Coping with a variable environment: evolvability and/vs regulation

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

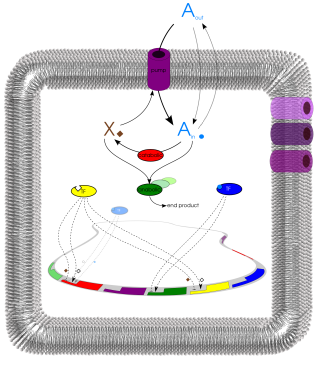
last time

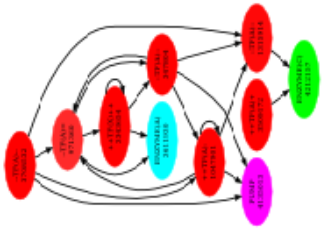
Simple Models of genome structure and GRN networks:
emergent properties (also seen in present day organisms)
- powelaw degree distribution of GRN
- "overrepresentation" of FFL by neutral evolution
emergent properties of short term evolution:
- Evolution of fast adaptation to environmental change
- through genome structuring and GP map structuring
- *"evolutiion of evolvability";" non-random/random mutations"*

- Experimental demonstration of these features

TODAY

Add metabolism ("virtual cell/ virtual microbe")
emergent properties of long term evolution (phylogeny)
++



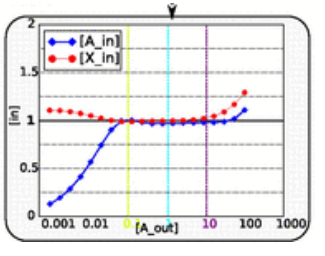


$$\left(\frac{d[A]}{dt} = \frac{[A]_{out}[X]Vmax_p[Prot]_p}{([A]_{out} + KA_p)([X] + KX_p)}\right) \quad (2)$$

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

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

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

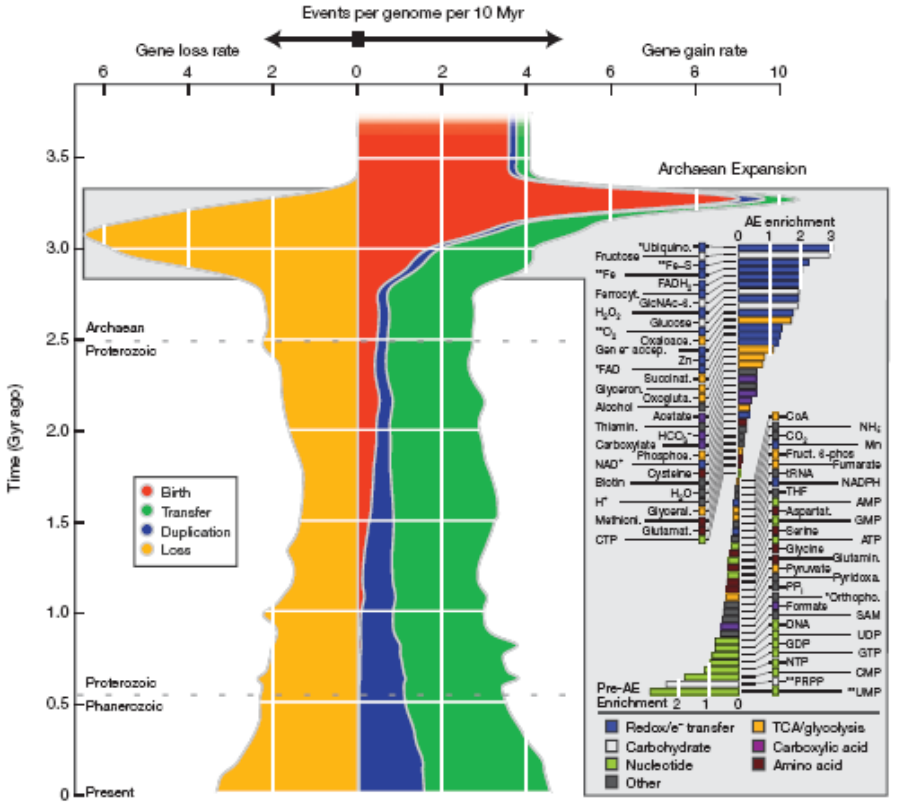


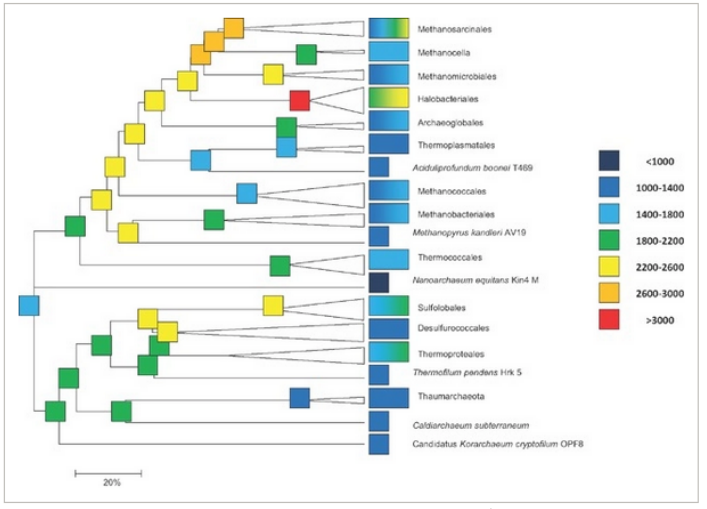
Some "surprising" (and debated) observations on the dynamics of evolution (of complexity) gleaned from phylogenetic analysis

- **Early complexity**
  - biological *Big Bangs*: major transitions in evolution
  - from phylogenies: large common ancestors
  - from phylogenies: closely packed early species radiations
  - important role of gene *LOSS* in adaptation
  - FECA to LECA: many gene duplications before species radiation
  - genes with "late" function often predate that function

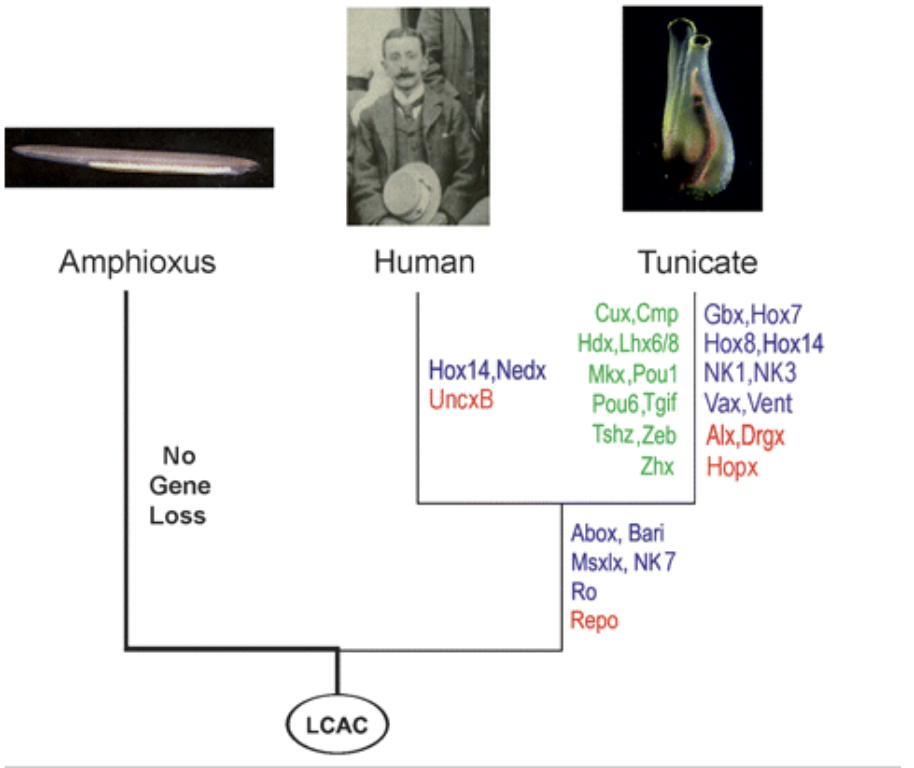
- **Whole Genome duplication rare but important**
  - occurs often (especially plants) but rarely fixed
  - at root of major radiation
  - during major environmental shifts (?)

