From Fisher to Feynman: nonequilibrium statistics of molecular evolution

Michael Lässig

Institute for Theoretical Physics
University of Cologne

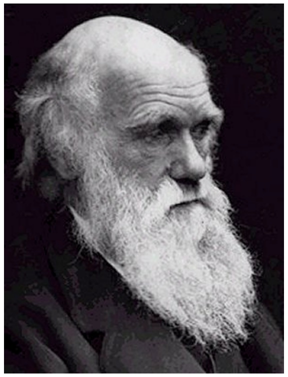
Thanks

Ville Mustonen

Curt Callan (Princeton)
Justin Kinney (Princeton)
Luca Peliti (Naples)

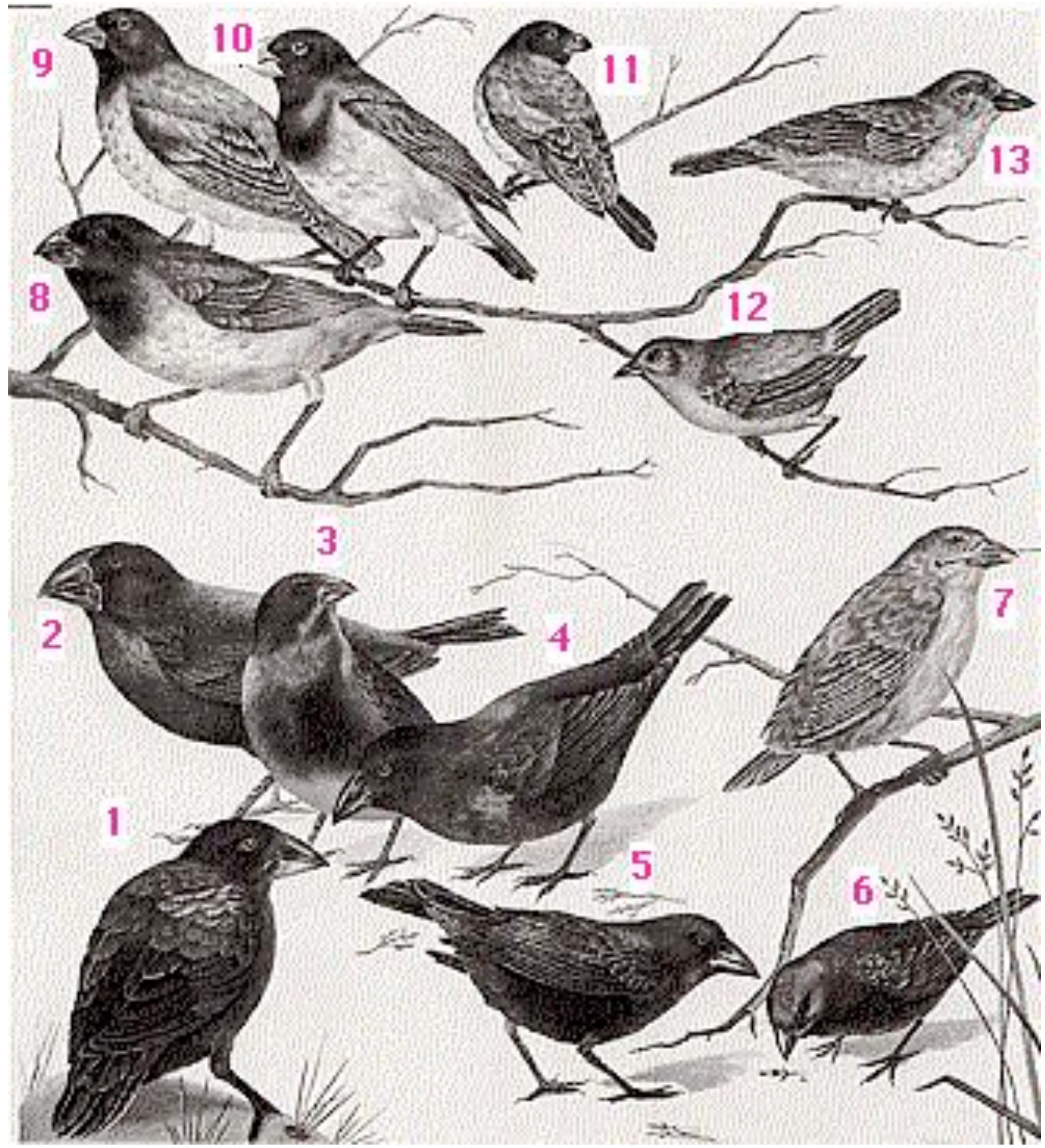
SFB TR 12
SFB 680
European Research Training Network STIPCO

