Quantifying the role of population subdivision in evolution on rugged fitness landscapes

Anne-Florence Bitbol
Collaborator: David J. Schwab

Biophysics Theory Group
Lewis-Sigler Institute for Integrative Genomics & Departments of Physics and Molecular Biology

KITP Program: Evolution of Drug Resistance – August 13, 2014
**Introduction**

- **Fitness landscape**
  Wright (1930s)

- **Origin of fitness valleys: epistasis**

Poelwijk, Kiviet, Weinreich and Tans (2007)
**Introduction**

- **Molecular example**

  Co-evolving systems → fitness valleys

  The *lac* operon:

  RNA polymerase  
  Repressor  
  Operator

  [Diagram: lac operon with RNA polymerase and repressor binding to operator]

  [Graph: Repression vs Mutation step, Dawid et al. (2010)]

- **Fitness costs in the evolution of antibiotic resistance**

  Evolution of streptomycin resistance in *E. coli*

  Schrag, Perrot and Levin (1997)
Introduction

- **Effect of population size on fitness valley crossing**

    Smaller population → stochasticity is more important
    Deleterious / neutral mutations can drift to fixation

- **Valley crossing time vs. population size: two regimes**

![Graph showing valley crossing time vs. population size]

*Weinreich and Chao (2005)*
*Weissman, Desai, Fisher and Feldman (2009)*
Population subdivision: a minimal model

Asexual population
Fixed size

Demes with identical size

→ Can subdivision with migration (alone) accelerate fitness valley crossing?
If yes, under what conditions, and how much?

N.B.: Wright's shifting balance theory (1930s)

Fitness landscape

- A single valley
- No backward mutations
- A single mutation rate $\mu +$ assume $N\mu < 1$

Here: No geographic structure
No extinction / founding
No environment heterogeneity
Constant migration rate
Best scenario

1. Valley crossing by the champion deme

2. Spreading by migration

At best: valley crossing time dominated by that of the champion (fastest) deme

→ Speedup in this best scenario?

→ Conditions?
Crossing by the champion among $D$ independent demes

1. Demes in the sequential fixation regime

Average crossing time for one deme:
\[
\tau = \tau_{01} + \tau_{12} = \frac{1}{N\mu_d p_{01}} + \frac{1}{N\mu_d p_{12}}
\]

Fixation probability of one “$j$” individual:
\[
p_{ij} = \frac{1 - e^{f_i - f_j}}{1 - e^{N(f_i - f_j)}}
\]

\[
\delta \ll 1, \ s \ll 1, \ N\delta \gg 1, \ Ns \gg 1 \rightarrow p_{01} = \frac{e^\delta - 1}{e^{N\delta} - 1} \approx \delta e^{-N\delta}
\]

\[
\tau_{01} \gg \tau_{12} \rightarrow \tau \approx \tau_{01} = \frac{1}{N\mu_d p_{01}} \approx \frac{e^{N\delta}}{N\mu_d \delta}
\]

Crossing time $\sim$ exponentially distributed

$\rightarrow$ Average for the champion among $D$ demes:
\[
\frac{\tau_c}{\tau_{id}} \approx \frac{1}{D}
\]

(\(c\): champion; \(id\): isolated deme)

2. Demes in the tunneling regime

In this case too,
\[
\frac{\tau_c}{\tau_{id}} \approx \frac{1}{D}
\]

\[
Dp_{01} \ll p_{12} \quad \text{(can be generalized)}
\]
**Best scenario**

- **Necessary conditions to obtain speedups**

  \[
  \tau_m \approx \tau_c \quad \text{with} \quad \frac{\tau_c}{\tau_{id}} \approx \frac{1}{D}
  \]

  Best scenario → \( \tau_m \approx \tau_c \) with \( \frac{\tau_c}{\tau_{id}} \approx \frac{1}{D} \)

  Hence, to have a speedup by subdivision (\( \tau_m < \tau_{ns} \)), we need \( \frac{\tau_{id}}{\tau_{ns}} < D \)

  - Slope needs to be larger (less negative) than -1
  - **Consequence:** Sequential fixation in individual demes is necessary in order to get speedups
  - **Reciprocally:** Demes in the sequential fixation regime → speedups *in the best scenario*

  → **Conditions under which the best scenario is attained?**
Best scenario (reminder)

1. Valley crossing by the champion deme

2. Spreading by migration

→ if demes are in the sequential fixation regime

At best: valley crossing time dominated by that of the champion (fastest) deme

→ Conditions?
**Condition 1: quasi-independence**

- The champion deme must be shielded from migration while in the deleterious state.

