A-CG-U, AC---GU, ... will find ACGU)

Treat T as U?

If checked, T and U will be treated as equal.
(ACGU will find ACGT and vice versa)

Ignore case?

If checked, a and A, c and C, ... are treated as equal (aCGu will find ACGU, AcgU, acgu, ...)

Search for complement?

If checked, search will go as well for complemented patterns.

Search for reverse?

If checked, search will go as well for reversed patterns.

Exact!

If checked, search will only go for the given combination of 'complement' and 'reverse'.

Example: If 'Exact', 'complement' and 'reverse' are checked, search will go only for complemented AND reversed patterns.

Allowed mismatches

Defines the minimum and maximum allowed number of non-matching base characters.

NOTES Found patterns hide possibly activated column statistics.

::

ebi_ift.hlp

::

TITLE NOTES: ebi

OCCURRENCE ARB_IMPORT

DESCRIPTION A well designed import filter for files from EBI

NOTES ebiwl is a special version of the ebi input format reader:
It reads all (!!!) comments. This may be a lot of data and
is normally not needed by simple users.

::

ecoliref.hlp

::

TITLE E. coli Reference

OCCURRENCE ARB_NT/ARB_EDIT/ETC/Reload Reference

DESCRIPTION The SAI entry 'ECOLI' is used by the editor and other
programs to detect and display the current cursor position
with respect to the homologous position within the
E. coli sequence.
If you have no SAI 'ECOLI' yet this creates one:

Search your ecoli sequence and select it. Press the menu
<species/Info(...)> -> SPECIES INFORMATION window pops up.
In this window press <SPECIES/COPY>, enter ECOLI in the
new name window, press GO, press <SPECIES/Convert to SAI>

NOTES After inserting or deleting gaps in the whole alignment (ARB_NT/
Sequence/Admin/INS/DEL CHAR) the reference SAI has to be
reloaded by choosing the 'Reload Reference' item from the
'ARB_EDIT/ETC' menu.

::

ed4_nds.hlp

::

TITLE Node Display Setup (NDS)

OCCURRENCE ARB_EDIT4/Properties/NDS

INFORMAL NOTE Read this text carefully. You won't need this function, but
it offers many many new possibilities.

DESCRIPTION Extracts data from the database entries of every species and builds a user-readable string from that data.
This string is used to show the species information at the left side of every sequence in the editor window.
It allows you to show part of the sequences, the full_name, the accession numbers and more.

Choose the 'NDS' item from the 'ARB_EDIT4/Properties' menu to display the 'NDS' window.

Used maximum group depth:
Here you define your common maximum group depth. It will be added automatically to the width (see below).

Enable field extraction:
Select the desired field extraction by selecting one of the toggles at the left border.

Description:
You may enter a description for each ACI PROGRAM.
The intention is to make it more easy for you to remember what every ACI PROGRAM does.

Generate new ACI program:
Type syntax (see 'HELP: Search and Replace Tool' and 'HELP: ARB Command Interpreter') to the 'ACI SRT PROGRAM' subwindow

Width:
Defines the width of the display in characters.

EXAMPLES @@@@@

BUGS The width of the output is limited to 4000 characters.

::
exec_bug.hlp
::

TITLE How to fix an 'exec' problem ?

OCCURRENCE NDS + ACI exec - command

DESCRIPTION If you use the exec command in the NDS window then you might have to wait extremely long:
number of species *
execution time of the command
(5000 species, 1.5 sec/exec) -> 2 hour waiting.

WHAT TO DO
The only way the to remove the exec command is:
run arb_panic on any shell
save the database to /tmp/panic.arb
edit /tmp/panic.arb

search the exec command
replace it by the 'echo' command
save file
test it with 'arb /tmp/panic.arb'
If everything works fine save it in your home directory
enter 'arb_clean' on any shell
remove /tmp/panic.arb

WARNINGS Test your database very carefully because it may not be consistent.
Check sequence checksum

::
export_format.hlp
::
TITLE Export File Formats

OCCURRENCE ARB_NT/File/Export Foreign Format

DESCRIPTION The export format description file (*.eft) describes the format.
It contains different sections:

SECTIONS
SUFFIX suffix ;The suggested file suffix
headerinfo ;Header information
BEGIN ... ;the main section

The main section is part of a 'SRT' (see 'HELP: Glossary') conversion.
(In fact the main section S is replaced by "*=S" and the SRT is started on an empty string)
That means:
1. All simple 'text' is written to the output file.
2. All references '*([ref][:|#modifier])' are replaced by the [modified] value of the reference
3. All lines containing the word '$$DELETE_LINE$$' are deleted.

SPECIAL SECTIONS
PRE_FORMAT xxx.eft
SYSTEM "command"

First a output is created using xxx.eft. Then command is used to convert the result into the wanted result.
In command use
$< as input-filename
$> as output-filename

INTERNAL type

used for formats hardcoded in ARB_NT.
Supported types:

xml_write Writes all none-hidden fields
to XML.

NOTES The best way to design new formats is:
1. Copy an existing format and modify it step by step.
2. Design a simple format and modify it step by step.

EXAMPLES Edit the files in '$ARBHOME/lib/export'

WARNINGS Don't change an '*.eft' file, if you don't understand how
it works.
Be careful when using ':' or "" symbols.
Escape them if those symbols don't have a meaning.

::
exportcursor.hlp
::
TITLE Synchronize Cursor

OCCURRENCE ARB_EDIT/ETC/Synchronize Cursor Position

DESCRIPTION Synchronizes the editor and other tools.

Choose the 'Synchronize Cursor Position' item from the
'ARB_EDIT/ETC' menu. The cursor position is recognized
by other tools such as 'ARB_EDIT/EDIT/Align Sequence' or
'ARB_NT/Sequence/Admin/INS/DEL CHAR'.

Choose the 'Don't Synchronize Cursor' item from the
'ARB_EDIT/ETC' menu to restore the default status.

NOTES Working with synchronized cursor reduces the performance (speed)
of the editor.

::
extended.hlp
::

TITLE SAI Sequence Associated Information

DESCRIPTION The main database devides all sequences into two groups:

I. Species sequences. ->Species
II. All other sequences ->SAI
(masks, filters, profiles)

You should keep the number of SAI's small for you cannot:
- search SAI's
- select a subset of SAI's
- exclude SAI's from the ARB_EDITor

IMPLEMENTED SAI's

ECOLI A copy of the E.coli sequence. It is used
to convert the absolute alignment position into
an E.coli-based position and vice versa.

HELIX_NR Helix Numbering
HELIX and Helix template
Note: Read the respective help file

FUTURE SAI's
COLORMASK a global colormask
INSERTS memory of all global alignment inserts

::
faligner.hlp
::
TITLE Fast Align Sequence

OCCURRENCE ARB_NT/ARB_EDIT/EDIT/Fast Align Sequence

DESCRIPTION Aligns or re-aligns a sequence using a new and faster algorithm.

Choose the 'Fast Align Sequence' item from the 'ARB_EDIT/EDIT'
menu to display the 'FastAligner' window.

ADJUSTMENTS Align Align selected or all marked sequences.
If you type 'Ctrl-A' in the main editor window
this option is set to selected species.

Reference The aligner needs a sequence as reference.
You can either select a fixed species, the
konsensus of the group containing the species
or the next relative found by the selected
PT-Server.

If you choose 'Auto search by pt_server', the
aligner will use the next relative to align.
If the next relative has gaps where the sequence
to align has bases, a mix of the nearest
relatives will be used. You can define the
maximum number of relatives.

Range Align only a part or whole sequence.
If you align only a part of the sequence, then
you have to enter the number of columns around
the cursor.

Example: If you align 10 columns around
position 100 then columns 90-110
get aligned.

Turn check The aligner is able to detect sequences which
were entered in the wrong direction. With this
switch you can select, if you like the aligner
to turn such sequences and if it should ask you.

NOTE: In two cases turn checking isn't
reasonable:

If you align only a part of a sequence or if you
do not search Reference via pt_server. In both
cases turn checking will be disabled.

Report The aligner can generate reports for the aligned
sequence and for the reference sequence. These
reports can be viewed with EDIT4, if you choose
File/Load Configuration/DEFAULT_CONFIGURATION

The report for the reference sequence (AMI)
contains a '>' for every position were the aligner
needed an insert in the reference sequence.

The report for the aligned sequence (ASC) contains
the following characters:
'-' for matching positions
'+' for inserts (in aligned sequence and in
reference sequence)
'~' for matching, but not equal bases (A aligned
to G, C aligned to T or U)
'#' for mismatching positions

NOTES This aligner knows about and uses all extended base characters
(ACGTUMRWSYKVHDN) for the alignment.
In other words: M aligned to R costs no penalty.

BUGS If you select the menu entry 'remove all aligner entries' ARB_EDIT4
crashes in most cases.
Workaround: 1. Close all groups containing species with aligner
entries, so that no aligner entries are visible.
2. Remove all aligner entries
3. Reload configuration

::
fasta_ift.hlp
::
TITLE NOTES: fasta

OCCURRENCE ARB_IMPORT

DESCRIPTION The fasta format is the most simple format available.
It looks like this:

>name_of_species some_additional_information
acgtuacgacgcaacg
acgcacgtutcat

```
>name_of_second_species
acgtuacgacgcaacg
```

NOTES The fasta format is relatively fool prove.

::
fastdnaml.hlp
::

fastDNAml 1.0

Gary J. Olsen, Department of Microbiology
University of Illinois, Urbana, IL
gary@phylo.life.uiuc.edu

Hideo Matsuda, Mathematics and Computer Science
Argonne National Laboratory, Argonne, IL
matsuda@mcs.anl.gov

Ray Hagstrom, Physics
Argonne National Laboratory, Argonne, IL
hagstrom@mcs.anl.gov

Ross Overbeek, Mathematics and Computer Science
Argonne National Laboratory, Argonne, IL
overbeek@mcs.anl.gov

Citing fastDNAml

If you publish work using fastDNAml, please cite the following publications:

Olsen, G. J., Matsuda, H., Hagstrom, R., and Overbeek, R. 1994. fastDNAml: A tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput. Appl. Biosci. 10: 41-48.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol. 17: 368-376.

What is fastDNAml

fastDNAml is a program derived from Joseph Felsenstein's version 3.3 DNAML (part of his PHYLIP package). Users should consult the documentation for DNAML before using this program.

fastDNAml is an attempt to solve the same problem as DNAML, but to do so faster and using less memory, so that larger trees and/or more bootstrap replicates become tractable. Much of fastDNAml is merely a recoding of the PHYLIP 3.3 DNAML program from PASCAL to C.

DNAML includes the following notice:

Why is fastDNAml faster?

Some recomputation of values has been eliminated (Joe Felsenstein has done much of this in version 3.4 DNAML).

The optimization of branch lengths has been accelerated by changing from an EM method to Newton's method (Joe Felsenstein has done much of this in version 3.4 DNAML).

The strategy for simultaneously optimizing all of the branches on the tree has been modified to spend less time getting an individual branch right before improving the other branches.

Other new features in fastDNAml

fastDNAml includes a checkpoint feature to regularly save its progress toward finding a large tree. If the program is interrupted, a minor change to the input file and adding the R (restart) option permits the work to be resumed from the last checkpoint.

The new R {restart) option can also be used for more rapid addition of new sequences to a previously computed tree (when new sequences are added to the alignment, it is best if the relative alignment of the previous sequences is not altered).

The G (global) option has been generalized to permit crossing any number of branches during tree rearrangements. In addition, it is possible to modify the extent of rearrangement explored during the sequential addition phase of tree building.

The G U (global and user tree) option combination instructs the program to find the best of the user trees, and then look for rearrangements that are better still.

The number of available rate categories has been raised from 9 to 35.

The weighting mask accepts values from 0 through 35.

The new B (bootstrap) option causes generation of a bootstrap sample, drawn from the input data.

The program includes "P4" code for distributing the problem over multiple processors (either within one machine, or across multiple machines).

Do DNAML and fastDNAml give the same answer?

Generally yes, though there are some reservations:

One or the other might find a better tree due to minor changes in the ways trees are searched. When sequence addition is replicated with different values of the jumble random number seed, they have about the same probability of finding the best tree, but any given seed might give different trees.

The likelihoods and branch lengths sometimes differ very slightly due to different criteria for stopping the optimization process.

An earlier version of fastDNAml had an error in generating tree rearrangements in the search for better trees. This did not affect the default (local) rearrangements, but could have caused the program to miss some global rearrangements. We think that it is now correct, but it is one of the most difficult program features to test.

Little has been done to check the confidence limits on branch lengths. There seem to be some instances in which they disagree, and we think that fastDNAml is correct. However, do not take the "significantly greater than zero" too seriously.

If you are concerned, you can supply a tree inferred by fastDNAml as a user tree to PHYLIP DNAML and let it (1) reoptimize branch lengths, (2) tell you the confidence limits and (3) tell you the tree likelihood. (It may be necessary to remove the quotation marks around the species names in the treefile.)

Features in the works

Test subtree exchanges (as well as moving a single subtree) in the search for better trees.

More quickly evaluating whether a tree is a good candidate for best tree.

Allowing the program to optimize any user-defined subset of branches when user lengths are supplied.

Maintaining a list of the several best trees, not just the (single) best.

Input and Options

Basics

The input to fastDNAml is similar to that used by DNAML (and the other PHYLIP programs). The user should consult the PHYLIP documentation for a basic description of the format.

This version of fastDNAml expects to get its input from stdin (standard input) and writes its output to stdout (standard output). (There are compile time options to modify this, for those who care to get into such things.)

On a UNIX system, it is a simple matter to redirect input from a file and output to a file:

```
fastDNAml < infile > outfile
```

On a VMS system it is only slightly more difficult. Immediately before running the program, one includes two commands that define the input and output files:

```
$ Define/User  Sys$Input   infile
```

```
$ Define/User  Sys$Output  outfile
$ Run fastDNAml
```

The default input data format is Interleaved (see I option). To help get data from a GenBank or similar format, the interleaved option can be switched off with the I o ption. Numbers in the sequence data (i.e., sequence position numbers) will be ignored, so they need not be stripped out.

(Note that the program also writes a file called checkpoint.PID. See the R option below for more description.)

1 -- Print Data

By default, fastDNAml 1.0 does not echo the sequence data to the output file. Option 1 reverses this.

3 -- Do Not Print Tree

By default, fastDNAml 1.0 prints the final tree to the output file. Option 3 reverses this.

4 -- Write Tree to File

By default, fastDNAml 1.0 does not write a machine readable (Newick format) copy of the final tree to an output file. Option 4 reverses this. The tree output file will be called treefile.PID (where PID is the process ID under which fastDNAml is running).

B -- Bootstrap

Generates a bootstrap sample of the input data. Requires auxiliary data line of the form:

```
B  random_number_seed
```

Example:

```
5  114  B
B  137
```

If the W option is used, only positions that have nonzero weights are used in computing the bootstrap sample. Warning: For a given random number seed, the sample will always be the same.

PHYLIP DNAML does not include a bootstrap option. (Use the DNABOOT program.)

C -- Categories

Requires auxiliary data of the form:

```
C  number_of_categories  list_of_category_rates
```

The maximum number of categories is 35. This line is followed by a list of the rates for each site:

Categories list_of_categories [per site, one or more lines]

Category "numbers" are ordered: 1, 2, 3, ..., 9, A, B, ..., Y, Z. Category zero (undefined rate) is permitted at sites with a zero in a user-supplied weighting mask.

Example:

```
5  114  C
C  12  0.0625  0.125  0.25  0.5  1  2  4  8  16  32  64  128
Categories  5111136343678975AAA894999556678888889AAAAAA9239898629AAAAA9
        6337922466244573642225748771888981329849634 99AA9899975
```

PHYLIP DNAML is limited to categories 1 through 9. Also, in PHYLIP version 3.3, the categories data came after all the other auxiliary data, but before the user-supplied base frequencies and sequence data. If you make the C line your last auxiliary data line, the programs will behave the same.

F -- Empirical Frequencies

Instructs the program to use empirical base frequencies derived from the sequence data. Therefore the input file should not include a base frequencies line preceding the data.

Example:

```
5  114  F
```

G -- Global

If the global option is specified, there may also be an [optional] auxiliary data line of form:

```
G  N1
```

or

```
G  N1  N2
```

N1 is the number of branches to cross in rearrangements of the completed tree. The value of N2 is the number of branches to cross in testing rearrangements during the sequential addition phase of tree inference.

N1 = 1: local rearrangement (default without G option)

1 < N1 < numsp-3: regional rearrangements (crossing N1 branches)

N1>= numsp-3: global rearrangements (default with G option)

N2 <= N1 the default N2 is 1, local rearrangements.

The G option can also be used to force branch swapping on user trees, that is,

a combination of G and U options.

If the auxiliary line is supplied, it cannot be the last line of auxiliary data. (It may be necessary to add the T option with an auxiliary data line of

```
 T 2.0
```

if no other auxiliary data are used.)

Examples:

Do local rearrangements after each addition, and global after last addition:

```
 5  114  G
```

Do local rearrangements after each addition, and regional (crossing 4 branches) after last addition:

```
 5  114  G T
G  4
T  2.0
```

Do no rearrangements after each addition, and local after last addition:

```
 5  114  G T
G  1 0
T  2.0
```

PHYLIP DNAML does not support the auxiliary data line or branch swapping on a user tree.

I -- Not Interleaved

By default, fastDNAml 1.0 expects data lines for the various sequences in an interleaved format (as did PHYLIP 3.3 DNAML). The I option reverses the expected format (to non-interleaved data, in which all the data lines for one sequence before the next sequence begins). This is particularly useful for editing a GenBank or equivalent format into a valid input file (note that numbers within the sequence data are ignored, so it is not necessary to remove them).

If all the data for each sequence are on one line, then the interleaved and non-interleaved formats are degenerate. (This is the way David Swofford's PAUP program writes PHYLIP format output files.) The drawback is that many programs do not handle long lines of text. This includes the vi and EDT text editors, many electronic mail programs, and some versions of FTP for VAX/VMS systems.

PHYLIP 3.3 DNAML expects interleaved data, and does not include an I option to alter this. PHYLIP 3.4 DNAML accepts an I option, but the default format is reversed.

J -- Jumble

Randomize the sequence addition order. Requires an auxiliary input line of the form:

```
J  random_number_seed
```

Example:

```
5  114  J
J  137
```

Note that fastDNAml explores a very small number of alternative tree topologies relative to a typical parsimony program. There is a very real chance that the search procedure will not find the tree topology with the highest likelihood. Altering the order of taxon addition and comparing the trees found is a fairly efficient method for testing convergence. Typically, it would be nice to find the same best tree at least twice (if not three times), as opposed to simply performing some fixed number of jumbles and hoping that at least one of them will be the optimum.

L -- User Lengths

Causes user trees to be read with branch lengths (and it is an error to omit any of them). Without the L option, branch lengths in user trees are not required, and are ignored if present.

Example:

```
5  114  U L
```

(The U is for user tree and the L for user lengths)

O -- Outgroup

Use the specified sequence number for the outgroup. Requires an auxiliary data line of the form:

```
O  outgroup_number
```

Example:

```
5  114  O
O 5
```

This option only affects the way the tree is drawn (and written to the treefile).

Q -- Quickadd

This option greatly decreases the time in initially placing a new sequence in the growing tree (but does not change the time required to subsequently test rearrangements). The overall time savings seems to be about 30%, based on a very limited number of test cases. Its downside, if any, is unknown. This will probably become default program behavior in the near future.

If the analysis is run with a global option of "G 0 0", so that no rearrangements are permitted, the tree is build very approximately, but very quickly. This may be of greatest interest if the question is, "Where does this one new sequence fit into this known tree? The known tree is provided with the restart option, below.

PHYLIP DNAML does not include anything comparable to the quickadd option.

R -- Restart

The R option causes the program to read a user-supplied tree with less than the full number of taxa as the starting point for sequential addition of the remaining taxa. Thus, the sequence data must be followed by a valid (Newick format) tree. (The phylip_tree/2, prolog fact format, is now also supported.)

The restart option can also be used to increase the range of the search for alternative (better) trees. For example, you can take a tree produced with only "local" tree rearrangements, and increase the rearrangements to "regional" or "global" by combining the appropriate global option with the restart option. If the starting tree was written by fastDNAml, then the extent of rearrangements is saved with the tree, and will be used as the starting point for the additional search. If the tree was already globally optimized, then no additional searching will be performed.

To support the R option, after each taxon is added to the growing tree, and after each round of rearrangements, the program appends a checkpoint tree to a file called checkpoint.PID, where PID is the process number of the running fastDNAml program. The last line of this file needs to be appended to the input file when the R option is used. (This should not be confused with the U (user tree) option, which expects a number followed by that number of trees. No additional taxa are added to user trees.)

The UNIX utility tail can be used to remove the last tree from the checkpoint file, and the utility cat can be used to append it to the input. For example, the following script can be used to add a starting tree and the R option to a data file, and restart fastDNAml:

```
#! /bin/sh
if test $# -ne 1
  then echo "Usage:  restart checkpoint_file"
  exit
fi
read first_line             # first line of data file
echo "$first_line R"        # add restart option
cat -                       # rest of data file
tail -1 $1                  # append last tree in checkpoint file
```

If this shell script is in the file called restart, then one might use the command:

```
restart  checkpoint.21312  < infile  | fastDNAml  > new_outfile
 ^script  ^checkpoint tree    ^data     ^dnaml program  ^output_file
```

If this is too opaque, don't worry about it, or talk with your local unix

wizard. In the mean time, this and other useful shell scripts are provided with the program.

PHYLIP DNAML does not write checkpoint trees and does not have a restart option.

T -- Transition/transversion ratio

Use a user-specified ratio of transition to transversion type substitutions. Without the T option, a value of 2.0 is used. Requires an auxiliary data line of the form:

```
 T  ratio
```

Example:

```
 5  114  T
 T  1.0
```

(Note that a T option with a value of 2.0 does nothing, but it can provide a last auxiliary data line following optional auxiliary data. See the examples for G and Y.)

U -- User Tree(s)

Read an input line with the number of user-specified trees, followed by the specified number of trees. These data immediately follow the sequence data.

The trees must be in Newick format, and terminated with a semicolon. (The program also accepts a pseudo_newick format, which is a valid prolog fact.)

The tree reader in this program is more powerful than that in PHYLIP 3.3. In particular, material enclosed in square brackets, [like this], is ignored as comments; taxa names can be wrapped in single quotation marks to support the inclusion of characters that would otherwise end the name (i.e., '(', ')', ':', ';', '[', ']', ',' and ' '); names of internal nodes are properly ignored; and exponential notation (such as 1.0E-6) for branch lengths is supported.

W -- Weights

Read user-specified column weighting information. This option requires auxiliary data of the form:

```
 Weights    list_of_weight_values    [per site, one or more lines]
```

Example:

```
 5  114  W
 Weights    11111111111100110000010001111110000000000000011000011000000
          11110111111111111111111111011100000111001011100000000011
```

It is necessary that the weight values not start before the 11'th character in the line, or some of them will be lost. Weights from 0 to 35 are indicated by

the series: 0, 1, 2, 3, ..., 9, A, B, ..., Y, Z.

PHYLIP DNAML does not support user weights with values other than 1 or 0. This limit has been removed in fastDNAml 1.0 to permit the use of user weights as a mechanism for representing a bootstrap sample (that is, only the auxiliary data lines change, not the body of the data file).

Y -- Write Tree

Output final tree to an output file called treefile.PID. By default the tree is in Newick format.

If an auxiliary input line of the form

```
 Y 2
```

is also included, then the tree output file is written as a prolog fact:

```
 pseudo_newick([Comment], (Subtree1, Subtree2, Subtree3):Length).
```

where each subtree is either

```
 (Subtree1,Subtree2):Length
```

or

```
 Label:Length
```

Because this auxiliary input line is optional, it cannot be the last auxiliary data line.

Examples:

```
 5  114  Y
```

```
 5  114  Y T
 Y 2
 T 2.0
```

PHYLIP DNAML does not append the PID (process ID) to the tree file name and does not support the prolog format output.

