Genome evolution: coding structures and evolvability

Course Computational Biology 2024; Paulien Hogeweg;
Theoretical Biology and Bioinformatics Grp Utrecht University

Evolution of coding structures (continued)

LAST TIME
Multilevel evolution in RNA world
- GP map based on Min. energy folding - fixed genome length
- spatial pattern formation

Evolution of multiple lineages (lower mutation rates)
Evolution of functional mutational neighborhood (at high mutation rates)
Evolution of multiple coding

ecosystem based solution to "do more" than individuals??

NOW
From prebiotic evolution to biotic evolution

Surprising observation explained from "first principles"

Evolution of genomes and transcription regulatory networks

Mutational operators Not only point mutations

evolution of evolvability (2)

Random mutation - non-random effects

Individual based vs ecosystem based "problem solving" predator-prey-scavenger coevolution

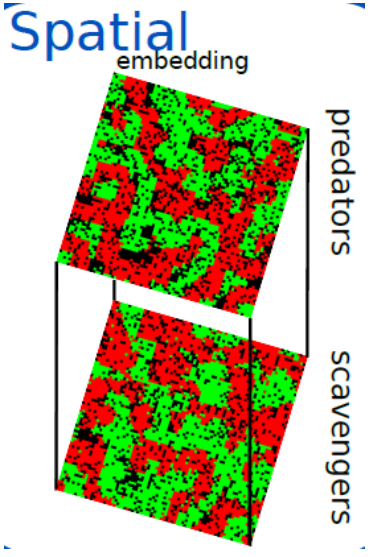
FK de Boer, P Hogeweg 2010

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

e.g.

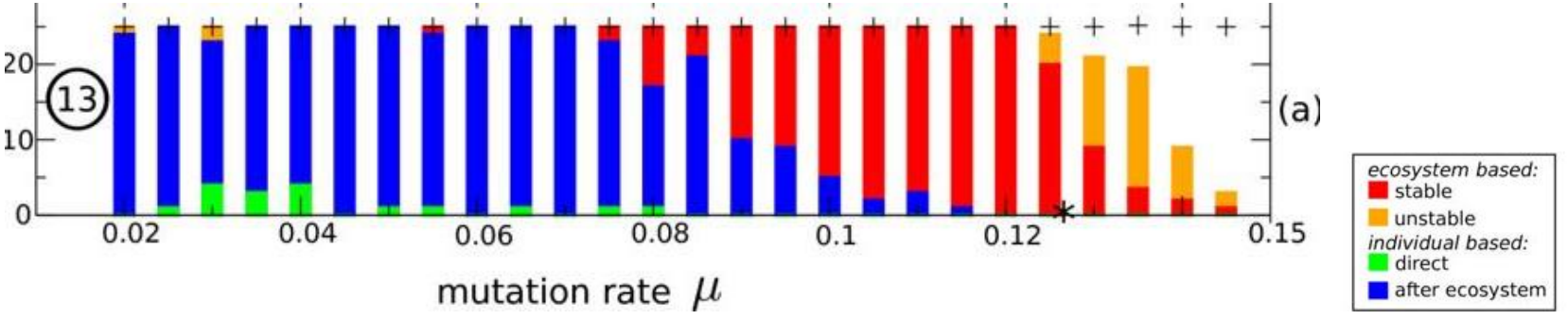
Evolutionary Target	Minimal Coding Example
$f(x, y) = x^3 + y^3 + 5x^2$	(+ (* (* (+ x 5) x) x) (* (* y y) y))

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

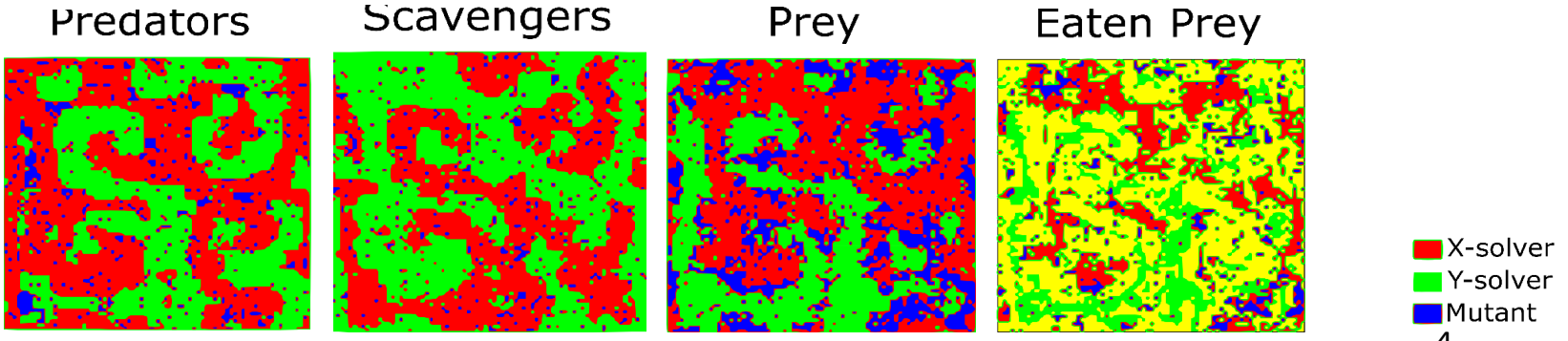


do individual predators, or does the ecosystems solve it

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

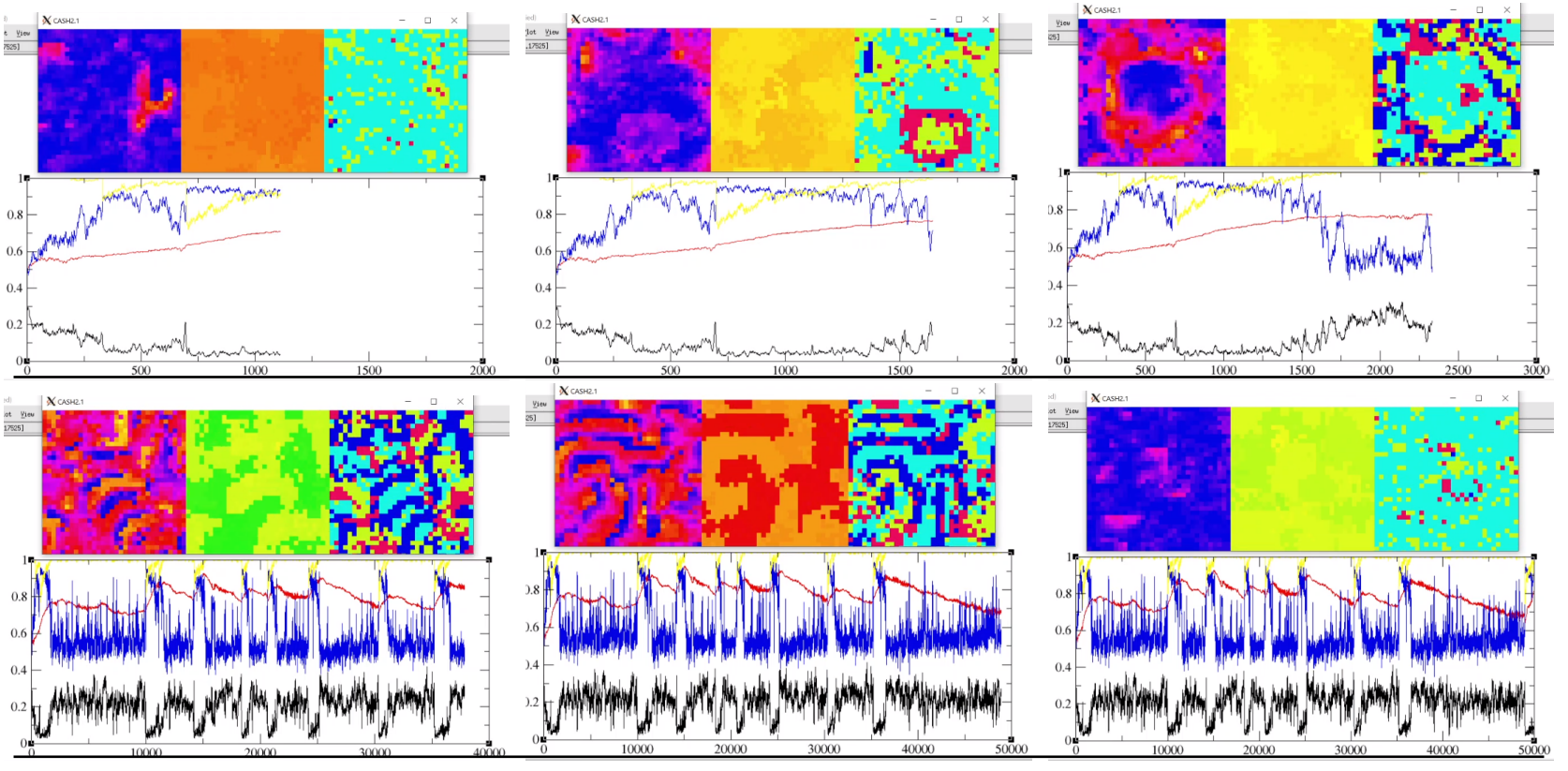


Two predator populations specializing on X or Y
Two scavenger populations specializing on X or Y
Two prey populations with high X or high Y values
Self organize in spiral waves,
X predator and Y scavenger pairs together digest prey fully
(i.e. encode the target function correctly)



Slightly beneficial 'genes' (bits) and individual vs ecosystem based solutions (gene conservation)

Host-Parasite coevolution, above the information threshold
Fitness target: bitstring(256), differential weighted fitness/bit
Wouter Ubbink 2021: the helpfull parasite



"the absolute of the host increases while its actual fitness decreases, and that of the parasite increases and the ecosystems looses bits"

black: host actual fitness; blue parasite actual fitness;
red: host absolute fitness; yellow: collective fitnses (bits in population)

conclusion

"Cooperation" (getting something done together) through spatial selforganization

Division of labor among predators

Coerces prey into certain types

See less $->$ can do more:
Cooperative solution of "all" problems, by "seeing" only a subset of problems No direct or indirect fitness benefit for predators to give scavengers an eatable bite.

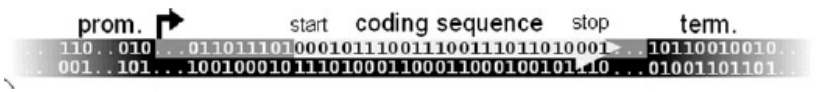
ecosystem based solution precedes individual based solution

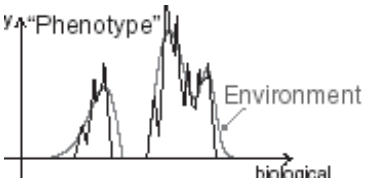
ecosystem based solution stable at high mutation rates

ecosystem based solution preserves slightly beneficial genes

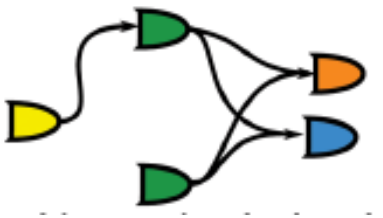
GENOME EVOLUTION: 3 (4) modeling frameworks
genome structure and genotype to phenotype/fitnessmap

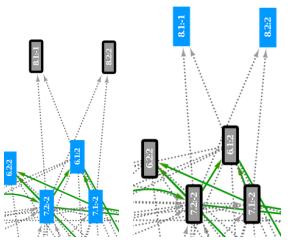
AEVOL bit (nucleotide) level coding of genome allows evolution of new genes!



