

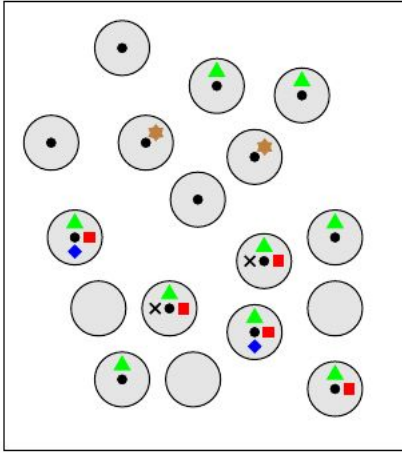
Compactly Representing a Set of Tumor Phylogenies

Brian Riccardi, Simone Moretti, **Murray Patterson**, Iman Hajirasouliha, Victoria Popic, Paola Bonizzoni, Simone Ciccolella, and Gianluca Della Vedova

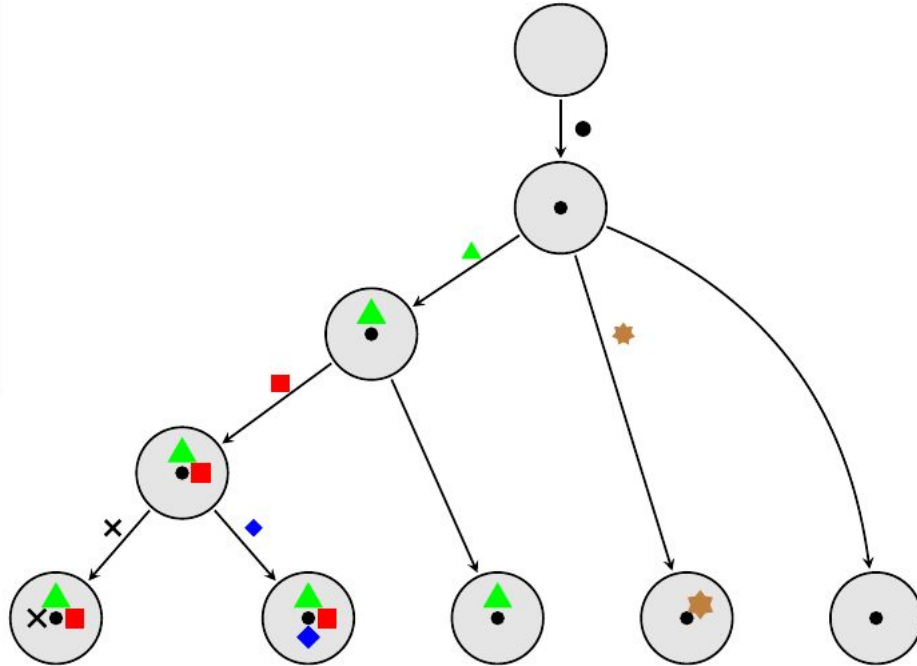
ICCABS 2026 North Carolina State University

