Comparing Metabolic Pathways through Potential Fluxes: a Selectively Open Approach

Marta Simeoni

Università Ca' Foscari Venezia

Workshop BioPPN 2013 - Milan, June 24, 2013

Joint work with
Paolo Baldan (Univ. Padova), Martina Bocci (Univ. Venezia) and
Nicoletta Cocco (Univ. Venezia)

Framework

Comparison of metabolic pathways of different species may be useful for

- understanding metabolic functions
- giving interesting information on their evolution

Framework

Comparison of metabolic pathways of different species may be useful for

- understanding metabolic functions
- giving interesting information on their evolution

CoMeta is a tool for comparing metabolic pathways of different organisms:

- KEGG used as a source of metabolic data
- metabolic pathways represented as Petri nets
- Petri net properties employed for the comparison

Metabolic pathways (MPs)

Metabolism: the chemical system which generates the essential components for life

Metabolic pathways:

- subsystems dealing with some specific function
- represented as a network of chemical *reactions* catalised by one or more *enzymes* where some molecules (*reactants or substrates*) are transformed into others (*products*)
- the *stoichiometric matrix* identifies the pathways components and their relations
- kinetics represented by the *rate equation* associated with each reaction

Representing MPs with Petri nets (PNs)

Metabolic pathways can be naturally modelled with PNs:

- Places are associated to molecular species (metabolites, compounds, enzymes)
- Transitions correspond to chemical reactions
  - Input places are substrates
  - Output places are products
- The incidence matrix of the PN is identical to the stoichiometric matrix of the system of chemical reactions
- The number of tokens in each place of the PN indicates the amount of substance associated with that place

Comparison in CoMeta

Comparison technique for MPs based on

- static aspects: by considering homology of enzymes/ reactions
- behavioural aspects: by considering a measure of the similarity of the potential fluxes in the pathways

Comparison in CoMeta

Comparison technique for MPs based on

- static aspects: by considering homology of enzymes/ reactions
- behavioural aspects: by considering a measure of the similarity of the potential fluxes in the pathways $\Rightarrow$ T-invariants

Comparison in CoMeta

Comparison technique for MPs based on

- static aspects: by considering homology of enzymes/ reactions
- behavioural aspects: by considering a measure of the similarity of the potential fluxes in the pathways $\Rightarrow$ T-invariants

Combined distance:

$$d_D(P_1, P_2) = \alpha \; d_R(P_1, P_2) + (1 - \alpha) \; d_I(P_1, P_2)$$

The weight $\alpha \in [0, 1]$ allow the analyst to move the focus between static ($\alpha = 1$) and behavioural ($\alpha = 0$) aspects.

Using T-invariants in the comparison

Why

- Minimal (semi-positive) T-invariants correspond to elementary flux modes of a metabolic pathway, i.e. minimal sets of reactions that can operate at a steady state

How

- The set of semi-positive T-invariants has a unique basis, the Hilbert basis, consisting of the minimal T-invariants $\Rightarrow$ characteristic of the net
- The invariant based distance is obtained by comparing the Hilbert bases of two pathways

The tool CoMeta

Its main features are:

- download of the information on the specified organisms and pathways from KEGG
- translate the MPs into corresponding PNs (MPath2PN)
- compute the combined distance for each pair of organisms and build the corresponding distance matrix (4ti2)
- build and display a phylogenetic tree (UPGMA or Neighbour Joining methods)

T-invariants in subnets

PNs corresponding to the metabolic pathways of an organism are subnets of a larger net representing its full metabolic network.

They can be considered as:

- *isolated* subnets $\Rightarrow$ interactions with the environment are ignored;
- *open* subnets $\Rightarrow$ input/output metabolites are open places where the environment can freely put/remove substances.

T-invariants in subnets

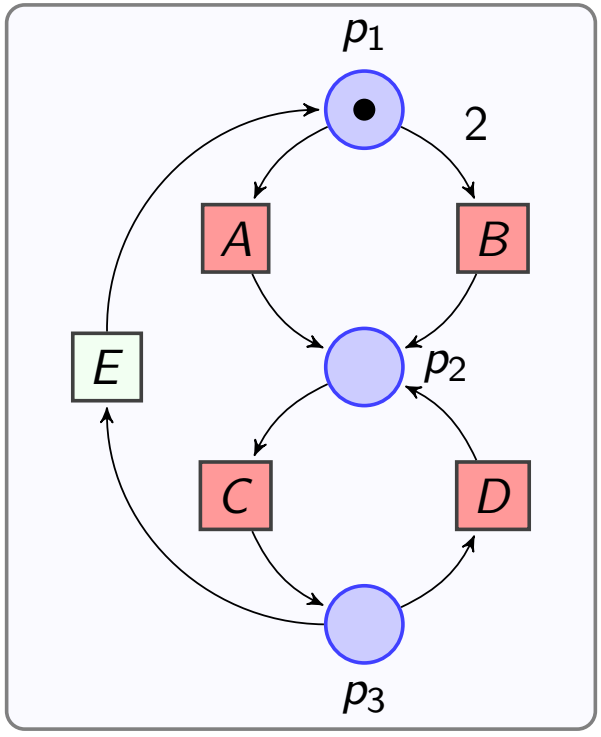
PNs corresponding to the metabolic pathways of an organism are subnets of a larger net representing its full metabolic network.

