Adaptive evolution of molecular quantitative traits

Michael Lässig

Institute for Theoretical Physics, University of Cologne

Thanks

Theory of quantitative traits:

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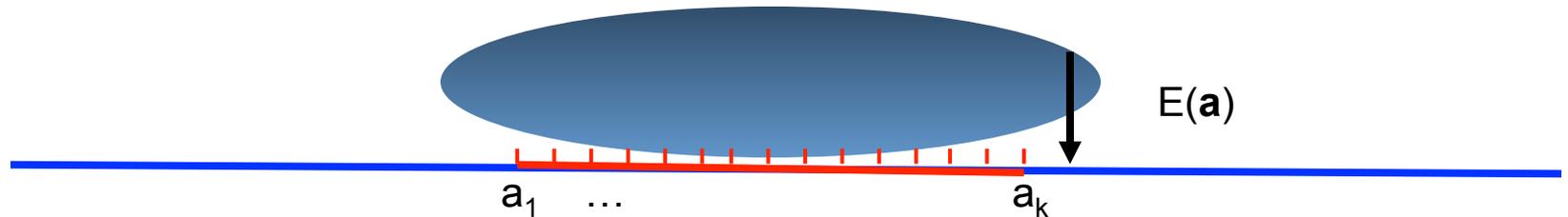Stephan Schiffels (Sanger)

Drosophila gene expression:

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

Examples of phenotypes

- **Molecular traits:**

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

$$E(\mathbf{a})$$

$$a_1 \quad \ldots \qquad\qquad a_k$$

- **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 statistics of quantitative traits

Population genetics

- **Stochastic equation of motion** for **population** frequencies:

$$\frac{dx}{dt} = s(x)g(x) + mx + \eta(x,t)$$

$$\langle \eta(x,t)\eta(x,t') \rangle = \frac{g(x)}{N}\delta(t-t')$$

selection   **mutations**   **reproductive fluctuations**
(genetic drift)

- **Equation of motion** for the **frequency distribution**:

$$\frac{\partial}{\partial t}P(x,t) = \nabla \left[\frac{1}{N}\nabla g(x) - s(x)g(x) - mx \right] P(x,t)$$
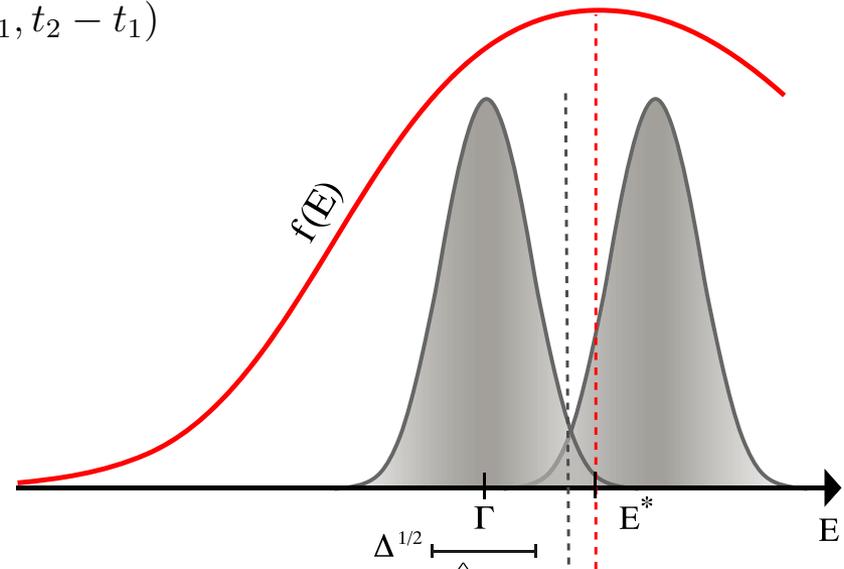
Evolutionary statistics of phenotypes