Tumor phylogeny

observation: cells



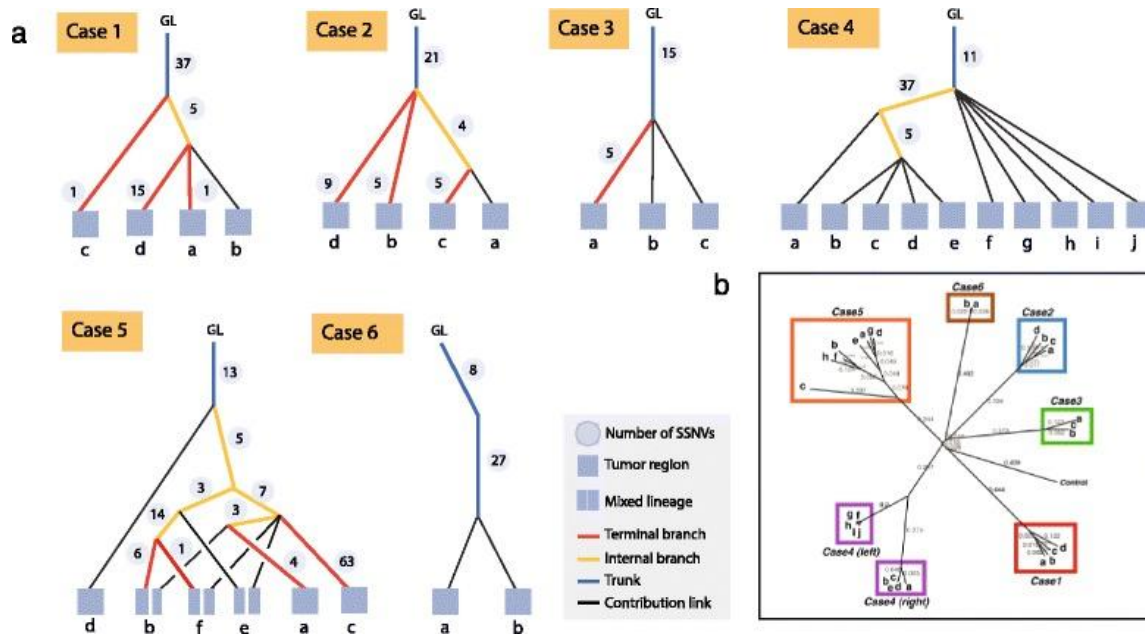
clonal evolution
(perfect phylogeny)



Many trees

There may be many different **equally probable** trees from different tools (or even by the same tool)

In **Popic et al., 2015**:
Comparison of lineage trees for HGSC Cases 1 to 6.
Phylogenetic trees obtained with LICHeE (a) are contrasted with the trees published in [Bashashati et al, 2013] (b) for each patient in the study. GL, germ line; HGSC, high-grade serous ovarian cancer; SSNV, somatic single nucleotide variant.



One thing to represent them all

Given a set of input phylogeny trees, we want a structure such that

- It contains all the input trees
- It does not contain any other tree
- It is compact

One thing to represent them all

Given a set of input phylogeny trees, we want a structure such that

- It contains all the input trees
- It does not contain any other tree
- It is compact

A tree is contained if it is an **arborescence** of the structure

One thing to represent them all

Given a set of input phylogeny trees, we want a structure such that

- It contains all the input trees
- It does not contain any other tree
- It is compact

Every arborescence is one of the input trees

One thing to represent them all

Given a set of input phylogeny trees, we want a structure such that

- It contains all the input trees
- It does not contain any other tree
- It is compact

Its size, e.g., number of nodes,
is minimized

One thing to represent them all

Given a set of input phylogeny trees, we want a structure such that

- It contains all the input trees
- It does not contain any other tree
- It is compact

But... this problem is difficult (maybe NP-hard?)

One thing to represent them all

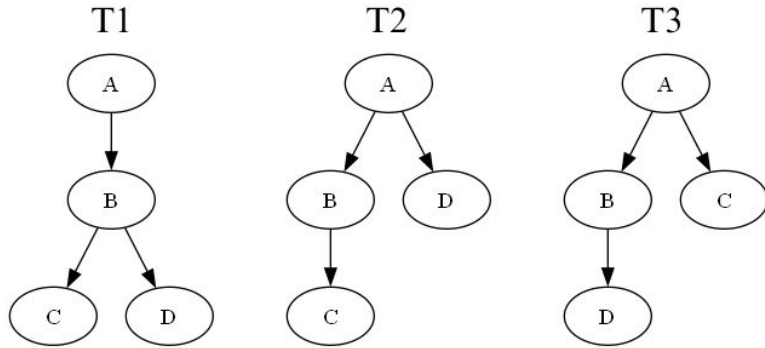
Given a set of input phylogeny trees, we want a structure such that

