

Coping with a variable environment: evolvability and/vs regulation

Course Computational Biology 2018/2019; Paulien Hogeweg;
Theoretical Biology and Bioinformatics Grp Utrecht University

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

- **Early complexity**

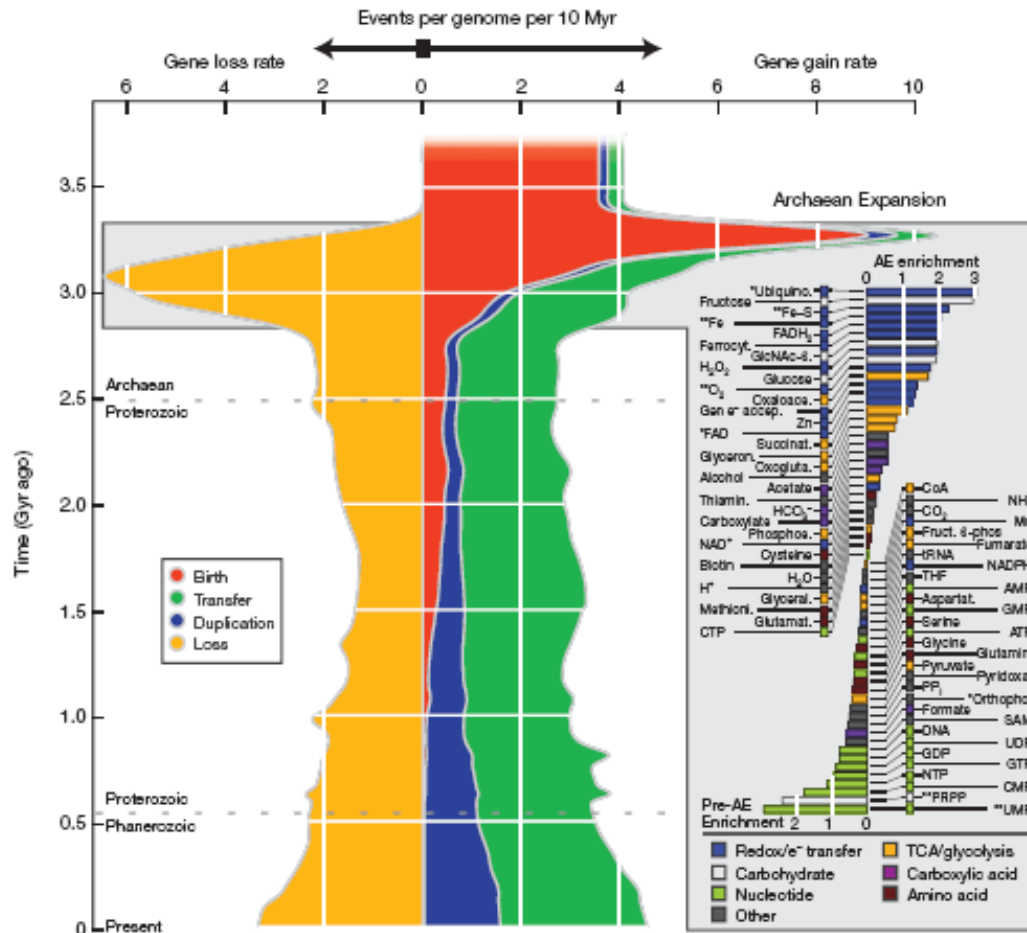
- biological *Big Bangs*: major transitions in evolution
- from pylogenies: large common ancestors
- from pylogenies: closely packed early species radiations
- important role of gene *LOSS* in adaptation
- FECA to LECA: many gene duplications before species radiation
- genes with “late” function often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

- **Whole Genome duplication rare but important**

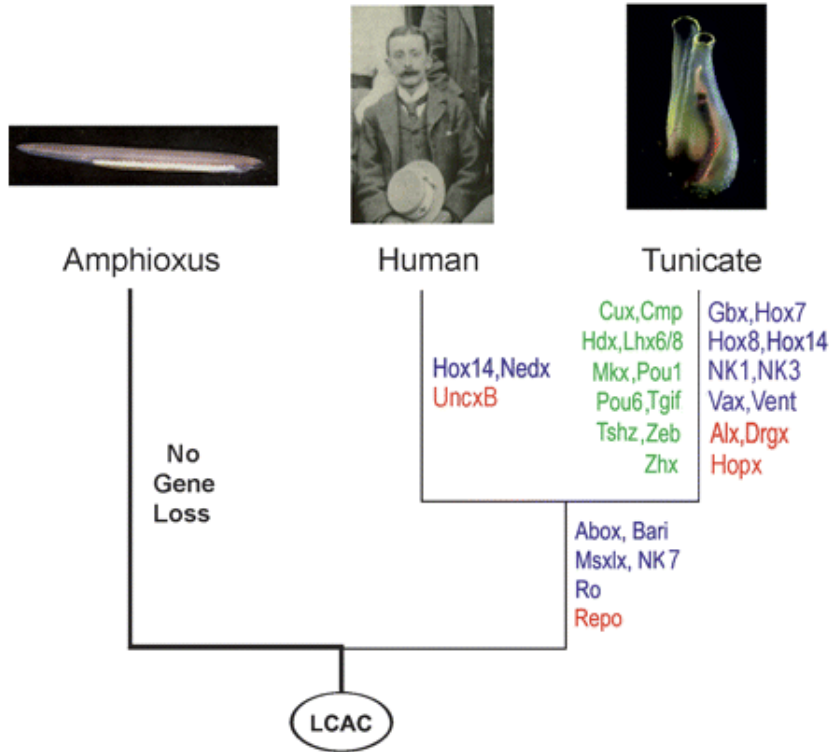
- occurs often (especially plants) but rarely fixed
- at root of major radiation
- during major environmental shifts (?)