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.**



Determinants of molecular evolution

- **Equation of motion** for the **population fractions** (frequencies) $x = (x_1, \ldots x_n)$ of phenotypes or genotypes in a population:

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

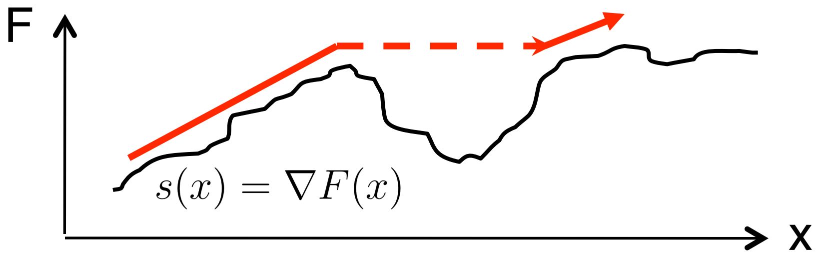
- $s(x)g(x)$: **selection**
- mx: **mutations**
- $\eta(x,t)$: **reproductive fluctuations** (genetic drift), with $\langle \eta(x,t)\eta(x,t') \rangle = \frac{g(x)}{N}\,\delta(t-t')$

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

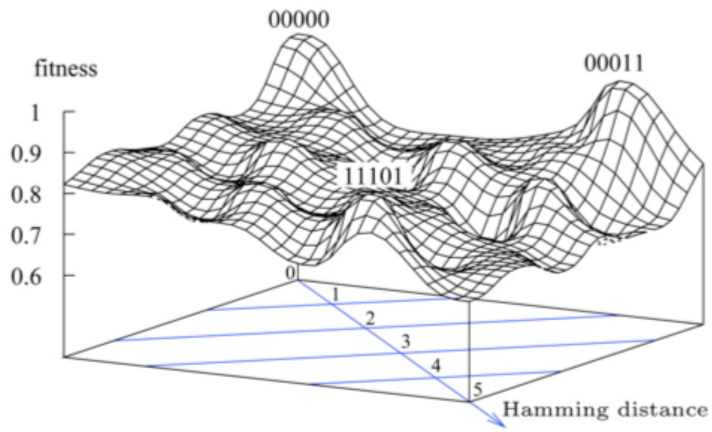
$$\frac{\partial}{\partial t}P(x,t) = \nabla \left[\frac{1}{N}\nabla g(x) - s(x)g(x) - mx \right] P(x,t)$$

