Journal of Theoretical Biology 380 (2015) 123–133

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

# Tipping the mutation–selection balance: Limited migration increases the frequency of deleterious mutants

Jacob D. Cooper [a,b,*], Claudia Neuhauser [c], Antony M. Dean [d], Benjamin Kerr [a,b]

[a] Department of Biology, University of Washington, Seattle, WA, United States
[b] BEACON Center for the Study of Evolution in Action, University of Washington, Seattle, WA, United States
[c] Biomedical Informatics and Computational Biology, University of Minnesota, Rochester, MN, United States
[d] College of Ecology and Evolution, Sun Yat-sen University, Guangzhau, China

## HIGHLIGHTS

- A birth-death model with migration is analyzed at mutation-selection balance.
- No assumptions are required about the strength of selection or mutation.
- Analytical approximations are tested against stochastic agent-based simulations.
- Limiting migration leads to more deleterious mutants at equilibrium.
- Limiting migration may lead to faster discovery of novel genotypes.

## ARTICLE INFO

*Article history:*
Received 18 October 2014
Received in revised form
29 April 2015
Accepted 4 May 2015
Available online 14 May 2015

*Keywords:*
Population genetics
Fitness landscape
Adaptive valley crossing
Spatial structure
Moment closure

## ABSTRACT

Typical mutation–selection models assume well-mixed populations, but dispersal and migration within many natural populations is spatially limited. Such limitations can lead to enhanced variation among locations as different types become clustered in different places. Such clustering weakens competition between unlike types relative to competition between like types; thus, the rate by which a fitter type displaces an inferior competitor can be affected by the spatial scale of movement. In this paper, we use a birth-death model to show that limited migration can affect asexual populations by creating competitive refugia. We use a moment closure approach to show that as population structure is introduced by limiting migration, the equilibrial frequency of deleterious mutants increases. We support and extend the model through stochastic simulation, and we use a spatially explicit cellular automaton approach to corroborate the results. We discuss the implications of these results for standing variation in structured populations and adaptive valley crossing in Wright's "shifting balance" process.



## 1. Introduction

Most mutations affecting fitness appear to be deleterious (see review by Eyre-Walker and Keightley, 2007). A deleterious mutation is expected to persist in a population at a level influenced by the rate at which it is generated and the strength of selection against it. This mutation–selection balance was first developed mathematically by Haldane and Fisher in the 1920's in models that assumed well-mixed populations (Fisher, 1930; Haldane, 1927). However, many natural populations are not well mixed: individuals may not disperse, and even if they do, dispersal or migration is often restricted to nearby locations (Evans et al., 2009; Howells et al., 2013; Martin and Canham, 2010). Such limited movement may influence the proportion of deleterious mutants at equilibrium in several ways. In mating diploid populations, the Wahlund effect (in which population-level heterozygosity is depressed when subpopulations differ in allele frequency) combines with dominance relationships among genotypes to influence the frequency of deleterious mutant alleles (Roze and Rousset, 2004; Whitlock, 2002). In haploid asexual models, limiting migration increases between-deme variation and decreases within-deme variation, but the extent to which this shift in variation affects evolution is unclear.

Limitations to migration are not predicted to affect the equilibrium frequency of deleterious mutants in asexual populations when fitness is independent of local composition and density. For instance, Whitlock (2002) finds no effect of migration under a "hard selection" scheme (in which absolute fitness is determined solely by genotype, and thus demes of different compositions may

* Correspondence to: University of Washington, Department of Biology, Box 351800, Seattle, WA 98195-1800, United States. Tel.: +1 206 221 7026.
*E-mail address:* yankel@uw.edu (J.D. Cooper).

http://dx.doi.org/10.1016/j.jtbi.2015.05.003
0022-5193/

differ in productivity). However, in "soft selection" regimes (in which relative fitness within a deme depends on genotype, but each deme's productivity is the same regardless of composition), demes enriched for mutants are as productive as demes enriched for wild types. Such mutant-rich demes may serve as competitive refugia. Thus, in soft selection schemes, limiting migration can increase the frequencies of deleterious mutants (Roze and Rousset, 2004; Whitlock, 2002).

