Towards a quantitative genetics of complex cellular traits

Michael Lässig

Institute for Theoretical Physics, University of Cologne

Thanks

Theory of quantitative traits:

Torsten Held (Cologne)
Daniel Klemmer (Cologne)
Armita Nourmohammad (Princeton)
Stephan Schiffels (Sanger)

Drosophila gene expression:

Armita Nourmohammad (Princeton)
Joachim Rambeau (Cologne, Montpellier)
Torsten Held (Cologne)
Johannes Berg (Cologne)

DFG Deutsche Forschungsgemeinschaft

sybacol

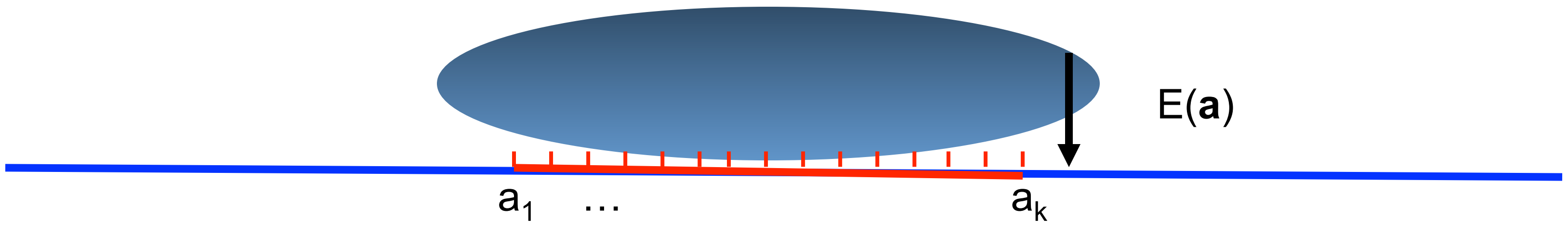
Federal Ministry of Education and Research

CORDIS
The European Commission
Community Research
Improving Human Research Potential & the Socio-economic Knowledge Base

Examples of phenotypes

- **Molecular traits:**

  Molecular binding interations, e.g., between transcription factors and regulatory DNA



- **Cellular traits:**

  metabolic, regulatory, and signaling pathways

- **Organismic traits:**

  body weight and size, shapes, metabolic rates
  fitness, longevity

➔ **Quantitative traits occur in coarse-grained descriptions of increasing level.**

Evolutionary theories

- Sequence level: population genetics
- Cellular level: systems biology
- Organismic level: quantitative genetics

→ **Evolutionary systems biology**

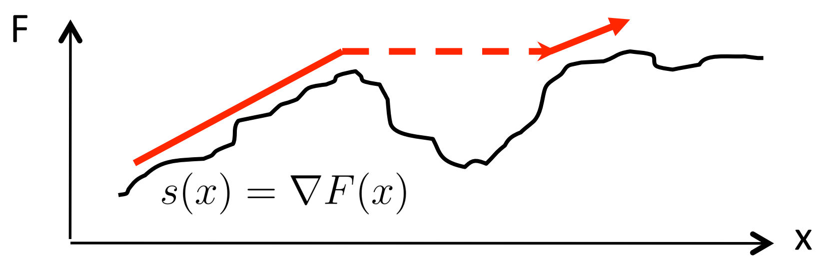
- **We need to rethink the quantitative genetics of molecular traits.**