They can be considered as:

- *isolated* subnets $\Rightarrow$ interactions with the environment are ignored;
- *open* subnets $\Rightarrow$ input/output metabolites are open places where the environment can freely put/remove substances.

What happens to the minimal T-invariants of the subnets in the two cases?

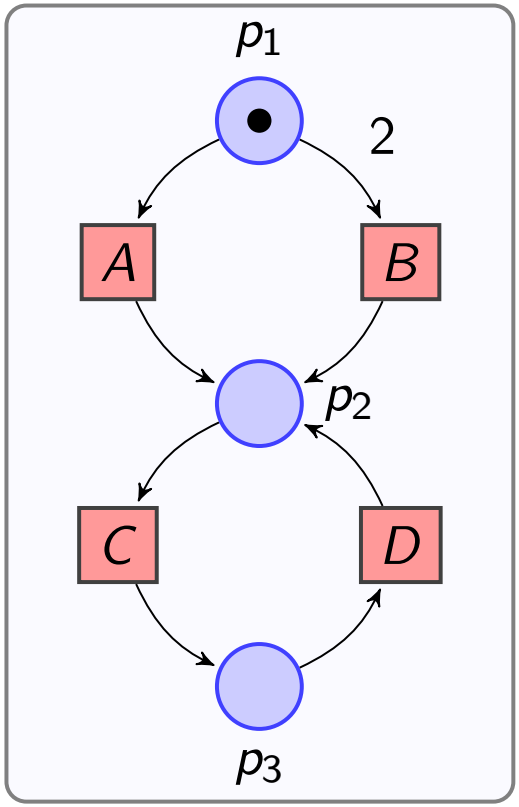
Example: a simple net



Minimal T-invariants:
$I_1 = \{A, C, E\}$, $I_2 = \{C, D\}$.

Note that $\{B, C, E\}$ is not an invariant, since B requires two tokens in p_1.

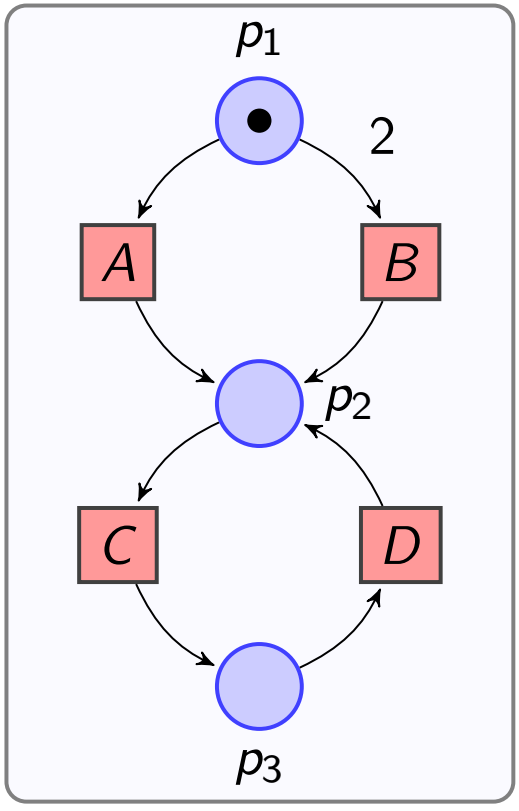
Example: isolated subnet



Minimal T-invariants:
$I_2 = \{C, D\}$

Invariant $I_1 = \{A, C, E\}$ is lost

Example: isolated subnet

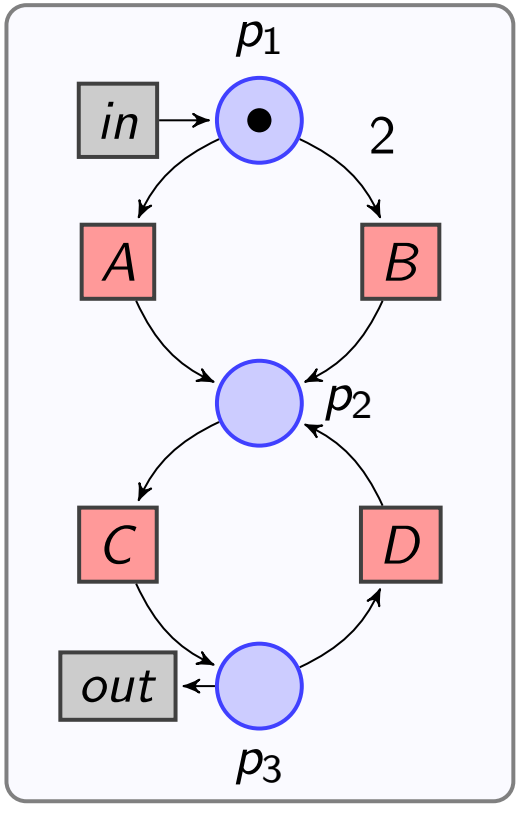


Minimal T-invariants:
$I_2 = \{C, D\}$

Invariant $I_1 = \{A, C, E\}$ is lost

Isolation guarantees correctness: minimal T-invariants of the subnet are minimal T-invariants of the full network