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

$$\Gamma \equiv \overline{E} = \int E\,\mathcal{W}(E)\,dE, \qquad \Delta \equiv \overline{(E - \Gamma)^2} = \int (E - \Gamma)^2\,\mathcal{W}(E)\,dE,$$
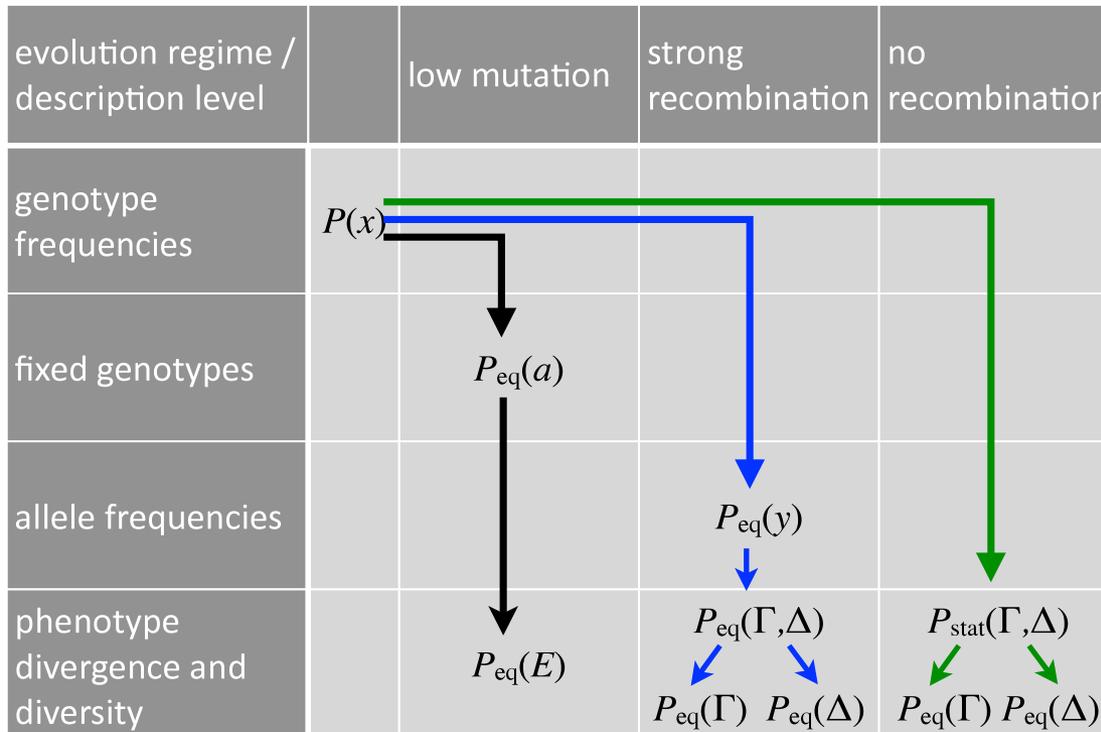
- **Trait divergence** between populations:

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

Coarse-graining to phenotypes

- **Genotype-phenotype mapping with and without recombination:**

evolution regime / description level		low mutation	strong recombination	no recombination
genotype frequencies	$P(x)$			
fixed genotypes		$P_{eq}(a)$		
allele frequencies			$P_{eq}(y)$	
phenotype divergence and diversity		$P_{eq}(E)$	$P_{eq}(\Gamma,\Delta)$ $P_{eq}(\Gamma)$ $P_{eq}(\Delta)$	$P_{stat}(\Gamma,\Delta)$ $P_{eq}(\Gamma)$ $P_{eq}(\Delta)$