- It contains all the input trees
- It does not contain any other tree
- It is compact

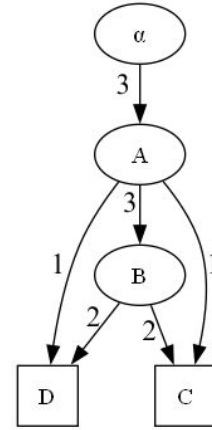
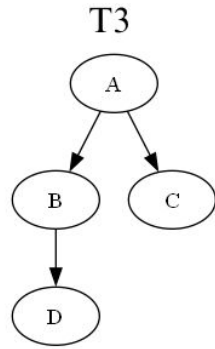
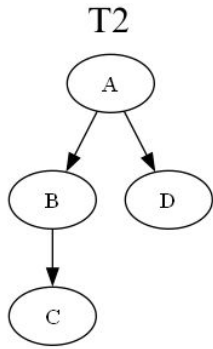
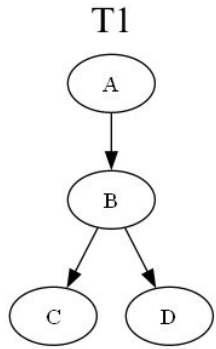
← Relaxing these makes it feasible

MegaTree

The MegaTree structure

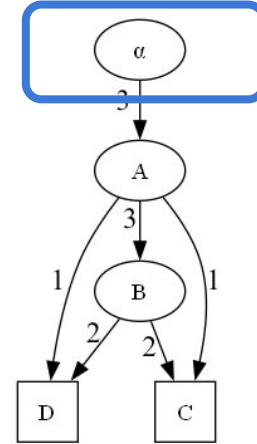


The MegaTree structure



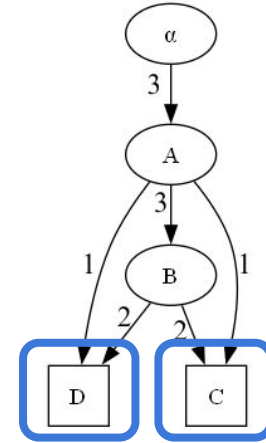
The MegaTree structure

- Super-root



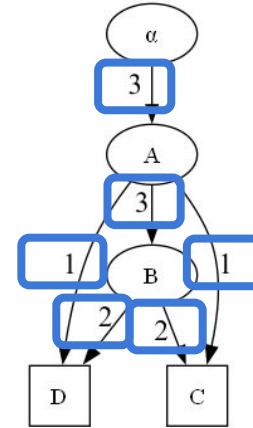
The MegaTree structure

- Super-root
- Terminal nodes



The MegaTree structure

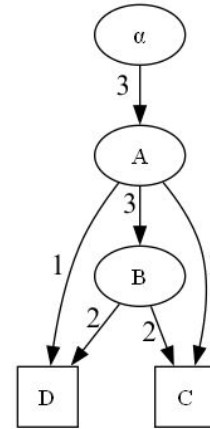
- Super-root
- Terminal nodes
- Optional weights



