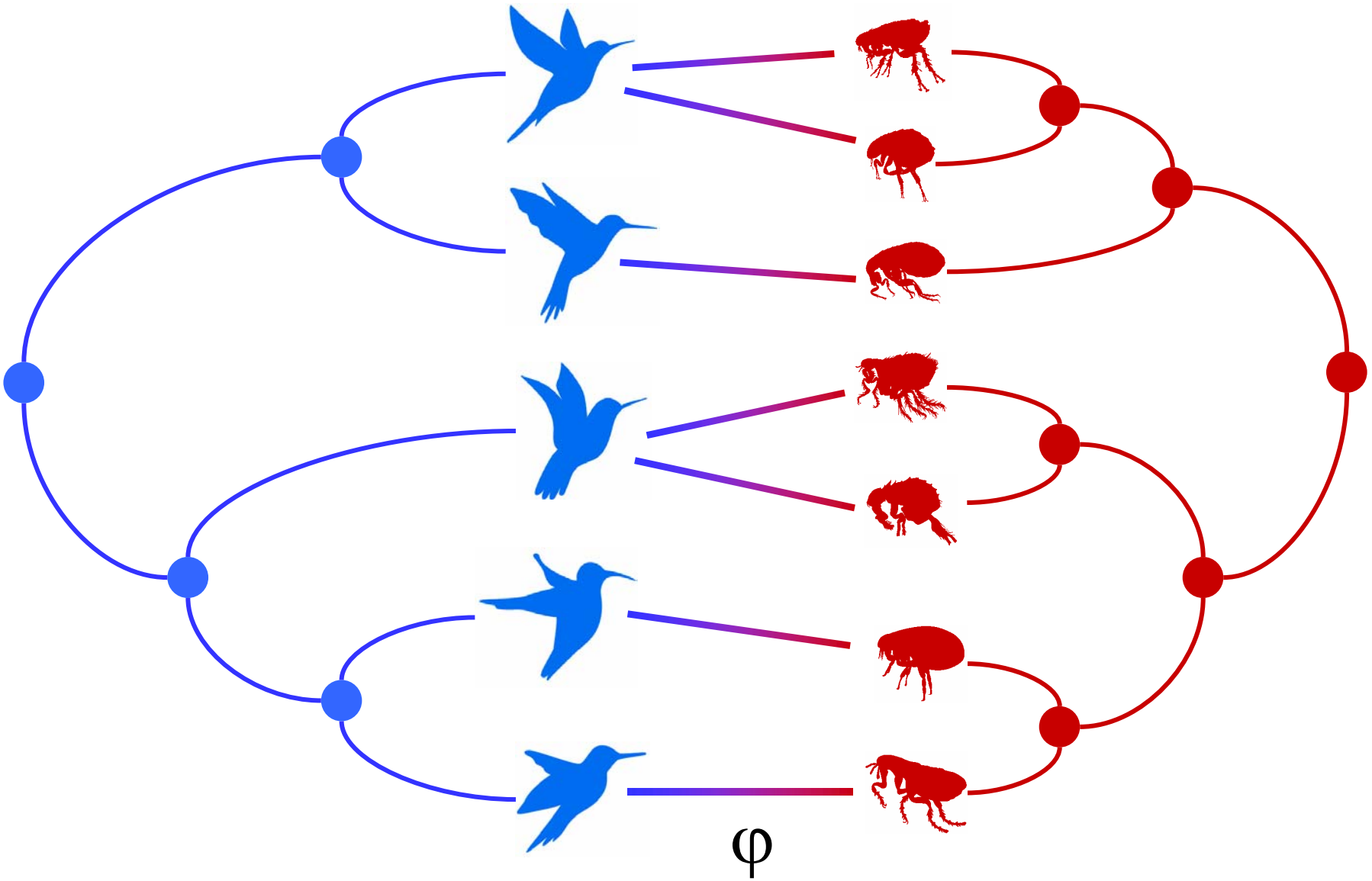
Visualizing Co-Phylogenetic Reconciliations

Tiziana Calamoneri[1], Valentino Di Donato[2], Diego Mariottini[2], and Maurizio Patrignani[2]

1 University of Rome "Sapienza", Rome, Italy
2 Roma Tre University, Rome, Italy

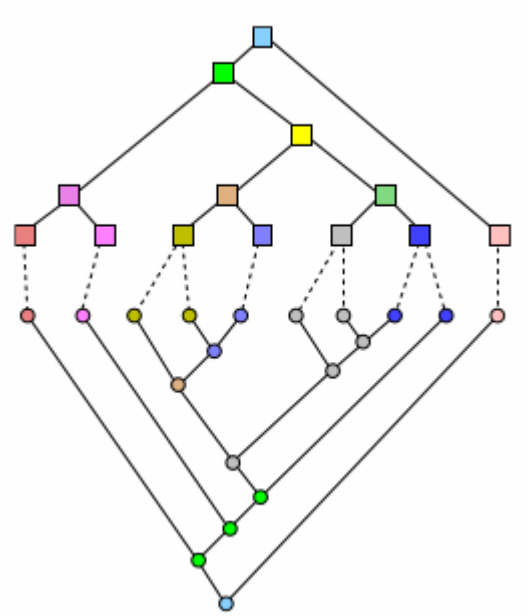
Co-phylogenetic trees



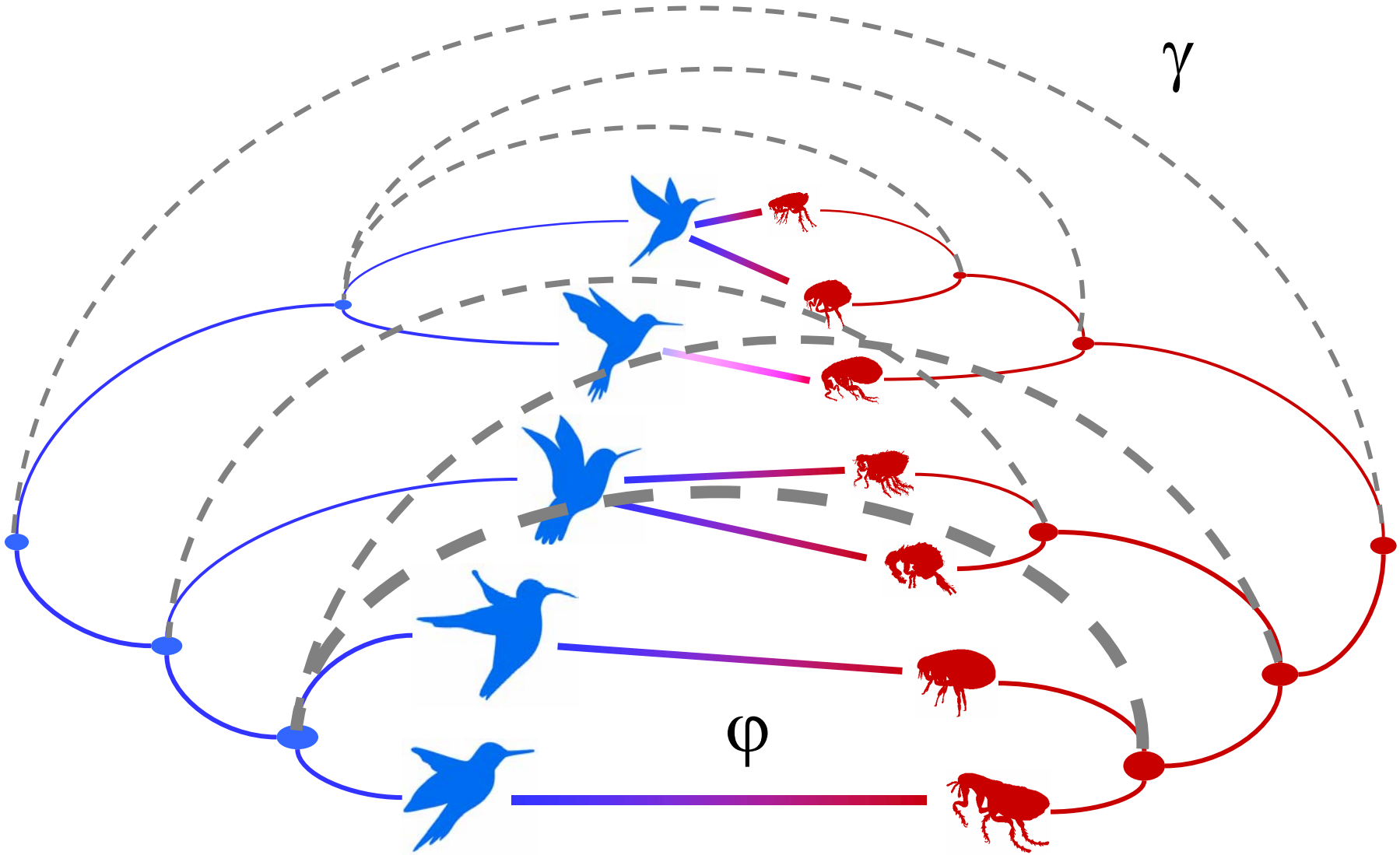
φ

Tanglegram drawings

- Tanglegram planarity
  - linear
    - [Fernau et al., J. Comput. Syst. Sci., 2010]
- Tanglegram crossing minimization
  - Polynomial for one-sided one-to-one leaf associations
    - [Dwyer and Schreiber, APVis 2004]
    - [Fernau et al., J. Comput. Syst. Sci., 2010]
  - NP-complete even for binary trees and for one-to-one leaf associations
    - [Fernau et al., J. Comput. Syst. Sci., 2010]
  - NP-complete even for complete binary trees and one-to-one leaf associations
    - [Buchin et al., Algorithmica 2012]

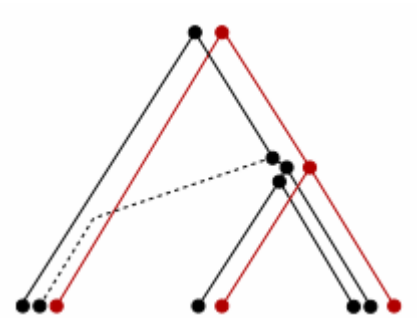


Co-phylogenetic reconciliation

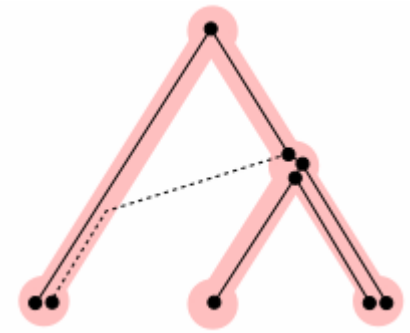


Representing reconciliations

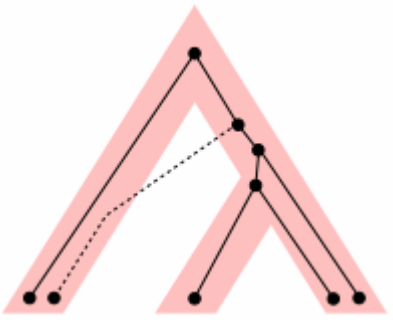
- Three main strategies
  1. representing two paired trees
  2. parasites are drawn inside their hosts
  3. host tree is made of pipes and parasites are drawn into the pipes



