Metabolism: flux balans analysis
Life is... energy/nutrient cycling

“The individual taxonomic units evolve and go extinct, yet the core machines survive surprisingly unperturbed.”

PG Falkowski et al, Science 2008
Metabolic networks: Exploiting constraints

metabolic network are evolved

However metabolic networks many physical/chemical constraints stochiometry, energetic constraints can/should be exploited

allows for model upscaling to complexity of present day organisms
stochiometric constraint, + equilibrium assumption allows calculation of (optimal) flux through large metabolic networks

KEGG database

metabolic network yeast
I. Reaction network formalism

II. FBA formulation

**Dynamic mass balance**

\[ \frac{dC}{dt} = Sv \]

- **C**: Concentration
- **t**: Time
- **S**: Stoichiometric matrix
- **v**: Flux vector

**Steady-state assumption**

\[ Sv = 0 \]

**LP formulation**

Objective: \( Z = v_5 \)

Constraints:

<table>
<thead>
<tr>
<th>A</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
<th>R7</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
\begin{bmatrix}
  v_1 \\
  \vdots \\
  v_7
\end{bmatrix} = 0 \quad 0 \leq v_1, \ldots, v_7 \leq 10
\]

III. Hypothetical flux distribution at steady-state

\( Z = 10 \)

\[ v = [6.67 \ 3.33 \ 6.67 \ 6.67 \ 0 \ 3.33 \ 6.67]^T \]

The flux values are shown in the diagram with values 6.67 and 3.33.
Highly redundant GP map:

there are more than $10^{800}$ metabolic networks with 2,000 reactions that can synthesize all the small biomass molecules of the bacterium E. coli using glucose as the sole carbon source.

Typically only 30% overlap of networks with same phenotype (core network)

add/ delete one reaction $\rightarrow$ percolating neutral networks seeing novel phenotypes
ONLY RNA...? (cont.)
no ...
also ‘real” metabolic network

Metabolic networks
and their evolution.
Wagner A.
(evolutionary systems biology)
Using FBA to reconstruct evolution of metabolic network of yeast after WGD
to cope with genome-size networks:
exploit constraints and use shortcut: optimal equilibrium flux

Yeast metab. network
Flux balance analysis (FBA) assume 'automatic' regulation such that flux in equilibrium and maximal growth

- FBA solution non-unique:
  - use secondary optimization, eq minimal total flux
- stochiometric matrix \((nA \rightarrow mB)\)
- reactions coupled to enzymes
- set maximum flux (when enzymes are present) ( OR, AND reactions))
- however actual flux not proportional to amount of enzymes
Examples of flux balance analysis

- How do fluxes (growth) change with change of environment (=input-flux)
- How do fluxes (growth) change with knock-outs?
- Reconstruction of ecosystem wide metabolome (cf Bas Dutilh)
- ....
- How do genomes reduce after whole genome duplication? (cf van Hoek and Hogeweg 2009)
automatic reconstruction of metabolic networks from annotated genomes cf Henry CS1, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. 2010
Figure S1: Overview of results and discoveries arising from the Model SEED pipeline.
evolution of metabolic flux after WGD
FBA assumptions

- WGD $\rightarrow$ volume increase (decrease surface/volume ratio)
  volume $\approx$ depends on genome size
- flux of metabolic reaction depends on gene expression, 
dosis effect: gene copy number
- max flux through each reaction preset to maximum needed 
  for optimal growth in sampled set of realizations of 10 environment types
- enzymes have multiple functions

- reactions need multiple enzymes 
  take into account OR, AND (AND/OR) relations
- flux transport reactions: depends on gene expression AND 
surface/volume ratio
- after gene deletion maxflux reduced accordingly
cell size scales with amount of DNA 
Cavallier Smith (e.g. 2005)
In Yeast diploide cells are:
\[ V = 1.89 \times \text{haploid cells} \]
surface: \[ 1.56 \times \text{haploid cells} \]
\[ V = N^{0.9}; \quad A = V^{0.7} \]
where \( N \) number of genes

MaxFlux change as function of area change (\( \alpha \)), volume change (\( \beta \)) and gene dosage change (\( \gamma \))

\[ F_{\text{max}}(i) = F_{\text{max},0}(i) \frac{\alpha \gamma(i)}{\beta} \frac{1 + x(i)}{\gamma(i)x(i) + \alpha} \]

\[ F_{\text{max}}(i) = F_{\text{max},0}(i) \frac{\gamma(i)}{\beta} \]
evolution of metabolic flux after WGD

• 9 types of environments (available nutrients).
  realized in different concentrations
• per generation 1 environment seen
• pop size 100: flux dependent replication
  death: nogrowth + random
• after wgd: only deletions
  or duplication + deletion (max 2 copies)
• no fitness advantage for genome shrinkage smaller than initial volume
evolution of metabolic flux after WGD

evolutionary dynamics: growth rate and genome reduction
evolution of metabolic flux after WGD
flux in the various environments
(max and mean concentration)
initial decrease – how/when does it happen in evolution
Genome shrinkage after whole genome duplication dynamics of use of pathways in anaerobic glucose environment (env 3)
Only in “new” environment - nodirect disadvantage of WGD