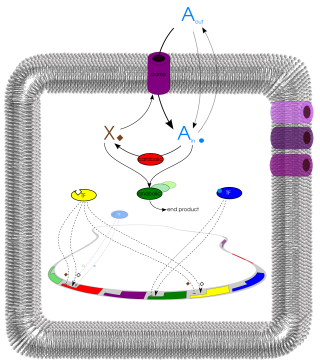


PoaS course grained genome allows multilevel (evolving) GPf mapping





PoaS + metabolism (Virtual cell) allows *indirect* sensing environment

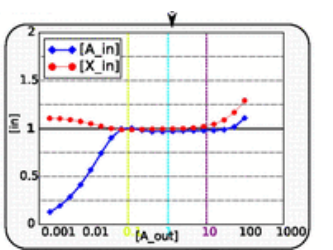


$$\left(\frac{d[A]}{dt} = \frac{[A]_{out}[X]Vmax_p[Prot]_p}{([A]_{out} + KA_p)([X] + KX_p)}\right) \quad (2)$$

$$\frac{d[X]}{dt} = \frac{-d[A]}{dt} \quad (3)$$

$$\left(\frac{d[A]}{dt} = \frac{-[Prot]_c[A]Vmax_c}{[A] + KA_c}\right) \quad (4)$$

$$\frac{d[X]}{dt} = \frac{-d[A]}{dt} \quad (5)$$



Genome evolution
evolution transcription regulation networks

Observations

Structural features of transcription regulation networks
FFL

Very fast evolutionary adaptation in the lab (e.g. YEAST)

Involves massive changes at transcriptome level (ca 10%)

Involves GCR - and the same ones in multiple experiments

Are these generic features, to be expected from random mutations (with or without selection)?

dynamics OF evolving and evolved Gene Regulatory Networks (GRN)

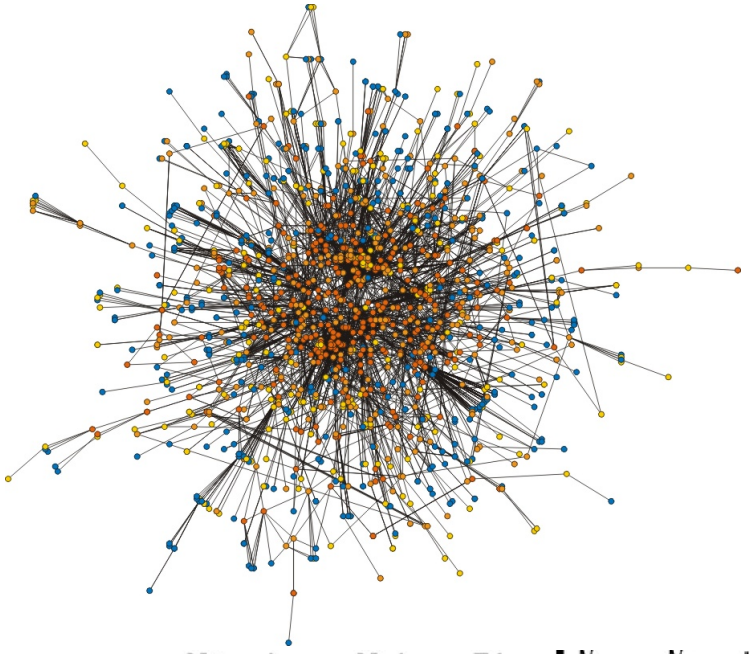
Structural features of transcription regulation networks
powerlaw and FFL

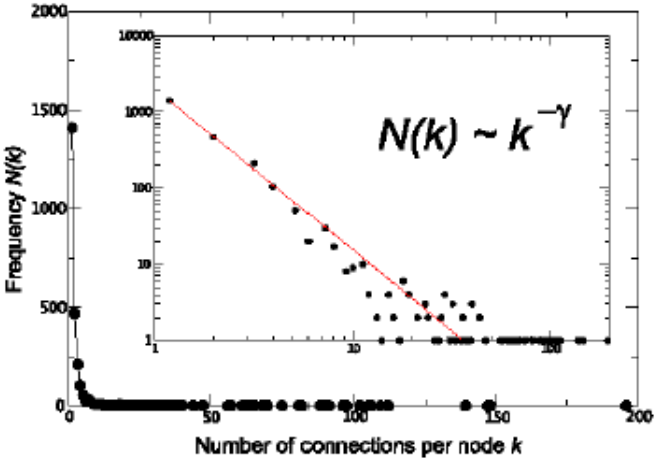
- Characterizing topology of GRN
- *What do we mean with "over-representation of..."*

- *Random mutations = / = randomization*

Importance of coding structure

Observed properties of the GRN of Yeast





Network	Nodes	Edges	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score
Gene regulation (transcription)					**Feed-forward loop**			**Bi-fan**
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13
*S. cerevisiae**	685	1,052	70	11 ± 4	14	1812	300 ± 40	41

(also many other networks: neural networks, computer networks, (but not Eco-networks))

Milo & Alon 2002

Genomic encoding of GRN
Modeling Mutational Dynamics

bag of genes with binding sites (BS)

BS deletion 8x10^{-3}

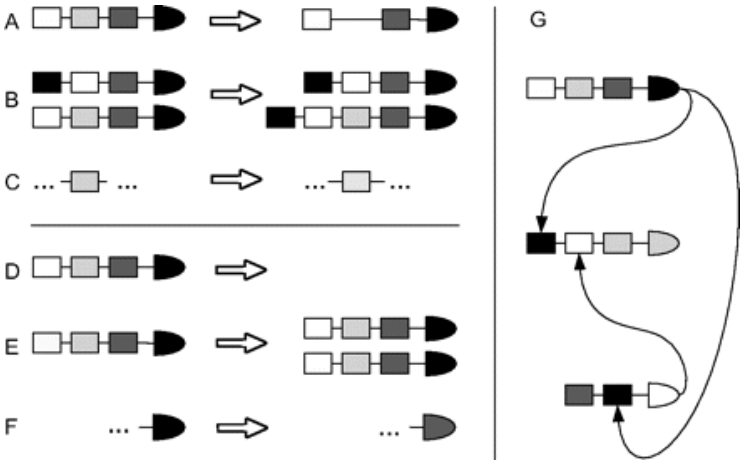
BS duplication 8x10^{-3}

Bs mutation 8x10^{-4}

Gene deletion 1x10^{-3}

Whole gene duplication $1x10^{-3}$

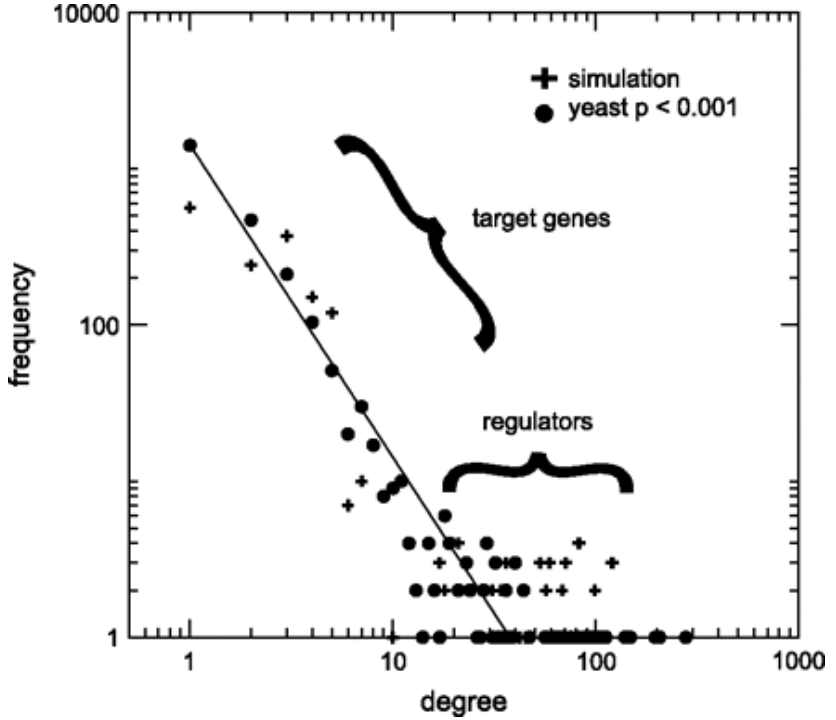
Protein divergence/innovation 5x10^{-3}



Parameters loosly chosen from literature, NOT FITTED.

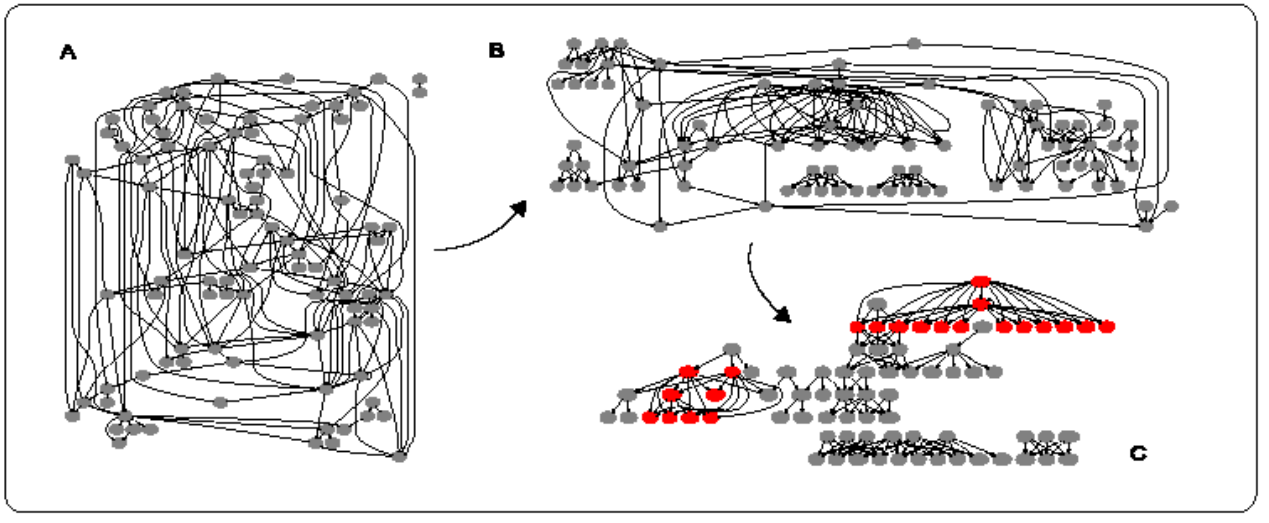
Feed-Forward Loop Circuits as a Side Effect of Genome Evolution Otto X. Cordero, Paulien Hogeweg MBE 2006

mutational dynamics WITHOUT selection leads to
Powerlaw distribution of connections
with similar γ as Yeast GRN
for similar number of genes and TF (2000, 100)



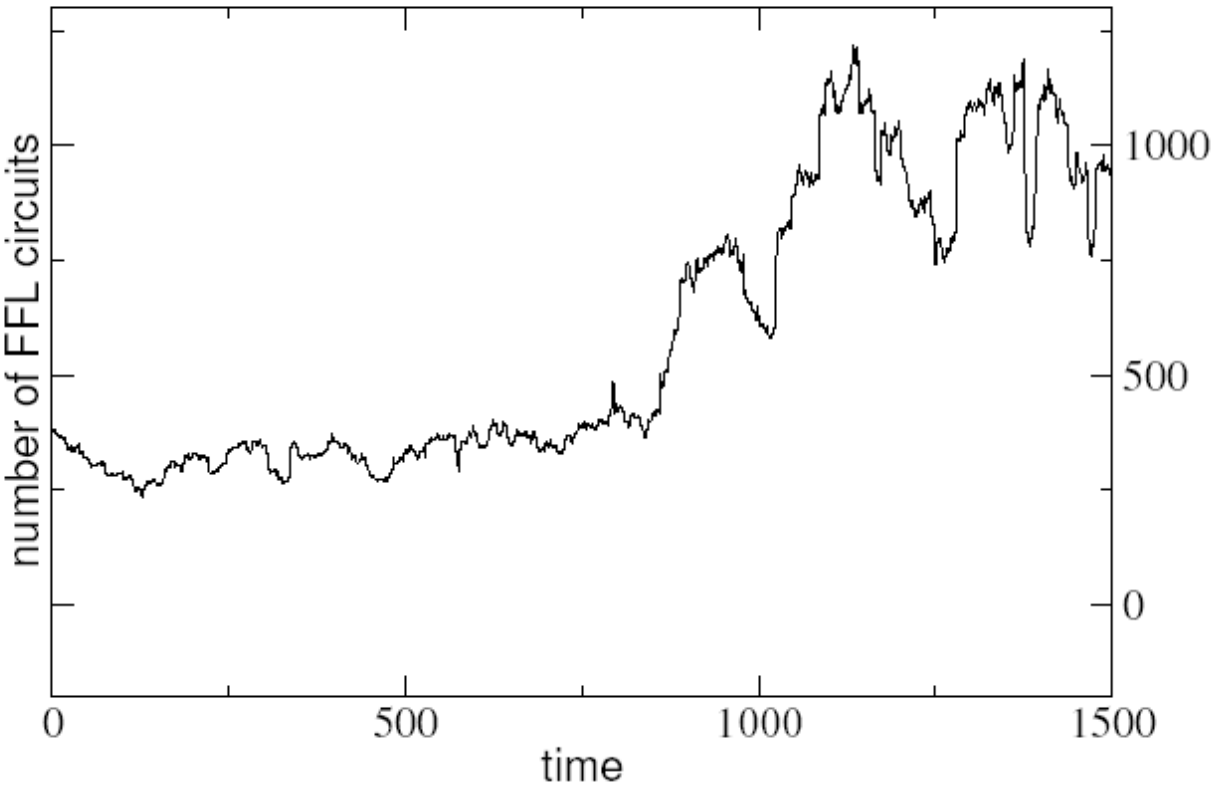
Toy model: Visualization of network restructuring during neutral evolution (hierarchical) structured network for free!

