Population Genetics and Evolution – I
The Mechanisms of Evolution: Reproduction and Selection

Luca Peliti

Dipartimento di Fisica and Sezione INFN
Università di Napoli "Federico II"

Helsinki / June 2013

I've got something to tell you before I start speaking.

R. SMULLYAN

Outline

Introduction

An Apology

I attempted mathematics, and even went during the summer of 1828 with a private tutor (a very dull man) to Barmouth, but I got on very slowly. The work was repugnant to me, chiefly from my not being able to see any meaning in the early steps in algebra. This impatience was very foolish, and in after years I have deeply regretted that I did not proceed far enough at least to understand something of the great leading principles of mathematics, for men thus endowed seem to have an extra sense.

CH. DARWIN

The Long-Term Evolution Experiment (LTEE) on *E. coli*

R. Lenski *et al.*, 1988–present

. . . the uniqueness of evolutionary history is itself amenable to careful experimental analysis, and. . . may be an inevitable consequence of the "laws" of microevolution.
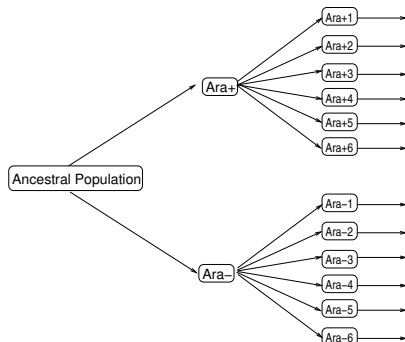
LENSKI & TRAVISANO, 1994

You have the luxury of making a prediction, and then you can test it. It's almost like physics.
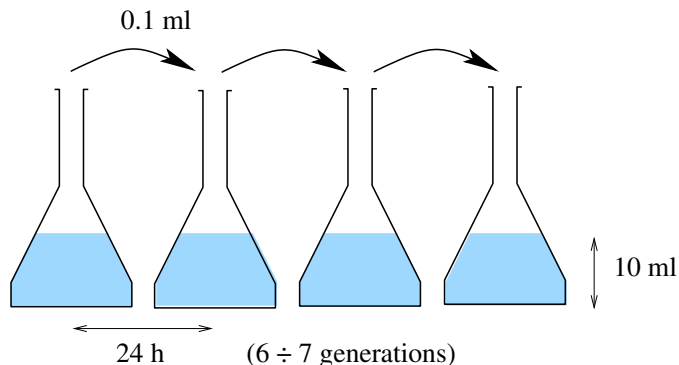
TRAVISANO, 1999

The scheme of the LTEE

The "twelve founding tribes"

12 independent *E. coli* populations, originating from a single clone, 6 Ara+ (can use ℓ-arabinose as a substrate) and 6 Ara-

The protocol of the LTEE

Parallel propagation of the 12 independent *E. coli* populations; standard density $\sim 5 \cdot 10^7$ cells per ml

Features of the LTEE

- 12 independently evolved populations in *glucose-poor* environment
- More than 40 000 generations till now
- Probably billions of "simple" mutations have occurred in each strain
- Only about 10 ÷ 20 mutations have reached *fixation* during the experiment
- Some evolutionary trends are common to all strains (e.g., larger and rounder cells, higher fitness on glucose)
- Four strains have evolved into hypermutators
- One major innovation (ability to metabolize citrate) evolved around generation 31 500 in only one strain

Features of the LTEE

- 12 independently evolved populations in *glucose-poor* environment
- More than 40 000 generations till now
- Probably billions of "simple" mutations have occurred in each strain
- Only about 10 ÷ 20 mutations have reached *fixation* during the experiment
- Some evolutionary trends are common to all strains (e.g., larger and rounder cells, higher fitness on glucose)
- Four strains have evolved into hypermutators
- One major innovation (ability to metabolize citrate) evolved around generation 31 500 in only one strain

