

Documentation for mergeDistMats Program

Ward C. Wheeler
Division of Invertebrate Zoology,
American Museum of Natural History,
200 Central Park West, New York, NY, 10024, USA;
wheeler@amnh.org

February 3, 2021

Running Title: mergeDistMats

0.1 Introduction

This is the first version of documentation for the program mergeDistMats. This program is designed to merge phylogenetic distance matrices for subsequent input to phylogenetic distance analysis programs such as Wag2020 (<https://github.com/wardwheeler/wag2020>).

All source code, precompiled binaries, test data, and documentation are available from <https://github.com/wardwheeler/mergeDistMats>.

This first version is brief.

0.2 Input Distance Matrix Format

Input distance matrices should be in CVS format (can be exported from spreadsheet programs), with first line of taxon (terminal) names and square matrix following. First column (from second row) contains distances, not repeat of names.

Matrices must be symmetrical and non-negative, but may have different (if overlapping) taxon (terminal) sets.

Beware of the MS-Excel non-conforming CVS output. A final line feed needs to be added to Excel created CVS files.

0.3 Output Formats

The program outputs (to stdout) a unified matrix in CVS format.

0.4 Command options

There is a single command line option to be specified. Either “normalize” or “weight”. This parameter specifies whether the input matrices (and output matrix) are scaled on [0,1] (“normalize”) to force an equal weighting on each input matrix. The alternative, “weight” leaves the input values as they are and simply adds them together.

The program requires a single argument and at least one input matrix file.

0.5 Program Use

The program is invoked from the command-line as in:

```
mergeDistMats normalize testData.csv testData2.csv > outputMatrix.csv
```

```
mergeDistMats weight testData.csv testData2.csv > outputMatrix.csv
```

Execution in Parallel

By default the program will execute using a single process core. By specifying the options ‘+RTS -NX -RTS’ where ‘X’ is the number of processors offered to the program. These are specified after the program as in (for 4 parallel threads):

`mergeDistMats +RTS -N4 -RTS other options...`

Acknowledgments

The author would like to thank NSF REU DBI-1358465, DARPA SIMPLEX N66001-15-C-4039, and Robert J. Kleberg Jr. and Helen C. Kleberg foundation grant “Mechanistic Analyses of Pancreatic Cancer Evolution” for financial support.