Evolution of coding structure

RNA world: sequence - structure- interactions
RNA genotype-phenotype mapping, summary

- “smoothness within ruggedness”
  single mutation can be neutral and can change ’everything’
- percolating and intercalating neutral networks
  *from smooth-rugged towards neutral networks*
- no local peaks: detours
- phenotypic vs genotypic information threshold
- diffusion on neutral networks (D prop.to $\lambda$)
- adaptive walk with majority of neutral mutations
- reconciliation neutral and adaptive evolution
- RNA landscape “ideal” for evolution
- Evolution of “robustness” (higher lambda) (“flattest”)
- Evolution of evolvability (iff innovation along NN)
Derived properties

**JUST RNA?**
or even just by wrongly computed (and 2 D) folding?

*percolating neutral path; innovations
evolution toward robustness*

NO......

similar (mutatis mutandis) properties in

Gene regulatory networks (A Wagner 2007a,b)

Protein folding A Wagner 2010

Metabolic networks (A. Wagner 2012)

see also books by A. Wagner
From paradigm systems to general conclusions

vs

Studying ‘‘all’’ cases

NK landscapes (Kauffman):
Class of models to study impact of GP mapping on evolutionary dynamics.

N: number of properties (e.g. sequence length)
K: number of ‘‘epistatic’’ interactions
most often 2 states per position

Fitness contribution of each $N.2^K$ states
chosen randomly. Fitness is sum of those

Calculate e.g. pathlength to local peak
height of optima reached (etc.)

NO percolating, intercalating neutral paths
and its evolutionary consequences

versions include neutrality .....
Neutrality and information accumulation (royal road)

information accumulation upto information threshold..
Genotype-phenotype mapping: Coding structure

3 questions/answers:

**Given code** – > which evolutionary dynamics?

eg RNA folding: punctuated evolution etc.

**Given problem** – > how to code?

expectation: smooth, non-redundant; found intertwining neutral paths

**Given evolutionary dynamics** – > which code?

towards robustness, hence evolvability
2 images of RNA world

individual complexity

ecosystem complexity

sequence to structure

replicator to wave-vesicle

X_{i+1} + X_i \rightarrow 2X_{i+1} + X_i
the RNA world

individual complexity

ecosystem complexity

RNA (without world)

world (without RNA)
Today: RNA in space

Themes

Structured based modeling

Individual and/or ecosystem based complexity
ecosystem diversification and mutation rate

Evolution of coding structures (cont)
multiple coding
mutational neighborhood

RNA even more evolvable than seen so far
RNA world: Preconceived networks vs evolving individuals, emerging species, emerging interactions

- **structured individuals**
  - here RNA sequences (+ and - strands)
  - if folding in predefined structure: replicase

- **no predefined target or fitness**

- **no predefined interactions**
  - but predefined reactions

**DO SPECIES/INTERACTION NETWORKS EVOLVE?**
**DOES EVOSYSTEM COMPLEXITY EVOLVE?**
genotype - phenotype - ecosystem mapping

feedback from higher levels to lower levels in evolving system

interacting RNA’s

Complex formation happens 5’-end → 3’-end

(“strong” altruism)

1. $X + Y \xrightleftharpoons[k_{-1}]{k_1} C_{X\sim Y}$ or $\xrightleftharpoons[k_{-2}]{k_2} C_{Y\sim X}$
2. $C_{X\sim Y} + \Phi \xrightarrow[\kappa]{\kappa} X + Y + Y^{-1}$
3. $X \xrightarrow[d]{d} \Phi$

only structure + reaction

no fitness function and no interaction predefined
Maximum mutation rates ($\mu = 0.015$): is only below information threshold for evolved coding structure ONE quasispecies

initial population dynamics with mutation

= after stopping mutation
High mutation rate ($\mu = .015$)
population structure of + strands

- Phylogeny reveals patterns in population of genotypes

- No clade patterns
- Population is supported by various genotypes
- One quasi-species

<table>
<thead>
<tr>
<th>Color</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyan</td>
<td>Catalyst</td>
</tr>
<tr>
<td>Red</td>
<td>Non-catalyst</td>
</tr>
</tbody>
</table>
High mutation rates ($\mu = .015$) sequence structure:
symmetry breaking: only $+$ strands catalytic

- Very high C frequency in 5’-end
- High G frequency in 3’-end
  - $\rightarrow$ many GC pairs
- Many interspersing U in 3’-end
  - $\rightarrow$ prevents base-pair formation \textit{in homo}
- No 5’-end in template strand
  - $\rightarrow$ prevents non-functional complex formation
Sequence is delicately tuned up

- Almost all base-pairs are GC
- Many other G and C that should not pair
  - → Difficult to form correct base-pairs
- High sequence conservation in all positions
  - Loop region must be tuned too
lowering mutation rates ($\mu - 0.13$): SPECIATION

$m = 0.014 \rightarrow m = 0.013$
lowering mutation rates: \( (\mu = 0.13) \) population structure

- Two quasi-species
  - distinct sequence classes
- Catalyst & Non-catalyst

<table>
<thead>
<tr>
<th>Color</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyan</td>
<td>Catalyst</td>
</tr>
<tr>
<td>Red</td>
<td>Non-catalyst</td>
</tr>
</tbody>
</table>
lowering mutation rates: $(\mu = .13)$ sequence structure