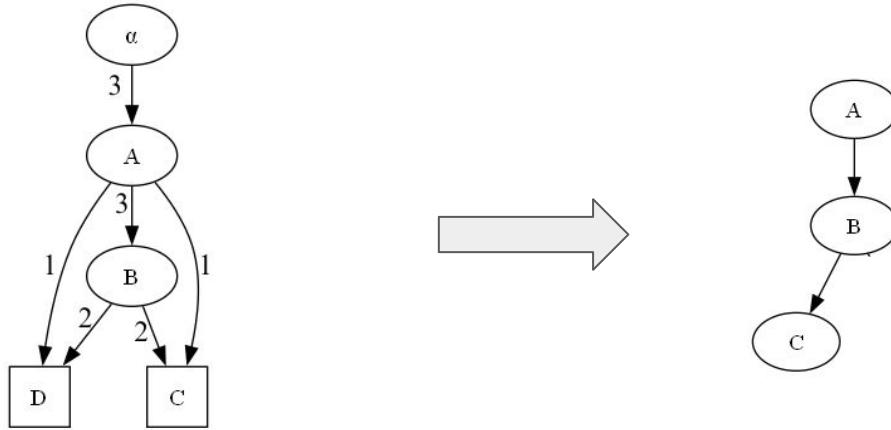
The MegaTree structure

- Super-root
- Terminal nodes
- Optional weights

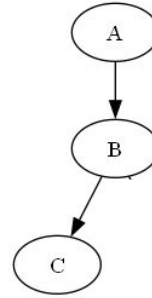
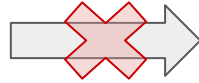
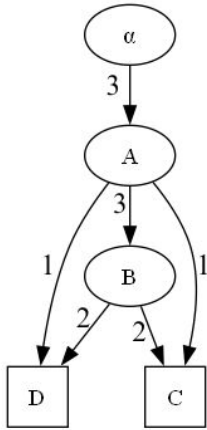
Every arborescence **must** contain **all** labels and have **only** terminals as leaves



Arborescences

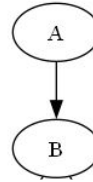
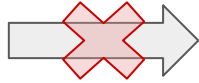
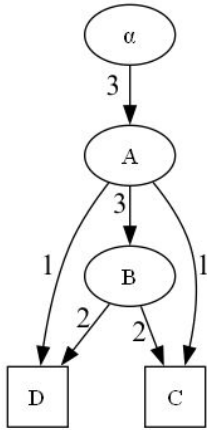