early gene innovation - and loss

Alm Nature2010



Gene loss as major evolutionary process



Metazoa
Loss of homeobox genes

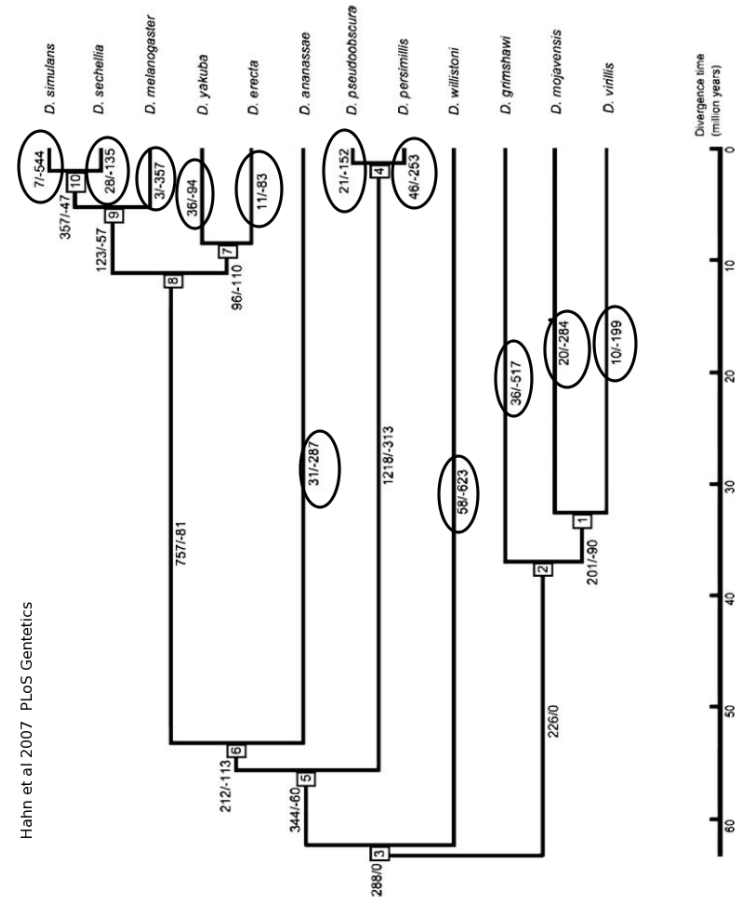
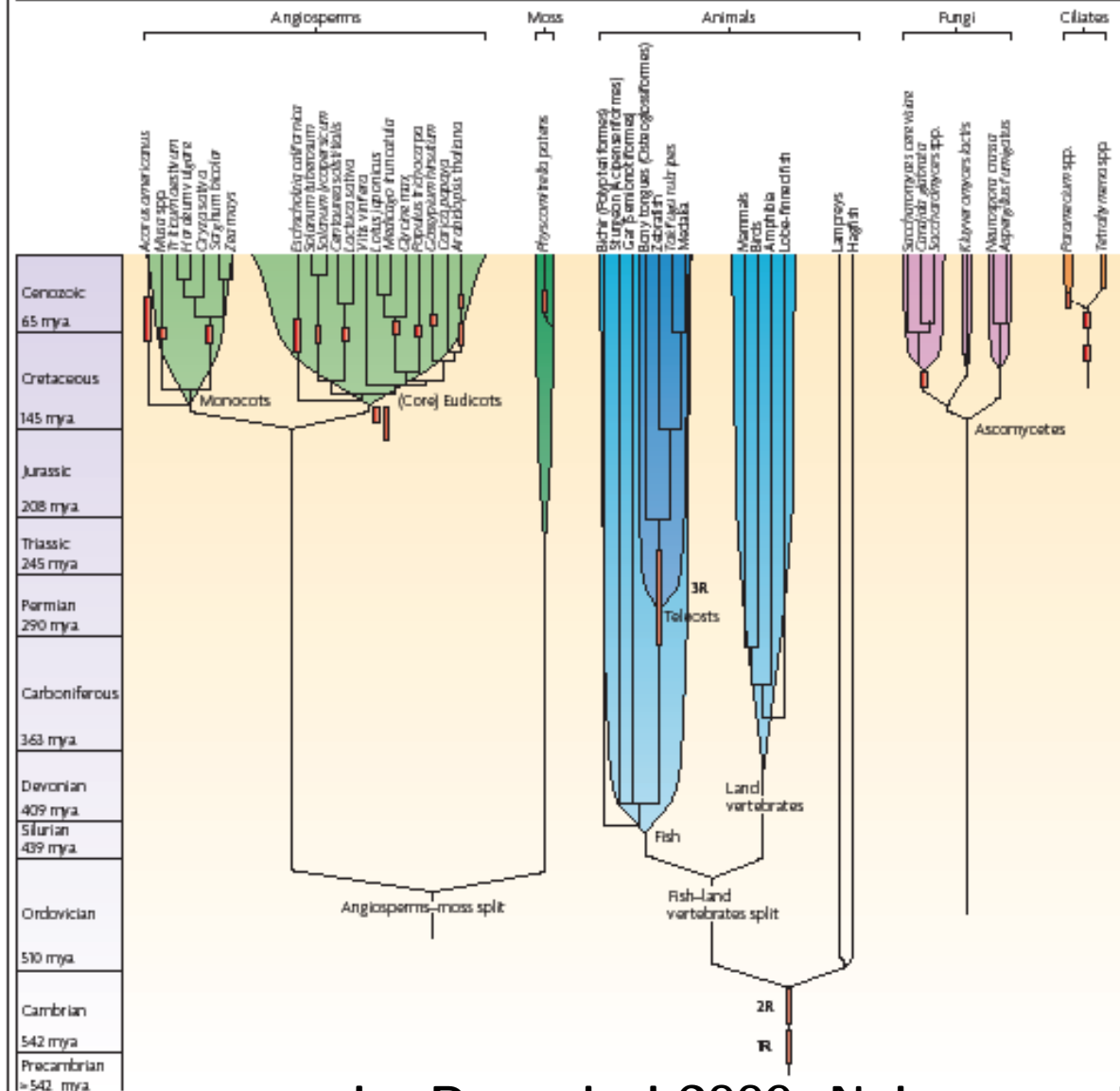