Strategy 1



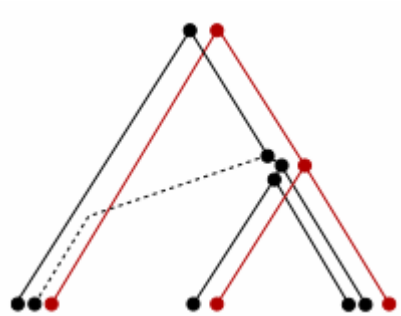
Strategy 2

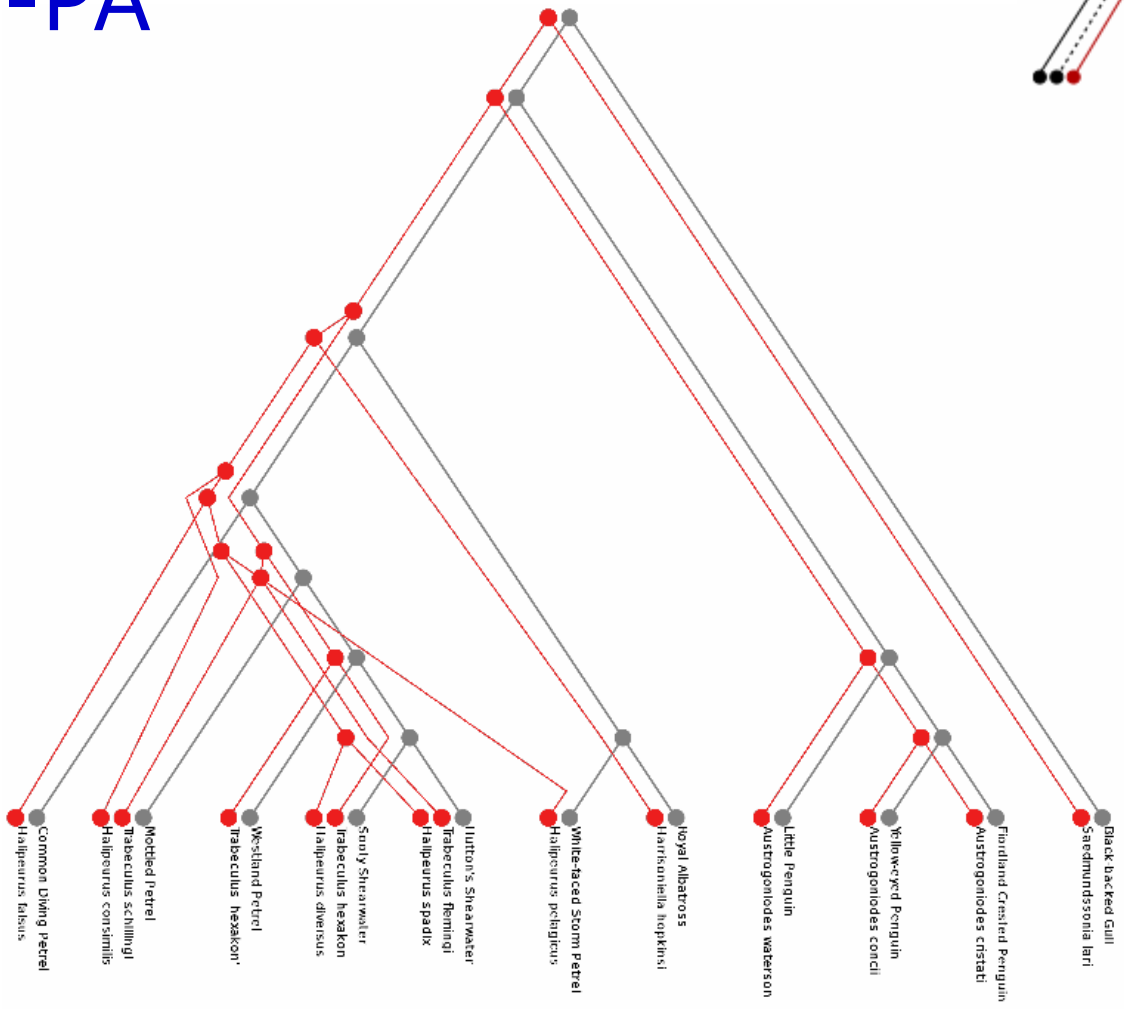


Strategy 3

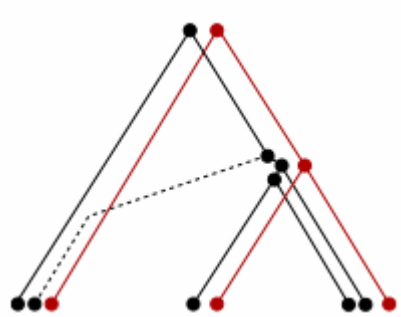
Example of the 1st strategy

- CoRe-PA

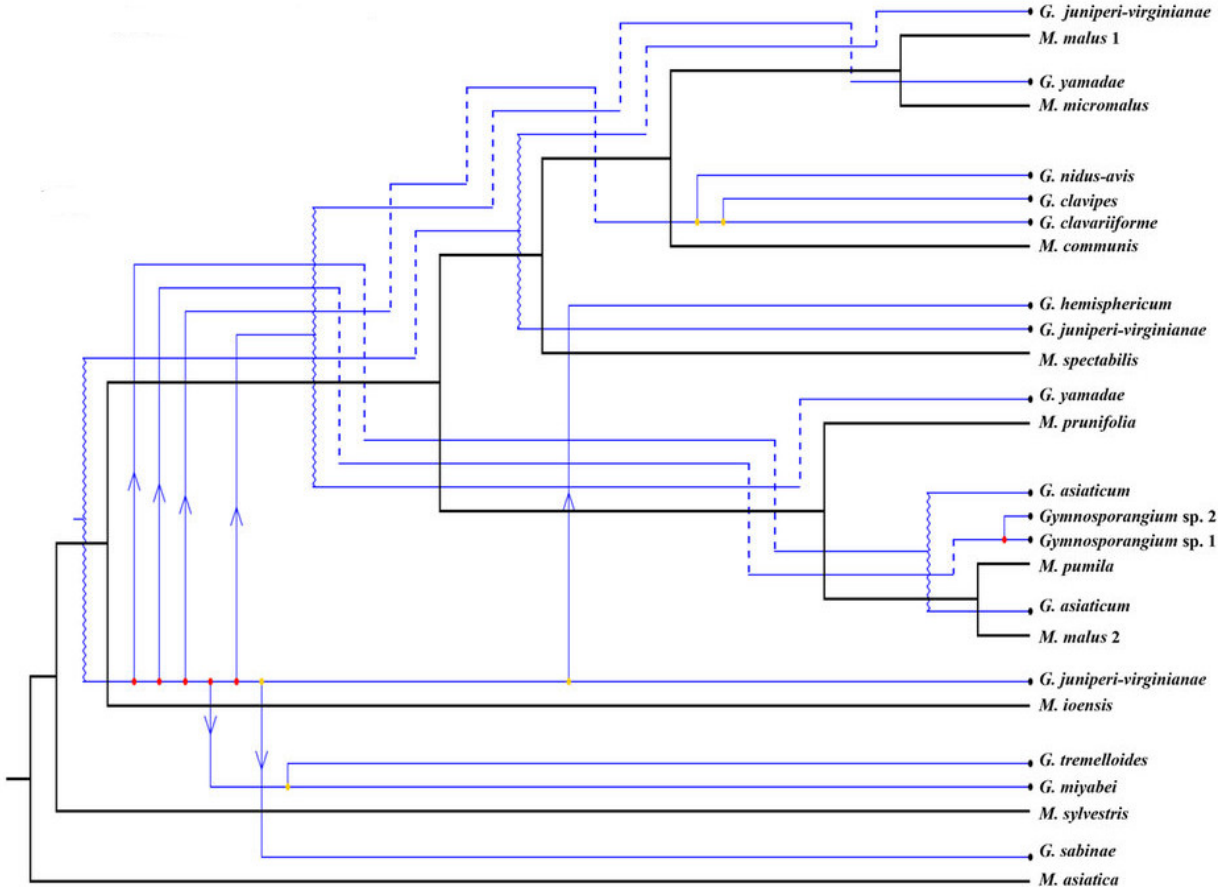




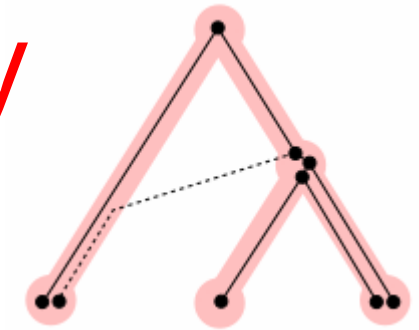
Example of the 1st strategy



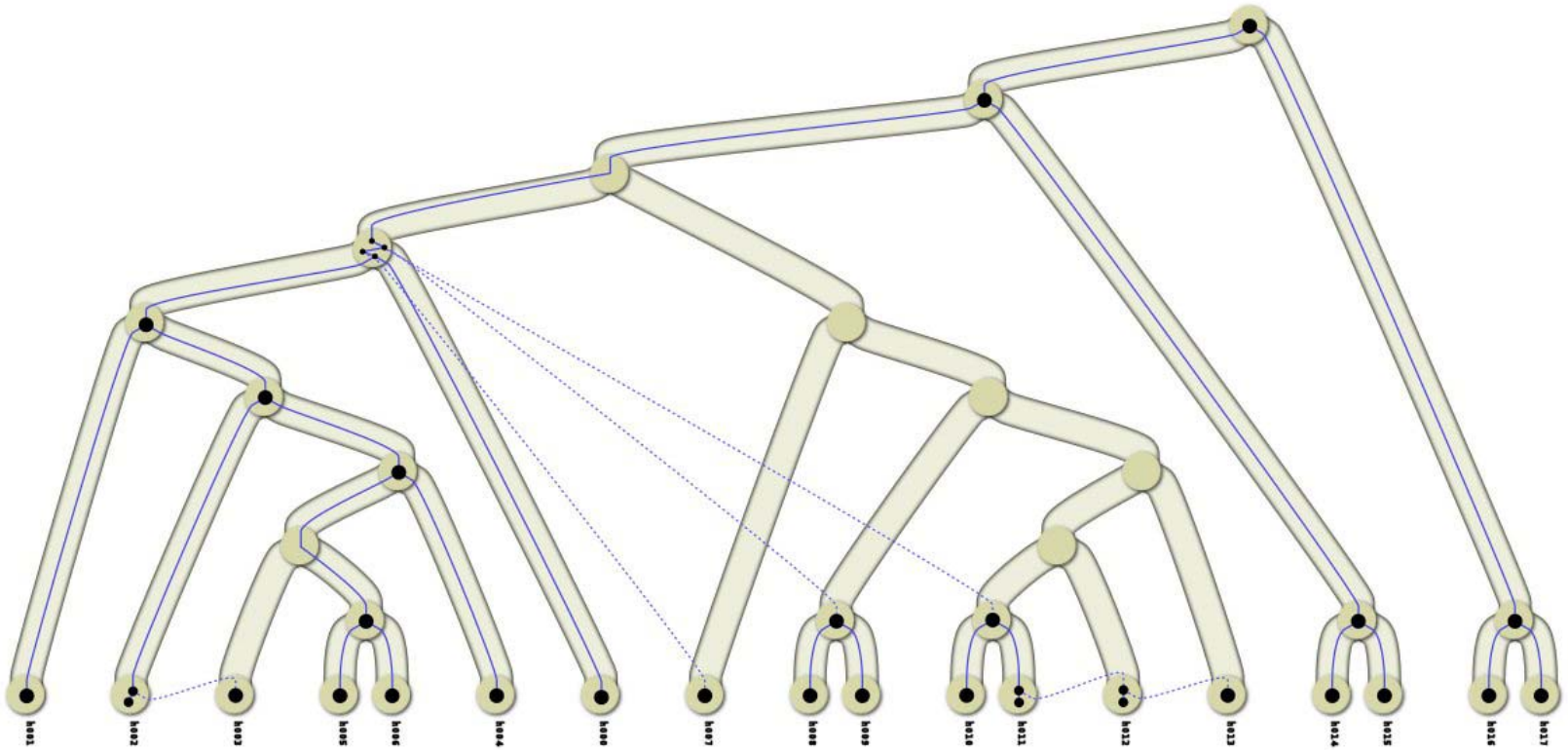
- Jane 4



Example of the 2nd strategy

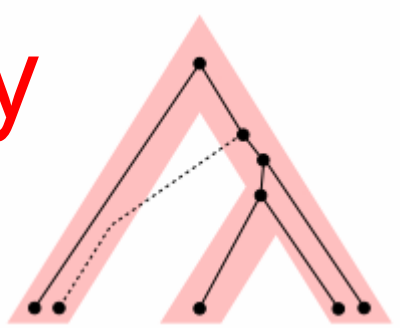


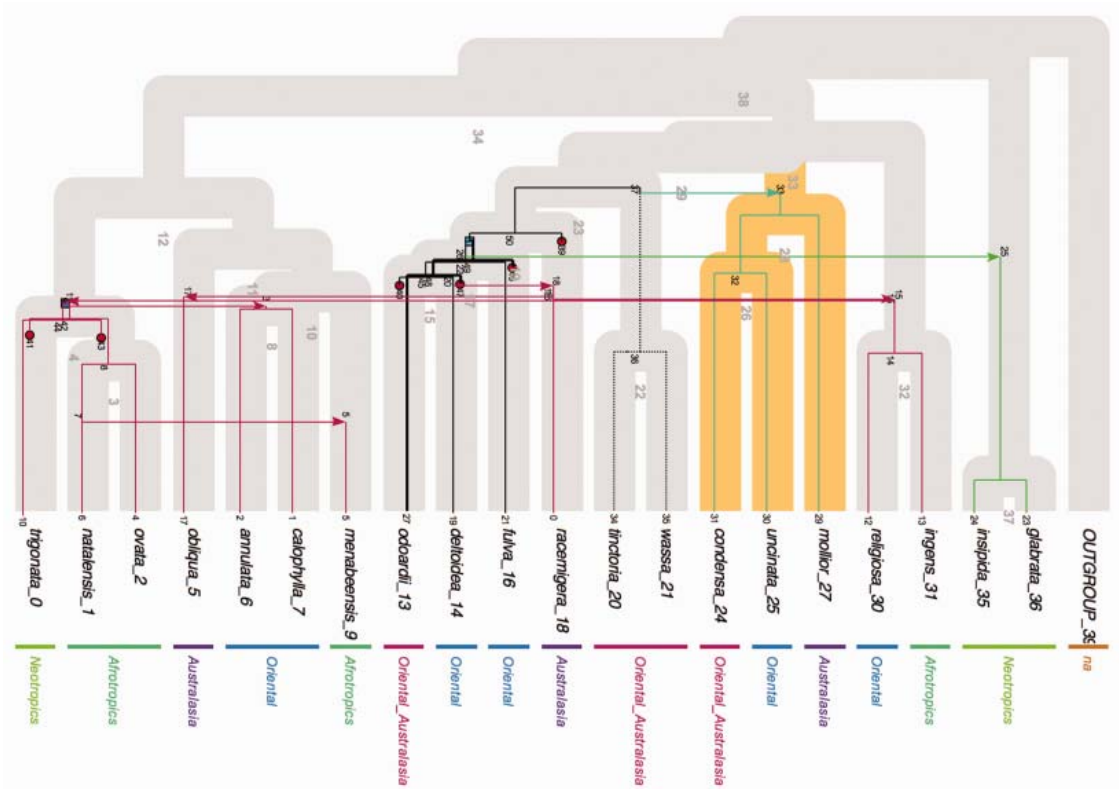
- CophyTrees

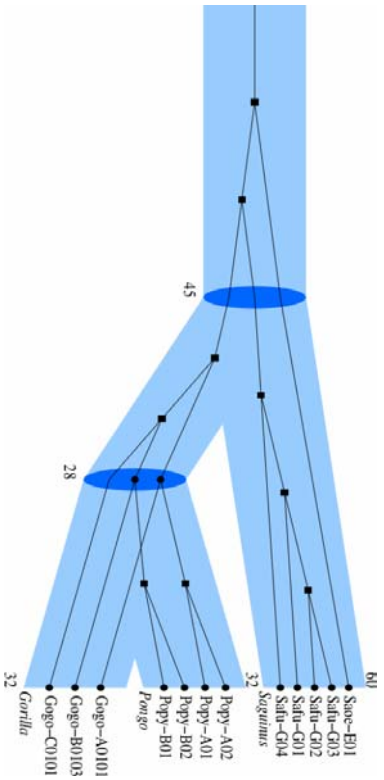


Examples of the 3rd strategy

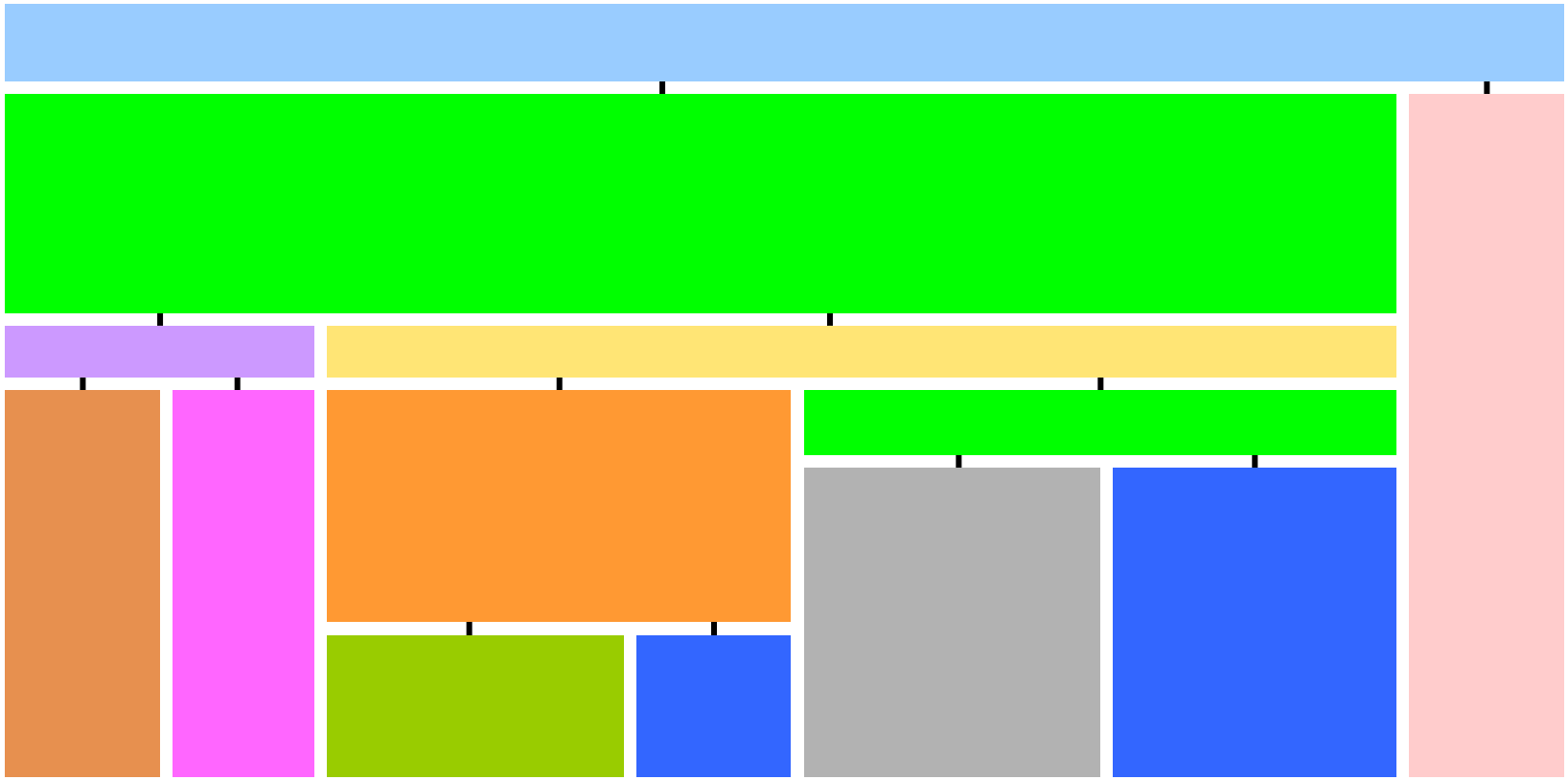
- SylvX & Primetv



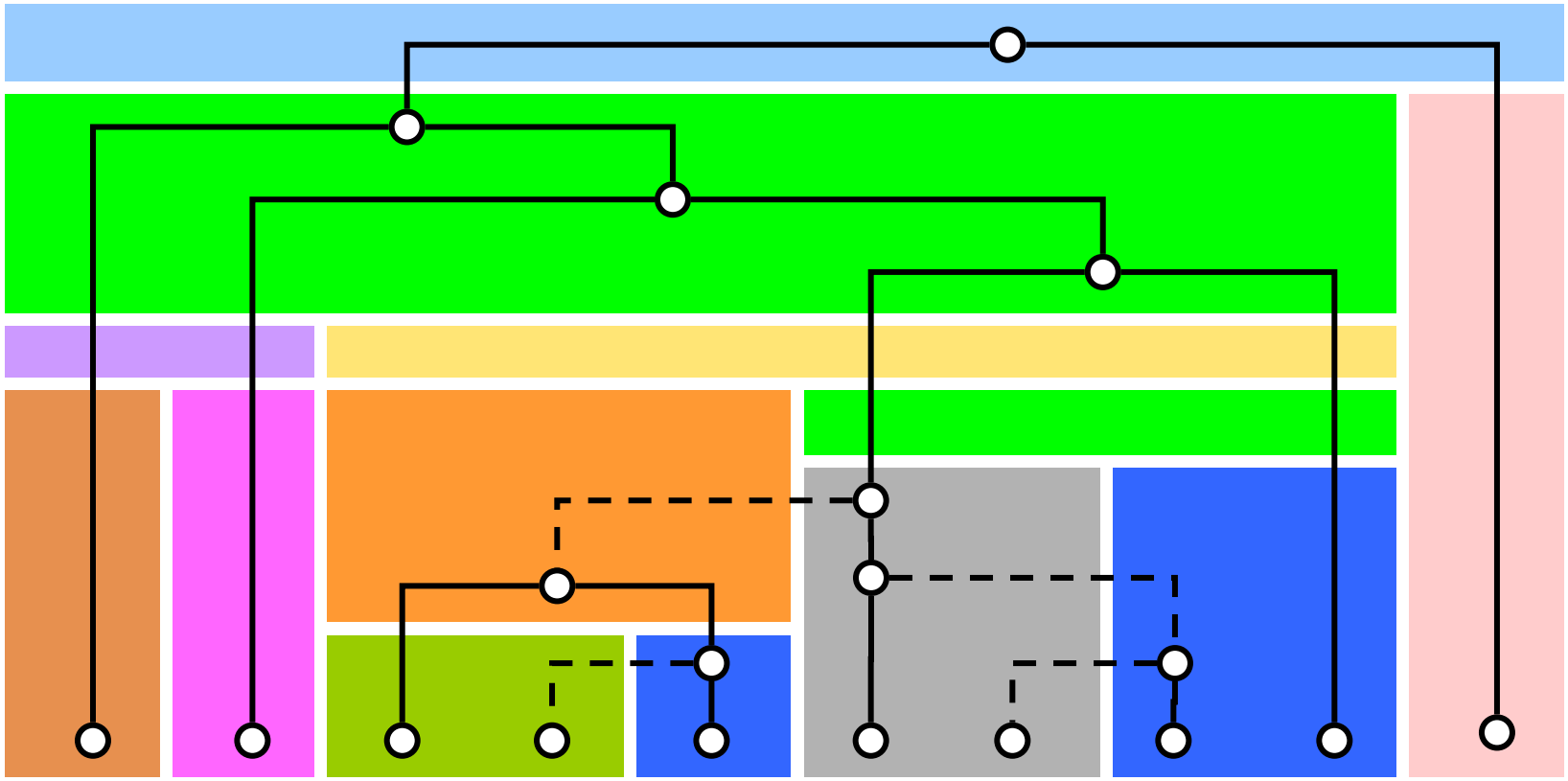




