

InferNetwork_MP

Description

Infers a species network(s) with a specified number of reticulation nodes under MDC criterion using parsimony-based method. The reticulation nodes in the inferred network will have inferred inheritance probabilities associated with them. To find the optimal network, steepest descent is used. The species network and gene trees must be specified in the [Rich Newick Format](#). However, only topologies of them are used in the method.

The input gene trees can be gene tree distributions inferred from Bayesian methods like MrBayes. See the second example below.

Usage

```
InferNetwork_MP (gene_tree_ident1 [, gene_tree_ident2...]) numReticulations [-a taxa map] [-b threshold] [-s startingNetwork] [-fs] [-n numNetReturned] [-m maxNetExamined] [-d maxDiameter] [-h {s1 [,s2...]}] [-w (w1,...,w6)] [-f maxFailure] [-x numRuns] [-pl numProcessors] [-di] [resultOutputFile]
```

<i>gene_tree_ident1</i> [, <i>gene_tree_ident2</i> ...]	Comma delimited list of gene tree identifiers. See details .	mandatory
<i>numReticulations</i>	Maximum number of reticulations to added.	mandatory
<i>-b threshold</i>	Gene trees bootstrap threshold. Edges in the gene trees that have support lower than <i>threshold</i> will be contracted.	optional
<i>-a taxa map</i>	Gene tree / species tree taxa association .	optional
<i>-s startingNetwork</i>	Specify the network to start search. Default value is the optimal MDC tree.	optional
<i>-fs</i>	Fix the start tree for search. If specified and give a start tree (-s), the search will fix the topology of the start tree.	optional
<i>-n numNetReturned</i>	Number of optimal networks to return. Default value is 1.	optional
<i>-m maxNetExamined</i>	Maximum number of network topologies to examined. Default value is infinity.	optional
<i>-d maxDiameter</i>	Maximum diameter to make an arrangement during network search. Default value is infinity.	optional
<i>-h {s1 [, s2...]}</i>	A set of specified hybrid species.	optional
<i>-w (w1, ..., w6)</i>	The weights of operations for network arrangement during the network search. Default value is (0.1,0.1,0.15,0.55,0.15,0.15).	optional
<i>-f maxFailure</i>	The maximum number of consecutive failures before the search terminates. Default value is 100.	optional
<i>-x numRuns</i>	The number of runs of the search. Default value is 5.	optional
<i>-pl numProcessors</i>	Number of processors if you want the computation to be done in parallel. Default value is 1.	optional
<i>-di</i>	Output the Rich Newick string of the inferred network that can be read by Dendroscope .	optional
<i>resultOutputFile</i>	Optional file destination for command output.	optional

It is mandatory to specify the number of reticulation nodes to added to the starting network. The *-s* option allows the users to specify a starting network (can be a tree) for network search. Then starting from this network, *numReticulations* number of reticulation nodes will be added during steepest descent search. If the starting network is not specified, the optimal tree under MDC (command [inferST_MDC](#)) will be used. If it is not binary, a random resolution will be used. By default, only the first optimal species network will be returned. However, users can use *-n* option to ask for multiple optimal networks.

Simple hill climbing is used for the search. Users can specify the weights of six operations for network arrangement through option *-w*. The six weights correspond to adding a reticulation node, deleting a reticulation node, relocating the head of a reticulation edge, relocating the tail of an edge, reversing the direction of a reticulation edge and replacing a reticulation edge, respectively. By default, the search terminates when a preset limit of consecutive failures is reached (Default is 100, but users can change it through option *-f*). However, option *-m* allows users to specify the maximum number of networks examined during the search. Once that number is reached, the program will terminate and return the optimal network found so far. On the other hand, users can use option *-d* to specify the maximum diameter of an operation for network rearrangement, like what local-SPR does. In order to avoid getting stuck at some local optimum, it is recommended to performed the search multiple times, which users can specify by option *-x*.

By default, it is assumed that only one individual is sampled per species in gene trees. However, the option [*-a taxa map*] allows multiple alleles to be sampled. If users have a prior knowledge of the hybrid species, they can specify them using option *-h*.

If users want to run the computation in parallel. Please specify the number of processors through option *-pl*.

Examples

```
#NEXUS

BEGIN TREES;

Tree geneTree1 = (C,((B,D),A));
Tree geneTree2 = (B,(D,(C,A)));
Tree geneTree3 = (D,(B,(C,A)));
Tree geneTree4 = (D,(B,(C,A)));

END;

BEGIN PHYLONET;

InferNetwork_MP (geneTree1,geneTree2,geneTree3,geneTree4) 1;

END;
```

```
#NEXUS

BEGIN TREES;

Tree geneTree1 = [&W 0.8] (C,((B,D),A));
Tree geneTree2 = [&W 0.2] (B,(D,(C,A)));
Tree geneTree3 = [&W 0.7] (D,(B,(C,A)));
Tree geneTree4 = [&W 0.3] (D,(B,(C,A)));

END;

BEGIN PHYLONET;

InferNetwork_MP (geneTree1,geneTree2,geneTree3,geneTree4) 1;

END;
```

Command References

- 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, N. Ristic and L. Nakhleh. Fast algorithms and Heuristics for Phylogenomics 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, Under Review.
- Z. Cao, J. Zhu and L. Nakhleh. Empirical performance of tree-based inference of phylogenetic networks. *WABI 2019*, vol. 143, 21: 1--21:13.

See Also

- [List of PhyloNet Commands](#)