early gene innovation - and loss
Alm Nature2010





Gene loss as major evolutionary process



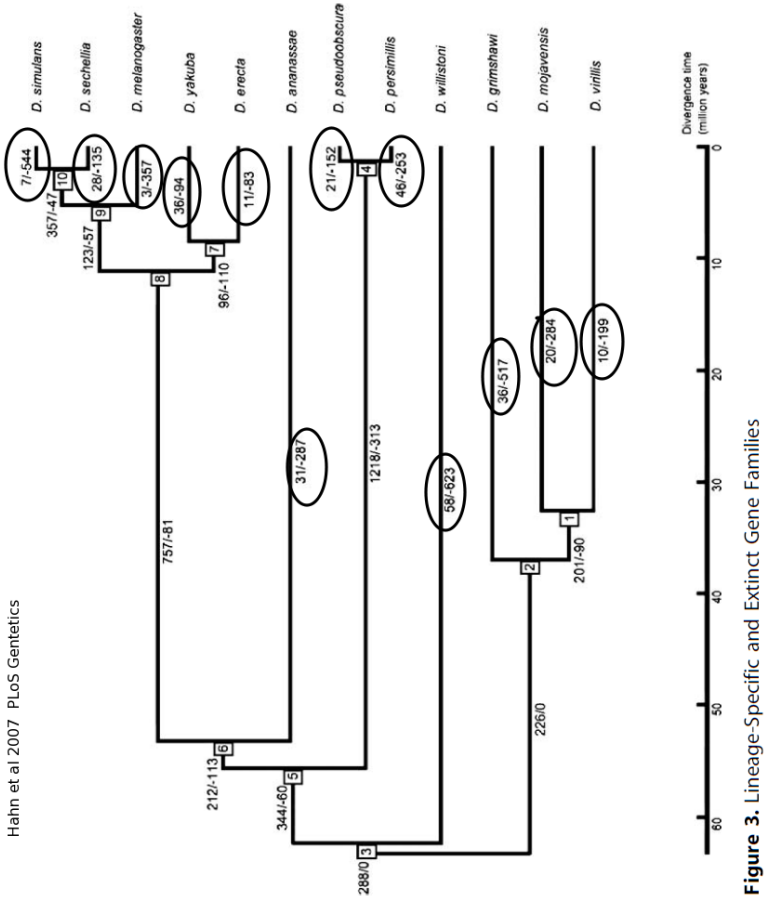
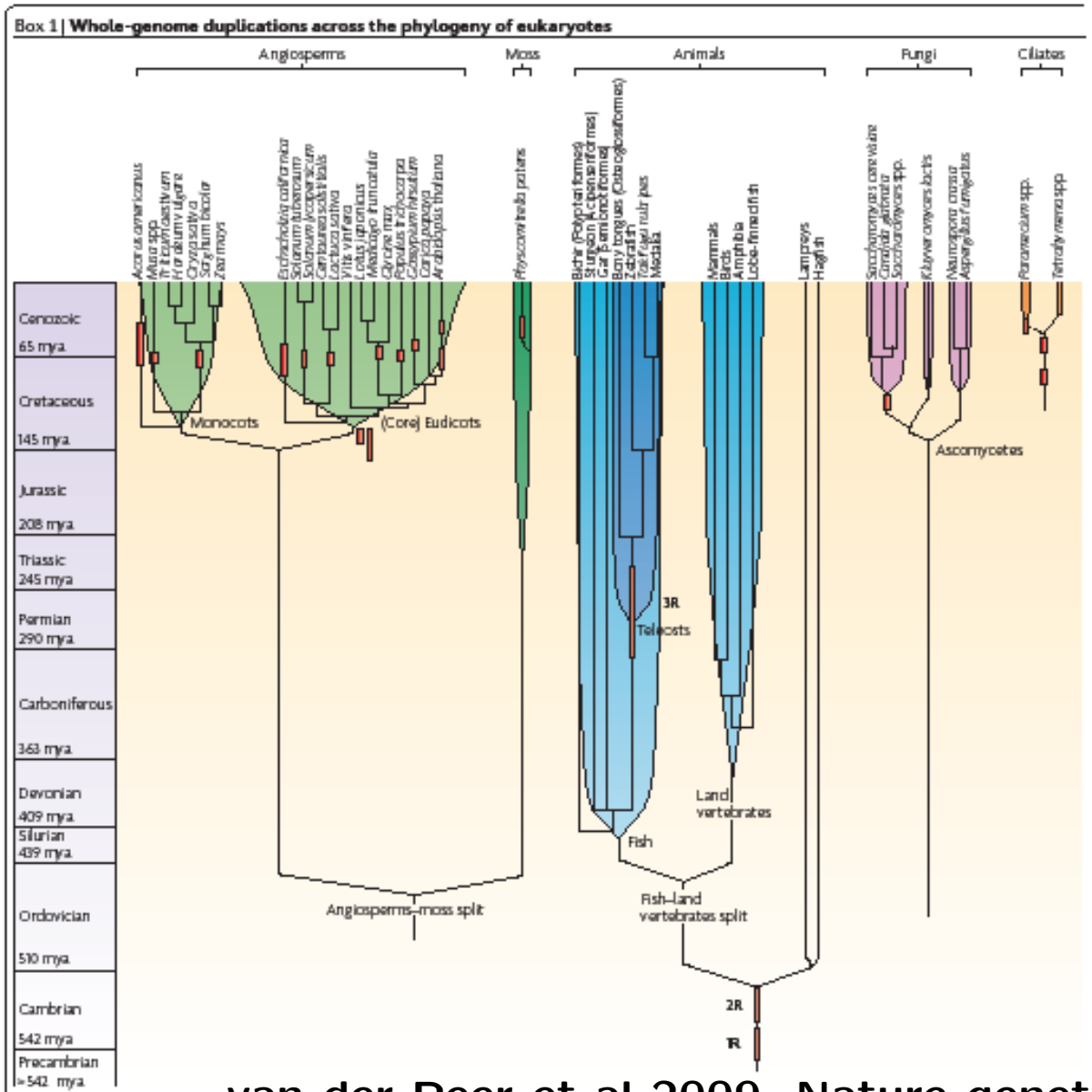


Figure 3. Lineage-Specific and Extinct Gene Families

Metazoa
Loss of homeoboxgenes

Drosophila species
gain/loss of genes



van der Peer et al 2009, Nature genetic reviews

Evolution in virtual cells: genome. GRN, metabolism

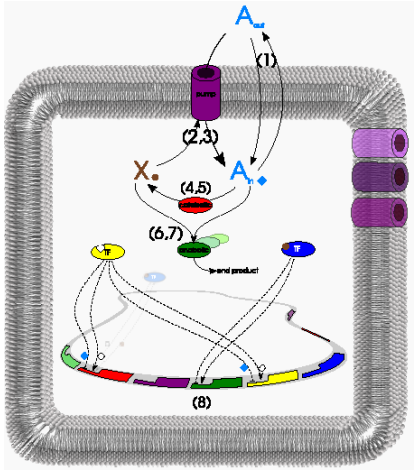
based on "plausible" *minimal* multilevel 'cell'
mutations segmental duplications/ deletions, pointmutations
fitness: *homestasis* (evolves regulatory adaptation)
evolving in varying environment

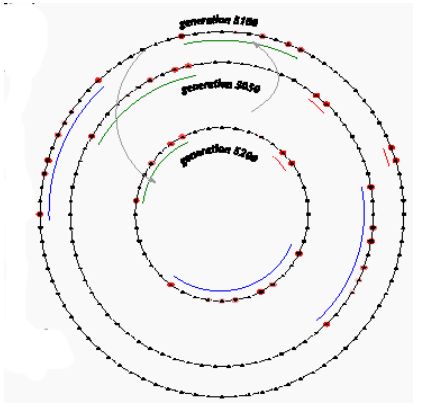
Questions

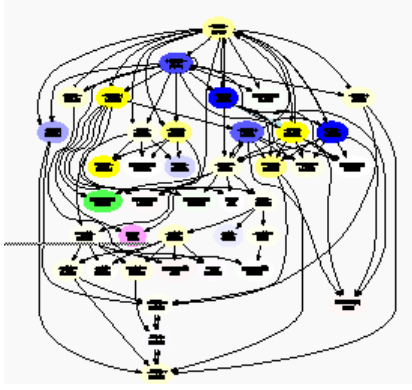
Are some of the features seen in phylogenetic analysis observable in evolution of such cells?
Early complexity, dominance of gene loss
whole genome duplication at "roots" of lineages
mutational/selectional enforced conservation

Cuypers & Hogeweg 2012, 2014, Cuypers, Rutten ő2017

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







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

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

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

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

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

cell genome(dupdel/gcr) GRN metabolism

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

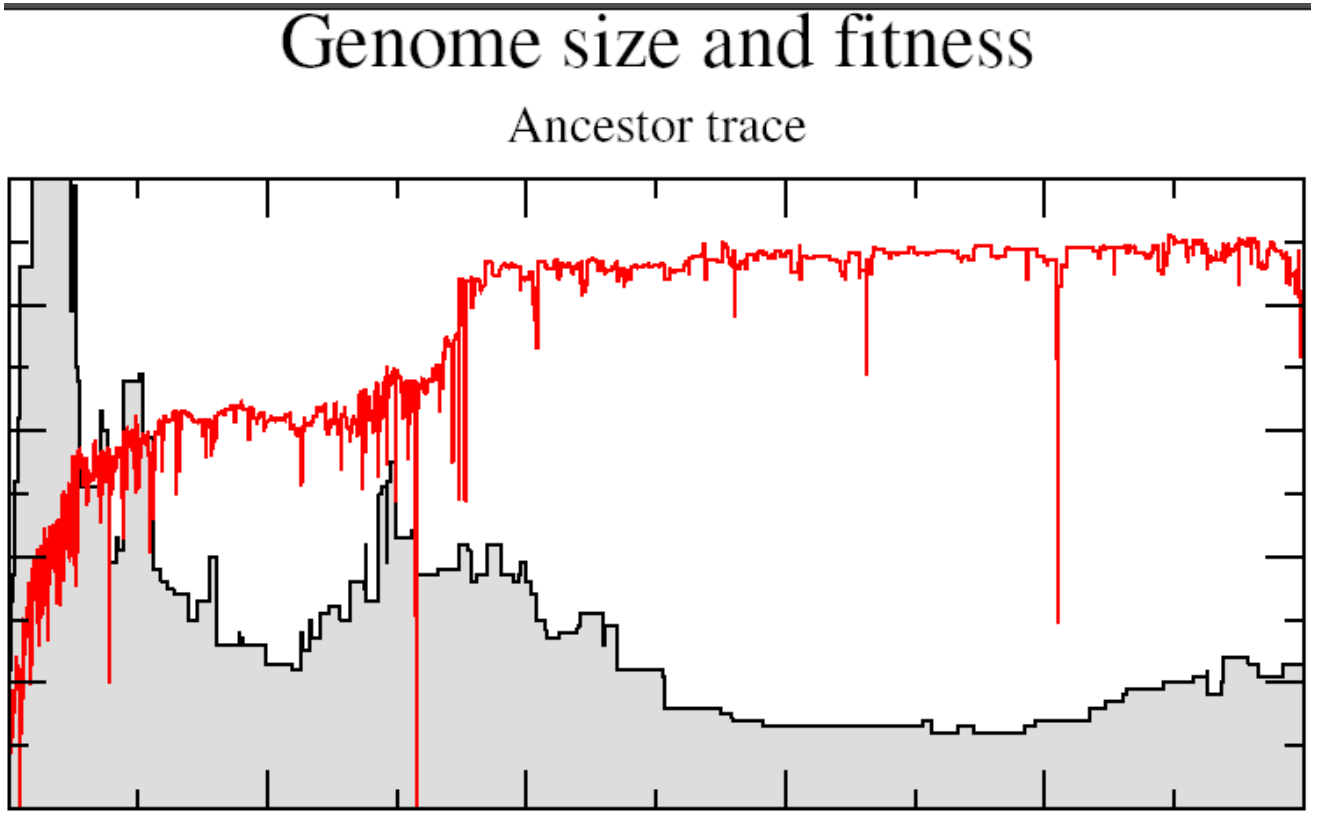
ecology and evolution of virtual cells

- **Environmental fluctuation of resource A**
  $[A_{out}]$ varies 4 orders of magnitude
  Cell 'sees' 1-3 randomly chosen concentration in lifetime
- **Fitness**: homeostasis
  distance to set value, average over lifetime
- **Population of cells** compete
  Replication probability proportional to fitness
- **Mutations** upon replication
  INDELS, LCR, values of parameters (Vmax, binding etc)

Analysis along ancestral lineage
evaluated in 3 standard environments

Note Differences with previous models
not on-off genes; fitness not expressed as gene expression but
as effect of gene expression, reacts on environment, allows regulatory adaptation

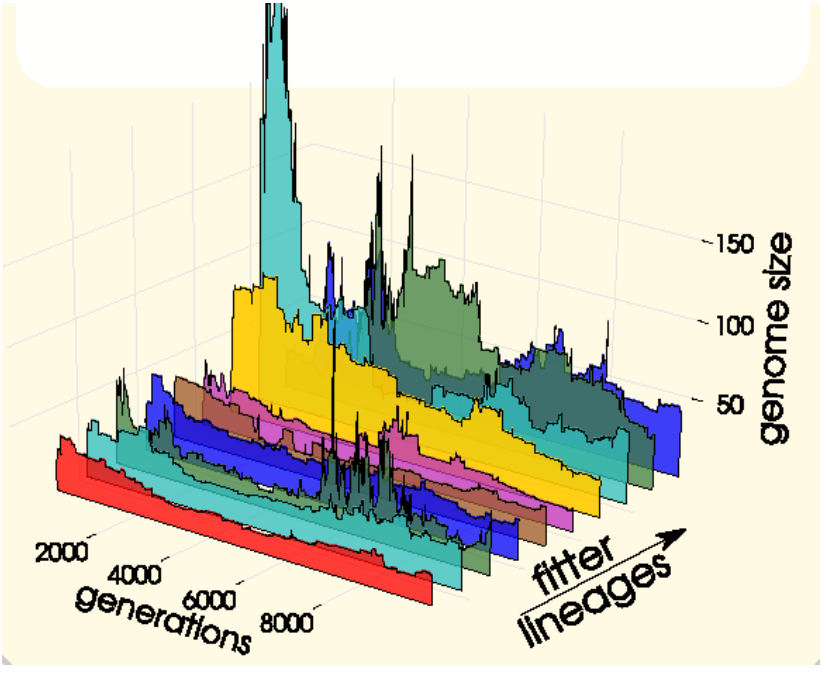
Typical evolutionary dynamics: Genome inflation(s) - followed by fitness increase - followed by stream lining - followed by genome size fluctuations



early genome inflation a "generic" pattern? Yes... in the sense that:

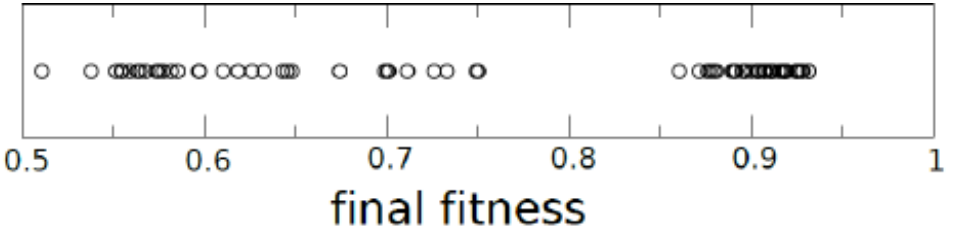
It occurs most pronounced in those runs which achieve high fitness eventually

It occurs most pronounced with mutational parameters which achieve often high fitness



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

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
(+)	+	$.95 - 1.05$	=	+	$.95 - 1.05$
-	-	$< .95$	=	-	$< .95$

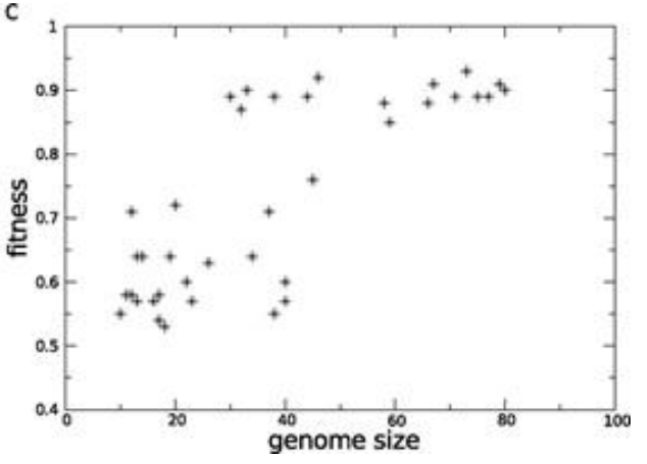
Genome Size			Fitness		
t=1-100	t=101-200		t=1-100	t=101-200	
+	+		=	=	

Why initial inflation?

Duplications more often advantageous than deletions

+ **hitchhiking** of other genes (which might **later** become functional)

higher degrees of freedom increases adaptability

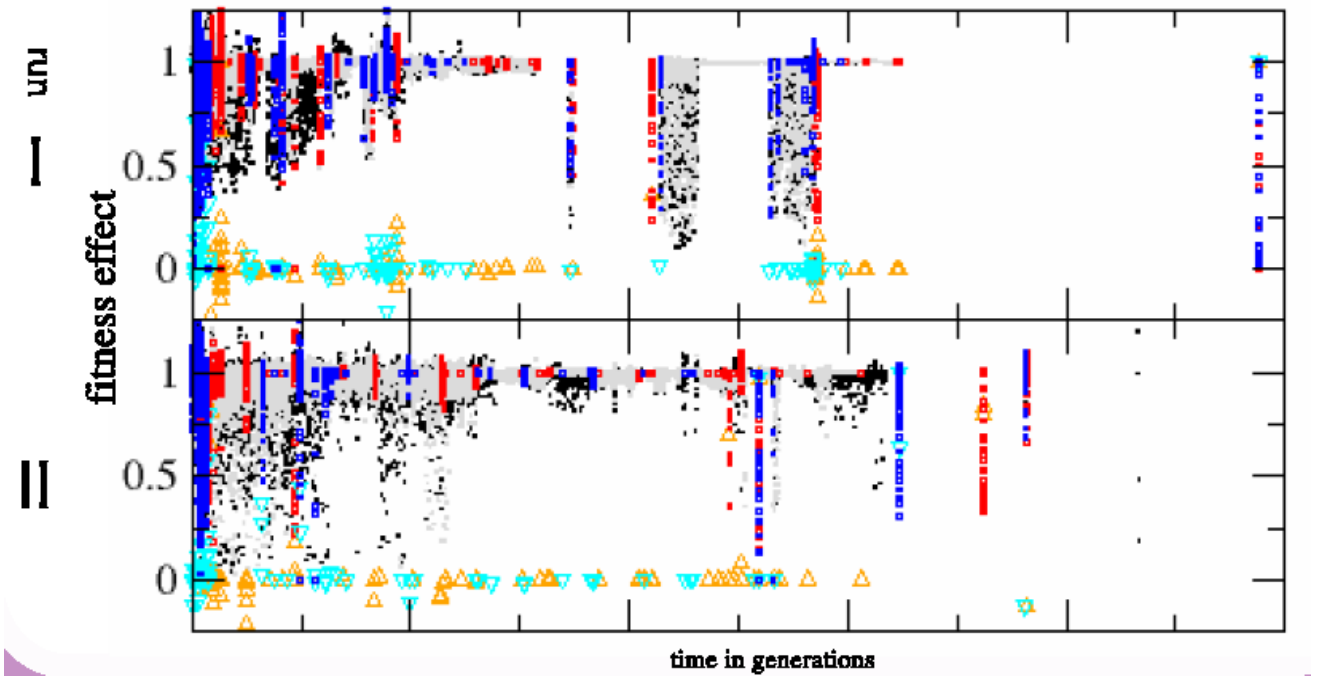


fitness reached increases with initial genome size

nevertheless streamlining

why streamlining?

gene loss decreases mutational load of *neutral* genes



Conclusion (1)

Surprising observations from bioinformatic data analysis of early genome inflation adaptation by gene loss
are
generic properties of Darwinian evolution

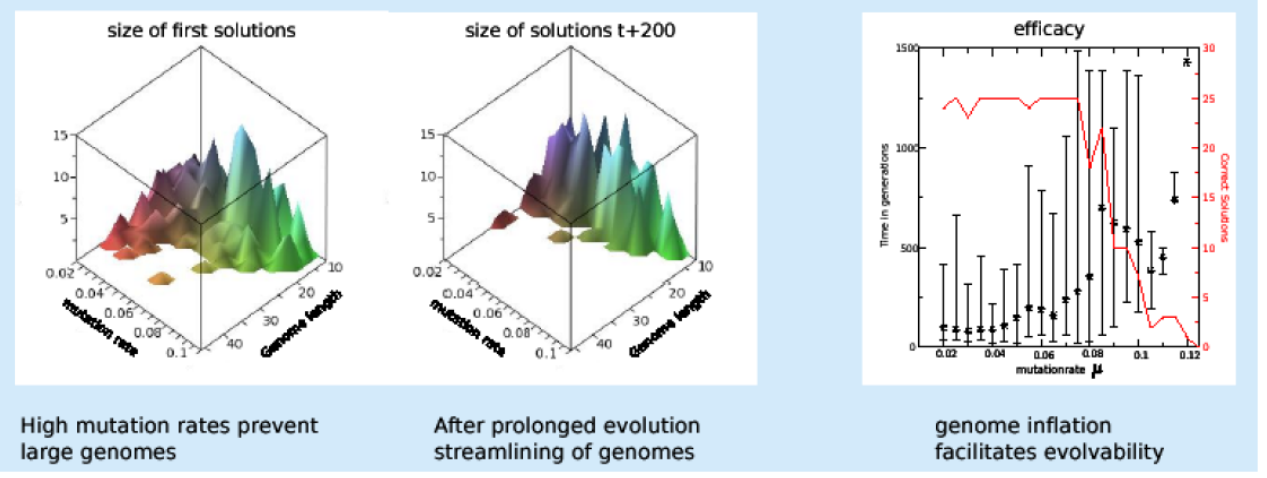
Models ++

e.g. AEVOL, Virtual Microbe, Function optimization, ...

Results ++

- Pattern of WGD fixation and subsequent evolution
- evolution of regulation vs evolution of evolution
- evolution of mutational neighborhood

Models ++
mutation rate, genome inflation and streamlining function optimization (de Boer & Hogeweg 2012)

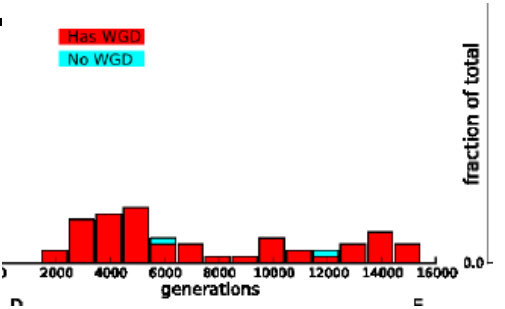


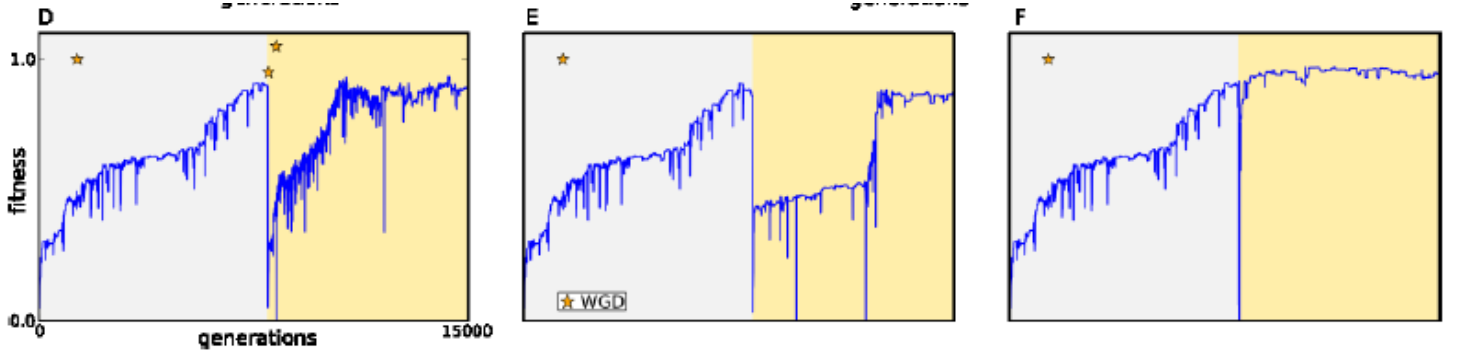
high mutation rate prevents genome expansion and compromises evolvability

High mutation prohibits genome expansion and therewith reduces evolvability

WGD in (adapted) virtual cell model ab initio evolution and re-adaptation switching to novel environment

