Precision Tango

Inria Lyon
and
University of Lyon / CNRS

Precision Tango

Motivation for the two terms of the title should appear clearer as the talk goes on

Precision ⟵ **Main, important reason: Interpretability**

Not the first time a researcher uses the word Tango in the title of a talk

Notice that the person I found out about afterwards is an excellent dancer which is not my case

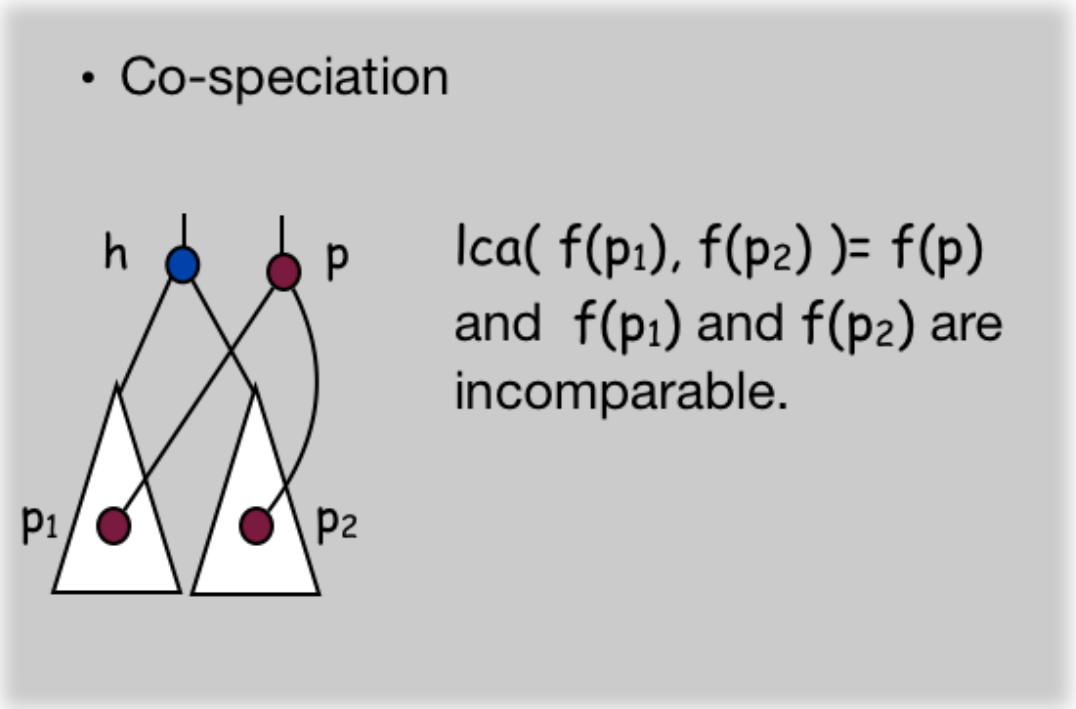
⟶ **Tango**

One problem – Cophylogeny reconciliation

Input:
Two undated phylogenetic trees *H* and *P*
A mapping *f* of the leaves of *P* onto the leaves of *H*
Output:
An optimal extension of *f* to the internal vertices of *P* onto the vertices of *H* given that *f* induces a partition of the vertices of *P* in three sets, each corresponding to an "event" with its own associated cost

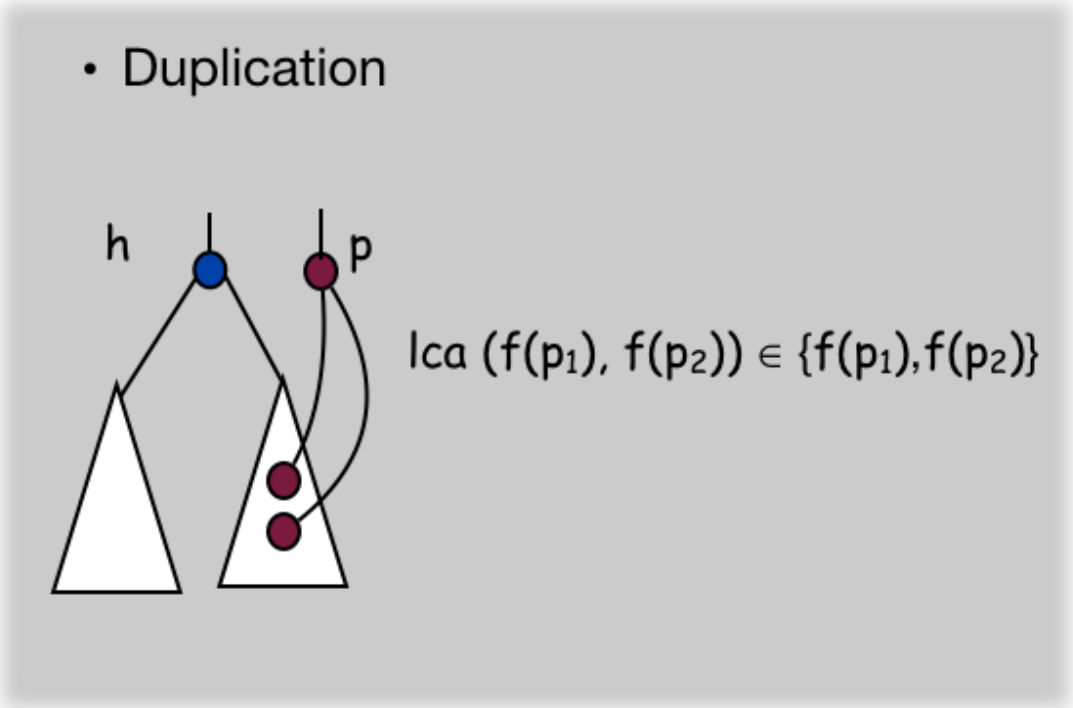
One problem – Cophylogeny reconciliation

Input:
Two undated phylogenetic trees *H* and *P*
A mapping *f* of the leaves of *P* onto the leaves of *H*
Output:
An optimal extension of *f* to the internal vertices of *P* onto the vertices of *H* given that *f* induces a partition of the vertices of *P* in three sets, each corresponding to an "event" with its own associated cost



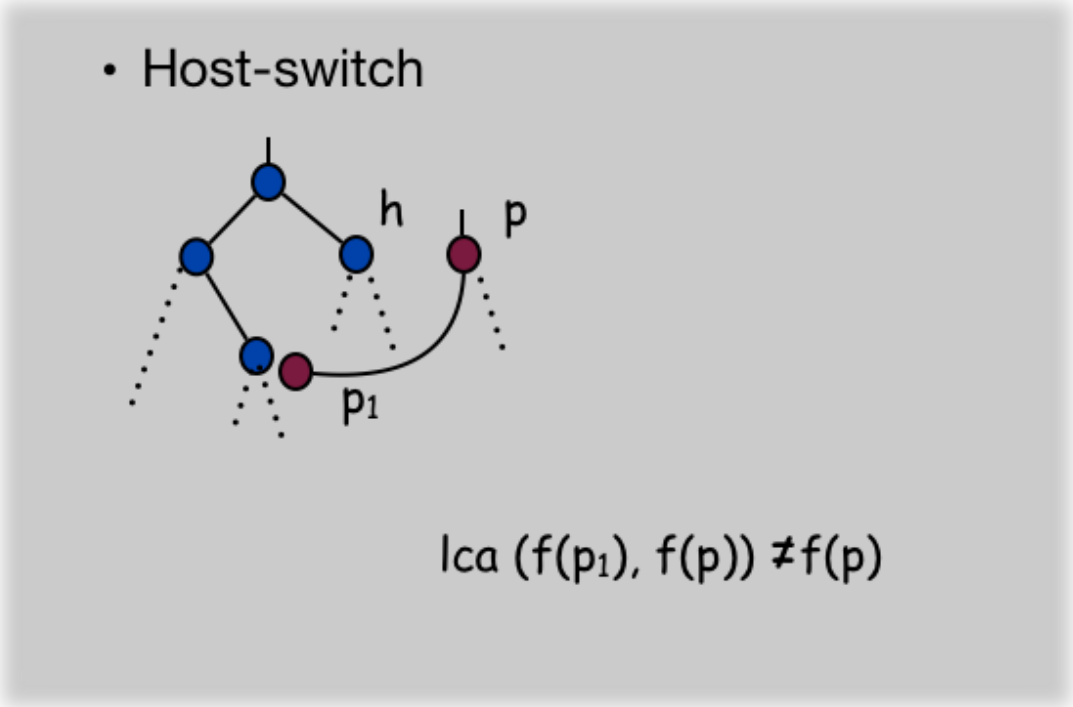
One problem – Cophylogeny reconciliation

Input:
Two undated phylogenetic trees *H* and *P*
A mapping *f* of the leaves of *P* onto the leaves of *H*
Output:
An optimal extension of *f* to the internal vertices of *P* onto the vertices of *H* given that *f* induces a partition of the vertices of *P* in three sets, each corresponding to an "event" with its own associated cost



One problem – Cophylogeny reconciliation

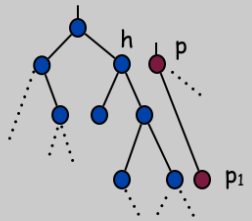
Input:
Two undated phylogenetic trees *H* and *P*
A mapping *f* of the leaves of *P* onto the leaves of *H*
Output:
An optimal extension of *f* to the internal vertices of *P* onto the vertices of *H* given that *f* induces a partition of the vertices of *P* in three sets, each corresponding to an "event" with its own associated cost



One problem – Cophylogeny reconciliation

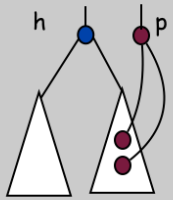
Input:
Two undated phylogenetic trees *H* and *P*
A mapping *f* of the leaves of *P* onto the leaves of *H*
Output:
An optimal extension of *f* to the internal vertices of *P* onto the vertices of *H* given that *f* induces a partition of the vertices of *P* in three sets, each corresponding to an "event" with its own associated cost
plus a cost for a fourth type of event – corresponding to a loss – induced by *f*

- Loss



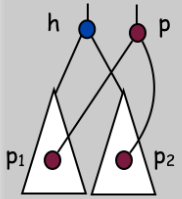
the edge (p, p_1) contributes with 1 loss.

- Duplication



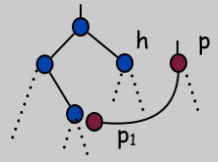
lca ($f(p_1)$, $f(p_2)$) ∈ {$f(p_1)$,$f(p_2)$}

- Co-speciation



lca($f(p_1)$, $f(p_2)$)= $f(p)$
and $f(p_1)$ and $f(p_2)$ are incomparable.

