Cophylogeny reconstruction in host-symbionts systems

#### Blerina Sinaimeri<sup>1</sup>, Laura Urbini<sup>2</sup>, Marie-France Sagot<sup>2</sup> and <u>Catherine Matias<sup>3</sup></u>

<sup>1</sup> LUISS University. Rome, Italy
<sup>2</sup> Inria Lyon, Université de Lyon, CNRS. Lyon, France
<sup>3</sup> Sorbonne Université, Université Paris Cité and CNRS. Paris, France

Juillet 2022

### Outline

1 Host-Symbionts co-evolution setup

2 Co-evolution models and reconciliation approaches

3 AmoCoala: simulating symbiont species trees along host species trees allowing multiple associations

### Outline

### 1 Host-Symbionts co-evolution setup

2 Co-evolution models and reconciliation approaches

3 AmoCoala: simulating symbiont species trees along host species trees allowing multiple associations

< ロ > < 同 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ >

## Co-evolution, Reconciliation and Cophylogeny

- **Co-evolution** is the study of ancestral relations between two groups of organisms that are ecologically related. *E.g. Host species and their symbiont species*
- Usually, one 'upper' group (e.g. Host species) and one 'lower' group (e.g. Symbiont species)
- **Reconciliation** acts on 2 different evolutionary histories of 2 co-evolving groups and aims at explaining one of these (in general lower group) w.r.t the other (in general, upper group)
- **Cophylogeny** is a reconciliation method that explains the phylogenetic tree of the 'lower' group w.r.t the one of the 'upper' group.

< ロ > < 同 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ >

## One setup, many different contexts I

Historical associations among genes, organisms and geographical areas share fundamental similarities.

#### Different contexts

. . .

. . .

- Hosts/Symbionts systems
  - Symbiont species co-speciate with their host species,
  - They also independently speciate (called duplication), or undergo hosts switches and 'losses'
- Species/Genes evolution differ, in particular because of
  - gene duplication, losses, horizontal transfers,
  - Incomplete lineage sorting,
- **Biogeography**: geographical areas and the distributional history of organisms that inhabit these areas.

イロト イポト イヨト イヨト

## One setup, many different contexts II

#### One common approach

Most reconciliation methods may be applied (almost) indifferently to any of these different contexts.

genes	symbionts
codivergence	cospeciation
duplication	independent speciation
horizontal transfer	host switch
loss (drift)	extinction

#### Subtle difference

- Hosts-Symbionts setup has the peculiarity that a same symbiont species may inhabit different host species. These are called **multiple** associations.
- This does not happen in the gene-species context where a gene is 'labeled' by the species in which it appears.

### Outline

Host-Symbionts co-evolution setup

#### 2 Co-evolution models and reconciliation approaches

3 AmoCoala: simulating symbiont species trees along host species trees allowing multiple associations

< ロ > < 同 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ >

## Cophylogeny reconciliation I



8/19

< □ > < 同 > < 回 > < 回 > < 回 >

# Cophylogeny reconciliation II

#### Data at hand

- H, S are the host and symbiont species trees
- φ : L(S) → P(L(H)) a function from the leaves of symbiont tree to subsets of leaves of host tree

### Goal of cophylogeny reconciliation

Construct λ : V(S) → P(V(H)) that extends φ to the nodes of symbiont tree and such that the image of a symbiont species s ∈ V(S) never contains an ancestor (in H) in the image of its mother symbiont species.

Note that H and S do not play a symmetric role.

## Cophylogeny reconciliation at a glance

#### Gene/species vs Host/symbiont contexts



Left picture from Hugo Menet's phD manuscript.

		-			
	hum	_	w/m	noel	
30			y y i i i i	posi	um

3

< ロ > < 同 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < 回 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ >

## Modelling co-evolution I

#### Co-evolution scenari

- The first step is to define co-evolution events
- **DTL model**: allows for co-speciation, duplication, transfer and loss [THL11]
- Beware: some events cannot be reconstructed ! Picture from Charleston



"Loss" may occur under different and undistinguishable situations

- extinction of parasite,
- failure to track both hosts after host divergence,

< □ > < 同 > < 回 > < 回 > < 回 >

• sampling error.

