Non linear genotype-phenotype mapping: RNA landscapes
UPTO NOW

Classical Population Dynamic models
  +
  Space or vesicles
  emergent or predefines mesoscale pattern
  +
  invasion OR ongoing PHENOTYPIC Mutations
  (parameters of the model)

Who persists/invades; outcompete/outevolve;

multilevel evolution: replicators and “above”
**HOWEVER.......**

**what we did “wrong” so far**

“Defining property of biotic systems”:

**Very high dimensional genotype space**

**complex genotype - phenotype mapping**

Therefore use of phenotypic mutations in few dimensions not appropriate

*no RNA in RNA world*

*RNA-like Replicators dimensionless points without physical/chemical properties*
Constructive Darwinian Evolution

Darwinian evolution as efficient design (optimization) tool

*Genetic Algorithms (Holland), evolutionary computation*

- population of (coded) structures/solutions/’cases’
- mutational operators
- fitness criterion
- reproduce/decay according to fitness

E.g.

- computer network design/ job scheduling
- robotic control / body design
  - **nano technology**

in vitro RNA evolution - ligase
search space ca $4^{120}$ - pop size $10^{10}$

**WHY do we find it??**

All positions essential!
Landscape important in hill climbing (and evolution of finite population(?))

\[ \text{genotype} \rightarrow \text{phenotype} \rightarrow \text{fitness} \]

Fitness Landscape

- smooth - no epistasis
- rugged epistasis

Coding structure!
RNA secondary structure as paradigm for 'natural' coding structure
genotype-phenotype map

- computable 'natural' genotype-phenotype map
- RNA world
- in vitro evolution efficient

**assume fitness depends on distance to predefined target sec. structure**

Fontana, Schuster, Hoffacker, Ancel, Flamm etc. (Vienna)
Huynen, van Nimwegen, Takeuchi, Hogeweg (Utrecht)
RNA Structure (tRNA-phe yeast)

Sequence (2D)

2D Structure

3D Structure
Computation of RNA genotype-phenotype mapping

- Secondary structure is planer

(Zuker & Stiegler '81!)

- Base pairs
\[
\begin{align*}
e(C, G) &= -3 \\
e(A, U) &= -2 \\
e(G, U) &= -1 \\
e(\text{others}) &= 0
\end{align*}
\]

- Minimum energy

If \( j - i < 4 \) (i.e. length < 5),
\[
E_{i,j} = 0
\]

Otherwise,
\[
E_{i,j} = \min \begin{cases} 
E_{i+1,j}, E_{i,j-1} & \text{(1)} \\
e(i,j) + E_{i+1,j-1} & \text{(2)} \\
\min_{i' = i+1}^{j-1}(E_{i,i'} + E_{i'+1,j}) & \text{(3)}
\end{cases}
\]
RNA secondary structure as paradigm for genotype-phenotype mapping computable??

16SRNA
Min. Energy folding vs Conserved folding

23S RNA

A. globiformis, Anabaena sp., A. tumefaciens, B. japonicum, E. coli B. subtilis, T. thermoph, Pir. marina, Rb. sphaero

RNA-landscape: multi-one genotype-phenotype mapping

Almost all sequences fold in 'typical shape'

but

Only small fraction of shapes is typical

Example: GC strings length 30:

1.07 \times 10^9 sequences
218830 shapes
22718 "typical"
93.4% seqs in typical shape


Nevertheless...

GCAU seqs. length 70: 999919 different structures in 1 million seqs
RNA Landscape: average phenotypic change by mutations
(local) RNA Landscape
relation distance in genotype and distance in phenotype

• single mutation: often NO change
ca 30% for length 70; saturates at 20% for longer seq.
• single mutation: sometimes NO similarity (max. distance)
• distance distribution of phenotypes independent of genotype distance for moderate to large genotype distances (small correlation length)

RUGGED

True for different measures of phenotypic distance
Hamming distance on string representation
# bond changes
Folding of Eukaryotic mRNA: major change by point mutation (5' vs 3' end)
2 different functional ribozymes

1 point mutation NO similarity in secondary structure

EA Schultes, DP Bartel - Science
RNA landscape: evolutionary consequences

shape of landscape important because of finite (localised) populaiton

- Rugged - small correlation length
- identical structures overrepresented 'closeby'
- single mutation can lead to complete change of structure

→ Stuck at local optima?...NO.....
Evolutionary dynamics of random RNA to prespecified target secondary structure
Rugged fitness landscape

Evolution “stuck on local optima??”

NO......
DETOURS!
Figure 10. Percolation of sequence space by neutral networks. A neutral path connects sequences of Hamming distance $h = 1$ (single base exchange) or $h = 2$ (base pair exchange) that fold into identical minimum free energy structures. The sketch shows a neutral path of length $h = 9$. The path ends because no identical structure was found with $h = 10$ and $h = 11$ from the reference.
Neutral Paths (Schuster and Fontana, 1994) typical shapes percolate through shape space

**Fig. 4:** Neutral paths. A neutral path is defined by a series of nearest neighbour sequences that fold into identical structures.
RNA landscape, neutral paths, information threshold

Error/ Information threshold (as defined):

\[ Q > \sigma^{-1} \]

\[ L < ln(\sigma)/(1 - q) \]

\(-- > L <= 0 \) if mutant has same fitness (phenotype)

\( == \) Genotypic information threshold

cf Phenotypic information threshold

\[ L < ln(\sigma)/((1 - q)(1 - \lambda)) \]

Takeuchi and H. 2005
Above the (genotypic) information threshold (?) (Adaptive vs) Neutral Evolution (neutral drift) (cf Kimura, theory of neutral molecular evolution)

In FLAT landscape: Diffusion through genotype space (Kimura):

\[ D = \frac{5ApL}{3 + 4pN} \]

A replication rate, \( p \) mutation rate, \( L \) length, \( N \) pop. size

On neutral network \( D' = \lambda D \)
evolution over neutral network is diffusion-like process

measured diffusion in RNA landscape (in target structure)
Higgs and Derrida: for finite populations
“speciation” in flat landscape

Fig. 1. Distribution of the elements of the matrix $T^{a\beta}$ in the OPM for a population of $M = 1000$ individuals. The distribution is shown at six times for the same population. There is a period of 50 generations between each successive pair of curves; therefore the peaks move a distance 50 to the right each time. Peaks fluctuate in size and eventually disappear.

