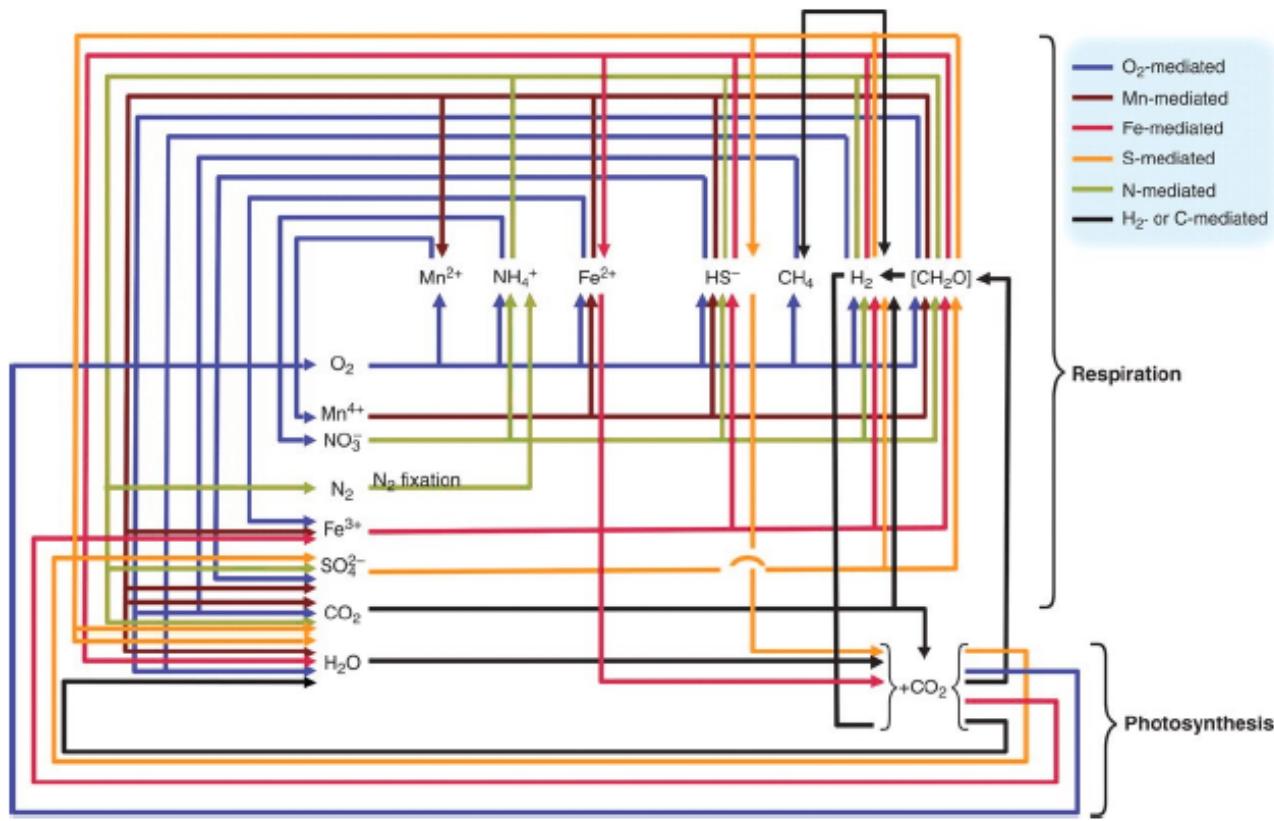


“Metabolism only”

Metabolic networks
how to cope with complexity of 'real' organisms

(1) Life is..... energy/nutrient cycling

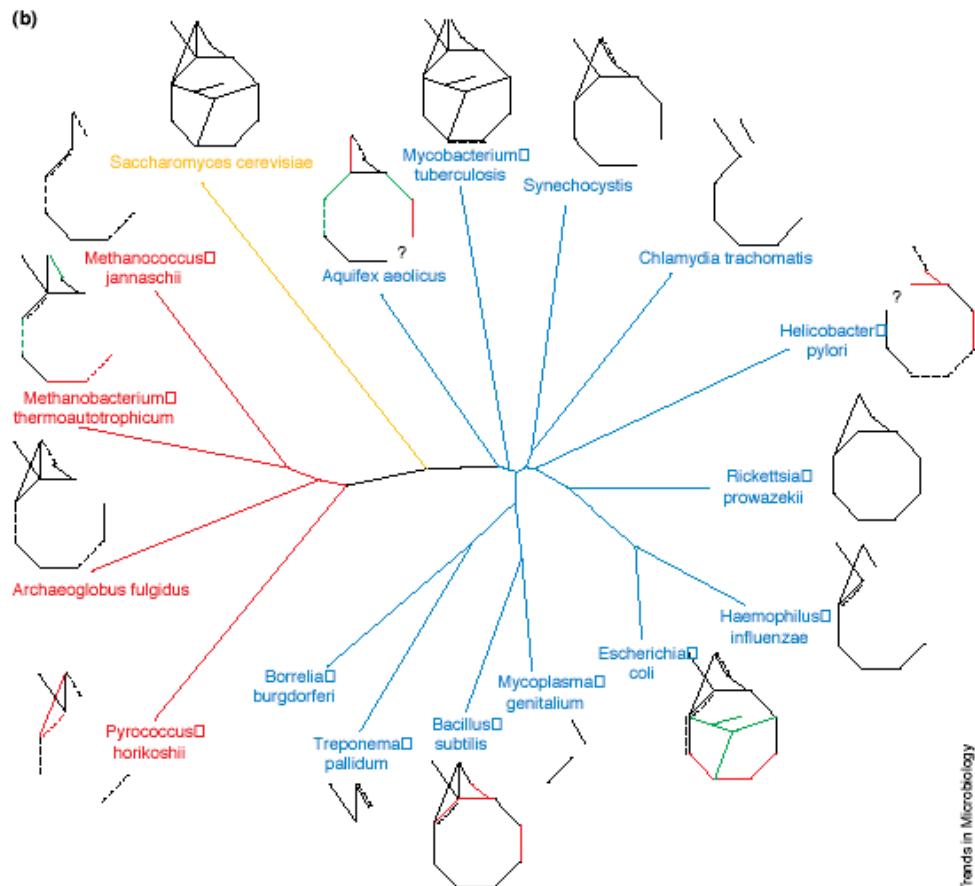


"The individual taxonomic units evolve and go extinct, yet the core machines survive surprisingly unperturbed."

PG Falkowski et al, Science 2008

energy/nutrient cycling: individual vs ecosystem based complexity

example fragments of TCA cycle



Huynen et al TMB 1999

Metabolic networks: Exploiting constraints

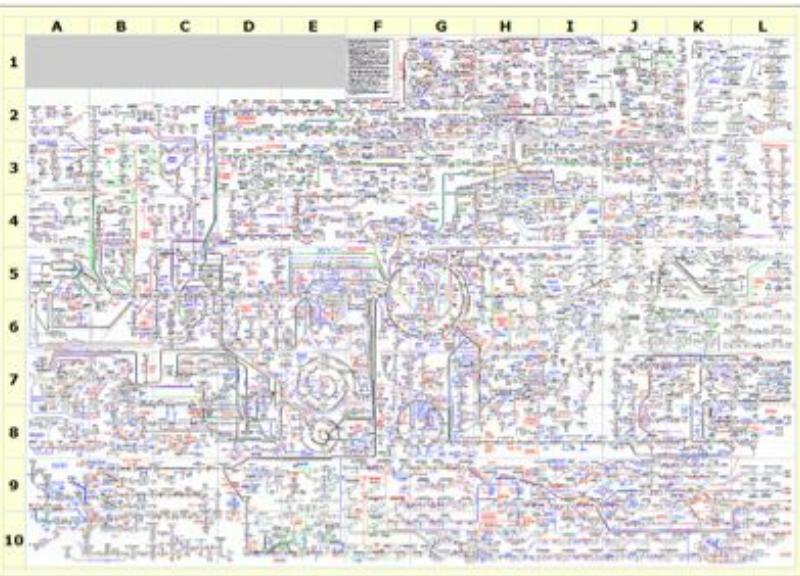
metabolic network are evolved

However metabolic networks many physical/chemical constraints
stochiometry, energetic constraints

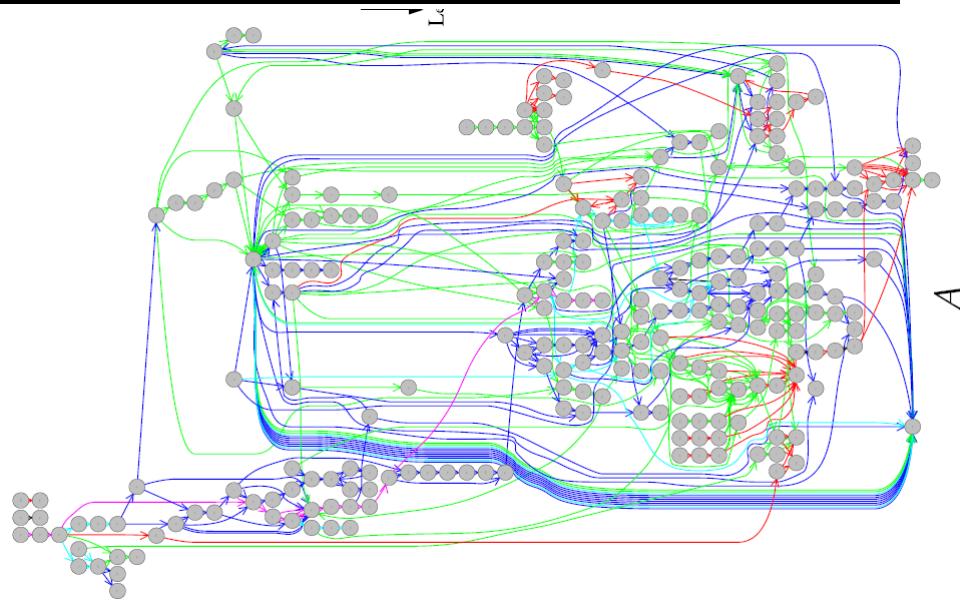
can/should be exploited

allows for model upscaling to complexity of present day organisms

**stochiometric constraint, + equilibrium assumption
allows calculation of (optimal) flux through large
metabolic networks**

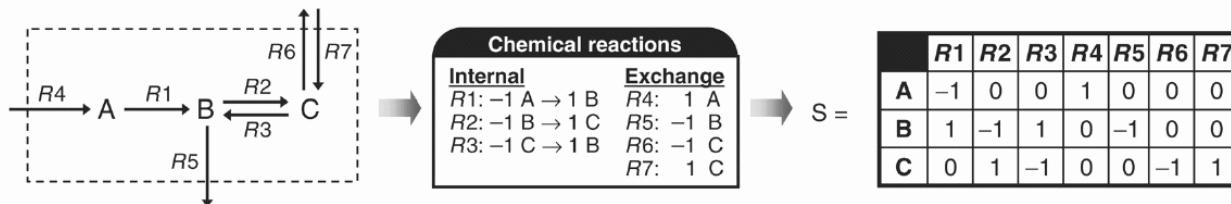


KEGG database



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: $\max Z = v_5$

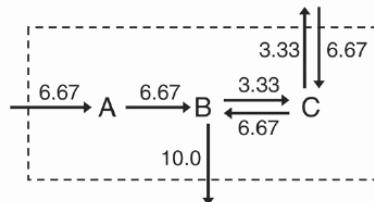
Constraints:

$$\begin{array}{l}
 \text{A } \begin{bmatrix} R1 & R2 & R3 & R4 & R5 & R6 & R7 \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_7 \end{bmatrix} = 0 \\
 \text{B } \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_7 \end{bmatrix} = 0 \\
 \text{C } \begin{bmatrix} 1 & -1 & 1 & 0 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_7 \end{bmatrix} = 0 \\
 \end{array} \quad 0 \leq v_1, \dots, v_7 \leq 10$$

III. Hypothetical flux distribution at steady-state

$$Z = 10$$

$$v = [6.67 \ 3.33 \ 6.67 \ 6.67 \ 10.0 \ 3.33 \ 6.67]^T$$



Using FBA to asses Evolvability of metabolic networks analysis of GP map in E coli (A Wagner)

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 -->
percolating neutral networks
seeing novel phenotypes

ONLY RNA...? (cont.)

no ...