# Modelling co-evolution II

#### Parsimony-based vs model-based approaches

• **Parsimony** methods assign a cost value to each event and find co-evolution scenarios with lowest cost. **Pbms**:

- output heavily depends on the choice of these costs
- there are no unique choices: costs should be inversely proportional to probabilities of events ; but probab. unknown !
- different datasets might require different costs
- Model-based methods
  - enable to estimate the probabilities of each event
  - However computing the likelihood is difficult because exploring the space of all possible reconciliations is not feasible,
  - ABC is a simulation tool that solves these issues

12/19

## Approximate Bayesian Computation (ABC) I

#### AmoCoala approach

- The goal is to consider a co-evolution model and estimate its parameter values for a pair of host/symbiont species trees.
- Without maximising a likelihood (likelihood-free approach).
- We rely on a symbiont tree generation algorithm:
  - ▶ input is a host species tree + parameter value for co-evolution model
  - output is a putative symbiont species tree that co-evolved with hosts according to given model
- Then use approximate Bayesian inference (ABC).

イロト 不得下 イヨト イヨト

## Approximate Bayesian Computation (ABC) II

#### ABC principle

Starting from an observed symbiont species tree (data  $D_0$ ), iterate

- Sample a parameter value  $\theta = (p_c, p_d, p_t, p_l)$ , from prior  $\pi$
- Generate dataset  $D_{\theta}$  from this parameter value (following parametric model above and using host tree H),
- Compute **discrepancy** between  $D_{\theta}$  and  $D_0$ :  $d(D_0, D_{\theta})$ . (This may be done through a distance between summary statistics of the data).

Keep  $\tau$ % of values  $\theta$  giving rise to smallest discrepancies.

ABC method belongs to the class of **rejection algorithms** and approximates the posterior  $\mathbb{P}(\theta|d(D_0, D_\theta) \leq \epsilon)$ , where  $\epsilon$  is a **tolerance threshold**.

< ロ > < 同 > < 回 > < 回 > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ >

### Outline

1 Host-Symbionts co-evolution setup

2 Co-evolution models and reconciliation approaches

3 AmoCoala: simulating symbiont species trees along host species trees allowing multiple associations

### Remember DTL model and its limits



- Only 4 events: co-speciation, duplication, transfer, loss
- Impossible to obtain multiple associations (same symbiont in different species)

47 ▶ ◀

### Spreads as fifth events

The main novelty in AmoCoala is the introduction of spread events.

- We introduce two kinds of spreads: vertical and horizontal spreads
- vertical spread = a spread of a symbiont in the entire subtree below a host species. It covers for the case of *failure to diverge* or *speciation as a generalist*.
- horizontal spread  $\simeq$  a "host switch" + 2 different vertical spreads, one occurring in the initial host subtree and the second in the new host subtree.
  - Explains the case where two phylogenetically distant host clades are associated with the same symbiont species.
  - horizontal spread ≠ host switch because in host switch the switching symbiont speciates and creates 2 different symbionts, each one associated to the initial and to the new host respectively.
  - In particular, host switch never induce multiple associations !

	<u> </u>	
lohim	-5vm	inosiiim
	<i>c</i> ,	poorann

### Spreads at a glance: vertical spread



In dotted lines, a reconciliation involving 2 cospeciations in s and  $s_6$  and 1 **vertical** spread of  $s_5$  at  $h_8$ . More precisely, the reconciliation is given by  $\lambda(s) = \{h\}; \lambda(s_6) = \{h_6\}$  and  $\lambda(s_5) = \{h_3, h_4, h_5, h_7, h_8\}$  (on the symbiont leaves, we have  $\lambda = \phi$ ).

