



Prunier: Detecting lateral gene transfers by statistical reconciliation of phylogenetic forests

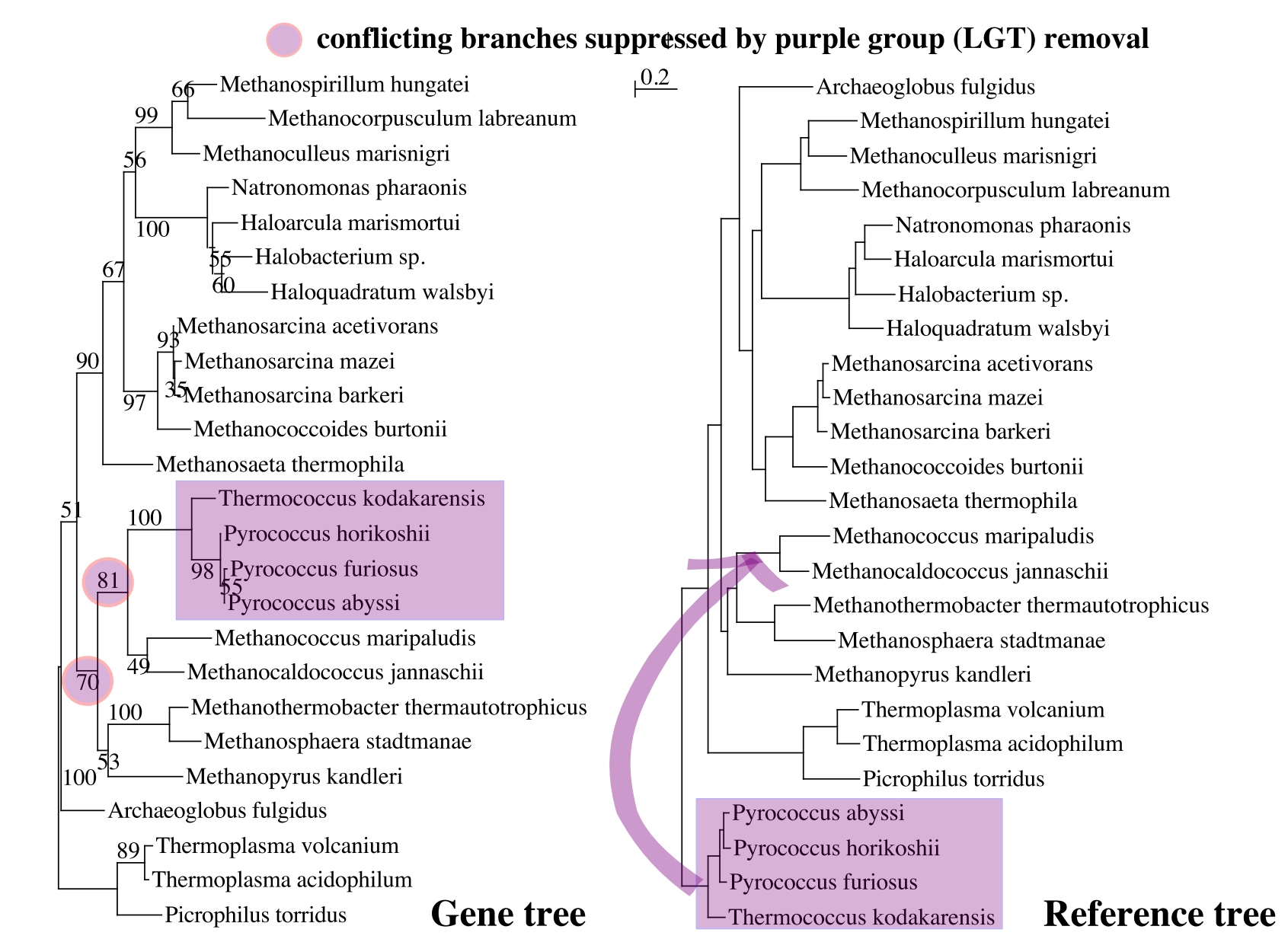
Sophie Abby, Eric Tannier, Manolo Gouy, Vincent Daubin