[Berg, Willman, ML, BMC Evol. Biol. 2004; Sella and Hirsh, PNAS 2005]
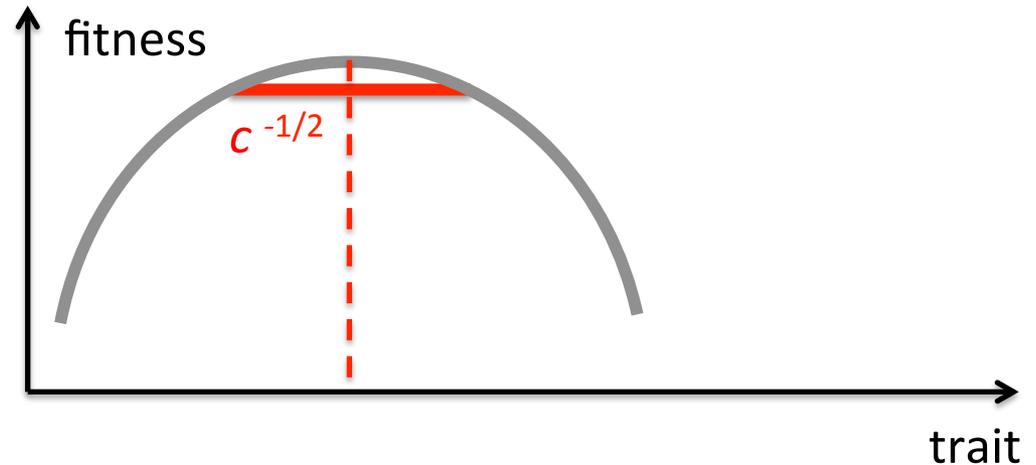
[Quantitative Genetics]

[Nourmohammad. Schiffels, ML, JSTAT 2013]
[Held, Nourmohammad, ML, JSTAT 2014]

Evolution under stablizing selection

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

$$\frac{d\Gamma}{dt} = -\langle\Delta\rangle\, 2c_0(\Gamma - E^*) - 2\mu(\Gamma - \Gamma_0) + \eta_\Gamma(t)$$

with noise generated by genetic drift: $\langle\eta_\Gamma(t)\eta_\Gamma(t')\rangle = \dfrac{\langle\Delta\rangle}{2N}$
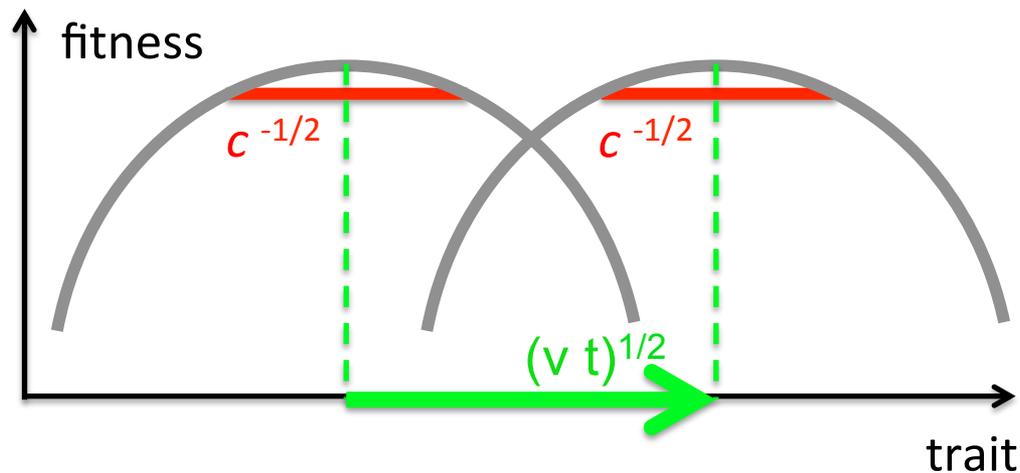
+ similar equation for Δ

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

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

[Nourmohammad. Schiffels, ML, JSTAT 2013]

Adaptive evolution

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

$$\frac{d\Gamma}{dt} = -\langle\Delta\rangle \, 2c_0(\Gamma - E^*(t)) - 2\mu(\Gamma - \Gamma_0) + \eta_\Gamma(t)$$

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

with noise generated by genetic drift and fitness peak displacements:

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

+ similar equation for Δ.