Example: open subnet

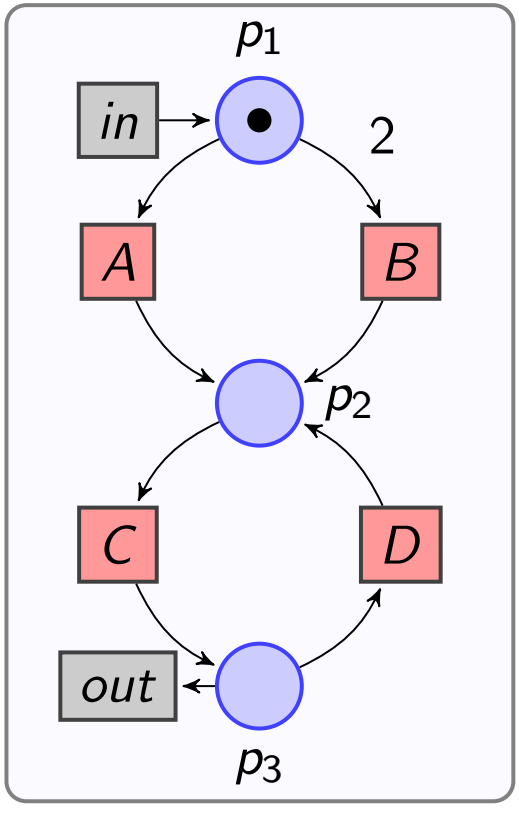


Minimal T-invariants:
$I_1' = \{in, A, C, out\}$, $I_2 = \{C, D\}$,
$I_3 = \{2 \cdot in, B, C, out\}$.

Invariant I_1' is the projection of $I_1 = \{A, C, E\}$ onto the subnet.

Invariant I_3 does not correspond to any invariant of the original net.

Example: open subnet



Minimal T-invariants:
$I_1' = \{in, A, C, out\}$, $I_2 = \{C, D\}$,
$I_3 = \{2 \cdot in, B, C, out\}$.

Invariant I_1' is the projection of $I_1 = \{A, C, E\}$ onto the subnet.

Invariant I_3 does not correspond to any invariant of the original net.

Opening the subnet guarantees completeness: any invariant of the full network, once projected onto the subnet, is an invariant of the open subnet

T-invariants in subnets: summary

Minimal T-invariants of the full network have clear relation with (minimal) T-invariants of a subnet:

- Isolation guarantees correctness: minimal T-invariants of the subnet are minimal T-invariants of the full network

T-invariants in subnets: summary

Minimal T-invariants of the full network have clear relation with (minimal) T-invariants of a subnet:

- Isolation guarantees correctness: minimal T-invariants of the subnet are minimal T-invariants of the full network ... but they capture only internal fluxes;

T-invariants in subnets: summary

Minimal T-invariants of the full network have clear relation with (minimal) T-invariants of a subnet:

- Isolation guarantees correctness: minimal T-invariants of the subnet are minimal T-invariants of the full network ... but they capture only internal fluxes;
- Opening the subnet guarantees completeness: any invariant of the full network, once projected onto the subnet, is an invariant of the open subnet

T-invariants in subnets: summary

Minimal T-invariants of the full network have clear relation with (minimal) T-invariants of a subnet:

- Isolation guarantees correctness: minimal T-invariants of the subnet are minimal T-invariants of the full network ... but they capture only internal fluxes;
- Opening the subnet guarantees completeness: any invariant of the full network, once projected onto the subnet, is an invariant of the open subnet ...but we may loose correctness.

T-invariants in subnets: summary

Minimal T-invariants of the full network have clear relation with (minimal) T-invariants of a subnet:

- Isolation guarantees correctness: minimal T-invariants of the subnet are minimal T-invariants of the full network ... but they capture only internal fluxes;
- Opening the subnet guarantees completeness: any invariant of the full network, once projected onto the subnet, is an invariant of the open subnet ...but we may loose correctness.

