Simulation of Cophylogenies using the Age Model

Stephanie Keller-Schmidt

Bioinformatics Group
Departement of Computer Science
University of Leipzig

TBI Winterseminar
Bled/Slovenia, 16 February 2011

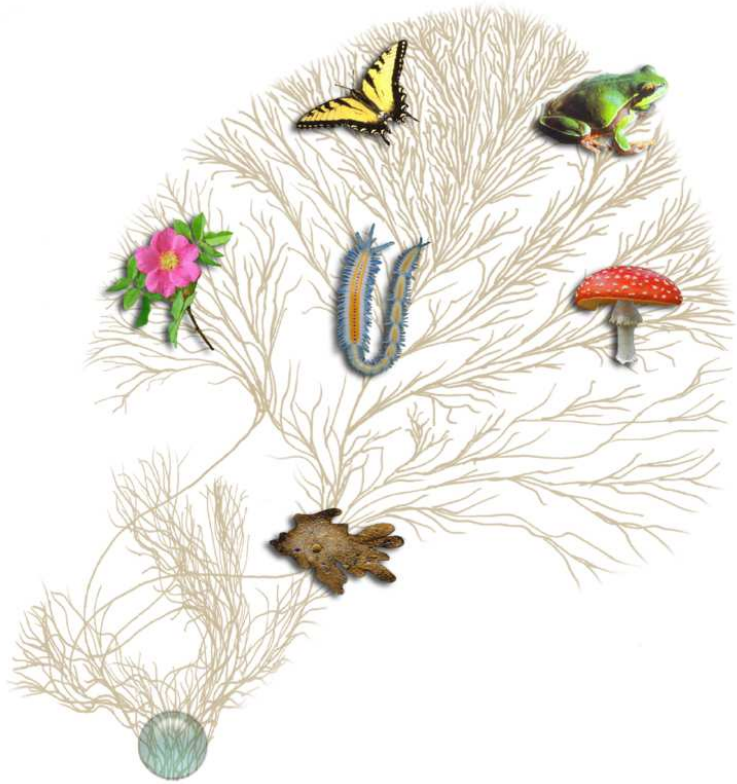
Outline

Phylogenetic Trees

- leaves, inner nodes
- estimated evolutionary relationship - evol. processes?
- reconstruction $\rightarrow$ assume macro-evolutionary models

Models

- consider tree shape
- infer how diversity has arisen by fitting stochastic models to tree data

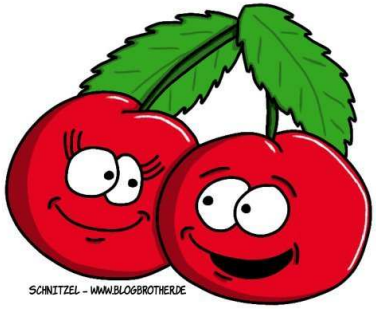


http://tolweb.org/

Generating a Tree

Choose leaf l with probability p given by model M.

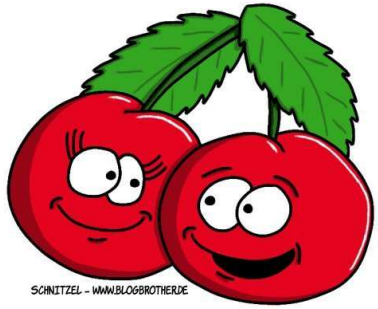
Replace leaf l by a cherry.

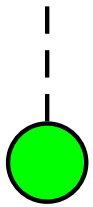


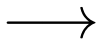
Generating a Tree

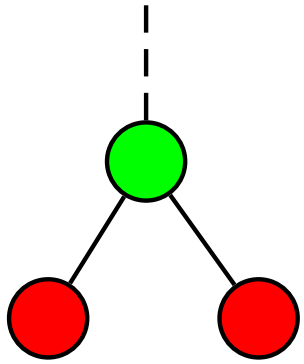
Choose leaf l with probability p given by model M.

Replace leaf l by a cherry.









Age model

Idea: The longer species i has not been involved in speciation, the less likely it will be.

Initialize: Set time $t = 0$, generate root node.

Iterate:

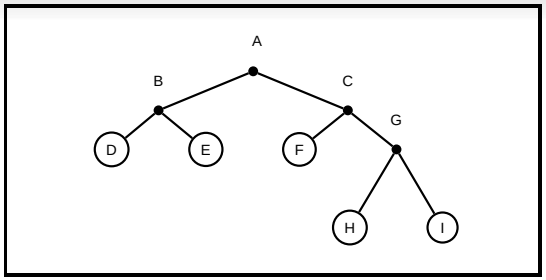
- Increment time t.
- From the set of leaves, choose leaf l with probability

$$p_i \propto (t - t_l)^{-1}$$

- Replace l by a cherry.

