PhyloNet Tutorial (Species Phylogeny Inference)

1. Introduction

PhyloNet is a tool designed mainly for analyzing, reconstructing, and evaluating reticulate (or non-treelike) evolutionary relationships, generally known as phylogenetic networks. Various methods that we have developed make use of techniques and tools from the domain of phylogenetic trees, and hence the PhyloNet package includes several tools for phylogenetic tree analysis. PhyloNet is released under the GNU General Public License. For the full license, see the file GPL.txt included with this distribution.

Species phylogeny inference is central in biology. PhyloNet considers the scenarios where the incongruence among gene trees are caused by incomplete lineage sorting (ILS) and hybridization (or horizontal gene transfer). More precisely, if assuming ILS is the only cause of gene tree incongruence, then given a set of gene trees, PhyloNet provides tools that can score a candidate species tree as well as infer a species tree under the criterion of minimizing deep coalescence (MDC). On the other hand, if assuming that the incongruence among gene trees are caused by both ILS and hybridization (or horizontal gene transfer), then given a set of gene trees, PhyloNet provides maximum parsimony, maximum likelihood, maximum pseudo-likelihood and Bayesian inferences of species networks. Furthermore, if you have some hypothesis of the species networks you want to compare, PhyloNet also provides tools to compute the parsimonious and likelihood score of given species networks.

2. Installation

System Requirements

In order to run the PhyloNet toolkit, you must have Java 1.8.0 or later installed on your system. All references to the java command assume that Java 1.8 is being used.

- To check your Java version, type "java -version" on your command line.
- To download Java 1.8, please go to website http://www.java.com/en/download/.
- To link to the new downloaded Java 1.8, for mac, try these two commands from command line:
  sudo rm /usr/bin/java
  sudo ln -s /Library/Internet\ Plug-Ins/JavaAppletPlugin.plugin/Contents/Home/bin/java /usr/bin

Downloading phylonet.jar

Acquire the current release of PhyloNet by downloading the most recent version of the PhyloNet JAR file. You will have a file named `PhyloNet_X.Y.Z.jar`, where X is the major version number and Y and Z are the minor version numbers.

Installing the file

Place the jar file in the desired installation directory. The remainder of this document assumes that it is located in `$PHYLONET PATH/jar`. Installation is now complete. In order to run PhyloNet, you must execute the file `PhyloNet_X.Y.Z.jar`, as described in the next section.

3. Usage

Basic Usage

The PhyloNet tool is executed by typing the following command into your console:

```
>java -jar $PHYLONET PATH/jar/PhyloNet_X.Y.Z.jar script.nex
```

Where `script.nex` is the NEXUS file containing the commands to be executed.

Scoring a candidate species phylogeny

- Parsimonious:
    - Scoring a species tree (ILS): DeepCoalCount_tree
    - Scoring a species tree/network (ILS+introgression): DeepCoalCount_network
- Probabilistic:
    - Scoring a species tree/network (ILS+introgression): CalGTProb

Inferring a species phylogeny

- Maximum parsimony:
    - Inferring a species tree (ILS): Infer_ST_MDC, Infer_ST_MDC_UR
    - Inferring a species tree/network (ILS+introgression): InferNetwork_MP
- Maximum likelihood:
    - Inferring a species tree/network (ILS+introgression): InferNetwork_ML
    - inferring a species tree/network (ILS+introgression) with cross-validation: InferNetwork_ML_CV
- Maximum pseudo-likelihood:
    - Inferring a species tree/network (ILS+introgression): InferNetwork_MPL
- Bayesian:
    - Inferring a species tree/network (ILS+introgression): MCMC_GT
    - Inferring a species tree/network (ILS+introgression) from multilocus data: MCMC_SEQ

For other tools in PhyloNet, please see here for a full list.

4. Yeast Data Set

Here is the yeast data set (Rokas et al., 2003) that you could try to analyze using our tools. It consists of 106 loci, each with a single allele sampled from seven Saccharomyces species *S. cerevisiae* (*Scer*), *S. paradoxus* (*Spar*), *S. mikatae* (*Smik*), *S. kudriavzevii* (*Skud*), *S. bayanus* (*S bay*), *S. castellii* (*Scas*), *S. kluyveri* (*Sklu*), and the outgroup fungus *Candida albicans* (*Calb*).