Adaptation to citrate

The front central flask (labelled A-3) has a higher turbidity than the others, since it has evolved to use the citrate present in the medium
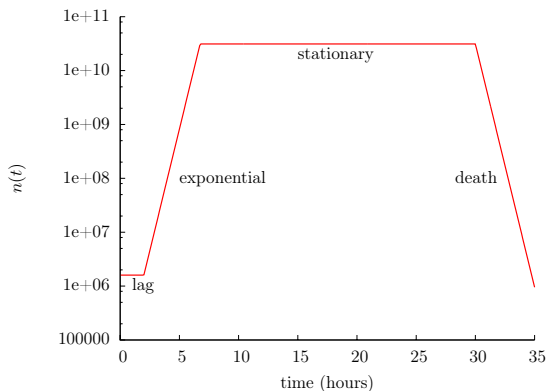
PHOTO COURTESY OF BRIAN BAER AND NEERJA HAJELA

The mechanisms of evolution

- Reproduction
- Selection
- Mutation

Reproduction

The bacterial growth curve

Schematic plot of the population size $n(t)$ vs. time t in a 10ml growth medium flask inoculated with 50 μl of *E. coli* culture.

Simple exponential growth

- Generation time $\tau \sim 20$ min
- Expected population at time t: $n(t) = n_0 2^{[t/\tau]}$
- This assumes that the cells are *synchronized* (discrete generations)
- Assume that cells reproduce *on average* once within τ: $n(t) = \lceil n_0\, 2^{t/\tau} \rceil$ (overlapping generations)
- We shall often switch between discrete time (discrete generations) and continuous time (overlapping generations)
- 72 generations in one day: from one individual, in two days, $2^{144} \simeq 2 \cdot 10^{43}$ individuals, $m_{\text{tot}} \simeq 2 \cdot 10^{28}$ kg ~ 3000 Earth

Simple exponential growth

- Generation time $\tau \sim 20 \ \mathrm{min}$
- Expected population at time t: $n(t) = n_0 2^{[t/\tau]}$
- This assumes that the cells are *synchronized* (discrete generations)
- Assume that cells reproduce *on average* once within τ: $n(t) = \left\lceil n_0 \, 2^{t/\tau} \right\rceil$ (overlapping generations)
- We shall often switch between discrete time (discrete generations) and continuous time (overlapping generations)
- 72 generations in one day: from one individual, in two days, $2^{144} \simeq 2 \cdot 10^{43}$ individuals, $m_{\mathrm{tot}} \simeq 2 \cdot 10^{28} \ \mathrm{kg} \sim 3000 \ \mathrm{Earth}$

Simple exponential growth

- Generation time $\tau \sim 20$ min
- Expected population at time t: $n(t) = n_0 2^{[t/\tau]}$
- This assumes that the cells are *synchronized* (discrete generations)
- Assume that cells reproduce *on average* once within τ: $n(t) = \left\lceil n_0 \, 2^{t/\tau} \right\rceil$ (overlapping generations)
- We shall often switch between discrete time (discrete generations) and continuous time (overlapping generations)
- 72 generations in one day: from one individual, in two days, $2^{144} \simeq 2 \cdot 10^{43}$ individuals, $m_{\text{tot}} \simeq 2 \cdot 10^{28}$ kg ~ 3000 Earth

Simple exponential growth

- Generation time $\tau \sim 20$ min
- Expected population at time t: $n(t) = n_0 2^{[t/\tau]}$
- This assumes that the cells are *synchronized* (discrete generations)
- Assume that cells reproduce *on average* once within τ: $n(t) = \lceil n_0 \, 2^{t/\tau} \rceil$ (overlapping generations)
- We shall often switch between discrete time (discrete generations) and continuous time (overlapping generations)
- 72 generations in one day: from one individual, in two days, $2^{144} \simeq 2 \cdot 10^{43}$ individuals, $m_{\text{tot}} \simeq 2 \cdot 10^{28}$ kg ~ 3000 Earth

Simple exponential growth

- Generation time $\tau \sim 20$ min
- Expected population at time t: $n(t) = n_0 2^{[t/\tau]}$
- This assumes that the cells are *synchronized* (discrete generations)
- Assume that cells reproduce *on average* once within τ: $n(t) = \left\lceil n_0\, 2^{t/\tau} \right\rceil$ (overlapping generations)
- We shall often switch between discrete time (discrete generations) and continuous time (overlapping generations)
- 72 generations in one day: from one individual, in two days, $2^{144} \simeq 2 \cdot 10^{43}$ individuals, $m_{tot} \simeq 2 \cdot 10^{28}$ kg ~ 3000 Earth