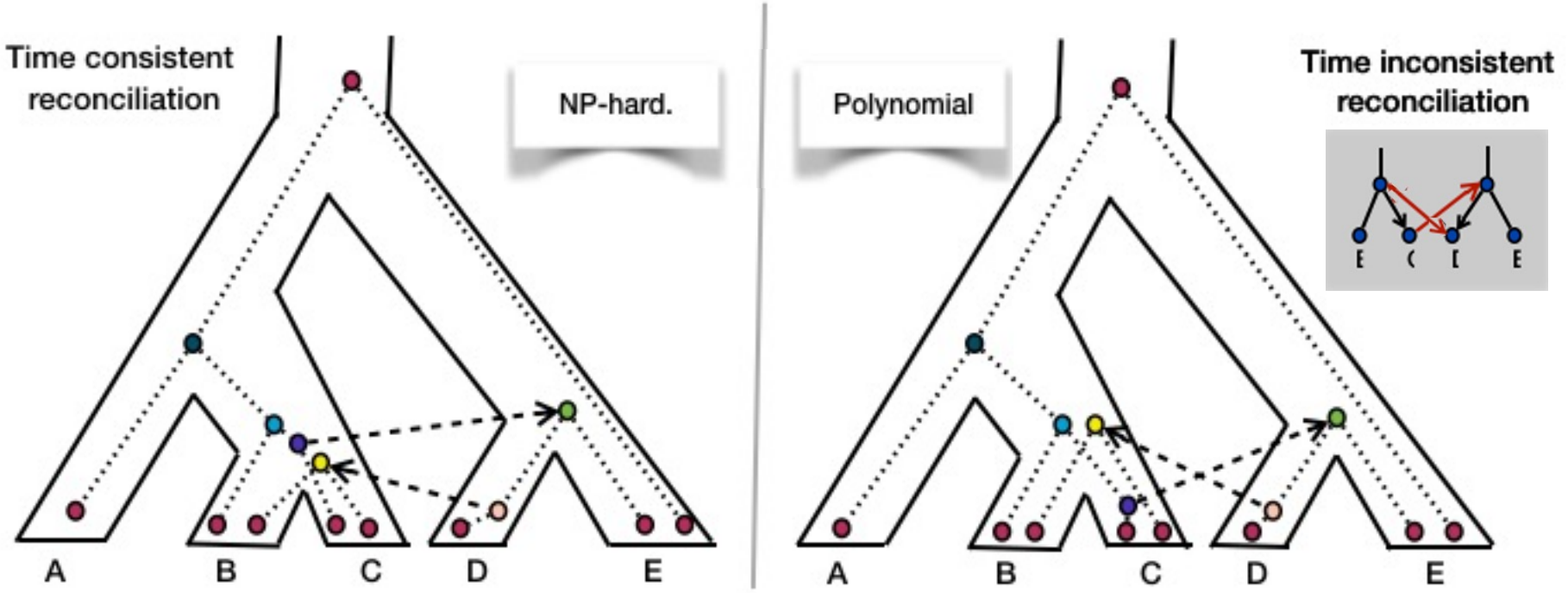
- Host-switch



lca ($f(p_1)$, $f(p)$) ≠ $f(p)$

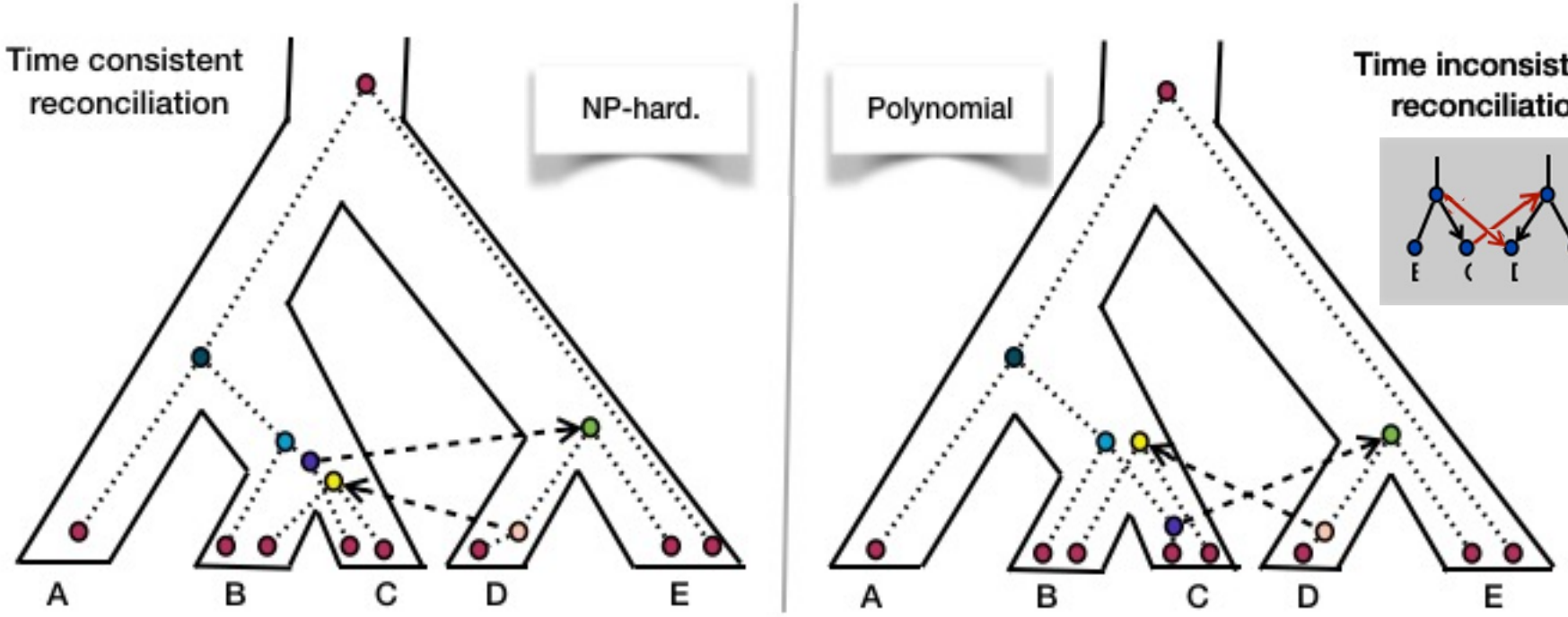
Complexity of this problem

Input:
Two undated phylogenetic trees *H* and *P*
A mapping *f* of the leaves of *P* onto the leaves of *H*
Output:
An optimal extension of *f* to the internal vertices of *P* onto the vertices of *H* given that *f* induces a partition of the vertices of *P* in three sets, each corresponding to an "event" with its own associated cost plus a cost for a fourth type of event – corresponding to a loss – induced by *f*



Counting also is easy

Complexity of this problem



Counting also is easy

There is however a way around it even in the case of undated trees that preserves exactness and works well in practice

This passes through enumeration (listing) in polynomial-time delay, perhaps also of suboptimal solutions and not only of optimal ones exclusively, combined with checking a posteriori time-consistency, which can be done efficiently (polynomial time, more precisely $O(n^2)$ where n is the number of vertices of H)

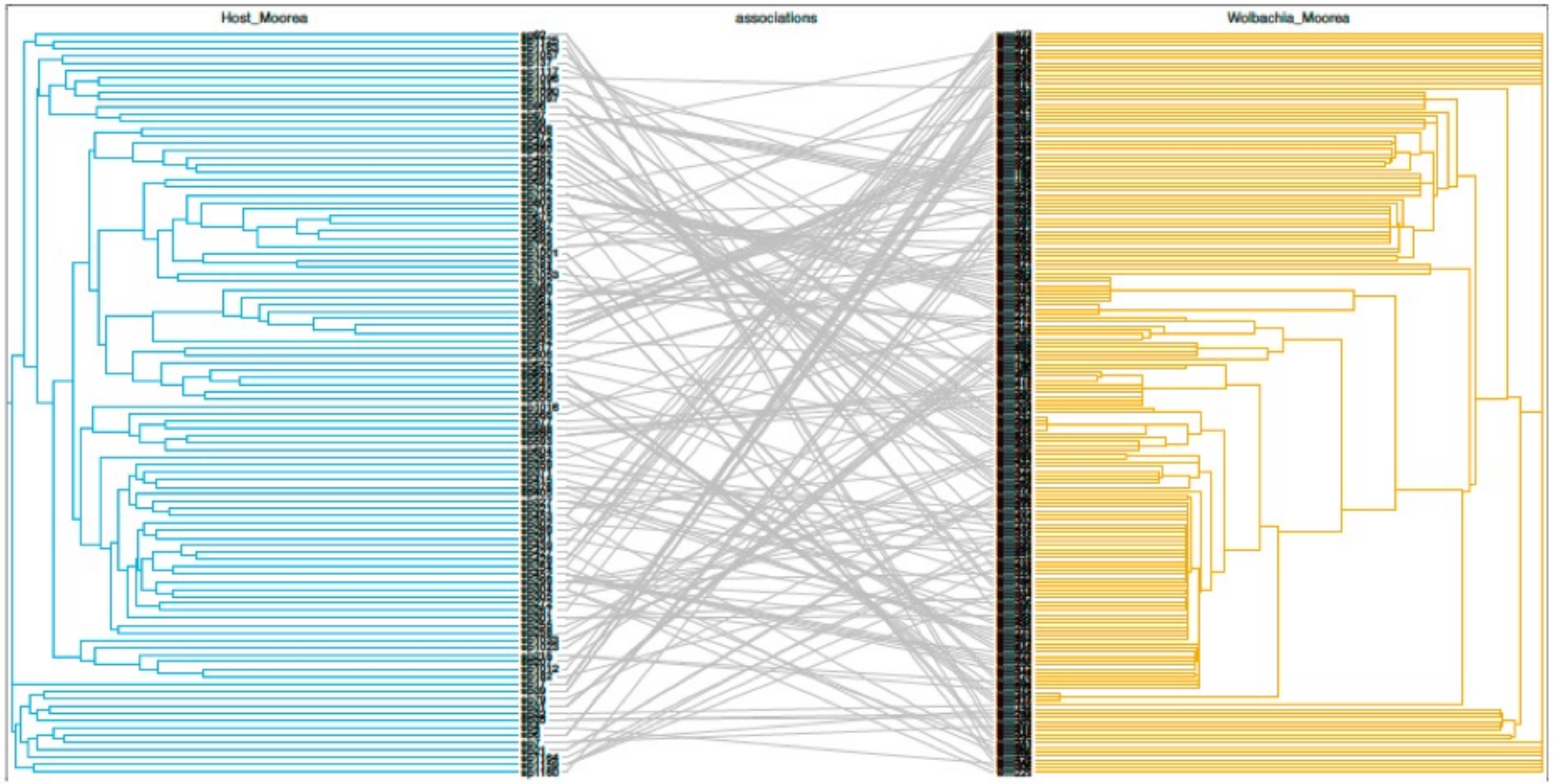
Enumeration

Obtaining time-consistent optimal solutions is however not the only reason that motivated enumeration

The first one was actually the following:

- **Two undated phylogenetic trees with 773 leaves each**
- **Number of solutions above 10^{42} depending on the cost assigned to each of the four events**

Tree *H*: Arthropods
Tree *P*: *Wolbachia*



Enumeration

Obtaining time-consistent optimal solutions is however not the only reason that motivated enumeration

The first one was actually the following:

- **Two undated phylogenetic trees with 773 leaves each**
- **Number of solutions above 10^{42} depending on the cost assigned to each of the four events**

Which to choose?

One possible reaction:

- **Who cares, just pick one!**

Problem:

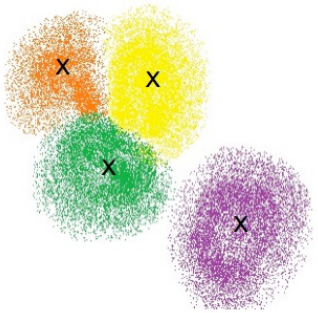
- **There may be big differences among the solutions, even in terms of just the number of each event**
- **Of course this depends on the cost assigned to each event, and on the trees**
- **However, these differences are observed even for quite small trees (approximately 10 leaves each)**

see B. Donati et al. AMB 2015, Y. Wang et al., Bioinformatics 2020 ,Y. Wang et al., WABI 2021

Possible solutions

First natural solution:

Agglomerative clustering that produces consensus or centroid solutions



Problem:

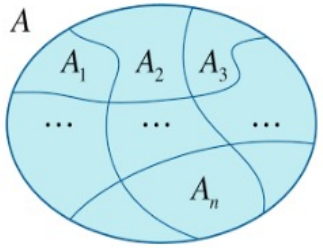
Requires some kind of distance or diversity measure between solutions

Might be less easy to interpret (in terms of biology/evolution)

Second solution:

Establish an a priori equivalence relation among solutions that makes sense (in terms of biology/evolution)

Enumerate the classes of equivalence only, meaning their characteristics and one representative per class

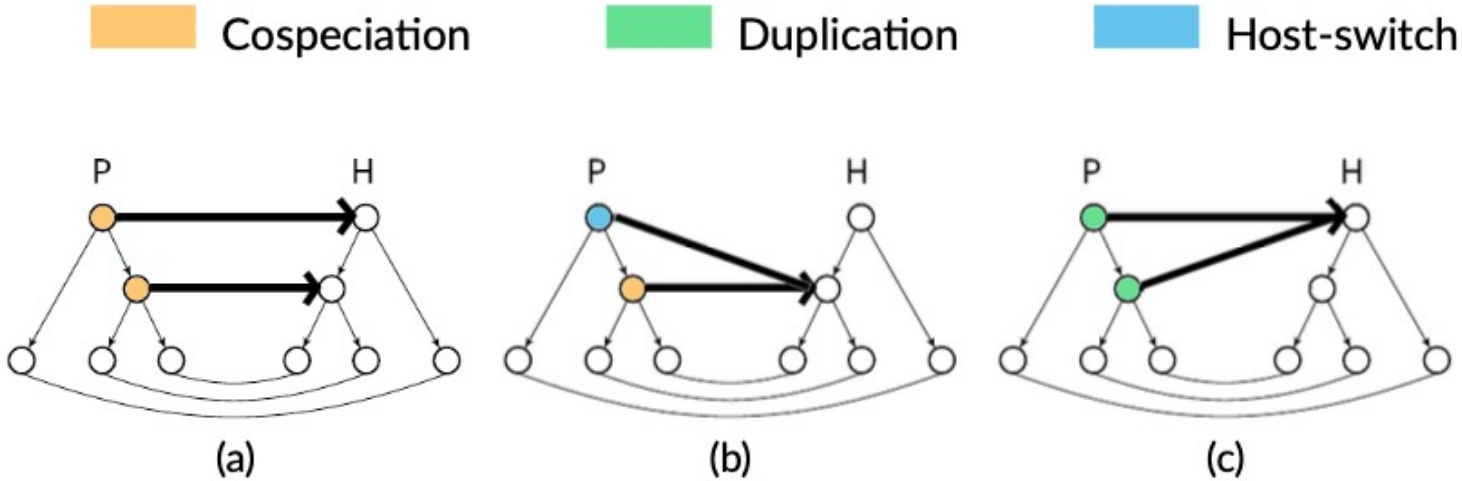


Equivalence relation/classes

First, one important observation:

Each internal vertex of the tree *P* is associated to one, and only one of 3 events

The tree P is thus fully coloured

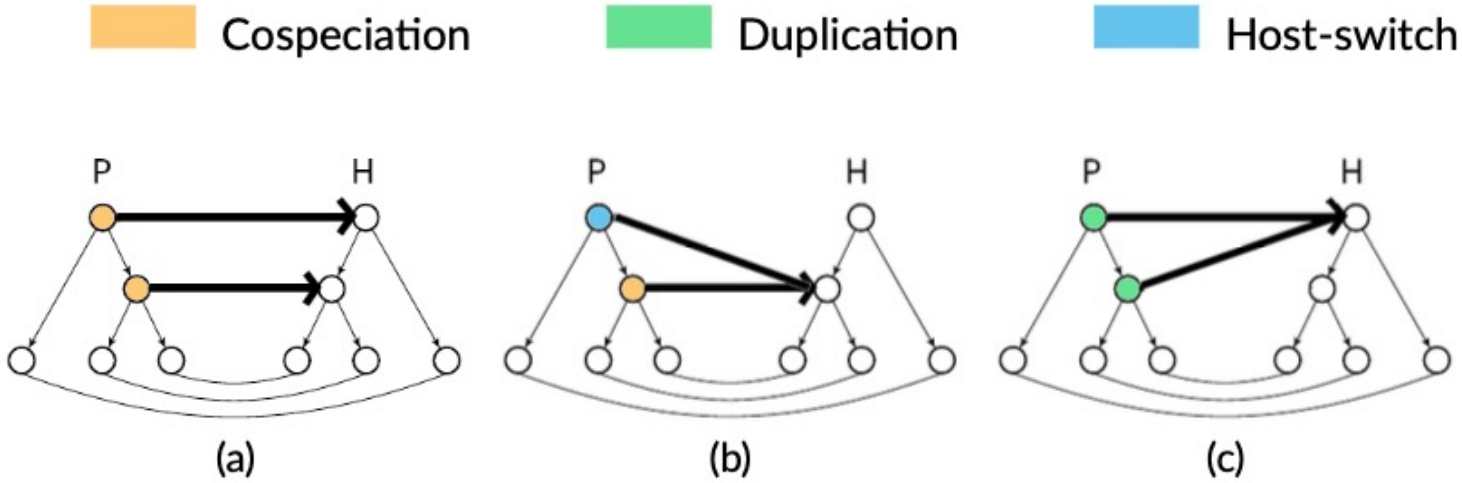


Equivalence relation/classes

First, one important observation:

Each internal vertex of the tree *P* is associated to one, and only one of 3 events

The tree P is thus fully coloured



Possible equivalence relations based on such colouring of the vertices of *P*

V-equivalence: 2 solutions are equivalent if the number of each colour is the same (same event vectors)

E-equivalence: 2 solutions are equivalent if the vertices of *P* are coloured the same way (but we do not care where they are mapped)

EL-equivalence: Same as E-equivalence plus the host-switch arcs are the same

DC-equivalence: Same as E-equivalence plus the co-speciation and duplication arcs are the same

Enumeration of equivalence classes / representatives per class

Does it make a difference in terms of complexity?