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

[Held, Nourmohammad, ML, JSTAT 2014]

Quantifying adaptation

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

$$\mathbf{x} = (x_0, \ldots, x_n) \qquad \text{at times} \qquad (t_0, \ldots, t_n).$$
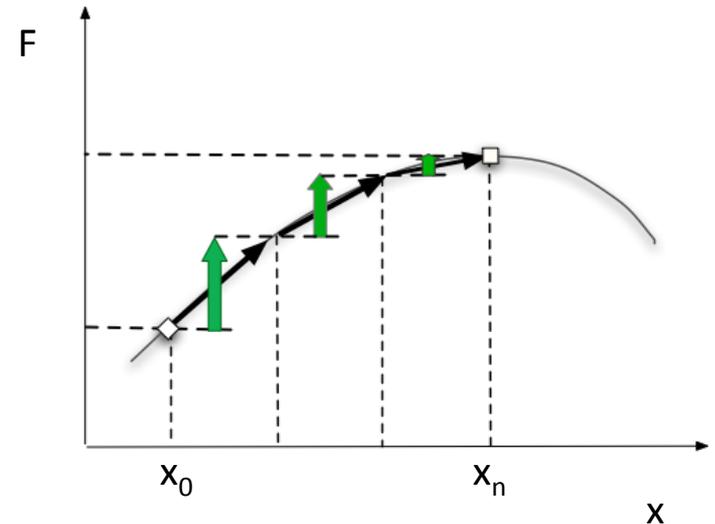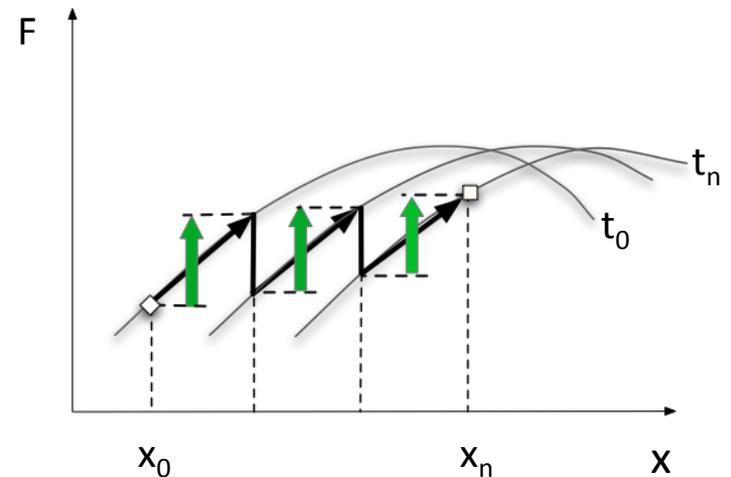
- The **fitness flux** of a population history is the cumulative **selective effect of frequency changes:**

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

- Flux in a **fitness landscape:**

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

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

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

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

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

$$\Phi(\mathbf{x}) \equiv \sum_{i=1}^{n} \Delta x_i\, s(x_i, t_i).$$
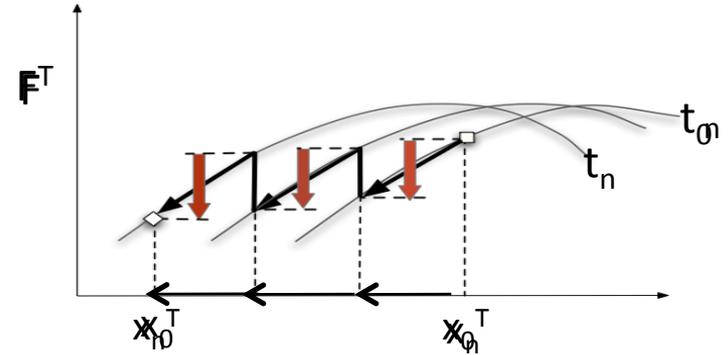
- Flux in a **fitness seascape:**

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

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

Quantifying adaptation

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

- **Fitness flux measures the amount of non-equilibrium / adaptation:**

$$\mathcal{P}(\mathbf{x}^T) = \mathcal{P}(\mathbf{x})\, e^{-N\Phi(\mathbf{x})} \;\; + \;\; \Delta\mathcal{H}(\mathbf{x})$$