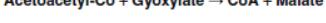
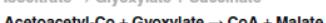
also 'real' metabolic network

Metabolic networks
and their evolution.

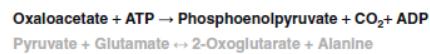
Wagner A.

Adv Exp Med Biol. 2012;751:29-52.
(evolutionary systems biology)

Metabolic genotype
(network of enzymatic reactions)

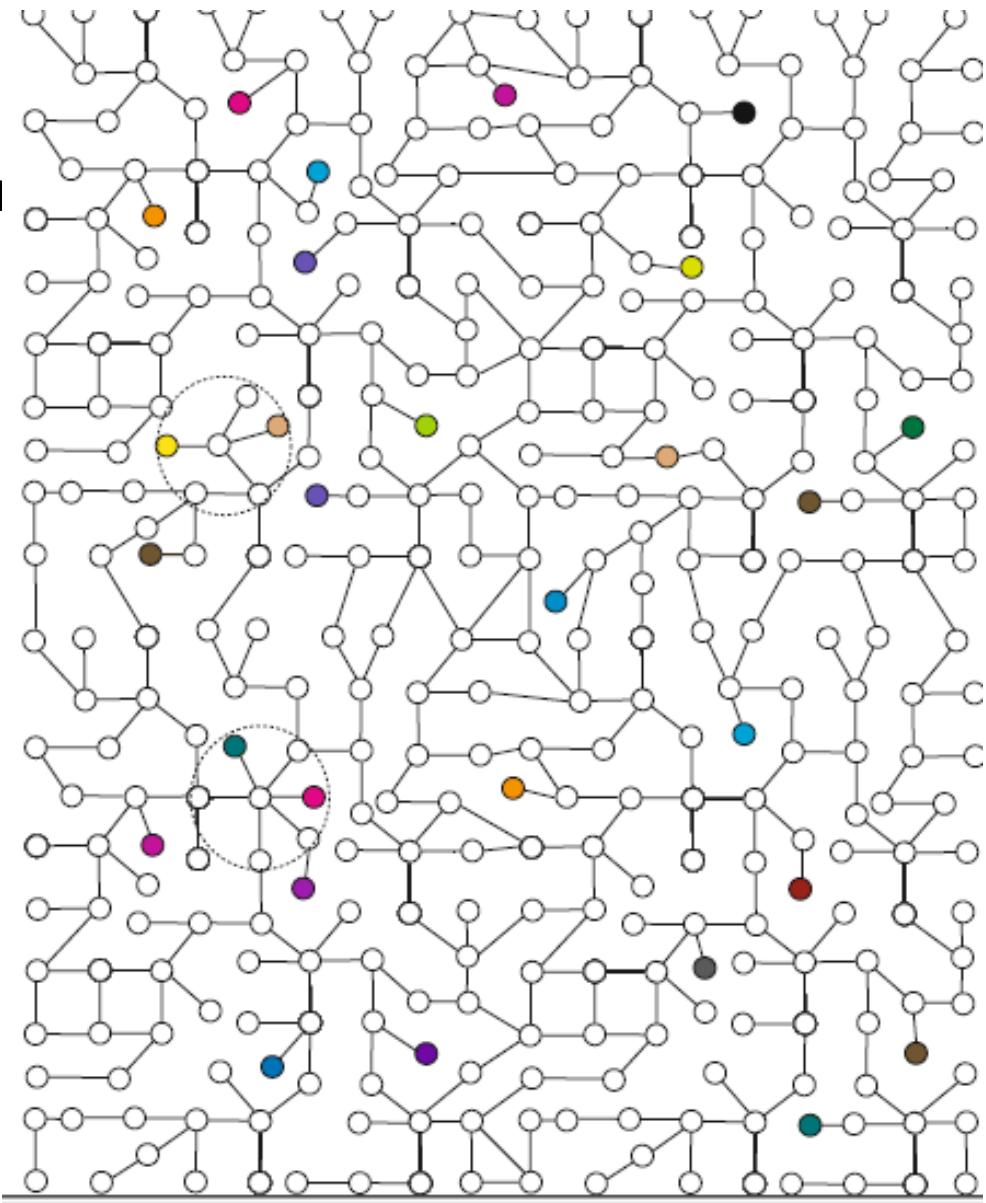
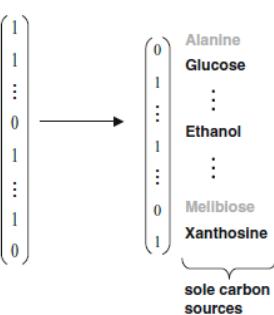


\vdots



>5000 biochemical reactions

Metabolic phenotype
(viability on carbon source)

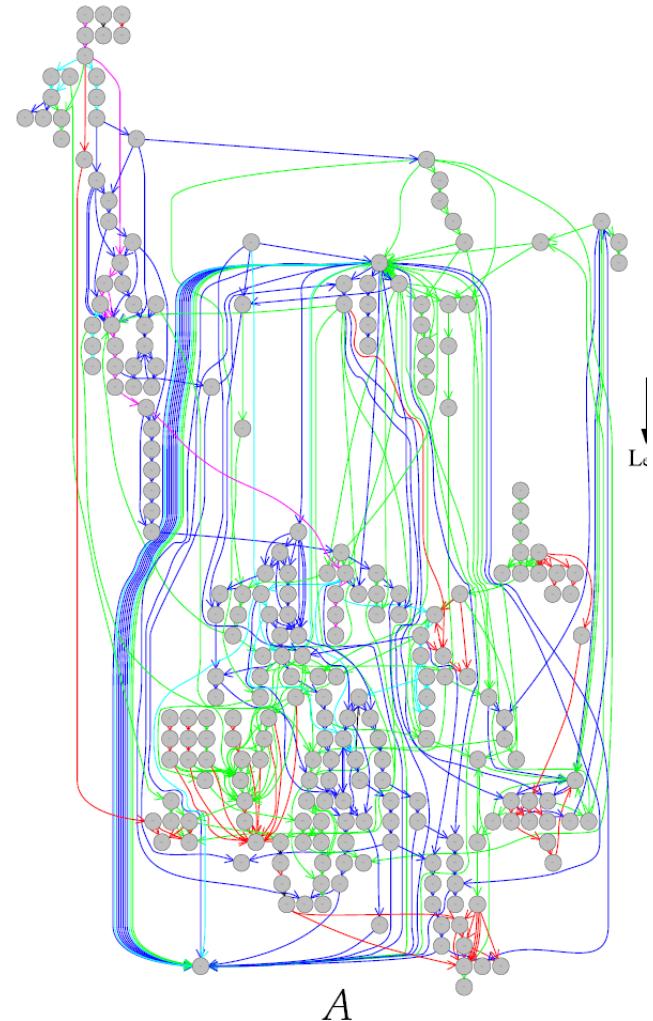


Using FBA to reconstruct evolution of metabolic network of yeast after WGD

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
- stoichiometric matrix ($nA -> mB$)
- reactions coupled to enzymes
- set maximum flux (when enzymes are present) (OR, AND reactions))
- however flux accrual 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?

How do fluxes (growth) change after endosymbiosis?
(cf Pal Papp Nature 2006)

How do genomes reduce after whole genome duplication?
cf van Hoek and Hogeweg 2009)

reconstruction of Ecosystem wide metabolomee (cf Bas Dutilh)

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 en-
vironment 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

WGD, cell size and fluxes

cell size scales with amount of DNA

Cavallier Smith (e.g. 2005)

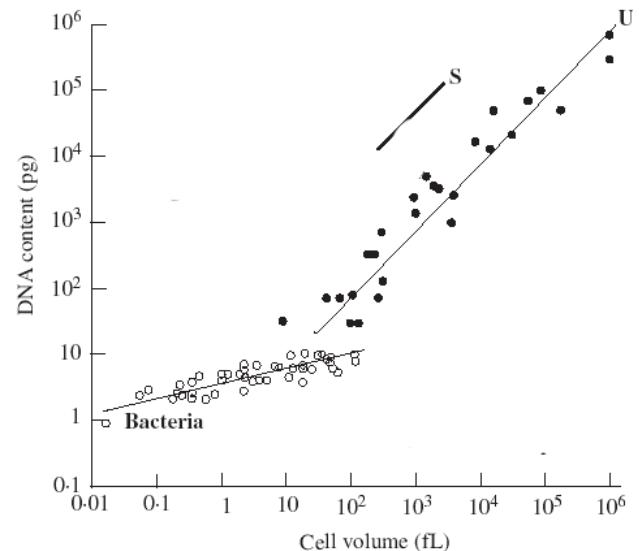
In Yeast diploide cells are:

$V = 1.89 * \text{haploid cells}$

surface: $1.56 * \text{haploid cells}$

$V = N^{.9}; A = V^{.7}$

where N number of genes



MaxFlux change as function of area change (α), volume change (β) and gene dosage change γ)

external flux

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

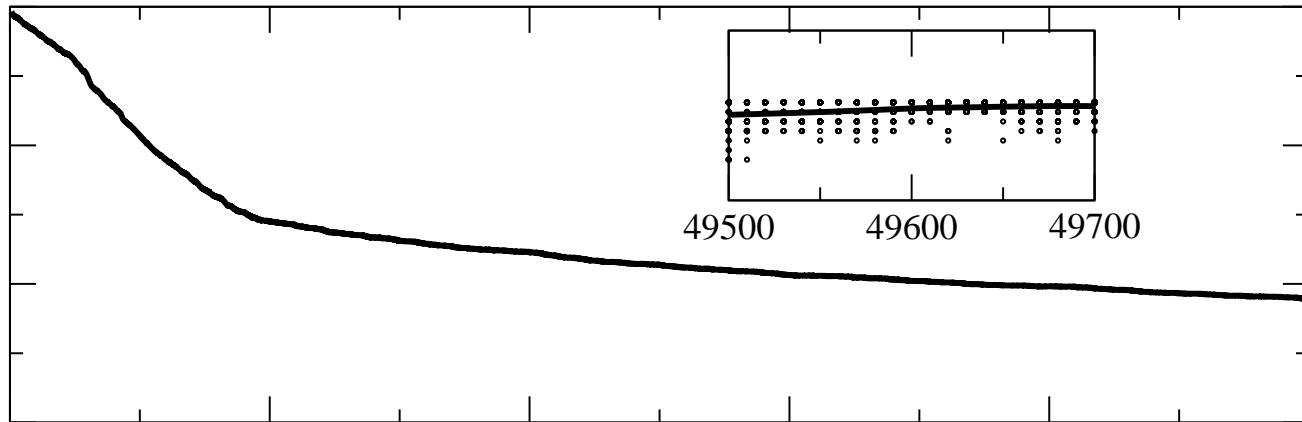
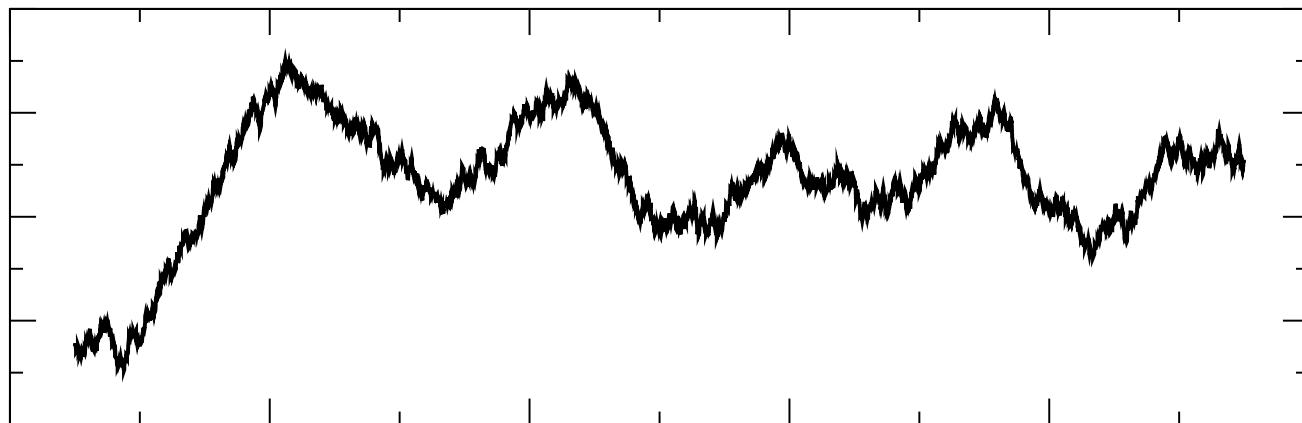
internal flux

