Fitness landscapes and seascapes

Michael Lässig

Institute for Theoretical Physics
University of Cologne
Thanks

- **Ville Mustonen:**
  Cross-species analysis of bacterial promoters,
  Nonequilibrium evolution of Drosophila

- **Johannes Berg, Stana Willmann**
  Adaptive evolution of transcription factor binding sites
Fitness effects govern the tempo of functional changes. Example: promoter evolution.
Selection takes place in a stochastic context

- **Selection**: deterministic frequency change of alleles with fitness difference $\Delta F$
  characteristic time: $1/ \Delta F$

- **Genetic drift**: stochastic change of allele frequencies
  characteristic time: $N$

- **Mutations**: random nucleotide changes in individuals
  characteristic time: $1/ \mu$

- **Evolutionary regimes**
  - $\mu N \ll 1$: substitutions, polymorphisms rare (Kimura-Ohta)
  - $\mu N \gg 1$: polymorphic population (Eigen quasispecies)
Substitution rates

- **Kimura-Ohta substitution rates** differ from individual mutation rates

\[
    u_{a \rightarrow b} = \mu_{a \rightarrow b} N \frac{1 - \exp[-2(F(b) - F(a))]}{1 - \exp[-2N(F(b) - F(a))]} 
\]

- **Ratio of forward and backward rates:**

\[
    \frac{u_{a \rightarrow b}}{u_{b \rightarrow a}} = \frac{\mu_{a \rightarrow b}}{\mu_{b \rightarrow a}} \exp[2N(F(b) - F(a))] 
\]
Fitness scales

- Fitness effects vary by orders of magnitude

  - lethal
  - deleterious
  - near-neutral
  - advantageous

  ![Fitness scales diagram](image)

- Experiments and genome statistics probe different fitness scales
Evolutionary equilibrium

- Under *neutral evolution or constant selection*, the frequency distribution reaches an *evolutionary equilibrium*, \( p_{eq}(x) \).

- **Detailed balance:**
  the numbers of deleterious and advantageous substitutions are equal.

- **Theoretical treatment:** allele frequency distribution \( p(x,t) \),
  diffusion approximation (Kimura).
Evolutionary equilibria in sequence space and fitness landscapes

- Given two families of loci,
  - background loci under neutral evolution with stationary distribution $P_0(a)$
  - functional loci under selection with stationary distribution $Q(a)$

the equilibrium fitness landscape $F(a)$ of functional loci is

$$2N F(a) = \log \left( \frac{Q(a)}{P_0(a)} \right) + \text{const.}$$

Measuring genomic fitness landscapes

- Example:
  CRP transcription factor binding sites in E. coli

- Binding energy $E$: molecular phenotype

- Fitness function
  \[ 2NF(E) = \log\left[\frac{Q(E)}{P_0(E)}\right] \]

- Functional sites are under substantial selection pressure:
  \[ 2N \Delta F \approx 10 \]

(see Mustonen's talk)
Nonequilibrium evolution and adaptation

- Under **time-dependent selection**, the allele frequency distribution reaches a **nonequilibrium stationary state** $p_{\text{stat}}(x)$.

- **Adaptation**: excess number of advantageous over deleterious **substitutions**.

- Adaptation is the **response** to new selection pressures.

- **Theoretical treatment**: models of fluctuating selection. Micro-evolutionary fluctuations $\rightarrow$ modified diffusion equations [Wright, Kimura, Ohta, Gillespie, …].
Inference of adaptation from genomic data

- From **increased rate of nonsynonymous over synonymous substitutions**: $u > \mu$ indicates adaptation (converse not true!).

- From the **spectrum of polymorphisms** (Tajima's D test and variants): increased fraction of high-number polymorphisms is consistent with adaptation.

- From the **relative weight of polymorphisms and substitutions** (McDonald-Kreitman test):
  \[
  \frac{\text{subst}}{\text{poly}} > \frac{\text{subst}_0}{\text{poly}_0} \approx \frac{\mu_0 t}{\mu_0 N}
  \]
  is consistent with adaptation.

Likelihood methods based on MK-test. [Bustamante et al., Nature 2002]

Confounding factors

- **Simultaneous presence of positive and negative selection.**

- **Partial linkage between loci.**

- **Demography** (population growth).

- **Lack of neutral reference** for intergenic regions, insertions/deletions (see Lunter's and Grün's talk), expression data.

- **Lack of model + solution for adaptive evolution.**

- **Lack of coherent scoring** for polymorphism spectra and substitutions.
Model-based inference of adaptation

- Phenotypic concept of Darwinian selection is dynamic: **newly arising** selection induces adaptation.

- Can we trace the **time-dependence of selection** together with the **adaptive response** in genomic data?