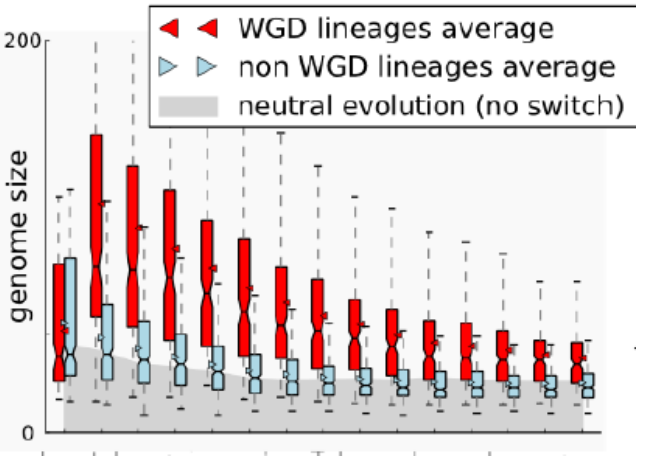
- almost all fit lineages had an early WGD and became fit much later
- minority of cases had WGD after switch
- NO WGD at intermediate times
- some VERY fast re-adaptation (no WGD) < 5 mutations

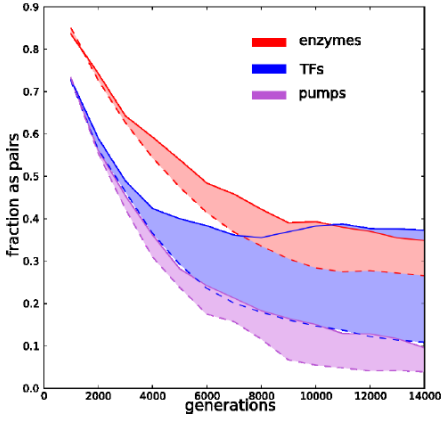




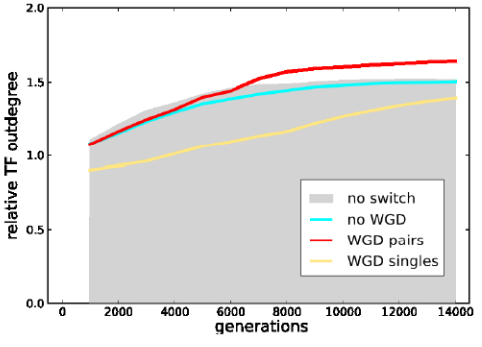
Differential gene loss after WGD :doses balance selection

- Streamlining, but larger genomes after WGD: "irremediable complexity"
- TF preferential kept
- with high connectivity
- NO sub-functionalization
- adaptation by peripheral TFs

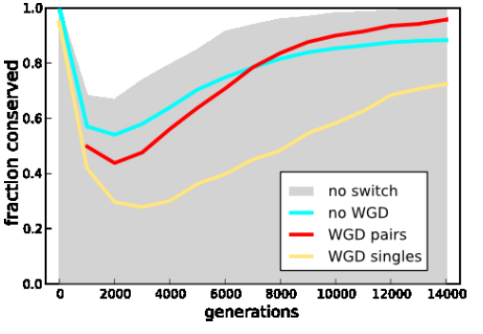




retained genes



their out-degree



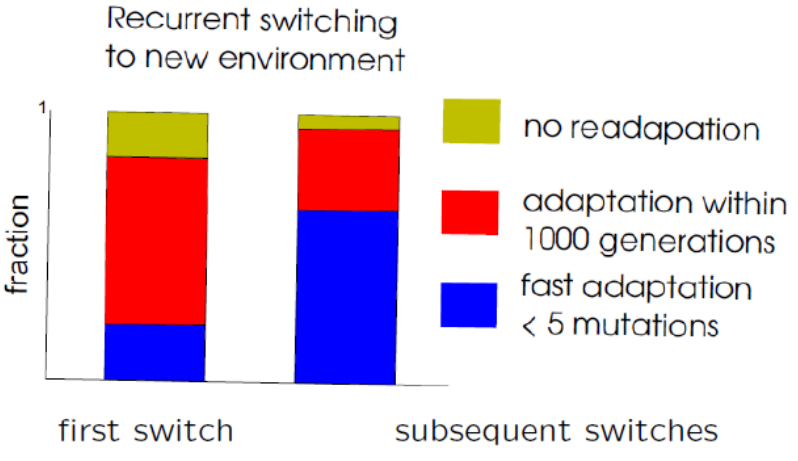
conserved binding

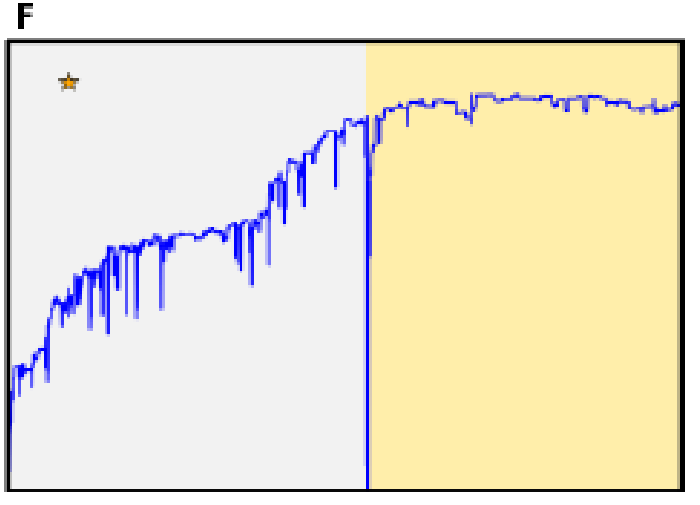
Use the evolved (fit) virtual cells to study short term evolution. Maintaining homeostasis in NOVEL environments

proxy for novel environments:

(2-4fold) in/decrease conversion factor. passive diffusion,decay

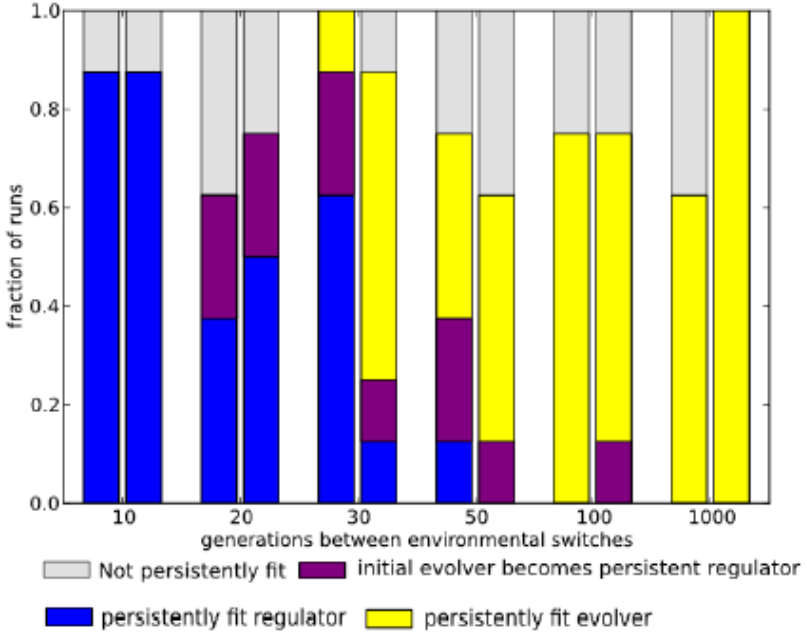
These change internal state (can be 'sensed')

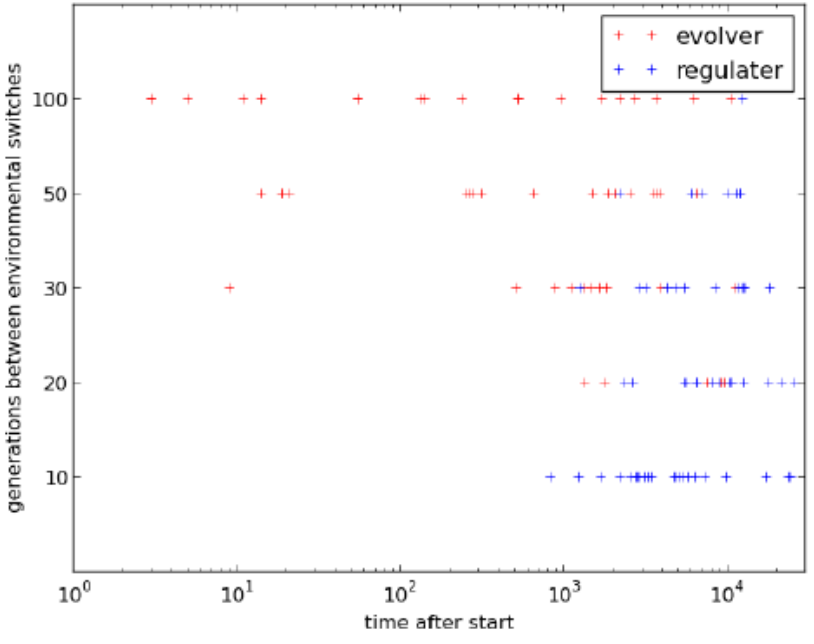




Cuypers, Rutten & Hogeweg 2017

regulation and evolvability alternative solutions
evolution of evolvability 'easier'





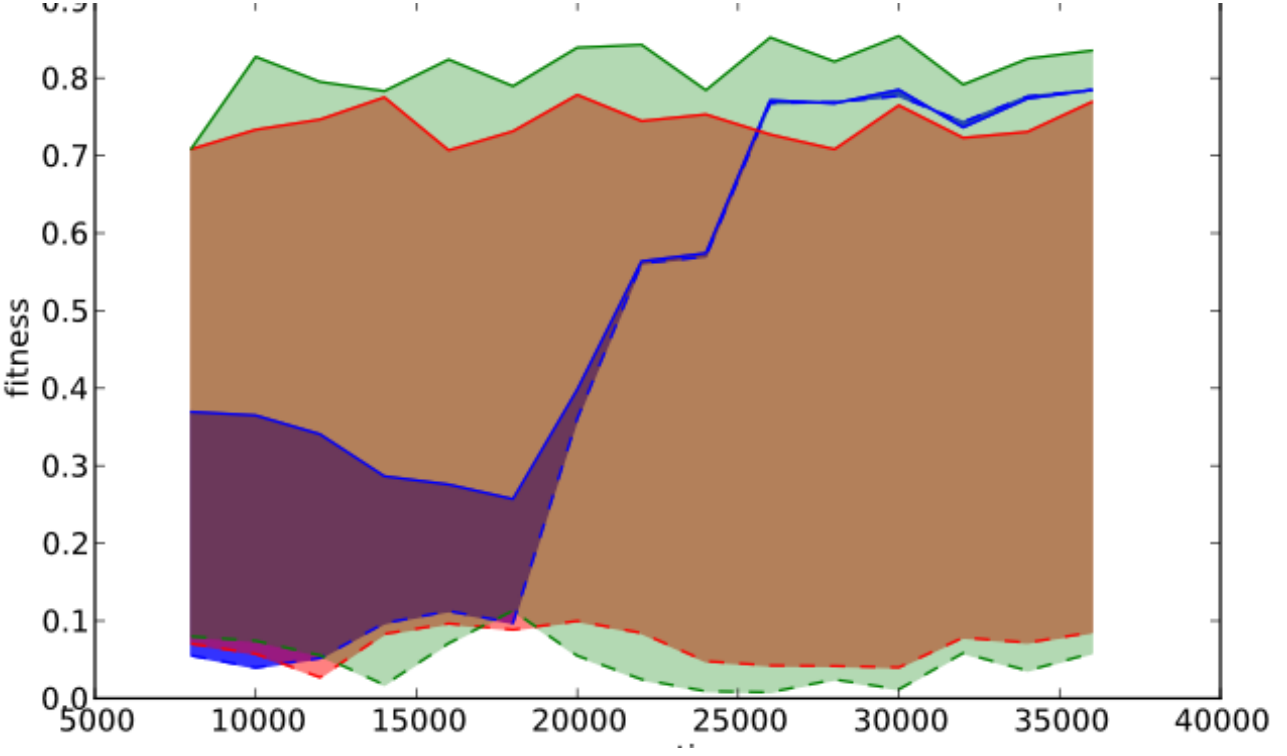
periodicity of switches

orange: evolver; blue regulator

2 different environments

regulation and evolvability alternative solutions of one WT

average fitness over 30 generations after switch



switch every 30 generations: dark blue: regulator; brown evolver
switch every 100 generations: light blue evolver

