Population Genetics and Evolution – IV

The Speed of Adaptation in Large Populations

Luca Peliti

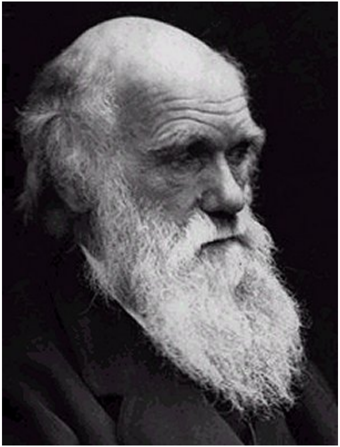
Dipartimento di Fisica and Sezione INFN
Università di Napoli “Federico II”

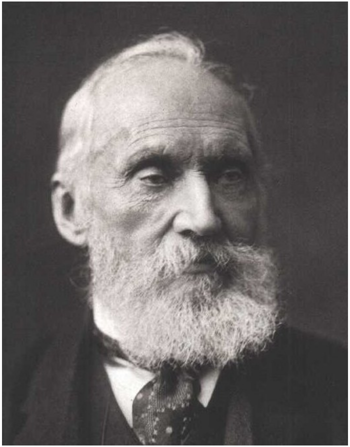
Helsinki / June 2013

Outline

The Battle of the Beards

Darwin vs. Kelvin



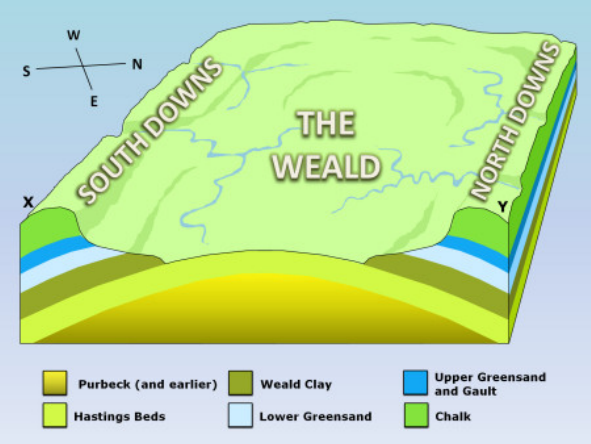


The question

Has there been time enough for evolution to act?

The denudation of the Weald

The denudation of the Weald



The denudation of the Weald

The Origin of Species:

> *Hence, under ordinary circumstances, I conclude that for a cliff 500 feet in height, a denudation of one inch per century for the whole length would be an ample allowance. At this rate, on the above data, the denudation of the Weald must have required 306,662,400 years; or say three hundred million years.*

CH. DARWIN (1859)

The denudation of the Weald

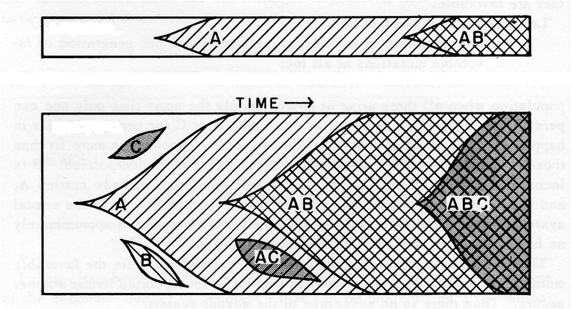
It must be supposed, therefore, that the sun always radiates away in heat something more than the Joule-equivalent of the work done on his contracting mass, by mutual gravitation of its parts[. . .]

From this it would follow with certainty that his temperature sinks 100° Cent. in some time from 700 years to 700,000 years.

What then are we to think of such geological estimates as 300,000,000 years for the "denudation of the Weald"? Whether is it more probable that the physical conditions of the sun's matter differ 1,000 times more than dynamics compel us to suppose they differ from those of matter in our laboratories; or that a stormy sea, with possibly Channel tides of extreme violence, should encroach on a chalk cliff 1,000 times more rapidly than Mr. Darwin's estimate of one inch per century?

LORD KELVIN (1862)

Adaptation Speed vs. Population Size



“Periodic”

“Multiple mutations”

The fitness distribution of a single mutation

- Organisms are “usually” well adapted
- Beneficial mutations are rare
- Thus beneficial mutations are mutations into the highest ranking genotypes
- For a large class of distributions, the “record” distribution is Gumbel’s one:

$$\begin{aligned}\rho(w) &= \frac{1}{W}\exp\left(e^{-(w-\mu)/W} - (w-\mu)/W\right)\\ &\approx e^{-(w-\mu)/W} \text{ for } w \gg \mu\end{aligned}$$

ORR, 2003

Clonal interference vs. multiple mutations

- In a small population, each mutation plays against the wild type
- In a large population, mutations play against a polymorphic population
- The fixation of a *single* beneficial mutation can be hindered by the appearance of a new, fitter, single mutation: *clonal interference*
- The fixation of a mutation cannot sometimes take place before new beneficial mutations appear on its background: *multiple mutations*

Contending mutations and clonal interference

FOGLE, NAGLE AND DESAI, 2008

- Wright-Fisher-like model with beneficial mutations
- Distribution of adaptation effect:

$$\rho(s) \propto \mathrm{e}^{-(s/\sigma)^{\beta}}$$

- Population size N, rate of beneficial mutations U_{b}

Distribution of mutations that fix

FOGLE, NAGLE AND DESAI, 2008

- Small populations: *Contending mutations*: Each mutation struggles *independently* against the wild type

$$\rho_{\mathrm{c}}(s) = \underbrace{\pi(s)}_{\text{fixation}}\ \underbrace{\rho(s)}_{\text{mutation}}$$

