Universiteit Utrecht

Evolutionary Systems Biology: multilevel evolution

Paulien Hogeweg
Theoretical Biology and Bioinformatics Grp UU

March 13, 2014

Biology is changing fast....

One of the most fundamental patterns of scientific discovery is the revolution in thought that accompanies a new body of data

Nigel Goldenfeld and Carl Woese
Biology's next revolution
Nature 445 (Jan 2007)

Biology faces a quantum leap into the incomprehensable: the complexity of biology information processing networks will bring us in a counterintuitive world

Paul Nurse
Four great ideas in Biology
gene; evolution;cell; **selforganization**
webvideo Guardian (nov 2010)

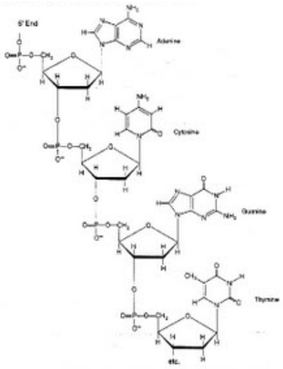
NEXT GENERATION BIOLOGY

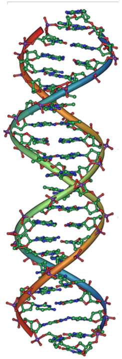
Evolutionary Systems Biology
multilevel evolution

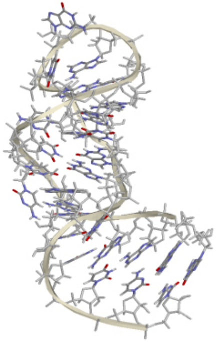
Using data 'tsunami' to reconstruct
what DID happen in evolution
bioinformatic data analysis

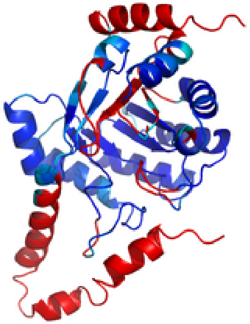
Using modeling to discover what DOES happen -
through mutation/selection process
very often very counterintuitive
in multilevel setting

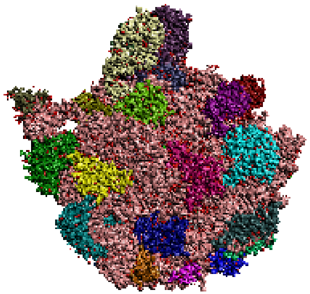
Experimental evolution + bioinformatic analysis of the data
+ modeling

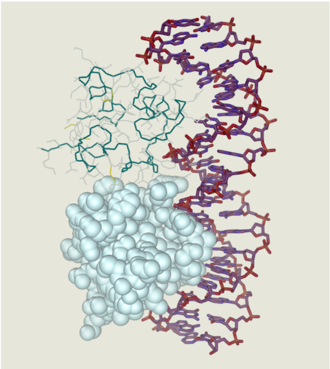

5' End
Adenine
Cytosine
Guanine
Thymine
etc.

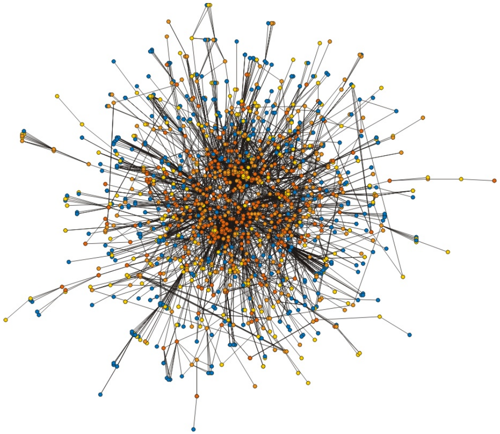


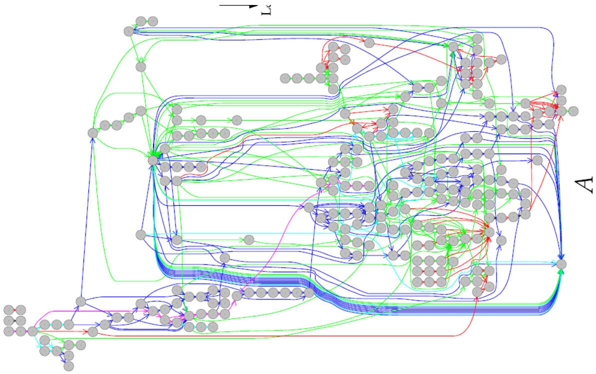


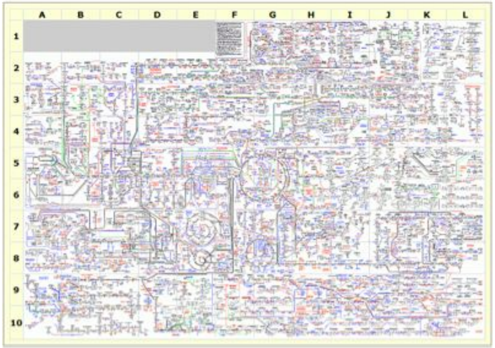


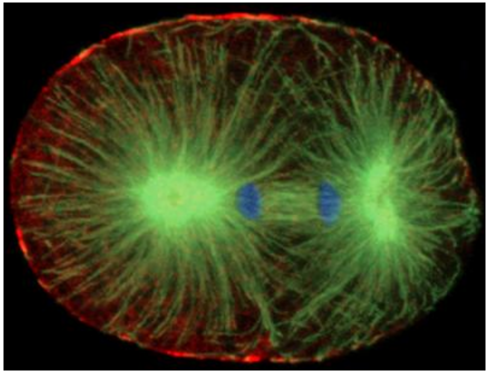


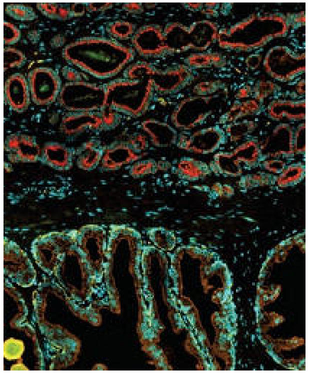


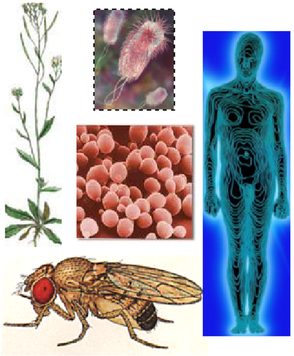



A


A
B
C
D
E
F
G
H
I
J
K
L
1
2
3
4
5
6
7
8
9
10







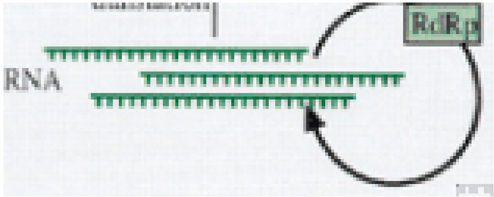
Today

---

eco-evolutionary dynamics:
emergence of new levels of selection
trough spatial pattern formation
- evolution of cooperation/altruism

Genome evolution
complex genotype-phenotype mapping help or hinder?
- observed long term trends in evolution *generic* property of (multilevel) darwinian evolution?
- evolution of evolvability.

“Life is a self-sustained chemical system capable of undergoing Darwinian evolution” G.F. Joyce, 1994



Simplest form: RNA-world
RNA both template and enzym

Joyce (and others) (back)evolve RNA world
e.g. evolve (engineer)
RNA which is RNA dependent RNA polymerase
(Wochner et al 2011: 95 nucleotides: not selfreplicating yet)

Here minimal model of minimal RNA world
study it dynamics *independent* from (bio)chemical properties

minimal model of RNA world: RP system

R replicase
L other RNA (“parasitic”)
replicated when unfolded
’functional’ when folded
fraction l in folded state

$$(a)\quad \mathrm{R}+\mathrm{R} \underset{1-k_R}{\overset{k_R}{\rightleftharpoons}} \mathrm{C_R} \overset{\kappa\theta}{\rightarrow} 2\mathrm{R}+\mathrm{R},$$

$$\mathrm{L}+\mathrm{R} \underset{1-k_L}{\overset{k_L(1-l)}{\rightleftharpoons}} \mathrm{C_L} \overset{\kappa\theta}{\rightarrow} 2\mathrm{L}+\mathrm{R},$$

$$(b)\quad \mathrm{R},\mathrm{L} \overset{d}{\rightarrow} \theta,$$

$$\mathrm{C_R} \overset{2d}{\rightarrow} \mathrm{R}+\theta,$$

$$\mathrm{C_L} \overset{d}{\rightarrow} \mathrm{R}+\theta,$$

$$\mathrm{C_L} \overset{d}{\rightarrow} \mathrm{L}+\theta,$$

$$(c)\quad \mathrm{L} \overset{ml}{\rightarrow} \mathrm{L}+x,$$

Evolve l and k_L

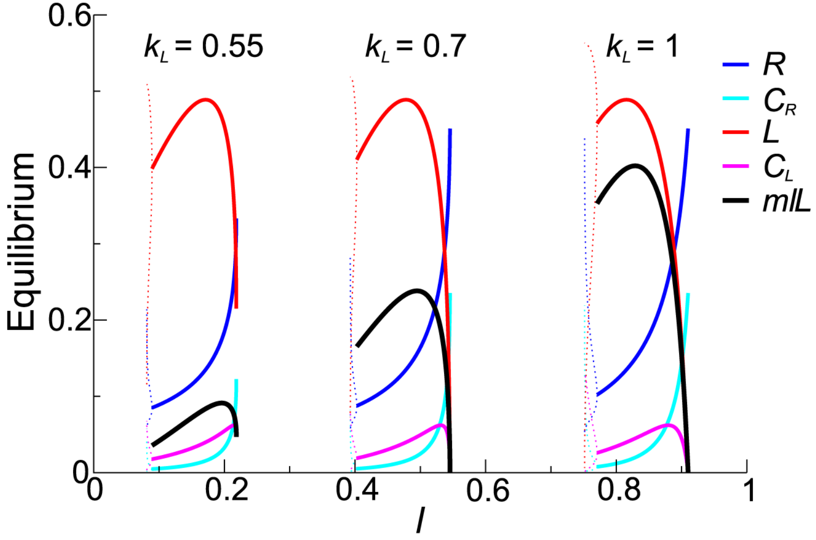
(i.e. multiple (infinite) L spe
one R species)

(Takeuchi & Hogeweg PLOS Comp Biol 2009)