t = number of leaves = current time; t_l = creation time of leaf l

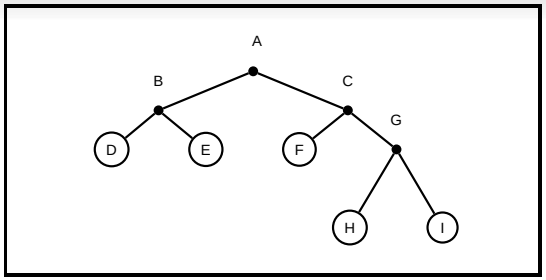
Age model - Example



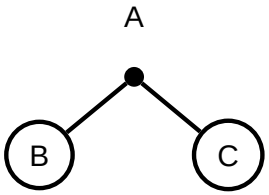
$t = 1$

$t_A = 0 \rightarrow p_A = 1$

Age model - Example

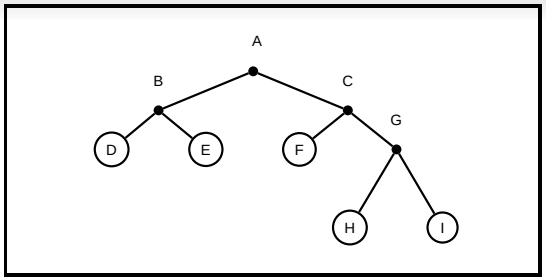


$t = 2$

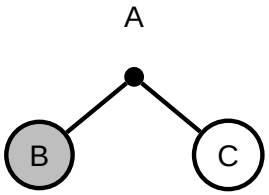


$$t_B = 1 \rightarrow p_B = \frac{1}{2}$$
$$t_C = 1 \rightarrow p_C = \frac{1}{2}$$

Age model - Example

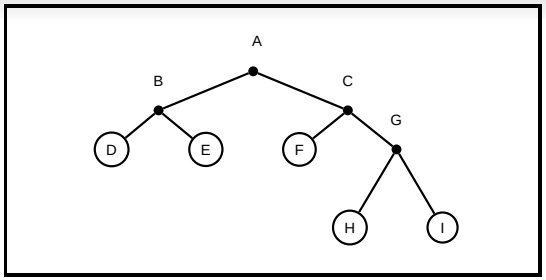


$t = 2$

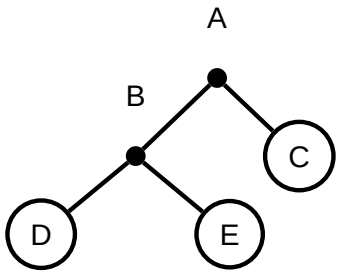


$$t_B = 1 \rightarrow p_B = \frac{1}{2}$$
$$t_C = 1 \rightarrow p_C = \frac{1}{2}$$

Age model - Example

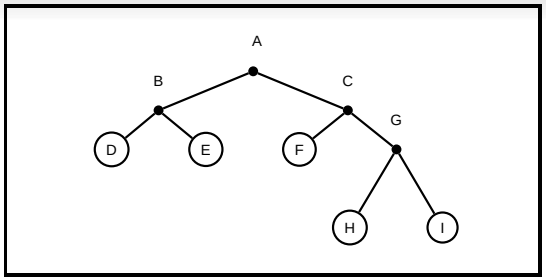


$t = 3$

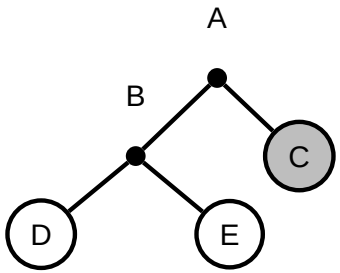


$$t_C = 1 \rightarrow p_C = \frac{1}{5}$$
$$t_D = 2 \rightarrow p_D = \frac{2}{5}$$
$$t_E = 2 \rightarrow p_E = \frac{2}{5}$$

Age model - Example

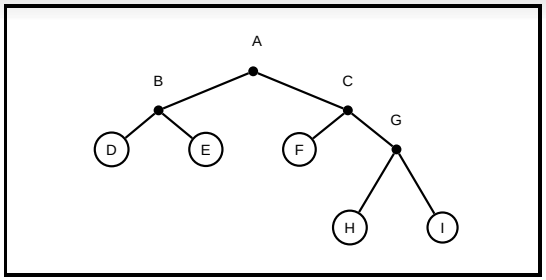


$t = 3$

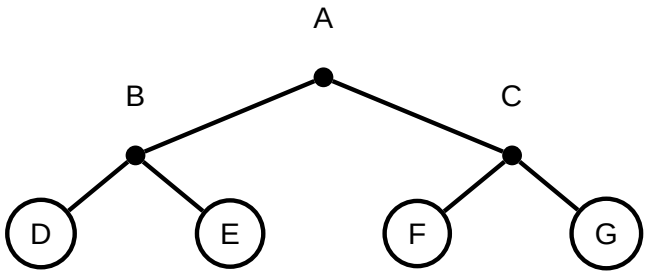


$$t_C = 1 \rightarrow p_C = \frac{1}{5}$$
$$t_D = 2 \rightarrow p_D = \frac{2}{5}$$
$$t_E = 2 \rightarrow p_E = \frac{2}{5}$$

Age model - Example

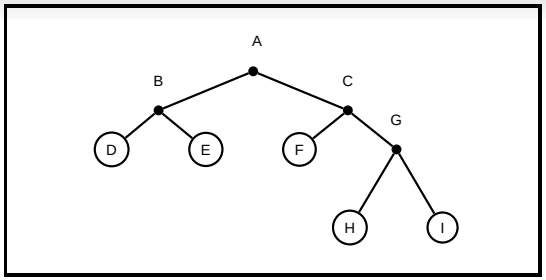


$t = 4$

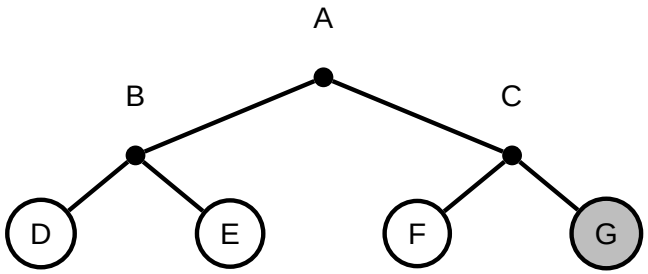


$$t_D = 2 \rightarrow p_D = \frac{1}{6}$$
$$t_E = 2 \rightarrow p_E = \frac{1}{6}$$
$$t_F = 3 \rightarrow p_F = \frac{2}{6}$$
$$t_G = 3 \rightarrow p_G = \frac{2}{6}$$

