Genome-scale prediction of metabolic fluxes

Standing on the shoulders of genomes

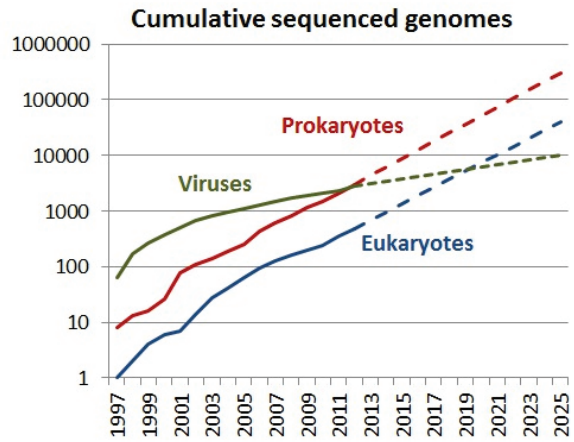
Marco Fondi
Dep. of Biology
University of Florence
marco.fondi@unifi.it
@mrcfnd

UNIVERSITÀ
DEGLI STUDI
FIRENZE

BIO
Dipartimento
di Biologia

florence
computational
biology group

dbefcb.unifi.it


Cumulative sequenced genomes
1000000
100000
10000
1000
100
10
1
Prokaryotes
Viruses
Eukaryotes
1997
1999
2001
2003
2005
2007
2009
2011
2013
2015
2017
2019
2021
2023
2025

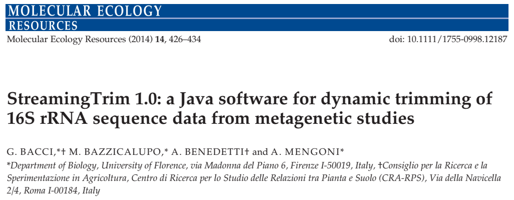

KEEP
CALM
AND
SEQUENCE
DNA

http://combo.dbe.unifi.it/medusa
https://github.com/combogenomics/medusa/releases/

http://combogenomics.github.io/DuctApe/

We are interested in:

- Microbial communities
- (Algorithms and tools for) Genomics
- Metabolic modeling



MOLECULAR ECOLOGY
RESOURCES

Molecular Ecology Resources (2014) **14**, 426–434 doi: 10.1111/1755-0998.12187

StreamingTrim 1.0: a Java software for dynamic trimming of 16S rRNA sequence data from metagenetic studies

G. BACCI,*† M. BAZZICALUPO,* A. BENEDETTI† and A. MENGONI*

**Department of Biology, University of Florence, via Madonna del Piano 6, Firenze I-50019, Italy, †Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Centro di Ricerca per lo Studio delle Relazioni tra Pianta e Suolo (CRA-RPS), Via della Navicella 2/4, Roma I-00184, Italy*

https://github.com/GiBacci/StreamingTrim/

Enly
Recycle your reads!

http://enly.sourceforge.net/

CONTIGuator web server
An online bacterial genomes finishing tool for structural insights on draft genomes

http://combo.dbe.unifi.it/contiguator

We can say something about:

Genome structure

Number of coding sequences

Preliminary functional annotation

Shared genes (comparative genomics)

Metabolic potential

We can say something about:

Genome structure

Number of coding sequences

Preliminary functional annotation

Shared genes (comparative genomics)

Metabolic potential

What we are (still) missing:

What’s the real functioning scheme of the cell?

What may happen if we remove or add a gene to the genome?

How can we “push” the metabolism towards some desired properties?

The metabolic influence of changes in gene expression

We can say something about:

- Genome structure
- Number of coding sequences
- Preliminary functional annotation
- Shared genes (comparative genomics)
- Metabolic potential

From genomes to models

Wha

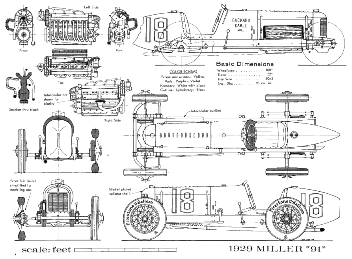
What

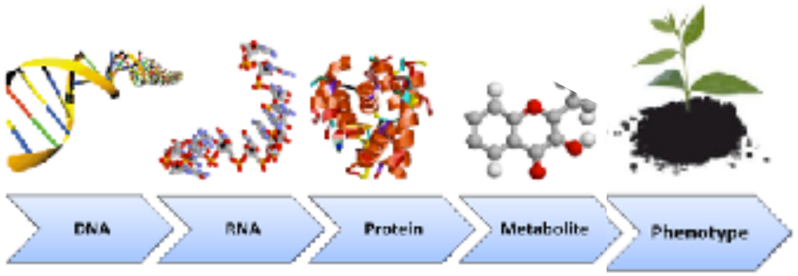
What

How

The r

Genome ⟶ Cell Functioning





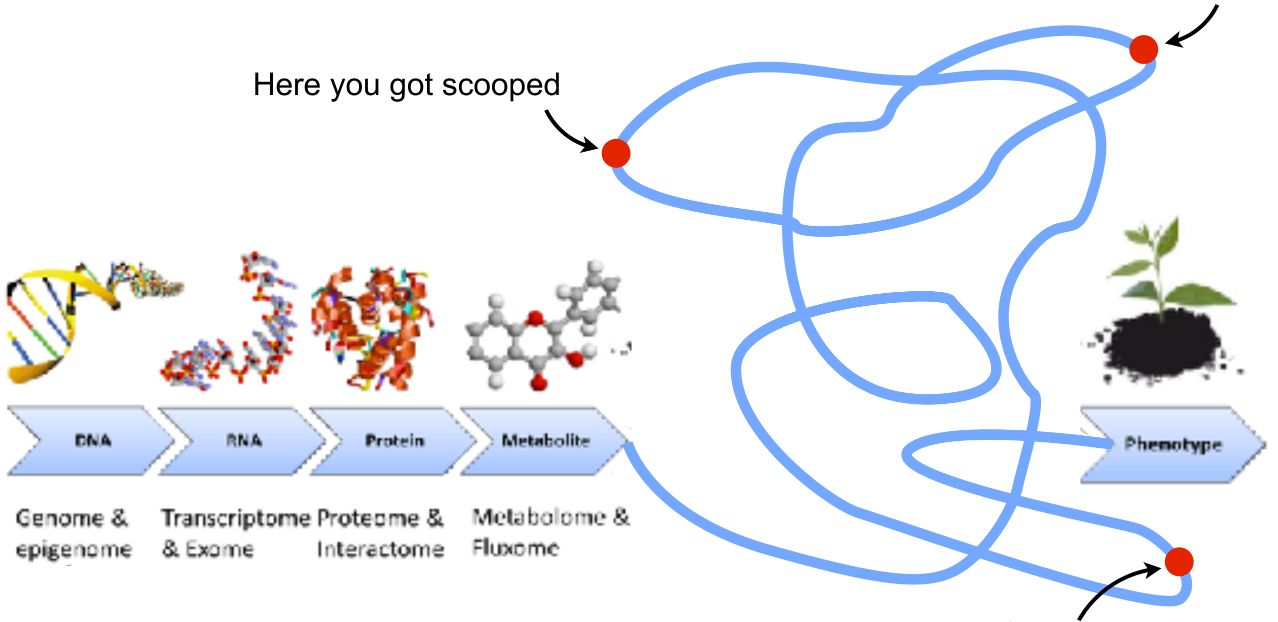
DNA → RNA → Protein → Metabolite → Phenotype

Genome & epigenome

Transcriptome & Exome

Proteome & Interactome

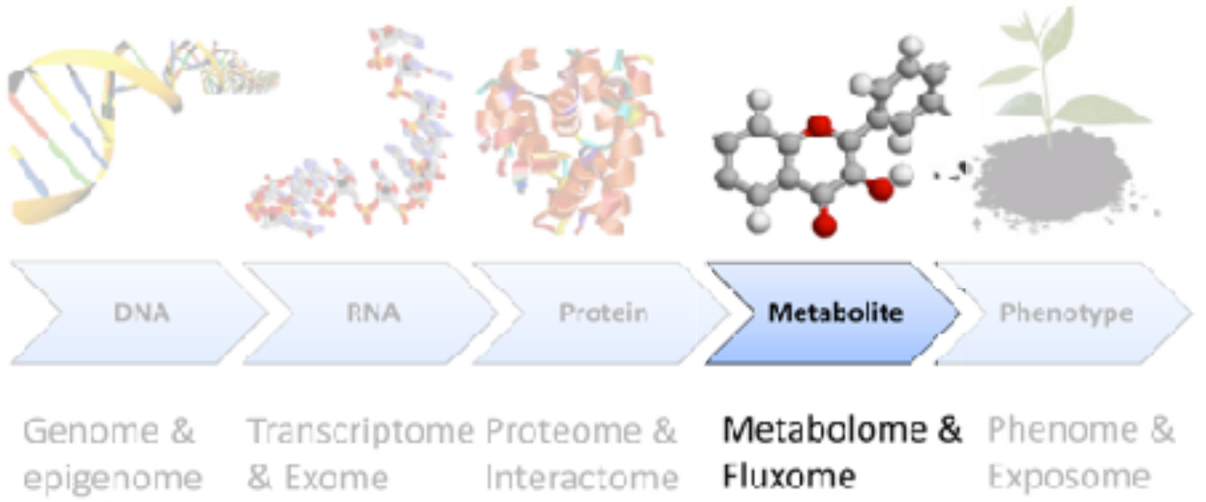
Metabolome & Fluxome



End of
your PhD

Here you got scooped

Nasty comments from reviewers



DNA

RNA

Protein

Metabolite

Phenotype

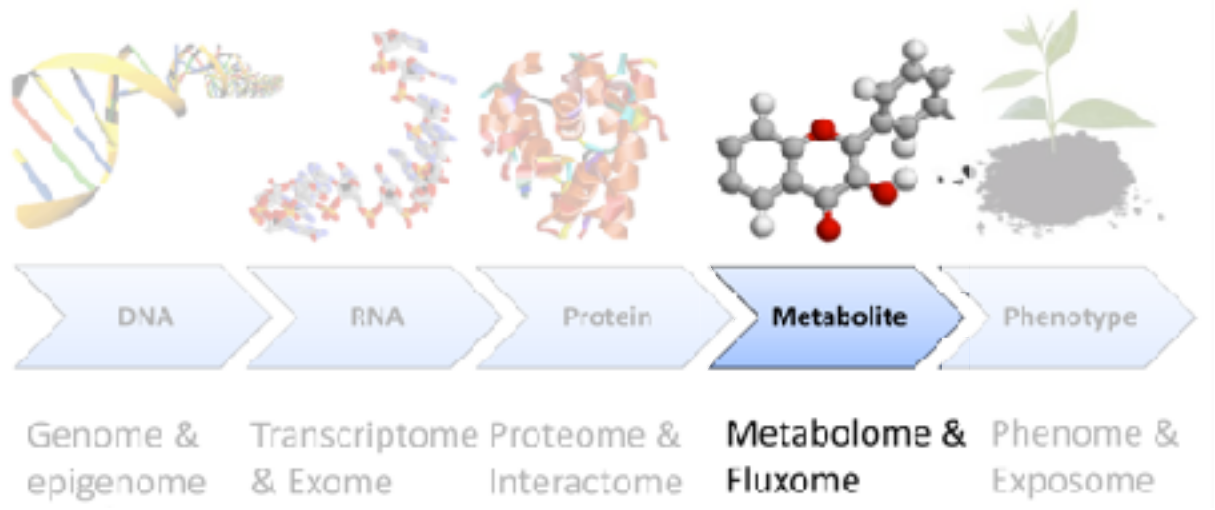
Genome & epigenome

Transcriptome & Exome

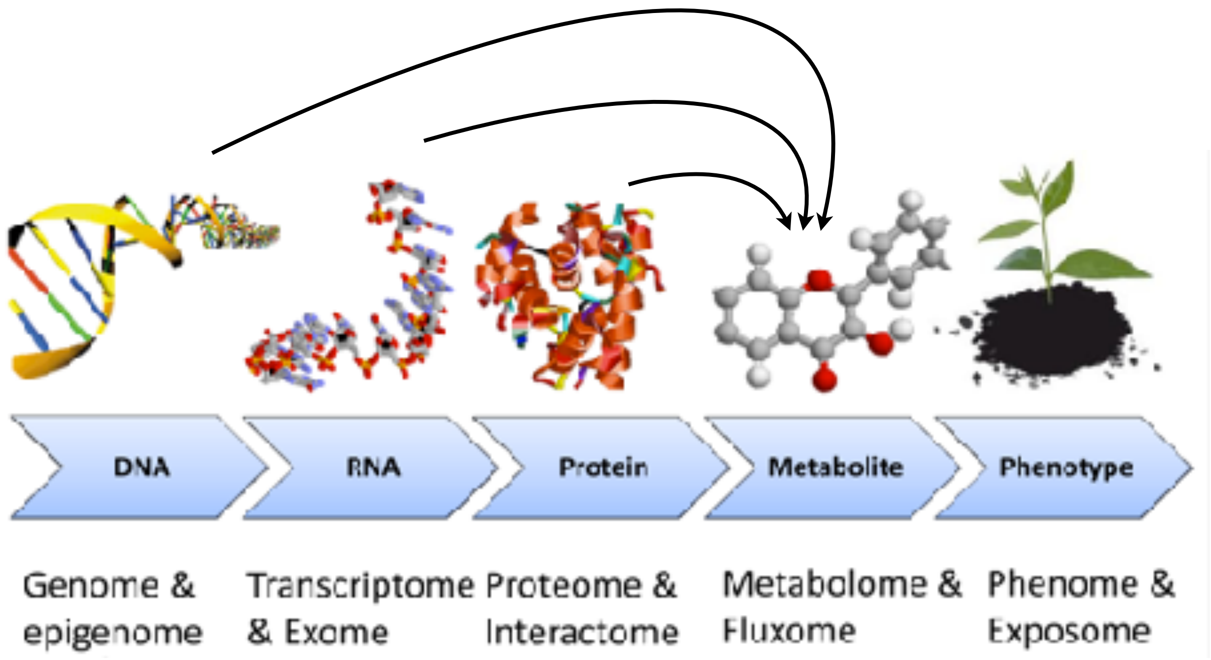
Proteome & Interactome

Metabolome & Fluxome

Phenome & Exposome

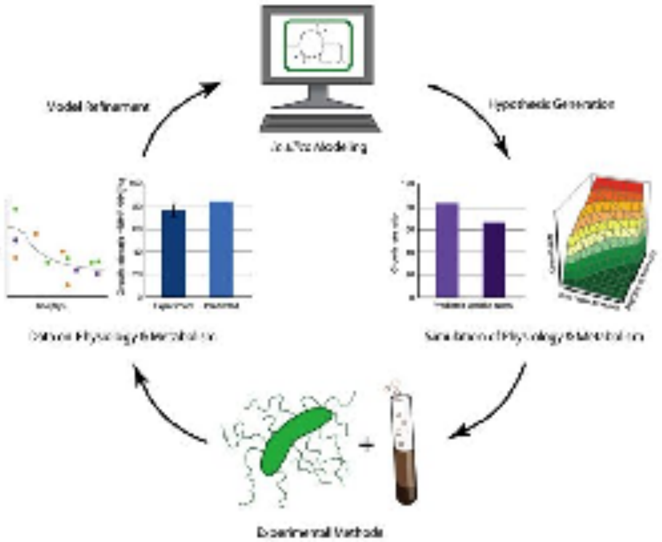


A reliable proxy for the cell phenotype


DNA
RNA
Protein
Metabolite
Phenotype
Genome & epigenome
Transcriptome & Exome
Proteome & Interactome
Metabolome & Fluxome
Phenome & Exposome

We are interested in:

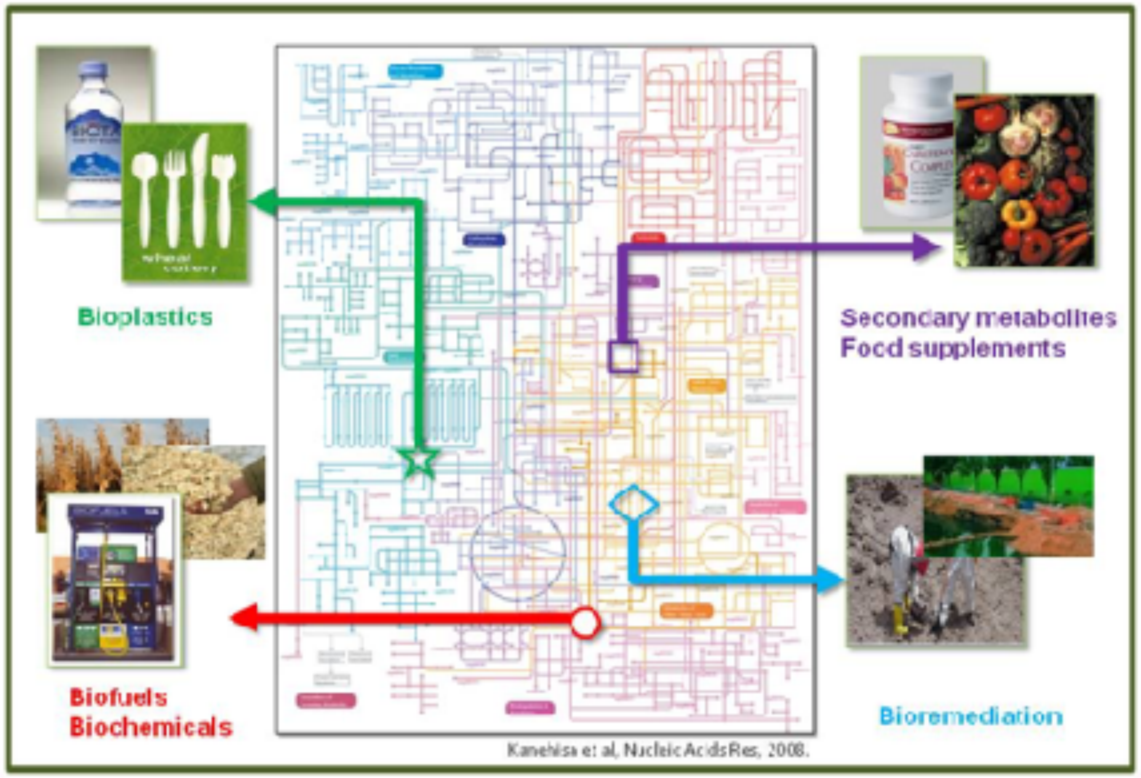
- Microbial communities
- (Algorithms and tools for) Genomics
- Metabolic modeling



the use of use of quantitative analysis methods to generate testable hypotheses and drive experimentation (possibly at whole-genome level).

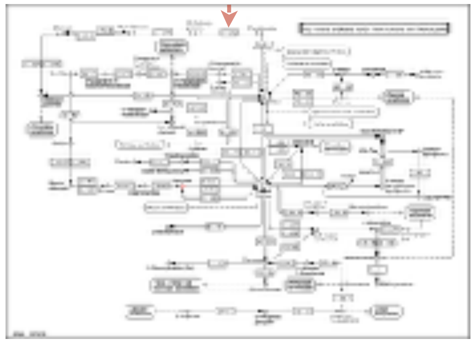
figure credit: http://www.pnl.gov/science/highlights/highlight.asp?id=982

Deciphering the architecture underlying these interconnected physicochemical processes remains one of the greatest challenges of our time.

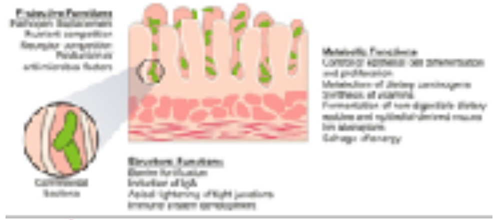


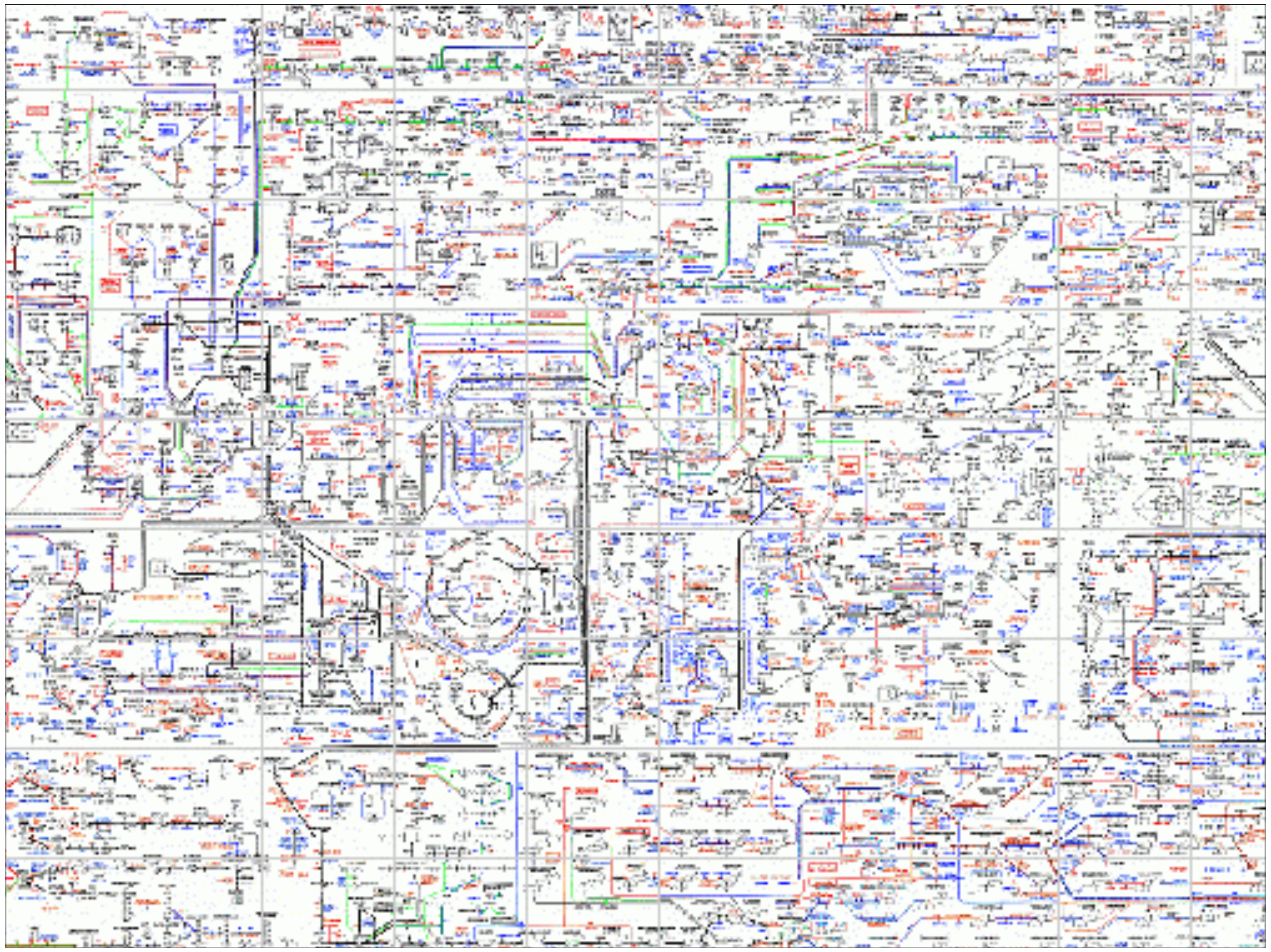
Kanehisa et al, Nucleic Acids Res, 2008.

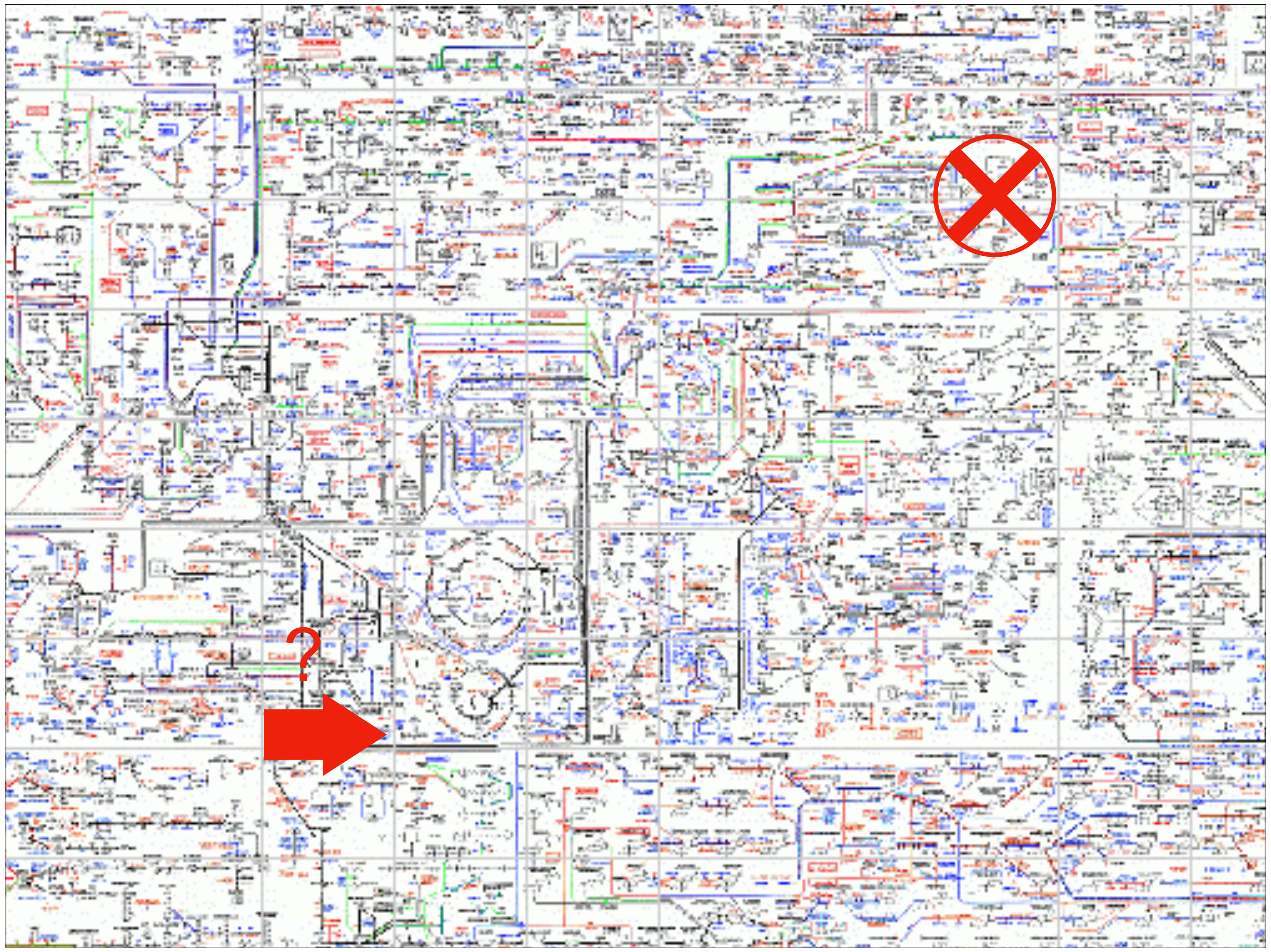
Drug target discovery



Symbiotic relationships







Aim: constructing a (working) computational representation of cellular metabolic processes for phenotypic space exploration and prediction capabilities

All models are wrong but some are useful
George E. P. Box

What is a genome-scale metabolic model?

The list of “all” the chemical reactions of a given organism

ATP + L-Glutamate + tRNA(Gln) <=> Pyrophosphate + AMP

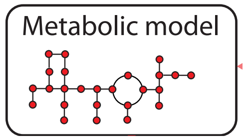
NADP+ + L-sulfolactate <=> NADPH + H+ + 3-Sulfopyruvate

(R)-Lactate + Menaquinone 6 => Pyruvate + Menaquinol 6

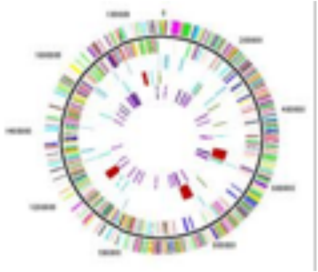
5-L-Glutamyl-L-alanine <=> L-Alanine + 5-Oxoproline

(2) H2O + GTP => (2) Orthophosphate + (2) H+ + GMP

……. ……….



(SBML format)



ATP + L-Glutamate + tRNA(Gln) <=> Pyrophosphate + AMP

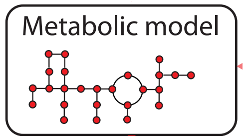
NADP+ + L-sulfolactate <=> NADPH + H+ + 3-Sulfopyruvate

(R)-Lactate + Menaquinone 6 => Pyruvate + Menaquinol 6

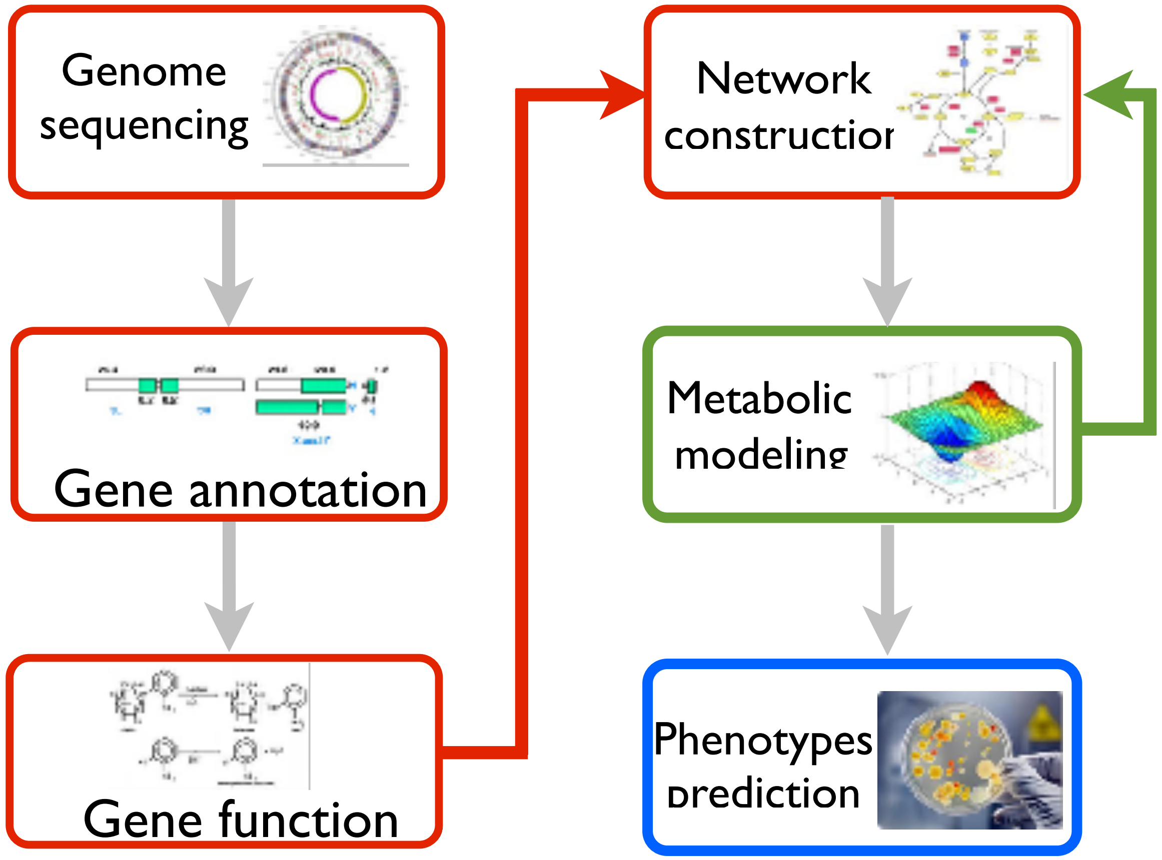
5-L-Glutamyl-L-alanine <=> L-Alanine + 5-Oxoproline

(2) H2O + GTP => (2) Orthophosphate + (2) H+ + GMP

……. ……….



