

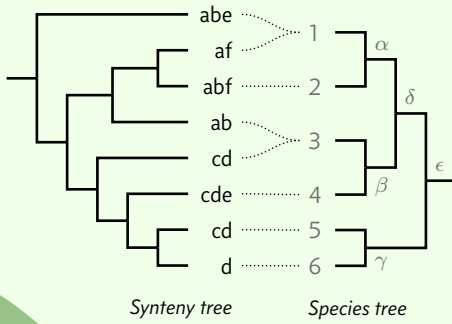
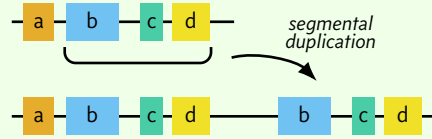
# INFERRING SCENARIOS FOR GENE SYNTENY-SPECIES COEVOLUTION THROUGH SEGMENTAL DUPLICATIONS, TRANSFERS, CUTS, GAINS AND LOSSES

Mattéo Delabre<sup>1</sup>  
Nadia El-Mabrouk<sup>1</sup>  
Yoann Anselmetti<sup>2</sup>

<sup>1</sup>LBIT, Université de Montréal, Canada  
<sup>2</sup>CoBIUS Lab, Université de Sherbrooke, Canada

## EVOLUTION OF GENE SYNTENIES

- Syntenies: groups of genes evolving together through *segmental events* (e.g., Cas systems, operons, ...)
- Also *coevolving* with their host species, similarly to individual gene families

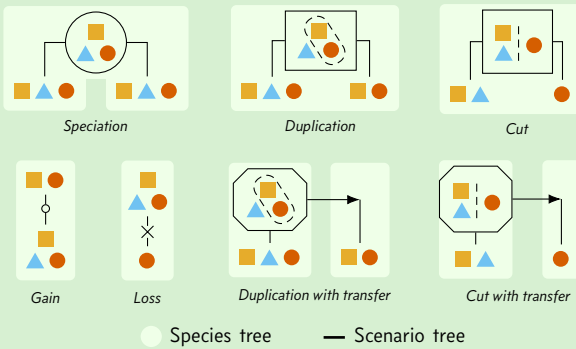


## (SUPER-)RECONCILIATIONS

- The gene-species evolutionary relationship creates *inconsistencies* between both phylogenies
- Both can be *reconciled* by postulating past events (duplications, lateral transfers, ...)
- These events form candidate *coevolution scenarios*
- *Super-reconciliation* extends this model to segmental events that describe *synteny evolution*

## EVOLUTION MODEL

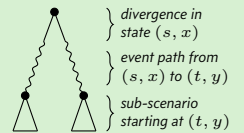
- Syntenies are *sets of genes* (disregarding gene orders)
- *Divergence events* create new synteny copies
- *Unary events* change existing syntenies



Given costs for each event type, can we find the most parsimonious (least costly) scenarios for a given tree pair?

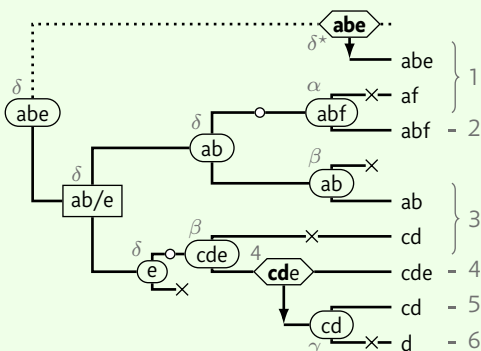
## BUILDING SCENARIOS WITH EVENT PATHS

- Joining *sub-scenarios* with *event paths* gives new scenarios
- Event paths start at a given state and have a single *visible leaf*: the target state
- The set of possible optimal event paths leading from  $(s, x)$  to  $(t, y)$  is of constant size



## SCENARIOS AND MISSING SPECIES

- Scenario: tree of events whose *leaves and topology* match those of the studied synteny tree
- Events happen along the specified species tree
- *Unsampled or extinct species* can be postulated



One of the optimal scenarios for the trees above  
(losses cost 1, duplications/cuts cost 2, transfers cost 3)

## DYNAMIC PROGRAMMING FORMULATION

- $c(v, s, x)$ : least cost of scenarios for the synteny tree below  $v$  where  $v$ 's host is  $s$  and contents are  $x$
- Only two contents need to be tried: the *minimum valid set* and any *bigger set*
- Compute  $c(v, s, x)$  recursively by evaluating the cost of each possible divergence event

## ONGOING AND FUTURE WORK

- Preliminary implementation available at [github.com/UdeM-LBIT/superrec2](https://github.com/UdeM-LBIT/superrec2)
- Ongoing evaluation on simulated and real data

Open questions:

- Build synteny trees from individual gene trees
- Integrate fusion events (reconciliation of networks)
- Minimize number of postulated unsampled species