

# SUPER-RECONCILIATION WITH HORIZONTAL GENE TRANSFERS

Mattéo Delabre Nadia El-Mabrouk  
matteo.delabre@umontreal.ca

University of Montreal

## EVOLUTION OF SYNTENIES

- *Synteny*: Genomic regions derived from a shared ancestral region.
- Gene families in synteny *do not evolve independently* from each other [1].
- Must be taken into account to reconstruct accurate syntenic histories.
- *Super-Reconciliation*: Method for reconstructing the history of synteny through segmental gain and loss events.

## PREVIOUS WORK AND LIMITATIONS

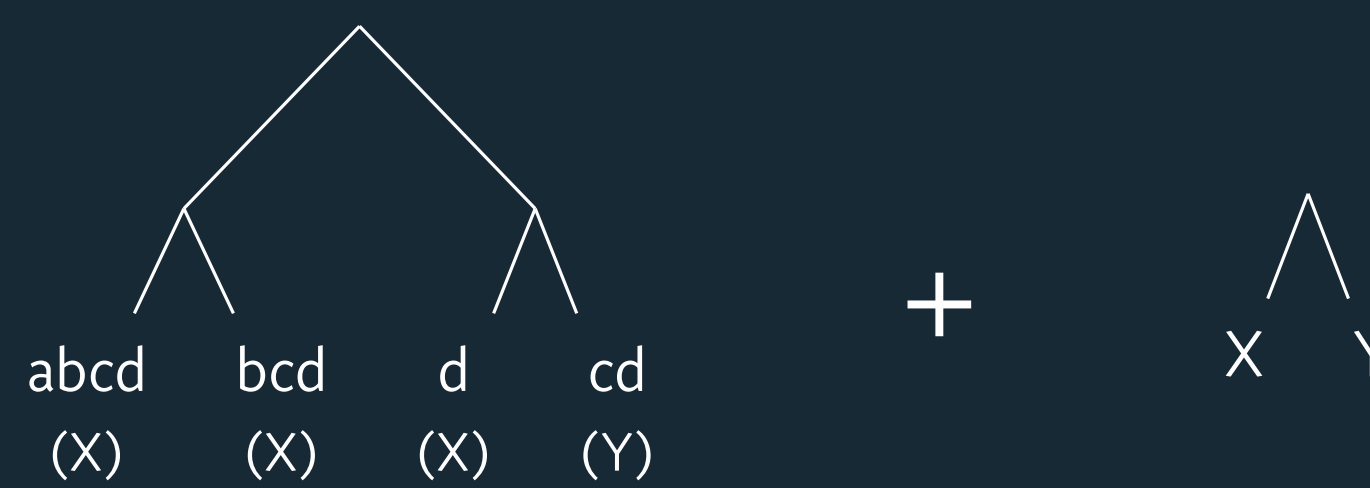
- Algorithms for Super-Reconciliation with *duplication and loss* events (DL), either with unlimited or without any gene rearrangements [2].
- Ignore important evolutionary mechanisms such as Horizontal Gene Transfers (HGTs) and tandem duplications.
- HGTs are essential to the study of the evolution of operons in bacteria.

## HOW TO EXTEND SUPER-RECONCILIATION TO INFER HGTs?

### SYNTENY TREE

Phylogenetic tree of the synteny of interest. Each synteny belongs to one of the species of the tree on the right. Computed, for example, from an *alignment* of the synteny or from a *supertree* of the individual gene trees.