Figure 3. Lineage-Specific and Extinct Gene Families

Drosophila species
gain/loss of genes

Box 1 | Whole-genome duplications across the phylogeny of eukaryotes



van der Peer et al 2009, Nature genetic reviews

Evolution in virtual cells: genome. GRN, metabolism

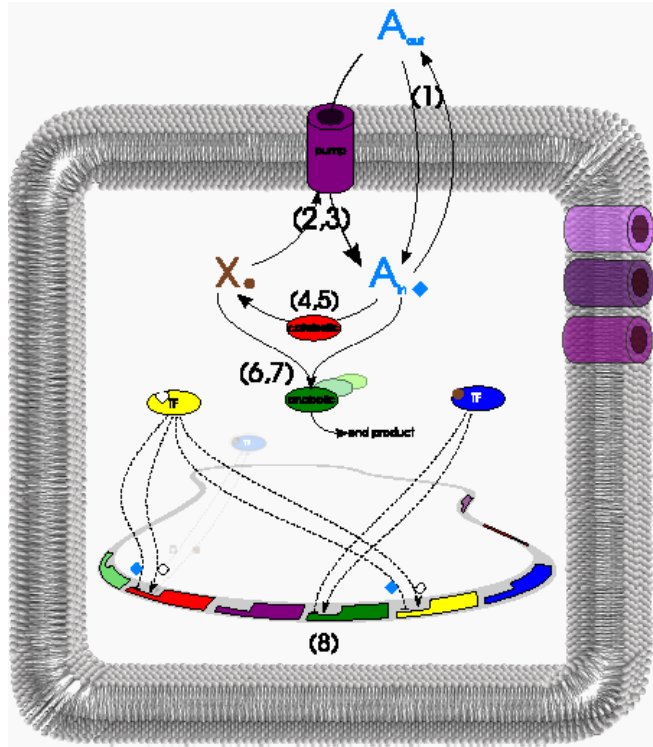
based on “plausible” *minimal* multilevel 'cell'
mutations segmental duplications/ deletions, pointmutations
fitness: *homestasis* (evolves regulatory adaptation)
evolving in varying environment

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

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



$$\begin{aligned} \frac{d[A]}{dt} &= ([A]_{out} - [A])P_{perm} & (1) \\ \left(\frac{d[A]}{dt} &= \frac{[A]_{out}[X]V_{max_p}[Prot]_p}{([A]_{out} + K_{A_p})([X] + K_{X_p})} \right. & (2) \\ \frac{d[X]}{dt} &= \frac{-d[A]}{dt} & (3) \\ \left(\frac{d[A]}{dt} &= \frac{-[Prot]_c[A]V_{max_c}}{[A] + K_{A_c}} \right. & (4) \\ \frac{d[X]}{dt} &= \frac{-d[A]}{dt} E_{conv} & (5) \\ \left(\frac{d[A]}{dt} &= \frac{-[Prot]_a[A][X]V_{max_a}}{([A] + K_{A_a})([X] + K_{X_a})} \right. & (6) \\ \frac{d[X]}{dt} &= \frac{d[A]}{dt} & (7) \\ \frac{d[Prot]}{dt} &= Pr \cdot Reg - Degr[Prot] & (8) \end{aligned}$$

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 (V_{max} , binding etc)

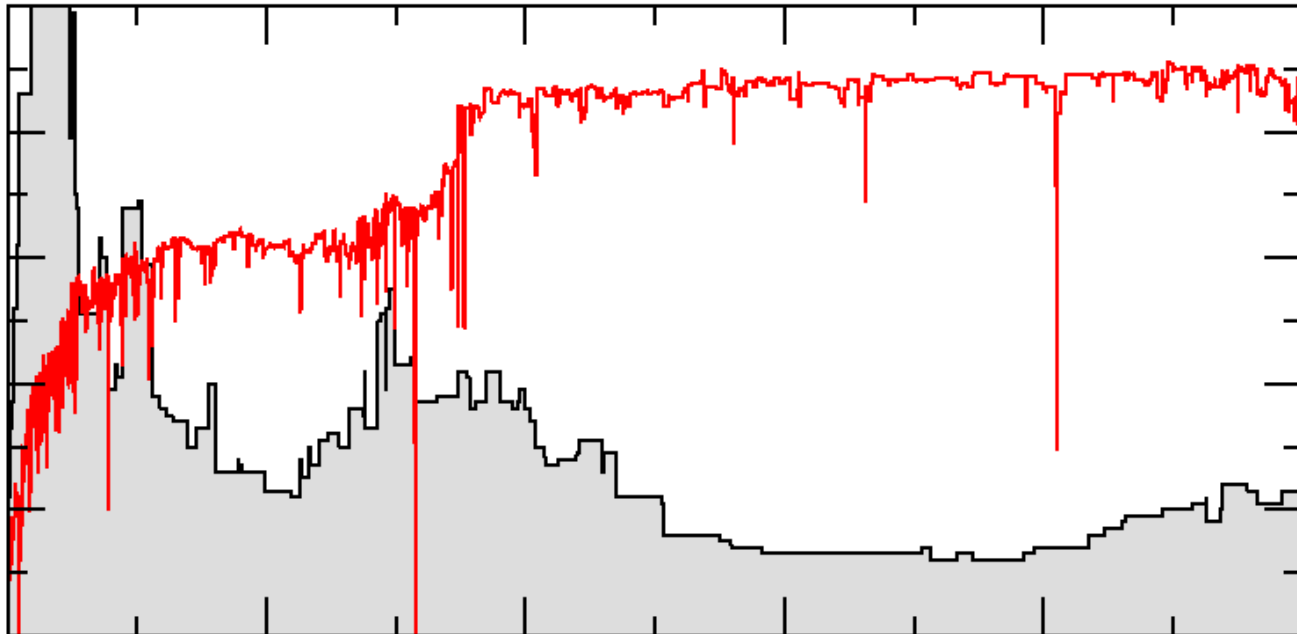
*Analysis along ancestral lineage
evaluated in 3 standard environments*

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**

Genome size and fitness

Ancestor trace

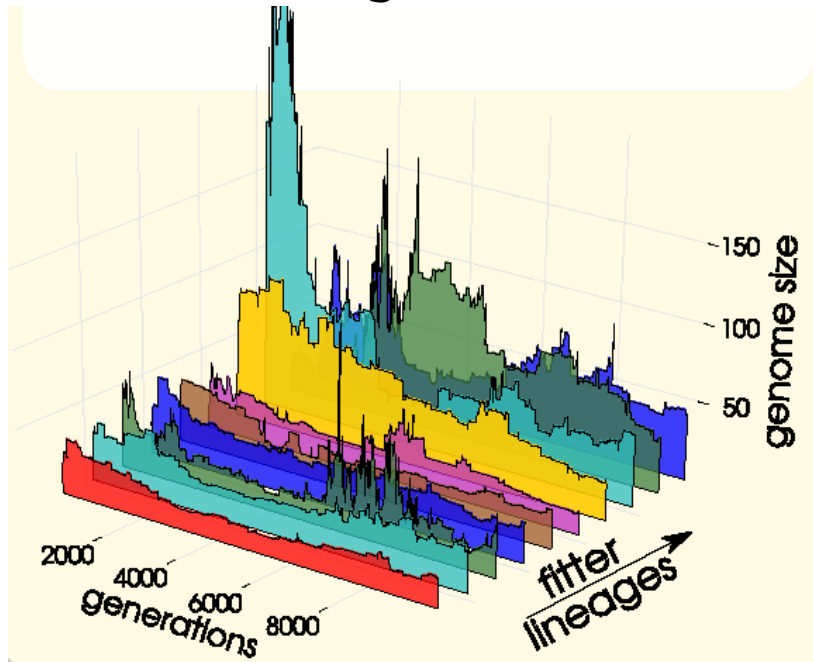


early genome inflation a “generic” pattern?