1. Evolutionary equilibrium of regulatory DNA

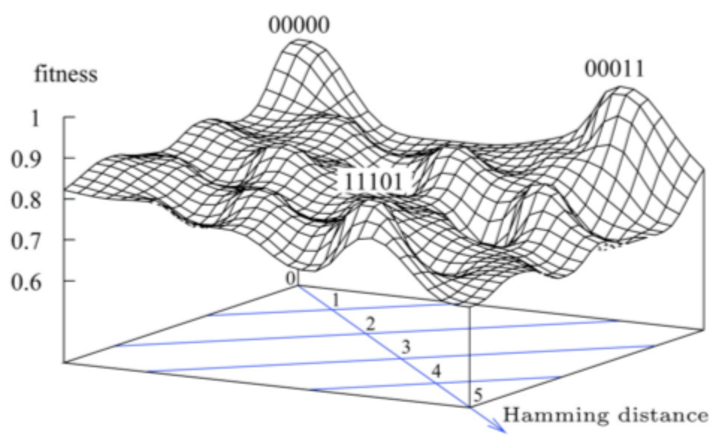
Fitness landscapes

- Evolution in a **fitness landscape** (S. Wright 1932) **:**

  - *interplay of selection and genetic drift*
  - *longer time intervals: mutations*



- Experimental fitness landscape in the fungus *Aspergillus niger*



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

Statistical mechanics of evolutionary equilibrium

- If the **neutral** process under mutations and genetic drift has an **equilibrium frequency distribution** $P_0(x)$, the process in an **arbitrary fitness landscape** also has an equilibrium

$$P_{\rm eq}(x) = P_0(x)\, \exp[2NF(x)] \,.$$

- **Average fitness** measures the **information of population states**:

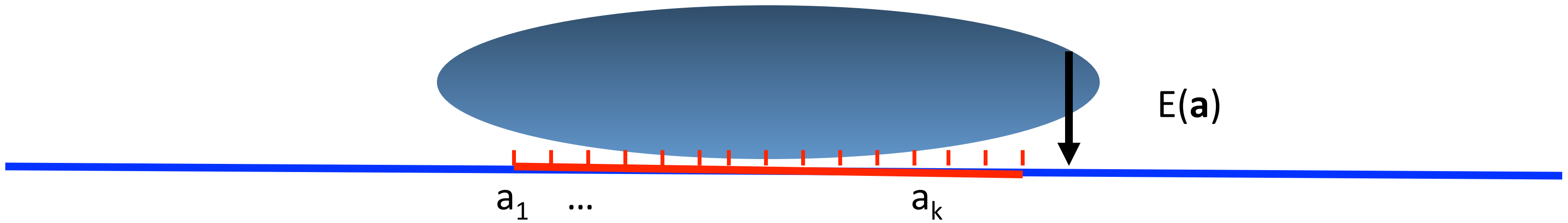
$$\langle 2NF \rangle = \int P_{\rm eq}(x)\, \log \frac{P_{\rm eq}(x)}{P_0(x)}\, dx = H(P_{\rm eq}|P_0) \;.$$

KL entropy

[J.Berg, S.Willmann, M.L., BMC Evol. Biol. 2004, V. Mustonen, M.L., PNAS 2010]

Cis-regulatory elements: from sequence to phenotype

- Transcription factor proteins bind to **DNA target sites,** catalyzing transcription.