(SBML format)



Genome sequencing

Gene annotation

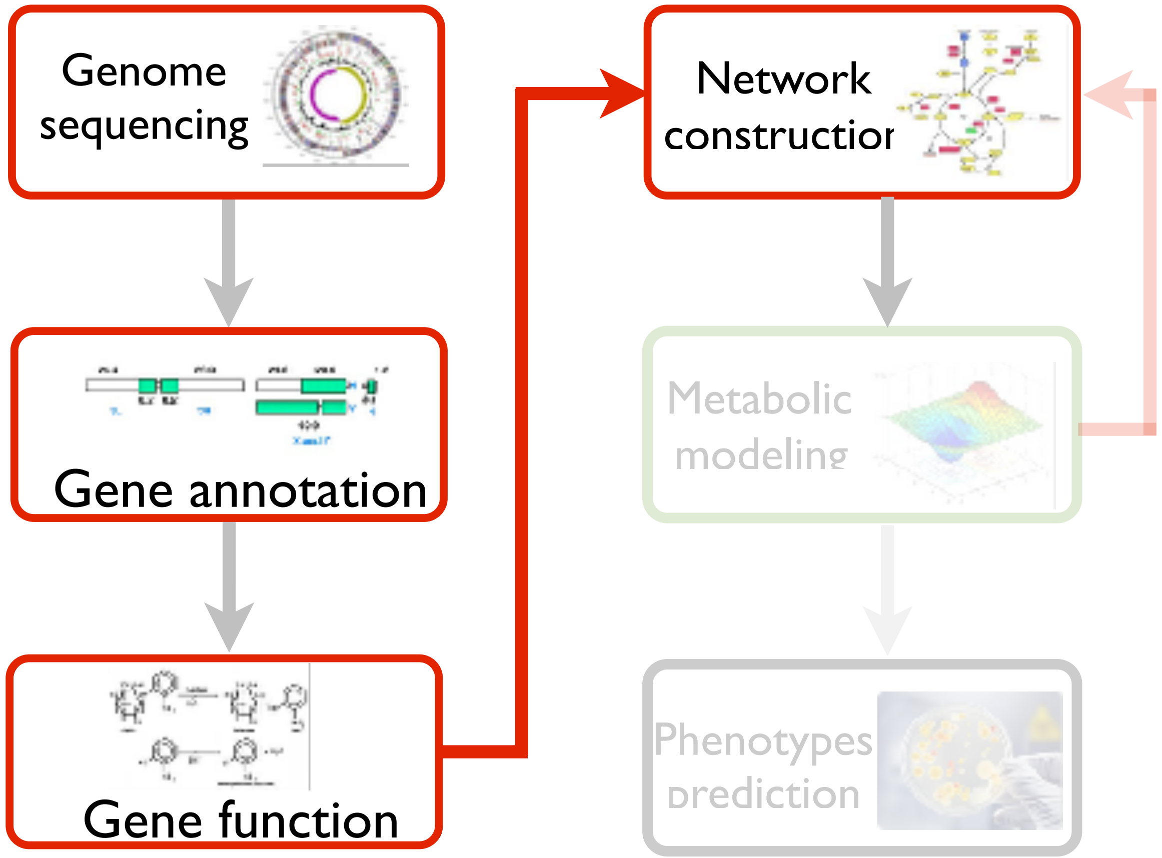
Gene function

Network construction

Metabolic modeling

Phenotypes prediction

Model reconstruction



Genome sequencing

Gene annotation

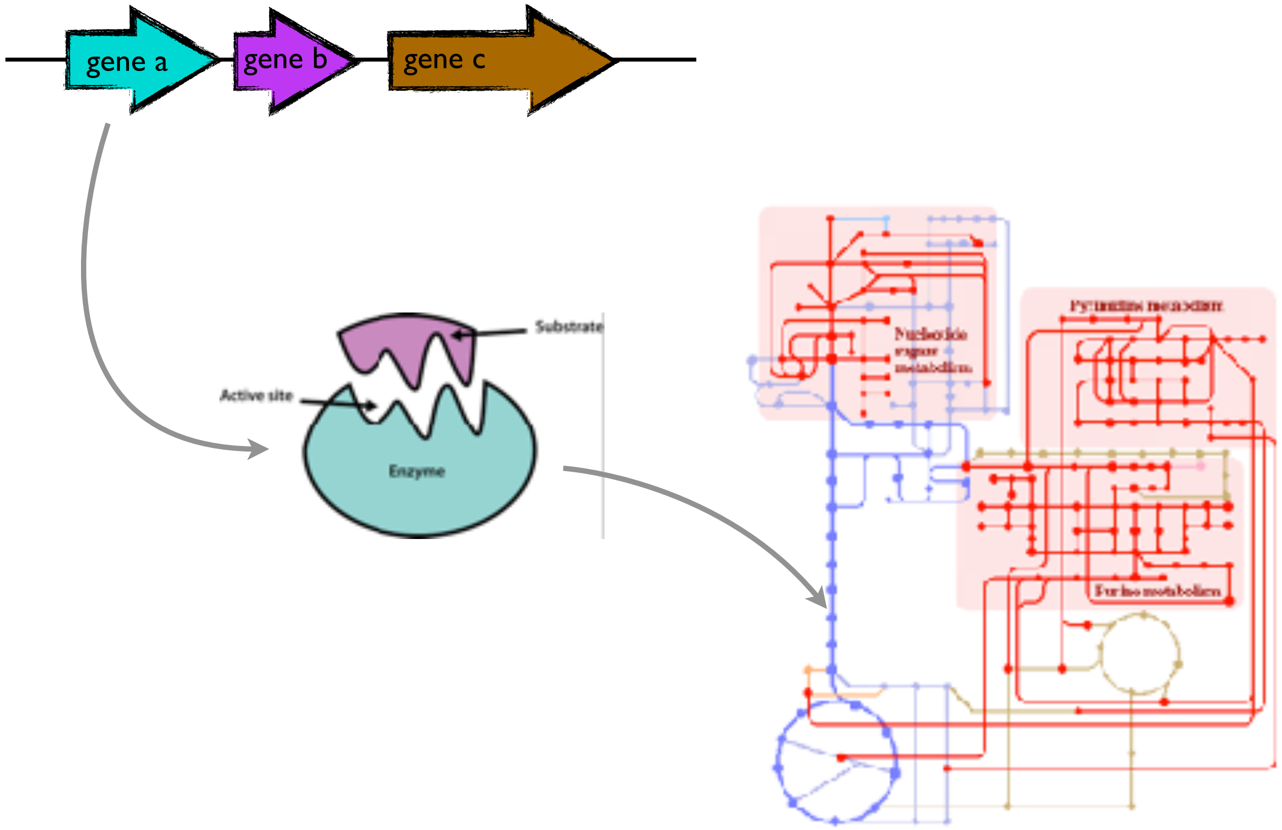
Gene function

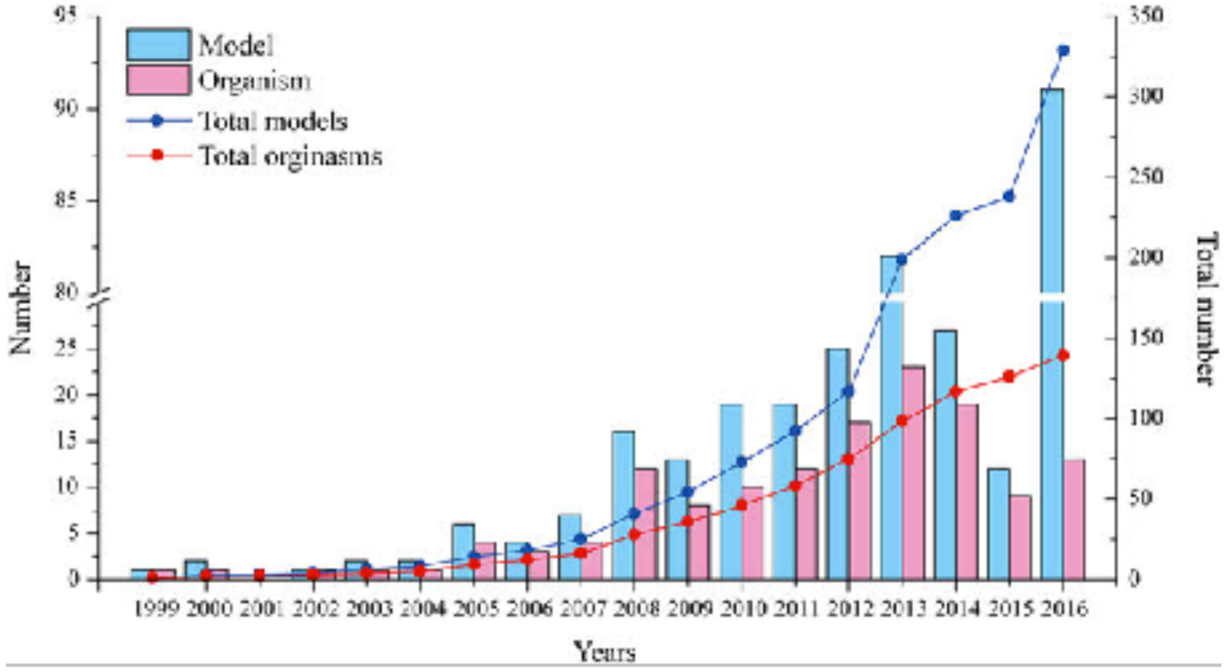
Network construction

Metabolic modeling

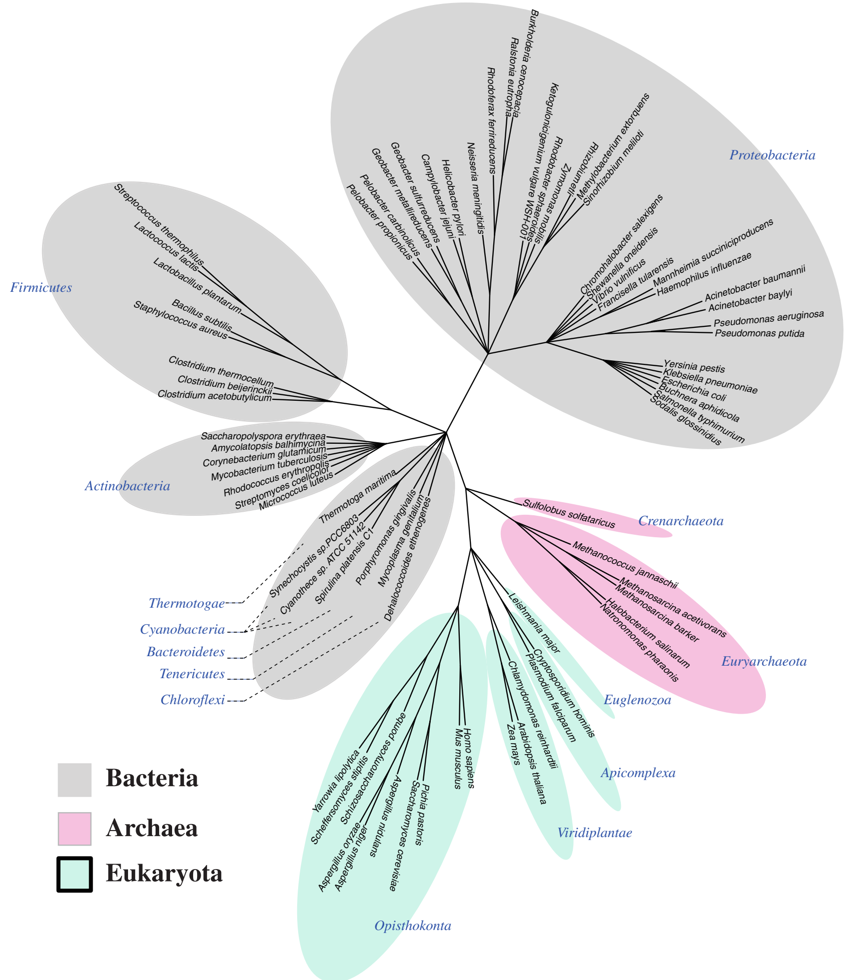
Phenotypes prediction

Genes - proteins - reactions (GPR)

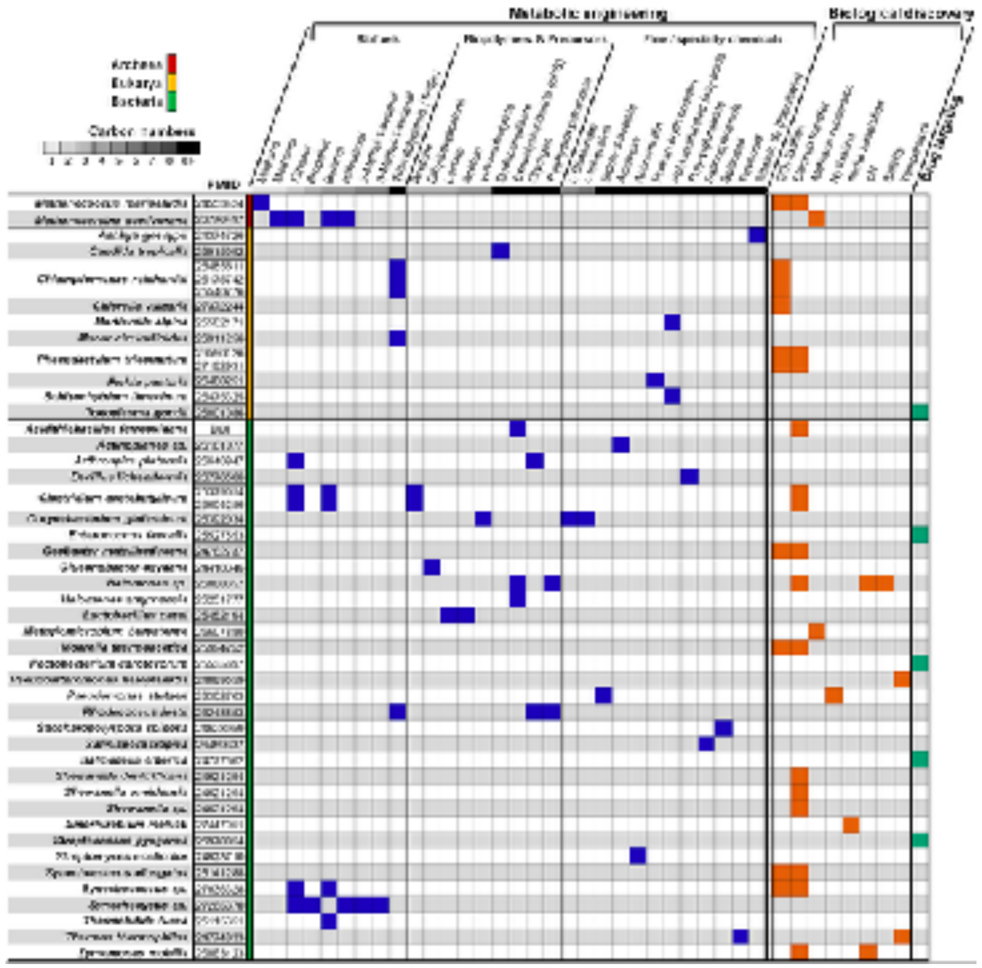




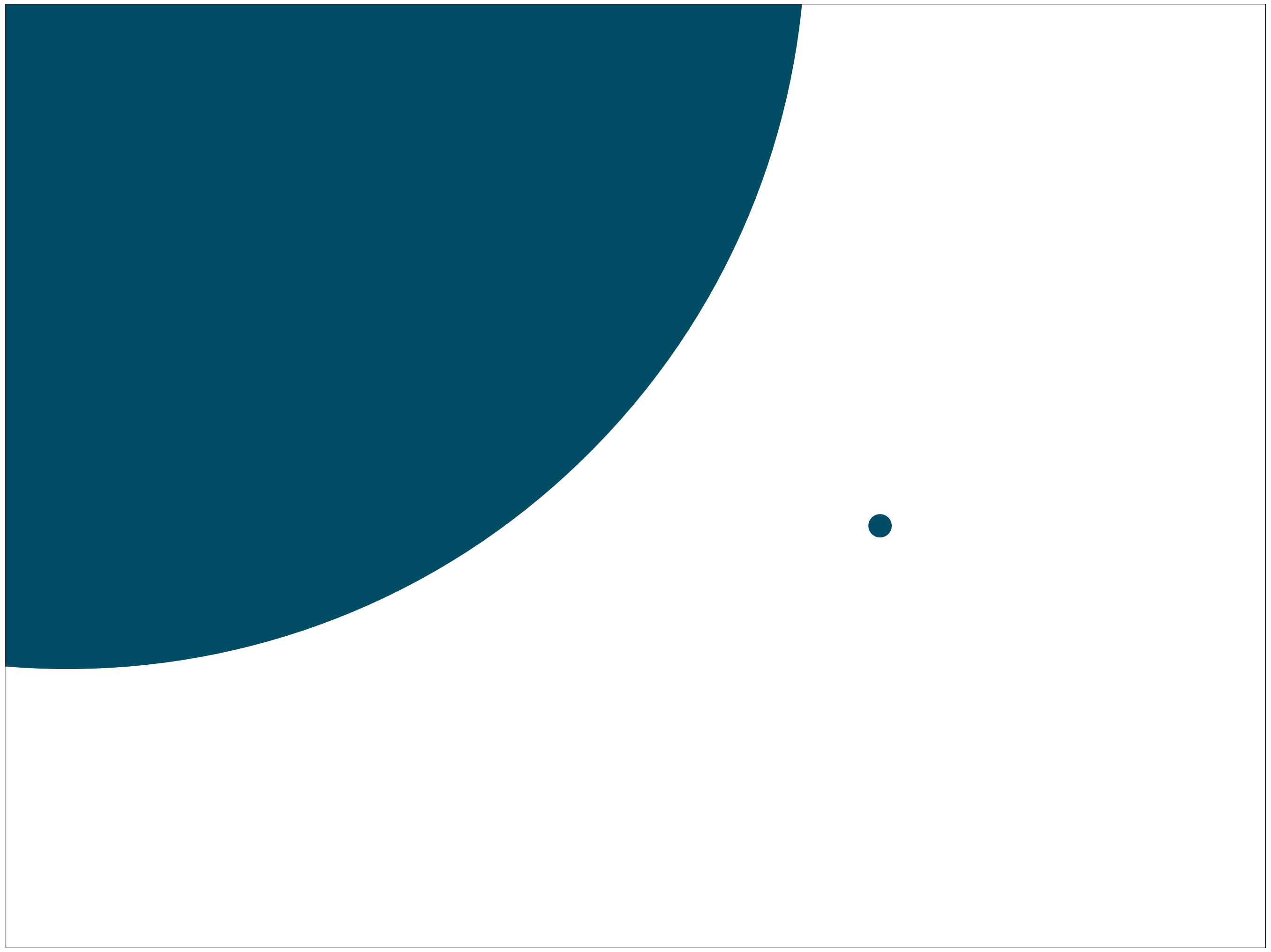
Chao et al. 2017, Scientific Reports



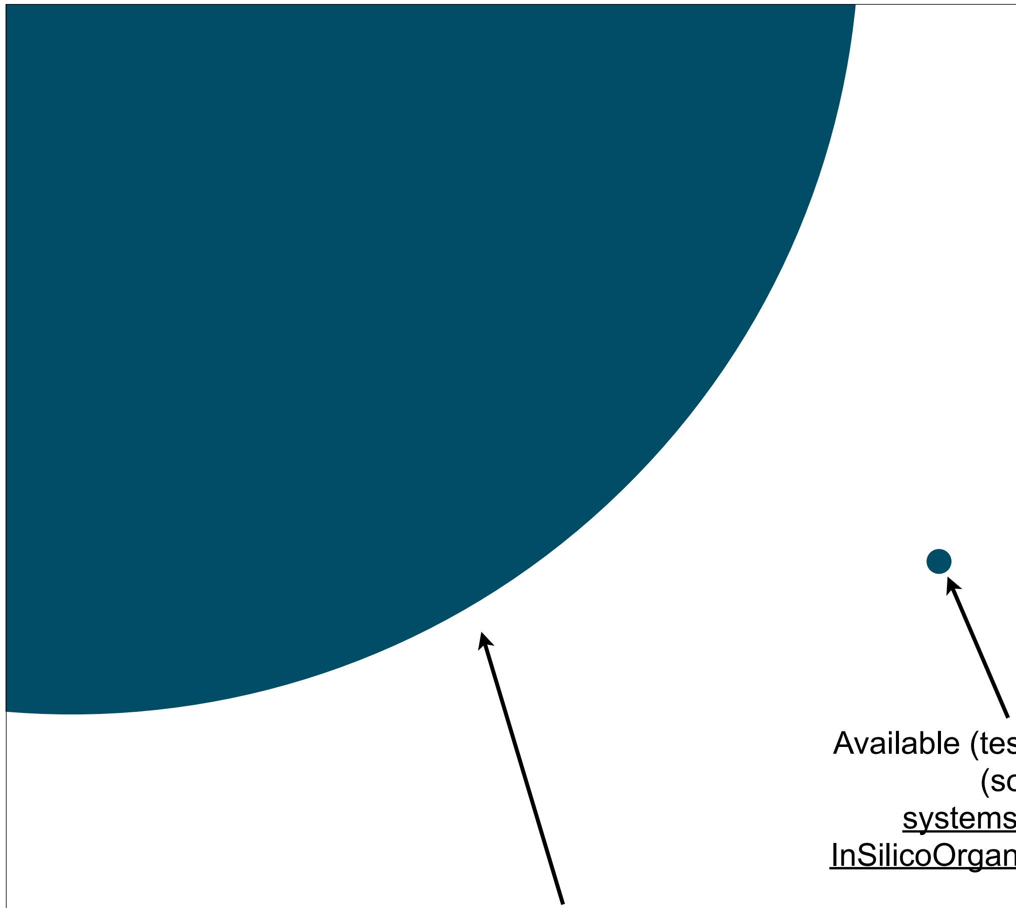
Xu et al. Appl Microbiol Biotechnol (2013) 97:519–539



Won JunKim et al. 2017

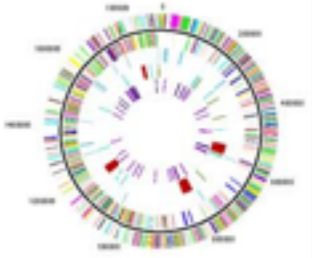


However, a huge gap exists between available genomes and working, experimentally tested metabolic models, as shown here by the size of these two circles.



Available (tested) metabolic models
(source: http://systemsbiology.ucsd.edu/InSilicoOrganisms/OtherOrganisms)

Complete and draft genomes (NCBI)



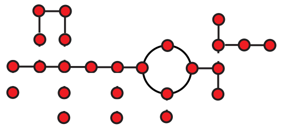
MetaCyc
A member of the BioCyc database collection

The SEED

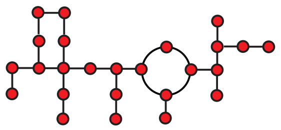
KBase
DOE Systems Biology Knowledgebase

KAAS

Draft metabolic model

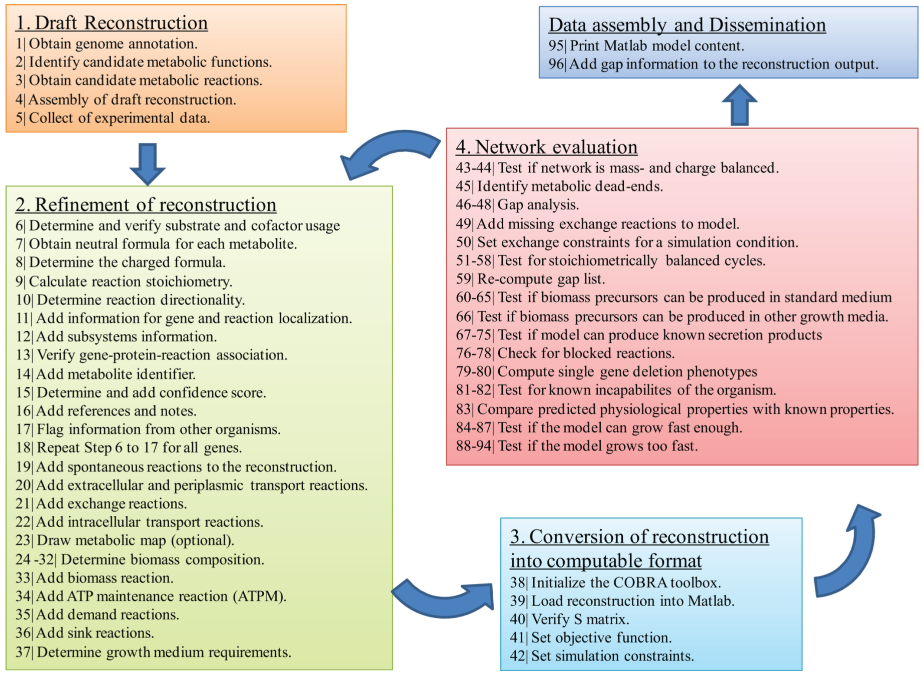


Complete metabolic model



A protocol for generating a high-quality genome-scale metabolic reconstruction

Ines Thiele[1, 2] & Bernhard Ø Palsson[1]



1. Draft Reconstruction

1| Obtain genome annotation.
2| Identify candidate metabolic functions.
3| Obtain candidate metabolic reactions.
4| Assembly of draft reconstruction.
5| Collect of experimental data.

2. Refinement of reconstruction

6| Determine and verify substrate and cofactor usage
7| Obtain neutral formula for each metabolite.
8| Determine the charged formula.
9| Calculate reaction stoichiometry.
10| Determine reaction directionality.
11| Add information for gene and reaction localization.
12| Add subsystems information.
13| Verify gene-protein-reaction association.
14| Add metabolite identifier.
15| Determine and add confidence score.
16| Add references and notes.
17| Flag information from other organisms.
18| Repeat Step 6 to 17 for all genes.
19| Add spontaneous reactions to the reconstruction.
20| Add extracellular and periplasmic transport reactions.
21| Add exchange reactions.
22| Add intracellular transport reactions.
23| Draw metabolic map (optional).
24 -32| Determine biomass composition.
33| Add biomass reaction.
34| Add ATP maintenance reaction (ATPM).
35| Add demand reactions.
36| Add sink reactions.
37| Determine growth medium requirements.

3. Conversion of reconstruction into computable format

38| Initialize the COBRA toolbox.
39| Load reconstruction into Matlab.
40| Verify S matrix.
41| Set objective function.
42| Set simulation constraints.

4. Network evaluation

43-44| Test if network is mass- and charge balanced.
45| Identify metabolic dead-ends.
46-48| Gap analysis.
49| Add missing exchange reactions to model.
50| Set exchange constraints for a simulation condition.
51-58| Test for stoichiometrically balanced cycles.
59| Re-compute gap list.
60-65| Test if biomass precursors can be produced in standard medium
66| Test if biomass precursors can be produced in other growth media.
67-75| Test if model can produce known secretion products
76-78| Check for blocked reactions.
79-80| Compute single gene deletion phenotypes
81-82| Test for known incapabilites of the organism.
83| Compare predicted physiological properties with known properties.
84-87| Test if the model can grow fast enough.
88-94| Test if the model grows too fast.

Data assembly and Dissemination

95| Print Matlab model content.
96| Add gap information to the reconstruction output.

Model reconstruction still represents a very labor and time intensive process. Speeding up this process is a priority for efficient and rapid in silico modeling of microbial metabolism

SYSTEMS BIOLOGY
RESEARCH GROUP

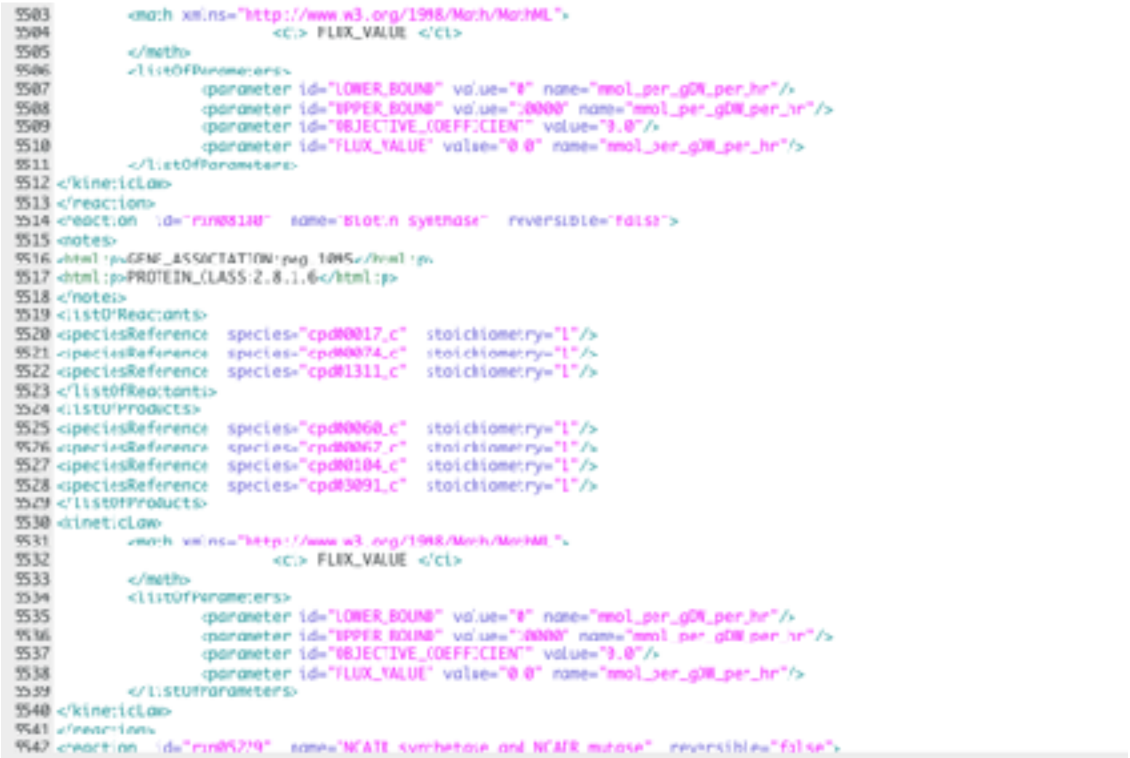
OTHER ORGANISMS

http://systemsbiology.ucsd.edu/InSilicoOrganisms/OtherOrganisms

- (the different models) are maintained by different groups and often incorporate different names for many chemical species and reactions, this step was not completely straightforward and could not be performed entirely automatically
- few highly curated and reliable models available

SBML Systems Biology Markup Language

SBML Systems Biology Markup Language

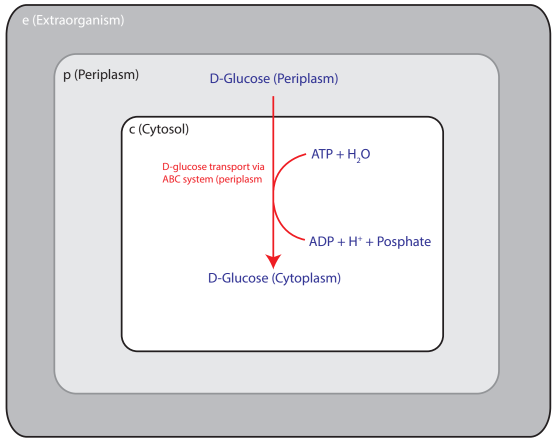


```
        <math xmlns="http://www.w3.org/1998/Math/MathML">
                        <ci> FLUX_VALUE </ci>
        </math>
        <listOfParameters>
                <parameter id="LOWER_BOUND" value="0" name="mmol_per_gDW_per_hr"/>
                <parameter id="UPPER_BOUND" value="10000" name="mmol_per_gDW_per_hr"/>
                <parameter id="OBJECTIVE_COEFFICIENT" value="0.0"/>
                <parameter id="FLUX_VALUE" value="0.0" name="mmol_per_gDW_per_hr"/>
        </listOfParameters>
</kineticLaw>
</reaction>
<reaction  id="rxn00130"  name="Biotin synthase"  reversible="false">
<notes>
<html:p>GENE_ASSOCIATION:peg.1095</html:p>
<html:p>PROTEIN_CLASS:2.8.1.6</html:p>
</notes>
<listOfReactants>
<speciesReference  species="cpd00017_c"  stoichiometry="1"/>
<speciesReference  species="cpd00074_c"  stoichiometry="1"/>
<speciesReference  species="cpd01311_c"  stoichiometry="1"/>
</listOfReactants>
<listOfProducts>
<speciesReference  species="cpd00060_c"  stoichiometry="1"/>
<speciesReference  species="cpd00067_c"  stoichiometry="1"/>
<speciesReference  species="cpd00104_c"  stoichiometry="1"/>
<speciesReference  species="cpd03091_c"  stoichiometry="1"/>
</listOfProducts>
<kineticLaw>
        <math xmlns="http://www.w3.org/1998/Math/MathML">
                        <ci> FLUX_VALUE </ci>
        </math>
        <listOfParameters>
                <parameter id="LOWER_BOUND" value="0" name="mmol_per_gDW_per_hr"/>
                <parameter id="UPPER_BOUND" value="10000" name="mmol_per_gDW_per_hr"/>
                <parameter id="OBJECTIVE_COEFFICIENT" value="0.0"/>
                <parameter id="FLUX_VALUE" value="0.0" name="mmol_per_gDW_per_hr"/>
        </listOfParameters>
</kineticLaw>
</reaction>
<reaction  id="rxn05219"  name="NCAIR synthetase and NCAIR mutase"  reversible="false">
```