fitness flux | **entropy difference of initial conditions**
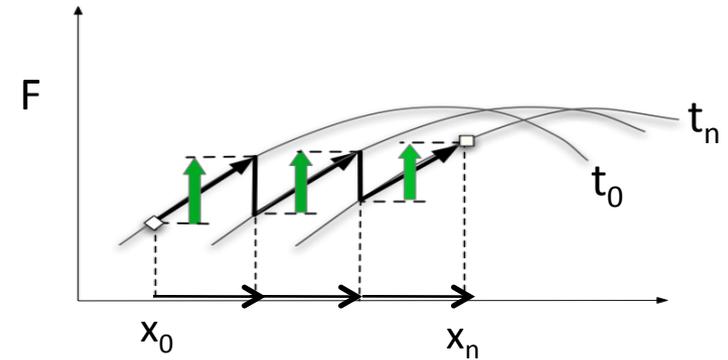
$$N\langle\Phi\rangle = H(\mathcal{P}|\mathcal{P}^T) \;-\; H(P(t_n)|P_0) + H(P(t_0)|P_0)$$

KL entropy | boundary terms

Quantifying adaptation

- **Phenotypic fitness flux**

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

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

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

.

- **Scaled divergence-diversity ratio:**

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

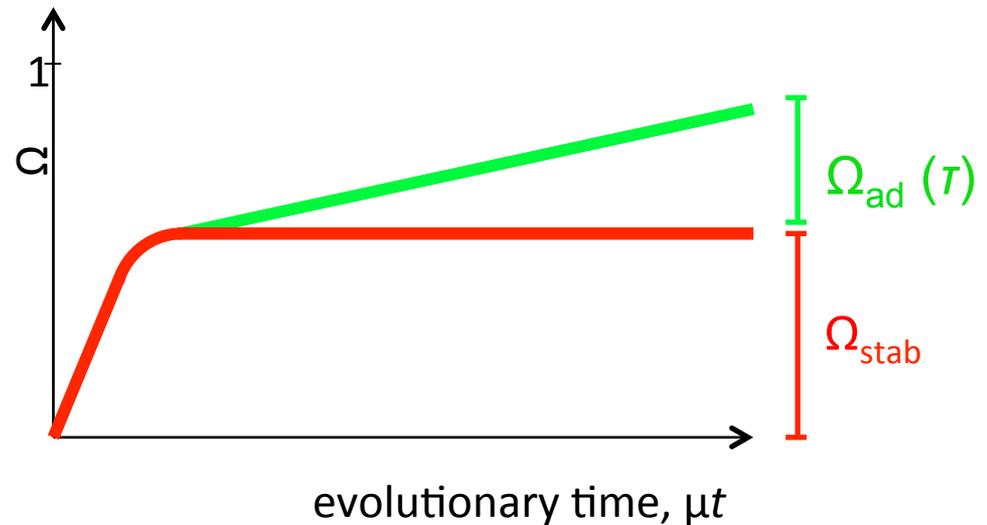- **Ω test:** The time-dependence of Ω measures:

  - *stabilizing strength*

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

  - *driving rate*

  $$v \approx \frac{2\Omega_{\text{ad}}(\tau)}{\tau}$$

  - *fitness flux*

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

- **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:
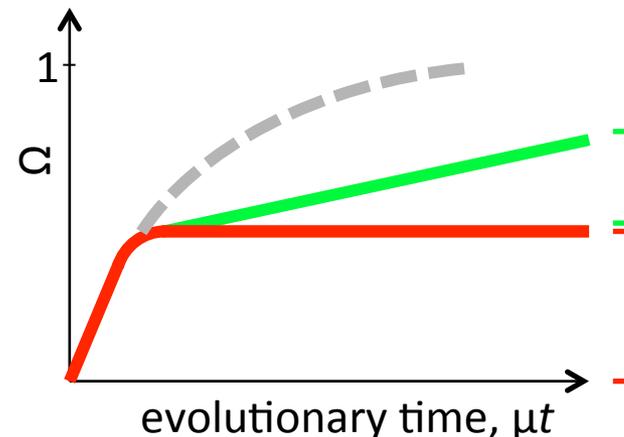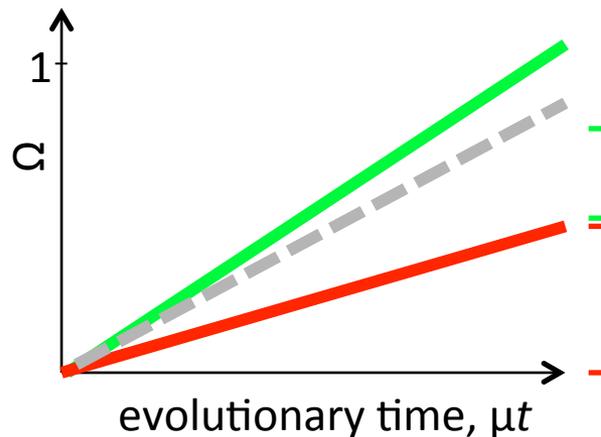   requires trait divergence data for at
   least two sufficiently large divergence times.