Still, minimal T-invariants of the full network can be obtained compositionally from those of the subnets [Pedersen, 2008]

Fully open approach

Opening information in KEGG:

- inter-pathways connections (relations of type maplink):
  - realised through *compounds*
  - not oriented
- sources and/or sinks (e.g. extracellular substances)

Opening the pathway in an automatic way means:

- opening the maplinks in input and output
- opening sources in input and sinks in output

Fully open approach

Opening information in KEGG:

- inter-pathways connections (relations of type maplink):
  - realised through *compounds*
  - not oriented
- sources and/or sinks (e.g. extracellular substances)

Opening the pathway in an automatic way means:

- opening the maplinks in input and output
- opening sources in input and sinks in output

However:

- experiments do not give good results with this choice
  (probably due to overestimation and imprecision of the boundaries)
- the size of the Hilbert basis increases significantly

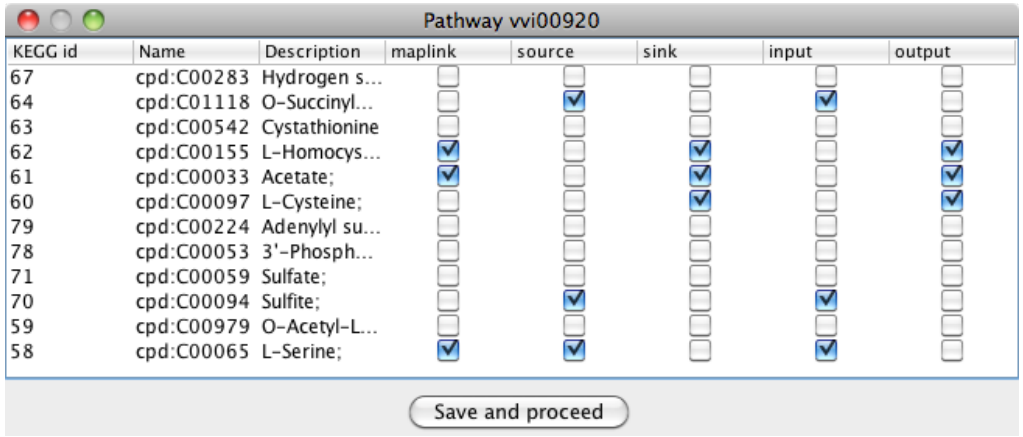
Selectively open approach

Idea: allow the user to freely select the compounds to be opened

For each specific pathway:

- all compouds are listed
- maplinks, sources and sinks are pointed out
- any compound can be opened in input and/or output

To ease the user, a *canonical choice* is offered: sources are opened in input and sinks in output



Experiments

- Goal: explore how the different treatment of the environment may affect the results of the comparison
- Only the invariant based distance is considered
- Three different approaches are compared:
  - isolated
  - fully open
  - selectively open with the *canonical choice*

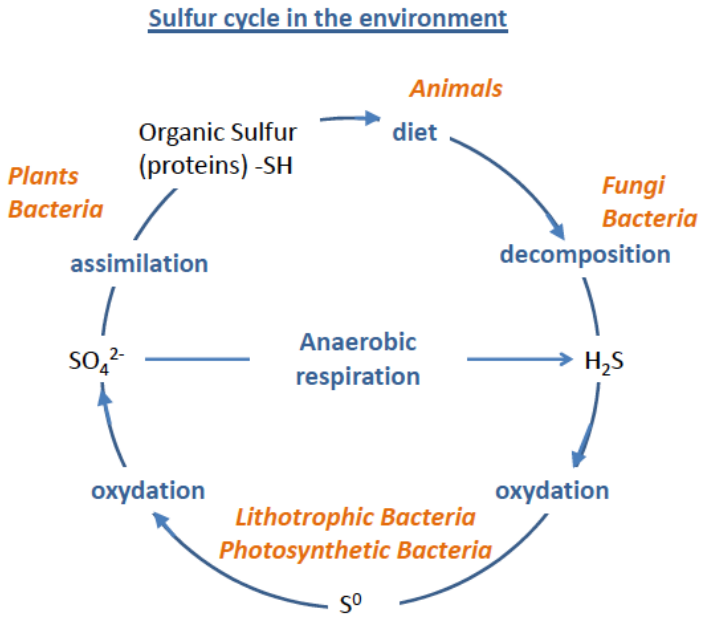
Common characteristic of the selected pathways: many irreversible reactions and few internal cicles

Experiments

- Goal: explore how the different treatment of the environment may affect the results of the comparison
- Only the invariant based distance is considered
- Three different approaches are compared:
  - isolated
  - fully open
  - selectively open with the *canonical choice*

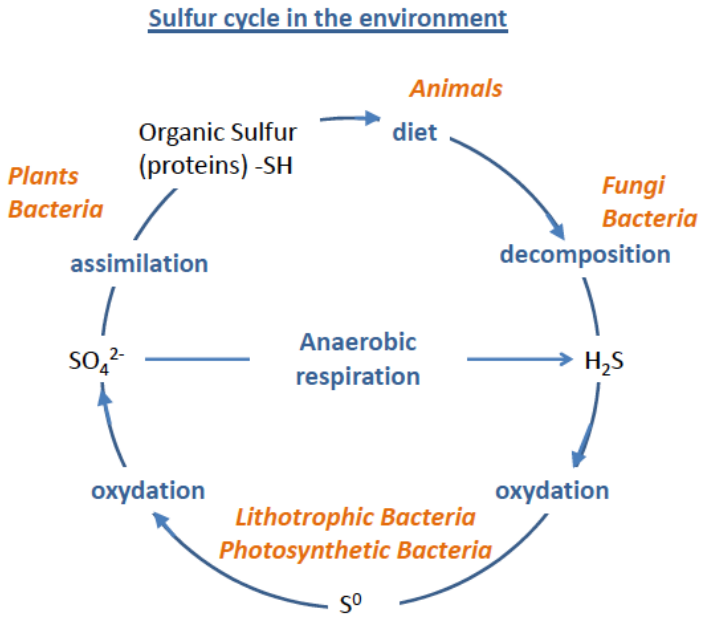
Common characteristic of the selected pathways: many irreversible reactions and few internal cicles $\Rightarrow$ few internal T-invariants