Note: higher fitness for less frequent switches
'better' adapted $->$ better evolvable

Conclusions evolution of virtual cells

- early genome inflations,
  increases degrees of freedom and therewith adaptability
- Intricate interplay of neutral and adaptive processes:
  adaptation $--> $ neutrality; neutrality $--$ adaptation
- Evolved genotype phenotype mapping maximizes
  neutrality AND selection
- Evolved genotype phenotype mapping increases evolvability
  to NOVEL conditions
- Evolvability and regulation 'equal' alternatives to cope with
  fluctuating environments
- Evolvability easier to evolve
- WGD frequent but rarely accepted
  only early in evolution or after environmental change

Conclusions:
Some "NON surprising" (and debated) observations generic properties of multilevel evolution

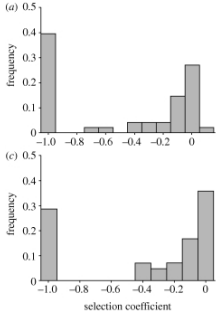
---

- **Early complexity**
  ** biological *Big Bangs*: major transitions in evolution
  ** large common ancestors
  ?* closely packed early species radiations
  ** important role of gene *LOSS* in adaptation
  ?* FECA to LECA: many gene duplications before species radiation
  ?? genes with "late" function often predate that function

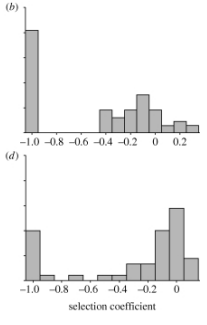
- **Whole Genome duplication rare but important**
  ** occurs often but rarely fixed
  ** at root of major radiation
  ** during major environmental shifts (?)

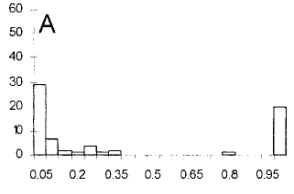
ResutIts ++ AND Models ++ (e.g. AEVOL)
Evolution of mutational neighborhood: U-shape

Flat and Steep;
Neutral and high Selection
Robust at individual and at population level
Evolvable at population level
Few slightly deleterious mutations

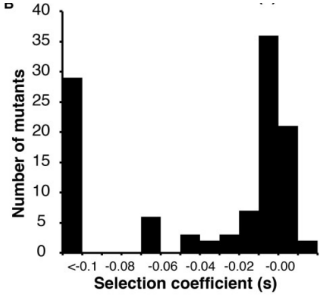


Viruses

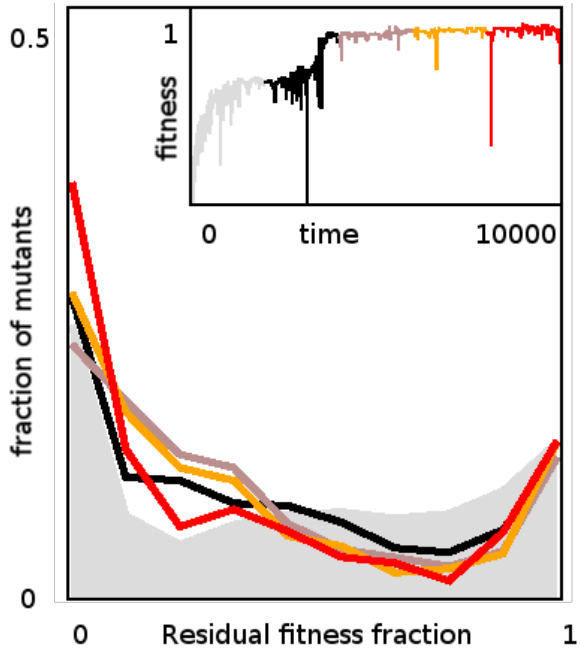




Yeast



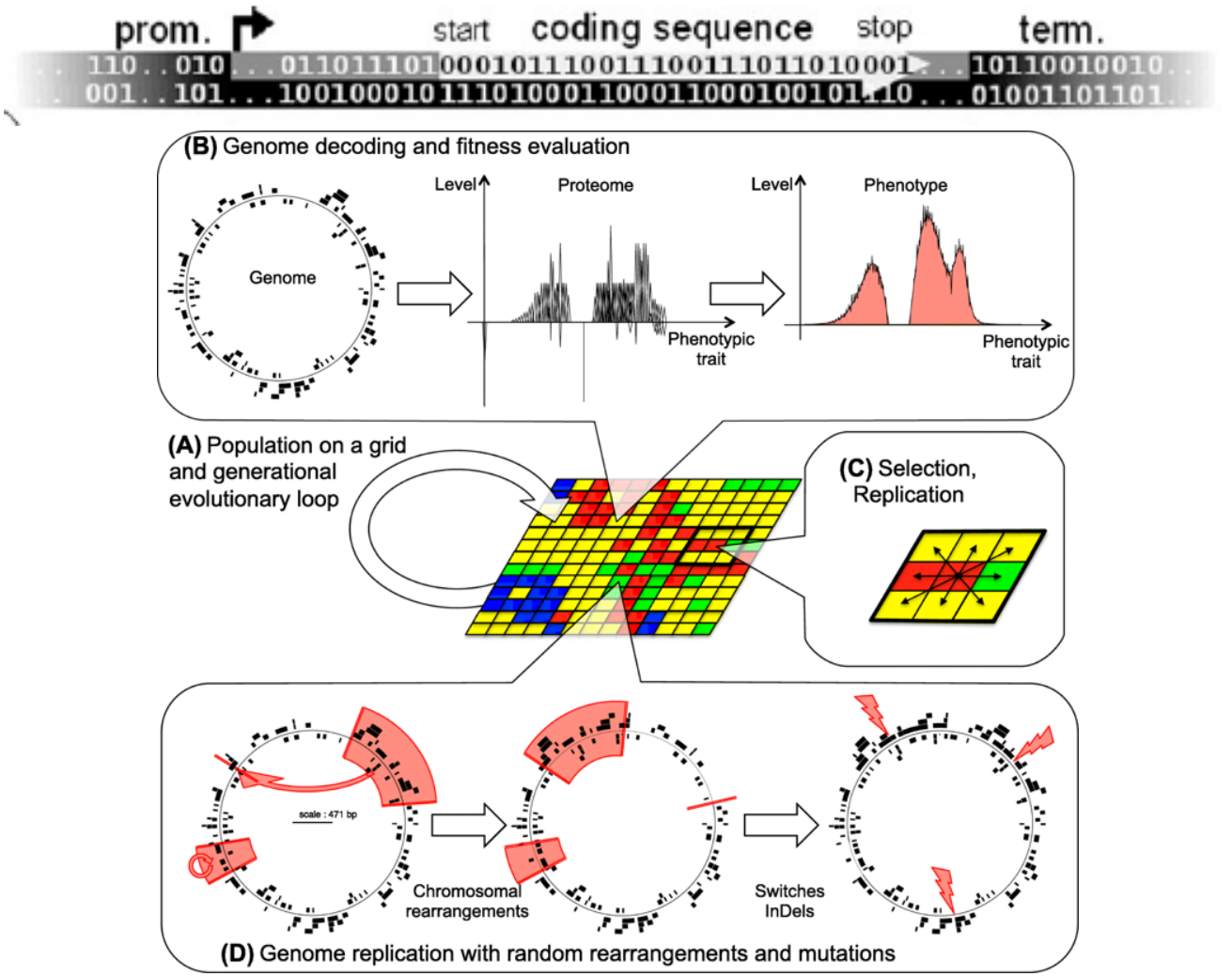
ARA-CDE



Virtual cell

U-shape: evolved property AND ideal for evolution

Aevol model structure (Beslon)



AEVOL "genetic" code

(A) Genetic code

000: Start
001: Stop
100: M_0
101: M_1
010: W_0
011: W_1
110: H_0
111: H_1

(B) Gene translation

RNA Seq. : ... RBS	000	010	101	100	011	111	011	101	101	110	001	...
Protein Seq. :	Start	W_0	M_1	M_0	W_1	H_1	W_1	M_1	M_1	H_0	Stop	

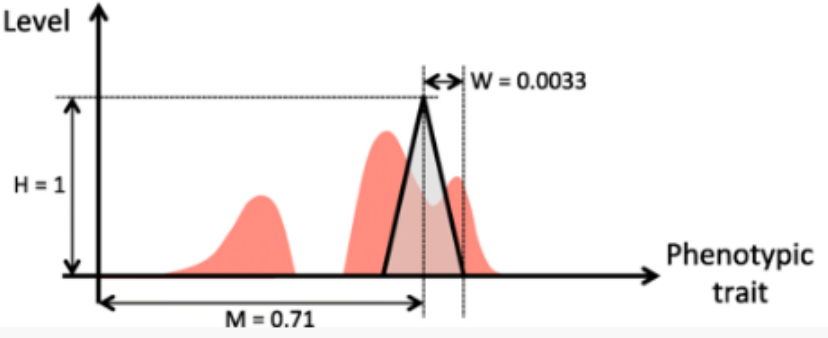
(C) Computation of the protein's function

$M_1M_1M_1 \rightarrow 111_{Gray} \rightarrow 5$
$MaxVal = 2^3-1 = 7$
$M_{range} : [0:1]$
$M = 5/7$

$W_0W_1 \rightarrow 01_{Gray} \rightarrow 1$
$MaxVal = 2^2-1 = 3$
$W_{range} : [0:W_{max}]$
$W = W_{max}.1/3$

$H_1H_0 \rightarrow 10_{Gray} \rightarrow 3$
$MaxVal = 2^2-1 = 3$
$H_{range} : [-1:1]$
$H = (2x3/3)-1$

(D) Graphical representation of the protein (with $W_{max} = 0.01$)



