

# CalGTProb Command

## Description

Computes the log likelihood of a phylogenetic network given a collection of gene trees. The network must be specified in the [Rich Newick Format](#).

The likelihood can be computed using only topologies of gene trees, or using both topologies and branch lengths of gene trees. The latter one requires the input gene trees to be ultrametric.

The input gene trees can be gene tree distributions inferred from Bayesian methods like MrBayes.

## Usage

```
CalGTProb network_ident geneTreeList [-a taxa map] [-b threshold] [-bl] [-o] [-p (rel,abs)] [-r maxRounds] [-t maxTryPerBr] [-i improveThreshold] [-l maxBL] [-x numRuns] [-pl numProcessors] [-m ac|mul] [resultOutputFile]
```

<i>network_ident</i>	The name of the network. See <a href="#">details</a> .	mandatory
<i>geneTreeList</i>	Comma delimited list of gene tree identifiers or comma delimited list of sets of gene tree identifiers. See <a href="#">details</a> .	mandatory
<i>-m ac or mul</i>	Specify the algorithm used for computation (see reference). The default value is <i>ac</i> .	optional
<i>-a taxa map</i>	Gene tree / species tree <a href="#">taxa association</a> .	optional
<i>-b threshold</i>	Specifies gene trees bootstrap threshold. Edges in the gene trees that have support lower than <i>threshold</i> will be contracted.	optional
<i>-bl</i>	The branch lengths of the input gene trees need to be considered in the computation	optional
<i>-o</i>	The network has only topologies, so the branch lengths and inheritance probabilities of it need to be optimized.	optional
<i>-p (rel, abs)</i>	The original stopping criterion of Brent's algorithm. Default value is (0.01, 0.001).	optional
<i>-r maxRound</i>	Maximum number of rounds to optimize branch lengths for a network topology. Default value is 100.	optional
<i>-t maxTryPerBr</i>	Maximum number of trial per branch in one round to optimize branch lengths for a network topology. Default value is 100.	optional
<i>-i improveThreshold</i>	Minimum threshold of improvement to continue the next round of optimization of branch lengths. Default value is 0.001.	optional
<i>-l maxBL</i>	Maximum branch lengths considered. Default value is 6.	optional
<i>-x numRuns</i>	The number of runs of optimizing branch lengths and inheritance probabilities. 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

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.

The *-m* option is used to specify the algorithm for computation, where *mul* stands for the algorithm based on MUL-trees (Yu Et. Al, 2012) and *ac* stands for the algorithm based on ancestral configurations (Yu and Nakhleh, under review). They produce exactly the same result, but the latter one is more efficiently in general cases.

By default, the inference uses only the topologies of gene trees, however, users can also use both topologies and branch lengths of the gene trees to do the inference, by specifying option *-bl*.

By default, it is assumed that the network has branch lengths and inheritance probabilities associated with each reticulation nodes. If the network has only topologies, users can use option *-o* so that branch lengths and inheritance probabilities will be inferred. During optimization, we use Richard Brent's algorithm (from his book "Algorithms for Minimization without Derivatives", p. 79) to optimize the branch lengths. Users can use different options to control this process. Option *-p* allows users to specify the original stopping criterion of Brent's algorithm. More precisely, *abs* and *rel* define a tolerance  $tol = rel |x| + abs$ . We optimize the branch lengths one by one. For every branch, it terminates when either *maxTryPerBr* (option *-t*) trials have been made or the Brent's algorithm suggests so. Users can put an upper bound of the branch lengths through option *-l*. Optimization of all branch lengths consists of a round. After every round, if the improvement in terms of likelihood score is greater than that from last round by at least *improveThreshold* (option *-i*), we starts next round. A maximum of *maxRound* (option *-r*) rounds will be tried. Since the optimization is a heuristic, it should be performed multiple times in order to avoid getting stuck at some local optimum (option *-x*; default is 5).

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

## Examples

```
#NEXUS

BEGIN NETWORKS;

Network net = ((A:2,((B:1,C:1):1)X#H1:0::0.3):1,(D:2,X#H1:0::0.7):1);

END;

BEGIN TREES;

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

END;

BEGIN PHYLONET;

CalGTProb net (geneTree1,geneTree2,geneTree3);

END;
```

```
#NEXUS

BEGIN NETWORKS;

Network net = (((B)#H1,E),(C,A)),(#H1,D));

END;

BEGIN TREES;

Tree geneTree1 = ((C:3,((B:1,D:1):1,A:2):1):1,E:4);
Tree geneTree2 = (B:4,(D:3,(C:2,(A:1,E:1):1):1):1);
Tree geneTree3 = (D:4,(B:3,((C:1,E:1):1,A:2):1):1);
Tree geneTree4 = (D:3,((B:1,E:1):1,(C:1,A:1):1):1);

END;

BEGIN PHYLONET;

CalGTProb net (geneTree1,geneTree2,geneTree3,geneTree4) -bl -o;

END;
```

## Command References

- 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*. 2012.
- 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, Maximum Likelihood Inference of Reticulate Evolutionary Histories, *Proceedings of the National Academy of Sciences*, vol. 111, no. 46, pp. 16448-16453, 2014.

## See Also

- [List of PhyloNet Commands](#)