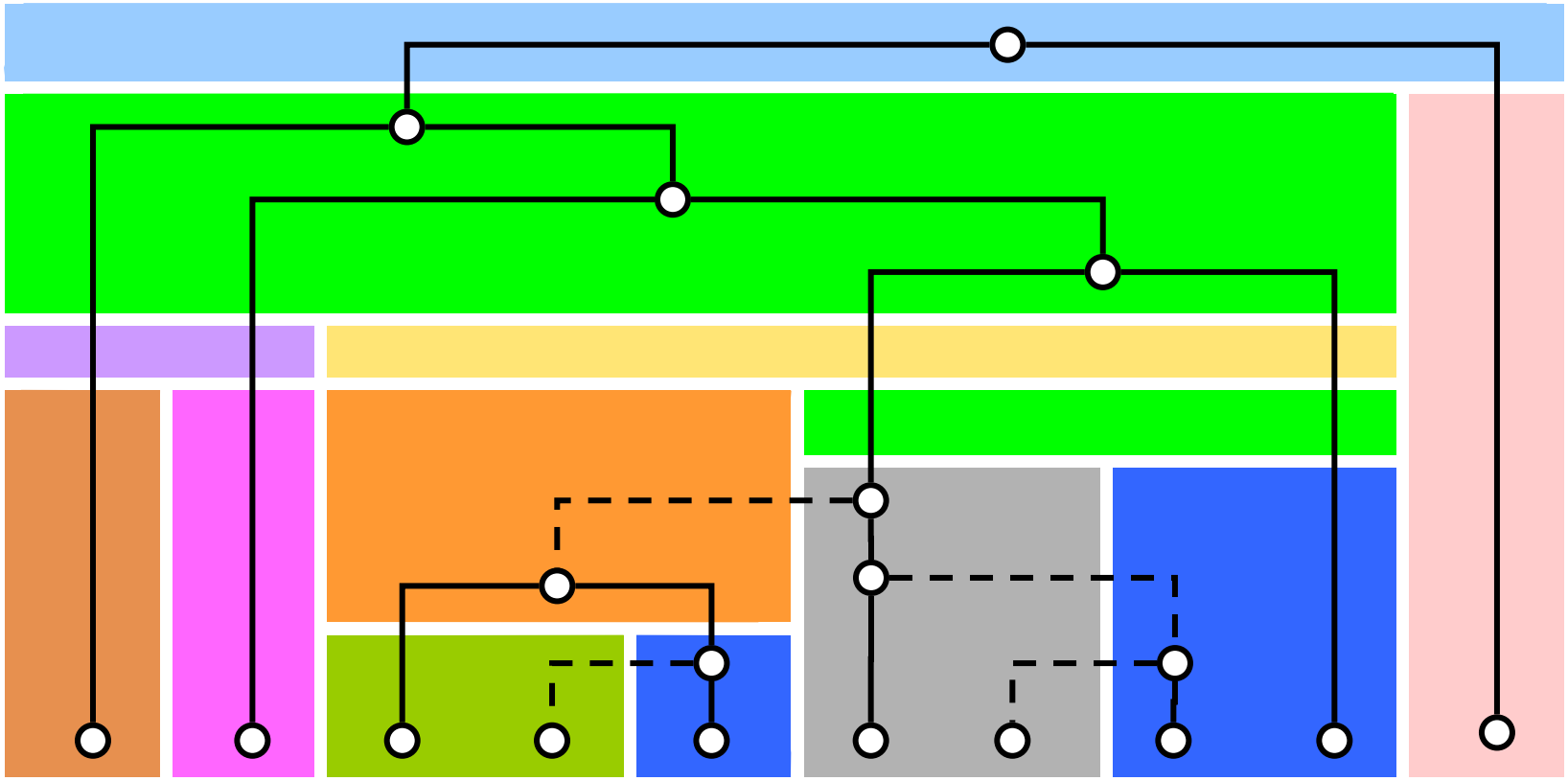
New metaphor: HP-drawings



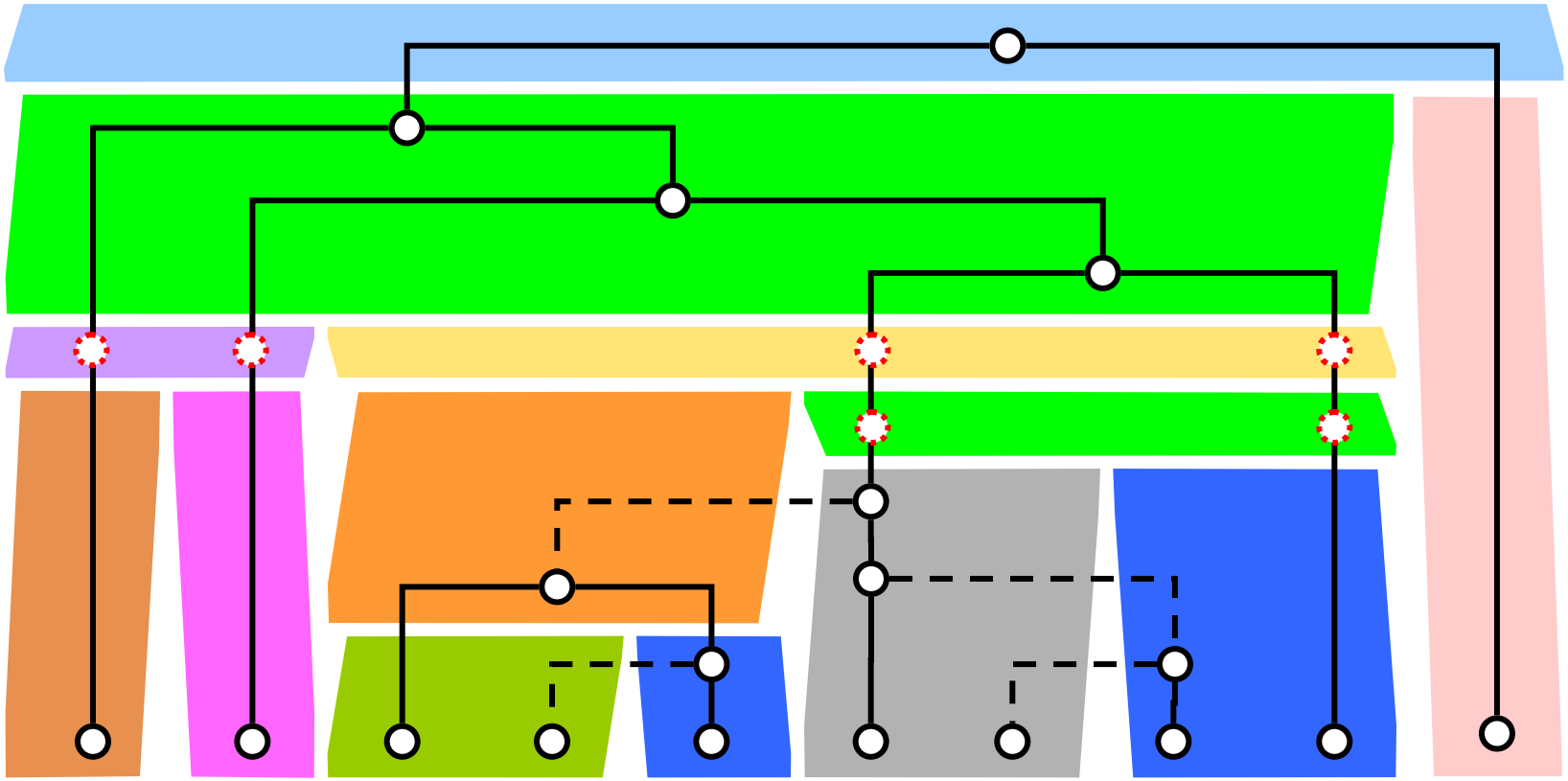
New metaphor: HP-drawings



New metaphor: HP-drawings

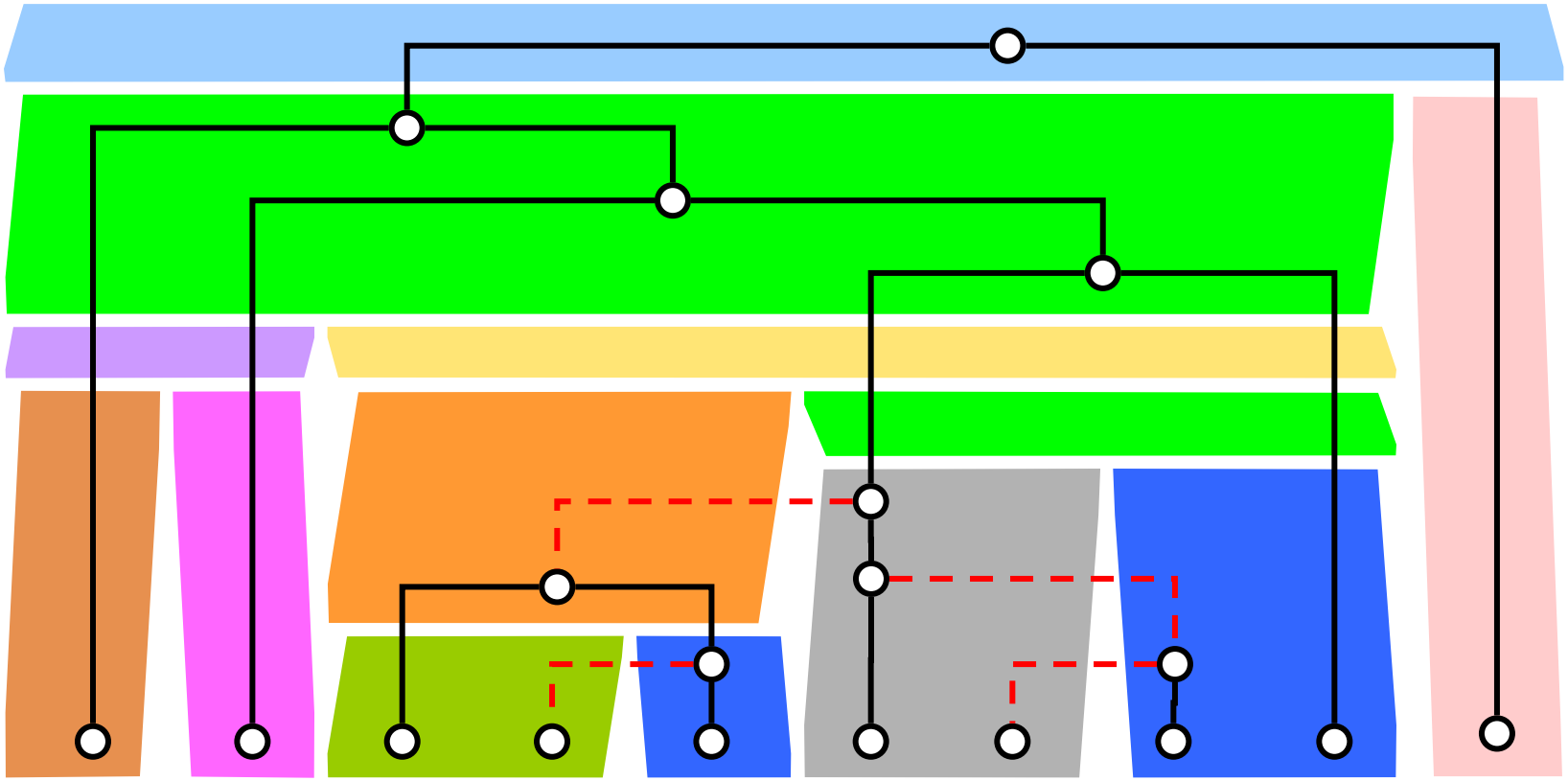


Evolutionary phenomena



- Loss:
  - a parasite is transmitted to one child but not to the other child

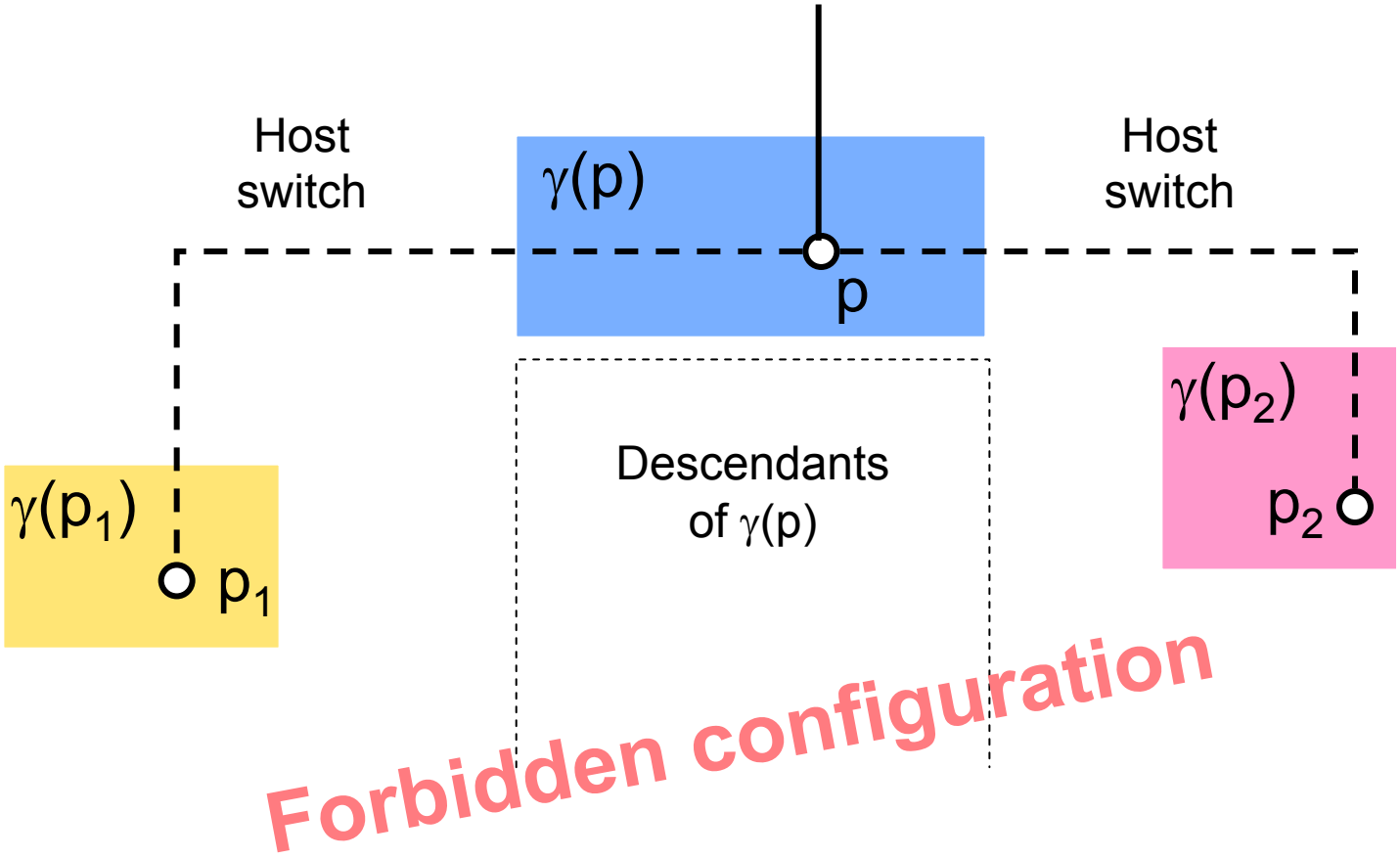
Evolutionary phenomena



- Host switch:
  - a parasite is transmitted to a host that is not a descendant of the current one

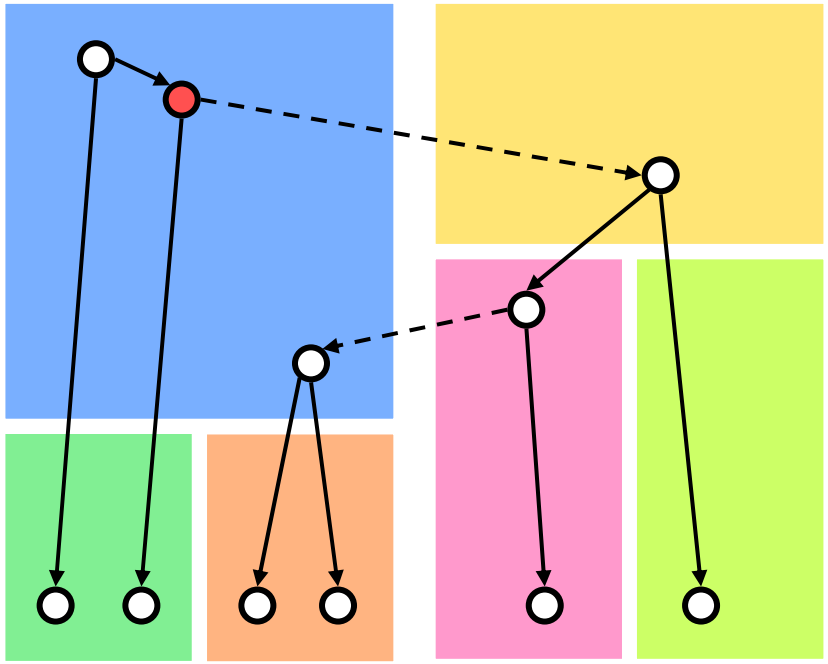
Reconciliation constraints

- No two host switches
  - the two children of a parasite p can not correspond both to host switches

