Evolutionary advantage of small populations on complex fitness landscapes

Su-Chan Park

NSPCS 2010, July 27, 2010

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

1. Introduction
   - Evolution in the lab
   - Evolution on simple and complex media

2. Advantage of small populations
   - Three-locus model
   - How generic is the three-locus model?

3. Summary
Outline

1. Introduction
   - Evolution in the lab
     - Evolution on simple and complex media

2. Advantage of small populations
   - Three-locus model
   - How generic is the three-locus model?

3. Summary
Evolution in action in the lab

R. Lenski, Michigan State University

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

\[ \frac{w_2}{w_1} \equiv \frac{w_2}{w_0} \div \frac{w_1}{w_0} \]
Measuring Fitness

\[
\frac{w_2}{w_1} \equiv \frac{w_2}{w_0} \div \frac{w_1}{w_0}
\]

cf: Kerr et al. (2002)
rock-paper-scissors game
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 $s = \frac{w}{w'} - 1$.
• fitness landscape $w(\sigma)$:

adaptation (natural selection): hill-climbing process
speed of adaptation (theory)

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

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


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

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

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


- a large population adapts faster than a small one.
speed of adaptation (theory)

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

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


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

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

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


- a large population adapts faster than a small one.
speed of adaptation (theory)

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

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


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

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

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


- a large population adapts faster than a small one.
Outline

1. Introduction
   - Evolution in the lab
   - Evolution on simple and complex media

2. Advantage of small populations
   - Three-locus model
   - How generic is the three-locus model?

3. Summary
mean fitness after 500 generations

Large : $2.5 \times 10^7$
Small : $5 \times 10^5$
Advantage of small populations

Evolution in the lab
Evolution on simple and complex media

**fitness trajectory (complex medium)**

---

**small population**

**large population**

---

![Graphs showing fitness trajectory](image-url)
Outline

1. Introduction
   - Evolution in the lab
   - Evolution on simple and complex media

2. Advantage of small populations
   - Three-locus model
   - How generic is the three-locus model?

3. Summary
Simple Model: Three-locus \((L = 3)\)

<table>
<thead>
<tr>
<th>000</th>
<th>001</th>
<th>010</th>
<th>011</th>
<th>100</th>
<th>101</th>
<th>110</th>
<th>111</th>
</tr>
</thead>
<tbody>
<tr>
<td>(w(000) = 1,)</td>
<td>(w(001) = w(010) = 1 + s_1,)</td>
<td>(w(100) = 1 + s_2,)</td>
<td>(w(011) = w(101) = w(110) = (1 + s_1)^2,)</td>
<td>(w(111) = (1 + s_1)^2)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

- 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 \( \sigma \) at generation \( t \).
- \( M(\sigma \leftarrow \sigma') \): mutation prob. from \( \sigma' \) to \( \sigma \).
- \( \bar{w}(t) = \sum_\sigma w(\sigma) f(\sigma; t) \): mean fitness
Background: Fixation

- Fixation probability $\pi(s) \approx \frac{1 - e^{-2s}}{1 - e^{-2sN}}$, $s = \frac{w_{\text{red}}}{w_{\text{white}}} - 1$
- $Ns \gg 1$ 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 \( \rightarrow \) “adaptive walk”
- Waiting time for (first event of) fixation
  - from \( w(000) = 1 \) to \( w(001) = w(010) = 1 + s_1 \), \( T_1 \approx \frac{1}{2\mu Ns_1} \)
  - from \( w(000) = 1 \) to \( w(100) = 1 + s_2 \), \( T_1 \approx \frac{1}{2\mu Ns_2} \)
- From local maximum to global maximum

\[
T_{esc} \approx \frac{s_{del}}{4N\mu^2s_{ben}} \\
s_{del} = \frac{w(100)}{w(101)} - 1 \approx s_2 - 2s_1 \\
s_{ben} = \frac{w(111)}{w(100)} - 1 \approx 3s_1 - s_2
\]
Path Probability

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

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

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

\[ N\mu \ln(Ns) \sim O(1) \]
Evolution on the fitness landscape

- Figure 2: Average fitness as a function of time on the fitness landscape defined by Eq. (1) for $\mu = 1 \times 10^{-5}$ and (left) $s_1 = 0$, $s_2 = 0.1$, (right) $s_1 = 0.02$, $s_2 = 0.05$.

- As a guide to the eyes, the maximum fitness values $W(111)$ for each case are also drawn. The data have been averaged over $10^5$ histories.

- Fitness increases with population size at short time and at long time scales, but in both cases this relationship is inverted for a range of population sizes at intermediate times.

- The expression [12] for the critical population size separating the two regimes shows that, for the parameters used in the present work, the simultaneous escape mode dominates for population sizes $N > 1000$. In the simultaneous mode the escape time is given approximately by $T_{\text{esc}} \approx s_{\text{del}} N \mu^2 s_{\text{ben}}$.

- Assuming all selection coefficients $s_1, s_2, s_{\text{ben}}, s_{\text{del}}$ to be of a similar magnitude, we see that $T_{\text{esc}} \approx T_1, T_2 \sim s_{\mu} \gg 1$ whenever $\mu \ll s$, which is expected under most conditions. In particular, it is true in the regime of strong selection and weak mutation (SSWM), where $N\mu \ll 1$ and $Ns \gg 1$ [16, 17]. The relation (6) implies that the evolution time rugged along a rugged path is dominated by the escape time $T_{\text{esc}}$, and is much larger than $T_{\text{smooth}}$.

- However, both expressions (3) and (5) share the same dependence on population size $N$, so once the evolutionary path is chosen, a large population is always at a relative advantage.
**Experiment**


---

**Conclusion**

The results show that small populations are better able to locate a diverse range of fitness peaks than large populations, and that advantages in terms of fitness are likely the result of the fact that rugged landscapes contain more fitness peaks. The simulations also reveal that variance in adaptive trajectories is higher for small than large populations, indicating that small populations follow more heterogeneous adaptive trajectories. Most interestingly, these simulation results show that small populations benefit from more effective landscape searching than large populations, despite the fact that they were obtained from an experiment and not from simulations.

---

**Summary**

In summary, our data provide experimental and theoretical evidence that small populations can reach higher fitness levels than large populations, and that this advantage is particularly pronounced on rugged adaptive landscapes. This suggests that the evolutionary response of small populations is contingent on the topography of the underlying fitness landscape, with epistatic interactions among beneficial mutations playing a crucial role.
Estimating the (beneficial) mutation probability

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

1. Introduction
   - Evolution in the lab
   - Evolution on simple and complex media

2. Advantage of small populations
   - Three-locus model
   - How generic is the three-locus model?

3. Summary
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)
\( P(t, N, N') \) : 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 finite $L$, there is a regime where a small population has advantage
  - cf: for infinite $L$, $\bar{w}(t) \sim \ln(NUt)$ (the larger, the faster)