**Course: Advance Bio Informatics**

**Module Title: PHYLIP**

**Module No: 62**

PHYLIP (PHYLogeny Inference Package) is a free computational phylogenetics package of programs for inferring evolutionary trees (phylogenies). It consists of 35 portable programs, i.e. the source code is written in C and precompiled executables are available for Windows (95/98/NT/2000/me/XP), Mac OS 8 and 9, Mac OS X, and Linux systems. Complete documentation is written for all the programs in the package and is part of the package.

Methods (implemented by each program) that are available in the package include parsimony, distance matrix, and likelihood methods, including bootstrapping and consensus trees. Data types that can be handled include molecular sequences, gene frequencies, restriction sites and fragments, distance matrices, and discrete characters.

Each program is controlled through a menu, which asks the users which options they want to set, and allows them to start the computation. The data is read into the program from a text file, which the user can prepare using any word processor or text editor (but it is important that this text file not be in the special format of that word processor—it should instead be in flat ASCII or Text Only format). Some sequence analysis programs such as the **ClustalW** alignment program can write data files in the PHYLIP format. Most of the programs look for the data in a file called **infile**—if they do not find this file they then ask the user to type in the file name of the data file.

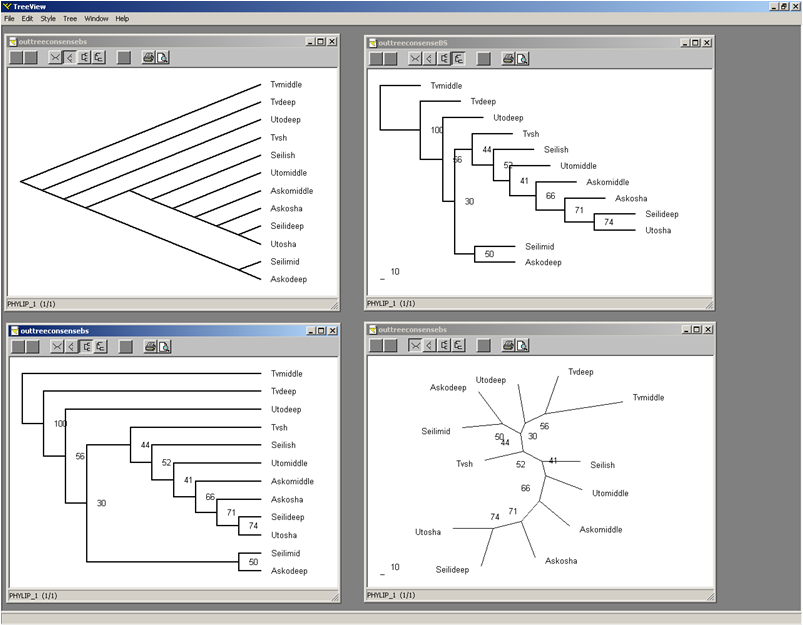
**Output** is written onto files with names like outfile and outtree. Trees written onto outtree are in the Newick format, an informal standard agreed to in 1986 by authors of a number of major phylogeny packages.

**Usefulness**

* Lots of people around who know how to use it Runs can be automated by using redirection and command lines.
* Support for phylip format files by other programs such as clustal, treeview.
* It has easy and transparent interface which means each program does one simple job

**Visualizing Tree**

**Treeview:** Change inthe graphic presentation of a tree (cladogram, rectangular cladogram, radial tree, phylogram), but not change the structure of a tree.