Yes... in the sense that:

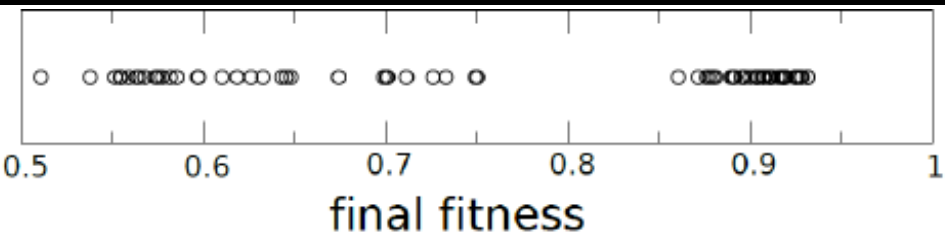
It occurs most pronounced in those runs which achieve high fitness eventually

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



| trend mutation type | 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?

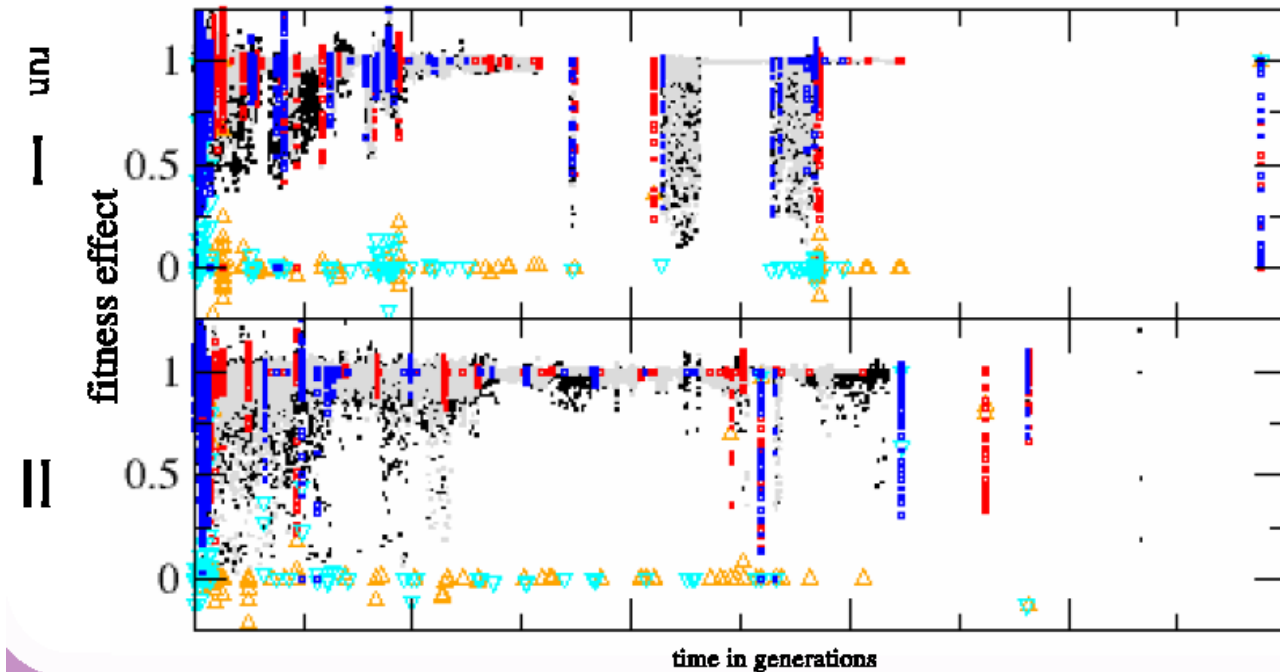
Duplications more often advantageous than deletions
+ **hitchhiking** of other genes (which might **later** become functional)

higher degrees of freedom increases adaptability

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 ++

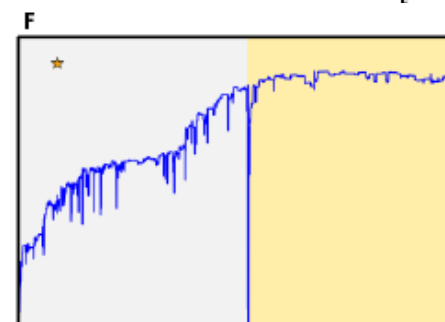
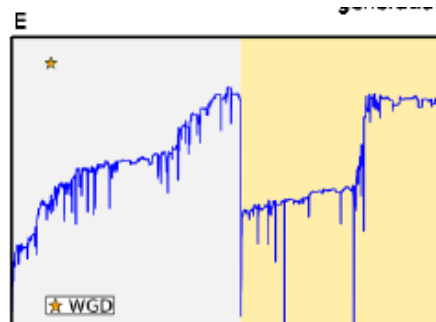
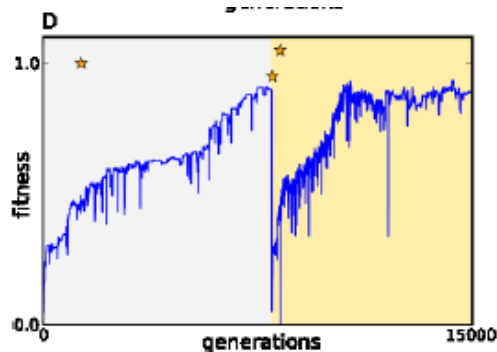
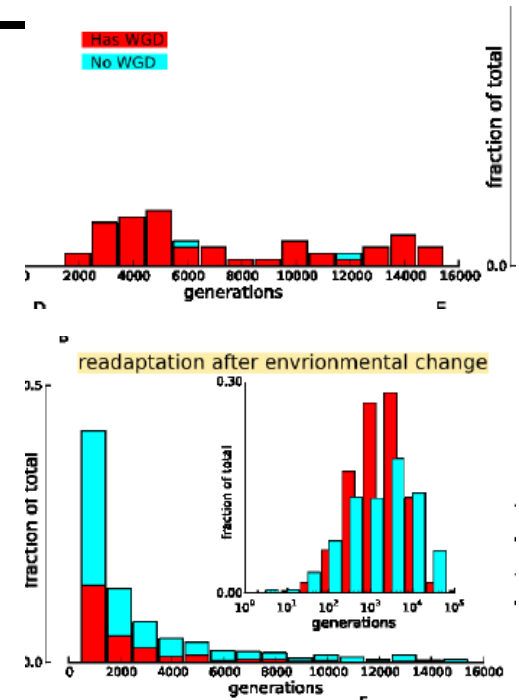
e.g. AEVOL, Virtual Microbe, Function optimization, ...