Timescale: \( \tau_{12} = \frac{1}{N \mu d p_{12}} \) with \( p_{12} = \frac{e^{-(\delta + s)} - 1}{e^{-N(\delta + s)} - 1} \approx \delta + s \)

\( \delta \ll 1, \ s \ll 1, \ N\delta \gg 1, \ Ns \gg 1 \)

Timescale: \( t_e = \frac{n_e}{DNm} \) where \( n_e = \) average number of migrations for “1” to get extinct.

Probability that a migration is relevant: \( p_r = \frac{2}{D} \)

Migrant fixation: \( p_{01} = \frac{e^\delta - 1}{e^{N\delta} - 1} \approx \delta e^{-N\delta} \) and \( p_{10} = \frac{e^{-\delta} - 1}{e^{-N\delta} - 1} \approx \delta \)

\[ n_e \approx \frac{1}{p_r p_{10} (1 - p_{01})} \approx \frac{D}{2\delta} \]

→ First condition: \( \tau_{12} < t_e \rightarrow \frac{m}{\mu d} < \frac{1}{2} \left( 1 + \frac{s}{\delta} \right) \) : upper bound on the migration rate.
Condition 2: fast spreading

- Spreading of the beneficial mutation must be faster than valley crossing by the champion deme

Timescale: \( t_s = \frac{n_s}{DNm} \) where \( n_s \) = average number of migrations for “2” to spread

\[
n_s = \sum_{i=1}^{D-1} n_{i \rightarrow i+1} = \sum_{i=1}^{D-1} \frac{1}{p_{i \rightarrow i+1}}
\]

\( p_{i \rightarrow i+1} = r_i p_{02} (1 - p_{20}) \approx r_i s \)

Probability that a migration is relevant: \( r_i = \frac{2i(D - i)}{D(D - 1)} \)

Hence, \( t_s \approx \frac{\log D}{Nsm} \)
**Condition 2: fast spreading**

- Spreading of the beneficial mutation must be faster than valley crossing by the champion deme

  **Timescale:** \( t_s \approx \frac{\log D}{N s m} \)

  Valley crossing by the champion deme

  **Timescale:** \( \tau_c \approx \frac{\tau_{id}}{D} \approx \frac{e^{N \delta}}{DN \mu d \delta} \)

  → **Second condition:** \( t_s < \tau_c \) → \( \frac{\delta e^{-N \delta}}{s} D \log D < \frac{m}{\mu d} \): lower bound on the migration rate

- **Prediction:**

  \[ \frac{\delta e^{-N \delta}}{s} D \log D \ll \frac{m}{\mu d} \ll \frac{1}{2} \left(1 + \frac{s}{\delta}\right) \] → optimal scenario, and \( \frac{\tau_m}{\tau_{id}} \approx \frac{1}{D} \)
Test: stochastic simulation

- Simulation (Gillespie algorithm) $\rightarrow$ crossing time vs. migration rate

Parameter values:
- $s = 0.3$
- $\delta = 0.006$
- $K = 357$
- $D = 7$
- $\mu = 8 \times 10^{-6}$
- $d = 0.1$

Minimum $\rightarrow \tau_m = (5.02 \pm 0.14) \times 10^5$

$\tau_{id} = (3.28 \pm 0.10) \times 10^6$ $\rightarrow$ factor of 6.54, close to $D = 7$
Test: stochastic simulation

- Valley crossing at the optimum

One realization:

End of the process:
Generalizing

- Beyond $N\delta \gg 1$: shallow valleys, plateaus, etc.

$N\delta \gg 1, \ Ns \gg 1 \rightarrow$ simple derivation of numbers of migrations until extinction or fixation

- A finite Markov chain

$i \in [0, D]$ : number of demes that have fixed the mutation (e.g., “1”)

At each migration step, $i$ can change
Outcome of the next migration only depends on current value of $i$

Two absorbing states: $i = 0$ and $i = D$

- Transition probabilities

$P_{i \rightarrow i+1} = r_i p_{01} (1 - p_{10})$

$P_{i \rightarrow i-1} = r_i p_{10} (1 - p_{01})$

$P_{i \rightarrow i} = 1 - (P_{i \rightarrow i+1} + P_{i \rightarrow i-1})$

Probability that a migration is relevant:

$r_i = \frac{2i(D - i)}{D(D - 1)}$

The matrix of transition probabilities is tri-diagonal $\rightarrow$ simple case!

