
PhylotreeDist

A C++ bioinformatic library that provides various phylogenetic distances
between pair of trees.

Manual

October 5, 2011

Contents

1	PhylotreeDist manual	2
1.1	Dependencies	3
1.2	Resources	3
1.3	Usage	3
1.3.1	Commandline executable	4
1.3.2	Library	6
1.4	Source code management	6
1.4.1	Directory structure	6
1.4.2	Compilation	7
1.5	Licence	7

Chapter 1

PhylotreeDist manual

PhylotreeDist is a C++ bioinformatic library that provides various phylogenetic distances between pair of trees.

PhylotreeDist is distributed with source code as well as binaries for Linux (x86) environment.

Available distances:

- **Perfect matching - on splits** for unrooted trees
- **Perfect matching - on clusters** for rooted trees
- **Perfect matching - on pairs** for rooted binary trees
- **Robinson-Foulds** for rooted and unrooted trees
- **Weighted Robinson-Foulds** for unrooted branch-weighted trees
- **Quartet** for unrooted trees
- **Triplets** for rooted trees
- **Nodal** for unrooted trees
- **Weighted Nodal** for unrooted branch-weighted trees

1.1 Dependences

Bio++ [2]

For required resources, visit <http://biopp.univ-montp2.fr/articles/download/> and download the packages of **exact version** as given below (as currently in July 2011 the newest version is invalid). The recommended sequence of installation is the order of the listing below.

1. bpp-core 2.0.0
2. bpp-utils 1.5.0
3. bpp-numcalc 1.8.0
4. bpp-seq 1.7.0
5. bpp-phyl 1.9.0

You can download the required libraries via <http://www.awek.com.pl/~andziaania/libbpp.tar.gz>
place the library files into `externals/lib` directory and add the `externals/lib` to the `LD_LIBRARY_PATH` variable (see section 1.3.1).

1.2 Resources

The library with sample application are available via Internet:

- Linux(x86) binaries
http://www.awek.com.pl/~andziaania/PhylotreeDist_LBin.tar.gz
- Source code with binaries for Linux(x86)
http://www.awek.com.pl/~andziaania/PhylotreeDist_src.tar.gz
The compilation process is described in section 1.4.2

You can unpack the resources e.i. using tar [4] archiving utility

```
tar -xvzf PhylotreeDist_XXX.tar.gz
```

1.3 Usage

The PhylotreeDist package can be used as a **library** or a **commandline executable**.

1.3.1 Commandline executable

1. Download Bio++ dependencies (section 1.1).

if needed - set LD_LIBRARY_PATH to them:

```
export LD_LIBRARY_PATH=\
$LD_LIBRARY_PATH:<path to Bio++ libraries>
```

2. Download and manage appropriate PhylotreeDist resources (see section 1.2).
3. Enter into the PhylotreeDist directory with binaries.
4. Show the library directory to the system:

```
export LD_LIBRARY_PATH=\
$LD_LIBRARY_PATH:<path to phylotreedist library e.i. current dir>
```

5. The executable is run according to the following pattern:

```
./phylotreeDistApp -i <inputFile.newick> -o <outputFile> \
[ -d <distance type> ] [ -m <comparison mode> ] [-c]
```

phylotreeDist executable arguments:

-i <inputFile.newick>

Compulsory.

Indicates the source file with trees in newick format.

-o <outputFile>

Compulsory.

Indicates the output file.

-m [p|m]

Default to p.

Optional.

Comparison mode - which input trees are to be compared together.

Options:

p - pair - every two neighboring trees in the input file are compared

m - matrix - every two trees in the input file are compared

-d [ms|mc|mp|rf|rffw|q|t|n|nw|nm|nmw|np|npw]

Default to rf.

Optional.

Distance choice - the algorithm used to compare the trees.

Options:

ms	unrooted trees	matching with splits as description elements
mc	rooted trees	matching with clusters as description elements
mp	rooted binary trees	matching with pairs as description elements
rf	two rooted or two unrooted trees	robinson-foulds (default if no metric is chosen)
rft	unrooted branch-weighted trees	weighted robinson-foulds
q	unrooted trees	quartets
t	rooted trees	triplets
nm or n	unrooted trees	nodal, computing distance between leaves with manhattan metric
nmw or nw	unrooted branch-weighted trees	weighted nodal, computing distance between leaves with manhattan metric
np	unrooted trees	nodal, computing distance between leaves with pythagorean metric
npw	unrooted branch-weighted trees	weighted nodal, computing distance between leaves with pythagorean metric

-c

Optional.

Check trees constraints (un/rooted, bi/multifurcating, the same leaves sets) and throw exception if anything is incorrect.

RESULTS

The distances results are available in the file given as the *-o* parameter value. The results are in the following format:

Metric:	<i>metric name</i>
Comparison mode:	<i>comparison mode type</i>
Input file:	<i>filename</i>
Output file:	<i>filename</i>
<hr/>	
<i>comparison id (1..n)</i>	<i><tabulation separator> distance value</i>

Example:

Metric:	Matching-Splits
Comparison mode:	pairs
Input file:	data/unrooted.newick
Output file:	msResult.dat
<hr/>	
1	137
2	25
3	4
4	144

1.3.2 Library

The interface to the library is provided in the file

`include\PhylotreeDist.h`

Also you can consult Anna Pawelczyk's "The Methods of Phylogenetic Trees Distance Definition" paper.

The dynamic library is available as `libPhylotreeDist.so`.

1.4 Source code management

1.4.1 Directory structure

bin\	binaries
bin\ GNU-Linux-x86\	Linux(x86) phylotreeDistApp and libPhylotreeDist.so binaries
data\	sample trees in Newick format
doc\	documentation

<code>externals\</code>	header files of Bio++
<code>include\</code>	header files of the PhylotreeDist library
<code>Makefile</code>	makefile
<code>makeDepend\</code>	dependences for the Makefile
<code>sampleapp\</code>	sample application <code>PhylotreeDistApp</code> source code and Makefile
<code>src\</code>	source files of the PhylotreeDist library
<code>test\</code>	unit test files for the PhylotreeDist library

1.4.2 Compilation

Enter into `PhylotreeDist` directory.

To compile the library `libPhylotreeDist.so`:

```
make
```

To compile the executable sample application `PhylotreeDistApp`

first compile the library and then:

```
cd PhylotreeDistApp
make
cd ..
```

The binaries (dynamic `libPhylotreeDist` library and executable `PhylotreeDistApp` sample application) are available at `bin\<platform>`.

1.5 Licence

Copyright (C) 2011, Anna Pawelczyk, Gdansk University of Technology

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see <http://www.gnu.org/licenses/>.

Bibliography

- [1] <http://www.cygwin.com/>.
- [2] <http://biopp.univ-montp2.fr/>.
- [3] <http://www.mingw.org/>.
- [4] [http://man.cx/tar\(1\)/](http://man.cx/tar(1)/).