First Experiment: Sulfur metabolism



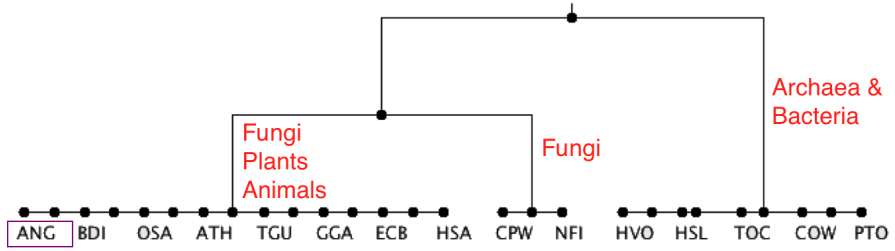
Code	Organism	Reign
hsa	*Homo sapiens*	*Mammals*
ecb	*Equus caballus*	*Mammals*
gga	*Gallus gallus*	*Birds*
tgu	*Taeniopygia guttataa*	*Birds*
ath	*Arabidopsis thaliana*	*Plants*
osa	*Oryza sativa japonica*	*Plants*
bdi	*Brachypodium distachyon*	*Plants*
nfi	*Neosartorya fischeri*	*Fungi*
ang	*Aspergillus niger*	*Fungi*
cpw	*Coccidioides posadasii*	*Fungi*
cow	*Caldicellulosiruptor owensensis*	*Bacteria*
toc	*Thermosediminibacter oceani*	*Bacteria*
hsl	*Halobacterium salinarum R1*	*Archaea*
hvo	*Haloferax volcanii*	*Archaea*
pto	*Picrophilus torridus*	*Archaea*

First Experiment: Sulfur metabolism

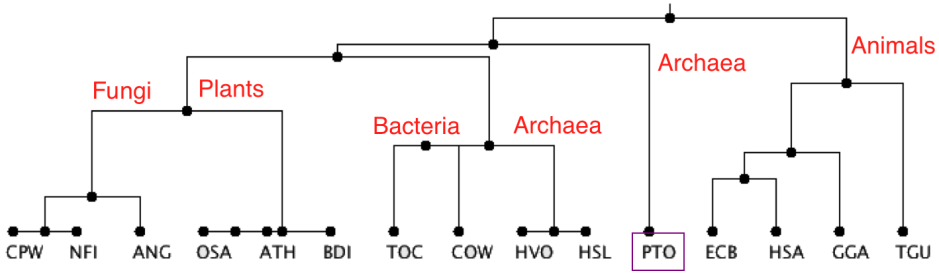


Code	Organism	Reign
hsa	*Homo sapiens*	*Mammals*
ecb	*Equus caballus*	*Mammals*
gga	*Gallus gallus*	*Birds*
tgu	*Taeniopygia guttataa*	*Birds*
ath	*Arabidopsis thaliana*	*Plants*
osa	*Oryza sativa japonica*	*Plants*
bdi	*Brachypodium distachyon*	*Plants*
nfi	*Neosartorya fischeri*	*Fungi*
ang	*Aspergillus niger*	*Fungi*
cpw	*Coccidioides posadasii*	*Fungi*
cow	*Caldicellulosiruptor owensensis*	*Bacteria*
toc	*Thermosediminibacter oceani*	*Bacteria*
hsl	*Halobacterium salinarum R1*	*Archaea*
hvo	*Haloferax volcanii*	*Archaea*
pto	*Picrophilus torridus*	*Archaea*