- Model for **fluctuating selection**:
  \[ \Delta F (t) = f \chi(t), \quad \chi(t) = \pm 1 \text{ with switching rate } \gamma. \]

- **Theoretical treatment** for macro-evolutionary fluctuations (\( \gamma << 1/N \)):
  quasistationary approximation for \( p(x,t) \),
  averaging over selection fluctuations (quenched disorder).
Bioinformatics of cross- and intra-species comparisons

- **Alignment** of multiple ingroup and outgroup sequences:

  AACTGTCCACGTTCCTTCCGATGCAGCCTGA
  AACTGTCCACGTTCCTTCCGATGCAGCCTGA
  AACTGTCCACGTTCCTTCCGATGCAGCCTGA
  AACTGTCCACGTTCCTTCCGATGCAGCCTGA
  AACTGTCCACGTTCCTTCCGATGCAGCCTGA
  AACTGTCCACGTTCCTTCCGATGCAGCCTGA
  AACTGTCCACGTTCCTTCCGATGCAGCCTGA
  AACTGTCCACGTTCCTTCCGATGCAGCCTGA

  AACTATCCACGTTCCTTCCGATGCAGTCTGA

  \[ \begin{array}{c}
  \text{ingroup sequences} \\
  \text{(m = 8)}
  \end{array} \]

  \[
  \begin{array}{c}
  \text{outgroup sequence(s)} \\
  \text{frequency count sequence } k_i
  \end{array}
  \]
Inference of adaptive evolution

- **Joint model-based scoring** of polymorphisms and substitutions:

  \[
  S(k) = \frac{Q(k \mid \mu t, N\mu, Nf, \gamma/\mu)}{P_0(k \mid \mu t, N\mu)}
  \]

  [V.Mustonen, M.L., PNAS 2007]

  → Bayesian inference of evolutionary parameters.

- Sensitivity gain compared to McDonald-Kreitman test: 
  fraction of adaptive substitutions \( \alpha \)

  - is correctly reproduced by model-based inference
  - is underestimated by MK estimate

  \[
  \alpha_{MK} = 1 - \frac{\text{subst}_0}{\text{poly}_0} / \frac{\text{subst}}{\text{poly}}
  \]

  [Smith and Eyre-Walker, Nature 2002]
Polymorphism-substitution spectra in *Drosophila*

- **Data:** 12 genomes from *D. melanogaster* aligned to outgroup *D. simulans.*
  

- **Frequency distribution** $Q(k)$ of polymorphisms and substitutions:

![Graphs showing frequency distribution](image)

- best equilibrium model ($N = \text{const}, \gamma = 0$):
  - poor fits except for four-fold synonymous changes

- best demographic model (time-dependent $N$, $\gamma = 0$):

- best fluctuating-selection model:
  - explains data in all categories

[V.Mustonen, M.L., 2007]
Non-equilibrium evolution in *Drosophila*

- **Selection is stronger than previously expected:**
  

- **Evolution is far from equilibrium:**
  selection switches occur at nearly the rate of neutral point mutations.

<table>
<thead>
<tr>
<th></th>
<th>2Nf</th>
<th>(\gamma/\mu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonsynonymous</td>
<td>115</td>
<td>0.1</td>
</tr>
<tr>
<td>Intergenic</td>
<td>17</td>
<td>0.3</td>
</tr>
<tr>
<td>Introns</td>
<td>14</td>
<td>0.4</td>
</tr>
<tr>
<td>UTRs:</td>
<td>40</td>
<td>0.4</td>
</tr>
</tbody>
</table>

- This ensures the **efficiency of adaptations:**

  \[ \frac{u}{\gamma} \sim 10 - 100. \]
In non-coding DNA, selection switches occur more frequently than for proteins.

⇒ Prediction of functional elements can less be based on sequence conservation.
Fitness seascapes

- **What drives the waves?**
  - **systems component:** correlations (epistasis) cause compensatory mutations.
  - **external component:** time-dependent environment
  - **genomic component:** linkage to other loci

- **Nonequilibrium + correlations:**
  one external change can trigger an avalanche of responses,
Conclusions

- **Equilibrium fitness landscapes can be measured** as a log-likelihood score:
  \[ S = 2N \Delta F + \text{const.} \]

- **Adaptations are a non-equilibrium response** to time-dependent selection.

- Substantial parts of the *Drosophila* genome evolve
  - far from equilibrium \((\gamma/\mu > 0.1)\)
  - far from neutrality \((2Nf > 10)\).

- Strong selection \(\rightarrow\) **efficiency of adaptations** as evolutionary mode.

- Selection is not only a **constraint**
  but also a major **driving force** of genomic change.

- **Non-equilibrium + epistasis:**
  one external change can cause an avalanche of responses.

- Evolutionary systems biology is largely uncharted territory.