

# **Genome evolution: coding structures and evolvability**

Course Computational Biology 2018/2019; Paulien Hogeweg;  
Theoretical Biology and Bioinformatics Grp Utrecht University  
<sub>1</sub>

# 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)

## NOW

## Evolution of genomes and transcription regulatory networks

Mutational operators Not only point mutations

emergent structure by random mutations (without selection)

evolution of mutational neighborhood continued:

mutational priming

–by evolution of genome structure (*mutational operators*)

–by evolution of genotype-phenotype map

*evolution of evolvability*

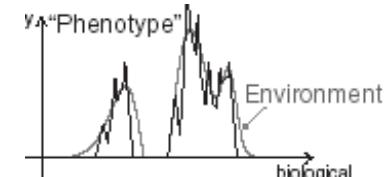
*Random mutation - non-random effects*

# GENOME EVOLUTION: 3 modeling frameworks

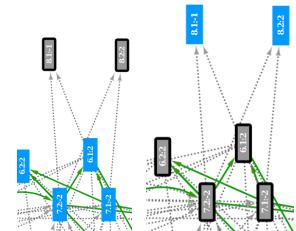
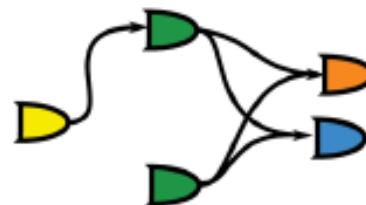
## genome structure and genotype to phenotype/fitnessmap

---

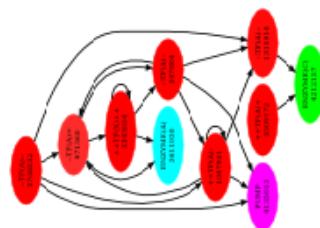
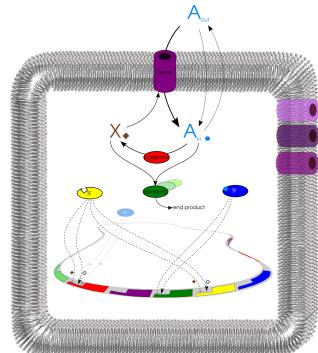
AEVOL bit (nucleotide) level coding of genome allows evolution of new genes!



PoS course grained genome allows multilevel (evolving) GPf mapping



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

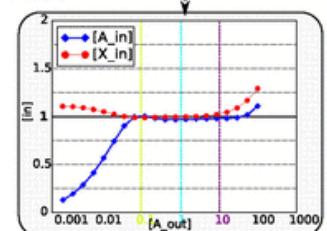


$$\frac{d[A]}{dt} = \frac{[A]_{out}[X]V_{max_p}[Prot]_p}{([A]_{out} + K_{Ap})([X] + K_{Xp})} \quad (2)$$

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

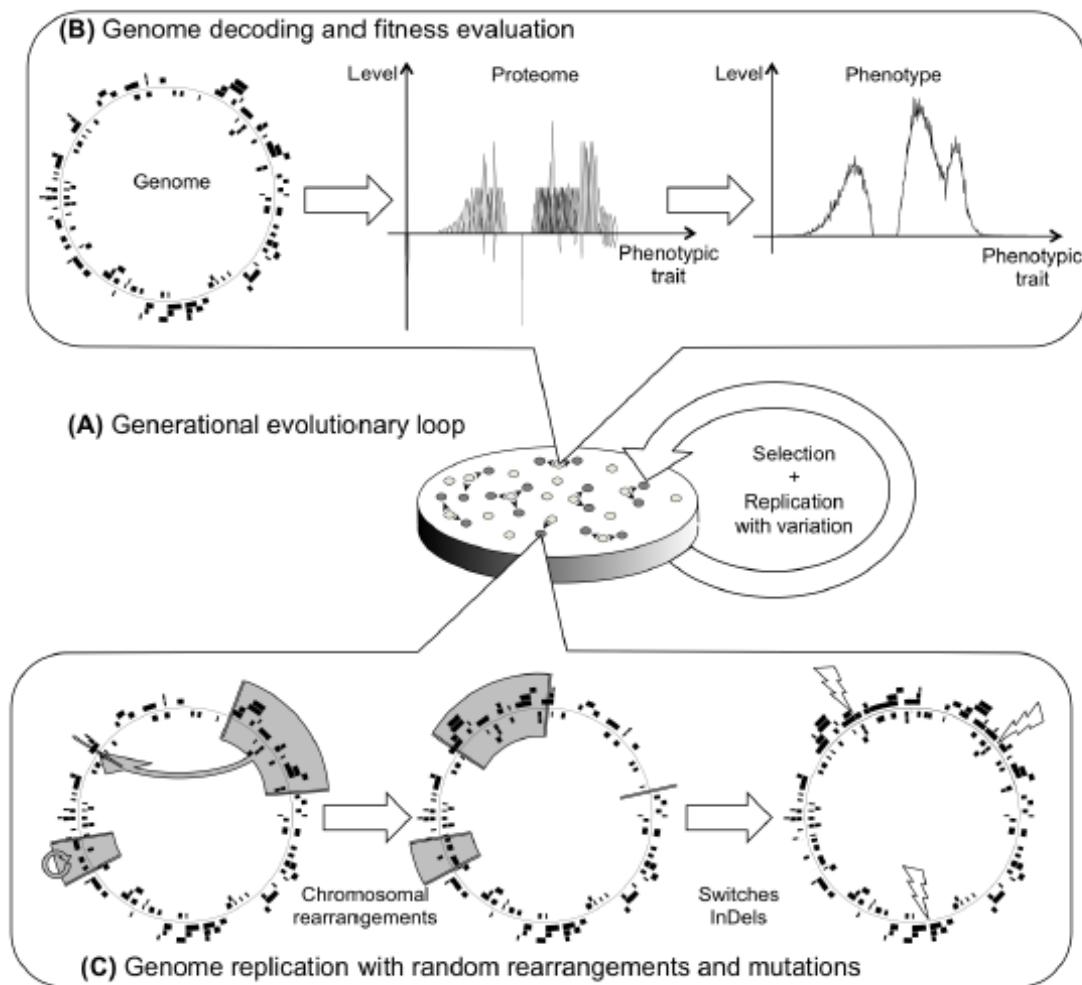
$$\frac{d[A]}{dt} = -\frac{[Prot]_c[A]V_{max_c}}{[A] + K_{Ac}} \quad (4)$$

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

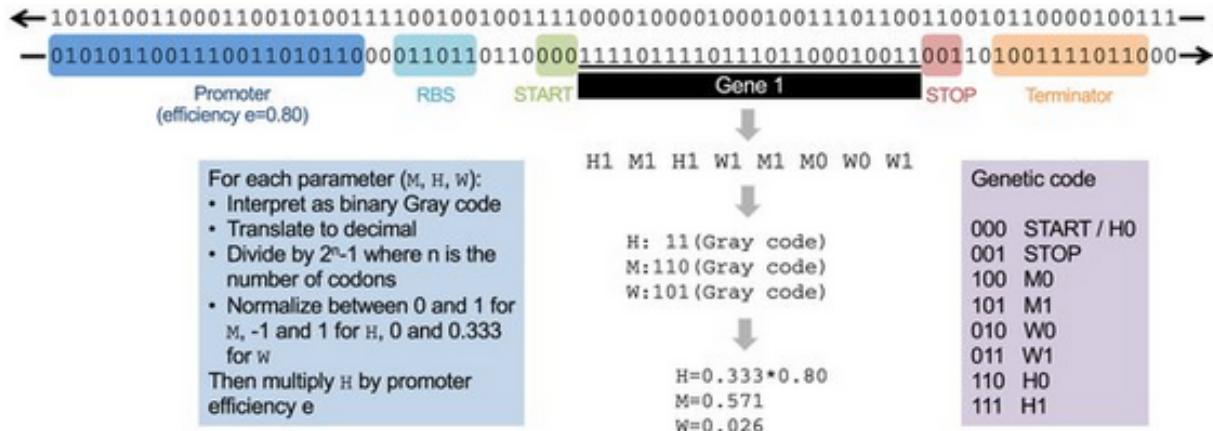




# Aevol model structure (Beslon)



# AEVOL “genetic” code



# Long term evolution of WT strains: Genome structure dependent on mutation rates

---

*evolved in constant environment; fine grained genome structure*

## Bacterium-like strains

compact genome closely packed genes

Aevol:

population on 40x40grid

pmut.rate=  $10^{-6}$  mut/bp/gen

indels =  $10^{-6}$  mut/bp/gen

LCR = $1^{-5}$  mut/bp/gen



## Virus-like strains

small genomes, overlapping genes, one start site'

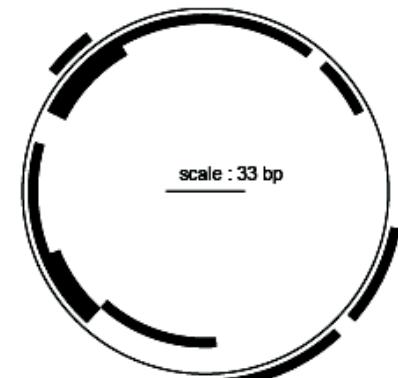
Aevol:

well mixed population 5000

pmut.rate=  $10^{-4}$  mut/bp/gen

indels =  $10^{-4}$  mut/bp/gen

LCR = $10^{-4}$  mut/bp/gen



# **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%)

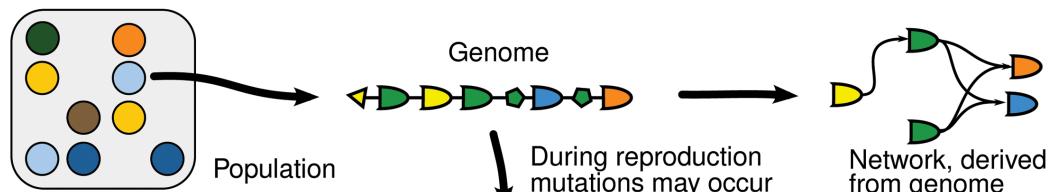
Involves GCR - and the same ones in multiple experiments

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

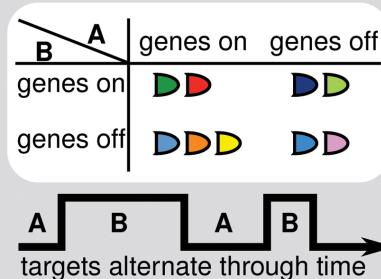
# basic course grained representation of a genome chain of genes, TF binding sites, transposon, ....

