Population Genetics and Evolution – I

The Mechanisms of Evolution: Reproduction and Selection

Luca Peliti

Bengaluru / December 2017

SMRI (Italy)
`luca@peliti.org`

Outline

In vitro evolution experiments

Reproduction

Selection

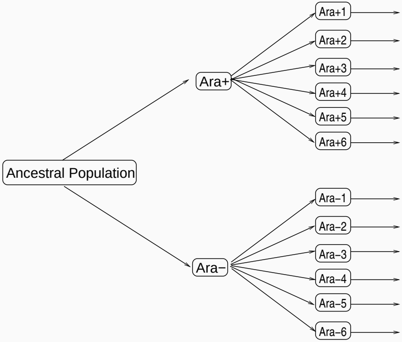
In vitro evolution experiments

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

> *… 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

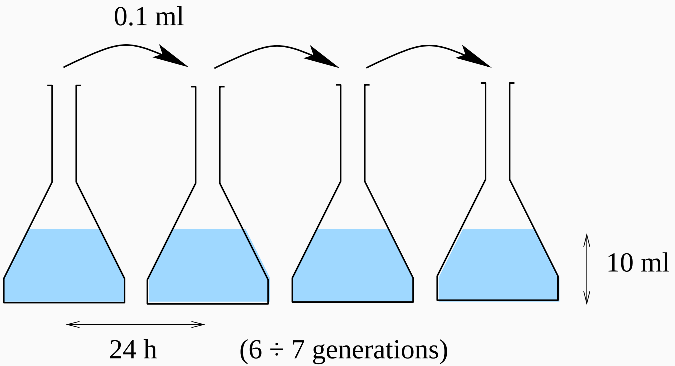
> *You have the luxury of making a prediction, and then you can test it. It’s almost like physics.*
>
> Travisano, 1999

The scheme of the LTEE



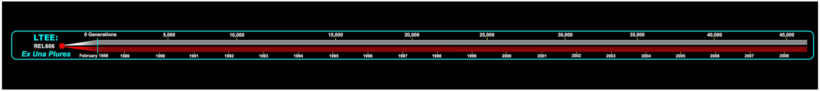
12 independent *E. coli* populations, originating from a single clone, 6 Ara+ (can use ℓ-arabinose as a substrate) and 6 Ara-

The protocol of the LTEE



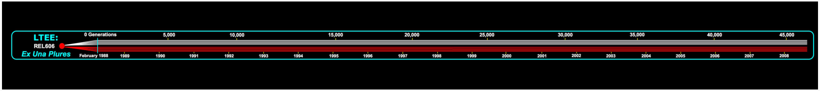
Parallel propagation of the 12 independent *E. coli* populations;
standard density $\sim 5 \cdot 10^7$ cells per ml

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

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

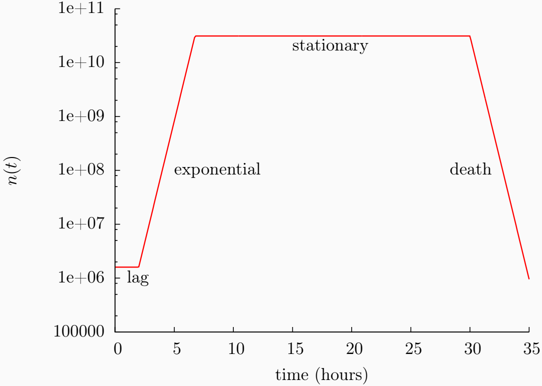
Photo courtesy of Brian Baer and Neerja Hajela

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 $\tau \sim 20\ \mathrm{min}$
- Expected population at time t: $n(t) = n_0 2^{\lfloor t/\tau \rfloor}$
- This assumes that the cells are *synchronized* (discrete generations)
- Assume that cells reproduce *on average* once within τ: $n(t) = \left\lfloor n_0\, 2^{t/\tau} \right\rfloor$ (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, $m_{\mathrm{tot}} \simeq 2 \cdot 10^{28}\ \mathrm{kg} \sim 3000\ \mathrm{Earth}$