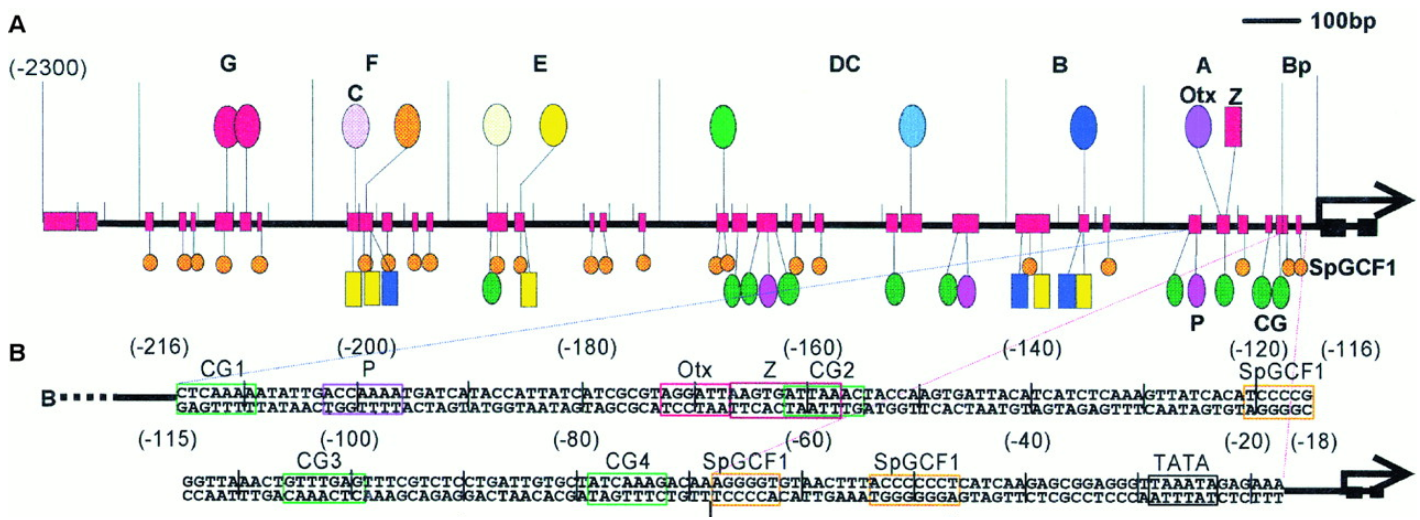


- **Binding energy** E(**a**) of a site **a** = (a_1,…,a_k) can be obtained from genomics and biophysical measurements.

- The binding energy E(**a**) is the **molecular phenotype** of a site, which **quantifies its functionality**.

Complexity of regulatory interactions

- Multiple binding sites allow for complex regulation of individual genes in higher organisms:



[Bolouri and Davidson, 2002]

- Input-output relation?

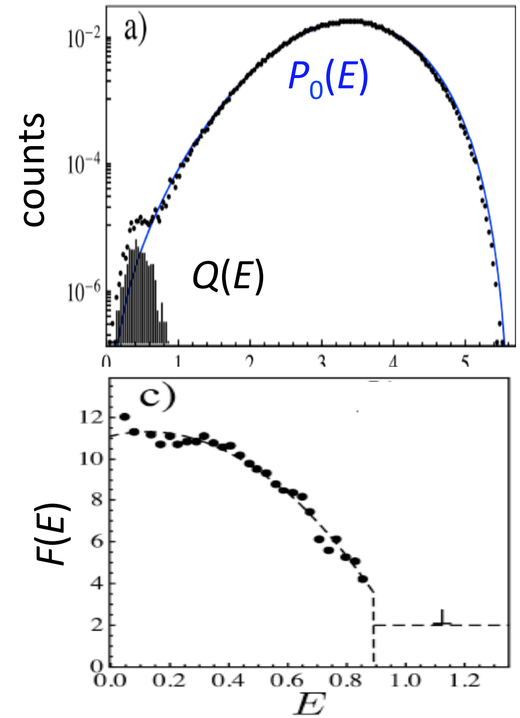
  Evolutionary dynamics?

Cis-regulatory elements: from phenotype to fitness

- Genomic analysis determines the distributions $P_0(E)$ and $Q(E)$.

- **Mesa fitness landscape** inferred at equilibrium

$$2NF(E) = \log \frac{Q(E)}{P_0(E)} + \text{const.}$$



Abf1 binding sites in S. cerevisiae

- This landscape successfully predicts binding site divergence across yeast species.

[Berg, Willmann, M.L., BMC Evol. Biol. 2004,
Mustonen, M.L., PNAS 2005, Mustonen, Kinney, Callen, M.L., PNAS 2008]

Conclusions (1)

- **Biophysical observables provide molecular phenotypes for evolutionary analysis.**

- **These can be used to infer phenotypic fitness landscapes.**

- **The mesa fitness landscape is shaped by thermodynamics of intra-cellular processes.**