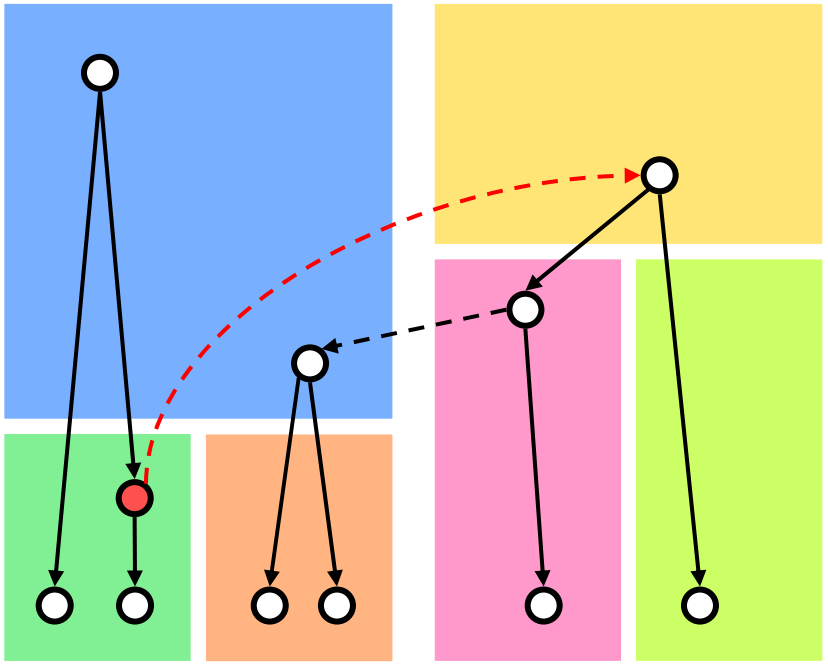


Reconciliation constraints

- Time consistency
  - intuitively, a reconciliation γ is time consistent if it does not imply the transmission of a parasite to a past host



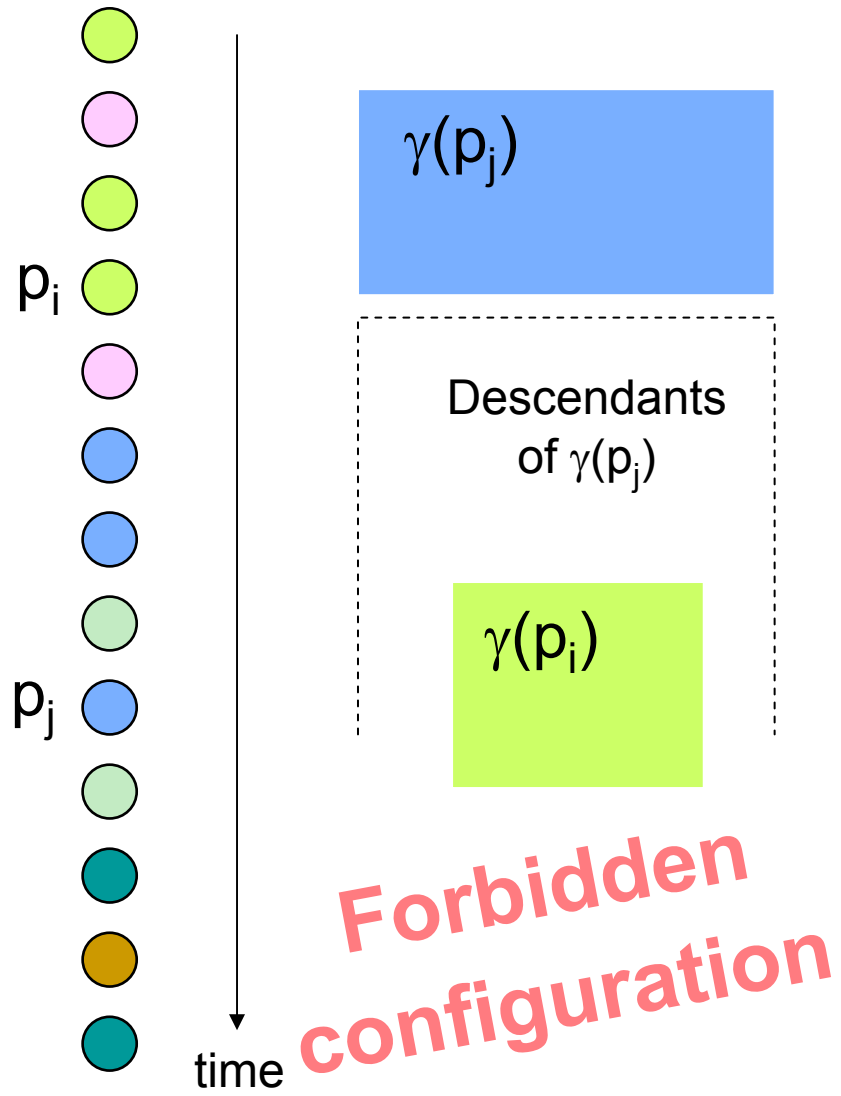
Time consistent reconciliation



Reconciliation that is not time consistent

Time consistency

- Formal definition
  - a reconciliation γ is *time consistent* if you can find a linear order of the parasites so that
    - each parasite comes before its descendant parasites
    - given two parasites p_i and p_j such that p_j comes after p_i, $\gamma(p_j)$ is not a proper ancestor of $\gamma(p_i)$

