Phenylpropanoids pathway:
Bioinformatics and modelling approaches to increase the production of Metabolites of interest
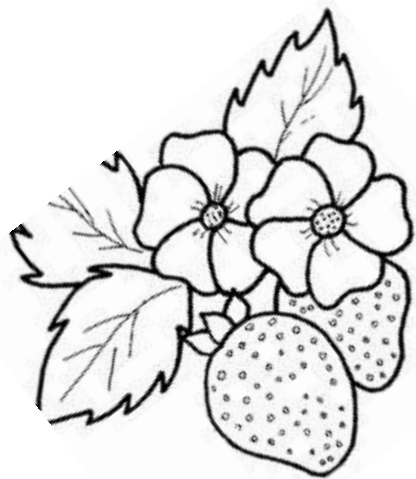
D. PARROT, A. JULIEN-LAFERRIERE, L. BULTEAU, R. ANDRADE,
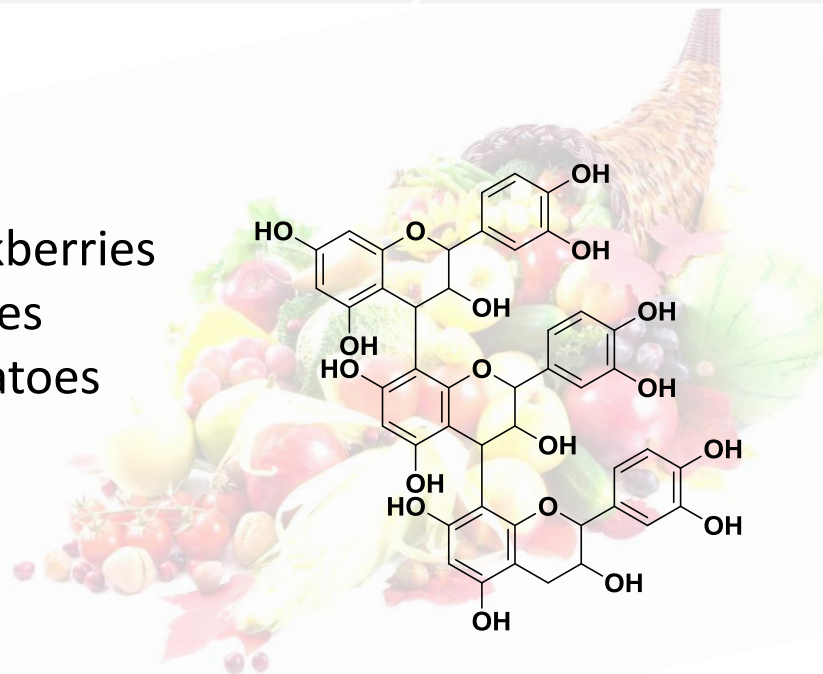R.S. COSTA, A. FERNANDES, A.A. HARTMANN, A. VERISSIMO,
S. VINGA, M-F. SAGOT

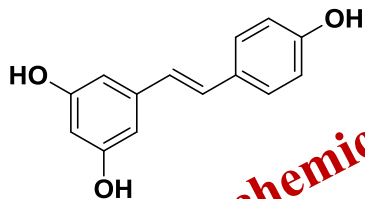*– AFERP/STOLON, Lyon, 16th July 2015 –*

•Several natural sources

- Cranberries
- Apples
- Strawberries

- Blackberries
- Grapes
- Tomatoes
Etc.

•Chemical diversity

- Anthocyanins
- Stilbenes
- Flavonols
- Phenolic acids
- Coumarines
 Etc.