Generalization to

- *higher genomic complexity?*
- *non-equilibrium adaptive processes?*

2. Evolutionary statistics of quantitative traits

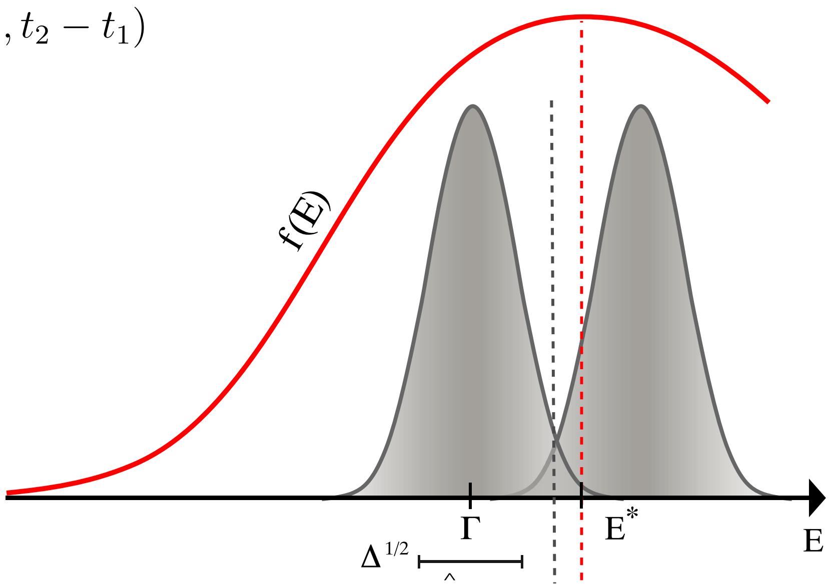
Evolutionary statistics of phenotypes

- **Trait mean and diversity** within one population:

$$\Gamma \equiv \overline{E} = \int E\,\mathcal{W}(E)\,dE, \qquad \Delta \equiv \overline{(E-\Gamma)^2} = \int (E-\Gamma)^2\,\mathcal{W}(E)\,dE,$$

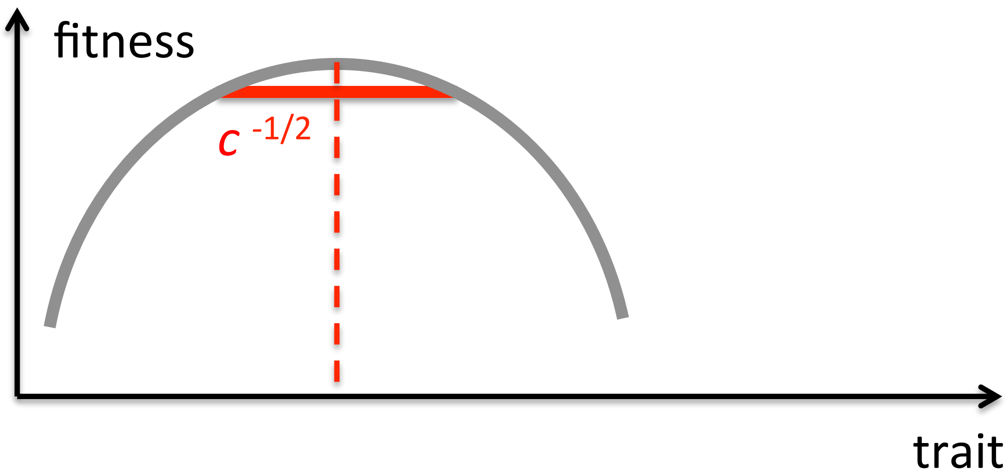
- **Trait divergence** between populations:

$$\left\langle (\Gamma(t_1) - \Gamma(t_2))^2 \right\rangle = \int d\Gamma_1 d\Gamma_2\; Q(\Gamma_2; \Gamma_1, t_2 - t_1)$$



Evolution under stablizing selection

- **Fitness landscape model:** Minimal dynamics under stabilizing selection



$$\frac{d\Gamma}{dt} = -\langle\Delta\rangle \frac{2c}{E_0^2}(\Gamma - E^*) - 2\mu(\Gamma - \Gamma_0) + \eta_\Gamma(t)$$

with noise generated by genetic drift: $\langle \eta_\Gamma(t)\eta_\Gamma(t')\rangle = \frac{\langle\Delta\rangle}{2N}\delta(t-t')$