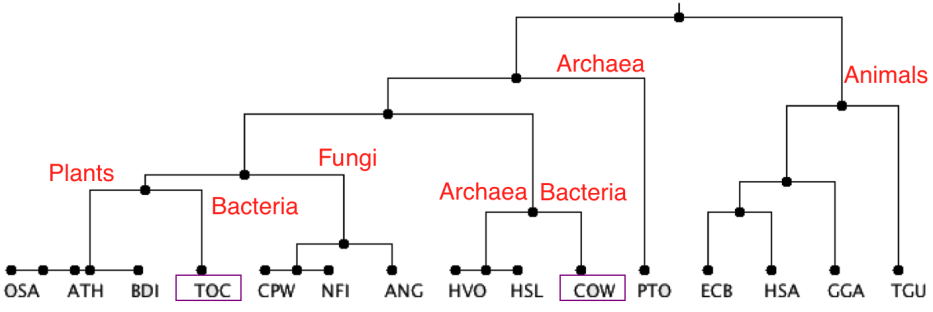
Isolated approach



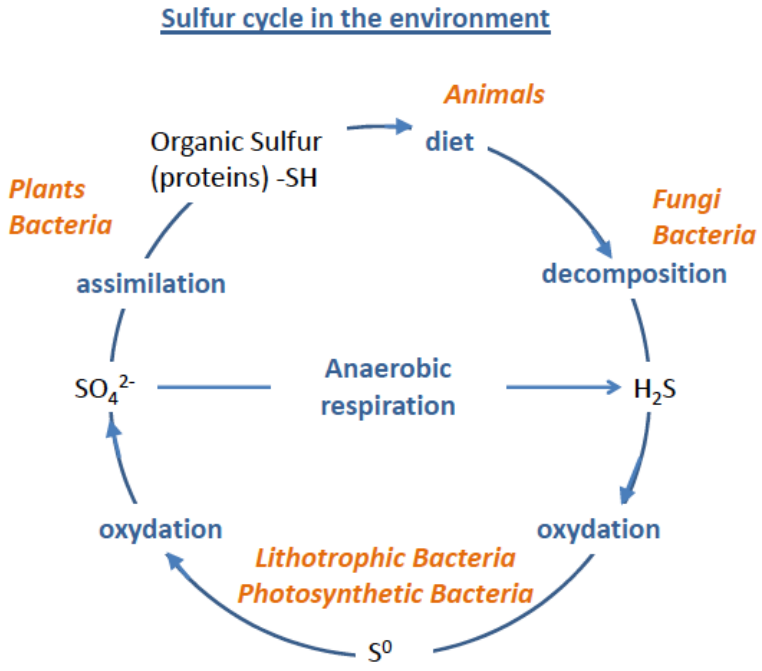
Selectively open approach



Fully open approach

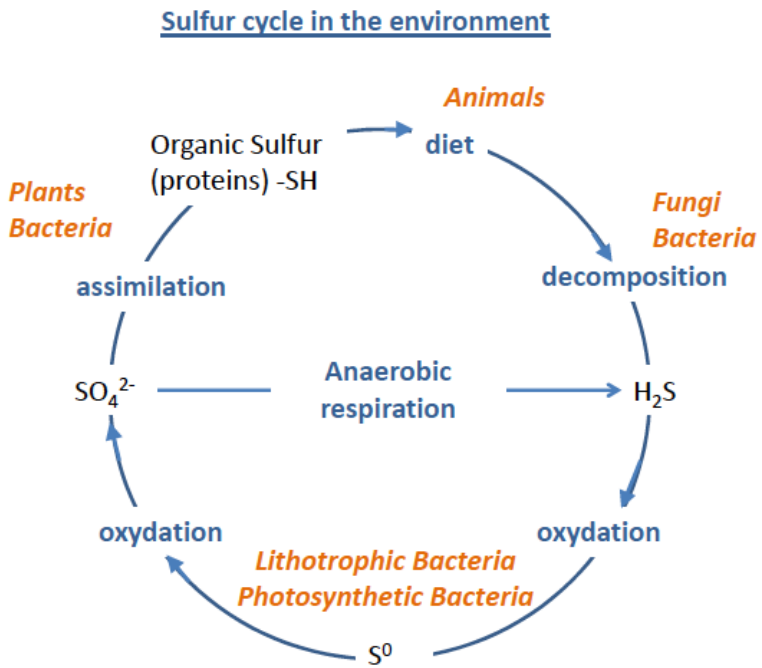


Second Experiment: Sulfur metabolism



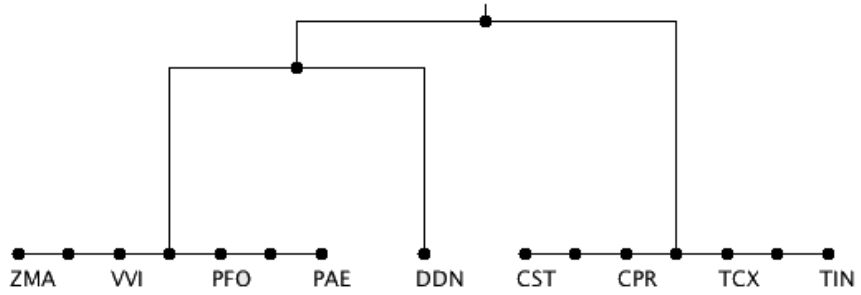
Code	Organism	Reign
pae	*Pseudomonas aeruginosa PAO1*	*Bacteria*
pfo	*Pseudomonas fluorescens Pf0-1*	*Bacteria*
tin	*Thiomonas intermedia*	*Bacteria*
tcx	*Thiomicrospira crunogena*	*Bacteria*
cpr	*Clostridium perfringens SM101*	*Bacteria*
cst	*Clostridium stricklandii*	*Bacteria*
ddn	*Desulfovibrio desulfuricans ND132*	*Bacteria*
vvi	*Vitis vinifera*	*Plants*
zma	*Zea mays*	*Plants*