Systems Biology Markup Language

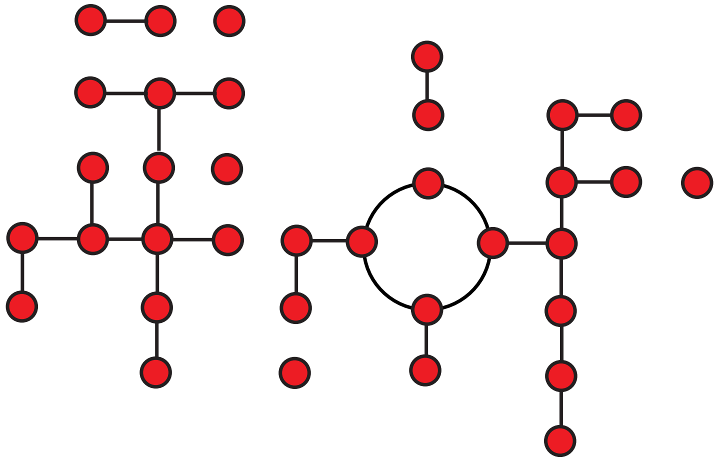
```
<reaction id="rxn03180" name="Biotin synthase" reversible="false">
<notes>
<html:p>GENE_ASSOCIATION:peg.1095</html:p>
<html:p>PROTEIN_CLASS:2.8.1.6</html:p>
</notes>
<listOfReactants>
<speciesReference species="cpd00017_c" stoichiometry="1"/>
<speciesReference species="cpd00074_c" stoichiometry="1"/>
<speciesReference species="cpd01311_c" stoichiometry="1"/>
</listOfReactants>
<listOfProducts>
<speciesReference species="cpd00060_c" stoichiometry="1"/>
<speciesReference species="cpd00067_c" stoichiometry="1"/>
<speciesReference species="cpd00104_c" stoichiometry="1"/>
<speciesReference species="cpd03091_c" stoichiometry="1"/>
</listOfProducts>
```

SBML file overall structure



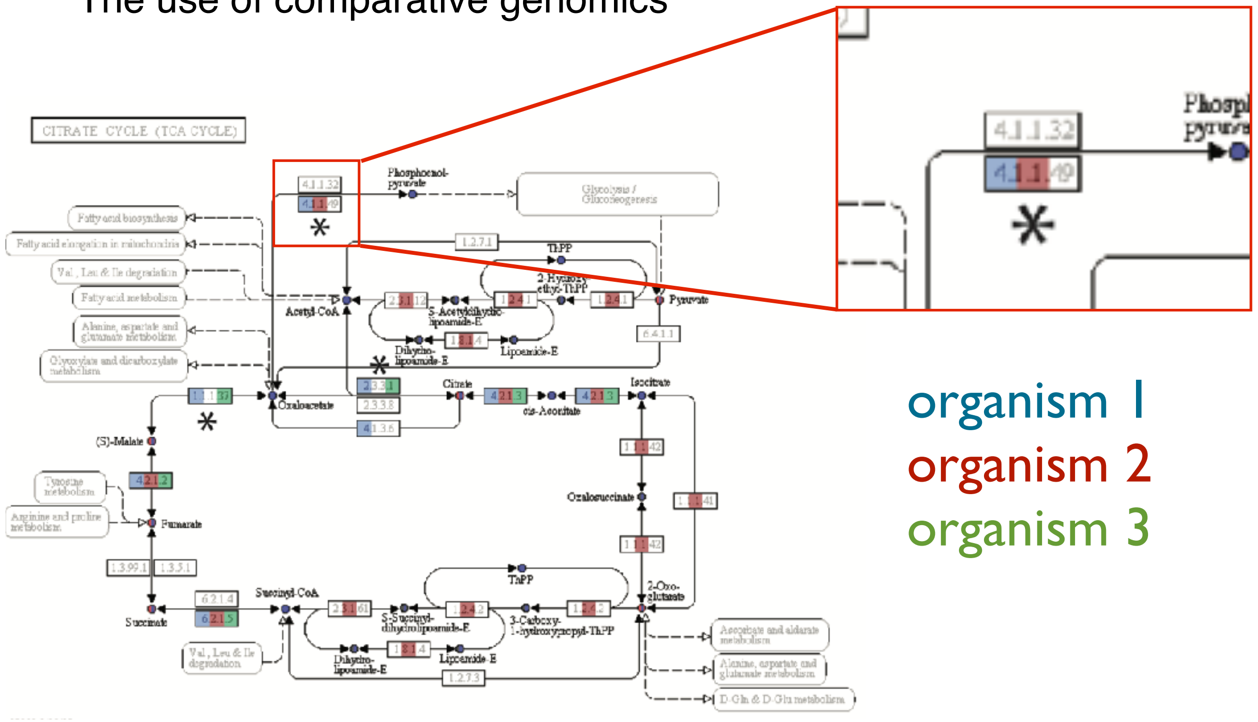
```xml
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level2" level="2" version="1" xmlns:html="http://www.w3.org/1999/xhtml">
<model id="MODELID_3307911" name="E. coli iAF1260">
<listOfUnitDefinitions>
<unitDefinition id="mmol_per_gDW_per_hr">
<listOfUnits>
<unit kind="mole" scale="-3" multiplier="1" offset="0" />
<unit kind="gram" exponent="-1" multiplier="1" offset="0" />
<unit kind="second" exponent="-1" multiplier="0.00027777" offset="0" />
</listOfUnits>
</unitDefinition>
</listOfUnitDefinitions>
<listOfCompartments>
<compartment id="C_e" name="Extraorganism"/>
<compartment id="C_c" name="Cytosol"/>
<compartment id="C_p" name="Periplasm"/>
</listOfCompartments>

<listOfSpecies>

<species id="M_glc_DASH_D_p" name="D-Glucose_C6H12O6" compartment="C_p" charge="0" boundaryCondition="false"/>
<species id="M_glc_DASH_D_c" name="D-Glucose_C6H12O6" compartment="C_c" charge="0" boundaryCondition="false"/>
<species id="M_atp_c" name="ATP_C10H12N5O13P3" compartment="C_c" charge="-4" boundaryCondition="false"/>
<species id="M_h2o_c" name="H2O_H2O" compartment="C_c" charge="0" boundaryCondition="false"/>
<species id="M_h_c" name="H+_H" compartment="C_c" charge="1" boundaryCondition="false"/>
<species id="M_pi_c" name="Phosphate_HO4P" compartment="C_c" charge="-2" boundaryCondition="false"/>
<species id="M_adp_c" name="ADP_C10H12N5O10P2" compartment="C_c" charge="-3" boundaryCondition="false"/>

</listOfSpecies>

<listOfReactions>

<reaction id="R_GLCabcpp" name="D-glucose transport via ABC system (periplasm)" reversible="false">
<listOfReactants>
<speciesReference species="M_atp_c" stoichiometry="1"/>
<speciesReference species="M_glc_DASH_D_p" stoichiometry="1"/>
<speciesReference species="M_h2o_c" stoichiometry="1"/>
</listOfReactants>
<listOfProducts>
<speciesReference species="M_adp_c" stoichiometry="1"/>
<speciesReference species="M_glc_DASH_D_c" stoichiometry="1"/>
<speciesReference species="M_h_c" stoichiometry="1"/>
<speciesReference species="M_pi_c" stoichiometry="1"/>
</listOfProducts>
<kineticLaw>
<math xmlns="http://www.w3.org/1998/Math/MathML">
<ci>FLUX_VALUE</ci>
</math>
<listOfParameters>
<parameter id="LOWER_BOUND" value="0" units="mmol_per_gDW_per_hr"/>
<parameter id="UPPER_BOUND" value="999999" units="mmol_per_gDW_per_hr"/>
<parameter id="OBJECTIVE_COEFFICIENT" value="0" />
<parameter id="FLUX_VALUE" value="0" units="mmol_per_gDW_per_hr"/>
</listOfParameters>
</kineticLaw>
</reaction>

</listOfReactions>

</model>
</sbml>
```

Unfortunately, at this stage, the reconstructed model may be incomplete and lack metabolic genes and/or functions. Thus, before starting modelling procedures, it is important to check possible sources of errors.



Unfortunately, at this stage, the reconstructed model may be incomplete and lack metabolic genes and/or functions. Thus, before starting modelling procedures, it is important to check possible sources of errors.

The use of comparative genomics

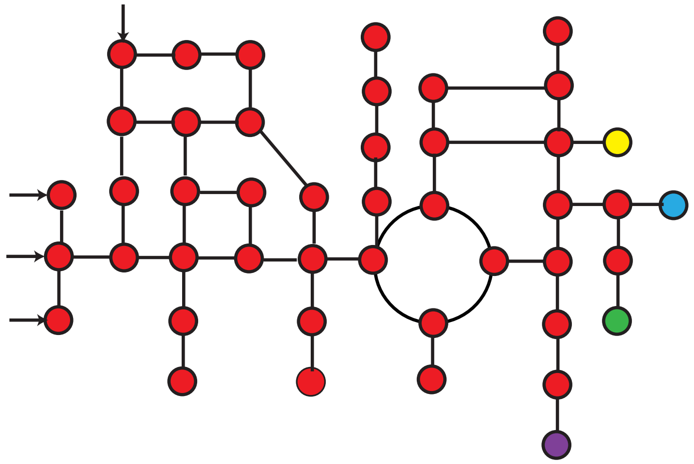


organism 1

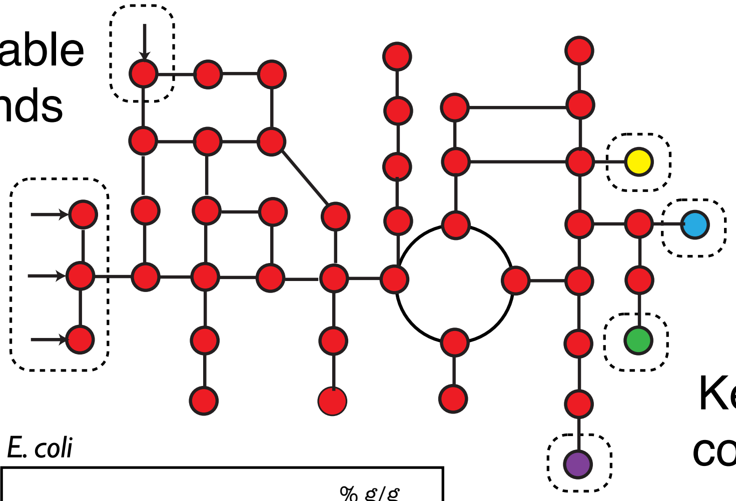
organism 2

organism 3

Towards a complete reconstruction

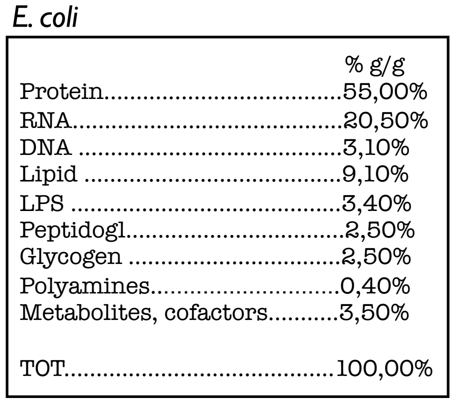


Towards a complete reconstruction



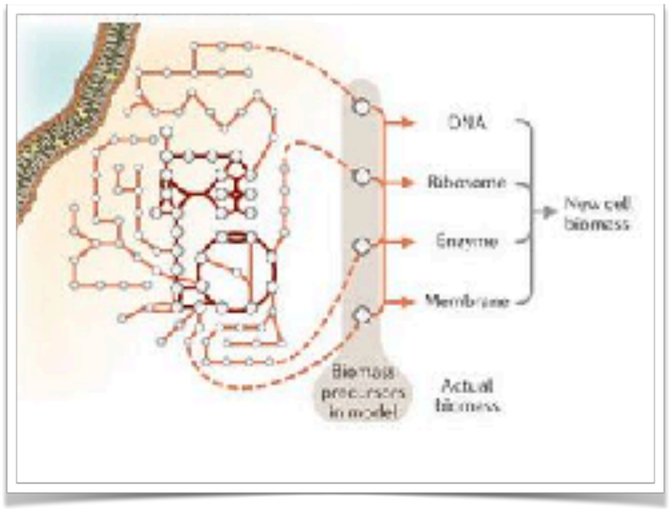
Transportable compounds

Key cellular components

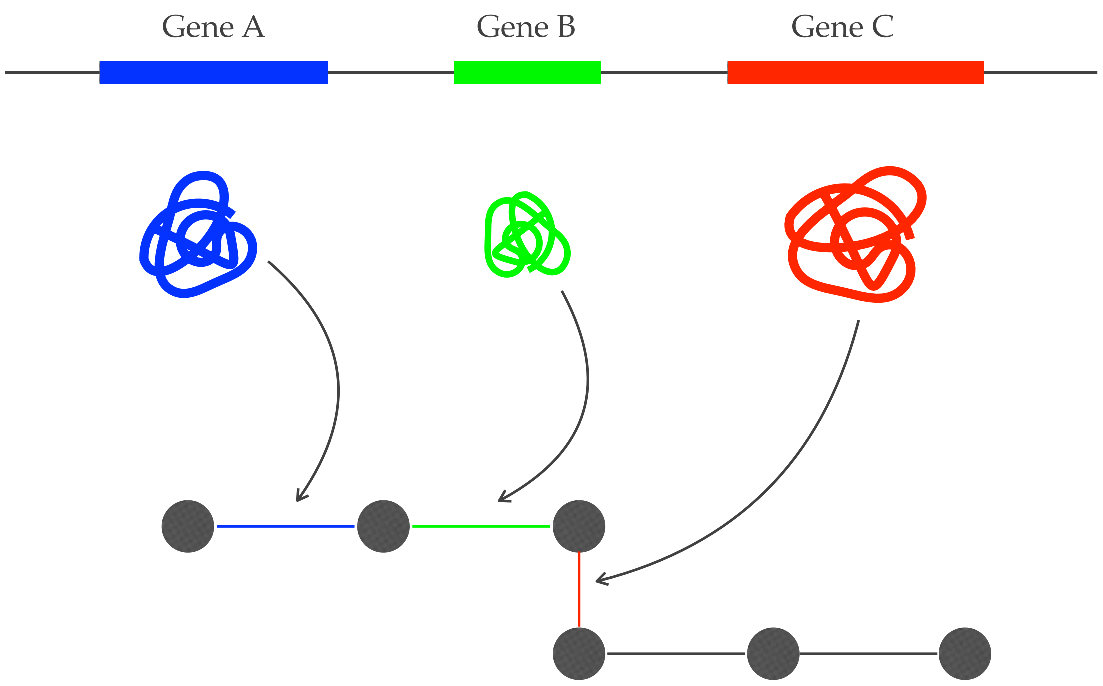


E. coli

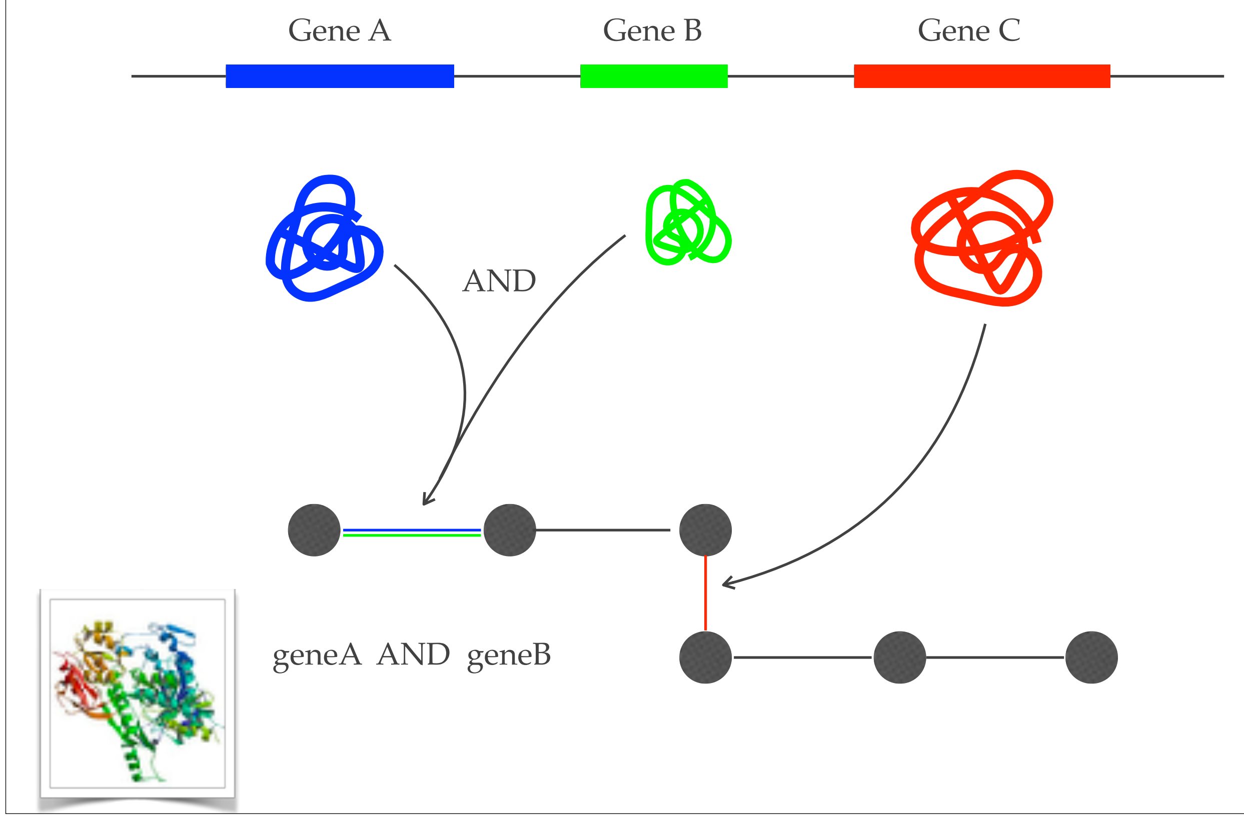
	% g/g
Protein	55,00%
RNA	20,50%
DNA	3,10%
Lipid	9,10%
LPS	3,40%
Peptidogl	2,50%
Glycogen	2,50%
Polyamines	0,40%
Metabolites, cofactors	3,50%
TOT	100,00%


DNA
Ribosome
Enzyme
Membrane
New cell biomass
Biomass precursors in model
Actual biomass

GPR - Gene - Protein Reaction association rules



GPR - Gene - Protein Reaction association rules



Gene A

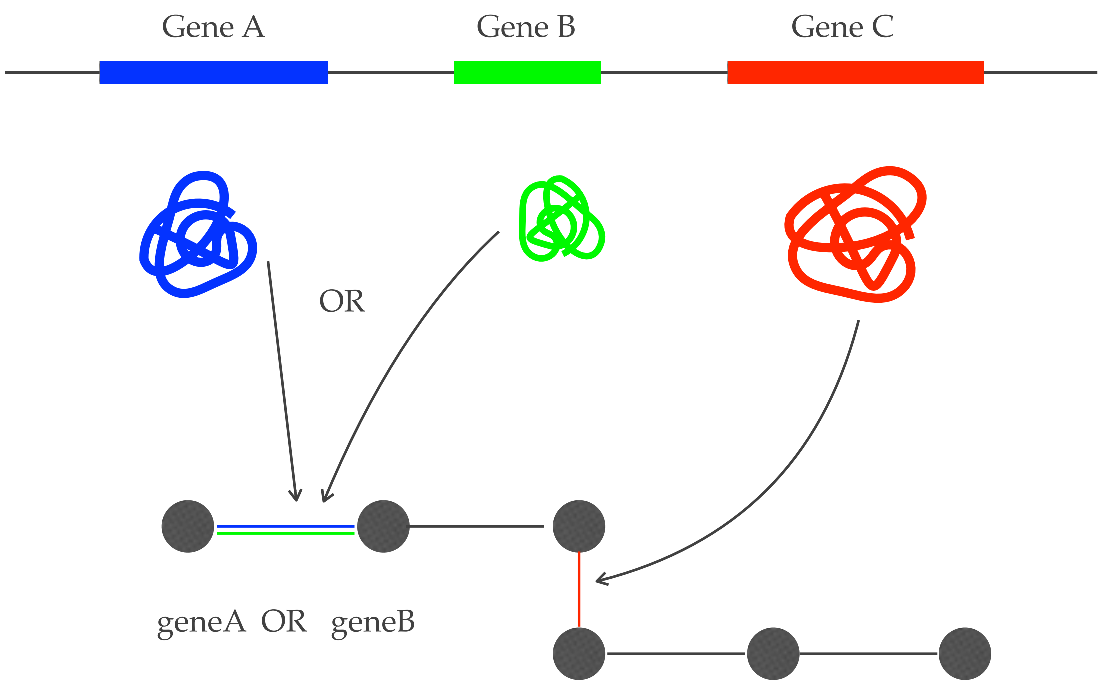
Gene B

Gene C

AND

geneA AND geneB

GPR - Gene - Protein Reaction association rules



Some bacterial reconstructions

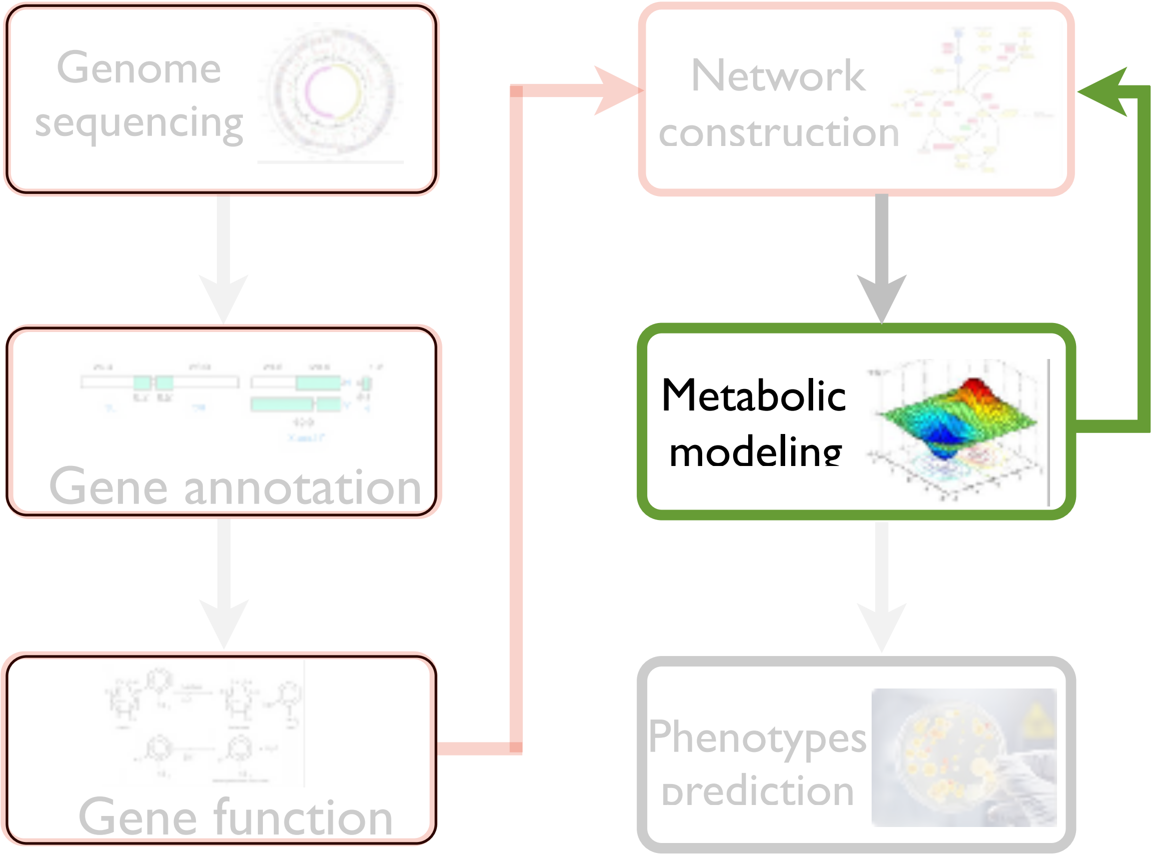
Organism	Strain	Model	Genes	Metabolites	Reactions	Reference
Escherichia coli	K12	iAF1260	1260	1039	2077	Feist et al.
Pseudomonas putida	KT2440	iNJ746	746	911	950	Nogales et al.
Salmonella typhimurium	LT2	STM_v1.0	1270	1119	2201	Thiele et al.
Klebsiella pneumoniae	MGH 78578	YL1228	1228	1658	1970	Liao et al.
Pseudoalteromonas haloplanktis	TAC125	iMF721	721	1133	1322	Fondi et al. 2014

Recon 2, human metabolic reconstruction: it accounts for 1,789 enzyme-encoding genes, 7,440 reactions and 2,626 unique metabolites distributed over eight cellular compartments (Thiele et al. 203)

What can we do with a metabolic model?

- optimized biosynthesis of compounds
- “disease” pathways
- metabolic engineering
- optimized growth of organism
- essential genes (i.e. good antimicrobial targets)
- well grounded wet lab experiments

Modelling phase



Genome sequencing

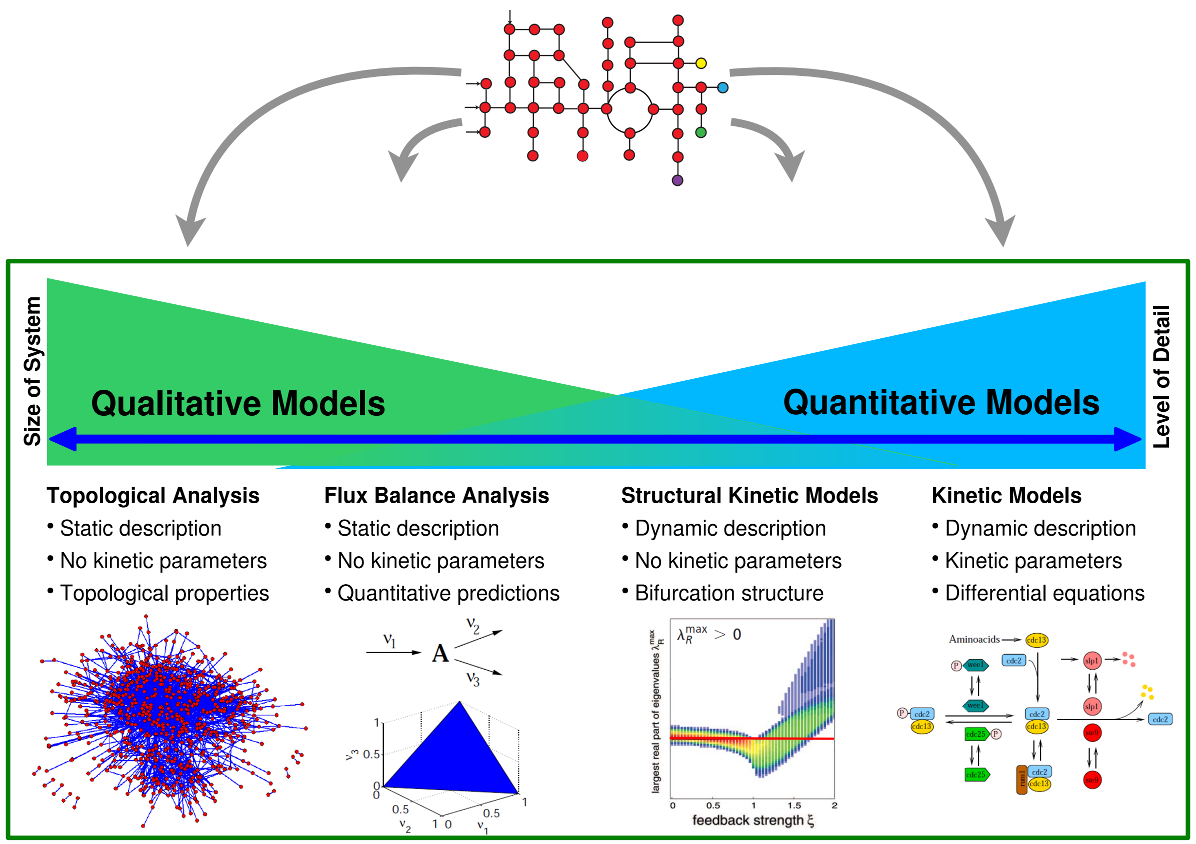
Gene annotation

Gene function

Network construction

Metabolic modeling

Phenotypes prediction



Size of System

Qualitative Models

Quantitative Models

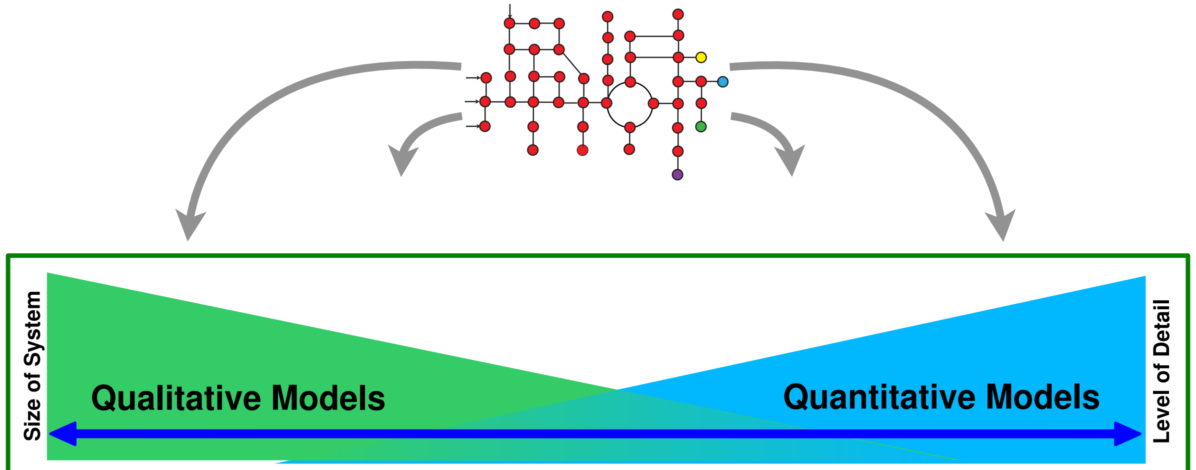
Level of Detail

Topological Analysis
- Static description
- No kinetic parameters
- Topological properties

Flux Balance Analysis
- Static description
- No kinetic parameters
- Quantitative predictions

Structural Kinetic Models
- Dynamic description
- No kinetic parameters
- Bifurcation structure

Kinetic Models
- Dynamic description
- Kinetic parameters
- Differential equations



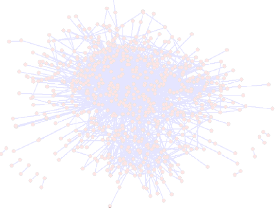
Size of System

Qualitative Models

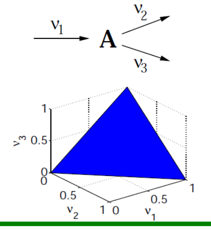
Quantitative Models

Level of Detail

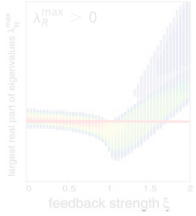
Topological Analysis
- Static description
- No kinetic parameters
- Topological properties