- Visualization of the network evolution (toy example):



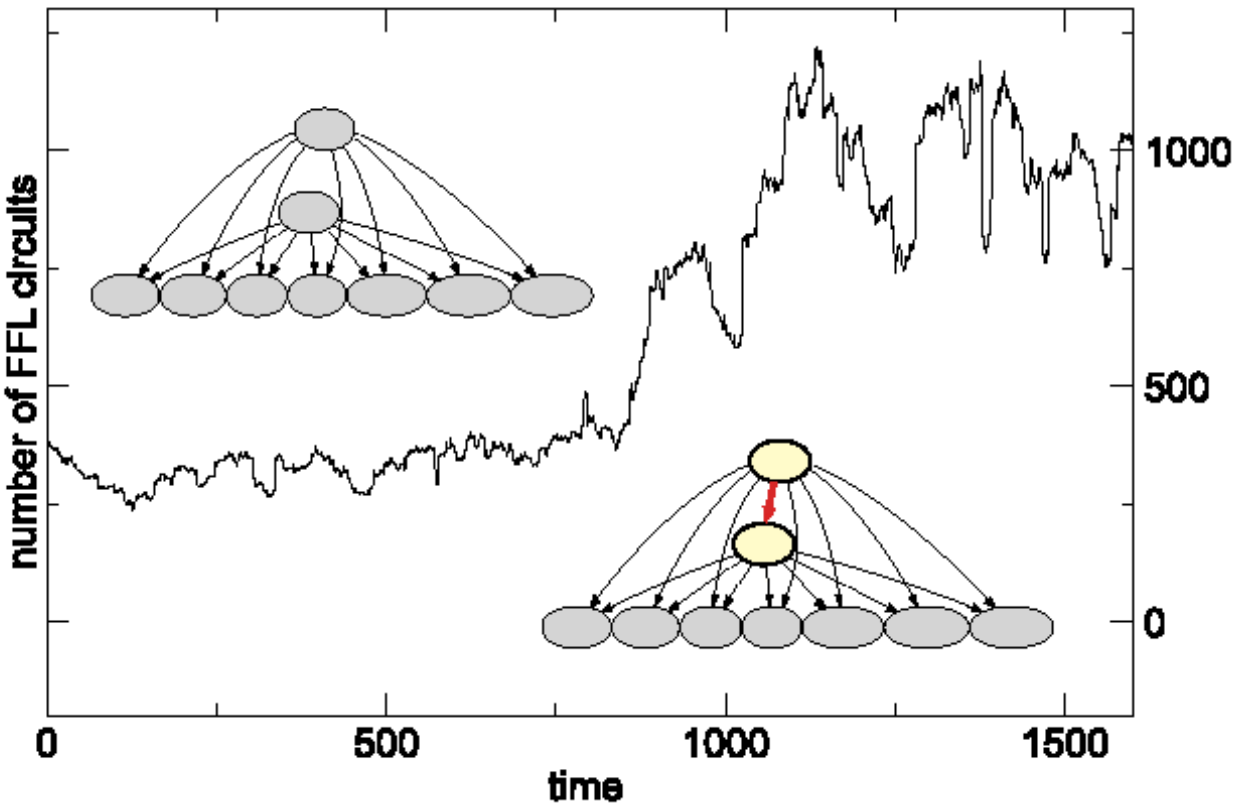
During evolution Sudden increase of FFL motifs: FFL as mutational signature

- Results of mutational dynamics at the microlevel:



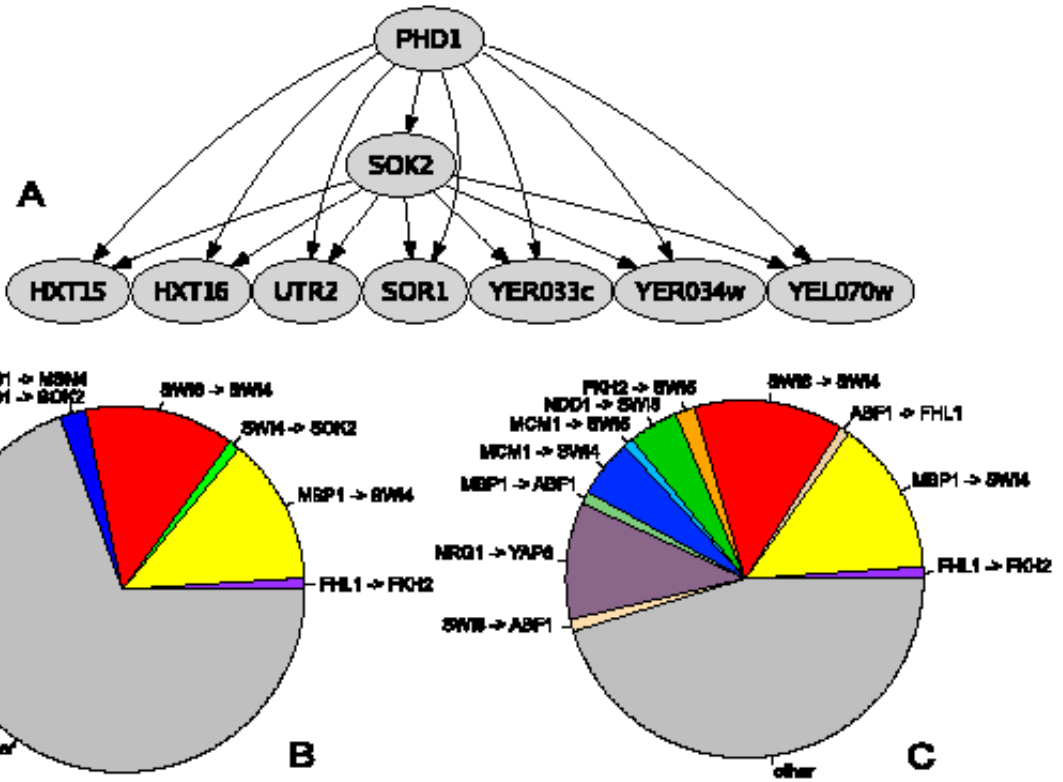
large increase of FFL motifs: Originate through duplication of hub + new connection

- Mechanics of massive FFL formation:



Over-representation of FFL motifs in Yeast: Duplication + connection of Hub genes

- Evidence in the yeast network:



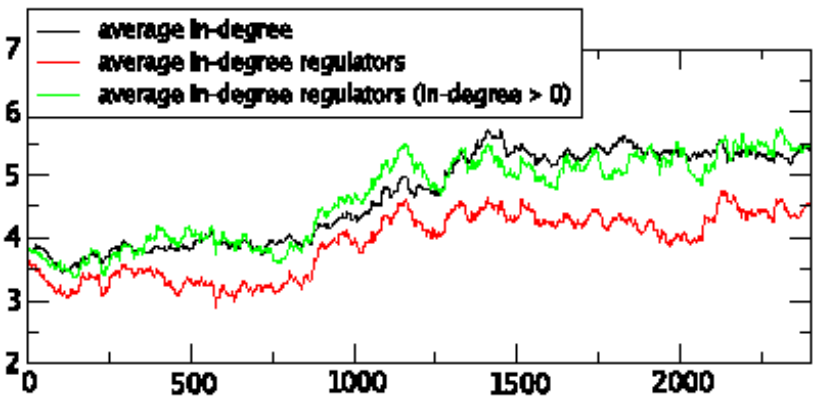
random mutations vs randomization

Randomization tests: keep everything the same EXCEPT feature to be tested

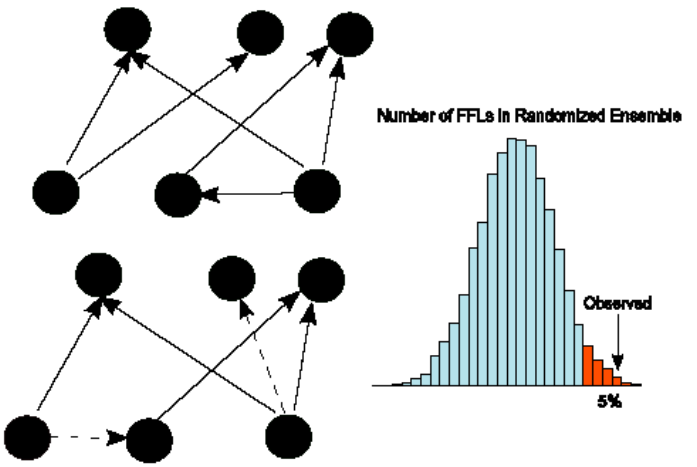
here: keep degree distribution
— test for FFL

BUT random mutations

do not conserve degree distribution



- Randomization test: swapping connections.

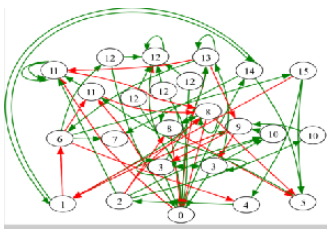


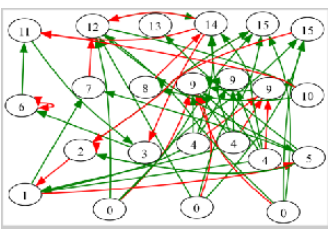
discussion/conclusions
Assessing relevance of observed evolved network structure in models or "real" life

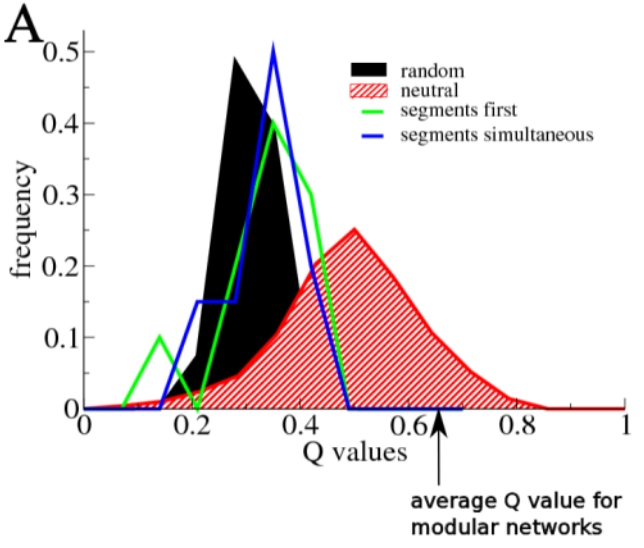
Random mutations lead to non-random structure
With or without selection of "something"

Random mutations as stochastic dynamical system goes to attractor

Example: Modularity in evolved GRN model Drosophila segmentation/differentiation







ten Tusscher & Hogeweg 2011: Evolution of Networks for Body Plan Patterning; Interplay of Modularity, Robustness and Evolvability

