in a separate line. Note that some programs (e.g., 
extract) require unrooted user trees whereas others (e.g., 
Newman) need rooted user trees instead. See the manuals for 
details.

PHYLIP does not provide programs to generate multiple 
alignment or distance matrices. These must be built by 
hand (remember to save the file in 'Text Only' or ASCII 
format) or through the use of other programs like, e.g. 
Clustal or TreeAlign (be sure to select PHYLIP output 
format for the data file, inspect it for correctness and rename 
it to infile). Many PHYLIP programs use the output of 
other programs in the package to further process the results. 
You should rename the appropriate output file (outfile or treefile) 
to infile before continuing.

A few programs are intended exclusively for converting 
or preprocessing PHYLIP input data for use with other 
programs:

- factor: recodes multistate characters into binary 
datasets.
- seqboot: bootstraps input data sets (molecular sequences, 
  binary characters, restriction sites or gene 
  frequencies).

Using the programs

When run, each program will display on your terminal its 
role, its version and a menu. Your screen will usually look 
like this:

```
Nucleic acid and protein Maximum Likelihood model, version 3.3
Settings for this run:
1. Maximum Likelihood method: Yes
2. bootstrap: no
3. output: replace
4. use missing data: no
5. Sequence of protein names: yes
6. Analyse multiple alignment: no
7. Display output: no
8. Write output: no

Press enter to execute.
```

To change any option, type the corresponding letter from 
the left column in the menu. You will be prompted for 
additional information as needed. Once you are happy 
with the options selected press "Y" to go on.

**PHYLIP** © 1986-1995 by Joseph Felsenstein and the 
University of Washington. 
http://evolution.genetics.washington.edu/

This document is © by José R. Valverde from EMBnet/
CNS and is being distributed by the PARc PC of EMBnet.

EMBnet is a bioinformatics support network situated 
primarily in Europe. Most countries have a national network 
which can provide training courses and user support. 
Further information on PHYLIP is available from your 
national node, which is reachable through

http://www.embnet.org/

If you have found this publication useful or have any ideas 
for similar documents, please contact us: emb-pro@dl.ac.uk

A Quick Guide to PHYLIP

1998

reprinted in 2003

---

**A Quick Guide to PHYLIP**

This is an introduction to the PHYLIP package of 
phylogeny inference programs. PHYLIP is a public domain 
package written by Joe Felsenstein and composed of a large 
number of programs, which makes it versatile and 
very powerful.

This Guide is only a simple roadmap to PHYLIP. Useful tips 
and reminders will be dropped along the way. A basic 
understanding of phylogeny, evolutionary mechanisms and the 
different theoretical approaches employed in evolutionary 
analysis is assumed.

**Preparing input data**

PHYLIP programs read their data from a file that should 
be named "infile". The general format of this file is:

```
x seqno
```
Constructing phylogenies
Next you have to select the program to use, depending on the type of characters with which you are working, what you want to find out and your preferred evolutionary theory/analysis method:

Molecular sequence data: DNA

- Parsimony analysis of phylogeny: DRAKER: unrooted parsimony, counting base changes by the method of Fitch. DRAFERN: finds most parsimonious trees by branch and bound.
- DNA2PAK: computes phylogenetic invariants (evolutionary parsimony) for four species using Cavender's method.

Maximum likelihood methods:
DNAML implements the maximum likelihood method. DNAMLK applies ML assuming trees must be consistent with a molecular clock, i.e., the leaves are all equidistant from the root.

Compatibility methods of phylogeny inference:
DNACOMP chooses the tree and topology to maximize the number of sites with minimum number of substitutions.

Compute distances between species:
DNADIST builds a distance matrix using either the Jukes-Cantor, Kimura, Jin-Nei or maximum likelihood methods.

Molecular sequence data: Proteins

- Parsimony analysis of phylogeny: PROTML: unrooted parsimony using a mixed approach of the methods of Fitch and Eick-Dayhoff.
- PROTML: unsupported program developed by © Jun Adachi and Masami Hasegawa for ML analysis of amino acid data.

Molecular sequence data: Restriction enzymes

- Maximum likelihood analysis of phylogeny: RAILE: phylogeny reconstruction from restriction sites (not restriction fragments) by data of Snolae and I.
- Gene frequencies and continuous characters
- Maximum likelihood analysis of phylogeny: CONTML: uses a Brownian model to reconstruct phylogeny by a restricted maximum likelihood method.

Distance Matrix data

- KITCHEN builds phylogenies using some methods such as Fitch but assuming an evolutionary clock and that all leaves of the tree are contemporaneous. USES neighbor-joining method of Nei and Saitou and the UPGMA methods to successively cluster lineages.

Displaying results

Most programs write a report into a file named "outtree" and a representation of the trees found in a file named "treefile". You should rename these files if you want to preserve the results, otherwise they will be overwritten by the next program you run.

Content of output varies from program to program and depending on the output options selected. Usually it contains the name of the program, the input data and the phylogenies plus associated information. Note that in most cases, the produced trees are unrooted.

The trees contained in the treefile are represented in a standard format that uses parentheses to indicate grouping of species. PHYLIP provides programs to construct, manipulate and print trees in graphic format. RXTREE allows interactive construction and manipulation of trees (topology, branch lengths, labels, etc.).

DRAKON draws cladograms or phenograms of a rooted tree.

DRAWTREE draws unrooted phylogenies on a variety of output devices.

After you select a text font, you will be able to play with the tree interactively using these programs; once you are satisfied with the tree, write it out to a "plotfile" in the format of your choice. This file can then be printed on the appropriate device.

Estimating phylogenies

Estimating phylogenies by hand:

- DRAKE: interactive reconstruction of DNA evolution using parsimony.
- MOVE: interactive reconstruction of 0-1 discrete data evolution using Wagner and Camin-Sokal mixed parsimony (as in MIX).

DOLPHINS: interactive phylogeny reconstruction using the Dollo and Polymorphism parsimony criteria.

BOOTSTRAPPING

- SUGAR: allows to resample data sets by the bootstrap, jackknife or permutation methods.
- CONSENS: finds the consensus of a number of trees using M. methods (including majority-rule and strict consensus).

To do a bootstrap analysis follow these steps:

1. Run SUGAR on the input dataset, select a shuffling method and specify at least 100 (better 1000 or more) replicates. Do not select option "I" to avoid interference with programs run at the next step. Rename outfile to infile.

2. Run the phylogeny analysis program of your choice depending on your data type, desired analysis and preferred method with your desired options. Select option "O" (multiple data sets) and enter the number of replicates generated with SUGAR.

3. If you generated distance matrices in the previous step you need to rename outfile to infile and run one of the Distance Matrix programs (KITCHEN, KITCHEN or KITCHEN) now to generate the trees.

4. Rename the treefile to infile and run CONSENS to evaluate the significance of your analysis.

Note that CONSENS gives the bootstrap support for branching order at each branch of the produced tree and not actual branch lengths.