Enumerating the classes / one representative per class can be done in polynomial-time delay (namely, in $O(n^2m)$ time where n is the number of vertices in H and m the number of vertices in P)

Does it make a difference in terms of the number of solutions?

V*: does not count losses

199 leaves each tree

773 leaves each tree

Dataset	Cost vector	#Optimal reconciliations	#V-(V*-)equiv. classes	#E-equiv. classes	#EL-equiv. classes	#CD-equiv. classes
COG4965	(−1, 1, 1, 1)	44800	5	13	23456	13
	(0, 1, 1, 1)	17408	2	4	17408	4
	(0, 1, 2, 1)	640	2	3	576	3
	(0, 2, 3, 1)	6528	3	5	448	5
	(0, 1, 1, 0)	907176	324 (10)	12	17	11958
WOLB	(−1, 1, 1, 1)	10^{47}	10	4080	*	24192
	(0, 1, 1, 1)	10^{48}	11	40960	*	76800
	(0, 1, 2, 1)	10^{47}	10	4080	*	24192
	(0, 2, 3, 1)	10^{42}	7	96	10^{36}	1152
	(0, 1, 1, 0)	10^{136}	* (74)	10^{27}	*	*

Characteristics that make enumeration of representatives polynomial-time delay

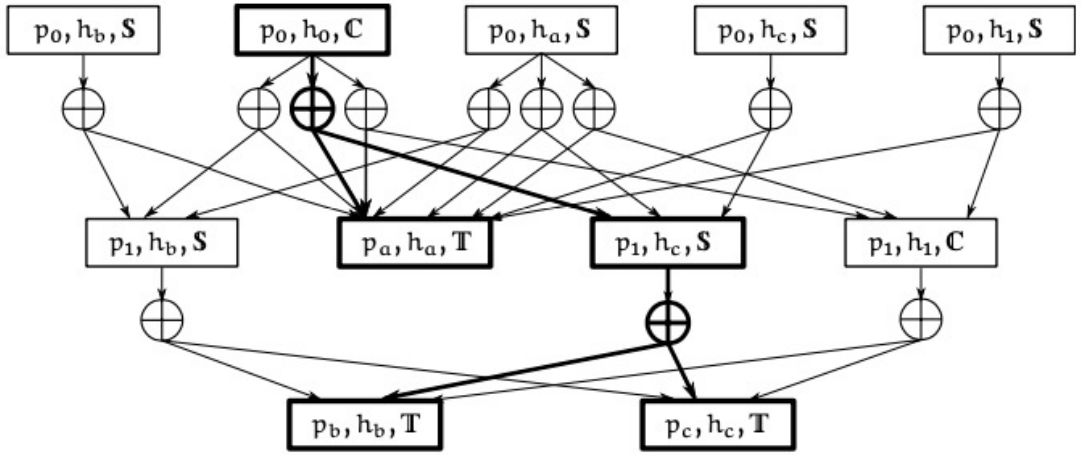
Algorithm is based on dynamic programming
Produces a graph structure which is a compact representation of all reconciliations of minimum cost
This corresponds to a directed AND-OR graph with some additional properties

- **The graph is bipartite (AND and OR nodes) where the children of an OR node are AND nodes, and the children of an AND node are OR nodes or goal nodes (*i.e.* OR nodes with out-degree zero)**
- **The graph is acyclic (it is a DAG)**
- **The graph is decomposable meaning that for any AND node, the sets of nodes that are reachable from each one of its child nodes are pairwise disjoint**

Reconciliation ad-AND/OR graph

Crossed circles are AND nodes
Rectangles are OR nodes

In bold is illustrated one solution



Eucalypt

Capybara

I'll skip the details of the algorithm

see *Y.Wang et al.,WABI 2021*

One initial comment

Algorithm is based on dynamic programming
Produces a graph structure which is a compact representation of all reconciliations of minimum cost
This corresponds to a directed AND-OR graph with some additional properties

Approach generalisable to other problems, not only in computational biology
Two examples:

- Frequency assignment (arising for instance in telecommunication networks)
- Alignment of a sequence (for instance gene sequence) on a tree (for example phylogenetic tree)

see *Y.Wang et al., ESA 2021*

One initial comment

Algorithm is based on dynamic programming
Produces a graph structure which is a compact representation of all reconciliations of minimum cost
This corresponds to a directed AND-OR graph with some additional properties

Approach generalisable to other problems, not only in computational biology
Two examples:

- Frequency assignment (arising for instance in telecommunication networks)
- Alignment of a sequence (for instance gene sequence) on a tree (for example phylogenetic tree)

see *Y.Wang et al., ESA 2021*

Open questions of this more general context

- Is the approach generalisable to other examples?
- Is it easy in such cases to establish relevant equivalence relations, and if yes, which ones?

And more importantly

- We heavily rely on the decomposability property of the structure of the solution space (ad-AND/OR graph). Is the problem of enumerating equivalence classes hard without this restriction?

Open problems in cophylogeny reconciliation including about things not mentioned

Open problems related to equivalence class enumeration:

- **Deal with time-consistency: easy for EL-equivalence (namely $O(nm)$ with n the number of leaves in H and m the number of leaves in P), to be established for others**
- **Modify/adapt the approach to deal with dated trees**
- **Explore other equivalence relations**

see *Nøjgaard et al., AMB 2018*

Open problems in cophylogeny reconciliation including about things not mentioned

Open problems related to equivalence class enumeration:

- **Deal with time-consistency: easy for EL-equivalence (namely $O(nm)$ with n the number of leaves in H and m the number of leaves in P), to be established for others** — see *Nøjgaard et al., AMB 2018*
- **Modify/adapt the approach to deal with dated trees**
- **Explore other equivalence relations**

Coala

Open problems / things not mentioned related to cophylogeny reconciliation more in general:

- **Event cost inference** — see *C. Baudet et al., Sys. Biol. 2021*
- **Improve current method based on Approximate Bayesian Computation**
- **Cophylogeny model**
  - **Current one allows a vertex of P to be mapped to one, and only one vertex of H: this makes sense when the context is of species/genes reconciliation, it does not make sense in the context of (host) species / (parasite/symbiont) species reconciliation: Spread and failure to diverge** — *Paper in preparation*
  - **Address multi-level cophylogeny**
  - **Deal with uncertainties in the phylogenetic trees given as input**

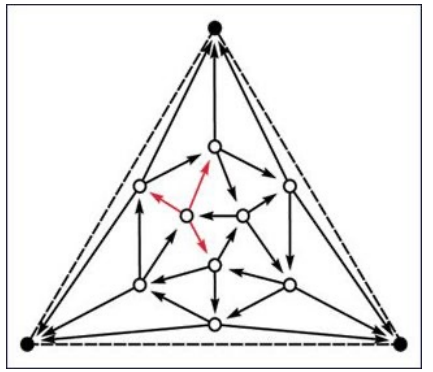
First travel back to the title – Precision Tango

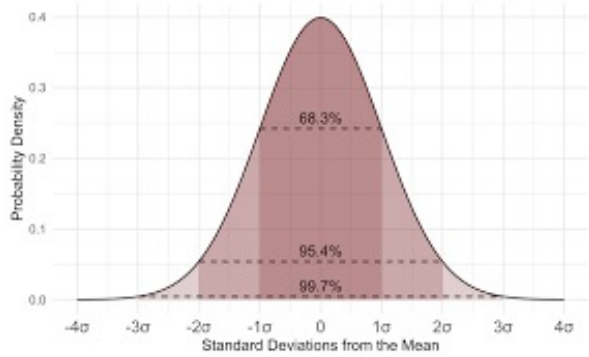
Precision is necessary from both sides; it is much more difficult to reach on the side of biology

What about tango?

Relatively obvious on the side of biology (collective tango if one considers the sets of species)

Required also on the methodological side *e.g.* tango between combinatorics & statistics



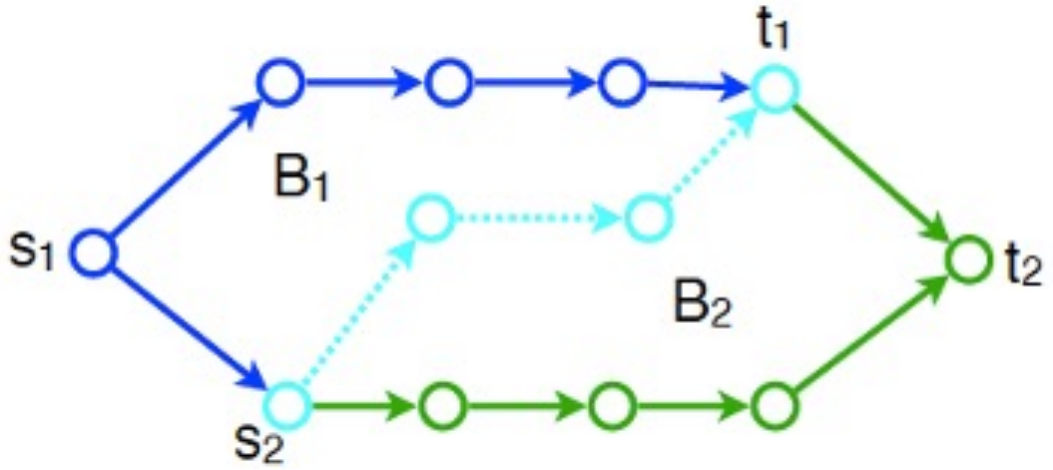


Second problem – Variant detection

Another way of dealing with many / diverse set of solutions
A genomics/transcriptomics context this time

Input:
Directed de Bruijn graph *G* obtained from a set of reads where the vertices are words of length *k* (*k*-mers) and there is an arc between 2 vertices with same suffix/prefix of length *k-1*
Output:
All vertex-disjoint pairs of paths between all pairs of vertices *s* and *t* – also called bubbles
Which correspond to genomic / transcriptomic variants

Small example:



Second problem – Variant detection

Input:

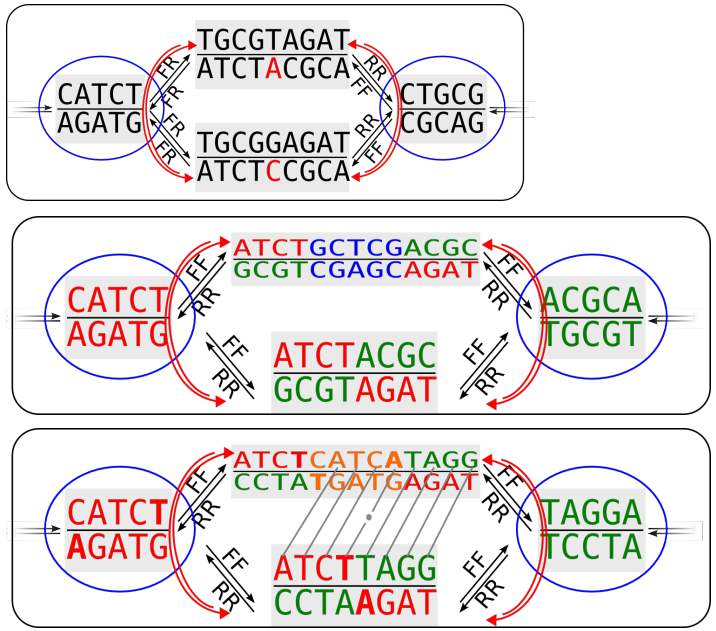
Directed de Bruijn graph G obtained from a set of reads where the vertices are words of length *k* (*k*-mers) and there is an arc between 2 vertices with same suffix/prefix of length *k-1*

Output:

All vertex-disjoint pairs of paths between all pairs of vertices *s* and *t*

The input may include further information on some desired constraints of the bubbles / variants sought

Examples:



Single Nucleotide Polymorphism:
2 paths of length 2k-1

Alternative Splicing:
1 path of length <= 2k-2

Repeats:
1 path of length at most 2k-2,
the two paths align

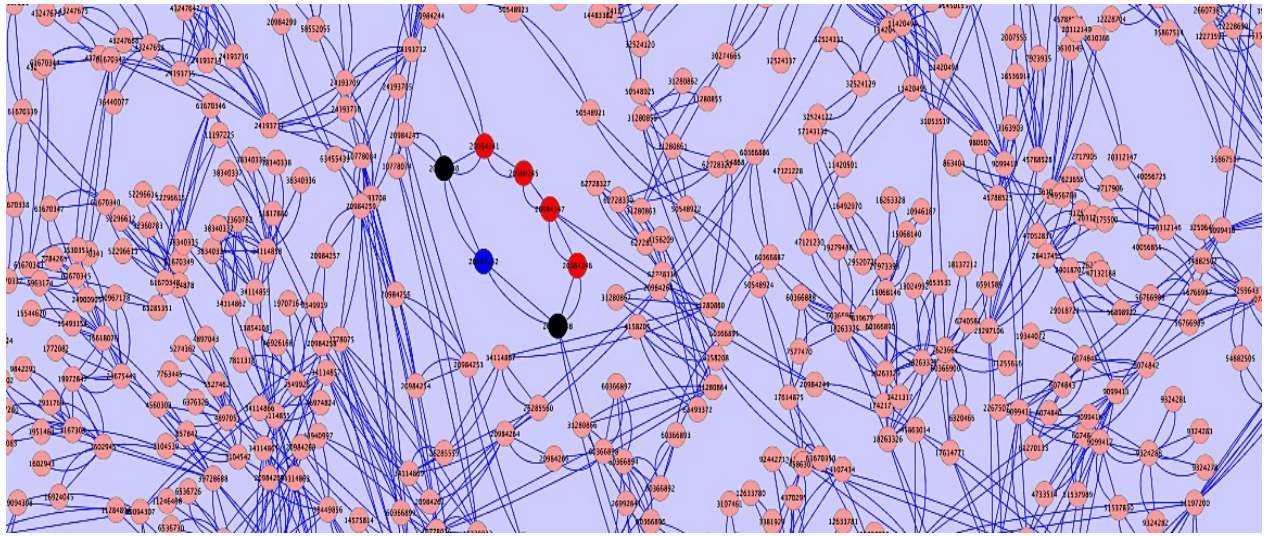
Enumeration

First, clearly here enumeration is required, it is actually an intrinsic part of the problem

