The dynamics of complex adaptation

Daniel Weissman

Mar. 20, 2014
People

- Michael Desai, Marc Feldman, Daniel Fisher
- Joanna Masel, Meredith Trotter; Yoav Ram
- Other relevant work: Nick Barton, Shahin Rouhani; Lin Chao, Dan Weinreich; Freddy Christiansen, Sally Otto, Aviv Bergman; Rick Durrett, Deena Schmidt, Jason Schweinsberg; Lilach Hadany; Rutger Hermsen, Terry Hwa; Yoh Iwasa, Natalia Komarova, Franziska Michor, Martin Nowak; Michael Lynch; Yannis Michalakis, Monty Slatkin; Richard Neher, Boris Shraiman; Erik van Nimwegen, James Crutchfield; Maria Serra, Patsy Haccou; Arjan de Visser, Su-Chan Park, Kavita Jain, Joachim Krug;...
Complex adaptation

- Need combination of $K \geq 2$ mutations for benefit
- “Fitness valley/plateau” / “Irreducible complexity”
Why do we care?

Specific cases: signal-receptor, cancer, . . .

Generally:

- When does evolution get stuck?
- Evolution by fittest mutations or fittest combinations?
  - Space of genotypes grows exponentially with $K$
Problems

Population has to:

1. Produce the combination
2. Fix it (incorporate it into everyone’s genome)
Start with the second problem:

When can a rare combination spread in a population?
Selection vs recombination

Frequency $x \ll 1$ of combination changes because of selection $s$, recombination $r$, etc

$$\dot{x} = (s - r)x + rf(\text{mutant allele frequencies}) + \text{stochasticity} + \ldots$$

$$\Rightarrow \begin{cases} 
\text{if } r \gg s : & \text{need } f(\text{allele freqs.}) \gtrsim x \text{ to get } \langle \dot{x} \rangle > 0 \\
\text{if } r < s : & \langle \dot{x} \rangle > 0 \text{ regardless of allele freqs.}
\end{cases}$$

(Simplest ($K = 2$) case: $f \equiv \text{product of mutant allele frequencies}$)
Selection vs recombination: numbers

Rare combination giving $s = 1\%$ more offspring/generation can spread faster than broken up by recombination if genes are within:

- Drosophila/human: 1Mb ($\sim 100$ genes in Drosophila, $\sim 10$ genes in humans)
- Yeast: whole genome??
- HIV within host: whole genome?
- *E. coli*: whole genome, all of the genes?
- Cancer: whole genome
Selection vs stochasticity

Trajectories of mutant lineages $n(t)$:

Near-critical branching process

- $\sim$ deterministic increase once $n \gtrsim 1/s$
- If alive at $t < 1/s$, usually $n \sim t$ descendants
- $P(\text{alive at time } t) \sim 1/t$ for $t < 1/s$

$\Rightarrow p_{\text{fix}}(s) \sim s$: If $s = 1\%$, need to produce combo $\sim 100\times$
Now address first problem:

How can a population find an adaptation that needs $K > 2$ mutations to function?
Moderate $K$: hard but possible?

- Have to do exhaustive search $\Rightarrow$ impossible for large $K$
- But what about moderate $K$?
  - Practically important: heterodimers, cancer, drug resistance...  
  - Number of potential genotypes also growing exponentially
- Population sizes, mutation rates, recombination rates vary over many orders of magnitude – need to know which parameter combinations are important
Simplest toy model

Focus on $K = 2$ mutants needed for beneficial combination, asexual

Population size $N$

Find the mean time $\tau$ for population to acquire combination*

*not the relevant statistic for cancer
Asexual dynamical regimes already complicated

Focus on “plateau” case: small $\delta$
Let $x_2(t) =$ frequency of double-mutants at time $t$

- $x_2(0) = 0$, $\dot{x}_2(t) = \mu^2 t + sx_2$

$\Rightarrow \frac{1}{\tau} \sim \frac{s}{\log(s/\mu)}$

- Cheated: what if $N x_2(t) < 1$? How can we select on nothing?