*"Pearls on a string model"*

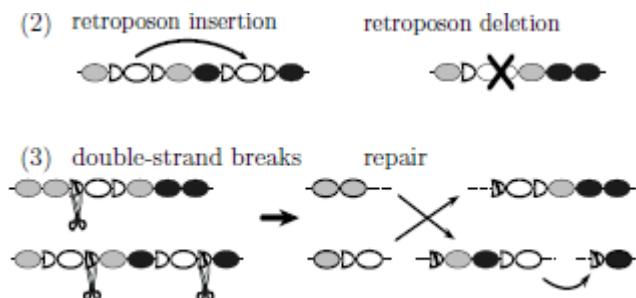
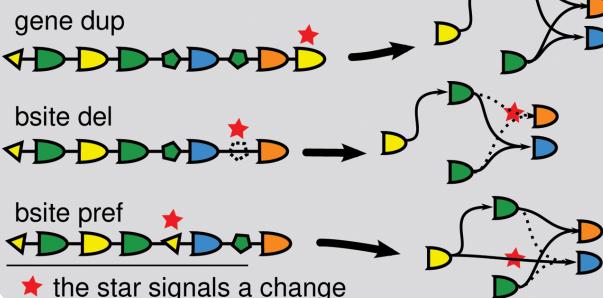
## A Overview of the model



## C Evolutionary targets



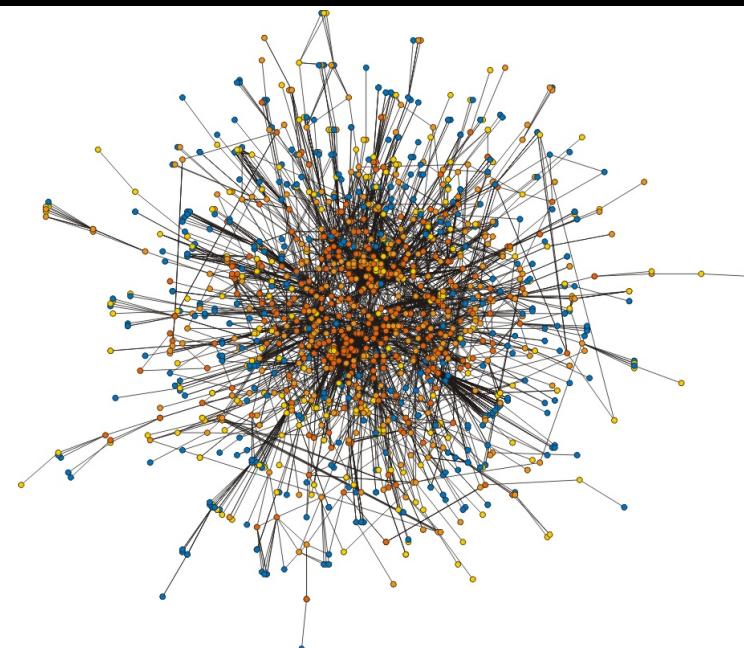
## B Effect of mutations



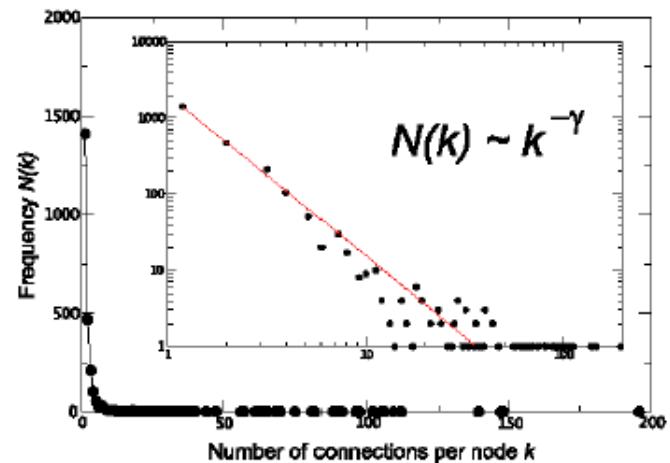
Genetic operators: beyond point mutations:  
(DupDel: single genes, TBS; LCR: random, targeted)

transcription networks: observed network structure  
global property: powerlaw of degree distribution  
local property: overrepresentation of motifs

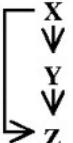
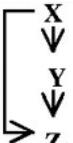
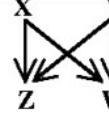
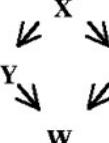
---



yeast network



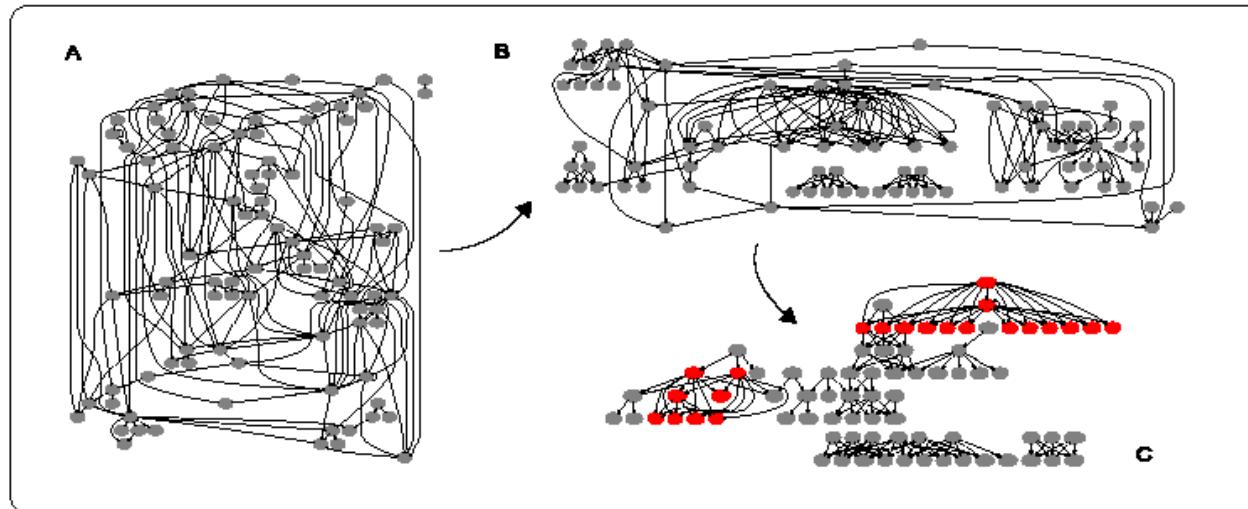
## network motifs (Alon)

Network	Nodes	Edges	$N_{\text{real}}$	$N_{\text{rand}} \pm \text{SD}$	Z score	$N_{\text{real}}$	$N_{\text{rand}} \pm \text{SD}$	Z score
<b>Gene regulation (transcription)</b>				Feed-forward loop			Bi-fan	
<i>E. coli</i>	424	519	40	$7 \pm 3$	10	203	$47 \pm 12$	13
<i>S. cerevisiae*</i>	685	1,052	70	$11 \pm 4$	14	1812	$300 \pm 40$	41
<b>Neurons</b>				Feed-forward loop			Bi-fan	
<i>C. elegans†</i>	252	509	125	$90 \pm 10$	3.7	127	$55 \pm 13$	5.3
<b>Food webs</b>				Three chain			Bi-parallel	
Little Rock	92	984	3219	$3120 \pm 50$	2.1	7295	$2220 \pm 210$	25
Ythan	83	391	1182	$1020 \pm 20$	7.2	1357	$230 \pm 50$	23
St. Martin	42	205	469	$450 \pm 10$	NS	382	$130 \pm 20$	12
Chesapeake	31	67	80	$82 \pm 4$	NS	26	$5 \pm 2$	8
Coachella	29	243	279	$235 \pm 12$	3.6	181	$80 \pm 20$	5
Skipwith	25	189	184	$150 \pm 7$	5.5	397	$80 \pm 25$	13
B. Brook	25	104	181	$130 \pm 7$	7.4	267	$30 \pm 7$	32

**Random mutations (duplications/deletion),  
NO SELECTION  
lead to (hierarchical) structured network for free!**

---

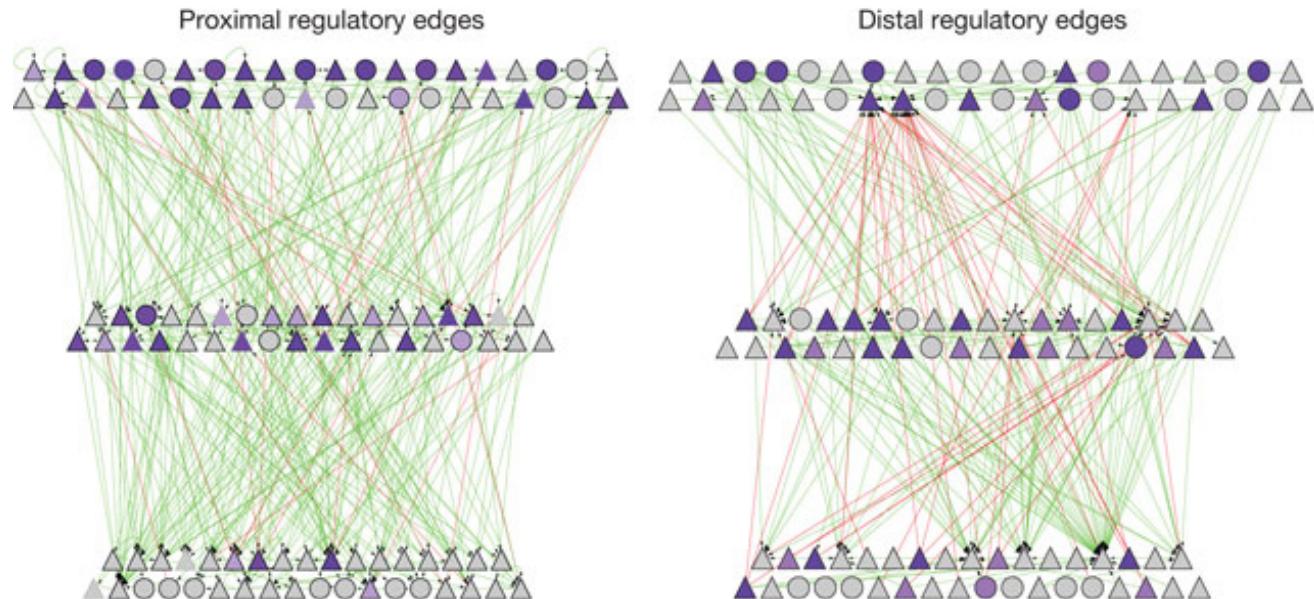
- ▶ Visualization of the network evolution (toy example):



mutation rates: gene dup/del  $10^{-3}$  BS dup/del  $8.10^{-3}$  TF mutation  $5.10^{-3}$  BS mutation  $8.10^{-4}$

global structure: powerlaw degree distribution

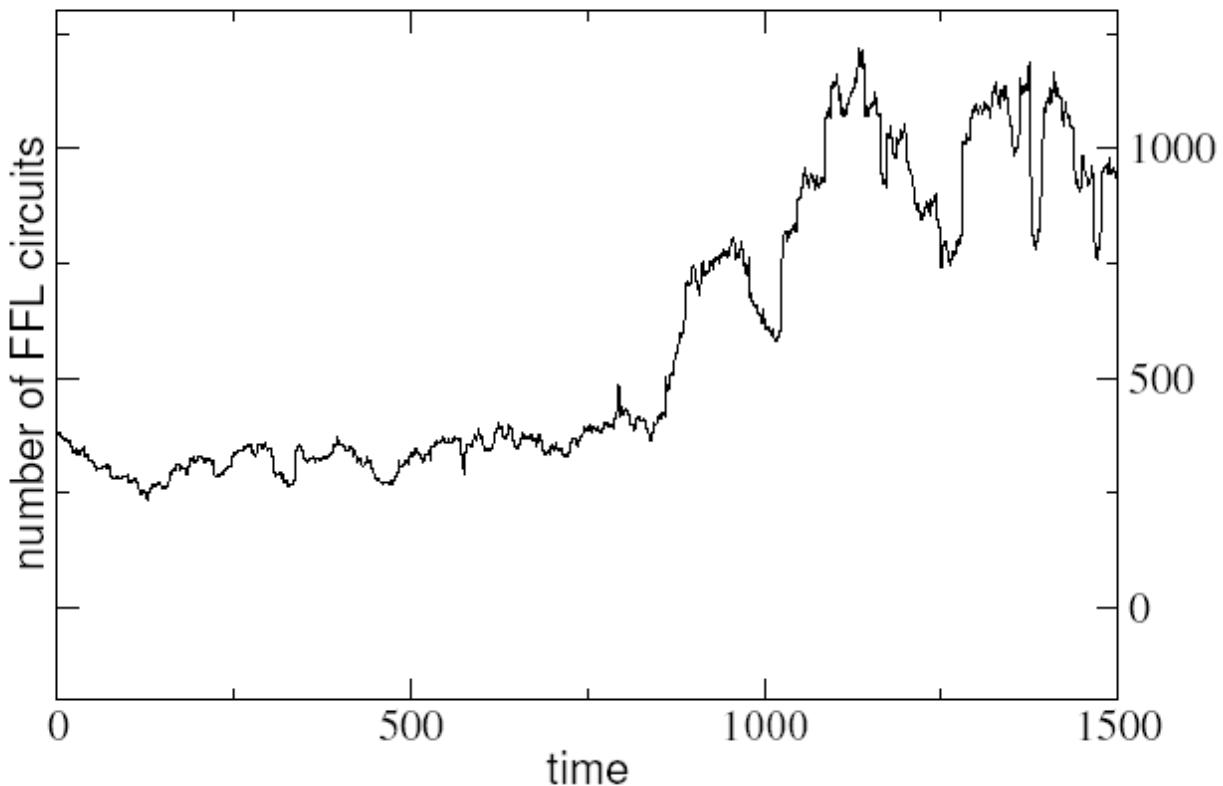
## Hierarchical structure seen in Human regulatory network, ENCODE results (Gerstein, 2012, Nature)