Age model - Example

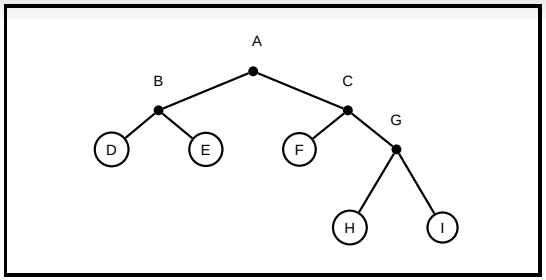


$t = 4$

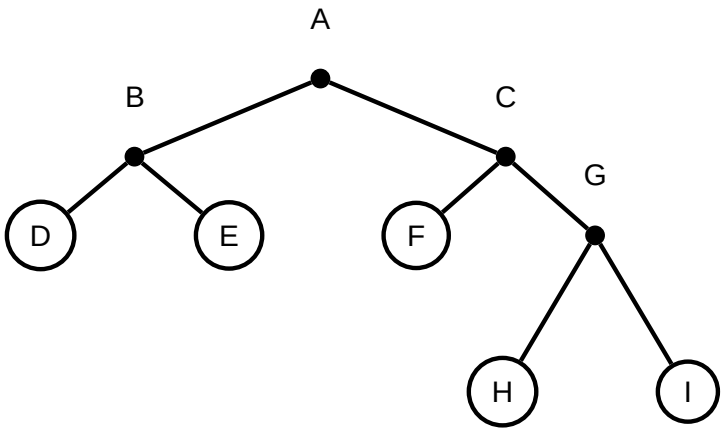


$$t_D = 2 \rightarrow p_D = \frac{1}{6}$$
$$t_E = 2 \rightarrow p_E = \frac{1}{6}$$
$$t_F = 3 \rightarrow p_F = \frac{2}{6}$$
$$t_G = 3 \rightarrow p_G = \frac{2}{6}$$