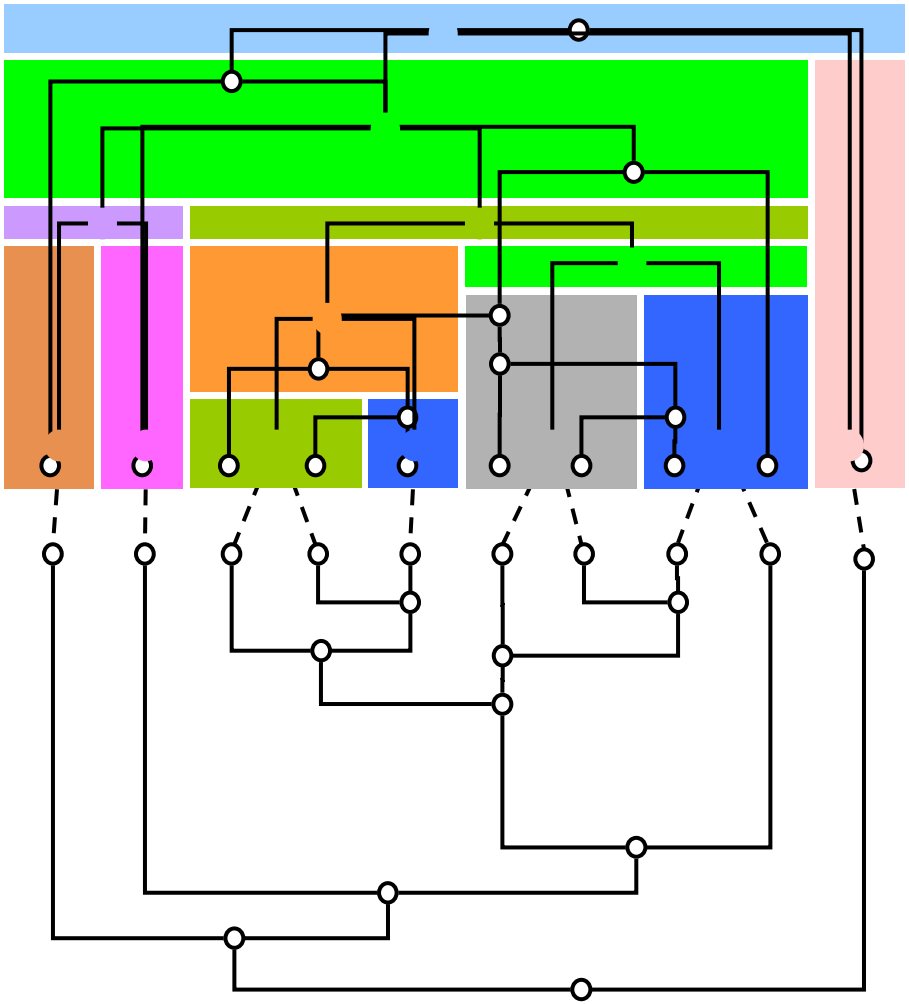


Planar reconciliations

- Theorem
  - Let (H,P,φ) be a co-phylogenetic tree and let γ be any time consistent reconciliation of it
  - (H,P,φ) admits a planar tanglegram drawing if and only if γ admits a planar downward HP-drawing

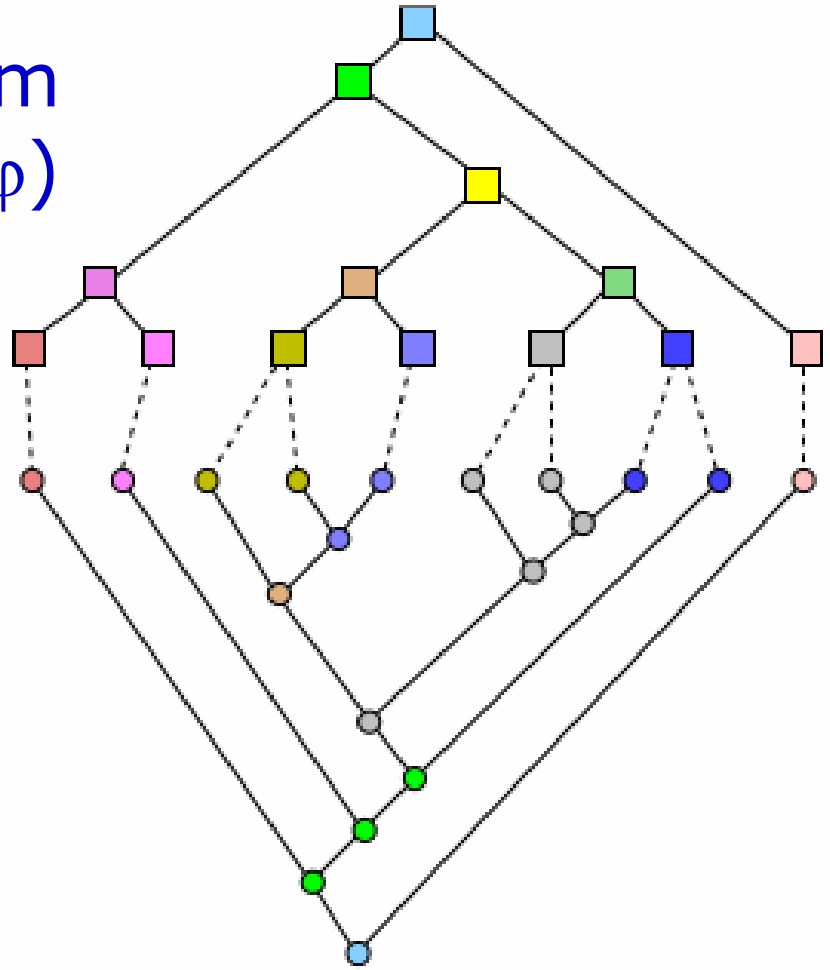
HP-drawing ⇒Tanglegram drawing

1. Vertically flip the drawing of the parasite tree
2. Join the leaves of the two trees
3. Represent the host tree in a node-link fashion



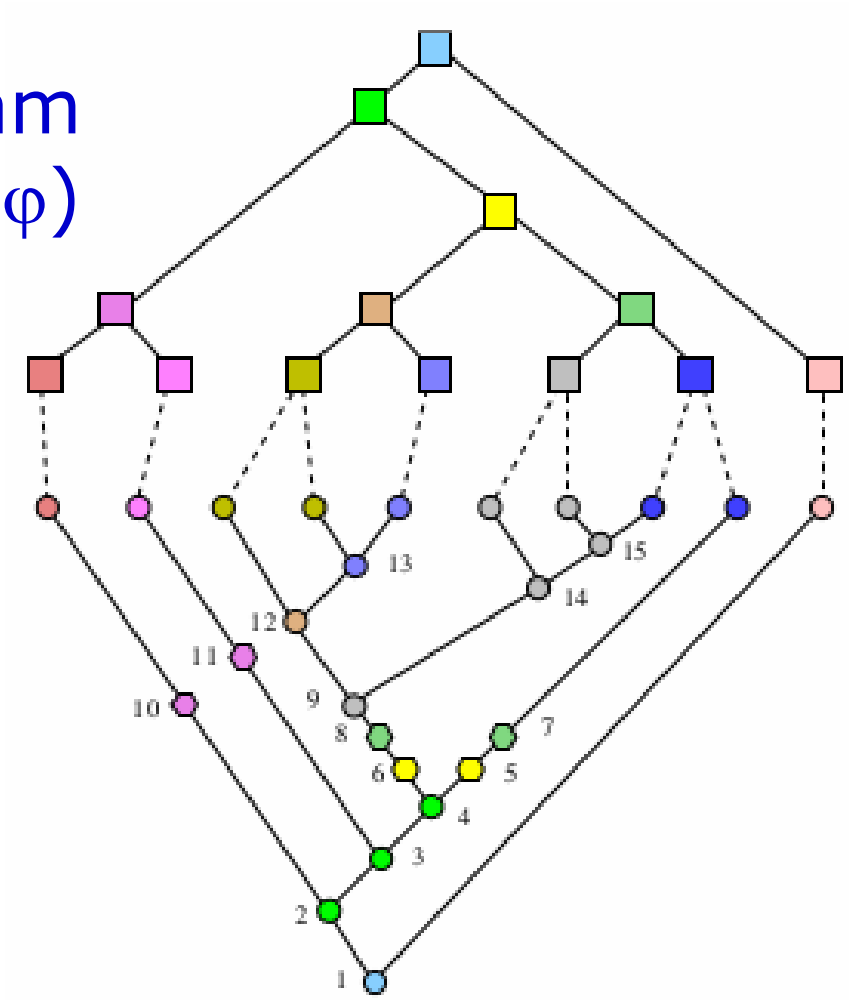
Tanglegram drawing ⇒ HP-drawing

1. Start from a planar tanglegram drawing of (H,P,φ)



Tanglegram drawing ⇒ HP-drawing

1. Start from a planar tanglegram drawing of (H,P,φ)
2. Add dummy nodes to the parasite tree to account for losses



Tanglegram drawing ⇒ HP-drawing

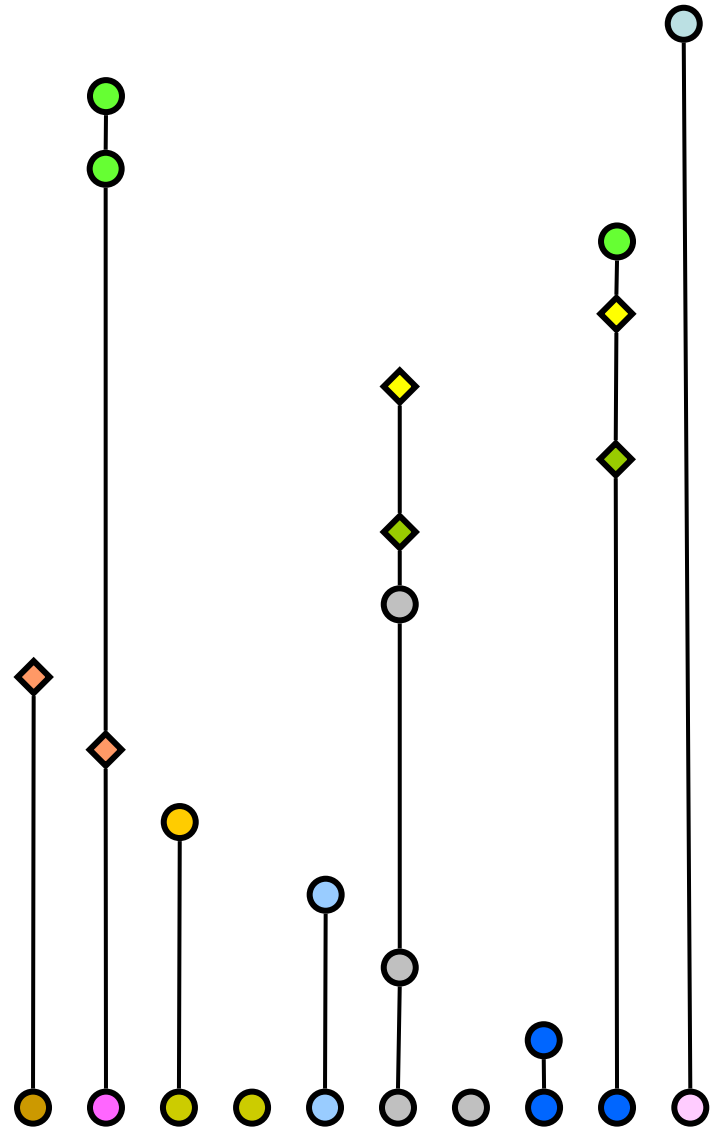
3. Copy the x-coords of the leaves from the tanglegram drawing
4. Assign y-coords to internal nodes based on a time consistent linear ordering

Tanglegram drawing ⇒ HP-drawing

3. Copy the x-coords of the leaves from the tanglegram drawing
4. Assign y-coords to internal nodes based on a time consistent linear ordering
5. Assign to each internal node the x-coord of one of its non-host-switch children

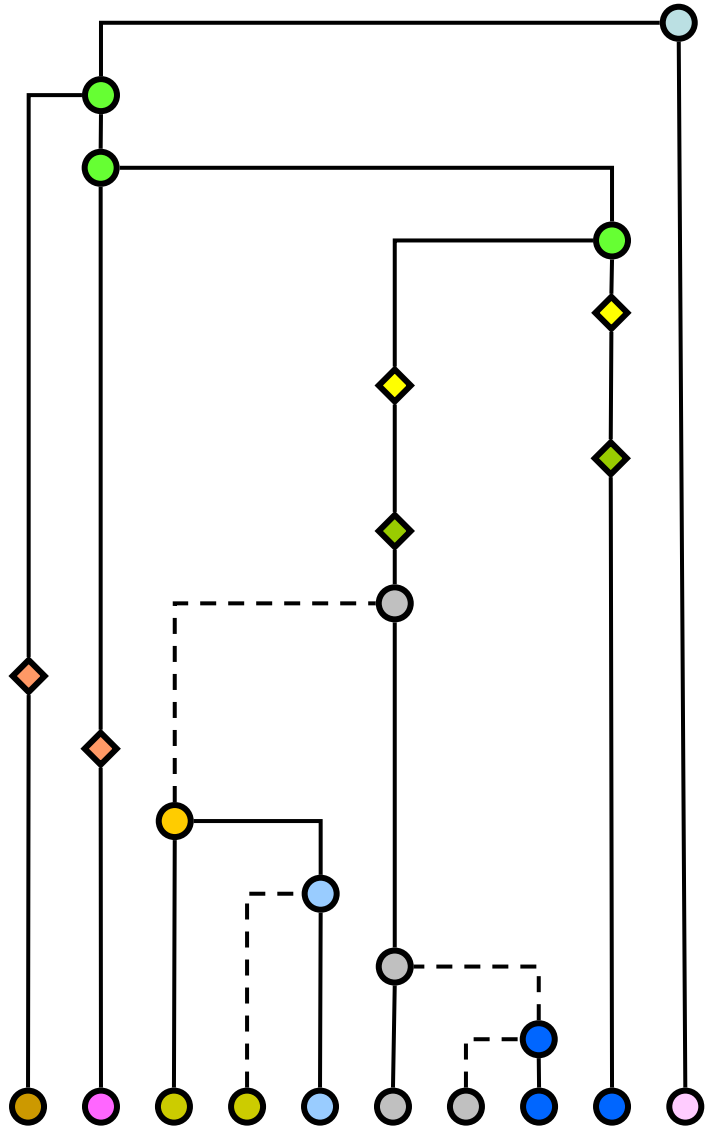
Tanglegram drawing ⇒ HP-drawing

3. Copy the x-coords of the leaves from the tanglegram drawing
4. Assign y-coords to internal nodes based on a time consistent linear ordering
5. Assign to each internal node the x-coord of one of its non-host-switch children