## Sudden increase of FFL motifs: evolutionary (mutational) signature

---

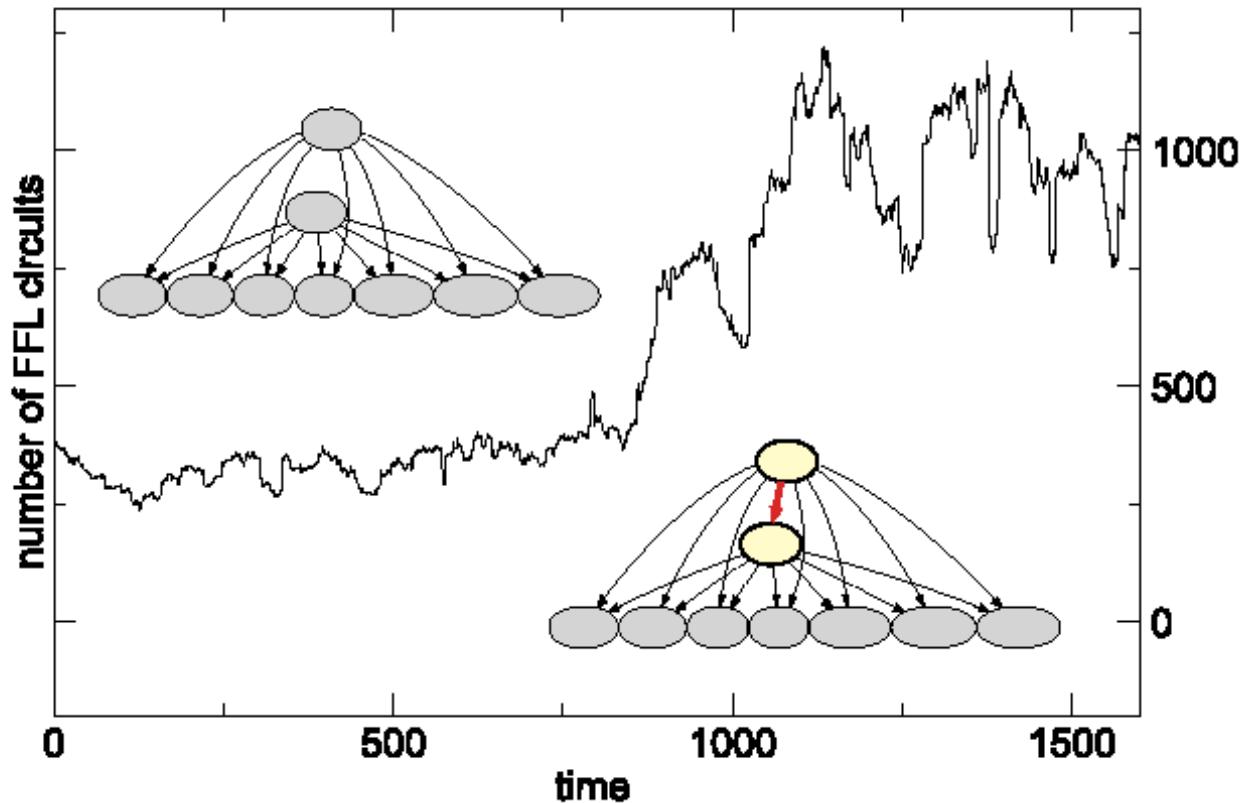
- ▶ Results of mutational dynamics at the microlevel:



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

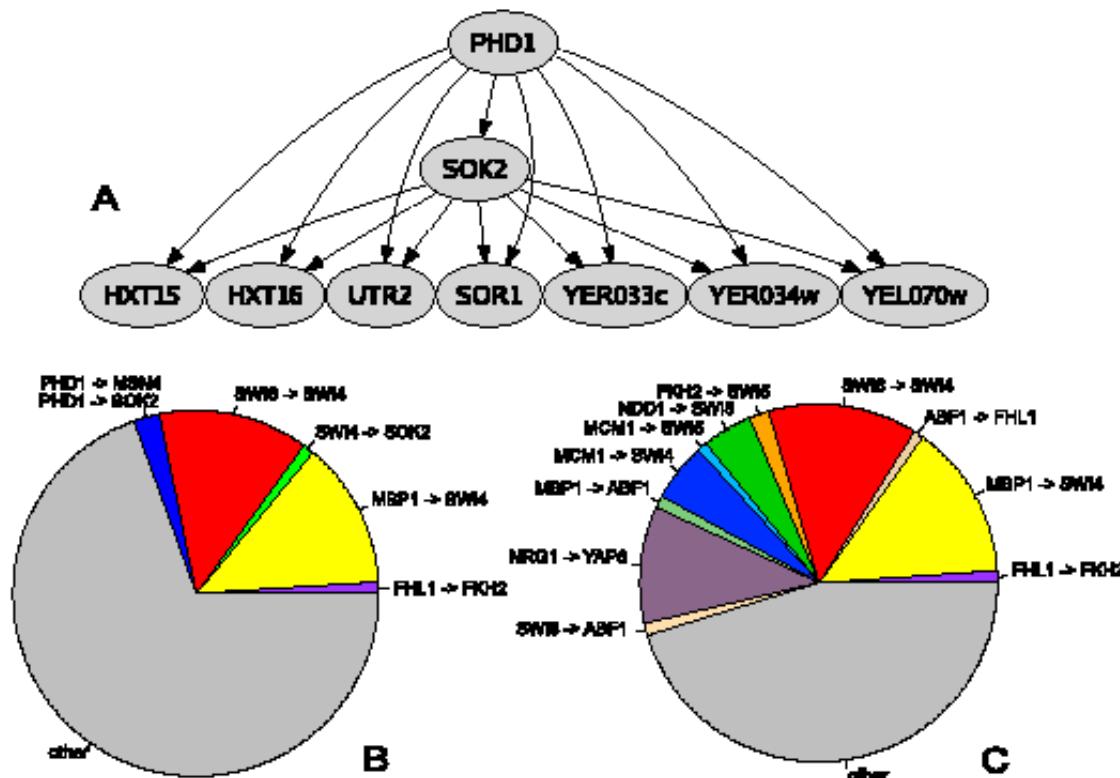
---

► Mechanics of massive FFL formation:



# Overrepresentation of FFL motifs in Yeast: Duplication + connection of Hub genes

- ▶ Evidence in the yeast network:



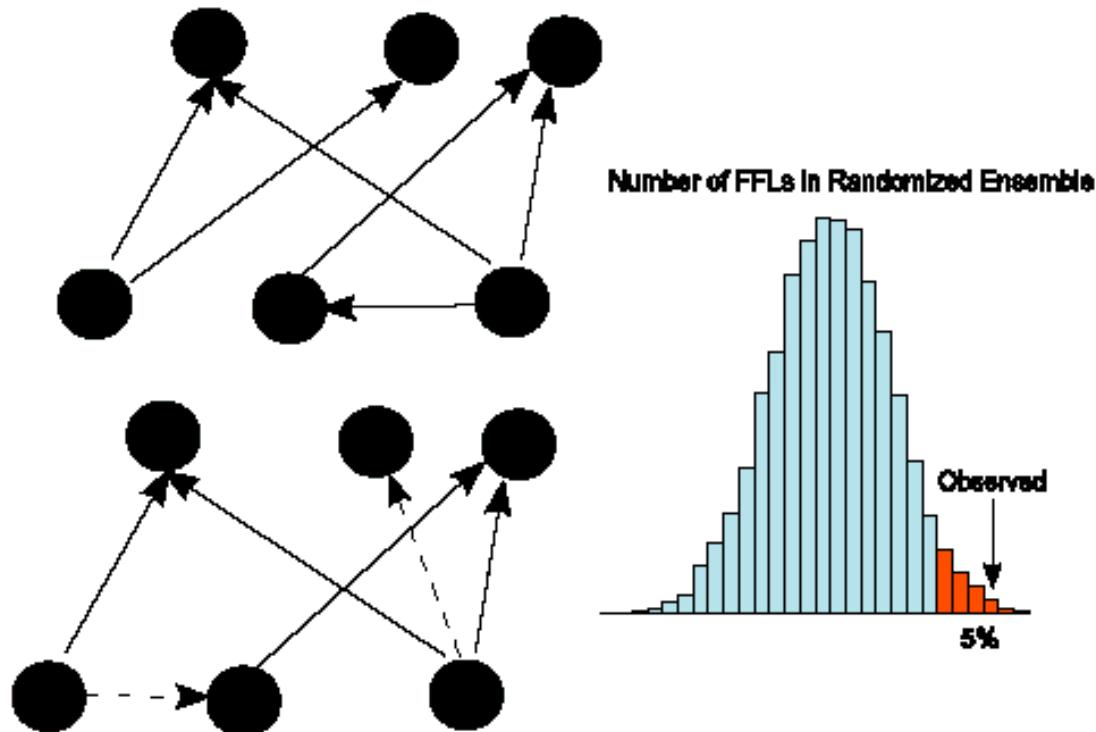
## **random mutations vs randomization**

---

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

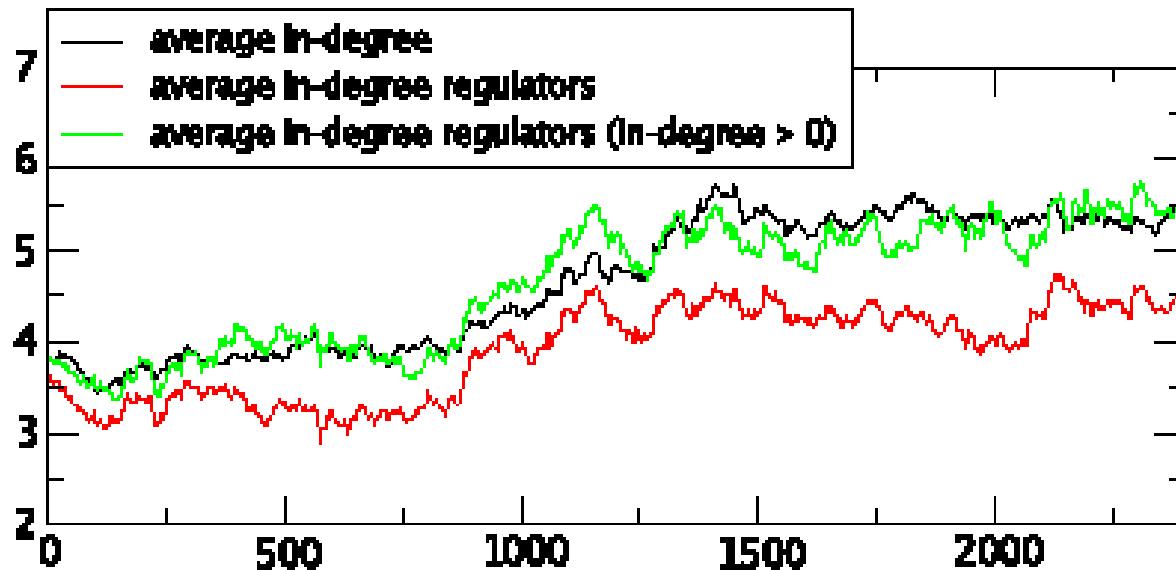
here: keep degree distribution — test for FFL

- ▶ Randomization test: swapping connections.