< □ > < 同 > < 回 > < 回 > < 回 >

### Spreads at a glance: horizontal spread



In dotted lines, a reconciliation involving a horizontal spread event is shown. The symbiont  $s_5$  makes a **horizontal** spread from  $h_6$  to  $h_7$  (or from  $h_7$  to  $h_6$ ) and thus is associated to the two subtrees  $H_{h_6}$  and  $H_{h_7}$  (*i.e.*  $\lambda(s_5) = H_{h_6} \cup H_{h_7}$ ). The symbiont *s* is associated to a duplication (and  $\lambda(s) = \{h\}$ ) and the symbiont  $s_6$  to a cospeciation (and  $\lambda(s_6) = \{h\}$ ).

< □ > < 同 > < 回 > < 回 > < 回 >

Computation of spreads probabilities

- AmoCoala computes the probabilities of spread events directly from the input  $(H, S, \phi)$
- These probabilities are **not constant** across the host tree: they tend to be larger at the bottom of the tree.
- These quantities heavily rely on the number of multiple associations. In particular
  - The spread probabilities at internal host h are set to 0 whenever the subtree rooted at h contains no multiple associations with the symbiont tree.
  - ► When (H, S, φ) contains no multiple associations, the method AmoCoala reduces to former Coala.

## What AmoCoala does (rough pseudo-code)

**Input data:** Host tree *H*, Symbiont tree *S*, (possibly multiple) associations  $\phi : L(S) \mapsto \mathcal{P}(L(H))$ .

**Parameters:** (with default values) that control the number of simulations.

**Output:** (approximate) posterior distribution of  $\theta = (p_c, p_d, p_t, p_l)$ .

**Pre-computation:** of the spreads probabilities  $p_{vs}(h)$ ,  $p_{hs}(h)$  at each host node h.

for a certain number of simulations do

Choose some  $\theta = (p_c, p_d, p_t, p_l)$ .

Simulate a symbiont tree  $\tilde{S}$  and a reconciliation  $\tilde{\lambda}$  along the host tree H with six events: co-speciation, duplication, transfer, loss (probabilities  $\theta$ ) and vertical and horizontal spreads.

Compare  $(H, \tilde{S}, \tilde{\phi})$  with  $(H, S, \phi)$  and keep  $\theta$  if distance is small end for

### Outline

1 Host-Symbionts co-evolution setup

2 Co-evolution models and reconciliation approaches

3 AmoCoala: simulating symbiont species trees along host species trees allowing multiple associations

- 4 同 ト 4 ヨ ト 4 ヨ ト

## Acacia & Pseudomyrmex (AP)

Acacia plants and Pseudomyrmex (ants), from[?]. 22 multiple associations (MA).



Jobim - Symposium	AmoCoala	Juillet 2022	23/19

Comparison of results obtained by Coala (without MA - first row) and AmoCoala (with MA - second rows) on AP



- Coala: very low cospeciation ; high transfer and loss. → high incongruence.
- AmoCoala: (adding 2-3 horizontal spreads / not shown) high cospeciation, very low duplication and transfer and lower loss → co-evolution.

Jobim - Symposium

### Conclusion

- We extended the DTL model to account for *spreads* events, thus allowing for multiple associations ;
- AmoCoala is a tool that infers the probabilities of co-speciation, duplication, transfer and loss ;
- These quantities may be transformed into proper costs and input in a parsimony reconciliation tool.
- Try it ! https://team.inria.fr/erable/en/software/amocoala/
- Look at the preprint https://hal.archives-ouvertes.fr/hal-03673256

## Bibliography

[THL11] A. Tofigh, M. T. Hallett, and J. Lagergren. Simultaneous identification of duplications and lateral gene transfers. *IEEE/ACM Trans. Comput. Biology Bioinform.*, 8(2):517–535, 2011.

3

イロト イヨト イヨト

#### Temporary page!

LATEX was unable to guess the total number of pages correctly. A was some unprocessed data that should have been added to the this extra page has been added to receive it. If you rerun the document (without altering it) this surplus page away, because LATEX now knows how many pages to expect for t

document.