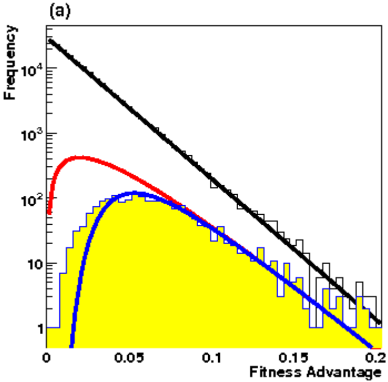
$$\pi(s) = \frac{1-\mathrm{e}^{-2s}}{1-\mathrm{e}^{2Ns}} \simeq \begin{cases} \frac{1}{N}, & \text{if } Ns \ll 1 \\ 2s, & \text{if } Ns \gg 1 \text{ and } s \ll 1 \end{cases}$$

- Larger populations: *Clonal interference*: Newly arising mutations struggle against each other
  Probability that a mutation fixes before a fitter one establishes (GERRISH AND LENSKI, 1998)

$$\begin{aligned} \rho_{\mathrm{f}}(s) &= \pi(s)\,\mathrm{e}^{-\lambda(s)}\,\rho(s) \\ \lambda(s) &\simeq \frac{NU_{\mathrm{b}}}{s}\ln N \int_s^{\infty} \mathrm{d}x\ \pi(x)\rho(x) \end{aligned}$$

Distribution of mutations that fix

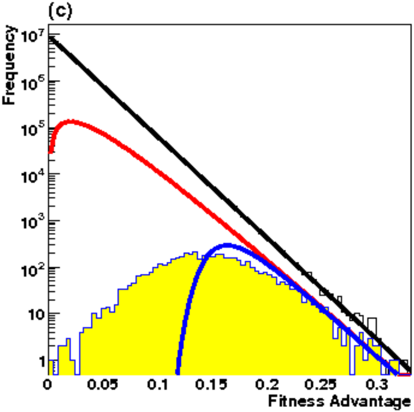
FOGLE, NAGLE AND DESAI, 2008



Small population: $N = 3 \cdot 10^4$, $U_b = 10^{-5}$, $\beta = 1$

Distribution of mutations that fix

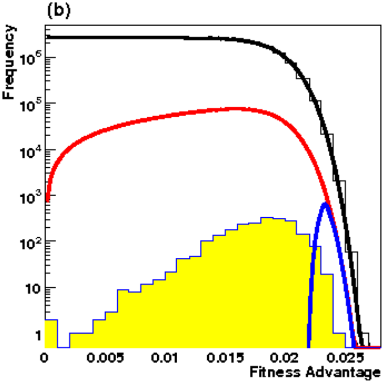
FOGLE, NAGLE AND DESAI, 2008



Large population: $N = 10^7$, $U_b = 10^{-5}$, $\beta = 1$

Distribution of mutations that fix

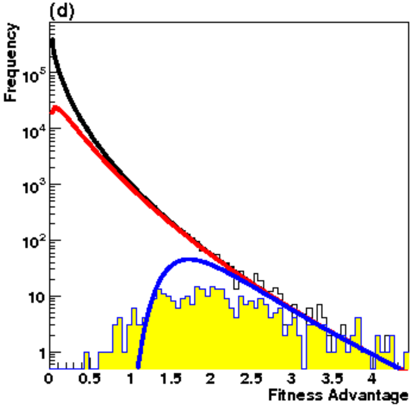
FOGLE, NAGLE AND DESAI, 2008



Large population, short tail: $N = 10^7$, $U_b = 10^{-5}$, $\beta = 10$

Distribution of mutations that fix

FOGLE, NAGLE AND DESAI, 2008



Large population, long tail: $N = 10^7$, $U_b = 10^{-5}$, $\beta = 0.5$

Adaptation in the infinite-site model

Representing the effects of multiple but comparable mutations:

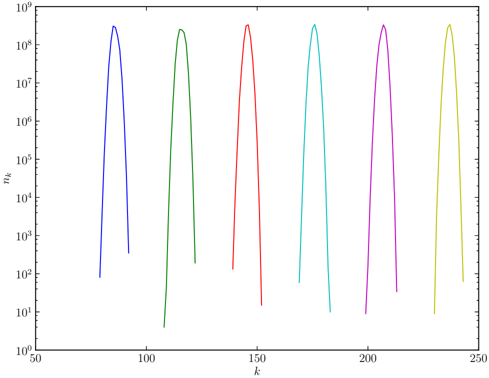
- Wright-Fisher model with N individuals
- Only beneficial mutations are considered
- Relative fitness with k mutations wrt ancestor:

$$W(k) = (1+s)^k$$

- Mutation rate U_{b}

Adaptation in the infinite-site model

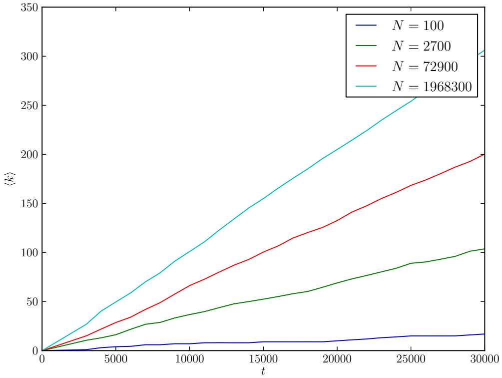
Travelling population wave



$N = 10^9$, $s = 0.01$, $U_\mathrm{b} = 0.0002$, $\Delta t = 2\,000$ generations

Adaptation in the infinite-site model

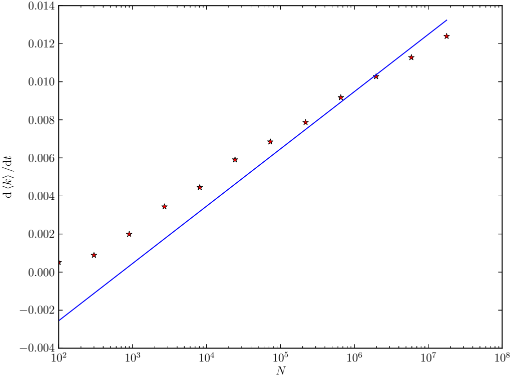
Adaptation speed vs. population size



$s = 0.01$, $U_{\mathrm{b}} = 0.0002$

Adaptation in the infinite-site model

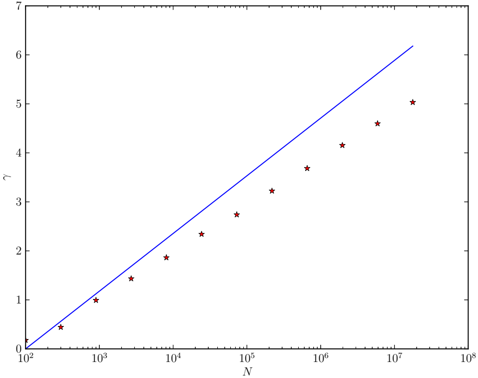
Adaptation speed vs. population size



$s = 0.01$, $U_\mathrm{b} = 0.0002$

Adaptation in the infinite-site model

The lead $\gamma = k_{\text{max}} - \langle k \rangle$ vs. population size



$s = 0.01$, $U_{\rm b} = 0.0002$

Approximate theory for the infinite-site model

WILKE, ROUZINE, BRUNET, DESAI, FISHER, . . . 2004–08

- Population size N
- Fitness $W_k = (1+s)^k$, $k = 0, 1, 2, \dots$
- Beneficial mutation rate U_b
- Fraction of population with k mutations: $x_k = n_k/N$
- Deterministic evolution equation ($N \to \infty$):

$$\frac{\mathrm{d}x_k}{\mathrm{d}t} = U_\mathrm{b}\left[x_{k-1} - x_k\right] + s\left(k - \langle k\rangle\right)x_k$$

Approximate theory for the infinite-site model

WILKE, ROUZINE, BRUNET, DESAI, FISHER, . . . 2004–08

- Assume $n_k s \gtrsim 1$: k-class is *established*
- Assume $U_\mathrm{b} \ll s$:

$$\frac{\mathrm{d}x_k}{\mathrm{d}t} \simeq s\,(k - \langle k \rangle)\,x_k$$

- Travelling-wave solution:

$$\begin{aligned} x_k &\propto \exp\left[-\frac{(k - \langle k \rangle_t)^2}{2\sigma^2}\right] \\ \frac{\mathrm{d}}{\mathrm{d}t}\langle k \rangle_t &= s\,\sigma^2 = v \qquad \text{(Fisher's F.T.)} \end{aligned}$$

Approximate theory for the infinite-site model

WILKE, ROUZINE, BRUNET, DESAI, FISHER, . . . 2004–08

- Self-consistency condition: $v = 1/\tau$, τ is the time needed for a new mutant class to get established
- Define the *lead* $\gamma = \max k - \langle k \rangle$ and set the origin at $k = \langle k \rangle$
- Once established, the fittest class grows at a rate γs and produces mutants at a rate $U_\mathrm{b} n_{k_\gamma}(t) \simeq \frac{U_\mathrm{b}}{\gamma s} \mathrm{e}^{\gamma s t}$
- New fittest mutants will be established when

$$n_{\gamma+1}(t) \simeq U_\mathrm{b} \int_0^\tau \mathrm{d}t \; n_\gamma(t) \sim \frac{1}{\gamma s}$$

- Thus $\tau \sim \frac{1}{\gamma s} \ln(s/U_\mathrm{b})$ and $v = s\sigma^2 = \gamma s / \ln(s/U_\mathrm{b})$

Approximate theory for the infinite-site model

WILKE, ROUZINE, BRUNET, DESAI, FISHER, . . . 2004–08

- By definition of the lead one has

$$\frac{N}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\gamma^2/2\sigma^2} \simeq \frac{1}{\gamma s}$$

- Thus we have the approximate equation for γ:

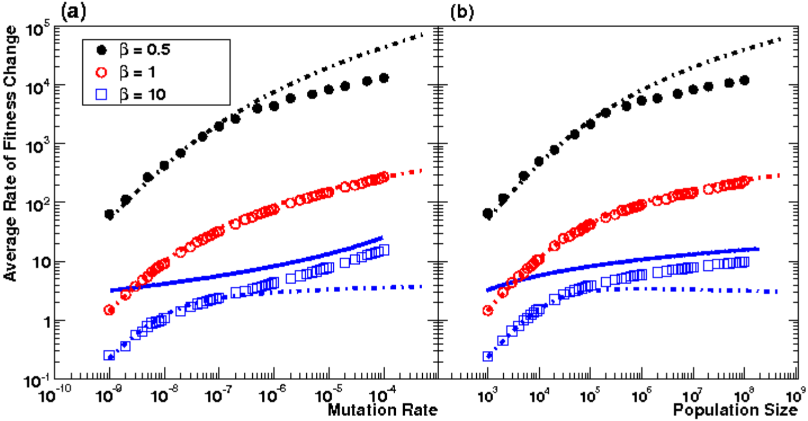
$$N\gamma s\,\mathrm{e}^{-\gamma^2/2\sigma^2} \sim Ns\,\mathrm{e}^{-\gamma\ln(s/U_\mathrm{b})/2} \sim 1$$

which yields

$$\begin{aligned} v &\approx \frac{2s\log(Ns)}{\log^2(s/U_\mathrm{b})} \\ \gamma &\approx \frac{2\log(Ns)}{\log(s/U_\mathrm{b})} \end{aligned}$$