- **Catalyst**

- **Non-catalyst** → **Parasite**

- No catalytic structure in both strands
- Long 3’-end with many G
- No 5’-end in both strands
- Sequence conservation patterns & secondary structure
Parasite invades in periphery of QS

- Population of Sequences
- Genotype & Phenotype
- Space & Time

C-catalyst:

G-parasite:

Hamm. dist. from master sequence
Lower mutation rate $\mu = 0.008$: 3 quasispecies

Population of Sequences

Genotype & Phenotype

Space & Time

A catalyst: HIGH neutrality (ca 50%
Lower mutation rate $\mu = .004$: 4 quasispecies
evolved 4 species system; evolved interaction topology

ECOsystem ($\mu = 0$)  

EVOL. system ($\mu = .004$)

Direct Interaction structure

<table>
<thead>
<tr>
<th></th>
<th>C-catalyst</th>
<th>A-catalyst</th>
<th>G-parasite</th>
<th>U-parasite</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CYAN</td>
<td>MAGENTA</td>
<td>RED</td>
<td>GREEN</td>
</tr>
<tr>
<td>C-cat</td>
<td>0.52</td>
<td>0.87</td>
<td>0.36</td>
<td>0.45</td>
</tr>
<tr>
<td>A-cat</td>
<td>0.39</td>
<td>0.05</td>
<td>0.50</td>
<td>0.77</td>
</tr>
</tbody>
</table>
From Coding structure to ecosystem based information accumulation

Catalytic function (BC)

Catalyst → Parasite
Catalyst 2 → Parasite 2

Genotype → Phenotype → Ecosystem

sequence → structure → space

C-cat → A-cat
G-par → U-par
Conclusion

Very stable multi-(quasi)species systems evolves

Interaction topology different from anything studied before.

Variability increases with decreasing mutation rate speciation

Ecosystem based “solution” only at lower mutation rates

EVOLVED genotype-phenotype-interaction-spatial structure mutual dependent (and “make sense” in relation to each other)

Evolved, niche dependent mutational landscape
Evolution of coding structure at high mutation rates
Mutational neighborhood
High mut. rate: 1 quasispecies: mutations along line(s) of descent
High mut. rate: 1 quasispecies LOW variability

mutational neighborhood of master seq.: STEEP

Colizzi & H. 2014

black replicator; blue rest
High mut. rate: 1 quasispecies LOW variability

mutational NB: STEEP and “special”

EVOLVED optimal repl av. random
black replicator; blue rest; yellow parasites
1 quasispecies: codes for multiple functions

**mutational NB: STEEP and “special”**

**EVOLVED**
- optimal repl
- av. random

black repl.; blue rest; yellow parasites; green helpers; red stallers; gray junk
mutational neighborhood at larger Hamming distances

Top follow replicases with $\geq$ replic rates masterseq. bottom follow replicases with $<\,$ replic rates masterseq.
Abundance of functional types at Hamming distance to master sequence
quasispecies composition in field

weakly reflects mutational neighbourhood
more replicators (because of replication,
less helpers, more stallers (like neighborhood of other replicases)

Replicases with 'good' mutational neighborhood overrepresented.
Helpers “help”
change in junk $\rightarrow$ extinction
change in empty $\rightarrow$ extinction
in simplified ODE model:
increases max $\mu$ without parasites
decreases max $\mu$ with parasites
**Stallers “stall”**

change in junk $\rightarrow$ increases density

BUT master seq. replaced

’pseudo stallers’ evolve

change into empty space

parasite lineage evolves!

in simplified ODE model:

protects against parasites

with parasites: x-axis: fraction staller-mutants
### Variability of evolved quasispecies

<table>
<thead>
<tr>
<th></th>
<th>mut. nei.</th>
<th>$\mu_{max}$</th>
<th>replic. rates</th>
<th>competition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$+/-$</td>
<td>7</td>
</tr>
<tr>
<td><strong>Steep quasispecies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>0.0164</td>
<td>0.902 0.914</td>
<td>0.858 0.831</td>
<td>✓</td>
</tr>
<tr>
<td>2.</td>
<td>0.0154</td>
<td>1.000 0.932</td>
<td>0.878 0.854</td>
<td>✓</td>
</tr>
<tr>
<td>3.</td>
<td>0.0151</td>
<td>1.000 0.870</td>
<td>0.777 0.744</td>
<td>✓</td>
</tr>
<tr>
<td>4.</td>
<td>0.0145</td>
<td>1.000 0.866</td>
<td>0.817 0.777</td>
<td>✓</td>
</tr>
<tr>
<td>5.</td>
<td>0.0151</td>
<td>1.000 0.818</td>
<td>0.777 0.731</td>
<td>✓</td>
</tr>
<tr>
<td>6.</td>
<td>0.0143</td>
<td>1.000 0.858</td>
<td>0.777 0.729</td>
<td>x</td>
</tr>
<tr>
<td><strong>Flat quasispecies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>0.0154</td>
<td>0.725 0.892</td>
<td>0.817 0.798</td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>0.0140</td>
<td>0.902 0.872</td>
<td>0.817 0.792</td>
<td></td>
</tr>
</tbody>
</table>

**Steep quasispecies**

**Flat quasispecies**

*Random colors* ’ majority function’ - dist. from masterseq
Mutational neighborhood of 2 functionally equivalent RNA’s

(c) $\mu = 0.008$

- $S$ -
  - optimal replicator
  - random replicator
  - black replicator; yellow parasite; green helper; red staller