## random mutations do not conserve degree distribution

---



## **conclusions: mutational vs functional signatures**

---

- Random mutations = / = randomization
- Evolved structure not necessarily selected
- (However, if mutational structure very deleterious, selection will counteract it)

# **Yeast regulatory network evolution**

## **Some “surprising” observations from short term evolution experiments**

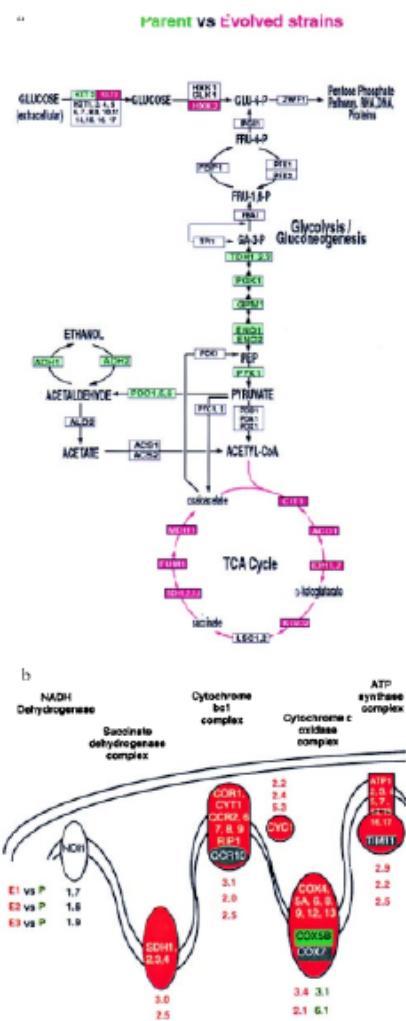
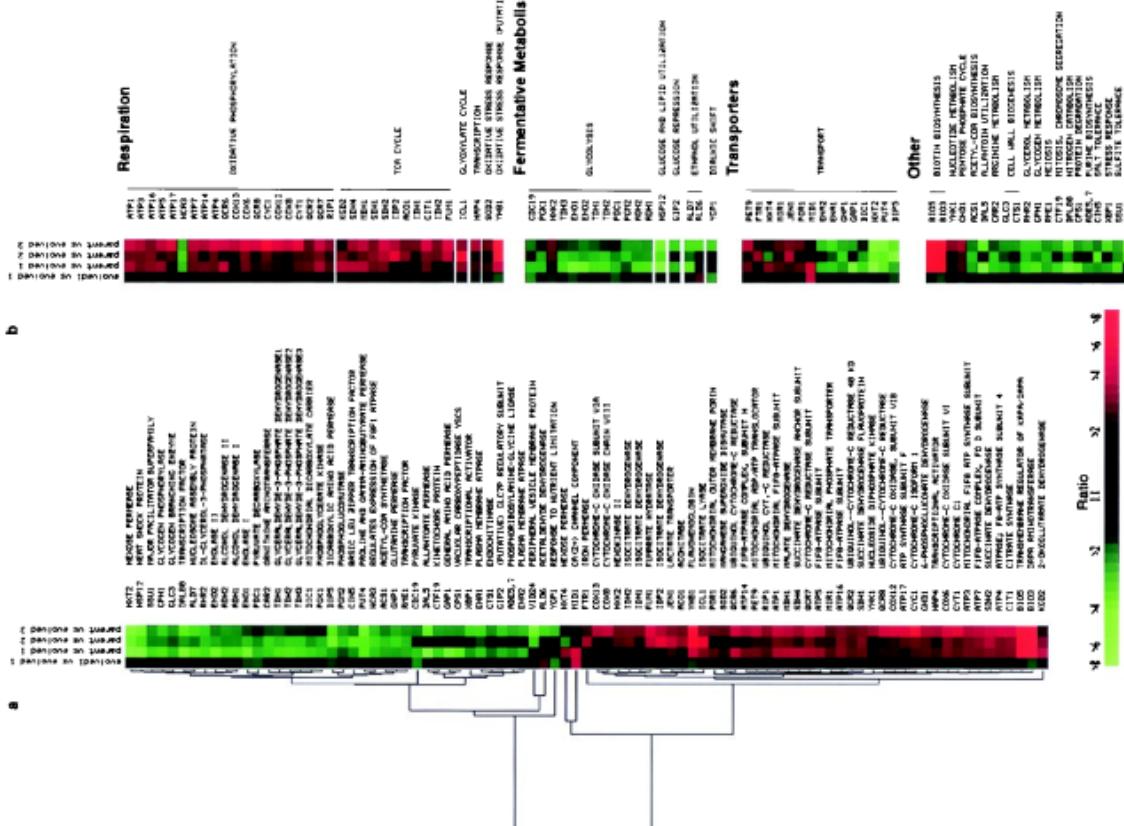
### **( Ferea et al 1999, Dunham et al 2002)**

---

- 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?**

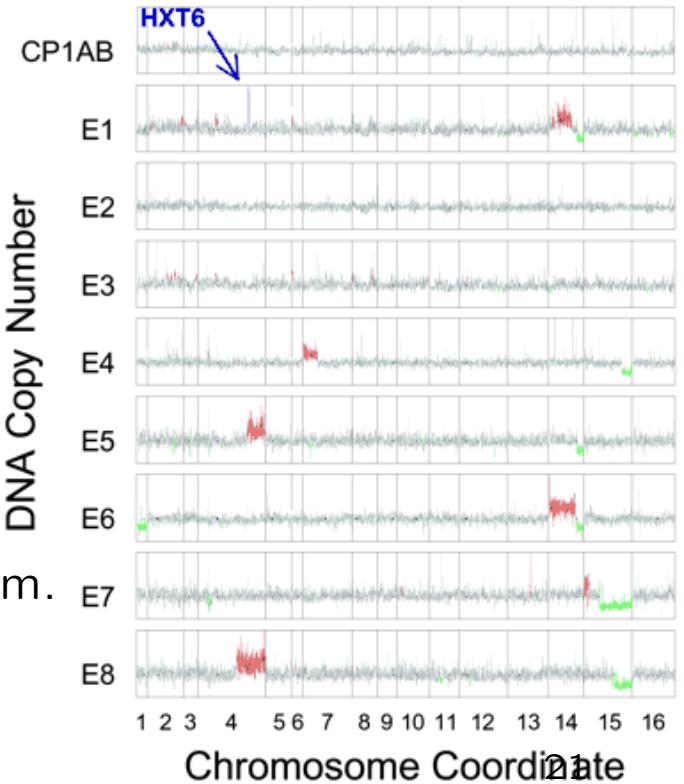
# regulatory and/vs evolutionary 'adaptation' gene expression changes in strains evolved on low glucose media



# “Mutational priming” seen in yeast evolution

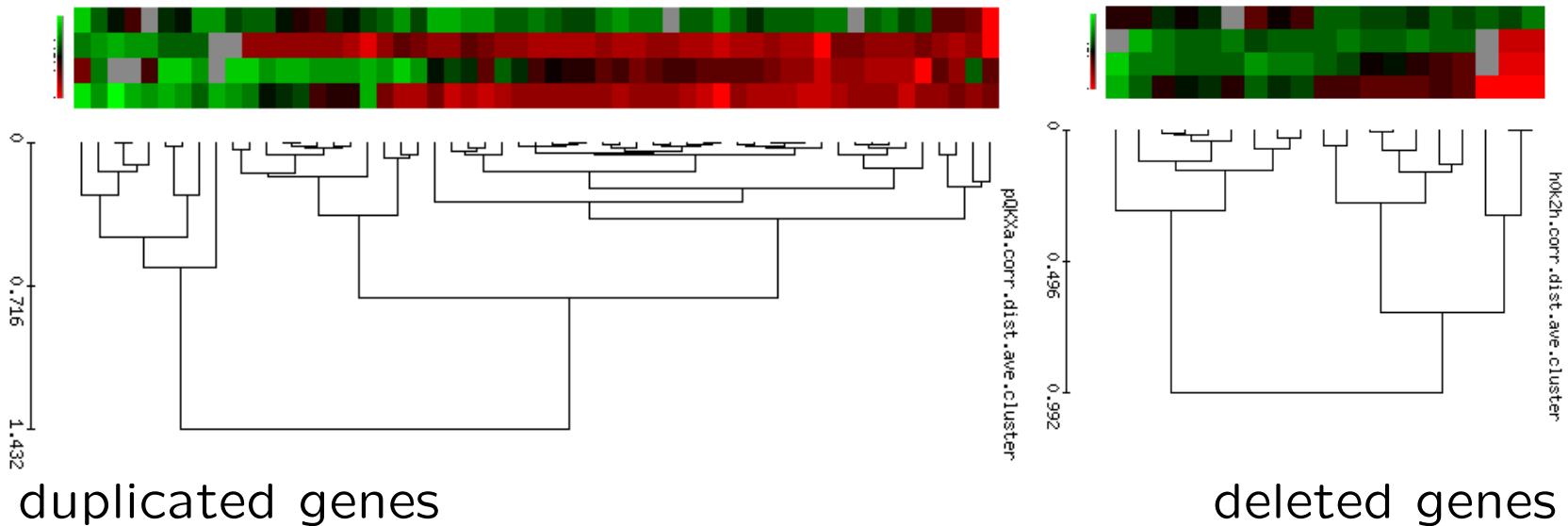
“Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*”  
(Dunham et al PNAS 2002)