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

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

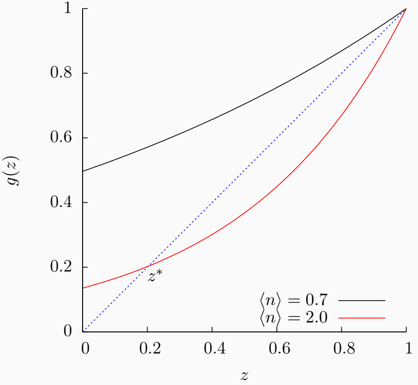
The generating function

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

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:

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

Generating functions

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

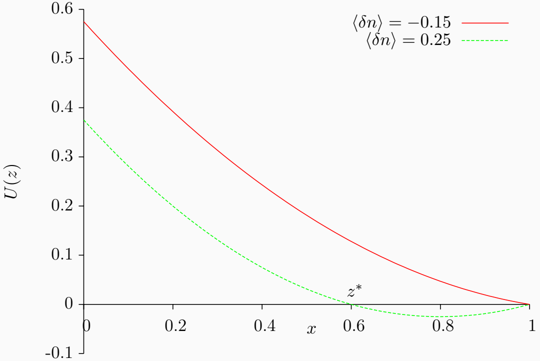
$$\Gamma(z, t + \mathrm{d}t) = (1 - \lambda\,\mathrm{d}t)\Gamma(z,t) + \lambda\,\mathrm{d}t \sum_k \pi_k \left(\Gamma(z,t)\right)^k + \mathrm{O}\left(\mathrm{d}t^2\right)$$

- 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:

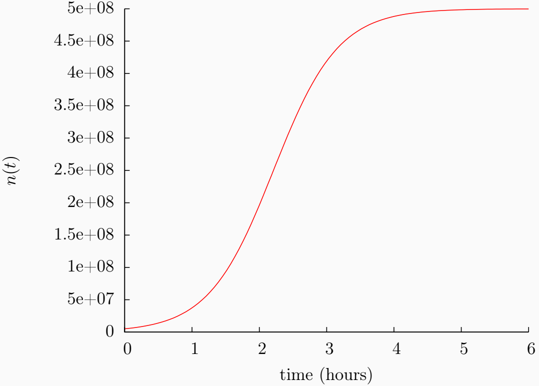
$$\frac{\mathrm{d}n}{\mathrm{d}t} = r\left(1 - \frac{n}{K}\right)n$$

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 $r = 3 \log 2 \, \mathrm{hour}^{-1}$, $n_0 = 5 \cdot 10^6$, $K = 5 \cdot 10^8$, 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 $n_{\mathrm{A}}(t)$ and $n_{\mathrm{B}}(t)$ at *discrete* generation t
- **Fitness:** *Expected* # of offspring of an individual of type i: W_i, $i \in \{\text{A, B}\}$
- 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

$$n_{\mathrm{i}}(t+1) = W_{\mathrm{i}} n_{\mathrm{i}}(t) \qquad i \in \{\text{A, B}\}$$

Change in composition

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

$$\begin{aligned} x(t+1) &= \frac{n_\mathrm{A}(t+1)}{n_\mathrm{A}(t+1) + n_\mathrm{B}(t+1)} = \frac{W_\mathrm{A} n_\mathrm{A}(t)}{W_\mathrm{A} n_\mathrm{A}(t) + W_\mathrm{B} n_\mathrm{B}(t)} \\ &= \frac{W_\mathrm{A} x(t)}{W_\mathrm{B} + (W_\mathrm{A} - W_\mathrm{B}) x(t)} = \frac{W_\mathrm{A}}{\langle W \rangle_x} x(t) \end{aligned}$$

where $\langle W \rangle_x = W_\mathrm{A} x_\mathrm{A} + W_\mathrm{B} x_\mathrm{B}$

- Setting $W_\mathrm{A}/W_\mathrm{B} = 1 + s$, 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

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

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, $k = 1, \dots, r$
- 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:

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

Selection in the Malthus-Verhulst regime

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

- Equation for $n_k(t)$:

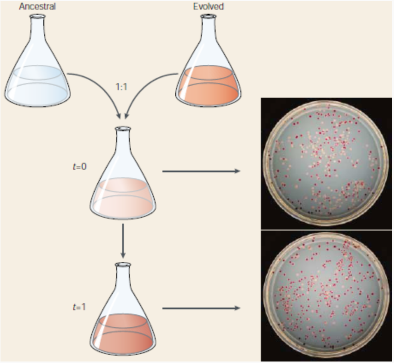
$$\frac{\mathrm{d}n_k}{\mathrm{d}t} = n_k \left(f_k - \frac{N(t)}{K} \langle f \rangle_{\boldsymbol{x}(t)} \right)$$

- 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: $\sum_k n_k \longrightarrow K$
    - Slow: selection of the “fittest”

Measuring fitness in the LTEE

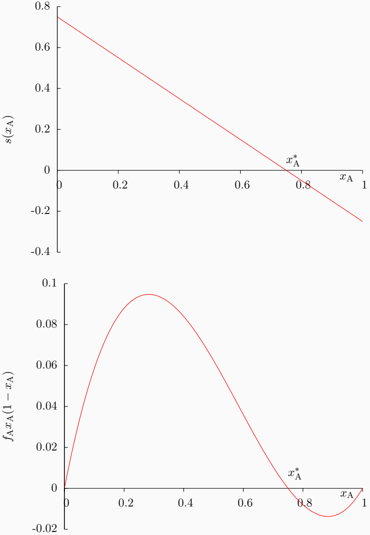


- Samples are mixed in a 1:1 ratio
- Densities are measured by plating on indicator agar
- Initial and final densities $n^{\mathrm{i,f}}_{\mathrm{A,B}}$
- Expression of the growth rates:

$$f_{\mathrm{A,B}} = \ln \frac{n^{\mathrm{f}}_{\mathrm{A,B}}}{n^{\mathrm{i}}_{\mathrm{A,B}}}$$

Elena & Lenski, 2003

Frequency-dependent selection

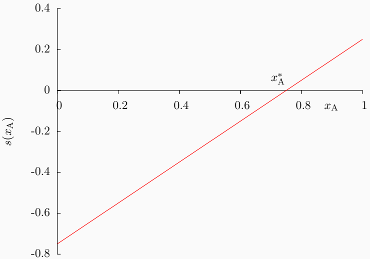


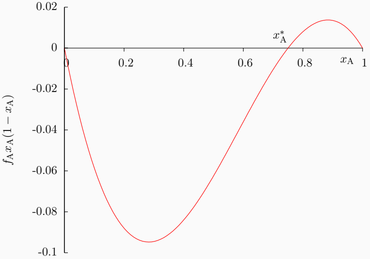
- *Negative* frequency-dependent selection: frequent types are selected against

$$\begin{aligned} s(x_\mathrm{A}) &= f_\mathrm{A}(x_\mathrm{A}) - f_\mathrm{B}(1-x_\mathrm{A}) \\ &= s_0 - s_1 x_\mathrm{A} \\ \frac{\mathrm{d}x_\mathrm{A}}{\mathrm{d}t} &= s(x_\mathrm{A}) x_\mathrm{A} x_\mathrm{B} \end{aligned}$$

- **Fixed point** at x_A^*: $s(x_\mathrm{A}^*) = 0$
- Negative frequency-dependent selection leads to long-term coexistence (*stabilizing selection*)

Positive selection





- *Positive* frequency-dependent selection: fitness increases with frequency

$$\begin{aligned} s(x_A) &= s_0 + s_1 x_A \\ \frac{\mathrm{d}x_A}{\mathrm{d}t} &= s(x_A) x_A x_B \end{aligned}$$

- **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 $W_{\mathrm{C}}(x) \leq W_{\mathrm{D}}(1-x)$, $0 \leq x \leq 1$ and the stable fixed point is $x_{\mathrm{D}} = 1$
- However $W_{\mathrm{D}}(1) = 1-c < W_{\mathrm{C}}(1) = 1$
- Then why is cooperation maintained in Nature?