Simple exponential growth

- Generation time $\tau \sim 20$ min
- Expected population at time t: $n(t) = n_0 2^{[t/\tau]}$
- This assumes that the cells are *synchronized* (discrete generations)
- Assume that cells reproduce *on average* once within τ: $n(t) = \left[n_0\, 2^{t/\tau} \right]$ (overlapping generations)
- We shall often switch between discrete time (discrete generations) and continuous time (overlapping generations)
- 72 generations in one day: from one individual, in two days, $2^{144} \simeq 2 \cdot 10^{43}$ individuals, $m_{\text{tot}} \simeq 2 \cdot 10^{28}$ kg ~ 3000 Earth

The Galton-Watson (GW) process

- Cells may *die* before reproducing
- Reproduction and death is a *random* process
- Discrete generations: p_n: probability that a cell has n offspring in the next generation
- p_0: probability of death; p_2: probability of reproduction
- Normalization: $\sum_n p_n = 1$
- Probability $P_n(t)$ that the population size equals n at time t (assuming $n(0) = 1$)
- Ultimate extinction probability:

$$Q = \lim_{t \to \infty} P_0(t)$$

Solving the GW process

- Consider the process after t generations, with pop size n
- If $n = 1$ for $t = 0$, then $n = k$ for $t = 1$ with probability p_k
- Then the probability that the issue of one of the k individuals present at $t = 1$ is equal to m is given by $P_m(t-1)$
- Thus $P_n(t)$ satisfies

$$P_m(t) = \sum_k p_k \sum_{\{m_1 \cdots m_k\}} \delta_{\sum_j m_j, n} \prod_{j=1}^k P_{m_j}(t-1)$$
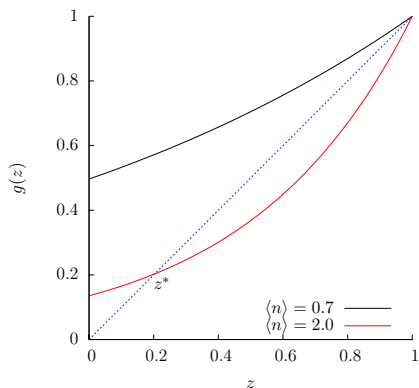
The generating function

- Define $\Gamma(z, t) = \sum_n z^n P_n(t)$ and $g(z) = \sum_n z^n p_n$

- Then

$$\Gamma(z, t) = \sum_k p_k \left[\sum_m z^m P_m(t - 1) \right]^k = g\left(\Gamma(z, t - 1)\right)$$

- Since $\Gamma(z, 0) = z$ we have $\Gamma(z, t) = \underbrace{g(g(\cdots g(z)))}_{t \text{ times}}$

- $Q = \lim_{t \to \infty} \Gamma(0, t) = \lim_{t \to \infty} \underbrace{g(g(\cdots g(0)))}_{t \text{ times}}$

Graphical solution

- $\langle n \rangle = \sum_n n p_n = g'(1)$
- If $\langle n \rangle < 1$, $\lim_{t \to \infty} \Gamma(z, t) = 1$ (**subcritical regime**)
- If $\langle n \rangle > 1$, $\lim_{t \to \infty} \Gamma(z, t) = z^* < 1$ (**supercritical regime**)
- In the supercritical regime, the average size grows exponentially with t

The Galton-Watson process in continuous time

Definition of the process:

- The pop size $n(t)$ is an integer $\forall t$; $n(0) = 1$
- Probability that a given individual is replaced by k
  ($k = 0, 2, 3, \ldots$) individuals in a short interval $\mathrm{d}t$: $\lambda \pi_k \, \mathrm{d}t$ with
  $\sum_k \pi_k = 1$
- Probability that it stays put: $1 - \lambda \, \mathrm{d}t$

The Galton-Watson process in continuous time

Definition of the process:

- The pop size $n(t)$ is an integer $\forall t$; $n(0) = 1$
- Probability that a given individual is replaced by k
  ($k = 0, 2, 3, \ldots$) individuals in a short interval $\mathrm{d}t$: $\lambda \pi_k \, \mathrm{d}t$ with
  $\sum_k \pi_k = 1$
- Probability that it stays put: $1 - \lambda \, \mathrm{d}t$

Solution:

- Events in a short initial interval:

$$
\begin{aligned}
P_n(t + \mathrm{d}t) = {} & (1 - \lambda \mathrm{d}t) P_n(t) \\
& + \lambda \mathrm{d}t \sum_k \pi_k \sum_{\{n_1, \ldots, n_k\}} \delta_{\sum_k n_k, n} \prod_{j=1}^{k} P_{n_j}(t) + \mathrm{O}\left(\mathrm{d}t^2\right)
\end{aligned}
$$

Generating functions

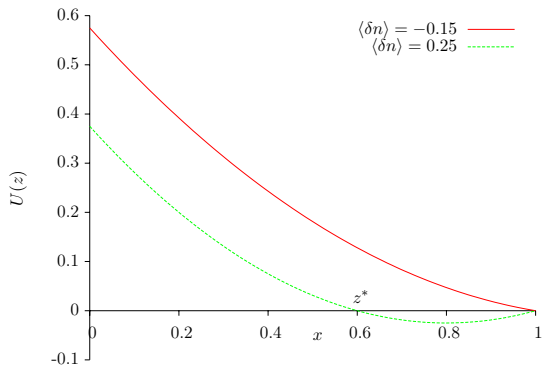- Define $\Gamma(z, t) = \sum_n z^n P_n(t)$ and $g(z) = \sum_k z^k \pi_k$

- Then

$$\Gamma(z, t+\mathrm{d}t) = (1 - \lambda\,\mathrm{d}t)\Gamma(z, t) + \lambda\,\mathrm{d}t \sum_k \pi_k \left(\Gamma(z, t)\right)^k + \mathrm{O}\left(\mathrm{d}t^2\right)$$

- Thus, with $U(z) = g(z) - z$

$$\frac{\partial \Gamma}{\partial t} = \lambda U\left(\Gamma(z, t)\right)$$

- Fixed point: $U(z^*) = 0$, $U'(z^*) < 0$

- $U'(1) = \langle \delta n \rangle$; if $\langle \delta n \rangle < 0$, $z^* = 1$ is the only stable fixed point; if $\langle \delta n \rangle > 0$, the stable fixed point is at $z^* < 1$

Graphical solution

Survival probability as a function of *n*

- What is the survival probability S_n of a population with size n?

- Assume that only π_0 and π_2 do not vanish, and define $s = 1 - (\pi_0/\pi_2)$

- Then S_n satisfies

$$S_n = \frac{1-s}{2-s}S_{n-1} + \frac{1}{2-s}S_{n+1}$$

- Boundary conditions: $S_0 = 0$; $\lim_{n\to\infty} S_n = 1$. Thus, for $s > 0$ (i.e., $\pi_2 > \pi_0$):

$$S_n = 1 - (1-s)^n$$

- For $s < 0$ one has $S_n = 0$, $\forall n$

- In particular, for $s > 0$, $S_1 = s$

Minimal population size

Find the minimal pop size n^* such that the survival probability S_n exceeds $1 - \gamma$

- Extinction probability for $n = 1$: z^*

- Probability that *all* n lineages get extinct: z^{*n}

- This must be smaller than γ, hence

$$n > n^* = \frac{\log \gamma}{\log z^*}$$

- When only π_0 and π_2 do not vanish, $z^* = \frac{\pi_0}{\pi_2} = 1 - s$

$$n^* = \frac{\log \gamma}{\log(1 - s)} \simeq \frac{|\log \gamma|}{s}$$

The Malthus-Verhulst equation

Assumptions:

- Continuous time t and *real values* of n
- Reproduction rate decreases linearly with n
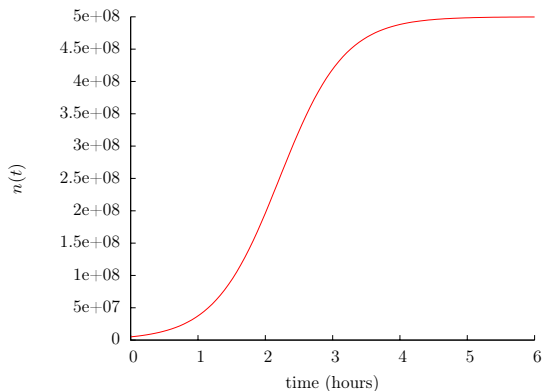- Fluctuations are neglected

Evolution equation for n:

$$\frac{\mathrm{d}n}{\mathrm{d}t} = r \left(1 - \frac{n}{K} \right) n$$

r is the basic reproduction rate, K the carrying capacity

Solution:

$$n(t) = \frac{K}{1 + \left(\frac{1}{n_0} - \frac{1}{K} \right) K \mathrm{e}^{-rt}}$$

The logistic function

Solution of the Malthus-Verhulst equation with
$r = 3 \log 2 \, \text{hour}^{-1}$, $n_0 = 5 \cdot 10^6$, $K = 5 \cdot 10^8$, as in a simple
description of the LTEE

Selection

Charles Darwin

Charles Darwin

Natural Selection

. . . can we doubt (remembering that many more individuals are born than can possibly survive) that individuals having any advantage, however slight, over others, would have the best chance of surviving and of procreating their kind? On the other hand, we may feel sure that any variation in the least degree injurious would be rigidly destroyed. This preservation of favourable variations and the rejection of injurious variations, I call Natural Selection. Variations neither useful nor injurious would not be affected by natural selection, and would be left a fluctuating element, as perhaps we see in the species called polymorphic.

CH. DARWIN, 1859

Natural Selection

My reflection when I first made myself master of the central idea of the Origin was, "How extremely stupid not to have thought of that."

T. H. HUXLEY

Fitness

- Two kinds of individuals, A and B, with populations $n_A(t)$ and $n_B(t)$ at *discrete* generation t
- **Fitness:** *Expected* # of offspring of an individual of type i: W_i, $i \in \{A, B\}$
- Assumptions:
    1. All individuals reproduce independently
    2. The offspring of an individual has the same value of W as its parent
    3. Fluctuations are neglected
- Then

$$n_i(t + 1) = W_i n_i(t) \qquad i \in \{A, B\}$$

Change in composition

- Fraction $x_i(t) = \frac{n_i(t)}{\sum_j n_j(t)}, \quad i, j \in \{A, B\}$

- Then

$$
\begin{aligned}
x(t+1) &= \frac{n_A(t+1)}{n_A(t+1) + n_B(t+1)} = \frac{W_A n_A(t)}{W_A n_A(t) + W_B n_B(t)} \\
&= \frac{W_A x(t)}{W_B + (W_A - W_B)x(t)} = \frac{W_A}{\langle W \rangle_x} x(t)
\end{aligned}
$$

  where $\langle W \rangle_x = W_A x_A + W_B x_B$

- Setting $W_A / W_B = 1 + s$, with $s > 0$,

$$
x(t+1) = x(t) \frac{1 + s}{1 + sx(t)} \geq x(t)
$$

The Fundamental Theorem

- $\boldsymbol{x}(t) = (x_1(t), \ldots, x_r(t))$, fitness W_k, $k = 1, \ldots, r$
- Mean fitness:

$$\langle W \rangle_{\boldsymbol{x}} = \sum_{k=1}^{r} W_k x_k$$

- Then

$$
\begin{aligned}
x_k(t+1) &= \frac{W_k x_k(t)}{\langle W \rangle_{\boldsymbol{x}(t)}} \\
\Delta \langle W \rangle_{\boldsymbol{x}(t)} &= \langle W \rangle_{\boldsymbol{x}(t+1)} - \langle W \rangle_{\boldsymbol{x}(t)} \\
&= \frac{1}{\langle W \rangle_{\boldsymbol{x}(t)}} \left(\sum_{k=1}^{r} W_k^2 x_k - \langle W \rangle_{\boldsymbol{x}(t)}^2 \right) \\
&= \frac{\langle W^2 \rangle_{\boldsymbol{x}(t)} - \langle W \rangle_{\boldsymbol{x}(t)}^2}{\langle W \rangle_{\boldsymbol{x}(t)}} \geq 0
\end{aligned}
$$

About the Fundamental Theorem

We may consequently state the fundamental theorem of Natural Selection in the form: **The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time.**
Professor Eddington has recently remarked that "The law that entropy always increases—the second law of thermodynamics—holds, I think, the supreme position among the laws of nature". It is not a little instructive that so similar a law should hold the supreme position among the biological sciences.