The difficulties in this case are:

- **The number, although we do not (seem to) reach those seen for the reconciliation problem**
- **The nature of the input graph**

Drowning in repeats



see *e.g. L. Lima et al., AMB 2017*

KisSplice

Often dealt with by doing some "graph cleaning"

The problem however is that the "graph cleaning" may also clean relevant biological information

Idea for simplifying the enumeration in this case

Enumerate a subset only of the bubbles that however would enable in a second step to recover all

In other words:

- **Define a basis for bubbles as had been previously defined for the cycles of a (di)graph**
- **Elaborate an efficient algorithm to find such basis**
- **Given such basis, elaborate an efficient algorithm to enumerate all bubbles of the input graph**

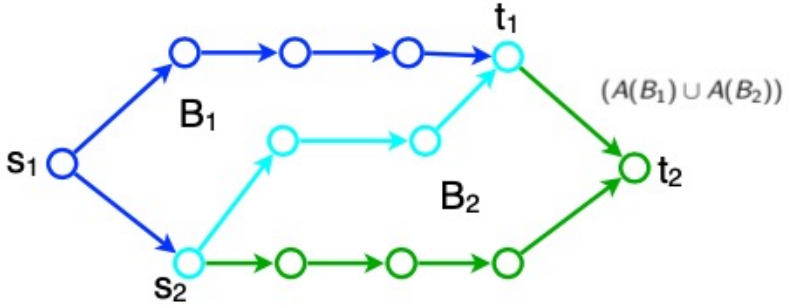
First a comment: the underlying ideas of the methods previously developed for cycle bases in (di)graphs cannot be applied, essentially because they would lead to a set of bubbles from which mathematical/biological objects other than bubbles could then be generated

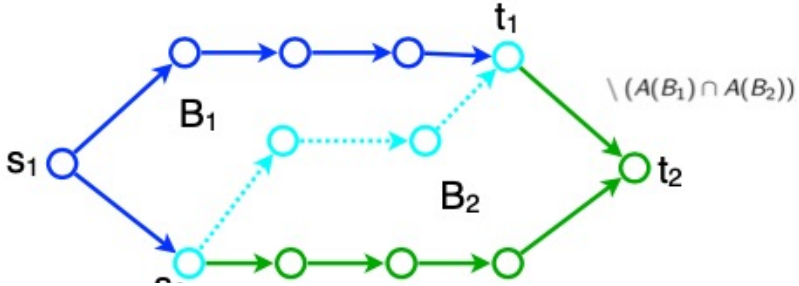
Our definition is then different, and it is not a basis, but a generator, meaning that it is not necessarily minimal except in the case of flow graphs where it is minimum

see *e.g. V. Acuña et al., Algoritmica 2019; V. Acuña et al., IWOCA 2020*
plus work in progress

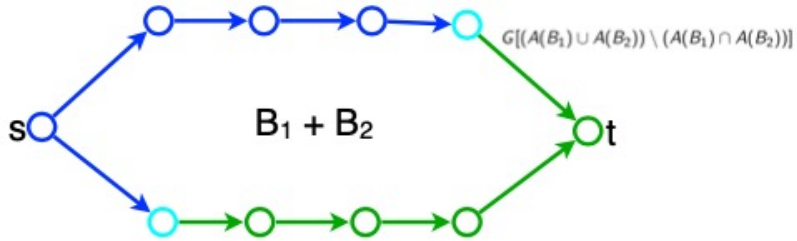
Generator

First, definition of an operator called the symmetric difference



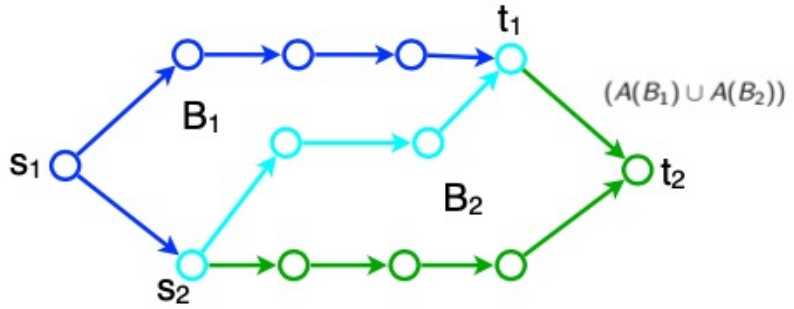


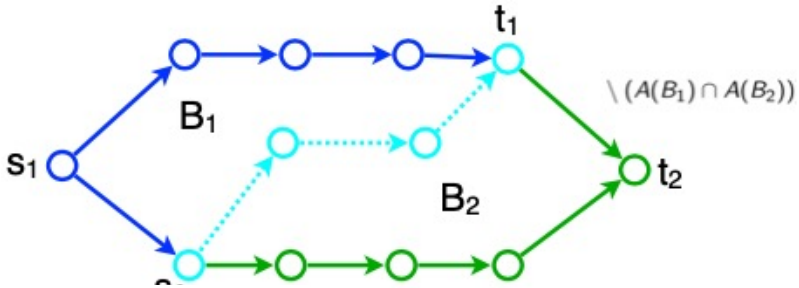
Symmetric difference of B_1 and B_2



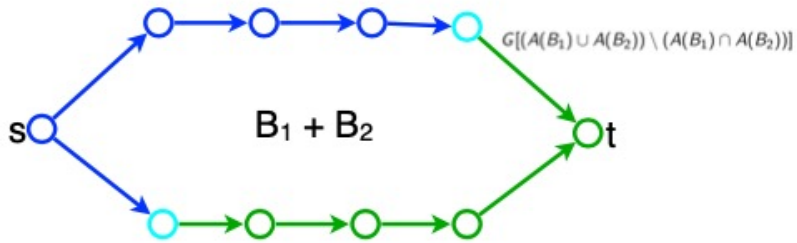
Generator

First, definition of an operator called the symmetric difference





Symmetric difference of B_1 and B_2



Now informal definition of a generator – It is a set of bubbles that satisfies two things:

- **The set contains only bubbles (notice that cycles are considered as a special type of bubble with only one path)**
- **Every remaining bubble of the graph can obtained by combining bubbles in the generator using the symmetric difference operator in such a way that at each step we always have bubbles and only bubbles**

High level view on the underlying method (for flow and for general graphs)

A generator $\mathcal{G}(G)$ satisfying the conditions previously indicated can be obtained from a spanning tree in the case of a flow graph, or from $|S|$ spanning trees in the case of a general graph with $|S|$ sources, where the spanning tree itself is found using DFS, BFS or any other type of graph visit

The size of $\mathcal{G}(G)$ is bounded by $|S|(m-n+1)$ (n=#vertices, m=#arcs, $|S|$=#sources of G), meaning it is bounded by $m-n+1$ for flow graphs

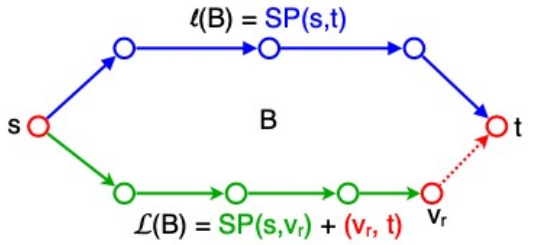
The complexity of the procedure is $O(n)$

Based on the symmetric difference operator, each bubble of G may then be obtained from $\mathcal{G}(G)$ in $O(n)$ time

see *e.g. V. Acuña et al., IWOCA 2020 plus work in progress*

As a comment, previously our definition of a generator was a set of bubbles satisfying two conditions on the vertex-disjoint paths from s to t:

- The shorter one should be the shortest from s to t in G
- The other path should be the shortest from s to the vertex just before t in G



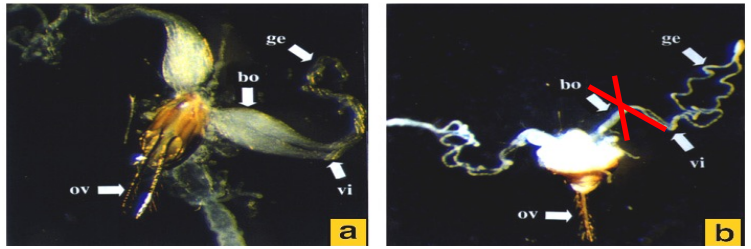
see *V. Acuña et al., Algoritmica 2019*

Tango or not tango?

One example of identifying variants in a context of species interactions

Detection of SNPs in pooled RNAseq was used to identify phenotypes of interest in 2 lines of the insect *Asobara tabida* in the presence or absence of its endosymbiont (tango partner) *Wolbachia*

Symbiotic



Aposymbiotic

KisSplice

see *e.g. VH. Lopez Maestre et al., NAR 2016*

Are however bubble generators nice tango (or any other dance) partners of biology?

Yes and no, not yet

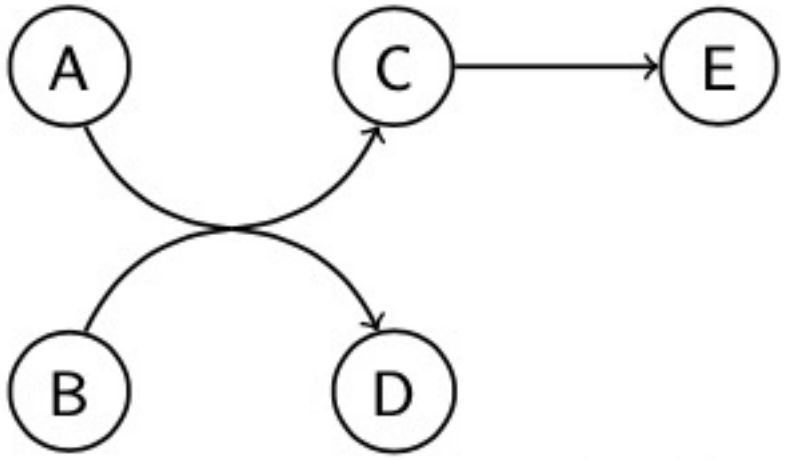
Work in progress

Third problem – Impact of an interaction on the metabolism

Input:
A metabolic network represented as a directed hypergraph

$$R_1 : 2A + B \longrightarrow C + D$$

$$R_2 : 3C \longrightarrow E$$



Directed hypergraph

	R_1	R_2
A	-2	0
B	-1	0
C	1	-3
D	1	0
E	0	1

Stoichiometric matrix

Third problem – Impact of an interaction on the metabolism

Input:
A metabolic network represented as a directed hypergraph
For each reaction in the metabolism (hyperarc), a list of its associated genes
Transcriptomic data, and more precisely the results of a different expression experiment, meaning for each gene:

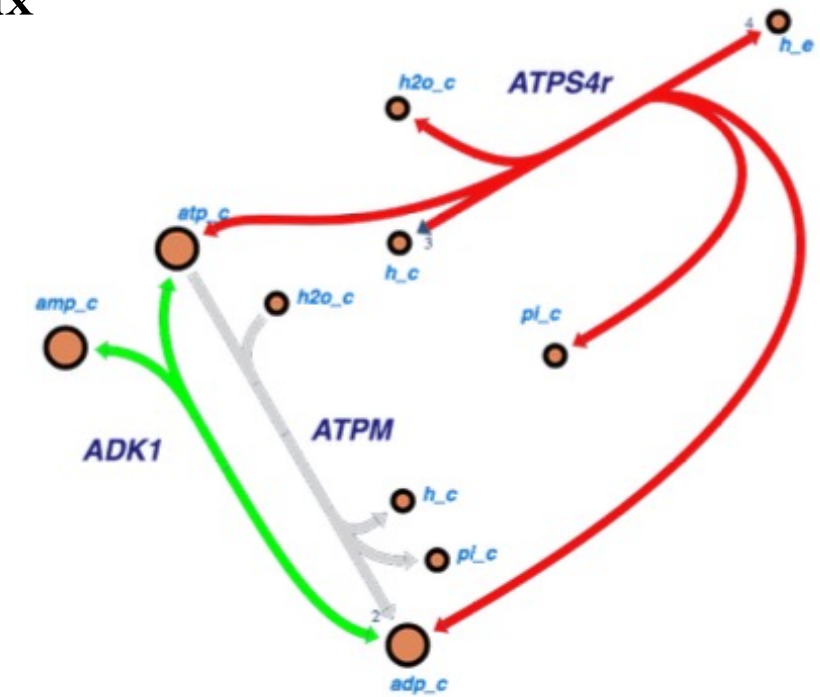
- **Direction of change (up or down)**
- **Probability of a significant change (posterior probability of differential expression - PPDE)**

Reaction	Gene ass.
r_1	$g_1 \text{AND} g_3$
r_2	$g_3 \text{OR} g_4$
r_3	$g_1 \text{AND} g_2$

Third problem – Impact of an interaction on the metabolism

Input:
A metabolic network represented as a directed hypergraph
For each reaction in the metabolism (hyperarc), a list of its associated genes
Transcriptomic data, and more precisely the results of a different expression experiment
Output:
A hypothesis of a metabolic shift, meaning a colouring of the network that defines the status of each reaction as:

- **green: increased flux**
- **red: decreased flux**
- **grey: no change**



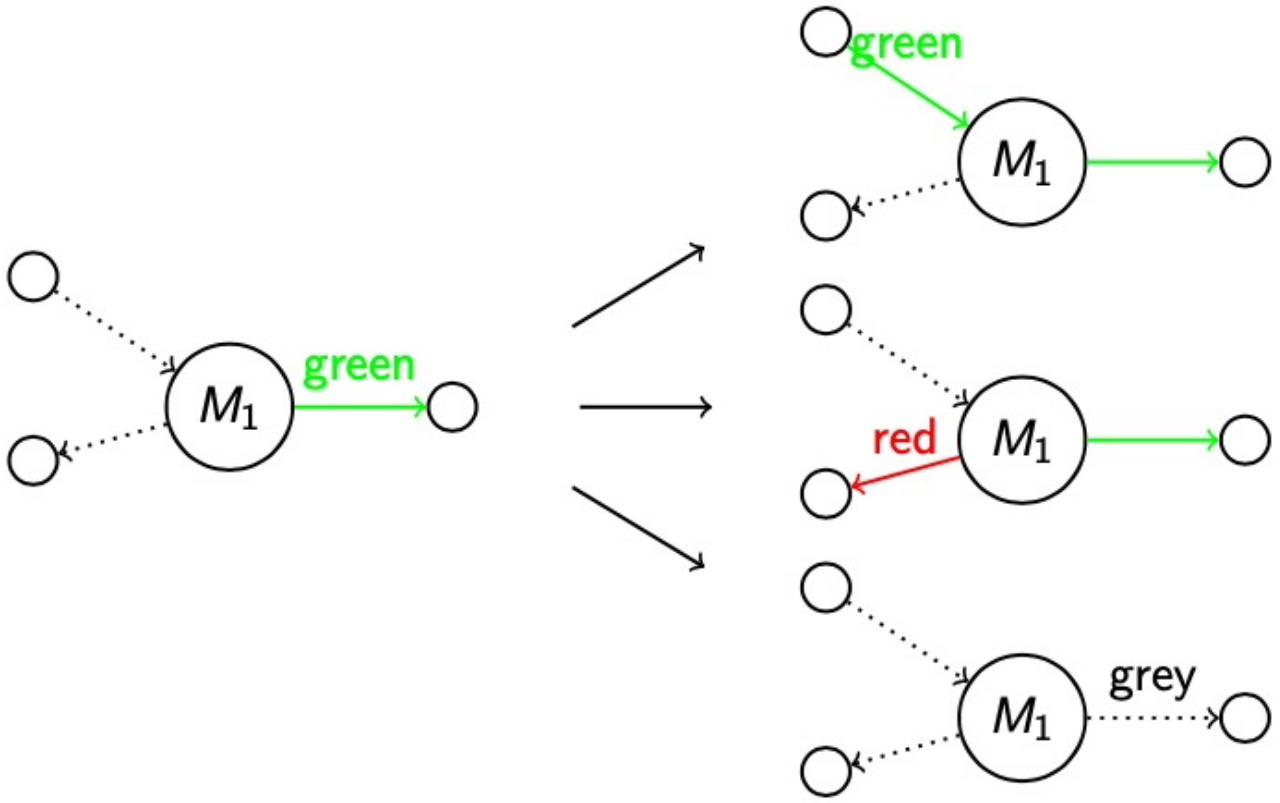
Reaction	Weight
r_1	green
r_2	grey
r_3	red

Solving the problem of providing a hypothesis of a metabolic shift

First, strong assumption made: the metabolic shifts are from one steady state to another (feasibility change assumption)

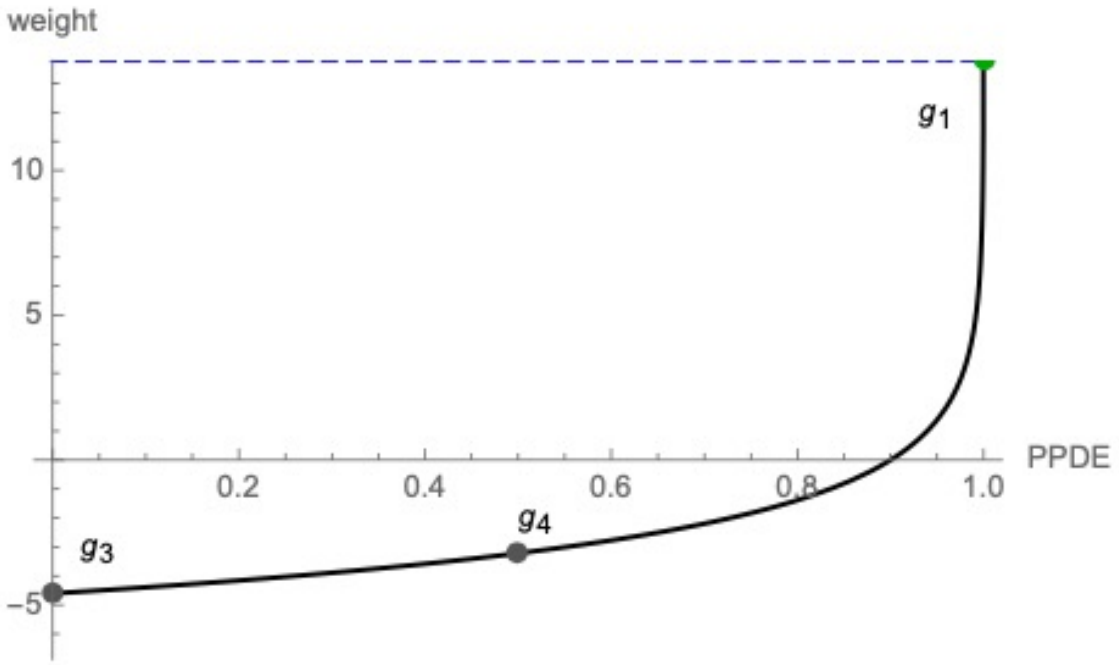
$$S \cdot \mathbf{v}_1 = 0 \text{ and } S \cdot \mathbf{v}_2 = 0 \Rightarrow S \cdot \Delta\mathbf{v} = 0$$

Feasible changes:



Solving the problem of providing a hypothesis of a metabolic shift

Then turn the PPDEs into weights and project the gene colours and weights onto the reactions



Gene	Fold change	PPDE	Colour	Weight
g_1	3	1	green	13
g_2	-2	0.95	red	1
g_3	0.1	0	grey	-5
g_4	0.5	0.5	grey	-3

Reaction	Gene association	Colour	Weight
r_1	g_1 AND g_3	green	13
r_2	g_3 OR g_4	grey	-3
r_3	g_1 AND g_2	grey	-3

Notice that some reactions may be assigned more than one colour, including both red and green (they are then considered as grey), while there may be many which are grey
The grey ones indicate that no change was detected either because indeed there was no change or because this was not measured, or yet because there is a contradictory assignment of both red and green

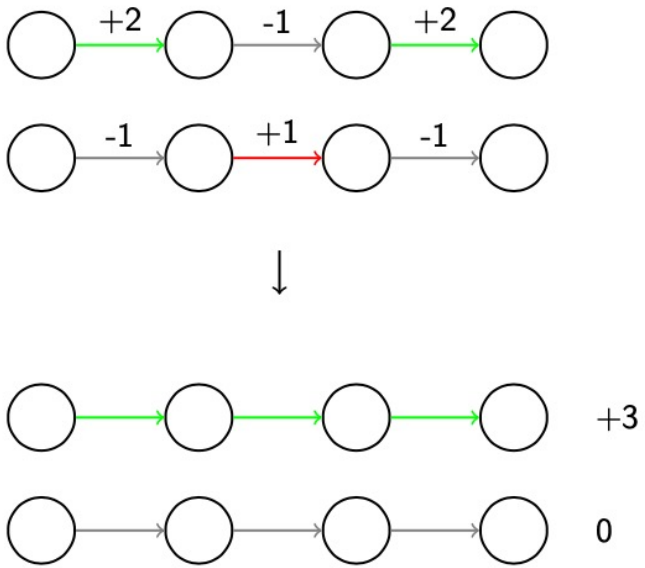
Solving the problem of providing a hypothesis of a metabolic shift

Reformulation of the output:

Maximise the sum of the weights of the reactions with an increased or decreased flux while ensuring that the feasibility change hypothesis is verified which may involve:

- **Changing an arc coloured red or green into grey; or the inverse**
- **Choosing a direction for the bidirectional hyperarcs**

Tiny example:



Feasibility of changes can be ensured using either topological or stoichiometric constraints

High level view of the method

The problem is solved using MILP (will skip all details) see *Pusa et al., Bioinformatics 2020*

$$\underset{(\mathbf{x}^+,\mathbf{x}^-)}{\text{maximise}} \quad \sum_{i=1}^{m}(\mathbf{x}^+{}_i + \mathbf{x}^-{}_i)w_i \qquad (1)$$

$$\text{s.t.} \quad \mathbf{x}^+{}_i + \mathbf{x}^-{}_i \leq 1, i = 1, \ldots, m \qquad (2)$$

$$\mathbf{x}_i^+ = 0 \quad \forall i \quad \text{s.t.} \quad \mathbf{x}_0^-{}_i = 1 \quad \text{and} \quad \mathbf{l}_i = 0 \qquad (3)$$

$$\mathbf{x}_i^- = 0 \quad \forall i \quad \text{s.t.} \quad \mathbf{x}_0^+{}_i = 1 \quad \text{and} \quad \mathbf{l}_i = 0 \qquad (4)$$

$$S \cdot \mathbf{v} = 0 \qquad (5)$$

$$\mathbf{v}_i + \mathbf{x}_i^+(\mathbf{l}_i - \varepsilon) \geq \mathbf{l}_i, i = 1, \ldots, m \qquad (6)$$

$$\mathbf{v}_i + \mathbf{x}_i^+\mathbf{u}_i \leq 0, i = 1, \ldots, m \qquad (7)$$

$$\mathbf{v}_i + \mathbf{x}_i^-(\mathbf{u}_i + \varepsilon) \leq \mathbf{u}_i, i = 1, \ldots, m \qquad (8)$$

$$\mathbf{v}_i + \mathbf{x}_i^-(-\mathbf{l}_i) \geq 0, i = 1, \ldots, m \qquad (9)$$

$$\mathbf{l}_i \leq \mathbf{v}_i \leq \mathbf{u}_i, i = 1, \ldots, m \qquad (10)$$

Here again, there may be more than one solution

Enumeration

In this case, no alternative found so far to enumerating all solutions
However, often here the number of solutions is smaller and all can be enumerated even though the problem of just finding one is NP-hard

One example: *Escherichia coli* mercury exposure see *Pusa et al., Bioinformatics 2020*

- **1684 alternative optima / solutions**
- **Across all these solutions, 1120 of 2578 reactions were present as coloured in all solutions, and 1338 in none**

In general:

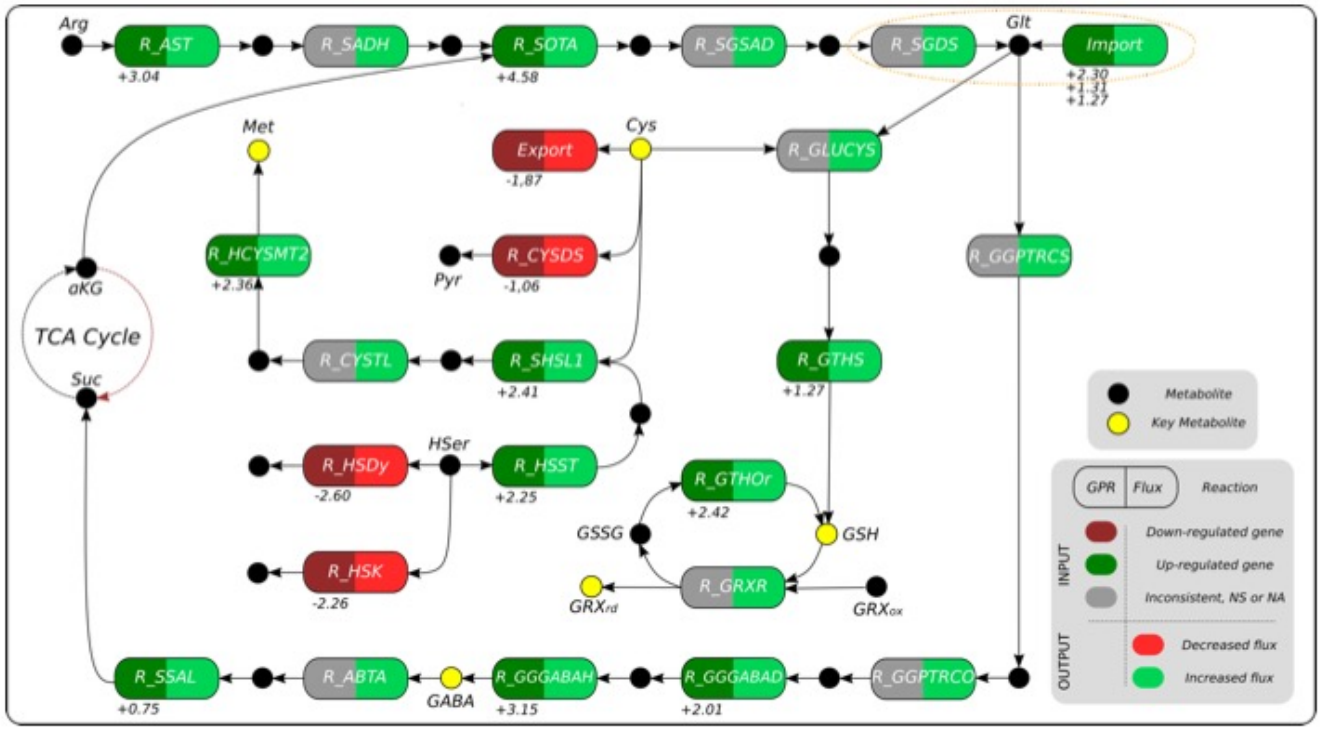
- **Small difference between ensuring feasibility of changes using either topological or stoichiometric constraints**
- **Two possible approaches for the enumeration problem:**
  - **As the algorithm is fast, enumerate all solutions even for bigger networks (*e.g.* Yeast)**
  - **As the consensus among solutions is great in all the cases studied so far, apply a heuristic that consists in enumerating a number of solutions only, say 1000**

High level illustration on the example of *Escherichia coli* exposure to mercury

Reminder

see *Pusa et al., Bioinformatics 2020*

- 1684 alternative optima
- 1120 of the 2578 reactions were present as coloured in all solutions, and 1338 in none
- Analysis of the results done using a consensus across all optima which for most reactions uniquely determined their colour



Key metabolites related to the mercury stress response are highlighted in yellow

Inferring impact on metabolism using other type of data

Metabolomic data

Again solved using MILP

see *Milreu et al., Bioinformatics 2014 plus paper in prep.*

$$\min_{\varphi,y} \quad \lambda \sum_{j=1}^{m} y_j + (1-\lambda) \sum |\mathcal{S} \cdot \varphi|_{\overline{x}} - (1-\lambda) \sum |\mathcal{S} \cdot \varphi|_{x}$$

$$\begin{aligned}
s.t \quad & \Delta^{\min} \leq \mathcal{S} \cdot \varphi \leq \Delta^{\max} \\
& 0 \leq \varphi_j \leq u_j && \forall j \in \mathcal{R} \\
& y_j = 0 \leftrightarrow \varphi_j = 0 && \forall j \in \mathcal{R} \\
& y_j + y_{\bar{j}} \leq 1 && \forall (j, \bar{j}) \in \mathcal{R} \\
& y_j \in \{0,1\}; \lambda \in (0,1); u_j, \varphi_j \in \mathbb{R}.
\end{aligned}$$