Results ++

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

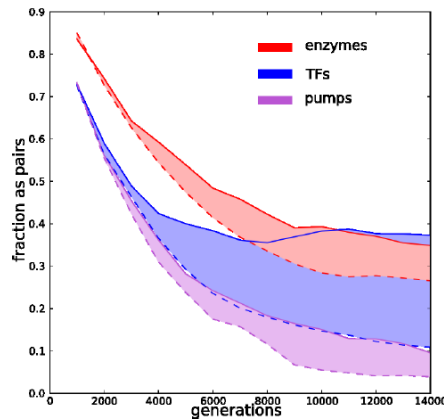
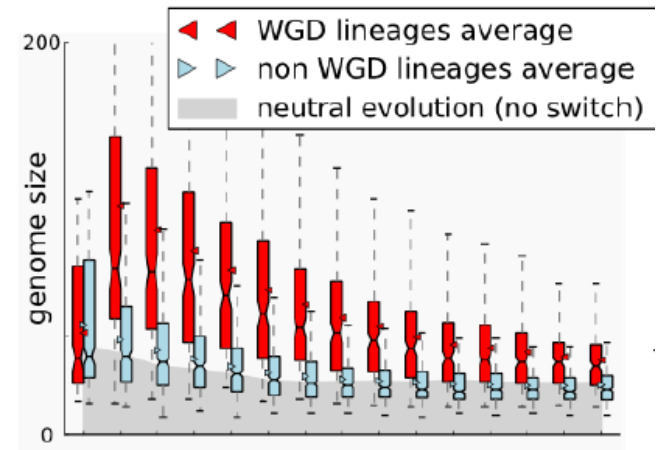
WGD in (adapted) virtual cell model ab initio evolution and re-adaptation switching to novel environment

- 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

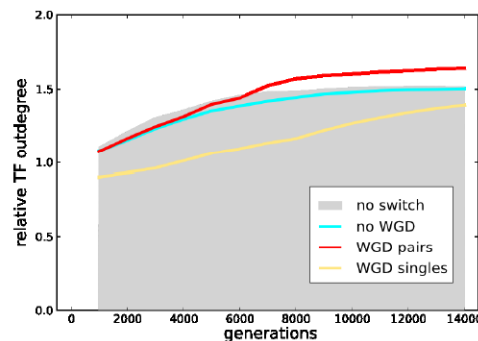


Differential gene loss after WGD :doses balance selection

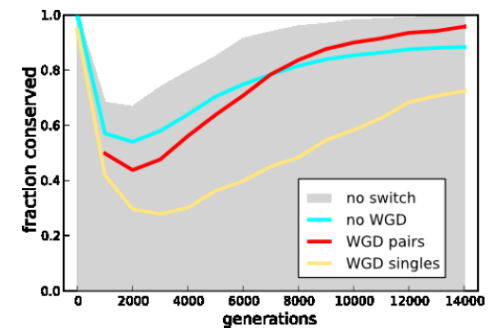
- Streamlining, but larger genomes after WGD: “irremediable complexity”
- TF preferential kept
- with high connectivity
- NO sub-functionalization
- adaptation by peripheral TFs



retained genes



their out-degree



conserved binding

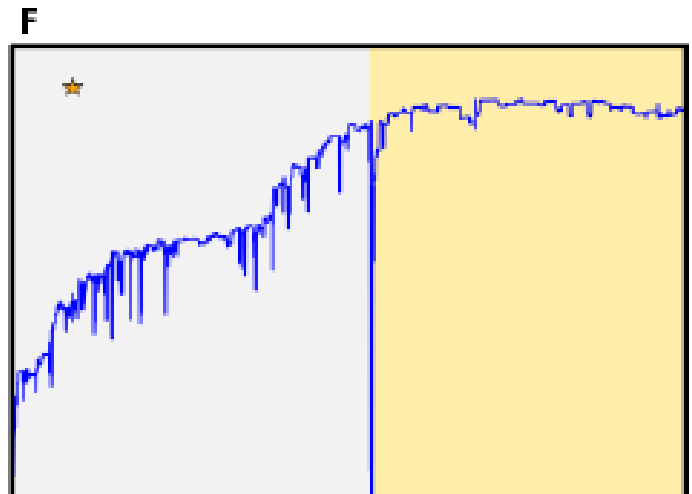
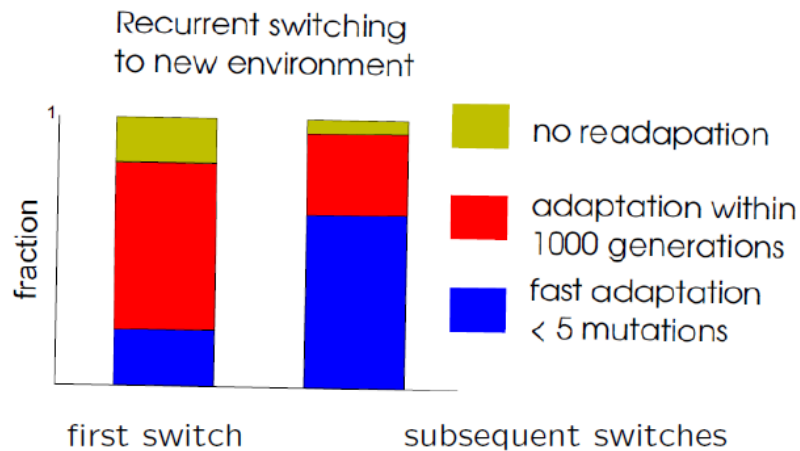
Use the evolved (fit) virtual cells to study short term evolution.