Fig. 2. Schematic representation of the genealogical OPM showing ultrametric property of the branching times $T^{\alpha\gamma}, T^{\beta\gamma}$. Cutting the tree at an arbitrary point in the past the population into families.
"punctuated evolution" ("epochal evolution")
Punctuated evolutionary dynamics

(vs “new synthesis” vs Gould)

- external environment change???
- “waiting for unlikely mutation”
  stuck on local optimum
- ecological equilibrium
  stable spatial patterns
- *phenotypic punctuated equilibria*
  stasis while on neutral path
Evolutionary dynamics: population structure

Fig. 3. Evolutionary optimization. A flow reactor with capacity $N = 1000$ is initialized with that many copies of a random sequence of length $r = 76$. The mutation rate is $\rho = 0.001$ and the target secondary structure is the tRNA$^{\text{Phe}}$ cloverleaf. The replication rate function is $A(d) = 1.06^{0.8-d}$, where $d$ is the tree-edit distance (9) to the target structure. The population average of the distance to the target is plotted against time (solid line) for a specific interval of the entire run (Inset). Superimposed series of dots render the evolution of the population structure over time. Dots at one time epoch are a one-dimensional projection (see Fig. 2 legend) of the population of sequences present in $>10$ copies at that time. Collecting all time slices yields a unique glimpse of the cluster dynamics. The same qualitative picture of punctuated equilibria occurs with all parameter settings and random target structures we tried for both linear and exponential fitness functions $A(d)$.
Fig. 2. Population structure in sequence space. The support of a population in sequence space is the set of sequences present in at least one copy. The population support can be pictured in two dimensions using some theorems from distance geometry (27). We compute the metric matrix M with entries $m_{ij} = (d_{ij} + d_{k} - d_{k}^2)/2$, where $d_{ij}$ is the Hamming distance between sequences $i$ and $j$ and $0$ is the center of mass of the support. Sequences are expressed in principal axes coordinates by diagonalizing $M$. Only the components corresponding to the largest two eigenvalues are kept, yielding a projection onto the plane that captures most of the variation. Dots represent a static snapshot of $N = 2000$ individuals after 135 time units replicating with $p = 0.002$. Among the 2000 individuals, 631 are different and among them 301 fold into different structures. To help correct for the distortions of the projection, the dots are connected by the edges of the minimum spanning tree. Edges connect closest points. Red (blue), Hamming distance less (more) than 6; dot size large (small), more (less) than four copies in the population; yellow (green), sequences that do (do not) fold into the tRNA target structure.
Novelty "seen" along the neutral path (Huynen 1998)

Figure 1: Perpetual Innovation along the Neutral Net.

Innovations’ Shadow of similar structures along neutral paths

Figure 2: Conservation along the neutral net.

Zuckerhandl "Neutral + adaptationist evolution reconciled"
(Kimura memorial lecture)
Shape of RNA fitness landscape percolating and intertwining Neutral Networks:

RNA Genotype - Phenotype mapping Ideal for evolution
(Schuster and Fontana, 1994)

**Fig. 5:** A sketch of the mapping from sequences into RNA secondary structures as derived here. Any random sequence is surrounded by a ball in sequence space which contains sequences folding into (almost) all common structures. The radius of this ball is much smaller than the dimension of sequence space.
MOREOVER: phenotype $\rightarrow$ function mapping

Alternative ligases (Ekland et al 1995)

'tyranny' of small motifs... or complex structures?
'drift' on neutral network not 'neutral':
(1) Longterm RNA evolution: fitness of mutants
(2) Evolution towards high lambda
redundant genotype-phenotype mapping: choice of coding

• Evolution towards 'flatter parts'
  == Mutational robustnes
  == high connectivity of neutral network
  == MAX EIGENVECTOR OF CONNECTION MATRIX (van Nimwegen 2000)

compare blind ant (moves with prob. rel neutral NB)
  -- > same freq in each node)

myopic ant (moves with fixed probability)
  -- > $D = \hat{d} + Var(d)/\hat{d}$
Evolution towards mutational robustness

== largest eigenvalue of connection matrix

van Nimwegen et al PNAS 1999

walk along neutral path not neutral
walk along neutral path not neutral.....
how neutral is neutral....
walk along neutral path not neutral.....
how neutral is neutral

neutral if above the information threshold!
Implications evolution towards higher robustness

• more robustness → more exploration ($D \lambda$)

• evolution of evolvability

high mutational load of recently evolved strains well known from traditional evolutionary experiments (Scharloo 1999: canalization)
Hsp90, 582-590 Effect on growth-rate of single point mutations from wild-type and from 7 (almost) neutral mutations

A systematic survey of an intragenic epistatic landscape Claudia Bank et al MBE 2014
Robustness, population diversity and evolutionary optimization

AVIDA: Self-replicating computer program (Adami et al)

Population variability per position (gene) $p_i \log(p_i)$