Age model - Example



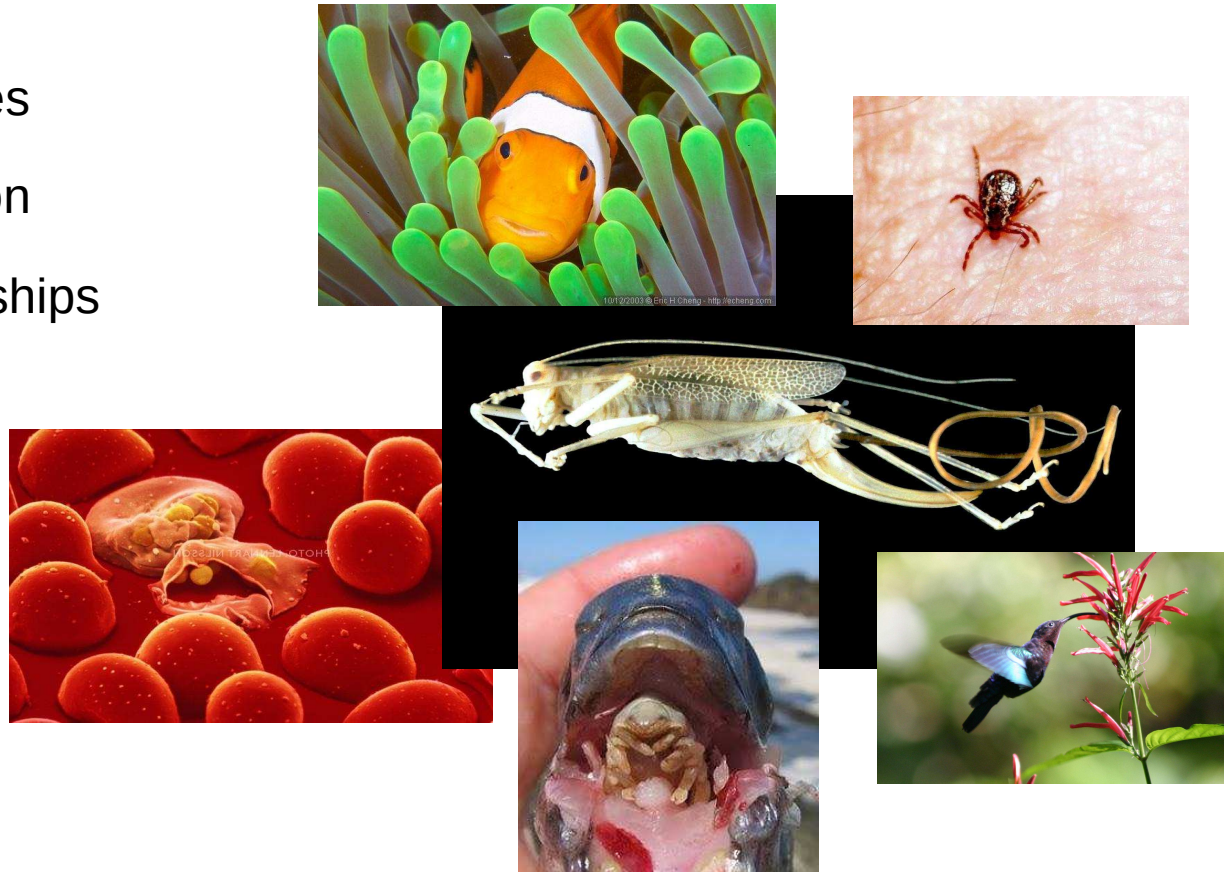
$t = 5$



Coevolutionary System

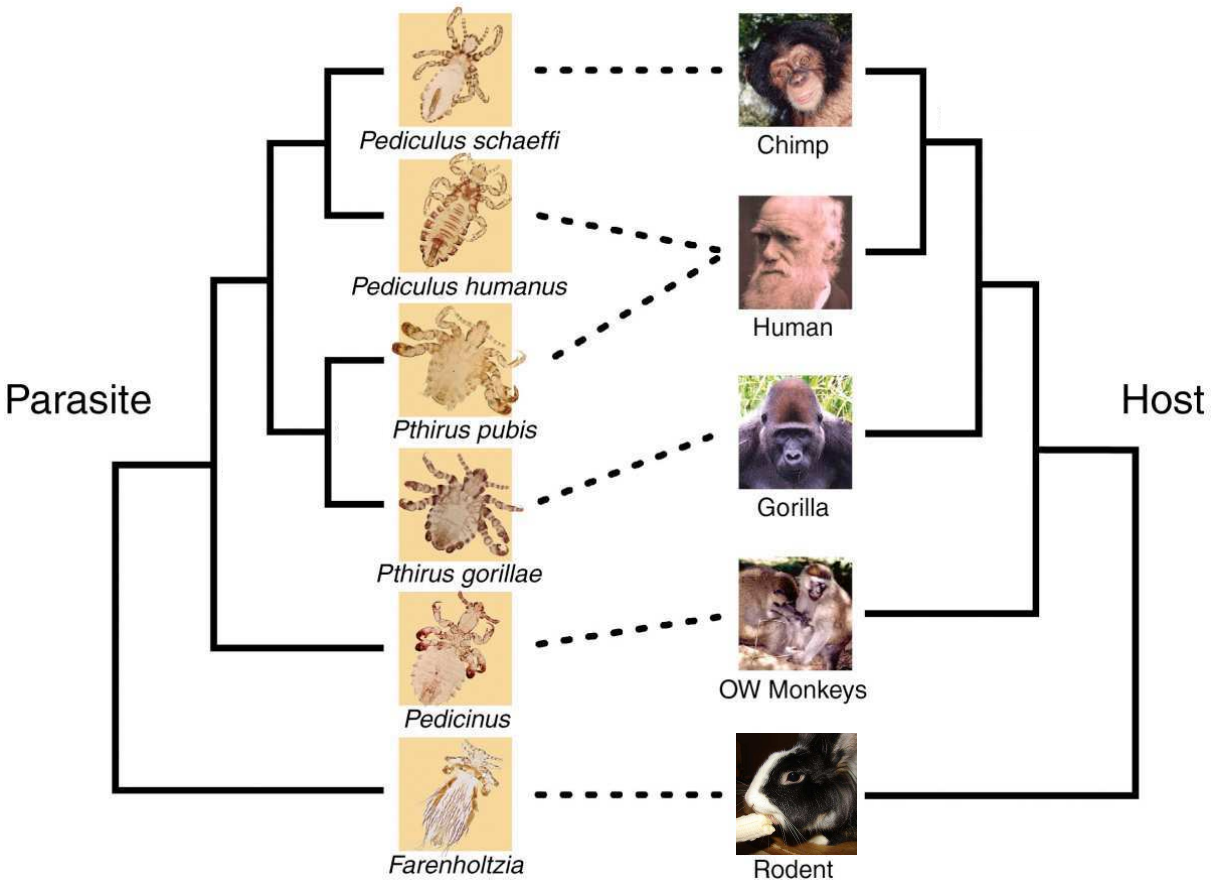
"change of a biological object triggered by the change of a related object"

- hosts and parasites
- insect-plant relation
- symbiotic relationships



Coevolutionary System

- hosts phylogeny
- parasite phylogeny
- leaf-to-leaf association



Coevolutionary Scenario

Goal:

Infer coevolutionary history based on given phylogenies of both groups.

Approach:

Use an evolutionary model that describes the set of possible types of events that happened during coevolution.

Coevolutionary Scenario

Goal:

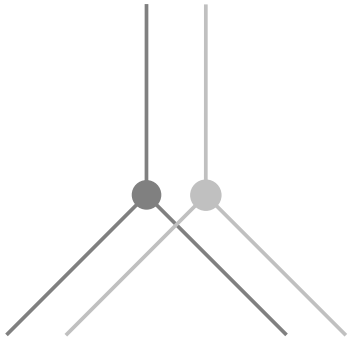
Infer coevolutionary history based on given phylogenies of both groups.

Approach:

Use an evolutionary model that describes the set of possible types of events that happened during coevolution.