Arborescences



D is missing

Arborescences

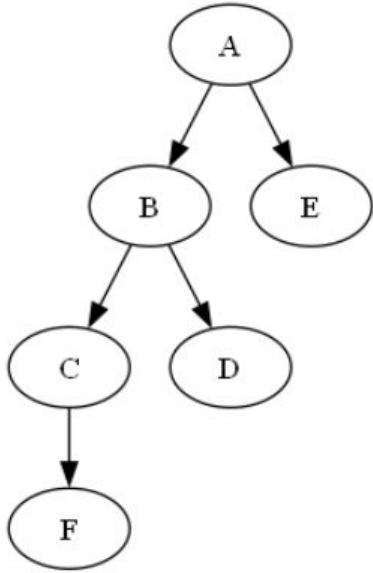


C,D is missing

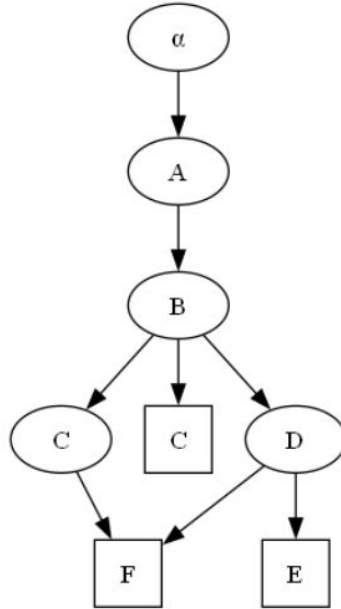
**B is not a
terminal node**

Heuristics

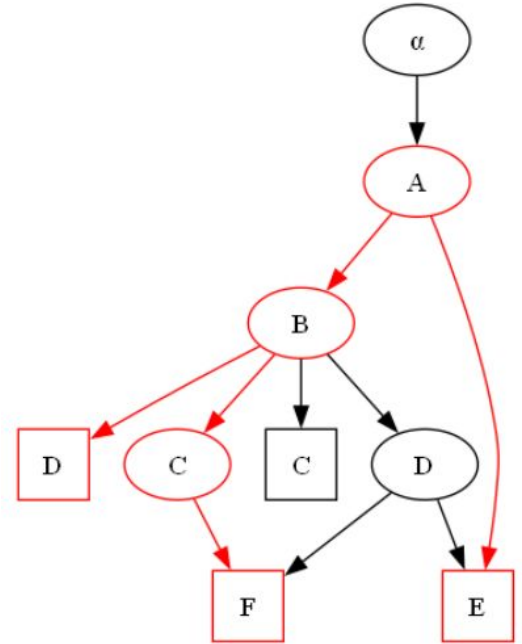
The Intersection Heuristic



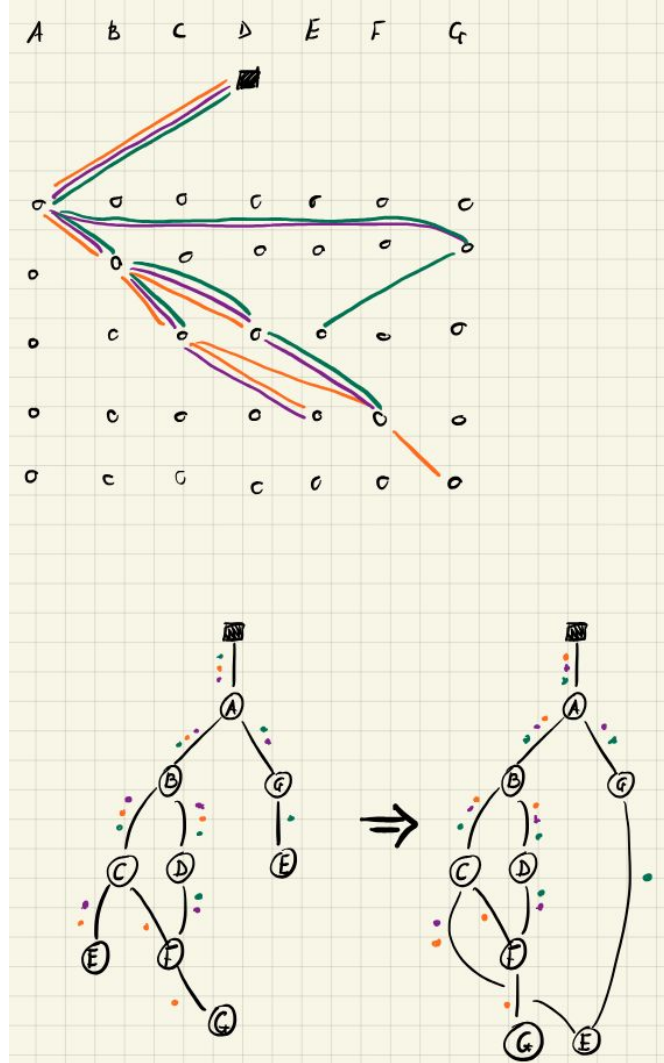
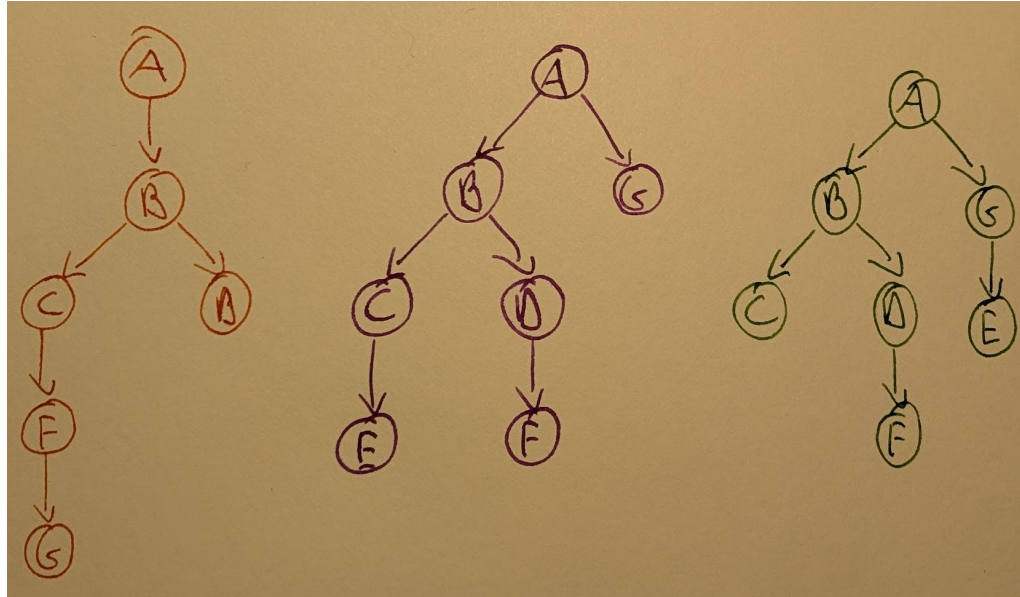
+