Classical problem
ODE model of RP system

evolutionary extinction because mutants of L which increase k_L and/or decrease of l will outcompete L, and eventually outcompete R)

$$
\begin{aligned}
\dot{R} &= -2k_R R^2 + [2(1-k_R) + 3\kappa\theta + 2d]C_R - k_L RL \\
&\quad + [(1-k_L) + \kappa\theta + d]C_L - dR, \\
\dot{L} &= -k_L(1-l)RL + [(1-k_L) + 2\kappa\theta + d]C_L - dL, \qquad (2) \\
\dot{C}_R &= k_R R^2 - [(1-k_R) + \kappa\theta]C_R - 2dC_R, \\
\dot{C}_L &= k_L(1-l)RL - [(1-k_L) + \kappa\theta]C_L - 2dC_L,
\end{aligned}
$$



$k_R = .6$

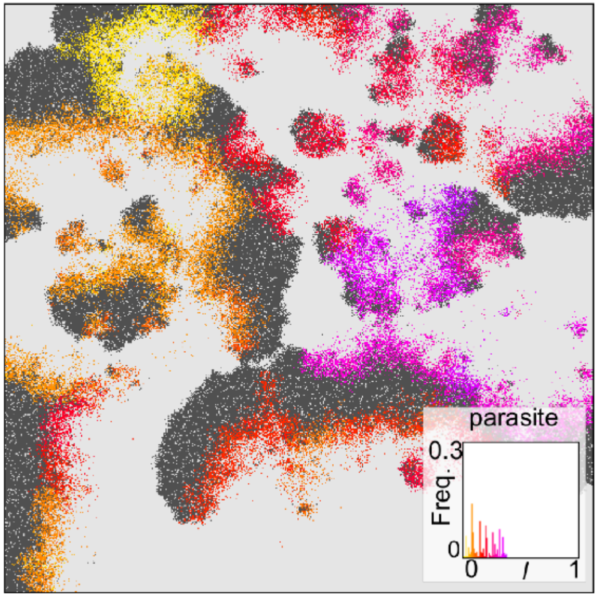
intrinsic advantage of parasite (L)

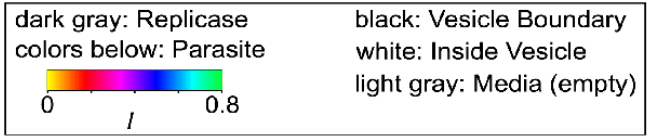
better way to model RP system
individual (particle) based, spatial model

better way to model evolving systems
because very many types possible
(less particles present than possible particles)

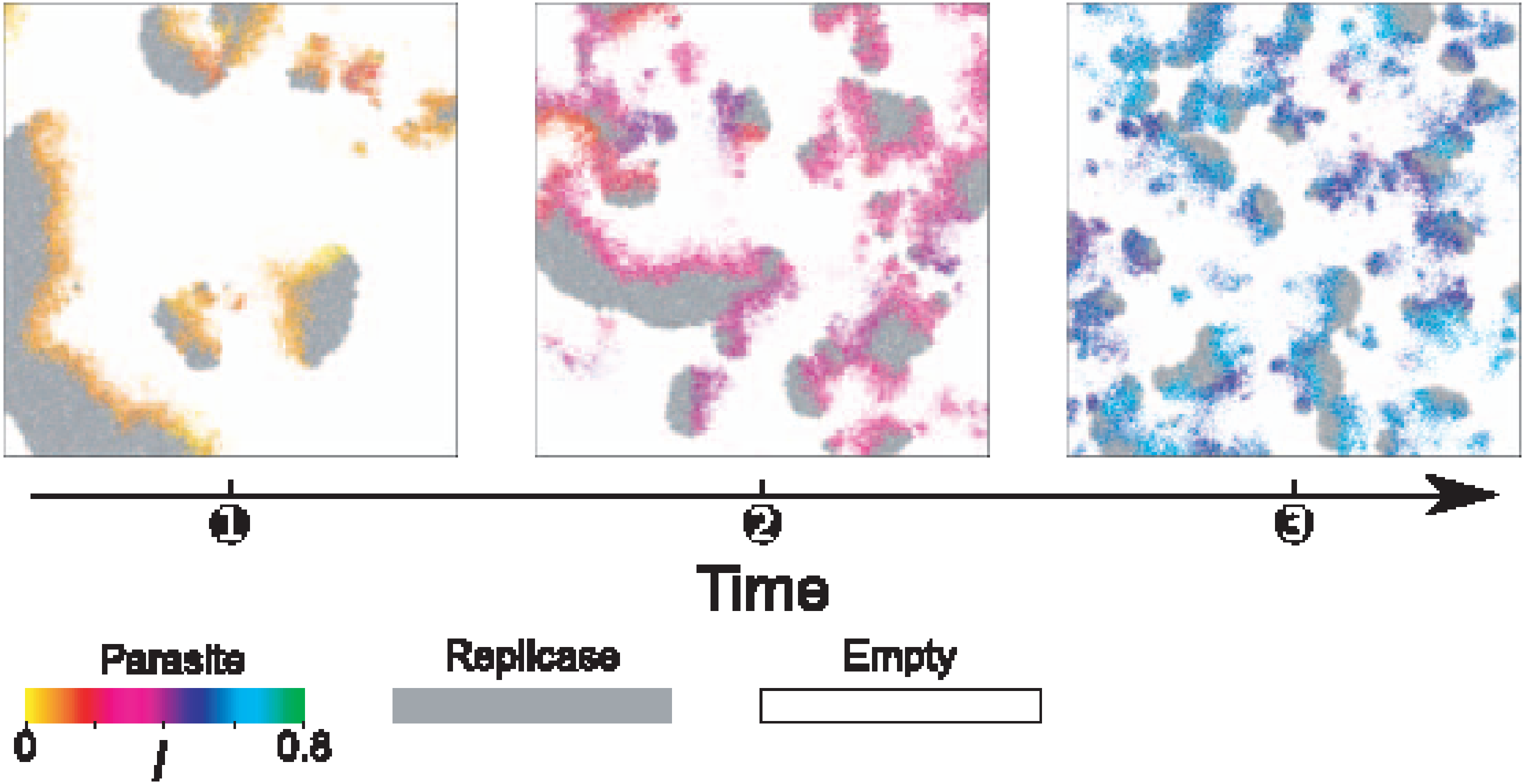
better way because spatial setting more 'natural'

grid based stochastic CA model
Monte Carlo step: N times
choose random patch and random NE
perform reaction or diffusion
with prob. according to
individual (evolving) parameters



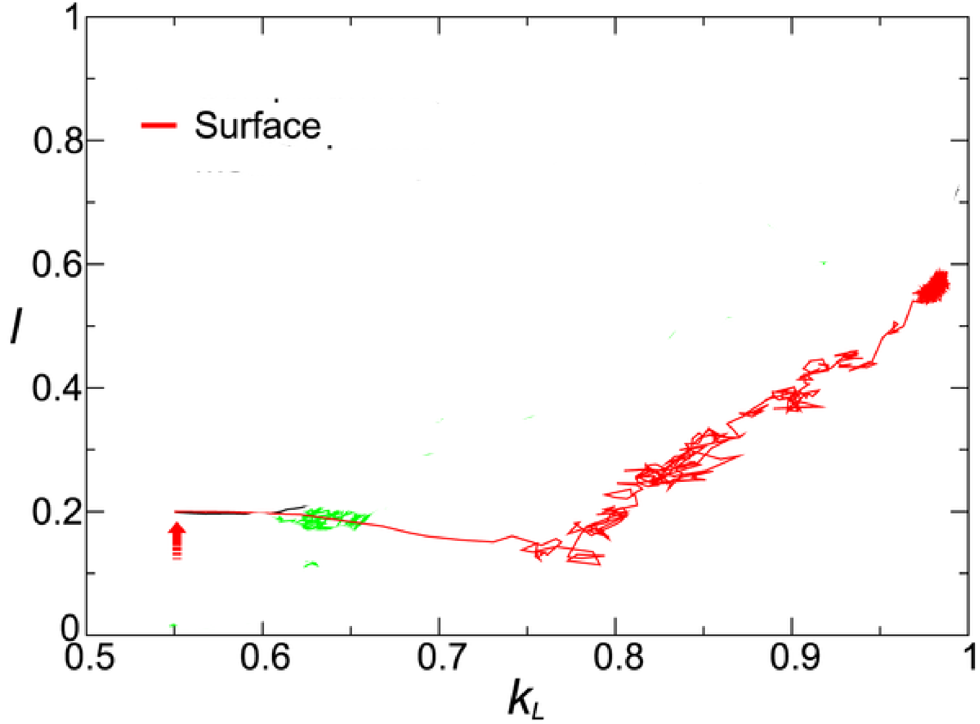

dark gray: Replicase
colors below: Parasite
black: Vesicle Boundary
white: Inside Vesicle
light gray: Media (empty)

long term evolution: towards smaller waves more folded L



Long term evolution (parameters)
emergent 'trade-off' k_L and I
Maximizing I : potential 'new' function