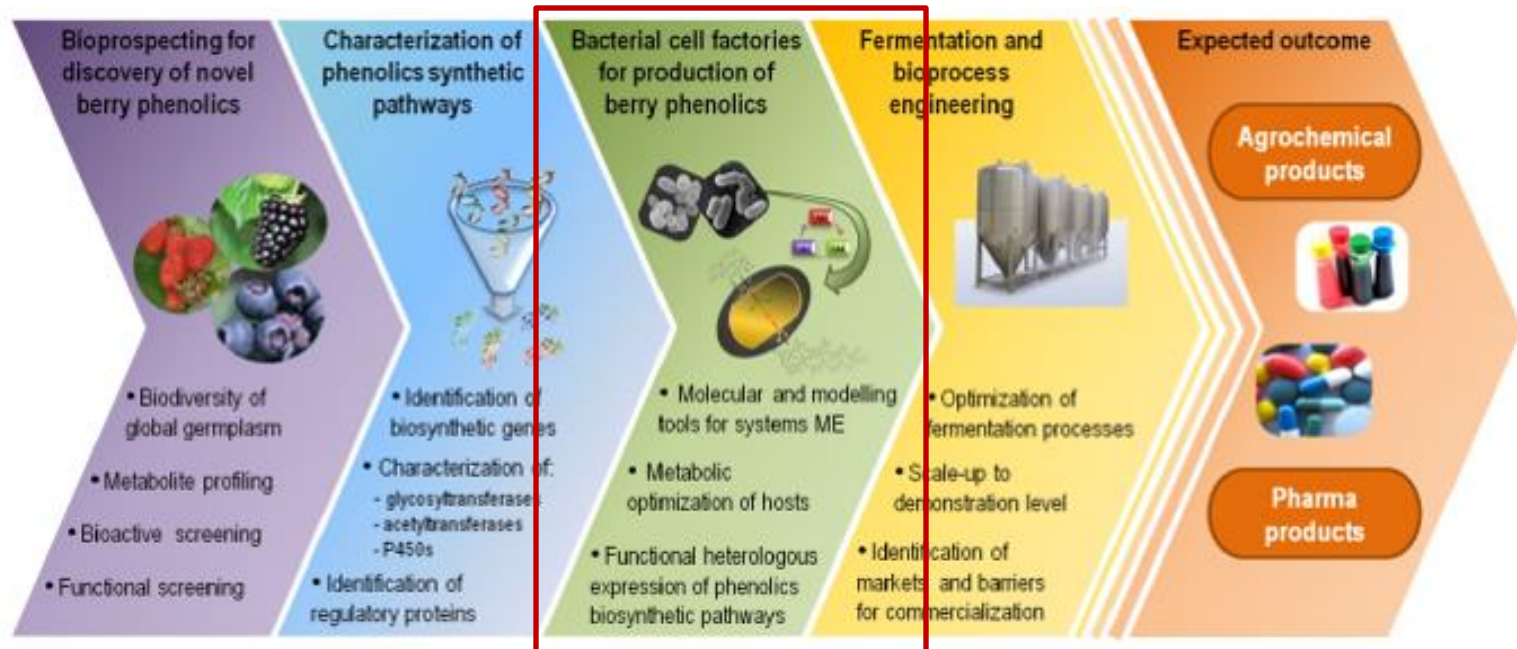> 8000 chemical structures identified

• Multiple properties

- Plants/environment relations
- Antioxidant
- Health-promoting
- Colours
Etc.