Maintaining homeostasis in NOVEL environments

proxy for novel environments:

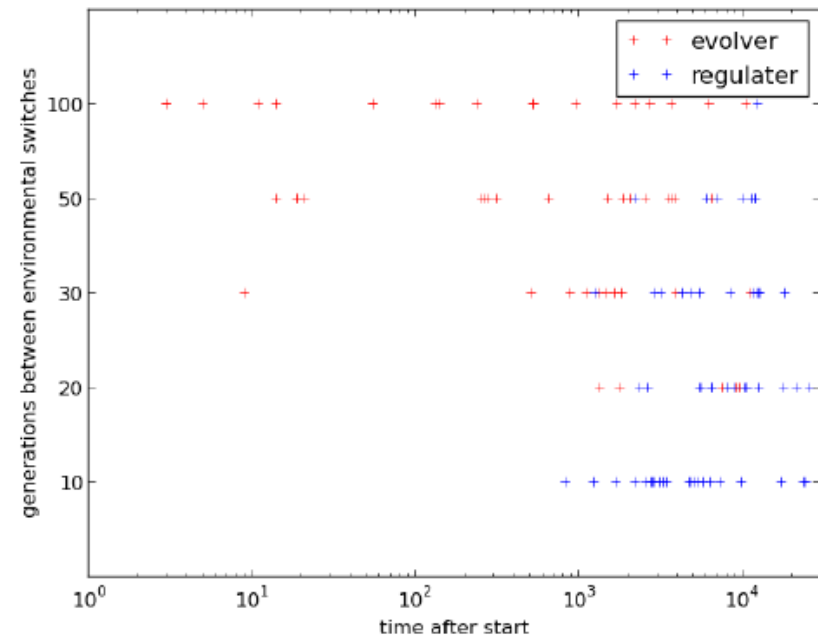
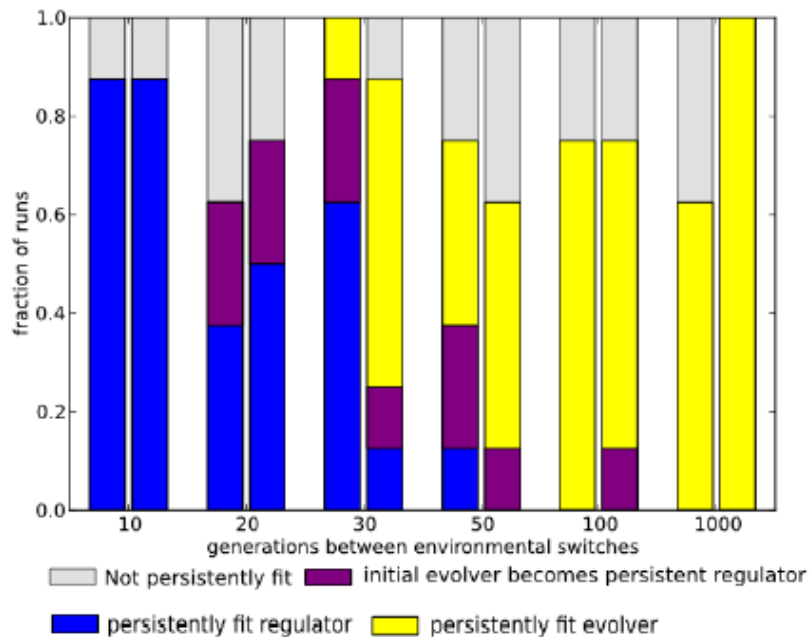
(2-4fold) in/decrease conversion factor. passive diffusion, decay

These change internal state (can be 'sensed')



regulation and evolvability alternative solutions

evolution of evolvability 'easier'