Ancestor trace

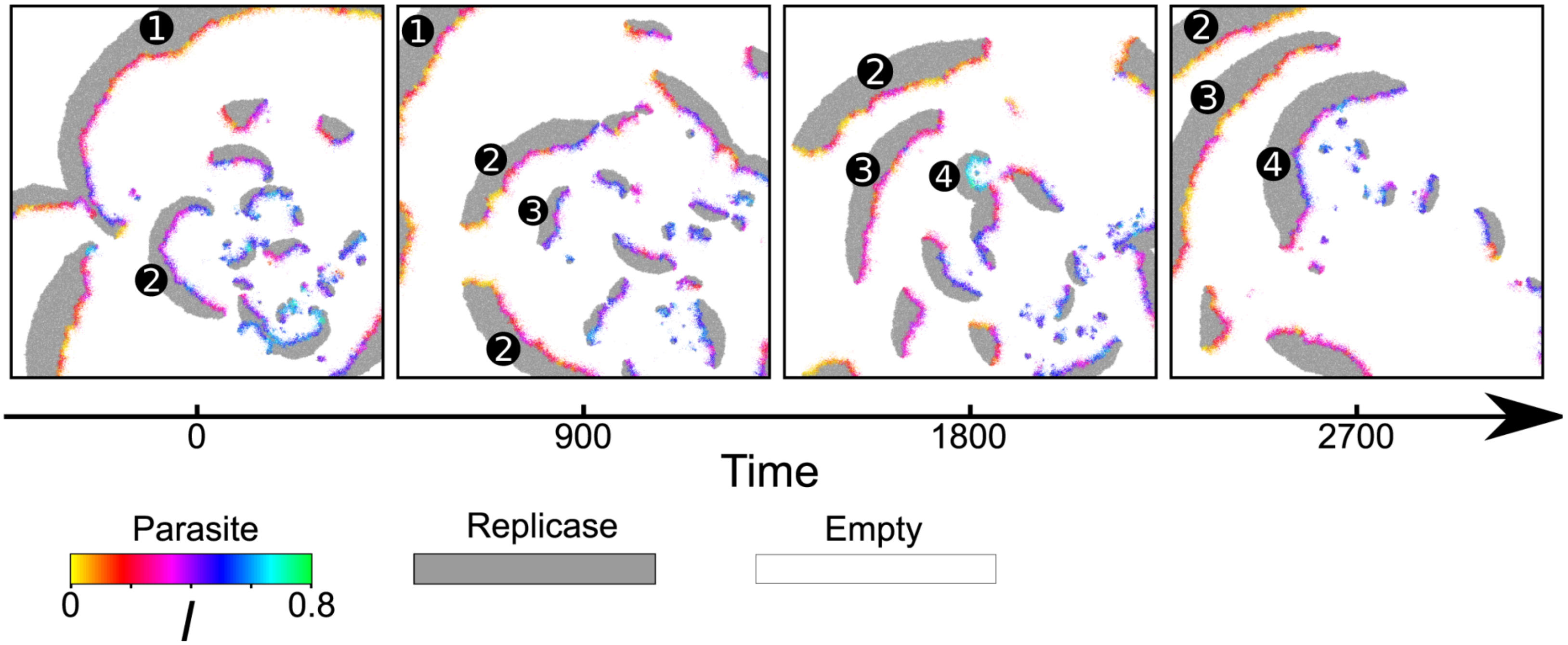


WHY?

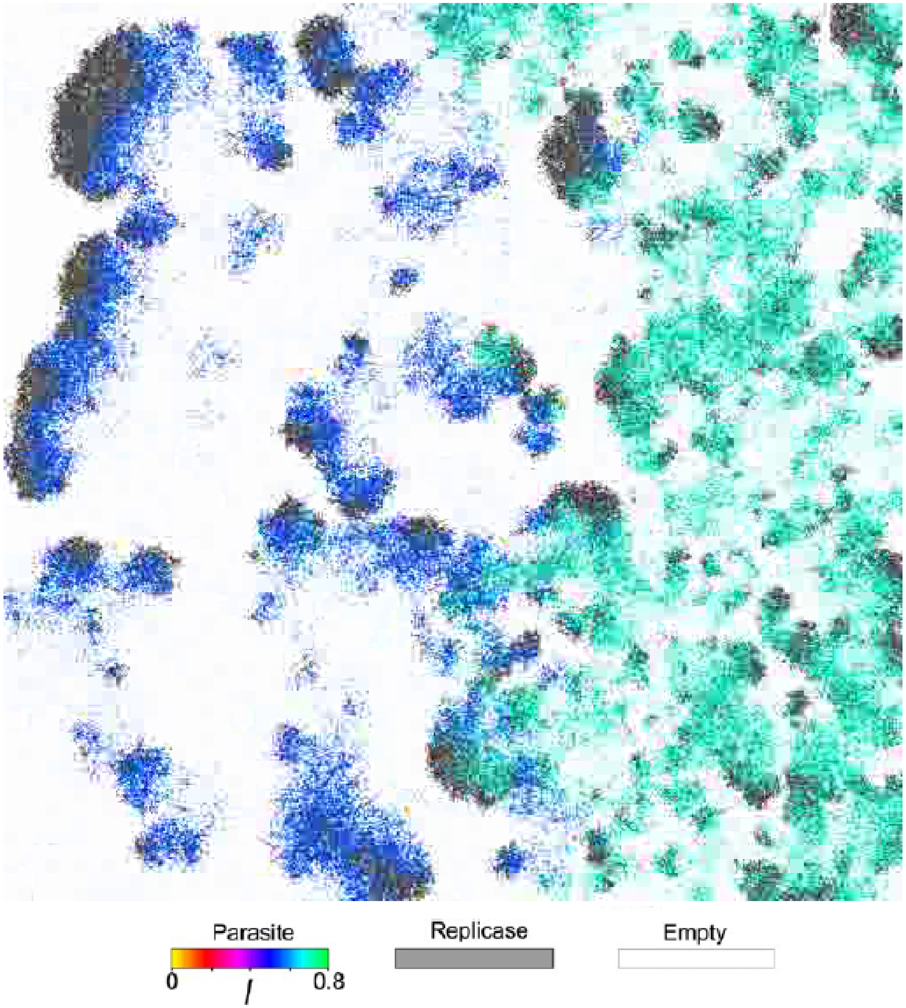
evolution of higher level entities

The waves of replicase and parasites are higher level "Darwinian" entities

Birth
Maturation
Death
Mutation
Selection
Competing *maximizing birthrate*



evolutionary attractor at "edge of chaos" ("border of order")



2 levels of Darwinian selection

Wave level selection

- Waves: long lived -
  (death not by parasites but by collision)
- Maximize Birthrate + growth rate of newborns
- Birthrate higher for high I ('escape')
- However higher birthrate $->$ more (smaller) waves
- $->$ increase collision! (= deathrate of waves))

Individual level selection

- Within waves: parasites evolve towards 'nastiness' (low I)
- However viability maintained $-->$
  "prudent" parasites
- because of higher level selection; which also
- 'frees' parasites to do other things (be folded)

through parasites
evolution of novel functionality

Not only “far away and long ago”
Similar for Evolution and cooperation
a classical problem in (too simplistic) evolutionary theory

---

why not cheat?

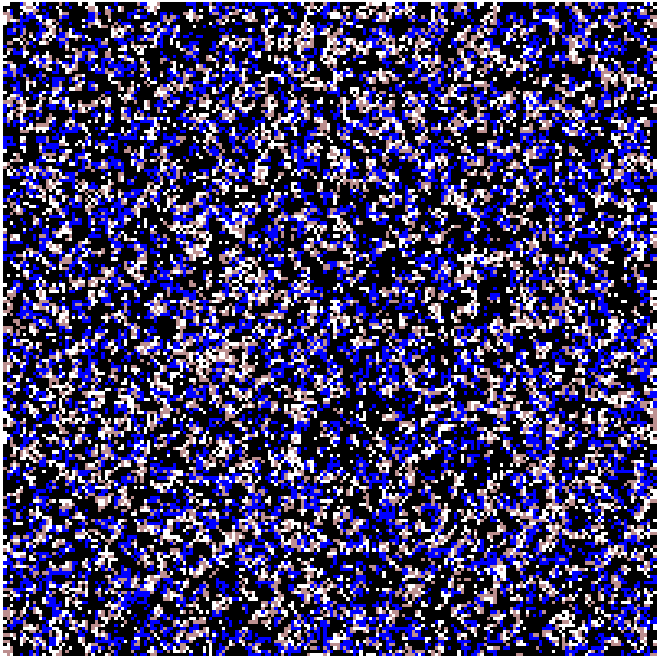
In simple ODE models cheaters destroy the cooperation

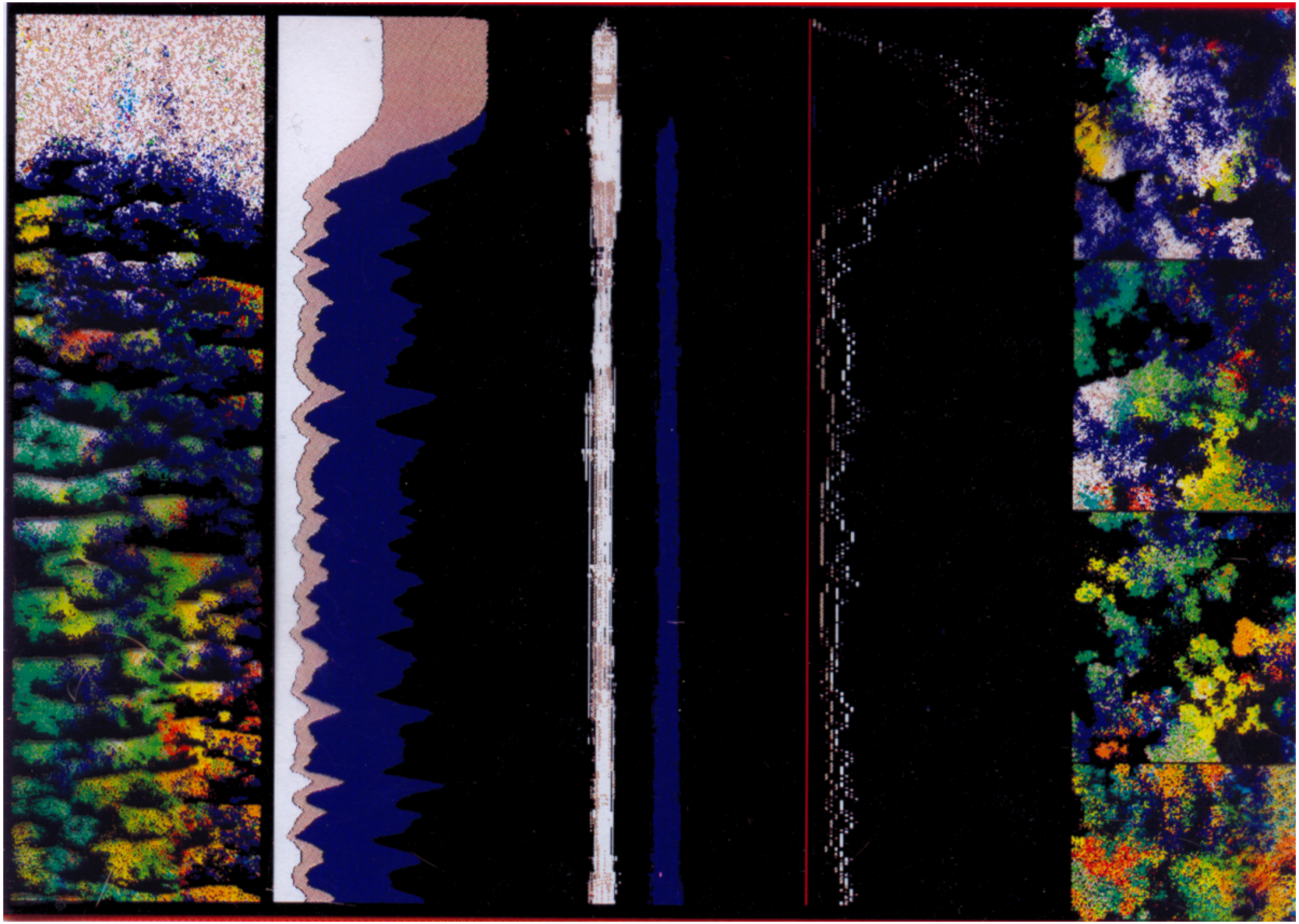
Nevertheless cooperation widespread
e.g. figs/figwasps, dictyostelium, social insects ....