BacHBerrY Project: BACterial Host for production of Bioactive phenolics from bERRY fruits
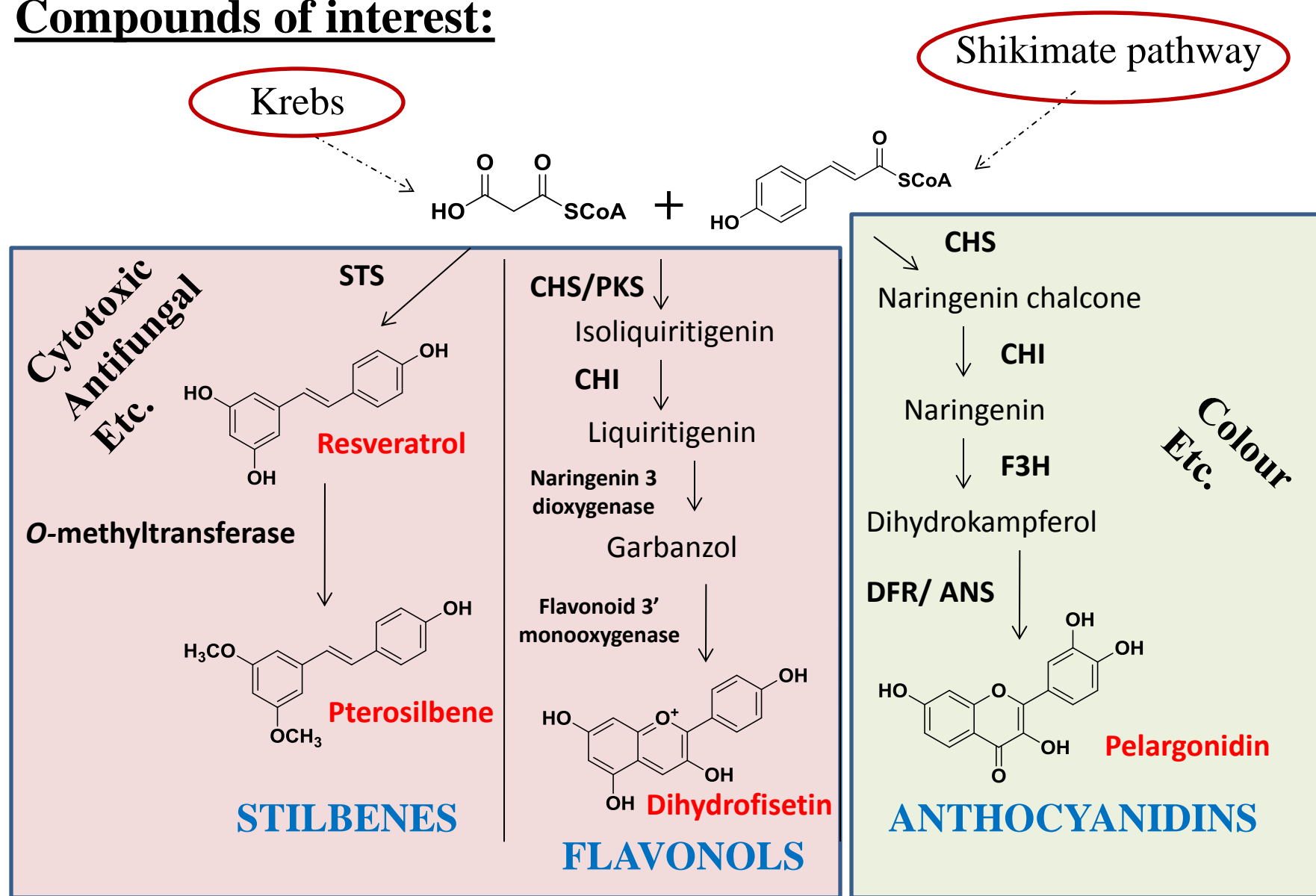
Chemistry of berries ⟶ Valorization

(http://www.bachberry.eu/)

GOAL

➜ Improve and optimize the production of metabolites of interest using bacteria

Compounds of interest:

Krebs

Shikimate pathway

Cytotoxic Antifungal Etc.

STS

Resveratrol

O-methyltransferase

Pterosilbene

STILBENES

CHS/PKS

Isoliquiritigenin

CHI

Liquiritigenin

Naringenin 3 dioxygenase

Garbanzol

Flavonoid 3' monooxygenase

Dihydrofisetin

FLAVONOLS

CHS

Naringenin chalcone

CHI

Naringenin

F3H

Dihydrokampferol

DFR/ ANS

Pelargonidin

Colour Etc.