R. A. FISHER, 1930

About the Fundamental Theorem

*So we see, in physics, disorder growing inexorably in
systems isolated from their surroundings: and in
biology, fitness increasing steadily in populations
struggling for life. Ascent here and degradation
there—almost too good to be true.*

K. SIGMUND, 1993

About the Fundamental Theorem

My own view is that it cannot play an important role in biology.

J. MAYNARD SMITH, 1989

Selection in continuous time

Assume simple Malthusian growth

- Population made of r "types", each with growth rate k, $k = 1, \ldots, r$

- Number $n_k(t)$ of individuals of type k at time t:

$$\frac{\mathrm{d}n_k}{\mathrm{d}t} = f_k n_k$$

- Fraction of individuals of type k: $x_k = n_k / \sum_j n_j$:

$$\frac{\mathrm{d}x_k}{\mathrm{d}t} = \left(f_k - \langle f \rangle_{\boldsymbol{x}(t)} \right) x_k$$

- Change in the mean growth rate:

$$\frac{\mathrm{d} \langle f \rangle_{\boldsymbol{x}(t)}}{\mathrm{d}t} = \left\langle f^2 \right\rangle_{\boldsymbol{x}(t)} - \langle f \rangle^2_{\boldsymbol{x}(t)}$$

Selection in the Malthus-Verhulst regime

Assume that the carrying capacity K is the same for all k:

- Equation for $n_k(t)$:

$$\frac{\mathrm{d}n_k}{\mathrm{d}t} = n_k \left(f_k - \frac{N(t)}{K} \langle f \rangle_{\boldsymbol{x}(t)} \right)$$

- Thus

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \underbrace{\langle f \rangle_{\boldsymbol{x}(t)}}_{\text{frequency dependent!}} N \left(1 - \frac{N}{K} \right)$$

- Separation of time scales:
  - Fast: $\sum_k n_k \longrightarrow N$
  - Slow: selection of the "fittest"

Measuring fitness in the LTEE
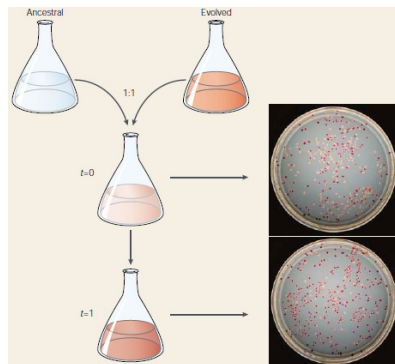
- Samples are mixed in a 1:1 ratio
- Densities are measured by plating on indicator agar
- Initial and final densities $n_{A,B}^{i,f}$
- Expression of the growth rates:

$$f_{A,B} = \ln \frac{n_{A,B}^{f}}{n_{A,B}^{i}}$$

ELENA & LENSKI, 2003

Frequency-dependent selection

- *Negative* frequency-dependent selection: frequent types are selected against
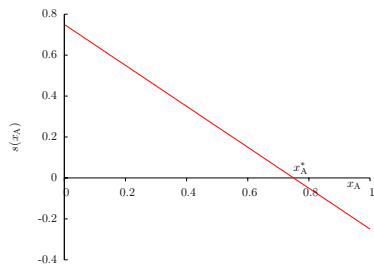
$$s(x_A) = f_A(x_A) - f_B(1 - x_A)$$
$$= s_0 - s_1 x_A$$
$$\frac{dx_A}{dt} = s(x_A) x_A x_B$$