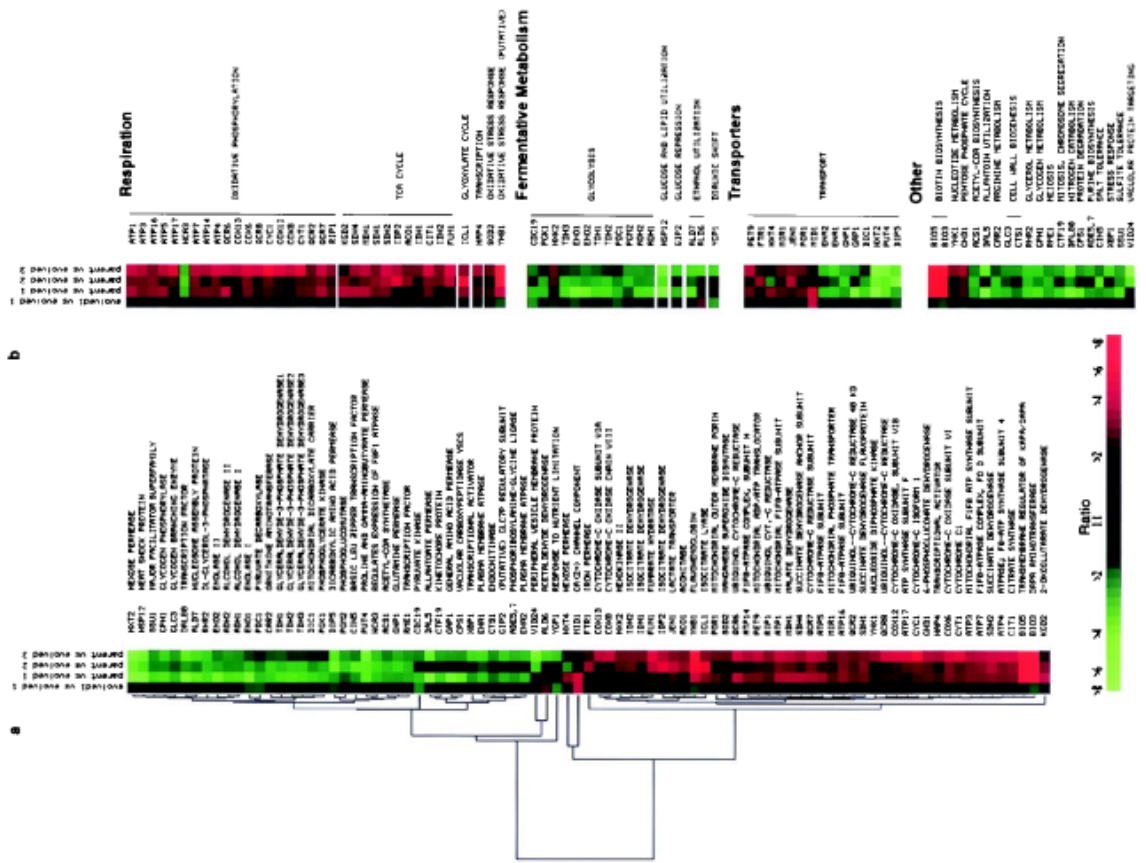
random networks as null model?

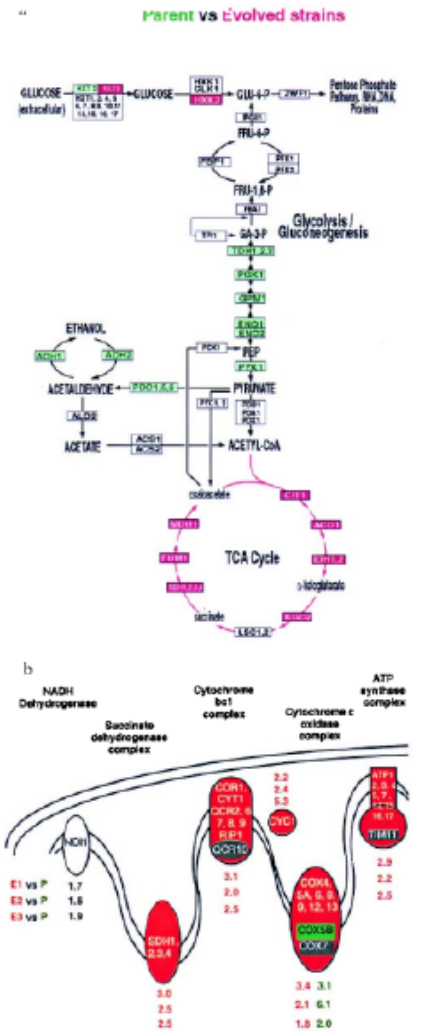
Yeast regulatory network evolution
Some "surprising" observations from short term evolution experiments
(Ferea et al 1999, Dunham et al 2002)

- 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?

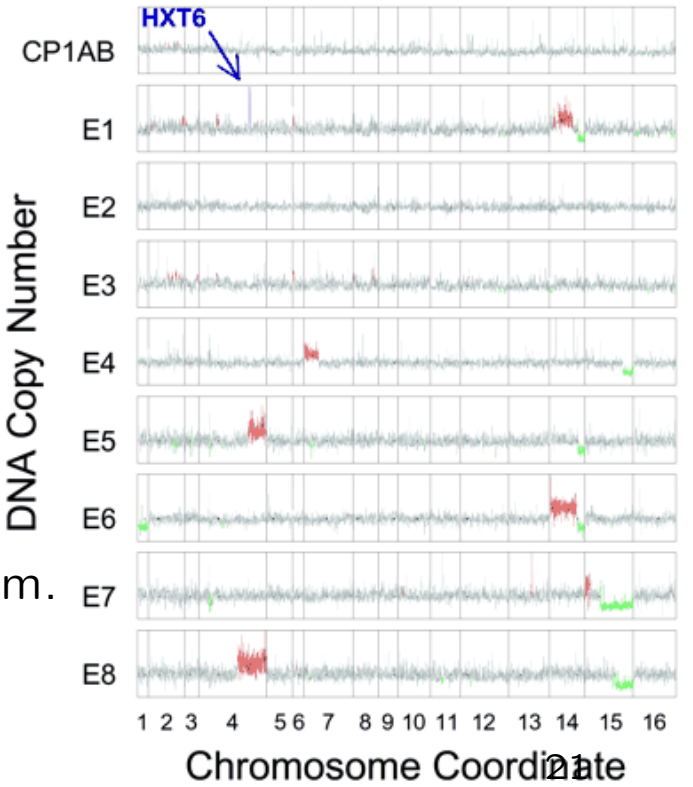
regulatory and/vs evolutionary 'adaptation' gene expression changes in strai evolved on low glucose med





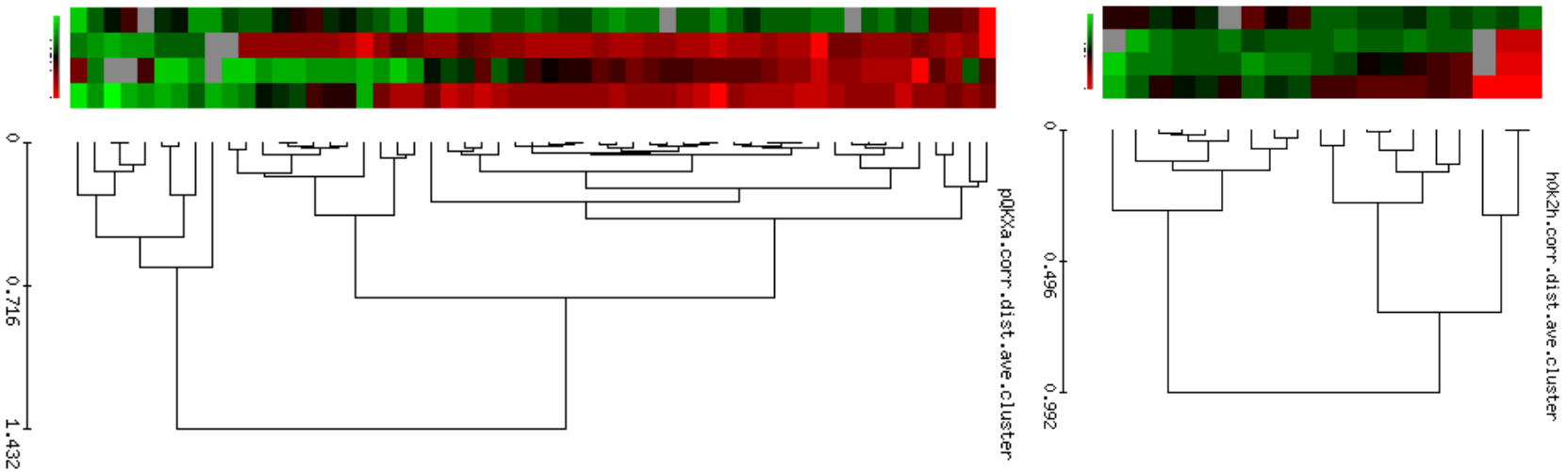
"Mutational priming" seen in yeast evolution

"Characteristic genome rearrangements in
experimental evolution of Saccharomyces cerevisiae"
(Dunham et al PNAS 2002)



repeated duplication and loss
at the same breakpoints
3* in C14 near CIT1 (citrate synthetase)
3* in C4 amplific. high-affinity hexose perm.
transposon-related sequences at
the breakpoints.

overexpression of deleted genes, underexpression of duplicated genes



duplicated genes

deleted genes

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 structure?

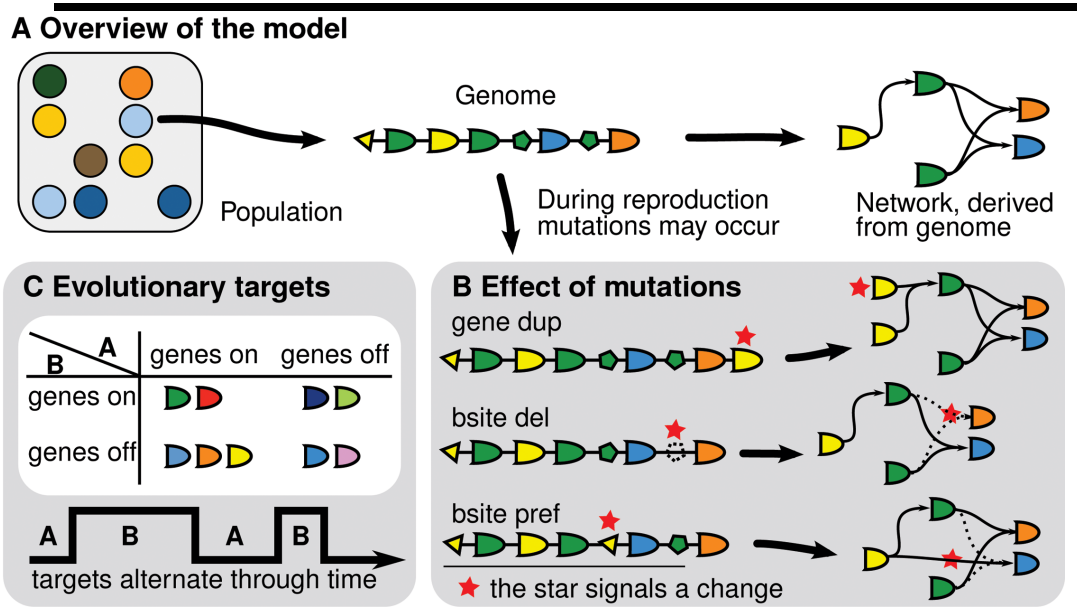
By evolution of transcriptome structure?

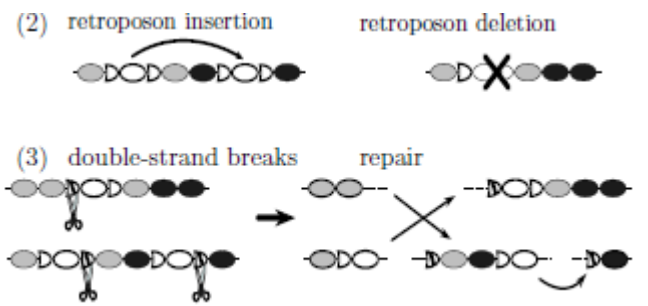
Crombach & H. 2007 MBE, 2008 PLOS-CompBio, Hajji Msc thesis 2019

basic course grained representation of a genome

chain of genes, TF binding sites, transposon, ...