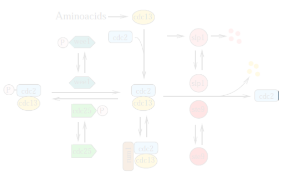
Flux Balance Analysis
- Static description
- No kinetic parameters
- Quantitative predictions

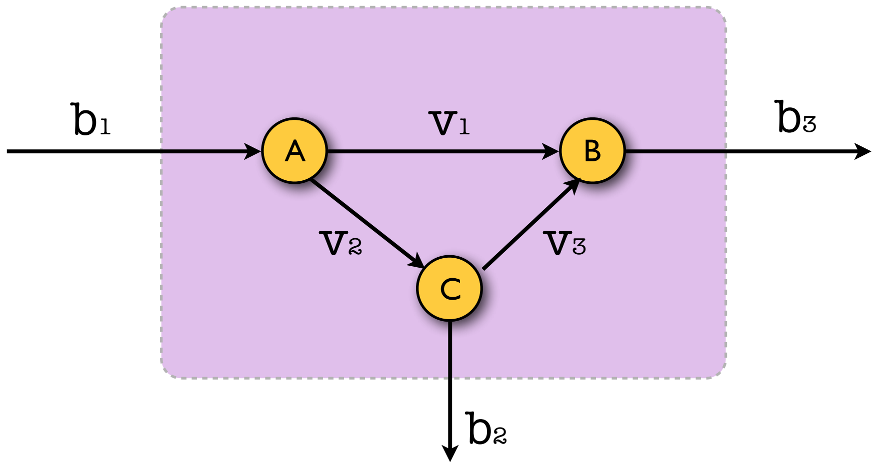


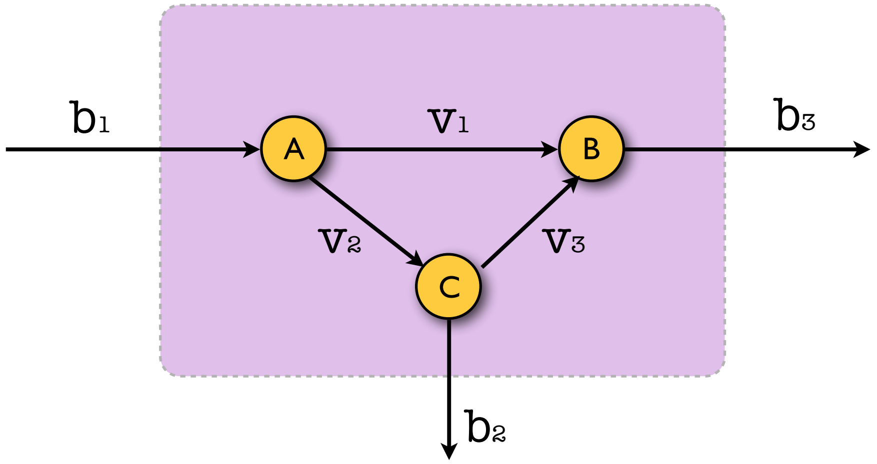
Structural Kinetic Models
- Dynamic description
- No kinetic parameters
- Bifurcation structure



Kinetic Models
- Dynamic description
- Kinetic parameters
- Differential equations




b1
A
v1
B
b3
v2
v3
C
b2

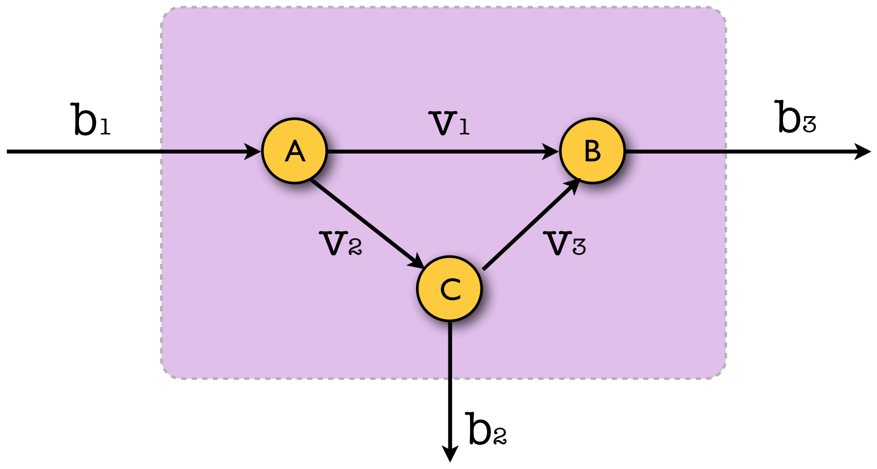


Material Balances

$dA/dt = -v_1 - v_2 + b_1$

$dB/dt = v_1 + v_3 - b_3$

$dC/dt = v_2 - v_3 - b_2$



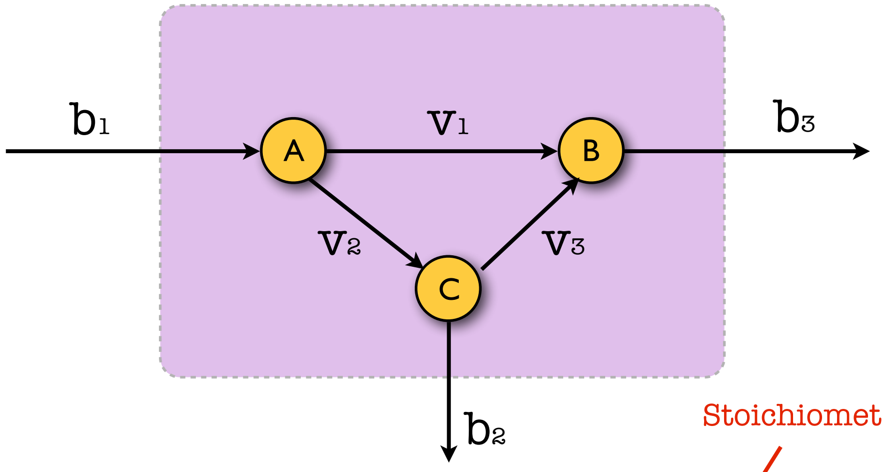
Material Balances

$dA/dt = -v_1 - v_2 + b_1$

$dB/dt = v_1 + v_3 - b_3$

$dC/dt = v_2 - v_3 - b_2$

$$\begin{bmatrix} dA/dt \\ dB/dt \\ dC/dt \end{bmatrix} = \begin{bmatrix} -1 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & -1 & 0 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$



Stoichiometric matrix (S)

Material Balances

$dA/dt = -v_1 - v_2 + b_1$

$dB/dt = v_1 + v_3 - b_3$

$dC/dt = v_2 - v_3 - b_2$

$$\begin{bmatrix} dA/dt \\ dB/dt \\ dC/dt \end{bmatrix} = \begin{bmatrix} -1 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & -1 & 0 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

Stoichiometric matrix alone does not provide sufficient information to uniquely determine all fluxes (i.e. the system is underdetermined), additional constraints are needed to determine meaningful flux distributions.

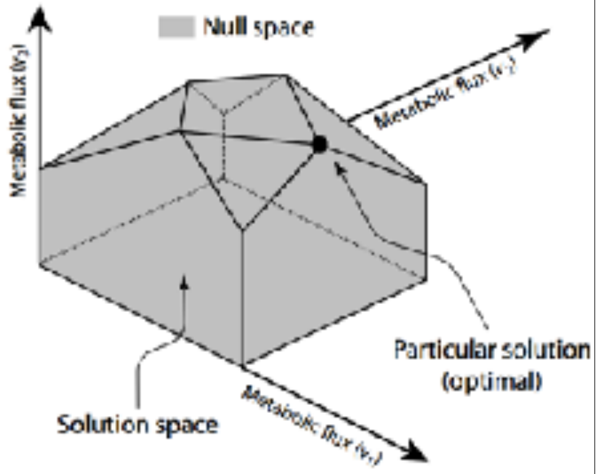
Introduction of constraints

(Constraint-based metabolic modelling)

1. Steady state assumption

2. Upper and lower bounds

3. Objective function

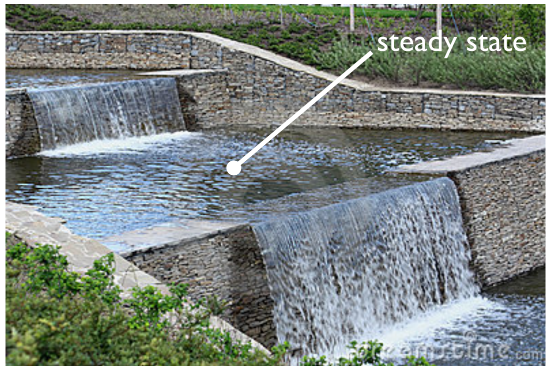


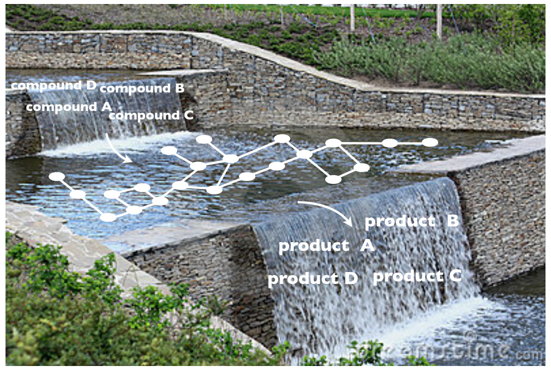
Introduction of constraints

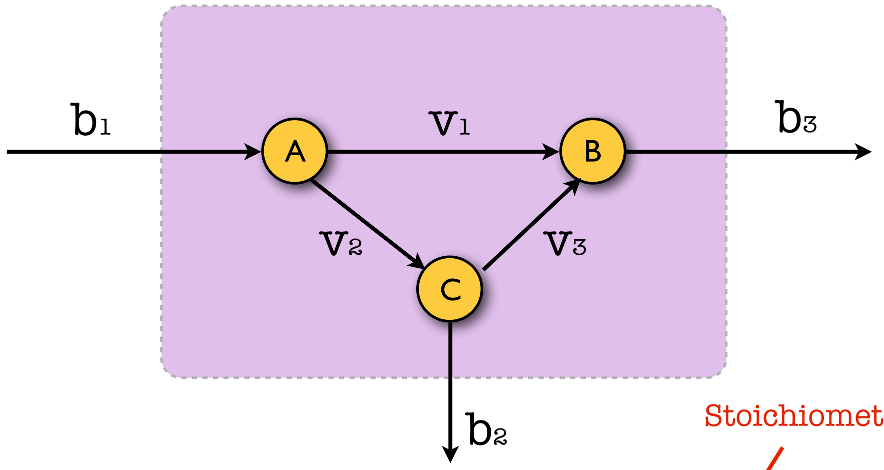
(Constraint-based metabolic modelling)

1. Steady state assumption
2. Upper and lower bounds
3. Objective function

1. Metabolism operates on a much faster time-scale than regulatory or cell division events. It is thus often reasonable to assume that metabolic dynamics have reached a quasi-or pseudo-steady state, where metabolite concentrations do not change. This leads to the metabolite balancing equation







Stoichiometric matrix (S)

Material Balances

$dA/dt = -v_1 - v_2 + b_1 = 0$

$dB/dt = v_1 + v_3 - b_3 = 0$

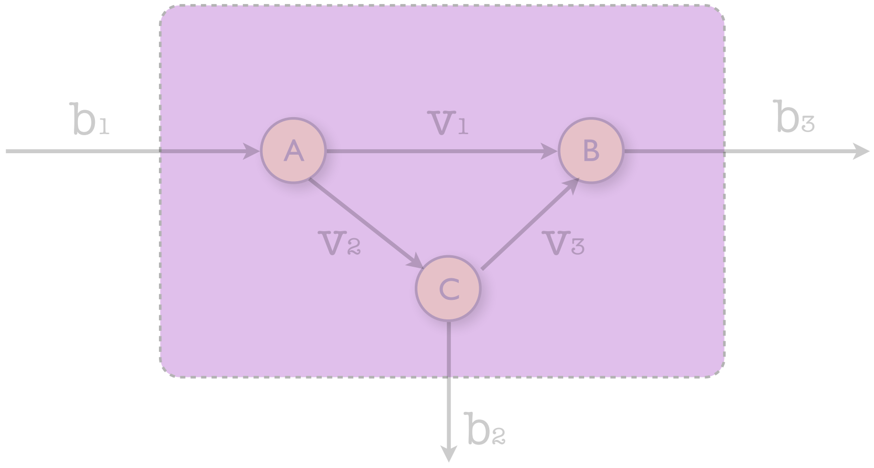
$dC/dt = v_2 - v_3 - b_2 = 0$

$$\begin{bmatrix} dA/dt \\ dB/dt \\ dC/dt \end{bmatrix} = \begin{bmatrix} -1 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & -1 & 0 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

Introduction of constraints

(Constraint-based metabolic modelling)

1. Steady state assumption
2. Upper and lower bounds
3. Objective function



Material Balances

$dA/dt = -v_1 - v_2 + b_1$

$dB/dt = v_2 - v_3 - b_3$

$dC/dt = v_1 + v_3 - b_2$

$$\begin{bmatrix} dA/dt \\ dB/dt \\ dC/dt \end{bmatrix} = \begin{bmatrix} -1 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

upper and lower bounds of fluxes through reactions

(all fluxes are expressed as mmol * g_{CDW}^{-1} * h^{-1} except for the flux through biomass assembly reaction, expressed as h^{-1})

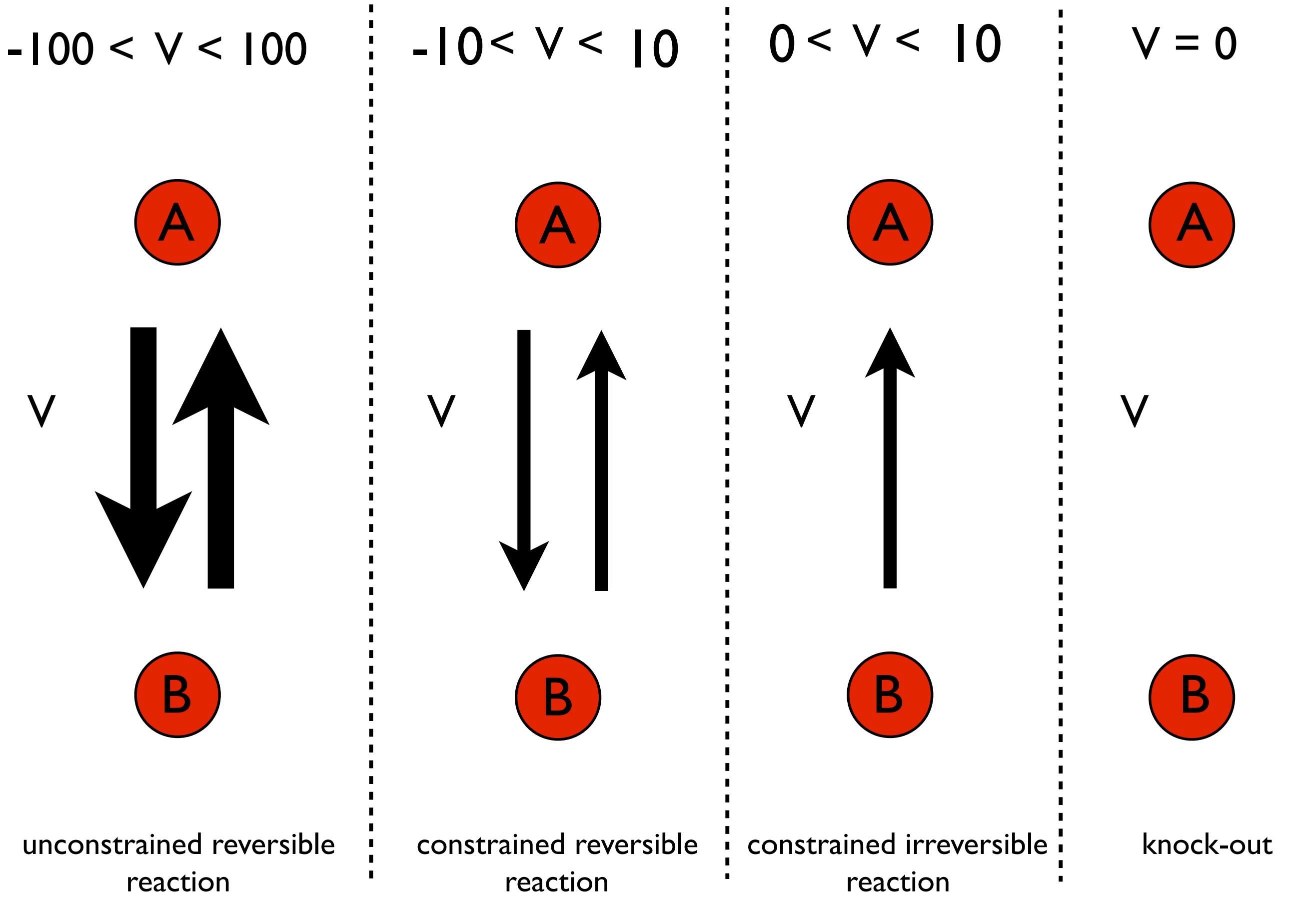
$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$$-1000 < v_1 < 1000$$

$$0 < v_2 < +$$

$$b_3 = 0$$

fluxes through exchange reactions
simulate growth medium composition



$-100 < V < 100$

unconstrained reversible reaction

$-10 < V < 10$

constrained reversible reaction

$0 < V < 10$

constrained irreversible reaction

$V = 0$

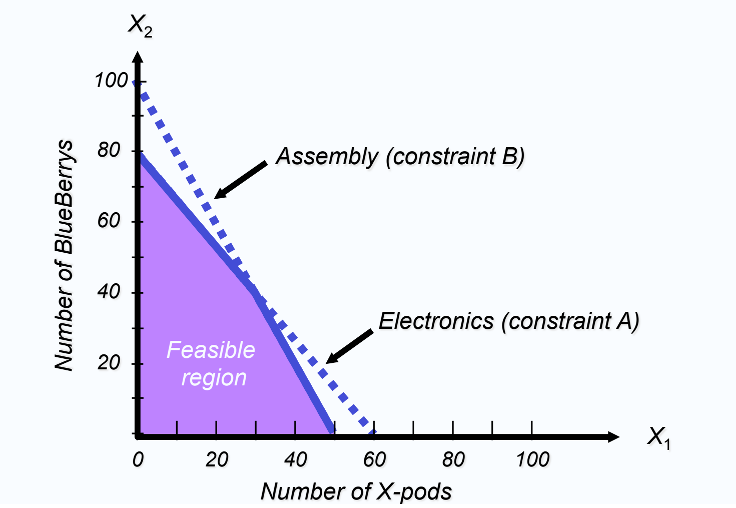
knock-out

all fluxes expressed in mmol/g*h

Introduction of constraints
(Constraint-based metabolic modelling)

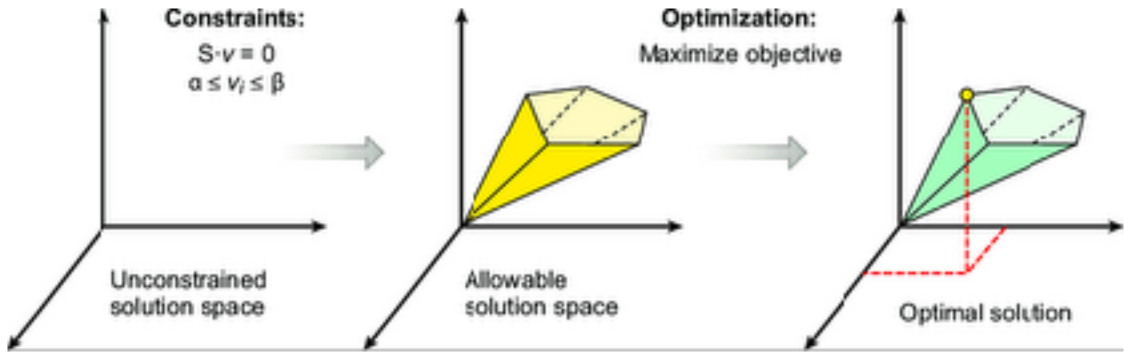
1. Steady state assumption
2. Upper and lower bounds
3. Objective function

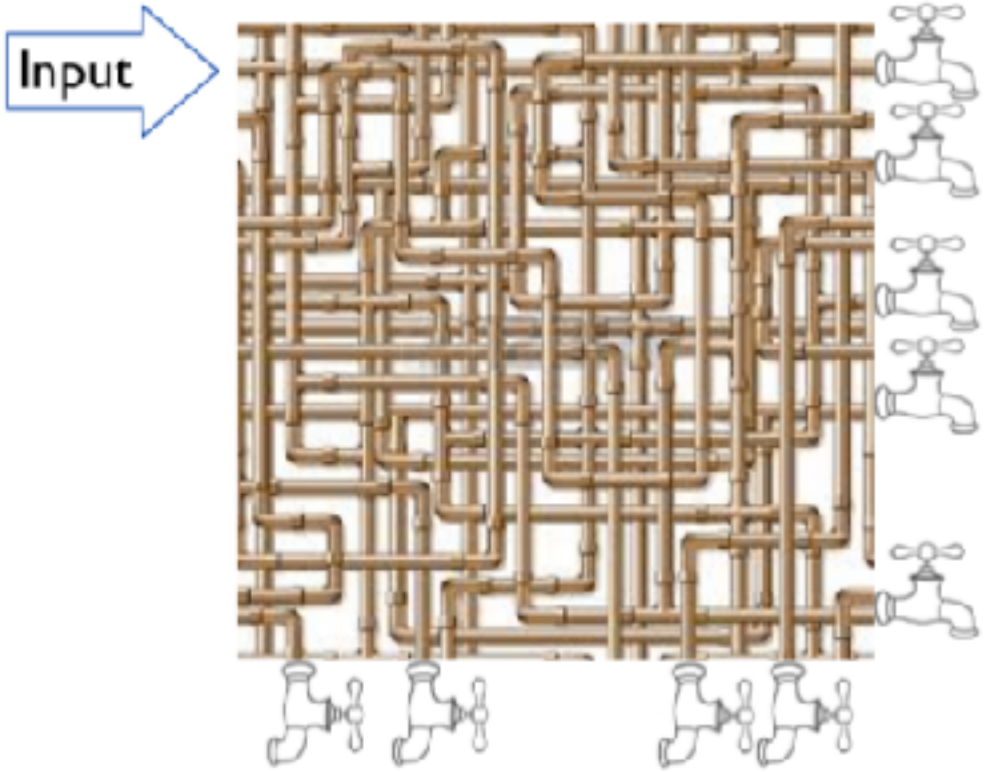
An objective function is necessary to compute an optimal network state and resulting flux distribution (unique or nonuni- que) in a constraint-based reconstruction as the solution space is often very large for genome-scale networks.



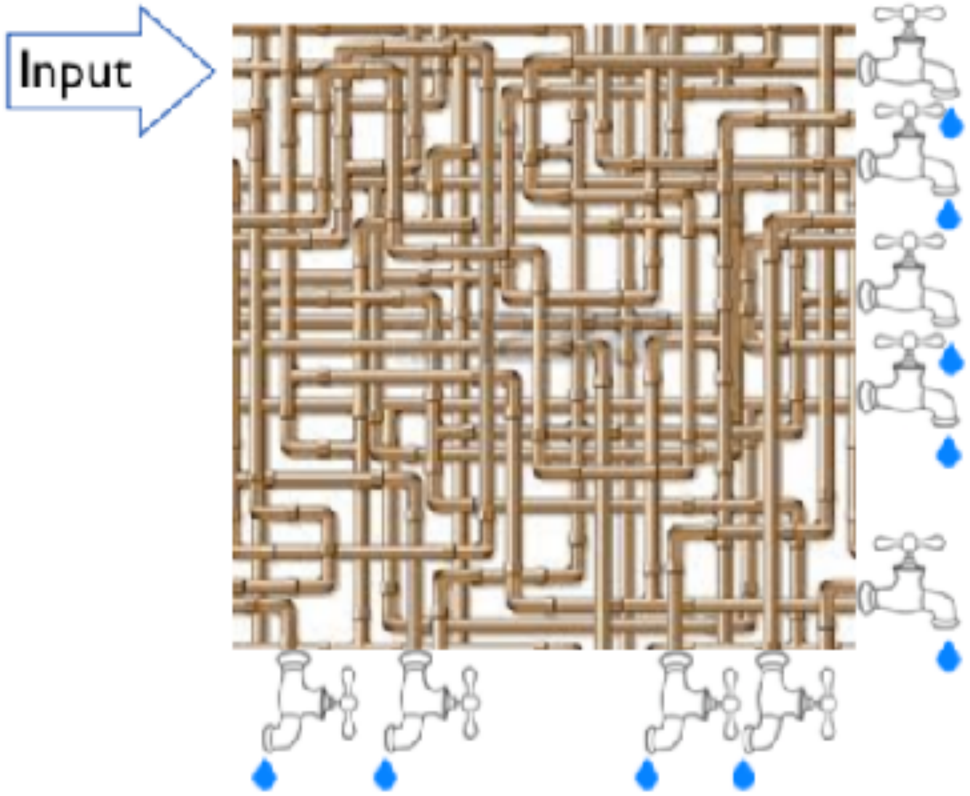
3. it is possible to solve for a flux distribution by assuming that the under-determined metabolic network is optimized with respect to a certain objective (Z)

$$Z = \sum_{i=1}^{r} c_i v_i \longrightarrow \max$$

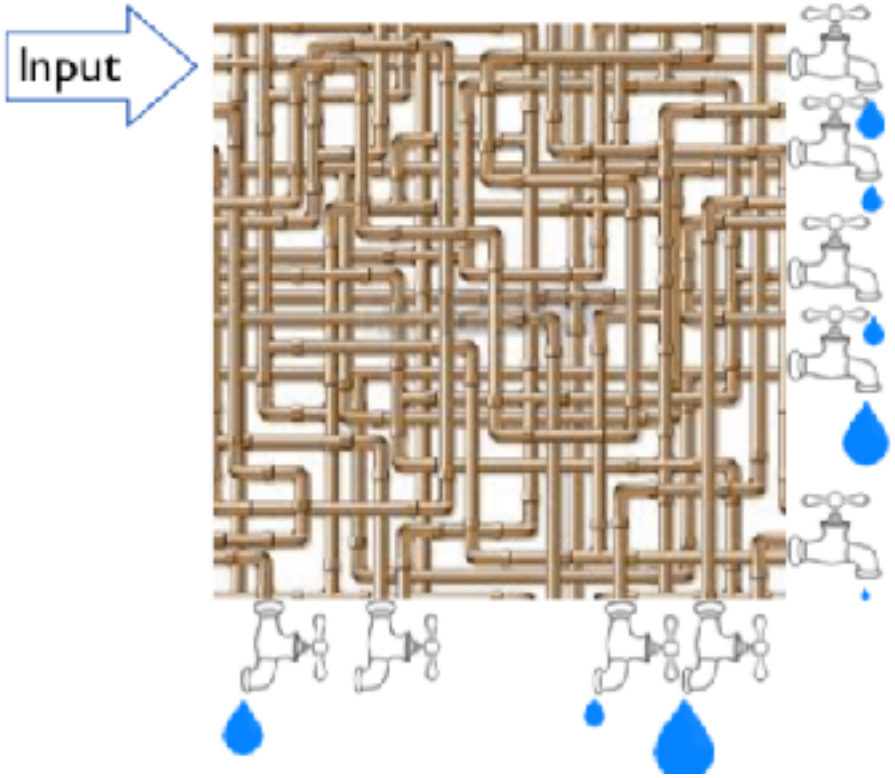

Constraints:
$S \cdot v = 0$
$\alpha \le v_i \le \beta$
Unconstrained solution space
Allowable solution space
Optimization:
Maximize objective
Optimal solution



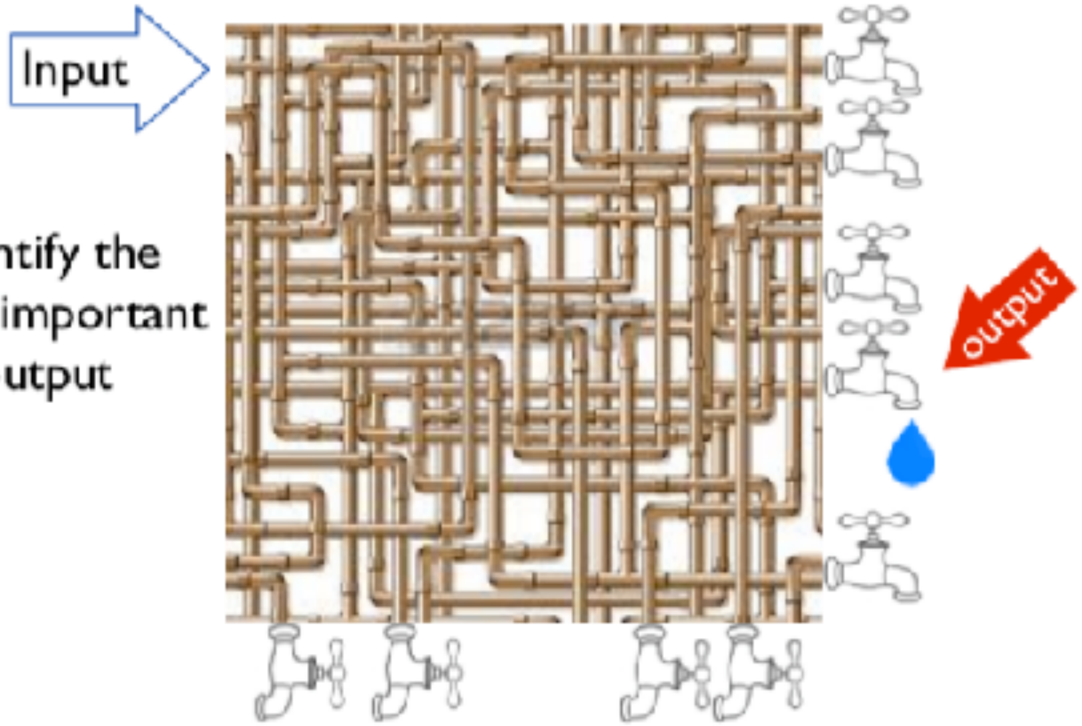
Input



Input

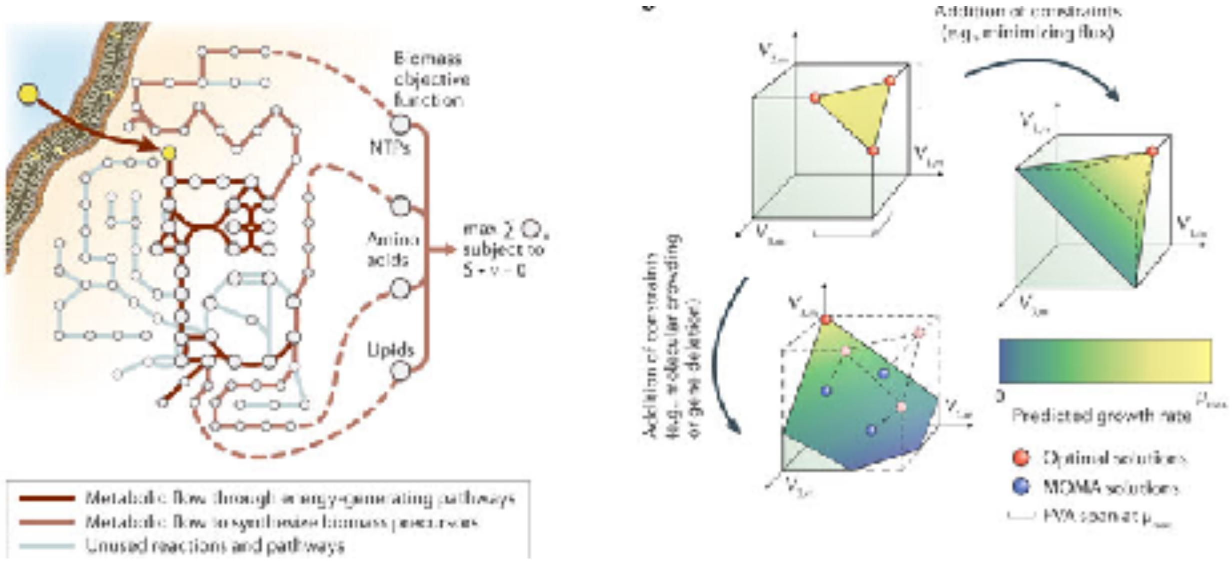


Input

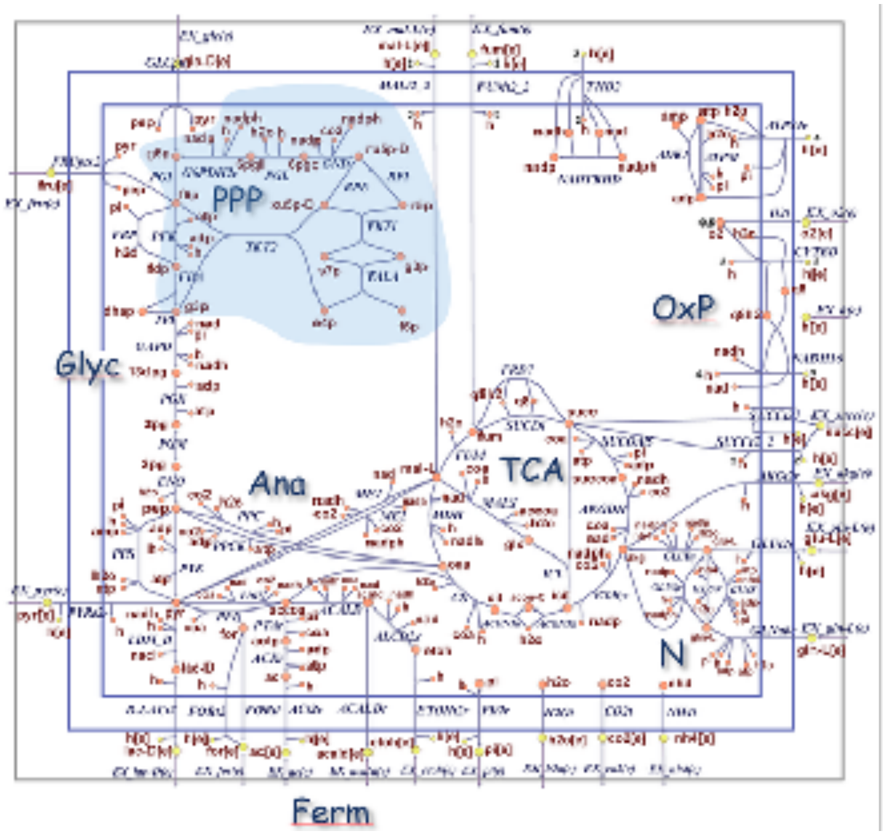


Identify the most important output

The biomass objective function describes an evolutionary pressure for microbial growth, and describes the metabolic demands to make the basic metabolite building blocks for all cellular components