Adaptative evolution

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



- 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 alone*

$$\frac{d}{dt} F(t) = s^2(x(t))$$

  - *evolution under time-dependent selection: nonequilibrium*

$$\frac{d}{dt} \Phi(t) = s^2(x(t), t)$$

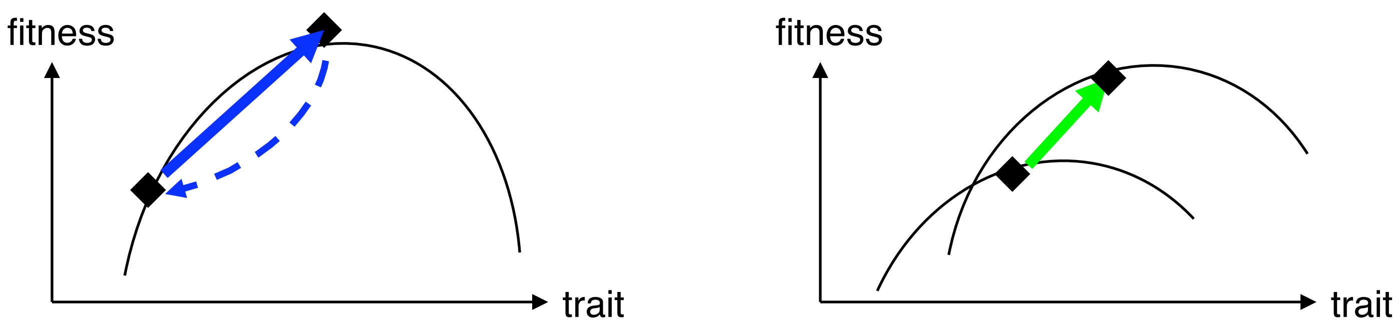
fitness flux

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

Adaptive evolution

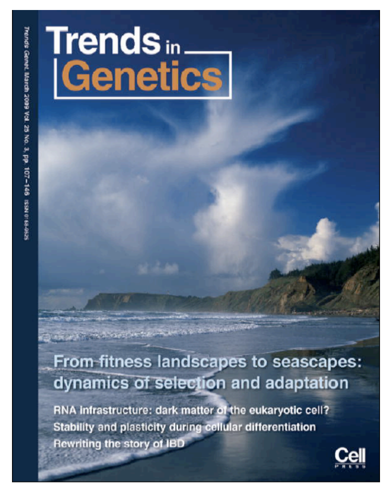
- A comprehensive **theory of molecular evolution** must include:

1. **stochastic forces** (mutations and genetic drift)
2. **evolutionary histories, correlations between genomic changes**:
   distinguish *compensatory evolution* from *adaptation:*



3. time-dependent selection:
   nonequilibrium **fitness seascapes**

V. Mustonen, M.L., March 2009



1. Fitness flux theorem

Population histories and fitness flux

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

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

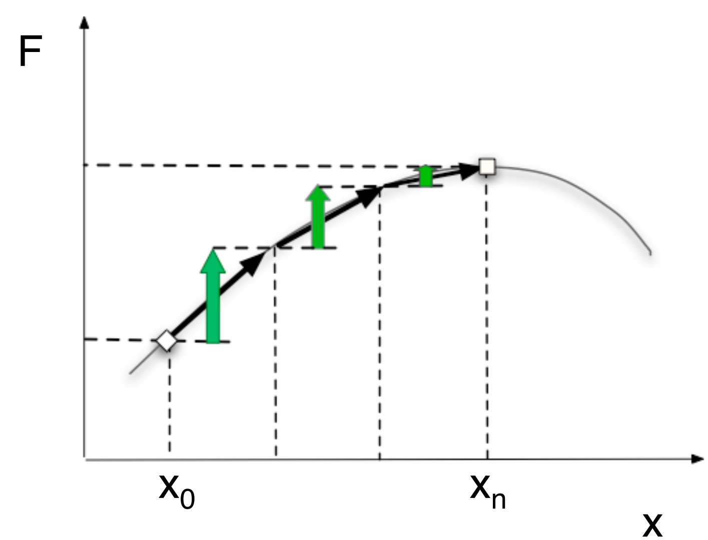
- The **fitness flux** of a population history is the cumulative **selective effect of frequency changes:**

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

- Flux in a **fitness landscape:**

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

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



Population histories and fitness flux

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

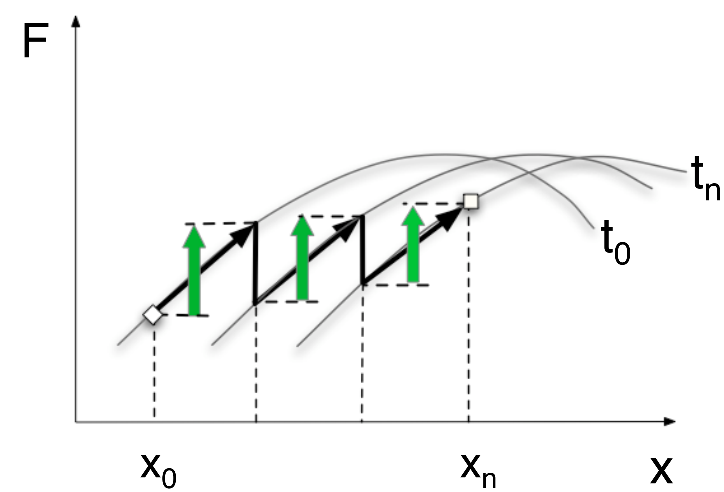
$$\mathbf{x} = (x_0, \ldots, x_n) \quad \text{at times} \quad (t_0, \ldots, t_n).$$

- 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:**
  $s(x,t) = \nabla F(x,t)$

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



Evolutionary equilibrium and fitness

- 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

$$P_{\rm eq}(x) = P_0(x)\, {\rm e}^{NF(x)} \;.$$

- Hence, **fitness** measures the **information of population states**:

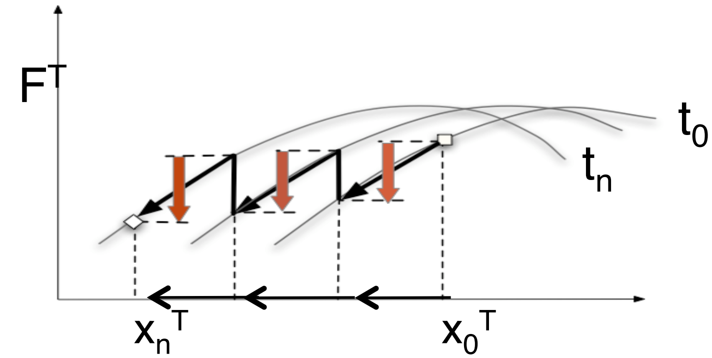
$$N\langle F\rangle_{\rm eq} = \int dx\, P_{\rm eq}(x) \log \frac{P_{\rm eq}(x)}{P_0(x)} \equiv H(P_{\rm eq}|P_0)\,.$$

KL entropy

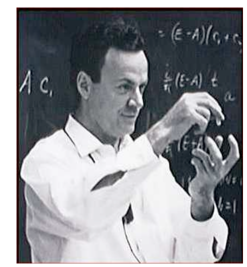
[J.Berg, S.Willmann, M.L., BMC Evol. Biol. 2004,
V. Mustonen, M.L., PNAS 2010, in press]

Fitness flux and time reversal

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



- The probabilities of forward and reverse history are related:



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

fitness flux

entropy difference of initial conditions

$$\mathcal{H}(x,t) \equiv \log[P(x,t)/P_0(x)]$$

$$\Delta\mathcal{H}(\mathbf{x}) = \mathcal{H}(x_n,t_n) - \mathcal{H}(x_0,t_0)$$

- Hence, fitness flux measures the **information of the evolution process:**

$$N\langle\Phi\rangle = H(\mathcal{P}|\mathcal{P}^T) \; - \; H(P(t_n)|P_0) + H(P(t_0)|P_0)$$

Fitness flux theorem

- **Theorem:** For an evolutionary process with mutations, genetic drift and selection given by an arbitrary fitness seascape,

$$\left\langle e^{-N\Phi+\Delta\mathcal{H}}\right\rangle = 1 \; .$$

$\langle\ldots\rangle$: average over population histories,
$\Delta\mathcal{H}(\mathbf{x}) = \mathcal{H}(x_n, t_n) - \mathcal{H}(x_0, t_0)$

.

- **Corollary: Φ increases almost universally,**

$$\langle\Phi\rangle \geq \Delta H \, .$$

$\Delta H = \langle\Delta\mathcal{H}\rangle = H(P(t_n)|P_0) - H(P(t_0)|P_0)$
entropy difference between initial and final state.

[V. Mustonen, M.L., PNAS 2010, in press]