Speed of adaptation

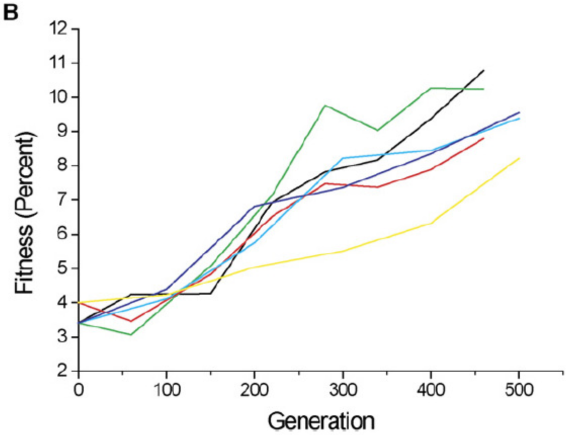
FOGLE, NAGLE AND DESAI, 2008



$\sigma = 0.02$ (a) $N = 10^7$ (b) $U_b = 10^{-5}$

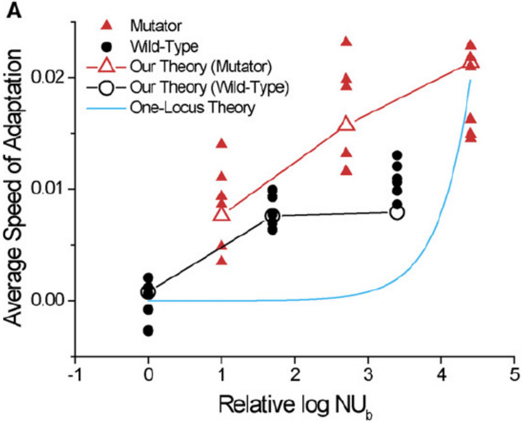
Experiments in Yeast

DESAI, FISHER AND MURRAY, 2007



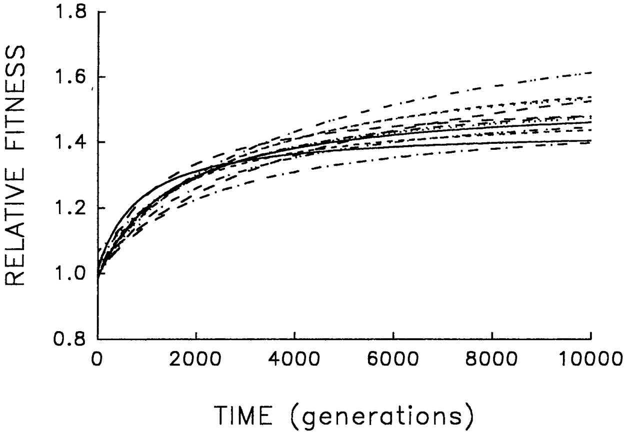
Experiments in Yeast

Desai, Fisher and Murray, 2007



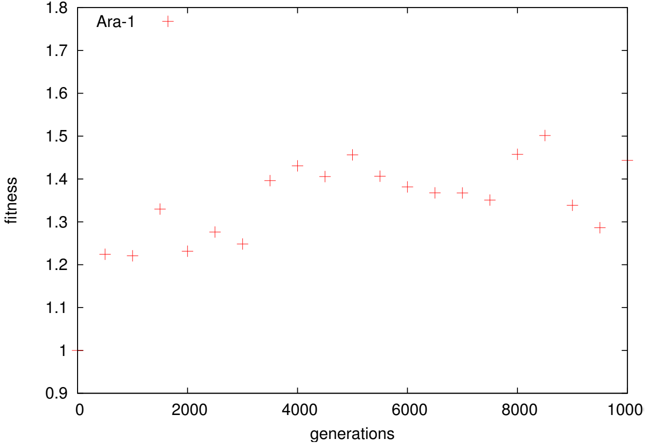
Fitness trajectories in the LTEE

Lenski & Travisano, 1994



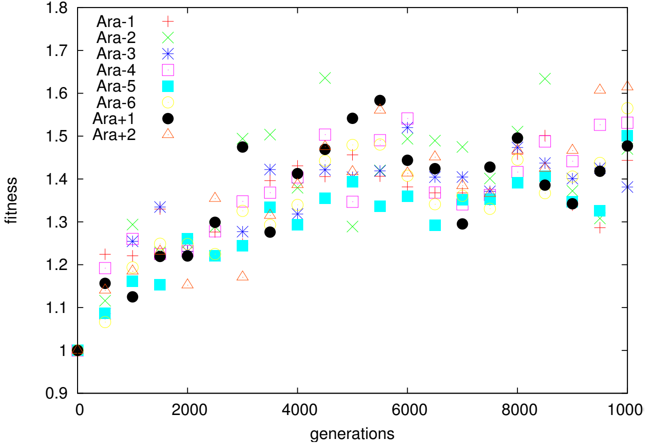
Fitness data

Lenski & Travisano, 1994



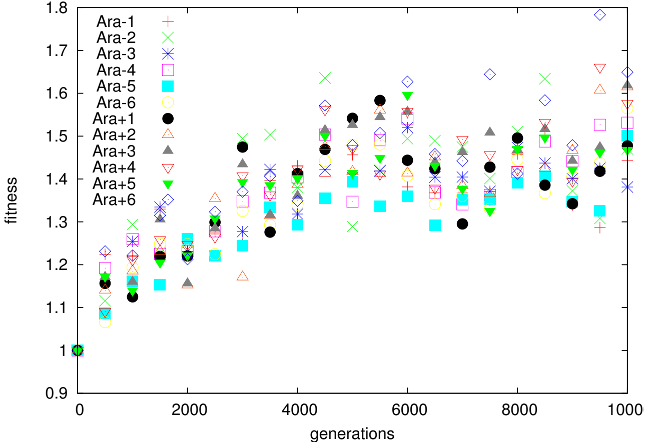
Fitness data

Lenski & Travisano, 1994



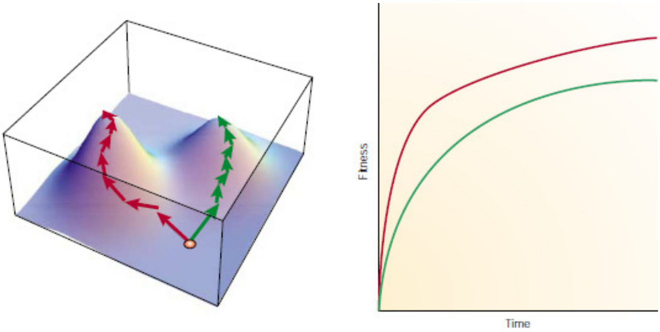
Fitness data

Lenski & Travisano, 1994



Fitness landscape & trajectories

Elena & Lenski, 2003



The infinite allele ("House of Cards") model

Kingman, 1978

- Each mutation yields a completely new genotype
- The fitness w of the genotype is independently extracted from a pdf $g(w)$
- Deleterious mutations are neglected

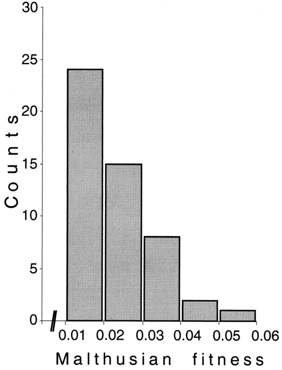
The infinite allele ("House of Cards") model

Kingman, 1978

- Each mutation yields a completely new genotype
- The fitness w of the genotype is independently extracted from a pdf $g(w)$
- Deleterious mutations are neglected

Beneficial mutation fitness in *E. coli*

Imhof & Schlötterer, 2001



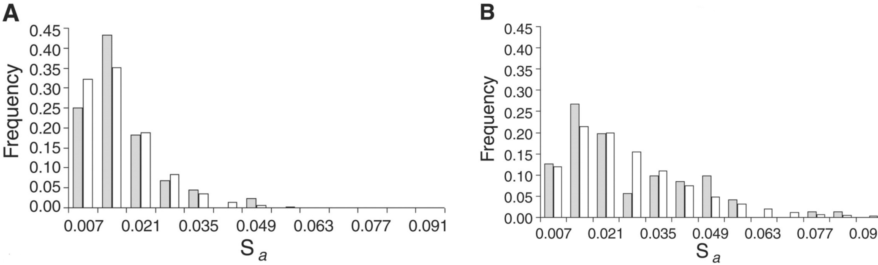
66 beneficial mutations

Estimated beneficial mutation rate

$U_\mathrm{b} \simeq 4\,10^{-9}$

Mutation fitness in *E. coli*

Perfeito *et al*, 2007



A: $N_e = 2\ 10^4$

B: $N_e = 10^7$

Estimated $U_b \simeq 2\,10^{-5}$

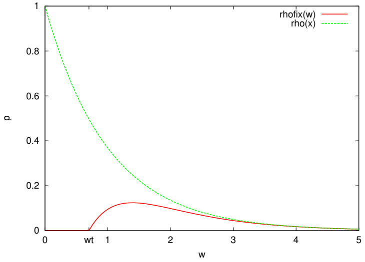
Theory: Finite population

Fixation probability of a mutant with relative selective advantage $s = (w - w_t)/w_t$ in a homogeneous population of fitness w_t (Kimura, 1962):