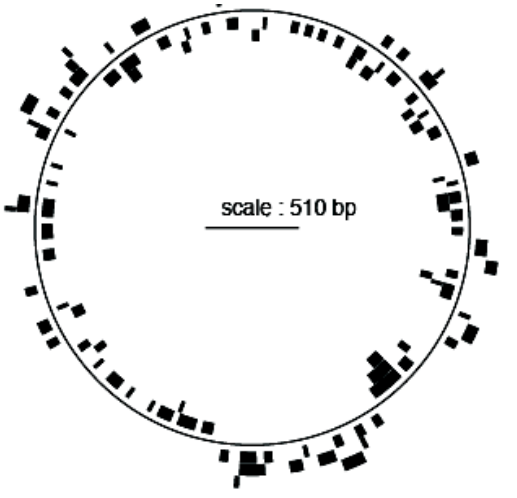
Long term evolution of WT strains: Genome structure dependent on mutation rates overlapping codes

evolved in constant environment; fine grained genome structure

Bacterium-like strains

compact genome closely packed genes

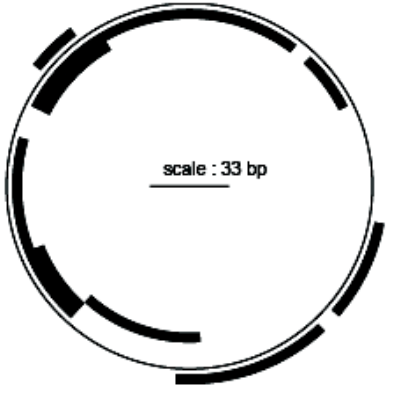
Aevol:
population on 40x40grid
pmut.rate= 10^{-6} mut/bp/gen
indels = 10^{-6} mut/bp/gen
LCR =1^{-5} mut/bp/gen



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



Mutation rates and Genome size vs error threshold

Rutten, Hogeweg & Beslon 2019

Mutator strains in E.coli e.g. 50% LTEE experiments

LTEE mutator populations are as fit or fitter than non-mutator strains

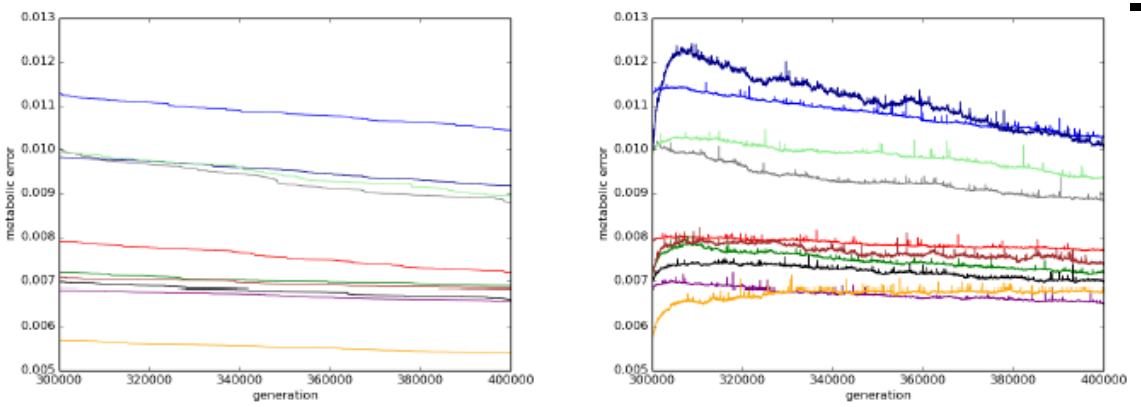
Question: how do populations evolve to cope with this?

Pre-evolve AEVOL populations with standard mutation rates.

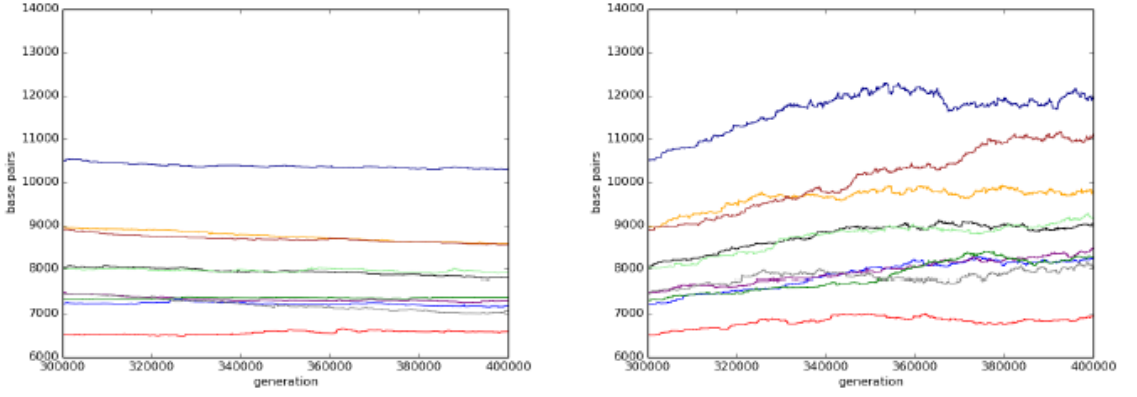
Create mutators strains (100 fold increse of point mutations)

Evolution of mutators and non-mutators

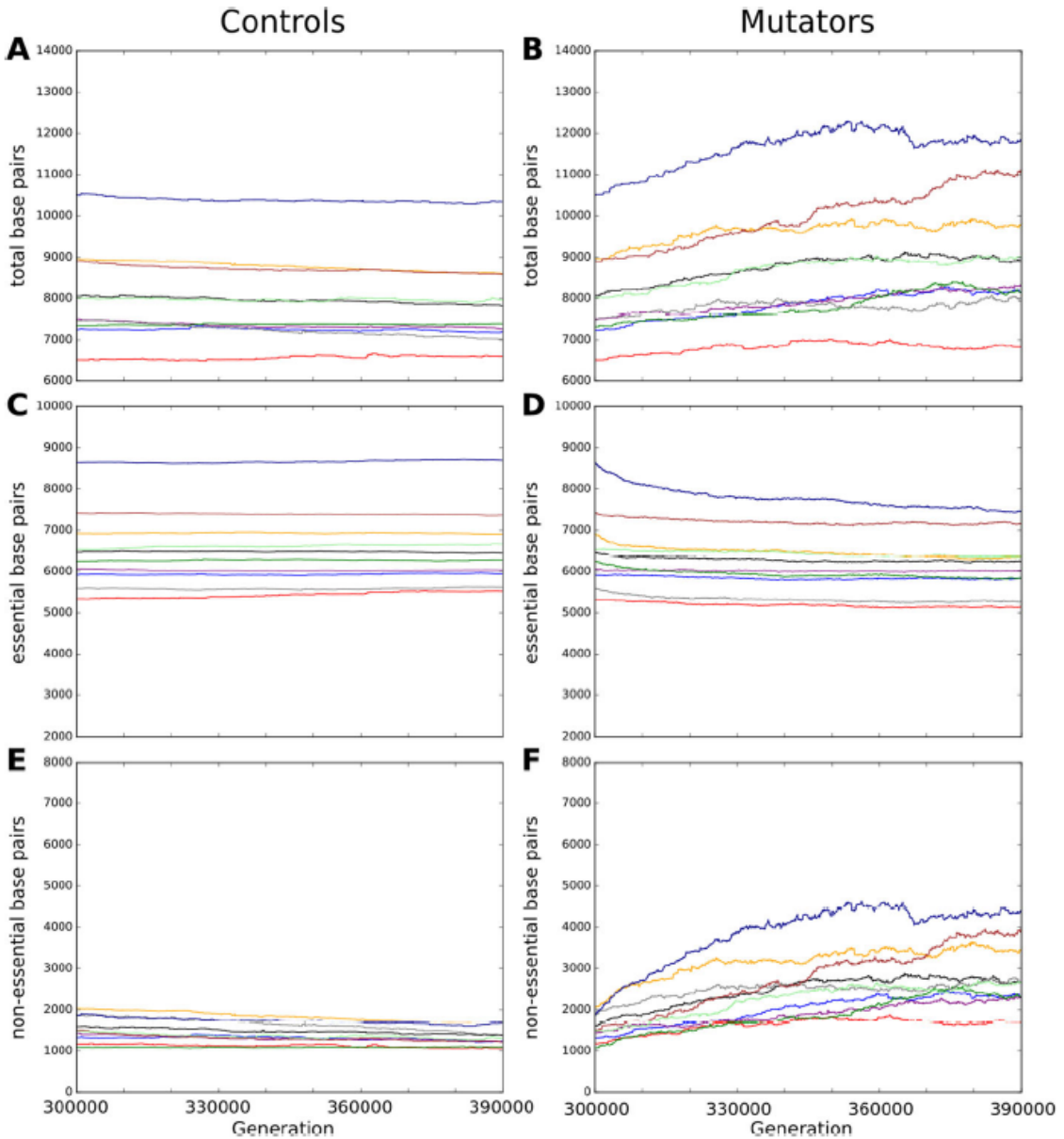
Mutators Increase genome size and recover fitness (ancestor lineage)



