

# Infer\_ST\_MDC Command

## Description

Infers the species tree using the “Minimize Deep Coalescence” (MDC) criterion. The input gene trees must be specified in the [Rich Newick Format](#). Gene trees must be rooted. Gene losses are allowed. The generated output trees will also be generated in the rich newick format.

The resulting species trees are displayed with the number of extra lineages in each branch. For example, consider the following inferred species tree:

```
((a:0,b:0):2,(c:0,d:0):1):0
```

In this species tree, there are two extra lineages in branch between node (a, b) and the root, and one extra lineage in branch between node (c, d) and the root. All other branches have 0 extra lineages.

## Usage

```
infer_ST_MDC (gene_tree_ident1 [, gene_tree_ident2...]) [-e proportion] [-x] [-b threshold] [-a taxa map] [-ur] [-t time] [result output file]
```

<i>gene_tree_ident1</i> [, <i>gene_tree_ident2</i> ...]	Comma delimited list of gene tree identifiers. See <a href="#">details</a> .	mandatory
<i>-e proportion</i>	Get optimal and sub-optimal trees.	optional
<i>-x</i>	Use all clusters in generation.	optional
<i>-b threshold</i>	Specifies 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 <a href="#">taxa association</a> .	optional
<i>-ur</i>	Allow non-binary species tree generation.	optional
<i>-t time</i>	Limit search time to <i>time</i> minutes.	optional
<i>result output file</i>	Optional file destination for command output.	optional

By default, the method returns the optimal tree. But the option *-e* allows the users to get the optimal tree and a set of sub-optimal trees. If the optimal tree has *n* extra lineages, all the sub-optimal trees that have extra lineages less than  $(1+proportion/100)*n$  will be returned with the optimal tree.

By default, the method uses clusters induced from gene trees to infer species tree. However, the option *-x* allows users to specify using all possible clusters to infer species tree.

If input gene trees have bootstrap values a threshold can be set with the *-b* option.

By default, the method will always return a binary species tree. But users can use option *-ur* to allow non-binary species tree. If the gene trees are not binary and the degree of resolution are low, it is recommended to use this option. Otherwise, the program will do some exhaustive search for a binary species tree. In this case, users can also use option *-t* to limit the search time. The time is in the unit of minutes.

By default, it is assumed that only one individual is sampled per species in gene trees. However, the option *-a* allows multiple alleles to be sampled.

## Examples

```
#NEXUS
```

```
BEGIN TREES;
```

```
Tree g1 = (((a,b),c),d),e);
```

```
Tree g2 = ((a,b),(c,e),d);
```

```
Tree g3 = ((a,c),(b,e),d);
```

```
END;
```

```
BEGIN PHYLONET;
```

```
Infer_ST_MDC (g1, g2, g3);
```

```
END;
```

```
#NEXUS
```

```
BEGIN TREES;
```

```
Tree g1 = (((a1::50,b1::50)::50,c::50)::50,d::50)::50,e::50)::50;
```

```
Tree g2 = ((a2::50,b2::50)::50,(c::50,e::50)::50,d::50)::50;
```

```
Tree g3 = ((a::50,c::50)::50,(b::50,e::50)::50,d::50)::50;
```

```
END;
```

```
BEGIN PHYLONET;
```

```
Infer_ST_MDC (g1, g2, g3) -b .5 -e .2 -x -ur -t 1 -a <z:a1,a2,a; y:b1,b2,b; c:c; d:d; e:e>;
```

```
END;
```

## Command References

- 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.

## See Also

- [List of PhyloNet Commands](#)