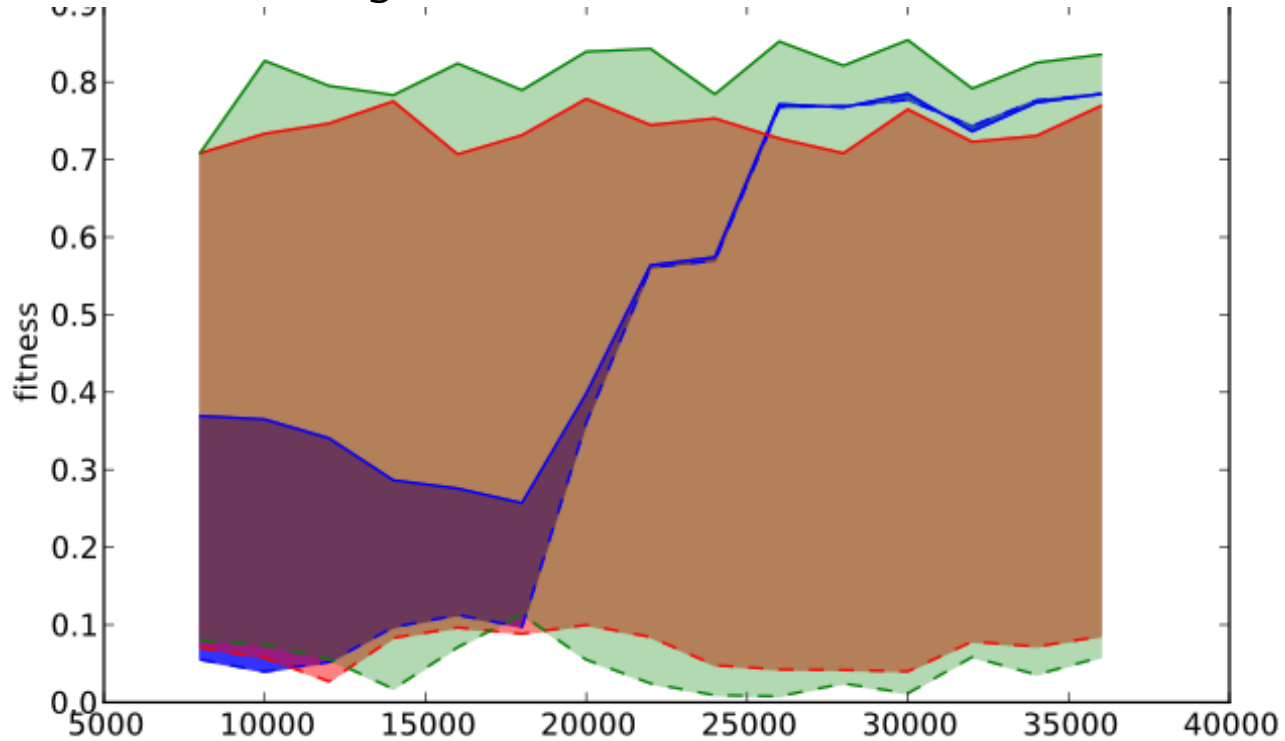
periodicity of switches

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

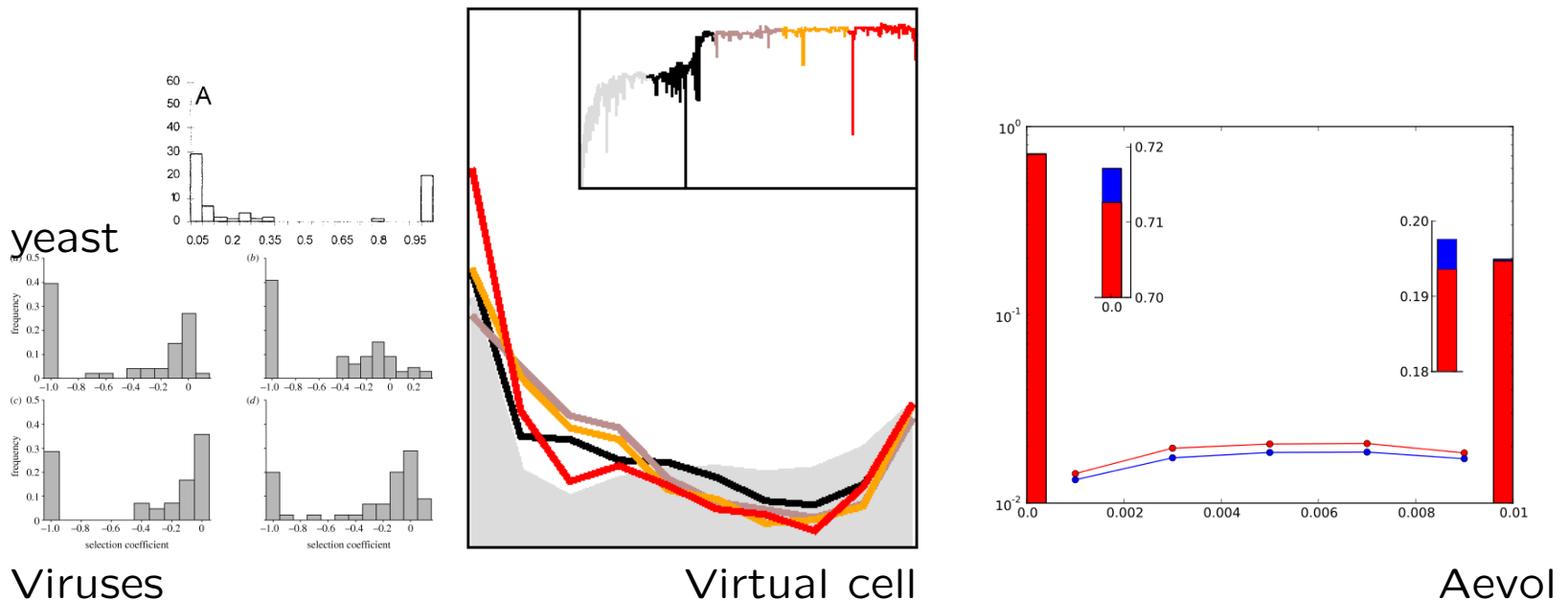
- **Whole Genome duplication rare but important**

- ** occurs often but rarely fixed
- ** at root of major radiation
- ** during major environmental shifts (?)

Results ++ AND Models ++

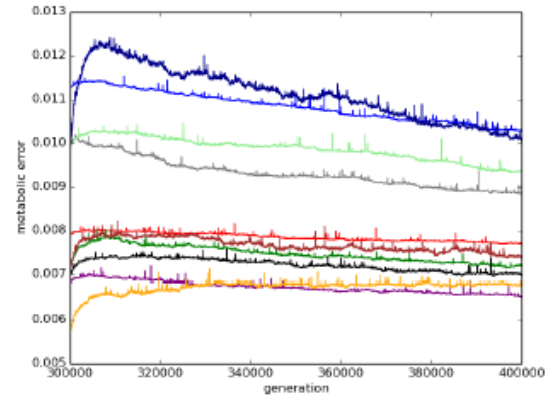
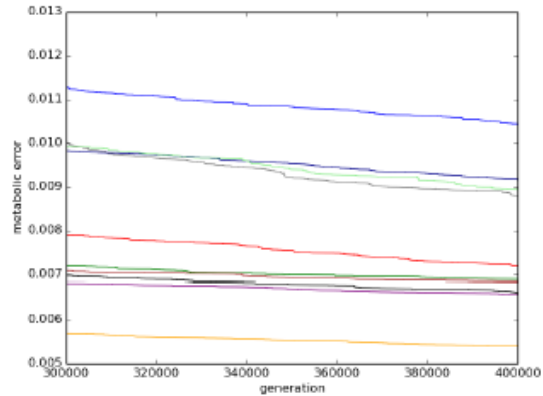
Evolution of mutational neighborhood: U-shape

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

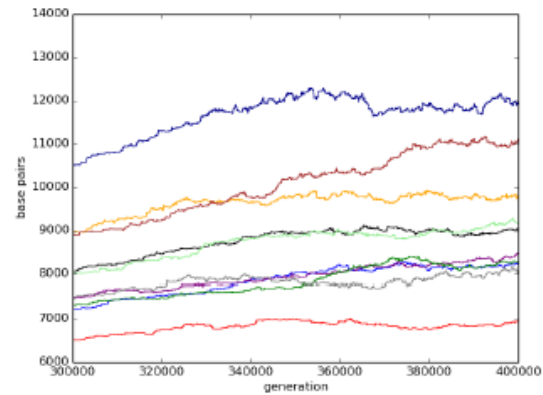
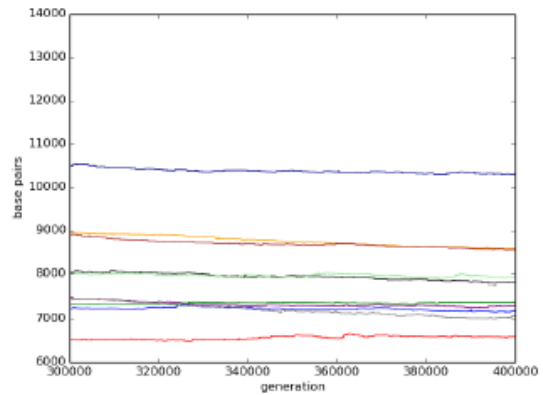