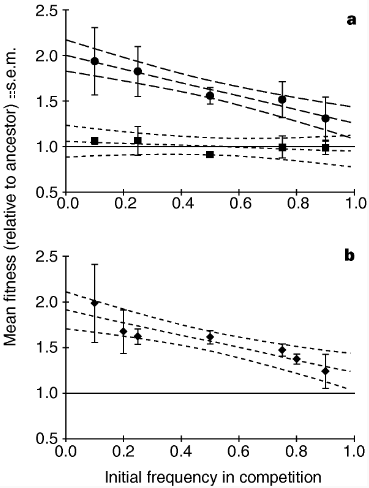
Prisoner's dilemma in an RNA virus

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

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

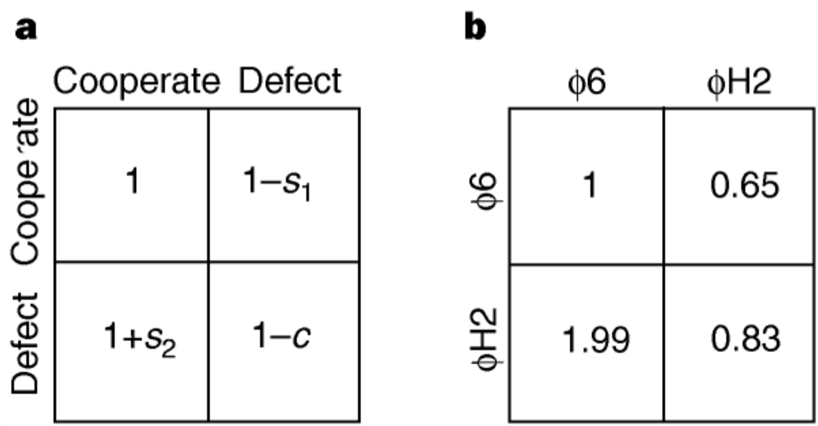
- Evaluation of $1 - c$: allow multiple infection of *only* C or D in the same bacterium: $W_{\mathrm{D}} = 1 - c$

Prisoner’s dilemma in an RNA virus



Fitness of two high-MOI derived strains relative to wild type as a function of initial frequency