Ω · evolutionary time, μt

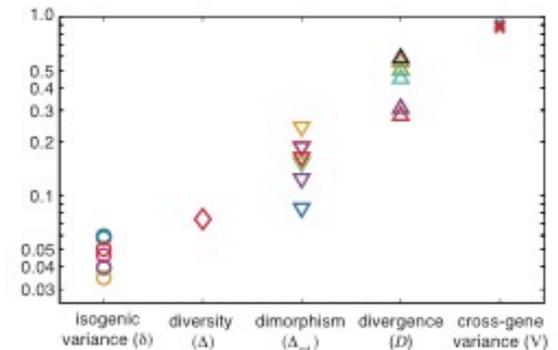Ω · evolutionary time, μt
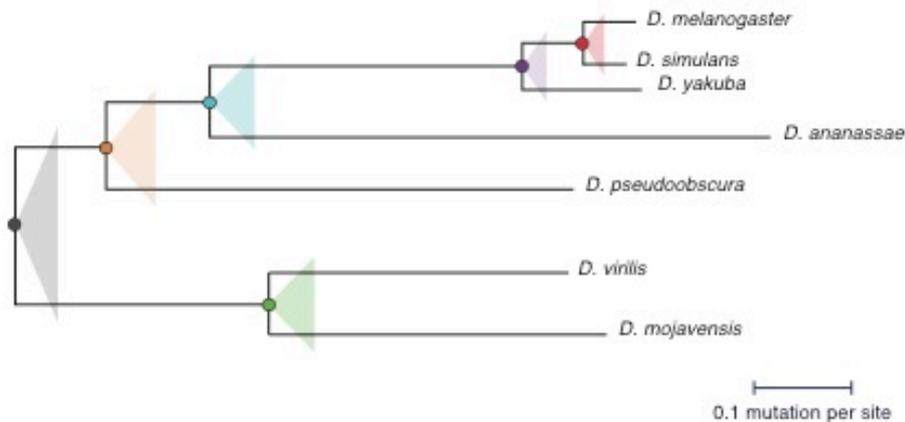
Conclusions (1)

- **The evolutionary statistics of 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.

2. Evolution of gene expression levels in *Drosophila*

Variation of gene expression across 7 Drosophila species

- **Comparative assay of genome-wide expression** levels   [Zhang et al, Nature 2007]
  - RNA levels of > 6000 genes from inbred lines,
  - 7 Drosophila species, 2 populations of D. simulans,
 - data from males and females.

clade	mel - sim	mel - yak	moj- vir	mel - ana	mel - pse	mel - moj (Dros.)
time, τ_o	0.11	0.27	0.71	1.12	1.17	1.38
species	D. melanogaster D. simulans	D. melanogaster D. simulans D.yakuba	D. mojavensis D. virilis	D. melanogaster D. simulans D.yakuba D. ananassae	D. melanogaster D. simulans D.yakuba D. ananassae D. pseudoobscura	D. melanogaster D. simulans D.yakuba D. ananassae D. pseudoobscura D. mojavensis D. virilis