ANTHOCYANIDINS
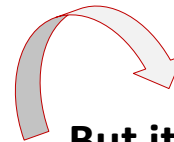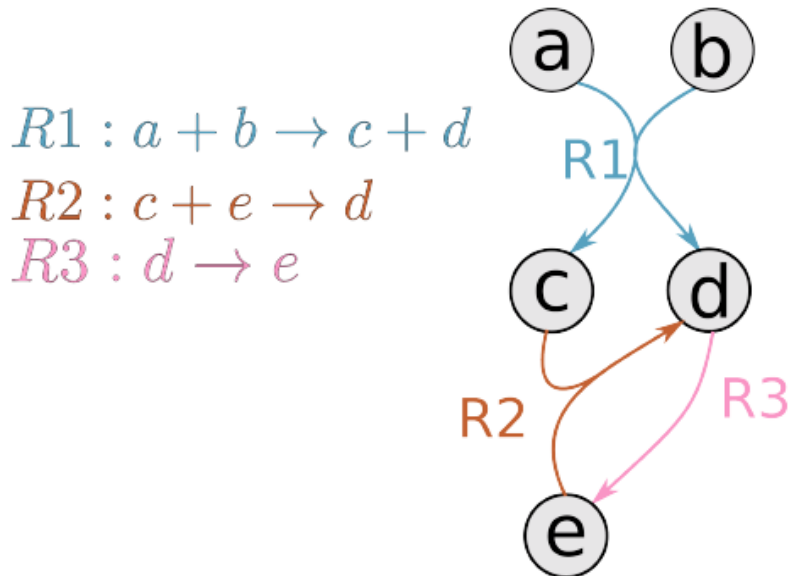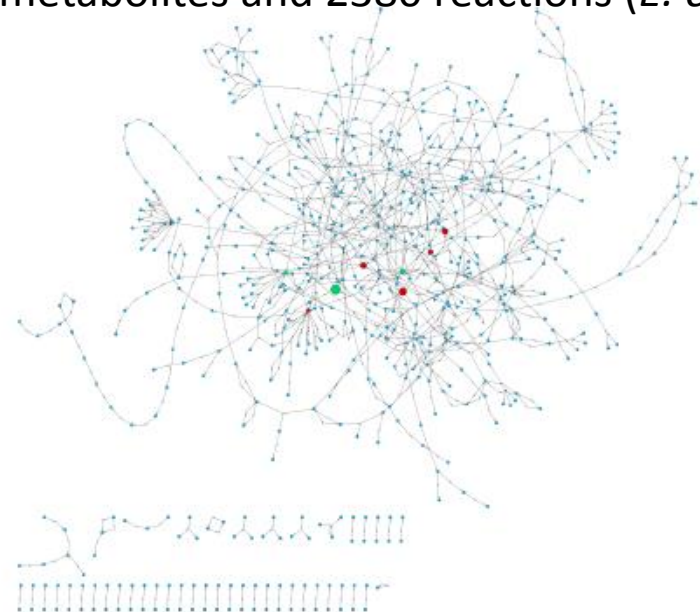
GOAL ➔ Improve and optimize the production of metabolites of interest using bacteria

Metabolic engineering:

But it's not easy …

- 600 metabolites and 573 reactions (*S. cerevisiae*)
- 1667 metabolites and 2380 reactions (*E. coli*)

$R1 : a + b \rightarrow c + d$
$R2 : c + e \rightarrow d$
$R3 : d \rightarrow e$

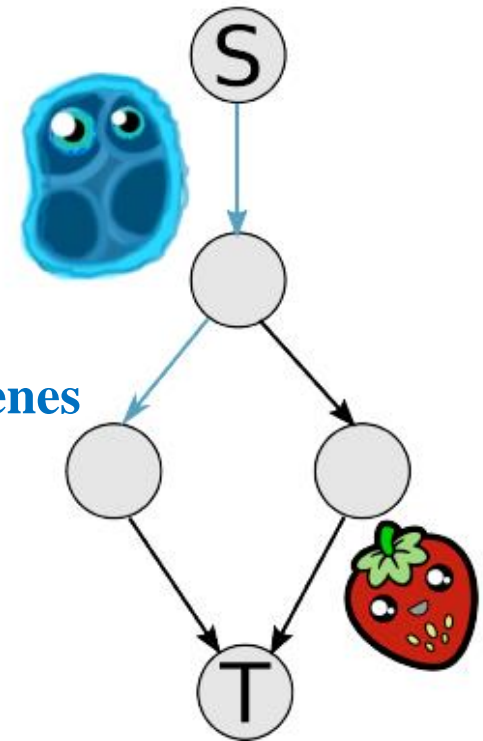Hypergraphs are used for modelling metabolic networks

GOAL

Selecting the optimal genes to introduce for the production of a target compound

Origin of the genes:

• The engineered species
  - *Lactococcus lactis*
  - *Coryneobacterium glutamicum*