$$F_{\max}(i) = F_{\max,0}(i) \frac{\gamma(i)}{\beta}$$

evolution of metabolic flux after WGD evolutionary protocol

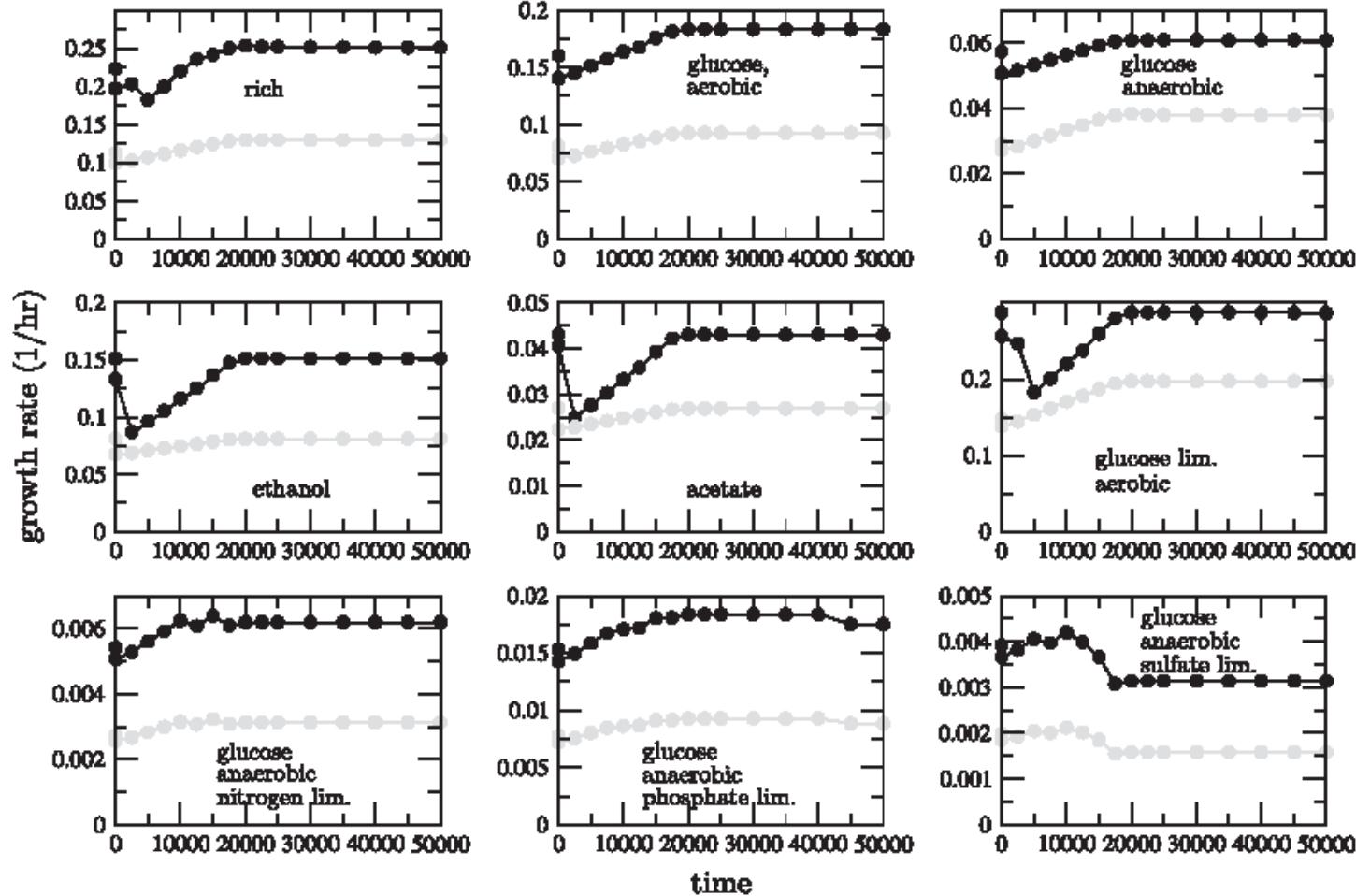
- 9 types of environments (available nutrients). realized in different concentrations
- per generation 1 environment seen
- pop size 100: flux dependent replication death: nogrowth + random
- after wgd: only deletions or duplication + deletion (max 2 copies)
- no fitness advantage for genome schrinkage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

evolution of metabolic flux after WGD
evolutionary dynamics: growth rate and genome reduction

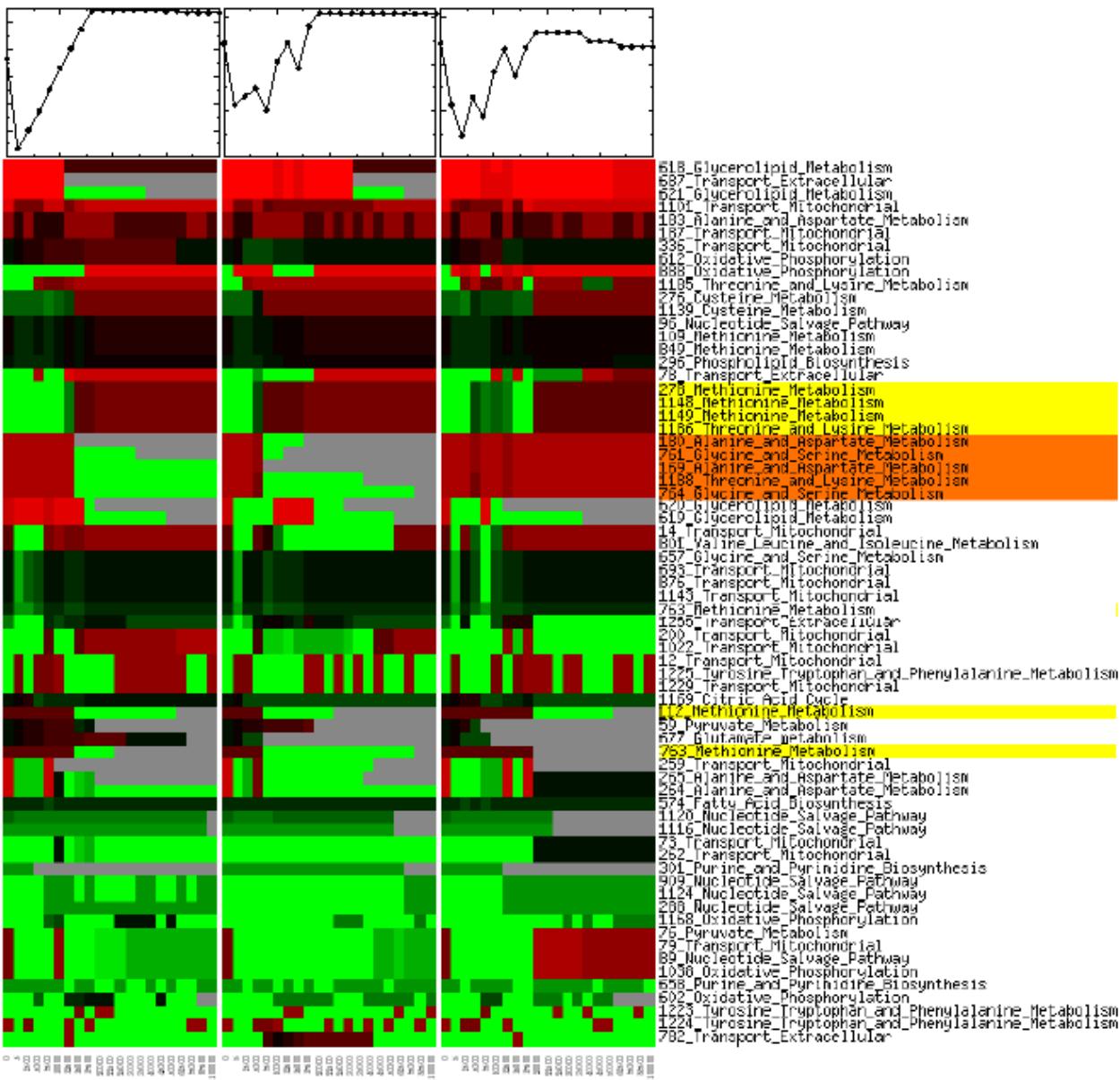


**evolution of metabolic flux after WGD
flux in the various environments
(max and mean concentration)**

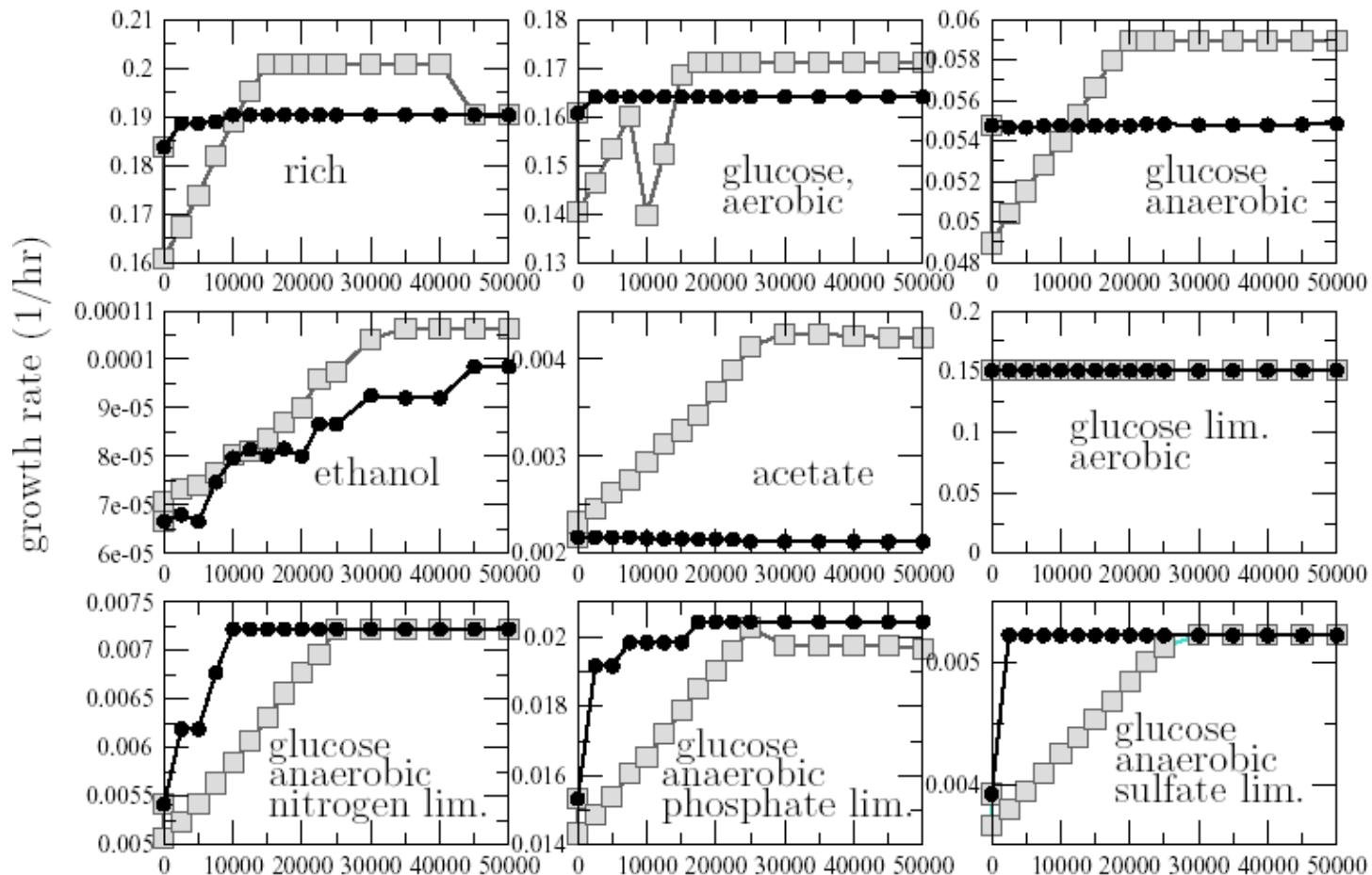
**initial decrease – how/when does it happen in
evolution**



