QuartetNet
User's Manual

Jialiang Yang,[1] Stefan Grünewald,[2] and Xiu-Feng Wan[1]

[1]Department of Basic Sciences, College of Veterinary Medicine, Mississippi State University, Mississippi State, MS 39762, USA;
[2] CAS-MPG Partner Institute for Computational Biology, Key Laboratory of Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

Contact: Xiu-Feng Wan wan@cvm.msstate.edu

1   Methods

QuartetNet is a quartet based algorithm, as well as an implementing software suite, to reconstruct phylogenetic networks from weighted triplets and quartets. It can be viewed as a quartet analogue of Split Decomposition [1, 2] and functions by agglomeratively decomposing all triplet and quartet weights into simple components based on full splits. The QuartetNet algorithm is canonical and no tie-breaking. It's running time is at most $O(n^8)$ and memory requirement is $O(n^5)$.

By far, QuartetNet is perhaps the best phylogenetic network reconstruction algorithm in consistency, which can precisely reconstructing all 2-weakly compatible split systems. Please refer to [3] for more details on the theory behind QuartetNet.

2   Installation

QuartetNet requires that C++ is installed and configured on the system. After unzipping the source code package, the user may use the command "make" to install QuartetNet for any UNIX/Linux/MacOSX system or add all the files to a c++ project and compile it for any Windows system. A script file or .bat file is generated after the installation or compilation,. QuartetNet can then be run from the command-line or the .bat file. See "Usage" for more information. Alternatively, the user can copy the script file (for UNIX/Linux/MacOSX systems) or .bat file (for Windows systems) in the source code package to a directory and run it there.

3 Usage

3.1 Use from command-line

QuartetNet accepts two kinds of input files at present, which include multiple sequence alignment (MSA) files in fasta format and .cweights files specifying all the weighted triplets and quartets in a prescribed format. Please refer to the Format section for the .cweights format. In addition, it provides two options to implement two methods, namely a minimum method and a second minimum method, to calculate the weights of full splits since both methods are consistent for all 2-weakly compatible split systems [3].

To use, run the following command:

./QuartetNet [-PARAMS] [-PARAMS]

PARAMS:

- -: help for the parameters and their usage

- a : read data from multiple alignment files in fasta format (Default).

- w : read data from a valid cweight files.

- m : calculate weights by the minimum method (Default).

- s : calculate weights by the second minimum method.

For example, "./QuartetNet -w -s" reads the input from a .cweight file and uses the second minimum method to generate weighted full splits. The program will ask for the directory and name of the input files and the directory and name of the files to output the full weighted splits.

3.2 Visualization

The final output will be a Nexus file, which can be viewed by many softwares including "SplitsTree" [4].

3.3 File format

A sample .cweights files reads:

```
taxanumber: 6;
description: artificial data;
taxon: 1 name: Bbro;
taxon: 2 name: Bmal;
taxon: 3 name: Bper;
```

```
taxon: 4 name: Bpse;
taxon: 5 name: Btha;
taxon: 6 name: Rsol;
quartet: 1 2 3 4 weights: 36 2434 0
quartet: 1 2 3 5 weights: 36 2296 0
quartet: 1 2 3 6 weights: 5 1074 1
quartet: 1 2 4 5 weights: 163 2 81
quartet: 1 2 4 6 weights: 15 2 657
quartet: 1 2 5 6 weights: 38 17 603
quartet: 1 3 4 5 weights: 2467 0 0
quartet: 1 3 4 6 weights: 1103 3 35
quartet: 1 3 5 6 weights: 1095 3 35
quartet: 1 4 5 6 weights: 25 16 784
quartet: 2 3 4 5 weights: 53 81 2
quartet: 2 3 4 6 weights: 11 660 36
quartet: 2 3 5 6 weights: 34 606 51
quartet: 2 4 5 6 weights: 64 0 128
quartet: 3 4 5 6 weights: 25 16 685
triplet: 1 2 3  weights: 11 660 36
triplet: 1 2 4  weights: 25 16 685
triplet: 1 2 5  weights: 265 182 366
triplet: 1 2 6  weights: 36 2296 0
triplet: 1 3 4  weights: 4000 5000 4000
triplet: 1 3 5  weights: 1233 2987 8
triplet: 1 3 6  weights: 3000 4000 3000
triplet: 1 4 5  weights: 5555 3453 2322
triplet: 1 4 6  weights: 4000 5000 4000
triplet: 1 5 6  weights: 3000 4000 3000
triplet: 2 3 4  weights: 3000 4000 3000
triplet: 2 3 5  weights: 38 17 603
triplet: 2 3 6  weights: 3000 4000 3000
triplet: 2 4 5  weights: 3000 4000 3000
triplet: 2 4 6  weights: 3000 4000 3000
triplet: 2 5 6  weights: 3000 4000 3000
triplet: 3 4 5  weights: 3000 4000 3000
triplet: 3 4 6  weights: 3000 4000 3000
triplet: 3 5 6  weights: 36 2296 0
triplet: 4 5 6  weights: 3000 4000 3000
```

The part for the triplets is optional. If the file does not specify any triplet weight, then QuartetNet will set all the triplet weights to be 0 by default.

4   Legal

- QuartetNet is released under the GNU General Public License as set down at

    `http://www.gnu.org/copyleft/gpl.html`

- The QuartetNet algorithm and software was developed by Jialiang Yang, Stefan Grünewald and Xiu-Feng Wan.

5   Citation

Yang J, Grünewald S, Wan X-F. 2012. QuartetNet: A quartet based method to reconstruct phylogenetic networks.

References

[1] Bandelt HJ, Dress AWM. 1992. A canonical decomposition theory for metrics on a finite set. Advances in mathematics. 92: 47-105.

[2] Bandelt HJ, Dress AWM. 1992 Split Decomposition: a new and useful approach to phylogenetic analysis of distance data. Molecular Phylogenetics and Evolution. 1: 242-252.

[3] Yang J, Grünewald S, Wan X-F. 2012 QuartetNet: A quartet based method to reconstruct phylogenetic networks. submitted.

[4] Huson D, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution. 23: 254-267. Software available from www.splitstree.org.