As mutation, selection and migration occur in a subdivided population, both first-order moments (i.e., the mean) and higher-order moments (i.e., variance, skew, kurtosis, etc.) of allele frequencies across demes can change. Previous models have estimated higher-order moments (or related quantities like $F_{ST}$) in terms of first-order moments under an assumption of weak selection. In this paper, we take a different approach. We build an ecological model of a subdivided population, in which higher-order moments are dynamic variables. No assumptions about the strength of selection or mutation are required. Using this model, we find that limited migration increases the fraction of mutants at mutation–selection balance. However, our moment-closure approach (in which we express third-order moments in terms of lower-order moments) is exact only under total migration. Thus, our analytical results are accurate when there is minimal subdivision. Similar moment closure approaches have been used to model ecological neutrality, competition, and stability (Bolker and Pacala, 1997; Haegeman and Loreau, 2011; Neuhauser, 2002; Vanpeteghem and Haegeman, 2010). We use computer simulations to confirm that the fraction of mutants at equilibrium increases under limited migration (where the mathematical analysis is approximate). The simulations also show spatial segregation of types, suggesting that mutant-rich areas act as competitive refugia.

## 2. Mutation–selection balance in a subdivided population

In our model, a population inhabits a metapopulation of patches. Space is implicit in this model; all patches are equally "far" from any given patch. Migration between patches occurs at birth with a specified probability. When the probability is one, every offspring migrates to a random patch, and the population is essentially well mixed. When the probability is lowered slightly from one, there is a small chance an offspring will stay in its natal patch, and thus a modicum of spatial structure is introduced.

### 2.1. Terminology and life cycle

Consider two genotypes $W$ and $M$, for wild type and mutant, respectively, inhabiting a metapopulation with an infinite number of patches. The population size of each patch is finite. In all that follows, genotype indices $i$ and $j$ will be used where $i,j\in\{W,M\}$ and $i\neq j$. The per capita birth rate of genotype $i$ is given by $F_i(n_i,n_j)=f_i-\beta_i(n_i+\alpha_{ij}n_j)$, where $n_i$ and $n_j$ are the numbers of genotype $i$ and $j$ in the patch, $f_i$ is the intrinsic growth rate of genotype $i$, $\beta_i$ measures the effect of intra-genotypic competition, and $\alpha_{ij}$ is an inter-genotypic conversion factor (i.e., one individual of genotype $j$ counts as $\alpha_{ij}$ individuals of genotype $i$). Genotype $i$ dies with rate $\delta_i$. Mutation from genotype $i$ to $j$ occurs during the birth process with probability $\mu_{i\to j}$. Migration also occurs at birth, when genotype $i$ migrates to a random patch with probability $m_i$. The population evolves stochastically in continuous time.

### 2.2. Moment dynamics

Let $N_i(t)$ be the expected number of genotype $i$ per patch at time $t$. For typographical convenience, we drop the explicit reference to time dependence in our notation for the terms and equations that follow (e.g., $N_i(t)$ is written $N_i$). In Appendix 1 we show that

$$\frac{dN_i}{dt}=(1-\mu_{i\to j})\,N_i\,F_i\left(N_{i|i},N_{j|i}\right)+\mu_{j\to i}\,N_j\,F_j\left(N_{j|j},N_{i|j}\right)-\delta_i\,N_i, \tag{1}$$

where $N_{i|j}$ is the expected number of individuals of genotype $i$ in the patch of a randomly chosen individual of genotype $j$, with $i,j\in\{W,M\}$.

It can be shown that $N_{i|i}=N_i+\sigma_i^2/N_i$, where $\sigma_i^2$ is the variance in the number of genotype $i$. When individuals of the given genotype are uniformly distributed (i.e., variance is zero), this reduces to the mean $N_i$. Similarly, $N_{i|j}=N_i+C/N_j$, where $C$ is the covariance between the numbers of genotypes $i$ and $j$. When the two genotypes are independently distributed (i.e., covariance is zero) this term reduces to the mean $N_i$. Covariance may be positive, indicating association between types, or negative, indicating segregation of types.

Thus the dynamics of the first order moments $N_i$ and $N_j$ rely on second order moments $\sigma_i^2$, $\sigma_j^2$, and $C$. The equations governing the dynamics of these second order moments involve third order moments, the differential equations for the third order moments involve fourth order moments, and so on. Our task is similar to Hercules' battle with the Hydra (in spirit, not magnitude!). With each Hydra head Hercules sliced off, new heads popped up in its place. For each moment dynamical equation we describe, the description of new, higher-order moment equations becomes necessary. We must find a way to stem the endless flow of higher-order moments. Hercules seared the necks of the Hydra to prevent the regrowth of the heads; we close our system of differential equations by a second-order moment closure technique. We approximate third-order moments in terms of lower-order moments (see Appendix 1 for details), thus sealing the endless flow. Our moment closure approximation is exact when migration is absolute (i.e., $m_W=m_M=1$), and we are not limited by assumptions of near neutrality (Neuhauser, 2002). With this approximation, the dynamics for the second order moments are given by:

$$\begin{aligned}\frac{d\sigma_i^2}{dt}=\frac{dN_i}{dt}&+2\delta_i(N_i-\sigma_i^2)\\&+2(1-m_i)(1-\mu_{i\to j})\left\{f_i\sigma_i^2-\beta_i(N_i+2N_i\sigma_i^2)-\beta_i\alpha_{ij}(N_iC+N_j\sigma_i^2)\right\}\\&+2(1-m_j)\,\mu_{j\to i}\left\{f_jC-\beta_j2N_jC-\beta_j\alpha_{ji}(N_iC+N_j\sigma_i^2)\right\}\end{aligned} \tag{2}$$

$$\begin{aligned}\frac{dC}{dt}=&-(\delta_i+\delta_j)C+(1-m_i)(1-\mu_{i\to j})\left\{f_iC-\beta_i2N_iC-\beta_i\alpha_{ij}\left(N_jC+N_i\sigma_j^2\right)\right\}\\&+(1-m_j)\mu_{j\to i}\left\{f_j\sigma_j^2-\beta_j\left(N_j+2N_j\sigma_j^2\right)-\beta_j\alpha_{ji}\left(N_jC+N_i\sigma_j^2\right)\right\}\\&+(1-m_j)(1-\mu_{j\to i})\left\{f_jC-\beta_j2N_jC-\beta_j\alpha_{ji}(N_iC+N_j\sigma_i^2)\right\}\\&+(1-m_i)\mu_{i\to j}\left\{f_i\sigma_i^2-\beta_i(N_i+2N_i\sigma_i^2)-\beta_i\alpha_{ij}(N_iC+N_j\sigma_i^2)\right\}\end{aligned} \tag{3}$$

### 2.3. Mutation–selection balance

Our dynamical system contains many parameters. To simplify matters, we assume $m_W=m_M=m$, $f_W=f_M=f$, $\beta_W=\beta_M=\beta$, $\alpha_{WM}=\alpha_{MW}=1$, $\mu_{W\to M}=\mu$, and $\mu_{M\to W}=0$. Thus, we assume our genotypes are identical in all parameters except their death rates, which define a $W$ to $M$ mutation as deleterious (i.e. $\delta_M>\delta_W>0$), and their mutation rates. Consequently, we only consider viability selection in this analysis, though we simulate other possibilities below. We have also assumed that intra-genotypic competition is identical to inter-genotypic competition (the $\alpha$ parameters are set to unity), and that back mutation does not occur. This might be realistic if the mutation from wild type to the mutant involves a deletion, but even if this mutation is a base substitution, the

density of mutants is often so low that back mutation does not greatly affect our results (see simulations below).

In Appendix 2, we derive the mutant fraction of the population at equilibrium under full migration ($m=1$). Because the fraction of mutants cannot be greater than one, there are parameter constraints on the analysis to ensure mutation does not "overwhelm" selection. Within those parameter constraints, the fraction of mutants at mutation–selection balance is

$$\Phi = \frac{\mu\delta_W}{(1-\mu)(\delta_M-\delta_W)}. \tag{4}$$

Like the classical result for a panmictic haploid population, for which $\Phi=(\mu/s)$ where $1-s$ is the fitness of a mutant relative to a wild-type (Crow and Kimura, 1970), our expression is proportional to $\mu$ (for small $\mu$) and inversely proportional to a measure of the selective disadvantage of the mutant ($(\delta_M/\delta_W)-1$).

How does this fraction change as spatial structure is introduced; that is, what happens to $\Phi$ as $m$ is lowered from unity? Since Eqs. (2) and (3) are exact for $m=1$, the partial derivative of $\Phi$ with respect to $m$ can be computed exactly at $m=1$ (Neuhauser, 2002). In Appendix 3, we derive the following:

$$\left.\frac{\partial\Phi}{\partial m}\right|_{m=1} = -\frac{\beta\mu}{(1-\mu)^2\delta_M\left\{(\delta_M/\delta_W)^2-1\right\}}. \tag{5}$$

This expression demonstrates that $\left.\partial\Phi/\partial m\right|_{m=1}<0$ for $0<\mu<1$, so the fraction of deleterious mutants at equilibrium always increases when a small amount of structure is introduced into the model.