Examples:

Data file with empirical frequencies (generic analysis):

```
5  114  F
Sequence1
ACACGGTGTCGTATCATGCTGCAGGATGCTAGACTGCGTCANATGTTCGTACTAACTGTG
Sequence2
ACGCGGTGTCGTGTCATGCTACATTATGCTAGACTGCGTCGGATGCTCGTATTGACTGCG
Sequence3
ACGCGGTGCCGTGTNATGCTGCATTATGCTCGACTGCGRCGGATGCTAGTATTGACTGCG
```

Sequence4
ACGCGCTGCCGTGTCATCCTACACGATGCYAGACAGCGTCAGCTGCTAGTACTGGCTGAG
Sequence5
ACGCGCTGTCGTGTCATACTGCAGGATGCTAGACTGCGTCAGCTGCTAGTACTGGCTGAG

AGCTCGATGATCGGTGACGTAGACTCAGGGGCCATGCCGCGAGTTTGCGATGCG
AGCACGGTGATCAATGACGTAGNCTCAGGRTCCACGCCGTGACTTTGTGATNCG
AGCACGATGACCGATGACGTAGACTGAGGGTCCGTGCCGCGACTTTGTGATGCG
ACCTCGGTGATTGATGACGTAGACTGCGGGTCCATGCCGCGATTTTGCGRTGCG
ACCTCGATGCTCGATGACGTAGACTGCGGGTCCATGCCGTGATTTTGCGATGCG

Data file with empirical frequencies and a random addition order:

5 114 F J
J 137
Sequence1
ACACGGTGTCGTATCATGCTGCAGGATGCTAGACTGCGTCANATGTTCGTACTAACTGTG
Sequence2
ACGCGGTGTCGTGTCATGCTACATTATGCTAGACTGCGTCGGATGCTCGTATTGACTGCG
Sequence3
ACGCGGTGCCGTGTNATGCTGCATTATGCTCGACTGCGRCGGATGCTAGTATTGACTGCG
Sequence4
ACGCGCTGCCGTGTCATCCTACACGATGCYAGACAGCGTCAGCTGCTAGTACTGGCTGAG
Sequence5
ACGCGCTGTCGTGTCATACTGCAGGATGCTAGACTGCGTCAGCTGCTAGTACTGGCTGAG

AGCTCGATGATCGGTGACGTAGACTCAGGGGCCATGCCGCGAGTTTGCGATGCG
AGCACGGTGATCAATGACGTAGNCTCAGGRTCCACGCCGTGACTTTGTGATNCG
AGCACGATGACCGATGACGTAGACTGAGGGTCCGTGCCGCGACTTTGTGATGCG
ACCTCGGTGATTGATGACGTAGACTGCGGGTCCATGCCGCGATTTTGCGRTGCG
ACCTCGATGCTCGATGACGTAGACTGCGGGTCCATGCCGTGATTTTGCGATGCG

Data file with empirical frequencies and a bootstrap resampling:

5 114 F B
B 137
Sequence1
ACACGGTGTCGTATCATGCTGCAGGATGCTAGACTGCGTCANATGTTCGTACTAACTGTG
Sequence2
ACGCGGTGTCGTGTCATGCTACATTATGCTAGACTGCGTCGGATGCTCGTATTGACTGCG
Sequence3
ACGCGGTGCCGTGTNATGCTGCATTATGCTCGACTGCGRCGGATGCTAGTATTGACTGCG
Sequence4
ACGCGCTGCCGTGTCATCCTACACGATGCYAGACAGCGTCAGCTGCTAGTACTGGCTGAG
Sequence5
ACGCGCTGTCGTGTCATACTGCAGGATGCTAGACTGCGTCAGCTGCTAGTACTGGCTGAG

AGCTCGATGATCGGTGACGTAGACTCAGGGGCCATGCCGCGAGTTTGCGATGCG
AGCACGGTGATCAATGACGTAGNCTCAGGRTCCACGCCGTGACTTTGTGATNCG
AGCACGATGACCGATGACGTAGACTGAGGGTCCGTGCCGCGACTTTGTGATGCG
ACCTCGGTGATTGATGACGTAGACTGCGGGTCCATGCCGCGATTTTGCGRTGCG
ACCTCGATGCTCGATGACGTAGACTGCGGGTCCATGCCGTGATTTTGCGATGCG

Data with weighting mask and rate categories:

```
5  114  F W C
Weights    1111111111110011000001000111111000000000000000110000110000000
        1111011111111111111111111011100000111001011100000000011
C  10  0.0625  0.125  0.25  0.5  1  2  4  8  16  32
Categories  5111136343678975AAA894999556678888889AAAAAA9239898629AAAAA9
        6337922466244573642225748771888981329849634 99AA9899975
Sequence1
ACACGGTGTCGTATCATGCTGCAGGATGCTAGACTGCGTCANATGTTCGTACTAACTGTG
Sequence2
ACGCGGTGTCGTGTCATGCTACATTATGCTAGACTGCGTCGGATGCTCGTATTGACTGCG
Sequence3
ACGCGGTGCCGTGTNATGCTGCATTATGCTCGACTGCGRCGGATGCTAGTATTGACTGCG
Sequence4
ACGCGCTGCCGTGTCATCCTACACGATGCYAGACAGCGTCAGCTGCTAGTACTGGCTGAG
Sequence5
ACGCGCTGTCGTGTCATACTGCAGGATGCTAGACTGCGTCAGCTGCTAGTACTGGCTGAG

        AGCTCGATGATCGGTGACGTAGACTCAGGGGCCATGCCGCGAGTTTGCGATGCG
        AGCACGGTGATCAATGACGTAGNCTCAGGRTCCACGCCGTGACTTTGTGATNCG
        AGCACGATGACCGATGACGTAGACTGAGGGTCCGTGCCGCGACTTTGTGATGCG
        ACCTCGGTGATTGATGACGTAGACTGCGGGTCCATGCCGCGATTTTGCGRTGCG
        ACCTCGATGCTCGATGACGTAGACTGCGGGTCCATGCCGTGATTTTGCGATGCG
```

Data with three user-specified tree branching orders:

```
5  114  F U
Sequence1
ACACGGTGTCGTATCATGCTGCAGGATGCTAGACTGCGTCANATGTTCGTACTAACTGTG
Sequence2
ACGCGGTGTCGTGTCATGCTACATTATGCTAGACTGCGTCGGATGCTCGTATTGACTGCG
Sequence3
ACGCGGTGCCGTGTNATGCTGCATTATGCTCGACTGCGRCGGATGCTAGTATTGACTGCG
Sequence4
ACGCGCTGCCGTGTCATCCTACACGATGCYAGACAGCGTCAGCTGCTAGTACTGGCTGAG
Sequence5
ACGCGCTGTCGTGTCATACTGCAGGATGCTAGACTGCGTCAGCTGCTAGTACTGGCTGAG

        AGCTCGATGATCGGTGACGTAGACTCAGGGGCCATGCCGCGAGTTTGCGATGCG
        AGCACGGTGATCAATGACGTAGNCTCAGGRTCCACGCCGTGACTTTGTGATNCG
        AGCACGATGACCGATGACGTAGACTGAGGGTCCGTGCCGCGACTTTGTGATGCG
        ACCTCGGTGATTGATGACGTAGACTGCGGGTCCATGCCGCGATTTTGCGRTGCG
        ACCTCGATGCTCGATGACGTAGACTGCGGGTCCATGCCGTGATTTTGCGATGCG
3
(Sequence1,(Sequence2,Sequence3),(Sequence4,Sequence5));
(Sequence2,(Sequence1,Sequence3),(Sequence4,Sequence5));
(Sequence3,(Sequence1,Sequence2),(Sequence4,Sequence5));
```

Data with transition/transversion ratio and base frequencies to
simulate Jukes & Cantor model:

```
5  114  T
T 0.501
0.25 0.25 0.25 0.25
```

Sequence1
ACACGGTGTCGTATCATGCTGCAGGATGCTAGACTGCGTCANATGTTCGTACTAACTGTG
Sequence2
ACGCGGTGTCGTGTCATGCTACATTATGCTAGACTGCGTCGGATGCTCGTATTGACTGCG
Sequence3
ACGCGGTGCCGTGTNATGCTGCATTATGCTCGACTGCGRCGGATGCTAGTATTGACTGCG
Sequence4
ACGCGCTGCCGTGTCATCCTACACGATGCYAGACAGCGTCAGCTGCTAGTACTGGCTGAG
Sequence5
ACGCGCTGTCGTGTCATACTGCAGGATGCTAGACTGCGTCAGCTGCTAGTACTGGCTGAG

AGCTCGATGATCGGTGACGTAGACTCAGGGGCCATGCCGCGAGTTTGCGATGCG
AGCACGGTGATCAATGACGTAGNCTCAGGRTCCACGCCGTGACTTTGTGATNCG
AGCACGATGACCGATGACGTAGACTGAGGGTCCGTGCCGCGACTTTGTGATGCG
ACCTCGGTGATTGATGACGTAGACTGCGGGTCCATGCCGCGATTTTGCGRTGCG
ACCTCGATGCTCGATGACGTAGACTGCGGGTCCATGCCGTGATTTTGCGATGCG

Non-interleaved data:

5 114 F I
Sequence1
ACACGGTGTCGTATCATGCTGCAGGATGCTAGACTGCGTCANATGTTCGTACTAACTGTG
AGCTCGATGATCGGTGACGTAGACTCAGGGGCCATGCCGCGAGTTTGCGATGCG
Sequence2
ACGCGGTGTCGTGTCATGCTACATTATGCTAGACTGCGTCGGATGCTCGTATTGACTGCG
AGCACGGTGATCAATGACGTAGNCTCAGGRTCCACGCCGTGACTTTGTGATNCG
Sequence3
ACGCGGTGCCGTGTNATGCTGCATTATGCTCGACTGCGRCGGATGCTAGTATTGACTGCG
AGCACGATGACCGATGACGTAGACTGAGGGTCCGTGCCGCGACTTTGTGATGCG
Sequence4
ACGCGCTGCCGTGTCATCCTACACGATGCYAGACAGCGTCAGCTGCTAGTACTGGCTGAG
ACCTCGGTGATTGATGACGTAGACTGCGGGTCCATGCCGCGATTTTGCGRTGCG
Sequence5
ACGCGCTGTCGTGTCATACTGCAGGATGCTAGACTGCGTCAGCTGCTAGTACTGGCTGAG
ACCTCGATGCTCGATGACGTAGACTGCGGGTCCATGCCGTGATTTTGCGATGCG

Non-interleaved data by editing a GenBank format (make sure that the names are
padded to at least ten characters with blanks):

5 114 F I
Sequence1
1 ACACGGTGTC GTATCATGCT GCAGGATGCT AGACTGCGTC ANATGTTCGT
ACTAACTGTG
61 AGCTCGATGA TCGGTGACGT AGACTCAGGG GCCATGCCGC GAGTTTGCGA TGCG
Sequence2
1 ACGCGGTGTC GTGTCATGCT ACATTATGCT AGACTGCGTC GGATGCTCGT
ATTGACTGCG
61 AGCACGGTGA TCAATGACGT AGNCTCAGGR TCCACGCCGT GACTTTGTGA TNCG
Sequence3
1 ACGCGGTGCC GTGTNATGCT GCATTATGCT CGACTGCGRC GGATGCTAGT
ATTGACTGCG
61 AGCACGATGA CCGATGACGT AGACTGAGGG TCCGTGCCGC GACTTTGTGA TGCG
Sequence4

```
      1 ACGCGCTGCC GTGTCATCCT ACACGATGCY AGACAGCGTC AGCTGCTAGT
ACTGGCTGAG
     61 ACCTCGGTGA TTGATGACGT AGACTGCGGG TCCATGCCGC GATTTTGCGR TGCG
Sequence5
      1 ACGCGCTGTC GTGTCATACT GCAGGATGCT AGACTGCGTC AGCTGCTAGT
ACTGGCTGAG
     61 ACCTCGATGC TCGATGACGT AGACTGCGGG TCCATGCCGT GATTTTGCGA TGCG

Data analysis restarted from a four-taxon tree (which happens to be wrong,
but it will be corrected by local rearrangements after the tree is read):

5  114  F R
Sequence1
ACACGGTGTCGTATCATGCTGCAGGATGCTAGACTGCGTCANATGTTCGTACTAACTGTG
Sequence2
ACGCGGTGTCGTGTCATGCTACATTATGCTAGACTGCGTCGGATGCTCGTATTGACTGCG
Sequence3
ACGCGGTGCCGTGTNATGCTGCATTATGCTCGACTGCGRCGGATGCTAGTATTGACTGCG
Sequence4
ACGCGCTGCCGTGTCATCCTACACGATGCYAGACAGCGTCAGCTGCTAGTACTGGCTGAG
Sequence5
ACGCGCTGTCGTGTCATACTGCAGGATGCTAGACTGCGTCAGCTGCTAGTACTGGCTGAG

          AGCTCGATGATCGGTGACGTAGACTCAGGGGCCATGCCGCGAGTTTGCGATGCG
          AGCACGGTGATCAATGACGTAGNCTCAGGRTCCACGCCGTGACTTTGTGATNCG
          AGCACGATGACCGATGACGTAGACTGAGGGTCCGTGCCGCGACTTTGTGATGCG
          ACCTCGGTGATTGATGACGTAGACTGCGGGTCCATGCCGCGATTTTGCGRTGCG
          ACCTCGATGCTCGATGACGTAGACTGCGGGTCCATGCCGTGATTTTGCGATGCG
(Sequence4:0.1,Sequence2:0.1,(Sequence1:0.1,Sequence5:0.1):0.1):0.0;

Data analysis restarted from a four-taxon tree (which is wrong, and which
will not be corrected after the tree is read due to the suppression of all
rearrangements by the global 0 0 option):

5  114  F R G T
G 0 0
T 2.0
Sequence1
ACACGGTGTCGTATCATGCTGCAGGATGCTAGACTGCGTCANATGTTCGTACTAACTGTG
Sequence2
ACGCGGTGTCGTGTCATGCTACATTATGCTAGACTGCGTCGGATGCTCGTATTGACTGCG
Sequence3
ACGCGGTGCCGTGTNATGCTGCATTATGCTCGACTGCGRCGGATGCTAGTATTGACTGCG
Sequence4
ACGCGCTGCCGTGTCATCCTACACGATGCYAGACAGCGTCAGCTGCTAGTACTGGCTGAG
Sequence5
ACGCGCTGTCGTGTCATACTGCAGGATGCTAGACTGCGTCAGCTGCTAGTACTGGCTGAG

          AGCTCGATGATCGGTGACGTAGACTCAGGGGCCATGCCGCGAGTTTGCGATGCG
          AGCACGGTGATCAATGACGTAGNCTCAGGRTCCACGCCGTGACTTTGTGATNCG
          AGCACGATGACCGATGACGTAGACTGAGGGTCCGTGCCGCGACTTTGTGATGCG
          ACCTCGGTGATTGATGACGTAGACTGCGGGTCCATGCCGCGATTTTGCGRTGCG
          ACCTCGATGCTCGATGACGTAGACTGCGGGTCCATGCCGTGATTTTGCGATGCG
(Sequence4:0.1,Sequence2:0.1,(Sequence1:0.1,Sequence5:0.1):0.1):0.0;
```

::
gde.hlp
::

TITLE GDE Interface and Editor

DESCRIPTION Starts the GDE Editor designed by Steven Smith.
See next chapter of this text for the original help text.
As GDE originally used its own built-in database, it had to be
slightly modified to run under ARB. So

**** READ THE WARNINGS/BUGS CAREFULLY ****

WARNINGS As soon as you start GDE, it creates a copy of the selected
sequences. That means that you may change the sequences
with either GDE or ARB, but not both. Therefore, if you have started
GDE, do nothing but sequence editing in GDE till you quit GDE.
To really save sequences to disc, you have to send the sequence
changes to ARB and then use ARB to save the ARB database.

BUGS Many functions, especially
-deleting,
-moving,
-duplicating,
-creating,
-importing,
species do not work correctly.

********* Part of the Original GDE HELPTEXT ******************

.c.Introduction

The Genetic Data Environment is part of a growing set of programs for manipulating and analyzing "genetic" data. It differs in design from other analysis programs in that it is intended to be an expandable and customizable system, while still being easy to use.

There are a tremendous number of publicly available programs for sequence analysis. Many of these programs have found their way into commercial packages which incorporate them into integrated, easy to use systems. The goal of the GDE is to minimize the amount of effort required to integrate sequence analysis functions into a common environment. The GDE takes care of the user interface issues, and allows the programer to concentrate on the analysis itself. Existing programs can be tied into the GDE in a matter of hours (or minutes) as apposed to days or weeks. Programs

may be written in any language, and still seamlessly be incorporated into the GDE.

These programs are, and will continue to be, available at no charge. It is the hope that this system will grow in functionality as more and more people see the benefits of a modular analysis environment. Users are encouraged to make modifications to the system, and forward all changes and additions to Steven Smith at smith@bioimage.millipore.com.

.c.What's New for this Release

GDE 2.2 represents a maintainence release. Several small bugs have been fixed, as well as new editing features and user interface elements. Also, I have tried to update all of the contributed external programs to their latest release. Updated programs include:

Phylip
Treetool
LoopTool
Readseq
Blast
Fasta

Improved versions of printing, and translate are included as well. As for new editing features, a useful "yanking" feature has been added by Scott Ferguson from Exxon Research, and the capability to export the colormap for a seqeunce (see appendicies A/C). Among the bugs fixed in this release are:

Selection mask problems when exporting to Genbank (fixed in 2.1)
Memory leaks (fixed in 2.1)
Correct handling of circular sequences
More liberal interpretation of Genbank formatted files. (not column dependent)

.c.System Requirements

GDE 2.2 currently runs on the Sun family of workstations. This includes the Sun3 and Sun4 (Sparcstation) systems. It was written in XView, and runs on Suns using OpenWindows 3.0 or MIT's X Windows. It runs in both monochrome, and color, and can be run remotely on any system capable of running X Windows Release 4. You should have at least 15 meg of free disk space available. The binay release for SparcStations was compiled under

SunOS 4.1.2 and Openwindows 3.0.

We are also supporting a DECStation version of GDE. This is running under XView 3.0/X11R5. We encourage interested people to port the programs to their favorite Unix platform. There are informal ports to the SGI line of unix machines.

.c.Note to Motif users

GDE2.2 can be run using different window managers. The most common alternative to olwm is the Motif window manager (mwm). The only problem in using another window manager is that the status line is not displayed. We have added a "Message panel" as an option under "File->Properties" which displays all of the information contained on the status line.

People using other window managers may also prefer using xterm, and xedit as default terminals and file editors. This can be accomplished by replacing all occurrences of 'shelltool' and 'textedit' with 'xterm -e' and 'xedit' in the $GDE_HELP_DIR/.GDEmenus file.

FastA and Blast need to have the properly formatted databases installed in the $GDE_HELP_DIR under the directories FASTA/PIR, FASTA/GENBANK, BLAST/pir BLAST/genbank. For FASTA, simply copy a version of PIR and Genbank into the proper directory. Alternately, the PIR and GENBANK files can be symbolic links to copies of Genbank held elsewhere on your system. You may need to look at the .GDEmenus file in $GDE_HELP_DIR to verify that you are using the same divisions for these databases.

Blast installation involves converting PIR and GENBANK to a temporary FASTA format (using pir2fasta and gb2fasta) and then using pressdb for nucleic acid, and setdb for amino acid to reformat the databases again into blast format. The .GDEmenus file is currently set up to search with blast using the following databases: pir, genpept, genupdate, and genbank. If you wish to divide these into subdivisions, then the .GDEmenus file will have to be edited.

The most up to date release of blast can be obtained via anonymous ftp to ncbi.nlm.nih.gov. The most recent release of FASTA can be obtained via anonymous ftp to uvaarpa.virginia.edu. It is strongly recommended that you retrieve these copies,

and become familiar with their setup.

.c.Using the GDE

It is assumed that the user is familiar with the Unix, and OpenWindows/Xwindows environments. It is also assumed that people running standard MIT X-Windows will be using the OpenLook window manager (olwm). Other window managers work with varied success. If you are not certain as to how your system is set up, please contact your systems administrator.

The GDE uses a menu description language to define what external programs it can call, and what parameters and data to pass to each function. This language allows users to customize their own environment to suite individual needs.

The following is how the GDE handles external programs when selected from a menu:

Each step in this process is described in a file .GDEmenus in the user's current or home directory.

The language used in this file describes three phases to an external function call. The first phase describes the menu item as it will appear, and the Unix command line that is actually run when it is selected. The second phase describes how to prompt for the parameters needed by the function. The third phase describes what data needs to be passed as input to the external function, and what data (if any) needs to be read back from its output.

The form of the language is a simple keyword/value list delimited by the colon (:) character. The language retains old values until new ones are set. For example, setting the menu name is done once for all items in that menu, and is only reset when the next menu is reached.

The keywords for phase one are:

menu:menu name
    Name of current menu
item:item name
    Name of current menu item
itemmeta:meta_key
    Meta key equivalence (quick keys)
itemhelp:help_file

Help file (either full path, or in

GDE_HELP_DIR)
itemmethod:Unix command

The item method command is a bit more involved, it is the Unix command that will actually run the external program intended. It is one line long, and can be up to 256 characters in length. It can have embedded variable names (starting with a '$') that will be replaced with appropriate values later on. It can consist of multiple Unix commands separated by semi-colons (;), and may contain shell scripts and background processes as well as simple command names. Examples will be given later.

The keywords for phase two are:

arg:argument_variable_name
Name of this variable. It will appear in the itemmethod: line with a dollar sign ($) in front of it.

argtype:slider,chooser,choice_menu or text
The type of graphic object representing this argument.

arglabel:descriptive label
A short description of what this argument represents

argmin:minimum_value (integer)
Used for sliders.

argmax:maximum_value (integer)
Used for sliders.

argvalue:default_value (integer)
It is the numeric value associated with sliders or the default choice in choosers, choice_menus, and choice_lists (the first choice is 0, the second is 1 etc.)

argtext:default value
Used for text fields.

argchoice:displayed value:passed value
Used for choosers and choice_menus. The first value is displayed on screen, and the second value is passed to the itemmethod line.

The keywords for phase three are as follows:

in:input_file
GDE will replace this name with a randomly generated temporary file name. It will then write the selected data out to this file.

informat:file_format
Write data to this file for input to this function. Currently support values are Genbank, and flat.

inmask:
This data can be controlled by a selection mask.

insave:
Do not remove this file after running the external function. This is useful for functions put in the background.

out:output_file
GDE will replace this name with a randomly generated temporary file name. It is up to the external function to fill this file with any results that might be read back into the GDE.

outformat:file_format
The data in the output file will be in this format. Currently support values are colormask, Genbank, and flat.

outsave:
Do not remove this file after reading. This is useful for background tasks.

outoverwrite:
Overwrite existing sequences in the current GDE window. Currently supported with "gde" format only.

Here is a sample dialog box, and it's entry in the .GDEmenus file:

Using the default parameters given in the dialog box, the executed Unix command line would be:

(tr '[a-z]' '[A-Z]' < .gde_001 >.gde_001.tmp ; mv .gde_001.tmp CAPS ; gde CAPS -Wx medium ; rm .gde_001) &

where .gde_001 is the name of the temporary file generated by the GDE which contains the selected

sequences in flat file format. Since the GDE runs this command in the background ('&' at the end) it is necessary to specify the insave: line, and to remove all temporary files manually. There is no output file specific because the data is not loaded back into the current GDE window, but rather a new GDE window is opened on the file. A simpler command that reloads the data after conversion might be:

item:All caps
itemmethod:tr '[a-z]' '[A-Z]' <INPUT > OUTPUT

in:INPUT
informat:flat

out:OUTPUT
outformat:flat

In this example, no arguments are specified, and so no dialog box will appear. The command is not run in the background, so the GDE can clean up after itself automatically. The converted sequence is automatically loaded back into the current GDE window.

In general, the easiest type of program to integrate into the GDE is a program completely driven from a Unix command line. Interactive programs can be tied in (MFOLD for example), however shell scripts must be used to drive the parameter entry for these programs. Programs of the form:

program_name -a1 argument1 -a2 arguement2 -f inputfile -er errorfile > outputfile

can be specified in the .GDEmenus file directly. As this is the general form of most one Unix commands, these tend to be simpler to implement under the GDE.

As functions grow in complexity, they may begin to need a user interface of their own. In these cases, the command line calling arguments are still necessary in order to allow the GDE to hand them the appropriate data, and possible retrieve results after some external manipulation.

.c.Appendix C, External functions

ClustalV - Cluster multiple sequence alignment

Author: Des Higgins.

Reference: Higgins,D.G. Bleasby,A.J. and
Fuchs,R. (1991) CLUSTAL V: improved software for multiple sequence alignment.
ms. submitted to CABIOS

Parameters:

k-tuple	pairwise search Word size for pairwise comparisons
Window size	Smaller values give faster alignments, larger values are more sensitive.
Transitions weighted	Can weight transitions twice as high as transversions (DNA only).
Fixed gap penalty	Gap insertion penalty, lower value, more gaps
Floating gap penalty	Gapextension penalty, lower value, longer gaps

Comments:

ClustalV is a directed multiple sequence alignment algorithm that aligns a set of sequences based on their level of similarity. It first uses a Lipman Peasron pairwise similarity scoring to find "clusters" of similar sequences, and prealigns those sequences. It then adds other sequences to the alignment in the order of their similarity so as to produce the cleanest alignment.

Warning: ClustalV only uses unambiguous character codes. It will also convert all sequences to upper case in the process of aligning. Clustal does not pass back comments, author etc. Be sure to keep copies of your sequences if you do not wish to lose this information.

MFOLD - RNA secondary prediction

Author: Michael Zuker
Reference: M. Zuker On Finding All Suboptimal Foldings of an RNA Molecule, Science, 244, 48-52, (1989)

J. A. Jaeger, D. H. Turner and M. Zuker
Improved Predictions of Secondary Structures for RNA.
Proc. Natl. Acad. Sci. USA, BIOCHEMISTRY, 86, 7706-7710, (1989)

J. A. Jaeger, D. H. Turner and M. Zuker
Predicting Optimal and Suboptimal Secondary Structure for RNA.
in "Molecular Evolution: Computer Analysis of Protein and
Nucleic Acid Sequences", R. F.Doolittle ed.
Methods in Enzymology, 183, 281-306 (1989)

Parameters:

Linear/circular RNA fold
ct File to save results

Comments:

MFOLD passes it's output to a program Zuk_to_gen that translates the secondary structure prediction to a nested bracket ([]) notation. This notation can then be used in the Highlight Helix, and Draw Secondary structure (LoopTool) functions.

MFOLD currently does not support much in the way of additional parameters. We hope to have all additional parameters available soon.

Blast - Basic Local Alignment Search Tool

Reference:

Karlin, Samuel and Stephen F. Altschul (1990). Methods for

assessing the statistical significance of molecular sequence features by using general scoring schemes, Proc. Natl. Acad. Sci. USA 87:2264-2268.

Altschul, Stephen F., Warren Gish,Webb Miller, Eugene W. Myers, and David J. Lipman (1990). Basic local alignment search tool, J. Mol. Biol. 215:403-410.

Altschul, Stephen F. (1991). Amino acid substitution matrices from an information theoretic perspective. J. Mol. Biol. 219:555-565.

Parameters: Which Database
Which nucleic or amino acid database to search.
Word Size Length
of initial hit. after locating a match of this length, alignment extension is attempted.
Blastn
Match score Score for matches in secondary alignment extension
Mismatch score Score for mismatches in secondary alignment extension

Blastx, tblastn, blastp, blast3
Substitution Matrix : PAM120 or PAM250

Comments: The report is loaded into a text editor. This should be saved as a new file as the default file is removed after execution. The latest version of blast can be obtained via anonymous ftp to ncbi.nlm.nih.gov.

FastA - Similarity search

Reference: W. R. Pearson and D. J. Lipman (1988), "Improved Tools for Biological Sequence Analysis", PNAS 85:2444-2448
W. R. Pearson (1990) "Rapid and Sensitive Sequence Comparison with FASTP and FASTA" Methods in Enzymology 183:63-98

Parameters:
Database
Which database to search
Number of alignments to report
SMATRIX
Which similarity matrix to use

Comments:
The FastA package includes several additional programs for pairwise alignment. We have only included a bare bones link to FastA. We hope to include a more complete setup for the actual 2.2release.

