Evolution at Molecular Resolution

Peter Schuster

Institut für Theoretische Chemie, Universität Wien, Austria
and
The Santa Fe Institute, Santa Fe, New Mexico, USA

EMBO Members‘ Meeting 2014

Heidelberg, 29.– 31.10.2014
Web-Page for further information:

http://www.tbi.univie.ac.at/~pks
Sewall Wrights fitness landscape as metaphor for Darwinian evolution

Sewall Wright. 1932. *The roles of mutation, inbreeding, crossbreeding and selection in evolution.*

Sewall Wright, 1889 - 1988

Sewall Wrights fitness landscape as metaphor for Darwinian evolution
The multiplicity of gene replacements with two alleles on each locus

\( Q_5 \): the space of binary sequences of chain length \( l = 5 \)

Binary sequences are encoded by their decimal equivalents:

- \( C = 0 \) and \( G = 1 \), for example,
  - "0" \( \equiv 00000 = CCCCC \),
  - "14" \( \equiv 01110 = CGGGC \),
  - "29" \( \equiv 11101 = GGGCG \), etc.
Fitness landscapes became experimentally accessible!


Evolution as a global phenomenon in genotype space

\[
\Phi: (Q, d_H) \Rightarrow (S, d_S) \\
\Psi: (S, d_S) \Rightarrow \mathbb{R}_+ \\
X \rightarrow S = \Phi(X) \rightarrow f = \Psi(S)
\]

sequence space \( Q \)

shape space \( S \)

parameter space \( \mathbb{R}_+ \)

genotype \( \rightarrow \) phenotype \( \rightarrow \) selection

Evolution as a global phenomenon in genotype space
The simplified model
\[
\frac{dx_j}{dt} = \sum_{i=1}^{n} W_{ji} x_i - x_j \Phi; \quad j = 1, 2, \ldots, n
\]

\[
W_{ji} = Q_{ji} \cdot f_i, \quad \sum_{i=1}^{n} x_i = 1, \quad \Phi = \sum_{i=1}^{n} f_i x_i
\]

**fitness landscape**

Mutation and (correct) replication as parallel chemical reactions

M. Eigen. 1971. *Naturwissenschaften* 58:465,
The error threshold in replication and mutation

Mutation rate and genome size
Master sequence

\[ p = 0 \]
$p = 0.25 \ p_{cr}$
$p = 0.50 \ p_{cr}$

Master sequence

Mutant cloud
$p = 0.75 \, p_{cr}$

Master sequence

Mutant cloud
$p > 1.00 \ p_{cr}$

Mutant cloud
Model fitness landscapes I

single peak landscape

step linear landscape
Error threshold on the single peak landscape
The linear fitness landscape shows no error threshold
AGCUUUACUUAGUCGCU

1 A-U
1 A-G
1 A-C
\[ f(S_j) = f_n + 2d (f_0 - f_n) \left( \eta_{j}^{(s)} - 0.5 \right) \]

\( j = 1, 2, \ldots, N; j \neq m \)

\( \eta \) … random number

\( s \) … seeds

![Graph showing rugged fitness landscapes](image)

„realistic“ landscape

Rugged fitness landscapes over individual binary sequences with \( n = 10 \)
Quasispecies with increasing random scatter $d$

Error threshold: Individual sequences

$n = 10$, $\sigma = 2$, $s = 491$ and $d = 0, 0.5, 0.9375$
Choice of random scatter:

\[ s = 637 \]

Error threshold on ’realistic‘ landscapes

\[ n = 10, \ f_0 = 1.1, \ f_n = 1.0, \ s = 637 \]
Choice of random scatter:

\[ s = 919 \]

Error threshold on 'realistic' landscapes

\[ n = 10, \ f_0 = 1.1, \ f_n = 1.0, \ s = 919 \]
Determination of the dominant mutation flow: $d = 1, s = 613$
Determination of the dominant mutation flow: \( d = 1, s = 919 \)
Predictions of the strong quasispecies concept

1. A strong quasispecies is dominated by a clan of mutationally coupled closely related sequences.

2. A four-membered clan consists of the master sequence being the fittest sequence, its fittest one error mutant, the fittest two-error mutant that has to lie in the one-error neighborhood of the fittest one-error mutant, and the fourth sequence completing the mutationally coupled quartet.

3. Strong quasispecies reproduce more efficiently, are stable to mutation, and should be favored by evolution.
Motoo Kimura’s population genetics of neutral evolution.


Pairs of neutral sequences in replication networks

Random fixation in the sense of Motoo Kimura


\[
\begin{align*}
\lim_{p \to 0} x_1(p) &= x_2(p) = 0.5 \\
\lim_{p \to 0} x_1(p) &= \alpha/(1 + \alpha) \\
\lim_{p \to 0} x_2(p) &= 1/(1 + \alpha) \\
\end{align*}
\]

\[
\begin{align*}
d_H &= 1 \\
d_H &= 2 \\
d_H &\geq 3
\end{align*}
\]
A fitness landscape including neutrality
Neutral network: Individual sequences

\[ n = 10, \sigma = 1.1, d = 1.0 \]
Neutral network: Individual sequences

\[ n = 10, \sigma = 1.1, d = 1.0 \]
Consensus sequences of a quasispecies of two strongly coupled sequences of Hamming distance $d_H(X_i,X_j) = 1$ and 2.

Consensus sequence: ACAUGCGAA
Conclusions

1. Realistic fitness landscapes sustain error thresholds.

2. Quasispecies may be centered around clans of sequences with high fitness, which provide evolutionary stability against increasing mutation rates.

3. Pairs of neutral sequences with Hamming distances one or two form clans and are not subjected to Kimura’s random selection.
Coworkers

Ivo L. Hofacker, Christoph Flamm, Universität Wien, AT

Peter Stadler, Universität Leipzig, DE

Walter Fontana, Harvard Medical School, MA

Christian Reidys, University of Southern Denmark, Odense, DK

Thomas Wiehe, Universität Köln, DE

Martin Nowak, Harvard University, MA

Stefan Bonhoeffer, ETH Zürich, CH

Christian Forst, Southwestern Medical Center, University of Texas, Dallas, TX

Erich Bornberg-Bauer, Münster, DE
Acknowledgement of support

Fonds zur Förderung der wissenschaftlichen Forschung (FWF)
Projects No. 09942, 10578, 11065, 13093
13887, and 14898

Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF)
Project No. Mat05

Jubiläumsfonds der Österreichischen Nationalbank
Project No. Nat-7813

European Commission: Contracts No. 98-0189, 12835 (NEST)

Austrian Genome Research Program – GEN-AU: Bioinformatics
Network (BIN)

Österreichische Akademie der Wissenschaften

Siemens AG, Austria

Universität Wien and the Santa Fe Institute
Thank you for your attention!
Web-Page for further information:

http://www.tbi.univie.ac.at/~pks

Preprint:

Santa Fe Institute Working Paper: # 12-06-12