InferNetwork_ML_CV Command

Description

Infers a species network(s) using maximum likelihood, where K-fold cross-validation is used to account for model complexity. Only topologies of gene trees are used in the inference and all gene trees need to be binary.

Usage

```
InferNetwork_ML_CV (gene_tree_ident1 [, gene_tree_ident2...]) numReticulations [-cv numFolds] [-a taxa map] [-b threshold] [-s startingNetwork] [-h {s1 [,s2...]}] [-w (w1,...,w7)] [-x numRuns] [-m maxNetExamined] [-md moveDiameter] [-rd reticulationDiameter] [-f maxFailure] [-o] [-p (rel,abs)] [-r maxRounds] [-t maxTryPerBr] [-i improveThreshold] [-l maxBL] [-pl numProcessors] [-di] [resultOutputFile]
```

gene_tree_ident1 [, *gene_tree_ident2*...]	Comma delimited list of gene tree identifiers. See details.	mandatory
numReticulations	Maximum number of reticulations to added.	mandatory
-cv *numFolds*	Number of folds in K-fold cross-validation. Default value is 10.	optional
-b *threshold*	Gene trees bootstrap threshold. Edges in the gene trees that have support lower than *threshold* will be contracted.	optional
-a *taxa map*	Gene tree / species tree taxa association.	optional
-s *startingNetwork*	Specify the network to start search. Default value is the optimal MDC tree.	optional
-h *{s1 [, s2...]}*	A set of specified hybrid species.	optional
-w *(w1, ..., w7)*	The weights of operations for network arrangement during the network search. Default value is (0.1,0.1,0.15,0.55,0.15,0.15,2.8).	optional
-x *numRuns*	The number of runs of the search. Default value is 5.	optional
-m *maxNetExamined*	Maximum number of network topologies to examined. Default value is infinity.	optional
-md *moveDiameter*	Maximum diameter to make an arrangement during network search. Default value is infinity.	optional
-rd *reticulationDiameter*	Maximum diameter for a reticulation event (the distance between two parents of a reticulation node). Default value is infinity.	
-f *maxFailure*	Maximum consecutive number of failures for hill climbing. Default value is 100.	optional
-o	If specified, during the search, for every proposed species network, its branch lengths and inheritance probabilities will be optimized to compute its likelihood. Default value is false.	optional
-p *(rel, abs)*	The original stopping criterion of Brent's algorithm. Default value is (0.01, 0.001).	optional
-r *maxRound*	Maximum number of rounds to optimize branch lengths for a network topology. Default value is 100.	optional
-t *maxTryPerBr*	Maximum number of trial per branch in one round to optimize branch lengths for a network topology. Default value is 100.	optional
-i *improveThreshold*	Minimum threshold of improvement to continue the next round of optimization of branch lengths. Default value is 0.001.	optional
-l *maxBL*	Maximum branch lengths considered. Default value is 6.	optional
-pl *numProcessors*	Number of processors if you want the computation to be done in parallel. Default value is 1.	optional

-di	Output the Rich Newick string of the inferred network that can be read by Dendroscope .	optional
resultOutputFile	Optional file destination for command output.	optional

By default, 10-fold cross-validation is used to account for model complexity. But users can change the number of folds using option *-cv*.

See command InferNetwork_ML for all other parameters.

Examples

```
#NEXUS

BEGIN TREES;

TREE gt0 = (D:7.07072,((C:3.56753,B:3.56753):1.76822,A:5.33575):1.73497);
TREE gt1 = (((B:1.97661,C:1.97661):2.23864,A:4.21524):1.77987,D:5.99511);
TREE gt2 = ((C:4.31675,(B:3.14621,A:3.14621):1.17054):2.09695,D:6.4137);
TREE gt3 = ((D:5.83927,A:5.83927):0.566624,(B:1.80987,C:1.80987):4.59603);
TREE gt4 = ((D:5.77537,(B:1.77451,C:1.77451):4.00086):0.810136,A:6.58551);
TREE gt5 = (D:6.80413,(A:3.82444,(C:2.31671,B:2.31671):1.50773):2.97969);
TREE gt6 = (D:7.61541,(C:4.41986,(A:2.52336,B:2.52336):1.8965):3.19554);
TREE gt7 = ((A:4.99068,(C:3.03372,B:3.03372):1.95696):0.782212,D:5.77289);
TREE gt8 = (D:5.95232,((C:1.86462,B:1.86462):3.20304,A:5.06765):0.88467);
TREE gt9 = ((A:3.84332,(C:3.74007,B:3.74007):0.103251):4.82743,D:8.67075);
TREE gt10 = ((A:4.19291,(C:1.73235,B:1.73235):2.46056):2.60752,D:6.80043);
TREE gt11 = (A:7.35563,((B:1.7592,C:1.7592):4.0683,D:5.8275):1.52813);
TREE gt12 = (((B:3.14432,A:3.14432):0.613233,C:3.75755):1.95882,D:5.71638);
TREE gt13 = (((C:2.6677,B:2.6677):2.66994,A:5.33764):4.71766,D:10.0553);
TREE gt14 = ((C:4.00286,(B:3.0957,A:3.0957):0.907165):1.94607,D:5.94893);
TREE gt15 = (D:5.89168,(A:3.77037,(B:2.13754,C:2.13754):1.63283):2.12131);
TREE gt16 = (D:6.43412,(A:3.85214,(B:2.6542,C:2.6542):1.19794):2.58198);
TREE gt17 = (D:6.27103,((B:1.80021,C:1.80021):1.91286,A:3.71307):2.55796);
TREE gt18 = (((C:1.72131,B:1.72131):4.04022,A:5.76154):2.49476,D:8.25629);
TREE gt19 = (((B:1.8262,C:1.8262):2.49809,A:4.32429):3.80882,D:8.13311);
TREE gt20 = (D:6.95436,((B:2.60072,C:2.60072):1.85771,A:4.45842):2.49594);

END;


BEGIN PHYLONET;

InferNetwork_ML_CV (gt0,gt1,gt2,gt3,gt4,gt5,gt6,gt7,gt8,gt9,gt10,gt11,gt12,gt13,gt14,gt15,gt16,gt17,gt18,gt19,
gt20) 3 -x 5;

END;
```

Command References

- 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