- **Fixed point** at x_A^*: $s(x_A^*) = 0$

- Negative frequency-dependent selection leads to long-term coexistence (*stabilizing selection*)
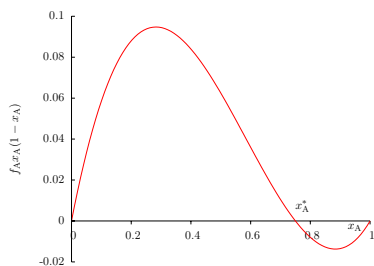
Positive selection
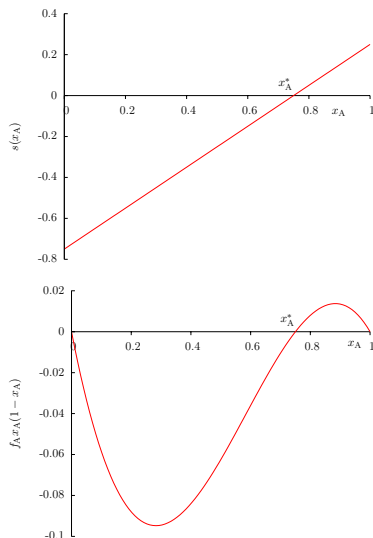
- *Positive* frequency-dependent selection: fitness increases with frequency

$$s(x_A) = s_0 + s_1 x_A$$
$$\frac{dx_A}{dt} = s(x_A)x_A x_B$$

- **Unstable** fixed point at x_A^*: $s(x_A^*) = 0$!

- Positive frequency-dependent selection leads to several possible equilibria (*disruptive selection*)

Frequency-dependent selection and optimization

- Prisoner's dilemma: Payoff matrix

	Cooperator	Defector
Cooperator	1	$1 - s_1$
Defector	$1 + s_2$	$1 - c$

  $(1 - c) > (1 - s_1)$

- Thus $W_C(x) \leq W_D(1 - x)$, $0 \leq x \leq 1$ and the stable fixed point is $x_D = 1$
- However $W_D(1) = 1 - c < W_C(1) = 1$
- Then why is cooperation maintained in Nature?

Prisoner's dilemma in an RNA virus
P. TURNER AND L. CHAO, 1999

- Phage $\phi 6$ at high multiplicity of infection (MOI)
- Evaluation of W_D/W_{wt} for different defector frequencies x

$$\left.\frac{W_D}{W_{wt}}\right|_x = \begin{cases} x \ll 1, & 1 + s_2; \\ x \simeq 1, & (1 - s_1)/(1 - c) > 1 \end{cases}$$

- Evaluation of $1 - c$: allow multiple infection of *only* C or D in the same bacterium: $W_D = 1 - c$

Prisoner's dilemma in an RNA virus

P. TURNER AND L. CHAO, 1999
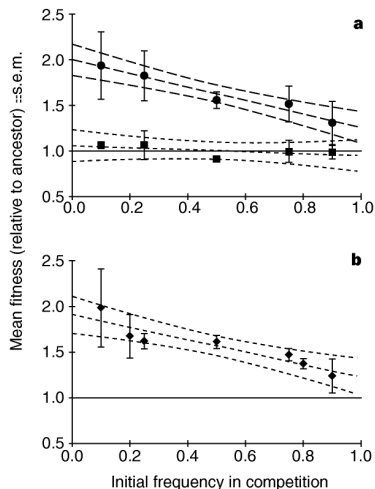
Fitness of two high-MOI derived strains relative to wild type as a function of initial frequency