Events

Cospeciation



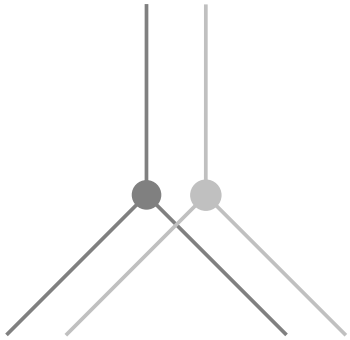
Duplication

Sorting

Host switch

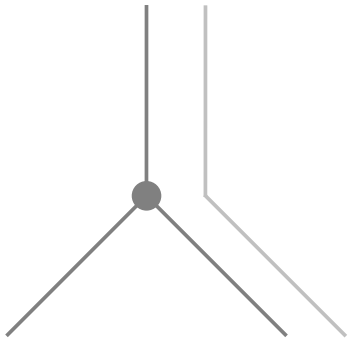
Events

Cospeciation



Duplication

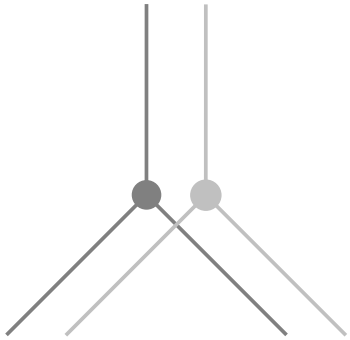
Sorting



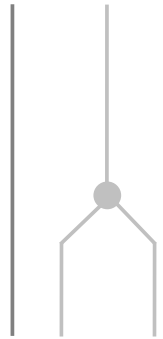
Host switch

Events

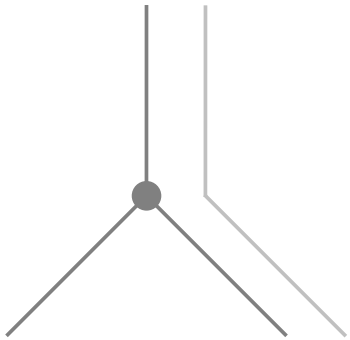
Cospeciation



Duplication



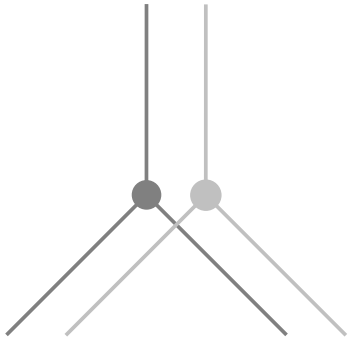
Sorting



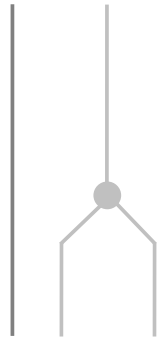
Host switch

Events

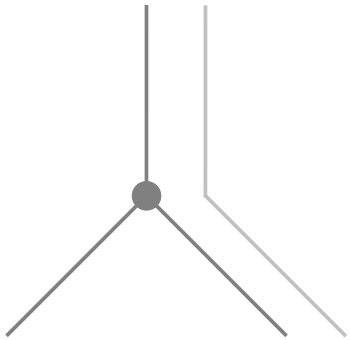
Cospeciation



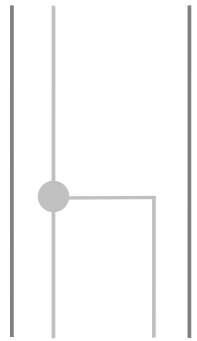
Duplication



Sorting



Host switch



Common Simulations

1. generate host tree of treesize using model
2. define event probabilities p_{event}
3. generate parasite tree using p_{event}

Disadvantages:

- treesize must be known in advance
- parasite tree modelfree
- 4 parameter
- parasite tree can get huge
- no chronological information

Common Simulations

1. generate host tree of treesize using model
2. define event probabilities p_{event}
3. generate parasite tree using p_{event}

Disadvantages:

- treesize must be known in advance
- parasite tree modelfree
- 4 parameter
- parasite tree can get huge
- no chronological information

Simulation using Age Model

Init: leave list for T_{host}, $T_{parasite}$ each, p_{event}

While size T_{host} or $T_{parasite}$ not reached **do**

With $p_{treetype=host}$ choose leave $l \in L_{host}$ else $l \in L_{parasite}$ according to age model probability