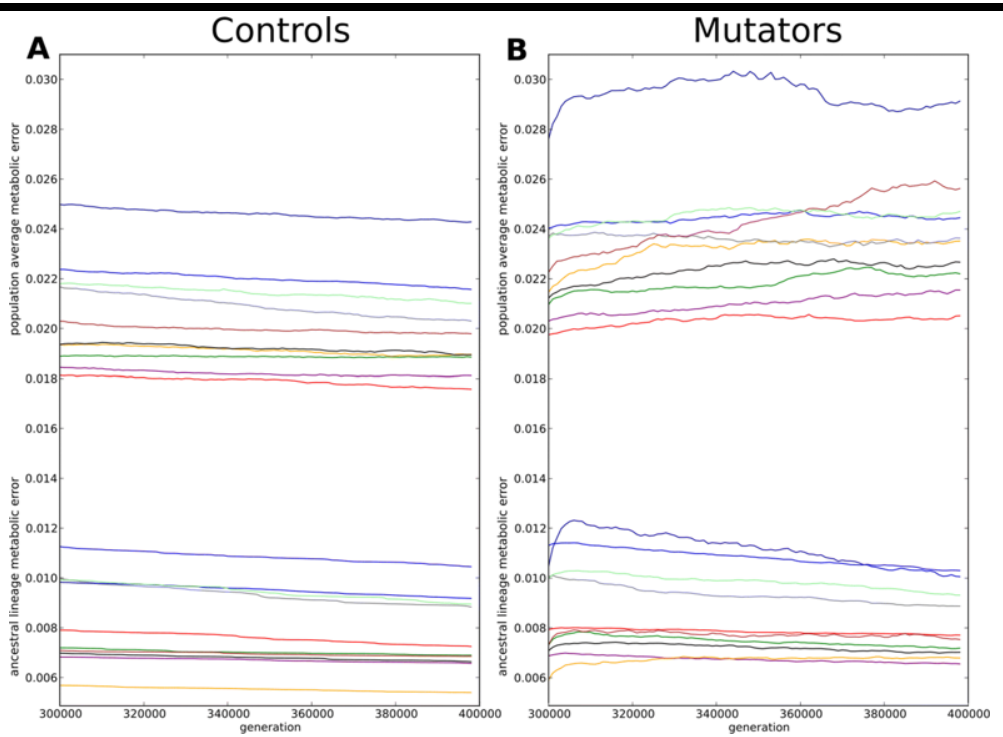
fitness: WT and Mutator



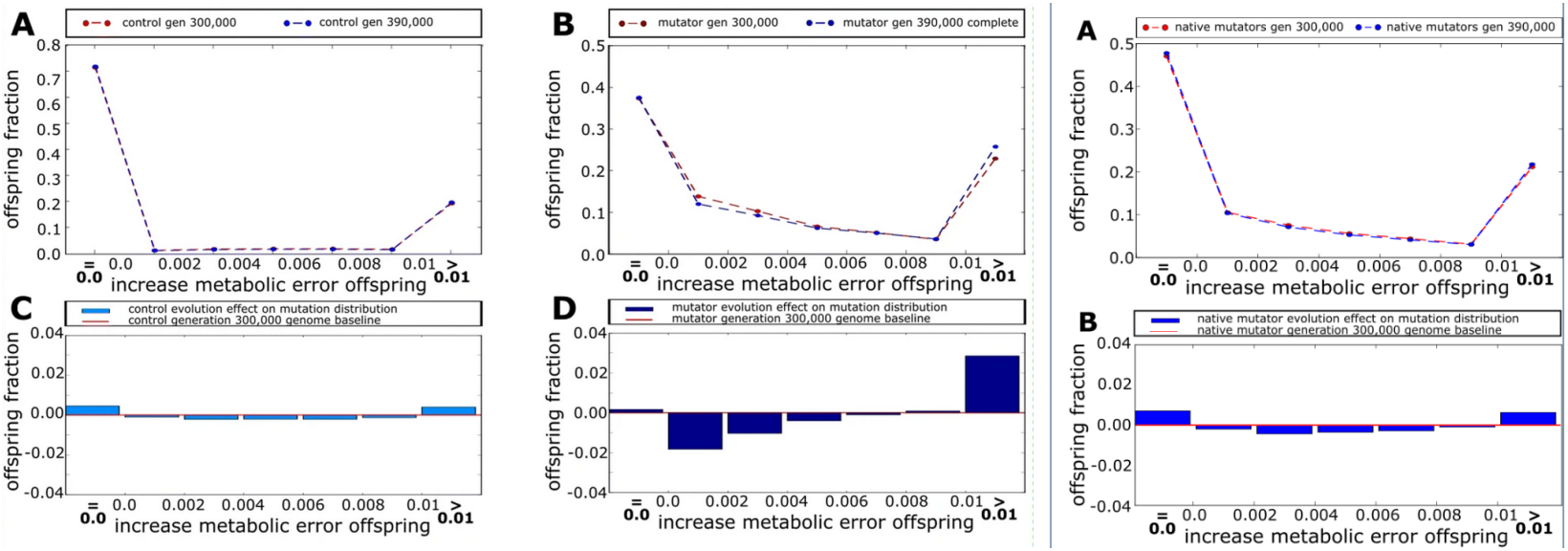
genome-size: WT and mutator


Controls
Mutators
A
B
C
D
E
F
total base pairs
essential base pairs
non-essential base pairs
Generation
300000
330000
360000
390000

Note: average fitness of mutator population decreases



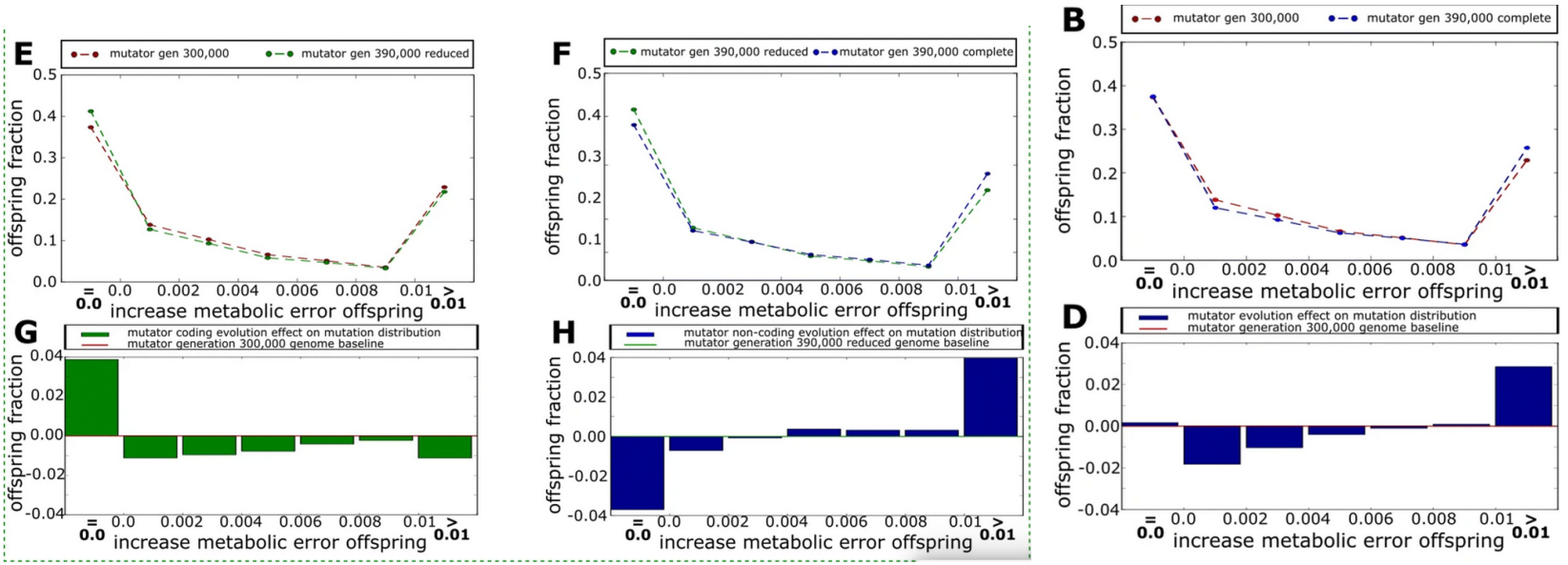
U shape mutational profile and mutator strains ancestor t=300.000 vs t=390.000



wildtype mutator from wildtype mutator

ongoing "deepening' of U shape; skewed toward deleterious after switch

Genome expansion and U shape mutational profile
Coding vs non coding sequences



in coding part in noncoding total

Reduce evolved mutator genome at t=390000 by deleting non-essential bases till original size is reached.

compare reduced vs full ancestor and reduced vs full evolved

Conclusions Mutational Neighborhood

- U-shaped mutational neighborhood:
  high neutrality AND high selection

- Genome size and mutation rate:
  high mutation rate: small genomes, overlapping genes (virusses
  Lower mutation rate: larger but compact genomes;
  BUT
  mutator strains *increase* genome size and regain fitness

- increased genome size due to increase non-coding regions
  (decrease of coding length
  leads to increase in "nonSNP's (LCR)
  and deleterious mutations
  skewed U-shape and stronger selection
  Compare RNA at high mutation rates!

Conclusions/Discussion
Non-supervised multilevel modeling

Generic properties from case studies?
(compare model organisms)

Not: All such are such in predefined universe
But: these patterns emerge in
. "arbitrary/plausible" universes

Not: What Did happen in evolution
But: What do we expect to happen by mutation/selection

search images

Regulation of ribosomal DNA amplification by the TOR pathway Carmen V. Jack et al 2015

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

How did genome architecture evolve to enbale this "reversed" causation, and what are the evolutionary consequences?

Genome organization of Ribosomal DNA and transcription/translation in YEAST

high nutrient condition enhance transcription rates

high transcriptional load induces mutations
(transcription/replication conflicts

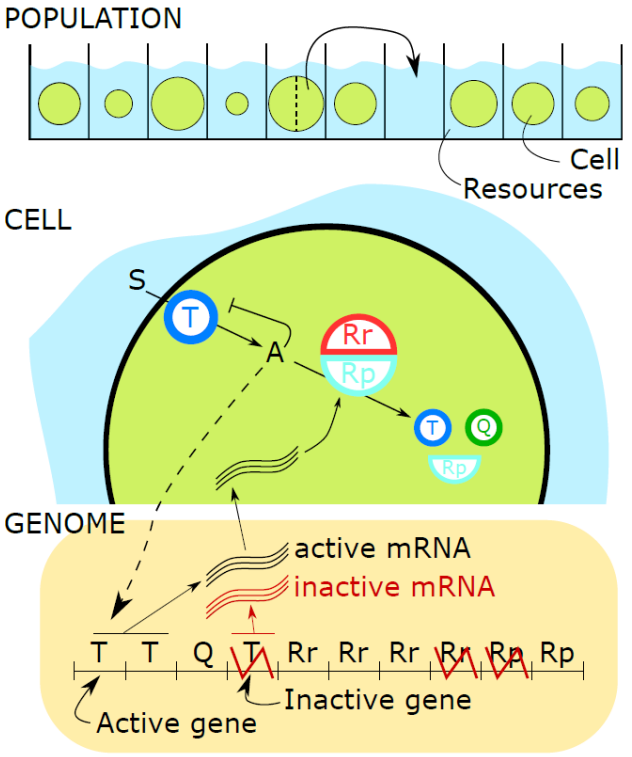
Replication Fork Barriers: in tandem Ribolomal DNA
divert these mutations to DupDels's (instead of SNP's

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

TOR pathway
biases toward double stranded break repair towards DUP's
Jack CV, et al.. PNAS 2015