Genome schrinkage after whole genome duplication dynamics of use of pathways in anaerobic glucose environment (env 3)



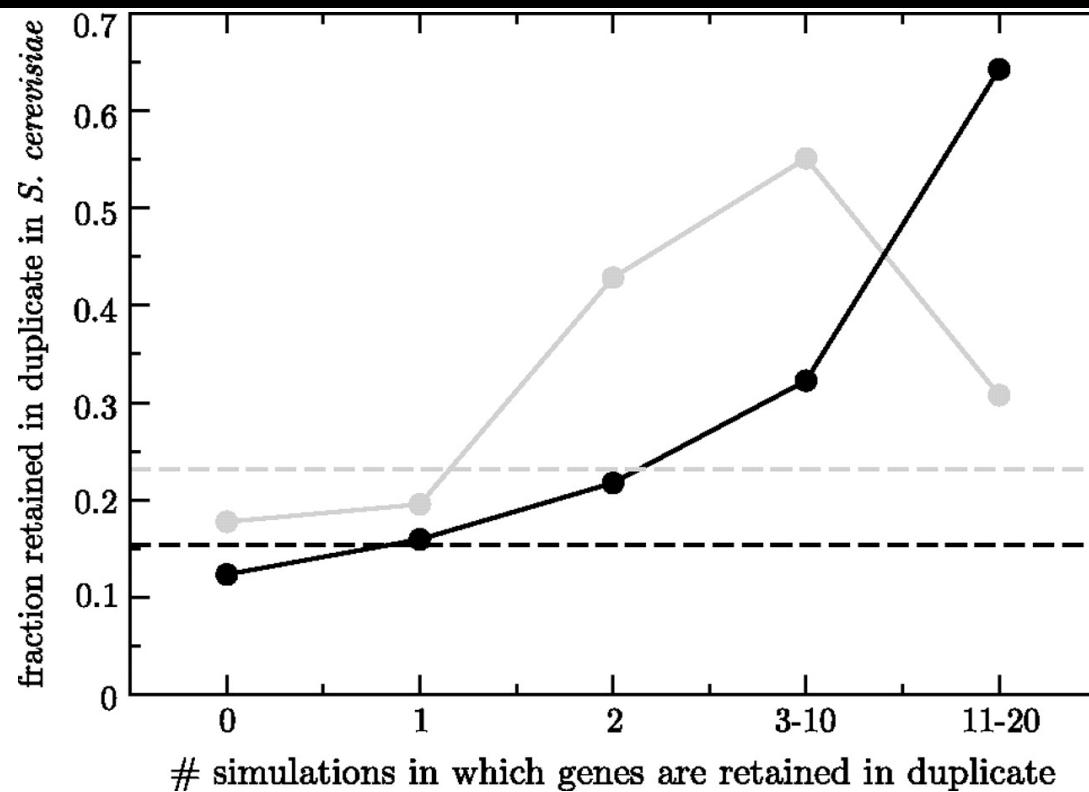
Only in “new” environment - no direct disadvantage of WGD



BUT single INDELS initially better Except 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

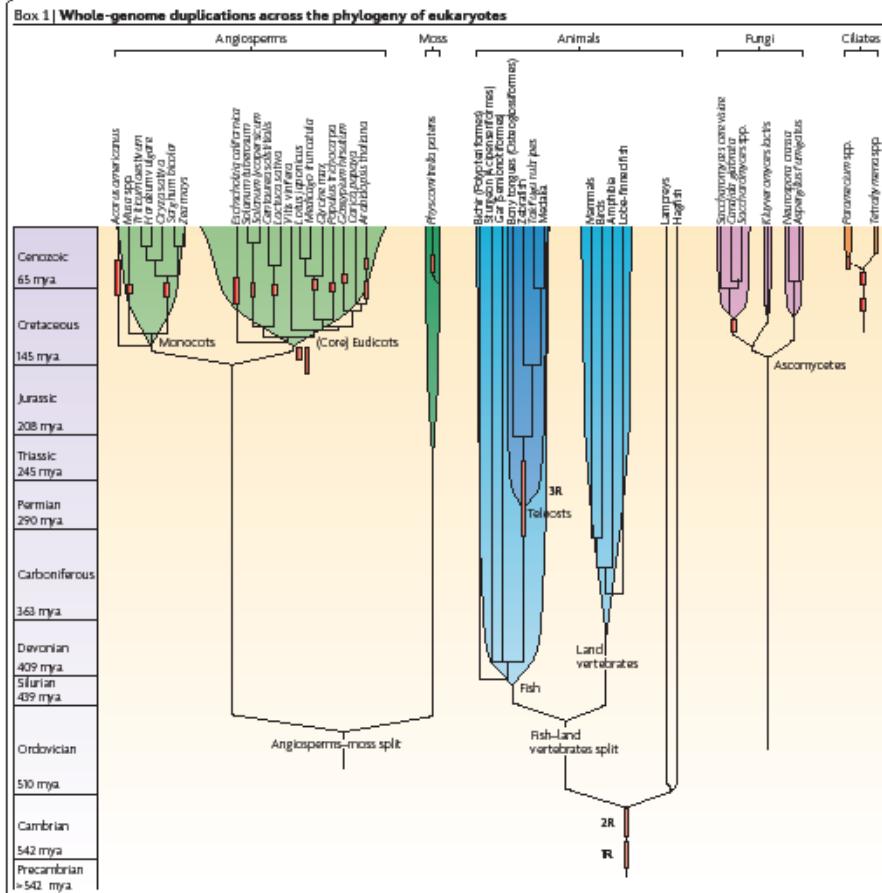
Evolution predictable!

conclusions

supervised vs non-supervised modeling of WGD in Yeast

“Supervised” Conan & Wolfe (2007) find genes preferentially retained <i>glycolysis pathway</i> Model glycolysis pathway assuming dosis effect of duplicated genes demonstrate WGD can lead to increased glycolic flux	“Non Supervised” van Hoek & H. (2009) take known interactions metab. net + DNA-volume relation model evolution find preferentially retained genes <i>glycolysis & transport</i> WGD mostly disadvantaeous initially except in “new” environments seldom better than single INDELS evolutionary outcome “deterministic”
WGD enabled to exploit high glucose resource during emergence of angiosperms	WGD enabled to exploit high glucose resource during emergence of angiosperms
observed outcome of WGD	expected outcome of WGD

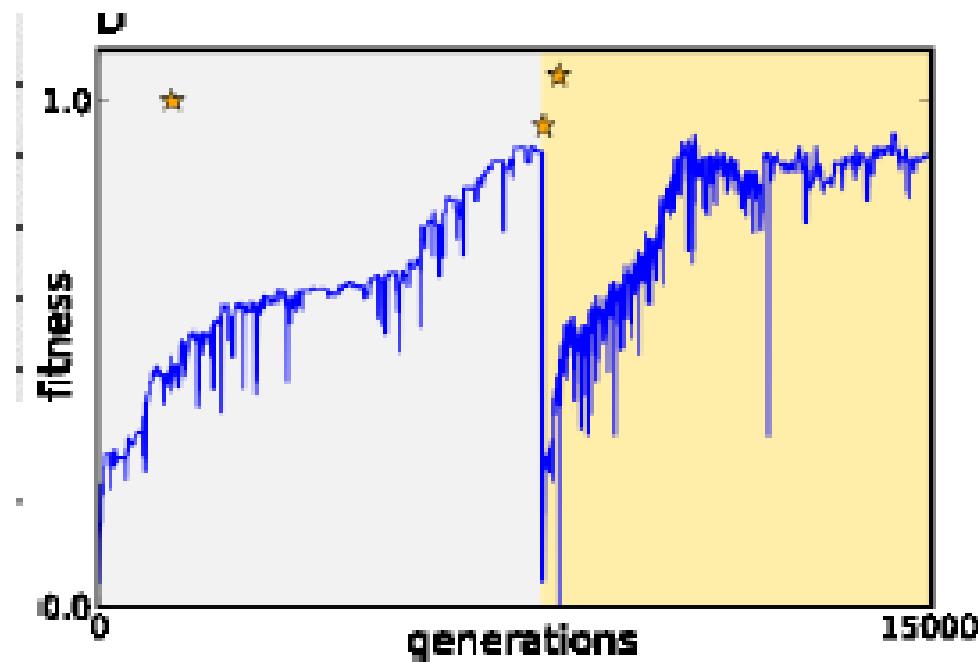
WGD observed in phylogeny at times of environmental shifts



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

**automatic reconstruction of metabolic networks from
annotated genomes cf Henry CS1, DeJongh M, Best
AA, Frybarger PM, Lindsay B, Stevens RL. 2010**

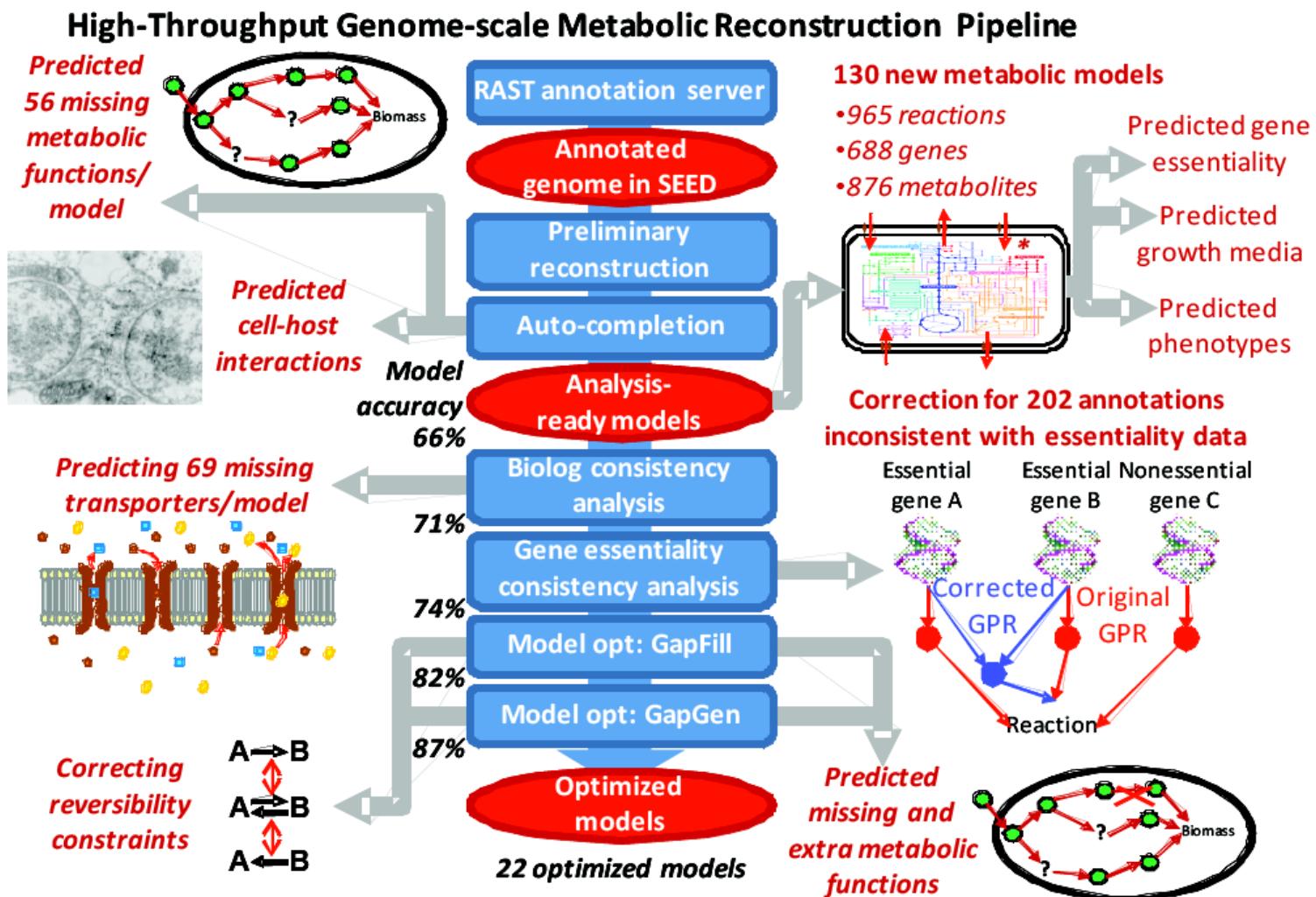


Figure S1: Overview of results and discoveries arising from the Model SEED pipeline.