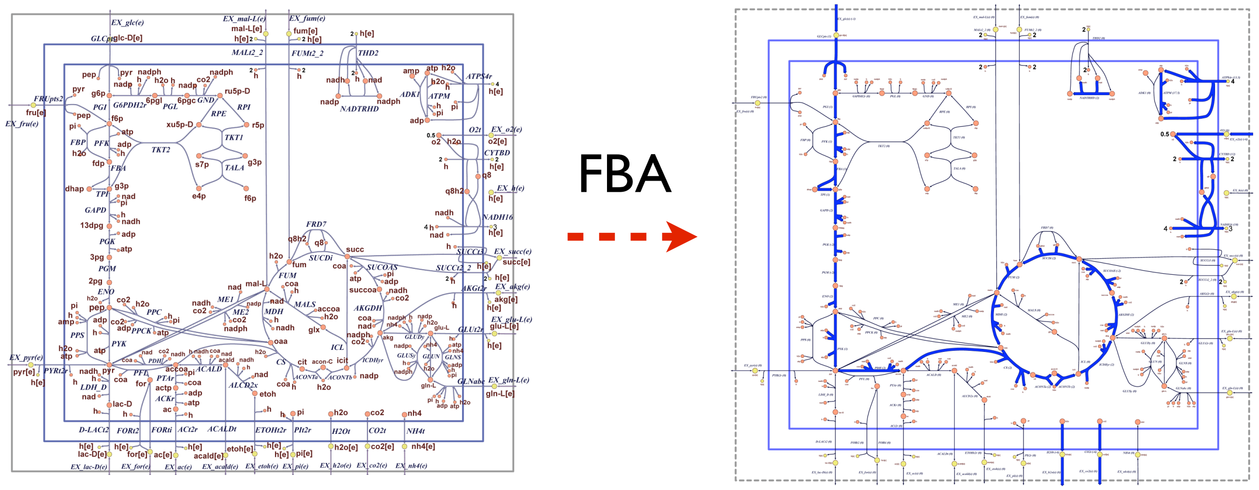


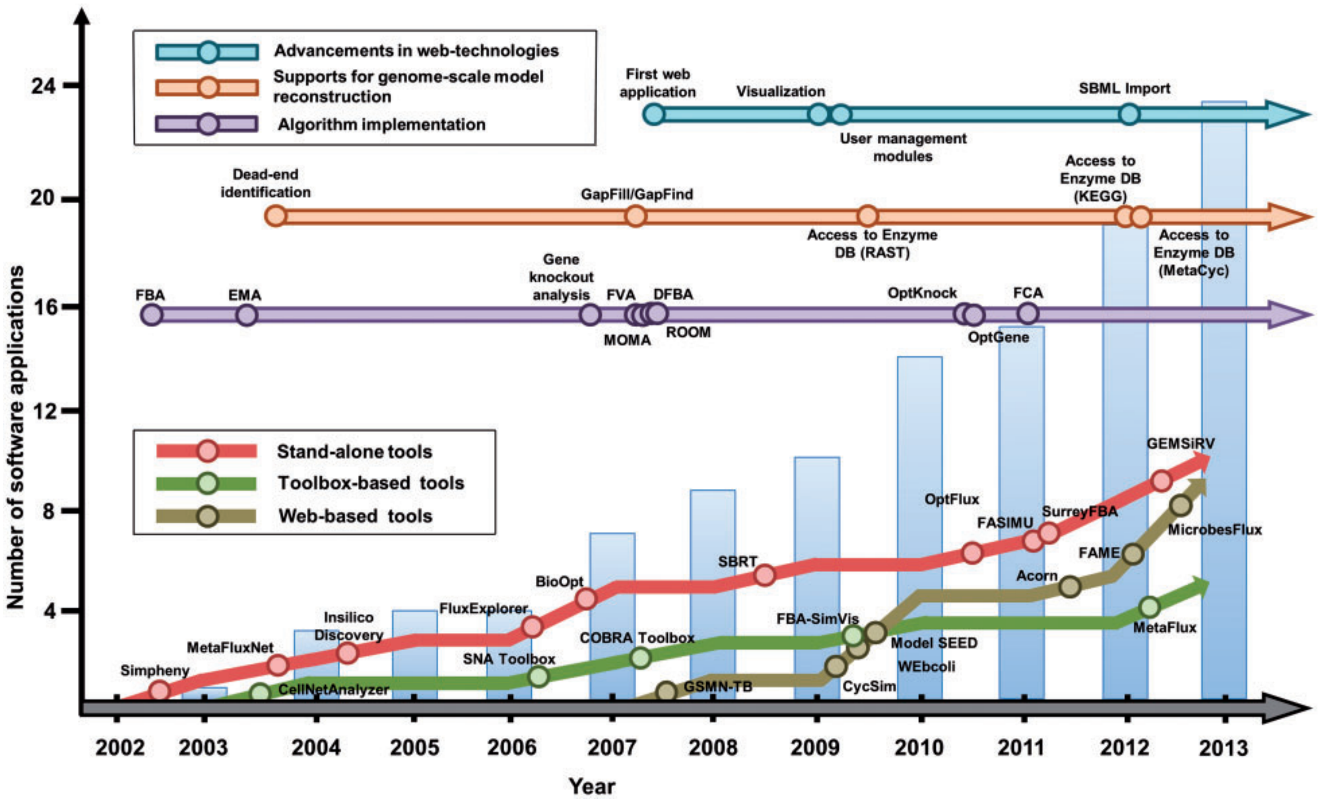
Reconstruction done, constraints applied, so what?



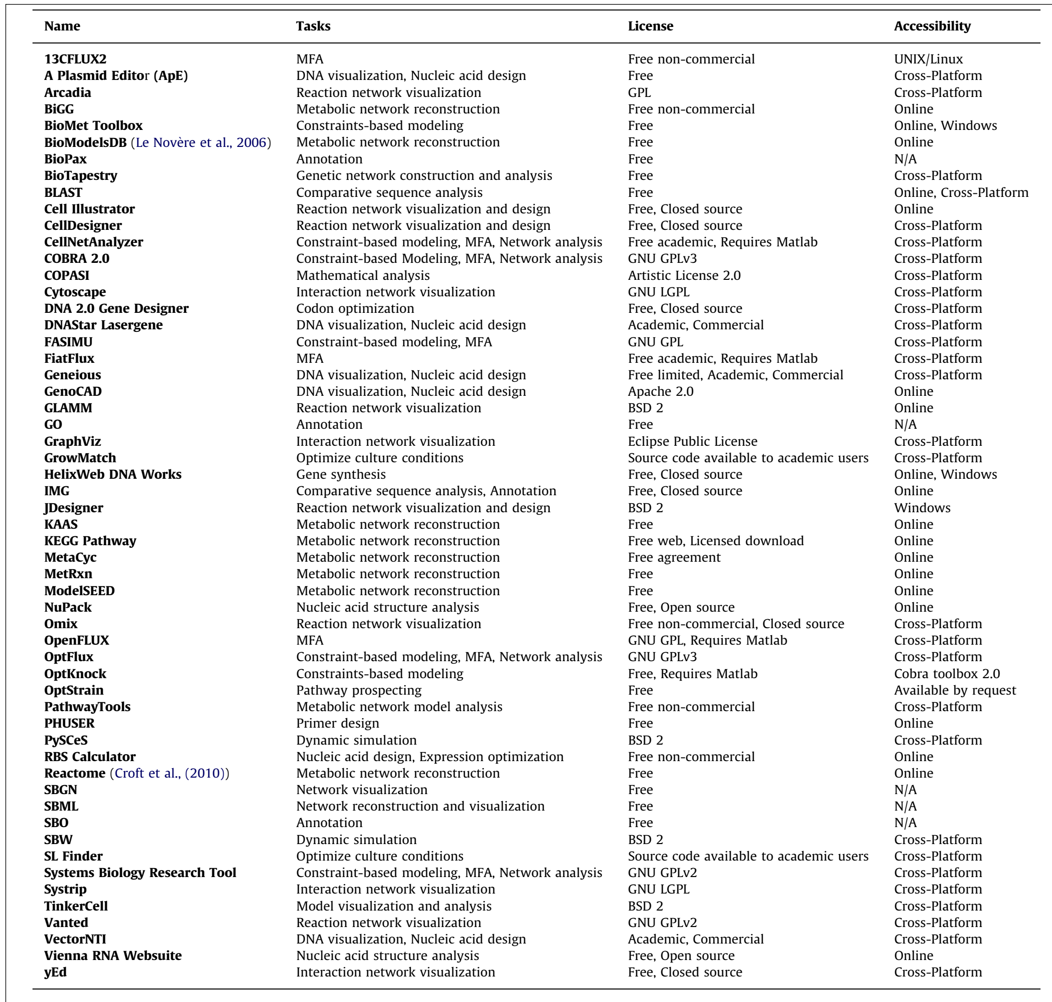
The model can now be resolved by means of linear programming (LP; also called linear optimization), a method to achieve the best outcome in a mathematical mode whose requirements are represented by linear equations. Used also for:

- economic analysis (profit maximization)
- pairwise sequence alignment
- constraints-based metabolic modelling





Lakshmanan et al. 2012



Name	Tasks	License	Accessibility
13CFLUX2	MFA	Free non-commercial	UNIX/Linux
A Plasmid Editor (ApE)	DNA visualization, Nucleic acid design	Free	Cross-Platform
Arcadia	Reaction network visualization	GPL	Cross-Platform
BiGG	Metabolic network reconstruction	Free non-commercial	Online
BioMet Toolbox	Constraints-based modeling	Free	Online, Windows
BioModelsDB (Le Novère et al., 2006)	Metabolic network reconstruction	Free	Online
BioPax	Annotation	Free	N/A
BioTapestry	Genetic network construction and analysis	Free	Cross-Platform
BLAST	Comparative sequence analysis	Free	Online, Cross-Platform
Cell Illustrator	Reaction network visualization and design	Free, Closed source	Online
CellDesigner	Reaction network visualization and design	Free, Closed source	Cross-Platform
CellNetAnalyzer	Constraint-based modeling, MFA, Network analysis	Free academic, Requires Matlab	Cross-Platform
COBRA 2.0	Constraint-based Modeling, MFA, Network analysis	GNU GPLv3	Cross-Platform
COPASI	Mathematical analysis	Artistic License 2.0	Cross-Platform
Cytoscape	Interaction network visualization	GNU LGPL	Cross-Platform
DNA 2.0 Gene Designer	Codon optimization	Free, Closed source	Cross-Platform
DNAStar Lasergene	DNA visualization, Nucleic acid design	Academic, Commercial	Cross-Platform
FASIMU	Constraint-based modeling, MFA	GNU GPL	Cross-Platform
FiatFlux	MFA	Free academic, Requires Matlab	Cross-Platform
Geneious	DNA visualization, Nucleic acid design	Free limited, Academic, Commercial	Cross-Platform
GenoCAD	DNA visualization, Nucleic acid design	Apache 2.0	Online
GLAMM	Reaction network visualization	BSD 2	Online
GO	Annotation	Free	N/A
GraphViz	Interaction network visualization	Eclipse Public License	Cross-Platform
GrowMatch	Optimize culture conditions	Source code available to academic users	Cross-Platform
HelixWeb DNA Works	Gene synthesis	Free, Closed source	Online, Windows
IMG	Comparative sequence analysis, Annotation	Free, Closed source	Online
JDesigner	Reaction network visualization and design	BSD 2	Windows
KAAS	Metabolic network reconstruction	Free	Online
KEGG Pathway	Metabolic network reconstruction	Free web, Licensed download	Online
MetaCyc	Metabolic network reconstruction	Free agreement	Online
MetRxn	Metabolic network reconstruction	Free	Online
ModelSEED	Metabolic network reconstruction	Free	Online
NuPack	Nucleic acid structure analysis	Free, Open source	Online
Omix	Reaction network visualization	Free non-commercial, Closed source	Cross-Platform
OpenFLUX	MFA	GNU GPL, Requires Matlab	Cross-Platform
OptFlux	Constraint-based modeling, MFA, Network analysis	GNU GPLv3	Cross-Platform
OptKnock	Constraints-based modeling	Free, Requires Matlab	Cobra toolbox 2.0
OptStrain	Pathway prospecting	Free	Available by request
PathwayTools	Metabolic network model analysis	Free non-commercial	Cross-Platform
PHUSER	Primer design	Free	Online
PySCeS	Dynamic simulation	BSD 2	Cross-Platform
RBS Calculator	Nucleic acid design, Expression optimization	Free non-commercial	Online
Reactome (Croft et al., (2010))	Metabolic network reconstruction	Free	Online
SBGN	Network visualization	Free	N/A
SBML	Network reconstruction and visualization	Free	N/A
SBO	Annotation	Free	N/A
SBW	Dynamic simulation	BSD 2	Cross-Platform
SL Finder	Optimize culture conditions	Source code available to academic users	Cross-Platform
Systems Biology Research Tool	Constraint-based modeling, MFA, Network analysis	GNU GPLv2	Cross-Platform
Systrip	Interaction network visualization	GNU LGPL	Cross-Platform
TinkerCell	Model visualization and analysis	BSD 2	Cross-Platform
Vanted	Reaction network visualization	GNU GPLv2	Cross-Platform
VectorNTI	DNA visualization, Nucleic acid design	Academic, Commercial	Cross-Platform
Vienna RNA Websuite	Nucleic acid structure analysis	Free, Open source	Online
yEd	Interaction network visualization	Free, Closed source	Cross-Platform

Copeland et al. 2012

http://opencobra.sourceforge.net

MATLAB

python

Copeland et al. 2012

Model validation

1. Is the model predicting a correct growth rate?

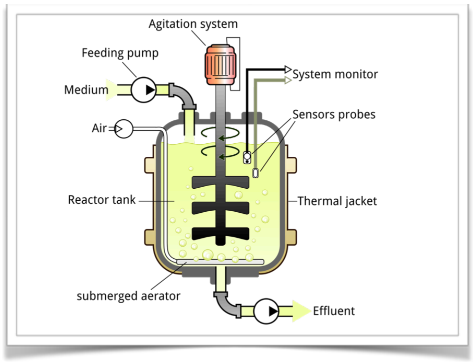
2. Is the model capable of metabolizing the compounds the organism is actually capable to *in vivo*?

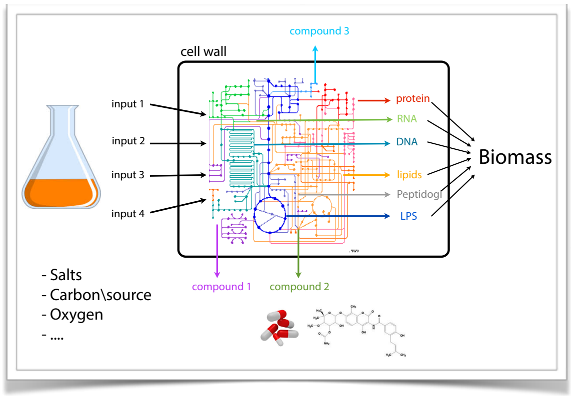
Model validation

1. Is the model predicting a correct growth rate?

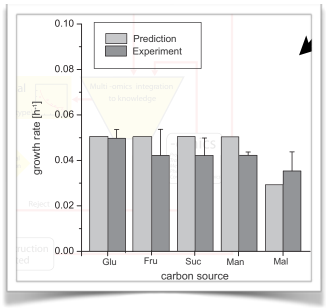
2. Is the model capable of metabolizing the compounds the organism is actually capable to *in vivo*?

Comparison withe experimental physiological data





μ

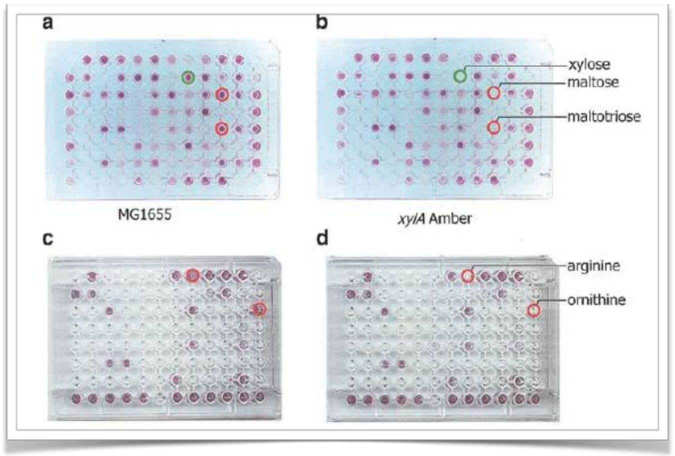


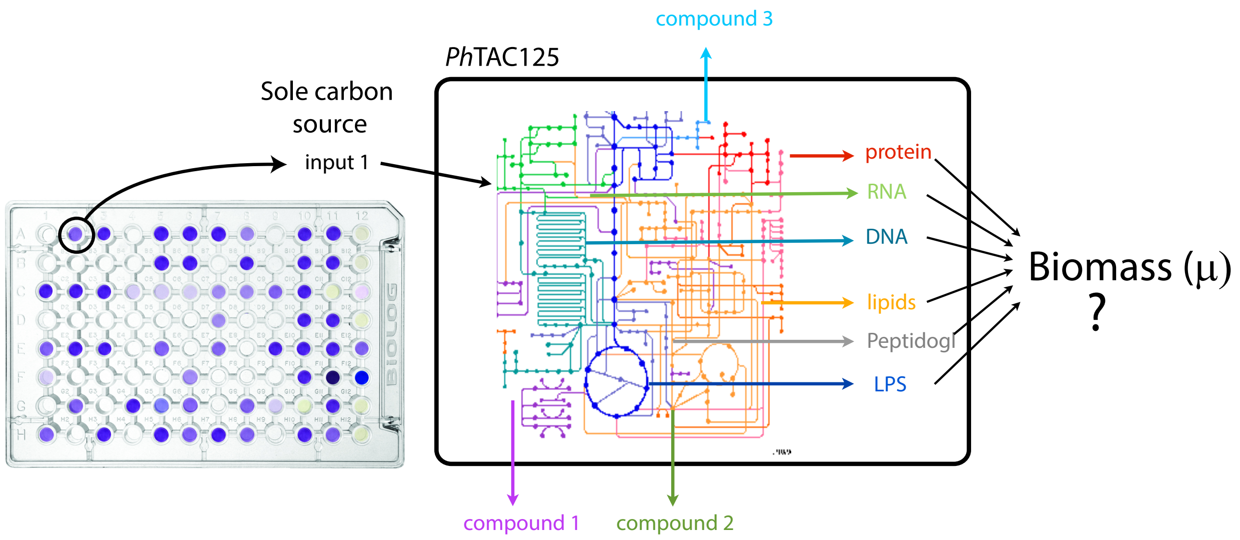
Model validation

1. Is the model predicting a correct growth rate?

2. Is the model capable of metabolizing the compounds the organism is actually capable to *in vivo*?

Large scale phenotypic assays





Sole carbon source

input 1

*Ph*TAC125

compound 3

protein

RNA

DNA

lipids

Peptidogl

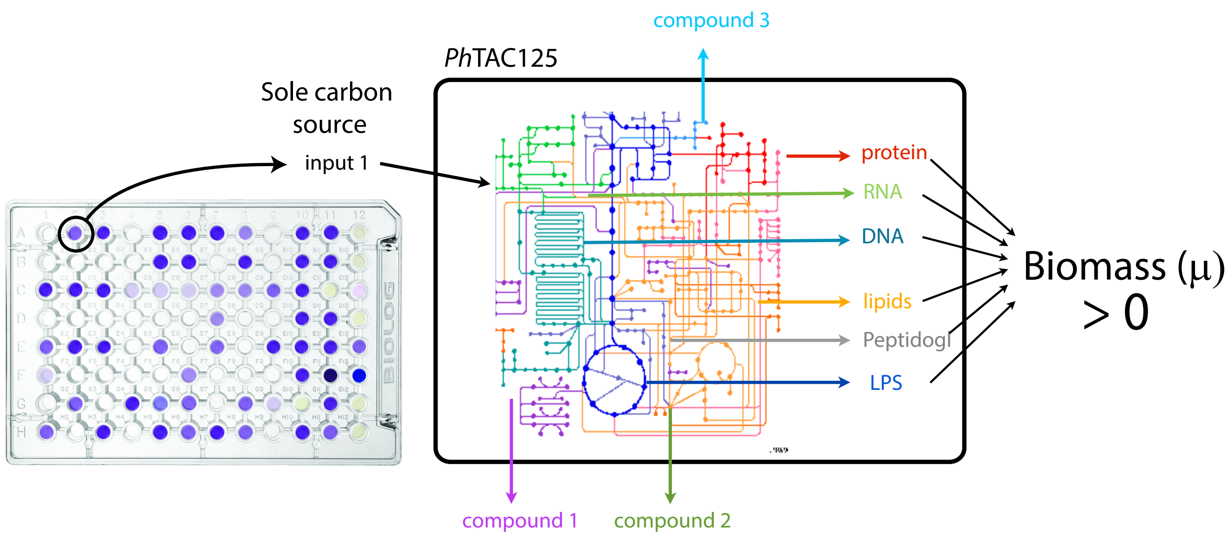
LPS

Biomass (μ)

?

compound 1

compound 2



Sole carbon source

input 1

*Ph*TAC125

compound 3

protein

RNA

DNA

lipids

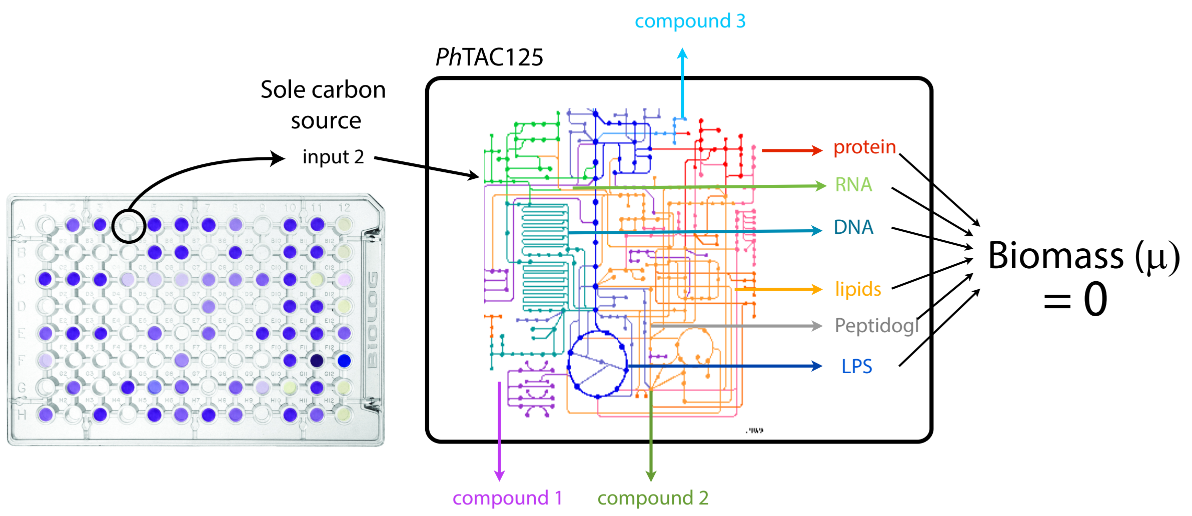
Peptidogl

LPS

compound 1

compound 2

Biomass (μ)
> 0



Sole carbon source

input 2

*Ph*TAC125

compound 3

protein

RNA

DNA

lipids

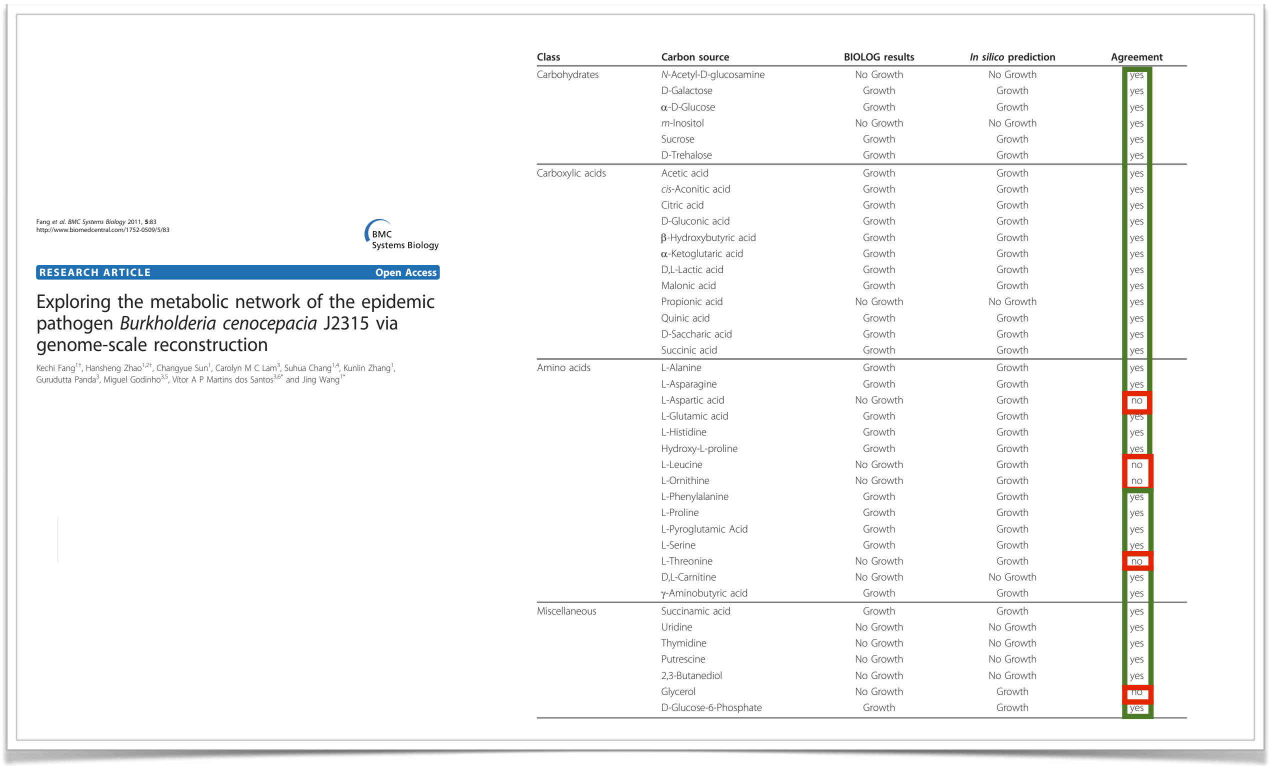
Peptidogl

LPS

Biomass (μ)
= 0

compound 1

compound 2



Fang *et al. BMC Systems Biology* 2011, **5**:83
http://www.biomedcentral.com/1752-0509/5/83

BMC
Systems Biology

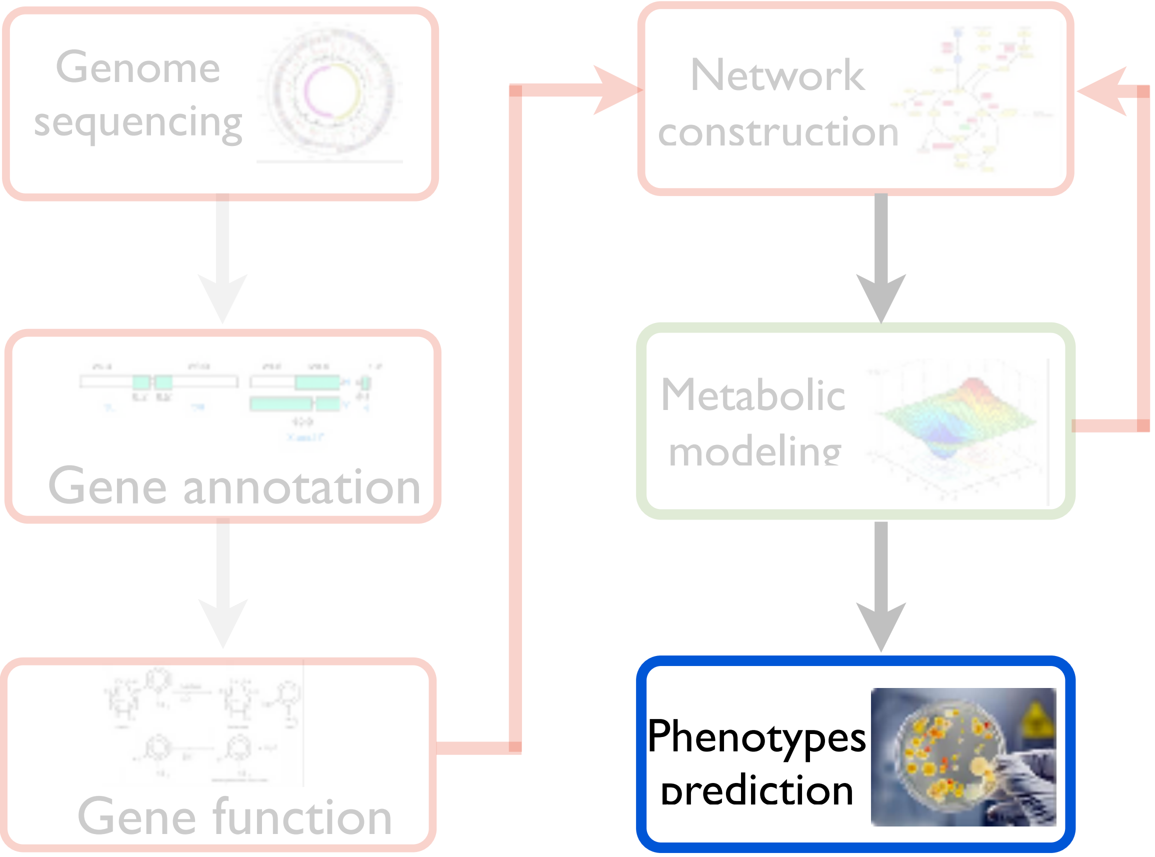
RESEARCH ARTICLE **Open Access**

Exploring the metabolic network of the epidemic pathogen *Burkholderia cenocepacia* J2315 via genome-scale reconstruction