Totoro

Minimizing the number of active reactions:

$$\min_{\varphi,y} \quad \lambda \sum_{j=1}^{m} y_j$$

Maximizing the changes in measured metabolites:

$$\max_{\varphi,y} \quad (1-\lambda) \sum |\mathcal{S} \cdot \varphi|_{x}$$

Minimizing the changes in unmeasured metabolites:

$$\min_{\varphi,y} \quad (1-\lambda) \sum |\mathcal{S} \cdot \varphi|_{\overline{x}}$$

where λ balances the objectives

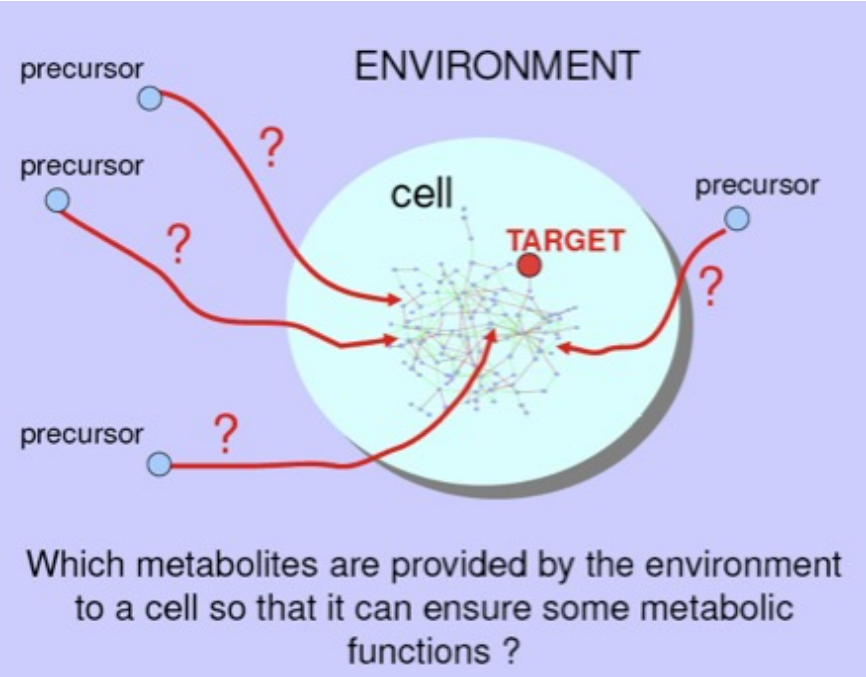
General perspective on analysing impact on metabolism

Integrating both type of data, plus possibly proteomics data

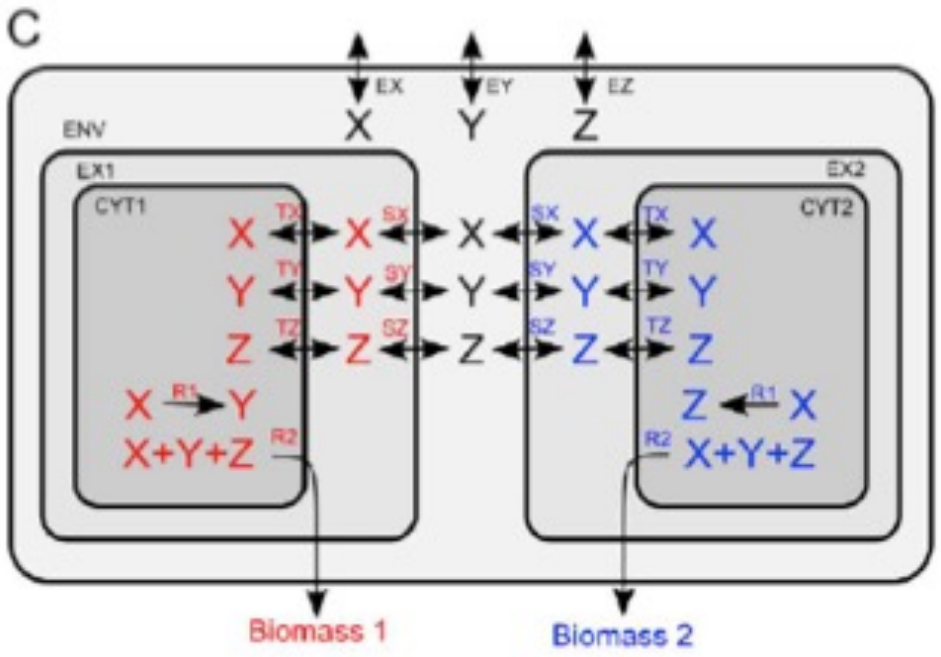
Not fully immediate already integrating metabolomics and transcriptomics

More directly related to interactions – Minimal stoichiometric precursor sets

Metabolic dialog



Environment could also be other species



see *Acuña et al. Bioinformatics 2012;*
Andrade et al., AMB 2016;

More directly related to interactions – Minimal stoichiometric precursor sets

Metabolic dialog

Input:

Metabolic network in the form of a directed hypergraph

A target or a set of targets

Output:

All minimal subsets of the sources of the network that enable to reach the target(s)

More directly related to interactions – Minimal stoichiometric precursor sets

Metabolic dialog

Input:

Metabolic network in the form of a directed hypergraph

A target or a set of targets

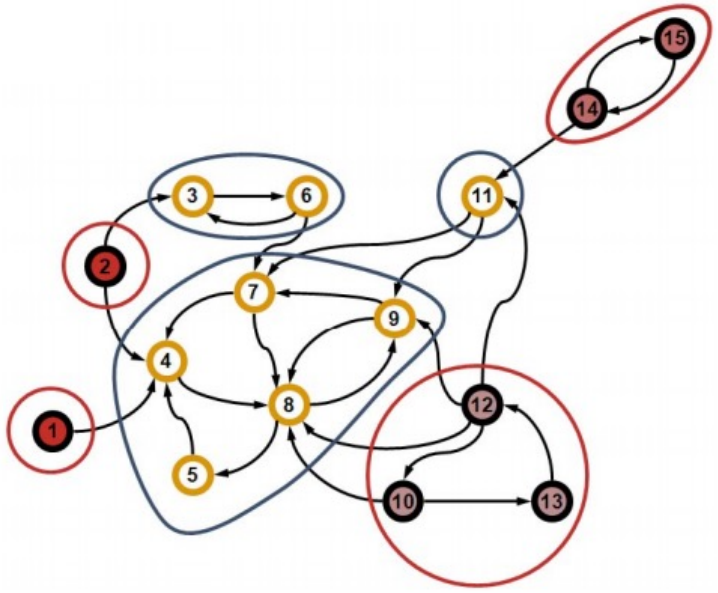
Output:

All minimal subsets of the sources of the network that enable to reach the target(s)

Sources of a metabolic network: strongly connected components at boundary

Can be computed in almost linear time up to a factor $\alpha(n)$
where α is the inverse of Ackermann function ($A(n,n)$)
and n is the number of vertices
Ackermann function

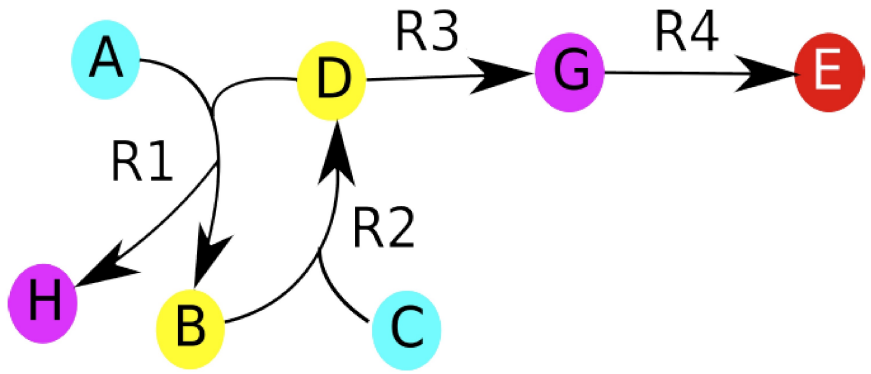
$$A(x, y) \equiv \begin{cases} y + 1 & \text{if } x = 0 \\ A(x - 1, 1) & \text{if } y = 0 \\ A(x - 1, A(x, y - 1)) & \text{otherwise} \end{cases}$$



Initial mathematical formalism represented main difficulty

Hypercycles

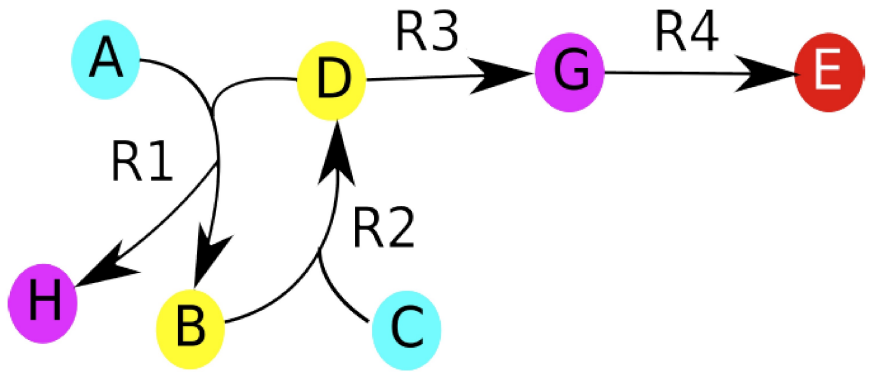
Blue: Sources
Red: Target

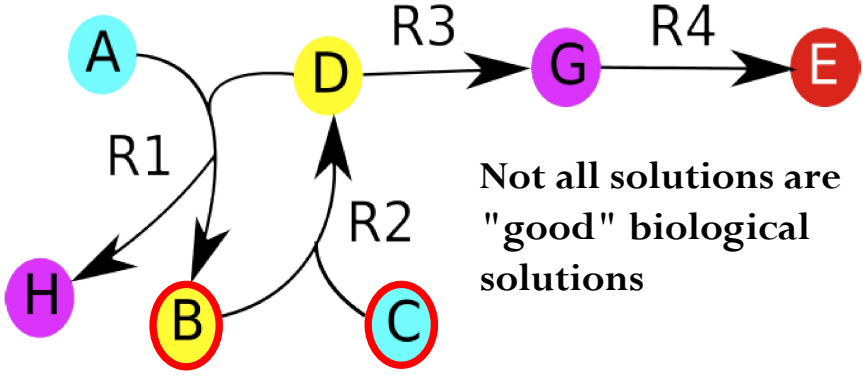


Initial mathematical formalism represented main difficulty

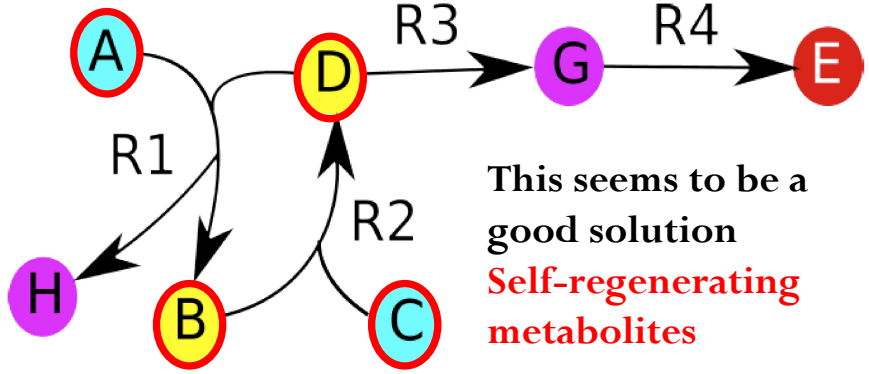
Hypercycles

Blue: Sources
Red: Target





Not all solutions are "good" biological solutions

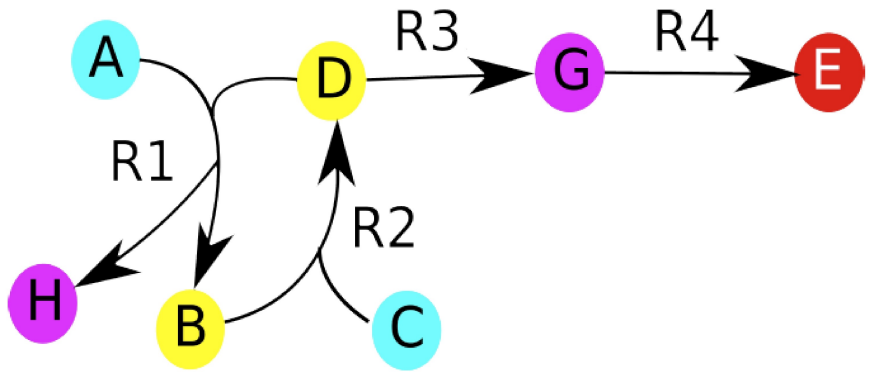


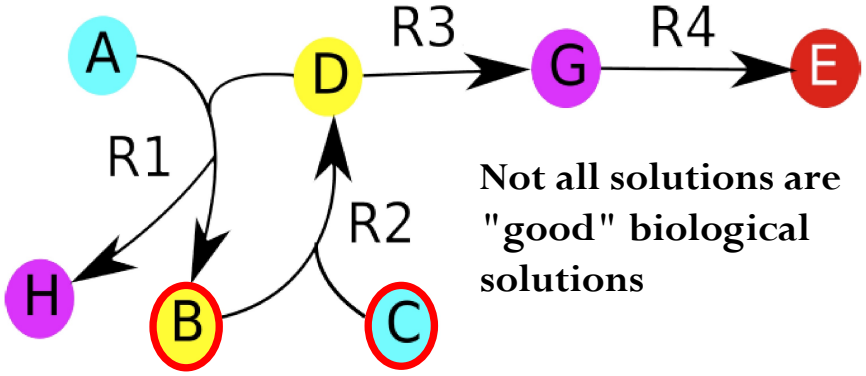
This seems to be a good solution
Self-regenerating metabolites