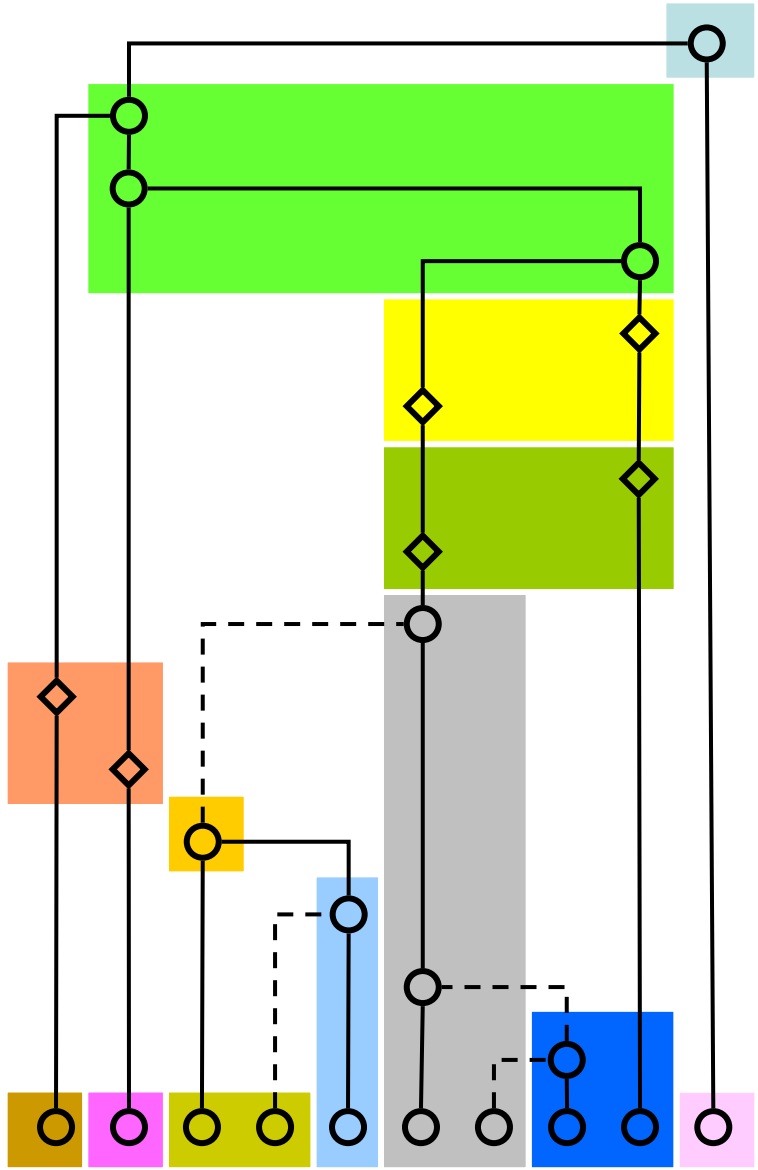
Tanglegram drawing ⇒ HP-drawing

6. Add the remaining edges



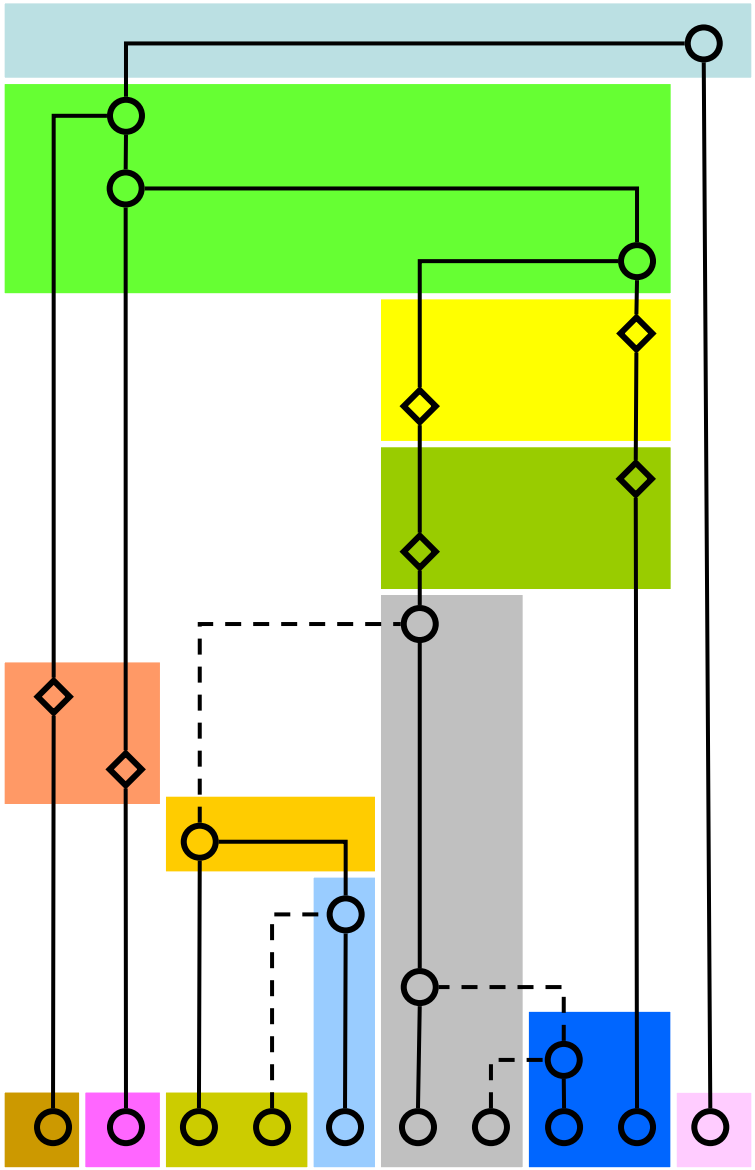
Tanglegram drawing ⇒ HP-drawing

6. Add the remaining edges
7. Draw each host as the smallest box containing its parasites



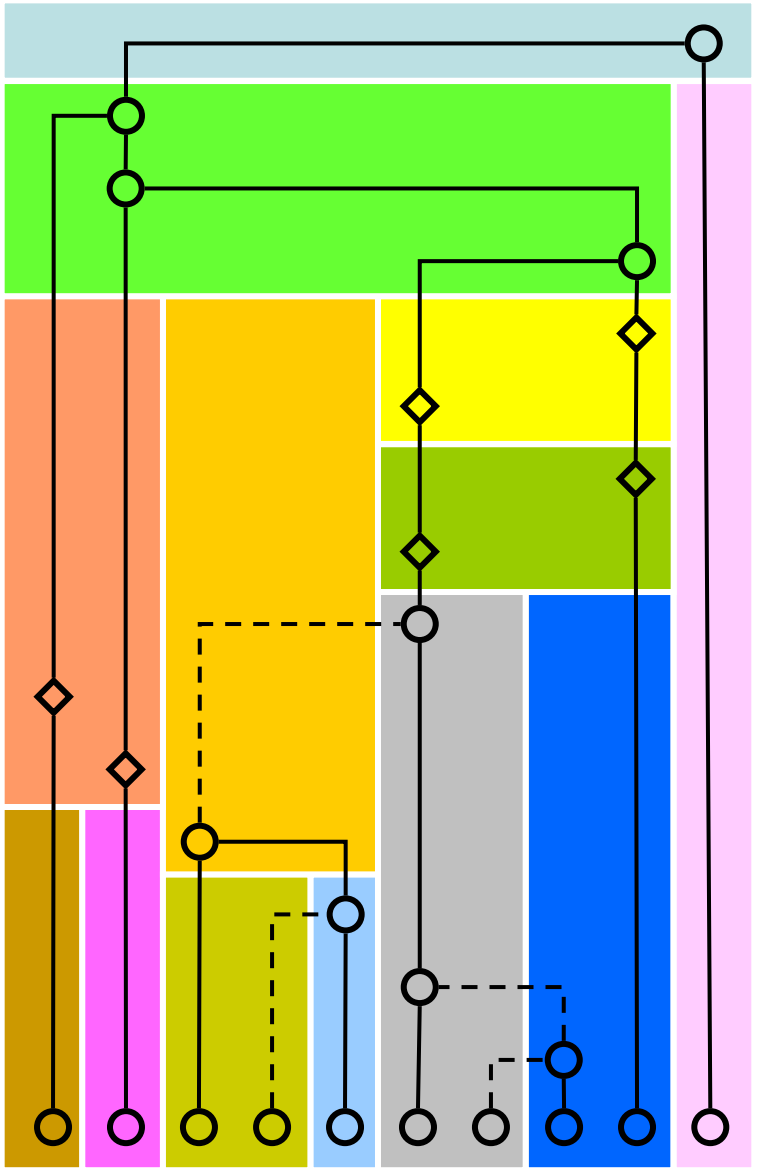
Tanglegram drawing ⇒ HP-drawing

6. Add the remaining edges
7. Draw each host as the smallest box containing its parasites
8. Horizontally enlarge the boxes to cover all non-host-switch descendants



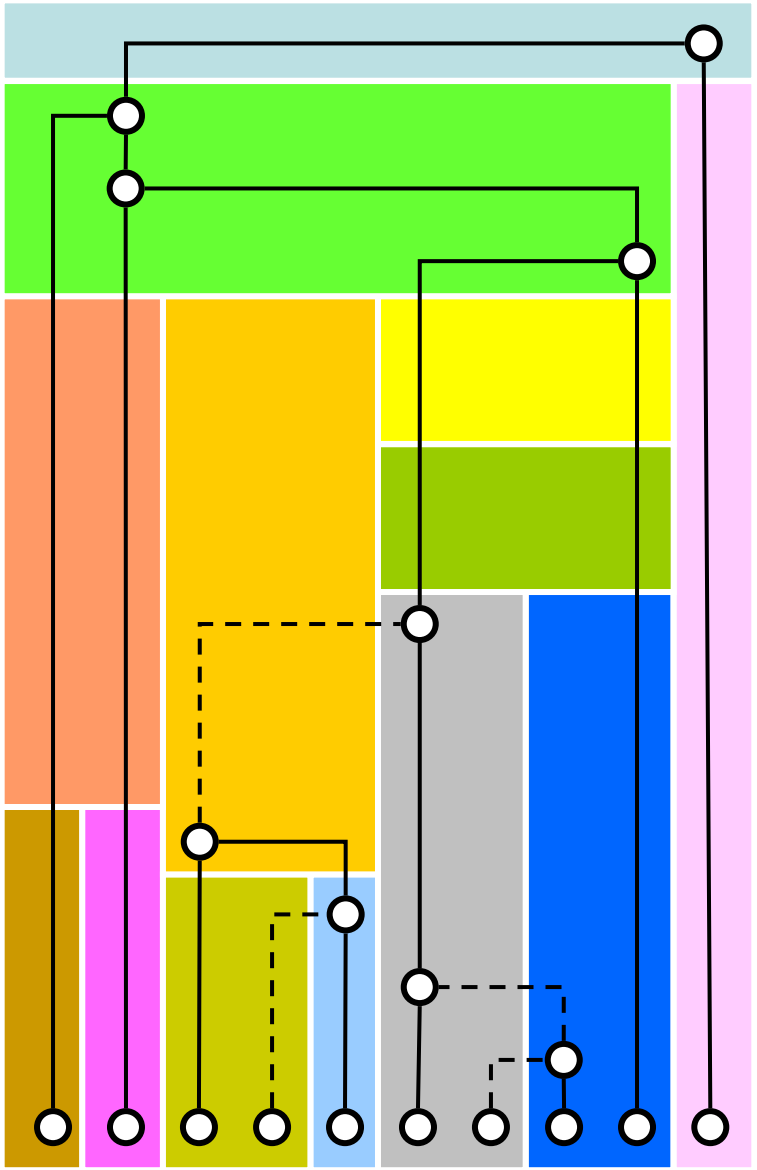
Tanglegram drawing ⇒ HP-drawing

6. Add the remaining edges
7. Draw each host as the smallest box containing its parasites
8. Horizontally enlarge the boxes to cover all non-host-switch descendants
9. Vertically enlarge the boxes to touch the boxes of their parents



Tanglegram drawing ⇒ HP-drawing

6. Add the remaining edges
7. Draw each host as the smallest box containing its parasites
8. Horizontally enlarge the boxes to cover all non-host-switch descendants
9. Vertically enlarge the boxes to touch the boxes of their parents
10. Remove dummy parasites

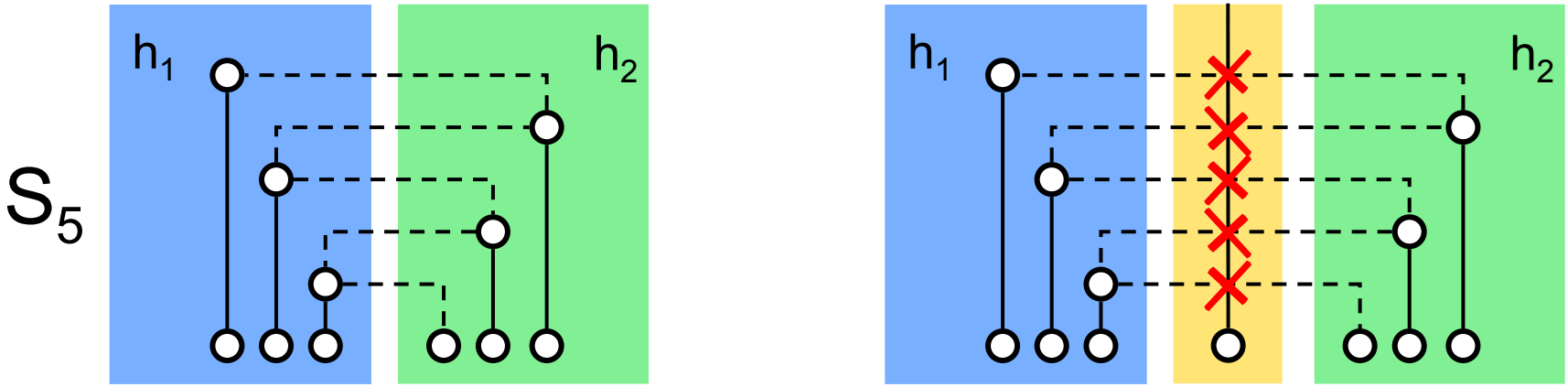


Crossing minimization is NP-complete

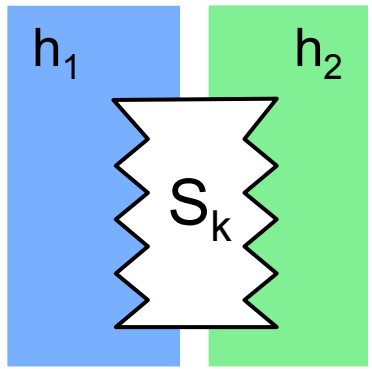
- Theorem
  - deciding whether a time-consistent reconciliation γ admits an HP-drawing with at most k crossings is NP-complete