Kechi Fang[1†], Hansheng Zhao[1,2†], Changyue Sun[1], Carolyn M C Lam[3], Suhua Chang[1,4], Kunlin Zhang[1], Gurudutta Panda[3], Miguel Godinho[3,5], Vítor A P Martins dos Santos[3,6*] and Jing Wang[1*]

Class	Carbon source	BIOLOG results	*In silico* prediction	Agreement
Carbohydrates	*N*-Acetyl-D-glucosamine	No Growth	No Growth	yes
	D-Galactose	Growth	Growth	yes
	α-D-Glucose	Growth	Growth	yes
	m-Inositol	No Growth	No Growth	yes
	Sucrose	Growth	Growth	yes
	D-Trehalose	Growth	Growth	yes
Carboxylic acids	Acetic acid	Growth	Growth	yes
	cis-Aconitic acid	Growth	Growth	yes
	Citric acid	Growth	Growth	yes
	D-Gluconic acid	Growth	Growth	yes
	β-Hydroxybutyric acid	Growth	Growth	yes
	α-Ketoglutaric acid	Growth	Growth	yes
	D,L-Lactic acid	Growth	Growth	yes
	Malonic acid	Growth	Growth	yes
	Propionic acid	No Growth	No Growth	yes
	Quinic acid	Growth	Growth	yes
	D-Saccharic acid	Growth	Growth	yes
	Succinic acid	Growth	Growth	yes
Amino acids	L-Alanine	Growth	Growth	yes
	L-Asparagine	Growth	Growth	yes
	L-Aspartic acid	No Growth	Growth	no
	L-Glutamic acid	Growth	Growth	yes
	L-Histidine	Growth	Growth	yes
	Hydroxy-L-proline	Growth	Growth	yes
	L-Leucine	No Growth	Growth	no
	L-Ornithine	No Growth	Growth	no
	L-Phenylalanine	Growth	Growth	yes
	L-Proline	Growth	Growth	yes
	L-Pyroglutamic Acid	Growth	Growth	yes
	L-Serine	Growth	Growth	yes
	L-Threonine	No Growth	Growth	no
	D,L-Carnitine	No Growth	No Growth	yes
	γ-Aminobutyric acid	Growth	Growth	yes
Miscellaneous	Succinamic acid	Growth	Growth	yes
	Uridine	No Growth	No Growth	yes
	Thymidine	No Growth	No Growth	yes
	Putrescine	No Growth	No Growth	yes
	2,3-Butanediol	No Growth	No Growth	yes
	Glycerol	No Growth	Growth	no
	D-Glucose-6-Phosphate	Growth	Growth	yes

Using the model



Genome sequencing

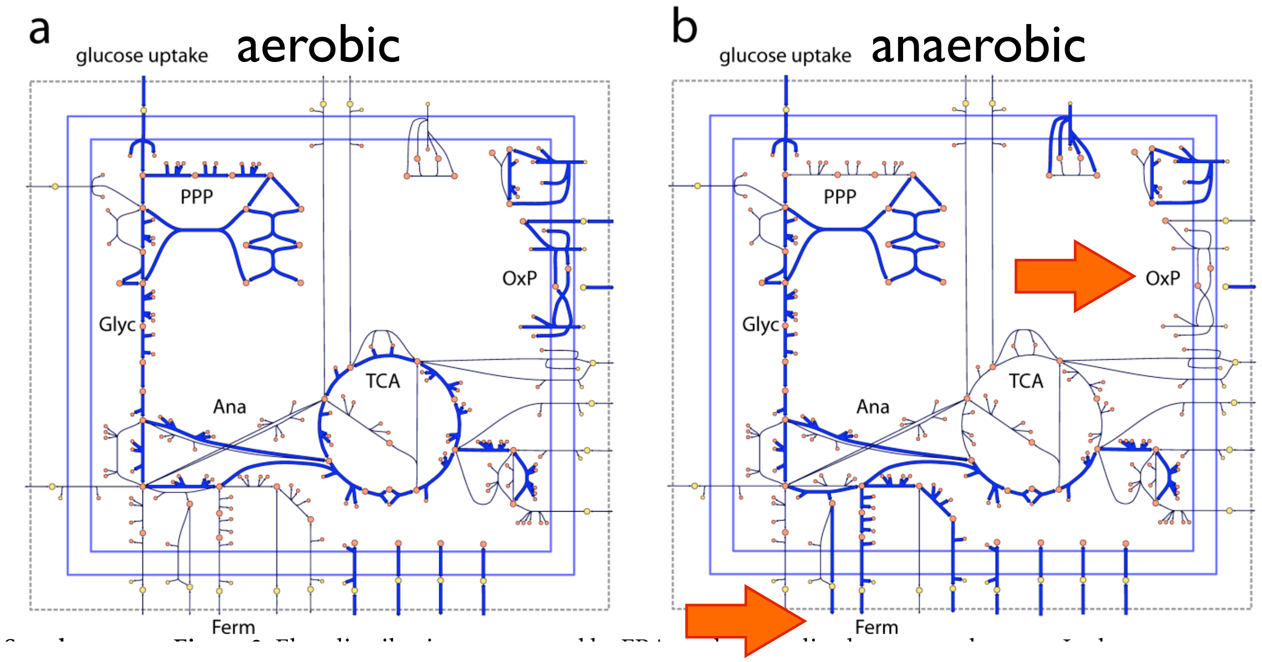
Gene annotation

Gene function

Network construction

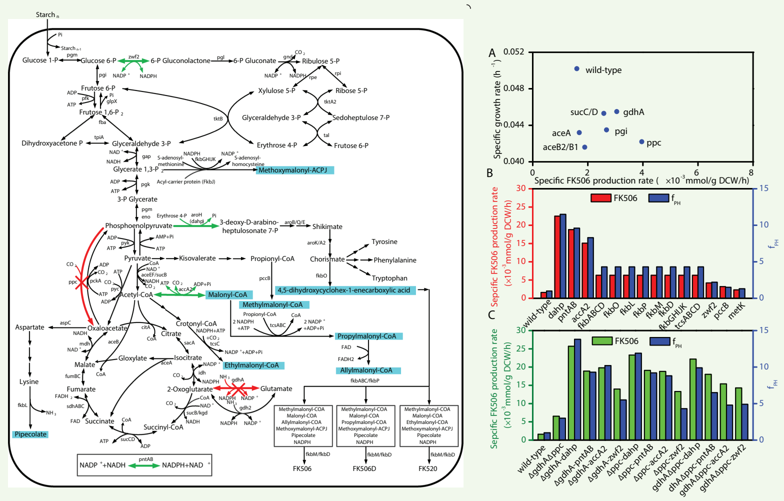
Metabolic modeling

Phenotypes prediction

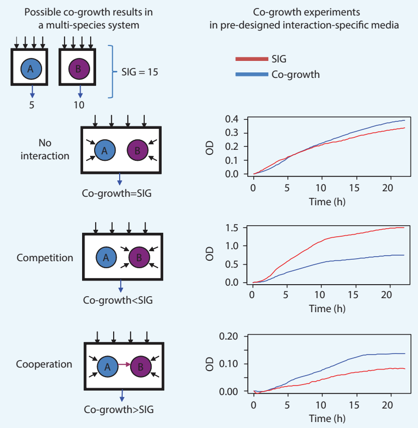


The flux distribution shows that oxidative phosphorylation is not used in these conditions, and that acetate, formate, and ethanol are produced by fermentation pathways

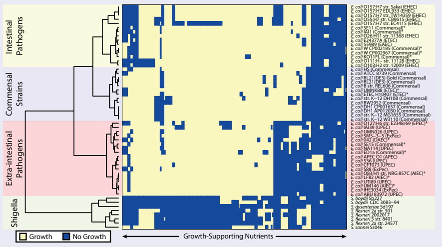
Flux distributions computed by FBA can be visualized on network maps. These maps show the state of the E. coli core model with maximum growth rate as the objective (Z) under aerobic (a) and anaerobic (b) conditions. The metabolic pathways shown in these maps are glycolysis (Glyc), pentose phosphate pathway (PPP), TCA cycle (TCA), oxidative phosphorylation (OxP), anaplerotic reactions (Ana), and fermentation pathways (Ferm).



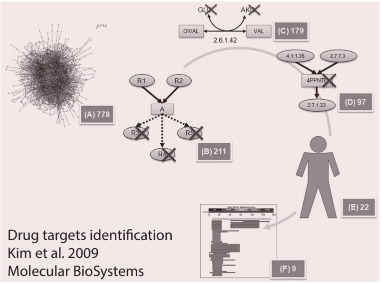
FK506 overproduction, Huang et al. 2013, Microbial Cell Factories



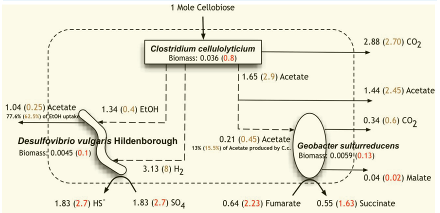
Competitive and cooperative patterns
Freilich et al. 2011, Nature Communications



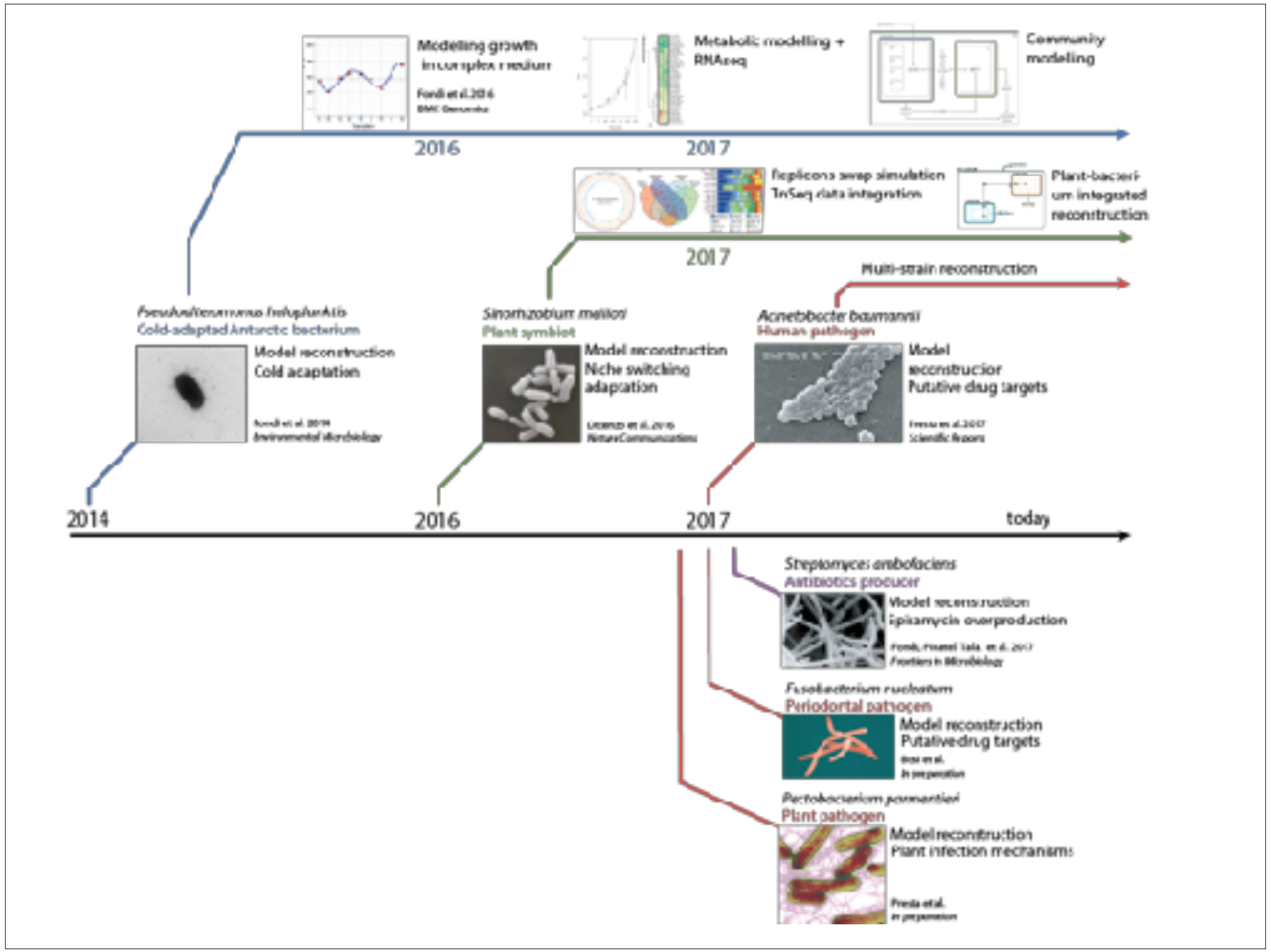
Multi-strain reconstruction, Monk et al. 2013, PNAS



Drug targets identification
Kim et al. 2009
Molecular BioSystems

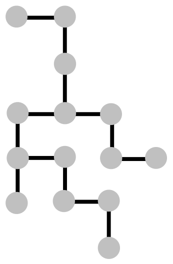


Community interaction
Zomorrodi and Maranas, 2012, Plos Comp Bio

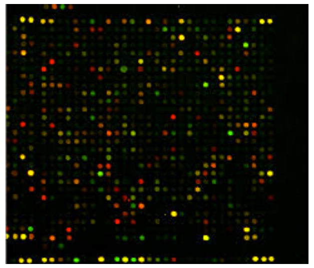

Modelling growth
in complex medium
Fondi et al. 2016
BMC Genomics
Metabolic modelling +
RNAseq
Community
modeling
2016
2017
Replicons swap simulation
TnSeq data integration
Plant-bacteri-
um integrated
reconstruction
2017
Multi-strain reconstruction
Pseudoalteromonas haloplanktis
Cold-adapted Antarctic bacterium
Model reconstruction
Cold adaptation
Fondi et al. 2014
Environmental Microbiology
Sinorhizobium meliloti
Plant symbiot
Model reconstruction
Niche switching
adaptation
Dicenzo et al. 2016
Nature Communications
Acinetobacter baumannii
Human pathogen
Model
reconstruction
Putative drug targets
Presta et al. 2017
Scientific Reports
2014
2016
2017
today
Streptomyces ambofaciens
Antibiotics producer
Model reconstruction
Spiramycin overproduction
Fondi, Pinatel, Talà, et al. 2017
Frontiers in Microbiology
Fusobacterium nucleatum
Periodontal pathogen
Model reconstruction
Putative drug targets
Bosi et al.
In preparation
Pectobacterium parmentieri
Plant pathogen
Model reconstruction
Plant infection mechanisms
Presta et al.
In preparation

Context specific metabolic models: integrating gene expression data

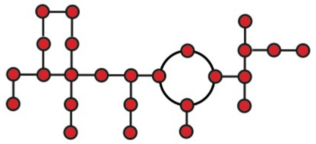
While genome-scale models aim at including the entirety of known metabolic reactions, mounting evidence has indicated that only a subset of these reactions is active in a given context, including: developmental stage, cell type, or environment. (Estevez and Nikoloski 2014)



gene expression data in condition x



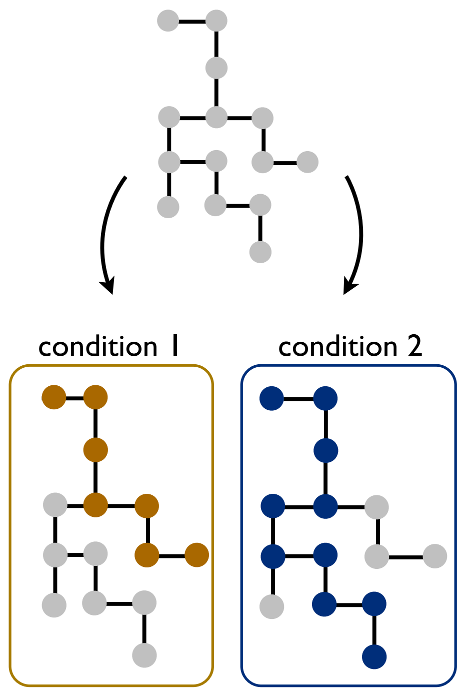
metabolic model



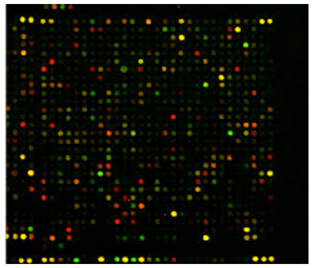
Given condition X (e.g. growth under pH/ temperature stress), how does the metabolic network re-wire or re-adjust to face it?

Context specific metabolic models: integrating gene expression data

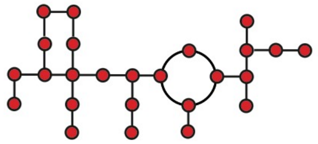
While genome-scale models aim at including the entirety of known metabolic reactions, mounting evidence has indicated that only a subset of these reactions is active in a given context, including: developmental stage, cell type, or environment. (Estevez and Nikoloski 2014)



gene expression data in condition x



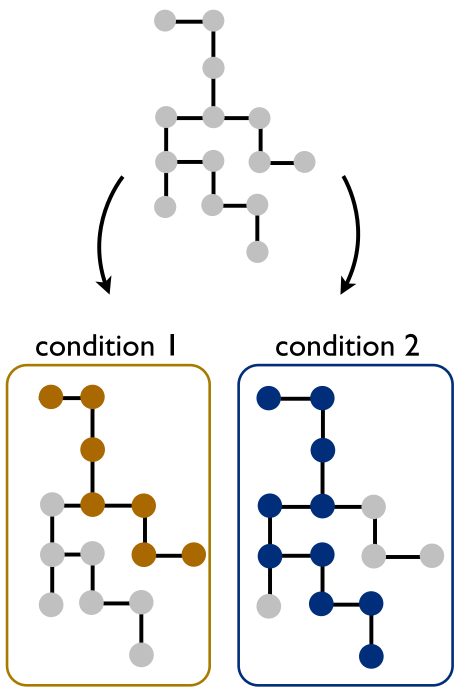
metabolic model

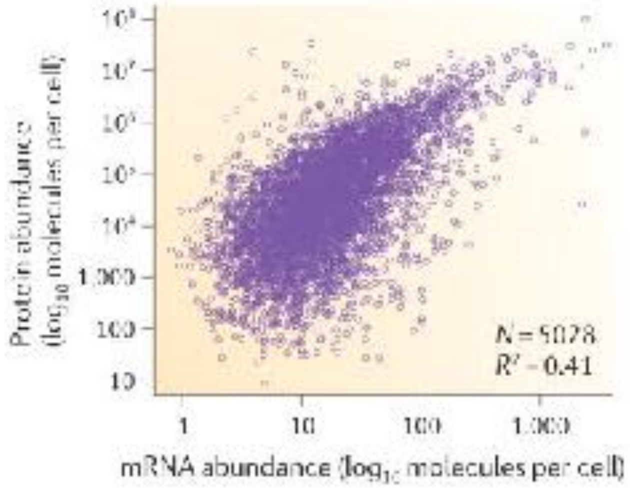


Given condition X (e.g. growth under pH/ temperature stress), how does the metabolic network re-wire or re-adjust to face it?

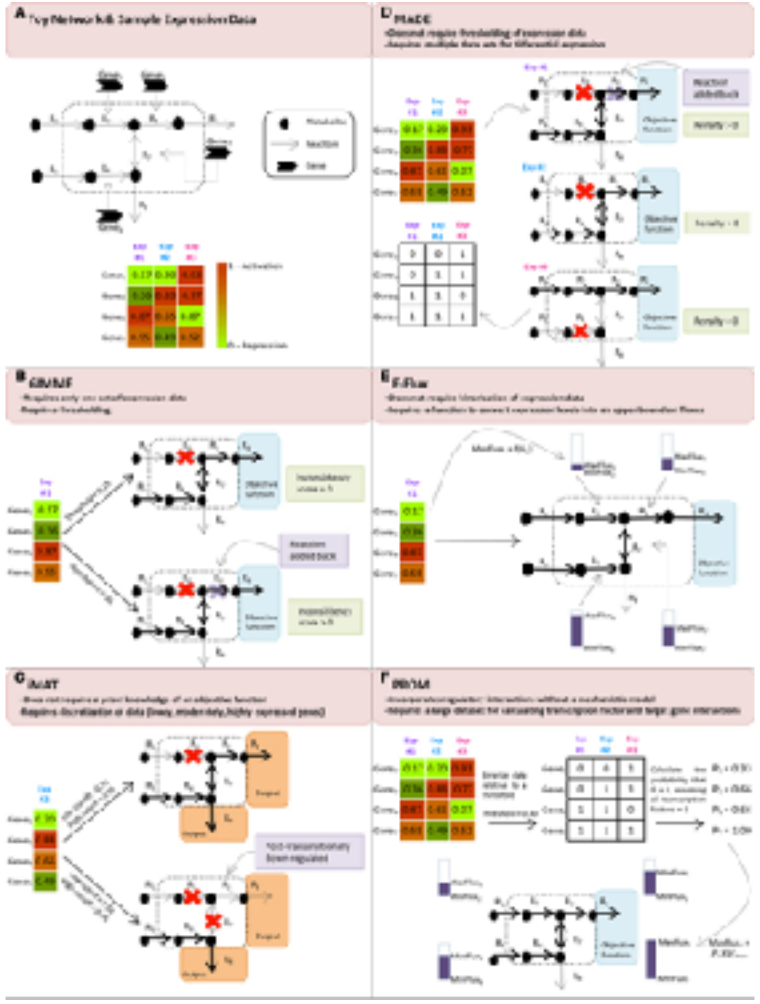
Context specific metabolic models: integrating gene expression data

While genome-scale models aim at including the entirety of known metabolic reactions, mounting evidence has indicated that only a subset of these reactions is active in a given context, including: developmental stage, cell type, or environment. (Estevez and Nikoloski 2014)





Vogel and Marcotte 2012


A Toy Network & Sample Expression Data
B GIMME
-Requires only one set of expression data
-Requires thresholding
C iMAT
-Does not require a priori knowledge of an objective function
-Requires discretization of data (lowly, moderately, highly expressed genes)
D MADE
-Does not require thresholding of expression data
-Requires multiple data sets for differential expression
E E-Flux
-Does not require binarization of expression data
-Requires a function to convert expression levels into an upper bound on fluxes
F PROM
-Incorporates regulatory interactions without a mechanistic model
-Requires a large dataset for calculating transcription factor and target gene interactions

Our model organism

Pseudoalteromonas haloplanktis TAC125

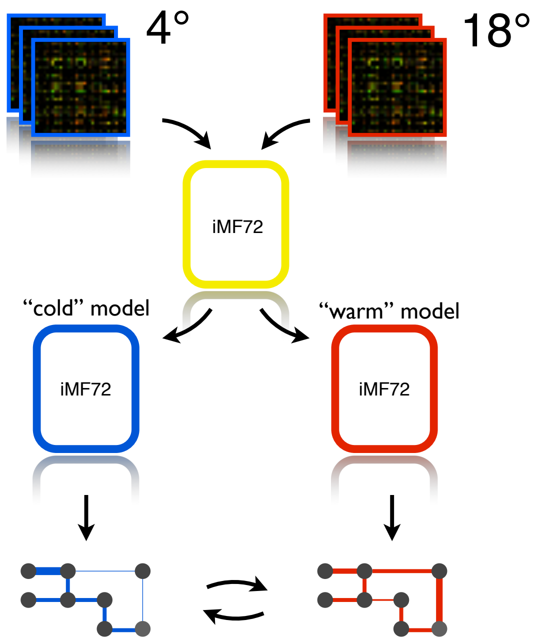
Cold-adapted Antarctic bacterium

- Biotechnological microorganism (grows fast and at temperature)
- New alternative expression host

The metabolic reprogramming following cold adaptation

- Integration of expression data

Piette et al. 2011



4°

18°

iMF72

"cold" model

"warm" model

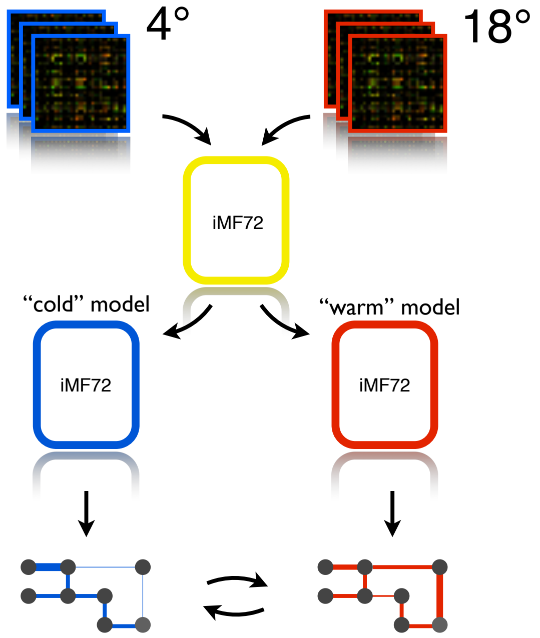
iMF72

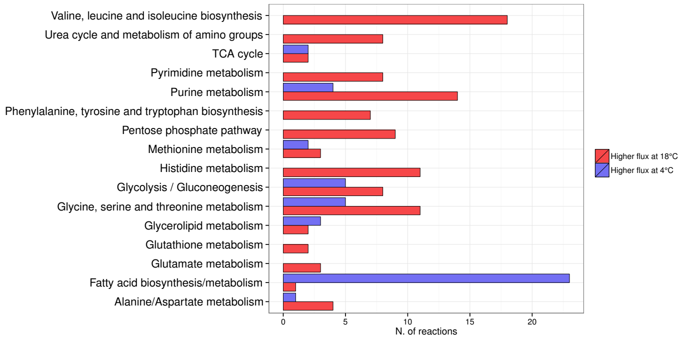
iMF72

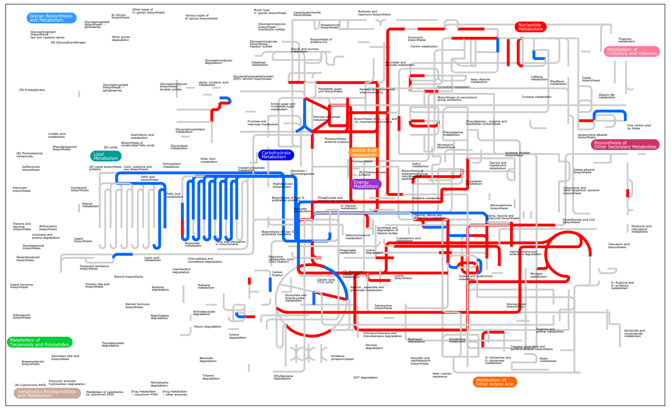
The metabolic reprogramming following cold adaptation

- Integration of expression data

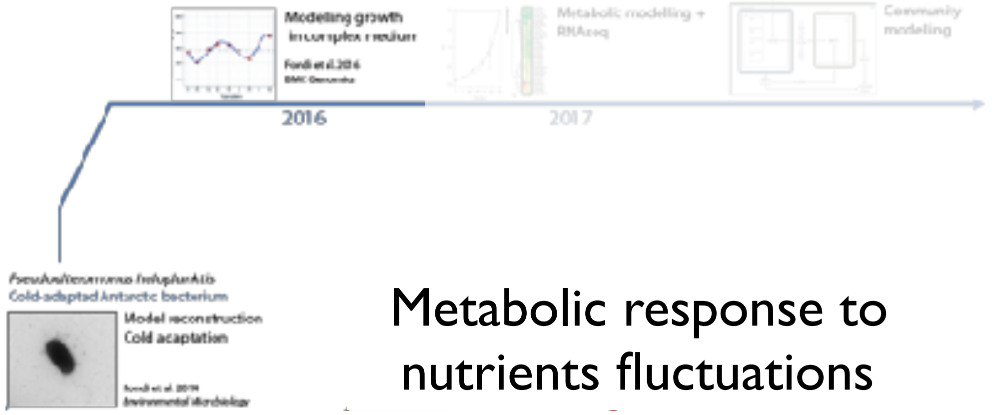
Piette et al. 2011



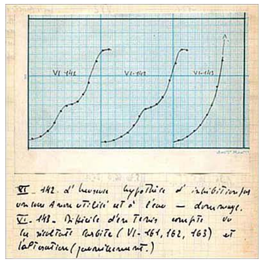




Fondi et al. 2014 Env Mic

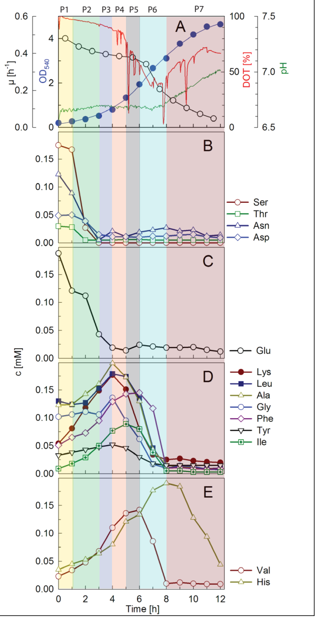


Metabolic response to nutrients fluctuations

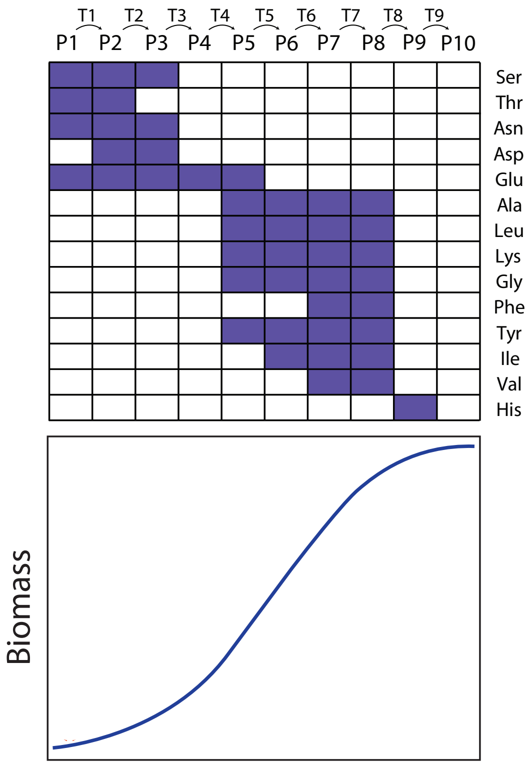


A bacterial cell does not simply take up various nutrients in proportion to their abundance, but actively controls their uptake by regulating its own enzymatic state ("di- auxie", Monod 1966)

The metabolic network of an organism has to continuously and dynamically adjust to optimally sustain cellular growth