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

---



# **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?

# selforganization of genomes by transposon mutational dynamics

## evolution of evolvability

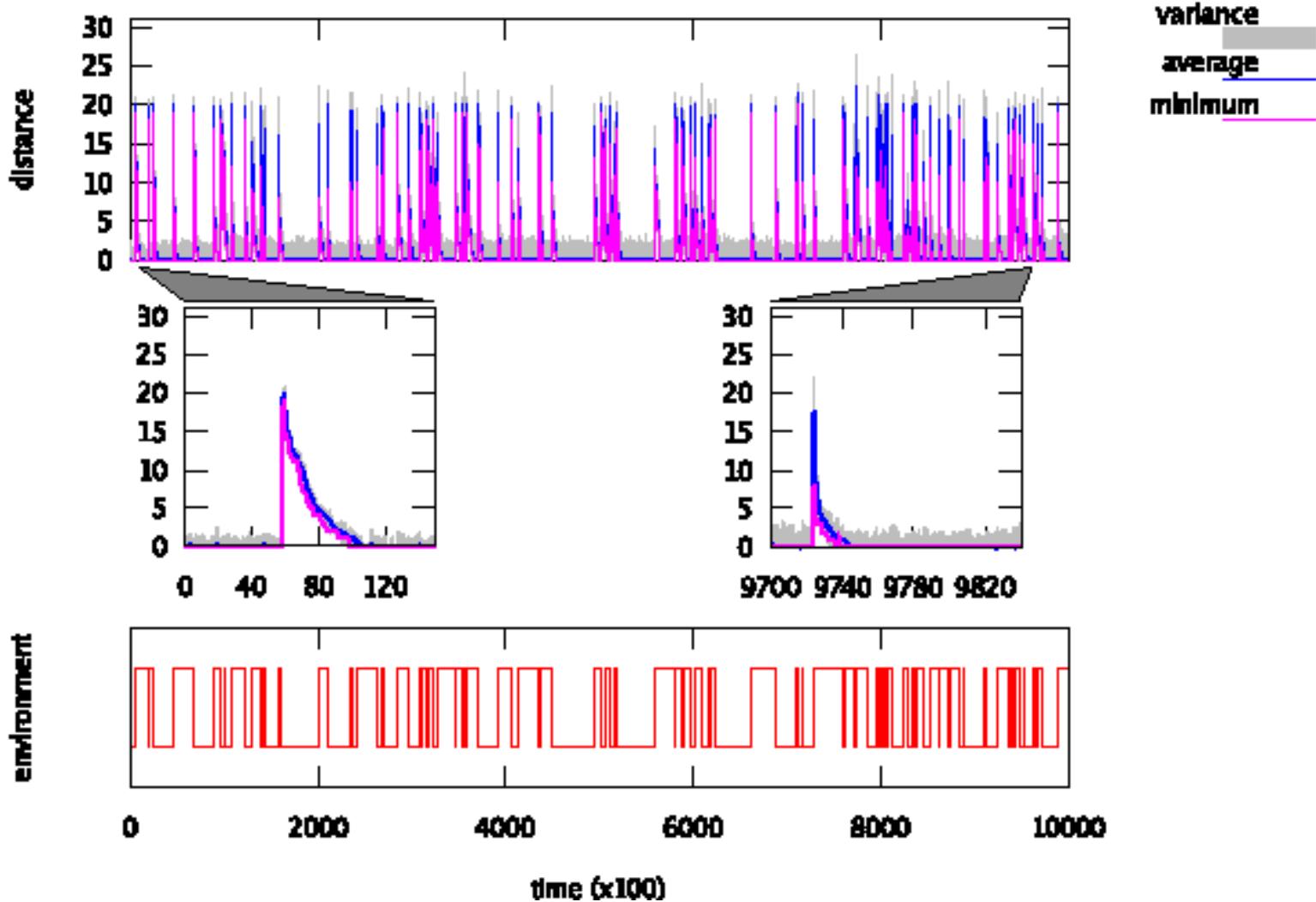
---

### 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



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

---

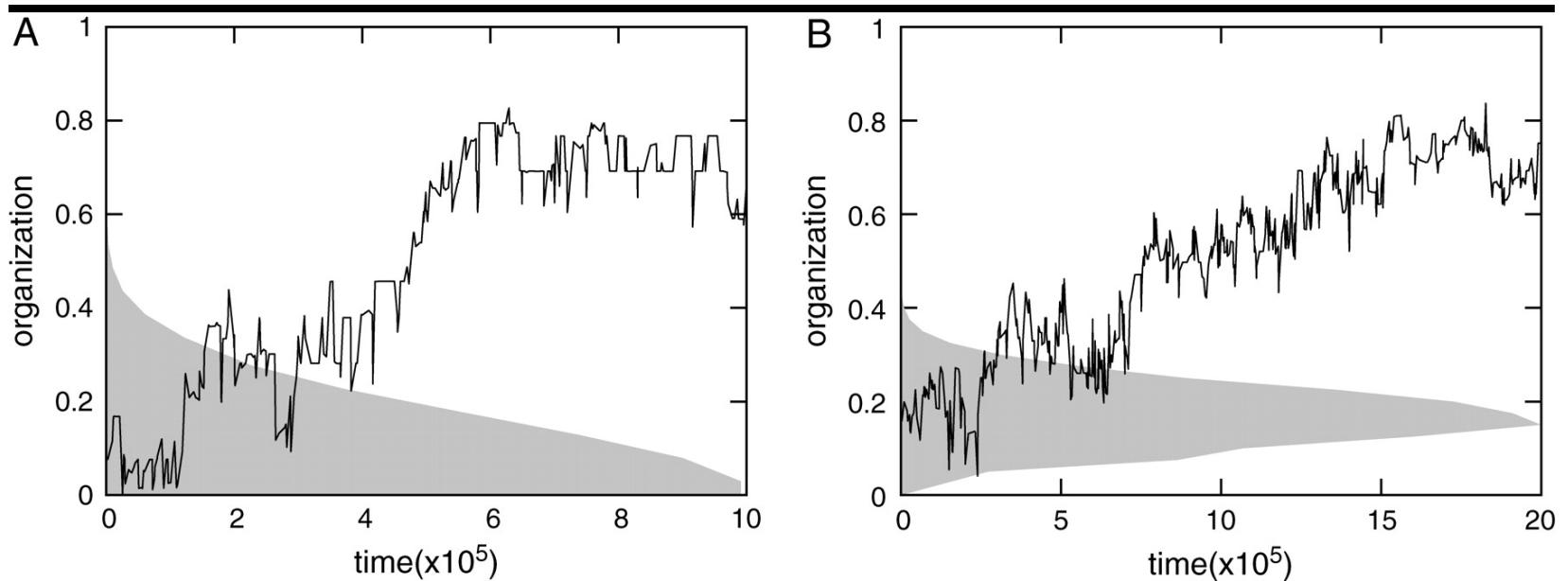
- Randomly generated genome (at start simulation):



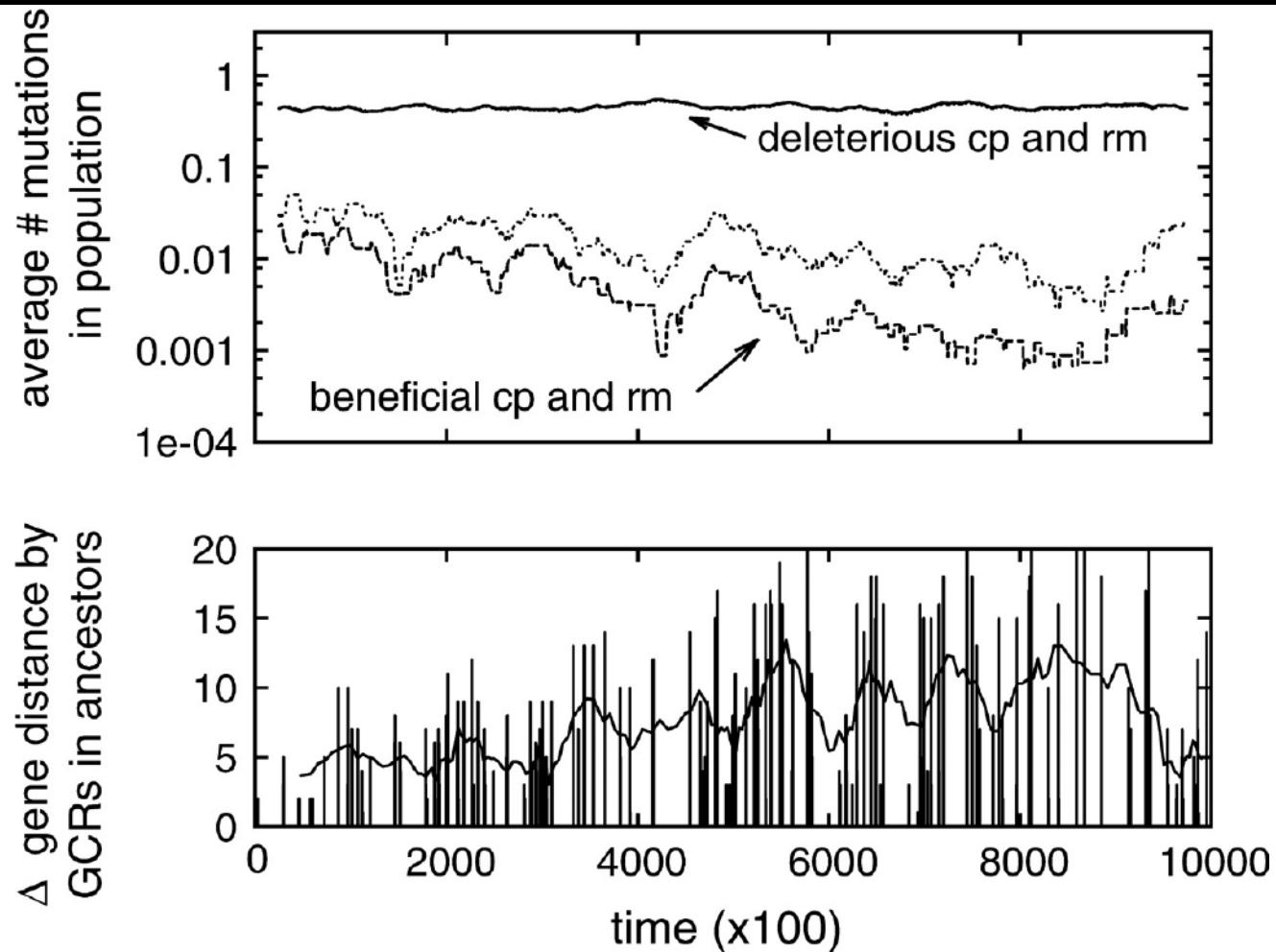
- Evolved genome, grouping of genes as a side effect:



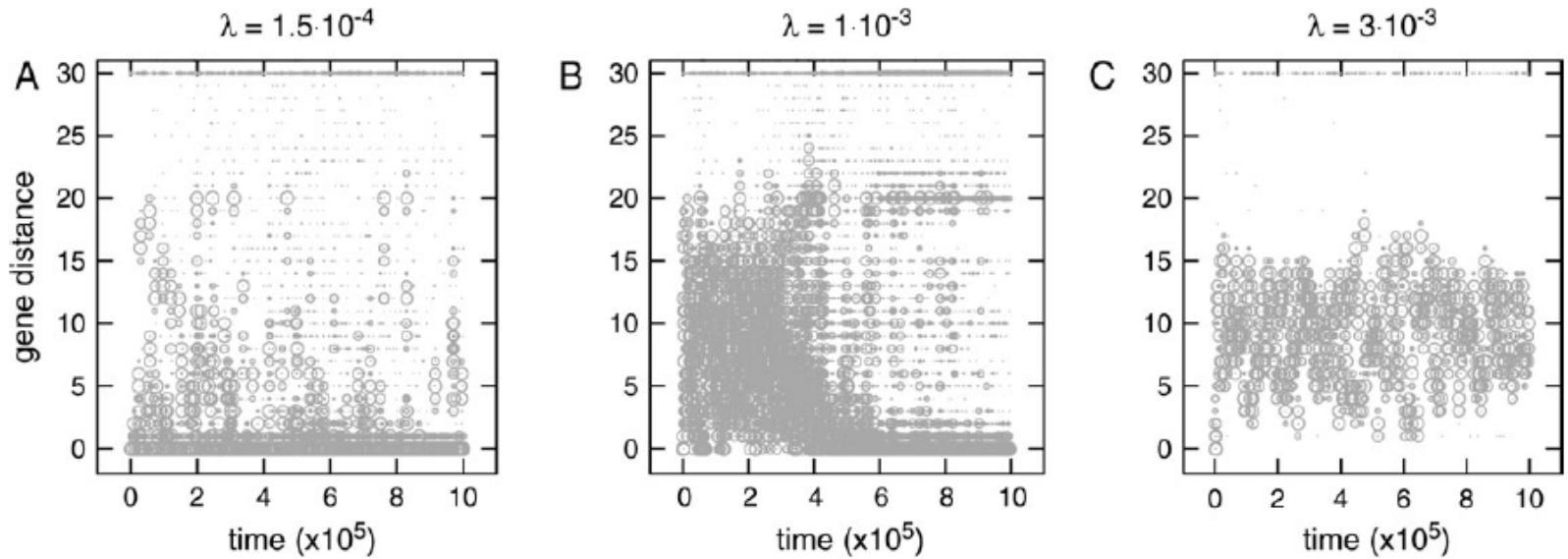
## genome organization over time



## 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**  
**Crombach and Hogeweg PLOS Comp Biol 2008**

---

## network dynamics and fitness

---

Network update:

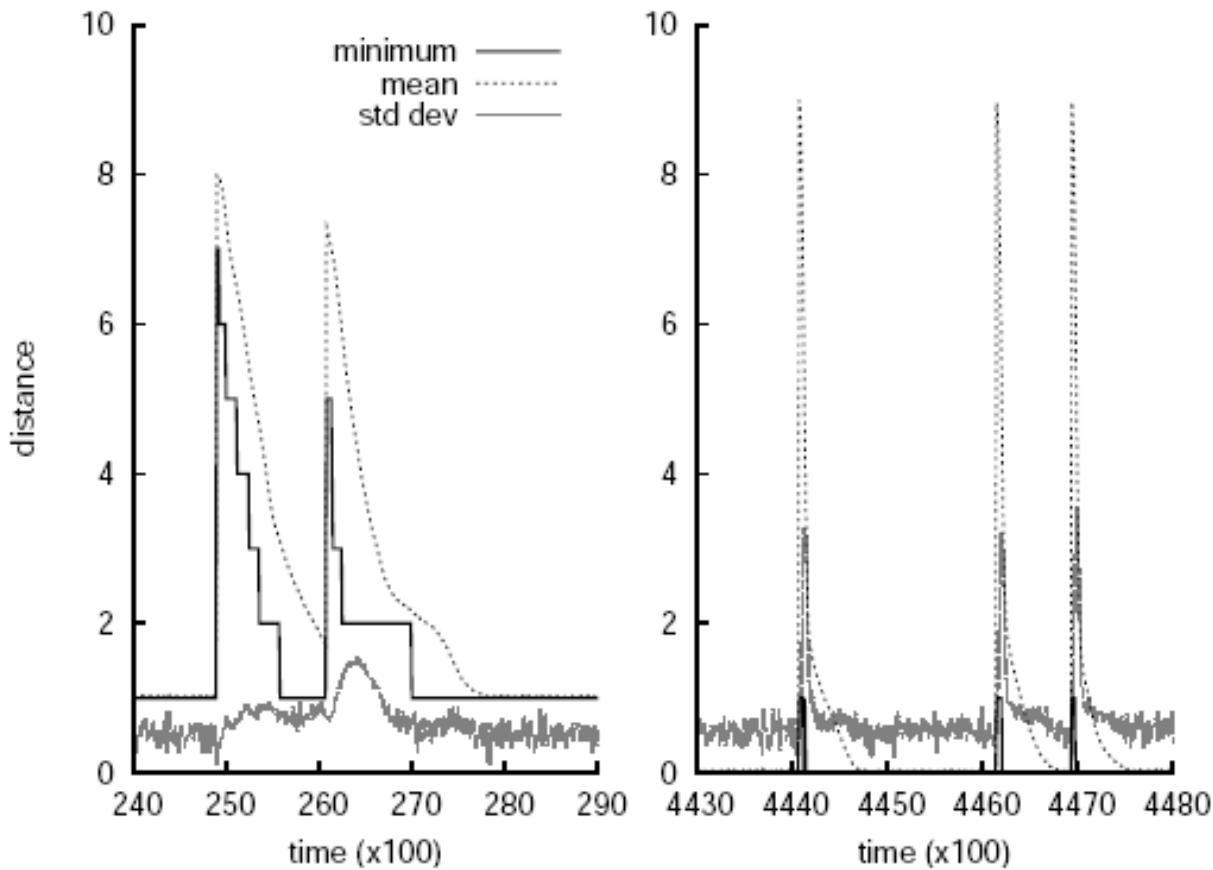
$$s_i^{t+1} = \begin{cases} 0 & \text{if } \sum_j w_{ij} s_j^t < \theta_i \\ s_i^t & \text{if } \sum_j w_{ij} s_j^t = \theta_i \\ 1 & \text{if } \sum_j w_{ij} s_j^t > \theta_i \end{cases}$$

fitness: distance to target

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

## improved evolvability observed

---

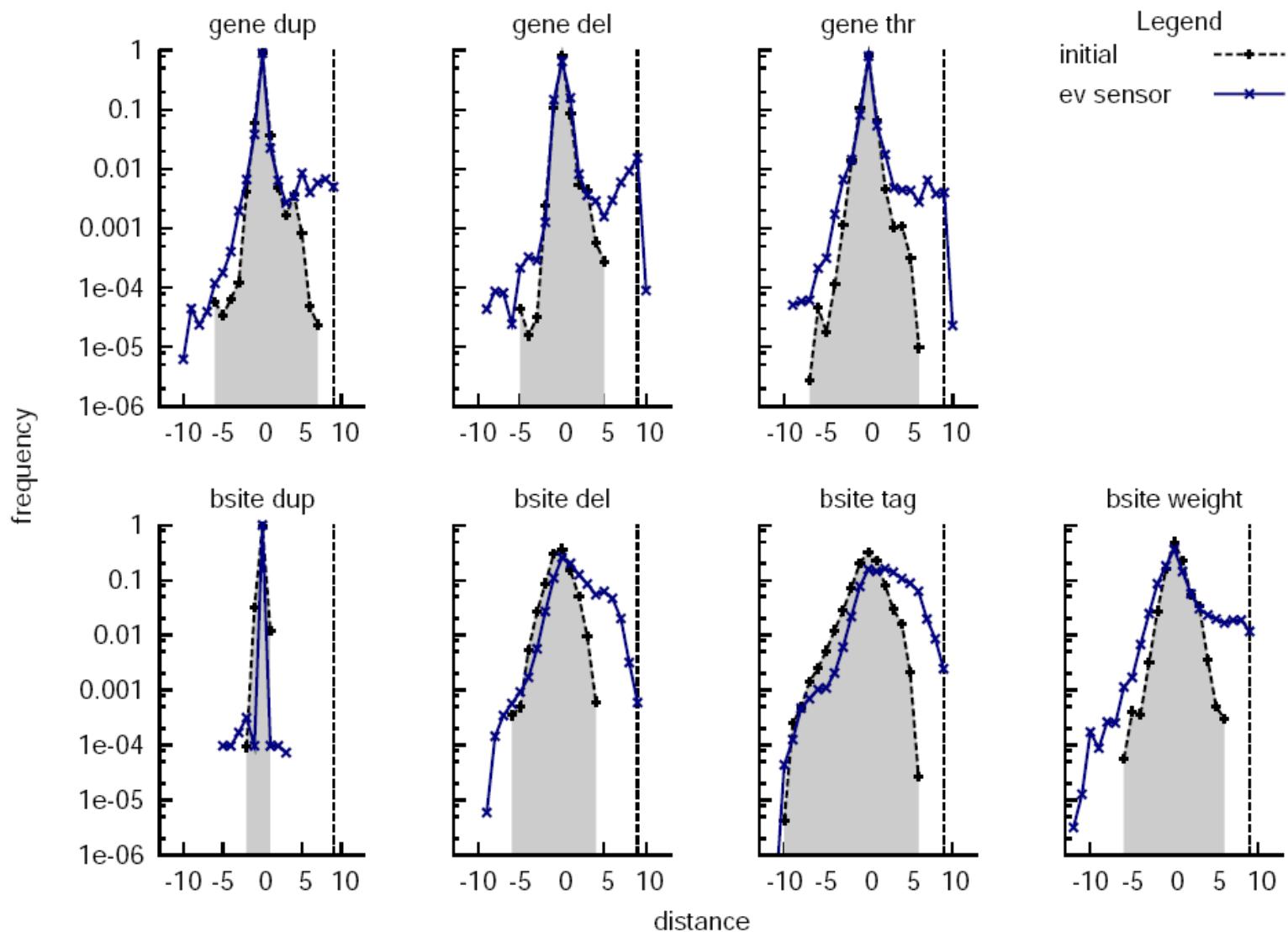


## **Hamming distance improvement to opposite target**

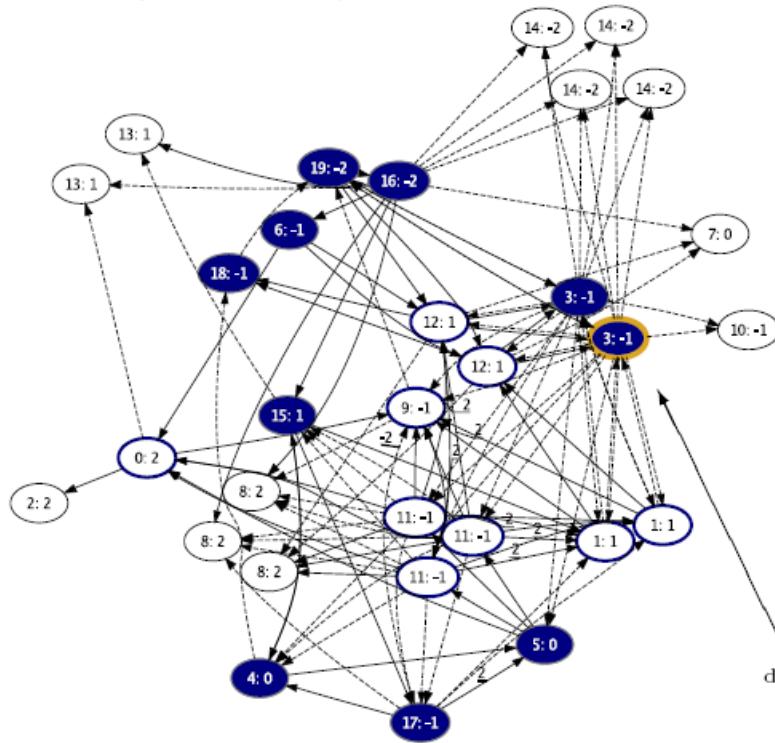
---

### **Regulatory Mutational Priming:**

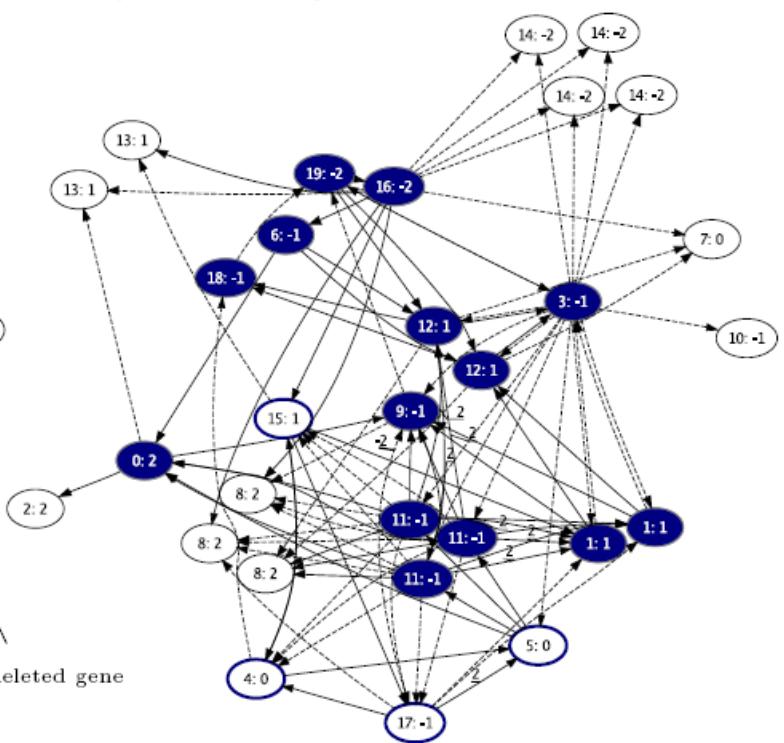
Many different mutations lead to “beneficial” adaptation