+ similar equation for Δ

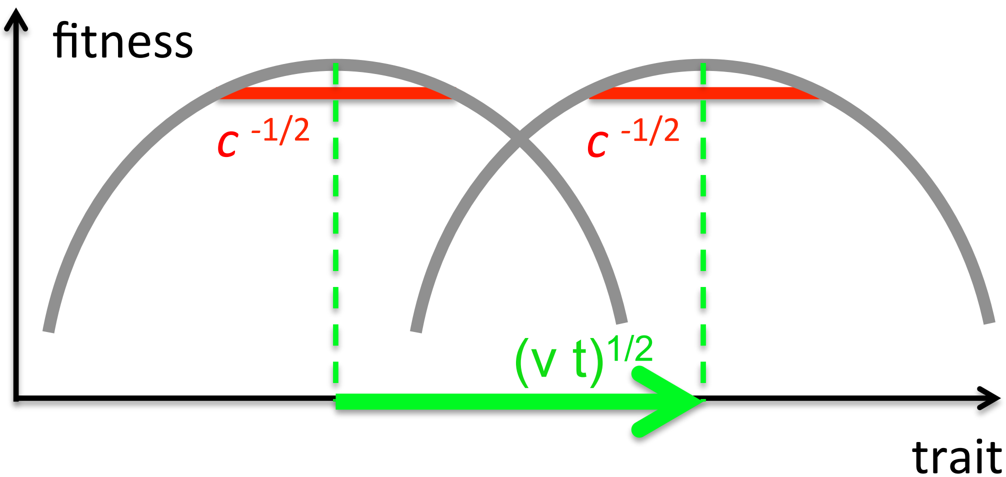
- **The landscape model generates equilibria of trait mean and variance:**

$$Q_{\rm eq}(\Gamma) = \frac{1}{Z}\tilde{Q}_0(\Gamma)\exp\left[2N\tilde{F}(\Gamma)\right], \qquad Q_{\rm eq}(\Delta) = \frac{1}{Z}Q_0(\Delta)\exp[2N\tilde{F}(\Delta)],$$

[Nourmohammad. Schiffels, ML, JSTAT 2013]

Adaptive evolution

- **Fitness seascape model:**
  Minimal adaptive dynamics



$$\frac{d\Gamma}{dt} = -\langle\Delta\rangle \frac{2c}{E_0^2}(\Gamma - E^*(t)) - 2\mu(\Gamma - \Gamma_0) + \eta_\Gamma(t)$$

$$\frac{dE^*}{dt} = \frac{\upsilon}{r^2}(E^* - \mathcal{E}) + \eta_{E^*}$$

with noise generated by genetic drift and fitness peak displacements:

$$\langle\eta_\Gamma(t)\eta_\Gamma(t')\rangle = \frac{\langle\Delta\rangle}{2N}\,\delta(t-t') \qquad \langle\eta_{E^*}(t)\eta_{E^*}(t')\rangle = \upsilon E_0^2\,\delta(t-t')$$

+ similar equation for Δ.

- **The seascape model generates non-equilibrium adaptive evolution.**

[Held, Nourmohammad, ML, JSTAT 2014]

Interlude: Quantifying adaptation

- A **population history** is a sequence of frequency measurements

$$\mathbf{x} = (x_0, \ldots, x_n) \quad \text{at times} \quad (t_0, \ldots, t_n).$$