Initial mathematical formalism represented main difficulty

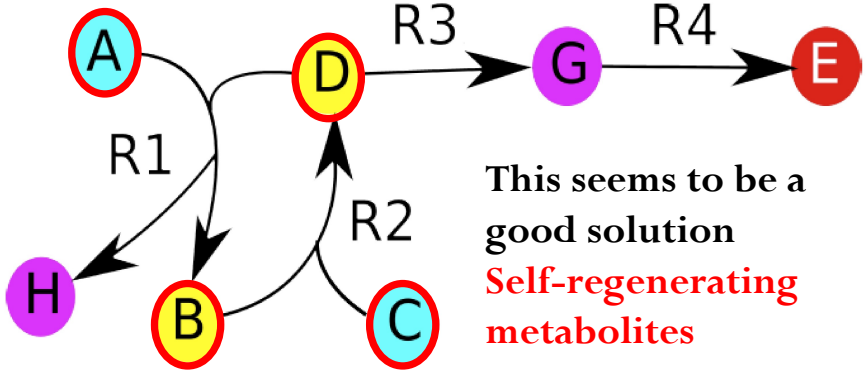
Hypercycles

Blue: Sources
Red: Target

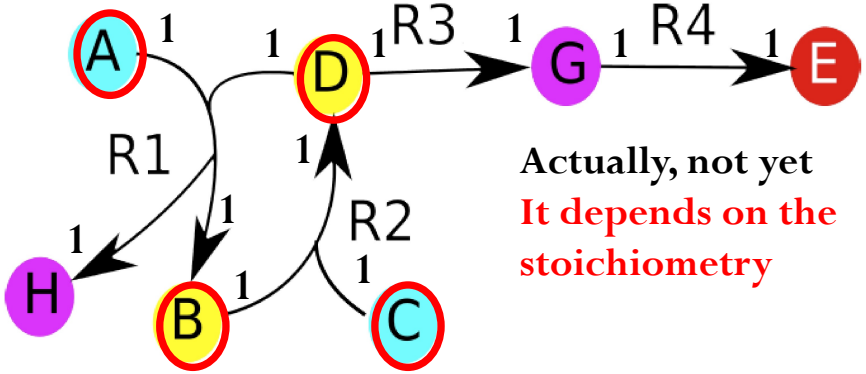




Not all solutions are "good" biological solutions



This seems to be a good solution
Self-regenerating metabolites

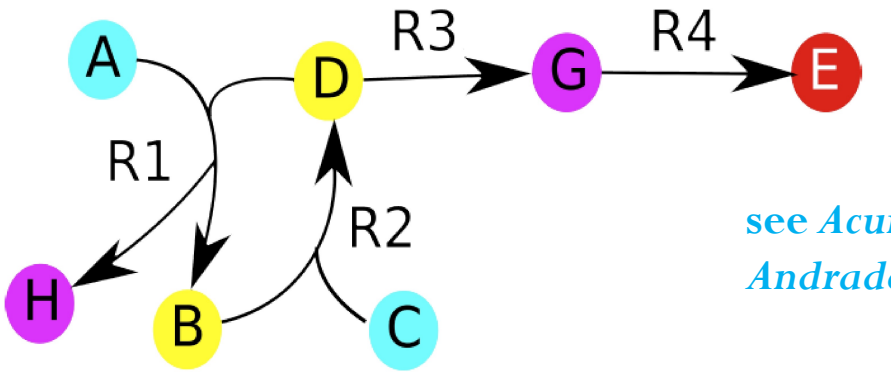


Actually, not yet
It depends on the stoichiometry

Initial mathematical formalism represented main difficulty

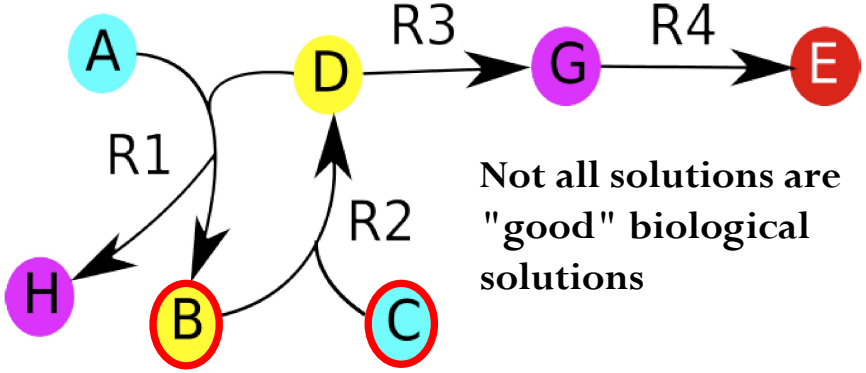
Hypercycles

Blue: Sources
Red: Target

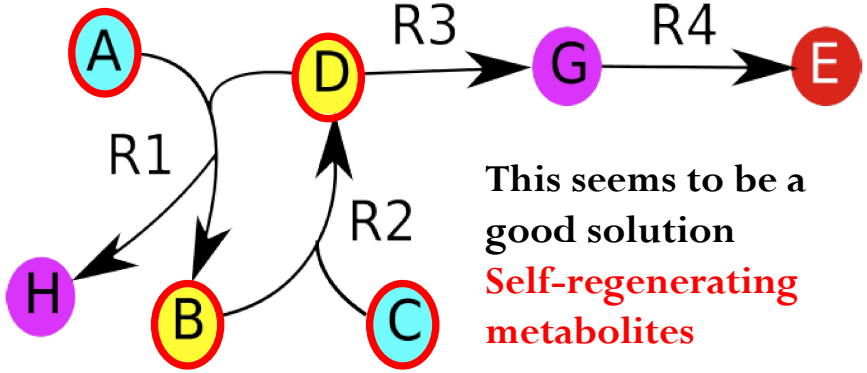


see Acuña et al. Bioinformatics 2012; Andrade et al., AMB 2016

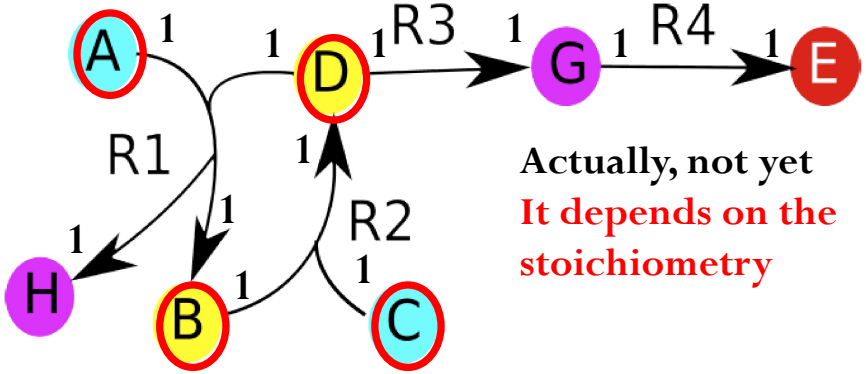
Sasita



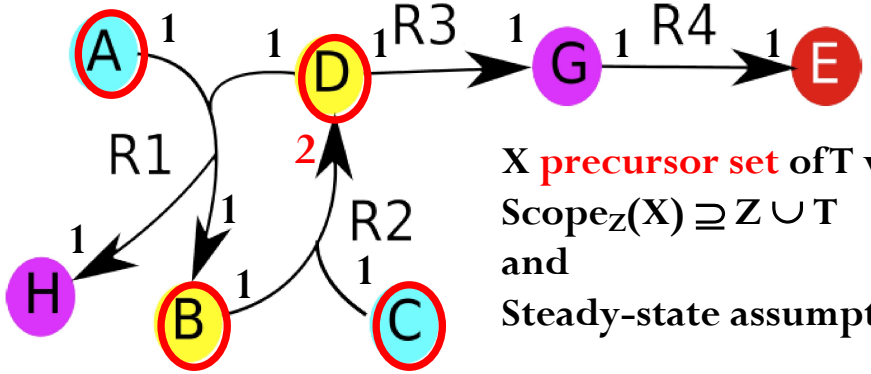
Not all solutions are "good" biological solutions



This seems to be a good solution
Self-regenerating metabolites



Actually, not yet
It depends on the stoichiometry



X precursor set of T with Z iff:
$Scope_Z(X) \supseteq Z \cup T$
and
Steady-state assumption fulfilled

This brings us back to precision, and also tango

Indeed, metabolism has since a long time been known to play an important role in the interaction (tango partnering) of an organism with its environment, which includes the other species present in it

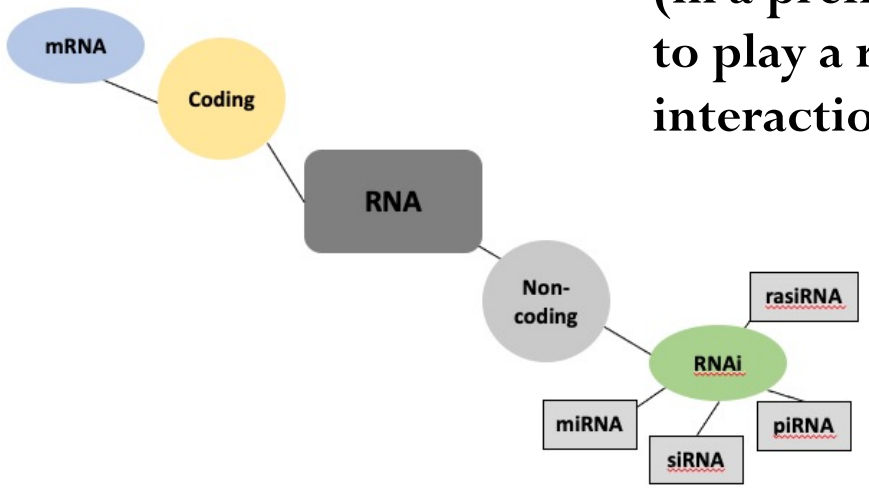
We've been exploring various other aspects of metabolism that I'll not have time to mention

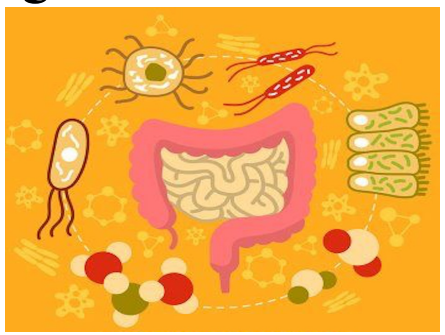
This brings us back to precision, and also tango

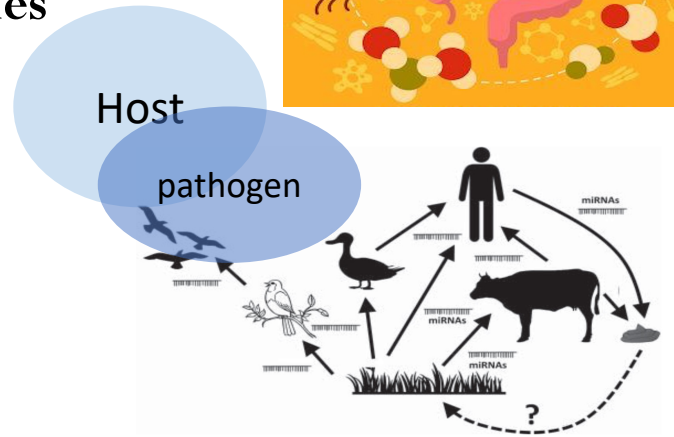
Indeed, metabolism has since a long time been known to play an important role in the interaction (tango partnering) of an organism with its environment, which includes the other species present in it
We've been exploring various other aspects of metabolism that I'll not have time to mention

There is also one other aspect that we have been investigating and which I have not mentioned and will not have time to mention, a very important one in my view as concerns interactions among species: this is the area of small non coding RNAs

This is an expanding world per se… it also has been shown (in a preliminary way) to play a role in species interactions



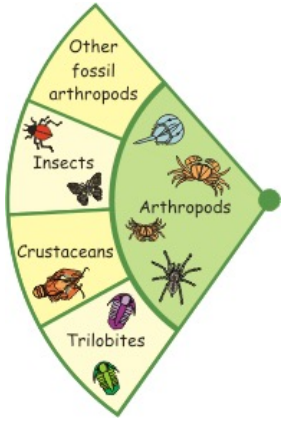




Something else I did not mention (actually, there are various others but no time…)

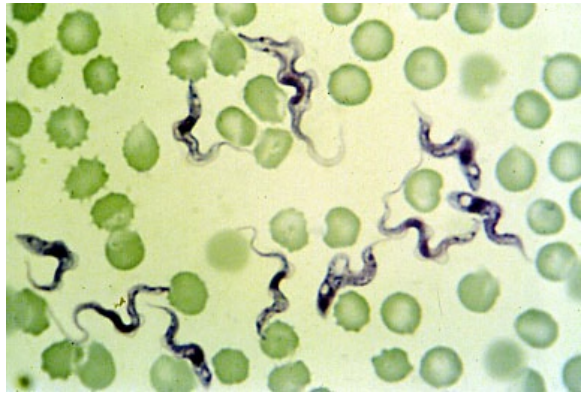
Different specific contexts where some of the aspects presented (cophylogeny/coevolution, genomics/transcriptomics, metabolism and regulation notably by small ncRNAs), sometimes more than one at the same time have been and still are explored

Arthropods (insects, arachnids, etc.) and their (colonies of) (endo)symbionts

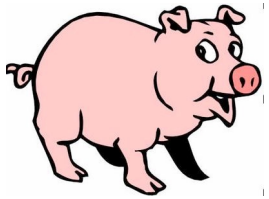


Gut microbiome

Trypanosomatids with/without endosymbiont and sometimes their hosts (3-level systems)

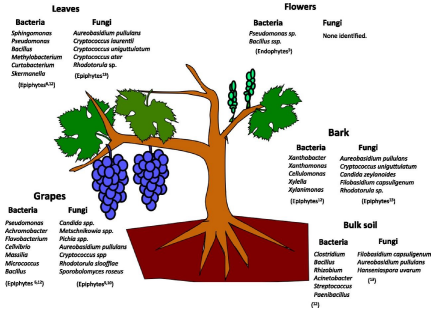


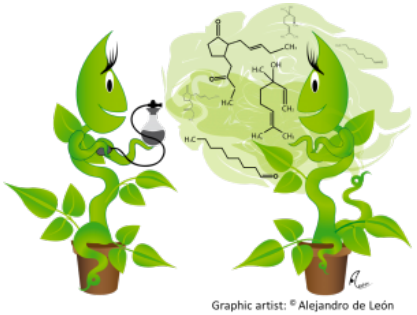
Swines and the bacterial colony in their respiratory tract



Mycoplasma hyopneumoniae attached to swine cilia. From the International Organization for Mycoplasmology.