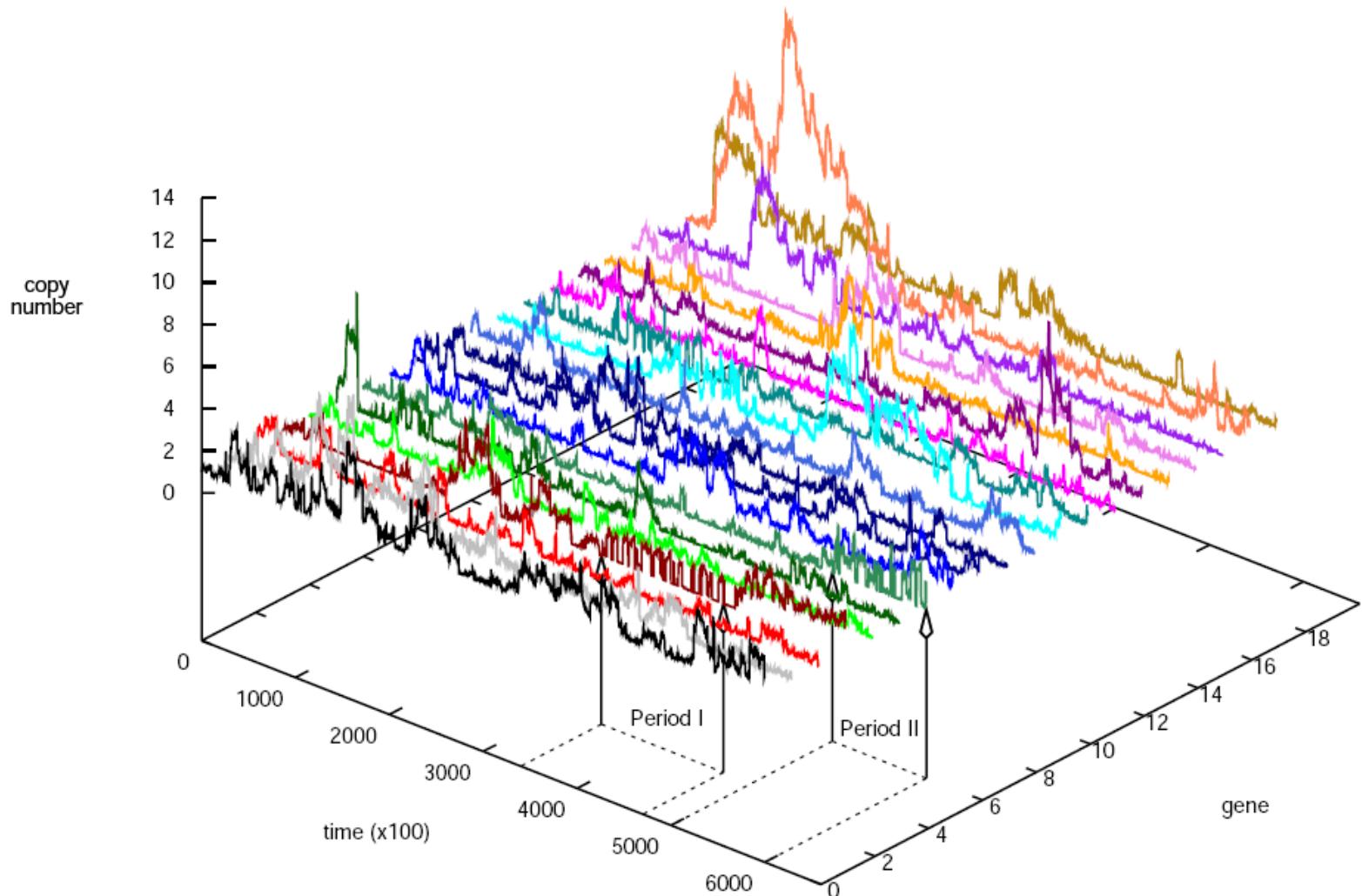


Parent ( $t = 457\ 755$ ):



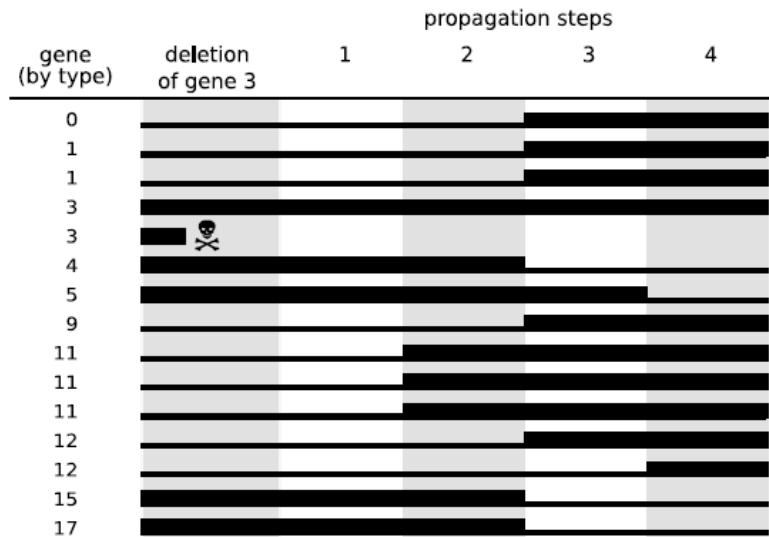
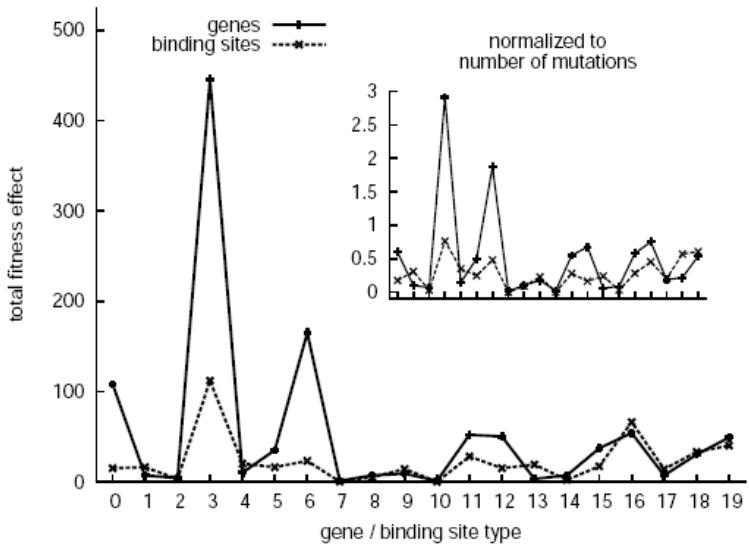
Child ( $t = 457\ 758$ ):