Second Experiment: Sulfur metabolism

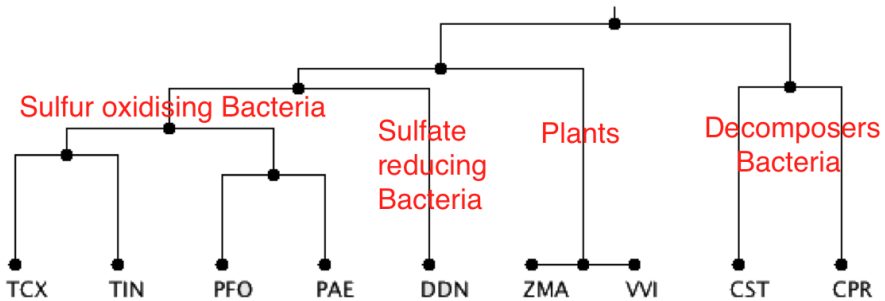


Code	Organism	Reign
pae	*Pseudomonas aeruginosa PAO1*	*Bacteria*
pfo	*Pseudomonas fluorescens Pf0-1*	*Bacteria*
tin	*Thiomonas intermedia*	*Bacteria*
tcx	*Thiomicrospira crunogena*	*Bacteria*
cpr	*Clostridium perfringens SM101*	*Bacteria*
cst	*Clostridium stricklandii*	*Bacteria*
ddn	*Desulfovibrio desulfuricans ND132*	*Bacteria*
vvi	*Vitis vinifera*	*Plants*
zma	*Zea mays*	*Plants*