Assemble Contigs - CAP Contig Assembly Program

Author - Xiaoqiu Huang
Department of Computer Science
Michigan Technological University
Houghton, MI 49931

E-mail: huang@cs.mtu.edu
Minor modifications for I/O by S.Smith

Reference -
"A Contig Assembly Program Based on Sensitive Detection of Fragment Overlaps" (submitted toGenomics, 1991)

Parameters:
Minimum overlap
Number of bases required for overlap
Percent match within overlap
Percentage match required in the overlap

region before merge is alowwed.

Comments:
CAP returns the aligned sequences to the current editor window. The sequences are placed into contigs by setting the groupid. Cap does not change the order of the sequences, and so the results should be sorted by group and offset (see sort under the Edit menu).

Lsadt - Least squares additive tree analysis

Author: Geert De Soete, 'C' implementation by Mike Maciukenas University of Illinois
Reference:LSADT, 1983 Psychometrika, 1984 Quality and Quantity

Parameters:
Distance correction to use in distance matrix calculations (see count below).
What should be used for initial parameters estimates
Random number seed
Display method (See TreeTool below)

Comments:
The program has been rewritten in'C' and will be included with the rRNA Database phylogenetic package being written at the University of Illinois Department of Microbiology.

Count is a short program to calculate a distance matrix from a sequence alignment (see below).

Count - Distance matrix calculator

Author: Steven Smith

Parameters:
Correction method
Currently Jukes-Cantor or none
Include dashed columns
Match upper case to lower

Comments:
Passes back a distance matrix in a format readable by LSADT.

Treetool - Tree drawing/manipulation

Author: Michael Maciukenas, University of Illinois

Comments:
See included documentation for TreeTool usage.

Readseq - format conversion program

Author: Don Gilbert

Parameters: Many, but can easily be run in interactive mdoe.

Comments:
Readseq is a very useful program for format conversion. The latest version supports over a dozen different file formats, as well as formating capabilities for publication. GDE makes of Readseq for importing and exporting seqeuences as well as a filtering tool to some external functions.

Lsadt - Least squares additive tree analysis

Author: Geert De Soete, 'C' implementation by Mike Maciukenas University of Illinois

Reference:LSADT, 1983 Psychometrika, 1984 Quality and Quantity

Parameters:
Distance correction to use in distance matrix calculations (see count below).
What should be used for initial parameters estimates
Random number seed
Display method (See TreeTool below)

Comments:
The program has been rewritten in'C' and will be included with the rRNA Database phylogenetic package being written at the University of Illinois Department of Microbiology.

Count is a short program to calculate a distance matrix from a sequence alignment (see below).

Count - Distance matrix calculator

Author: Steven Smith

Parameters:
Correction method
Currently Jukes-Cantor or none
Include dashed columns
Match upper case to lower

Comments:
Passes back a distance matrix in a format readable by LSADT.

Copyright Notice

::
gde_flat_ift.hlp
::
TITLE NOTES: gde_flat

OCCURRENCE ARB_IMPORT

DESCRIPTION A very simple format. Looks like

```
#name_of_species
acgtcgacgactgactgacacgacg
acgaacgacggacgc
#next_name
acgacgugactgacugacgcguagcuacgagcacucg
```

NOTES Simple and fool prove format
Autodetection may fail

::
gde_ift.hlp
::
TITLE NOTES: gde

OCCURRENCE ARB_IMPORT

DESCRIPTION This is just an experimental GDE file reader. It only reads the name and sequence. You should not use it if you can avoiD it.

::
gene_extract.hlp
::
TITLE Extract genes to gene-species

OCCURRENCE ARB_GENE_MAP/Genes/Extract marked

DESCRIPTION For every marked gene this command creates a so called 'gene-species'.

Just enter the name of the alignment you like to create or to add species to.

::
gene_hide.hlp
::
TITLE Hiding genes

OCCURRENCE ARB_NT/ARB_GENE_MAP/Hide

DESCRIPTION You can hide genes.
Hidden genes will not be displayed in ARB_GENE_MAP.

::
gene_info.hlp
::
TITLE GENE INFORMATION

OCCURRENCE ARB_GENEMAP/<INFO> button (on the left side)
ARB_NT/Genes/Info
ARB_NT/Genes/Search
ARB_GENEMAP/Genes/Info
ARB_GENEMAP/Genes/Search

DESCRIPTION Displays gene information stored within the 'fields' (see 'HELP: Glossary'). The particular 'gene' (see 'HELP: Glossary') can be 'marked' or 'unmarked' (see 'HELP: Glossary') by pressing the checkbox after the 'Marked?' prompt.
Editing of 'field' entries is enabled or prevented by pressing the 'Edit enabled?' Checkbox.
The entries of a 'field' are modified by choosing it from the 'DATABASE FIELDS' subwindow and modifying the entries displayed in the 'Edit box' subwindow.

NOTES The 'SEARCH' window can be displayed by pressing the <SEARCH> button.

For modification of 'field' entries, a protection level

has to be selected from the Protection menu of the main window (ARB_NT/Protection) equal to or higher than that assigned to the selected 'field'

Cut and paste of the window system can be used in the 'Edit box' subwindow. This provides is an easy way to export/import data.

STANDARD FIELDS The following fields have a special meaning in ARB:

name unique name for the gene/annotation
pos_begin start position of the gene
pos_end end position of the gene
Note: pos_begin <= pos_end !!!
pos_uncertain if either pos_begin or pos_end is uncertain this field contains a string build from <,>,=
The first character of the field corresponds to pos_begin, the second character to pos_end.
< means the position may be lower
> ----------"---------- higher
= means: the position is certain
pos_joined If this field exists, it contains a number X>1
In this case the gene is joined from X parts and it contains further fields pos_begin2, pos_begin3, ... (up to pos_beginX) containing the start positions of the other parts. The same applies to the fields pos_end and pos_uncertain.
complement if complement is 1, the coding direction of the gene is from right to left.

WARNINGS It is recommended to reset the protection level after modifying entries to prevent unintentional modification or loss of data.

::
gene_map.hlp
::
TITLE ARB_GENEMAP - Gene map editor

OCCURRENCE ARB_NTREE/Genes/Gene map

DESCRIPTION ARB_GENEMAP displays a map of all genes of the selected organism (species).

::
gene_mark.hlp
::
TITLE Mark genes

OCCURRENCE ARB_NT/Genes/
ARB_NT/ARB_GENE_MAP/Genes/

DESCRIPTION Marking genes works similar to marking species.

You can select to
- mark all
- unmark all
- invert all marked

genes of
- the current
- all marked
- all

species.

::
gene_mode.hlp
::

TITLE ARB_GENEMAP modes

OCCURRENCE ARB_GENEMAP/mode buttons

DESCRIPTION ZOOM MODE Mode to zoom the displayed gene map.

You can zoom in by dragging a rectangle with
the left mouse button. You can zoom out with
the right mouse button.

INFO MODE Mode to get info about a specific gene.

Click on the gene to get info about.

::
gene_options.hlp
::

TITLE Standard help file form

OCCURRENCE ARB_GENE_MAP/Properties/Options

DESCRIPTION None yet

BUGS Not working yet.

::
gene_pseudo.hlp
::

TITLE Pseudo gene-species

DESCRIPTION In order to work with alignments of genes you first have to
extract the wanted genes into some kind of pseudo-species
called 'gene-species'.

These gene-species store additional information about the
organism and the gene they originated from.

They only contain that part of the whole organism-sequence
the gene corresponds to.

::
gene_search.hlp
::
TITLE Search Database for Genes

OCCURENCE ARB_NT/Genes/search
ARB_GENEMAP/Genes/search

DESCRIPTION Searches for a (set of) genes
that match (dont match) a query or are marked.

The database is scanned for 'genes' which contain (or do not contain) the search string within the specified 'field'.
The corresponding genes and the respective 'field' entries are listed in the 'HIT LIST' subwindow. The number of hits is displayed after the 'Hits:' prompt.

Define whether matching or non-matching genes should be listed by pressing the appropriate combination of buttons in the top area.

Define whether genes are searched in the current, in all marked or in all species.

When performing multiple searches, define whether the list of 'genes' should be replaced by the new results, and whether newly found 'genes' should be removed from or appended to the existing list.

Select a 'field' from the 'Search Field' subwindow.

Type the search string in the 'Search string' subwindow.

Press the <SEARCH> button.

WARNINGS If the hitlist becomes too long, it will be truncated.

::
genes.hlp
::
TITLE What are Genes ?

DESCRIPTION One gene is one database entry conaining many database entries describing the gene. Every gene belongs to a specific species, but one species may have many genes.

[Database field name] [Comment]

name Unique name for a species
max 8 characters long

pos_begin The start position in the
main-sequence of the species.

pos_end The end position in the main-sequence of the species.

NOTE The main-sequence of a organism(species) is stored in the alignment 'ali_genom'. You should not change this alignment.

WARNINGS The alignment 'ali_genom' should NOT contain any gaps!!!!

::
glossary.hlp
::
TITLE Glossary

ALIGNMENT Subdatabase containing homologous sequences (different 'alignments' can be stored along with the species information)

ACI ARB Command Interpreter. ACI is a simple programing language for calculating special species information (eg. G+C Content..)

EDGE Branch within a tree defined by nodes

FIELD Container for 'species' associated information

HELIX

MARKED SPECIES Marking of 'species' defines the datasubset which is analyzed by the ARB tools.

NAMES Identifier for 'species' entries. Can be automatically generated

NDS Node Display Setup: defines informatio which is displayed at tree nodes.

NODE Internal: connects branches of a tree
Terminal: free end of branches representing 'species'

@@@ PROTECTION LEVEL

PT_SERVER Server using a special form of the ARB database for rapid similarity searching (automated alignment, probe design/check)

SAI Sequence Associated Information

SELECTED SPECIES 'species' highlighted (selected) using the 'ARB_NT/Species/ Search' or 'ARB_INTRO <MERGE TWO ARB DATABASES> tools.

SPECIES Database entry containing a sequence and associated information. Not necessarily consistent with a taxonomic species

SRT Search and Replace Tool allows to search substrings in a string and replace them by another substring.

::
helix.hlp

::

TITLE Helix

OCCURRENCE ARB_EDIT/ETC/Reload Helix

DESCRIPTION The 'SAI' (see 'HELP: Glossary') entry 'HELIX_NR' contains numbering of potential higher order structure elements. The paired 5' and 3' helix halves are indicated by identical numbers. The SAI entry 'HELIX' contains symbols ([<(and)>]) indicating positions which usually are base paired. This information is used by the editor to check the sequences for (correct) base pairing. User-defined symbols indicating base pairing and mispairing are shown below the particular sequences.

FORMAT OF HELIX/HELIX_NR:

When the 'HELIX/HELIX_NR SAI's are analyzed the following steps are performed:

1. HELIX_NR contains the helix numbering. A helix number is a number between 0 and 100 optionally followed by one letter:
   e.g. 11a, 23g, 0i, 6, 23,
   A helix number is valid from the position of the first digit to the position just before the next helix number.

```
eg.   .........1a.........23b.........
      .........||||||||||............
```

   The '|' symbols indicate the positions for '1a'

2. All '<[{(' symbols in HELIX are replaced by '('
   All '>]})' symbols in HELIX are replaced by ')'
   eg. ...[<<[..]>>].. is the same as ...((((..))))..

3. For each helix number 'HN' a temporary helix is created that contains all '()' symbols of the 'SAI HELIX' at all positions where 'HN' is valid.

```
eg. HELIX_NR   .....1a.....23b.......1a....23b.....
      HELIX    .....[<<[...[<[.......]>>]..]>].....
      -> 1a    .....((((.............)))).........
      -> 23b   ............(((.............))).....
```

4. The (usually) paired positions can simply be defined by recursively removing the innermost bracket pairs in the temporary helices:

```
eg. HELIX_NR   .....1a.....23b.......1a....23b.....
      HELIX    .....[<<[...[<[.......]>>]..]>].....
      Pos      123456789012345678901234567890123456
      -> 1a    .....((((.............)))).........
      -> 23b   ............(((.............))).....
      1a 9-23  ........+-------------+............
```

1a 8-24 ....+---------------+...........
1a 7-25 ......+-----------------+..........
1a 6-26 .....+-------------------+.........
23b 15-29 ....+------------+.......
23b 14-30 ....+--------------+......
23b 13-31 ....+----------------+.....

NOTES After inserting or deleting gaps in the whole alignment (ARB_NT/ Sequence/Admin/INS/DEL CHAR) the helix information has to be:
1. Checked (inserting a '.' between '1' and '2' destroys the original '12').
2. reloaded by choosing the 'Reload Helix' item of the 'ARB_EDIT/ETC' menu.

::
helixsym.hlp
::
TITLE Define Helix Symbols

OCCURRENCE ARB_NT/ARB_EDIT/Props/Helix Symbols

DESCRIPTION @@@@@ Eigentliche Hilfe sollte hier stehen
Paerchen durch ' ' getrennt
Reihenfolge egal
Gross / Kleinschrift egal

Falls Paerchen auftritt -> Symbol auf der rechten Seite
Falls nicht gefunden -> Default
Non_Standard 0-9

Different base pairs can be indicated by user-defined symbols.

Press the 'Helix Symbols' button of the 'ARB_NT/ARB_EDIT/Props' menu to display the 'HELIX_PROPERTIES' window. Type the base pairs in the large subwindows and the corresponding symbols in the small subwindows.

NOTES Upper and lower case letters are not differentiated.

::
help.hlp
::
TITLE Using online help

DESCRIPTION

Format:

"text" indicates an important term.

'text' indicates ARB-specific terms, paths, tools and prompts.

<text> indicates buttons.

Terminology:

Screen:

"Windows" are the workstation work areas. The names are displayed at the very top of the window.

"Subwindows" (indented) are display and input areas within windows.

"Menu bars" at the top of windows (if present) are used to expose menu choises. The menus are displayed after clicking on the prompts of the menu bar.

"Menu buttons" are raised rectangular buttons containing a small raised rectangle at the right. Menu options are exposed after clicking on the button.

"Menus" initiate actions after clicking on the items.

"Scroll bars" at the exterme right and bottom of windows and subwindows allow to move through the display area in various increments.

Mouse buttons communicate commands to the program:

"Left mouse button" is used exclusively for clicking on prompts, checkboxes and buttoms.
!!! Exception: current version of 'ARB_PHYL' (see 'HELP: Matrices, Masks, Profiles V1.0.!!!

"Middle and right mouse buttons" are used for modifying the tree display (see "HELP: MODES')

"Press" means hold down the mous button while completeing an operation.

"Release" means to let go of the mouse button.

"Click on" means positioning the cursor on a location, prompt, button or checkbox, pressing and immediately releasing the mouse button.

"Choose" from a menu means clicking on the desired menu option.

OLINE HELP is available by clicking on the 'HELP' prompts (flat) or buttons (raised) located in the upper right of the windows.

Clicking on the <HELP> button displays the 'HELP WINDOW' containing online help.

Clicking on the 'HELP' prompt displays the 'HELP' menu. Choose

the 'Press this menu and the button you want help for'
to switch to the online help mode (indicated by the
'open question cursor'). Then click on any button
or prompt to display the 'HELP WINDOW' containing
online help.

There are two formates of online 'HELP' texts:

Normally, ARB uses simple ASCII texts displayed within the
'HELP WINDOWS'.
The titles of help texts on related main- and subtopics
linked to the current help file are displayed in the
subwindows in the lower part of the 'HELP WINDOW' and
are activated by clicking on the tile prompts.

Press the <BACK> button in the upper part of the 'HELP
WINDOW' to display the former help text.

Some 'HELP' texts are stored as postscript files and displayed
using the public domain software 'ghostview'.

Press <Ghostview/Page/Next> to display the folowing
page(s), <Ghostview/File/Print> to print the
help text. <Ghostview/File/Quit> to leave
ghostview.

NOTES The 'HELP' texts for foreign programs or tools which have been
incorporated into ARB (see 'HELP: ARB') are usually (parts of)
the original documentation or help texts.

The ASCII texts can be edited by clicking on the <EDIT> button
of the 'HELP WINDOW' and can be modified by the user. Write
permission is needed for '$ARBHOME/lib/help'.

BUGS In the current release, the 'open question pointer' is reset to
the original symbol when moved to a menu item.

::
importift.hlp
::
TITLE How to define new import formats

OCCURRENCE ARB_NT

BRIEF DES. All description files of import formats are located in
the directory '$ARBHOME/lib/import'.
Each of these files describe how to analyze the input files.
A basic import description file (.ift) looks like this:

```
[AUTODETECT     "Matchpattern"]
BEGIN           "Matchpattern"
[KEYWIDTH       #Columnnumber]
[MATCH          "Matchpattern"
        [SRT    "SRT_STRING"]
```

```
        [ACI    "ACI_STRING"]
        [WRTIE  "DB_FIELD_NAME"]
        [APPEND "DB_FIELD_NAME"]]*
    SEQUENCESTART   "Matchpattern"
    SEQUENCECOLUMN  #Columnnumber
    [SEQUENCESRT    "SRT_STRING"]
    [SEQUENCEACI    "ACI_STRING"]
    SEQUENCEEND     "STRING"
    [CREATE_ACC_FROM_SEQUENCE]
    [DONT_GEN_NAMES]
    END             "STRING"

    or it can pipe the data through any external program PROGRAM to
    convert it to an already existing format 'exformat'
    using the following basic design:

    [AUTODETECT     "Matchpattern"]
    SYSTEM          "PROGRAM $< $>"
    NEW_FORMAT      "lib/import/exformat.ift"

    $< will be replaced by the input file name
    $> will ve replaced by the intermediate file name

DESCRIPTION     First of all the converter appends all import files maching
        the filepattern into one file. The files are seperated by the
        string defined with the keyword  SEQUENCEEND.
        1.   Search  the first line matching the pattern defined
             by BEGIN
        2.   Try to match all MATCH_patterns.
             for all lines that match do:
             2.1 append all following lines, which start after
                 column KEYWIDTH;
             2.2 if an SRT_command is defined, start the
                 string replace tool.
             2.3 if an ACI_command is set, run the Arb
                 command interpreter
             2.4 Write the result into the database using the
                 fieldname 'WRTIE' or append it to a database
                 field with the name 'APPEND'
        3.   If the line matches SEQUENCESTART_pattern, assume that
             all following lines to and except the line
             matching SEQUENCEEND_pattern.
        4. GOTO 1

        Postprocesses:
        CREATE_ACC_FROM_SEQUENCE:       Generate a checksum for
                        all sequences with no accession
                        entry ('acc' -field) and write
                        it as the accession number;
        DONT_GEN_NAMES:                 Do not try to generate
                        unique short names for the species
                        using the full_name field.

COMMANDS
```

EXAMPLES Look at the files in '$ARBHOME/lib/import'

WARNINGS Format detection does not always work

::
insdelchar.hlp
::
TITLE Insert / Delete Column

OCCURRENCE ARB_NT/Sequence/Admin/INS/DEL CHAR

DESCRIPTION Inserts or deletes columns in an alignment.

Select alignment position by typing it in the 'Sequence Position' subwindow or by putting the cursor at the desired position in the editor.

Type the number of columns to insert or delete from the defined position towards the (3'-) ends of the sequences.

Specify the characters to be deleted by typing in the 'Delete Only' window

NOTES @@@ Beachte SAIS

EXAMPLES @@@ Delete Example mit Fehlern

BUGS @@@ Enden im Editor u.U. nicht sichtbar

::
iupac-codes.hlp
::

TITLE Primer Design/IUPAC codes

OCCURRENCE ARB_NT/ETC/Primer Design

DESCRIPTION

```
R = A  G
M = A C
S =   CG
Y =  TC
K =  T G
W = AT
V = A CG
B =  TCG
D = AT G
H = ATC
N = ATCG
```

NOTES written by Wolfram Förster 2001

::

join_species.hlp

TITLE Standard help file form

OCCURRENCE ARB_NT/Species/Destroy Species/Join Marked Species

DESCRIPTION This function allows to join to set of species.
For example you have a 23s and a 16s database and you want
to append the 16s to the 23s sequences than you have to do:
- merge the 16s to the 23s sequences.
- create a field 'species_name' which holds the
real name of the species.
That means the species_name of the 16s sequence
should be the same as of the 23s seq.
Do not use the 'names' field because it is
used as a UNIQUE id to the database.
- ARB_NT/Species/Destroy Species/Join Marked Species
and use the field 'species_name'.

BUGS Time will show

macro.hlp

TITLE Macros

OCCURRENCE ARB_NT

DESCRIPTION Macros are used to combine a set of menu-actions. They work like
a tape recorder, which records all buttons presses.

To execute an existing macro select a macro and press <EXECUTE>
To record a new macro, go to a directory where the new macro
should be placed, enter a macro name and press <RECORD>. The
button label will switch to <STOP>. Do all actions, then press
<STOP>.

NOTES Read the BUGS
Macros may be edited using a normal text editor.
Macros may call submacros
Create only small macros, not more than 5 actions, use
submacros instead.

WARNINGS Close all subwindows first.

BUGS - Only buttons, values and menus are recorded
- When recording a macro you have to open a subwindow
before you can press the buttons within.
Hint: Close all subwindows before you start to build a
macro.
- Only value changes are recorded. If you open a subwindow
and just press the go button, all user changeable
parameters are unchanged. If you want a parameter set
to a certain value, you have to change it. Sometimes it

is necessary to change it to any value and than back to
the real value, just to ensure that it is recorded in
the macros.
- Only actions within the main application are recorded.

::
mark.hlp
::
TITLE What are Marked Species ?

DESCRIPTION All species have a flag which can be set to FALSE or TRUE.
All species with the flag set to TRUE are called marked
species. Normally this flag is used to specify a subgroup of
species for editors, tree building programs ...

EXAMPLES Mark one known species:
1. Search for the species: <ARB_NT/Species/Search>
2. Select the species
3. Mark it
Mark a list of species
1. Search for the species: <ARB_NT/species/search>
2. Press <Mark Listed/Unmark Rest>
Mark all species in selected tree:
1. Unmark all species: <ARB_NT/species/unmark all sp>
2. <ARB_NT/Species/Mark Species in Tree>
Mark a subtree:
1. Select the <MARK MODE>
2. Press the left mouse button at a subtree.

2. Press the left mouse button at a subtree.

::
mark_genes.hlp
::
TITLE What are Marked Genes ?

DESCRIPTION All genes have a flag which can be set to FALSE or TRUE.
All genes with the flag set to TRUE are called marked
species. Normally this flag is used to specify a subgroup of
genes and then perform some actions on these genes.

EXAMPLES Mark one known gene:
1. Search for the Gene: <ARB_NT/Genes/Search>
2. Select the gene
3. Mark it
Mark a list of genes:
1. Search for the gene: <ARB_NT/Genes/search>
2. Press <Mark Listed/Unmark Rest>

WARNINGS Be careful: The marked genes can be spread over several
species w/o showing this in the ARB_GENEMAP window.

::

mark_list.hlp

TITLE MARK LISTED SPECIES/GENES

OCCURRENCE ARB_NT/Species/Search: MARK LISTED UNMARK REST
ARB_NT/Genes/Search: MARK LISTED UNMARK REST

DESCRIPTION Marks all listed species/genes,
unmarks all species which are not listed

WARNINGS Species list may be truncated

BUGS none

mark_long_branches.hlp

TITLE Mark long branches

OCCURRENCE ARB_NT

DESCRIPTION Badly aligned sequences often result in long branches in the tree. Being able to identify those branches quickly helps to find those sequences.
The programs allows to mark all species having an extra long branch compared to it's nearest neighbour.

NOTES Play around and try different numbers.

BUGS The program does not show all extra long branches because then the output won't be user readable. So after fixing the alignment of sequences and recalculating the branch lengths, there might be new sequences with extra long branches.

max_freq.hlp

TITLE Calculate the Percentage of the Most Frequent Base

OCCURRENCE ARB_NT/SAI/create SAI/Max Frequency

DESCRIPTION Finds the most frequent base in each column for all marked species. Than the number of all sequences with this base are devided by:
a: the number of all marked sequences, if not ignoring gaps
b: the number of bases in this column, if ignoring gaps
The resulting percentage is divided by ten and the second last digit taken:

0% - 25% will never occur

25% - 29% -> '2'
30% - 39% -> '3'
...
90% - 99% -> '9'
100% -> '0'

NOTE The result can be used as a conservation profile and filter:
W.L.'s rule:
The higher the number the more conserved the position.

EXAMPLES Say one column contains 7 A's 4 G's and 5 Gaps.
Ignoring Gaps will result in 7/11 == 64 %
which is converted to '6'.
Otherwise we get 7/16 == 44% which will be indecated by a
'4' in the target sequence.

NOTE '-' @#@ are regarded as gaps.
'.' are always ignored.

::
mg_alignment.hlp
::
TITLE Check and modify the consistency of two alignments

OCCURRENCE ARB_INTRO/MERGE TWO DATABASES <Check Consistency of Alignments>
button

DESCRIPTION Checks the types of data stored within 'alignments' (see 'HELP:
Glossary') for consistency to prevent merging of non homologous
data such as 16S and 23S rRNA or protein and DNA sequence data.

Press the 'Check Consistency of Alignments' button to display
the 'MERGE ALIGNMENTS' window

To save modifications pres the 'Save result' button of the
'ARB_MERGE' window.

NOTES If any alignment exists only in database I then the alignment
will be created in database II (but no sequences will be
copied).

If the alignment exists in both databases then the sequence_type
will be compared

If any inconsistencies are detected the user has to resolve them
by pressing the 'MODIFY' button to display the 'ALIGNMENT
CONTROL' window. This tool allows to modify the alignments.

::
mg_extendeds.hlp
::
TITLE Transfer SAI Entries

OCCURRENCE ARB_INTRO <MERGE TWO ARB DATABASES> ARB_MERGE <Transfer SAIs>

DESCRIPTION Transferres 'SAI' entries (see 'HELP: Glossary') from database I to database II.

