Evolution of evolvability and/or regulation
Models ++

genome inflation and streamlining
e.g. in function optimization (de Boer & Hogeweg 2010)

Evolution of LISP function (genetic programming)

<table>
<thead>
<tr>
<th>m</th>
<th>Evolutionary Target</th>
<th>Minimal Coding Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) 13</td>
<td>$f(x, y) = x^3 + y^3 + 5x^2$</td>
<td>(+ (* (* (+ x 5) x) (* (* y y) y)))</td>
</tr>
</tbody>
</table>

25: $(- (* (y y) y) (* (- (- x x) x) (+ x (* (+ 3 (- 1 (- x x) x))) x))))$

21: $(- (* (y y) y) (* (- (- x x) x) (+ x (* (+ 3 (+ 1) x))) x)))$

19: $(- (* (y y) y) (* (- (- x x) x) (+ x (* x (+ 4 x))))))$

17: $(+ (* (y y) y) (* x (+ x (* x (+ 3 (+ x 1))))))$

15: $(+ (* (y y) y) (* x (+ x (* x (+ 4 x))))))$

high mutation rate prevents genome expansion and compromises evolvability
Results + (1)
occurrence of WGD after environmental shifts

- 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
Results ++ (2) Streamlining after WGD but “Irremediable complexity”

relatively small deletion events, many point mutations lead to adaptation
Which genes are retained/deleted? 
with or without WGD, with or without environmental switch

Intact ohnologs as a fraction of conserved WGD content
solid line: observed in evolution.
Dotted line when same number randomly deleted

their conserved binding

their out-degree
Summary

- Streamlining, but larger genomes after WGD: “irremediable complexity”
- TF preferential kept
- with high connectivity
- NO sub-functionalization
- adaptation by peripheral TFs
- dosage balance selection
Results ++ (3)

Short term evolution of fit evolved virtual cells
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

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 $\rightarrow$ neutrality; neutrality $\rightarrow$ 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 ++ (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)
(A) Genetic code

- 000: Start
- 001: Stop
- 100: M₀
- 101: M₁
- 010: W₀
- 011: W₁
- 110: H₀
- 111: H₁

(B) Gene translation

| RNA Seq. | 000 | 010 | 100 | 011 | 111 | 011 | 101 | 101 | 110 | 001 | ...
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein Seq.</td>
<td>Start</td>
<td>W₀</td>
<td>M₁</td>
<td>M₀</td>
<td>W₁</td>
<td>H₁</td>
<td>W₁</td>
<td>M₁</td>
<td>M₁</td>
<td>H₀</td>
</tr>
</tbody>
</table>

(C) Computation of the protein’s function

\[ M₁M₁M₁ \rightarrow 111_{\text{Gray}} \rightarrow 5 \]
\[ W₀W₁ \rightarrow 01_{\text{Gray}} \rightarrow 1 \]
\[ H₁H₀ \rightarrow 10_{\text{Gray}} \rightarrow 3 \]

\[ \text{MaxVal } = 2^3 - 1 = 7 \]
\[ \text{MaxVal } = 2^2 - 1 = 3 \]
\[ \text{MaxVal } = 2^2 - 1 = 3 \]

\[ M_{\text{range}} : [0:1] \]
\[ W_{\text{range}} : [0:W_{\text{max}}] \]
\[ H_{\text{range}} : [-1:1] \]

\[ M = 5/7 \]
\[ W = W_{\text{max}}^{1/3} \]
\[ H = (2x3/3) - 1 \]

(D) Graphical representation of the protein (with \( W_{\text{max}} = 0.01 \))
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 = $10^{-5}$ 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
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 increase 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
Note: average fitness of mutator population decreases
U shape mutational profile and mutator strains
ancestor t=300.000 vs t=390.000

ongoing “deepening’ of U shape; skewed toward deleterious after switch
Genome expansion and U shape mutational profile
Coding vs non coding sequences

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 (viruses) 
  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
Recurring evolutionary signatures
in structured (non-supervised) models

Genome inflation during “environmental” change
(however can be limited by
- cost (cf cell cycle)
- mutation rate (cf function optimization)

Evolution of mutational neighborhood
- U-shape: neutrality AND strong selection
- special - mutational “priming”

Mutation rate shapes coding structure - RNA world / function optimization / AeVOL

Convergence to known but ill understood phenomena
- e.g. prefential retention TF ’hubs’
Search images for further research
- no subfunctionalization of ohnologs, but adaptation by change of singles

How to make sense of observed (and thus evolved) coding structure
in the light of evolution