NAME

weighbor − weighted neighbor joining

SYNOPSIS

weighbor [−L *sequenceLength]* [−b *alphabetSize]* [−i *fileName]* [−o *fileName]* [−v[v[v]]] [−V]

DESCRIPTION

Weighbor takes an input file of pairwise distances in *Phylip* format (see section **FILE FORMATS** below) and computes the phylogentic tree that best corresponds to those distances. The output is the tree in the lisp-like 'Newick' format. Currently the input is read from **stdin** and the output is written to **stdout.** Note, several inputs can be sent to **weighbor** and it will process each in order and print the corresponding tree. Additional information about how the tree was built can be output to a log file (currently **weighbor.out**). The amount of information written to the log file can be controlled.

These options are available:

−L *sequenceLength*

Sets the length of the sequences. Default is 500. This is the effective sequence length equal to the number of varying sites. Note if the **−L** option is not used then the program will print a warning message to stderr indicating that it is using this default length.

−b *alphabetSize*

Sets the size of the alphabet of characters (number of bases) **b. 1/b** is equal to the probablity that there will be a match for infinite evolution time. The default value for **b** is 4.

−i *fileName*

Read input from the filename given. Default is to read from stdin.

−o *fileName*

Write output to the given file. Default is to write to stdout.

−v Turns on logging and creates the logfile *weighbor.out* The amount of information can be adjusted by using **−v, −vv** or **−vvv** to get progressively more detailed information in the logfile. At level one **(−v)** the the logfile will contain the pair of taxa that was joined at each iteration of the algorithm along with the taxon that was the next best choice. Also printed is the p-value of this triplet.

−V Print out the version number.

FILE FORMATS

Weighbor tries to be as flexable as possible in the format of the input file it will allow, while maintaining PHYLIP compatibility. A file that conforms *EXACTLY* to the PHYLIP format (as defined in the PHYLIP package) will work with *PROVIDED* the first character after the 10-character name (i.e., the 11th character on the line) is a space (' '), tab or newline.

Additionally we have added the following extensions.

Long taxon names.

Taxon names can be up to 128 characters long. There is one caveat, spaces embedded in the file name can only exist from position 2-9. A space before pos 2 will be discared and after position 9 will terminate the name. Any non-white space character can be used in the name.

Tab and newline delimiters

Weighbor will allow both tabs and newlines to delimit taxon names so the following will be read correct even though it is not a valid PHYLIP format because the names are not padded to 10 characters.

3
Name1
0 1 2
Name2
1 0 2

Name3
1 2 0

Lower triangular form

Weighbor will also a distance matrices in lower triangular form. For example the following matrix will result in the same distance matrix as the previous sample

3
Name1
0
Name2
1 0
Name3
1 2 0

Symmetry sanity checks

Weigbhor will force symmetry on non symmetric inputs. It does so by averaging the given input with its transpose.

DEVELOPMENT TEAM

W. Bruno <*wjb@dnamining.com*>

A. Halpern <*halpern@ender.unm.edu*>

N. D. Socci <*soccin@rockefeller.edu*>