Evolutionary advantage of small populations on complex fitness landscapes

Su-Chan Park

가톨릭대학교
THE CATHOLIC UNIVERSITY OF KOREA

NSPCS 2010, July 27, 2010

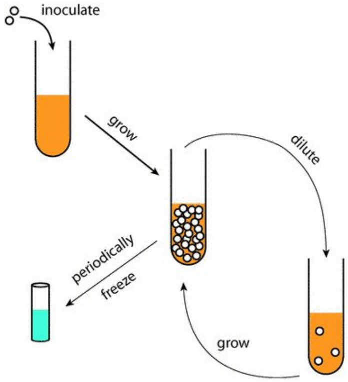
Joint work with J. Krug and K. Jain (arXiv:1003.5380)

Outline

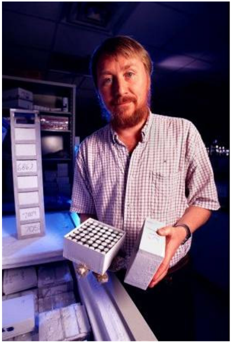
Outline

Evolution in action in the lab

S. F. Elena and R. E. Lenski, Nat. Rev. Genetics **4**, 457 (2003).



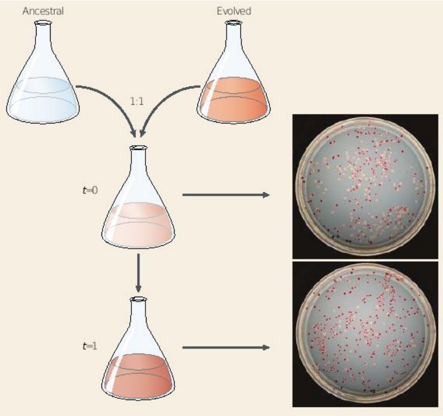
R. Lenski, Michigan State University



50,000 generations in Feb. 2010 (6 generations a day)

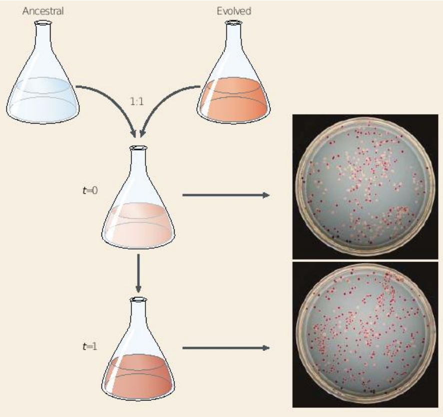
Measuring Fitness

S. F. Elena and R. E. Lenski, Nat. Rev. Genetics **4**, 457 (2003).



Measuring Fitness

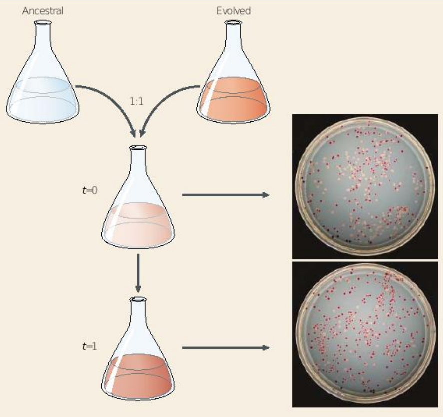
S. F. Elena and R. E. Lenski, Nat. Rev. Genetics **4**, 457 (2003).