Press the 'Transfer SAIs' button of the 'ARB_MERGE' window to display the 'MERGE SAIS' window.

Select source 'SAI' from the left (database I) 'SAI' subwindow and the destination from the right (database II) 'SAI subwindow.

Press the 'Transfer SAI' button between the 'SAI' subwindows.

To delete or rename 'SAI's in the individual databases press the corresponding buttons below the 'SAI' subwindows.

To save modifications pres the 'Save result' button of the 'ARB_MERGE' window.

NOTES The names of source and destination 'SAI' may be different.

If an 'SAI' should be transferred which does not exist within database II and no other destination 'SAI' has been specified the 'SAI' will be created in database II.

::
mg_names.hlp
::
TITLE Check and Update Names

DESCRIPTION Merging two databases species names assigned to the corresponding accession numbers and other information have to be identical in both databases.

To enshure this, press the 'RENAME DATABASE' buttons to display the 'AUTORENAME SPECIES' window.

To save modifications pres the 'Save result' button of the 'ARB_MERGE' window.

NOTES Creates automatically unique names (=identifiers) for the species entries in the database. The entries are identified by their accession numbers (public databases). The names are given using the 'full_name' information. Usually, the first three letters are taken from the genus designation, the remaining letters from the species name.

If there are duplicated entries (same accession number - different 'full_name'; no accession number - same 'full_name') the different versions are indicated by appending running numbers separated from the 'name' by a dot.

WARNINGS If any species has lost is accession number it will not be renamed correctly.

::
mg_spec_sel_field.hlp
::
TITLE Transfer one field of selected species

OCCURRENCE ARB_MERGE/Transfer Species/Transfer Field ...

DESCRIPTION Copies just one field from all listed species (left side) to the corresponding species (== equal 'name') of database II. If the corresponding species does not exist then it will be created.
There are two modes:

1. Copy Mode:
   Delete all fields of all corresponding species before copying data.

2. Append Mode (Strings Only)
   Append all fields of listed species in db I to corresponding field and species of db II

NOTES Transfering fields that are not of string type in append mode does not work

::
mg_species.hlp
::
TITLE Compare and Transfer Species Entries

OCCURRENCE ARB_INTRO <MERGE TWO ARB DATABASES> ARB_MERGE <Transfer Species>

DESCRIPTION Allows # database searching,
comparison of the two databases,
transfering data from left to right
realigning sequences to new alignment

Database Searching:

To perform database searching within the individual databases use the left or right part of the 'TRANSFER SPECIES' window for database I (source) and database II (destination), respectively.

The database is scanned for 'species' (see 'HELP: Glossary') which contain (or do not contain) the search string within the specified 'field' (see 'HELP: Glossary'). The corresponding 'species' and the respective 'field' entries are listed in the 'HIT LIST' subwindow. The number of hits is displayed after the 'Hits:' prompt.

Define whether matching or non matching species should be listed by pressing the appropriate combination of left and right buttons in the 'QUERY TYPE' area. Performing multiple searches, define whether the list of 'species' should be replaced by the new results, whether newly found 'species' should be removed from or appended to the existing list.

Select a 'field' from the 'Fields' subwindow.

Type the search string to the 'Search string' subwindow.

Press the 'RUN QUERY' button of the results area.

Data Transfer:

To transfer a single 'species' and all its 'field' entries, select the species from the list displayed in the 'HIT LIST' subwindow and press The 'TRANSFER SELECTED SPECIES' button in the middle column of the 'TRANSFER SPECIES' window.

To transfer all listed 'species' and all corresponding 'field' entries, press the 'TRANSFER ALL LISTED SPECIES' button in the middle column of the 'TRANSFER SPECIES' window.

To transfer the entries of a 'field' selected from the 'Search field' subwindow for all listed 'species, press the 'TRANSFER A FIELD FOR ALL LISTED SPECIES' button in the middle column of the 'TRANSFER SPECIES' window.

To save modifications pres the 'Save result' button of the 'ARB_MERGE' window.

Preserve Alignment:

ARB Merge tries to keep the alignment correct. Normally people have inserted new gaps in either the left or right database. By entering the name of some reference species in the input box in the upper center of the TRANSFER SPECIES window, the program will try to find those species in both databases, create a coloumn reference list, and realign all transfered sequences. If you do not want this feature, turn off the 'Preserve Alignment' togge.

EXAMPLES For examples for database searching see 'HELP: Search Database for Species'.

WARNINGS The data will always be transferred from 'Database I' to 'Database II'.
If you want to align sequences during transfer, it is recommended that the left database has fewer gaps than the right one.

!!! If there are entries (for selected data) in both databases,

those in 'Database II' will be overwritten.!!!

::
mg_trees.hlp
::
TITLE Transfer Trees

OCCURRENCE ARBB_INTRO <MERGE TWO ARB DATABASES> ARB_MERGE <Transfer Trees>

DESCRIPTION Transferres trees.

Press the 'Transfer Trees' button of the 'ARB_MERGE' window to display the 'MERGE TREES' window.

Select source tree from the left (database I) 'Trees' subwindow and the destination from the right (database II) 'Trees' subwindow.

Press the 'Transfer Tree' button between the 'Tree' subwindows.

To delete or rename trees in the individual databases press the corresponding buttons below the 'Tree' subwindows.

To save modifications pres the 'Save result' button of the 'ARB_MERGE' window.

NOTES If a 'Tree' should be transferred which does not exist within database II and no other destination tree has been specified the tree will be created in database II.

::
mod_field_list.hlp
::
TITLE MODIFY FIELDS OF LISTED SPECIES/GENES

OCCURRENCE ARB_NT/Species/Search/More Functions/Modify fields of listed
ARB_NT/Genes/Search/More Functions/Modify fields of listed
ARB_NT/Tree/NDS

DESCRIPTION Finds and replaces substrings within fields/tagged subfields of all listed species/genes. The entries within the selected fields of all listed species/genes can be modified either individually or globally.

Three different languages can be used to modify an entry:

SRT: indicated by a leading ':' character
ACI: indicated by a leading '|' character
REG: indicated by sourrounding '/' characters

Details:

REG: Simple Regular Expressions (not for beginners)

'/Seach RegExpr/Replace String/' or
'/Seach RegExpr/'
(see REG-help text for more details)

SRT: Replaces substrings
Syntax: ':old_string=new_string'
see SRT help text for more details
example: remove all spaces -> SRT ': ='

Different search/replace commands can be performed
simultanously and have to be seperated by ':'
':search1=replace1:search2=replace2: ... :searchn=replacen'.

* and ? are wild cards for multiple and single
characters, respectively.

ACI: More sophisticated string manipulations
(Read help text for more information)

NOTES You may add new commands by editing one of the files:
$ARBHOME/lib/sellists/mod_fields.sellst
$ARBHOME/lib/sellists/mod_gene_fields.sellst
You should save this file to another location when
installing new versions of ARB

EXAMPLES ':p?r=p?1w' replaces par to paw
pbr to pbw
pcr to pcw ...
':p??r=p?2?1r' swaps the two letters between p and r

':a*=b*1' replaces only the first 'a' by 'b'
':?* *=?1. *2' Replaces the first word by its first
letter + '.'
':\:=\n' replaces all ':' by <newline>
':*=*1 *(key1)' appends the database field <key1>
':*=*1 *(key1|nothing found)'
appends the database field <key1>
if <key1> does not contain entries
append 'nothing found'

1. Global modification: Replace 'spec.' by 'sp.' within
the field full_name of all listed species:

Press: 'MODIFY FIELDS OD LISTED SPECIES'

Select Field: 'full_name'
Type Command: ':spec.=sp.'
Press: 'GO'

2. Individual modification: Append the particular entries
of fields 'title' and 'journal' to that of the
fields 'author' of all listed species if there
are any entries:

Press: 'MODIFY FIELDS OD LISTED SPECIES'

Select Field: 'author'
Type Command: ':*=*1 *(title) *(journal)'
Press: 'GO'

NOTE Undo does work.

WARNINGS Be careful if search or replace string contain special characters (such as ':').

::
mode.hlp
::
TITLE MODES

OCCURRENCE ARB_NT <buttons in the first column>

DESCRIPTION Pressing one of the buttons in the left column of the 'ARB_NT' window activates tree display and modification functions.

NOTES Short descriptions of the 'MODEs' and the functions assigned to the mouse buttons are given in the fourth line of the 'ARB_NT' window. To execute the respective functions, the cursor has to be located within the tree display area while pressing the mouse buttons.
You may also use the function keys to select a mode. F1 selects the topmost mode, F2 the next and so on.

::
mode_angle.hlp
::
TITLE ANGLE MODE [***Although I like the sound of "angel mode"***]

OCCURRENCE ARB_NT < button in the left column>

DESCRIPTION The 'ANGLE MODE' allows you to increase or decrease the angles of adjacent terminal branches or all angles within subtrees.

Click on the eighth button in the left column of the 'ARB_NT' window to activate the 'ANGLE MODE'.

Change the angle of terminal branches:

Move the cursor to nay position along the edge connecting the two terminal branches and the rest of the tree. Click on the left or right mouse button repeatedly to gradually increase or decrease the angle.

Change all angles of a subtree:

Move the cursor to any position along the edge defining the subtree and proceed as described above.

The subtree-defining branch is that located closest to the indicated root (open square) of the whole tree which is directly or indirectly connected to all branches of the subtree.

NOTES Short descriptions of the 'MODE' and the functions assigned to the mouse buttons are given in the fourth line of the 'ARB_NT' window.

The displayed tree topology is exported to the drawing tool xfig.

The middle mouse button (placed anywhere within the tree display area) can be used to move the whole tree.

The scale bar and its label can be moved by positioning the cursor anywhere on the bar, keeping the left mouse button pressed, moving the cursor to the desired position and subsequently releasing the mouse button. The label of the bar can be moved independently.

The previous angle(s) can be restored using the 'RESET MODE' ('RESET' button on the 'ARB_NT' window).

::
mode_group.hlp
::
TITLE GROUP MODE

OCCURRENCE ARB_NT (<GROUP> button of the left column)
F3 on the keyboard

DESCRIPTION A group is a subtree with a name assigned to its root node. The 'GROUP MODE' allows display of subtrees (groups of species) as triangles (radial tree) or rectangles (dendrogram). You also can convert a subtree into a group.

Click on the <GROUP> button of the 'ARB_NT' window or the 'F3' key on the keyboard to activate the 'GROUP MODE'. In the latter case, the cursor has to be placed within the tree display area.

Group species:
Move the cursor to that internal node of the displayed tree (section) which defines the subtree and click on the left mouse button. The subtree-defining node is that located closest to the indicated root (open square) root of the whole tree which directly or indirectly connects all branches of the subtree.

Ungroup species:
Press the left mouse button while the cursor is over the triangle of a group.

Transform a subtree into a group and assign a name to it:
Proceed as described above using the right mouse button.

NOTES To store the groupings, assign any information ('SAI': name, fullname, ...) to them using the 'SPECIES INFORMATION' tool.

The middle mouse button (placed anywhere within the tree display area) can be used to move the whole tree.

::
mode_info.hlp
::
TITLE INFO MODE

OCCURRENCE ARB_NT <INFO> button in the left column
F6 on the keyboard

DESCRIPTION The 'INFO MODE' allows you to show or modify information assigned to terminal or
internal nodes (i.e. species or groups of species).

Press the <INFO> button of the 'ARB_NT' window or the 'F6' key on the keyboard to activate the 'INFO MODE'. In the latter case, the cursor has to be placed within the tree display area. The 'SPECIES INFORMATION' window is displayed.

Select a species or an existing group:

Move the cusor to the respective node and press the left mouse button. The corresponding information is displayed within the 'SPECIES INFORMATION' window.

Create and select a group:

Move the cusor to any corner or edge of the triangle (radial tree) or rectangle (dendrogram) representing the desired group and press the right mouse button. The corresponding information is displayed within the 'SPECIES INFORMATION' window.

NOTES The selected node is indicated by an open square.

::
mode_kernlin.hlp
::
TITLE K.L. Optimization of the Tree

OCCURRENCE ARB_NT/Tree/Parsimony/<K.L.>

DESCRIPTION Searches for a better (more parsimonious) tree by swapping

subtrees which are separated by up to 15 edges.

NOTES The current parsimony value (number of base changes needed) is displayed after the 'Current Par' prompt in the upper part of the 'ARB_PARSIMONY' window and the 'Message' window (pops up while calculating).

On cycle of swapping is performed.

::
mode_length.hlp
::
TITLE LENGTH MODE

OCCURRENCEE ARB_NT <button in the left column>

DESCRIPTION The 'LENGTH MODE' allows you to increase or reduce the length of terminal and internal branches.

Click on the tenth button in the left column of the 'ARB_NT' window to activate the 'LENGTH MODE'.

Change branch length:
The new branch length is defined by the internal node closest to the displayed root (open square) and the current position of the cursor!!!

Move the cursor to any position along the branch, keep the left mouse button pressed, move the cursor to the desired position, release the button.

NOTES Short descriptions of the 'MODE' and the functions assigned to the mouse buttons are given in the fourth line of the 'ARB_NT' window.

The displayed tree topology is exported to the drawing tool xfig.

The middle mouse button (placed anywhere within the tree display area) can be used to move the whole tree.

The scale bar and its label can be moved by positioning the cursor anywhere on the bar, keeping the left mouse button pressed, moving the cursor to the desired position and sub-sequently releasing the mouse button. The label of the bar can be moved independently.

The previous branch length can be restored using the 'RESET MODE' ('RESET' button on the 'ARB_NT' window).

EXAMPLE None

::
mode_lzoom.hlp

::

TITLE LOGICAL ZOOM MODE

OCCURRENCE ARB_NT <LZoom> button in the left column
F5 on the keyboard

DESCRIPTION The 'LOGICAL ZOOM MODE' allows you to display subtrees within the tree display area of the 'ARB_NT' window.

Click on the <LZoom> button of the 'ARB_NT' window or press the 'F5' key on the keyboard to activate the 'LOGICAL ZOOM MODE'. In the latter case, the cursor has to be placed within the tree display area.

Display a subtree:
Move the cursor to that internal node of the displayed tree (section) which defines the subtree and click on the left mouse button. The subtree-defining node is that located closest to the indicated root (open square) of the whole tree which directly or indirectly connects all branches of the subtree.

NOTES Short descriptions of the 'MODE' and the functions assigned to the mouse buttons are given in the fourth line of the 'ARB_NT' window.

To restore the initial scale of the display choose the 'Reset Logical Zoom' item from the 'ETC' menu of the 'ARB_NT' window.

The middle mouse button (placed anywhere within the tree display area) can be used to move the whole tree.

::

mode_mark.hlp

::

TITLE MARK MODE

OCCURRENCE ARB_NT <MARK> button (second button of the left column)>
F2 on the keyboard

DESCRIPTION The 'MARK MODE' allows you to mark and unmark individual species, or all species of a subtree.

Click on the <MARK> button (first button in the left column) of the 'ARB_NT' window or the 'F2' key on the keyboard to activate the 'MARK MODE'. In the latter case, the cursor has to be placed within the tree display area.

Mark single species:
Move the cursor to any terminal branch or the left part of the information displayed at the terminal node and press the left mouse button.

Mark all species of a subtree:

Move the cursor to any position along the edge defining the subtree and proceed as described above.

The subtree-defining branch is that located closest to the indicated root (open square) of the whole tree directly or indirectly connected to all branches of the subtree.

NOTES The marked species are indicated by filled squares at the terminal nodes.

Short descriptions of the 'MODE' and the functions assigned to the mouse buttons are given in the fourth line of the 'ARB_NT' window.

The middle mouse button (placed anywhere within the tree display area) can be used to move the whole tree.

::

mode_move.hlp

::

TITLE MOVE MODE

OCCURRENCE ARB_NT < button in the left column>

DESCRIPTION The 'MOVE MODE' allows to change the topology of the displayed tree by moving subtrees.

Click on the <MOVE> button in the left column of the 'ARB_NT' window.

Move subtree:

Move the cursor to any position along the edge defining the respective subtree, keep the left mouse button pressed while moving the cursor to the desired position on another edge, release the button to fix the position of the subtree.

The subtree defining branch is that located closest to the indicated root (open square) of the whole tree directly or indirectly connected to all branches of the subtree.

Move Group Info: @@@@@

Within the 'ARB_PARSIMONY' window, this mode can be used to test tree topologies with respect to maximum parsimony criteria by moving subtrees. The current parsimony value is shown after the 'Current Par' pormpt and can be compared with the 'Optimum Par' value of the iinitial tree.

NOTES Short descriptions of the 'MODE' and the functions assigned to the mouse buttons are given in the fourth line of the 'ARB_NT' window.

The displayed tree topology is exported to the drawing tool xfig.

The middle mouse button (placed anywhere within the tree display area) can be used to move the whole tree.

The scale bar and its label can be moved by positioning the cursor anywhere on the bar, keeping the left mouse button pressed, moving the cursor to the desired position and subsequently releasing the mouse button. The label of the bar can be moved independently.

WARNINGS !!! No 'undo' function available yet !!!

It is recommended to copy the tree before using the 'MOVE MODE'.

::

mode_nni.hlp

::

TITLE Nearest Neighbour Interchange

OCCURRENCE ARB_NT/Tree/Parsimony/<NNI>

DESCRIPTION Searches for a better (more parsimonious) tree by swapping the relative positions of subtrees separated by not more than one edge.

NOTES The current parsimony value (number of base changes needed) is displayed after the 'Current Par' prompt in the upper part of the 'ARB_PARSIMONY' window and the 'Message' window (pops up while calculating).

On cycle of swapping is performed.

::

mode_optimize.hlp

::

TITLE Nearest Neighbour Interchange and K.L. Optimization

OCCURRENCE ARB_NT/Tree/Parsimony/<NNI + K.L.>

DESCRIPTION Searches for a better (more parsimonious) tree by performing alternating cycles of 'Nearest Neighbour Interchange' and 'K.L.' optimizations.

NOTES The current parsimony value (number of base changes needed) is displayed after the 'Current Par' prompt in the upper part of the 'ARB_PARSIMONY' window and the 'Message' window (pops up while calculating).

::

mode_pzoom.hlp

::

TITLE ZOOM MODE

OCCURRENCE ARB_NT <PZoom> button in the left column
F4 on the keyboard

DESCRIPTION The 'ZOOM MODE' allows you to magnify selected square sections of the tree display area within the 'ARB_NT' window.

Click on the <PZoom> button of the 'ARB_NT' window or press the 'F4' key on the keyboard to activate the 'ZOOM MODE'. In the latter case, the cursor has to be placed within the tree display area.

Define and magnify a section:
Position the cursor to define the first corner of the square to magnify. Keep the left mouse button pressed and move the cursor to define size and positon of the region to magnify. Release the button.

Reduce the zoom scale:
To gradually reduce the zoom scale click on the right mouse button while the cursor is located within the tree display area of the 'ARB_NT' window.

NOTES Short descriptions of the 'MODE' and the functions assigned to the mouse buttons are given in the fourth line of the 'ARB_NT' window.

To restore the initial display choose the 'Reset Physical Zoom' item from the 'ETC' menu of the 'ARB_NT' window.

The middle mouse button (placed anywhere within the tree display area) can be used to move the whole tree.

::
mode_reset.hlp
::
TITLE RESET MODE

OCCURRENCE ARB_NT <RESET> button in the left column

DESCRIPTION The 'RESET MODE' allows to restore the previous tree topology after using the 'ROTATE', 'LENGTH', and 'MOVE MODE' pressing The left, middle, or right mouse buttons respectively.

Click on the 'RESET' button to activate the 'RESET MODE'.

NOTES Short descriptions of the 'MODE' and the functions assigned to the mouse buttons are given in the fourth line of the 'ARB_NT' window.

WARNINGS !!! Only the most recent previous topology can be restored!!!

::
mode_rotate.hlp

::

TITLE ROTATE MODE

OCCURRENCE ARB_NT <eighth button in the left column>
F9 on the keyboard

DESCRIPTION The 'ROTATE MODE' allows you to rotate branches or subtrees of the displayed tree.

Click on the <eighth button in the left column> of the 'ARB_NT' window or press the 'F9' key on the keyboard to activate the 'ROTATE MODE'. In the latter case, the cursor has to be placed within the tree display area.

Rotate terminal branch:
Move the cursor to any position along the branch, keep the left mouse button pressed while moving the cursor, release the button to fix the position of the branch.

Rotate subtree:
Move the cursor to any position along the edge defining the subtree and proceed as described above.

The subtree-defining branch is that located closest to the indicated root (open square) of the whole tree which is directly or indirectly connected to all branches of the subtree.

NOTES Short descriptions of the 'MODE' and the functions assigned to the mouse buttons are given in the fourth line of the 'ARB_NT' window.

The displayed tree topology is exported to the drawing tool xfig.

The middle mouse button (placed anywhere within the tree display area) can be used to move the whole tree.

The scale bar and its label can be moved by positioning the cursor anywhere on the bar, keeping the left mouse button pressed, moving the cursor to the desired position and subsequently releasing the mouse button. The label of the bar can be moved independently.

The previous topology can be restored using the 'RESET MODE' ('RESET' button on the 'ARB_NT' window).

::

mode_select.hlp

::

TITLE SELECT MODE

OCCURRENCE ARB_NT <SEL> button (first button of left column)

F1 on keyboard

DESCRIPTION Like in the info mode you can select species by clicking them, but no info window pops up.

When clicking on groups this mode works like the group mode.

::
mode_set_root.hlp
::
TITLE SET ROOT MODE

OCCURRENCE ARB_NT <S.ROOT> (button in the left column)

DESCRIPTION The 'SET ROOT MODE' allows to position the root of the displayed tree to any of the edges.

Click on the <S.ROOT> button to activate the 'SET ROOT MODE'.

Set the root:
Move the cursor to any position along the desired edge
and click on the left mouse button.

NOTES Short descriptions of the 'MODE' and the functions assigned to the mouse buttons are given in the fourth line of the 'ARB_NT' window.

The root is always placed to the middle of the edge.

::
mode_swap.hlp
::
TITLE SWAP MODE

OCCURRENCE ARB_NT <button in the left column>

DESCRIPTION The 'SWAP MODE' allows you to exchange the positions of adjacent branches of the displayed tree.

Click on the ninth button in the left column of the 'ARB_NT' window to activate the 'SWAP MODE'.

Swap branches:

Move the cursor to any position along the edge
connecting the two branches and the rest of the tree
(with respect to the indicated [open rectangle] root)
and click the left mouse button.

NOTES Short descriptions of the 'MODE' and the functions assigned to the mouse buttons are given in the fourth line of the 'ARB_NT' window.

The displayed tree topology is exported to the drawing

tool xfig.

The middle mouse button (placed anywhere within the tree display area) can be used to move the whole tree.

Swapping adjacent subtrees, the topology within the subtrees is not changed

The scale bar and its label can be moved by positioning the cursor anywhere on the bar, keeping the left mouse button pressed, moving the cursor to the desired position and subsequently releasing the mouse button. The label of the bar can be moved independently.

::
mode_width.hlp
::
TITLE LINE MODE

OCCURRENCE ARB_NT <'WIDTH'> button in the left column
F8 on the keyboard

DESCRIPTION The 'LINE MODE' allows you to gradually change the line width of the edges of the displayed tree individually.

Click on the 'WIDTH' button of the 'ARB_NT' window or press the 'F8' key on the keyboard to activate the 'LINE MODE'. In the latter case, the cursor has to be placed within the tree display area.

Change line width:
Move the cursor to the desired edge of the displayed tree and gradually increase or decrease the line width by repeatedly clicking on the left or right mouse button.

NOTES Short descriptions of the 'MODE' and the functions assigned to the mouse buttons are given in the fourth line of the 'ARB_NT' window.

The selected line widths are exported to the drawing tool xfig.

The middle mouse button (placed anywhere within the tree display area) can be used to move the whole tree.

::
mode_www.hlp
::
TITLE WWW MODE

OCCURRENCE ARB_NT (WWW button in leftmost column)
F7 on the keyboard

DESCRIPTION Starts default Web-Query in your browser.

NOTES Configure this feature in Properties/WWW

::
mp_params.hlp
::
TITLE Multiprobe parameters for calculation

Further parameters the user has to specify:
- select how big the combination should be
- select if the complement should be checked, too
- select the minimum mismatch that unmarked species should have in the combination
- for weighting mismatches there are three possiblities:
    1. all mismatches are weighted equally
    2. mismatches are weighted depending on which kind of mismatch occured
    3. mismatches are weighted stronger the nearer it occures at the center of the probe

- maximum non-group hits which are allowed for the calculation
- last but not least the pt_server has to be chosen. Remember to use an up to date server !!!

::
multiprobe.hlp
::
TITLE Multiprobe Calculation

DESCRIPTION Finds an optimal probe-triple given a set of probes. Often a group of species cannot be detected by a single oligo-nucleic probe. So several (normally 3) probes are combined to get a better result. Each probe is labeled with a unique color (red green blue).

WHAT DOES MULTI PROBE DO NOT It does not generate new probes !

WHAT DOES IT DO It combines a given set of probes (eg. the results of several probe design processes).

The MULTI PROBE main window shows two selection list. The upper left one is just used as a temporary clipboard. The upper right shows all probes that will be used in further calculations.

INPUT As input a list of precomputed probes is expected. Normally you simply use the result from a former probe design session, you transfer the data using save and load.
If you plan to use your probe design to get probes for MULTI PROBE, you should loosen the parameter set:
SET high values for 'Max non group hits' (+- 200)
and low values for 'Min group hits (%)' (+- 30%)

PARAMETERS

PT_SERVER See PT_SERVER online help for more details

Build select number of probes in target set

Check complement check also the complement. This should be the default selection, if you are mixing probe swith target sequences.

Weight mismaches If set, minor mismatches and mismatches at the ends of the probe are down weighted.

Max. non group hits As you can never be sure that your tree is absolutely correct, you allow a few non group hits. If you set this parameter to a too small value, you will not get any or good results.

Min. mismatches for non group the better your technical assistant the lower the value of this parameter.