In spatial (CA) model cooperation does persist!

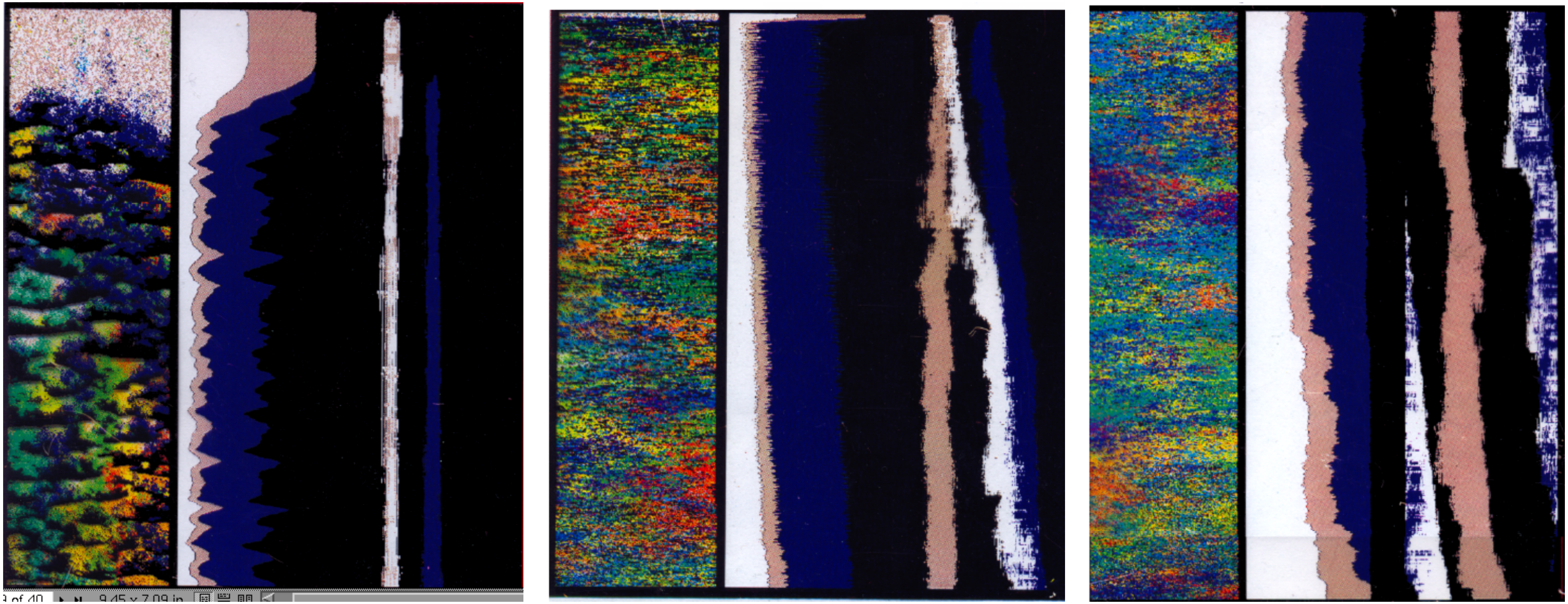
Persistance of cooperation

$A \rightleftarrows B \longrightarrow P$ Persistence in Ecological Time





long term evolution: extinction of cheater selection on spatial patterns, brown species B NO selection for getting more help



$A \rightleftarrows B \rightarrow P$ Persistence in Ecological Time

Phylogenetic reconstruction shows: Gene loss plays major role in evolution

(reconstructed) Ancestral Genomes relatively large

Genes often present before their known present day function is realized.

Example HOX genes before differentiated bodyplan

Example Cell differentiation genes before multicellularity (cf Volvox)

Are these counterintuitive observations inherent to evolutionary processes?

Study by modeling basic evolutionary processes

phylogenetic reconstruction of metabolic enzymes David and Alm, Nature 2010

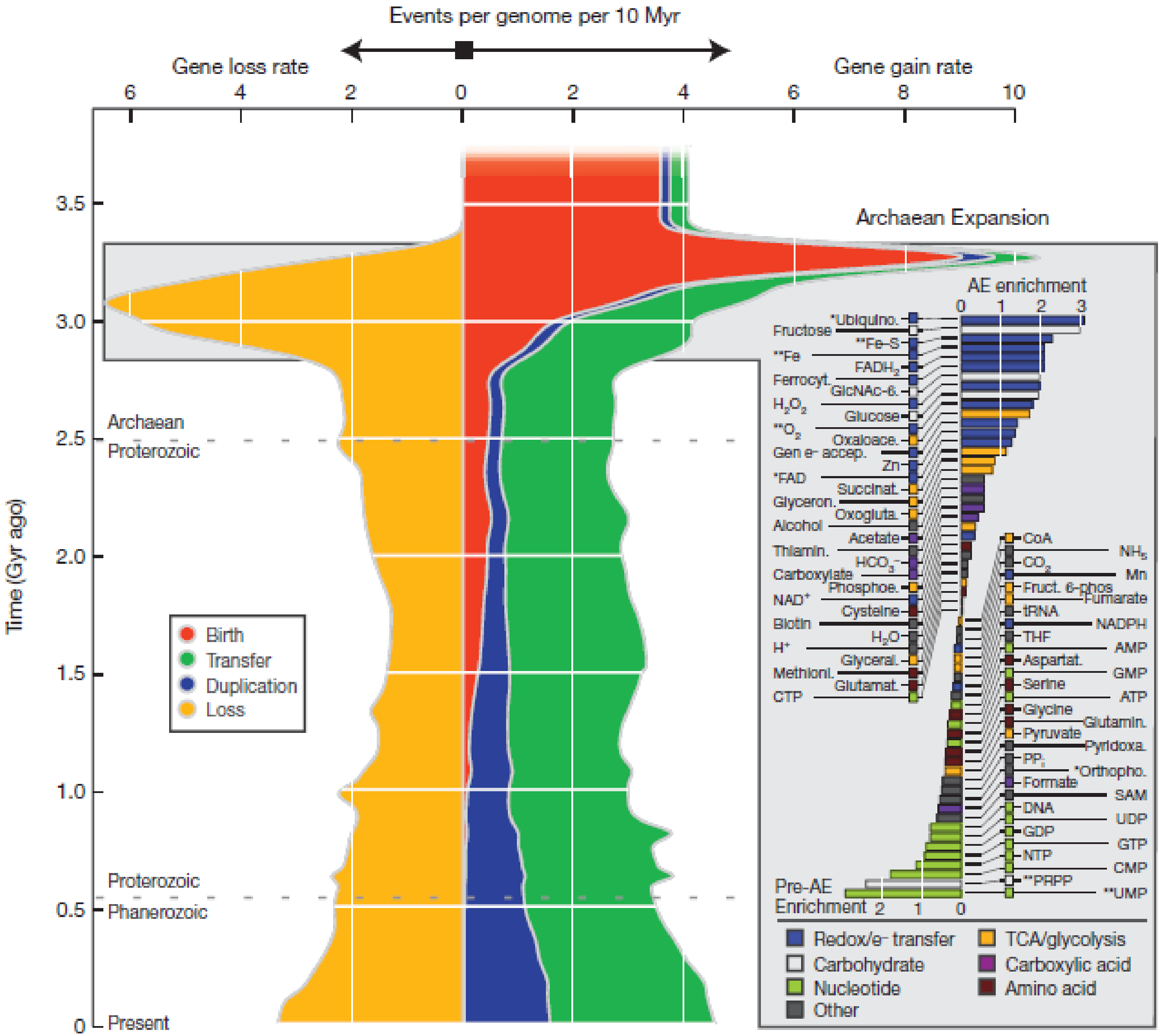
- make all gene trees (3983)

- reconcile gene trees on species tree
minimizing number of 'events': innovation, loss, HGT, duplication and changes in genome sizes along the tree

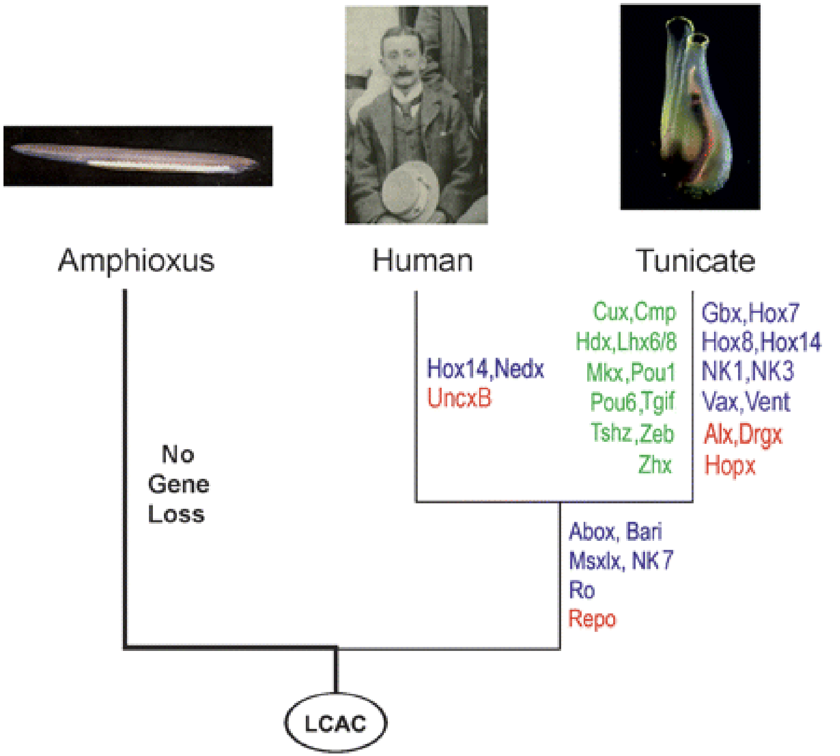
- callibrate timing on fossil record

How did tot biospere metabolism change over the history of life?