Plants (vine and others) and their colonies of fungi and bacteria





Graphic artist: © Alejandro de León

Tango everywhere, and then a story that turned into an obsession

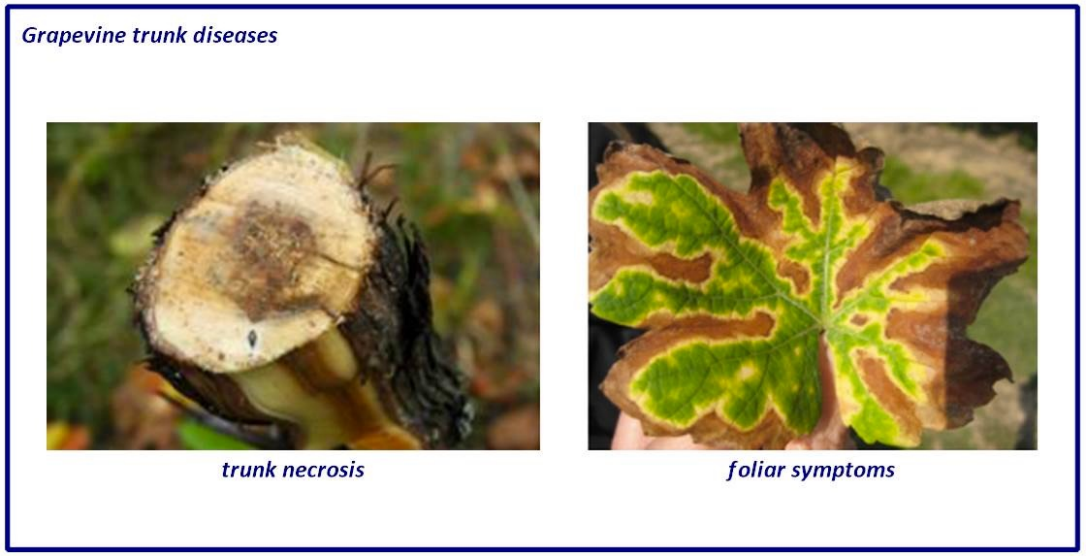
Two stories actually, both related to plant diseases caused by fungi or bacteria

This first is the Esca disease of the grapevine trunk
Little is known, apart from the fact that 3 fungi are possibly involved:

- ***Phaeomoniella chlamydospora***
- ***Phaeoacremonium aleophilum***
- ***Fomitiporia mediterranea***

Ricardo B. Ferreira

Plus others (*Epicoccum*) that may act as antagonists!

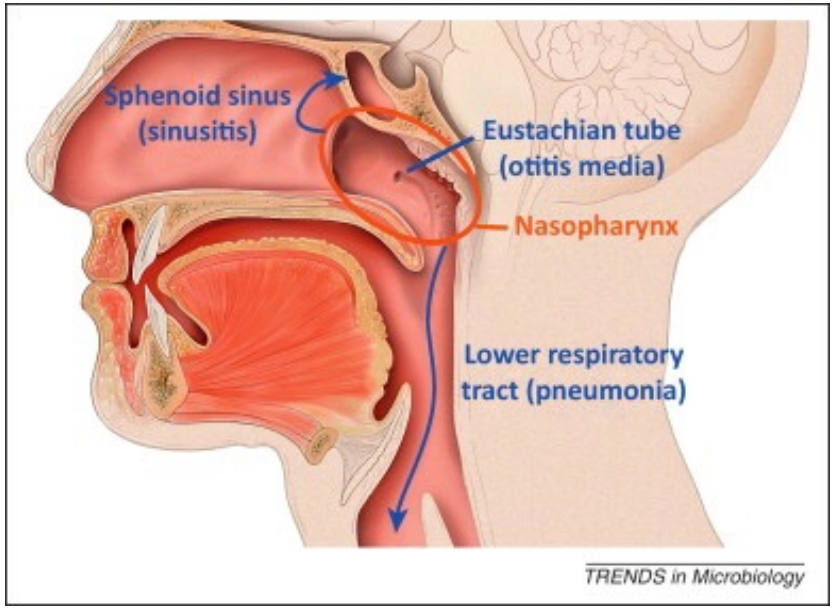


The second is Pierce's disease involving the bacterium *Xylella fastidiosa*

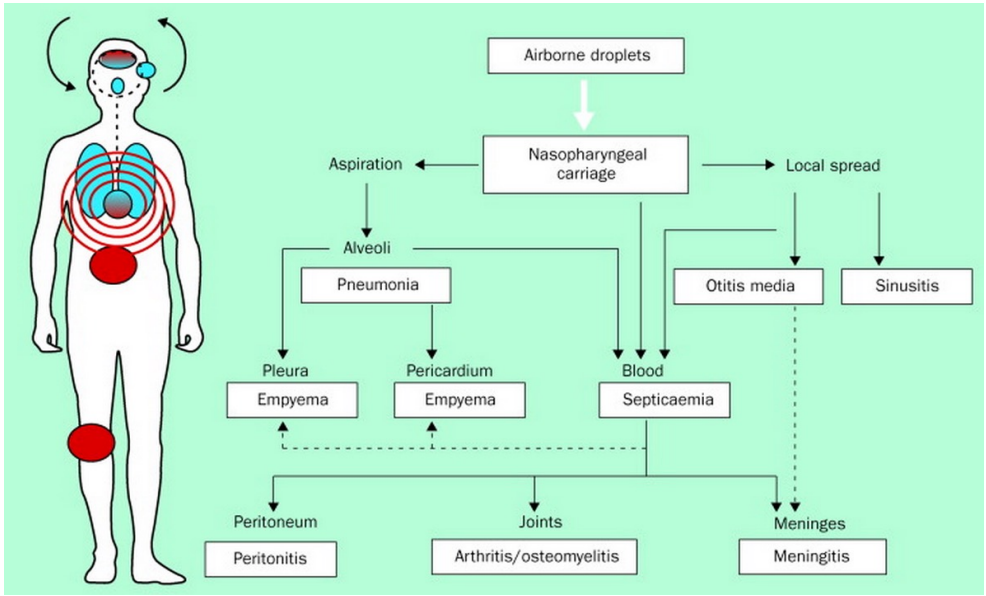
Lars H. Hansen

Actually the obsession started earlier

The initial story that was told to me by a collaborator in Portugal / Denmark, Ana Rute Neves, and started haunting me was related to *Streptococcus pneumoniae*, a bacterium that is commensal in the nasopharynx and pathogen in the lungs



Ana Rute Neves



How these two problems are usually treated

Usual treatment for plant diseases: Insecticides

Usual treatment for pneumoniae: Antibiotics

WHO

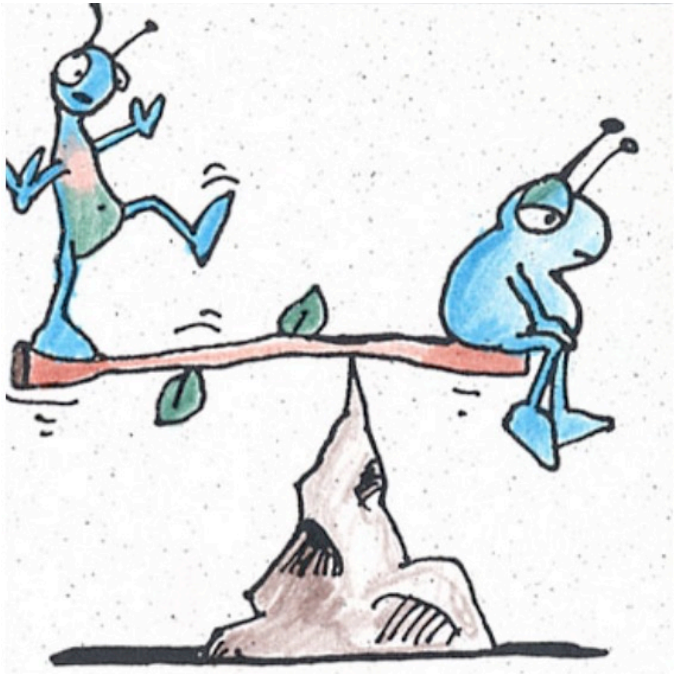
Antimicrobial resistance is not a new problem but one that is becoming more dangerous; urgent and consolidated efforts are needed to avoid regressing to the pre-antibiotic era.

For World Health Day 2011, WHO introduced a six-point policy package to combat the spread of antimicrobial resistance.

What then became the obsession? It takes two to tango

Informally, the expression "it takes two to tango" means that the two persons involved in a dance are responsible for it

Which may be extended to any other type of situation, good or bad, with two or more partners, any type of partners

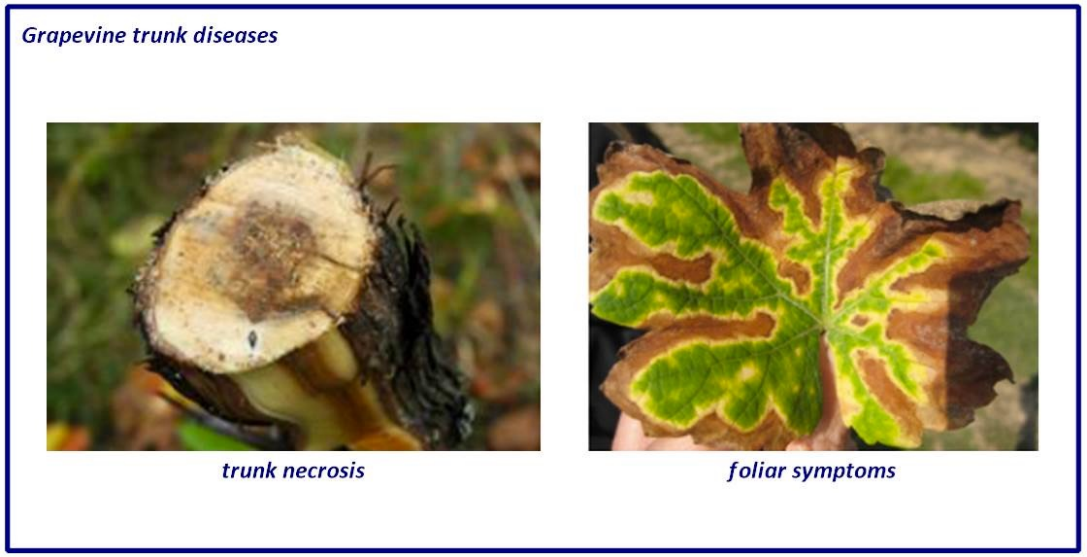


What then became the obsession? It takes two to tango

Informally, the expression "it takes two to tango" means that the two persons involved in a dance are responsible for it

Which may be extended to any other type of situation, good or bad, with two or more partners, any type of partners

Obsession: "Non-aggressive" interventions



Is this crazy?

Perhaps, most probably, no doubt

Certainly there is still a lot to be done to show

- **Whether this is possible even at a very small scale – for instance, Esca disease**
- **And if it is possible, how far / fast can we then extend this type of approach?**

Certainly, in order to do this, three things are crucial

- **Not look at only one aspect of biology – Tango of different areas within a discipline**
- **Not look only at biology but also, *e.g.*, ecology – Tango of different disciplines within life sciences**
- **And then also Tango with other disciplines outside life sciences**

Which means, it is absolutely necessary to tango with other researchers

Erable Team – Current members

Alberto Marchetti-Spaccamela
Arnaud Mary
Blerina Sinaimeri
Giuseppe Italiano
Leen Stougie
Luca P. Sciarria
Mariana Ferrarini
Marianne Borderes
Nicolas Homberg
Roberto Grossi
Scheila Mucha
Vincent Lacroix
Yishu Wang

Plus others!

Erable Team – Ex-members

Alex di Genova
Alice Julien-Laferrière
André Veríssimo
Augusto Vellozo
Beatrice Donati
Carol Moraga
Cecilia Klein
Christian Baudet
Christian Gautier
Delphine Parrot
Emmanuel Prestat
Gustavo Sacomoto
Irene Ziska
Laura Urbini
Laurent Bulteau
Leandro Lima
Lilia Boucinha
Martin Wannagat
Pierluigi Crescenzi
Ricardo Andrade
Paulo Milreu
Taneli Pusa
Vicente Acuña

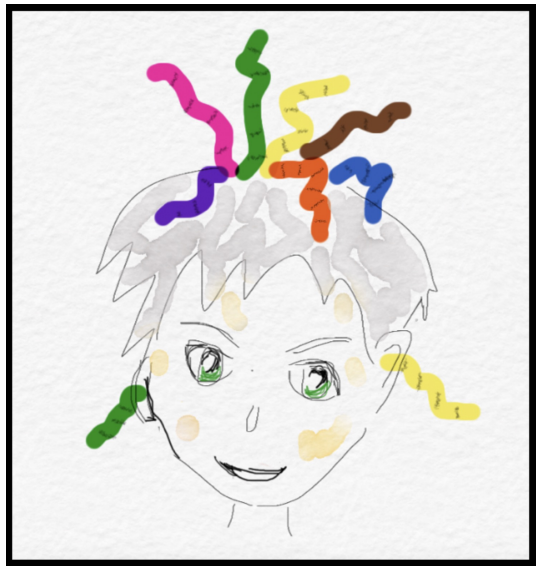
Collaborators

Alexandra Carvalho
Ana Tereza Vasconcelos
Andréa Ávila
Ariel Silber
Arnaldo Zaha
Catherine Matias
Christine Gaspin
Claudia N. Santos
Elena Vidal
Franciele Siqueira
Helisson Faoro
Henrique B. Ferreira
Lars H. Hansen
Katharina Huber
Nuno Mira
Ricardo B. Ferreira
Romeo Rizzi
Susana Vinga
Tiziana Calamoneri
Vincent Moulton

Plus others!

Other members of the LBBE - Lyon
Chr. Hansen / MaatPharma

Thanks!



Funding over the years

ANR

erc
European Research Council
Established by the European Commission

HORIZON 2020

cnrs

UNIVERSITÉ DE LYON

Lyon 1

Inria