$\Rightarrow$ Need $N \mu^2 \gg s$

- Generally: $N \mu^K \gg K! s^{K-1}$
Deterministic for very large population sizes

\[ 1 / \tau \sim s \log(s / \mu) \]

- Neutral sequential fixation
- Neutral stochastic tunneling
- Neutral semi-deterministic tunneling
- Deleterious tunneling
- Deleterious sequential fixation

Population size, \( N \)

Selective disadvantage of single-mutants, \( \delta_1 \)
Second guess: need to treat double-mutants stochastically

- $\tau \sim$ time to produce first successful double-mutant
- Single-mutant frequency $x_1(t) \sim \mu t$, so $\tau$ satisfies:

$$N\mu^2\tau^2 \sim 1/s$$

$\Rightarrow 1/\tau \sim \mu\sqrt{Ns}$

- Ignored stochasticity in the single-mutants – is this ok?
  - Need $\langle x_1(\tau) \rangle \gg$ fluctuations
- Third guess: treat all mutants stochastically
Single-mutant lineage

Total # of individuals (area) = # of mutational opportunities

\[
\text{Prob(success)} \sim (\# \text{ double-mutants produced}) \times p_{\text{fix}}(s) \\
\sim \text{area} \times \mu \times p_{\text{fix}}(s)
\]
Distribution of total progeny

\[ \text{Prob(success | area)} \sim \text{area} \times \mu \times p_{\text{fix}}(s) \]

Critical branching process:

- If alive at \( t \ll N \), usually \( n \sim t \) descendants
- \( P(\text{alive at time } t) \sim 1/t \) for \( t \ll N \)
  \[ \Rightarrow P(\text{area} > a) \sim P(\text{alive at time } \sqrt{a}) \sim 1/\sqrt{a} \]
  - Long-tailed distribution of progeny – large fluctuations
  \[ \Rightarrow \text{Prob(success)} \sim 1/\sqrt{\mu s} \]

Most likely path to success: rare lineage that persists for \( t \sim 1/\sqrt{\mu s} \); occurs with prob \( \sim \sqrt{\mu s} \)

\[ \Rightarrow 1/\tau \sim N\mu \sqrt{\mu s} \]
Most likely path to success: one big lineage

\[
\text{Prob}(\text{success} \mid \text{area}) \sim \text{area} \times \mu s; \quad \text{Prob}(\text{area} > a) \sim 1/\sqrt{a}
\]

So wait for one big lineage that persists for \( t \sim 1/\sqrt{\mu s} \);
occurs with \( \text{prob} \sim \sqrt{\mu s} \)

\[
\Rightarrow \frac{1}{\tau} \sim N\mu \sqrt{\mu s}
\]

\[K > 2 : \frac{1}{\tau} \sim (N\mu^2)(s/\mu)^{1/2K-1}\]
Range of behaviors over different population sizes

- **Selective disadvantage of single-mutants, $\delta_1$**
- **Population size, $N$**
- **$1/\tau \sim s \log(s/\mu)$**
- **Deterministic**
- **Neutral semi-deterministic tunneling**
  
  $$1/\tau \sim \mu \sqrt{N s}$$

- **Neutral stochastic tunneling**
  
  $$1/\tau \sim N \mu \sqrt{\mu s}$$

- **Neutral sequential fixation**

- **Deleterious tunneling**

- **Deleterious sequential fixation**

- **$N = \frac{\delta_1}{\mu_0 \mu_1}$**

- **$N = \frac{2\delta_1^2}{\pi \mu_0 \mu_1 s}$**

- **$N = \frac{1}{\delta_1 \log \left[ \frac{\delta_1^2}{\mu_1 s} \right]$**

Selectively disadvantage of single-mutants, $\delta_1$
When is complex adaptation likely?

- At least medium-sized population: \( N > \frac{1}{\sqrt{\mu s}} \)
- Neutral single mutants: \( \delta < \sqrt{\mu s} \)
  - condition on \( \delta \) relaxed for larger \( N \)
To be able to “see” combo of two individually neutral point mutations with $s = 0.01$, need $N > 10/\sqrt{\mu}$

- “neutral”: $\delta < \sqrt{\mu}/10$
- *E. coli*: $\mu \sim 10^{-10} \Rightarrow N \gtrsim 10^6$ ($\sim 10^{11}$ in you)
- RNA virus: $\mu \sim 10^{-4} \Rightarrow N \gtrsim 10^3$
What about sex?
Sex helps for $r, \delta < \frac{s}{2}$
Putting it all together

\[ K = 2, \delta = 10^{-3}, \mu = 5 \times 10^{-7} \]

Complicated, but understandable
Conclusion

▶ Summary:
  ▶ Adaptation can spread without intermediate genotypes if advantage $s > \text{recombination rate } r$
  ▶ Moderately complex adaptation is easy if:
    ▶ Population is large ($N > 1/\sqrt{\mu s}$, $N > 1/\mu$, etc)
    ▶ Intermediate genotypes not too deleterious ($\delta < \sqrt{\mu s}$, etc)
    ▶ Moderate recombination $r \lesssim s$
  ▶ No reason why it shouldn’t be happening in natural populations
▶ Questions:
  ▶ Effect of sex for $K > 2$?
  ▶ Interaction with simple adaptation?
  ▶ Real populations/fitness landscapes?

Thanks for listening!