Laboratoire Biométrie et Biologie Evolutive
UMR 5558 CNRS; Université de Lyon; FRANCE



Introduction

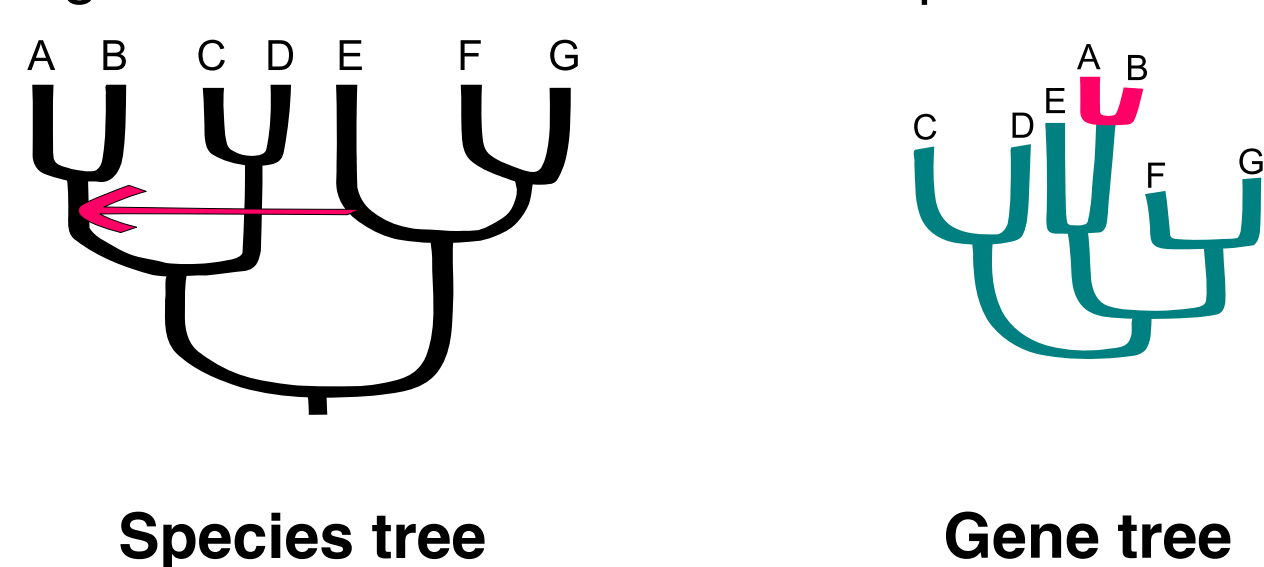
The availability of complete prokaryotic genomic sequences revealed an unforeseen number of lateral gene transfers (LGT). They are thought to blur the vertical signal that can be found in molecular markers, and thus to question the existence of the tree of life. An important challenge of phylogeny is the ability to detect them. Phylogenetic methods of LGT detection are among the most efficient. They are based on the comparison of a gene tree and a given species tree. Most of them consider gene trees as fully resolved, but gene trees are often partially unresolved, because of the lack of signal and/or stochastic effects in tree reconstruction process. We propose a new method for LGT detection, Prunier, that considers that gene trees can be partially unresolved and where LGT search is led by branch support in gene trees.



Example of a LGT detected by Prunier (Euryarchaeal tree of ribosomal protein S7)