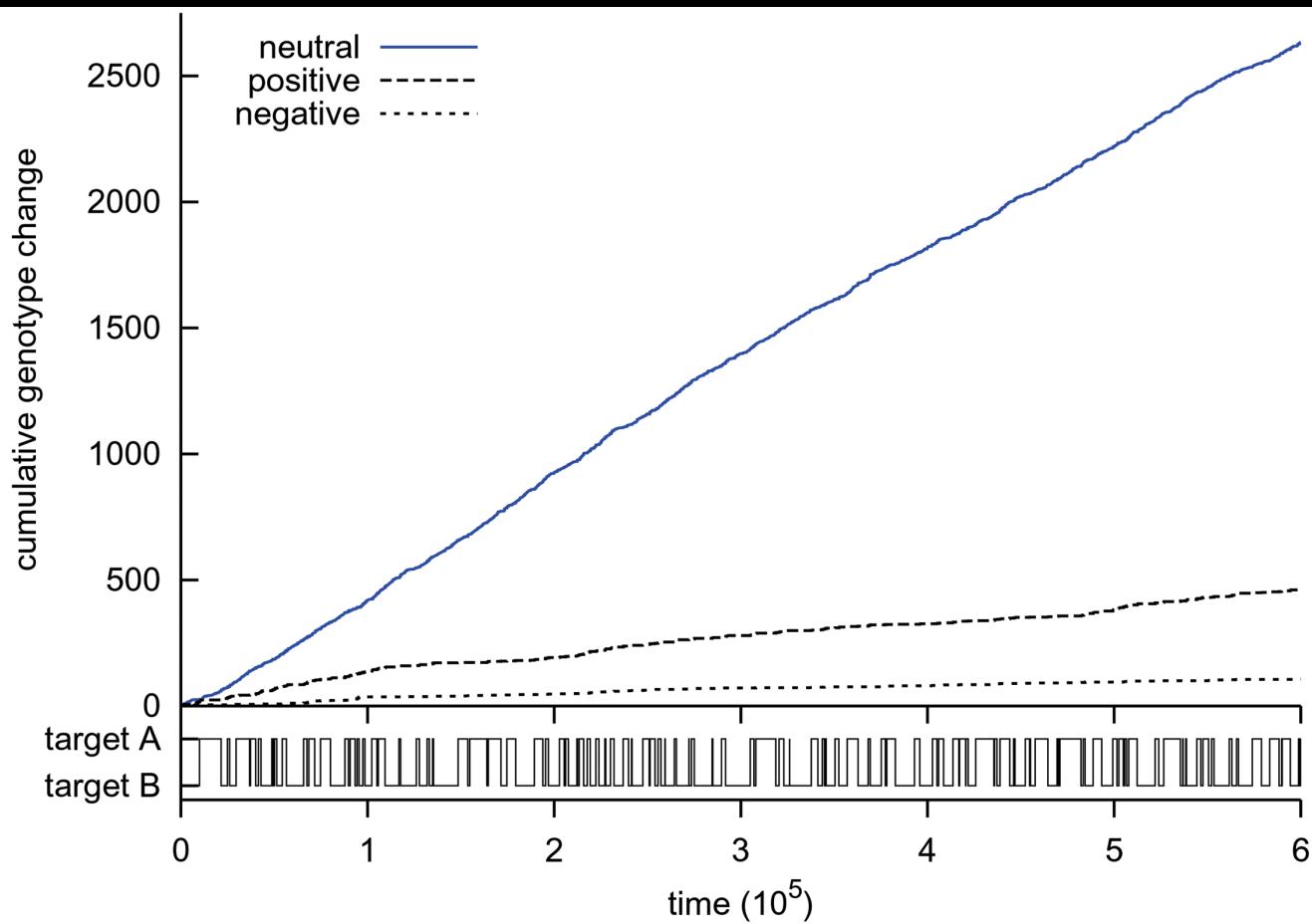
**“evolutionary sensor”**

# Duplication/Deletion of single gene(s) and regulatory effect



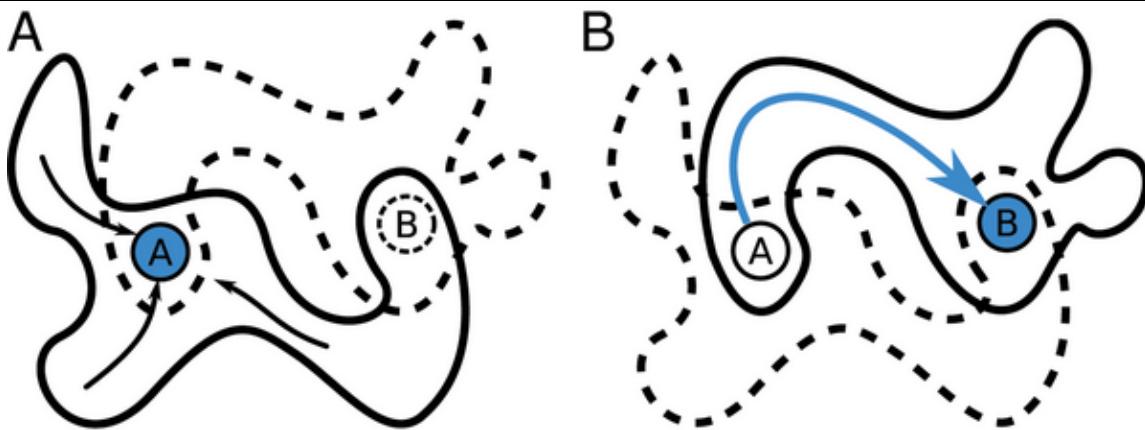
# Neutral drift far greater than adaptive change!

---



## 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*

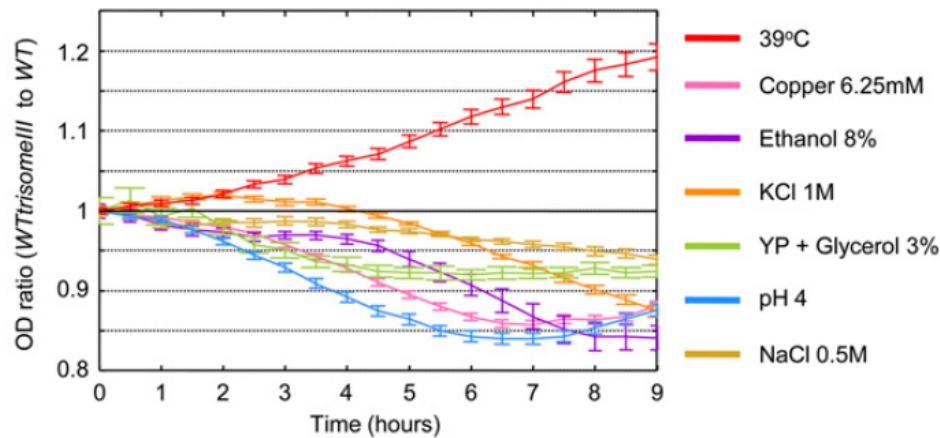
- Transposon dynamics structures genomes creating hotspots for mutations and genome ordering. Long term evolution leads to genome structures such that short term evolutionn is facilitated
- Genotype to phenotype mapping through gene regulatory networks evolves such that (advantageous) attractor switching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

*Both these mechanisms appear to occur in Yeast*

# Yeast adaptation to high temperature: short term vs long term effects

---

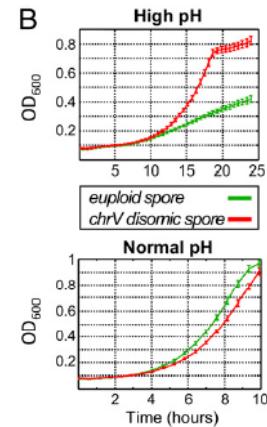
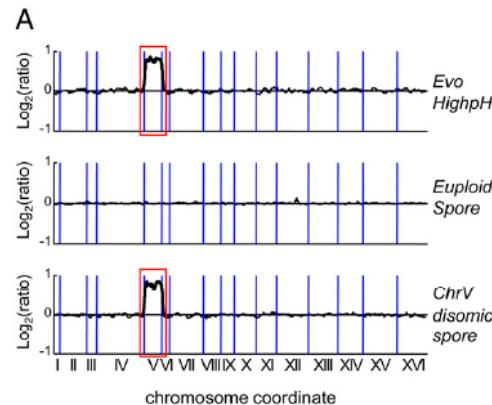
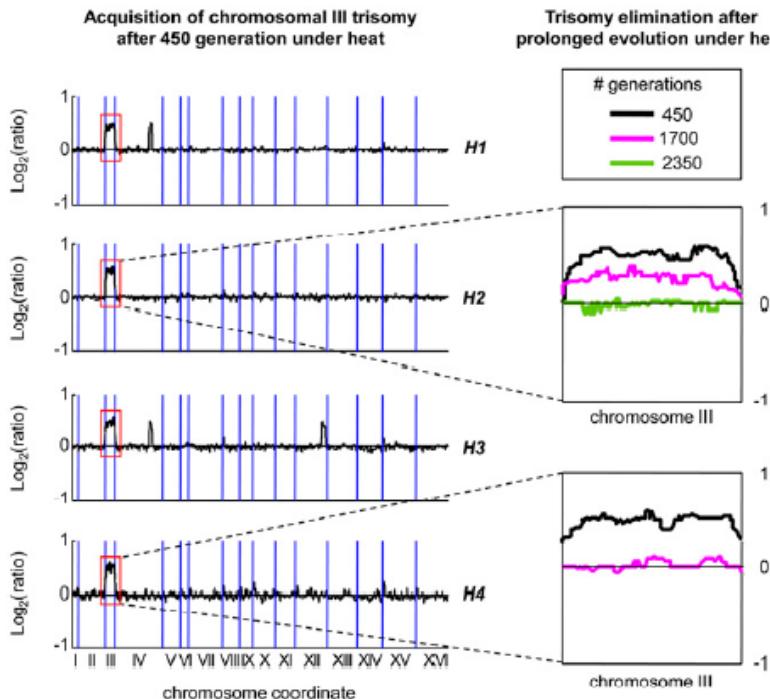
Diploid yeast adapted to 'normal' temperture of 30 C placed in 39 C. After 450 hr:  
Increased growthrate (specific for temperature)



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)

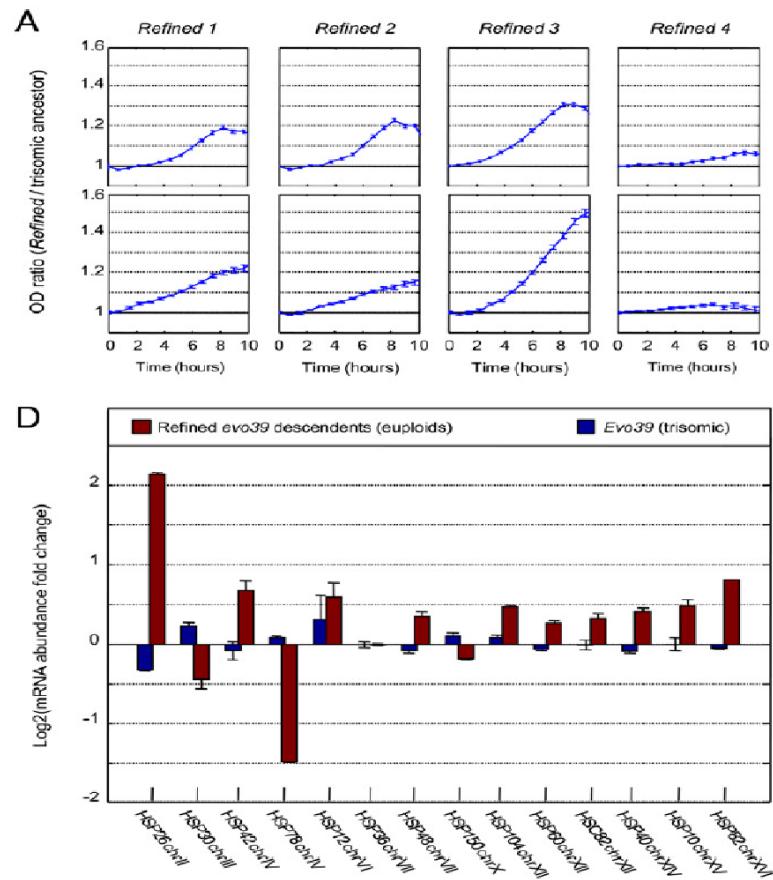


trisomy not retained

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

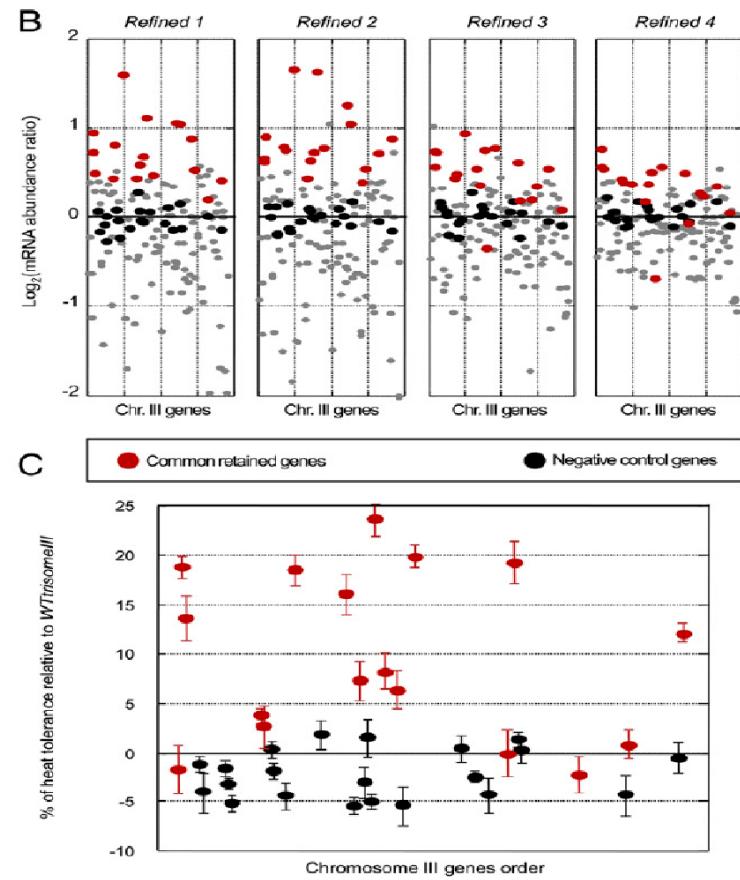
**chromosomal duplication is followed by regulatory adaptation**

continued increase growthrate



overextreeion HSP only later

overexpression retained of some genes



retained genes increase growthrate in ancestor