BUT single INDELS initially better Exept in ethanol env

WGD mostly better end result than single INDELS
WGD: Simulated evolution and vs yeast duplication of yeast vs duplication of ancestor of yeast (±hgt)

 Preferential retained genes: Glycolysis pathway and Transporters
Evolution predictable!
conclusions
supervised vs non-supervised modeling of WGD in Yeast

<table>
<thead>
<tr>
<th>“Supervised”</th>
<th>“Non Supervised”</th>
</tr>
</thead>
<tbody>
<tr>
<td>find genes preferentially retained glycolysis pathway</td>
<td>take known interactions metab. net + DNA-volume relation</td>
</tr>
<tr>
<td>Model glycolysis pathway assuming dosis effect of duplicated genes</td>
<td>model evolution</td>
</tr>
<tr>
<td>demonstrate WGD can lead to increased glycolic flux</td>
<td>find preferentially retained genes glycolysis &amp; transport</td>
</tr>
<tr>
<td>WGD enabled to exploit high glucose resource during emergence of angiosperms</td>
<td>WGD enabled to exploit high glucose resource during emergence of angiosperms</td>
</tr>
<tr>
<td>observed outcome of WGD</td>
<td>expected outcome of WGD</td>
</tr>
</tbody>
</table>
WGD observed in phylogeny at times of environmental shifts

van der Peer et al 2009, Nature genetic reviews
WGD observed in virtual cell model at times of environmental shifts

WGD ongoing mutation, but only fixed in population EARLY in evolution OR after SOME (severe?) environmental changes

and WGD leads to high fitness much later

Cuypers & Hogeweg 2014
Explicit model of Insuline signaling pathway

Random sampling of 15 kinetic parameters $10^{-3} – 10^3$ and evolving populations by mutating these parameters

Generate many “healthy” and “sick” individuals (pathway instantiations)

Classifying behavior as “normal” $V = 0.076 \times 10^{-4}$ or ”deseased” $V = 0.33 \times 10^{-4}$
(based on glucase uptake-curve in time) Determine sensitivity of parameters in different populations and during evolution.
**Fig 1. Insulin signaling model, input and output.**

a) Molecular interactions in the signaling pathway modeled here. Briefly, extracellular insulin leads to phosphorylation of the insulin receptor, which promotes the phosphorylation of IRS1 to yield IRS1P. The latter molecule associates with PI3K in a complex that triggers production of the second messenger PI34P3, which activates the protein kinases Akt and PKCζ. These kinases then promote the translocation of the glucose transporter GLUT4 to the membrane, where it helps import glucose into the cell. Mass-action parameters that determine the rates of the respective reactions are indicated by a 'k' followed by a subscript. Activated PKCζ and Akt exert feedback on the production of two different phosphorylated forms of IRS1 (IRS1SP and IRS1P). The strength of this feedback is encapsulated by parameters $f_{PKCζ}$ and $f_{AMP}$, respectively. See Methods for details.

b) The model's first three equations are:

\[
\frac{d[IRS1]}{dt} = -k_{F,IRS1P}[IRS1][SIGNAL] + f_{Akt}k_{B,IRS1P}[Akt][IRS1P] + k_{F,IRS1SP}[IRS1SP] - f_{PKCζ}[PKCζ][IRS1] \\
\frac{d[IRS1P]}{dt} = k_{F,IRS1P}[IRS1][SIGNAL] - f_{Akt}k_{B,IRS1P}[Akt][IRS1P] + k_{B,IRS1P,PI3K}[IRS1P,PI3K] - k_{F,IRS1P,PI3K}[PI3K][IRS1P] \\
\frac{d[IRS1SP]}{dt} = f_{PKCζ}[PKCζ][IRS1] - k_{B,IRS1SP}[IRS1SP]
\]
**Fig 1. Insulin signaling model, input and output.** a) Molecular interactions in the signaling pathway modeled here. Briefly, extracellular insulin leads to phosphorylation of the insulin receptor, which promotes the phosphorylation of IRS1 to yield IRS1P. The latter molecule associates with PI3K in a complex that triggers production of the second messenger P_{i}^{3,4,5}P_{3}, which activates the protein kinases Akt and PKCZ. These kinases then promote the translocation of the glucose transporter GLUT4 to the membrane, where it helps import glucose into the cell. Mass-action parameters that determine the rates of the respective reactions are indicated by a ‘k’ followed by a subscript. Activated PKCZ and Akt exert feedback on the production of two different phosphorylated forms of IRS1 (IRS1SP and IRS1P). The strength of this feedback is encapsulated by parameters $f_{PKCZP}$ and $f_{AMP}$, respectively. See Methods for details. b)
Importance of parameter varies greatly depending on parameter set (= genetic background)

sensitivity of parameter discrimination healthy/sick
very high neutrality of 'gene'deletions but very different in different parameter sets (instantiations).

neutrality of deletions likelihood of deleterious effects
Rapid “Causal drift”

rapid change of sensitivity to parameter changes (mutations) due to neutral drift

“genetic background”

“cause of disease”

cf GWAS studies 50% “explained”

Mouse models