Prisoner's dilemma in an RNA virus
P. TURNER AND L. CHAO, 1999

Rock-scissors-paper game in *Uta stansburiana*
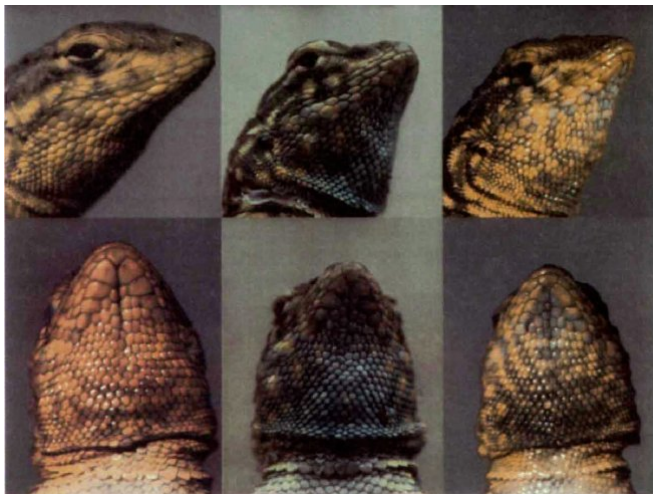
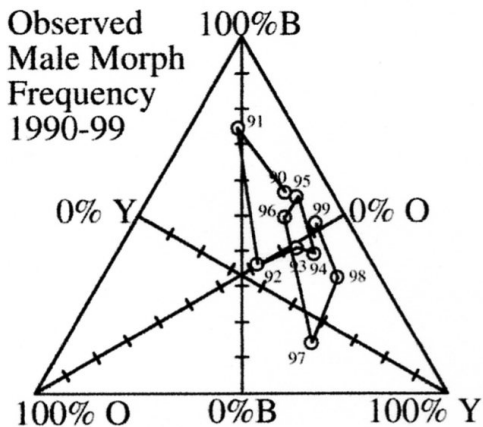B. SINERVO AND C. LIVELY, 1996

Games lizards play
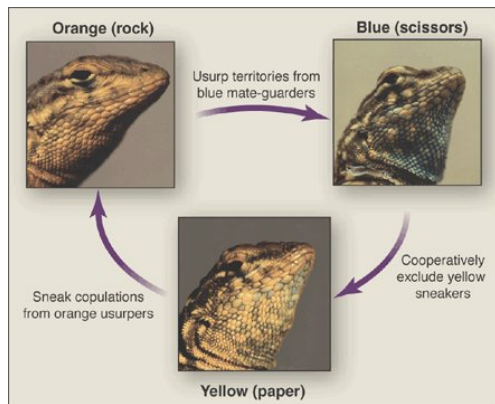
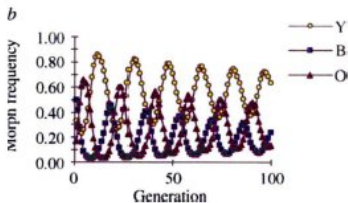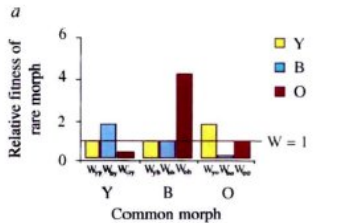B. SINERVO AND C. LIVELY, 1996

Field data
De Finetti diagram

B. SINERVO AND C. LIVELY, 1996

Fitness model

B. SINERVO AND K. R. ZAMUDIO, 2001

The Fundamental Theorem revisited

- r types, frequency-dependent fitness $f_k(\boldsymbol{x})$, $\boldsymbol{x} = (x_1, \ldots, x_r)$
- Evolution equation for $\boldsymbol{x}$:

$$\frac{\mathrm{d}x_k}{\mathrm{d}t} = \left(f_k(\boldsymbol{x}) - \langle f \rangle_{\boldsymbol{x}(t)} \right) x_k$$

- Change in $\langle f \rangle_{\boldsymbol{x}(t)}$:

$$\frac{\mathrm{d}\langle f \rangle_{\boldsymbol{x}(t)}}{\mathrm{d}t} = \sum_k \left[\underbrace{f_k \frac{\mathrm{d}x_k}{\mathrm{d}t}}_{\langle f^2 \rangle_{\boldsymbol{x}} - \langle f \rangle_{\boldsymbol{x}}^2} + \sum_j \frac{\partial f_j}{\partial x_k} \frac{\mathrm{d}x_k}{\mathrm{d}t} x_j \right]$$

Comments

What Fisher's theorem tells us is that natural selection (in his restricted meaning involving only additive effects) at all times acts to increase the fitness of a species to live under the conditions that existed an instant earlier. But since this standard of "fitness" changes from instant to instant, this constant improving tendency of natural selection does not necessarily get anywhere in terms of increasing "fitness" as measured by any fixed standard, and in fact M [mean fitness] is as likely to decrease under natural selection as to increase.

G. PRICE, 1972