### 2.4. Simulation of the spatial model

Our analysis is exact when all offspring migrate, but becomes approximate as soon as some offspring remain in their natal patches. How well do the approximations capture actual dynamics? Here, we explore the model via simulation.

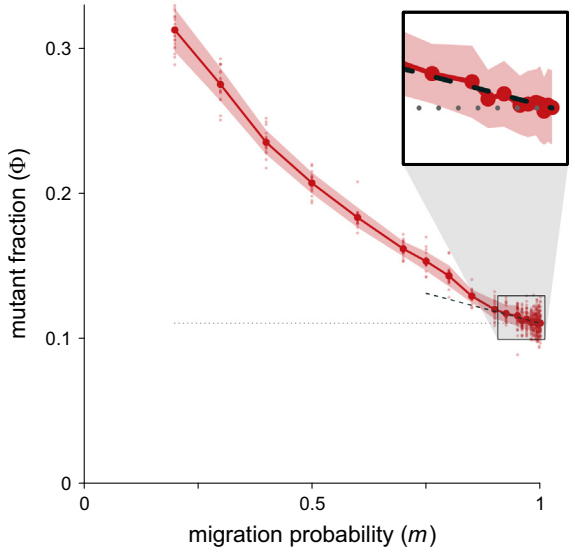


**Fig. 1.** Stochastic simulation under various migration rates. Frequency of deleterious mutants at mutation–selection balance across various probabilities of migration as found by simulation (red circles), compared to the $m=1$ derivative of our analytical model (black dashed line). The analytically calculated mutant fraction at full migration is given as a gray dotted horizontal line for comparison. At high migration probabilities, the simulation results agree well with our analytical model (see inset). As the probability of migration decreases further from unity, the fraction of deleterious mutants increases faster than the analytical extrapolation. Large data points and shading represent mean values and standard deviation of 20–40 replicate simulations (small data points) using parameter values $P=10^4$, $f=0.5$, $\beta=0.2$, $\mu=0.1$, $\delta_W=0.05$, $\delta_M=0.1$, corresponding to a per-patch carrying capacity of approximately $\frac{f-\delta_W}{\beta}=2.25$.

In the simulation, we seed a finite (but large) number of patches $P$ with wild type and mutant individuals, and simulate evolution using a Gillespie algorithm in which birth and death events occur stochastically (see Appendix 4 for details). In simulation runs with absolute migration, all first order moments and second order moments approach our analytic predictions as equilibrium is reached, even when initialized far from the calculated equilibrium (Sup. Fig. 1). This is expected, as our analysis is exact when migration is absolute. As the probability of migration is lowered from unity, our analysis becomes approximate. Fig. 1 shows simulation results across a range of migration probabilities, and the analytical prediction extrapolated from Eq. (5). At high levels of migration, the simulation corresponds well with the analysis, with $\Phi$ following its derivative calculated at $m=1$ (Fig. 1, inset). As migration drops further, the mutant frequency rises faster than the linear extrapolation from our analytical model (Fig. 1). The correspondence between our finite simulation and our deterministic analysis for high migration indicates that the number of patches $P$ is large enough for the metapopulation to behave deterministically. Moreover, results are not appreciably affected when fewer patches are used (Sup. Fig. 2B). Results are also not appreciably affected when back mutation is allowed (Sup. Fig. 2A).

Parameters for Fig. 1 were chosen to illustrate a large effect of limited migration on the mutant fraction at mutation–selection balance. When the competition parameter $\beta$ is decreased, $\mathrm{abs}\left(\left.\partial\Phi/\partial m\right|_{m=1}\right)$ is proportionately decreased (see Eq. 5) and the per-patch carrying capacity is increased, but the simulated mutant frequency still rises faster than the linear extrapolation from our model (Sup. Fig. 2C). Similar results to those shown in Fig. 1 occur when both the mutation rate and the selective disadvantage of mutants are decreased (Sup. Fig. 2D).

To see why limited migration increases the fraction of deleterious mutants in a population, we follow a simulation (Fig. 2) as it transitions from absolute migration to limited migration (i.e., $m=1$ to $m=0.5$). We see that when limited migration is introduced, the mutant frequency increases (Fig. 2A), the variance in mutant density increases (i.e., the mutants become more clumped), and the covariance between the densities of the two genotypes becomes negative (i.e., patches with many wild-type genotypes tend to have fewer mutant genotypes, and vice versa) (Fig. 2B).

