The arrow of time in evolutionary biology

Michael Lässig

Institute for Theoretical Physics
University of Cologne
Thanks

Ville Mustonen (Sanger Institute)
Natalja Strelkowa
Stephane Ghozzi
Marta Luksza

Curtis Callan (Princeton)
Justin Kinney (Princeton)

SFB TR 12
SFB 680
European Research Training Network STIPCO
Life is irreversible at multiple scales

- **Cellular processes**
  seconds – minutes
  individual cells

- **Development / Ageing**
  1 generation
  individual organisms

- **Evolution**
  many generations
  population
**Darwinian evolution and adaptation**

- **Adaptative evolution** of phenotypes in a population occurs due to **natural variation** and **natural selection**.

- Adaptive evolution is an ongoing process, because **selection pressures keep changing**.
Forces of molecular evolution

- The population fractions (frequencies)
  \[ x = (x^1, \ldots, x^k) \]

of phenotypes or genotypes in a population change by

1. selection
2. mutations
3. reproductive fluctuations (genetic drift)
1. Evolution and irreversibility
   (Classical views and questions)
Adaptative evolution

- Evolution in a **fitness landscape** (S. Wright 1932):
  - *interplay of selection and genetic drift*

\[ s(x) = \nabla F(x) \]

- Example: fitness landscape in the fungus *Aspergillus niger*

[A. de Visser, SC. Park, J. Krug 09]
Adaptative evolution

- **Fundamental Theorem of Natural Selection** (R.A. Fisher 1930):

  - deterministic evolution under *time-independent* selection

    \[
    \frac{d}{dt} F(t) = s^2(t)
    \]

  - deterministic evolution under *time-dependent* selection

    \[
    \frac{d}{dt} \Phi(t) = s^2(t)
    \]

    *fitness flux*
Is there an entropy principle of biological evolution?
(Schrödinger, *What Is Life* 1943)
Adaptative evolution

- Stochastic evolution in fitness seascapes

V. Mustonen, M.L., March 2009
2. **Fitness flux increases**  
   (A fluctuation theorem on evolution)
A population history is a sequence of frequency measurements
\[ x = (x_0, \ldots, x_n) \text{ at times } (t_0, \ldots, t_n). \]

The fitness flux of a population history is the cumulative selective effect of frequency changes:
\[ \Phi(x) \equiv \sum_{i=1}^{n} \Delta x_i s(x_i, t_i). \]

In a fitness landscape, the fitness flux equals the fitness difference between initial and final state.

\[ s(x) = \nabla F(x) \]
Population histories and fitness flux

- **A population history** is a sequence of frequency measurements
  \[ x = (x_0, \ldots, x_n) \text{ at times } (t_0, \ldots, t_n). \]

- The **fitness flux** of a population history is the cumulative **selective effect of frequency changes**:
  \[ \Phi(x) \equiv \sum_{i=1}^{n} \Delta x_i s(x_i, t_i). \]

- In a **fitness seascape**, the fitness flux does **not** equal the fitness difference between initial and final state.
Fitness flux theorem

- For a large class of processes with mutations, genetic drift, and selection, the probabilities of forward and reverse history are related:

\[ P(x^T) = P(x) e^{-N\Phi(x)} + \Delta H(x) \]

\( \text{fitness flux} \quad \text{entropy difference of initial conditions} \)

- **Fitness flux theorem**

\[ \left\langle e^{-N\Phi + \Delta H} \right\rangle = 1 \]

- **Corollary**: \( \Phi \) increases almost universally,

\[ \left\langle \Phi \right\rangle \geq \Delta H \]

\( \Delta H \) : entropy difference between initial and final state.

[V. Mustonen, M.L., PNAS 2010]
3. **Evolution and entropy**  
(Schrödinger’s problem)
Thermodynamics | Biological evolution

- **(-) energy**
  \[ -E(x, t) \]
  \[ F(x, t) \]

- **heat flux**
  \[ Q(x) = \sum_{i=1}^{n} \Delta x_i \left(- \nabla E\right)(x_i, t_i) \]
  \[ \Phi(x) = \sum_{i=1}^{n} \Delta x_i \nabla F(x_i, t_i) \]

- **Second Law**
  \[ \beta\langle Q \rangle + \Delta S = \Delta S_{\text{tot}} \geq 0 \]
  \[ N\langle \Phi \rangle - \Delta H \geq 0 \]

[Seifert 05, cf. Jarzynski 97, Crooks 99].
Thermodynamics

Biological evolution
<table>
<thead>
<tr>
<th>Thermodynamics</th>
<th>Biological evolution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First Law:</strong> conservation of energy</td>
<td>No tsunamis on fitness seascapes!</td>
</tr>
<tr>
<td>[ E(x, t) \rightarrow ] [ E(x, t) + E_0 ]</td>
<td>[ F(x, t) \rightarrow F(x, t) + F_0(t) ]</td>
</tr>
<tr>
<td><strong>Second Law:</strong> increase of total entropy</td>
<td>Adaptation can decrease the system's entropy!</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4. Measuring irreversibility
(The length of time’s arrow)
Influenza as a model system

- **Sequence data:**
  Haemagglutinin coding sequence data available over 40 years.

- **Protein structure and host interactions:**

  - viral epitopes (containing receptor-binding site)
  - viral haemagglutinin (surface protein)
  - human antibody (containing antigen-binding fragment)
Sequence tree of influenza

- **Strain tree** based on HA1 sequences of 1971 influenza H3N2 strains between 1969 and 2006

[cf. Bush and Fitch, 1999]
Sequence tree of influenza

- **Strain tree** based on HA1 sequences of 1971 influenza H3N2 strains between 1969 and 2006

  - allows reconstruction frequency time-series for
    - synonymous mutations,
    - nonsynonymous non-epitope mutations,
    - nonsynonymous epitope mutations.

[N. Strelkowa, M.L.]
Inference of adaptive evolution

- **Punctuated selection**: In a small population, beneficial mutations are rare and independent.
Inference of adaptive evolution

- **Clonal interference:** In a large population, beneficial mutations are frequent and compete for fixation.

- In 40 years, we find **at least 45 beneficial** and **at most 5 deleterious** substitutions:
  - $\Phi$ steadily increases.

  
  [N. Strelkowa, M.L.]
5. Evolution and complexity
Genomic information

- Transcription factors bind to **DNA target sites**.

- Target sites have a more **specific sequence** than background DNA.

- **Information gain (entropy loss)** in the formation of a new site (in bacteria or yeast):

  \[ \Delta H \approx 15 \text{ bytes} \]

Genomic complexity

- **This loss of entropy increases regulatory complexity.**

- **Information content Complexity of the entire genome / organism?**
  - Gene number, genome length etc. are not good measures of functional complexity.
  - Adaptive evolution does not imply increase in complexity.
Conclusions

- Adaptive evolution is a **stochastic nonequilibrium process** quantified by **fitness flux** $\Phi$.

- **Fitness flux theorem:**
  
  *Increase of $\Phi$ is a nearly universal evolutionary principle.*

- *Influenza* evolution:
  
  adaptive dynamics with positive $\Phi$.

- Adaptive evolution *can* increase genomic information and complexity.