### PROBLEM INPUT



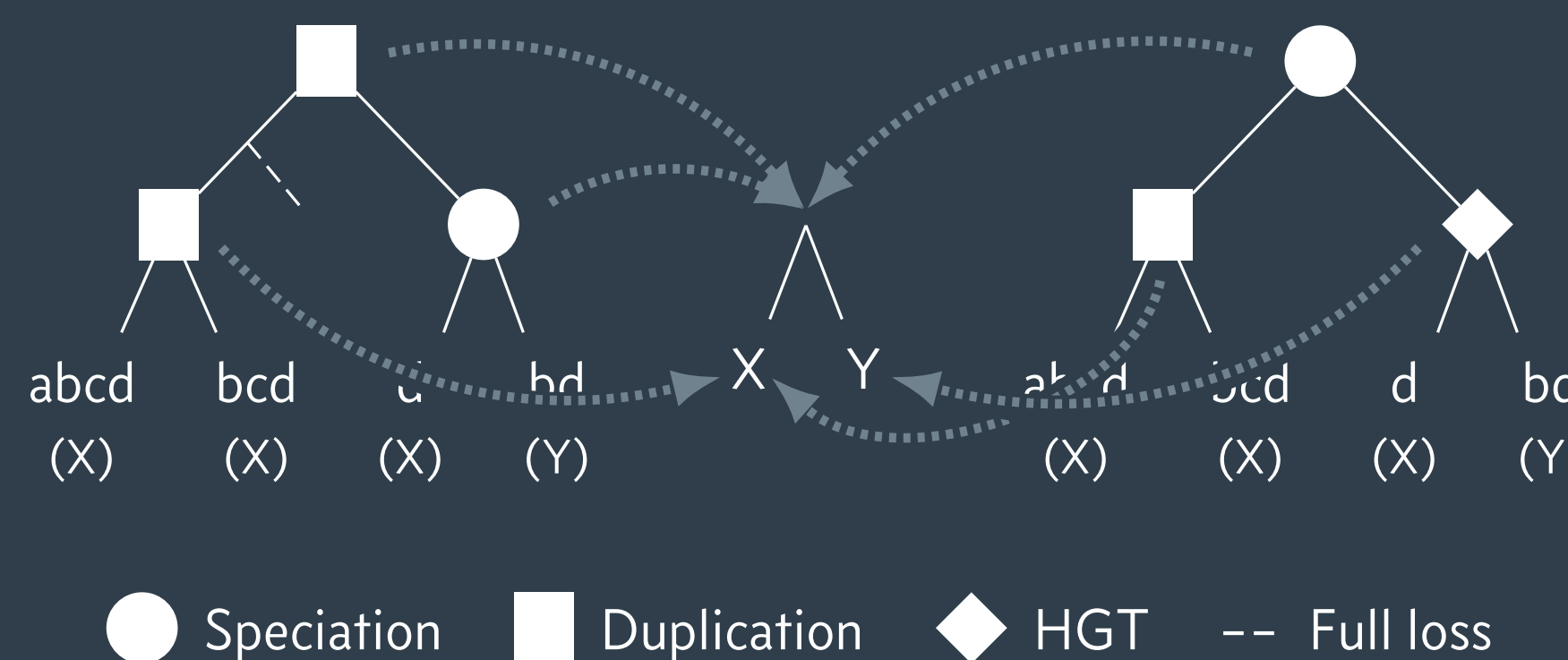
### SPECIES TREE

Phylogenetic tree of the genomes that contain the synteny of interest.

### WITHOUT HGTs

Only speciations, duplications, and full losses are allowed, so the only solution minimizing the number of non-speciation events (the "cost") is the one obtained via *LCA-mapping*. It can be computed in time  $O(n)$ , where  $n$  is the number of nodes in the synteny tree. This example has two duplications and one full loss, so its cost is three.

### 1 EVENT LABELLING



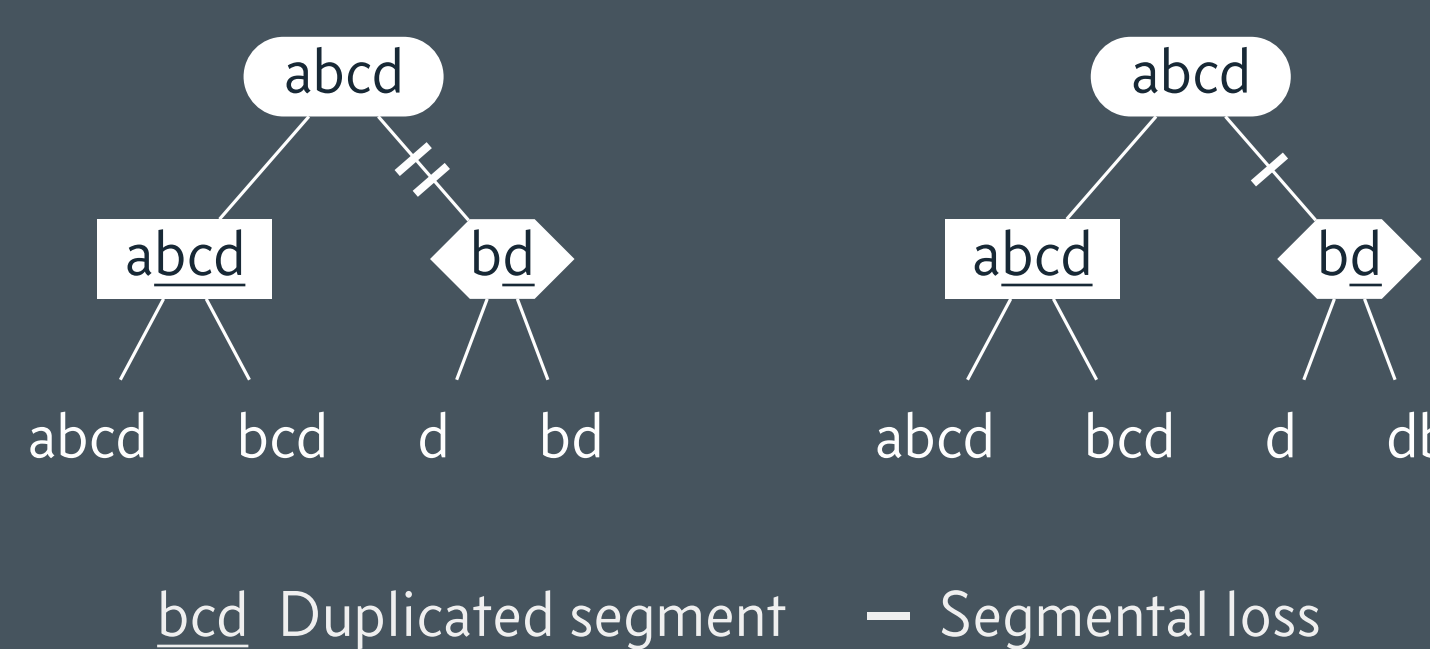
### WITH HGTs

More than one mapping can be minimal for the number of non-speciation events. *LCA-mapping* is not always minimal. This example has one duplication and one HGT, so its cost is two. An optimal solution can be found in time  $O(nm)$ , with  $m$  the number of nodes in the species tree [3].