Max mismatches for group Often a small mismatch (GU instead of GC, or GA instead of TA) does not destabilize a probe.

COMPUTATION Normally there are too many combinations of probes to be tested. So the program uses a heuristic approach to find a good but not optimal solution. The program never stops unless the user stops the computation by hand pressing the kill button.

RESULT Read the result help text.

NOTES You get much better results if all sequences are full sequences. Maybe you should delete all short sequences from the dataset, and create a new pt_server index file.

The pt_server index file and the currently loaded database should be nearly identical.

The program never stops. If you think you cannot wait any longer press kill and inspect the results.

The target group and nothing else should be marked. Be sure that you don't forget species to mark, especially if you are not working with the complete tree.

The buttons at the bottom of the window:
- 'Compute' calculates possible results
- 'Open result window' : This button can be chosen to go directly to the result window without calculation(i.e. to load an old result list)

The colors for the probes can be specified in the ARB Properties->Tree: Color and Fonts

::
multiproberesults.hlp
::

TITLE Multiproberesults

DESCRIPTION Shows the best multi-probes. Each multi-probe has a score, the higher the score the better.
The user may add any string as a comment (or one of the three default comments)

The selected multi-probe can be used to set the colors at the tips at the tree. First probe will be red, the next green, blue, red, green, blue and so on.

NOTE If you do not see any colors, you are probably using an old default file. You should delete the file '~/.arb_prob/ntree.arb' and restart arb.

::
ne_align_seq.hlp
::
TITLE Align Sequence

OCCURRENCE ARB_NT/ARB_EDIT/EDIT/Align Sequence

DESCRIPTION Aligns or realigns selected sequence with the database sequences. The most similar database sequences are found, and the new sequence is aligned with these according to primary and secondary structure similarity.

Choose the 'Align Sequence' item from the 'ARB_EDIT/EDIT' menu to display the 'ALIGNER' window.

Select a sequence by positioning the cursor on the name and double clicking the left mouse button.

Choose the appropriate 'PT_SERVER' database from the menu displayed after pressing the <Search relatives in> button. (For example, to align a 16S-like rRNA select 16s_rrna_aligned.arb.)

Define whether the most closely related species of the database should be marked or not by pressing the <Mark used relatives> button.

Press the <ALIGN> button.

NOTES The search for the most similar sequences is performed by the 'PT-server'. This server needs a special external database for rapid search. This database is not synchronized with the current database and has to be updated from time to time by a computing-time expe
nsive procedure ('ARB_NT/ETC/Probe Admin').

Marking the closest relatives allows manual optimization of the alignment of the new species with its closest relatives.

WARNINGS The aligner version available with the present package often

does not properly align short partial sequences.

BUGS The aligner needs lots of computer memory, sometimes more than you have. In this case it may not align some sequences.

The aligner does not align the last 3-6 bases of short sequences correctly. Look at the end of your sequences.

::
ne_compl.hlp
::
TITLE Complement Reverse

OCCURRENCE ARB_NT/ARB_EDIT/EDIT/Complement

DESCRIPTION Complements and/or reverses (part of) sequences. This is helpful if sequences were entered the wrong way round or if potential higher order structure elements should be studied.

Select a sequence by positoning the cursor on its name and pressing the left mouse button.

Choose the 'Complement' item from the 'ARB_EDIT/EDIT' menu to display the 'COMPLEMENT and REVERSE' window.

Define whether the sequence should be reversed, complemented or both by choosing from the menu displayed after pressing the <COMPLEMENT/REVERSE> button.

Define whether the position of gap symbols should be maintained (default) or the gaps should be removed (press the <remove> button).

Type left and right positions to the respective subwindows if part of the sequence should be complemented and/or reversed. In this case, press the <SEQUENCE PART> button.

Press <REST SEQUENCE> or <REST EDITOR> to complement or reverse the sequence from the curent cursor position to the (right) end of the sequence or the (right) end of the last (bottom) edited sequence, respectively.

Press <SEQUENCE> or <ALL> to complement the selected sequence or all edited sequences (independent of the cursor position).

WARNINGS !!! It is recommended to copy the sequence data before performing the operations. Currently there is no way to prevent the export of the modified sequences to the database. Saving the database means saving the complemented sequences. !!!!

Deleting gaps takes a lot of time.

::
ne_copy_sequence.hlp
::
TITLE Copy Sequence

OCCURRENCE ARB_NT/ARB_EDIT/EDIT/Copy Sequence

DESCRIPTION Copies a sequence entry.

Choose the 'Copy Sequence' item from the 'ARB_EDIT/EDIT' menu to display the 'COPY SELECTED SEQUENCE' window.

Type a name to the 'COPY SELECTED SEQUENCE' window and press <GO>.

NOTES The new sequence is displayed within the 'ARB_EDIT' window.

::
ne_new_sequence.hlp
::
TITLE Create Sequence

OCCURRENCE ARB_NT/ARB_EDIT/EDIT/Create

DESCRIPTION Creates a new sequence entry.

Type a name to the 'CREATE SEQUENCE' window and press <GO>.

The new sequence is displayed within the 'ARB_EDIT' window.

WARNINGS Save database first.

BUGS Sometimes ARB crashes.
Save database before doing this operation.

:::::::: :: ::
ne_pretty.hlp
::
TITLE Printing from Eitor

OCCURRENCE ARB_NT/ARB_EDIT/File/Pretty Print

DESCRIPTION The data displayed by the editor cannot be printed directly but written to ASCII file, which can be printed.

The corresponding tool is taken from the public domain software GDE and adopted for the special needs of the ARB editor.

Press the 'Pretty Print' button of the 'ARB_NT/ARB_EDIT/File' menu to display the 'GDE/FILE/Pretty Print' window.

Define whether SAI and/or helix symbols should be printed with

the sequences by pressing the 'Print SAI' and/or 'Print helix' buttons, respectively.

Define whether the data in the fixed (if) top and/or fixed (if) bottom part and/or (scrollable) middle part of the editing area should be printed by pressing the 'top area', 'middle area' or 'bottom area' buttons, respectively.

Compress aligned data:

Select a filter (gap symbols and zeros in SAI entries define alignment columns to delete) from the submenu displayed after pressing the button after the 'Filter:' prompt.

Press one of the 'Compression' buttons. Pressing the 'vertical gap' button removes columns containing gap symbols common to all edited sequences. Pressing the 'all gaps' button removes all gap symbols and destroys the alignment!

NOTES You need an ASCII printer or an ASCII to POSTSCRIPT converter.

::
ne_replace.hlp
::
TITLE Replace Character String

OCCURRENCE ARB_NT/ARB_EDIT/EDIT/Replace

DESCRIPTION Replaces characters or strings within the selected sequence or SAI entry.

Select a sequence by positoning the curser on the name and pressing the left mouse button. Position the cursor within the SAI or sequence entry.

Choose the 'Replace' item from the 'ARB_EDIT/EDIT' menu to display the 'EDIT SEARCH' window.

Type the search and replace strings to the 'SEARCH-STRING' and 'REPLACE-STRING' subwindows, respectively.

Type the number of accepted mismatches to the 'Maximum mismatches' subwindow.

Define whether lower and upper case symbols or T and U should be treated as matches or mismatches by pressing the <Case> or <T=U?> buttons respectively.

Define whether gap symbols should be taken into account comparing search string and edited sequences by choosing from the menu displayed after pressing the <Consider

Gaps> button.

Press one of the <FIND ...> buttons to see the string within the sequence or SAI entry. Press <REPLACE>. Alternatively press one of the <... REPLACE ...> buttons.

NOTES 'REPLACE TO END OF SEQUENCE' and 'REPLACE TO END OF EDITOR' replace all search strings from the current cursor position to the (right) end of the entry or the (right) end of the last (bottom) entry, respectively.

::
ne_search.hlp
::
TITLE Search Character String

OCCURRENCE ARB_NT/ARB_EDIT/EDIT/Search

DESCRIPTION Searches for character strings within the selected sequence or SAI entry.

Select a sequence by positioning the cursor on the name and pressing the left mouse button. Position the cursor within the SAI or sequence entry.

Choose the 'Search' item from the 'ARB_EDIT/EDIT' menu to display the 'EDIT SEARCH' window.

Type the search string to the 'SEARCH-String' subwindow.

Define number of accepted mismatches in the 'Maximum mismatches' subwindow.

Define whether lower and upper case symbols or T and U should be treated as matches or mismatches by pressing the 'Case' or 'T=U?' buttons, respectively.

Define whether gap symbols should be taken into account comparing search string and edited sequences by pressing the 'Consider Gaps' button.

Press one of the 'FIND ...' buttons to execute the search.

The cursor is positioned to the left of the detected string.

NOTES Gaps within the aligned sequences should only be considered if the gap symbols are contained in the search string.

::
nekey_map.hlp
::
TITLE Customize Keyboard

OCCURRENCE ARB_NT/ARB_EDIT/Props/Key Mappings

DESCRIPTION Nucleotide (amino acid) and gap symbols can be assigned to any of the letter and symbol keys of the keyboard. This makes it easier to import sequence and SAI (sequence associated information) data by typing.

Press the 'Key Mappings' button of the 'ARB_NT/ARB_EDIT/Props' menu to display the 'KEY MAPPINGS' window. Assign symbols to keys by typing the real key and symbol to the 'Map keyboard' and 'to ASCII' subwindows, respectively.

Keymapping is only performed if 'Enable Mapping' is activated.

WARNINGS The user-defined key mappings are not available for typing of search and replace strings using the 'ARB_NT/ARB_EDIT/EDIT/Search' and 'ARB_NT/ARB_EDIT/EDIT/Replace' tools. Be careful with ' ' symbols. They can also be mapped but are not visible to the user.

::
neprops.hlp
::
TITLE Line Spacing

OCCURRENCE ARB_NT/ARB_EDIT/Props/ETC

DESCRIPTION Line spacing of sequence data and helix symbols can be adjusted independently.

Press the 'ETC' button of the 'ARB_NT/ARB_EDIT/Props' menu to display the 'EDIT_PROPERTIES' window.

Define whether helix symbols should be shown with SAI (sequence associated information) entries by pressing the 'Show Helix for SAI'. This may be of interest with consensus sequences.

Define line spacing for sequence and SAI entries by typing to the 'Line Space' subwindow.

Define line spacing for Sequence entries and the corresponding helix symbols by typing to the 'Seq. - Helix Space' subwindow.

NOTES The helix symbols are shown below the corresponding sequence entries and can be displayed in a different colour (ARB_NT/ARB_EDIT/Props/Data:).

::
neprops_data.hlp
::
TITLE Setup of Editing Area

OCCURRENCE ARB_EDIT/Props/Sequences Colors and Fonts

DESCRIPTION Allows definition of colours and fonts for the editing area of the 'ARB_EDIT' window

Choose the 'Sequences Colors and Fonts' item from the 'ARB_EDIT/Props' menu and select colours and fonts by choosing from the submenus or typing in the subwindows of the 'COLORS AND FONTS' window.

NOTES Different colours and fonts can be used to indicate selected and non-selected sequences or the user-defined helix symbols.

Line spacing for sequences and helix symbols can be defined independently (ARB_EDIT/Props/ETC).

::

next_neighbours.hlp

::

TITLE Search the nearest relatives

OCCURRENCE ARB_NT

DESCRIPTION Scans a selected PT_SERVER database for the nearest relatives of the sequence of the selected species.

ALGORITHM Splits the sequence into oligos of size 10. Those oligos are 'Probe Matched' in the PT_SERVER database. The more hits within one foreign sequence, the more equal the forein sequence is.

WARNINGS Names may be different in private and PT_SERVER database.

::

no_tree.hlp

::

TITLE How to get an initial tree

DESCRIPTION The main idea of ARB is to manage database access via a tree.

[You don't have a tree, but you may access all species data using 'Species/Search and Query'.]

There are several ways to construct an initial tree:

Align the sequence data:

1. Mark all 'species' (see 'HELP: Glossary')
   Choose the item 'Mark all Species'
   from the menu 'Species'
   of the 'ARB_NT' main window (this program)
   (short: ARB_NT/Species/Mark all Species);

2. Select an alignment:

<ARB_NT/3rd big Button in top area>
(see online help)

3. Start the editor:
<ARB_NT/Sequence/Edit marked Sequences>
(see online help)

4. Align the data (by hand)

Alternatively use ClustalV for aligning (few sequences only):

Choose <ARB_NT/ETC/GDE> to activate the GDE extension.

Choose <DNARNA/Clustal...> or <Protein/Clustal...> to

Click on <help>: read the help and set all parameters

Choose an 'alignment' from the Alignment subwindow.

Click on <GO>

After some time (1min - several hours) The aligned
data will be shown by the GDE editor.

Save the data as: transfer
type: genbank

Quit the GDE program

Reimport the data into ARB_NT:

Set the protection level to 6 by clicking ob the
<Protection> button and choosing from the
displayed menu.

Choose the 'Import ... (using GDE/readseq)' item
from the 'File' menu to display the
'GDE/File/Import...' window.

Type "transfer" to the 'Name of foreign file' subwindow

Click on <GO>

WARNING: Sequence names have to contain at least 3 characters.

Reconstruct an initial tree:

To get a good tree, you should use different treeing
methods.

To quickly get an initial tree we recommend to use
neighbour joining:

DNA/RNA or Protein sequence:

Choose the 'Neighbour joining' item from the 'Tree' menu to display the 'NEIGHBOUR JOINING' window. See help for setting parameters and tree reconstruction

Show tree:

Click on the <tree_*> (2nd big rectangular) button in top area and choose the respective tree

Save everything:

Choose the 'Save Whole Database as' item from the 'File' menu.

::
nt_align_select.hlp
::
TITLE Select an Alignment

OCCURRENCE ARB_NT/(3rd broad rectangular button in top-area)

DESCRIPTION Different 'alignments' (see 'HELP: Glossary) each comprising aligned sequences of different genes or nucleic acid and predicted amino acid sequences, respectively, assigned to the same species can be stored in one database.
The name of the currently acessible alignment (ali_*) is shown in the respective button (ARB_NT/3rd broad rectangular button in top-area)
To switch to another dataset press the button and select one from the 'SELECT AN ALIGNMENT' window.

WARNINGS Running editors are not informed about a change of that entry. You have to restart them.

::
nt_props_data.hlp
::
TITLE Tree Display

OCCURRENCE ARB_NT/Properties/Data:

DESCRIPTION Allows to define colours and fonts for the tree display area of the 'ARB_NT' window.

Press the 'Tree:' button of the 'ARB_NT/Properties' menu and select colours and fonts by using the respective submenus or typing in the respective subwindows of the 'COLORS AND FONTS' window.

NOTES Different colours and fonts can be used to indicate marked and unmarked species as well as branches or groups containing marked and unmarked species.

::
nt_props_tree.hlp
::
TITLE Standard help file form

OCCURRENCE ARB_NT

DESCRIPTION

::
nt_tree_select.hlp
::
TITLE SELECT A TREE

OCCURRENCE ARB_NT<second broad rectangular button in top area>

DESCRIPTION Different trees for the same data (sub) sets can
be stored in the database.
The name of the currently edited tree (tree_*) is
shown in the button <second broad rectangular button
in top area>.
To display another tree press the button and select
a tree from 'SELECT A TREE' window.

::
nt_tree_settings.hlp
::
TITLE Tree Settings

OCCURRENCE ARB_NT/Properties/Tree Settings

DESCRIPTION Defines line width of the branches and vertical distance of the
terminal nodes (dendrogram).

Type values to the respective subwindows.

The changes can be saved to files 'arb_prop*' located in the
users home directory by choosing the "Save Defaults'.

@@@ Einzelne Werte

NOTES The line width of individual branches can be changed
independently using the 'LINE WIDTH MODE'.

::
optimize.hlp
::
TITLE Optimize Database

OCCURRENCE ARB_NT

DESCRIPTION Sequence data normally need's a lot of memory. To be able to
handle thousands of sequences we implemented an online

compression. All data is compressed most of the time and only uncompressed on demand. As a user you only find smaller database files, that's all.
Without understanding the data, the program can compress data only by a limited factor. With the help of a tree aligned sequences can be compressed much better by storing only the differences to a consensus sequence.
Once a sequence is compressed using a tree, it will keep the good compression method until it is changed. Then only the older method is used.
As long as you change only a few (up to 100) sequences, the database won't grow very much.

To compress the entire database, the program needs a tree, which should cover most of the sequences. The larger and better the tree, the better the compression.

EXAMPLE 10000 aligned 16s sequences need 50 mega-bytes of memory. Without your help ARB will reduce them to 10 mega-bytes, and given a tree not more than 2 mega-bytes will be needed.

NOTES Any major database update, especially inserting or deleting gaps in an alignment, should be followed by a new optimization step.

::
pa_add.hlp
::
TITLE Add Marked Species and Local Optimization

OCCURRENCE ARB_NT/Tree/Parsimony/Tree/Add Marked Species + NNI

DESCRIPTION All 'marked species' (see 'HELP: Glossary') are positioned according to maximum parsimony criteria. Local optimization of the current tree is performed by swapping subtrees which are separated by not more than one edge.

NOTES The current parsimony value (number of base changes needed) is displayed after the 'Current Par' prompt in the upper part of the 'ARB_PARSIMONY' window and the 'Message' window (pops up while calculating).

WARNINGS No global optimization is performed

::
pa_add_sel.hlp
::
TITLE Add Selected Species and Local Optimization

OCCURRENCE ARB_NT/Tree/Parsimony/Tree/Add Selected Species + NNI

DESCRIPTION The 'selected species' (see 'HELP: Glossary') is positioned within the current tree according to maximum parsimony criteria. Local optimization of the current tree is performed by swapping

subtrees which are separated by not more than one edge.

NOTES The current parsimony value (number of base changes needed) is displayed after the 'Current Par' prompt in the upper part of the 'ARB_PARSIMONY' window and the 'Message' window (pops up while calculating).

WARNINGS No global optimization is performed

::
pa_bootstrap.hlp
::
TITLE Estimation of Bootstrap by Parsimony

OCCURRENCE ARB_NT/Tree/Parsimony

DESCRIPTION Given a large tree, traditional ways to calculate bootstrap values are by magnitudes to slow. So a faster algorithm was developed:
the bootstrap values for each branch is caclualted under
the assumption that all other branches have a 100% value.
Doing this we get an upper limit for the real bootstrap values.

NOTES The program does not use the traditional Monte Carlo method to estimate the bootstrap values but calculates them correctly under the assumption that the tree changes only locally.
Try different filters and see the effekt on the tree.

ALGORITHM For each branch B do:

```
a              b
 \            /
  >-----------<
 /      B     \
c              d
```

```
exchange a with b      ( or a with d )
and count all coloums in the alignment
with a greater/smaller/equal minimal number of mutations
than the original tree.
result:         n_plus, n_minus, n_equal
        freq_n_plus = n_plus/ (seq_len)
        ...
Bootstrap value = sum of
    for all i = 1.. seqlen do
        for all combinations of np, nm,ne with np - nm == i do
            sum +=  freq_n_plus  ^ np *
                freq_n_minus ^ nm *
                freq_n_equal ^ ne *
                seq_len! / np! /nm! /ne!
        done
```

PUBLIC This algorithm is not published and I am not going to publish it. If you feel the strong need to to this, please don't forget

to mention me (strunk@mikro.biologie.tu-muenchen.de)

WARNINGS Use filters to exclude superfluous gaps and to
increase bootstrap values

BUGS Does not work with weights
Does not work with proteins

::
pa_optimizer.hlp
::
TITLE Global Optimization

OCCURRENCE ARB_NT/Tree/Parsimony/Tree/Global Optimization

DESCRIPTION Performs global optimization of the current tree.

WARNINGS !!! Not implemented yet.!!!

::
pa_quick.hlp
::
TITLE Add Marked Species

OCCURRENCE ARB_NT/Tree/Parsimony/Tree/Add Marked Species

DESCRIPTION All 'marked species' (see 'HELP: Glossary') are
positioned according to maximum parsimony criteria. The current
tree topology is not changed.

NOTES The current parsimony value (number of base changes needed) is
displayed after the 'Current Par' prompt in the upper part of
the 'ARB_PARSIMONY' window and the 'Message' window (pops up
while calculating).

This tool should be used for the positioning of 'species' for
which only partial or preliminary sequence data are availble.

WARNINGS The phylogenetic information conferred by the new sequence(s) is
not used for global treee optimization.

::
pa_quick_sel.hlp
::
TITLE Add Selected Species

OCCURRENCE ARB_NT/Tree/Parsimony/Tree/Add Selected Species

DESCRIPTION The 'selected species' (see 'HELP: Glossary') is positioned
within the current tree according to maximum parsimony criteria.
The current tree topology is not changed.

NOTES The current parsimony value (number of base changes needed) is
displayed after the 'Current Par' prompt in the upper part of

the 'ARB_PARSIMONY' window and the 'Message' window (pops up while calculating).

This tool should be used for the positioning of 'species' for which only partial or preliminary sequence data are availble.

WARNINGS The phylogenetic information conferred by the new sequence is not used for global treee optimization.

::

pa_ranchlengths.hlp

::

TITLE Calculate Branch Lengths

OCCURRENCE ARB_NT/Tree/Parsimony/Tree/Calculate Branch Lengths

DESCRIPTION Calculates branch lenghts for the current tree. Branch swapping is used to estimate the significance of tree topologies.

NOTES The branch lengths reflect the significance of edges rather than the number of changed residues.

::

pars.hlp

::

TITLE ARB Parsimony

OCCURRENCE ARB_NT/Tree/Parsimony

DESCRIPTION: In this release, 'ARB Parsimony' cannot be used to reconstruct new trees.

Existing trees can be optimized at different levels (depths) according to maximum parsimony criteria.

New sequences (species) can be placed rapidly into an existing tree. This can be done without changing that tree or allowing optimization at different levels according to maximum parsimony criteria including the new information for calculation.

Filters for the in- or exclusion of alignment columns ('ARB_NT/SAI; ARB_NT/Tree/Dist Matrix) can be seleted.

Choose the 'Parsimony' option from the 'Tree' menu to display the 'SET PARSIMONY OPTIONS' window.

Select a tree (to be modified) from the 'Tree:' subwindow.

Select an 'alignment' (see 'HELP: Glossary') from the 'alignment:' subwindow.

To select a filter display the 'Select Filter' window pressing the button after the 'Filter:' prompt (see 'HELP: Select Filter').

Press the <GO> button to display the 'ARB_PARSIMONY' window.

::
parser.hlp
::
TITLE SRT Search and Replace Tool

MODULE The String Parser
is used to search and replace substrings.

WHERE 1. Parse field entries: 'TREE/Species/Search:PARSE FIELD'
2. Parse tip infos: 'TREE/Proporties/NDS:

SYNTAX 'search=replace' means search all occurrences of 'search'
and replace it by 'replace'

Different search/replace commands can be seperated by ':':

'search1=replace1:search2=replace2: ... :searchn=replacen'

SPECIAL CHARACTERS:

Search && Replace string:
: seperates two commands
= seperates the search from the replace string
\ Escape symbol
\\ the '\' symbol itself
\n newline
\t tabulator
\: ':'
\= '='
\? '?'
* '*'

Search string:
? single letter wildcard
* multi letter wildcard
Replace string:
? a reference to the corresponding single
letter wildcard in the search string
(if no digit or '(' follows).
?n n = { 1,...,9 }
a reference to the n'th single letter wildcard
in the search string
* a reference to the corresponding multi
letter wildcard in the search string
(no digit or '(' follows).
*n n = { 1,...,9 }
a reference to the n'th multi letter wildcard
in the search string
*(key) the value of a database field named <key>
if this field does not exits then the string ''
*(key#mystring) the value of a database field named <key>

```
            if this field does not exits then the string
            'mystring'
        *(key\:nsrt) invokes the SRT recursively on the value of the
            database field 'key'
            NOTE: The ':' have to be 'escaped'
        *([key]|ACI) starts the ARB COMMAND INTERPRETER on the
            optional value of 'key'

EXAMPLES:
        'p?r=p?w'       replaces all    par to paw
                                        pbr to pbw
                                        pcr to pcw ...
        'p??r=p?2?1r'   swaps the two letters between p and r

        'a*=b*'         replaces only the first 'a' by 'b'
        '?* *=?. *2'    Replaces the first word by its first
                        letter + '.'
        '\:=\n'         replaces all ':' by <newline>
        '*=* *(key1)'   appends the database field <key1>
                        if <key1> does not exists append nothing
        '*=* *(key1#no info)'
                        appends the database field <key1>
                        if <key1> does not exists append
                        'no info'
        '*=*(key2\: =)' The value of 'key2' with all spaces removed

        '*=*(key2|remove(.-))'
                        the value of the database entry 'key2',
                        but all '.' and '-' characters removed
WARNINGS:
        Be careful when search  or replace string contain
        special characters (such as ':'). Avoid to write to
        complicated commands.


:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
pd_expert.hlp
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
TITLE       Standard help file form

OCCURRENCE      ARB_NT

DESCRIPTION

:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
pd_result.hlp
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
TITLE       Standard help file form

OCCURRENCE      ARB_NT

DESCRIPTION

:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
```

pd_spec_param.hlp

TITLE Weighting of Base Pairings

OCCURRENCE ARB_NT/ETC/Probe Design/PROBE DESIGN/EXPERT

DESCRIPTION Allows to adjust values indicating the relative strengths of base (mis) pairings. These values are used by ARB_PROBE for the ranking of potential probe target sites.

Define weights:
Type the values to the corresponding subwindows of the 'Relative Strength of Base Pairings' matrix.

Define threshold for splitting melting domains:
Type a value to the 'Theshold for Splitting' subwindow. This value is substracted from the mean of all pairing values of the potential probe target hybrid. If the resulting value is lower than that of a particular mispairing, it is assumed that the remaining base paired stretches melt independently. The dissociation temperature of the whole hybrid is that of the more stable part of it.