model these properties and study their effect on long term evolution

Model

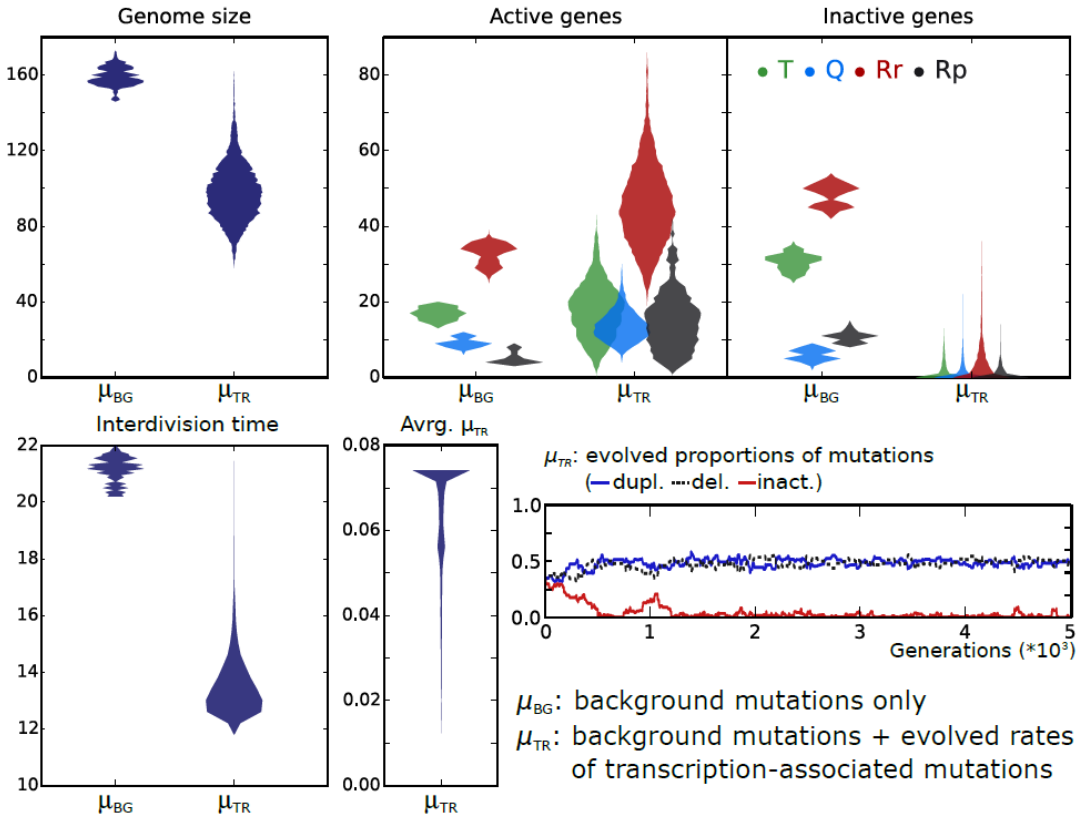


$$F_{transcr(x)} = k_0(x) + k_1(x)A$$
k_0, k_1 evolvable

$$\begin{aligned}
\dot{S} &= S_{in} - d_s S - F_{\text{met}} \\
\dot{A} &= -d_a A + F_{\text{met}} - F_{\text{transl}}(m_T) - F_{\text{transl}}(m_Q) - F_{\text{transl}}(m_{R_p}) \\
\dot{R}_r &= F_{\text{transcr}}(R_r) - d_{R_r} \\
\dot{m}_T &= F_{\text{transcr}}(T) - d_{m_T} \\
\dot{m}_Q &= F_{\text{transcr}}(Q) - d_{m_Q} \\
\dot{m}_{R_p} &= F_{\text{transcr}}(R_p) - d_{m_{R_p}} \\
\dot{T} &= F_{\text{transl}}(m_T) - d_t T \\
\dot{Q} &= F_{\text{transl}}(m_Q) - d_q Q \\
\dot{R}_p &= F_{\text{transl}}(m_{R_p}) - d_{R_p} R_p
\end{aligned}$$

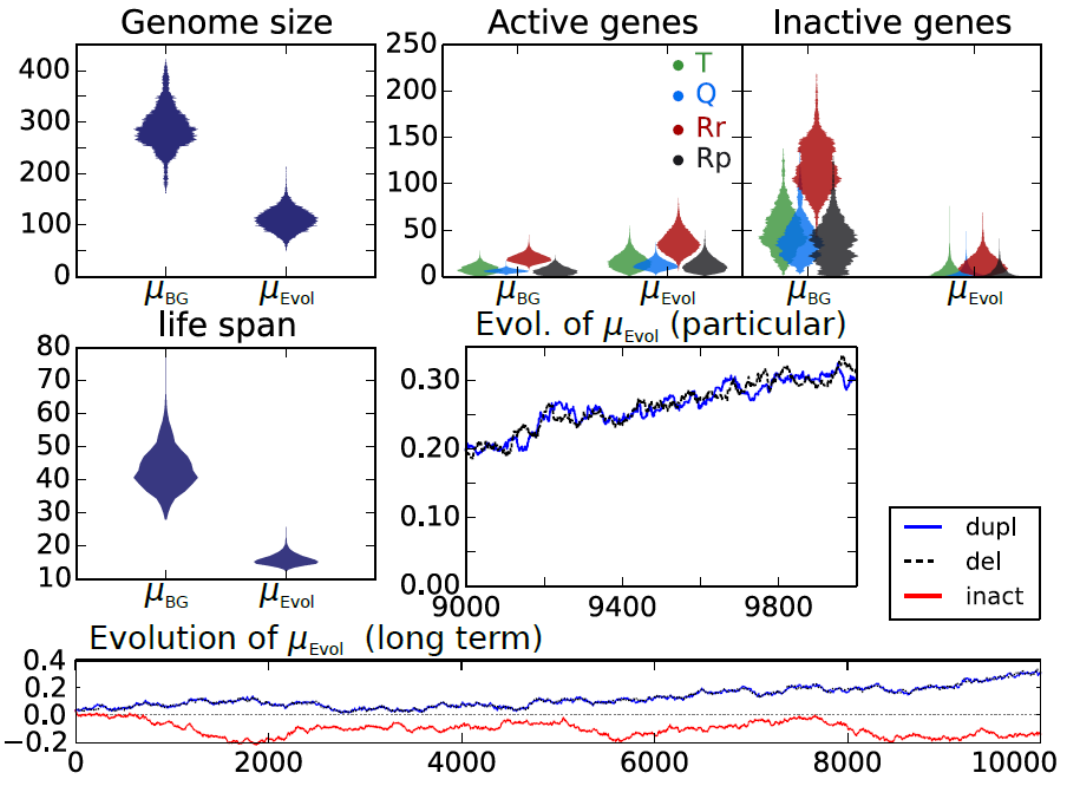
Background mutations + transcription induced mutations
inactivations, DupDels
Volume: number of macromolecules
Division Volume scales with genome size
total transcription polymerase limited (no dosage effect)

Given transcription translation conflicts bias towards DUPDELS evolves and prevents genome ans fitness degradation

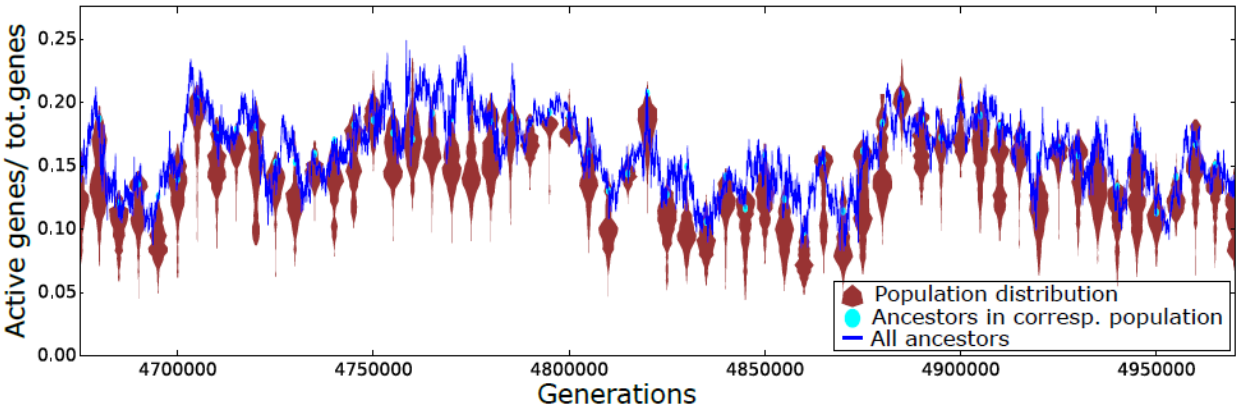


High transcription load evolves, and therwith much transcription-translation conflicts

No pre-imposed transcription induced mutations
high level of background mutations $- - - >$
high levels of transcription mutation evolve



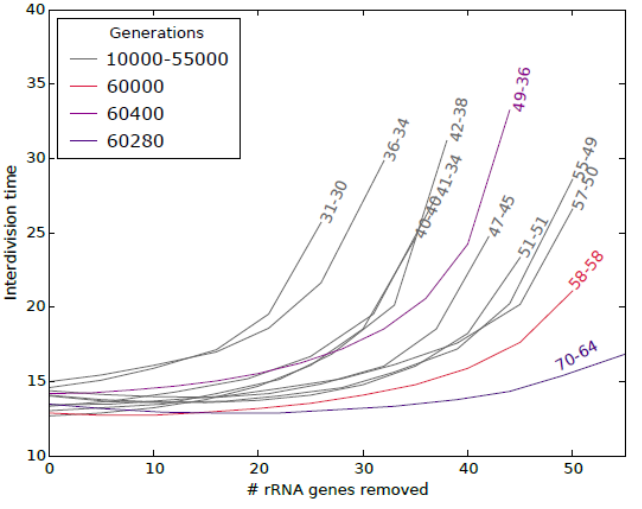
short term evolution favors duplication of active genes leads to higher per genome mutation rates, and genome degradation



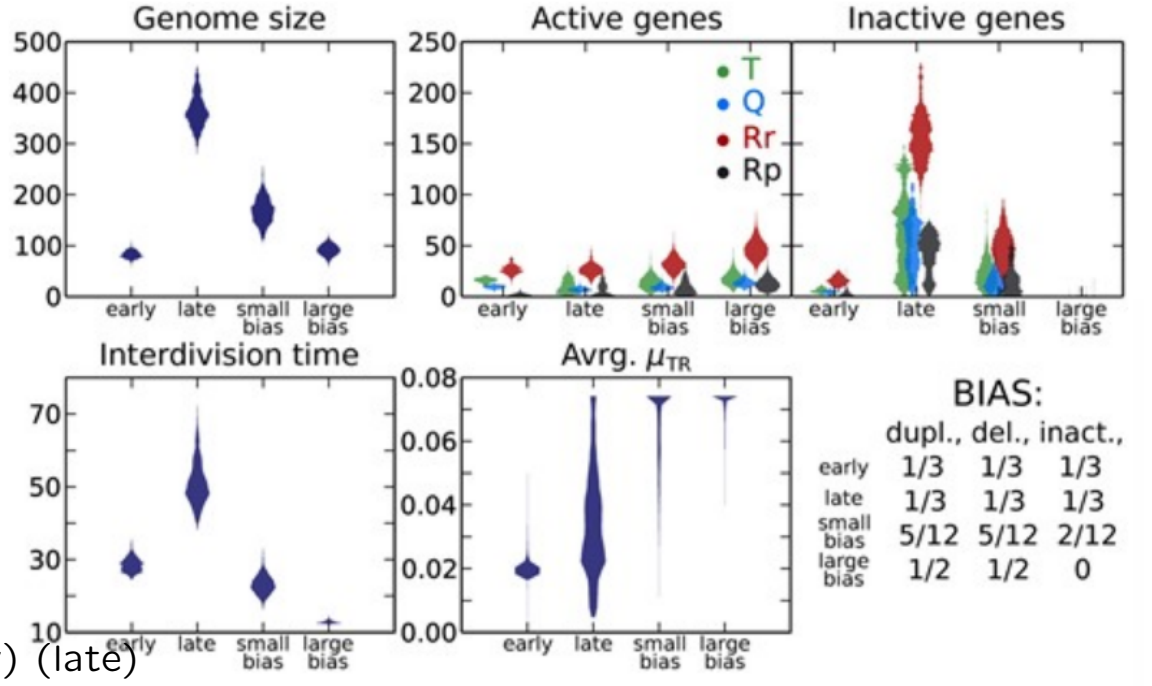
No (little) fitness effect for deletin

rRNA genes (like in Yeast)

$- - - >$ no selection fo deleting gene



feedback evolving transcription initiation and (inactive) gene accumulation



evolution of k0, k1: (early) (late)

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

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

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

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

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

"Not all mutations are created equal"

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

Number of "good genes" preserved by intercellular competiton

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

Genome organization has evolved in yeast to do so
(i.e. bias to DUPDELS of transcription induced mutations by replication fork bariers/TOR pathway