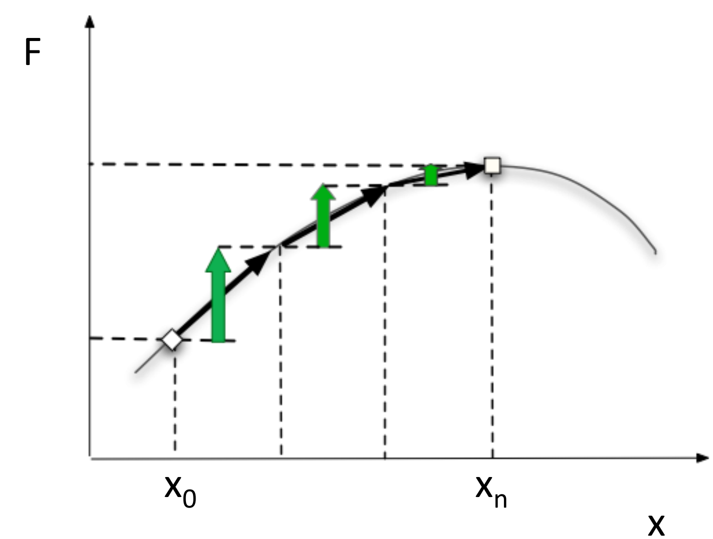
- The **fitness flux** of a population history is the cumulative **selective effect of frequency changes:**

$$\Phi(\mathbf{x}) \equiv \sum_{i=1}^{n} \Delta x_i \, s(x_i, t_i).$$

- Flux in a **fitness landscape:**

$$s(x) = \nabla F(x)$$

$$\begin{aligned}\Phi(\mathbf{x}) &= \sum_{i=1}^{n} \Delta x_i \, \nabla F(x_i) \\ &= F(x_n) - F(x_0).\end{aligned}$$



Interlude: Quantifying adaptation

- A **population history** is a sequence of frequency measurements

$$\mathbf{x} = (x_0, \ldots, x_n) \quad \text{at times} \quad (t_0, \ldots, t_n).$$

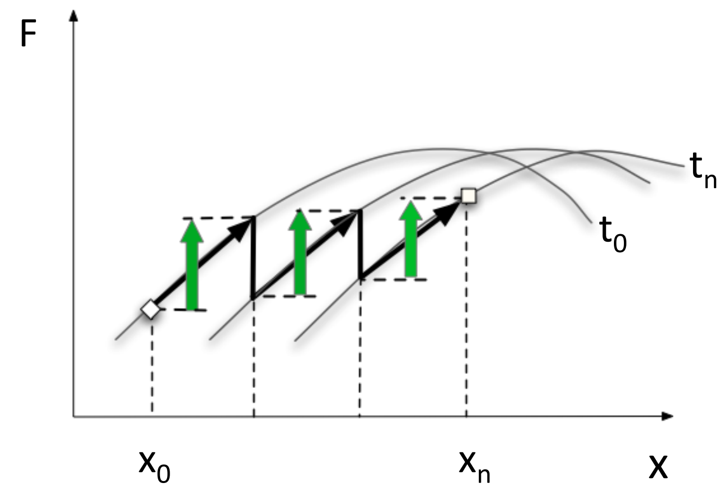
- The **fitness flux** of a population history is the cumulative **selective effect of frequency changes:**

$$\Phi(\mathbf{x}) \equiv \sum_{i=1}^{n} \Delta x_i \, s(x_i, t_i).$$

- Flux in a **fitness seascape:**

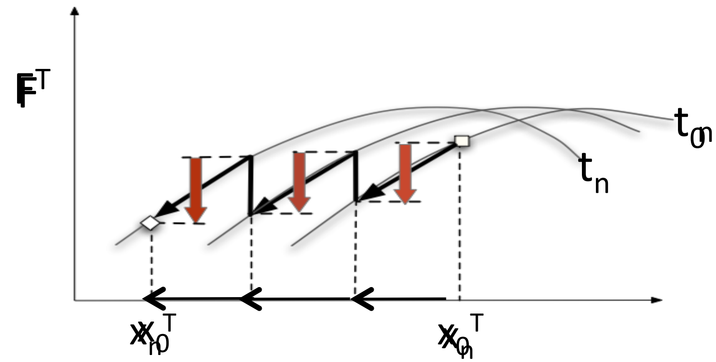
$$s(x,t) = \nabla F(x,t)$$

$$\begin{aligned}\Phi(\mathbf{x}) &= \sum_{i=1}^{n} \Delta x_i \, \nabla F(x_i, t_i) \\ &\neq F(x_n, t_n) - F(x_0, t_0).\end{aligned}$$