Such spatial segregation leads to an increase in the fraction of inhabited patches that house mutant-only populations (Sup. Fig. 3A). The fraction of mutants in mutant-only patches also increases as the migration rate decreases (Sup. Fig. 3B). Notably, the fraction of mutants in patches that also house wild type genotypes does not increase as the migration rate decreases (Sup. Fig. 3B). Thus, the increase of mutants in mutant-only patches may suffice to explain the overall increase in mutants at limited migration rates.

We conclude that limited migration leads to a higher mutant frequency at mutation–selection balance because the less fit mutant is able to escape competition with the wild type due to spatial segregation. Thus, mutant-rich patches are competitive refugia that allow the mutant genotype to persist in relative isolation from the competitively superior wild type. If this explanation is correct, limited migration should safeguard deleterious mutants regardless of whether selection occurs via differences in viability or fecundity, whether space is explicit or implicit, and whether the spatially distributed units are populations or individuals.

## 3. A lattice-based approach

Our next approach considers individuals that are embedded in a lattice. Here, unlike our first approach, (a) space is explicit,

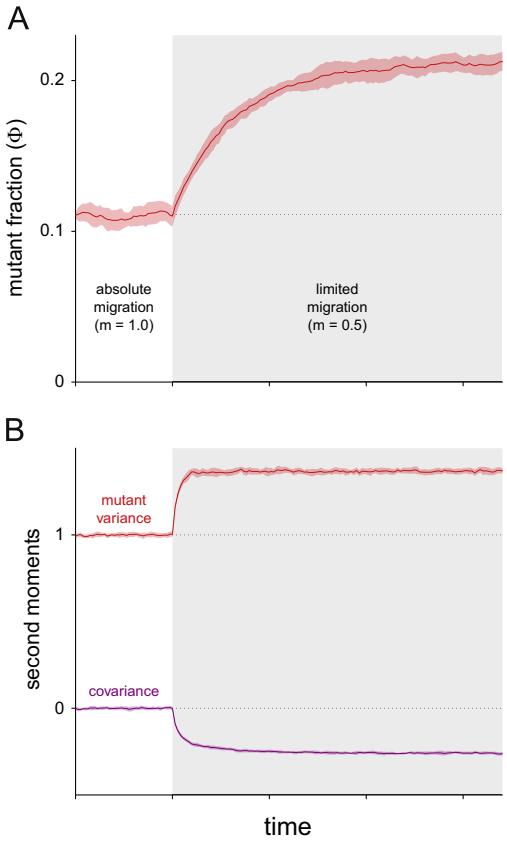


**Fig. 2.** A simulated shift in the probability of migration. When migration becomes limited (grey-shaded portion of plots), the increase in deleterious mutant frequency (A) coincides with an increase in the variance in mutant density and a decrease in the covariance between mutant and wild type densities (B, variances are divided by the means of their corresponding variables, and covariance is divided by the product of the roots of the two means). Time units are relative, and defined by a Gillespie algorithm described in Appendix 4. Solid lines and colored shading represent the mean ± SD of 16 replicate simulations using the parameter values listed in Fig. 1.

(b) population structure varies with dispersal distance, (c) the "patches" house individuals rather than subpopulations, and (d) we consider both viability and fecundity selection. A similar lattice-based approach has been used to explore many eco-evolutionary aspects of spatially structured populations, including the invasion of rare types, species coexistence, host-parasite evolution, spatial structuring of communities, and evolutionary trajectories (Débarre et al., 2012; Durrett and Levin, 1997; Hauert and Doebeli, 2004; Kerr et al., 2002).

In our simulation, we consider two haploid asexual genotypes: wild type ($W$) and mutant ($M$). These genotypes occupy an $L \times L$ regular square lattice with periodic boundaries (i.e., a toroidal geometry). Each lattice point may take one of three states: empty, wild type, or mutant. At each update, a point is chosen at random. If this focal point is "filled" with a wild type, the wild type dies with probability $\delta_W^*$, giving a transformation to the empty state. Likewise, a mutant that is chosen will die with probability $\delta_M^*$ (where $\delta_M^* \geq \delta_W^* > 0$).