### ORDERED SYNTENIES

Rearrangements are forbidden, so gene families must keep the same relative order in all of the synteny. The number of segmental losses induced by a labelling is added to the overall cost of a candidate solution, so this example has an overall cost of four. Computing a labelling that minimizes the overall cost is a NP-complete problem that can be solved in time  $O(nt2^t \log t + t)$ , where  $t$  is the number of gene families [2].

### 2 SYNTENY LABELLING

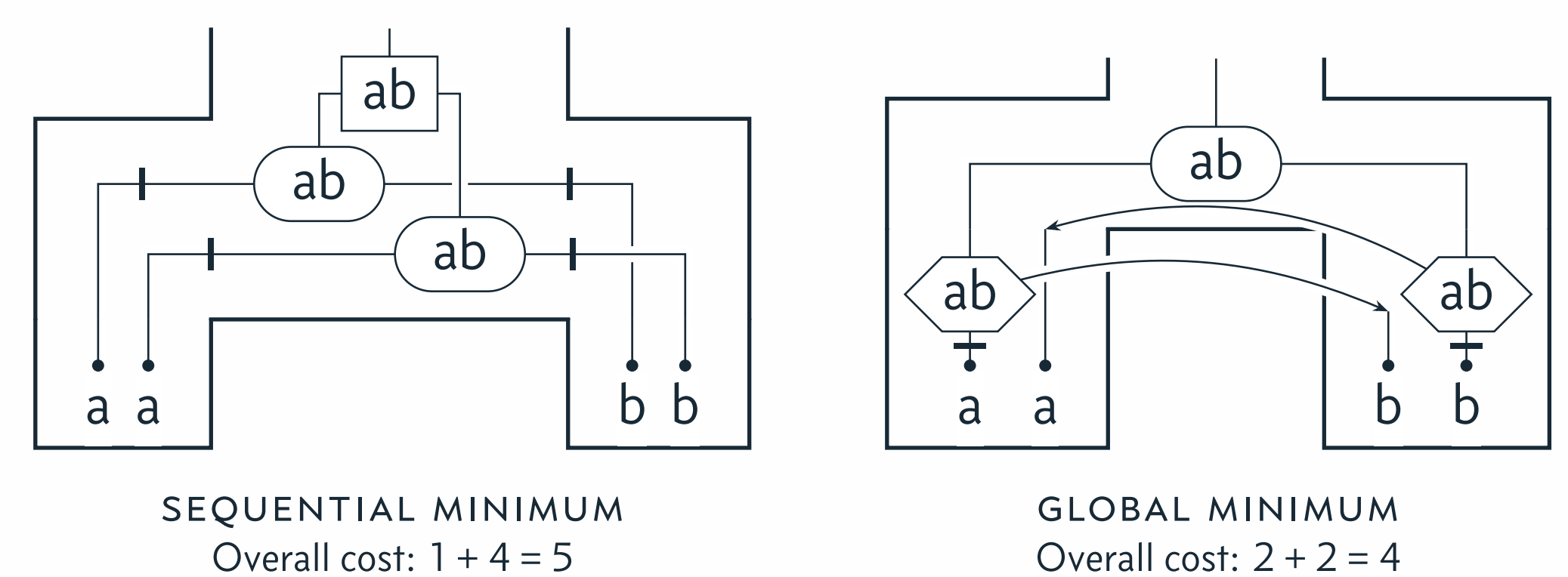


### UNORDERED SYNTENIES

Rearrangements are allowed but not minimized. Note the lower-right synteny of this example, whose gene families' order differs from the others. As a consequence of this, each edge can have at most one segmental loss. This example has an overall cost of three. A solution that minimizes the overall cost can be found in time  $O(nt)$  [2].

### UPDATED ALGORITHMS

- Without HGTs, both parts of the super-reconciliation can be optimized sequentially.
- When HGTs are allowed, we need to explore all combinations to find a global minimum (see the example on the right).
- Via dynamic programming:  $O(nm^2t^2 \log t + t)$  (ordered) or  $O(nm^2t)$  (unordered).



### FUTURE WORK

- Validation of the model via simulations.
- Application to real-world synteny evolution studies.
- Further extensions: tandem duplications, gene gains.

### REFERENCES

- [1] El-Mabrouk. "Predicting the evolution of synteny—An algorithmic review" (May 2021).
- [2] Delabre et al. "Evolution through segmental duplications and losses: a Super-Reconciliation approach" (May 2020).
- [3] Bansal et al. "Efficient algorithms for the reconciliation problem with gene duplication, horizontal transfer and loss" (2012).