Interlude: Quantifying adaptation

- Each population history **x** has a **reverse history x^T,** in which all frequency transitions have opposite fitness effects:



- **Fitness flux measures adaptation as deviation from evolutionary equilibrium**

$$\mathcal{P}^T(\mathbf{x}^T) = \mathcal{P}(\mathbf{x}) \, \exp[-2N\Phi(\mathbf{x}) + \Delta\mathcal{H}(\mathbf{x})]$$

fitness flux

entropy difference of initial conditions

- **Average fitness flux** measures the **information of the adaptive process**:

$$\langle 2N\Phi \rangle = H(\mathcal{P}|\mathcal{P}^T) + \text{boundary terms}$$

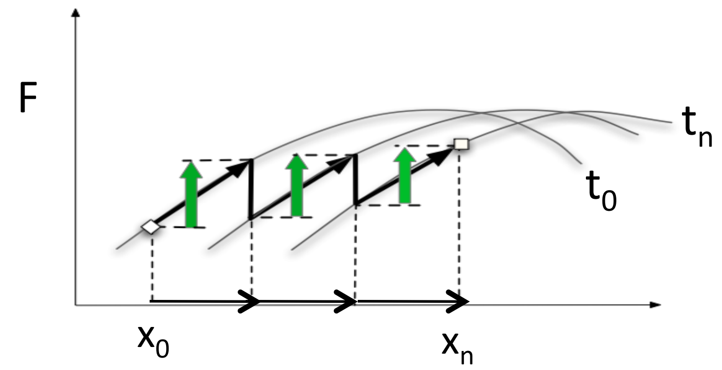
KL entropy

[Mustonen, M.L., PNAS 2010]

Interlude: Quantifying adaptation

- **Phenotypic fitness flux**

$$\Phi(\tau) = \int_0^\tau \frac{\partial F(\Gamma,t)}{\partial \Gamma}\,\frac{d\Gamma}{dt}\,dt$$



- **Average fitness flux in the seascape model:**

$$\langle 2N\Phi(\tau)\rangle = c\,\upsilon\,\tau$$

.

Inferring adaptation of quantitative traits

- **Scaled divergence-diversity ratio:**

$$\Omega(\tau) = 2\mu N \frac{\langle D(\tau) \rangle}{\langle \Delta \rangle}$$

- **Ω test:** The time-dependence of Ω measures evolution in a seascape:

- stabilizing strength

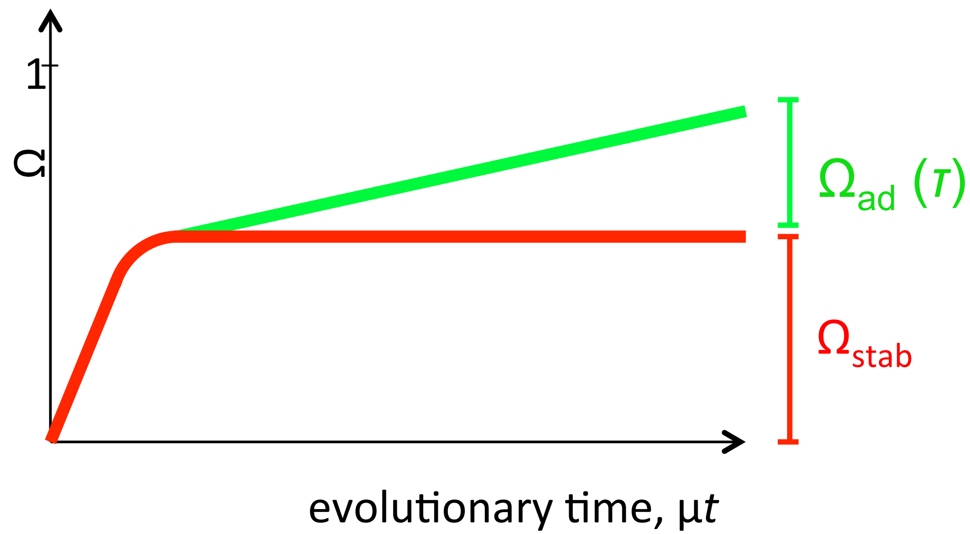
$$c \approx \frac{1}{\Omega_{\mathrm{stab}}}$$