$$\frac{w_2}{w_1} \stackrel{?}{=} \frac{w_2}{w_0} \Big/ \frac{w_1}{w_0}$$

Measuring Fitness

S. F. Elena and R. E. Lenski, Nat. Rev. Genetics **4**, 457 (2003).

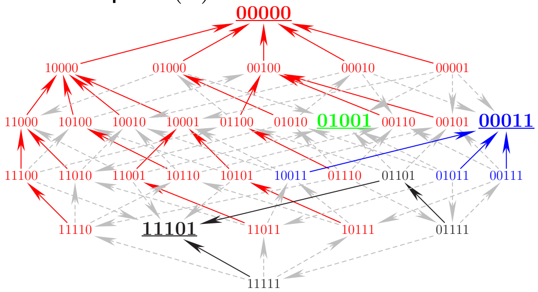


$$\frac{w_2}{w_1} \stackrel{?}{=} \frac{w_2}{w_0} \bigg/ \frac{w_1}{w_0}$$

cf: Kerr et al. (2002)
rock-paper-scissors game

Fitness and fitness landscape

- fitness $w(\sigma)$: average number of offspring
- genotype $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_L)$ ($\sigma_i = 1$ or 0)
- fitness is a function of genotypes (and environment).
- selection coefficient $s = \dfrac{w}{w'} - 1$.
- fitness landscape $w(\sigma)$:



- adaptation (natural selection): hill-climbing process

speed of adaptation (theory)

- non-epistatic fitness landscape (Levine's talk)

$$w(\sigma) = \prod_{i=1}^{L} \exp(s_i \sigma_i) \Rightarrow \frac{d \ln \bar{w}(t)}{dt} \sim \ln(NU)$$

Review : SCP, D. Simon and J. Krug, JSP **138**, 381 (2010)

- house-of-cards model with infinite number of sites

$$w(\sigma) = \text{random number drawn from } p(w)$$

if $p(w) = \exp(-w)$ and $\rightarrow \bar{w}(t) \sim \ln(NUt)$

SCP and J. Krug, J. Stat. Mech. P04014 (2008).

- a large population adapts faster than a small one.

speed of adaptation (theory)

- non-epistatic fitness landscape (Levine's talk)

$$w(\sigma) = \prod_{i=1}^{L} \exp(s_i \sigma_i) \Rightarrow \frac{d \ln \bar{w}(t)}{dt} \sim \ln(NU)$$

Review : SCP, D. Simon and J. Krug, JSP **138**, 381 (2010)

- house-of-cards model with infinite number of sites

$$w(\sigma) = \text{random number drawn from } p(w)$$

$$\text{if } p(w) = \exp(-w) \text{ and } \rightarrow \bar{w}(t) \sim \ln(NUt)$$

SCP and J. Krug, J. Stat. Mech. P04014 (2008).

- a large population adapts faster than a small one.

speed of adaptation (theory)

- non-epistatic fitness landscape (Levine's talk)

$$w(\sigma) = \prod_{i=1}^{L} \exp(s_i \sigma_i) \Rightarrow \frac{d \ln \bar{w}(t)}{dt} \sim \ln(NU)$$

Review : SCP, D. Simon and J. Krug, JSP **138**, 381 (2010)

- house-of-cards model with infinite number of sites

$$w(\sigma) = \text{random number drawn from } p(w)$$

$$\text{if } p(w) = \exp(-w) \text{ and } \rightarrow \bar{w}(t) \sim \ln(NUt)$$

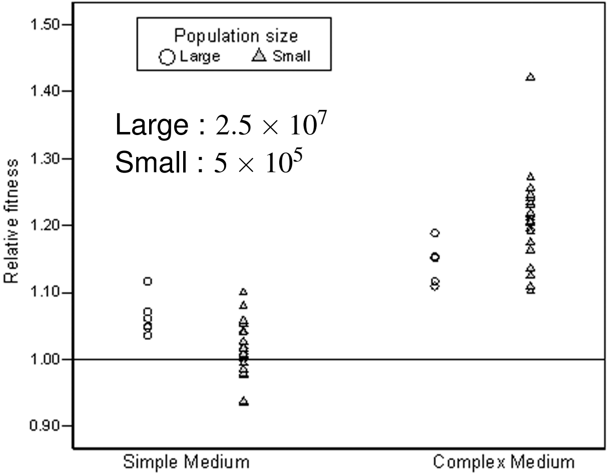
SCP and J. Krug, J. Stat. Mech. P04014 (2008).

- a large population adapts faster than a small one.