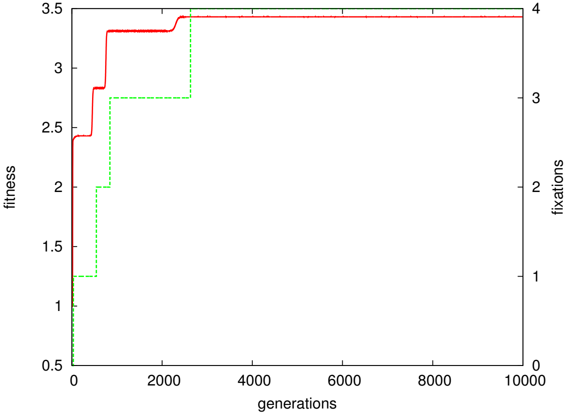
$$P_{\text{fix}}(s) = \frac{1 - \mathrm{e}^{-2s}}{1 - \mathrm{e}^{-2Ns}} \simeq \theta(s)(1 - \mathrm{e}^{-2s})$$

Distribution of fitness for mutations which fix in a homogeneous population of fitness w_t:

$$\rho_{\text{fix}}(w) = P_{\text{fix}}\left(\frac{w - w_t}{w_t}\right)\rho(w)$$

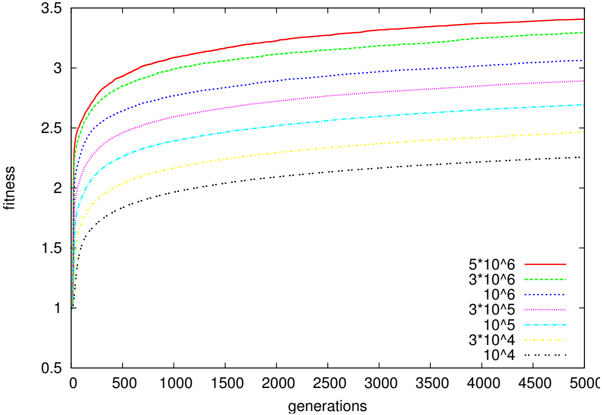


Exponentially distributed fitness

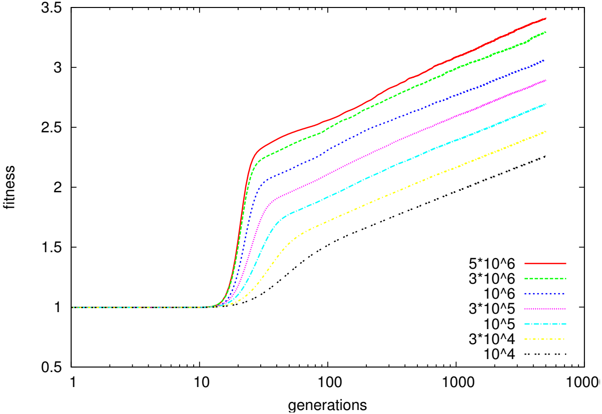


$N = 5\,10^6$; Mutation rate: $u_0 = 0.0033$; $g(w) = \mathrm{e}^{-w/W}/W$;
W: $P(w > 1) = u_\mathrm{a}/u_0$; $u_\mathrm{a} = 2\,10^{-5}$

Average behavior

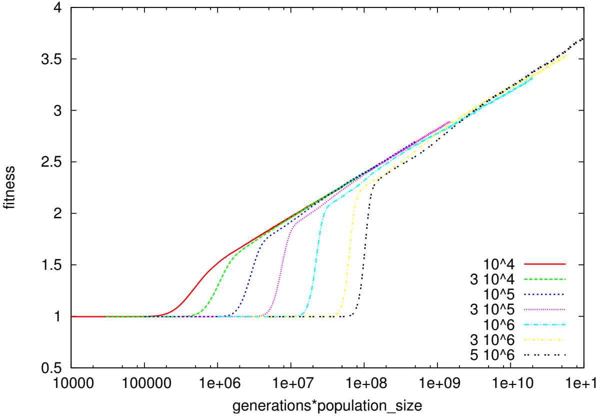


Average behavior



Speed and population size

$$\mathrm{d}w_t/\mathrm{d}t \propto N$$



Asymptotic behavior: Mean-field Theory

Park & Krug, 2008

At long times, beneficial mutations become rarer and rarer
Population is monomorphic almost all the time
Current fitness value w_t: Probability of a fixation

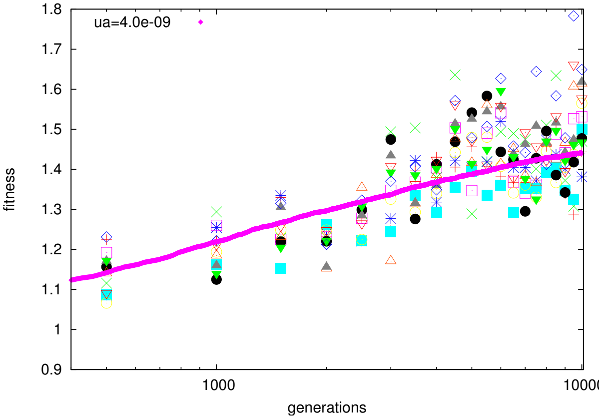
$$p_t = \int_{w_t}^{\infty} \mathrm{d}w \; \rho_{\mathrm{fix}}(w) \sim \frac{2W\mathrm{e}^{-w_t/W}}{2W + w_t}$$

$$\langle \Delta w \rangle = \int_{w_t}^{\infty} \mathrm{d}w \; (w - w_t) \, \frac{\rho_{\mathrm{fix}}(w)}{p_t(w_t)} = \frac{2W(W + w_t)}{2W + w_t}$$