Modes of fitness evolution

- **Equilibrium:**

$$\langle \Phi \rangle = 0$$

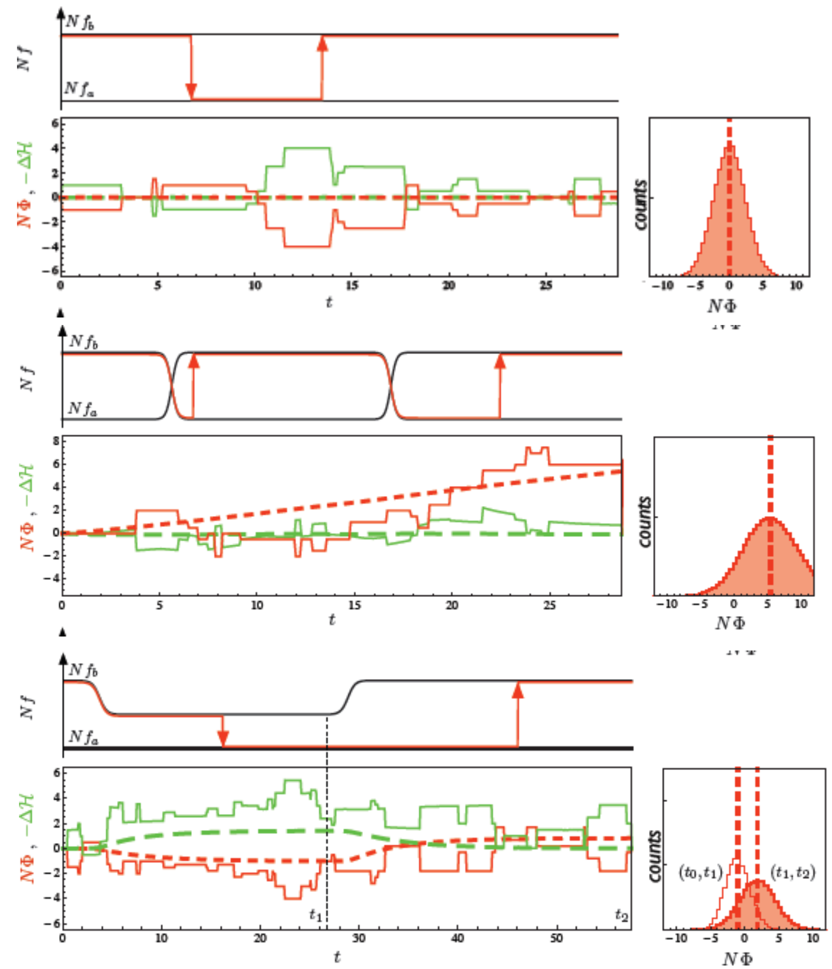
- **Stationary nonequilibrium:**

$$\langle \Phi \rangle > 0$$

- **Demographic nonequilibrium:**

$\langle \Phi \rangle < 0$ declining pop. size

$\langle \Phi \rangle > 0$ recovery



- **Strong-selection limit:** Fundamental theorem of natural selection

$$\frac{d}{dt}\Phi(t) = s^2(x(t), t)$$

2. Evolution and entropy

Thermodynamics | Biological evolution

- **(-) energy**

$$-E(x,t)$$

- **thermodynamic equilibrium**

$$P_{\mathrm{eq}}(x) = C\,\mathrm{e}^{-\beta E(x)}$$

- **heat flux**

$$Q(\mathbf{x}) = \sum_{i=1}^{n} \Delta x_i\,(-\nabla E)(x_i,t_i)$$

- **local entropy**

$$\mathcal{S}(x,t) = -\log P(x,t)$$

- **fluctuation theorem**

$$\left\langle \mathrm{e}^{-\beta Q - \Delta\mathcal{S}} \right\rangle = 1$$

$$\beta\langle Q\rangle + \Delta S = \Delta S_{\mathrm{tot}} \geq 0$$

[Seifert 05, cf. Jarzynski 97, Crook 99].