Outline

mean fitness after 500 generations

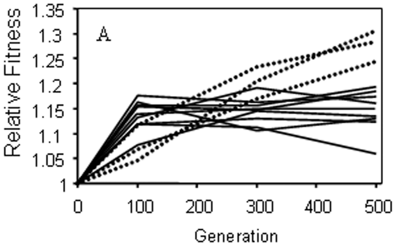
Rozen et al., PLoS one **3**, e1715 (2008)



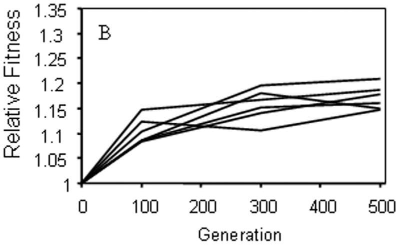
fitness trajectory (complex medium)

Rozen et al., PLoS one **3**, e1715 (2008)

small population

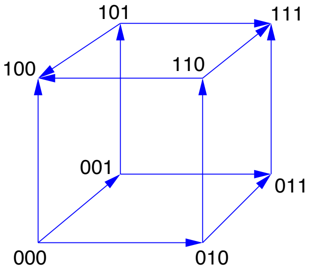


large population



Outline

Simple Model : Three-locus ($L = 3$)

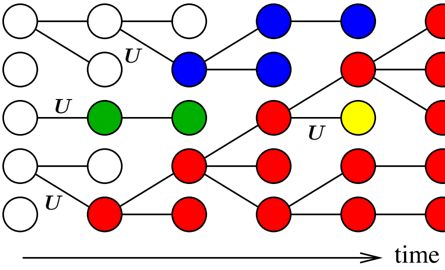


$$w(000) = 1,$$
$$w(001) = w(010) = 1 + s_1,$$
$$w(100) = 1 + s_2,$$
$$w(011) = w(101) = w(110) = (1 + s_1)^2,$$
$$w(111) = (1 + s_1)^2$$

$$(1 + s_1)^2 < 1 + s_2 < (1 + s_1)^3$$

- 000 : global minimum, 111 : global maximum
- smooth and rugged paths to the global maximum

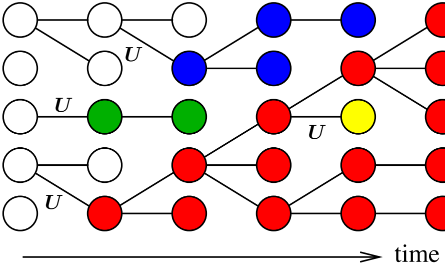
Evolving population: Wright-Fisher Model



$$f(\sigma; t+1) = \sum_{\sigma'} M(\sigma \leftarrow \sigma') \frac{w(\sigma')}{\bar{w}(t)} f(\sigma'; t)$$

- multinomial distribution
- $f(\sigma, t)$: frequency of the genotype σ at generation t.
- $M(\sigma \leftarrow \sigma')$: mutation prob. from σ' to σ.
- $\bar{w}(t) = \sum_{\sigma} w(\sigma) f(\sigma; t)$: mean fitness

Background: Fixation



- fixation probability $\pi(s) \approx \dfrac{1-e^{-2s}}{1-e^{-2sN}}, \quad s = \dfrac{w_{\text{red}}}{w_{\text{white}}} - 1$
- $Ns \gg 1$ and $0 < s \ll 1$, $\pi(s) \approx 2s$
- $Ns \gg 1$: strong selection regime

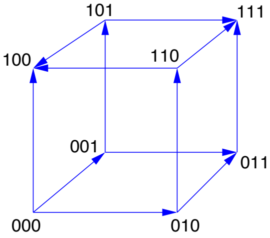
Strong selection-weak mutation (SSWM) regime

- $NU \ll 1$: weak mutation regime $\to$ “adaptive walk”
- Waiting time for (first event of) fixation
  - from $w(000) = 1$ to $w(001) = w(010) = 1 + s_1$, $T_1 \approx \frac{1}{2\mu N s_1}$
  - from $w(000) = 1$ to $w(100) = 1 + s_2$, $T_1 \approx \frac{1}{2\mu N s_2}$
- From local maximum to global maximum

$$T_{\text{esc}} \approx \frac{s_{\text{del}}}{4N\mu^2 s_{\text{ben}}}$$

$$s_{\text{del}} = \frac{w(100)}{w(101)} - 1 \approx s_2 - 2s_1$$

$$s_{\text{ben}} = \frac{w(111)}{w(100)} - 1 \approx 3s_1 - s_2$$



Path Probability

- $P_r(N)$: prob. that a population of size N takes the rugged path
- Reduced problem : three alleles (types) model
  $w(A) = 1$, $w(B) = 1 + s_1$, $w(C) = 1 + s_2$ $(s_2 > s_1)$.
  $M(A \to B) = 2\mu$, $M(A \to C) = \mu$
- $N\mu \ll 1$: fate of mutations is independent from each other