U-shape: evolved property AND ideal for evolution

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

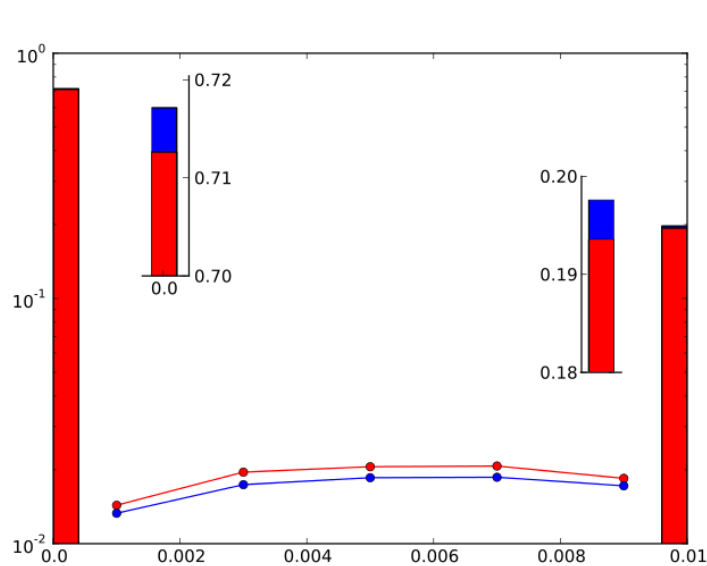


fitness: WT and Mutator



genome-size: WT and mutator

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

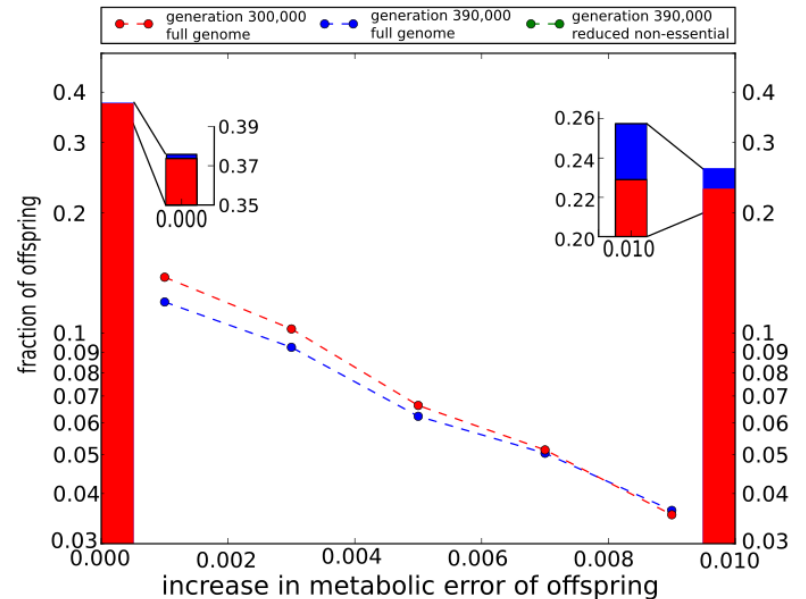


neutral

unfit

Wildtype

decrease genome size



neutral

unfit

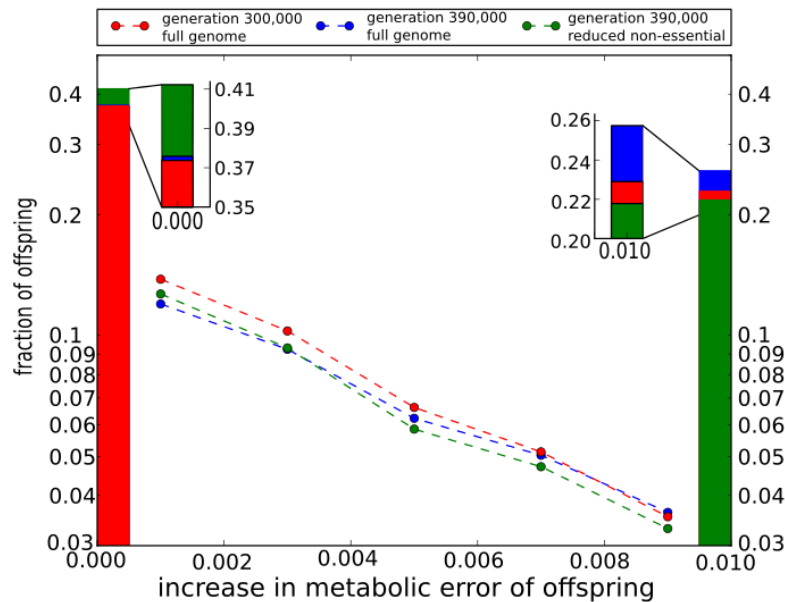
WT - Mutator ($\mu * 100$)

increase genome size

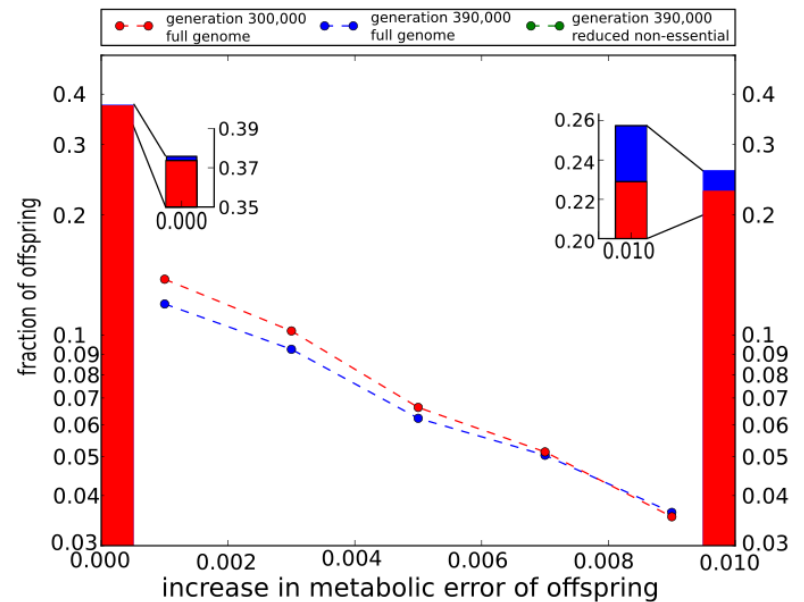
final fitness similar

U shape mutational profile and mutator strains

ancestor t=300.000 vs t=390.000



'cropped' mutator



wt - mutator ($\mu * 100$)

Increase of genome size
to increase deleterious mutations, to regain fitness

Conclusions Mutational Neighborhood

- U-shaped mutational neighborhood: high neutrality AND high selection
- Genome size and mutation rate:
high mutation rate: small genomes, overlapping genes (viruses)
Lower mutation rate: larger but compact genomes; operon, little overlap
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