“big bang” in metabolic explansion and radiation


Events per genome per 10 Myr
Gene loss rate
Gene gain rate
6
4
2
0
2
4
6
8
10
Time (Gyr ago)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
Archaean
Proterozoic
Proterozoic
Phanerozoic
Present
Birth
Transfer
Duplication
Loss
Archaean Expansion
AE enrichment
0
1
2
3
*Ubiquino.
Fructose
**Fe-S
**Fe
$FADH_2$
Ferrocyt.
GlcNAc-6.
H_2O_2
Glucose
$^{**}O_2$
Oxaloace.
Gen e- accep.
Zn
*FAD
Succinat.
Glyceron.
Oxogluta.
Alcohol
Acetate
Thiamin.
HCO_3^-
Carboxylate
Phosphoe.
NAD^+
Cysteine
Biotin
H_2O
H^+
Glyceral.
Methioni.
Glutamat.
CTP
CoA
NH_3
CO_2
Mn
Fruct. 6-phos
Fumarate
tRNA
NADPH
THF
AMP
Aspartat.
GMP
Serine
ATP
Glycine
Glutamin.
Pyruvate
Pyridoxa.
PP_i
*Orthopho.
Formate
SAM
DNA
UDP
GDP
GTP
NTP
CMP
**PRPP
**UMP
Pre-AE Enrichment
2
1
0
Redox/e- transfer
TCA/glycolysis
Carbohydrate
Carboxylic acid
Nucleotide
Amino acid
Other

Gene loss as major evolutionary process



Hahn et al 2007 PLoS Gentetics

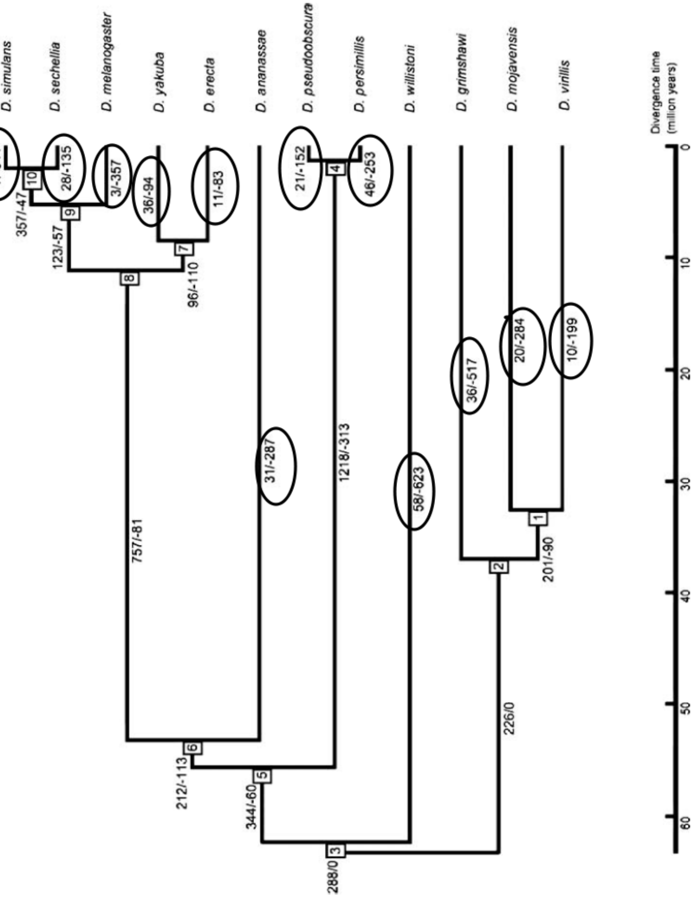


Figure 3. Lineage-Specific and Extinct Gene Families

Metazoa
Loss of homeoboxgenes

Drosophila species
gain/loss of genes

Modeling genome evolution

NOT like in ecological/immunological models in the course populations of identical individuals.

But (through mutations) all individuals may be unique.

Not ODE, but individual oriented models (like above) but moreover

Individuals: genotype - phenotype - fitness mapping
can be dynamical system ODE (gene regulation, metabolism)

birth/death dependent on fitness

mutational operators: INDELS, substiutions (and/or parameter changes)

Evolution of genome size in virtual cells

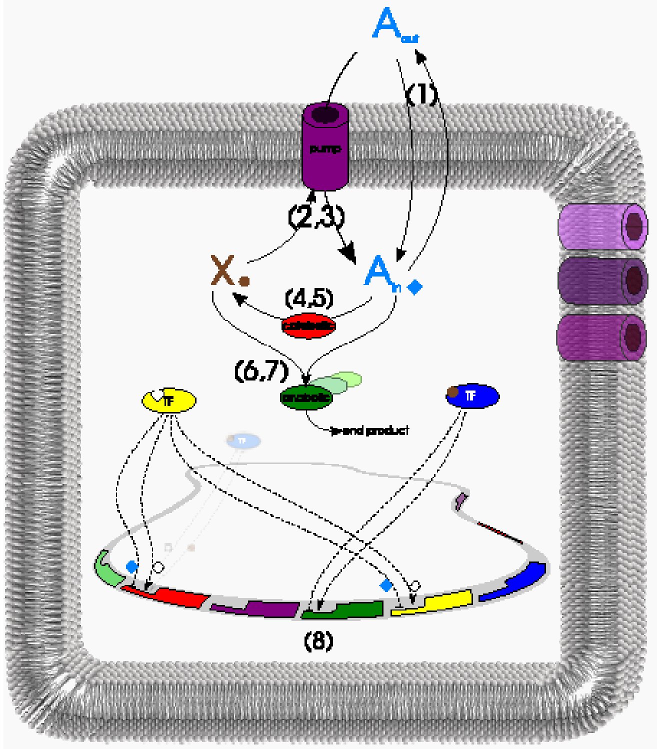
based on “plausable” *minimal* multilevel ’cell’
mutations segmental duplications/ deletions, pointmutations
fitness: *homestasis* (evolves regulatory adaptation)
evolving in varying environment

Questions

Are some of the features seen in phylogenetic analysis observable in evolution of such cells?
Early complexity, dominance of gene loss

Cuypers & Hogeweg 2012

virtual cell model (adapted from Neyfakh et al 2009 Biol Direct)



$$\frac{d[A]}{dt} = ([A]_{out} - [A])Perm \qquad (1)$$

$$\left(\begin{array}{ll} \frac{d[A]}{dt} = \frac{[A]_{out}[X]\mathrm{Vmax_p}[Prot]_p}{([A]_{out} + \mathrm{KA_p})([X] + \mathrm{KX_p})} & (2) \\ \frac{d[X]}{dt} = \frac{-d[A]}{dt} & (3) \end{array}\right)$$

$$\left(\begin{array}{ll} \frac{d[A]}{dt} = \frac{-[Prot]_c[A]\mathrm{Vmax_c}}{[A] + \mathrm{KA_c}} & (4) \\ \frac{d[X]}{dt} = \frac{-d[A]}{dt}\ \mathrm{E_{conv}} & (5) \end{array}\right)$$

$$\left(\begin{array}{ll} \frac{d[A]}{dt} = \frac{-[Prot]_a[A][X]\mathrm{Vmax_a}}{([A] + \mathrm{KA_a})([X] + \mathrm{KX_a})} & (6) \\ \frac{d[X]}{dt} = \frac{d[A]}{dt} & (7) \end{array}\right)$$

$$\frac{d[Prot]}{dt} = Pr \cdot Reg - Degr[Prot] \qquad (8)$$

Processes modelled in the cell:

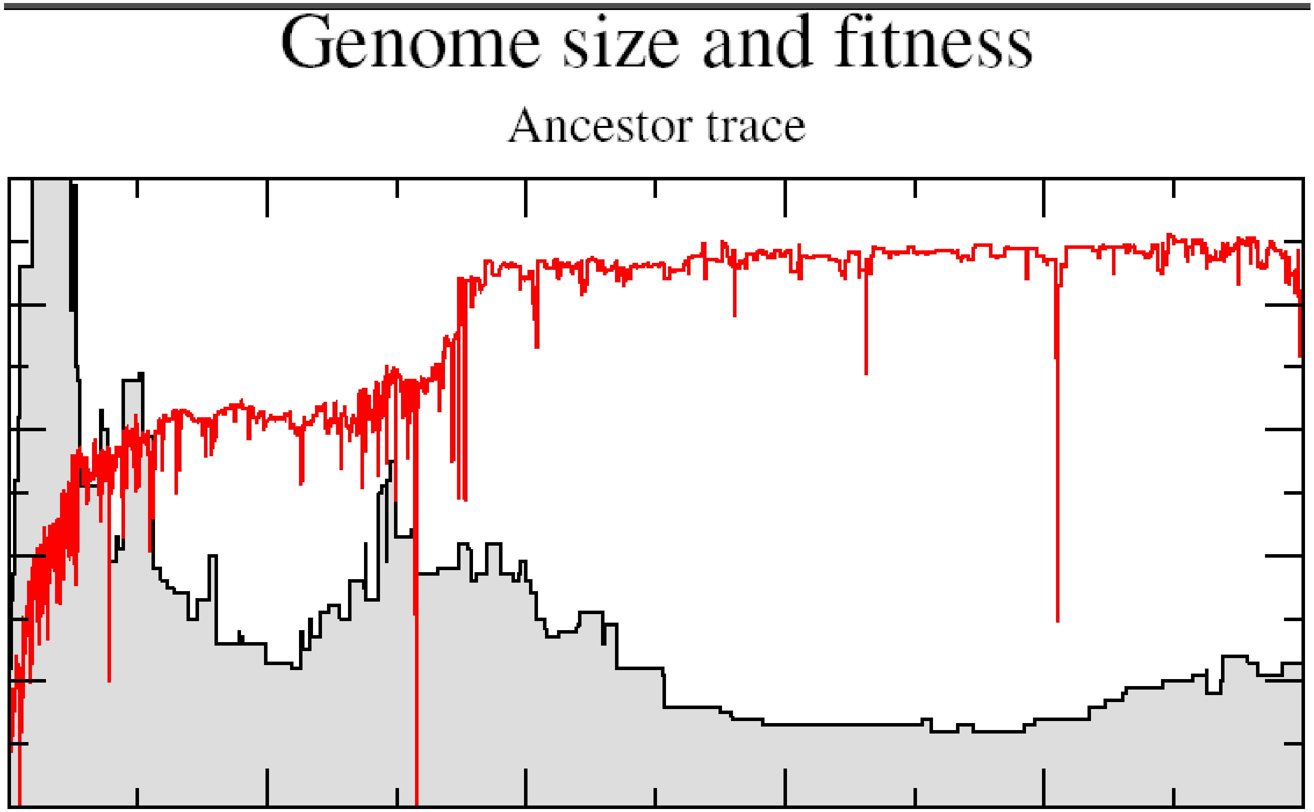
- diffusion (1) : A follows the gradient over the cell membrane
- pumping (2,3) : pump enzymes consume X to import A
- catabolism (4,5) : catabolic enzymes convert resource (A) into energy (X)
- anabolism (6,7) : anabolic enzymes consume A and X to produce building blocks
- protein production and degradation (8) : TFs regulate the rate of transcription of proteins; degradation takes place at a constant rate