Endogenous genes

• The organism of origin
• Other organisms

Exogenous genes

Nodes = metabolites

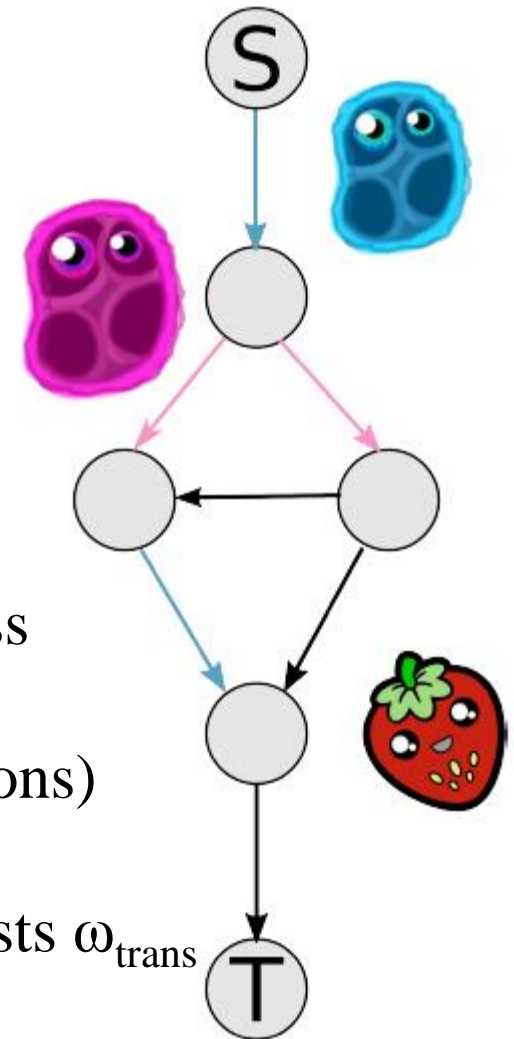arcs = enzymatic reactions

MAIN IDEAS

- Combine all metabolic network by **putting together all possible reactions available** (e.g: in MetaCyc)

- Introduction & heterologous expression of an exogenous gene **has a cost**

- Expression of an endogenous gene is **free**

- If there are several organisms, going from one to another in the metabolic network has a **cost (use of transporters)**

Bacteria 1

A
B → C ····> C ····> C → D

Bacteria 2

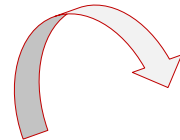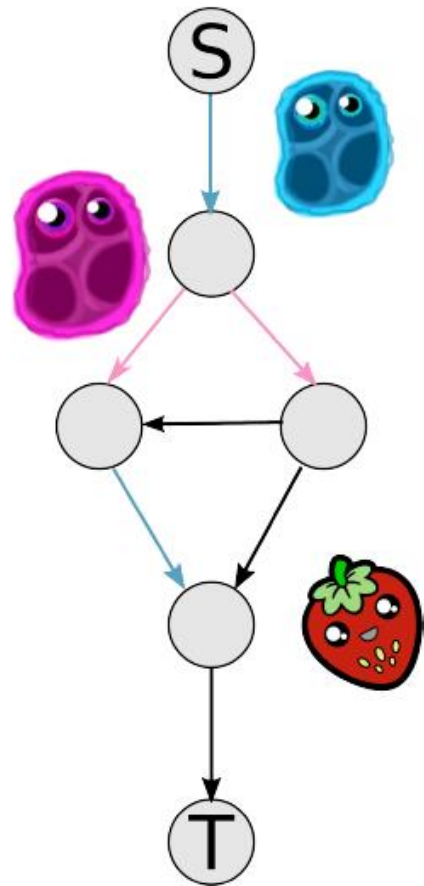*Transporters*
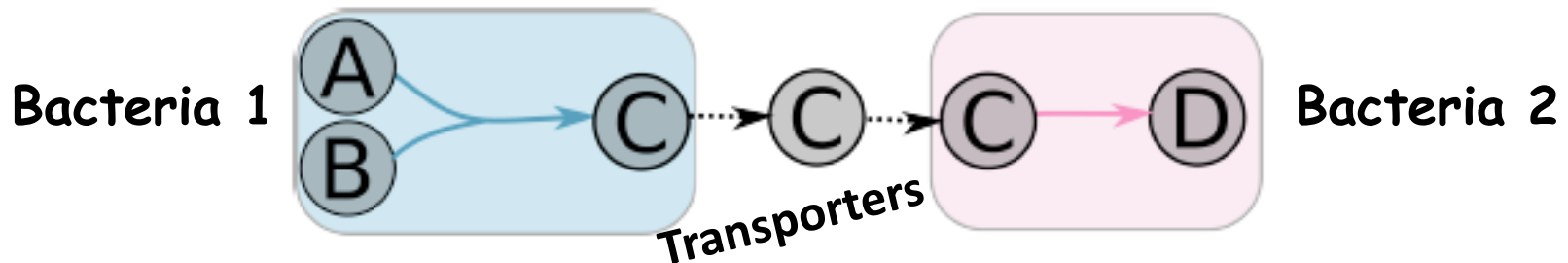
Production T (Target) from S (Substrat)

Conditions:

- All possible ways to produce T from S
-A succession of arcs of a same colour ➔ cost less

