Evolution to understand evolved signaling and metabolic pathways;

*evolution predictable???
“Our results reveal how a signaling pathway can orchestrate specific genome changes and demonstrate that the copy number of repetitive DNA can be altered to suit environmental conditions”.

How did genome architecture evolve to enable this “reversed” causation, and what are the evolutionary consequences?
Genome organization of Ribosomal DNA and transcription/translation in YEAST

high nutrient condition enhance transcription rates

high transcriptional load induces mutations (transcription/replication conflicts

Replication Fork Barriers: in tandem Ribosomal DNA divert these mutations to DupDels’s (instead of SNP’s

variable length tandem repeats of rRNA-genes
no direct fitness effect of number of genes

TOR pathway
biases toward double stranded break repair towards DUP’s

Jack CV, et al.. PNAS 2015

model these properties and study their effect on long term evolution
Background mutations + transcription induced mutations inactivations, DupDels

Volume: number of macromolecules
Division Volume scales with genome size
total transcription polymerase limited (no dosage effect)
Given transcription translation conflicts bias towards DUPDELS evolves and prevents genome ans fitness degradation

High transcription load evolves, and therewith much transcription-translation conflicts
No pre-imposed transcription induced mutations

high level of background mutations $\rightarrow$

high levels of transcription mutation evolve
short term evolution favors duplication of active genes leads to higher per genome mutation rates, and genome degradation

No (little) fitness effect for deleting rRNA genes (like in Yeast)

\[ \text{Population distribution} \]

\[ \text{Ancestors in corresp. population} \]

\[ \text{All ancestors} \]
feedback evolving transcription initiation and (inactive) gene accumulation

evolution of $k_0$, $k_1$: (early) (late)

enzymes (0.099, 0.014), (0.54, 0.39)

housekeeping (0.11, 0.0078), (-0.24, 0.33)

ribosomal RNA (0.12, 0.031), (0.13, 0.69)

ribosomal prot. (0.11, 0.025) (0.14, 0.21)
conclusions:

Higher mutation rates (DUPDELS) evolve to prevent genome and fitness degradation (cf error threshold)

“Not all mutations are created equal”

Intra-cellular genome dynamics (DUPDELS) can counteract genome deterioration

Number of “good genes” preserved by intercellular competition

Increase of “bad genes” not prohibited by intercellular competition alone need high DEL rates (intra cellular dynamics) to compensation inactivations.

Genome organization has evolved in yeast to do so (i.e. bias to DUPDELS of transcription induced mutations by replication fork bariers/TOR pathway)
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” $V = 0.076x10^{-4}$ or ”deseased” $V = 0.33x10^{-4}$
(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 PI3P, 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 (IRS1IP and IRS1SP). The strength of this feedback is encapsulated by parameters \( f_{PKCZP} \) and \( f_{AMP} \), respectively. See Methods for details. b)
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 \( P_{345} P_3 \), 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)
sensitivity of outcome to parameter changes for different “healthy” instantiations of the pathway

sensitivity of parameter discrimination healthy/sick for different neutral instantiations
very high neutrality of 'gene' deletions but very different in different parameter sets (instantiations).
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
“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
\[ \frac{dC}{dt} = S\nu \]

- \( C \): Concentration
- \( t \): Time
- \( S \): Stoichiometric matrix
- \( \nu \): Flux vector

Steady-state assumption
\[ S\nu = 0 \]

LP formulation

Objective: \( \max Z = \nu_5 \)

Constraints:
\[ \begin{bmatrix} R1 & R2 & R3 & R4 & R5 & R6 & R7 \\ \hline A & -1 & 0 & 0 & 1 & 0 & 0 \\ B & 1 & -1 & 1 & 0 & -1 & 0 \\ C & 0 & 1 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \nu_1 \\ \vdots \\ \nu_7 \end{bmatrix} = 0 \quad 0 \leq \nu_1, ..., \nu_7 \leq 10 \]

III. Hypothetical flux distribution at steady-state

\[ Z = 10 \]
\[ \nu = [6.67 \ 3.33 \ 6.67 \ 6.6710.0 \ 3.33 \ 6.67]^T \]
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 → volume increase (decrease surface/volume ratio)
  volume = 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 (\( \gamma \))

external flux

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

internal flux

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

evolutionary protocol

- 9 types of environments (available nutrients). realized in different concentrations
- per generation 1 environment seen
- pop size 100: flux dependent replication
dead: 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 <em>glycolysis pathway</em></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 <em>glycolysis &amp; transport</em></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><strong>Observed Outcome of WGD</strong></td>
<td><strong>Expected Outcome of WGD</strong></td>
</tr>
</tbody>
</table>

WGD enabled to exploit high glucose resource during emergence of angiosperms.

During emergence of angiosperms, WGD usually advantageous except in “new” environments. Seldom better than single INDELS. Evolutionary outcome “deterministic.”
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