evolution of virtual cells

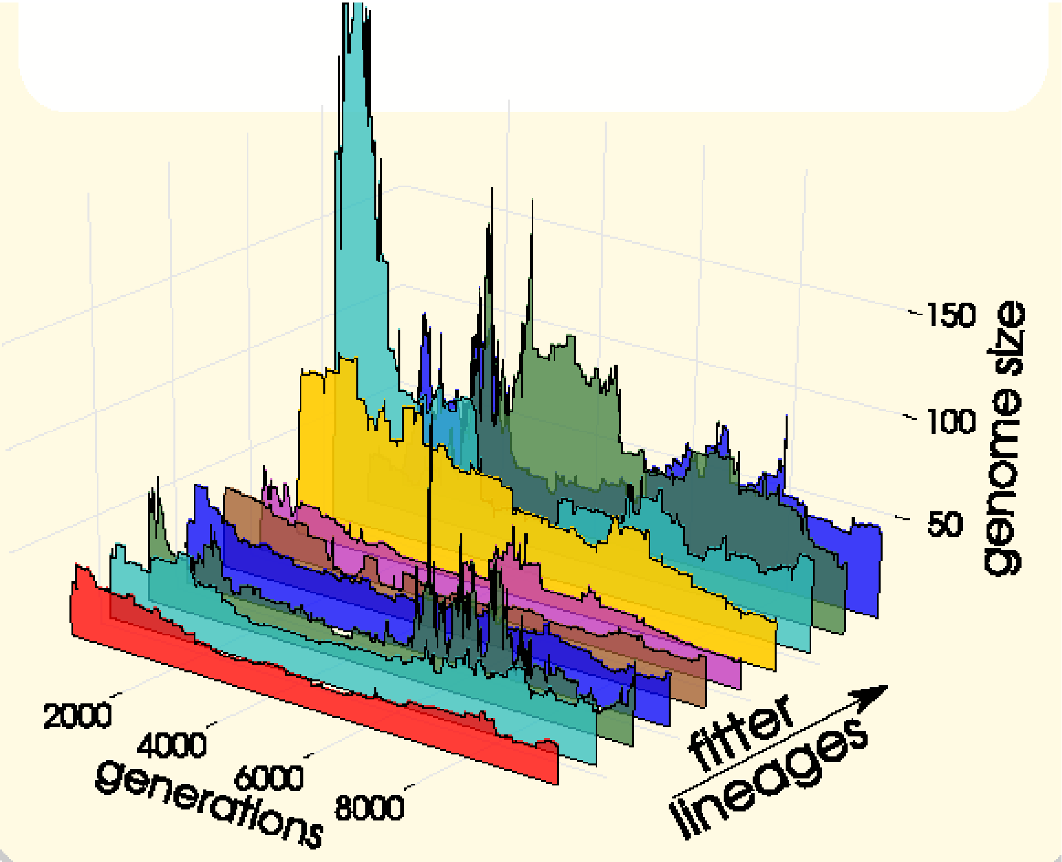
- Population of 1000 cells, 10000 generations
- external concentration of resource **A** fluctuates between .003 and 30

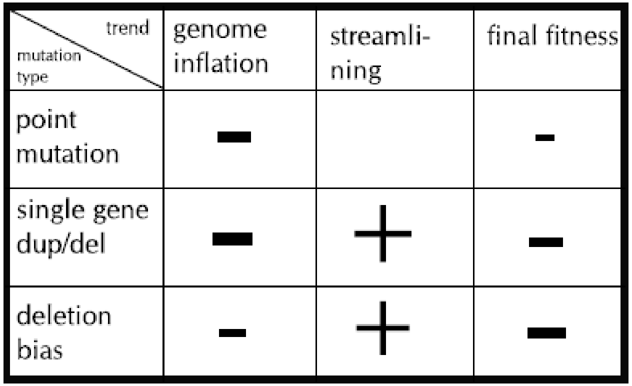
- *homeostasis:* Internal concentration should be kept at 1.
- Initial genome size ca 10 genes
- Mutational operators: duplication / deletion / rearrangement / point mutations
- ('sees' (only) 1-3 environments in lifetime - adapts to 'all')

Typical evolutionary dynamics: Genome inflation(s) - followed by fitness increase - followed by stream lining - followed by genome size fluctuations



early genome inflation “generic” pattern



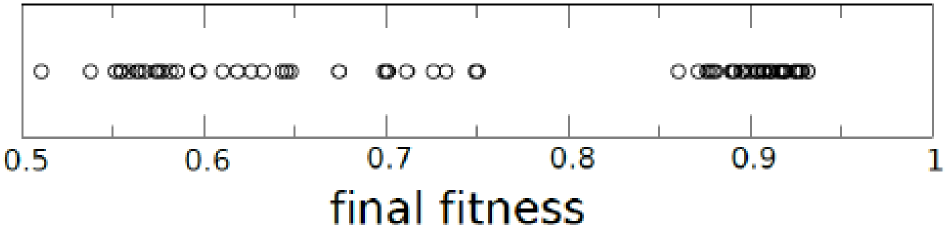


trend / mutation type	genome inflation	streamli-ning	final fitness
point mutation	–		-
single gene dup/del	–	+	–
deletion bias	-	+	–

occurs in “better’ runs
in one param. setting

occurs in parameter settings
which lead to “better” results

Local landscapes, genome expansion and future fitness



Duplications			Deletions		
t=1-100	t=101-200	ΔF	t=1-100	t=101-200	ΔF
+	(+)	> 1.05	=	=	> 1.05
(+)	+	$.95 - 1.05$	=	+	$.95 - 1.05$
-	-	$< .95$	=	-	$< .95$

Genome Size			Fitness		
t=1-100	t=101-200		t=1-100	t=101-200	
+	+		=	=	

Conclusions evolution of virtual cells

- early genome inflations,
  increases degrees of freedom and therewith adaptability
- followed by streamlining: fitness gain through gene loss
- Intricate interplay of neutral and adaptive processes:
  adaptation $-->$ neutrality; neutrality $--$ adaptation

- also other observables, eg effect of mutations, e.g.
  Evolved genotype phenotype mapping maximizes
  neutrality AND selection

interesting (unexpected) but generic behaviour
of mutation/selection

Rugged fitness landscape

Evolution “stuck on local optima??”

NO......

DETOURS!

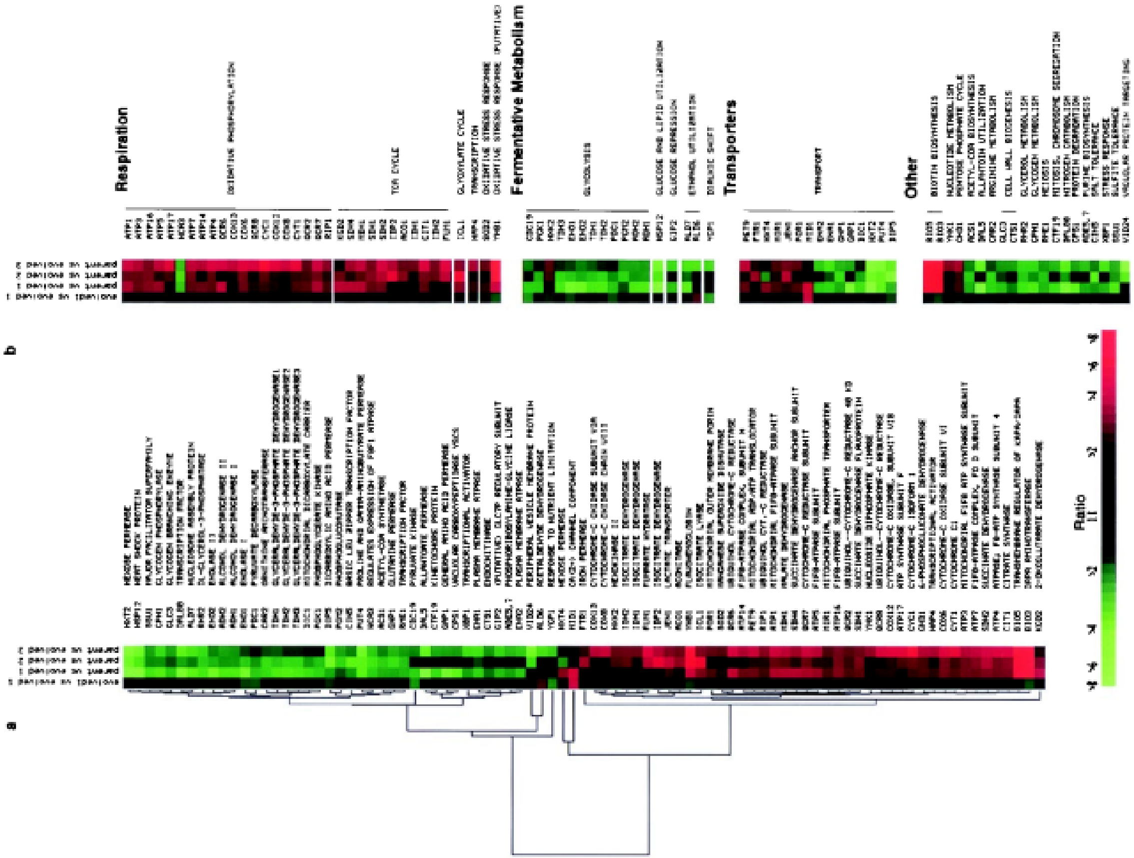
Evolution not ”far away and long ago” New insights through experimental evolution, high throughput data, bioinformatic analysis and evolutionary modeling