Prisoner’s dilemma in an RNA virus



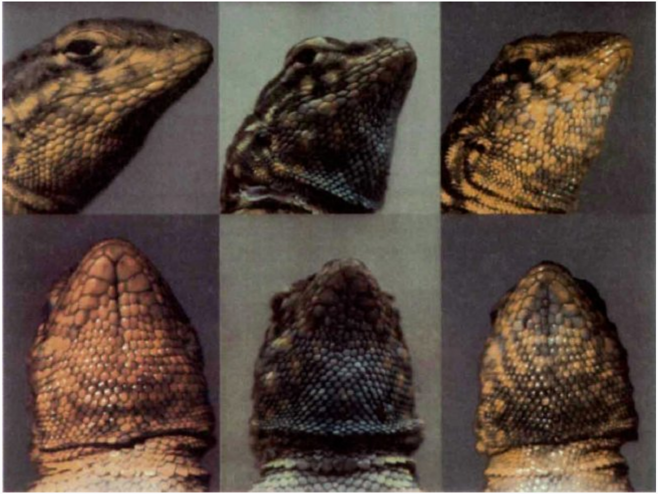
a

	Cooperate	Defect
Cooperate	1	$1-s_1$
Defect	$1+s_2$	$1-c$

b

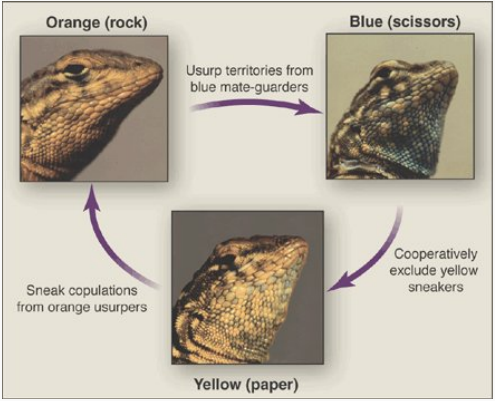
	ϕ6	ϕH2
ϕ6	1	0.65
ϕH2	1.99	0.83

Rock-scissors-paper game in *Uta stansburiana*



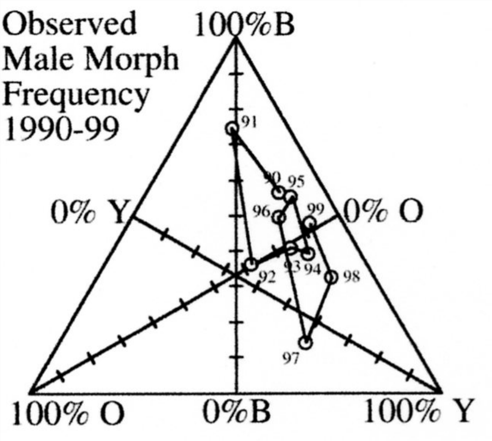
B. Sinervo and C. Lively, 1996

Games lizards play



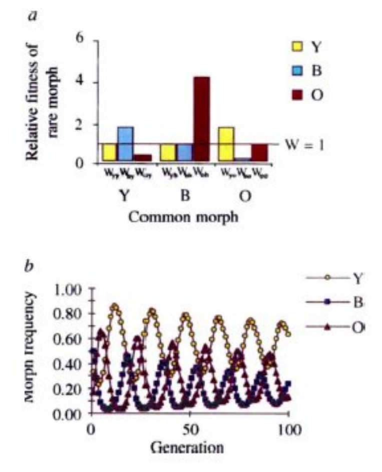
B. Sinervo and C. Lively, 1996

Field data



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 $\boldsymbol{x}$:

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

- 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

Thank you!

References i

These are some of the texts I used in the preparation of these lectures. No attempt at completeness has been made.

- Ch. Darwin, *Autobiography*, in: Ch. Darwin, T. H. Huxley, *Autobiographies* (G. De Beer, ed.) (Oxford: Oxford U. P., 1983)
- T. Appenzeller, Test tube evolution catches time in a bottle. Science, **284** 2108–2110 (1999)
- R. E. Lenski and M. Travisano, Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations, *Proc. Natl. Acad. Sci. USA* **91**, 6808-6814 (1994)
- Zachary D. Blount, Christina Z. Borland, and Richard E. Lenski, Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*, *Proc. Natl. Acad. Sci. USA* **105** 7899-7906 (2008)

References ii

- J. D. Murray, *Mathematical Biology, I. An Introduction* (3rd ed.) (New York/Berlin: Springer, 2002)
- H. W. Watson and F. Galton, On the probability of the extinction of families, *Journal of the Anthropological Institute of Great Britain* **4** 138–144 (1875)
- P.-F. Verhulst, Notice sur la loi que la population suit dans son accroissement, *Correspondance mathématique et physique* **10** 113–121 (1838)
- P.-F. Verhulst, Recherches mathématiques sur la loi d'accroissement de la population, *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles* **18** 1–42 (1845)
- Ch. Darwin, *On the Origin of Species by Means of Natural Selection* (London: J. Murray, 1859)

References iii

- R. A. Fisher, *The Genetical Theory of Natural Selection*. (Oxford: Oxford U. P., 1930). Reprinted by Dover Publications (N.Y.) (1958)
- K. Sigmund, *Games of Life: Explorations in Ecology, Evolution and Behaviour* (Oxford: Oxford U. P., 1993)
- J. Maynard Smith, *Evolutionary Genetics* (Oxford: Oxford U. P., 1989)
- S. F. Elena and R. E. Lenski, Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation, *Nature Reviews Genetics* **4** 457–469 (2003)
- P. E. Turner and L. Chao, Prisoner's dilemma in an RNA virus, *Nature* **398** 441-443 (1999)
- B. Sinervo and C. M. Lively, The rock-paper-scissors game and the evolution of alternative male strategies, *Nature* **380** 240-243 (1996)

References iv

- W. A. Ewens, An interpretation and proof of the fundamental theorem of natural selection, *Theor. Pop. Biol.*, **36** 167–180 (1989)
- S. A. Frank and M. Slatkin, Fisher's fundamental theorem of Natural Selection. *Trends in Ecology & Evolution* **7** 92–95 (1992).
- G. R. Price, Fisher's “fundamental theorem” made clear, *Annals of Human Genetics* **36** 129–140 (1972)