WARNINGS The criteria used for probe target site ranking are based on experimental experience. However, optimum conditions have to be determined experimentally for the particular probe target hybrids.

ph_export_markerline.hlp

TITLE Export Filter to ARB @@@@@@@@@@@@

OCCURRENCE ARB_NT

DESCRIPTION

phyl.hlp

TITLE Matrices, Masks, Profiles V1.0

OCCURRENCE ARB_NT/Tree/Dist Matrix V 1.0

DESCRIPTION This tool allows to calculate distance and similarity matrices for the marked species. Conservation profiles can be established and used as filters for column selection by other programs.

1. Selection of columns:

Select part of the alignment to analyze by typing first and last column numbers after the 'start at column:' and 'stop at column:' prompts, respectively, and press 'Return' on

the keyboard.

Select minimum and maximum similarities for the individual columns to be included for similarty or distance matrix calculation by typing the values (50 means the most frequent base at a particular position is shared by at least 50% of all marked sequences (species)) after the 'minimum similarity:' and 'maximum similarity:' prompts, respectively, and press 'Return' on the keyboard.

Define whether alignment gaps and ambiguities within individual marked sequences (species) should be taken into account:

Use the rigth mouse button to display the submenus associated to the items below the 'markerline:' prompt by pressing the respective buttons.

don't count:
Calculate conservation from unambiguous bases only

don't use column if maximal:
Exclude column if the respective symbol is present in the majority of the marked sequences.

exclude column:
Exclude column if the respective symbol is present in any of the marked sequences.

treat as abiguous:
Take the respective symbol as an unambiguous residue.

2. After selecting columns define how to treat ambiguities for distance calculations:

Use the rigth mouse button to display the submenus associated to the itemss below the 'distance matrix:' prompt by pressing the respective buttons.

don't count:
The particular position is not included for binary distance calculations if the symbol is present in one or both sequences.

use distance table:
the symbols are treated as unambiguous residues

3. Calculate profiles and matrices:

Use the right mouse button to display the 'CALCULATE' menu and select 'markerline' (profile) or 'distance matrix' by releasing the mouse button while the curser is positioned on the respective menu button.

4. Display results:

Use the right mouse button to display the 'VIEW' menu and select 'species', 'markerline' or 'distance matrix' by releasing the mouse button while the curser is positioned on the respective menu button. The names, the aligment of the marked sequences and the conservation profile, or the distance matrix are shown within the display area, respectively.

The profile:

The fraction of sequences sharing most frequent residue at a particular alignment position is shown as a number to read bottom down. Alternatively, the profile can be displayed as a curve by pressing the <toggle> button in the left part of the window. It can be smoothed by selecting a number from the 'smooth' menu (left part of the window).

Editing the aligned sequences and profiles, a name can be selected by moving the cursor to it and pressing the left mouse button. Pressing the <reference> button, the respective sequence is used as a filter superimposed to profile. This allows to exclude further positions from subsequent calculations which are not occupied by bases in the reference.

Tha matrix:

Editing the matrix, mean values can be calculated for groups of organisms displayed as triangles (radial tree) or rectangles (dendrogram) in a tree stored in the database. The grouping currently or most recently displayed is used for selecting sequences for the calculation of mean values. Select a tree by pressing the <grouping> button in the left area of the window.

5. Save results:

To export distances and profiles (not graphs!!) to ascii files, use the right mouse button to display the 'SAVE' menu and select the corresponding menu item by releasing the mouse button while the curser is positioned on it.

uncoded mline:
The positional conservation is encoded by a sequence of letters: A = 1%,

B = 5%, ..., Z = 100%.

bin mline:
Included and excluded columns are indicated by 1 and 0, respectively.

% dif. matrix:
Dissimilarity values.

% sim. matrix:
Similarity values.

phyl/prot distmat:
Corrected distances according Jukes and Cantor (nucleic acid sequences only).

pos. vari:
Number of different residues at the particular alignment position.

6. Reconstruct a tree:

Press the <NEIGHBOR> button. The tree is reconstructed using the neighbour joining method and is stored in the database as 'tree_neighbour'.

7. Store profile in the database:

Press the <ARBSAVE> button to store any calculated profile 0 / 1 encoded in the database (SAI). The profile can be used as filter for other ARB tools.

NOTES A new version of the tool is under development.

It is recommended to calculate profiles here, to save them in the database, and to reconstruct distance matrix trees using 'ARB_NT/Tree/Neighbour joining' in combination with the profile as filter.

WARNINGS Whenever text is typed to the window, press 'Return' on the keyboard, to ensure that the information is recognized by the program!!!

::
phylo.hlp
::
TITLE Matrices, Masks, Profiles V2.0 @@@@@@@ siehe phyl.hlp

OCCURRENCE ARB_NT/SAI/Create Filter/by Base Frequncy

DESCRIPTION None

NOTES Not yet implemented

::
pos_var_pars.hlp
::
TITLE Column - Statistik

OCCURRENCE ARB_NT/SAI/Create SAI from Sequences/Positional Variability ...

DESCRIPTION Calculates the base and frequencies positional variability for
each column independently.
It uses the parsimony method to find the minimum number of
mutations for each site.

The result can be used by:

- Parsimony to weight the characters properly
- Neighbour joining to estimate the distances
more accurately.
- Filter

@@@@
Result:
'.' Less than 10% valid characters
'0123456789ABCDE...'
The higher the number the more conserved
+2 half mutations
eg .'7' half number of mutations than '5'

NOTES Use the biggest tree you have.

NOTE @@@ Compare it to consensus and max frequency.

WARNINGS if you have only small trees (<100 species), forget this
function.

::
primer.hlp
::
TITLE Primer Design @@@@@@@

DESCRIPTION nur markierte Sequencen
Daten muessen aligned sein,
sucht alignmentabschnitte mit moeglichst wenig verschiedenen
Kombinationen
fals ihre Anzahl < Wert -> moeglicher Prime

Die Ausgabe erfordert viel viel Handarbeit

normalerweise mehrere hundert Seiten

ctrl-C bricht das Programm ab

PraxisErfahrung:
Schlechte Sequencen bringen den Alg. durcheinander
-> entmarkieren

```
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
primer_new.hlp
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
TITLE          Primer Design

OCCURRENCE     ARB_NT/ETC/Primer Design

DESCRIPTION    Searches for pairs
               within the given physical and biological ranges

               1. select a species (ARB_NT <INFO button>)
               2. choose ARB_NT/Etc/Primer Design...
               3. specify searchparameters
               4. mark a result, it will be displayed in ARB_EDIT

NOTES          written by Wolfram Förster 2001

:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
primer_parameters.hlp
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
TITLE          Primer Design/Search Parameters

OCCURRENCE     ARB_NT/ETC/Primer Design

DESCRIPTION    left/right position:
               specifies the start of primer ranges
               (leftmost base position)
               [right.min > left.min]

               left/right number of bases
               specifies how many bases a primer range includes

               primer length:
               number of bases within a primer
               [length > 0]

               primer distance:
               min/max number of bases between leftmost base in left primer
               and rightmost base in right primer

               G+C ratio:
               range of GC-ratio of primers in percent
               (countG+countC / countBases * 100)

               temperature:
               range of temperature of primers
               (4*(countG+countC) + 2*(countA+countT))

               minimal distance to next match of primer:
               if a primer is found again outside this range it is ignored
               example: primerpos     100
                        min.dist      50
                        2nd occurence  30 => primer remains in list
```

3rd occurence 140 => primer is removed from list

expand IUPAC-codes:
switches expansion of IUPAC-codes on/off
example: TTRC is expanded to TTAC and TTGC

maximal number of results:
how many possible pairs you want ?

G+C-factor / temperature-factor:
weight of GC / temperature while evaluation of
possible primerpairs

NOTES written by Wolfram Förster 2001

::
probe_param.hlp
::

TITLE Probe Parameters

OCCURRENCE ARB_NT/ETC/Probe Design/PROBE DESIGN

DESCRIPTION Allows to select the 'PT_SERVER' and its database, to customize
the presentation of the results, define the stringency of target
search and physical characteristics of the probes.

Select a 'PT_SERVER':
Select the appropriate 'PT_SERVER' from the menu
displayed after pressing the button below the
'PT_SERVER' prompt. The 'PT_SERVER' database (*.arb)
has to be consistent with the current. (see 'Main
Topics': 'Probe Design'; 'PT_SERVER What Why and How')

Customize output of results:
Type the number of probe target proposals to be
displayed in the 'PD RESULT' window by typing it to the
'Length of output' subwindow.

Define number of accepted non-targets:
Type the number of non-target species which you would
accept to be detected by the probe to the 'Max. non
group hits' subwindow.
This helps not to miss potential target sites in case
that species belonging to the particular specificity
group had been overlooked while marking (see: 'Main
Topics': 'Probe Design').

Define number of accepted inter- and intraprobe base pairings:
Type the number of potential base pairings within or
between probe molecules.
!!! Not yet implemented !!!

Define minimum number of target species:

Type the fraction (%) of marked species which have to share the target site to the 'Min. group hits (%) subwindow.
This helps to design multiple probes in case that common target sites are not present in all species of the particular specificity group.

Define length of the probe:
Type minimum and maximum number of nucleotides to the corresponding 'Length of probe' subwindows to define probe length.

Define dissociation temperature:
A range of allowed dissociation temperatures (= 4xGC + 2xAU; centigrade) can be defined by typing minimum and maximum values to the corresponding subwindows.

Define the G+C content:
A range of allowed G+C fractions (%) can be defined by typing minimum and maximum values to the corresponding subwindows.

Define target region:
A preferred sequence (alignment) region can be defined for the probe target sites by typing the nucleotide numbers of the homologous positions within the E. coli molecule.
This require:
1. Your (pt_server) data is aligned.
2. There is a SAI named 'ECOLI' which includes the reference sequence.

Press the 'EXPERT' button to display the 'PD SPECIAL' window if base pairings should be individually weighted (see 'HELP: Weighting of Base Pairings).

Press the 'GO' button to start the calculations.

NOTES Increasing of the 'Max. non group hits' and reducing 'Min. group hits (%)' values as well as increasing of the difference of the minimum and maximum values for 'Length of probe', 'Temperature' as well as 'G+C content' reduces the performance (speed) of the program.

The results will be shown within the 'PD RESULT' window which can be displayed by pressing the 'RESULT' button. The window is automatically displayed when the probe search is completed.

::
probeadmin.hlp
::
TITLE PT_SERVER Administration

OCCURRENCE ARB_NT/ETC/PT Admin

DESCRIPTION Allows to start and kill 'PT_SERVER' and to update 'PT_SERVER' databases.

Select a 'PT_SERVER' and database from the menu displayed after pressing the button below the 'PT_SERVER' prompt.

Start or kill the selected 'PT_SERVER' by pressing the corresponding buttons.

Press the 'UPDATE SERVER' button to export the current database to the selected 'PT_SERVER' database.

NOTES Read PT_SERVER: What Why and How if you do not understand this helptext.

The probe designing and matching tools ('ARB_PROBE') and the aligner of the editor ('ARB_EDIT') rely on 'PT_SERVER' databases and servers. Update the 'PT_SERVER' databas when new sequence (species) entries or sequence modifications (base changes) have been introduced into the current data base.

::

probedesign.hlp

::

TITLE Probe Design

OCCURRENCE ARB_NT/ETC/Probe Design

DESCRIPTION Searches for potential probe target sites within the sequence entries of the corresponding 'PT_SERVER' (not the current) database.

Mark (single or group of) species (ARB_NT <MARK button>; ARB_NT/Species/...) for which probe target sites should be found.

Choose the 'Probe design' item from the 'ARB_NT/ETC' menu to display the 'PROBE DESIGN' window.

PARAMETERS (press Probe-Parameter help to get detailed information)

Select a PT_SERVER: Probe design is not run on your database
but on the pt_server's.
See 'PT_SERVER What Why and How'.
To work on a consistent database you
should 1. update the pt_server or
2. run arb on the databases used
by the pt_server
($ARBHOME/lib/pts/*.arb)

Length of output: Clip too long output lists

Max. non group hits: Maximum number of species, that
are not marked but matches a probe

Max. hairpin bonds: Currently not implemented.

Min. group hits: Minimum percentage of marked species
to match a probe.

ECOLI-position: Restricts the target position.
This require:
1. Your (pt_server) data is aligned.
2. There is a SAI named 'ECOLI' which
includes the reference sequence.

OUTPUT

NOTES The 'PT_SERVER' database ('*.arb' and '*.arb.pt') stored in '$ARBHOME/lib/pts' is used for probe target searching not the current database.
The 'PT_SERVER' database has to be updated ('ARB_NT/ETC/Probe Admin') if species entries should be considered for probe target searching which have been added or modified (sequence symbols) later than the date of the most recent 'PT_SERVER' database update.
Probe target searching does not depend on correctly aligned sequences and is not affected by any modifications of database entries except changes of sequence residues.

WARNINGS Take care that only and all species are marked for which a group specific probe has to be designed! If the displayed tree is used for species marking, consider that species belonging to the specificity group but not contained in this tree will remain unmarked and treated as non-targets.
Consequently, useful target sites may be not detected.
Similarily, marked species not related to the target species and not contained in the displayed tree will be treated as targets.

::
probedesignresult.hlp
::
TITLE Potential Probe Targets

OCCURRENCE ARB_NT/ETC/Probe Design/PROBE DESIGN/PD RESULT

DESCRIPTION The 'PD RESULT' window is displayed automatically after finishing the probe target searching or manually by pressing the 'RESULT' button of the 'PROBE DESIGN' window.

The parameters defined by in the 'PROBE DESIGN' window are listed in the first line of the display area of the 'PD

RESULT' window.

The target and probe sequences are given in the first and last columns of the display area, respectively.

The length of the psoposed probe is given in column 2.

Columns 3 and 4 indicate the 5'-positions of the target sites within the alignment and the E.coli sequence, respectively

The number of species targeted (perfectly paired) by the probe is given in column 5.

The G+C contents and predicted dissociation temperatures (4xGC + 2xAT) are listed in columns 6 and 7, respectively.

The numbers of the next columns (from left to right) indicated how many non-target species would be detected by the particular probe if the temperature would be gradually lowered.
!!! The columns do only represent virtual temparature shifts and cannot be assigned to centigrades.!!!

To write the results to an ascii file press the 'SAVE' button to display the 'SELECTION BOX' window.

SORTING The programm brings the best probes to the front of the list.
Best means the produkt
group_hits * (min. mismatches to non group sp.)
It does not take G+C, temperature or ECOLI position into account.
It is up to the user to study the list carefully and choose his candidates. (In our lab we often found the 20th probe the best).

::
probematch.hlp
::
TITLE Probe Match

OCCURRENCE ARB_NT/ETC/Probe Match

DESCRIPTION Finds and displays all occurrences of a given target and/or probe sequence within any specified 'PT_SERVER' database. The species, targets and additional information are ranked and displayed according to the degree of similarity.

Select a 'PT_SERVER' from the menu dispalyed after pressing the 'PT_SERVER' button of the 'PROBE MATCH' window.

Define whether similar (not perfectly matched) sites should be displayed by pressing the 'Search depth' button and selecting the number of mismatches (1 - 5) from the menu.

Press the 'Use weighted mismatches' button to use or not the weights defined in the 'PD SPECIAL' window (see 'HELP: Weighing of Base Pairings').

Define whether the species which contain the target or probe (or similar) sequence should be marked (see 'HELP: MARKED SPECIES') by pressing the 'Mark in database' button.

Define whether probe and/or target sequence should be searched by pressing the 'Check complement too' button.

Press the 'EXPERT' button to display the 'PD SPECIAL' window if base pairings should be individually weighed (see 'HELP: Weighing of Base Pairings').

Press the 'MATCH' button to perform the search.

The results will be displayed within the display area, ranked according to the degree of similarity between probe string and database entries.

Name and full name of the species are shown in the first and second columns, respectively.

The number of mismatches, pairings of ambiguous residues (N), and mismatch weights are listed in columns 3 - 5, respectively.

Column 6 indicates whether the probe string (0) or its reversed complement (1) was found.

Sequence sections containing stretches identical or similar to the probe string or its reversed complement are shown between dashes in the last column.
Perfectly matched positions are indicated by double dashes, mismatches by base symbols.
The (hybrid destabilizing) quality of mismatches is indicated by upper and lower case letters.

To write the results to an ascii file press the 'SAVE' button to display the 'SELECTION BOX' window.

NOTES Unlike the 'ARB_PROBE_DESIGN' tool, the 'ARB_PROBE_MATCH' tool does not depend on the consistancy of the current and the 'PT_SERVER' database. Any 'PT_SERVER' database containing homologous or non-homologous, aligned or crude data can be searched for potential probe target matches.

BUGS Ambiguities do not work yet.
(Will be changed in future)

::

props_frame.hlp
:::
TITLE Frame Properties

OCCURRENCE ARB_NT/Properties/Menu:
ARB_NT/Tree/Neighbour joining/Properties
ARB_EDIT/Properties/Menu:
ARB_SECEDIT/Properties/Menu:

DESCRIPTION Setup of the window. Choose the 'Menu:' item from the 'Properties' menu and select colours and fonts by using the respective submenus or typing in the respective subwindows of the 'WINDOW_PROPERTIES' window.

NOTES The selected properties are saved to files '.arb/prop*' in the users home directory by choosing the 'Save Defaults' item from the 'FILE' menu.

Changes are only shown when the program is restarted.

:::
props_nds.hlp
:::
TITLE Node Display Setup (NDS)

OCCURRENCE ARB_NT/Tree/NDS
ARB_GENE_MAP/Properties/NDS

INFORMAL NOTE Read this text carefully. You won't need this function, but it offers many many new possibilities.

DESCRIPTION Extracts data from the database entries of every species and builds a user-readable string from that data.

This string can be used in many different functions, especially to show the species information at the tips of the tree or to show gene information at the tips of the genes in the gene map.

It allows you to show part of the sequences, the full_name, the accession numbers right in the tree. You may even generate tables with all kinds of useful information, like probe match results for a set of oligo probes.

Choose the 'NDS' item from the 'ARB_NT/Tree' menu to display the 'NDS' window.

Enable field extraction:
Press the corresponding 'SHOW' checkbox to display the 'field' entry at the nodes of the tree shown in the tree display area of the 'ARB_NT' window.

Select 'FIELD':

The order of the data shown at the tree nodes
(from left to right) corresponds to that in the 'NDS'
window (top to bottom). To select a 'field' enter the field name
by hand or press the respective <S> button and select it
from the pop upped list. (Press <RESCAN> to display all
fields if fields had been deleted from the list
[ARB_NT/Species/Info/Delete Fields in List])

Show field entries assigned to internal nodes: (see 'HELP:
Glossary')
Press the 'INH.' (Inherit) checkbox to display information
assigned to internal nodes of the tree ('ARB_NT/INFO
MODE') at all terminal nodes of the subtree defined by
the internal node. This function only works when using NDS
with a tree.

Define number of characters:
Type the number of characters (including blanks) to be
displayed to the 'WIDTH' subwindow.

Display predefined SRT/ACI:
The 'Search and replace Tool (SRT)' and the 'ARB Command
Interpreter (ACI)' in combination with 'Nodes Display
Setup (NDS)' allow information to be extracted and/or modified
from the 'Field' entries and displayed at the nodes.

Press the corresponding <SRT> button to display the
'SRT_ACI_SELECT' window and select a procedure
(see 'HELP: Predefined SRT/ACI).
The corresponding syntax (see 'HELP: Search and
Replace Tool' and 'HELP: ARB Command
Interpreter') is displayed in the 'ACI SRT
PROGRAM' subwindow. The displayed strings can be
modified by typing to the subwindows.

Generate new program:
Type syntax (see 'HELP: Search and Replace Tool' and
'HELP: ARB Command Interpreter') to the 'ACI SRT
PROGRAM' subwindow

NOTES !!! Strings generated using 'SRT' or 'ACI' and displayed at the
nodes cannot be stored in the database!!!

EXAMPLES @@@@@ Beispiele !!!!!!

BUGS The width of the output is limited to 4000 characters.

::
props_www.hlp
::
TITLE Standard help file form

OCCURRENCE ARB_NT

DESCRIPTION The WWW-Interface can be used to search the Web for information
taken from the ARB-Database.

EXAMPLES Here are some search examples (URL-Entries):

Retrieve from EMBL-Database:
"http://www.ebi.ac.uk/cgi-bin/emblfetch?";readdb(acc)

Search Medline for related information:
"http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=";readdb
(medline_id);"&form=6&db=m&Dopt=b"

Search the Web for full-name:
"http://search.metacrawler.com/crawler?general=%22";readdb(full_n
ame)|srt(" =+");"%22&method=0&format=1&region=0&rpp=20&timeout=15&hpe=10"

WARNINGS Look at the bottom line of the WWW configuration window
and check out whether the command line there calls your
preferred internet browser.

::
pt_server.hlp
::
TITLE PT_SERVER: What Why and How

OCCURRENCE <ARB_NT/ETC/probe*>
<ARB_EDIT/EDIT/align_sequence ..>

DESCRIPTION Probe design, probe matching, and searching the nearest relatives require a
lot of database searching.
Simply scanning the whole database for a specific target
takes several tens of seconds. Probe design and searching for
nearest relatives start thousands of search
operations. Waiting several weeks to get a result from the computer
makes every user hate the program, so better algorithms have
to be developed.

The basic ARB SEARCH ALGORITHM:

- When any (calling) program wants to scan a database, it
does not do the job itself but calls a special
SEARCH_PATTERNS_IN_A_BIG_DATABASE_PROGRAM.
We named this programm PT_SERVER (PREFIX TREE SERVER).
The PT_SERVER searches for patterns in special database files
and sends all matches back to the calling program.

- Different databases have different PT_SERVERs.
You must choose a PT_SERVER to match the database
you wish to search. The file $ARBHOME/lib/arb_tcp.dat defines
all possible choices.

- If there is no PT_SERVER running, one is automatically started.
A PT_SERVER does not scan the database of the calling
program, but the database in $ARBHOME/lib/pts/*.arb.

That means:
!!!! If you have just entered a new sequence this sequence will not be found by the the nearest-relative search. This is normally very useful, as you only want to use old sequences as a reference in the aligning process. For generating probes there is an important exception: new sequences are added to the group sequences of the pt_server.

*** {Oliver: Do you mean
"for generating probes, new sequences should be added to the PT_SERVER (see "PT_SERVER Administration)"?}******

Creating a new PT_SERVER template:

- Edit the file $ARBHOME/lib/arb_tcp.dat.
  <ARB_NT/ETC/PT_SERVER Admin/CREATE TEMPLATE>
  Note: Make a copy of your changes because a new ARB installation may reinstall $ARBHOME/lib/arb_tcp.dat.

- Restart arb

----> Create a database for an existing template:

- Start ARB with the database you want to send to the PT_SERVER.
- Open the PT_SERVER ADMIN Window <ARB_NT/ETC/PT_SERVER Admin>.
- Select a template
- Press <UPDATE SERVER> : ARB will save the DB into
  $ARBHOME/lib/pts/name.arb
- Wait (SUN Sparc 10: 10 minutes/megabyte_sequence_data)
  ARB generates an index file
  $ARBHOME/lib/pts/name.arb.pt

  If any '*.arb' file in $ARBHOME/lib/pts is newer than the corresponding '*.arb.pt' (== prefix tree) file the '*.arb.pt' file will be updated as soon as the PT_SERVER is started.

---> Updating a server: see 'Create a database for an existing template'.

NOTES - Once started a PT_SERVER never stops. The only ways to stop a PT_SERVER are:
  - SOFT KILL (everybody) (only idle PT_SERVERS):
    - <ARB_NT/ETC/PT_SERVER Admin/KILL SERVER>

  - HARD KILL (supervisor) (all servers):
    - become superuser
    - enter 'ps -auxww |grep pt_server' at any shell
    - enter 'kill -9 PID'
    (to get help enter 'man kill' or 'man ps')

- WHY DOESN'T ARB USE BLAST ?
***"Automata" is plural, "automaton" is singular (Latin, sorry) ***
The basic idea of blast is to create a finite automata for all search patterns and do the database search only once.

If someone wants to start a fuzzy search (allowing mismatches)
the size of this automata increases exponentially.
For example:
Three mismatches in a 20-base pattern would yield
about 100,000 states in the automaton per search
pattern. If probe design creates 1000 search
patterns, and every state needs 10 bytes of
computer memory, the search operation will require
1 gigabyte RAM.

- You will need a lot of swap space to run multiple PT_SERVER.

WARNINGS Do not modify any databases in $ARBHOME/lib/pts/ except
by the <PT_SERVER Admin> tools.
The reason is to keep the write protections of those files
consistent.

BUGS Sometimes some bugs

::
quit.hlp
::
TITLE Quit

OCCURRENCE ARB_NT/File/Quit

DESCRIPTION To exit ARB without saving the most recent changes press
the 'Quit' button in the 'File' menu of the 'ARB_NT' window.
This restores the database to the state it was in the last
time it had been saved.

::
rdp_ift.hlp
::
TITLE NOTES: rdp

OCCURRENCE ARB_IMPORT

DESCRIPTION This is a well designed reader for files from RDP.
It reads a lot of additional informations.

NOTES Destination fields are tagged by [RDP]

::
realign_dna.hlp
::
TITLE Standard help file form

OCCURRENCE ARB_NT @@@@@

DESCRIPTION

::
regexpr.hlp

:::

TITLE REG Regular Expressions

OCCURRENCE Everywhere

DESCRIPTION Standart Regular Expressions:

There are two possibilities to use regular expressions:

[1] /Search Regexpr/Replace String/
[2] /Search Regexpr/

[1] searches the input for occurances of 'Search Regexpr' and replaces every occurance with 'Replace String'.
[2] searches the input for the FIRST occurance of 'Search Regexpr' - if not found it returns an empty string

Excerpt from UNIX man pages:

Regular Expressions
supports a limited form of regular-expression notation, which can be used in a line address to specify lines by content. A regular expression (RE) specifies a set of character strings to match against - such as "any string containing digits 5 through 9" or "only lines containing uppercase letters." A member of this set of strings is said to be matched by the regular expression.

Where multiple matches are present in a line, a regular expression matches the longest of the leftmost matching strings.

Regular expressions can be built up from the following "single-character" RE's:

c Any ordinary character not listed below. An ordinary character matches itself.

\ Backslash. When followed by a special character, the RE matches the "quoted" character. A backslash followed by one of <, >, (,), {, or }, represents an operator in a regular expression, as described below.

