QNet
User's Manual

Kristoffer Forslund

November 6, 2006

1 Methods

1.1 Quartet-based phylogenetic network reconstruction

QNet, short for Quartet Network, is an algorithm to combine quartet phylogenies into a phylogenetic network, as well as the name of the software suite implementing the method. It functions by first finding a cyclic ordering between the taxa in the input set, then estimate via nonnegative least squares the set of split weights that best recreates the original set of quartet weights. Its running time and memory requirements are on the order of number of possible quartets, that is, $\binom{n}{4}$.

After creating a QNet network, a subapplication, Filterer, may be used to exclude splits that are poorly supported for ease of visualisation. This is done by requiring that for a split to be displayed, its weight must exceed a user-defined threshold ratio times the weight of the strongest split that conflicts with this split.

For more details on the theory behind QNet, see ([4]).

2 Installation

QNet requires that Java is installed and configured on the system. For details and for downloading that software, se `http://java.sun.com`.

Unzip the distribution `.zip` file ("qnet.zip") in a directory where you wish to place the program files. In this directory you will then find a set of script files (for UNIX/Linux/MacOSX systems) or `.bat` files (for Windows systems) that start the QNet applications. Alternately, QNet may be run completely manually from the command-line. See "Usage" for more information.

3 Usage

3.1 Generating quartet weights from an alignment file using GWeight

GWeights is a program that will produce quartet weights for an alignment of amino acid or nucleotide sequences, in the weights format used by Stefan Grünewald and associates. The method used for generating weights is that of sequence space distance classes as used in statistical geometry ([3], [2], [7]). Both standard sequence space distance classes and the weighted version introduced by Nieselt-Struwe and von Haeseler are supported, to allow handling of protein sequences.

GWeights should have been installed along with QNet. To use, run the following command in the directory where QNet has been installed:

```
java gweight.GWeight <input file in FASTA format> <output file in weights format> <method code>
```

where `<method code>` currently should be one of the following:

- SG: standard statistical geometry
- BL62: BLOSUM62 weighted statistical geometry
- AK: AlmostKimura, not-quite-perfected implementation of the Kimura 2-parameter distance correction

3.2 Generating quartet weights from an alignment using the LengthWeighter Tree-Puzzle front-end

LengthWeighter uses Tree-Puzzle ([9]) to generate quartet weights, then weighs these values based on the Tree-Puzzle branch lengths. To use, run the following command in the directory where QNet has been installed:

```
java -cp .  chopper.LengthWeighter <alignment infile> <quartet weights outfile> <param file>
```

where the alignment must be in a format acceptable to Tree-Puzzle. The path to Tree-Puzzle will be read from the parameter file, which should be edited to suit your configuration (a sample file, `LengthWeighter.ini` should be situated in the directory where QNet has been installed). The first line should hold the path to Tree-Puzzle and the second its executable name, correct as per your configuration of Tree-Puzzle. See provided example file for details.

3.3 Conversion of data sources using Chopper

Another QNet subapplication, Chopper, enables conversion of several types of phylogenetic data into quartet phylogeny information.

Current datatypes that Chopper supports (to some degree; please report bugs to the author as you encounter them) as source sets includes:

- qweights files, see syntax below - simple, monopurpose internal format used by QNet and Chopper, originally appeared in ([10])
- Nexus files with st_quartets (old format) or Quartets blocks ([6], [5])
- Nexus files with st_splits (old format) or Splits blocks ([6], [5])
- Nexus files with distance blocks ([6], [5])
- Treebase syntax Nexus files with TREES blocks ([6])
- Newick trees, with or without branch lengths [8]

To convert data into suitable quartet weights files from other formats or from data such as splits or Newick trees, the Chopper subapplication is used.

To use the provided script file (which is "chopper.bat" for Windows users and "chopper.sh" otherwise), run the following command in the directory where QNet has been installed:

```
<chopper.bat/chopper.sh> <type> <infile> <outfile>
```

To run manually, instead run:

```
java -cp .  chopper.Chopper <type> <infile> <outfile>
```

`<type>` should be "newick", "qweight", "nexus:st_quartets", "nexus:distances", "nexus:trees" or "nexus:st_splits", corresponding to what type of file `<infile>` points to.

If file type is "newick", what is expected is a file containing a single line describing a standard Newick format tree.

If file type is "qweight", what is expected is a standard qweights file (see below for specifications).

If file type is "nexus:st_quartets", "nexus:distances", "nexus:trees" or "nexus:st_splits", Nexus files with `st_quartets`, `distances`, `trees` or `st_splits` (alternately `Splits`) blocks are expected.

`<outfile>` is the name of the resulting qweights file.