Waiting time $\langle \Delta t \rangle$ till the next fixation: $\langle \Delta t \rangle \simeq 1/(Nu\, p_t(w_t))$

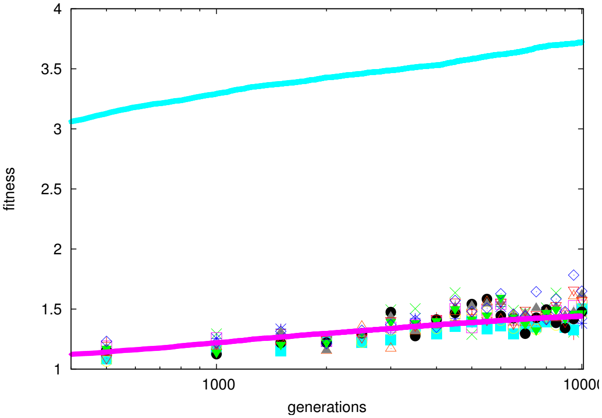
$$\begin{aligned} t(w) &\simeq \int^{w} \mathrm{d}w_t \, \frac{\langle \Delta t \rangle}{\langle \Delta w \rangle} \\ &= \frac{1}{4NeuW} \left[\mathrm{e}^{1+w/W}(2W + w) + W\mathrm{Ei}\left(1 + \frac{w}{W}\right) \right] + t_0 \end{aligned}$$

Data vs. the simple model



$N_e = 1.5\,10^7$, $u = 0.0033$, $u_a = 4.0\,10^{-9}$

Data vs. the simple model



$N_\mathrm{e} = 1.5\,10^7$, $u = 0.0033$, $u_\mathrm{a} = 4.0\,10^{-9}\ , 2.0 10^{-5}$

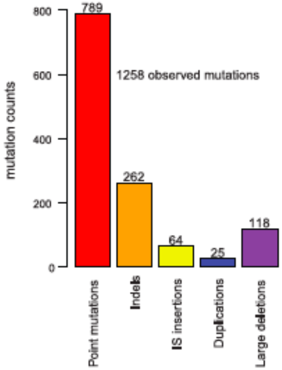
Adaptive convergence in *E. coli*

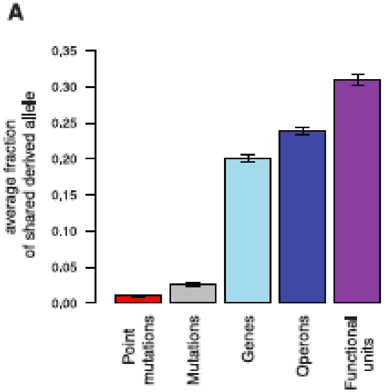
TENAILLON *et al.*, 2012

- 115 independent populations of *E. coli* at high temperature for 2000 generations
- One clone for each line sequenced
- $\sim$ 1200 *different* mutations, affecting $\sim$ 600 different sites
- $\sim$ 400 different genes
- $\sim$ 50 “modules”
- 2 distinct pathways: RNA polymerase or the termination factor ρ