FBA and the prediction of environmental metabolome

From metagenomic data - reconstruct species-abundances
Match species to reference genomes
reconstruct metabolic network of these species
Sample environments (semi-markov chain sampling)
predict flux toward growth of all species
equate relative growth to relatice abundnce
optimize correlation of observed abundances
with predicted abundaces by environmental sampling

(note: efflux of metabolites not used)

“Towards predicting the environmental metabolome from metagenomics with a mechanistic model” DR Garza, MC Verk, MA Huynen, BE Dutilh - Nature microbiology, 2018

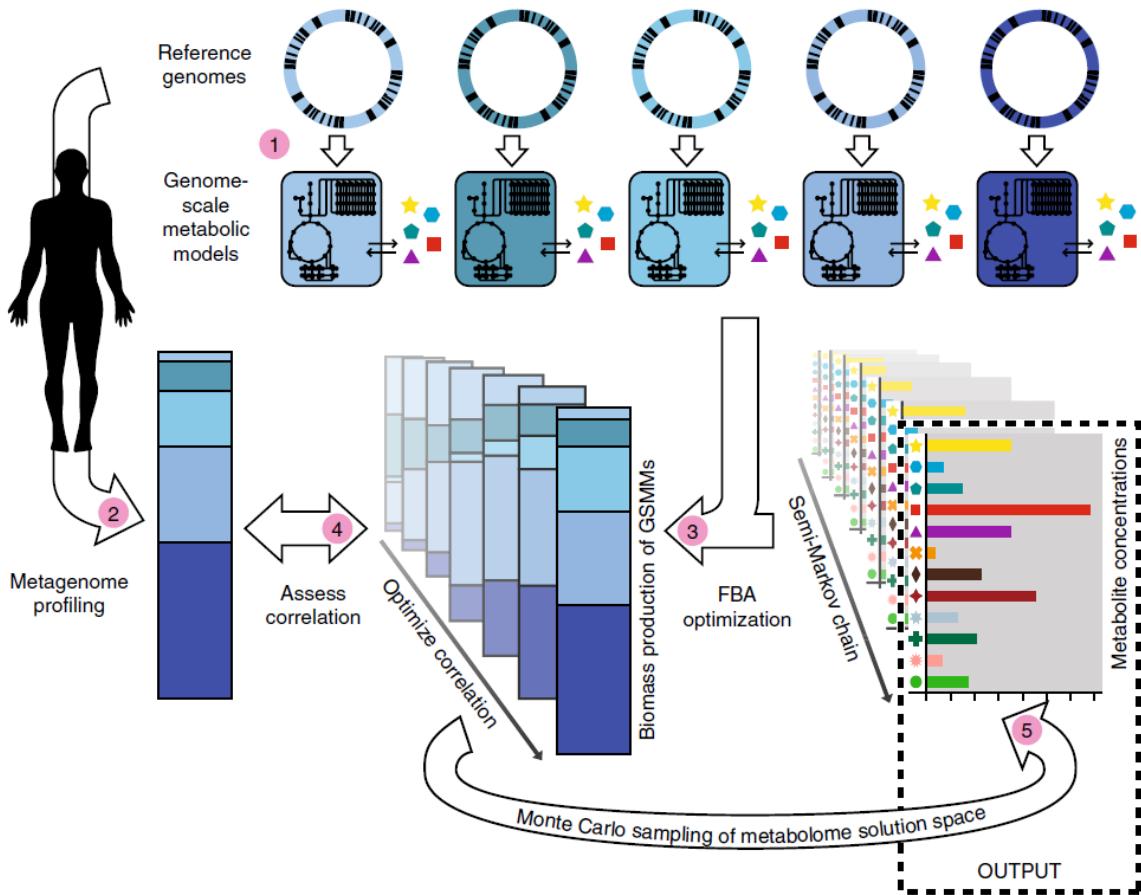
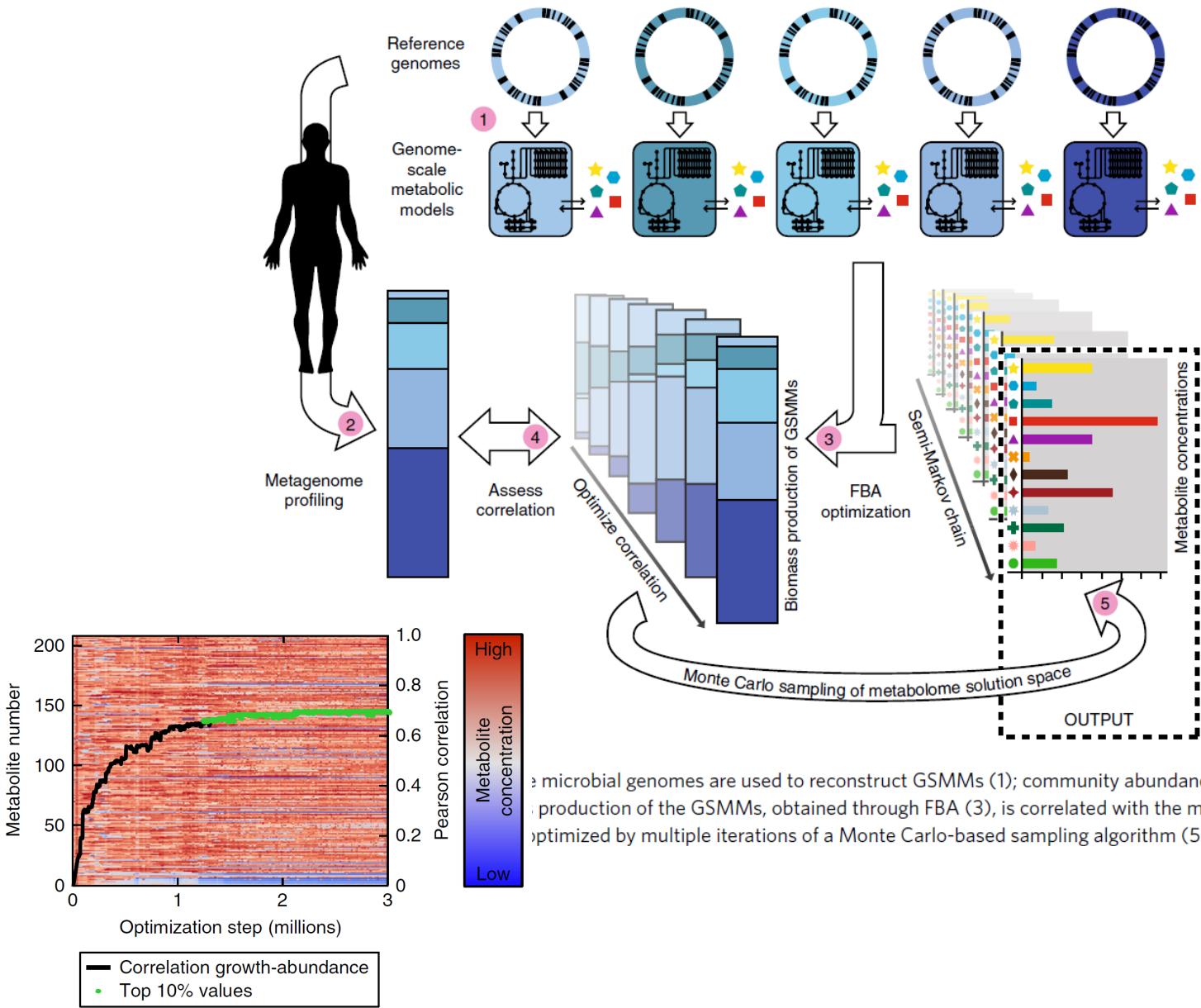
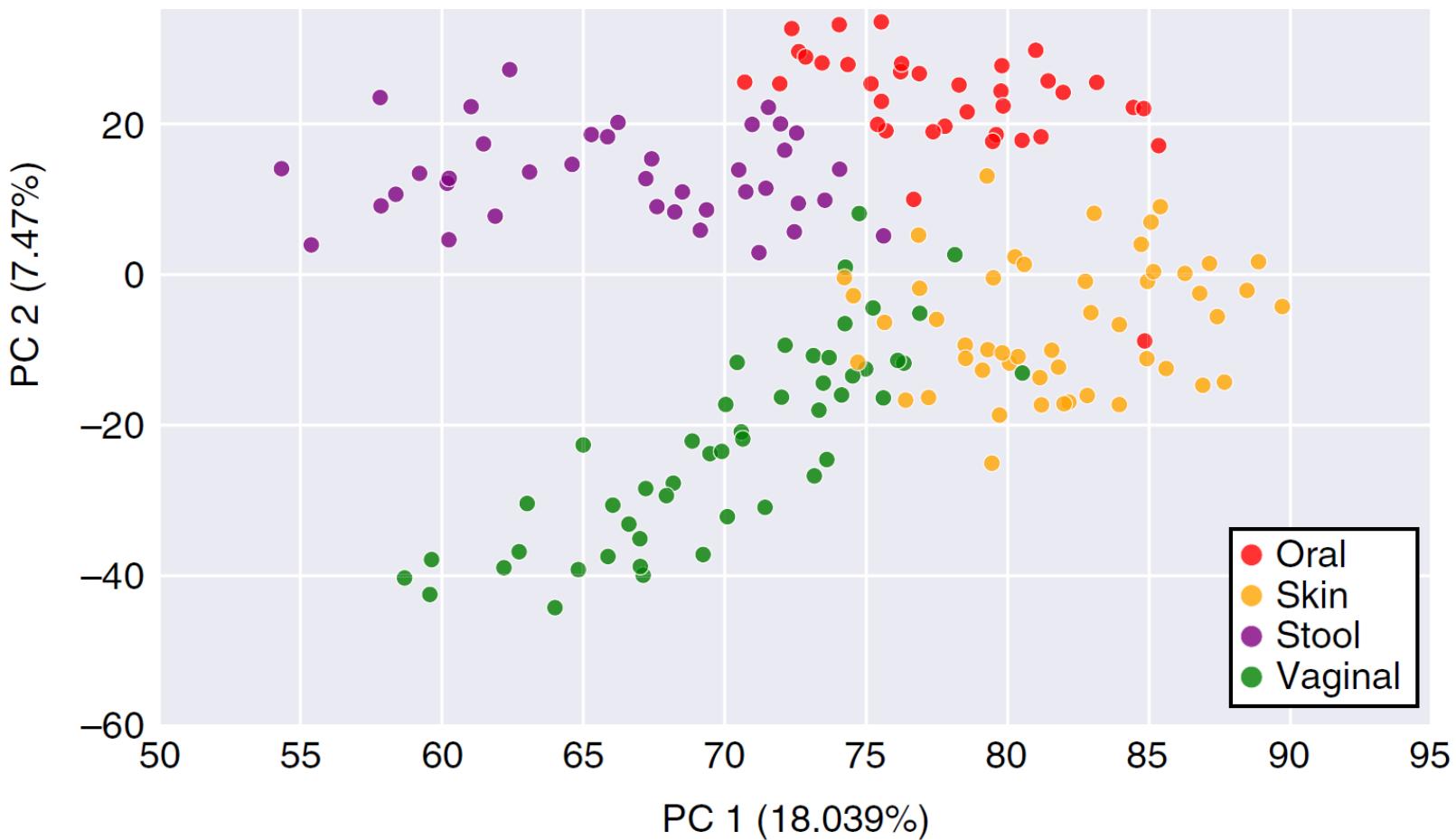