The followings are the input NEXUS file you could try:

- ILS:
    - Score a species tree:
        - Count the number of extra lineages given a species tree: ScoringTreeUnderParsimony
        - Compute the probability of the gene trees given a species tree: ScoringTreeUnderLikelihood
    - Infer a species tree:
        - Infer the species tree under parsimony: InferringTreeMP
        - Infer the species tree using maximum likelihood: InferringTreeML
- ILS + introgression
    - Score a species network:
        - Count the number of extra lineages given a species network: ScoringNetworkUnderParsimony
        - Compute the probability of the gene trees given a species network: ScoringNetworkUnderLikelihood
    - Infer a species network:
        - Infer the species network under parsimony: InferringNetworkMP
        - Infer the species network using maximum likelihood: InferringNetworkML
        - Infer the species network using maximum pseudo-likelihood: InferringNetworkMPL
        - Infer the species network using Bayesian: InferringNetworkMCMC
        - Infer the species network using Bayesian from multilocus data: InferringNetworkMultilocusDataMCMC

5. Visualizing a Phylogenetic Network

Phylogenetic network in Rich Newick string can be visualized in Dendroscope. However, Dendroscope can not recognize inheritance probabilities (branch lengths and supports are fine). You have to remove those probabilities manually from the Rich Newick string, or use option "-di" so that PhyloNet returns the network that Dendroscope takes directly. For example, the following network is what PhyloNet returns when the input NEXUS file InferringNetworkMP is used:

```
((((((Skud,(Sbay)#H1:::0.36082474226804123),((Spar,Scer),Smik)),#H1:::0.63917525773195
87),Scas),Sklu),Calb);
```

In order to visualize this phylogenetic network in Dendroscope, please remove the inheritance probabilities like follows, and Dendroscope will be able to read it. Or you can use option "-di".

```
((((((Skud,(Sbay)#H1),((Spar,Scer),Smik)),#H1),Scas),Sklu),Calb);
```

6. Tutorial

PDF. tutorial.pdf

7. Website

To see a full description of PhyloNet, please visit its website, where you can get the latest version of PhyloNet as well as detailed descriptions of all available commands.

8. References

- Than C, Ruths D, Nakhleh L. PhyloNet: A Software Package for Analyzing and Reconstructing Reticulate Evolutionary Relationships. *BMC Bioinformatics*, 9:322, 2008
- C. Than and L. Nakhleh. Species tree inference by minimizing deep coalescences. *PLoS Computational Biology*, 5(9):e1000501, 2009.
- Y. Yu, T. Warnow, and L. Nakhleh. Algorithms for MDC-based multi-locus phylogeny inference. *Proceedings of the 15th Annual International Conference on Research in Computational Molecular Biology (RECOMB)*, LNBI 6577, 531-545, 2011.
- Y. Yu, T. Warnow, and L. Nakhleh. Algorithms for MDC-based multi-locus phylogeny inference: Beyond rooted binary gene trees on single alleles. *Journal of Computational Biology*, 18(11):1-18, 2011.
- Y. Yu, J.H. Degnan, and L. Nakhleh. The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. PLoS Genetics, 8(4):e1002660, 2012.
- Y. Yu, R.M. Barnett, and L. Nakhleh. Parsimonious inference of hybridization in the presence of incomplete lineage sorting. Systematic Biology, vol. 62, no. 5, pp. 738-751, 2013.
- Y. Yu and L. Nakhleh. Fast algorithms for reconciliation under hybridization and incomplete lineage sorting. BMC Bioinformatics, vol. 14, no. Suppl 15, p. S6, 2013.
- Y. Yu, J. Dong, K. Liu, and L. Nakhleh, "Probabilistic inference of reticulate evolutionary histories," Proceedings of the National Academy of Sciences, vol. 111, no. 46, pp. 16448-16453, 2014.
- Y. Yu and Nakhleh, L., "A Maximum Pseudo-likelihood Approach for Phylogenetic Networks", BMC Genomics, vol. 16, no. Suppl 10, p. S10, 2015.
- D. Wen, Yu, Y., Hahn, M. W., and Nakhleh, L., "Reticulate Evolutionary History and Extensive Introgression in Mosquito Species Revealed by Phylogenetic Network Analysis", Molecular Ecology, vol. 25, pp. 2361-2372, 2016.
- D. Wen, Yu, Y., and Nakhleh, L., "Bayesian inference of species phylogenies under the multispecies network coalescent", PLoS Genetics, vol. 12, no. 5, p. e1006006, 2016.
- D.Wen and L. Nakhleh. Co-estimating reticulate phylogenies and gene trees on sequences from multiple independent loci. *Submitted*.