- Minimize the number of uncoloured arc (insertions)
➔ costs ω_u
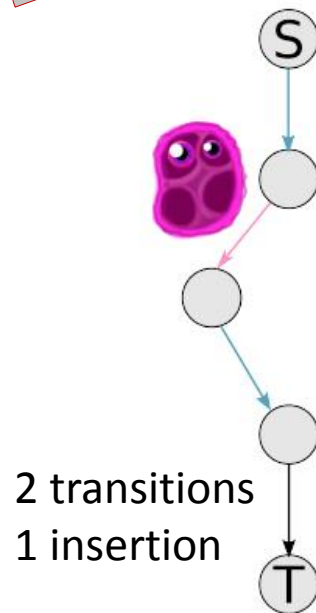- Minimize the number of colour transition ➔ costs ω_{trans}

⬤ Nodes = metabolites

⟶ arcs = enzymatic reactions

3 possible pathways to produce T

1 transition
2 insertions
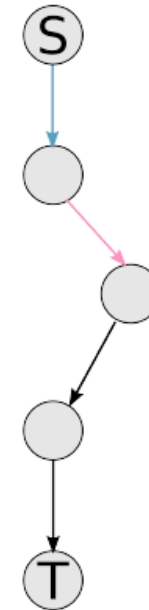
2 transitions
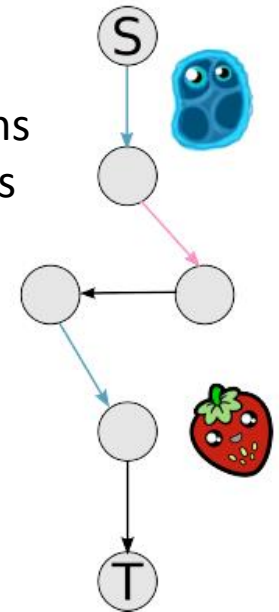2 insertions

2 transitions
1 insertion

(a) Cost : $\omega_u + 2 * \omega_{trans}$
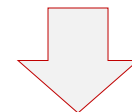
(b) Cost : $2 * \omega_u + \omega_{trans}$

(c) Cost : $2 * \omega_u + 2 * \omega_{trans}$
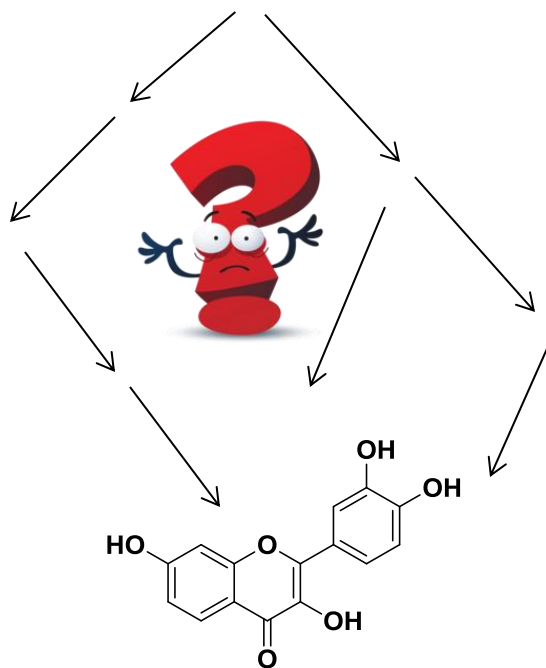
Nodes = metabolites

⟶ arcs = enzymatic reactions

Most « expensive » option

CONCLUSION

Obtain **an algorithm** that allows to select the optimal genes to introduce for the production of a target compound

Sources

Compound of interest

Origin of the genes:
- Exogenous genes
- Endogenous genes

Criteria ?
- Enzymatic activities
- Thermodynamic limits
- Stoechiometry
- Toxicity

. . .

➔ **But … still theorical work to do**

DINGHY: Dynamic Interactive Navigator for General Hypergraphs in biologY

GOAL

☐Rapid display of small metabolic networks
☐Takes into account meta data:
→ Metabolic pathways, compartments, cofactors, reaction "value", etc.