If $l \in T_{host}$ do host speciation

If $p_{cospeciation}$ do cospeciation

Else do sorting

If $l \in T_{parasite}$

If p_{switch} do switch

Else do duplication

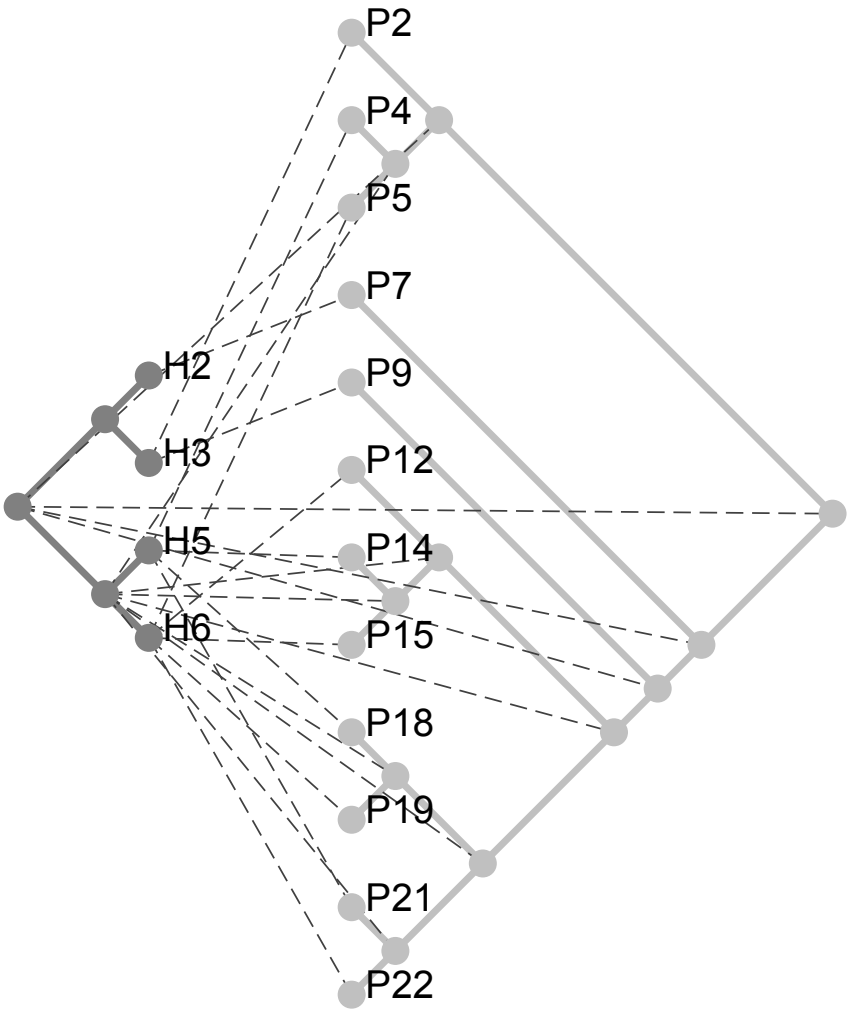
Update L_{host} or $L_{parasite}$

Alternative: merged leave list $L_{host,parasite}$

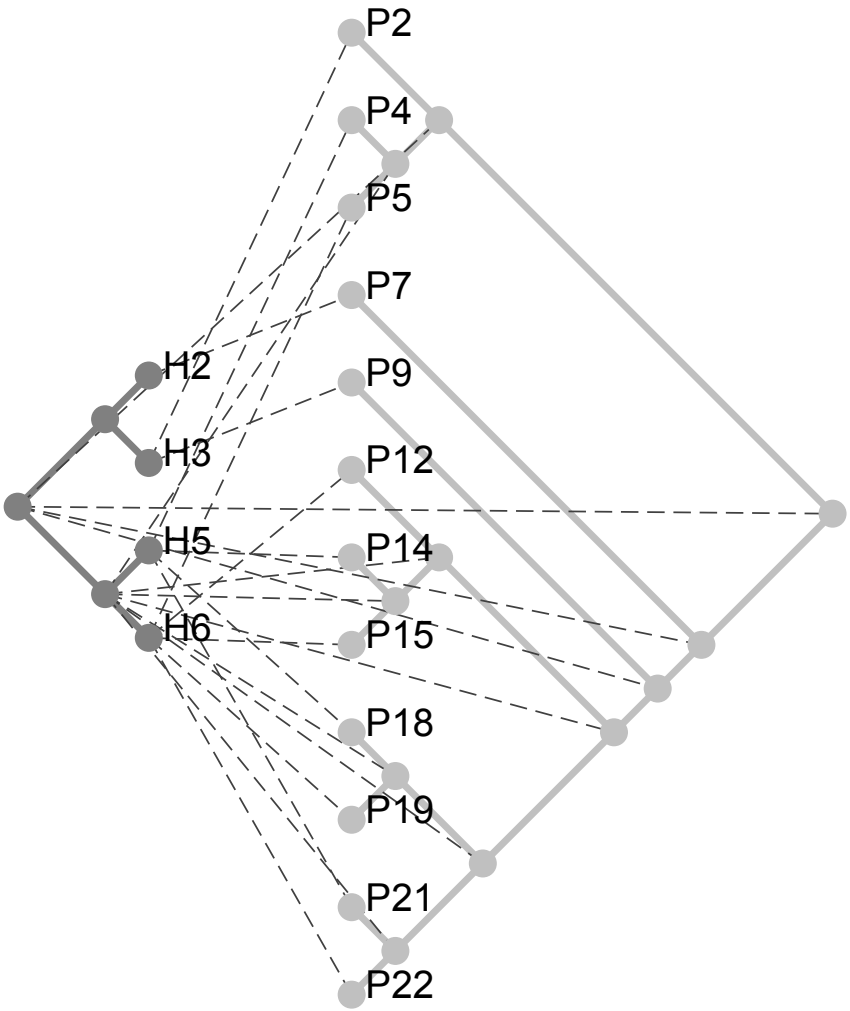
Reconstruction

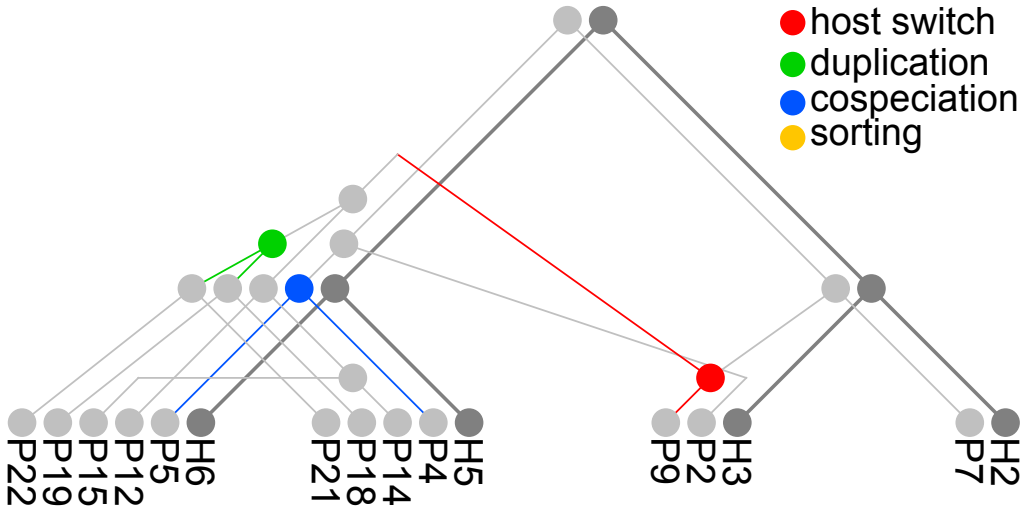
- mapping from parasite node to host nodes
- n^m combinations ($n = \#\text{host}, m = \#\text{parasites}$)
- specify cost for each event type
- rate reconstruction by the sum of all weights
- from all find reconstruction with find minimal costs