. Dot. Matches any single character except NEWLINE.

^ As the leftmost character, a caret (or circumflex) constrains the RE to match the leftmost portion of a line. A match of this type is called an "anchored match" because it is "anchored" to a specific place in the line. The ^ character loses its special meaning if it appears in any position other than the start of the RE.

$ As the rightmost character, a dollar sign constrains the RE to match the rightmost portion of a line. The $ character loses its special meaning if it appears in

any position other than at the end of the RE.

^RE$ The construction ^RE$ constrains the RE to match the entire line.

\< The sequence \< in an RE constrains the one-character RE immediately following it only to match something at the beginning of a "word"; that is, either at the beginning of a line, or just before a letter, digit, or underline and after a character not one of these.

\> The sequence \> in an RE constrains the one-character RE immediately following it only to match something at the end of a "word."

[c...]
A nonempty string of characters, enclosed in square brackets matches any single character in the string. For example, [abcxyz] matches any single character from

the set `abcxyz'. When the first character of the string is a caret (^), then the RE matches any character except NEWLINE and those in the remainder of the string. For example, `[^45678]' matches any character except `45678'. A caret in any other position is interpreted as an ordinary character.

[]c...]
The right square bracket does not terminate the enclosed string if it is the first character (after an initial `^', if any), in the bracketed string. In this position it is treated as an ordinary character.

[l-r]
The minus sign, between two characters, indicates a range of consecutive ASCII characters to match. For example, the range `[0-9]' is equivalent to the string `[0123456789]'. Such a bracketed string of characters is known as a character class. The `-' is treated as an ordinary character if it occurs first (or first after an initial ^) or last in the string.

d Delimiter character. The character used to delimit an RE within a command is special for that command (for example, see how / is used in the g command, below).

The following rules and special characters allow for constructing RE's from single-character RE's:

A concatenation of RE's matches a concatenation of text strings, each of which is a match for a successive RE

in the search pattern.

* A single-character RE, followed by an asterisk (*) matches zero or more occurrences of the single-character RE. Such a pattern is called a closure. For example, [a-z][a-z]* matches any string of one or more lower case letters.

\{m\}
\{m,\}
\{m,n\}
A one-character RE followed by \{m\}, \{m,\}, or \{m,n\} is an RE that matches a range of occurrences of the one-character RE. The values of m and n must be nonnegative integers less than 256; \{m\} matches exactly m occurrences; \{m,\} matches at least m occurrences; \{m,n\} matches any number of occurrences between m and n, inclusively. Whenever a choice exists, the RE matches as many occurrences as possible.

\(...\)
An RE enclosed between the character sequences \(and \) matches whatever the unadorned RE matches, but saves the string matched by the enclosed RE in a numbered substring register. There can be up to nine such sub-strings in an RE, and parenthesis operators can be nested.

\n Match the contents of the nth substring register from the current RE. This provides a mechanism for extract-ing matched substrings. For example, the expression ^\(..*\)\1$ matches a line consisting entirely of two adjacent non-null appearances of the same string. When nested parenthesized substrings are present, n is determined by counting occurrences of \(starting from the left.

::
registration.hlp
::
TITLE Register Yourself / Bug Report

REGISTRATION Please fill in your name, department, and what do you want to do with ARB. So we can inform you about new program releases, database updates and bugs in old versions.

BUG REPORT When you find bugs in the program, (and you will), please send a short description of the bug and the conditions under which it happened. Most bugs need only an hour to fix and will automatically be part of the new version.
In most cases we will fix it and send you a notice that you can download the new ARB.
If we need more information we will send you our questions.

NOTE if arb@mikro.biologie.tu-muenchen.de does not work
try arb@BioGate.com

::
rename.hlp
::
TITLE Autorename Species

OCCURRENCE ARB_NT/Species/Rename
ARB_INTRO <MERGE TWO ARB DATABASES> ARB_MERGE <Check Consistency of Names> CHECK NAMES <RENAME DATASBASE .>

DESCRIPTION Starts the 'NAME SERVER' to update names within the specified database according to the information storted in the file '$ARBHOME/lib/nas/names.dat'.

Press the 'GO' button.

::
resorttree.hlp
::
TITLE Sort Tree Topology

OCCURRENCE ARB_NT/Tree/Sort Tree

DESCRIPTION Arranges branches and groups (triangles [radial tree] and rectangles [dendrogram]) according to depth and size.

NOTES This facilitates the comparison of tree topologies resulting from different treeing approaches.

::
result.hlp
::
TITLE ARB_HELP_SEARCH

OCCURRENCE anywhere in ARB-Help

DESCRIPTION The subtopics contain all helpfiles in which the specified word was found in.

::
rl_dna_prot.hlp
::
TITLE Realign DNA Sequences According Amino Acids Alignment.

OCCURRENCE ARB_NT/Sequence/Realign DNA According to Aligned Pro.

DESCRIPTION Adapts nucleic acid sequence alignments to that of the corresponding amino acid sequences.
Select DNA and protein alignments from the respective subwindows press the 'REALIGN' button.

::
rst_log_zoom.hlp
::
TITLE Reset Logical Zoom

OCCURRENCE ARB_NT/ETC/Reset Logical Zoom

DESCRIPTION Displays the full tree.
To display the full tree choose the 'Reset Logical Zoom'
item from the 'ETC' menu of the 'ARB_NT' window.

NOTES To display subtrees (logical zoom) activate the 'LOGICAL
ZOOM MODE' by clicking on the <Zoom> button of the 'ARB_NT'
window or pressing the 'F5' key on the keyboard. In the latter
case, the cursor has to be placed within the tree display area.

::
rst_phys_zoom.hlp
::
TITLE Reset Physical Zoom

OCCURRENCE ARB_NT/ETC/Reset Physical Zoom

DESCRIPTION Restores the initial scale of the tree display.

NOTES To magnify sections of the tree display (physical zoom) activate
the 'ZOOM MODE' by clicking on the <PZoom> button of the
'ARB_NT' window or pressing the 'F4' key on the keyboard. In the
latter case, the cursor has to be placed within the tree display
area.

EXAMPLES To restore the initial scale of the tree display choose the
'Reset Physical Zoom' item from the 'ETC' menu of the
'ARB_NT' window.

::
save.hlp
::
TITLE Save Database

OCCURRENCE ARB_NT/File/Save

DESCRIPTION To save the entire database, choose 'Save Whole Database'
from the 'File' menu of the 'ARB_NT' window.

You may also save only the changes to the database
with the 'Save Changes' or 'Save Changes as'
functions. 'Save Changes' saves all changes to the database
'data.arb' to a file 'data.aXX' where X is a number
between 0 and 99 starting with 1. {**huh? Between 1 and 99?**}
This is normally much faster than saving the entire database and
results in much less disc consumption.
Each time you save your changes, X is increased by one.

The last 5 old data.aXX are not deleted and allow you to
step to an old state. All data.aXX are independent and you may
delete old changes files without losing data (as long as the
latest is unchanged)
When you start arb you can choose among the different files:
enter 'arb filename.arb' to search and load the latest changes,
enter 'arb filename.aXX' to load a user-defined version.

BUGS Not all changes to the database can be saved into a small file -
especially, changing the order of database entries disables
a quick save.
Hopefully no real bugs.

::
save_def.hlp
::
TITLE Save Defaults

OCCURRENCE ARB_NT/File/Save Defaults
ARB_NT/Tree/Neighbour joining/File/Save Defaults
ARB_NT/ARB_EDIT/File

DESCRIPTION Saves the userdefined setup (colours, fonts, ...) to files

::
save_matrix.hlp
::
TITLE Save Matrix to Disc

OCCURRENCE ARB_DIST @@@

DESCRIPTION

FORMATS Phylip Format: Format to be reused in Phylip Programs
Readable (using NDS): Use NDS to replace the short names
by a user defined selection of fields
and generate a pretty view of the
matrix

WARNING: If you select to many fields by nds your matrix will be hardly
readable

BUGS: You cannot select the number of digits after '.'.

::
saveas.hlp
::
TITLE Save As

OCCURRENCE ARB_NT/File/Save As

DESCRIPTION Copies the current version of the database including any
changes that have been made since the last save to the name
specified.

Choose the 'Save as' option from the 'File' menu to display the
'SAVE ARB DB' window.

Select a file name from the 'Directories and Files' subwindow or
type the full path and file name to the subwindow
'FILE NAME:', press the <SAVE> button and wait until the
'SAVE ARB DB' window disappeares.

NOTES The suffix shown in the 'SUFFIX' subwindow is appended to the
file name. It has to be '.arb' to make the file readable to
the ARB programs.

WARNINGS !!! All local file-permissions are removed.!!!

::
savedef.hlp
::
TITLE Save Defaults

OCCURRENCE ARB_NT/File/Save Defaults
ARB_NT/Tree/Neighbour joining/File/Save Defaults
ARB_NT/ARB_EDIT/File
ARB_EDIT4/Properties/Save defaults
ARB_SECEDIT/Properties/Save defaults
ARB_GENEMAP/Properties/Save defaults

DESCRIPTION Saves the userdefined setup (colours, fonts, ...) to files
'~/.arb_prop/*' located in the users home directory.

NOTES To reset ARB to it's defaults simply remove one/all *.arb
files from the directory '~/.arb_prop'

The files in ~/.arb_prop are textfiles. Advanced users may
remove or change single sections or entries.

You can hand over your properties to another user by giving him
your property-files.

WARNINGS Better make backups of these files before you change them.

::
scandb.hlp
::
TITLE Scan Database for all Fields

OCCURRENCE ARB_NT/Species/Info/FIELDS/Scan Database for all fields
ARB_NT/Genes/Info/FIELDS/Scan Database for all fields

DESCRIPTION This function scans the database for all existing fields
in the species/genes database.
Normally it is not necessary to run this procedure unless
some external programs generate new and yet unknown fields.

NOTES May take a long time.

::
sec_mode.hlp
::
TITLE ARB_SECEDIT modes

OCCURRENCE ARB_SECEDIT/mode buttons

DESCRIPTION ZOOM MODE Mode to zoom the displayed secondary structure.

You can zoom in by dragging a rectangle with the
left mouse button.
You can zoom out with the right mouse button.

HELIX MODE Mode to build and destroy helix regions.

Left mouse button converts the clicked loop region
into a helix region (if possible).
Right mouse button removes the clicked helix region.

SET ROOT MODE Mode to set the root of the secondary structure.

Left click on the loop which should be the root of your
secondary structure. The root loop will not move if you
create or remove helix region nor if you rotate branches
of the secondary structure.

ROTATE MODE Mode to rotate single branches of the structure.

@@@

CONSTRAINT MODE @@@

::
secedit_imexport.hlp
::
TITLE Export & import secondary structure

OCCURRENCE ARB_SECEDIT/Files/Import Structure
/Export Structure

DESCRIPTION Export secondary structure to file and import secondary
structure from file.

WARNINGS After you have changed the length of the alignment, it's no
longer possible to import your exported secondary structure.

::
security.hlp
::
TITLE Protection Level

OCCURENCE ARB_NT/Protection

ARB_EDIT/EDIT/Set Protection Level ...

DESCRIPTION An individual protection level (0 - 6) can be assigned to all types of database entries (sequences and additional information stored in a particular 'field').
To modify any entries, a protection level has to be selected from the 'Protection' menu of the main window equal to or higher than that assigned to the data.

Default Protection Levels:

Sequence names: 5

You can see the Protection level of fields in the species/gene information window.

WARNINGS It is recommended to reset the protection level after performing operations to prevent unintentional modification or loss of data.

::
sel_box.hlp
::
TITLE Write to File

OCCURRENCE ARB_NT/ETC/Probe Design/PROBE DESIGN <RESULT/SAVE>

DESCRIPTION Writes data to an ascii file.

Select a file name from the 'Directories and Files' subwindow of
the displayed 'SELECTION BOX' or
Type a file name to the 'File Name' subwindow.
A suffix typed or displayed in the 'Suffix' subwindow
is used as a filter for the file names to be displayed
and is automatically appended to a file name typed.
The length (number of ranked probe or target proposals) can be
defined by selecting a value from the menu displayed
after pressing the 'How many Lines' button.

::
sel_fil.hlp
::
TITLE Select Filter

OCCURRENCE ARB_NT/Tree/Neighbour joining/Select Filter

DESCRIPTION Any sequence of symbols stored as 'sequence associated information' ('SAI') can be selected from the 'Select a Filter' subwindow and used as a filter for the in or exclusion of alignment columns for treeing.
If any species is selected, you may use it's sequence for filter. It's name will be displayed just before the SAI names.

The characters of the filter which define columns to exclude

have to be defined by typing to the respective subwindow.

The selected filter is displayed in the subwindow on the bottom.

All bases may be simplified, leaving only transversions
and simplidfied amino-acid-groups, allowing transversion
parsimony/fdnaml/distmethods

NOTES Any nucleotide sequence can be copied to SAI (sequence associated information) (ARB_NT/Species/Info/SPECIES/Convert to SAI) and then used as a filter.

EXAMPLES Include only positions which are occupied by a residue within the E. coli sequence:

1. Select 'ECOLI' from the 'Select a Filter' subwindow.

2. Type non-nucleotide symbols to the 'Exclude Column' subwindow (.-).

Include only non-base paired positions:

1. Select 'HELIX' from the 'Select a Filter' subwindow.

2. Type base pair symbols to the 'Exclude Column' subwindow ([<>]).

Include only positions which have been unambiguously determined within a particular sequence:

1. Select the species and convert it to 'SAI' (ARB_NT/Species/Info/SPECIES/Convert to SAI)

2. Select the new 'SAI' from the 'Select a Filter' subwindow.

3. Type non-nucleotide and ambiguoity symbols to the 'Exclude Column' subwindow (.-acguRYS ....).

::
sel_spec.hlp
::
TITLE Select species

OCCURRENCE ARB_NT/Species/Search

DESCRIPTION An Individual species can be selecetd from 'Hit list'.
The 'SPECIES INFO' window showing all 'DATABASE FIELDS' and the corresponding entries is initiated or updated.
The 'name' of that species appears within the species button of the main window (fourth broad rectangular button in the upper part) and is also highlighted in an activated sequence editor if present.

NOTES This tool is helpful to mark or unmark individual species using the 'SPECIES INFO' window and to rapidly find its position in the tree edited in the main window pressing the 'PSEARCH' button of that window.

::
selected.hlp
::
TITLE Selected Species and Cursor Position

OCCURRENCE ARB_NT ARB_EDIT ARB_PARS

DESCRIPTION An individual species (not SAI) can be selected. That means:

1. Its database entries are displayed in the 'SPECIES INFORMATION' window <ARB_NT/Species/Info>

2. ARB_PARS can insert the selected species into an existing tree.

3. The sequence of the selected species can be aligned <ARB_EDIT/EDIT/Align Sequence>.

4. It can be used for filter generation.

The selected species is shown at
<ARB_NT/4th big button in top area>.

There are several ways to select a species:

1. Press <ARB_NT/Species/Search>
Search for species
Select a species

2. Press <ARB_EDIT/ETC/Synchronize Cursor Position>
Select one species in the editor.

3. Double click mouse at species name in the arb sequence editor

::
selected_gene.hlp
::
TITLE Selected gene

OCCURRENCE ARB_NT/Genes/Info

DESCRIPTION An individual gene can be selected.

There are several ways to select a gene:

1. Search for a gene using ARB_NT/Genes/Search and then select one gene from the result list.

NOTE: Selecting a gene of another species changes
the current species!

2. Click on a gene in ARB_NT/Gene Map

When a gene is selected the content of the GENE INFORMATION window changes.

::

sequence_colors.hlp

::

TITLE Change Colors Of Sequences

OCCURRENCE ARB_EDIT4

DESCRIPTION The ARB_EDIT4 editor uses two tables to translate the original sequence into a colored displayed sequence:

First Step: Each character in each sequence is translated into a new character and a color index (0..9) using the <OPTIONS/Sequence Colors> options. This window shows different sets (S_0 .. S_#) of color shemes, one coloumn for one set. In the first row you may choose one set which is used for translation.

The first coloumn shows the character which should be translated/replaced. The nth + 1 coloumn holds the data for translation set n. Each of its fields has two characters:
1. the character which should replace the original value, or '=' if no translation should be performed.
2. A color index for this character, between 0 and 9.

Second Step:
To get the real color, the color index is translated to a color using the <Properties/Data> window.

EXAMPLES You may use this feature to show:

- A simplified version of your amino acid alignment.
- Only YR instead of ACGTU
- Only ambigous symbols
- ...

::

set_meny_fields.hlp

::

TITLE Standard help file form

OCCURRENCE ARB_NT/Species/Search/WRITE TO FIELDS OF LISTED SPECIES

DESCRIPTION Sets one field in all species shown in the hitlist of the species search window.

EXAMPLES All species which have no accession number should get

a dummy string 'no accession':

1. Open the species search window
2. Search species that don't match the query
   Search field = acc
   Search string = *
   Press 'Search'
3. Press WRITE TO FIELDS OF LISTED SPECIES
3.a Select field name = 'acc'
3.b Field value = no accession
3.c Press SET ALL EMPTY FIELDS

WARNINGS You may easily destroy valuable data.

::
set_protection.hlp
::
TITLE Set Protection Level of Field of Listed Species

OCCURRENCE <ARB_NT/Species/Search/Do On Listed/Set Protection ...>

DESCRIPTION Set the protection level of one field of listed species.

EXAMPLES Disable editing of the full_name field of all entries which came from GENBANK/EMBL:

1. Search for species with an accession number:
   Press <ARB_NT/Species/Search>
   Select: Search Species that match the query
   Search Field: 'acc'
   Search String: '*'
   Press 'SEARCH'
2. Press <Do_On_Listed/ Set Protection ..>
   Select 'full_name'
   Select '6'
   Press 'PROTECT'

WARNINGS see BUGS

BUGS You can only change the protection level of existing entries. New entries will automatically get the p. level 0.

::
sp_count_mrk.hlp
::
TITLE Count Marked Species

OCCURRENCE ARB_NT/Species/Count Marked Species

DESCRIPTION Displays the number of marked species in the 'MESSAGE' window.

::
sp_del_mrkd.hlp
::

TITLE Delete Marked Species

OCCURRENCE ARB_NT/Species/Delete Marked Species

DESCRIPTION Deletes all marked species entries.

NOTES To check the number of marked species choose 'Count Marked Species' from the 'ARB_NT/Species' menu.

A protection level has to be set on the 'ARB_NT' window equal or higher than that assigned to the species entries.

::
sp_info.hlp
::

TITLE SPECIES INFORMATION

OCCURRENCE ARB_NT/<INFO> button: left area, fifth from top
ARB_NT/Species/Info
ARB_NT/Species/Search
ARB_NT/Editor

DESCRIPTION Displays species information stored within the 'fields' (see 'HELP: Glossary'). The particular 'species' (see 'HELP: Glossary') can be 'marked' or 'unmarked' (see 'HELP: Glossary') by pressing the checkbox after the 'Marked?' prompt.
Editing of 'field' entries is enabled or prevented by pressing the 'Edit enabled?' Checkbox.
The entries of a 'field' are modified by choosing it from the 'DATABASE FIELDS' subwindow and modifying the entries displayed in the 'Edit box' subwindow.

NOTES The 'SEARCH' window can be displayed by pressing the <SEARCH> button.

For modification of 'field' entries, a protection level has to be selected from the Protection menu of the main window (ARB_NT/Protection) equal to or higher than that assigned to the selected 'field'

Cut and paste of the window system can be used in the 'Edit box' subwindow. This provides is an easy way to export/import sequences.

EXAMPLES Update the 'full_name' 'Pseudomonas cepacia' to 'Comamonas cepacia':
Select FIELDS: 'full_name'
Move cursor into the 'Edit box' subwindow
Delete: 'Pseudomonas'
Type: 'Comamonas'

WARNINGS It is recommended to reset the protection level after modifying entries to prevent unintentional modification or loss of data.

::
sp_info_species.hlp
::
TITLE Standard help file form

OCCURRENCE ARB_NT

DESCRIPTION

::
sp_invert_mrk.hlp
::
TITLE Swap Marked and Unmarked Species

OCCURRENCE ARB_NT/Species/Swap Marked Species

DESCRIPTION Interchanges the status of marked and unmarked species.

NOTES To check the number of marked species choose 'Count Marked Species' from the 'ARB_NT/Species' menu.

::
sp_mrk_all.hlp
::
TITLE Mark all Species

OCCURRENCE ARB_NT/Species/Mark all Species

DESCRIPTION Marks all species entries of the database.

NOTES To check the number of marked species choose 'Count Marked Species' from the 'ARB_NT/Species' menu.

::
sp_mrk_tree.hlp
::
TITLE Mark Species in Tree

OCCURRENCE ARB_NT/Species/Mark Species in Tree

DESCRIPTION Marks all species of the tree currently shown in the tree display area of the 'ARB_NT' window.

NOTES To check the number of marked species choose 'Count Marked Species' from the 'ARB_NT/Species' menu.

WARNINGS All species of the tree are marked not only those visible on the screen (zoom!)

Marked species which are not represented in the current tree remain marked!

::

sp_rename.hlp

TITLE Create Names

OCCURRENCE ARB_NT/Species/Create names

DESCRIPTION Creates automatically unique names (=identifiers) for the species entries in the database. The entries are identified by their accession numbers (public databases). The names are given using the 'full_name' information. Usually, the first three letters are taken from the genus designation, the remaining letters from the species name.

If there are duplicated entries (same accession number - different 'full_name'; no accession number - same 'full_name') the different versions are indicated by appending running numbers separated from the 'name' by a dot.

Press the 'Rename' button of the 'Species' menu to display the 'AUTORENAME SPECIES' window.

NOTES The names are stored with the database. The names can be changed by the user (ARB_NT/Species/Info/SPECIES/Rename). However, the names are protected to prevent to assign the same name to different sequences.

It is recommended to assign accession numbers to species entries (sequences) which are not from public databases.

sp_search.hlp

TITLE Search Database for Species

OCCURENCE ARB_NT/Species/search
ARB_NT/<species button: 4th broad rectangular button in top area>

DESCRIPTION Searches for a (set of) species (not SAIs) that match (dont match) a query or are marked.

The database is scanned for 'species' (see 'HELP') which contain (or do not contain) the search string within the specified 'field' (see 'HELP: Glossary'). The corresponding species and the respective 'field' entries are listed in the 'HIT LIST' subwindow. The number of hits is displayed after the 'Hits:' prompt.

Define whether matching or non-matching species should be listed by pressing the appropriate combination of left and right buttons in the top area. When performing multiple searches, define whether the list of 'species' should be replaced by the new results, and whether newly

found 'species' should be removed from or appended
to the existing list.

Select a 'field' from the 'Search Field' subwindow.

Type the search string in the 'Search string' subwindow.

Press the <SEARCH> button.

EXAMPLES 1. Search for a species called 'Pseudomonas tolaasii'
Select: - List all species that
- match the query
Select search field: 'full_name'
Type search string: 'pseu*tol*'
* and ? are wild cards for multiple and
single characters, respectively
Press: 'SEARCH' to start the operation

2. Search for all species that are marked:
Select: - List all species that
- are marked
Press: 'SEARCH' to start the operation

3. Search for all species that are marked and for which
an entry is present in the field 'reference'

Select: - List all species that
- are marked
Press: 'SEARCH' to start the operation 1

Select: - Keep species that
- match the query
Select search field: 'reference'
Type search string: '?*' (More than one character)
Press: 'SEARCH' to start the operation 2

WARNINGS If the hitlist becomes too long, it will be truncated.

::
sp_sort_fld.hlp
::
TITLE Sort Species According Database Entries

OCCURRENCE ARB_NT/Species/Sort Species According Database Entries

DESCRIPTION Arranges the species entries according to species associated
field entries. Up to three fields can be specified in a
hierarchical order. The arrangement is done in the alphabetical
or numerical order of the field entries.

Choose the 'Sort Species According Database Entries' item from
the 'ARB_NT/Species' menu to display the 'SORT DATABASE' window.
Select the fields from the 'Primary, Secondary' and 'Last Sort
Key' subwindows.

::
sp_sort_phyl.hlp
::
TITLE Sort Species According to Phylogeny

OCCURRENCE ARB_NT/Species/Sort Species According to Phylogeny

DESCRIPTION Arranges the species entries in the order of their ranking within the tree currently shown in the tree display area of the 'ARB_NT' window

NOTES This order is used by other tools displaying species data (sequence editor, pretty print, export foreign format)

::
sp_sp_2_ext.hlp
::
TITLE Convert Species to SAI

OCCURRENCE ARB_NT/Species/Info/SPECIES/Convert to SAI

DESCRIPTION Converts the sequence to a SAI entry.

Choose the 'Convert to SAI' item from the 'SPECIES' menu.

NOTES Converting a species entry to a SAI entry allows to use the sequence as a filter for treeing and other procedures.

::
sp_umrk_all.hlp
::
TITLE Unmark all Species

OCCURRENCE ARB_NT/Species/Unmark all Species

DESCRIPTION Unmarks all species entries of the database.

NOTES To check the number of marked species choose 'Count Marked Species' from the 'ARB_NT/Species' menu.

::
sp_umrk_tree.hlp
::
TITLE Unmark Species in Tree

OCCURRENCE ARB_NT/Species/Unmark Species in Tree

DESCRIPTION Unmarks all species of the tree currently shown in the tree display area of the 'ARB_NT' window.

NOTES To check the number of marked species choose 'Count Marked Species' from the 'ARB_NT/Species' menu.

WARNINGS All species of the tree are unmarked not only those visible on the screen (zoom!)

Marked species which are not represented in the current tree remain marked!

::
spa_copy.hlp
::

TITLE Copy Species/Gene

OCCURRENCE ARB_NT/Species/Info/SPECIES/Copy
ARB_NT/Gene/Info/SPECIES/Copy

DESCRIPTION Copies a species/gene entry.

Choose the 'Copy' item from the 'SPECIES/GENE' menu to display the 'SPECIES/GENE COPY' window. Type a name to this window and press <GO>

NOTES In case of copying species the ARB name server ensures that no duplicated names are created.