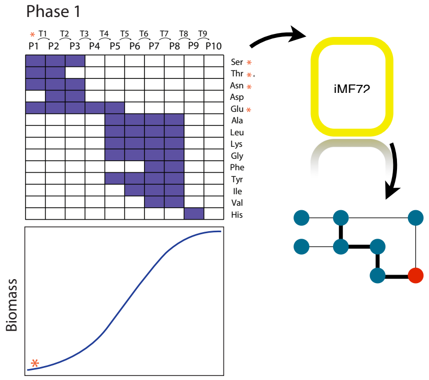


Wilmes et al. 2010



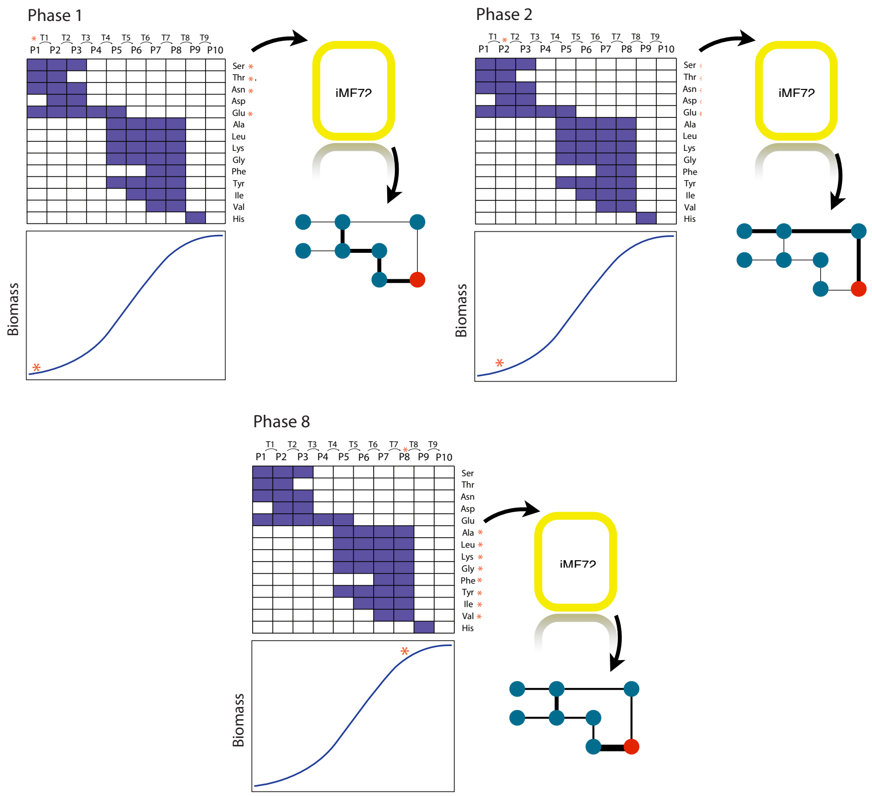
P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	
■	■	■								Ser
■	■									Thr
■	■	■								Asn
	■	■								Asp
■	■	■	■	■						Glu
				■	■	■	■			Ala
				■	■	■	■			Leu
				■	■	■	■			Lys
				■	■	■	■			Gly
						■	■			Phe
				■	■	■	■			Tyr
					■	■	■			Ile
						■	■			Val
								■		His

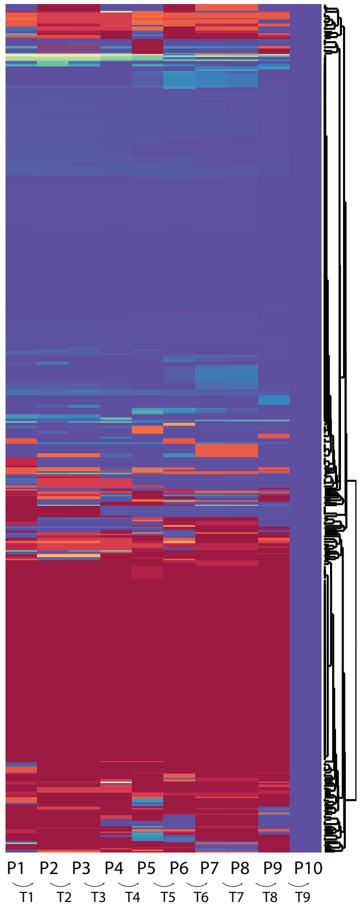
Multi-step FBA to understand the required metabolic reprogramming following nutrients switching

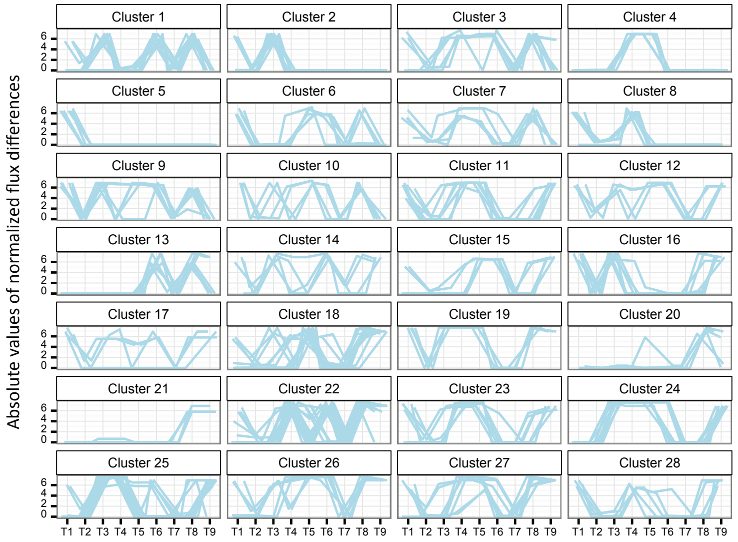


P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	
■	■	■								Ser *
■	■									Thr *
■	■	■								Asn *
	■	■								Asp
■	■	■	■	■						Glu *
				■	■	■	■			Ala
				■	■	■	■			Leu
				■	■	■	■			Lys
				■	■	■	■			Gly
						■	■			Phe
				■	■	■	■			Tyr
					■	■	■			Ile
						■	■			Val
								■		His

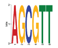
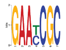
Multi-step FBA to understand the required metabolic reprogramming following nutrients switching



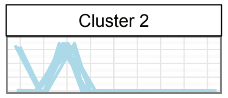




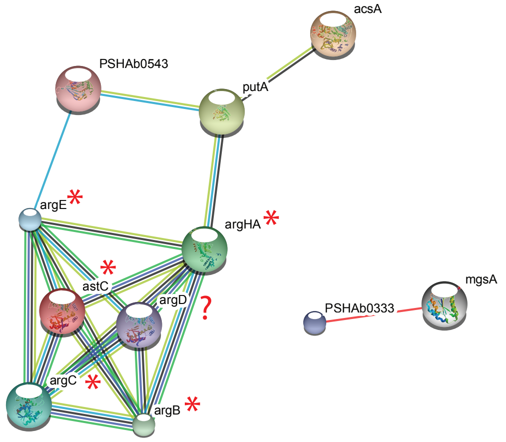
Gene co-regulation?

Cluster name	Motif name	Genes	Weblogo
Cluster 2	ArgR	PSHAa0194, PSHAa0698,PSHAa2175, PSHAa2287, PSHAa2290, PSHAa2291, PSHAa2292, PSHAb0333, PSHAb0428, PSHAb0543	
Cluster 3	CcpA	PSHAa0189, PSHAa0609, PSHAa0740, PSHAa1167, PSHAa1648, PSHAa1649, PSHAa1650, PSHAa1651, PSHAa2167, PSHAb0082, PSHAb0345	
Cluster 6	GalR	PSHAa0603, PSHAa0871, PSHAa1364, PSHAa1767, PSHAa2301, PSHAb0295	

Fondi et al. 2016 BMC genomics

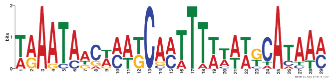


Evidence network (String DB)

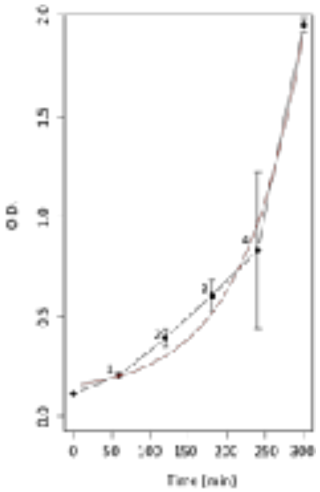


* Inferred Arg regulon for PhTAC125 (regprecise DB)

Shared motif upstream with the other genes in the regulon

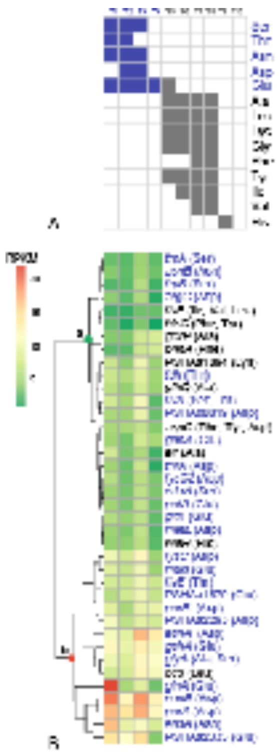


RNAseq of P. haloplanktis TAC125 growth in a nutritionally complex environment

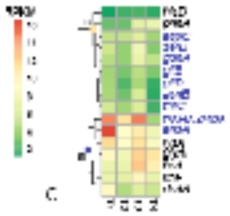


- 2 biological and 3 technical replicates
- ~ 25 M reads/sample, 75 bp

AA degradation genes



AA biosynthetic genes

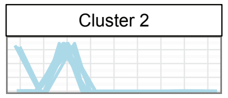


Gene expression data:

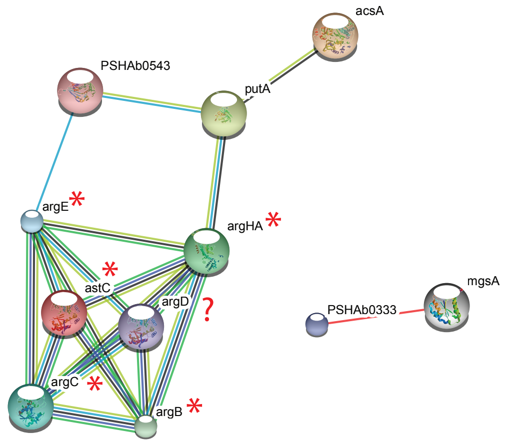
- are consistent with observed nutrient assimilation patterns

- illustrate preferential amino acids degradation pathways

Cluster 2

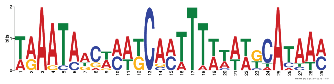


Evidence network (String DB)

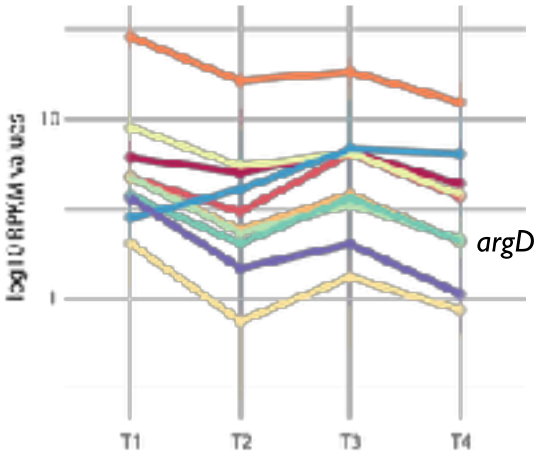


* Inferred Arg regulon for PhTAC125 (regprecise DB)

Shared motif upstream with the other genes in the regulon

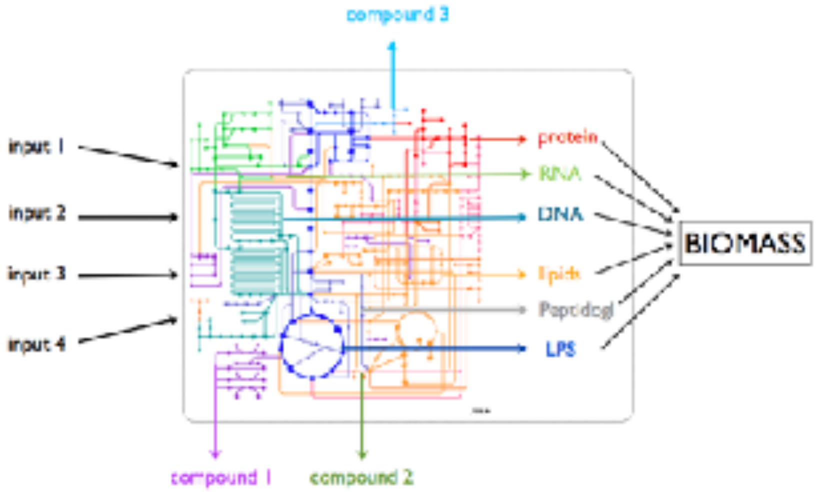


Co-expressed genes



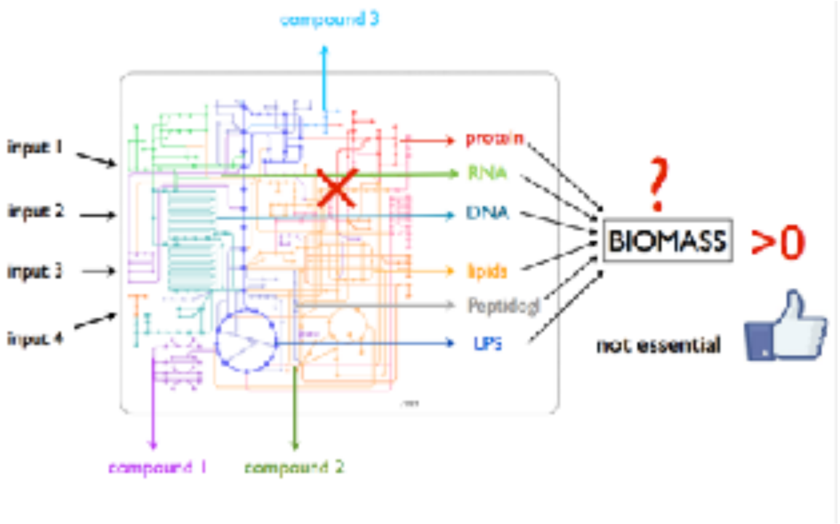
Essential genes prediction

A gene whose removal leads to a non vital phenotype or to a largely impaired phenotype



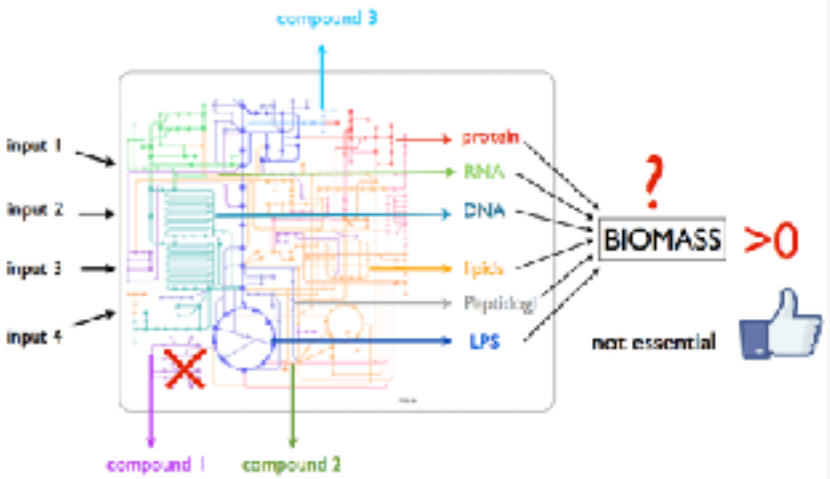
Essential genes prediction

A gene whose removal leads to a non vital phenotype or to a largely impaired phenotype



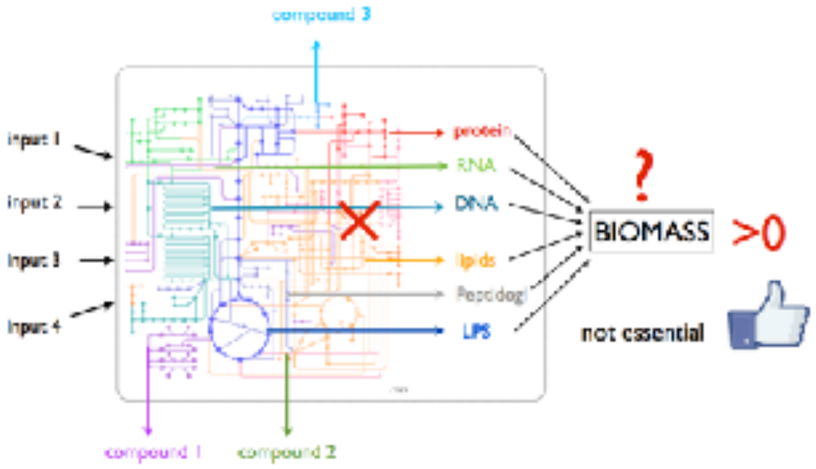
Essential genes prediction

A gene whose removal leads to a non vital phenotype or to a largely impaired phenotype



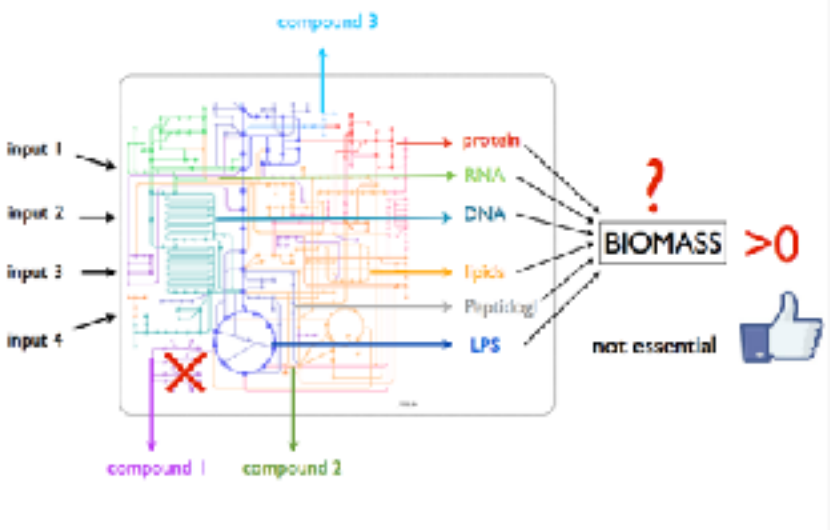
Essential genes prediction

A gene whose removal leads to a non vital phenotype or to a largely impaired phenotype



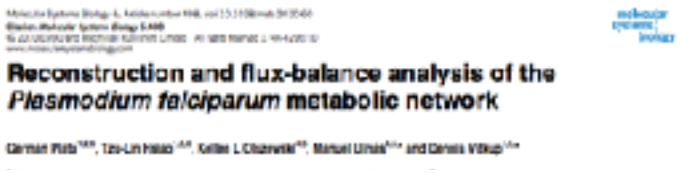
Essential genes prediction

A gene whose removal leads to a non vital phenotype or to a largely impaired phenotype

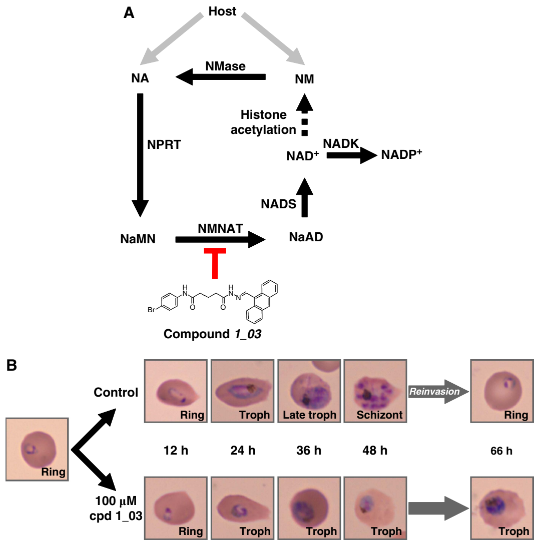


Essential genes prediction

A gene whose removal leads to a non vital phenotype or to a largely impaired phenotype



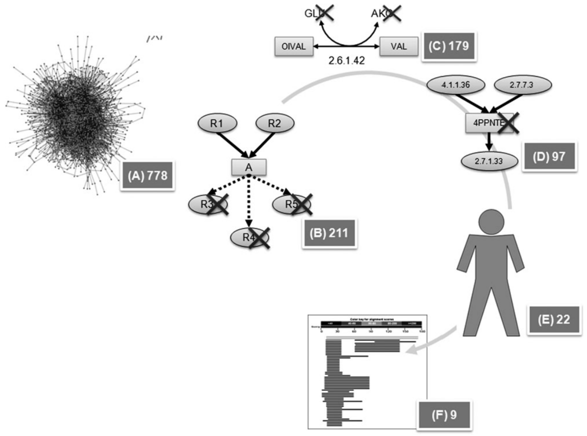
Reconstruction and flux-balance analysis of the *Plasmodium falciparum* metabolic network



SCIENTIFIC REPORTS

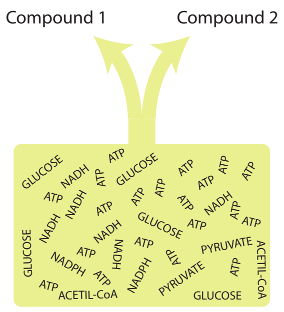
Constraint-based modeling identifies new putative targets to fight colistin-resistant *A. baumannii* infections

Presta et al. 2017



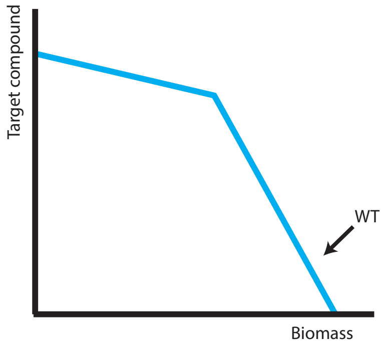
Kim et al. 2009

Metabolic engineering



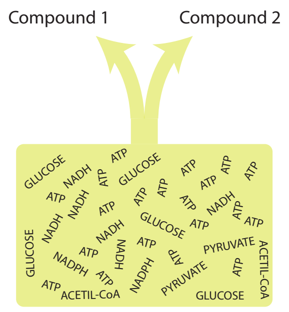
Microbial metabolic production is usually lower than theoretical yield. This may happen because:

- the objective (target) of microbial metabolism is often other than the desired one
- distinct biosynthetic pathways compete for common pool of intermediates



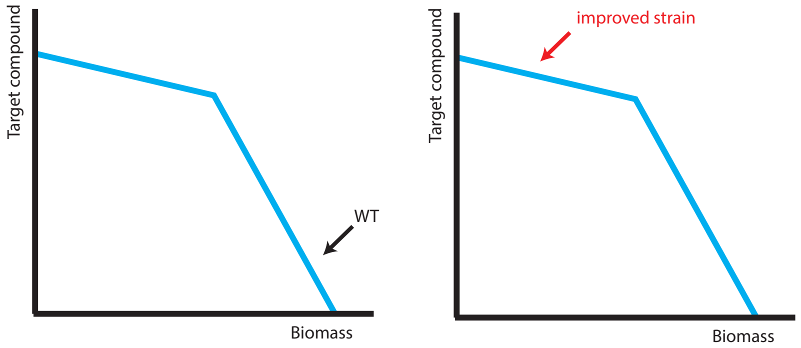
Pareto front

Metabolic engineering

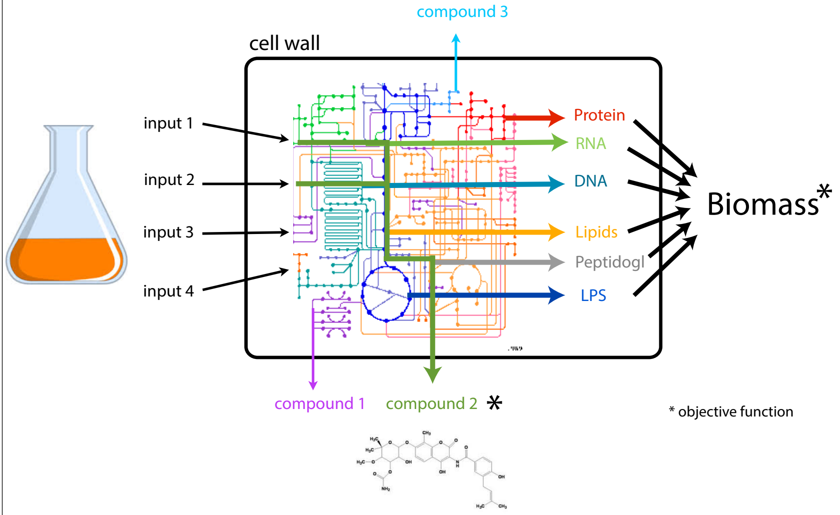


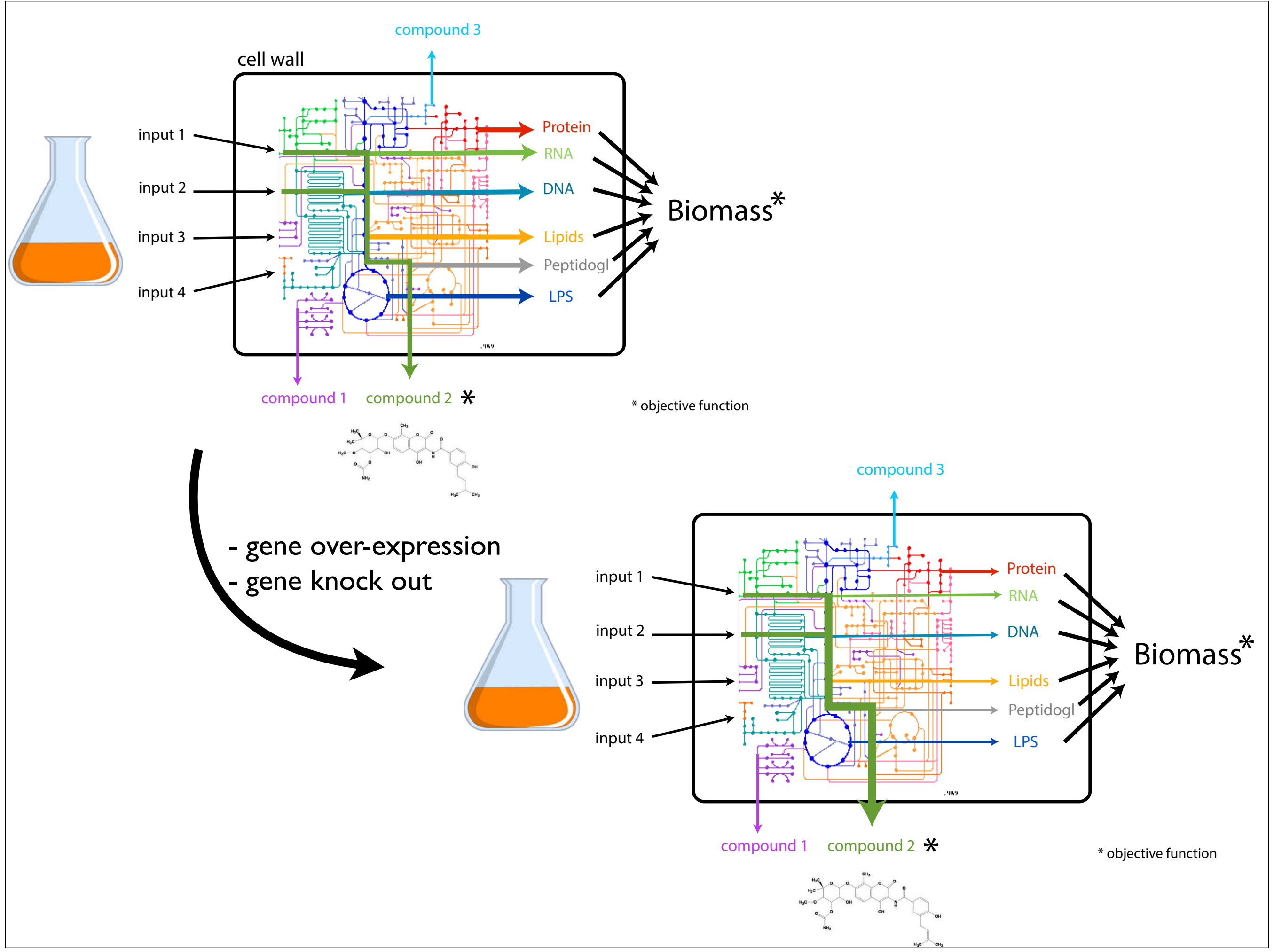
Microbial metabolic production is usually lower than theoretical yield. This may happen because:

- the objective (target) of microbial metabolism is often other than the desired one
- distinct biosynthetic pathways compete for common pool of intermediates



Pareto front


compound 3
cell wall
input 1
input 2
input 3
input 4
Protein
RNA
DNA
Lipids
Peptidogl
LPS
Biomass*
compound 1
compound 2 *
* objective function



compound 3

cell wall

input 1
input 2
input 3
input 4

Protein
RNA
DNA
Lipids
Peptidogl
LPS

Biomass*

compound 1
compound 2 *

* objective function

- gene over-expression
- gene knock out

compound 3

input 1
input 2
input 3
input 4

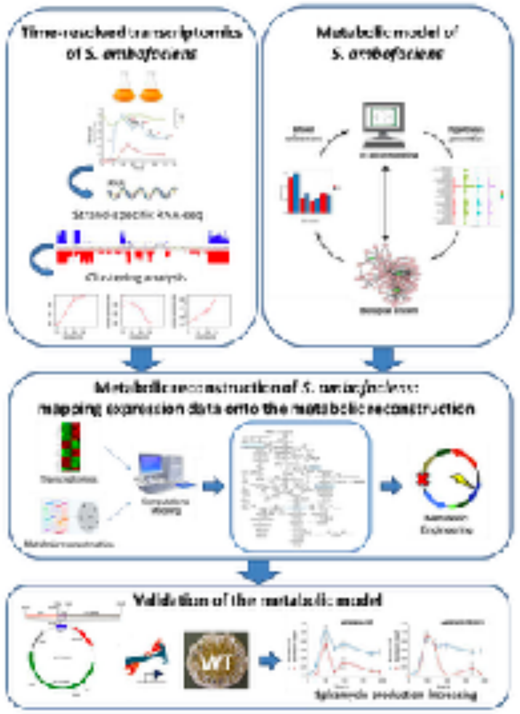
Protein
RNA
DNA
Lipids
Peptidogl
LPS

Biomass*

compound 1
compound 2 *

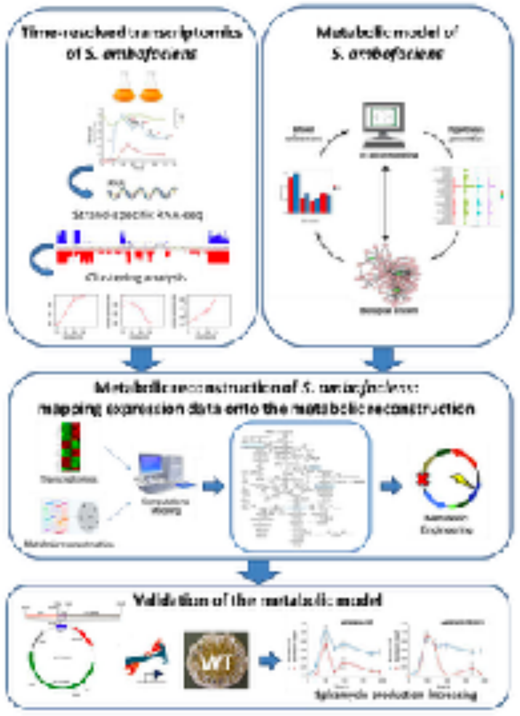
* objective function

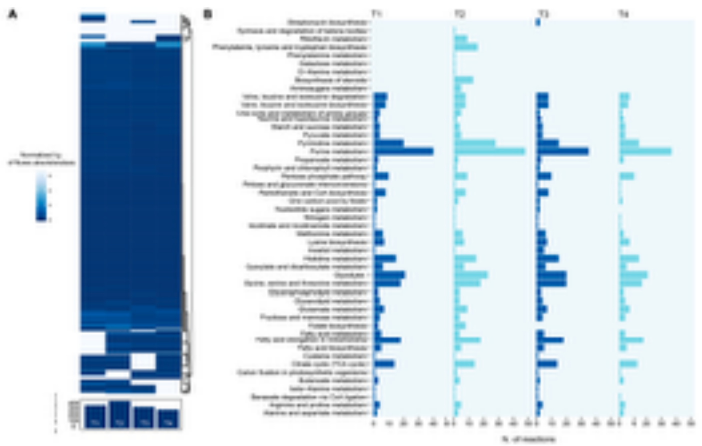
Increasing Spiramycin Production in *Streptomyces ambofaciens*

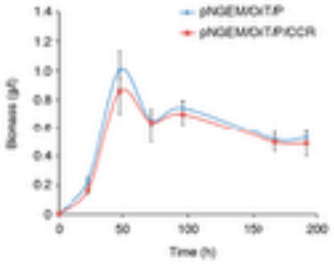


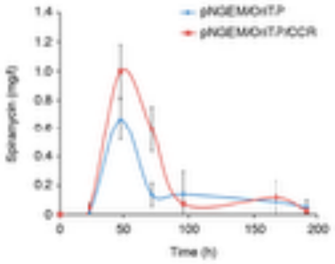
Fondi et al. 2016

Increasing Spiramycin Production in *Streptomyces ambofaciens*









Fondi et al. 2016

Appl Microbiol Biotechnol (2012) 94:637–649
DOI 10.1007/s00253-011-3773-6

BIOTECHNOLOGICAL PRODUCTS AND PROCESS ENGINEERING

In silico aided metabolic engineering of Streptomyces roseosporus for daptomycin yield improvement