☐Dynamic placement: the user can interact with the layout
→ response to node drag and drop
→ gravity, magnetism, springs

☐Smart drawing of hyper arcs
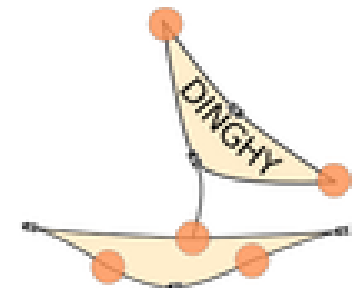
Contributors:
Laurent BULTEAU
Alice JULIEN-LAFERRIERE
Delphine PARROT

Pipeline

Whole network (.SBML)

Reactions of interest (.txt)

```
1   <?xml version="1.0" encoding="UTF-8"?>
2   <sbml  level="2" version="1">
3     <model id="Ec_iAF1260_flux2_xml">
4       <listOfUnitDefinitions>
```

```
R_23CAMPtex -2.03371-
R_23CUMPtex -1.54514*
R_23PDE2pp 1.54514*
R_23PDE7pp 2.03371*
```

Cofactors (.txt)

Nodes of interest (.bn)

```
M_atp_c
M_adp_c
M_nadh_c
M_nad_c
M_h_c
M_h_e
M_nadph_c
M_nadh_c
M_h2o_c
```

Pre-processing (Python)

Pyruvate
Fumarate
Cytosine
…

Summary (.json)

Display (html + javascript)
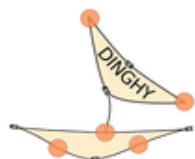
DINGHY: Dynamic Interactive Navigator for General Hypergraphs in Biology

| HOME | DOCUMENTATIONS | UPLOAD | CONTACT |

Settings

Drawing elements

6	Arc Unit Thickness
1	Arc Min Thickness
20	Arc Max Thickness
1	Arc Curve Scale
1.4	Node Scale
1	Separate Components

Two input formats are possible:

Biological data

Two data loading modes are available: either by loading a file or writing directly :
be download here ExtractJSON script

2 files are required in addition to the extractJSON script, and 2 more are optional:

- System SBML (.xml) ⟵
- List of reactions of interest (.txt) ⟵ } Files required
- List of highlighted nodes (.bn)
- List of cofactors (.txt)

Text informations

| 0 | Trim Id Right |
| 2 | Trim Id Left |

☑ Nodes not Overlaps
☐ Edges not overlaps

Positioning

1	Rounds
0.7	Coeff Gravity
1	Coeff Charge
5	Coeff Arc Distance
1	Coeff Reaction Repulsion

Generic hypergraph

Example:
```
a + b -> c
b + c -> d + e
```

Submit

Tools

Visualization window

DEMO
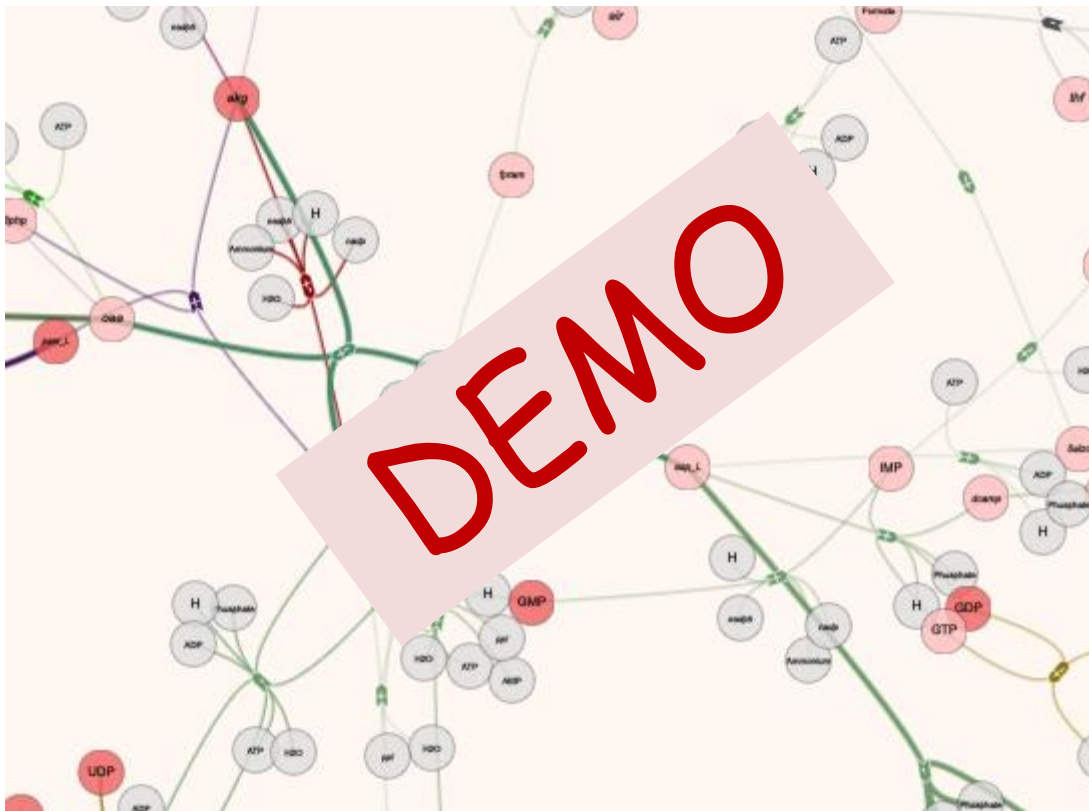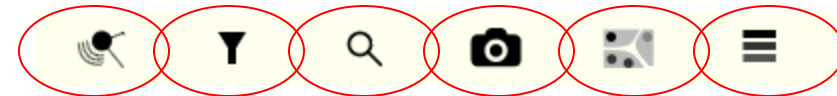
1 – Node dynamics