3.4 Reconstruction of a phylogenetic network using QNet

To use the provided script file (which is "qnet.bat" for Windows users and "qnet.sh" otherwise), run the following command in the directory where QNet has been installed:

```
<qnet.bat/qnet.sh> <lin/log> <infile> <outfile>
```

To run manually, instead run:

```
java -cp .  qnet.QNet <lin/log> <infile> <outfile>
```

As seen in this example, the program requires three command-line parameters. The first enables taking the logarithm of the quartet weights in the input data if the parameter is given as "log", if given as "lin" the weights will be used as-is. The second is the name (or path to + name) of the quartet weights input file, which should be according to the monopurpose internal format that Chopper outputs, or a Nexus file with a `taxa` and `st_quartets` block. The last parameter is the name of the Nexus file to which results (a taxa and a `st_splits` block) should be saved.

3.5 Filtering a splits file

Once a network has been computed, it may be desirable to filter out weakly supported splits that may result from numerical error or noise. This can be done using the Filterer subapplication.

To use the provided script file (which is "filterer.bat" for Windows users and "filterer.sh" otherwise), run the following command in the directory where QNet has been installed:

```
<filterer.bat/filterer.sh> <infile> <outfile> <threshold>
```

To run manually, instead run:

```
java -cp .  qnet.Filterer <infile> <outfile> <threshold>
```

`<infile>` and `<outfile>` should be Nexus files with `st_splits` blocks, whereas `<threshold>` is a number. Only those splits whose weight is higher than the threshold number times the weight of the most highly weighted conflicting split are retained from `<infile>` to `<outfile>`.

3.6 Visualizing the results

The final output will be a Nexus file, which can be viewed by a variety of applications, among them SplitsTree ([1]).

3.7 The qweights internal format - specification

A sample file reads:

```
0   taxanumber:  6;
1   description: artificial data;
2   sense: max;
3   taxon:   001   name: a;
4   taxon:   002   name: b;
5   taxon:   003   name: c;
6   taxon:   004   name: d;
7   taxon:   005   name: e;
8   taxon:   006   name: f;
9   quartet: 001 002 003 004 weights: 200 0 200;
10  quartet: 001 002 003 005 weights: 200 0 200;
11  quartet: 001 002 003 006 weights: 200 0 200;
12  quartet: 001 002 004 005 weights: 210 0 210;
13  quartet: 001 002 004 006 weights: 210 0 210;
14  quartet: 001 002 005 006 weights: 410 0 410;
15  quartet: 001 003 004 005 weights: 10 0 10;
16  quartet: 001 003 004 006 weights: 10 0 10;
17  quartet: 001 003 005 006 weights: 210 0 210;
18  quartet: 001 004 005 006 weights: 200 0 200;
19  quartet: 002 003 004 005 weights: 10 0 10;
20  quartet: 002 003 004 006 weights: 10 0 10;
21  quartet: 002 003 005 006 weights: 210 0 210;
22  quartet: 002 004 005 006 weights: 200 0 200;
23  quartet: 003 004 005 006 weights: 200 0 200;
```

4 Legal

The QNet algorithm was developed by Stefan Grünewald.

The QNet software was written by Kristoffer Forslund.

References

[1] Andreas Dress, Daniel Huson, and Vincent Moulton. Analyzing and visualizing sequence and distance data using splitstree. *Discrete Applied Mathematics*, 71:95–109, 1996.

[2] Manfred Eigen and Ruthild Winkler-Oswatitsch. Statistical geometry on sequence space. *Methods in enzymology*, 183:505–530, 1990.

[3] Manfred Eigen, Ruthild Winkler-Oswatitsch, and Andreas Dress. Statistical geometry in sequence space: A method of quantitative comparative sequence analysis. *PNAS*, 85:5913–5917, 1988.

[4] S. Grunewald, K. Forslund, A. Dress, and V. Moulton. Qnet: An agglomerative method for the construction of phylogenetic networks from weighted quartets. submitted.

[5] Daniel H. Huson and David Bryant. Splitstree manual. http://www-ab.informatik.uni-tuebingen.de/software/splitstree4/V4_6/manual.pdf.

[6] David R. Maddison, David L. Swofford, and Wayne P. Maddison. Nexus: An extensible file format for systematic information. *Systematic Biology*, 46:590–621, 1997.

[7] Kay Nieselt-Struwe. Graphs in sequence spaces: a review of statistical geometry. *Biophysical Chemistry*, 66:111–131, 1997.

[8] Gary Olsen. The "newick's 8:45" tree format standard. http://evolution.genetics.washington.edu/phylip/newick_doc.html.

[9] H. A. Schmidt, K. Strimmer, M. Vingron, and A. von Haeseler. Tree-puzzle: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*, pages 502–504, 2002.

[10] Jan Weyer-Menkhoff. New quartet methods in phylogenetic combinatorics. Ph.D. thesis, Universität Bielefeld, 2003.