"Pearls on a string model"





Genetic operators: beyond point mutations:

(DupDel: single genes, TBS; LCR: random, targeted)

selforganization of genomes by transposon mutational dynamics
evolution of evolvability

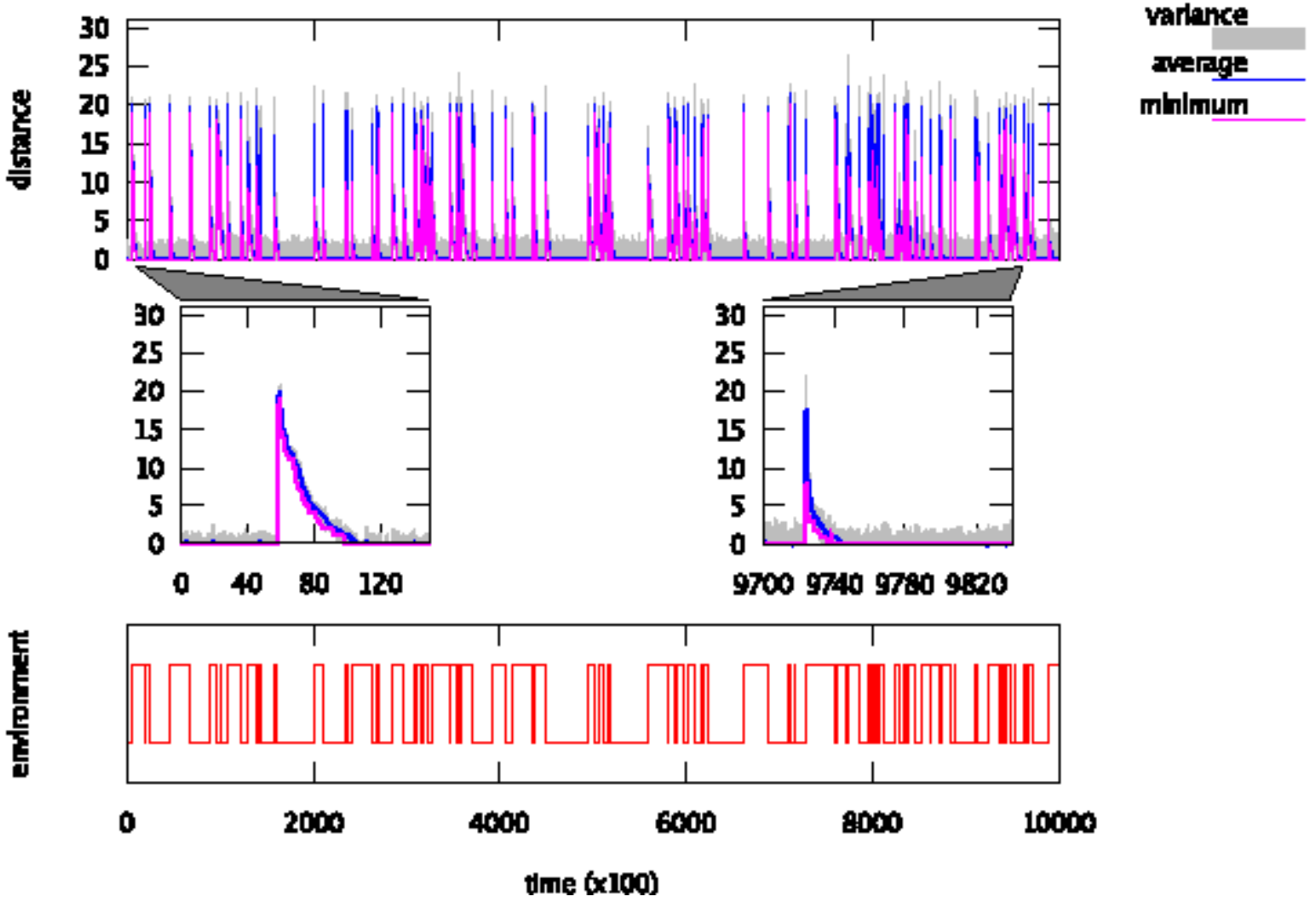
mutational dynamics

- gene duplication; gene deletion.
- transposon duplication;
- transposon deletion; leaves breakpoints
- double stranded breaks and repair
  −> gross chromosomal rearrangement

selection

- fluctuating environment
- need 2 copies of part of the genes in one environment

Crombach and Hogeweg MBE 2007

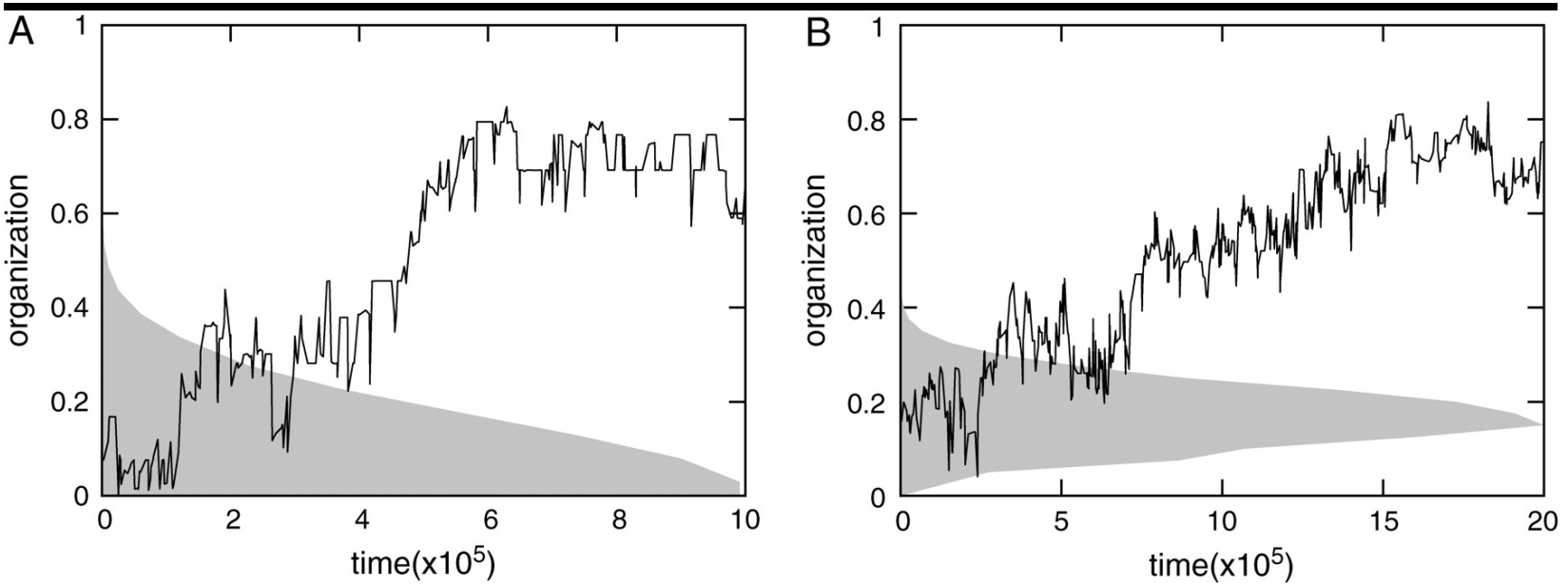

variance
average
minimum
distance
environment
time (x100)
0
5
10
15
20
25
30
0
40
80
120
9700
9740
9780
9820
0
2000
4000
6000
8000
10000

self organization of the genomes
clustering of genes which need to be duplicated

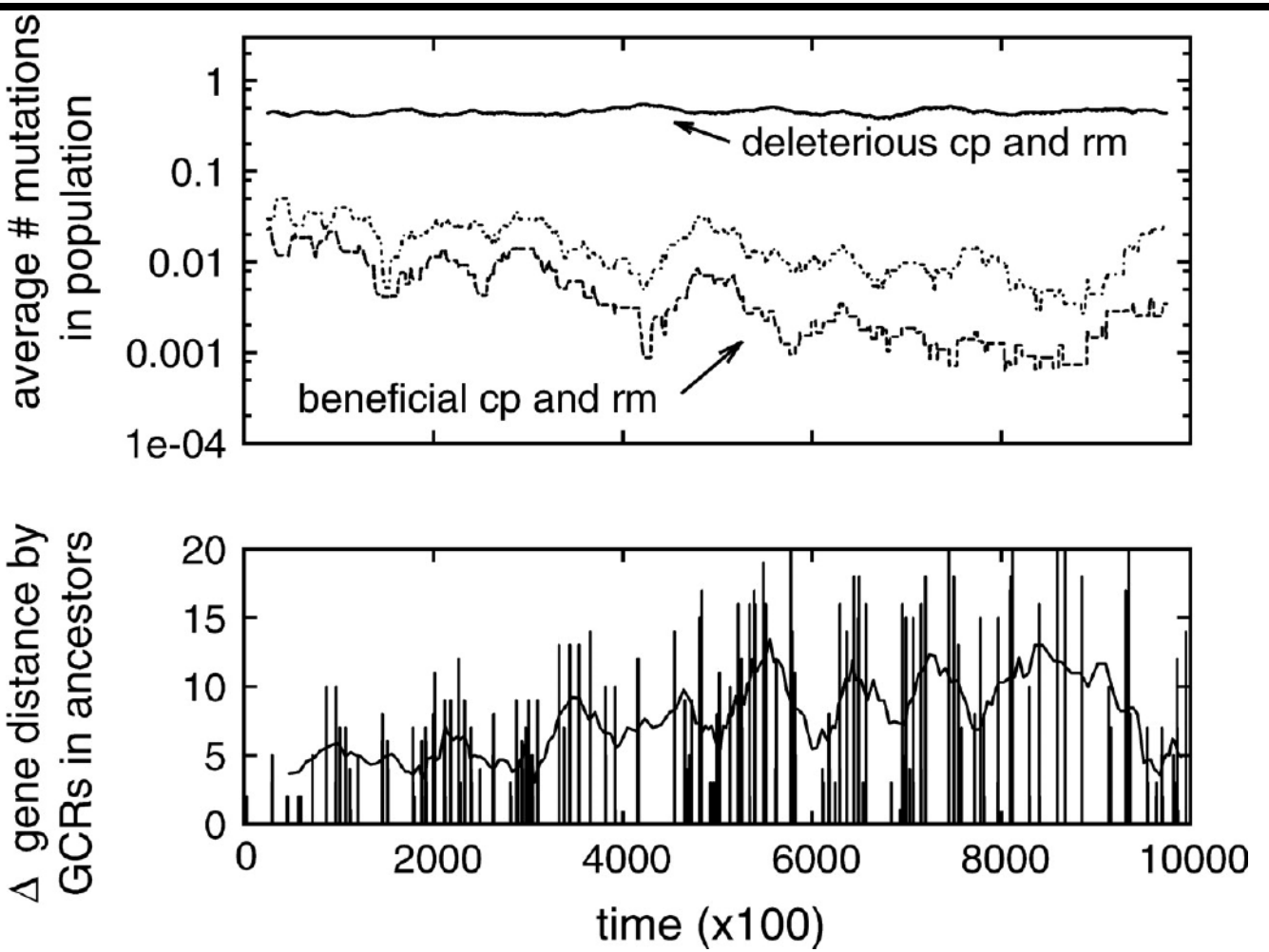
- Randomly generated genome (at start simulation):

- Evolved genome, grouping of genes as a side effect:

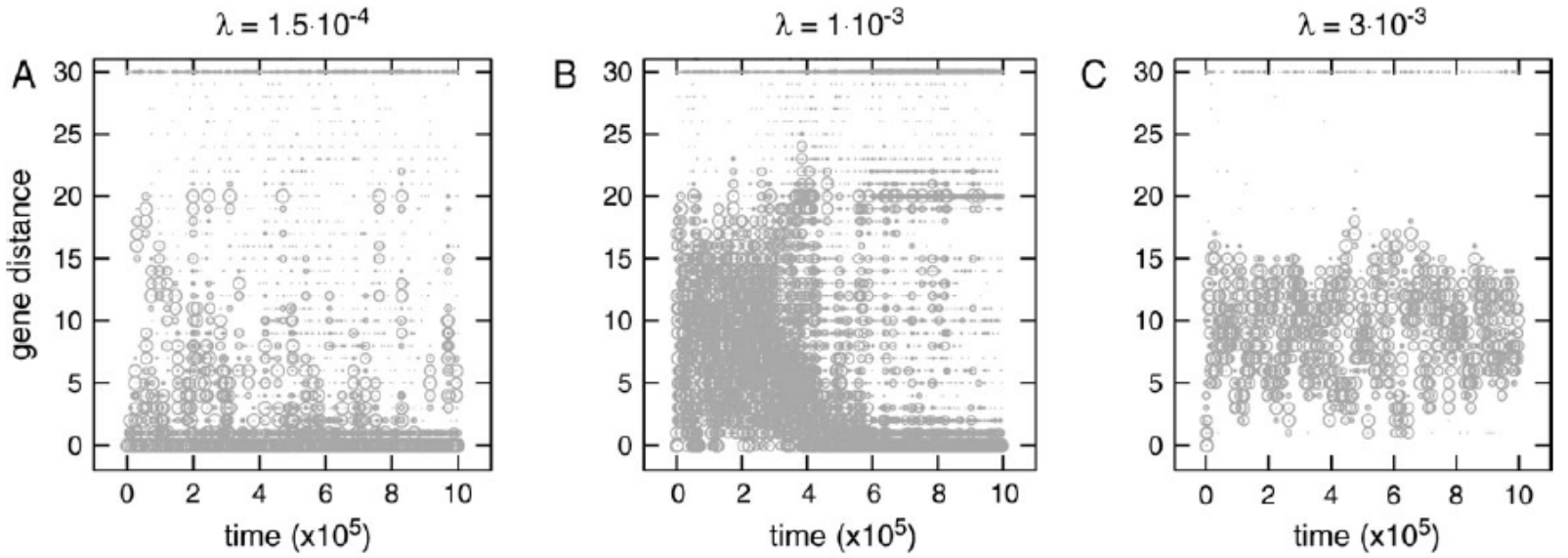
genome organization over time



mutations over time



Evolution of evolvability: timescales



genome organization evolves when no adaptation is possible - and so enables adaptation

conclusions

Very simple demonstration of mutational priming through genome structuring

Yeast example also transposon remnants on breakpoints

Much pattern analysis research:

observation:
older transposons often in "important" (e.g. regulatory) regions

Evolution of Regulation based mutational priming

v

network dynamics and fitness

Network update:

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

fitness:

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

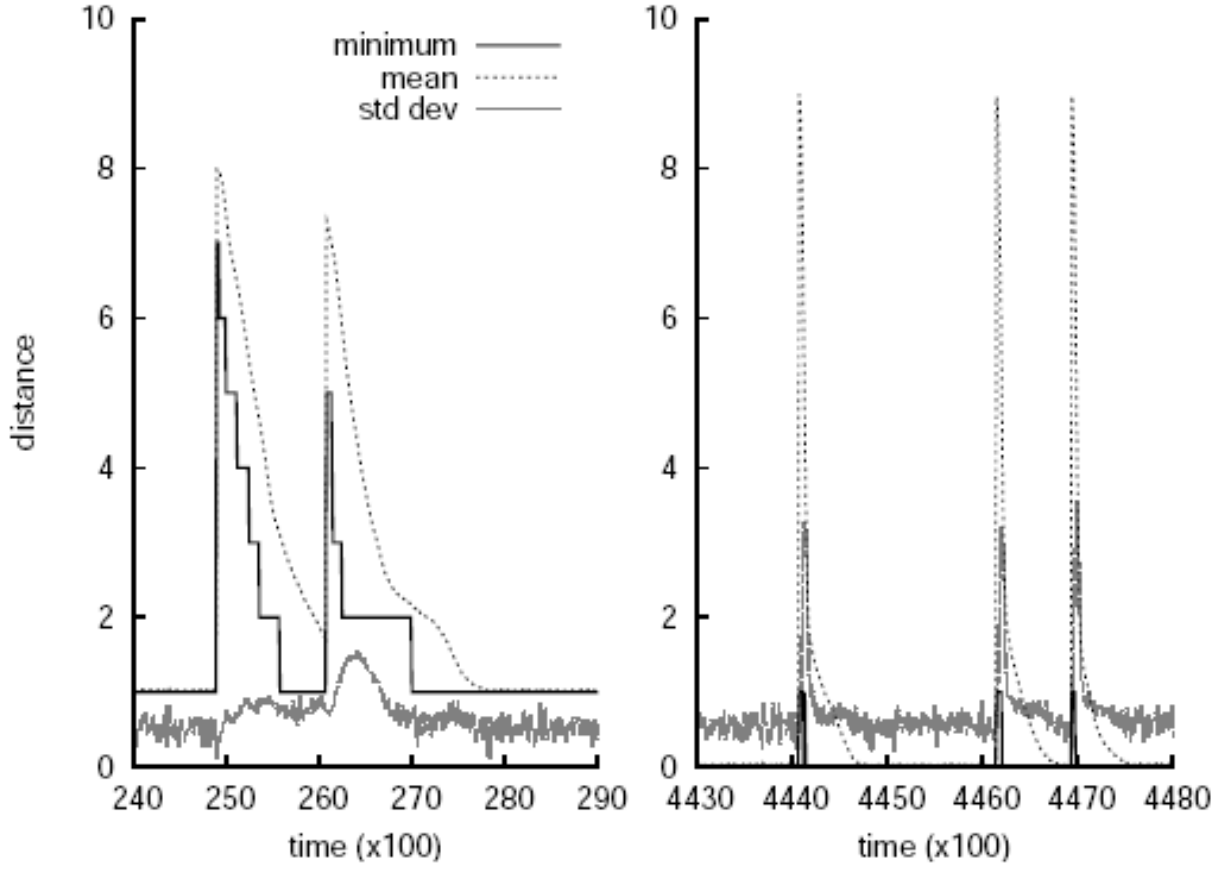
Switching environment

Gene expression in attractor of GRN defines fitness

each environment defines in terms of target gene expression

Crombach & H 2008

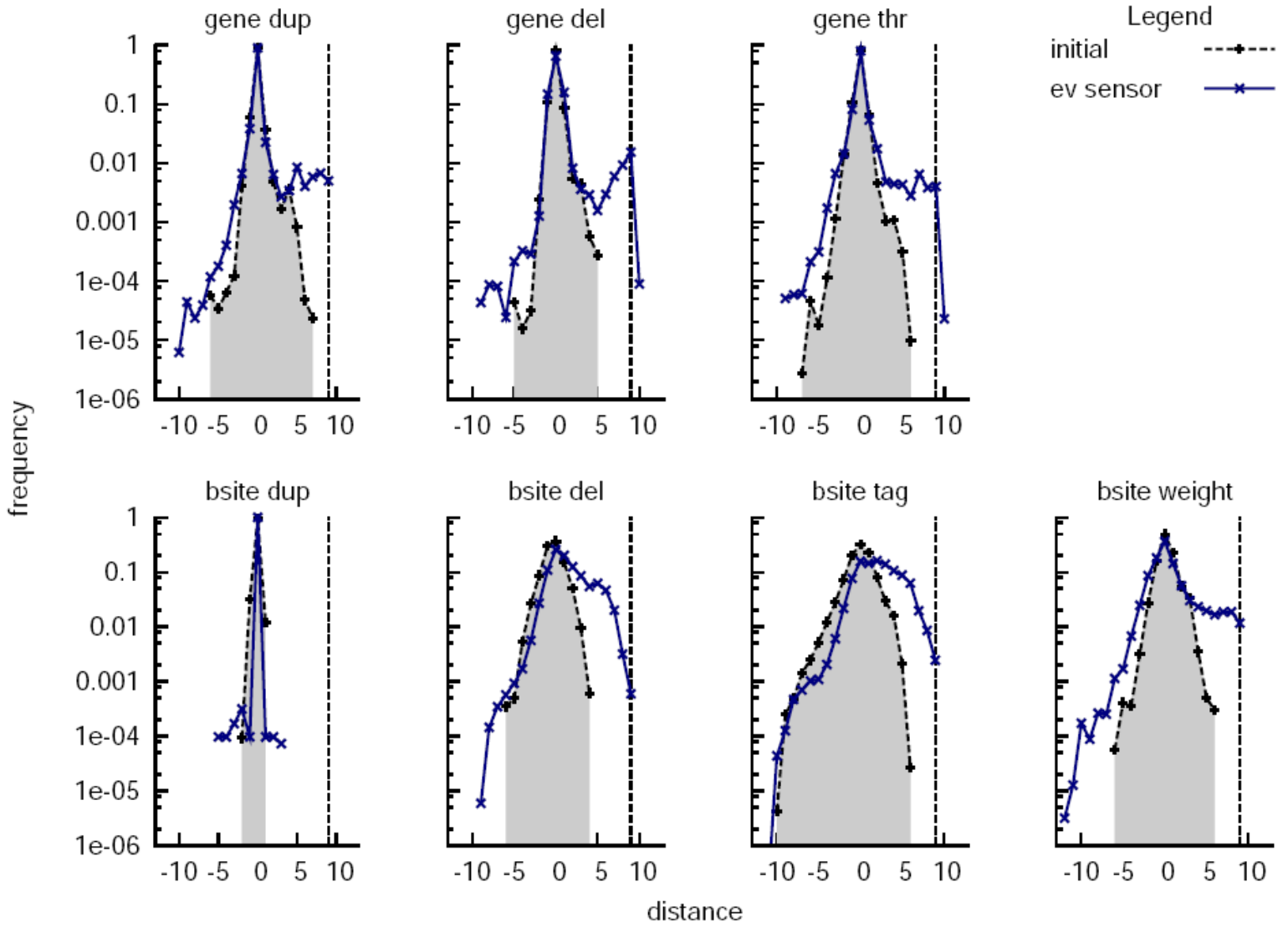
improved (faster) evolvability observed



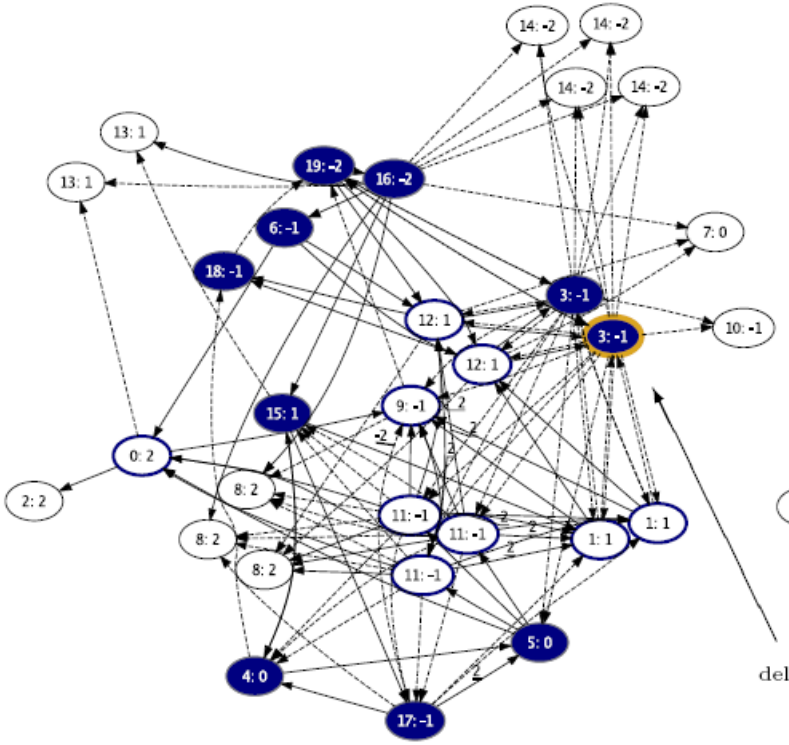
Hamming distance improvement to opposite target

Regulatory Mutational Priming:

Many different mutations lead to "beneficial" adaptation

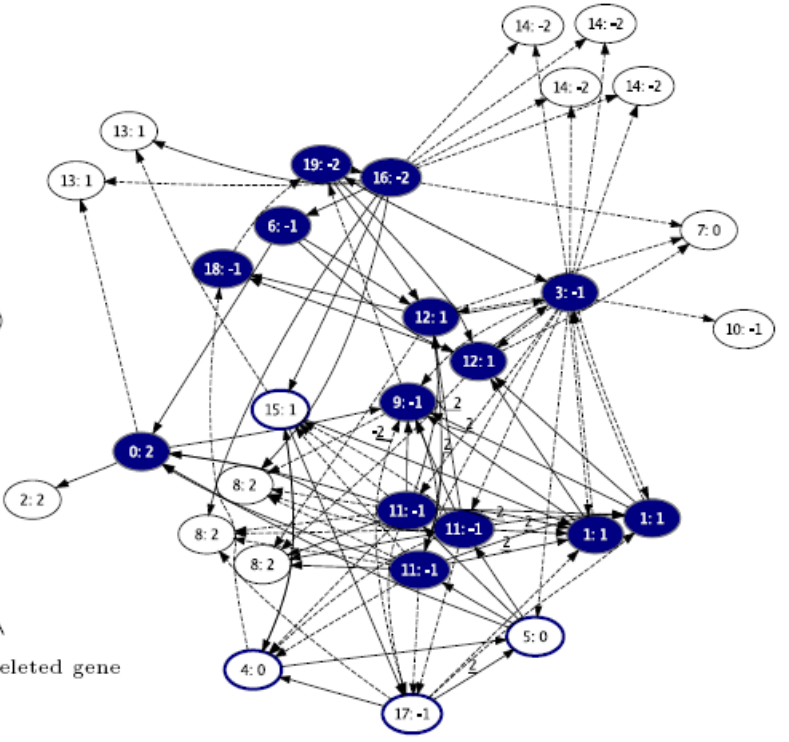

gene dup
gene del
gene thr
Legend
initial
ev sensor
bsite dup
bsite del
bsite tag
bsite weight
frequency
distance
1
0.1
0.01
0.001
1e-04
1e-05
1e-06
-10 -5 0 5 10