To overwrite the entries of an existing species/gene a protection level equal or higher than that assigned to this entry has to be set on the 'ARB_NT' window

::
spa_create.hlp
::

TITLE Create Species/Gene

OCCURRENCE ARB_NT/Species/Info/SPECIES/Create
ARB_NT/Genes/Info/SPECIES/Create

DESCRIPTION Creates a new species/gene entry.

Choose the 'Create' item from the 'SPECIES/GENE' menu to display the 'SPECIES/GENE CREATE' window. Type a name to this window and press <GO>

NOTES In case of creating a new species the ARB name server ensures that no duplicated names are created.

::
spa_delete.hlp
::

TITLE Delete Species/Gene

OCCURRENCE ARB_NT/Species/Info/SPECIES/Delete
ARB_NT/Genes/Info/SPECIES/Delete

DESCRIPTION Deletes a species/gene entry.

Choose the 'Delete' item from the 'SPECIES/GENE' menu.

NOTES To delete a species/gene entry, a protection level equal or higher than that assigned to this entry has to be set on the 'ARB_NT' window.

::
spa_rename.hlp
::
TITLE Rename Species/Gene

OCCURRENCE ARB_NT/Species/Info/SPECIES/Rename
ARB_NT/Genes/Info/SPECIES/Rename

DESCRIPTION Renames a 'species/gene' (see 'HELP: Glossary) entry.

Choose the 'Rename' item from the 'SPECIES/GENE' menu to display the 'SPECIES/GENE RENAME' window. Type a name to this window and press <GO>

NOTES In case of renaming species the ARB name server ensures that no duplicated names are created.

To change the name of 'species/gene' entries, a protection level equal or higher than that assigned to this entry has to be set on the 'ARB_NT' window.

::
spaf_create.hlp
::
TITLE Create a Field

OCCURRENCE ARB_NT/Species/Info/FIELDS/Create a field
ARB_NT/Genes/Info/FIELDS/Create a field

DESCRIPTION Allows to create new fields.

Choose the 'Create field' item from the 'FIELDS' menu to display the 'CREATE A NEW FIELD' window. Type a name to the 'FIELD NAME' subwindow, define the type by pressing the corresponding button and press <GO>.

NOTES If the field name contains a / a hierarchical key is created.

::
spaf_delete.hlp
::
TITLE Delete Fields

OCCURRENCE ARB_NT/Species/Info/FIELDS/Delete fields in list
ARB_NT/Genes/Info/FIELDS/Delete fields in list

DESCRIPTION Allows to delete/hide fields.

Choose the 'Delete field in list' item from the 'FIELDS' menu to display the 'DELETE FIELD' window. Select a field from the 'Fields' subwindow and press the HIDE FIELD>
(the field and its entry are no loger displayed in the 'SPECIES/GENE INFORMATION' window) or the <DELETE FIELD> (field and its entry are deleted) button.

NOTES A protection level has to be set on the 'ARB_NT' window equal or higher than that assigned to the species entry.

::
spaf_reorder.hlp
::
TITLE Reorder Fields

OCCURRENCE ARB_NT/Species/Info/FIELDS/Reorder fields
ARB_NT/Genes/Info/FIELDS/Reorder fields

DESCRIPTION Allows changing the order of the fields within the 'DATABASE FIELDS' subwindow of the 'SPECIES/GENE INFORMATION' window.

Choose the 'Reorder fields' item from the 'FIELDS' menu to display the 'REORDER FIELDS' window. Select fields from the subwindows and press the 'PUT SELECTED ITEM ..' to change the fields display on the 'SPECIES/GENE INFORMATION' window.

::
spaf_scandb.hlp
::
TITLE Show all Fields

OCCURRENCE ARB_NT/Species/Info/FIELDS/Show all fields

DESCRIPTION To display all fields and entries stored in the database, press the 'Show all fields' button of the 'FIELDS' menu.

::
species.hlp
::
TITLE What are Species ?

DESCRIPTION One species is one database entry conaining many database entries describing the species and one or more sequences. Every sequence belongs to a different alignment:

[Database field name] [Comment]

name Unique name for a species
max 8 characters long

acc Accession number

full_name Name for a species

strain

author

journal

ali_16s/data a 16s sequence (part of the 16s alignment
)
ali_23s/data a 23s sequence
ali_.../data a ... sequence

The description of the alignment, e.g. type or length, is not stored in every
species but in an alignment description.

::
species_info.hlp
::
TITLE View and edit species information

HOW TO OPEN - <ARB_NT/MODE INFO>
- <ARB_NT/species/search ... select a species>
- <ARB_NT/species/info>

DESCRIPTION This window allows you to show and modify species information.
To change an entry, select it from the bottom box and
edit it in the 'Edit box'.

NOTES The edit box can be used as a cut and paste field.
It is a very easy way to export/import sequences.

EXAMPLES

::
st_ml.hlp
::
TITLE Column Statistic (Prototype)
Detailed column statistic (see below)

OCCURRENCE ARB EDITOR

DESCRIPTION Highlites unlikely bases in an alignment using the maximum
likelihood technique.
As soon as the go button is pressed, the selected tree (by
ARB_NT) and all marked sequences are read. Then the relative
likelihood for each base is calculated and transformed to
a number between 0 and 9. This number is translated into a color,
using the colors 'RANGE 0 ... RANGE 9'. The higher the number,
the more unlikely a particular base.
You may improve the output by selecting a valid column statistic,
which holds information about column dependent rates and base
frequencies.

NOTES The colors are not set correctly. Please set different 'CS' colors
<Props:Sequences: Colors and Fonts> and save them.
Without a powerfull computer only a small number of sequences can
be viewed.
The program assumes that your tree is correct.

WARNINGS This is only a prototype, don't expect something perfect.
All sequences which should be analysed should be marked and
in the tree shown by ARB_NT !!!!!

BUGS The colors are not set correctly by default.
The programm can only be started once.

DETAILED COLUMN STATISTIC This special mode displays 4 rows below the sequence containing the likelihood of each base character.

Each row consists of two rows of digits displayed in the same color (which actually is the color normally used for displaying the appropriate base character). The upper of these two rows is the first, the lower the second digit of the likelihood.
[You may like to use the cursor to simplify reading]

Special character used:

SPACE = likelihood is 20%
^ = likelihood is 100%
? = can't determine likelihood for that column

Background colors used:

Normally the normal background color is used.

If a column has a significant likelihood for one base character (or for the sum of two base characters), all four rows are displayed in color 'Range 0'.

[significance is 90% (hardcoded) - will be made utilisable soon]

The single (or two) base character(s) responsive for the significance will be displayed in colors 'Range 1' to 'Range 8' (the higher the Range-number is, the higher is the likelihood)

::
submission.hlp
::
TITLE Write to Submission Forms

OCCURRENCE ARB_NT/Species/Submission

DESCRIPTION !!! Currently no help available.!!!

NOTES This is one of the oldest windows of arb and therefore
badly designed.

@@@@

::
sv_def.hlp
::
TITLE Save Defaults

OCCURRENCE ARB_NT/File/Save Defaults

DESCRIPTION Saves changes of the setups (font, colour, ...) to the file
'.arb_prop.ntree' located in the users home directory

::
tags.hlp
::
TITLE TAGS: Subfields

OCCURRENCE ARB_NT: Merge Tools / Modify Database Fields

DESCRIPTION Tags are used to subdivide fields into subfields, which
often have the same value. E.g. After merging the 16s databases
from RDP and DeWachter all species have two full_names:
The RDP and DeWachter version which should be equal but which
are often not. So we 'TAG' the dewachter database using the
tag 'DEW' and the RDP with the tag 'RDP'. Say one species has
the full_names escherichia_RDP and escherichia_DEW.
Using the tags mechanism the final field will look like this:
' [DEW] escherichia_DEW [RDP] escherichia_RDP'.
Tags are sourrounded by brackets and put in front of the
corresponding field value. If both subfields have the same
value, (like escherichia_coli) fields are merged:
' [DEW,RDP] escherichia_coli'

NOTES The Modify Database Fields tools allows to modify only single
subfields.
If there are no tags used yet, the default tag is taken.

BUGS Square Brackets in the fields are replaces by '{}'

::
tgroupall.hlp
::
TITLE Group All

OCCURRENCE ARB_NT/Tree/Group All

DESCRIPTION Shows the ranking subtrees (groups of species; see 'HELP: GROUPE
MODE') of the currently displayed tree as triangles (radial
tree) or rectangles (dendrogram).

NOTES The lengths of the sides connected by the internal node reflect
the longest and shortest overall branch lengths within the
subtree.

::
tgroupnmrkd.hlp
::
TITLE Group All Except Marked

OCCURRENCE ARB_NT/Tree/Group All Except Marked

DESCRIPTION Shows the ranking subtrees (groups of species; see 'HELP: GROUPE MODE') of the currently displayed tree which do not contain marked species as triangles (radial tree) or rectangles (dendrogram).

NOTEST The lengths of the sides connected by the internal node reflect the longest and shortest overall branche lengths within the subtree.

::
tkeep_mrkd.hlp
::
TITLE Keep Marked

OCCURRENCE ARB_NT/Tree/Keep Marked

DESCRIPTION Removes all unmarked species and zombies from the currently displayed tree.

WARNINGS !!! No 'undo' function available yet !!!

It is rekommended to copy the tree before modifying it.

::
tr_export.hlp
::
TITLE Export Tree to File

OCCURRENCE ARB_NT/Tree/Copy\Delete\Export\Import/EXPORT

DESCRIPTION Writes a tree to file in Newick formate

Press the 'EXPORT' button to display the 'TREE SAVE' window.

Select a tree file from the 'Directories and Files' subwindow or type the file name to the 'FILE NAME' subwindow.

Select 'PLAIN FORMAT' or 'USE NDS' after pressing the 'PLAIN FORMAT' button to write a Newick tree containg the 'names' only or the full information displayed at the terminal nodes of the specified tree, respectively.

PLAIN FORMAT Include only short names in the newick tree

USE NDS Export only. Replace short names by nds selected fields.

ORS TANSFER BINARY FORMAT

Special format used by our database administrators.
You need a special program to use it.

NOTES If a suffix is displayed in or typed to the 'SUFFIX' subwindow, only the corresponding file names will be displayed. The suffix is automatically appended to the file name typed to the 'FILE NAME' subwindow.

WARNINGS !!! Tree data written with the 'USE NDS' option cannot be reimported to the database.!!!

::
tr_import.hlp
::
TITLE Load a Tree

OCCURRENCE ARB_NT/Tree/Copy\Delete\Export\Import/IMPORT

DESCRIPTION Allows to import a tree written in Newick formate.

Press the 'IMPORT' button to display the 'TREE LOAD' window.

Select a tree file from the 'Directories and Files' subwindow or type the file name to the 'FILE NAME' subwindow.

Specify a tree name ('tree_*') by typing it to the 'tree_name:' subwindow.

Press the 'LOAD' button.

NOTES If a suffix is displayed in or typed to the 'SUFFIX' subwindow, only the corresponding file names will be displayed. The suffix is automatically appended to the file name typed to the 'FILE NAME' subwindow.

WARNINGS !!! The names associated with terminal nodes have to be consistent with those of the current database.!!!

::
tr_jump.hlp
::
TITLE Search & Logical Zoom

OCCURRENCE ARB_NT/ETC/<JUMP button>

DESCRIPTION Searches the selected species in the tree shown in the tree display area of the 'ARB_NT" window and displays the subtree containing the selected species.

NOTES The selected species is indicated by an open square at the respective terminal node of the tree shown in the tree display area of the 'ARB_NT" window.

::

tr_type_list.hlp

::

TITLE SHOW DENDROGRAM

OCCURRENCE ARB_NT<second small rectangular button in top area>

DESCRIPTION Trees can be displayed as dendrograms or radial trees.
To display a dendrogram press the button <second small
rectangular button in top area>

::

tr_type_radial.hlp

::

TITLE SHOW RADIAL TREE

OCCURRENCE ARB_NT<first small rectangular button in top area>

DESCRIPTION Trees can be displayed as dendrograms or radial trees.
To display a radial tree press the button <first small
rectangular button in top area>

::

trans_anal.hlp

::

TITLE Transversion Analyses

OCCURRENCE in all filter windows

DESCRIPTION There is no special transversion analysis. But using the filter
nearly all programs allow to convert DNA/RNA sequences to
A/G sequences on the fly, simply by using a filter and
setting the <simplify your data> selector to TRANSVERSIONS ONLY

WARNINGS Don't use the Jukes Cantor distance transformation, use
Felsenstein instead

BUGS maybe/ maybe not

::

translate_dna_2_pro.hlp

::

TITLE Translate DNA to Protein

OCCURRENCE ARB_NT/Sequence/Translate

DESCRIPTION Translates nucleic acid sequences. The alignment of the amino
acid sequences is adapted to that of the nucleic acids. The one
letter code is used.

1. Select source and destination alignment from the respective
subwindows.

2. Select reading frame by pressing the 'Start at alignment
position' button and selecting first, second or third

absolut position (alternatively you can position the cursor
in ARB_EDIT4 at start of the reading frame).

3. Press the 'TRANSLATE' button.

Example:
DNA: ---UGG...GUAUGGUUA
-> PRO: -Y.LYG

WARNING: The program does begin at the first three bases, but at the first three alignment positions. That means that all your three letter codons should start at every third position.

Example: (### codon for aminoacid #)

DNA: ...111...222...333444
PRO: .1.2.34

DNA: ..111...222....333444
PRO: XXXX.34 // 1 2 are out of sync

DNA: ...111...22.2..333444
PRO: .1.XX34 // bad alignment for 2

WARNINGS see above

::
tree2file.hlp
::
TITLE Print to File

OCCURRENCE ARB_NT/Tree/Print to File

DESCRIPTION Exports a ONE page screenshot of the displayed tree to a file. Different file formates can be written.

Select:
- Language:
  Press the button after the 'Language' prompt
  and select from the submenu.
- Orientation:
  Press the button after the 'Orientation' prompt
  and select from the submenu.
- Magnification (%) (Postrscript only):
  Type number to the 'Magnification' subwindow
- Print What:
  Press the button after the 'Orientation' prompt
  and select from the submenu.
  Screen = Tree or section of the tree
  shown in the tree display window
  Total Tree = Full tree

Select a file name from the 'Directories and Files' subwindow

or type it to the 'File Name' subwindow.

Save and/or edit the data:
Press one of the buttons in the last line of the
'Export TREE TO FILE' window

- SAVE: exports the data
- S & XFIG Writes the data to the specified file and edits the file. (xfig language has to selected)
- S & GHOST Writes the data to the specified file and displays the preview. (postscript language has to be selected). The previewer allows tree printing.

NOTES The suffix shown or typed in the 'Suffix' subwindow is used as a filter for the displayed file names and is automatically appended to the file name in the 'File Name' subwindow.

Ghostview and xfig are public domain software.

BUGS Magnification and orientation is only used for postscript output.

::
tree2prt.hlp
::
TITLE Print a Graphic to a Printer

OCCURRENCE ARB_NT/Tree/Print to Printer

DESCRIPTION Multi Page Printer

*** Clip at Screen: Show only those graphics that are drawn at the ARB_NT main window

Show Handles: Show/Hide root and mark symbols

*** Graphic Size: X * Y size of Graphic in inches.
When you change the tree while this window stays open
(eg. zoom in/out ...) press
<Get Graphic Size> to update graphic size

Magnification:

Paper Size: X * Y inches

Orientation: Landscape or Portrait mode

Pages: X * Y Number of resulting pages
0.7 2.3 means 1*3 == 3 pages
If you modify the X/Y page field,

magnification will be adjusted to fit
the graphic into the number of pages.

Page Range: Not implemented

*** Destination: Printer: Use print command to print
File: Use File Name as destination file name
Preview:

NOTES

EXAMPLES Print a long list tree:
1. Select list tree style in ARB_NT main window
2.

WARNINGS printing cannot be stopped

BUGS This print utility is based on 'fig2dev', which sometimes generates too many pages.
Use preview to check the printout, and in case of many empty pages use preview.

::
tree_cmp.hlp
::
TITLE Copy node info

OCCURRENCE ARB_NT/Tree/Copy_Del.../Copy Node Info

DESCRIPTION Move the internal group labels from one tree to another. If the trees are different, the program will try to guess where to put the inner node labels. If a node cannot be places optimal, then a message is generated.

Select the (source) tree with the node labels in the left selection list and the destination tree in the right selection list.

If you press 'Copy node info' all nodes in the destination tree get deleted and the new nodes are inserted instead.

If you press 'Add node info' the nodes in the destination tree are renamed ('newname [was: oldname]') and non-existing nodes from the source tree are inserted.

If you check 'only info containing marked species' only groups containing at least one marked species are moved to the destination tree.

WARNINGS It's slow..

::
treeadm.hlp
::

TITLE TREE ADMINISTRATION

OCCURRENCE ARB_NT/Tree_Copy_Delete_Import_Export

DESCRIPTION Trees stored in the database can be deleted, renamed, copied, exported to ascii file, imported from ascii file. Individual protection can be assigned to the trees.

To perform one of the operations
1. select 'Tree_Copy_Delete_Import_Export' from the 'Tree' menu,
2. select a tree from the 'TREE ADMINISTRATION' window
3. press the respective button

IMPORT Loads a NEWICK formatted tree. By default only files with suffix '.tree' are shown in the file selection box. If you want to see all, delete the suffix in the upper right corner.

EXPORT Save tree in NEWICK format.

MOVE NODE INFO Move the internal group labels from one tree to another. If the trees are different, the program will try to guess where to put the inner node labels.

NOTES A protection level equal or higher than that assigned to the particular tree has to be set on the 'ARB_NT' window to allow to delete a tree.

::
trees.hlp
::
TITLE Trees

OCCURRENCE ARB_NT

DESCRIPTION A tree consists of two types of 'nodes':

- tips or terminal nodes: species
  connected to exactly one father
- inner nodes: ancestors
  connected to one father
  (and in the binary case to two children).
  A group_name may be assigned to an inner node; in this case, the node and all its children become a group.
  < ARB_NT/INFO_MODE/M >

There are significant differences between
phylogenetic trees and binary trees:

unrooted	rooted
inner nodes have n children	all inner nodes have two children

Since binary trees are much easier to handle, ARB converts
phylogenetic trees to binary trees.

::
trm_del.hlp
::
TITLE Remove Zoombies

OCCURRENCE ARB_NT/Tree/Remove Zoombies

DESCRIPTION Removes 'deleted' species from the currently displayed tree.

NOTES Nodes and names of species which have been deleted from the
database ('ARB_NT/Species/Delete Marked Species'; 'ARB_NT/
Species/Search/DELETE LISTED SPECIES'; 'ARB_NT/Species/
Info/SPECIES/Delete') are maintained in the trees and indicated
as deleted (<name>)

::
trm_mrkd.hlp
::
TITLE Remove Marked

OCCURRENCE ARB_NT/Tree/Remove Marked

DESCRIPTION Removes all marked species from the currently displayed tree.

::
tungroupall.hlp
::
TITLE Ungroup All

OCCURRENCE ARB_NT/Tree/Ungroup All

DESCRIPTION Shows all internal and terminal nodes of the currently displayed
tree

::
undo.hlp
::
TITLE Undo/Redo

OCCURRENCE ARB_NT ARB_EDIT4 ARB_SECEDIT

DESCRIPTION The undo/redo feature works on your database.
This means it will only affect actions which change the database.
We cannot give you a detailed list of what affects the
database (DB) and what does not - just a few examples:

- changing tree or sequence data affects DB
(you've guessed that!)
- changing the cursor position in ARB_EDIT4 does not affect DB
- changing dialogs does _sometimes_ affect DB (this depends
on whether the value is temporary or permanent).

Note that there is only _one_ undo/redo queue for all applications
together, cause they all work on the same DB!
If you e.g. change a tree in ARB and then change a sequence
in ARB_EDIT4 you cannot undo your tree change by clicking UNDO in
ARB, which would on the contrary undo your sequence change because
it was your last DB-change.

After undo-ing something be careful: If you change anything (that
affects the DB) your current redo-list will be deleted - this
means you can no longer redo your undos!

NOTES The undo-depth depends on memory usage.

::
universal_ift.hlp
::
TITLE NOTES: universal dna

OCCURRENCE ARB_IMPORT

DESCRIPTION This input format reader should read every sequence format,
but:
- removes all other information
- maybe adds additional words to the beginning
or end of the sequence
The final name of the species will be 'spec#', where '#' is
a numerical number. You should rename the species as soon
as possible.

INTERNAL This format scans the file for long strings containing more
than fifty percent ACGTUN-. characters. All other words
are deleted.

NOTES There is no autodetection of this format
No additional information is read
No alignment is preserved
Subsequences with more then 30% ambiguities are removed

::
unmark_list.hlp
::
TITLE UNMARK LISTED SPECIES/GENES

OCCURRENCE ARB_NT/Species/Search: UNMARK LISTED MARK REST
ARB_NT/Genes/Search: UNMARK LISTED MARK REST

DESCRIPTION Unmarks listed species/genes.

WARNINGS Result list may be truncated.

::
user_matrix.hlp
::

TITLE User Defined Distance Matrix

OCCURRENCE ARB_DIST

DESCRIPTION Allows the user to define distances between different bases.
E.g. An A-T should get a smaller difference
as an A-G, so the corresponding matrix elements should be set
to 1.0 and 2.0.

NOTE Before the algorithm starts, all matrix values will be
multiplied by a factor, so that the average matrix value
(except the diagonal) gets a value of one !!!

To save a user matrix use <Properties/Save Properties> menu
item.

BUGS Works only with some distance corrections:

none similarity jukes_cantor

Works only with dna

If enabled only ACGTU/GAPS bases are calculated.

::
version.hlp
::
TITLE Version Info

Version Date Comments
(b=beta)

1.0: 1993-94 Openwin Version

2.0.0b Juni 95 Full Motif Version,
Phylip included,
Final database but untested
Only online help available

2.0.1b July Undo Redo,
Phylip running fully in background
save/load branch labels to NEWICK Format

2.1.0 Nov GDE editor can save
NEW ALI editor from Niels, Fogt ...
a lot of bug fixes:
- database fixes
- recover from corrupt database
...
2.1.1 Jan GDE is working now
compress matrixes is possible

April Vacation

Mai Save Changes as
Parsimony inserts species sorted by sequence length
bug fixes: - import by readseq improved

2.1.2 July USA visit.
Simple Import Function implemented
Perl Interface is running

2.1.3 August PT_server can be updated on the fly

2.2b August Fast Load File implemented (uff a lot of work)
bug fixes: Consensus, import sequences
new: set protection of database fields

2.2.2b August more bug fixes in Fast Load File

2.3b September much better sequence compression
bug fix: probes for groups

2.4b Oktober Tags implemented, Tags can be used to subdivide fields
resize of most windows does work

2.5b November Overall Compression
Incremental fastdnaml
phylips dnapars is running twice as fast now
Search - nearest neighbours
- Equal Fields
December Linux Version

Januar Macros

February One major bug fixed in the database system
Prototype of the new editor

March OSF Alpha Version
Bootstrap for Neighbour Joining

April Fast Aligner
Merge preserves alignment

Feb'99 ARB_EDIT4 now is the default editor.
Code for amino-DNA-translation/realignment
completely rewritten.

::
write_field_list.hlp
::
TITLE WRITE TO FIELDS OF LISTED SPECIES/GENES

OCCURRENCE ARB_NT/Species/Search: WRITE TO FIELDS OF LISTED
ARB_NT/Genes/Search: WRITE TO FIELDS OF LISTED

DESCRIPTION Writes the same text or integers to the selected fields of
all listed species/genes.

NOTES A protection level has to be selected from the Protection menu of the main window (ARB_NT/Protection) equal or higher than that assigned to the selected alignment (ARB_NT/ali_*; third broad rectangular button in the upper part of the main window).

EXAMPLES 1. Write a date of sequence modification (03.27.95) to the fields 'date' of all listed species and replace former entries of these fields:

Select 'Field' by pressing: 'date'
Type 'Text or integer': '03.27.95'
Press: 'WRITE'

2. Add the initials of the modifying person (OS) to the entries stored in the fields 'date' of all listed species:

Select 'Field' by pressing: 'date'
Type 'Text or integer': 'OS'
Press: 'APPEND'

WARNINGS Exept for the 'name' field, there are no different protection levels for different fields. Take care not to write to fields which should contain unique entries for the corresponding species/genes such as accession numbers.

arb-help/help>
help/seer> more *.hlp

::
login.hlp
::
TITLE Standard help file form

OCCURRENCE SEER

DESCRIPTION login into seer system

WARNINGS to change login name, quit seer and start new sesssion

::
main_help.hlp
::
TITLE Seer Main Panel

OCCURRENCE ARB_NT

DESCRIPTION Controls data flow from and to seer database.

HELP: This file

CREATE SKELETON SPECIES: In some cases the tree contains more species than the database. So there are unresolved

links at the tree tips. Pressing this button
creates a dummy species for each unresolved
tree tip

GET MARKED FROM SEERS Convert marked skeleton species into real species

QUERY SEERS AND GET Read data from seers into arb

At the end of the session you should press (in this order)

SAVE SOME TO SEERS Save data back into the Seers dbms

STRIP AND SAVE Delete all species and SAI,
and save the result to an arb file

QUIT quit whole program

WARNING Without saving data back to seers,
you will loose all your work

::
select_attributes.hlp
::
TITLE Seer select attribute

OCCURRENCE SEER

DESCRIPTION At the beginning of each seer session you have to select
all attributes you later want to access or modify during
the session.

WARNINGS There is no way in changing this attribute list afterwards.
(A new session have to be started)

::
select_upload.hlp
::
TITLE Upload Data back to SEER

OCCURRENCE SEER

DESCRIPTION Sends data back to seers

You may only save changes made back to SEER:

Jesus Christ's birth: Send the whole database, don't
check for update stamps
last full arb file: Send all recently changes +
all changes that are in the
delta arb files (xxx.axx)
(see help file saving the database)
Login Send only changes during this
seer session

BUGS does not keep track of already uploaded data. multiple uploads results in multiple database changes.