The pattern of gene expression divergence

- **Ω ratio for clades of different divergence times**
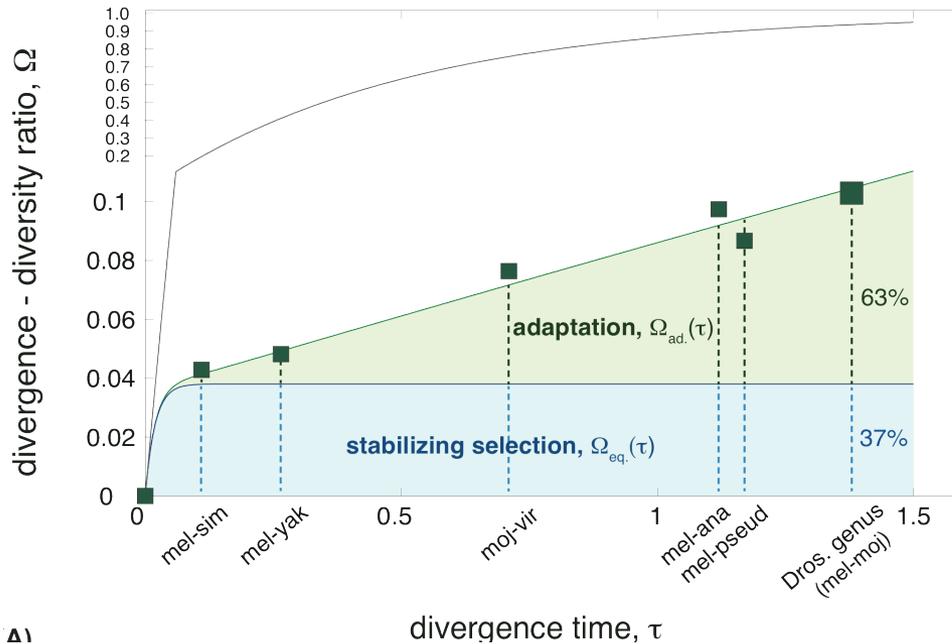
The Ω data are   - incompatible with neutral evolution.
   - incompatible with evolution under time-independent stabilizing selection.
   - **consistent with evolution in a single-peak fitness seascape.**

[Nourmohammad, Rambeau, Held, Berg, ML, submitted]

Inferring adaptive evolution

- **Aggregate analysis**

  Between   *mel* and *sim*,   about   8%   of the observed expression divergence is adaptive.
                 mel and *pse*,   64%

- **Probabilistic inference for individual genes:**

  - 54% of all genes are significantly adaptive ($\Phi_{ML} > 4$).

  - Less adaptation in genes with specific codon usage and in genes with high expression level.

  - More adaptation in genes with broad codon usage, genes with specific functions (sensory perception, regulation, neural maturation, regulation of growth, aging and morphology),
    genes with male-biased expression.

fitness flux, *2N* Φ

[Nourmohammad, Rambeau, Held, Berg, ML, submitted]
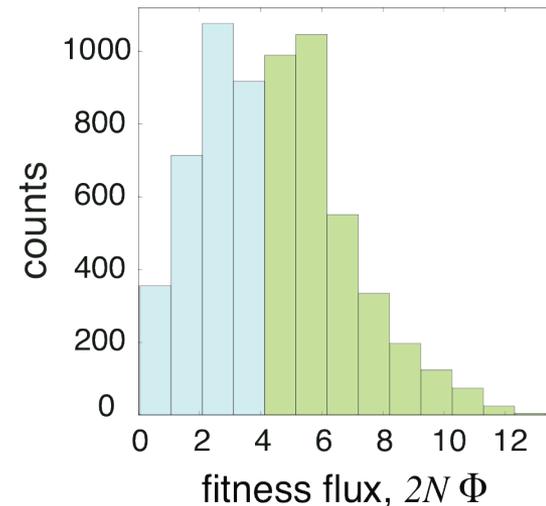
Inferring adaptive evolution

- **Testing alternative evolutionary scenarios:**

  - *Lineage-specific demography.*

  - *Lineage- and gene-specific relaxation of stabilizing selection:*
    stochastic gene loss.

  - *Punctuated fitness seascape:*
    large peak shifts with small rate.

[Nourmohammad, Rambeau, Held, Berg, ML, submitted]

- **Adaptive evolution of gene expression in *Drosophila* is pervasive.**

- **The adaptive pattern of expression has two molecular clocks.**

- **Quantitative genetics can be put to use for evolutionary systems biology.**