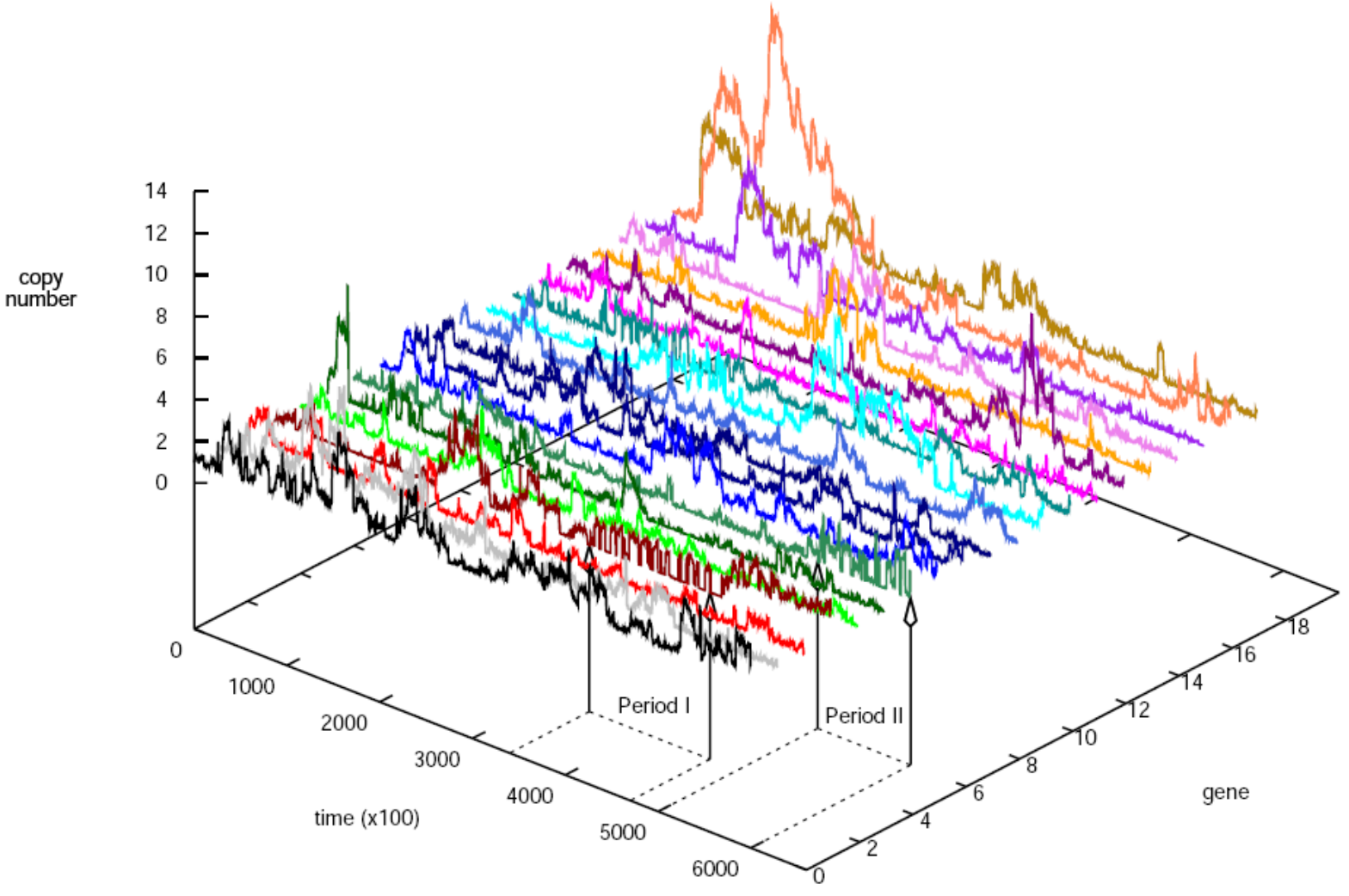
Parent (t = 457 755):



deleted gene

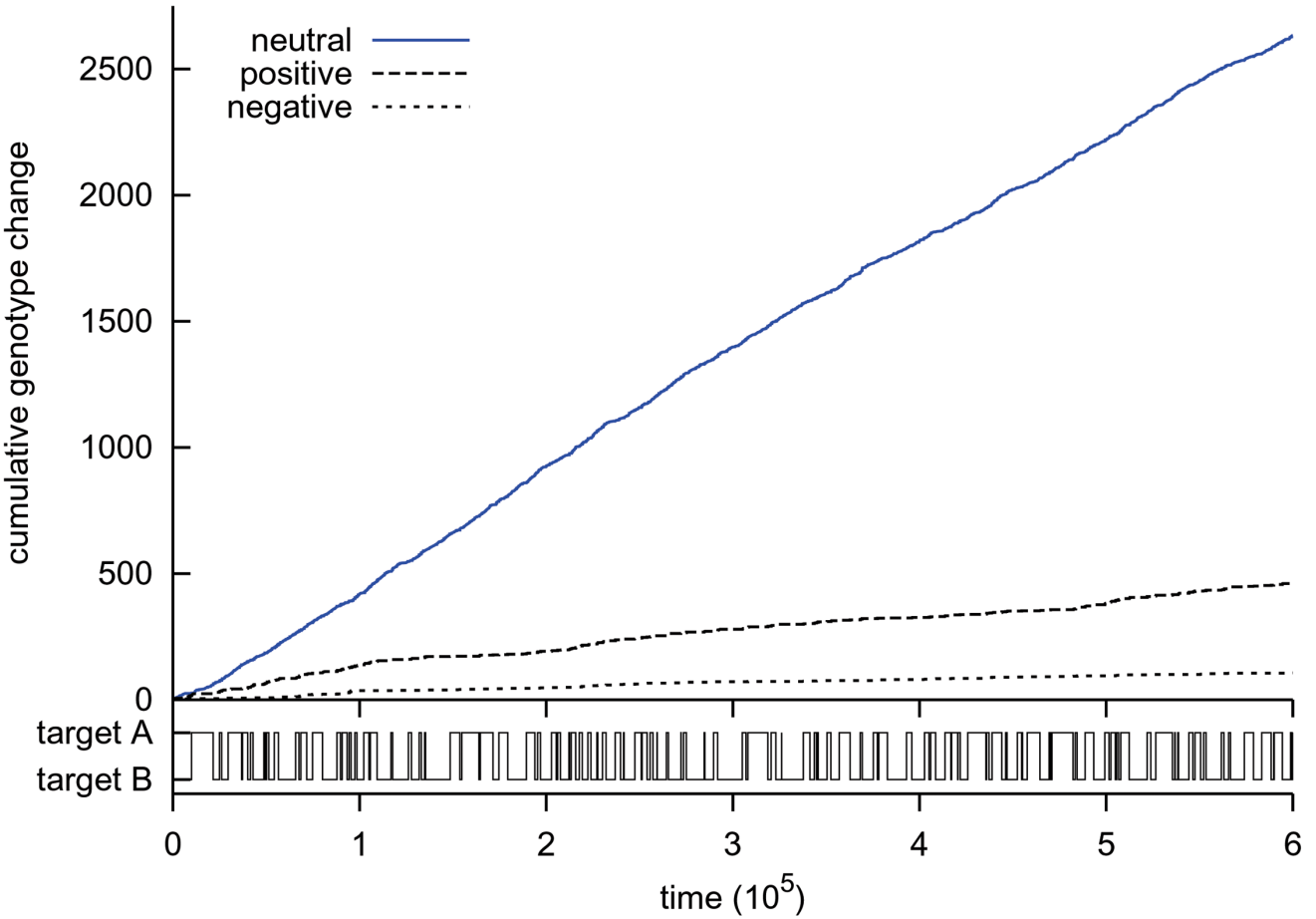
Child (t = 457 758):



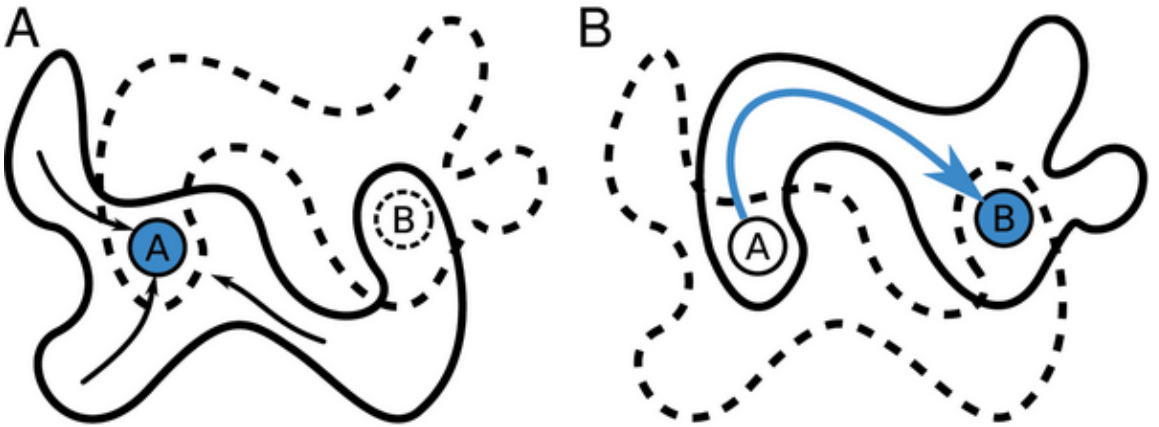


“evolutionary sensor”

Neutral drift far greater than adaptive change!



evolution of evolvability and bases of attraction



single/few Mutations destabilize attractor of env 1
and becomes state in domain of attraction of of env 2

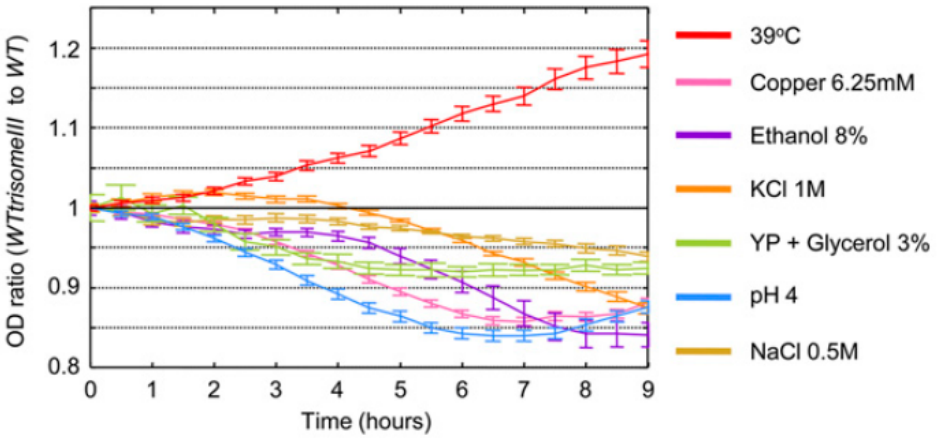
conclusions
Evolution of genomes and gene regulatory networks
evolution of evolvability

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)
- State space may remain very similar despite attractor switch, but can also change drastically
- *"individual vs population based" evolution of evolvability*

Mutational Priming in Yeast: Genome organization and regulion adaptation to high temperature: short term vs long term effects