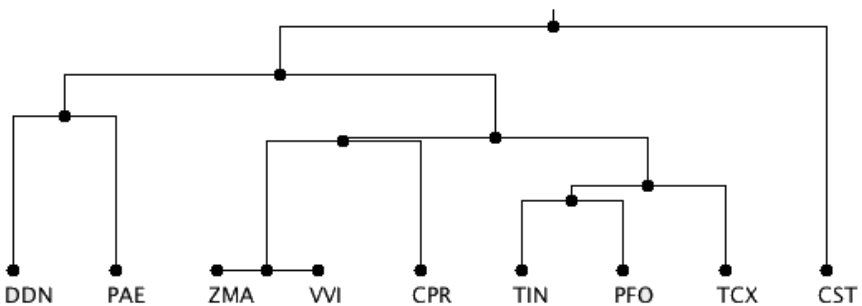
Isolated approach



Selectively open approach

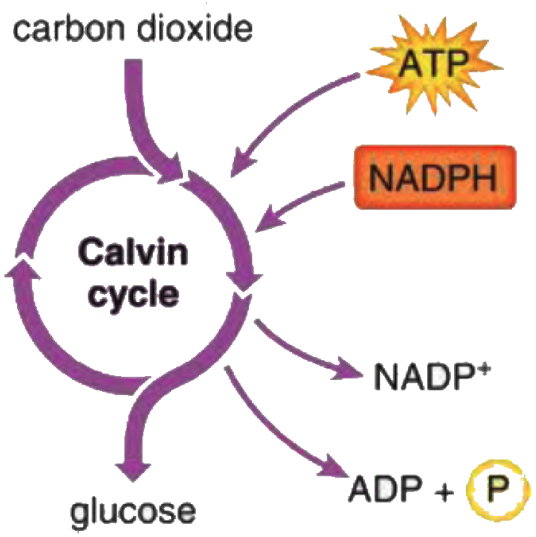


Fully open approach



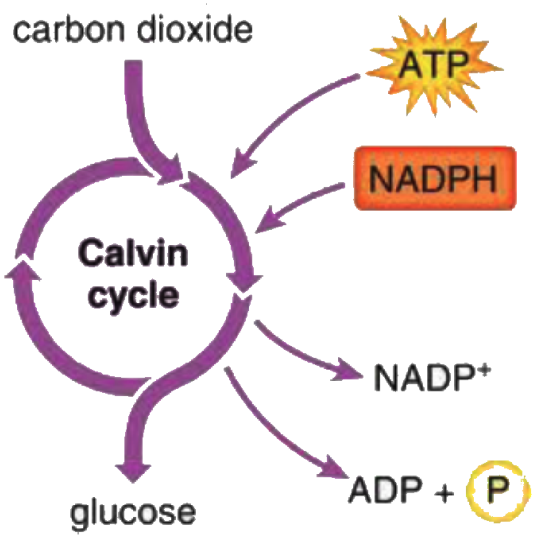
Third Experiment: Carbon metabolism

Carbon fixation pathway in photosynthetic organisms



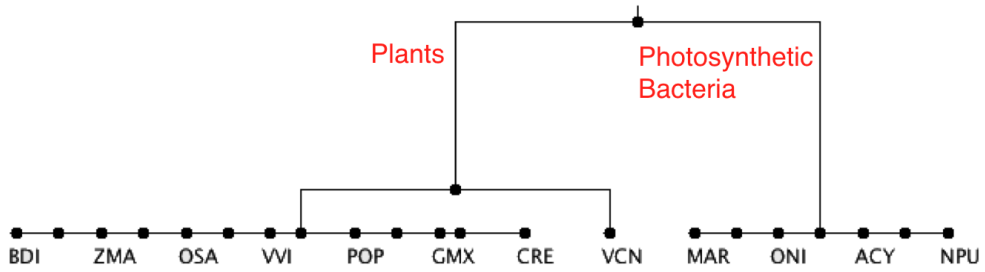
Code	Organism	Reign
gmx	*Glycine max*	*Plants, Eudicots*
pop	*Populus trichocarpa*	*Plants, Eudicots*
vvi	*Vitis vinifera*	*Plants, Eudicots*
osa	*Oryza sativa japonica*	*Plants, Monocots*
zma	*Zea mays*	*Plants, Monocots*
bdi	*Brachypodium distachyon*	*Plants, Monocots*
cre	*Chlamydomonas reinhardtii*	*Plants, green algae*
vcn	*Volvox carteri f. nagariensis*	*Plants, green algae*
npu	*Nostoc punctiforme*	*Bacteria*
acy	*Anabaena cilindrica*	*Bacteria*
oni	*Oscillatoria nigro-viridis*	*Bacteria*
mar	*Microcystis aeruginosa*	*Bacteria*

Third Experiment: Carbon metabolism

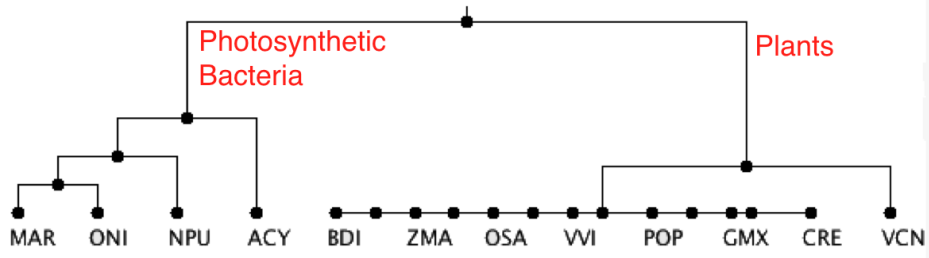


Code	Organism	Reign
gmx	*Glycine max*	*Plants, Eudicots*
pop	*Populus trichocarpa*	*Plants, Eudicots*
vvi	*Vitis vinifera*	*Plants, Eudicots*
osa	*Oryza sativa japonica*	*Plants, Monocots*
zma	*Zea mays*	*Plants, Monocots*
bdi	*Brachypodium distachyon*	*Plants, Monocots*
cre	*Chlamydomonas reinhardtii*	*Plants, green algae*
vcn	*Volvox carteri f. nagariensis*	*Plants, green algae*
npu	*Nostoc punctiforme*	*Bacteria*
acy	*Anabaena cilindrica*	*Bacteria*
oni	*Oscillatoria nigro-viridis*	*Bacteria*
mar	*Microcystis aeruginosa*	*Bacteria*

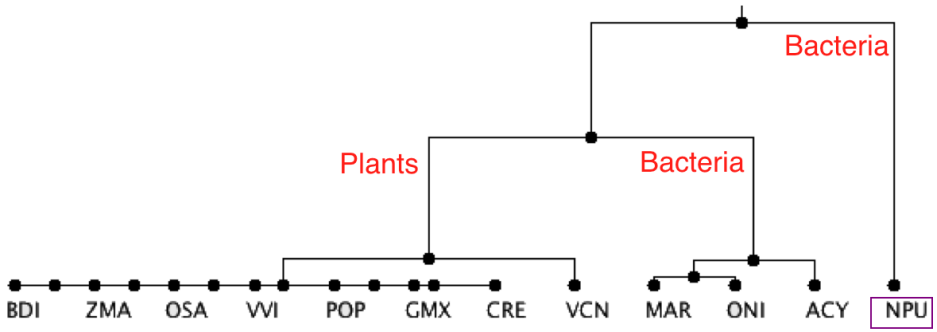
Isolated approach



Selectively open approach



Fully open approach



Concluding remarks

- Considering the environment in PN models of metabolic pathways:
  - Isolated ⇒ correctness of minimal T-invariants
  - Fully open ⇒ completeness of minimal T-invariants

  Neither of them is definitively better than the other:
  - Isolated: works well in most cases, but only internal fluxes are captured
  - Fully open: increase the size of the Hilbert basis without guaranteeing a better characterisation. Links between pathways become relevant... but KEGG links are imprecise
- We propose the Selectively open approach, where the user can freely decide the compounds to be opened
- The performed experiments suggest the appropriateness of the *canonical choice* for opening the model

Future Works

Future works will deal with:

- further experimenting with the selectively open approach
- extending the comparison to whole metabolic networks

However, the size of the Hilbert basis can be exponential in the size of the network

Two possible ways to ensure scalability of the approach:

- incrementality: compare networks obtained by merging a number of pathways of interest
- network simplification: detect portions of the whole network which are not active under some specific conditions and crop the network accordingly