- **fitness**

$$F(x,t)$$

- **evolutionary equilibrium**

$$P_{\mathrm{eq}}(x) = P_0(x)\mathrm{e}^{NF(x)}$$

- **fitness flux**

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

- **local entropy**

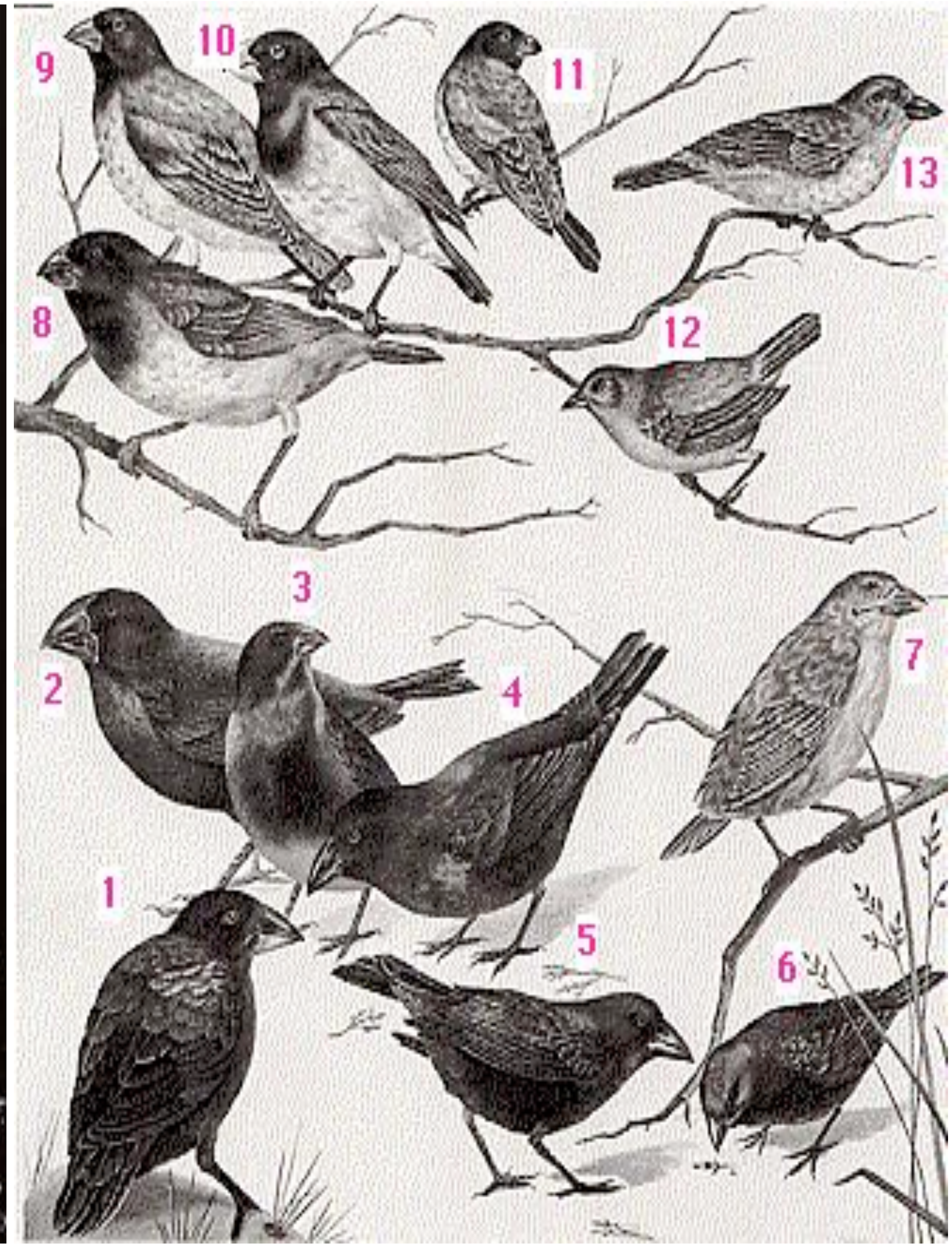
$$\mathcal{H}(x,t) = \log \frac{P(x,t)}{P_0(x)}$$