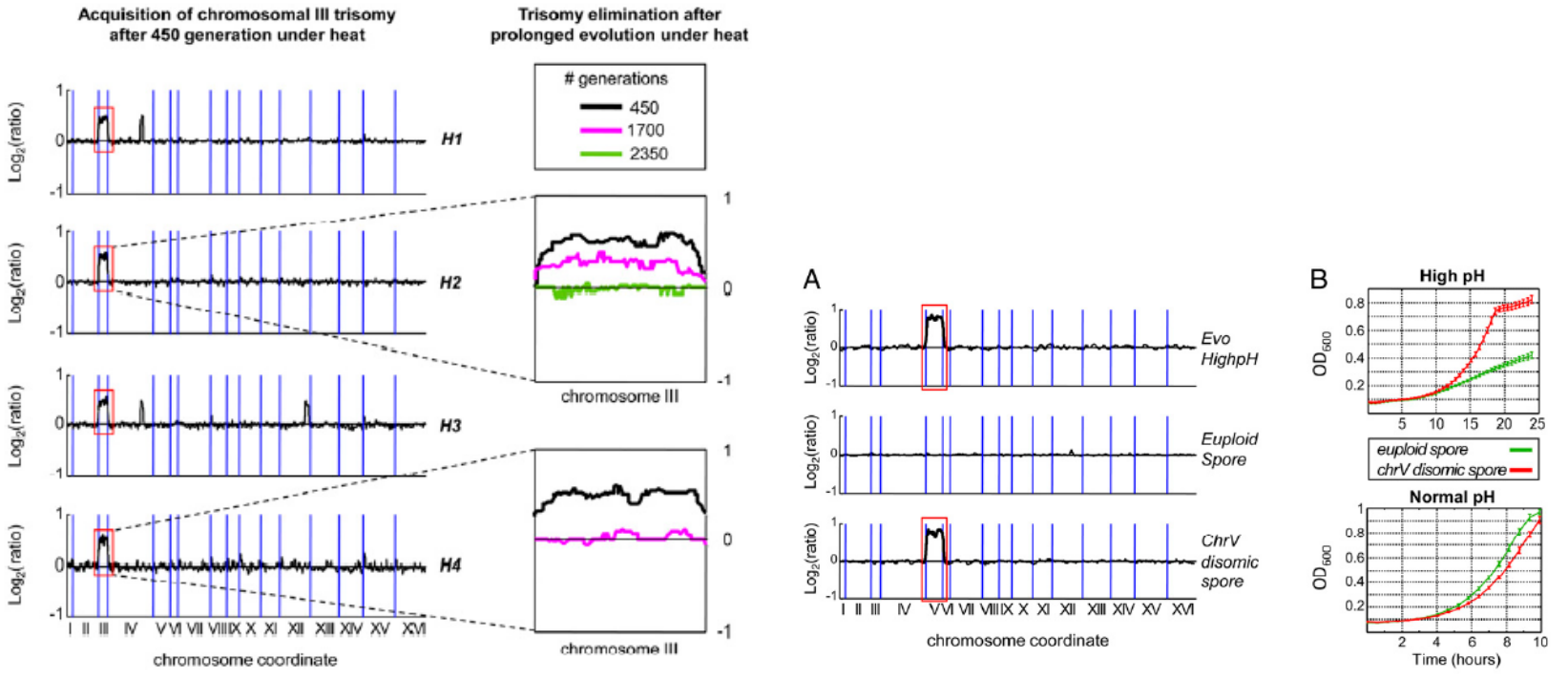
Diploid yeast adapted to 'normal' temperture of 30 C placed in 39 C. After 450 generations:
Increased growthrate (specific for temperature)



Next continued at 39 C

Yona et al, PNAS 2012: Chromosomal duplication is a transient evolutionary solution to stress

Yeast adaptation to high temperature by duplication of resp chromosome 3 (and at high ph chromosome 5)



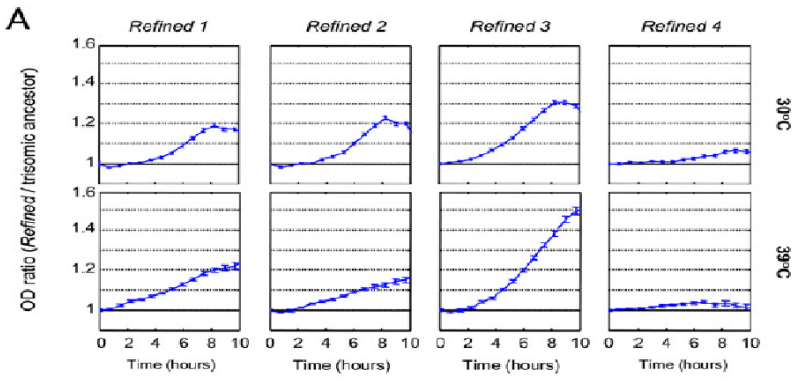
trisomy not retained

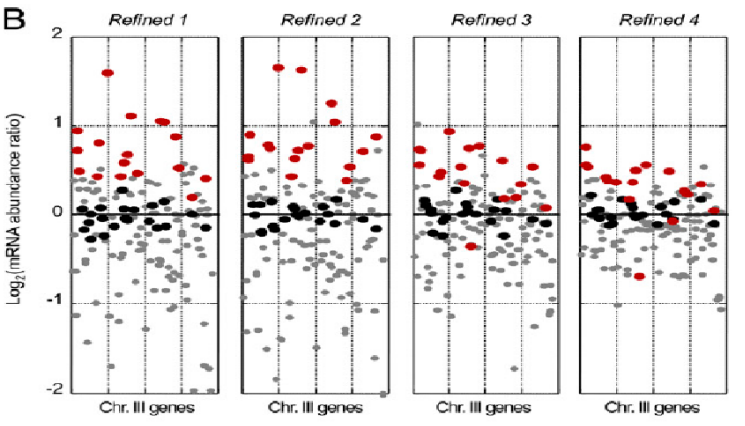
Yona et al, PNAS 2012: Chromosomal duplication is a transient evolutionary solution to stress

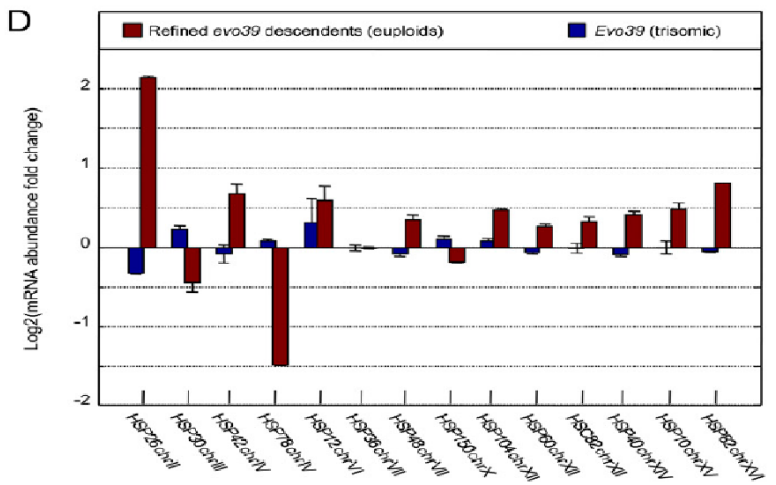
chromosomal duplication is followed by regulatory adaptation

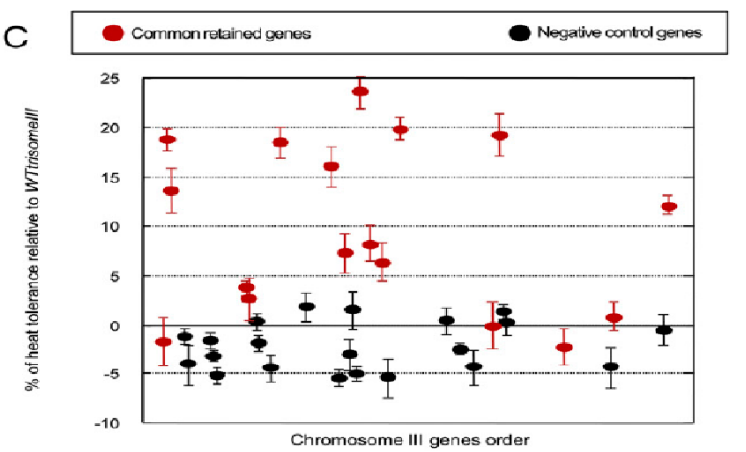
continued increase growthrate

overexpression retained of some genes







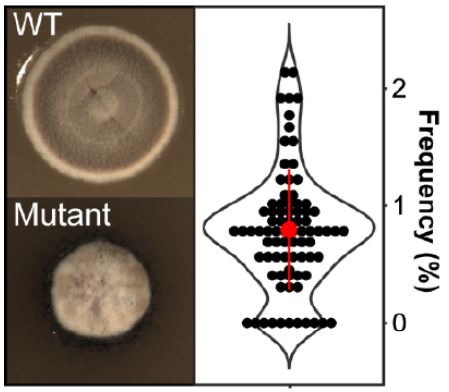


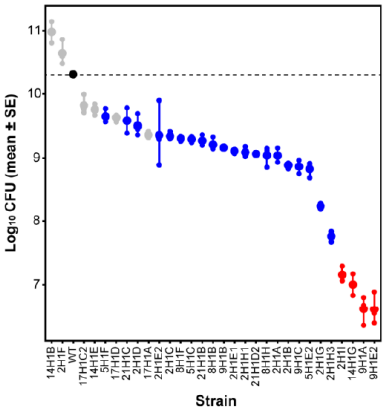
overextreeion HSP only later

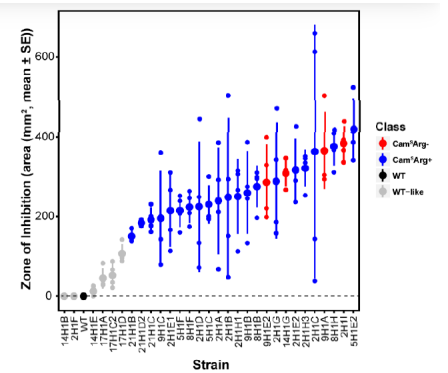
retained genes increase growthrate in ancestor

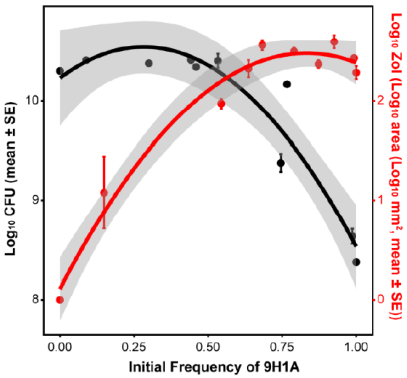
Mutational priming in Streptomyces: quasispecies

Antibiotic production in Streptomyces is organized by a division of labor through terminal genomic differentiation. Zhang ....Rozen 2020



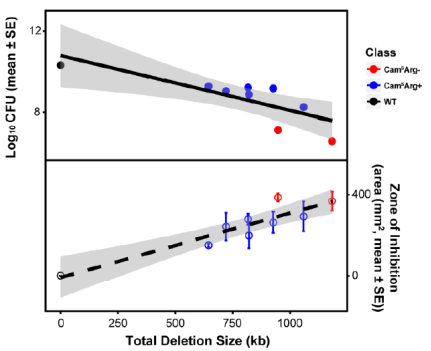


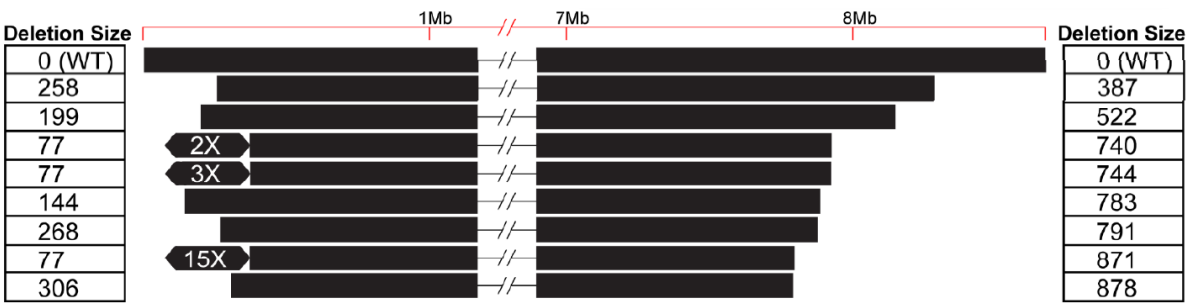




High Mut. Rate Mut. fitness antibiotics production QS fitness

only $> 50\%$ of mutants lowers colony fitness





Genome structure and targeted mutations