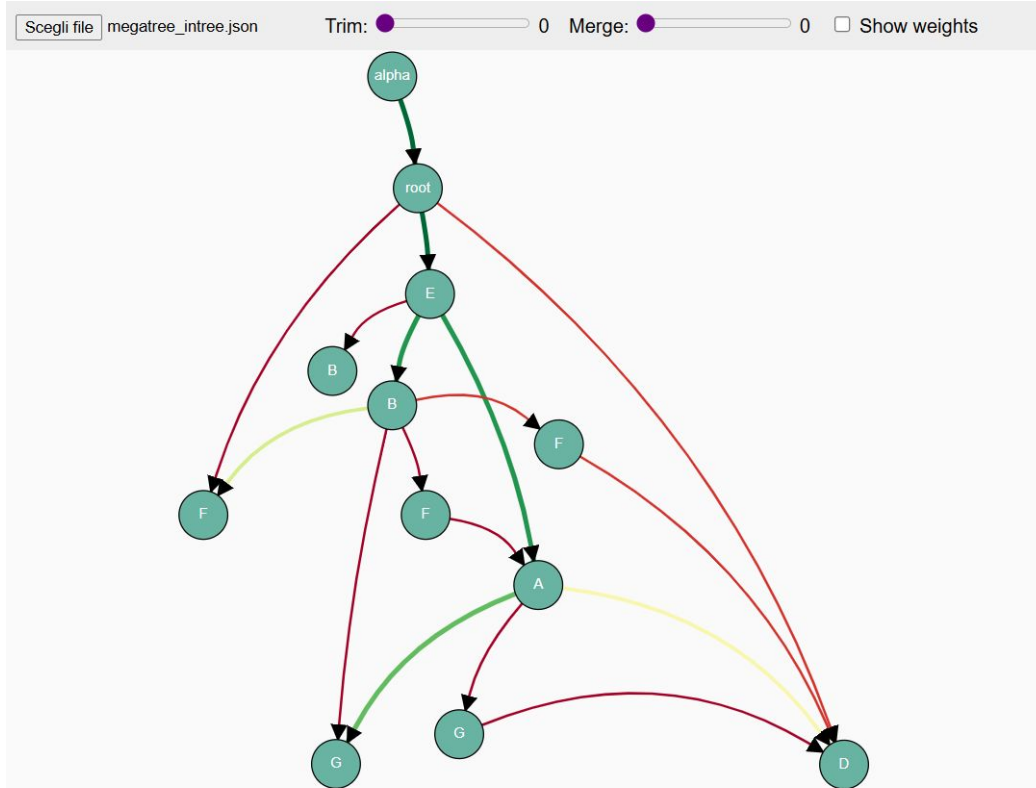
=



The Level-Based Heuristic



Implementation and visualization tool



- Interactive structure
- Color based on weights
- Trim nodes (based on weights)
- Merge nodes (based on weights)

Thank you!