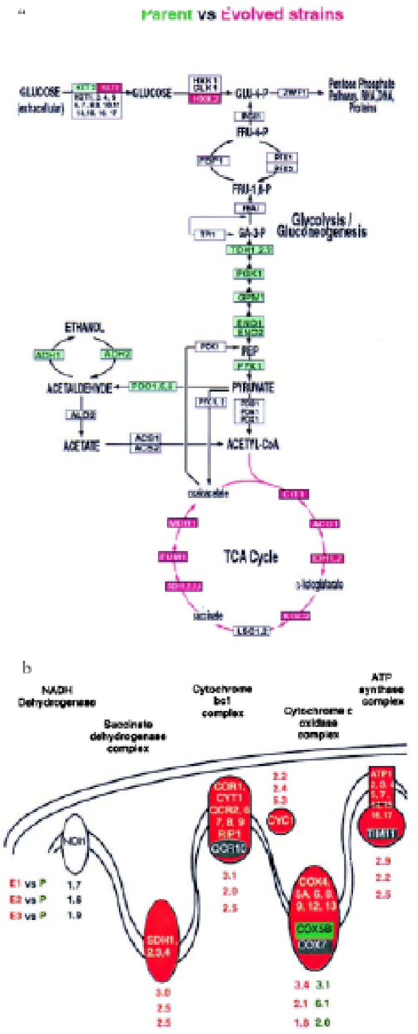
Minor (?) transitions in evolution
Yeast regulatory network evolution
Some "surprising" observations from short term evolution experiments
(Ferrea et al 1999, Dunham et al 2002)

- very efficient adaptation in short period
- major changes in gene expression in short evolutionary time: ca 600 genes differentially expressed in period that no more than 7 mutations expected
- changes in gene expression make "sense" with respect to adaptation
- resemble regulatory adaptation
- many gross chromosomal rearrangement (GCR)
- similar GCR in duplicate evol experiment

evolved evolvability?

regulatory and/vs evolutionary 'adaptation' gene expression changes in strai evolved on low glucose med



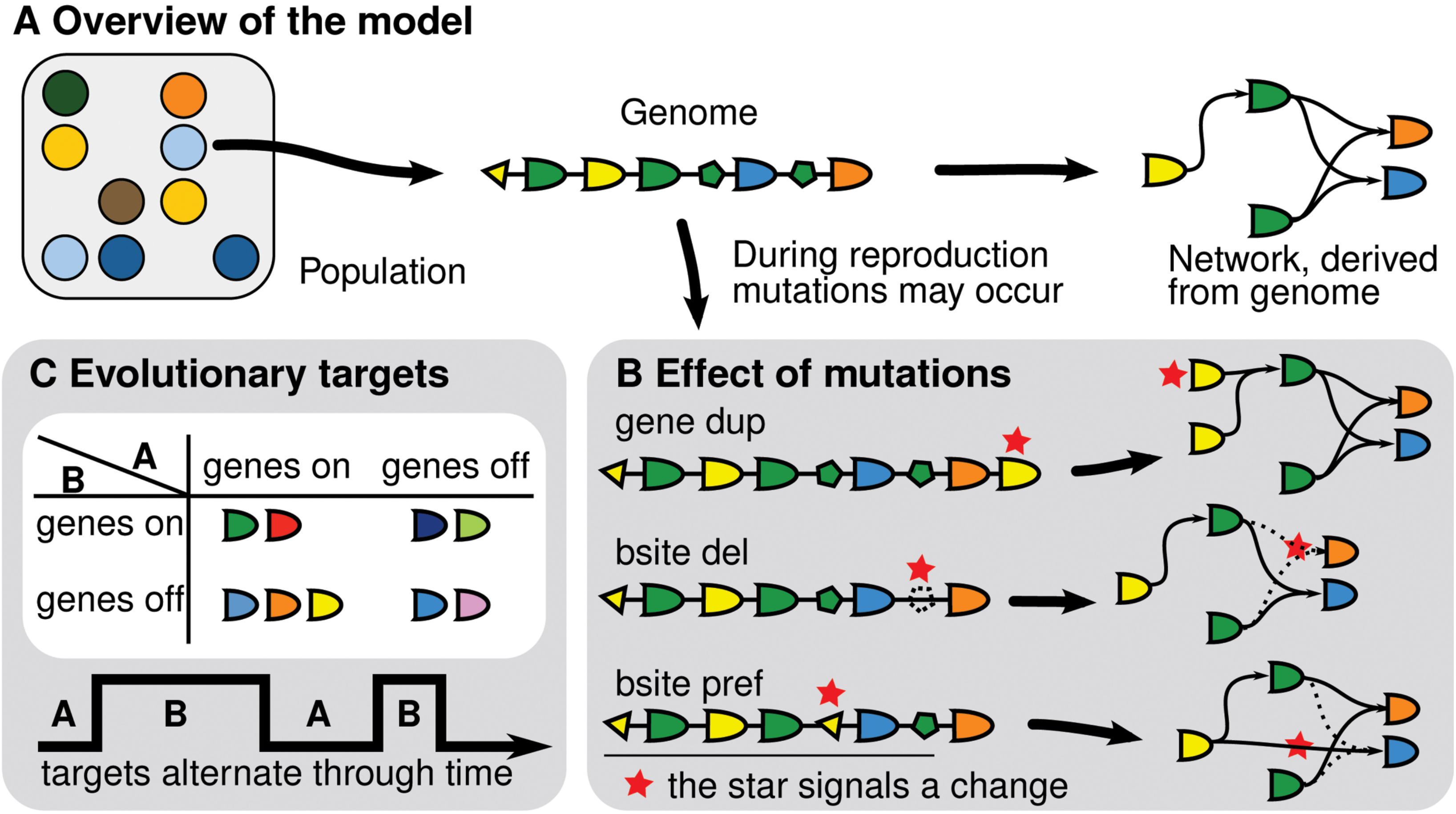


Are these properties of short term evolution a generic property of mutation/selection in evolving systems with explicit genome-network mapping?

By evolution of genome/transcriptome structure?

Crombach & H. 2007 MBE, 2008 PLOS-CompBio

Evolution of Regulation based mutational priming
Crombach and Hogeweg PLOS Comp Biol 2008



A Overview of the model

Genome

Population

During reproduction mutations may occur

Network, derived from genome

C Evolutionary targets

targets alternate through time

B Effect of mutations

gene dup

bsite del

bsite pref

★ the star signals a change

network dynamics and fitness

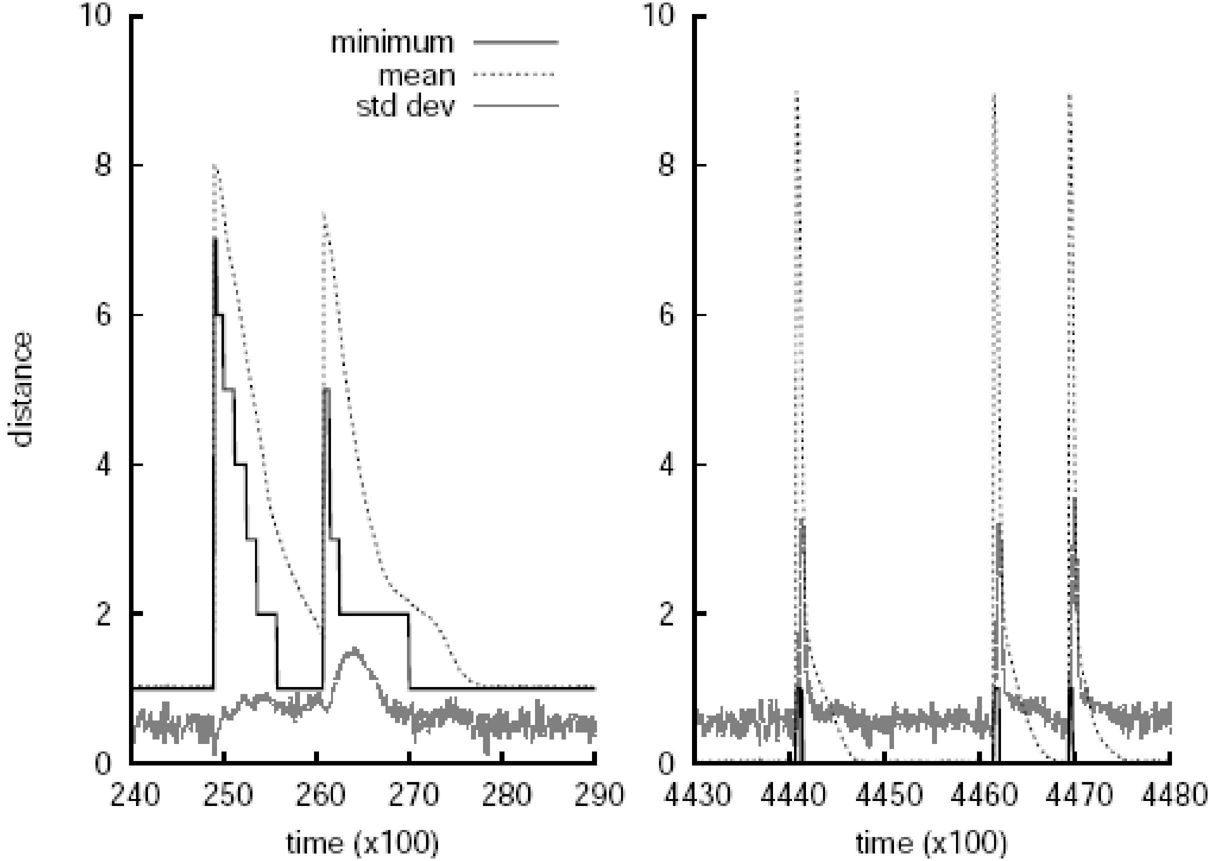
Network update:

$$s_i^{t+1} = \begin{cases} 0 & \mathit{if} \ \sum_j w_{ij} s_j^t < \theta_i \\ s_i^t & \mathit{if} \ \sum_j w_{ij} s_j^t = \theta_i \\ 1 & \mathit{if} \ \sum_j w_{ij} s_j^t > \theta_i \end{cases}$$