Fig. 1 | Overview of the MAMBO algorithm. Reference microbial genomes are used to reconstruct GSMMs (1); community abundance profiles are obtained through reference mapping (2); and biomass production of the GSMMs, obtained through FBA (3), is correlated with the metagenomic community abundance profile (4). This correlation is optimized by multiple iterations of a Monte Carlo-based sampling algorithm (5) (see main text)

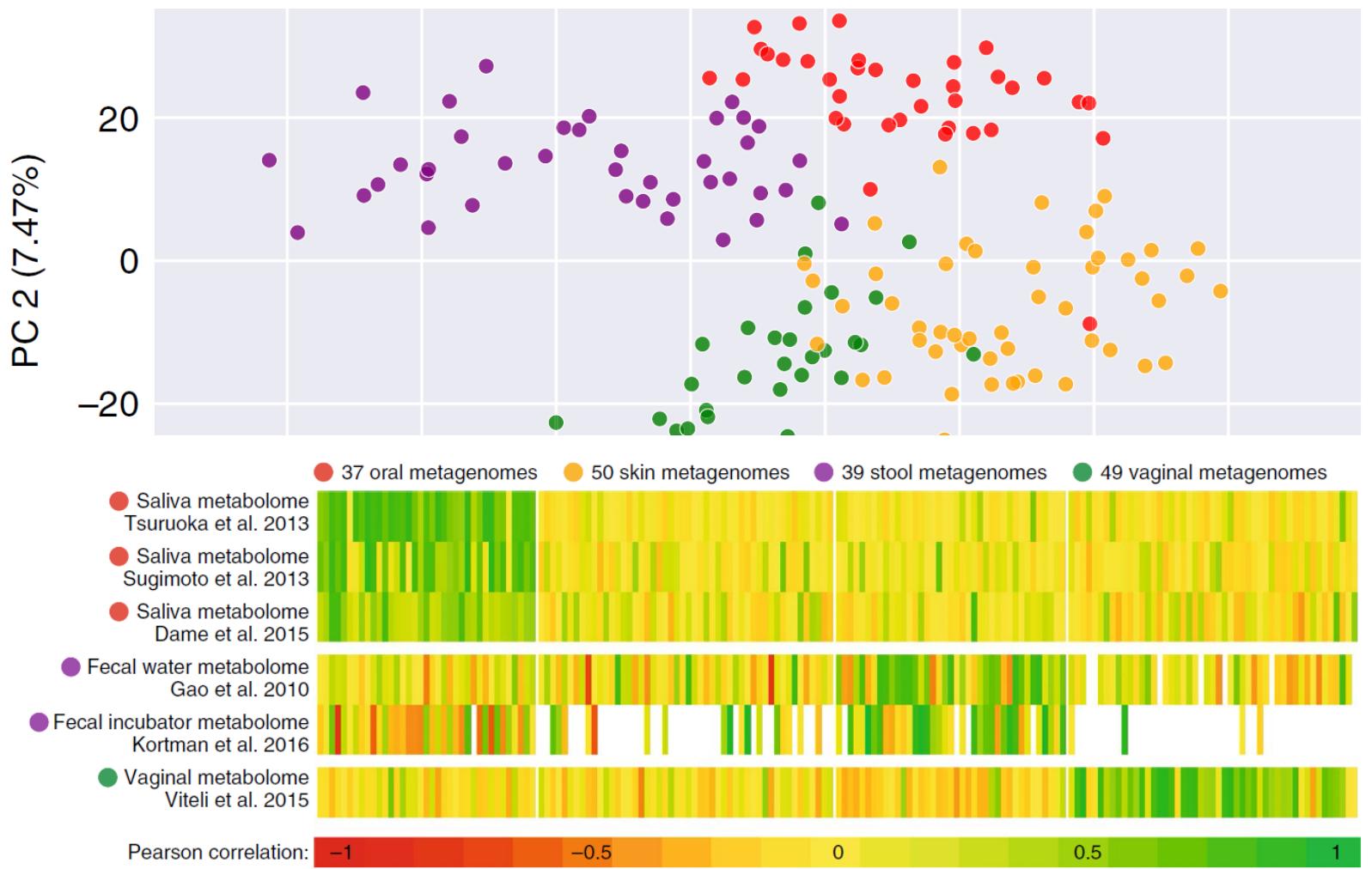


Microbial genomes are used to reconstruct GSMMs (1); community abundance profiles are produced through FBA optimization (3). The production of the GSMMs, obtained through FBA (3), is correlated with the metagenomic abundance profiles (4). This correlation is optimized by multiple iterations of a Monte Carlo-based sampling algorithm (5) (see main text).

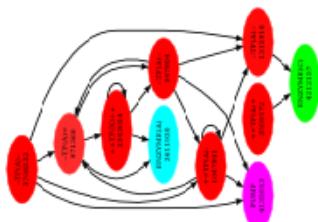
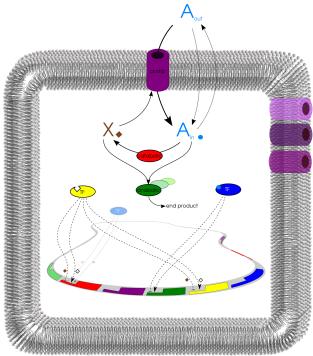
predicted metabolomes of different body parts cluster



predicted metabolomes correlate with measured ones

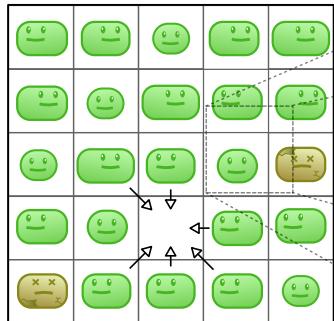


evolution of metabolomes: Virtual Microbes

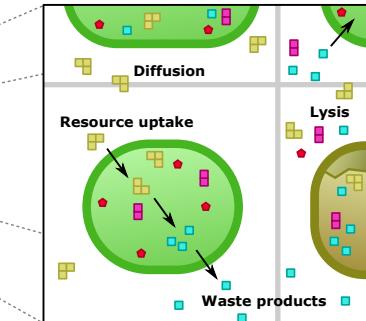


$$\begin{aligned} \frac{d[A]}{dt} &= \frac{[A]_{out}[X]V_{max,p}[Prot]_p}{([A]_{out} + K_{Ap})([X] + K_{Xp})} \quad (2) \\ \frac{d[X]}{dt} &= -\frac{d[A]}{dt} \quad (3) \\ \frac{d[A]}{dt} &= -\frac{[Prot]_c[A]V_{max,c}}{[A] + K_{Ac}} \quad (4) \\ \frac{d[X]}{dt} &= -\frac{d[A]}{dt} K_{new} \quad (5) \end{aligned}$$

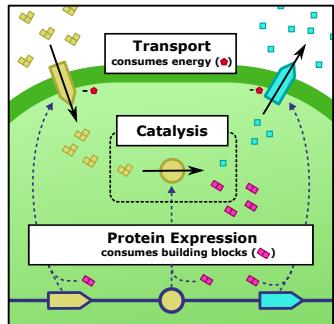
A Growth and division



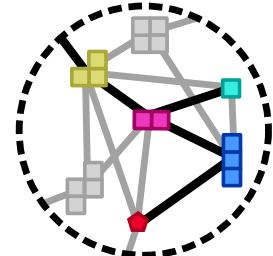
B Interplay with environment



C Cellular processes



D Microbes encode subset of reactions



extension rel to virtual cells

metabolic universe
import/export metabolites
spatial embedding
modification of environment
NO predefined fitness

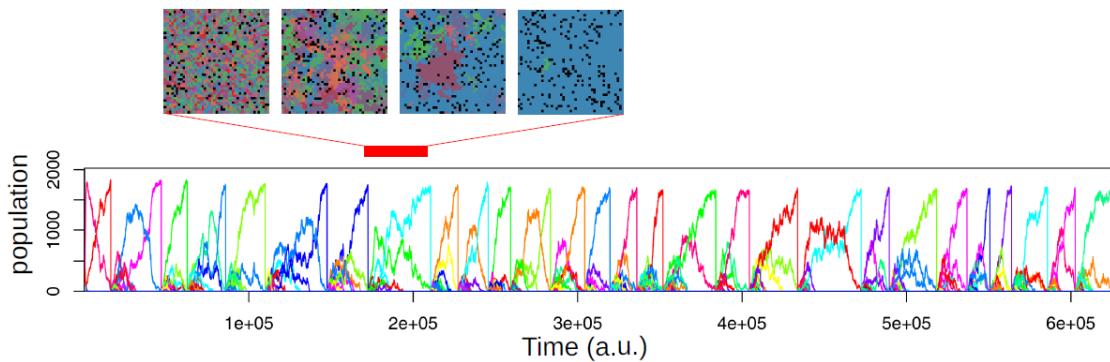
Individual vs ecosystem based metabolism

de novo evolution of Vmicrobes; constant homogeneous environment

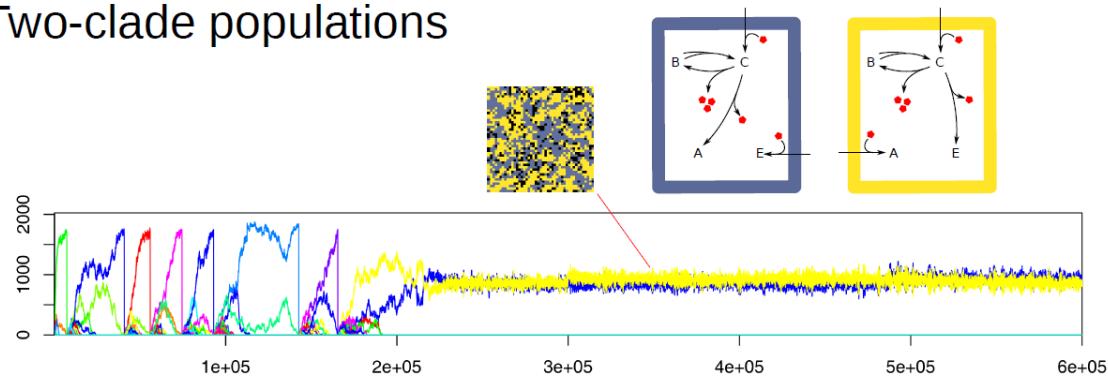
20 replicates identical except for mutational random seed

2 eco-evolutionary
attractors

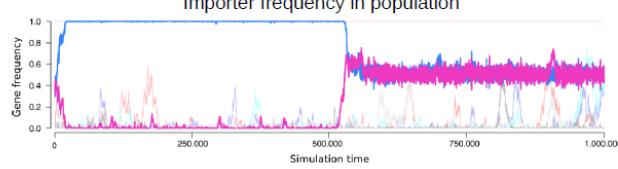
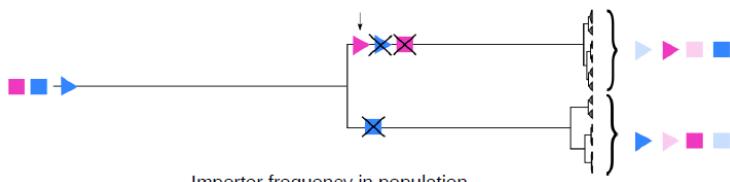
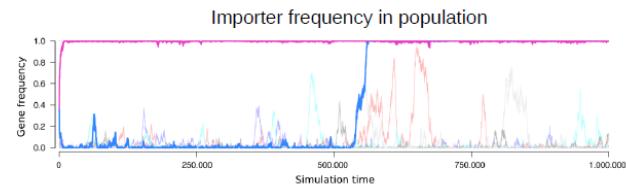
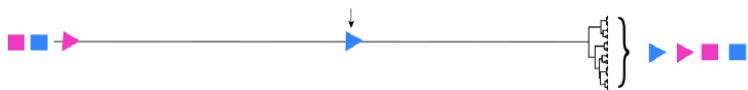
Single-clade populations