2 – Filters

3 – Research:
 - Node
 - Paths

4 – Snapshots

5 – Compartment separation

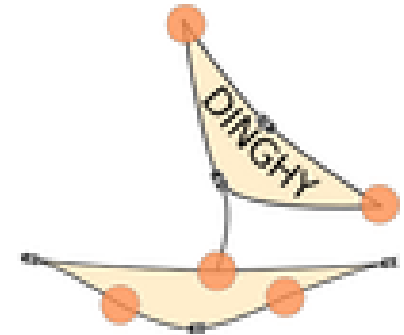6 – Metabolic pathways

http://dinghy.gforge.inria.fr/
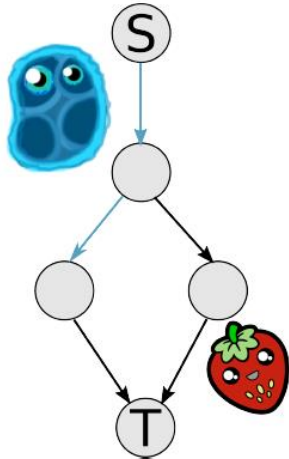
Conclusion

Rainbow steiner problem …

Under work

BacHBerrY Project

Always in improvement but already available

http://dinghy.gforge.inria.fr/

Perspectives

➔ Propose an algorithm to facilite and increase the production of molecules of interest using bacteria
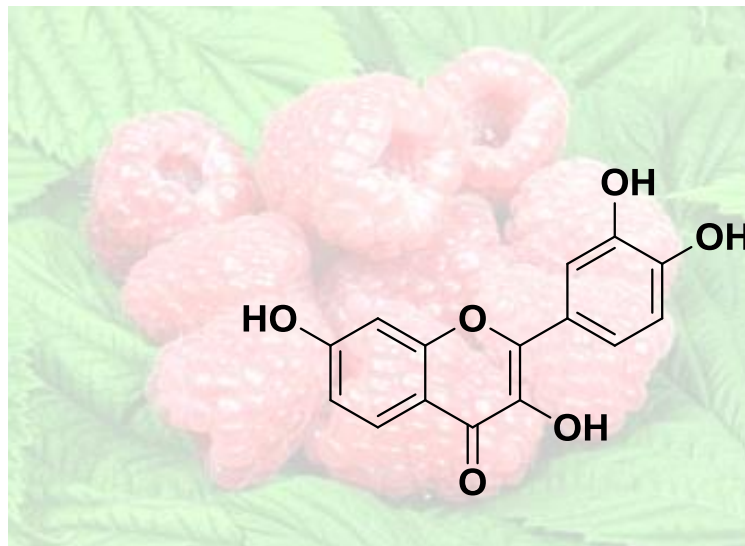
Acknowledgments:

Phenylpropanoids pathway:
Bioinformatics and modelling approaches to increase the production of Metabolites of interest

D. PARROT, A. JULIEN-LAFERRIERE, L. BULTEAU, R. ANDRADE,
R.S. COSTA, A. FERNANDES, A.A. HARTMANN, A. VERISSIMO,
S. VINGA, M-F. SAGOT

– AFERP/STOLON, Lyon, 16th July 2015 –