Non linear genotype-phenotype mapping: evolution on RNA landscapes
RNA landscape: evolutionary consequences

- 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!
Percolation of sequence space by neutral networks (Schuster)

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.
Error/Information threshold (as defined):

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

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

\[ \Rightarrow L \leq 0 \text{ if mutant has same fitness (phenotype)} \]

\[ \Rightarrow \text{Genotypic information threshold} \]

\[ \text{cf Phenotypic information threshold} \]

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

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 $T^{\alpha\alpha}, T^{\beta\beta}$. 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 environmental change???
- “waiting for unlikely mutation”
  stuck on local optimum
- ecological quilibrium
  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 $A(d) = 1.06^{0.50 - d}$, 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 ($Insert$). 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 $A(d)$. 
Population Structure: landscape sampling

**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. We compute the metric matrix $M$ with entries $m_{ij} = (d_{ij} + d_{ii} - d_{jj})/2$, 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 $r = 0.002$. 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 -> 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

1-q=0.039

- Total
- Master
- Mutant

\[ \lambda \]

Time
redundant genotype-phenotype mapping: choice of coding

- Evolution towards 'flatter parts'
  - Mutational robustness
  - High connectivity of neutral network
  - MAX EIGENVECTOR OF CONNECTION MATRIX
  - $D = \text{Max eigen value}$
  (van Nimwegen 2000)

compare blind ant (moves with prob. rel neutral NB)
$\rightarrow$ same freq in each node

myopic ant (moves with fixed probability)
$\rightarrow D = \hat{d} + \text{Var}(d)/\hat{d}$
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

v neutral if above the information threshold!
Experimental determination of tRNA GP map (23,284 muts robustness vs fitness: many more fit mutational neighbors)

However WT very robust

Fitness Landscape Analysis of a tRNA Gene Reveals that the Wild Type Allele is Sub-optimal, Yet Mutationally Robust Gabzi, Pilpel, Friedlander, MBE 2022
example of intra-molecular evolved landscape
negative epistasis

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) $p_i \log(p_i)$
Neutrality and information accumulation (royal road)

information accumulation upto information threshold..
Implications evolution towards higher robustness

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

• evolution of evolvability *at level of population*

high mutational load of recently evolved strains well known from traditional evolutionary experiments (Scharloo 1999: canalization)
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 lamda) (“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

Manrubia et al From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics Physics of life 2021
Empirical derived GP map of yeast promoter sequences (selection/robustness/evolvability)
Vaishnavi .... Aviv Regev 2022, Nature

- Use deep neural network model to learn GP map
  (= promoter sequence to gene expression in Yeast (in 2 media))
- Teaching set: promoter - measured expression pairs
- Test set: random promoters - predicted expression - measured expression in yeast
- Correlation: Pearson’s $r = 0.960$, $P < 5 \times 10^{-324}$, $n = 61,150$;
Use now calculatable GP map to characterize GP landscape

- **designing promotor sequence for desired expression level**
  (using evolutionary search)
  *only few mutation needed from arbitrary sequence
  but harder for differential expression*

- **determining selection pressure on each promotor**
  compare expression levels of natural occurring orthologs
  (5,569 S. cerevisiae genes
  using the natural variation observed
  across over 4.73 million orthologous
  promoter sequences from the 1,011
  S. cerevisiae isolates)
  with random mutations of the promotor
  $\sigma_c$ expression variation
  of random mutational NB
  $\sigma_b$ expression variation
  of orthologs
• Robustness defined as number of (near) neutral mutants of natural promotor sequences \((\lambda\) as defined before)

• correlates with selection regime and conservation at various levels

"Thus, genes with expression levels that are under stabilizing selection have regulatory sequences that tend to be more robust to the effects of mutations, which may reflect their history and constrain their future"
Alternatively: evolvability (malleability) distribution of expression changes in mutational neighborhood

linear phenotype space

map in archetype space

robustness

expression

density
Mutational neighborhood of mutants of malleable promoter left (DBP7) and of robust promotors (right)

evolvability at the individual level alternative functionality ”nearby”