- **fitness flux theorem**

$$\left\langle \mathrm{e}^{-N\Phi + \Delta\mathcal{H}} \right\rangle = 1$$

$$N\langle\Phi\rangle - \Delta H \geq 0$$

Thermodynamics | Biological evolution

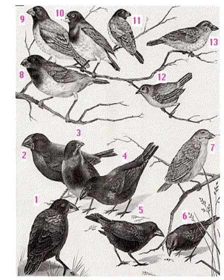


Thermodynamics

- **Second Law**

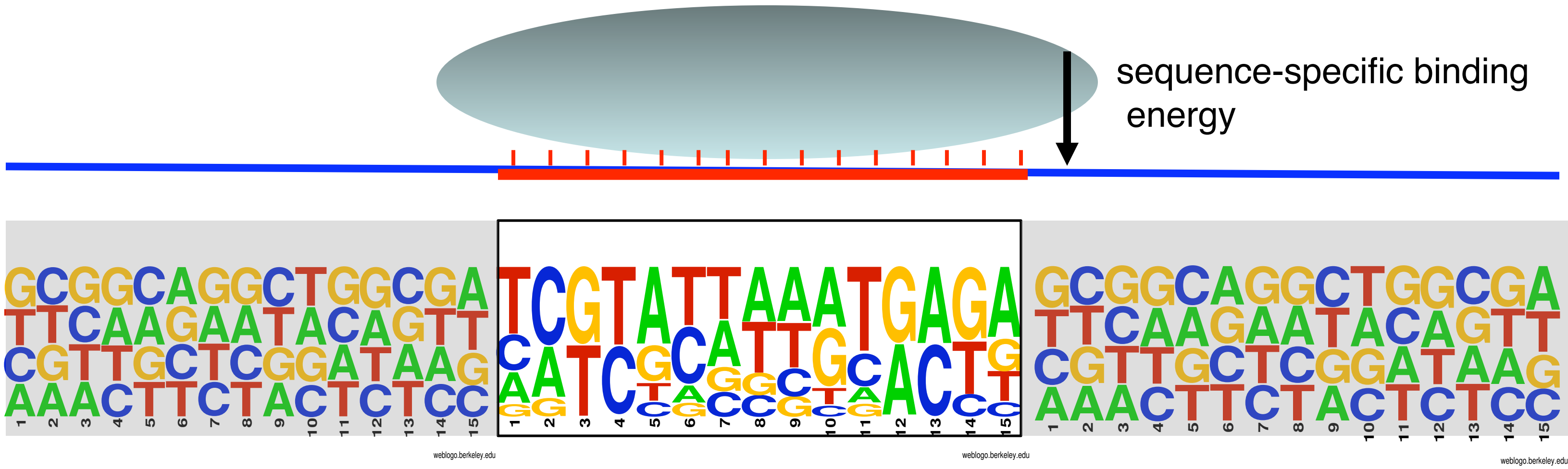
Biological evolution

- **Adaptation can decrease entropy!**



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 adaptive formation of a new site (in bacteria or yeast):

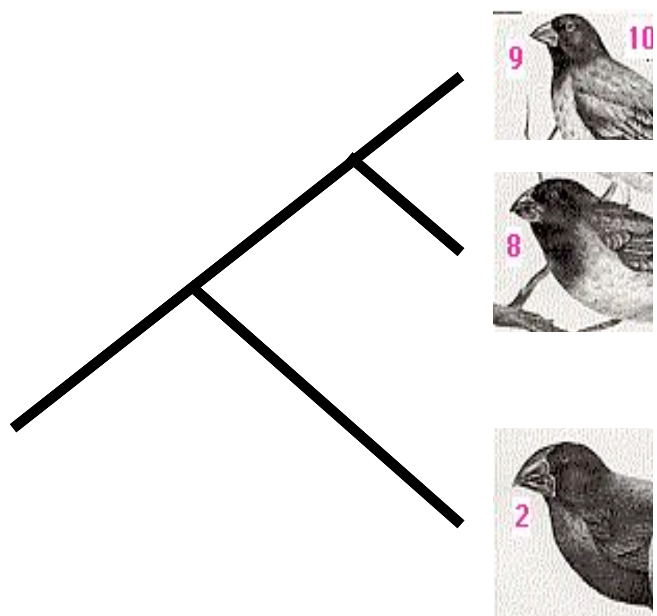
$$\Delta H \approx 15 \text{ bytes.}$$

[Mustonen, M.L., PNAS 2005, Mustonen, Kinney, Callen, M.L., PNAS 2008]

- **Information content of the entire genome?**

3. Irreversibility of evolution

(The length of time's arrow)



Adaptation and fitness flux in flies

AAGTCAGTCGATCAGTTCTCGAATAAGTCAGTCCTTCAGTTCTCGAAT
AAGTCAGTCGATCAGTTCTCGAATAAGTCAGTCCTTCAGTTCTCGAAT
AAGTCAGTCGATCAGTTCTCGAATAAGTCAGTCCATCAGTTCTCGAAT
AAGTCAGTCGATCAGTTCTCGAATAAGTCAGTCCATCAGTTCTCGAAT
AAGTCAGTCCATCAGTTCTCGAATAAGTCAGTCCATCAGTTCTCGAAT *Drosophila melanogaster*

AAGTCAGTCCATCACTTCTCGAATAAGTCAGTCCATCAGATCTCGAAT *Drosophila simulans*

[Glinka et al 2003, Ometto et al 2005]

- Cross-species data and polymorphisms within species are used to infer **rate** and **average selection coefficient** of **substitutions.**

- Adaptation is quantified by a **positive fitness flux:**

  Φ **=** (*rate*) x (*average selection coefficient of substitutions*).

- This indicates a **stationary nonequilibrium process:**
  - **Selection coefficients at individual sites fluctuate** at nearly the rate of neutral substitutions.
  - There is a **surplus of beneficial over deleterious substitutions.**

[Mustonen and M.L, PNAS 2007]

- **Amount of genome-wide adaptation?**

Conclusions

- Adaptive evolution is a **stochastic nonequilibrium process** quantified by **fitness flux** Φ.

- *Drosophila* genomes:
  evidence for adaptive evolution driven by **fitness seascapes.**

- Fitness flux theorem:
  Increase of Φ is a nearly universal evolutionary principle.