Programs which are part of phylip packages are

|  |  |
| --- | --- |
| **Program Name** | **Description** |
| **protpars** | Estimates phylogenies of protein sequences using the [Parsimony](https://en.wikipedia.org/wiki/Maximum_parsimony) Method |
| **dnapars** | Estimates phylogenies of DNA sequences using the parsimony method. |
| **dnapenny** | DNA parsimony branch and bound method. Finds all of the most parsimonious phylogenies for nucleic acid sequences by branch-and-bound search |
| **dnamove** | Interactive construction of phylogenies from nucleic acid sequences, with their evaluation by DNA parsimony method, with compatibility and display of reconstructed ancestral bases. |
| **dnacomp** | Estimates phylogenies from nucleic acid sequence data using the compatibility criterion. |
| **dnaml** | Estimates phylogenies from nucleotide sequences using the [maximum likelihood](https://en.wikipedia.org/wiki/Maximum_likelihood) method. |
| **dnamlk** | DNA maximum likelihood method with molecular clock. Using both dnaml and dnamlk together permits a [likelihood-ratio test](https://en.wikipedia.org/wiki/Likelihood-ratio_test) for the [molecular clock](https://en.wikipedia.org/wiki/Molecular_clock) hypothesis. |
| **proml** | Estimates phylogenies from protein amino acid sequences by using the maximum likelihood method. |
| **promlk** | Protein sequence maximum likelihood method with molecular clock. |
| **restml** | Estimation of phylogenies by maximum likelihood using restriction sites data (not from restriction fragments but from the presence or absence of individual sites). |
| **dnainvar** | For nucleic acid sequence data on four species, computes Lake's and Cavender's phylogenetic invariants, which test alternative tree topologies. |
| **dnadist** | DNA distance method which computes four different distances between species from nucleic acid sequences. The distances can then be used in the distance matrix programs. |
| **protdist** | Protein sequence distance method which computes a distance measure for protein sequences, using maximum likelihood estimates based on the Dayhoff [PAM matrix](https://en.wikipedia.org/wiki/Point_accepted_mutation), Kimura's 1983 approximation to it, or a model based on the genetic code plus a constraint on changing to a different category of amino acid. |
| **restdist** | Distances calculated from restriction sites data or restriction fragments data. |
| **seqboot** | Bootstrapping/Jackknifing program. Reads in a [data set](https://en.wikipedia.org/wiki/Data_set), and produces multiple data sets from it by bootstrap resampling. |
| **fitch** | Fitch-Margoliash [distance matrix](https://en.wikipedia.org/wiki/Distance_matrix) method. Estimates phylogenies from distance matrix data under the "additive tree model" according to which the distances are expected to equal the sums of branch lengths between the species. |
| **kitsch** | Fitch-Margoliash distance matrix method with molecular clock. Estimates phylogenies from distance matrix data under the "ultrametric" model which is the same as the additive tree model except that an evolutionary clock is assumed. |
| **neighbor** | An implementation of the [Neighbor-Joining](https://en.wikipedia.org/wiki/Neighbor-Joining) method and the [UPGMA](https://en.wikipedia.org/wiki/UPGMA) method. |
| **contml** | Maximum likelihood continuous characters and gene frequencies. Estimates phylogenies from gene frequency data by maximum likelihood under a model in which all divergence is due to genetic drift in the absence of new mutations. This program can also do maximum likelihood analysis of continuous characters that evolve by a Brownian Motion model, assuming that the characters evolve at equal rates and in an uncorrelated fashion. Does not take into account the correlations of characters. |
| **contrast** | Reads a tree from a tree file, and a data set with continuous characters data, and produces the independent contrasts for those characters, for use in any multivariate statistics package. |
| **gendist** | Genetic distance program which computes one of three different genetic distance formulas from gene frequency data |
| **pars** | Unordered multistate discrete-characters parsimony method. |
| **mix** | Estimates phylogenies by some parsimony methods for discrete character data with two states (0 and 1). Allows use of the Wagner parsimony method, the Camin-Sokal parsimony method, or arbitrary mixtures of these. |
| **penny** | Branch and bound mixed method which finds all of the most parsimonious phylogenies for discrete-character data with two states, for the Wagner, Camin-Sokal, and mixed parsimony criteria using the branch-and-bound method of exact search. |
| **move** | Interactive construction of phylogenies from discrete character data with two states (0 and 1). Evaluates parsimony and compatibility criteria for those phylogenies and displays reconstructed states throughout the tree. |
| **dollop** | Estimates phylogenies by the Dollo or polymorphism parsimony criteria for discrete character data with two states (0 and 1). |
| **dolpenny** | Finds all most parsimonious phylogenies for discrete-character data with two states, for the Dollo or polymorphism parsimony criteria using the branch-and-bound method of exact search. |
| **dolmove** | Interactive construction of phylogenies from discrete character data with two states (0 and 1) using the Dollo or polymorphism parsimony criteria. Evaluates parsimony and compatibility criteria for those phylogenies and displays reconstructed states throughout the tree. |
| **clique** | Finds the largest clique of mutually compatible characters, and the phylogeny which they recommend, for discrete character data with two states (0 and 1). The largest clique (or all cliques within a given size range of the largest one) are found by a very fast branch and bound search method. |
| **factor** | Character recoding program which takes discrete multistate data with character state trees and produces the corresponding data set with two states (0 and 1). |
| **drawgram** | Rooted tree drawing program which plots rooted phylogenies, cladograms, and phenograms in a wide variety of user-controllable formats. The program is interactive and allows previewing of the tree on PC or Macintosh graphics screens, and Tektronix or Digital graphics terminals. |
| **drawtree** | Unrooted tree drawing program similar to DRAWGRAM, but plots unrooted phylogenies.. |
| **consense** | Consensus tree program which Computes consensus trees by the majority-rule consensus tree method, which also allows one to easily find the strict consensus tree. Is not able to compute the Adams consensus tree |
| **treedist** | Computes the [Robinson-Foulds](https://en.wikipedia.org/wiki/Robinson-Foulds_metric) symmetric difference distance between trees, which allows for differences in tree topology. |
| **retree** | Interactive tree rearrangement program which reads in a tree (with branch lengths if necessary) and allows you to reroot the tree, to flip branches, to change species names and branch lengths, and then write the result out. Can be used to convert between rooted and unrooted trees. |