Di Huang & Jianping Wen & Guoying Wang &
Guanghai Yu & Xiaoqiang Jia & Yunlin Chen

Bioprocess Biosyst Eng
DOI 10.1007/s00449-013-1027-y

ORIGINAL PAPER

Dynamic flux balance analysis of batch fermentation: effect of genetic manipulations on ethanol production

K. P. Lisha · Debasis Sarkar

Huang et al. Microbial Cell Factories 2013, 12:52
http://www.microbialcellfactories.com/content/12/1/52

MICROBIAL CELL FACTORIES

RESEARCH Open Access

Genome-scale metabolic network guided engineering of Streptomyces tsukubaensis for FK506 production improvement

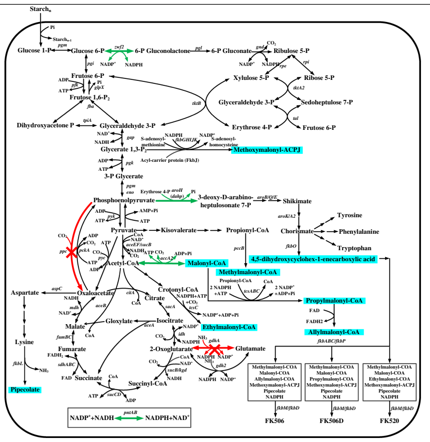
Di Huang[1,3], Shanshan Li[1], Menglei Xia[1], Jianping Wen[1,2*] and Xiaoqiang Jia[1,2]

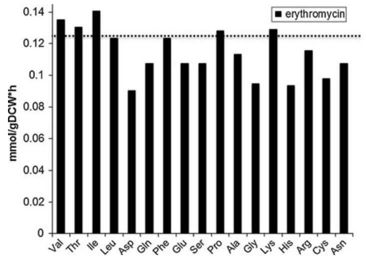
Santala et al. Microbial Cell Factories 2011, 10:36
http://www.microbialcellfactories.com/content/10/1/36

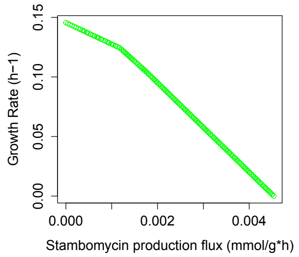
MICROBIAL CEL

RESEARCH Open Access

Improved Triacylglycerol Production in Acinetobacter baylyi ADP1 by Metabolic Engineering







Metabolic symbiosis

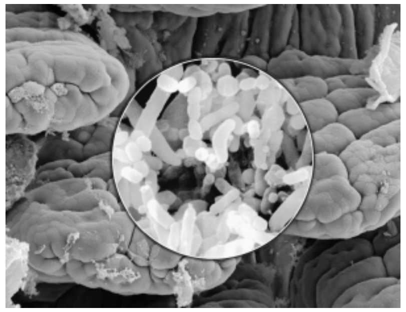
Gut Microbes 4:1, 28–40; January/February 2013; © 2013 Landes Bioscience

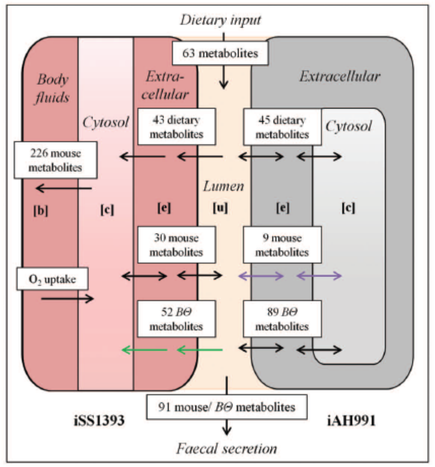
Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut

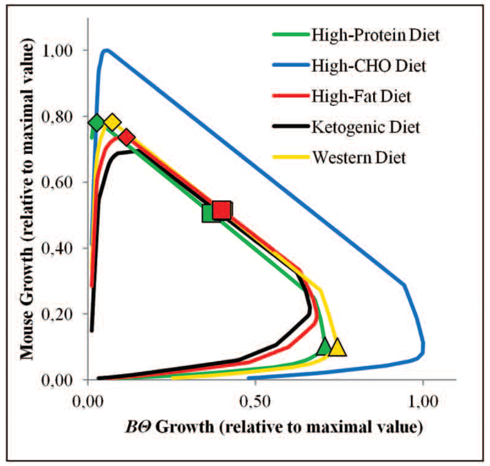
Almut Heinken,[1] Swagatika Sahoo,[1] Ronan M. T. Fleming,[1,2] and Ines Thiele[1,3,*]

[1]Center for Systems Biology; University of Iceland; Reykjavik, Iceland; [2]Department of Biochemistry and Molecular Biology; Faculty of Medicine; University of Iceland; Reykjavik, Iceland; [3]Faculty of Industrial Engineering; Mechanical Engineering and Computer Science; University of Iceland; Reykjavik, Iceland

Keywords: systems biology, computational modeling, metabolism, host-microbe interactions, *Bacteroides thetaiotaomicron*, constraint-based modeling, Mus musculus







Conclusions and outlooks

- Genome sequence is not enough. Modelling tools to account for the emergence of complex behaviors. Even better, if integrated with -omics data
- Constraint-based approaches can be used to identify the effects of environmental perturbations on the overall physiology of the cell
- Experimental tests are crucial for model validation and testing (positive feedback between computational biologists and experimentalists)
- What’s next? Dynamic modeling (kinetic modelling (ODEs) + FBA) and community models

Acknowledgements

Dep. of Biology, Unifi
Prof. Renato Fani
Alessio Mengoni
Elena Perrin
Giovanni Bacci
Luana Presta
Emanuele Bosi (present affiliation: Dipartimento di Scienze Biomediche Sperimentali e Cliniche, Unifi)

Ente CRA
(Biolog data)
Stefano Mocali

Università di Napoli Federico II
(PhTAC125 physiology)
Maria Luisa Tutino
Ermenegilda Parrilli

Università di Verona
(PhTAC125 transcriptomics)
Barbara Cardazzo
Lisa Carraro
Nadia Andrea Andreani

Funding

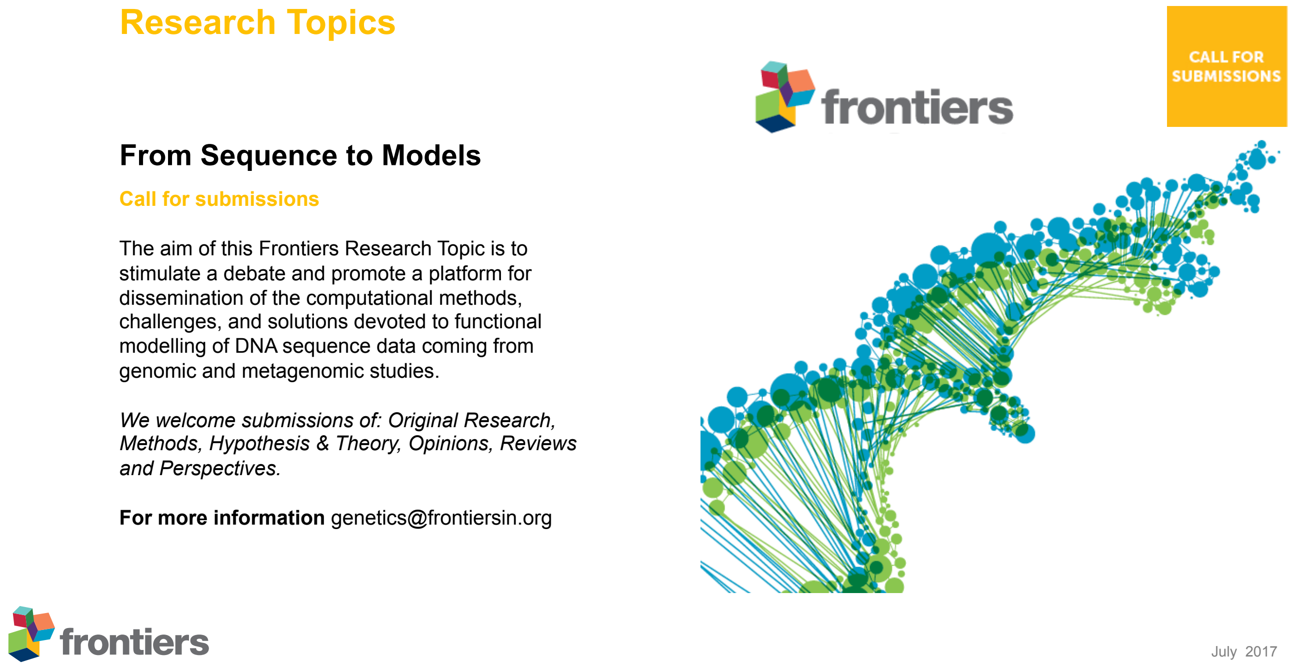
FEDERATION OF EUROPEAN MICROBIOLOGICAL SOCIETIES
FEMS

Long term fellowship
2012-14

PROGRAMMA NAZIONALE DI RICERCHE IN ANTARTIDE
ITALIANTARTIDE

PNRA2013-B0.4-2
PNRA2016_000246-A1

We are currently accepting submission for an open research topic in Frontiers series



Research Topics

From Sequence to Models

Call for submissions

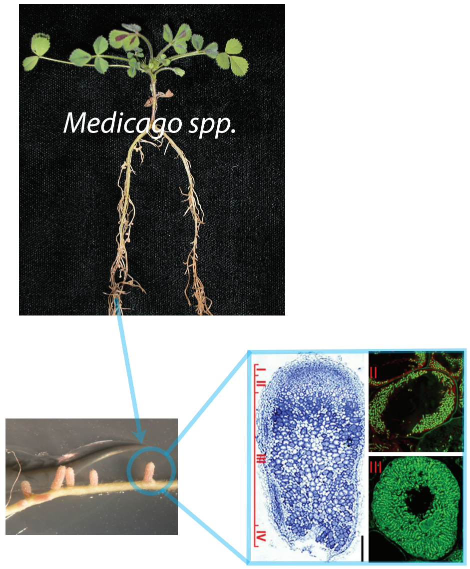
The aim of this Frontiers Research Topic is to stimulate a debate and promote a platform for dissemination of the computational methods, challenges, and solutions devoted to functional modelling of DNA sequence data coming from genomic and metagenomic studies.

We welcome submissions of: Original Research, Methods, Hypothesis & Theory, Opinions, Reviews and Perspectives.

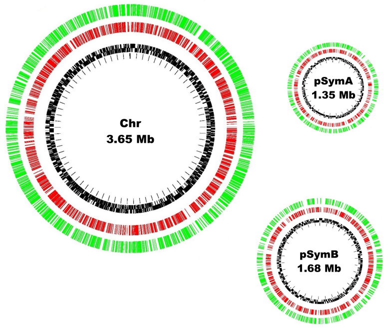
For more information genetics@frontiersin.org

July 2017

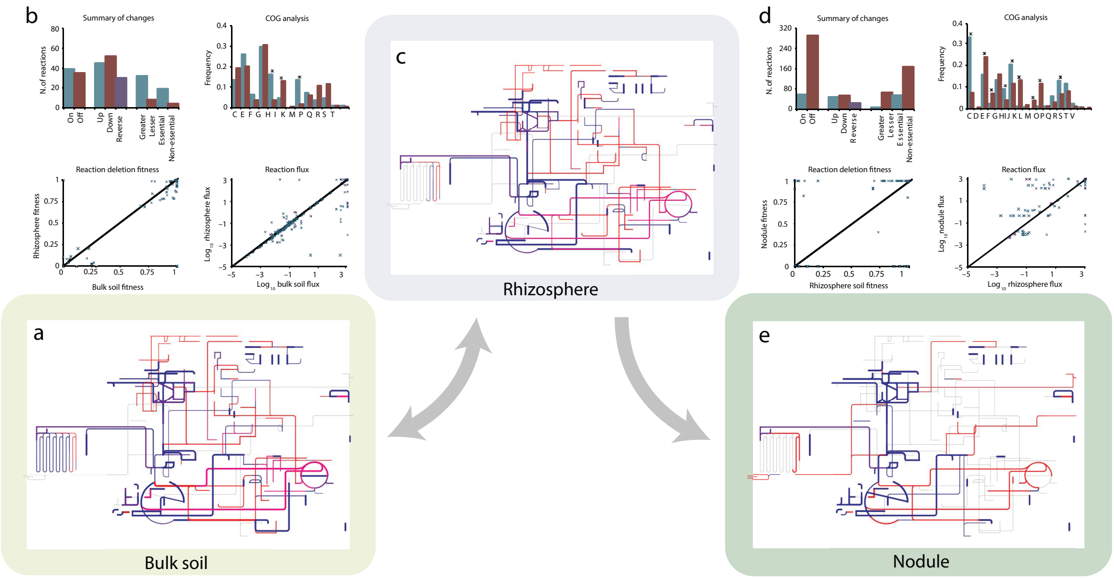
The nitrogen fixing bacterium *Sinorhizobium meliloti*



- *S. meliloti* promotes legume growth via N_2 fixation
- A model organism for the study of bacterial multipartite genome function and evolution (10% of microbes possess such a feature)

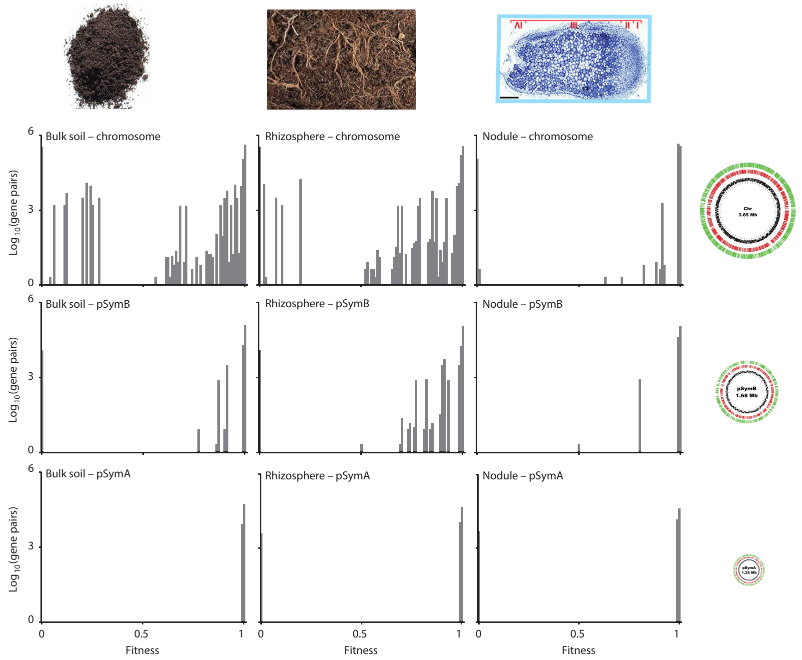


The metabolic shifts experienced by S. meliloti during transition between bulk soil, the rhizosphere and the nodule were modelled using in silico representations of the nutritional composition of each environment.

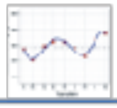


Rhizosphere colonization required a metabolic refinement.

Dicenzo et al. 2016, *Nature Communications*

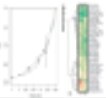


The chromosome is not metabolically specialized for a particular niche. pSymB is metabolically specialized for the rhizosphere, helping S. meliloti to adapt to this environment and utilize the newly available substrate



Modelling growth in complex medium

Fondi et al. 2016
BMC Genomics



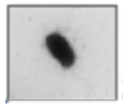
Metabolic modelling + RNAseq

Community modeling

2016

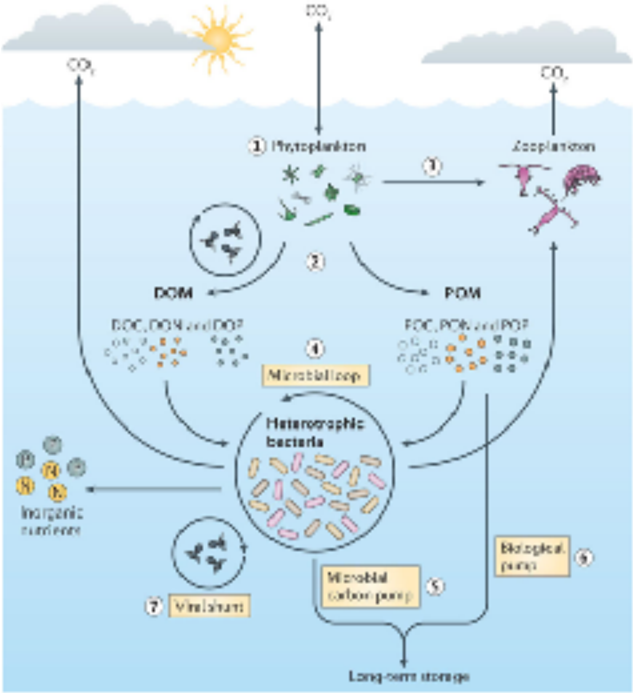
2017

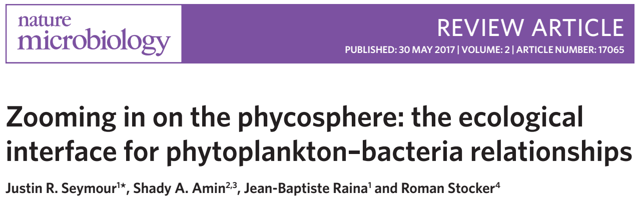
Pseudoalteromonas haloplanktis
Cold-adapted Antarctic bacterium



Model reconstruction
Cold adaptation

Fondi et al. 2014
Environmental Microbiology

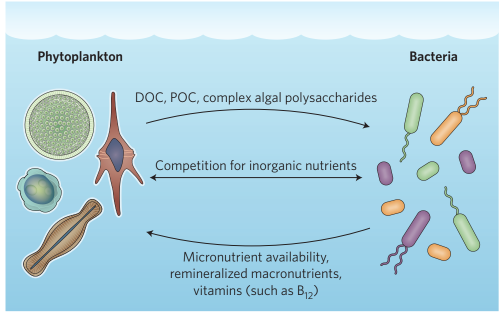


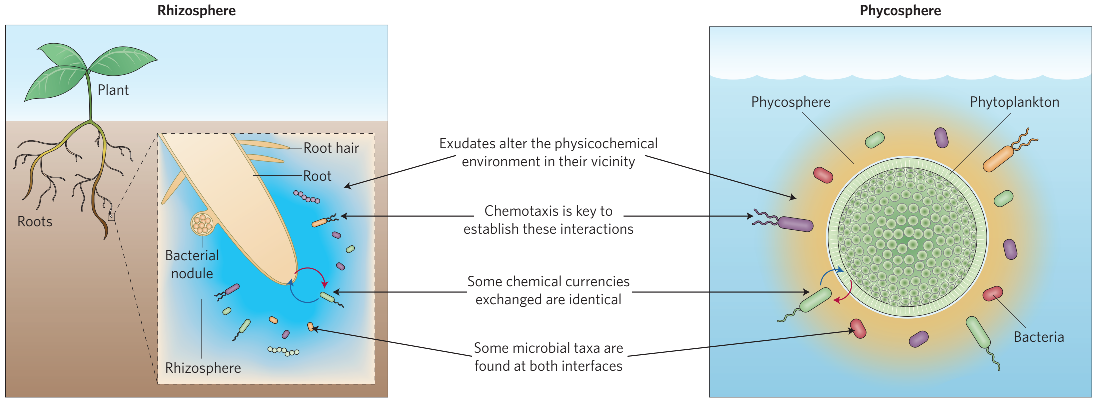


nature microbiology
REVIEW ARTICLE
PUBLISHED: 30 MAY 2017 | VOLUME: 2 | ARTICLE NUMBER: 17065

Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships

Justin R. Seymour[1]*, Shady A. Amin[2,3], Jean-Baptiste Raina[1] and Roman Stocker[4]





A “crowded” phycosphere

MMBR

Interactions between Diatoms and Bacteria

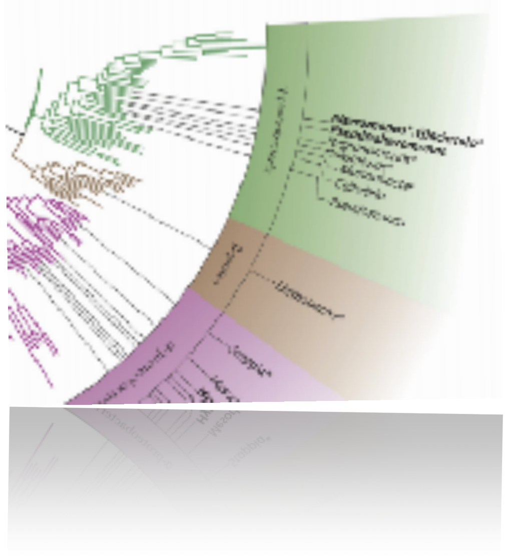
Shady A. Amin, Micaela S. Parker, and E. Virginia Armbrust

School of Oceanography, University of Washington, Seattle, Washington, USA

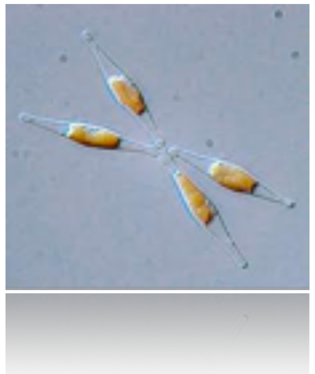
- Diatoms and bacteria have developed specific interactions over hundreds of millions of years.

- Understanding interactions between diatoms and bacteria is of prime importance to deciphering oceanic nutrient fluxes and biogeochemical cycles

- Metabolic interactions (mutualism, competition) have been deeply described

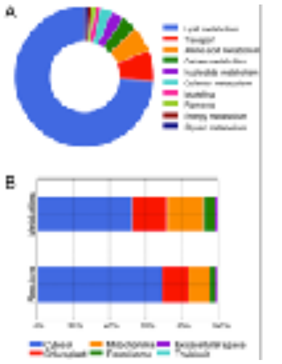


Introducing the marine diatom Phaeodactylum tricornutum



- Complete genome sequence available since 2008
- It has been used in laboratory-based studies of diatom physiology for several decades
- Transformable

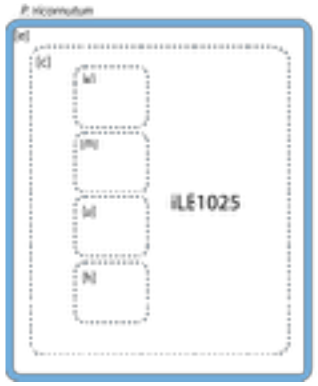
Many genome-scale metabolic reconstructions are available for this organism



Levering et al. 2016, Plos ONE

Levering et al. 2017, mSystems

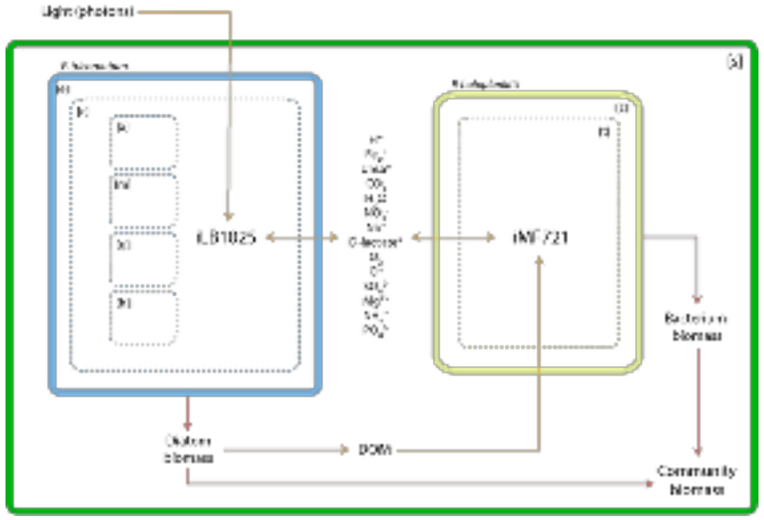
1,027 genes associated with 4,456 reactions and 2,172 metabolites distributed across six compartments

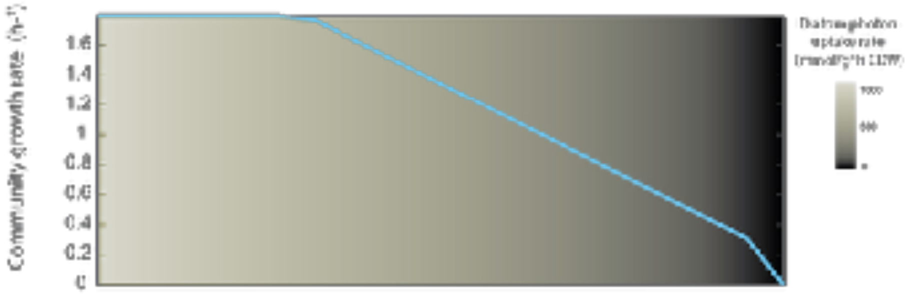


An integrated metabolic reconstructions to study phtotroph-heterotroph metabolic interactions

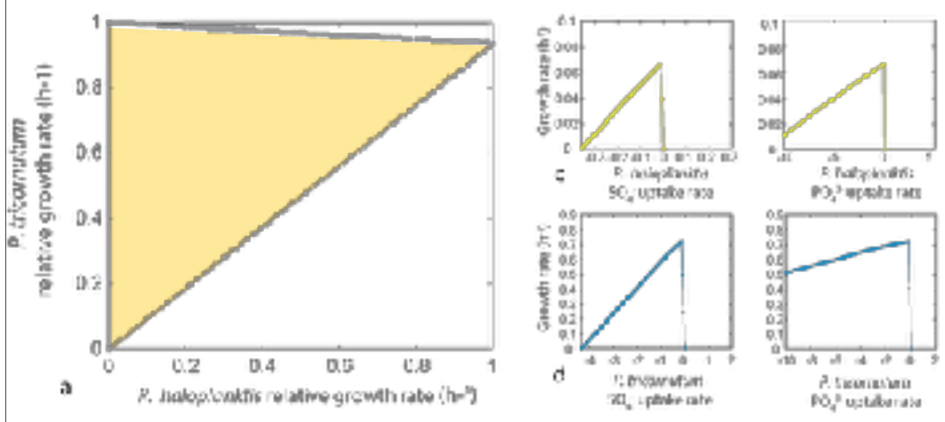
- Combined two existing metabolic reconstructions (namely iLPB1025 and iMF721) in a single, integrated model

- The combined model should be able to capture basic metabolic relationships as competition and commensalism, which are commonly observed in diatom-heterotroph associations





A community-level metabolic reconstruction captures known diatom-bacteria interaction

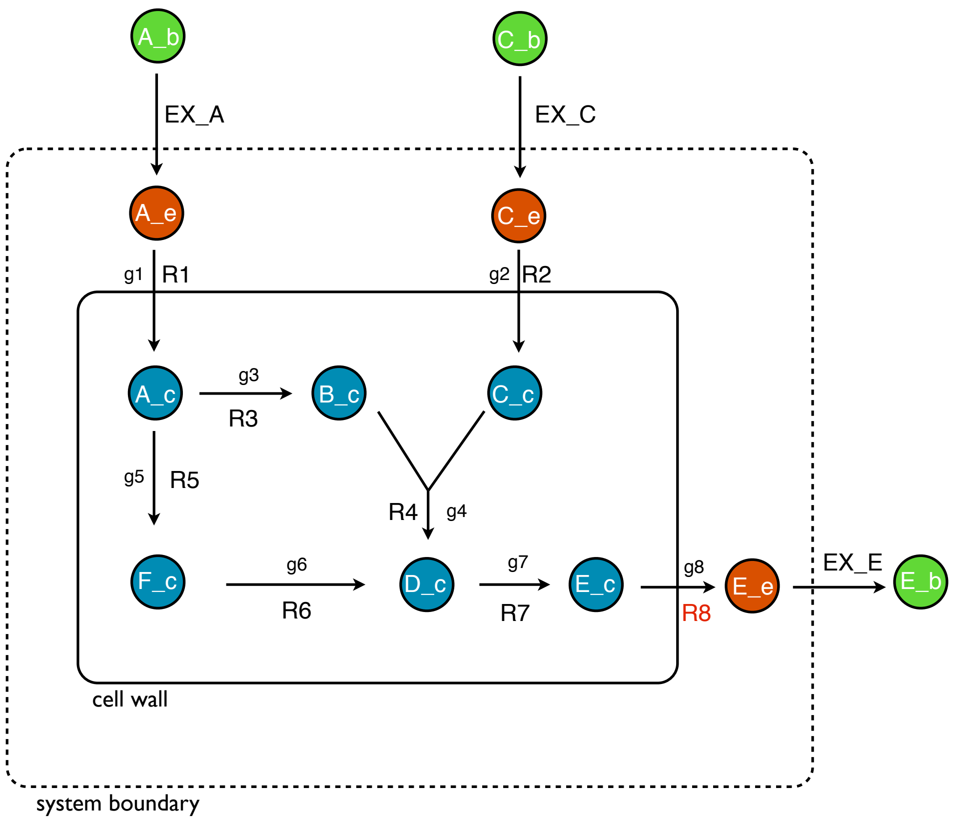


Pareto front analysis

Robustness analysis

- competition
- commensalism
- PO_4^{3-} limitation

A toy model



Metabolites

C_b A_b E_b	boundary
C_e A_e E_e	extracellular
C_c A_c B_c F_c D_c E_c	intracellular

Reactions

R1 to R8

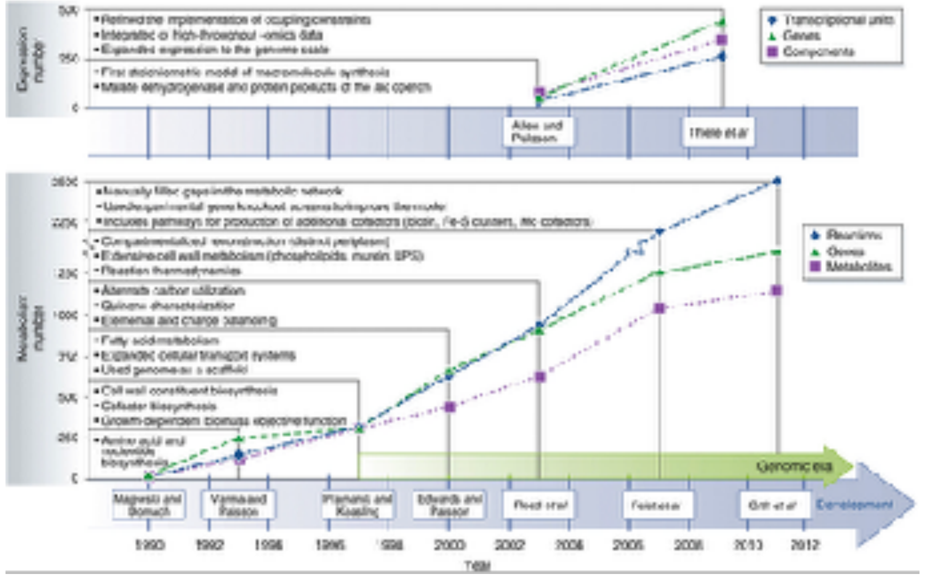
Genes

g1 to g8

Objective function

R8 (or EX_E)

A bit on *E. coli* model and the protocol



iOJ1366

Component	Count
Metabolites	1805
Reactions	2583
Genes	1367

- Download the code and the *E. coli* reconstruction from
- dbefcb.unifi.it, section “Talks, slides and posters” and uncompress the archive.
- Use Matlab command line to move in that folder (use “cd” command)
- Run the script line by line

Section 1, understanding the model

. Initializing the cobra toolbox
. Importing the model

Section 2, playing with growth conditions

. Changing substrate uptake
. Changing uptake rate

Section 3, playing with *E. coli* genes

. Identification of essential genes
. Identification of essential genes on different growth media