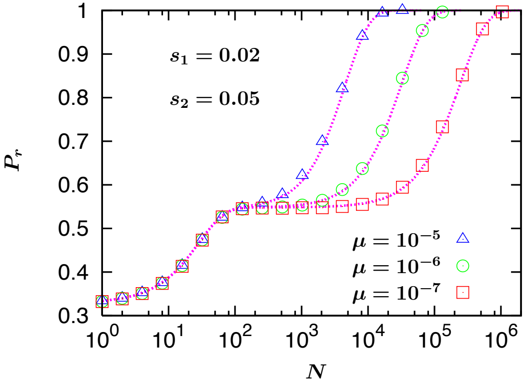
$$P_r|_{N\mu \ll 1} \approx \frac{\pi(s_2)}{\pi(s_2) + 2\pi(s_1)}$$

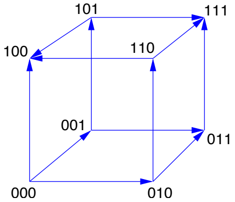
- $N\mu \gg 1$: type C should appear with large numbers in one generation

$$\lim_{N\to\infty} P_r = 1$$

Path probability : approximate formula

$$P_r = 1 - \frac{2\pi(s_1)}{\pi(s_2) + 2\pi(s_1)} \exp\left(-N\mu \ln(Ns_1)\pi(s_2)/s_1\right).$$

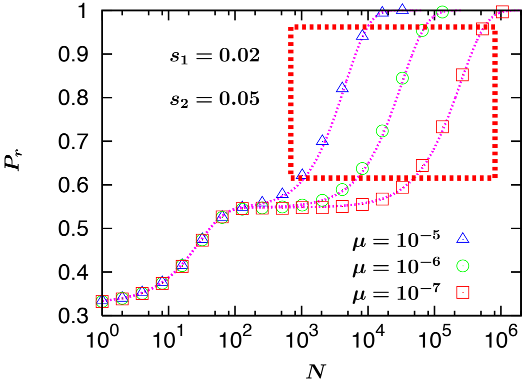


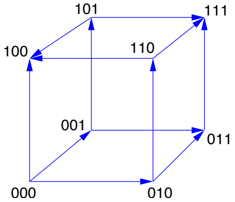


$$\pi(s) = \frac{1 - e^{-2s}}{1 - e^{-2sN}}$$

Path probability : approximate formula

$$P_r = 1 - \frac{2\pi(s_1)}{\pi(s_2) + 2\pi(s_1)} \exp\left(-N\mu \ln(Ns_1)\pi(s_2)/s_1\right).$$