Example

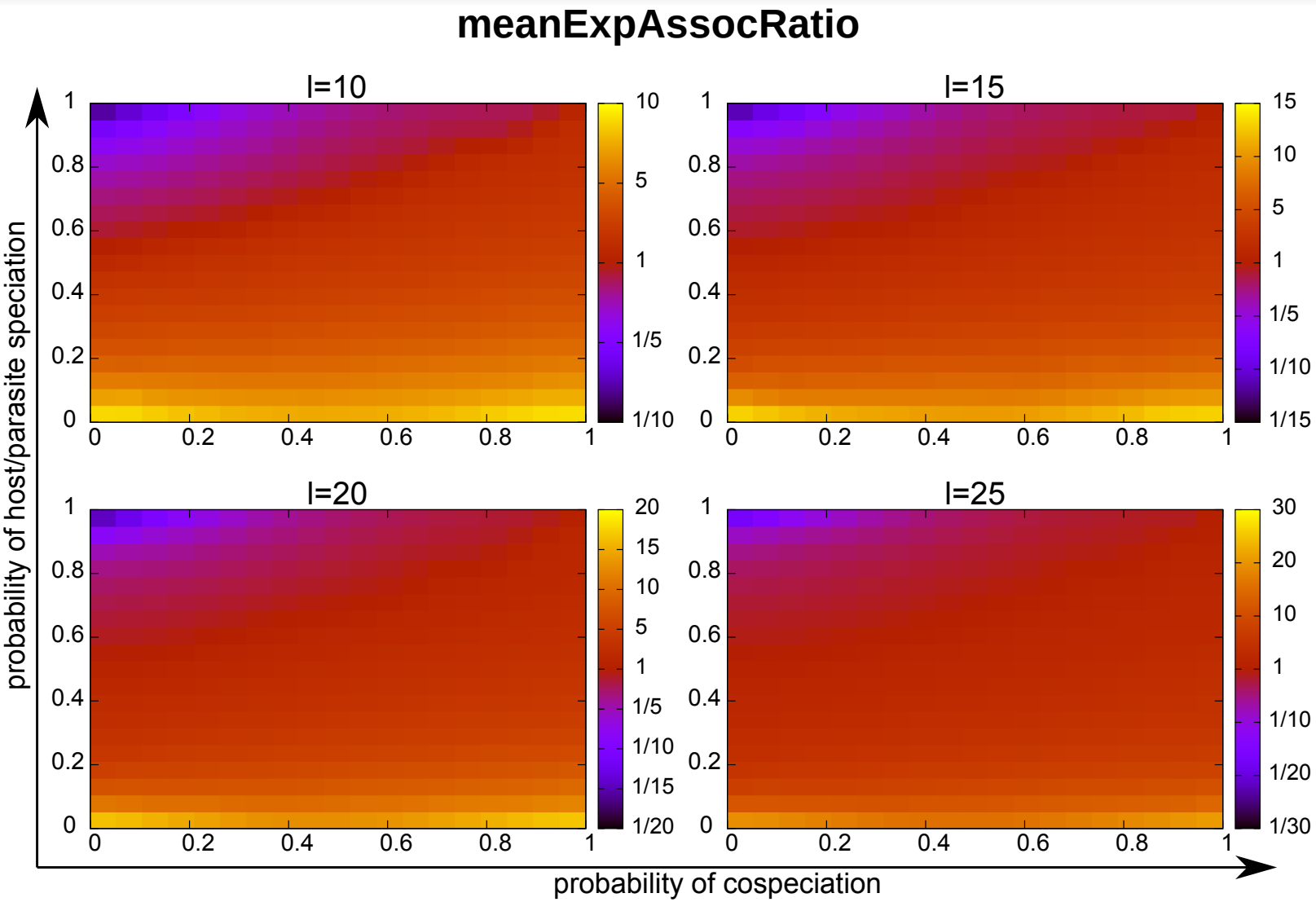


Example

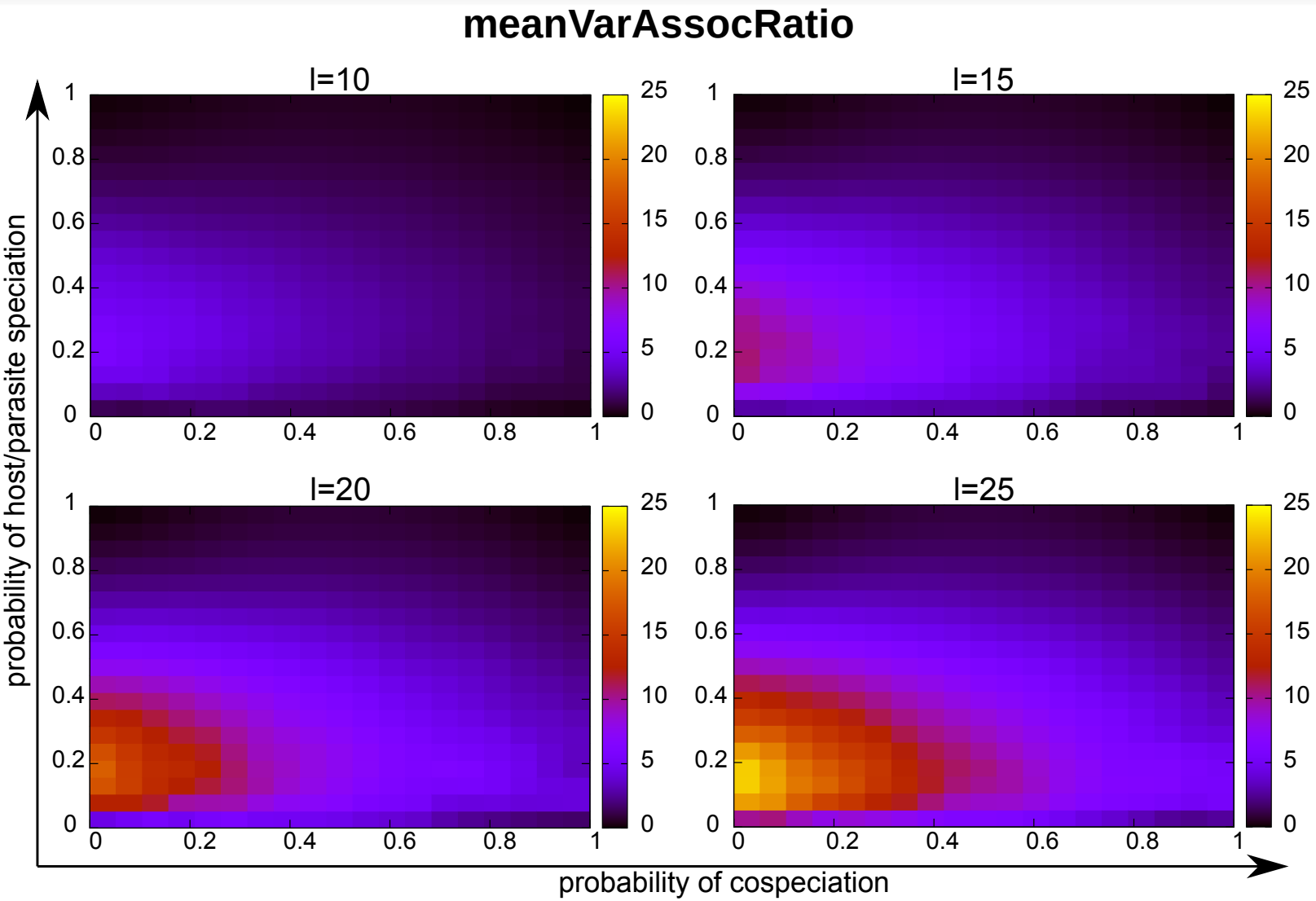

P2
P4
P5
P7
H2
P9
H3
P12
H5
P14
H6
P15
P18
P19
P21
P22


host switch
duplication
cospeciation
sorting
P22
P19
P15
P12
P5
H6
P21
P18
P14
P4
H5
P9
P2
H3
P7
H2

Results


meanExpAssocRatio
l=10
l=15
l=20
l=25
probability of host/parasite speciation
probability of cospeciation
0
0.2
0.4
0.6
0.8
1
10
5
1
1/5
1/10
15
1/15
20
1/20
30
1/30

Results


meanVarAssocRatio
l=10
l=15
l=20
l=25
probability of host/parasite speciation
probability of cospeciation
0
0.2
0.4
0.6
0.8
1
0
5
10
15
20
25

Summary & Outlook

Summary

- no known models to simulate coevolution
- advantages of Age model
  - tree size not necessary in advance
  - generating host and parasite tree simultaneously
  - 3 parameters (sw./dup.; cospec./sort.; host/parasite)
  - chronological information for better simulation

Outlook

- parameter $\rightarrow$ further analysis of reconstructions
- compare with reconstructions using other models
- improve alternative merged list
- different event types

Summary & Outlook

Summary

- no known models to simulate coevolution
- advantages of Age model
  - tree size not necessary in advance
  - generating host and parasite tree simultaneously
  - 3 parameters (sw./dup.; cospec./sort.; host/parasite)
  - chronological information for better simulation

Outlook

- parameter $\rightarrow$ further analysis of reconstructions
- compare with reconstructions using other models
- improve alternative merged list
- different event types

Thanks to
Konstantin
and
Nic

&