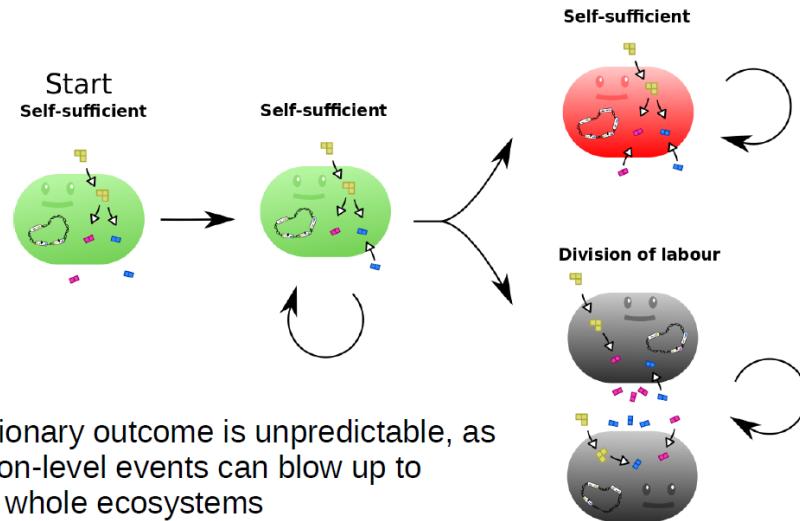
Two-clade populations



mutational cascade leading to alternative attractors

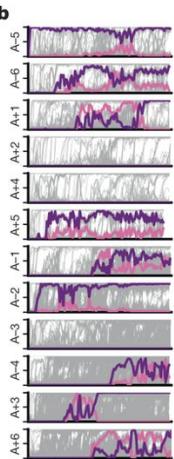
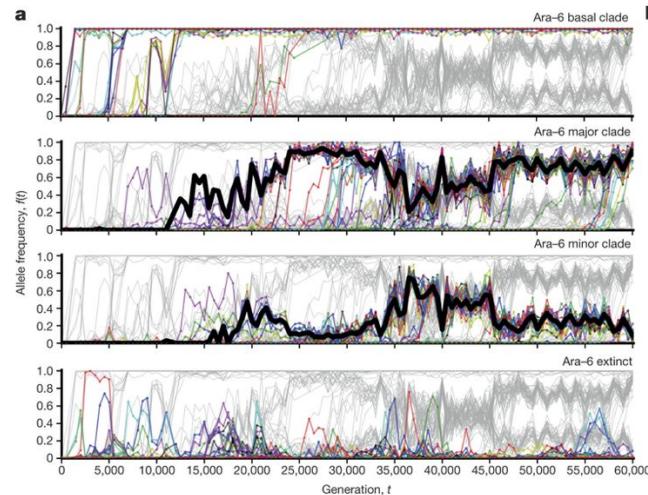
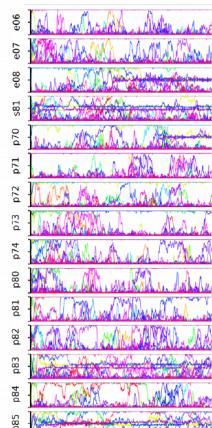
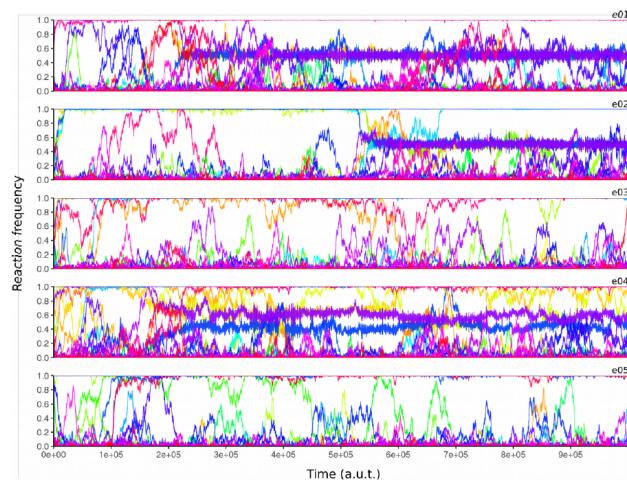


Many genotypes, but few eco-evolutionary attractors

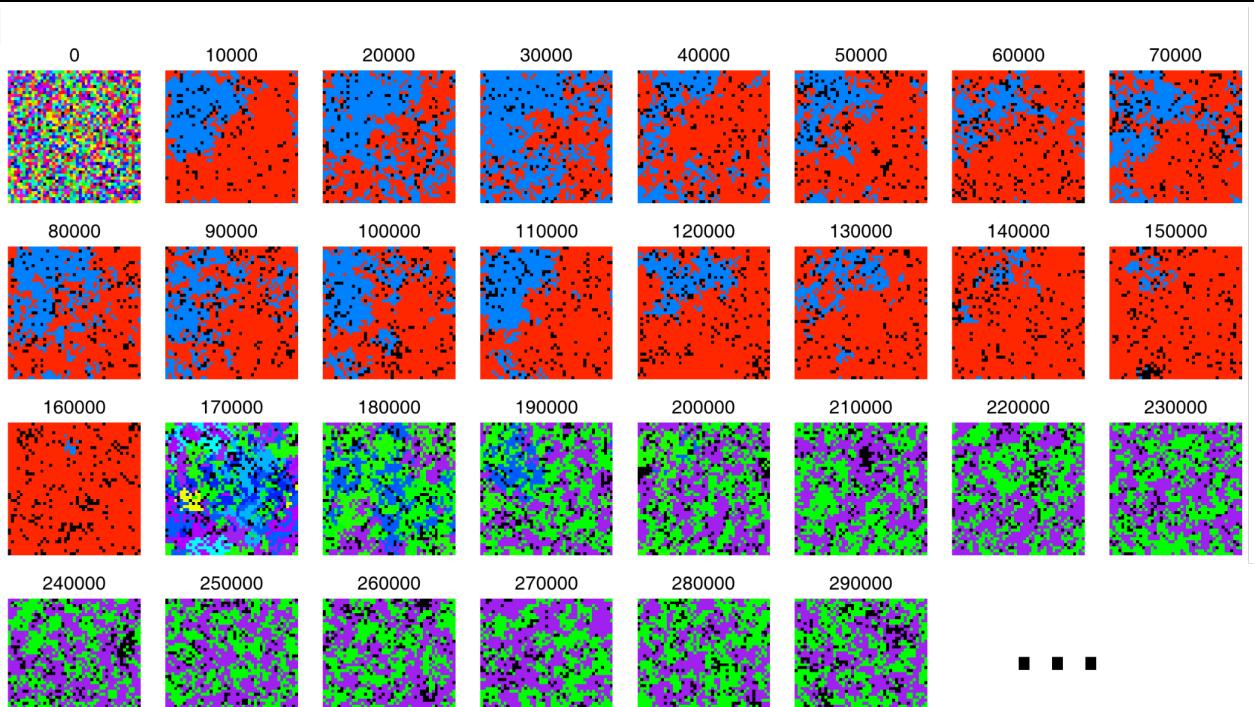


similar dynamics seen in LTEE experiment (Good 2017)

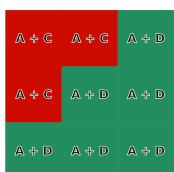
From: [The dynamics of molecular evolution over 60,000 generations](#)



Heterogeneous environment: evolution from niche specialization to niche deconstructions



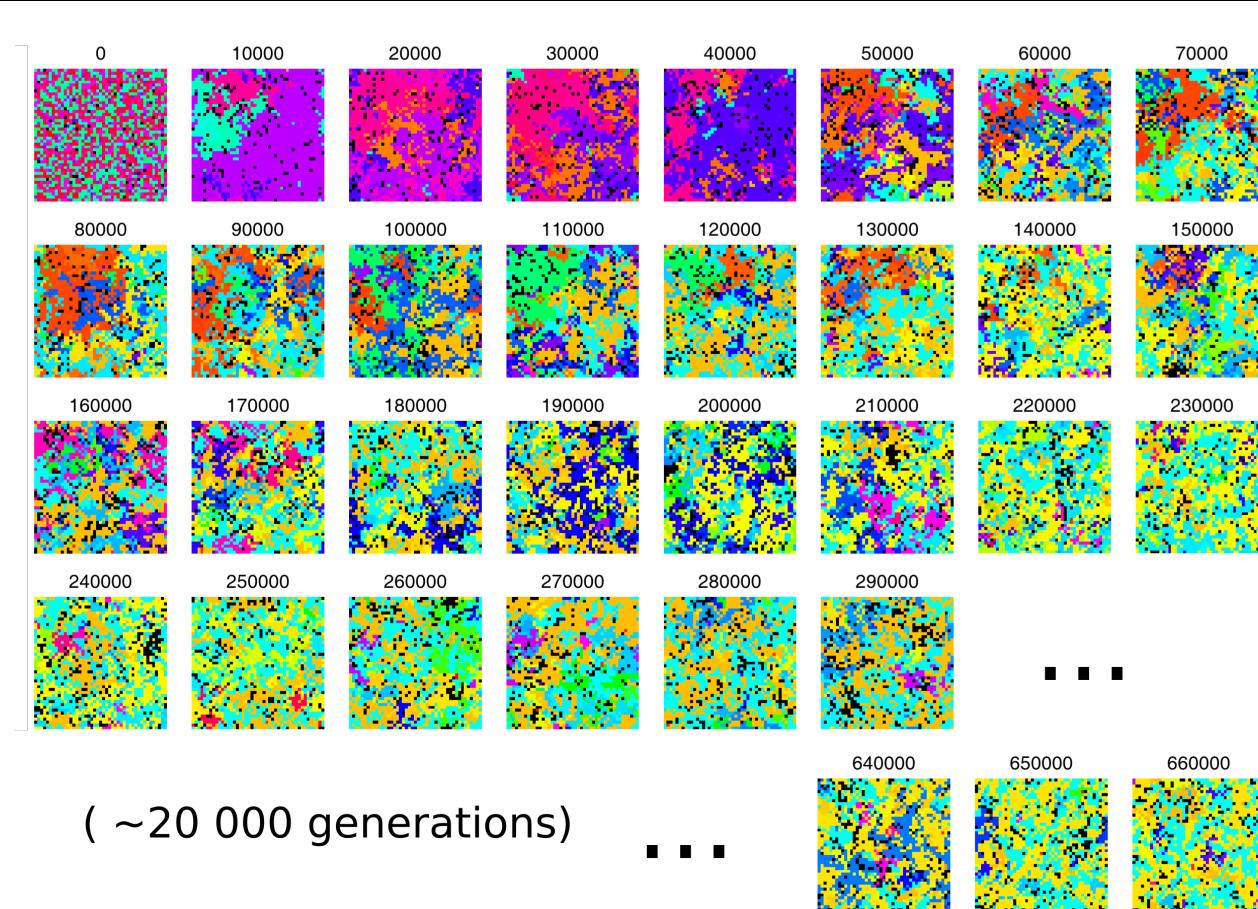
2 different niches
defined by
combinations of resources