- driving rate

$$\upsilon \approx \frac{2\Omega_{\mathrm{ad}}(\tau)}{\tau}$$

- fitness flux

$$\langle 2N\Phi(\tau) \rangle \approx \frac{2\Omega_{\mathrm{ad}}(\tau)}{\Omega(\tau) - \Omega_{\mathrm{ad}}(\tau)}$$



[Held, Nourmohammad, ML, JSTAT 2014]

- **Comparison with other methods**

1. *Q_{st} / F_{st} analysis:*
   infers directional selection if it is the dominant part of selection on short time scales ($\Omega > \Omega_0$) .

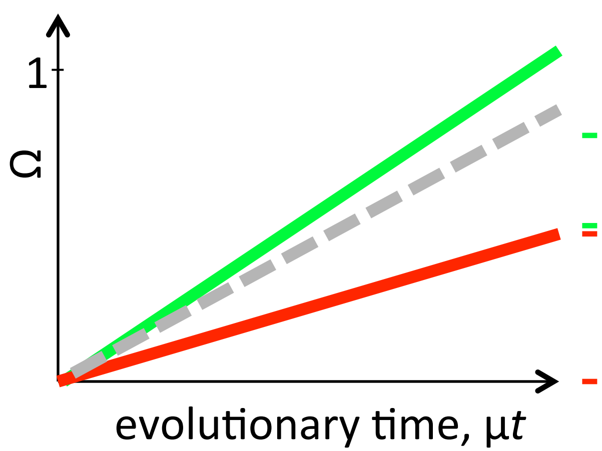
   Ω test:
   joint inference of directional and stabilizing selection (also for $\Omega > \Omega_0$) .

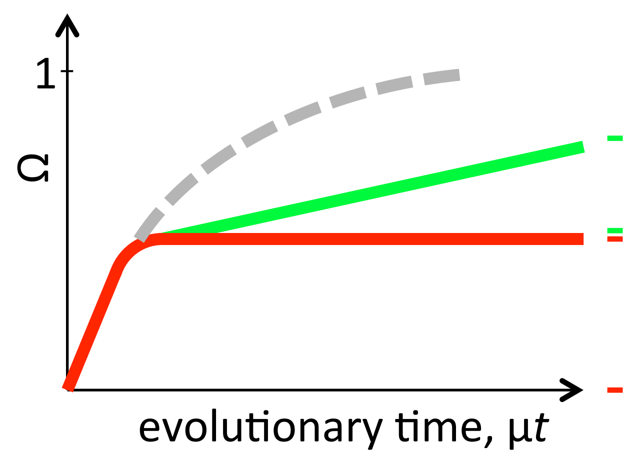
2. *Ornstein-Uhlenbeck model:*
   heuristics for equilibrium under stabilizing selection.

   Landscape model:
   determines Ornstein-Uhlenbeck parameters in terms of quantitative genestics

3. *McDonald-Kreitman test:*
   requires divergence data from query sequence vs. neutral sequence at one divergence time.

   Ω test:
   requires trait divergence data for at least two sufficiently large divergence times.





Conclusions (2)

- **The evolutionary statistics of complex quantitative traits shows *universal* characteristics.**

- **These can be used to infer phenotypic fitness** land- and **seascapes.**

- **The single-peak seascape - unifies stabilizing and directional selection.**
  - provides a minimal model for adaptive evolution.