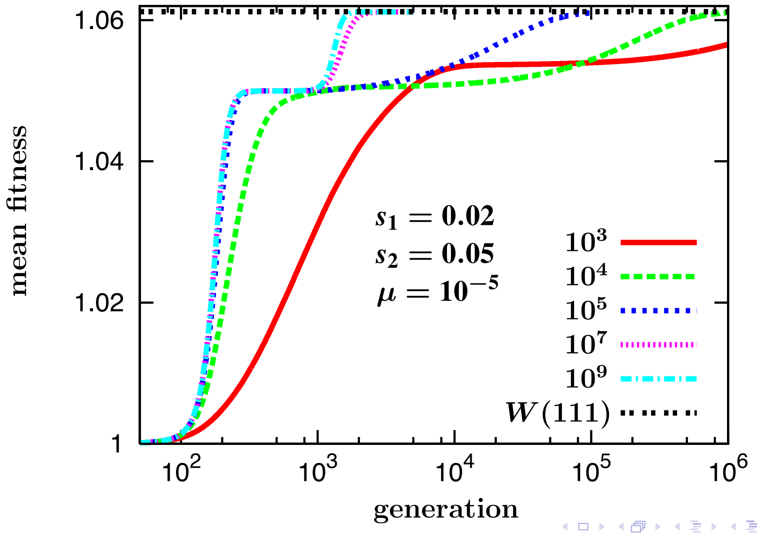




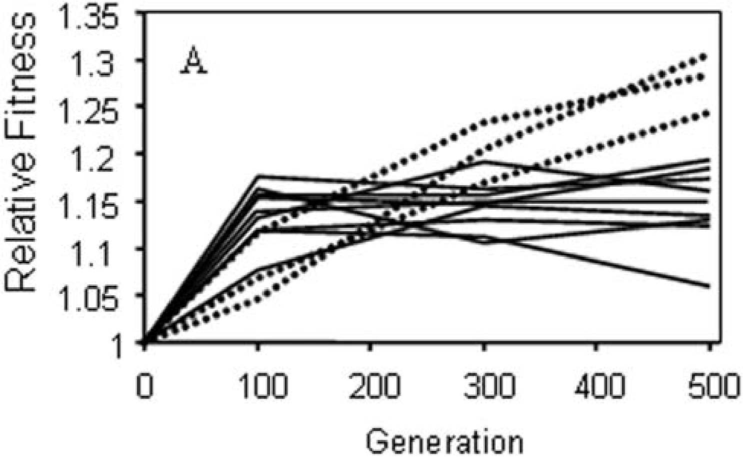
$$\pi(s) = \frac{1 - e^{-2s}}{1 - e^{-2sN}}$$

$$N\mu \ln(Ns) \sim O(1)$$

Evolution on the fitness landscape



Experiment
Rozen et al., PLoS one **3**, e1715 (2008)



Estimating the (beneficial) mutation probability

- size of small population in the experiment $= 5 \times 10^5$
- size of selection coefficient ≈ 0.1
- Criterion of the small population advantage

$$N\mu \log(Ns) \sim 1 \rightarrow \mu \approx 10^{-6}$$

cf: Perfeito et al. (2007) : 10^{-5}

Outline

House-of-cards model

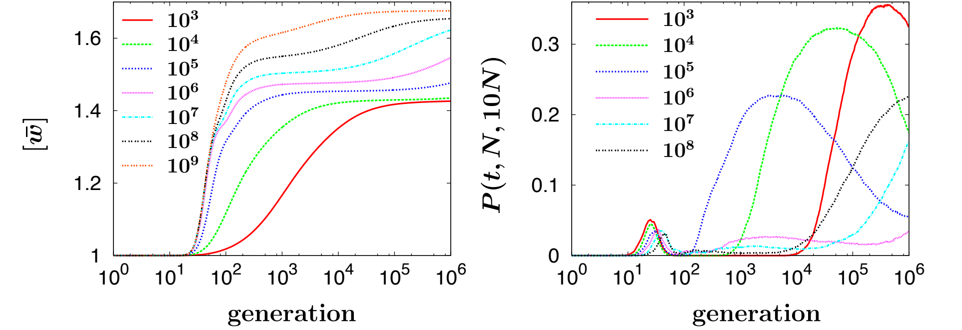
- House-of-cards model

$$w(\sigma) = 1 + Sx$$

S : a parameter which controls the strength of selection (0.1)
x : a random number drawn from $p(x) = e^{-x}$.
$w(000\ldots) = 1$ (global minimum)

Probability of being advantageous

- $P(t,N,N')$: probability that a random landscape confers larger mean fitness to a population of size N than to that with size N' at t.



Summary

- Three-locus model with smooth and rugged evolutionary paths
  - analytic expression for taking rugged path
  - criterion for the small population advantage $N\mu \ln N \sim O(1)$
- House-of-Cards model
  - With a certain probability, the three-locus model behavior is observed
  - For finte L, there is a regime where a small population has advantage
  - cf: for infinite L, $\bar{w}(t) \sim \ln(NUt)$ (the larger, the faster)