Coping with a variable environment: evolvability and vs regulation
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”

TODAY

Add metabolism (“virtual cell/ virtual microbe”) emergent properties of long term evolution (phylogeny)
++
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 Nature 2010
Gene loss as major evolutionary process

**Metazoa**
Loss of homeobox genes

**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: homeostasis (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 2017
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

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
Local landscapes, genome expansion and future fitness

<table>
<thead>
<tr>
<th>Duplications</th>
<th>Deletions</th>
</tr>
</thead>
<tbody>
<tr>
<td>t=1-100</td>
<td>t=101-200</td>
</tr>
<tr>
<td>t=1-100</td>
<td>t=101-200</td>
</tr>
<tr>
<td>ΔF</td>
<td>ΔF</td>
</tr>
<tr>
<td>+</td>
<td>=</td>
</tr>
<tr>
<td>(+)</td>
<td>=</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>&gt; 1.05</td>
<td>&gt; 1.05</td>
</tr>
<tr>
<td>.95 – 1.05</td>
<td>.95 – 1.05</td>
</tr>
<tr>
<td>&lt; .95</td>
<td>&lt; .95</td>
</tr>
<tr>
<td>Genome Size</td>
<td>Fitness</td>
</tr>
<tr>
<td>t=1-100</td>
<td>t=1-100</td>
</tr>
<tr>
<td>t=1-100</td>
<td>t=101-200</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>=</td>
<td>=</td>
</tr>
</tbody>
</table>
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
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 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

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

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.: ... RBS | 000 | 010 | 101 | 100 | 011 | 111 | 011 | 101 | 101 | 110 | 001 | ...
Protein Seq.: Start | M₀ | M₁ | M₀ | W₁ | H₁ | W₁ | M₁ | M₁ | H₀ | Stop |

(C) Computation of the protein’s function
\[ M₁M₁M₁ → 111_{\text{Gray}} → 5 \]
MaxVal = \( 2^3 - 1 = 7 \)
\[ W₀W₁ → 01_{\text{Gray}} → 1 \]
MaxVal = \( 2^2 - 1 = 3 \)
\[ H₁H₀ → 10_{\text{Gray}} → 3 \]
MaxVal = \( 2^2 - 1 = 3 \)

\[ M_{\text{range}} : [0:1] \]
\[ W_{\text{range}} : [0:W_{\text{max}}] \]
\[ H_{\text{range}} : [-1:1] \]
\[ M = 5/7 \]
\[ W = W_{\text{max}}^{1/3} \]
\[ H = (2 \times 3/3) - 1 \]

(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

neutral unfit neutral unfit

Wildtype WT - Mutator (\(\mu \times 100\))
decrease genome size increase genome size

final fitness similar

Rutten, Hogeweg & Beslon 2019
U shape mutational profile and mutator strains
ancestor t=300.000 vs t=390.000

Neutral unfit neutral unfit
'cropped' mutator

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

wt - mutator ($\mu \times 100$)
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/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