Gadget: sewing trees

- Recursively defined



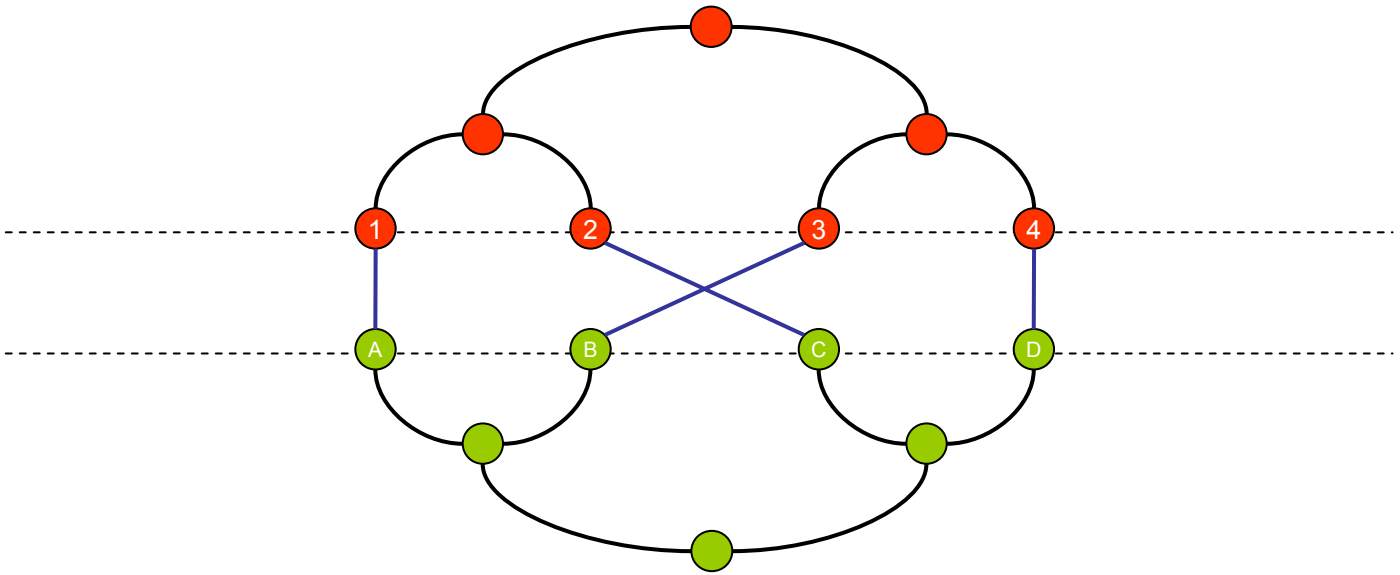
- A sewing tree of suitable size can make extremely costly inserting a node between the two leaves h_1 and h_2



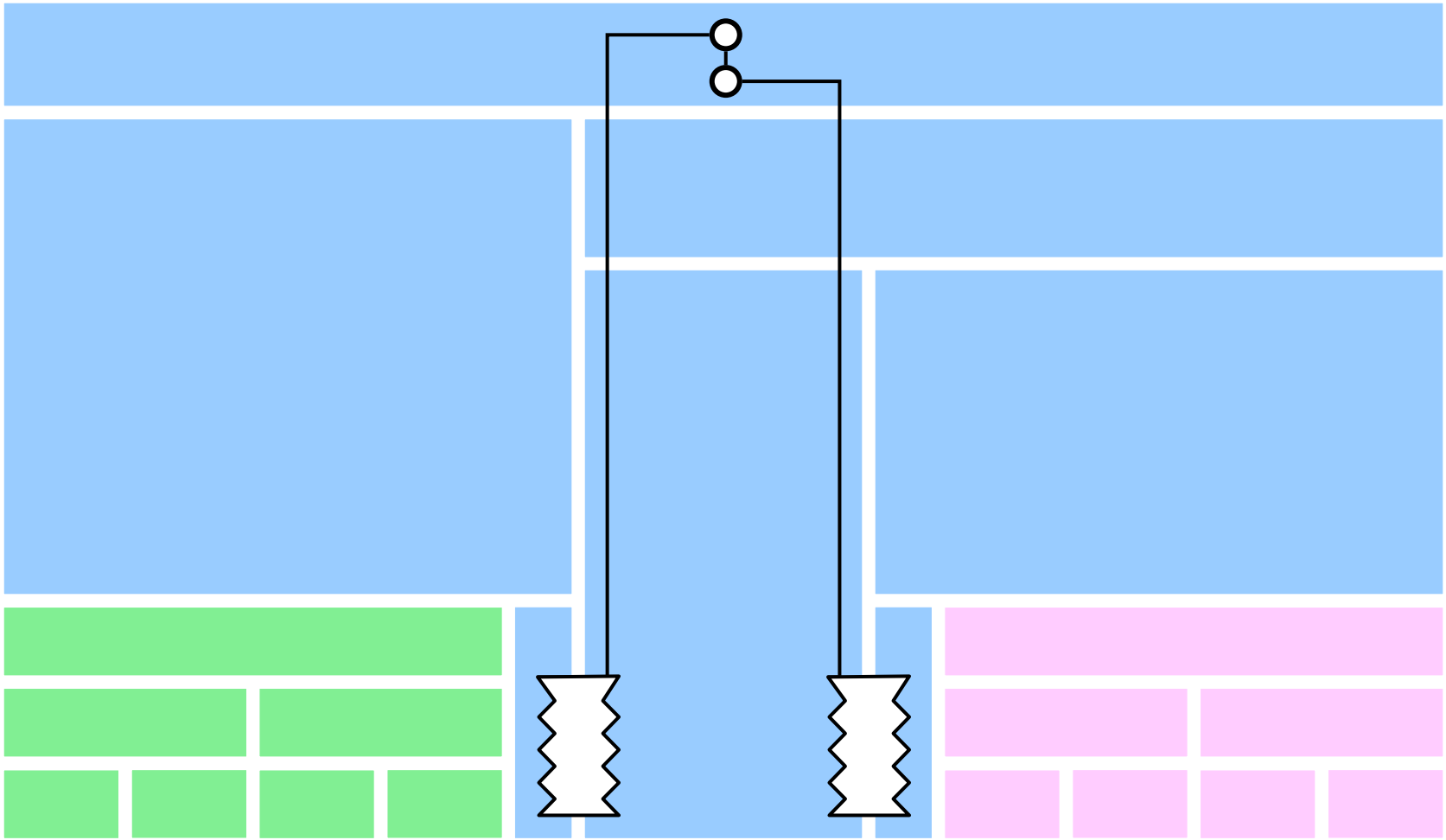
Iconic representation

The reduced problem

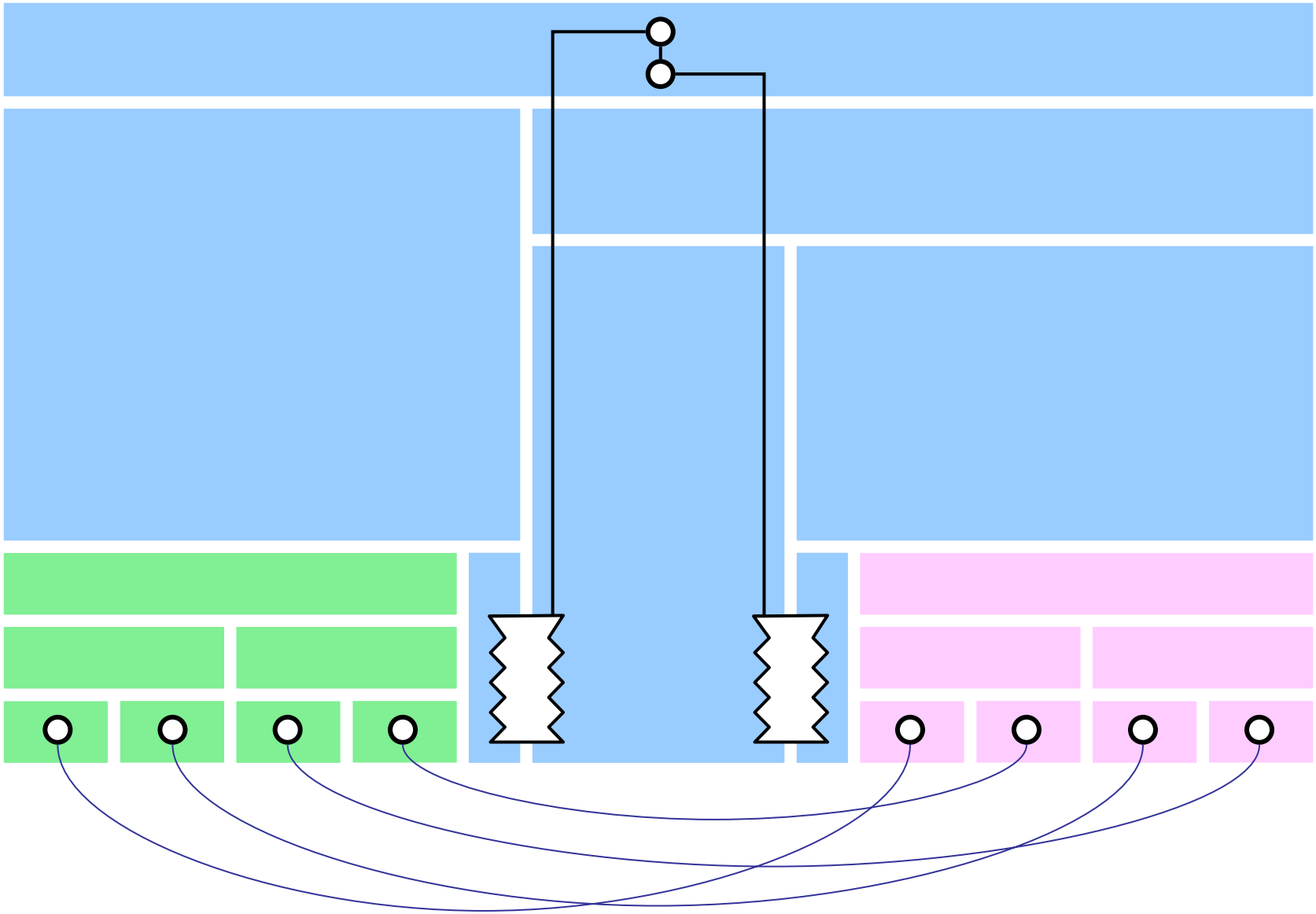
- We reduce crossing minimization for tanglegram drawings of two complete binary trees with one-to-one leaf associations
  - NP-complete [Buchin et al., Algorithmica 2012]