Adaptive convergence in *E. coli*

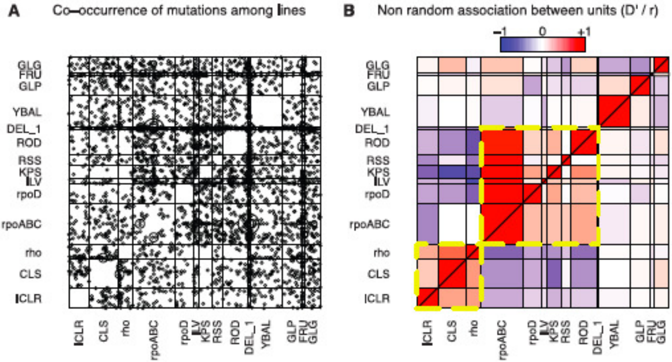
TENAILLON *et al.*, 2012





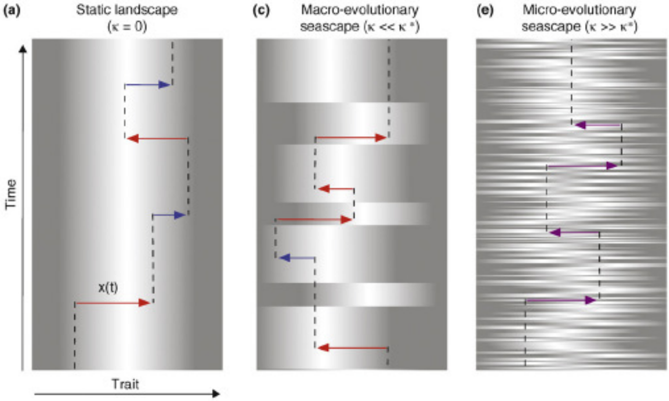
Adaptive convergence in *E. coli*

TENAILLON *et al.*, 2012



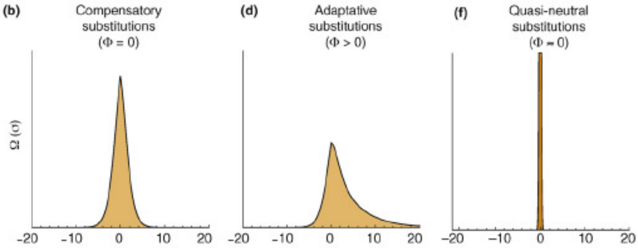
Adaptation in a changing environment

MUSTONEN AND LÄSSIG, 2007–2009



Adaptation in a changing environment

MUSTONEN AND LÄSSIG, 2007–2009



Adaptation flux

MUSTONEN AND LÄSSIG, 2010

- A fixed fitness function $\boldsymbol{f} = (f_k)$ defines a *fitness landscape* $\langle f \rangle_{\boldsymbol{x}} = \sum_k x_k f_k$
- $P^{\mathrm{eq}}(\boldsymbol{x}) = P_0(\boldsymbol{x})\, \mathrm{e}^{N\langle f \rangle_{\boldsymbol{x}}}$
- *At equilibrium* $[\langle f \rangle]_{\mathrm{av}}$ does not change: improvements are counterbalanced by deteriorations
- In time-changing environment (seascape) there is more room for improvement at each instant

Adaptation flux

MUSTONEN AND LÄSSIG, 2010

- Define the *cumulative fitness flux* $\Phi(\boldsymbol{x}) = \sum_{k=1}^{r} \Delta x_k \cdot \partial_k \langle \boldsymbol{f} \rangle$
- The fitness flux is in principle observable even if the fitness itself is not
- In a fixed landscape, $\int \mathrm{d}t \; \dot{\Phi}(\boldsymbol{x}_t) = \langle f \rangle_{x_T} - \langle f \rangle_{x_0}$
- In a seascape, $\mathrm{d} \langle f \rangle_{\boldsymbol{x}} = \mathrm{d}\Phi(\boldsymbol{x}) + \mathrm{d}t \; \langle \partial_t f \rangle_{\boldsymbol{x}_t}$

Adaptation flux

MUSTONEN AND LÄSSIG, 2010

- Consider the history $\hat{\boldsymbol{x}} = (x_0, x_1, \ldots, x_T)$ with fitness $\boldsymbol{f}(t)$, $0 \le t \le T$
- Consider the *reverse history* $\tilde{\boldsymbol{x}} = (\tilde{x}_0 = x_T, \ldots, \tilde{x}_T = x_0)$ with fitness $\tilde{\boldsymbol{f}}(t) = \boldsymbol{f}(T - t)$
- Then one can prove the *fluctuation relation*

$$\mathcal{P}(\tilde{\boldsymbol{x}}) = \mathcal{P}(\hat{\boldsymbol{x}})\, \mathrm{e}^{-N\Phi(\hat{\boldsymbol{x}}) + \Delta\mathcal{H}(\hat{\boldsymbol{x}})}$$

where

$$\mathcal{H}(\boldsymbol{x}, t) = \log \frac{P(\boldsymbol{x}, t)}{P_0(\boldsymbol{x})} \qquad \Delta\mathcal{H}(\hat{\boldsymbol{x}}) = \mathcal{H}(\boldsymbol{x}_T, T) - \mathcal{H}(\boldsymbol{x}_0, 0)$$

- In particular

$$\left[\mathrm{e}^{-N\Phi + \Delta\mathcal{H}}\right]_{\mathrm{av}} = 1$$

hence

$$\boxed{[N\Phi]_{\mathrm{av}} \ge [\Delta\mathcal{H}]_{\mathrm{av}}}$$

General bibliography

These are some of the texts I used in the preparation of these lectures. The bibliography makes no attempt at fairness nor completeness.

- John H. Gillespie, *Population Genetics: A Concise Guide* (2nd ed.) (Baltimore: Johns Hopkins Press, 2004)
- John Wakeley, *Coalescent Theory: An Introduction* (Greenwood Village (CO): Robertis & Co., 2009)
- Jotun Hein, Mikkel H. Schierup, Carsten Wiuf, *Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory* (Oxford: Oxford U.P., 2005)
- Fabio A.C.C. Chalub, José Francisco Rodrigues (eds.), *The Mathematics of Darwin's Legacy* (Basel: Birkhäuser, 2011)
- JSTAT: Journal of Statistical Mechanics: Theory and Experiment, special issue January 2013
- Nathanaël Berestycki, Recent Progress in Coalescent Theory, Ensaios Matemáticos **16**, 1–103 (2009)

General bibliography

- Su-Chan Park and Joachim Krug, Evolution in random fitness landscapes: the infinite sites model, *J. Stat. Mech.* **4** P04014 (2008)
- Su-Chan Park and Joachim Krug, Clonal interference in large populations, *PNAS* **104** 18135–18140 (2007)
- Michael M. Desai, Daniel S.Fisher, Andrew W. Murray, The Speed of Evolution and Maintenance of Variation in Asexual Populations, *Current Biology* **17**, 385–394 (2007)
- Craig A. Fogle, James L. Nagle, and Michael M. Desai, Clonal Interference, Multiple Mutations, and Adaptation in Large Asexual Populations, *Genetics* **180** 2163–2173 (2008)
- Éric Brunet, Igor M. Rouzine and Claus O. Wilke, The Stochastic Edge in Adaptive Evolution, *Genetics* **179** 603—620 (2008)
- Ville Mustonen and Michael Lässig, From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation, *Trends in Genetics* **25** 111–119 (2009)
- V. Mustonen and M. Lässig, M. Adaptations to fluctuating selection in Drosophila, *PNAS* **104** 2277—2282 (2007)
- Ville Mustonen and Michael Lässig, Fitness flux and ubiquity of adaptive evolution*PNAS* **107** 4248–4253 (2010)