fitness: distance to target

$$f = \left(1 - \frac{D}{D_{max}}\right)^p$$

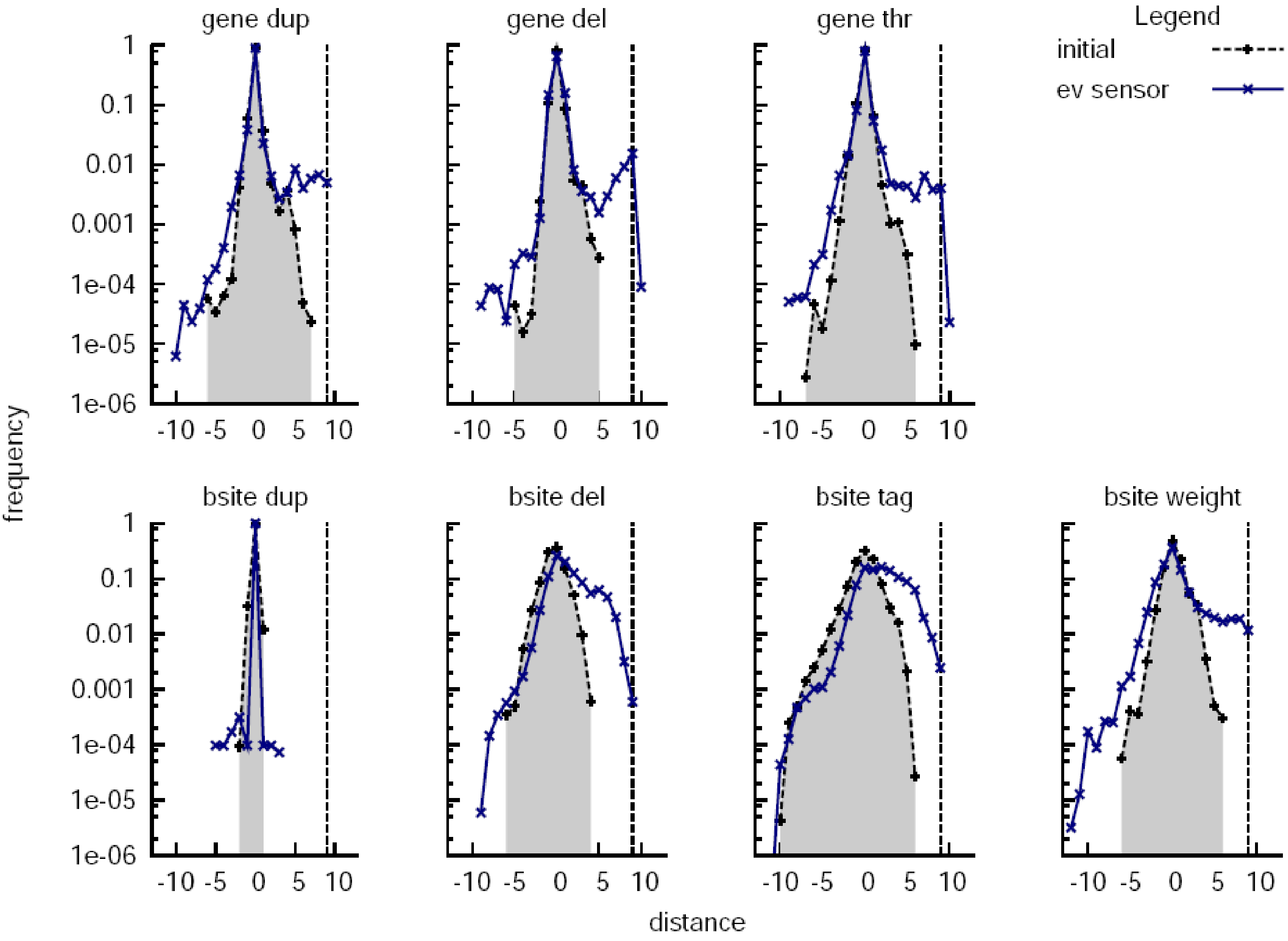
improved evolvability observed



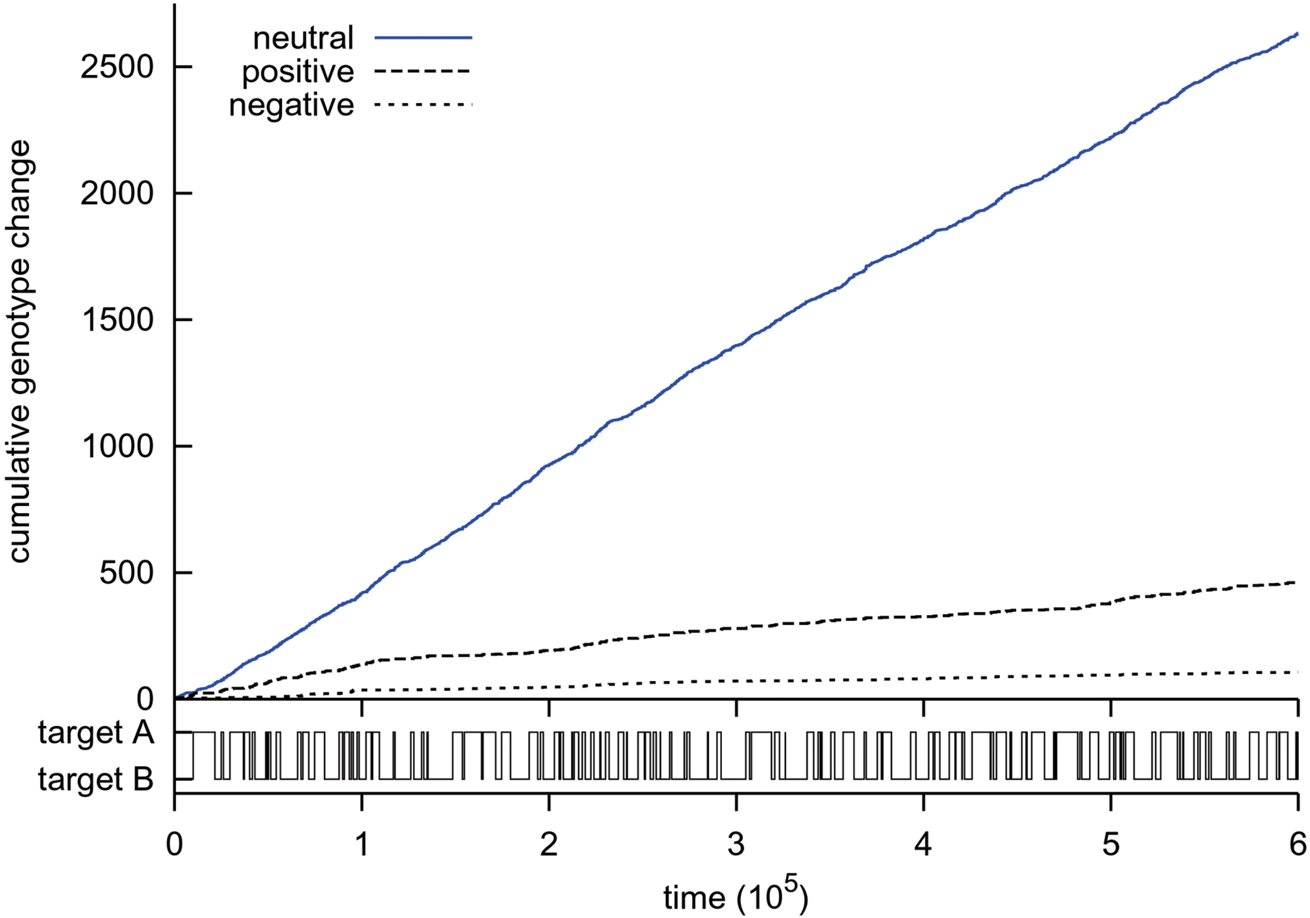
Hamming distance improvement to opposite target

Regulatory Mutational Priming:

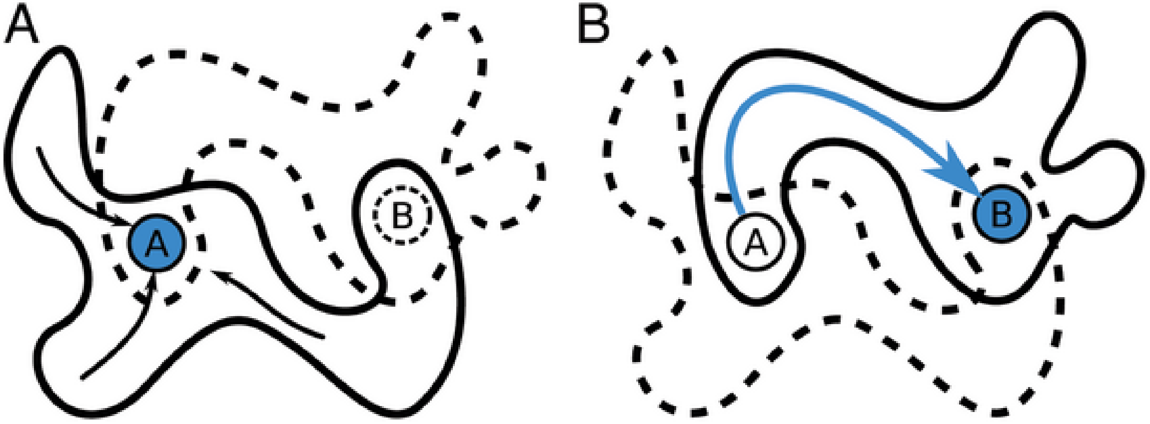
Many different mutations lead to “beneficial” adaptation


gene dup
gene del
gene thr
Legend
initial
ev sensor
bsite dup
bsite del
bsite tag
bsite weight
frequency
distance
1
0.1
0.01
0.001
1e-04
1e-05
1e-06
-10
-5
0
5
10

Neutral drift far greater than adaptive change!



evolution of evolvability and bases of attraction



conclusions
Evolution of genomes and gene regulatory networks
evolution of evolvability

Random mutations are not “random”
in EVOLVED genomes

- Transposon dynamics structures genomes creating hotspots for mutations and genome ordering. Long term evolution leads to genome structures such that short term evolutionn is facilitated
- Genotype to phenotype mapping through gene regulatory networks evolves such that (advantageous) attractor switching occurs (blow up of single mutations to large scale effects)

Both these mechanisms appear to occur in Yeast

Overall summary/conclusions

Nurse: Biology faces a quantum leap into the incomprehensable: the complexity of biology information processing networks will bring us in a counterintuitive world

Indeed but: **+ bioinformatic modeling**
renders the counterintuitive comprehensable

$--->$ profound new insight is major transitions in evolution
$--->$ current functioning of organisms

Next part of the course

Woese: "One of the most fundamental patterns of scientific discovery is the revolution in thought that accompanies a new body of data"

Indeed High throughput data
+ bioinformatic data analysis

$--->$ profound new insight is major transitions in evolution
$--->$ current functioning of organisms