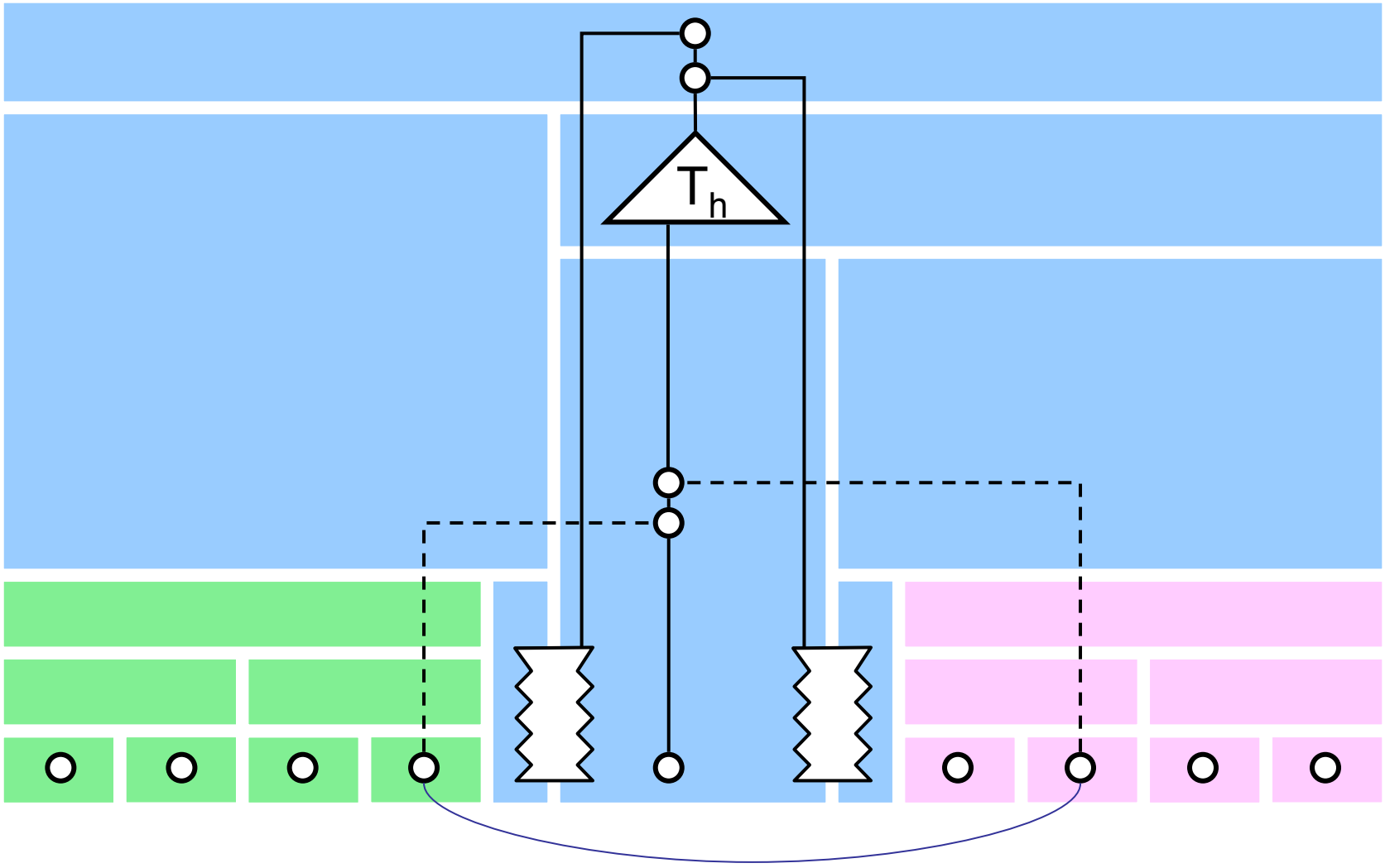
The reduction



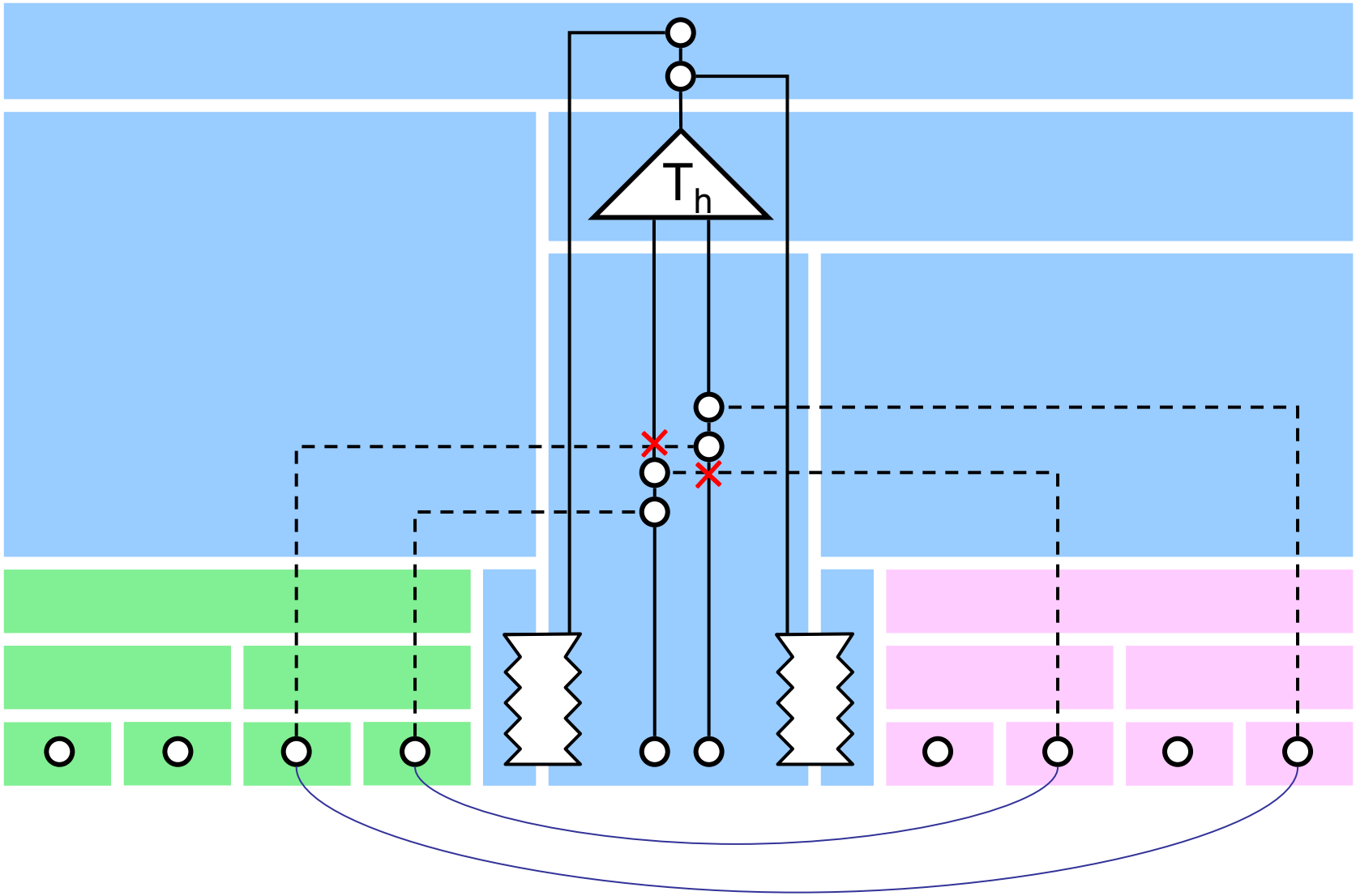
The reduction



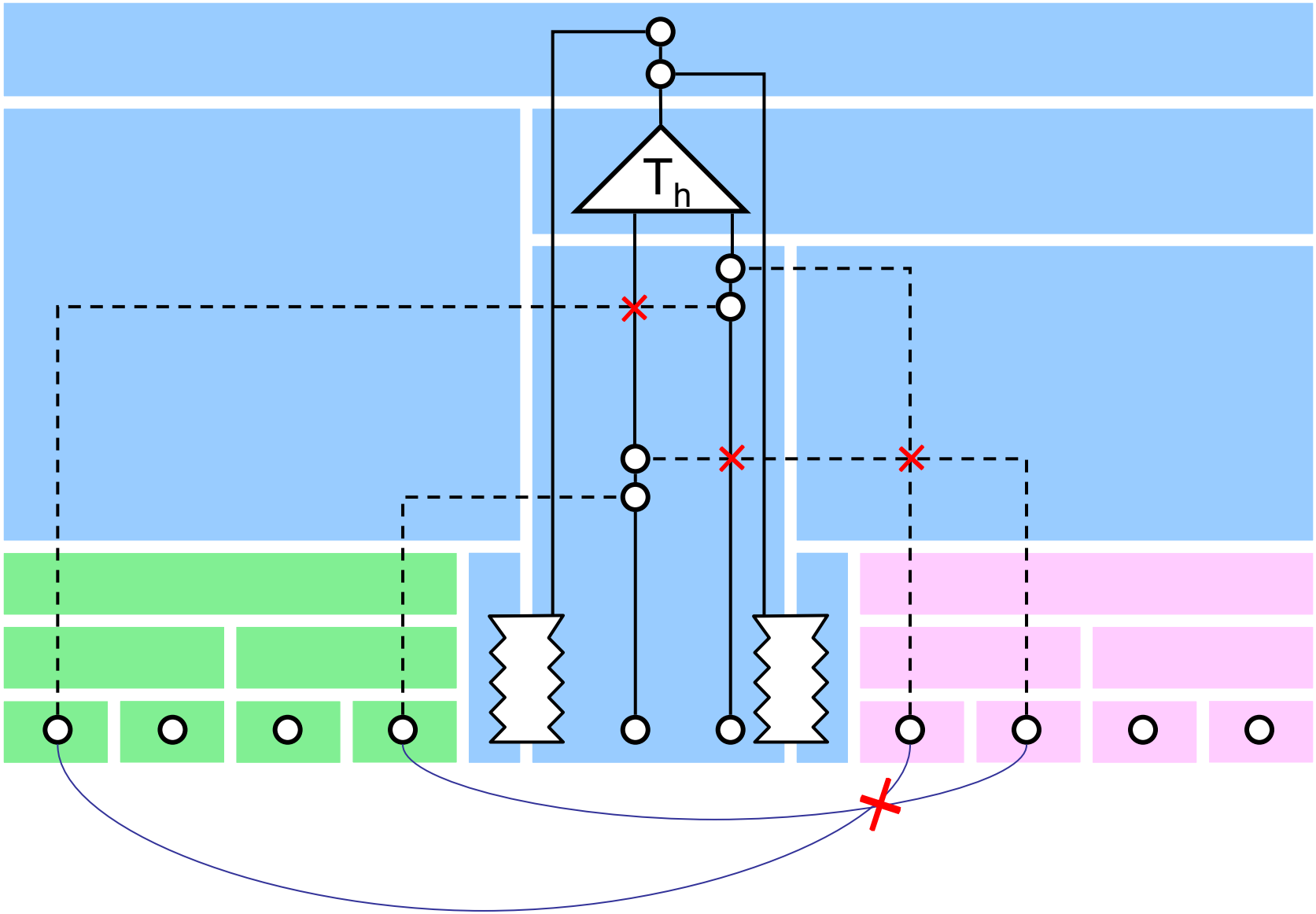
The reduction



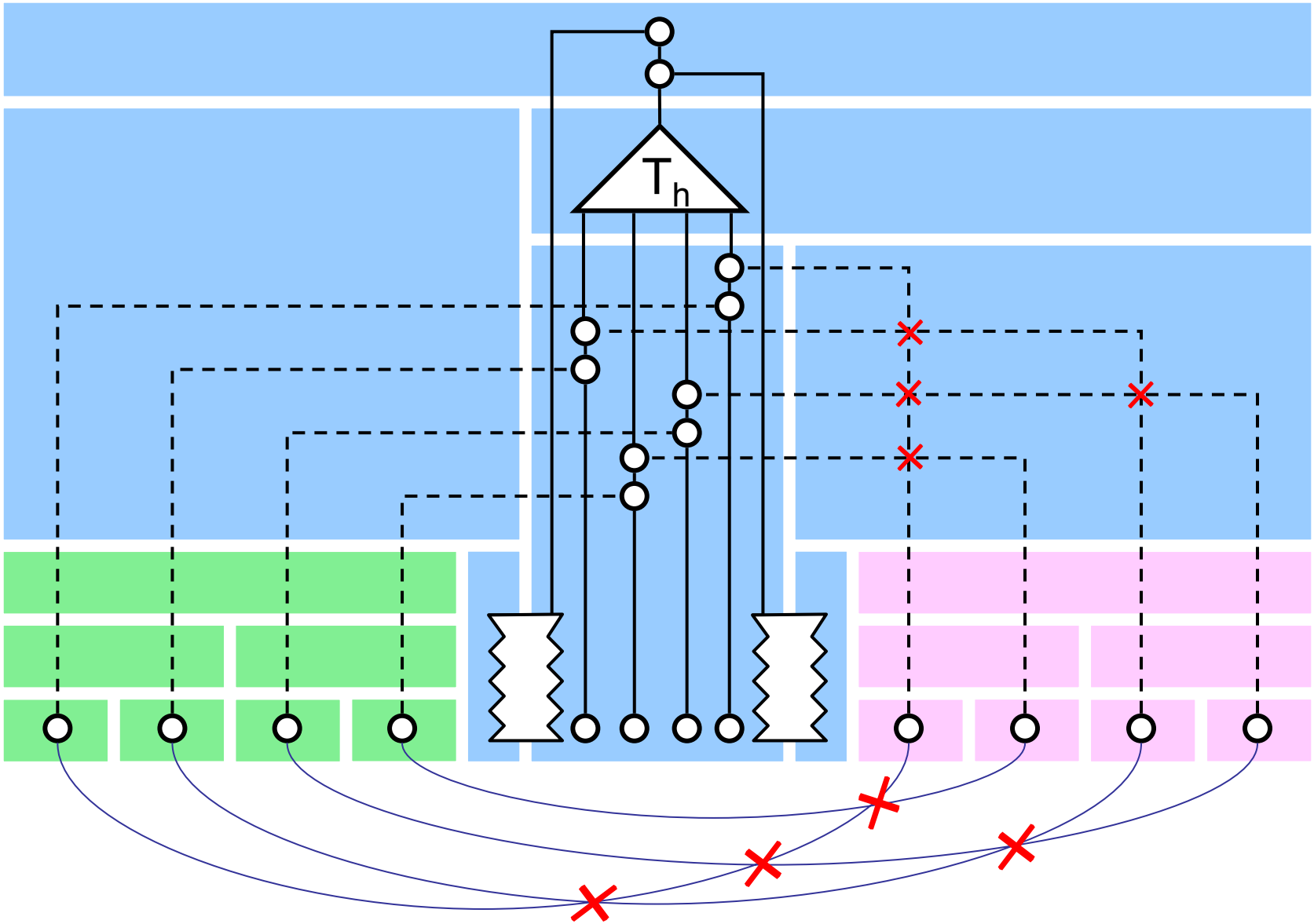
The reduction



The reduction



The reduction



Two heuristics for crossing reduction

- Heuristic SearchMaximalPlanar
  - Construct a maximal planar instance and the add the remaining edges in a post-processing step
- Heuristic ShortenHostSwitch
  - Embed the host tree in such a way to reduce the length of the host-switch links among parasites

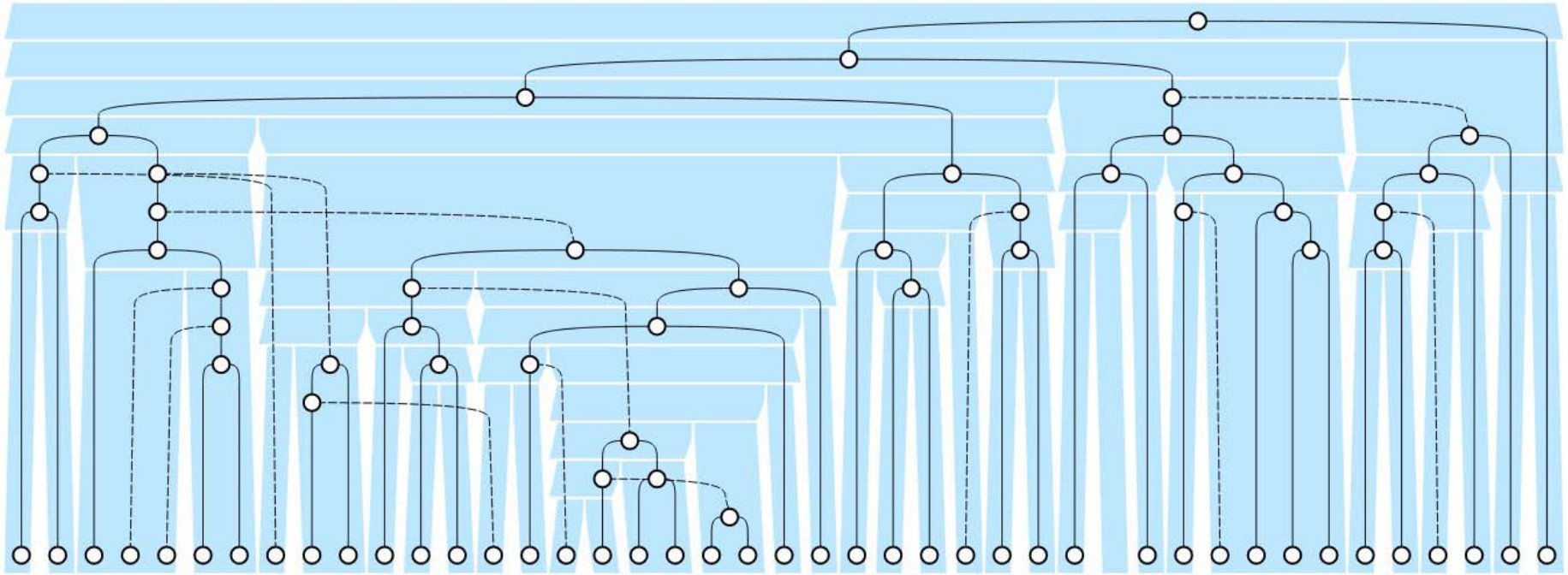
Experiments

		SearchMaximalPlanar		ShortenHostSwitch	
instance	#rec.	avg. #cross.	avg. ms	avg. #cross.	avg. ms
CM	64	16	485	21	0.5
FD	80	91	4596	69	1
GL	2	2	67	1	0
PP	72	1	1154	3	1
RH	100	12	1701	11	2
RP	3	3	195	3	1
SC	1	4	166	6	0
SFC	16	15	355	17	0
C0G2085	100	82	17270	70	7
C0G4965	100	58	5636	97	8

Conclusions and future work

- We introduced a new metaphor that takes advantage both of the space-filling and of the node-link visualization paradigms
- Future work
  - address the problem of visually exploring and analyzing sets of reconciliations of the same co-phylogenetic tree
  - adapt heuristics for the reduction of the crossings of tanglegram drawings to HP-drawings
  - perform user-tests to assess readability

Thanks!



Any question?