The number of migration steps before absorption can be expressed analytically

Ewens (1979)
Generalizing

- **Optimal parameter range**

\[ n_s p_{01} \ll \frac{m}{\mu d} \ll \frac{n_e p_{12}}{D} \]

Exact expressions for \( n_s \) and \( n_e \) (number of migration steps before absorption)

Case of the plateau \((\delta = 0)\): optimal speedup is obtained for

\[ \frac{1}{N s} D \log D \ll \frac{m}{\mu d} \ll \frac{N s}{2} \log D \]

- **Effectively neutral intermediates**

Effectively neutral intermediate: \(|\delta| < \max(\sqrt{\mu s}, 1/N)\): includes weakly beneficial ones

\(\rightarrow\) plateau results hold

Weissman et al. (2009)

**Example:**

Parameter values:

- \( s = 0.5 \)
- \( N = 130 \)
- \( D = 10 \)
- \( \mu = 5 \times 10^{-7} \)
- \( d = 0.1 \)
Heatmaps

Parameters:
- $s = 0.3$
- $K = 50$
- $D = 10$
- $\mu = 5 \times 10^{-6}$
- $d = 0.1$

Tunneling starts dominating for:
- Isolated demes
- The nonsubdivided population
**Highest possible speedup by subdivision**

Optimal case → speedup gained by subdividing a population: \[
\frac{\tau_m}{\tau_{ns}} = \frac{\tau_c}{\tau_{ns}}
\]

**Assume:**
- isolated deme in the sequential fixation regime
- nonsubdivided population in the tunneling regime

At fixed \(N\), this ratio is minimal for \(\tilde{\delta} \approx \frac{1.594}{N}\) (→ importance of general calculations)

Its minimal value is \[
\frac{\tau_m}{\tau_{ns}} \approx 1.544 N^2 \mu s
\]

Heatmaps → optimal valley depth: \(\tilde{\delta} \approx 0.035 \approx 10^{-1.45}\)

**A trade-off in the choice of \(D\)**

Fixed \(N = ND\) → highest speedup: \[
\frac{\tau_m}{\tau_{ns}} \approx 1.544 \frac{N^2 \mu s}{D^2}
\]

Increase \(D\) → gain more speedup

But \[
\frac{\delta e^{-N\delta}}{s} D \log D \ll \frac{m}{\mu d} \ll \frac{1}{2} \left(1 + \frac{s}{\delta}\right)
\]

Increase \(D\) → narrower optimal parameter range
Varying the degree of subdivision

Parameter values:

- $s = 0.3$
- $\delta = 0.006$
- $DK = 2500$
- $\mu = 8 \times 10^{-6}$
- $d = 0.1$

$m = 10^{-5}$

$m/(\mu d) = 12.5$
Application

- **An example**

  \[ E. \text{coli} \rightarrow \mu \approx 8.9 \times 10^{-11} \]  
  Wielgoss et al. (2011)

  Take \( N = 5 \times 10^4 \)  
  (small but realistic)  
  Rozen et al. (2008)

  \( D = 100 \)  
  (96-well plates)

  Plateau → sequential fixation below \( N_x = 1/\sqrt{\mu s} \)

  \( s = 10^{-2} \) → isolated demes in the sequential fixation regime

  for \( 0 \leq \delta \lesssim 2.2 \times 10^{-4} \)

  The optimal range of migration rates spans 2 to 4 orders of magnitude depending on \( \delta \)

  Speedup factor from 18 to \( 2.7 \times 10^2 \)

- **More generally**

  For given \( N \) and \( D \), we can predict:
  - for which valleys subdivision speeds up crossing
  - the highest speedups obtained
  - the range of migration rates for which they are reached
Conclusion

**Summary**

- Subdivision with migration (alone) can significantly accelerate fitness valley & plateau crossing
- Sufficiently small demes (performing sequential fixation) are necessary
- Effect of varying the degree of subdivision

**Some related experimental studies**

- Kryazhimskiy, Rice and Desai (2012) → evolution of subdivided populations of yeast → no evidence of any advantage of subdivided populations

- Nahum, Godfrey-Smith, Harding, Marcus, Carlson-Stevermer and Kerr (BioRXiv 2014) → evolution of subdivided populations of bacteria → some advantage of subdivision

→ Importance of understanding quantitatively the conditions under which subdivision is beneficial
Conclusion

- Perspectives
  - More complex population structure (different sizes)
    → already treated: large population + islands
  - Case of sexual populations (recombination)
  - Spatial structure (expanding front)
  - Effect of population subdivision on the evolution of antibiotic resistance

- Acknowledgements

  David J. Schwab
  
  Ned S. Wingreen
  
  The Princeton Biophysics Journal Club

- For more information

  DOI: 10.1371/journal.pcbi.1003778
  Preprint: ArXiv 1308.0278
Thanks!