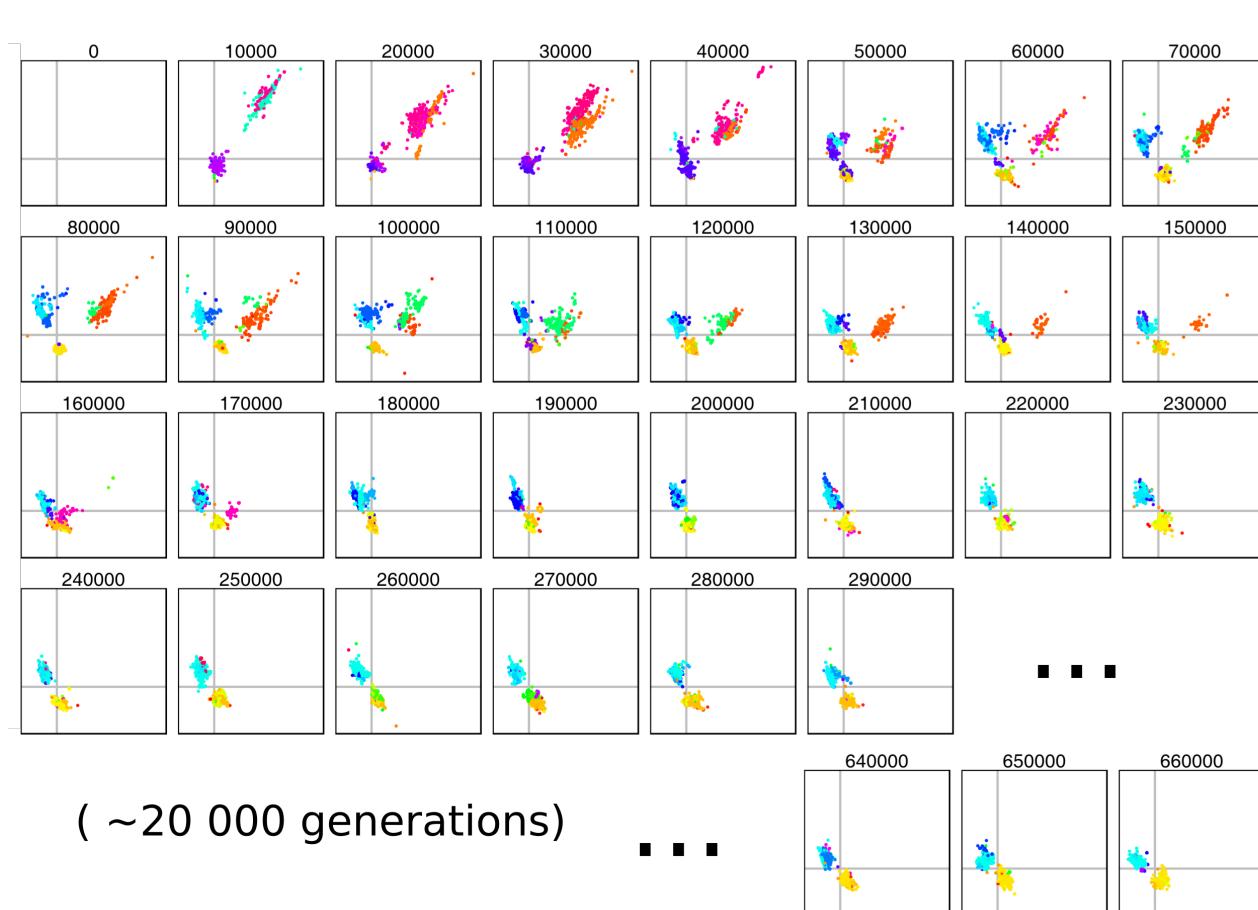
(~20 000 generations)

■ ■ ■

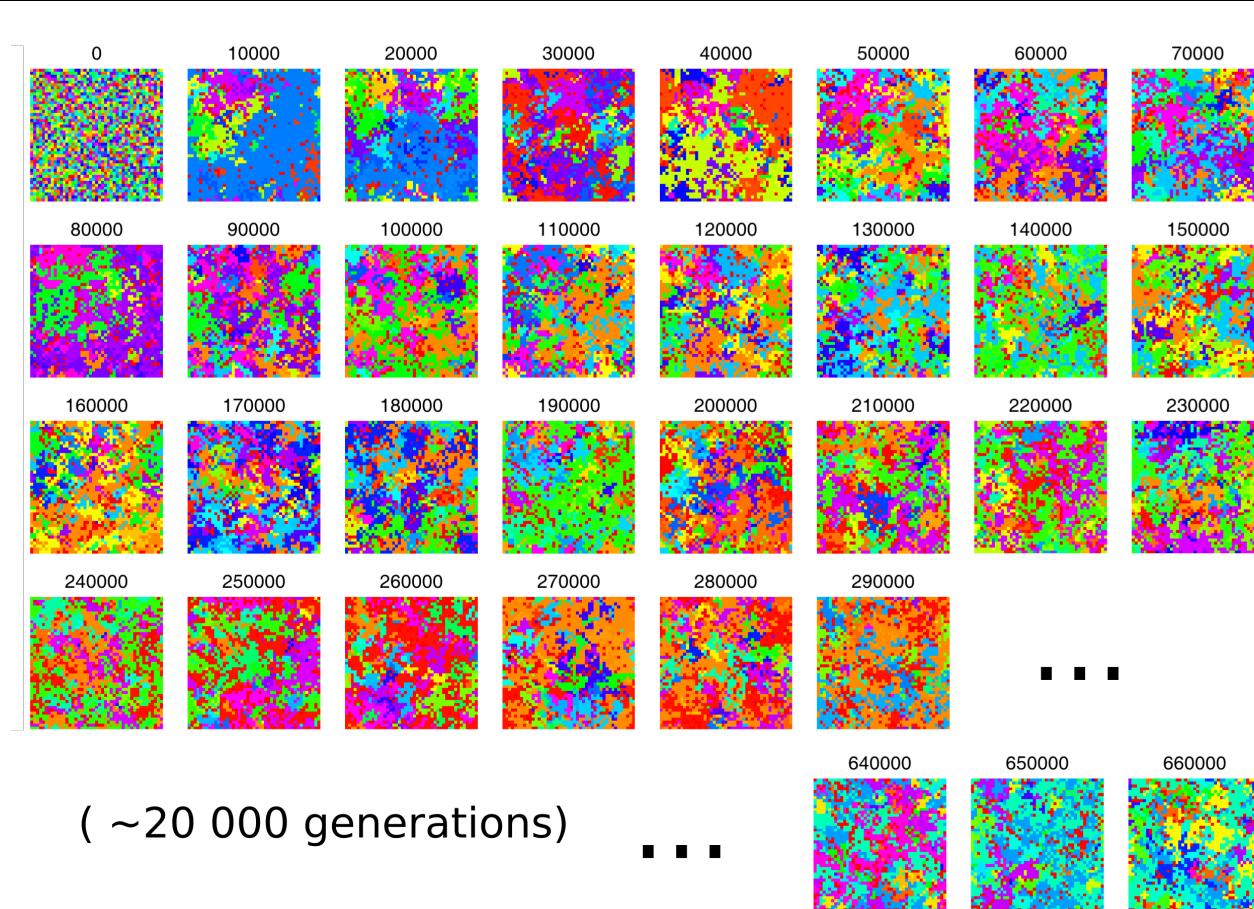
Niche deconstruction: Phenotypes (metabolic types)



Niche deconstruction: Phenotypes (metabolic types)



Niche deconstruction: Genotypes

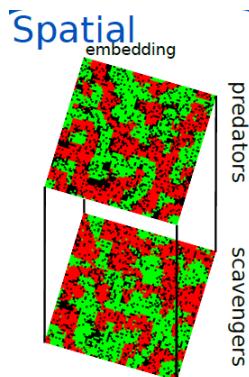


conclusions: individual vs ecosystem based metabolism
“doing it alone or together”
alternative attractors by random mutational (ordered) events

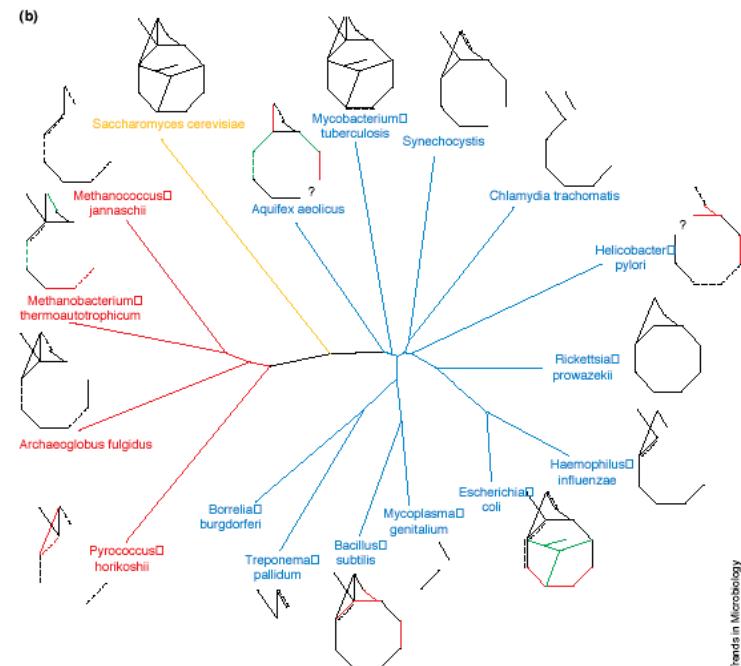
models ++

Pre-evolved Vmicrobes in LTEE (van Dijk et al in prep)

Cooperative Problem solving
(LISP function approximation
cooperative solution precedes
individual solution OR only solution
(high mutation rates)



de Boer & Hogeweg 2010]
Huynen et al TMB 1999



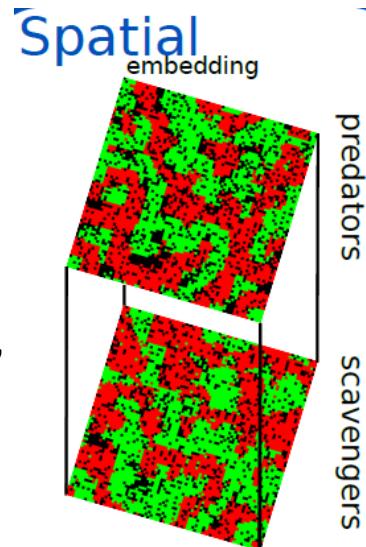
Individual based vs ecosystem based “problem solving” predator-prey-scavenger coevolution

Problem: solve “function” - fully digest all possible prey
prey 2 continuous properties: $0 < X, Y < K$
Fully eaten when predator calculates $f(X, Y)$ correctly

Evol. Target	Minimal Coding Example
$f(x, y) = x^3 + y^3 + 5x^2$	$(+ (* (* (+ x 5) x) x) (* (* y y) y))$

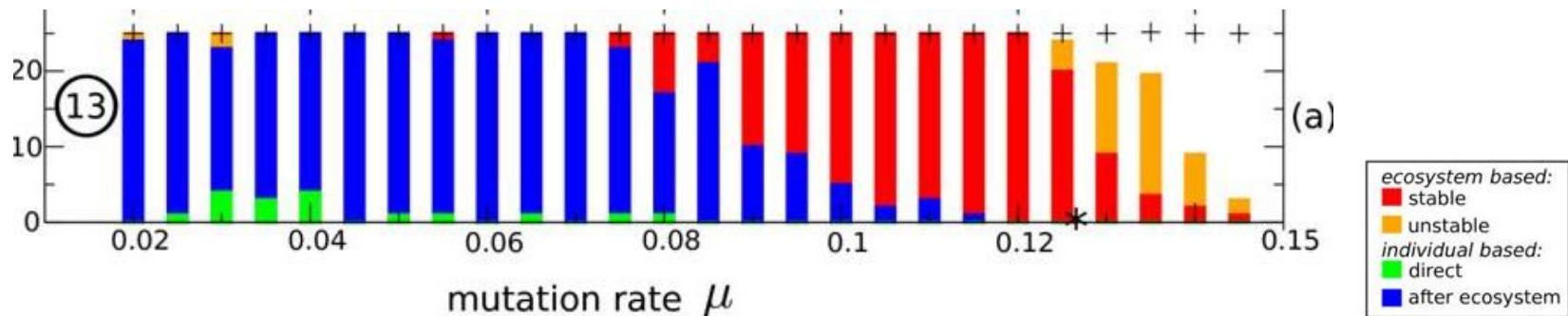
e.g.

Fitness predator: how well it solves “its” prey
Fitness prey: how badly predator solves it
Fitness scavenger: How it solves “what is left”



do individual predators, or does the ecosystems solve it

Ecosystem based solution 'easier' to evolve preceeds individual based solution



Two predator populations specializing on X or Y

Two scavenger populations specializing on X or Y

Two prey populations with high X or high Y values

Self organize in spiral waves,

X predator and Y scavenger pairs together digest prey fully

(i.e. encode the target function correctly)