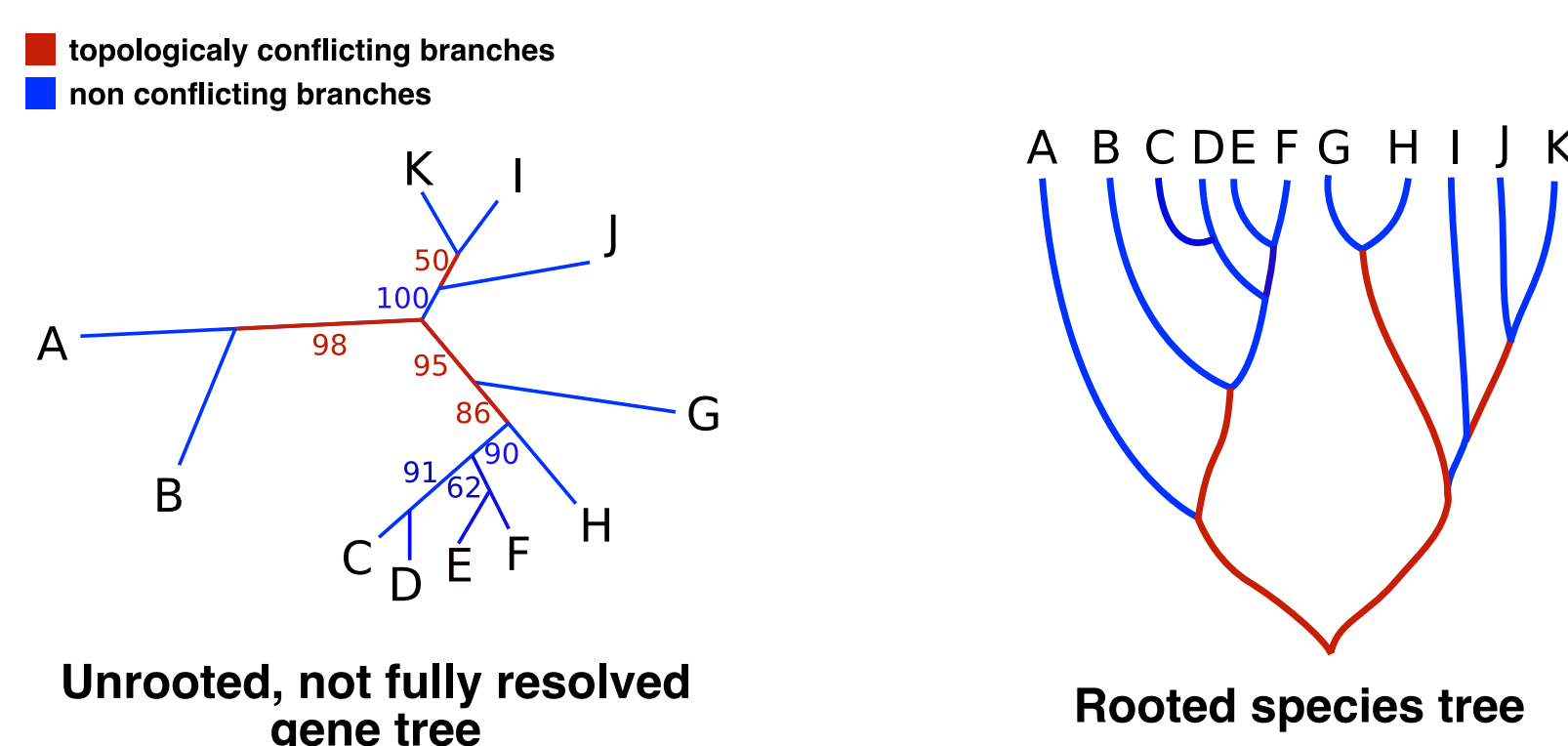
Phylogenetic detection of LGT

LGT make gene trees different from the species tree:

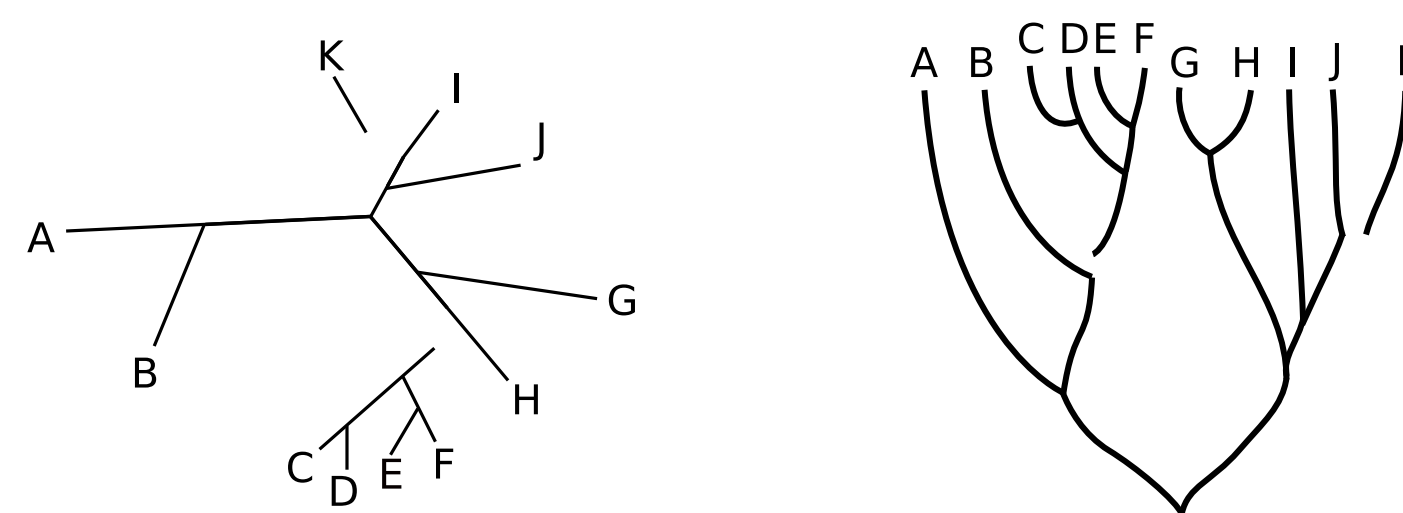


Phylogenetic methods of LGT detection try to sort out groups that look misplaced regarding a species tree.

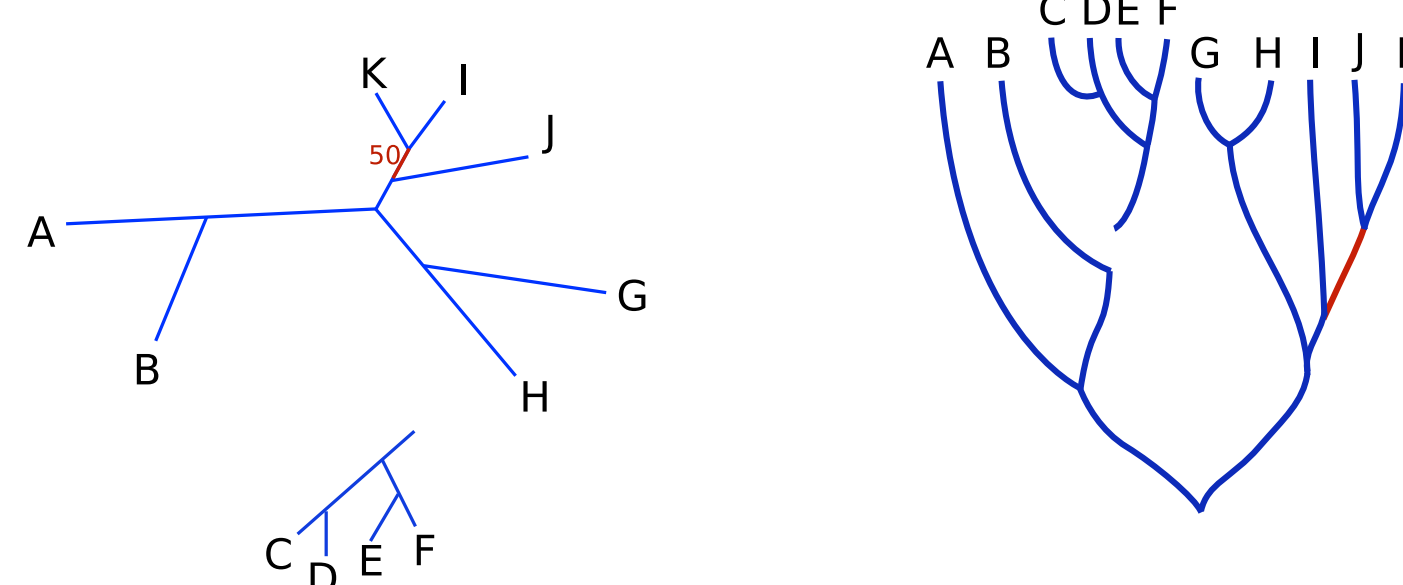
The data available typically look like this:



The Maximum Statistical Agreement Forest



Maximum Agreement Forest (Rodrigues 2001): find the minimum number of branches to cut to obtain **topologically** agreeing forests. Any topological difference is viewed as a LGT. But gene trees can be unresolved. We propose to **relax the topological criterion** in a derived version of the MAF :

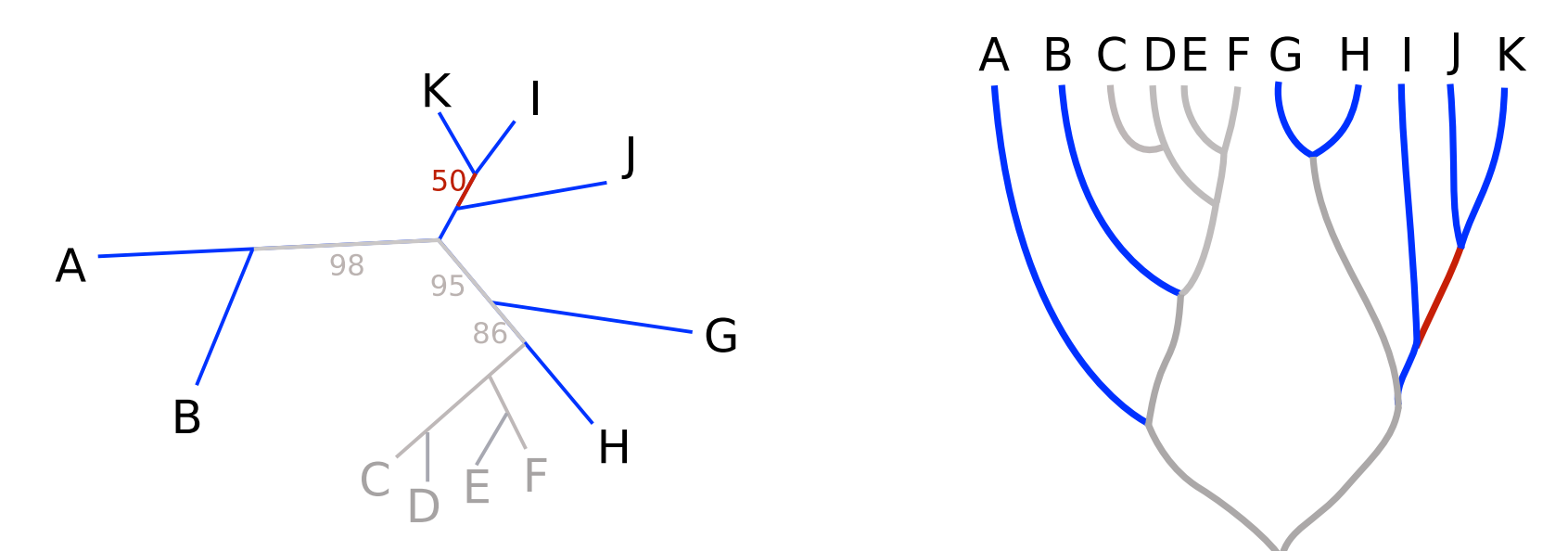


Maximum Statistical Agreement Forest (MSAF): find the minimum number of branches to cut to obtain **statistically** agreeing forests. Only significant differences (for ex. support > 80%) are viewed as LGT.

Prunier, a heuristic to search for the MSAF

We implemented Prunier, a **heuristic** to search for the MSAF.

We define a **conflict score** for each subtree, that depends on the support and number of conflicting branches its position produces:

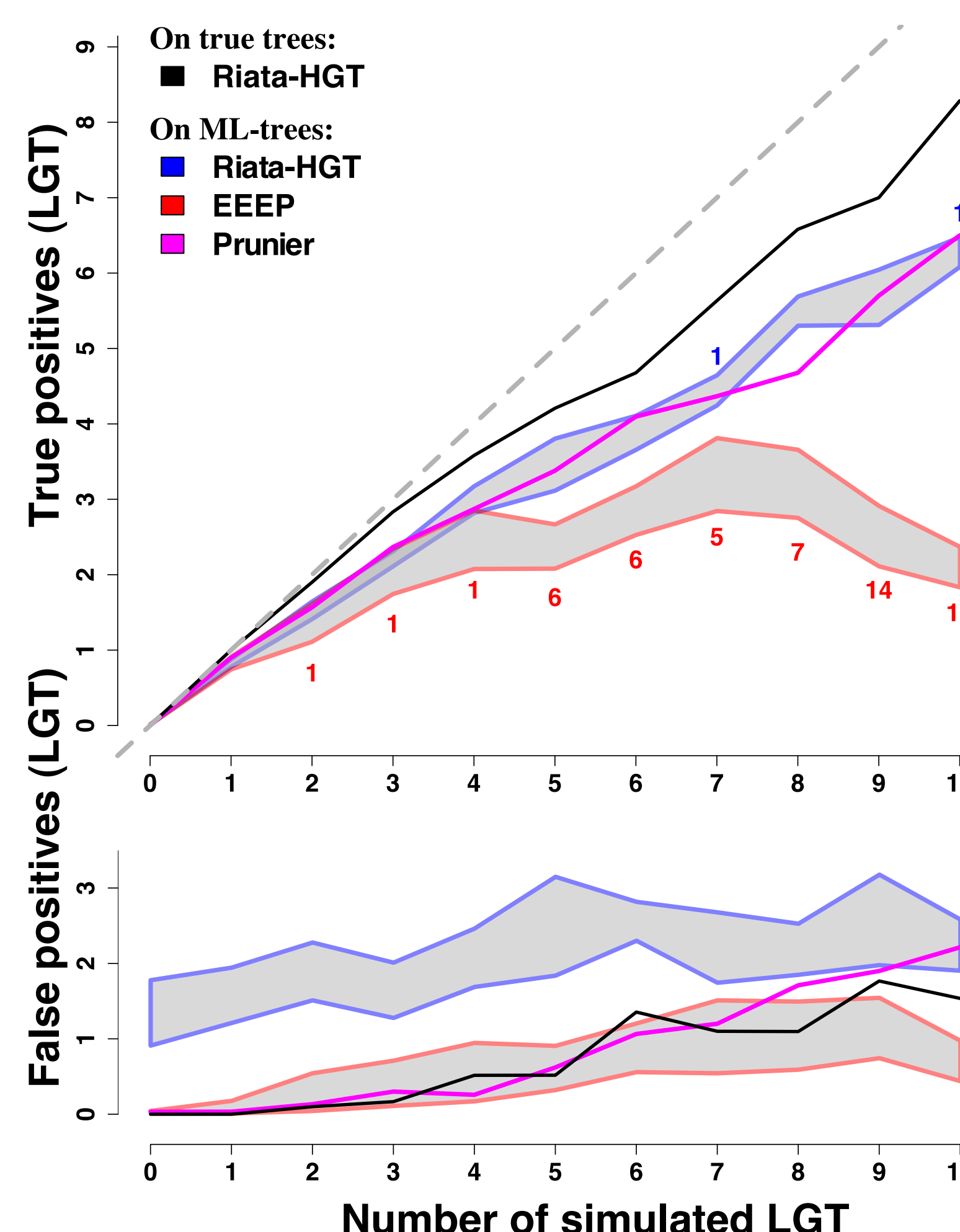
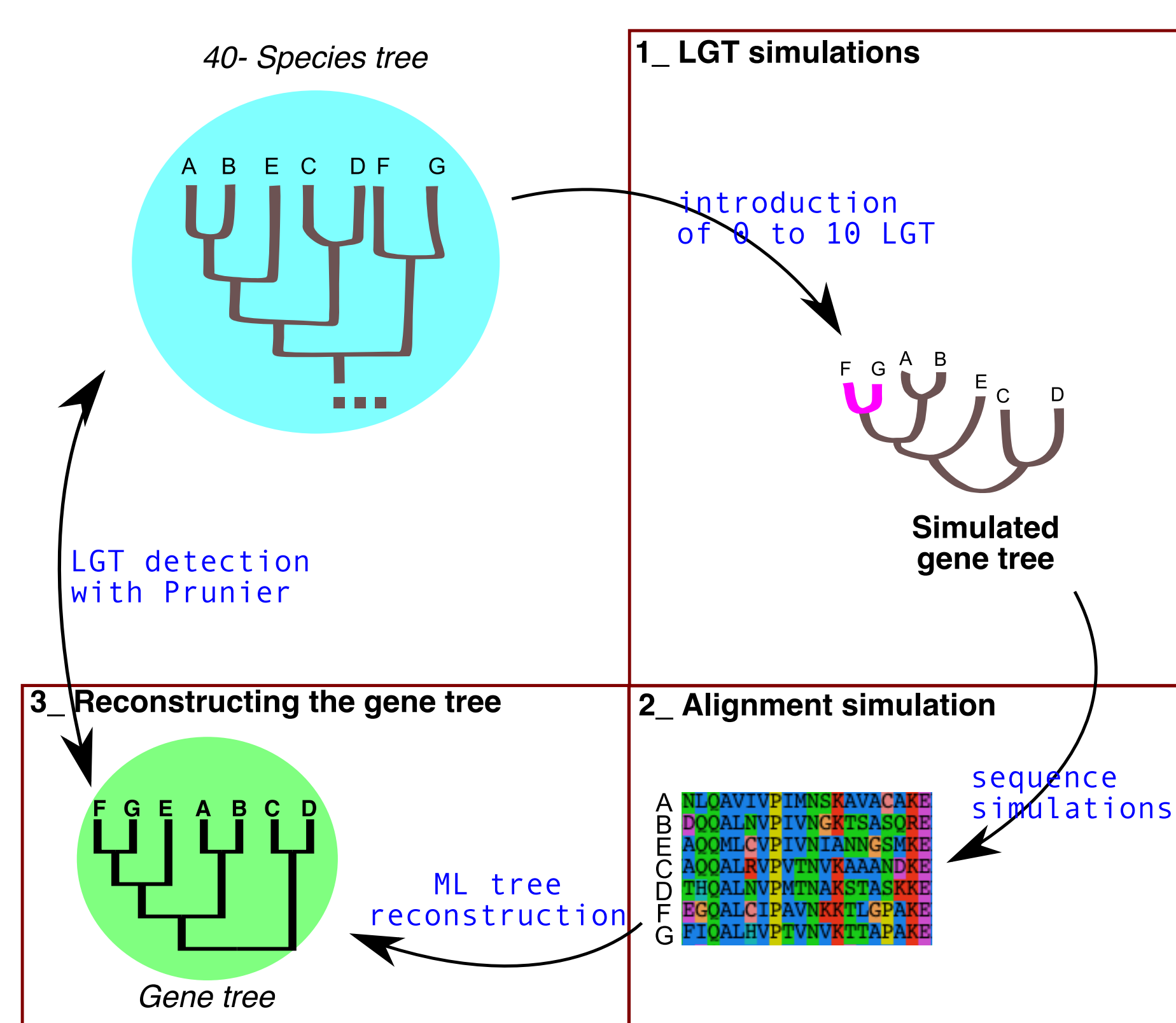


Removing the subtree {C,D,E,F} eliminates three conflicting branches, supported by 98, 95 and 95% respectively.

Prunier algorithm

- ▶ While the two forests statistically disagree
- ▶ Prune the subtree with the maximum conflict score

Prunier: results and conclusions



Conclusions:

- ▶ Prunier is more accurate, in particular it infers less false positives
- ▶ Prunier provides a unique scenario, but of high quality
- ▶ Demonstration of the relevance of guiding LGT search by branch support
- ▶ Prunier is able to deal with large biological data sets (data not shown)

Availability:

<http://pbil.univ-lyon1.fr/software/prunier/>

Reference:

Detecting lateral gene transfers by statistical reconciliation of phylogenetic forests
Abby SS, Tannier E, Gouy M, Daubin V
BMC Bioinformatics 2010, **11**:324

Acknowledgments:

Simon Penel, Bastien Boussau, Alain Guénoche, Nicolas Galtier, Pascal Calvat, Stéphane Delmotte, Lionel Humblot, Bruno Spataro