If the focal point is already empty, then a birth event can occur, where an individual in a pre-defined neighborhood of the focal point produces an offspring that fills the focal point (giving a transformation to a filled state). Let $x_W$ and $x_M$ be the fraction of the focal point's neighborhood occupied by wild type and mutant lattice points, respectively. Then with probabilities $f_W^* x_W$ and $f_M^* x_M$ the parent of the individual "born into" the focal point is wild type and mutant, respectively. The parameters $f_W^*$ and $f_M^*$ represent the fecundities of wild-type and mutant individuals (where $0 \leq f_M^* \leq f_W^* \leq 1$). The focal point stays empty with probability $1 - f_W^* x_W - f_M^* x_M$. Mutation occurs at birth: from wild type to mutant with probability $\mu_{W\to M}^*$, and from mutant to wild type with probability $\mu_{M\to W}^*$. The degree of population structure is controlled by adjusting the size of the neighborhood around any focal point (effectively altering the distribution of distance at dispersal). We focus on three cases: a von Neumann neighborhood (where the lattice points immediately to the north, east, south and west of the focal point comprise the neighborhood), a Moore neighborhood (where the eight lattice points nearest the focal point constitute the neighborhood), and a Global neighborhood (where the entire lattice, minus the focal point, comprises the

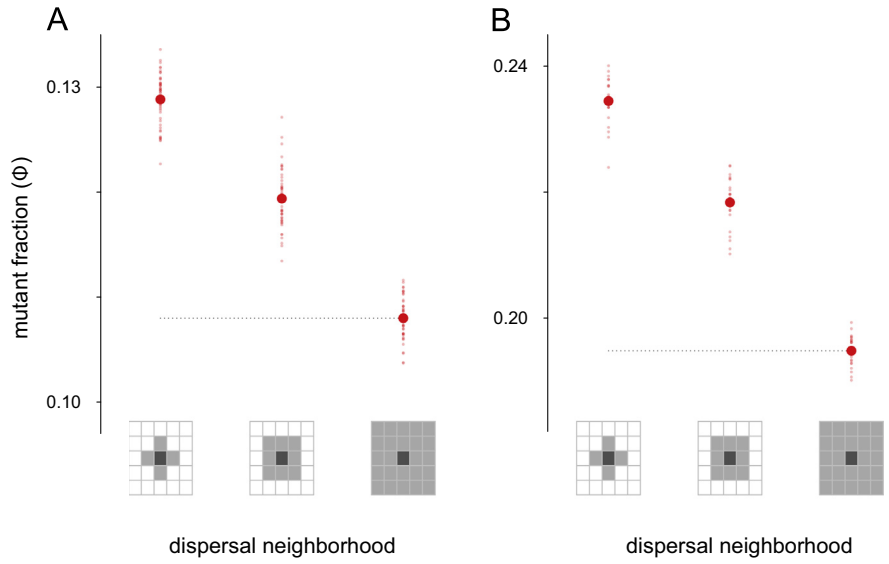


**Fig. 3.** Lattice-based simulation results. The frequency of deleterious mutants at mutation–selection balance across various neighborhood sizes in lattice-based simulations with viability selection (A) or fecundity selection (B). As dispersal is limited to smaller neighborhoods, the frequency of deleterious mutants increases. Large data points represent mean values of 24 replicate simulations (small data points) using parameter values $L = 200$, $\mu_{W\to M}^* = 0.1$, $\mu_{M\to W}^* = 0.02$, and either viability selection (A, with $\delta_W^* = 0.1$, $\delta_M^* = 0.2$, $f_W^* = f_M^* = 1$) or fecundity selection (B, with $\delta_W^* = \delta_M^* = 0.1$, $f_W^* = 1$, $f_M^* = 0.5$).

neighborhood). Thus, the evolving population can range from highly structured (von Neumann neighborhood) to effectively well mixed (Global neighborhood).

Fig. 3 shows that smaller dispersal neighborhoods lead to higher mutant frequencies at equilibrium, corroborating our prior analysis. This pattern holds under both pure viability selection ($\delta_M^* > \delta_W^*$ and $f_M^* = f_W^*$) and pure fecundity selection ($\delta_M^* = \delta_W^*$ and $f_M^* < f_W^*$).

## 4. Discussion

We find analytically and computationally that limited migration increases the frequency of deleterious mutants, and this increase is not restricted to a specific form of space or mode of selection. Prior models of selection in metapopulations have shown effects of limited migration when local interactions are defined by mating, or when mutant-rich demes have similar productivity to wild type enriched demes, and can thus act as competitive refugia for mutants (Glémin et al., 2003; Roze and Rousset, 2004; Whitlock, 2002). In this paper, we embedded competition in an explicitly ecological framework, which allows us to manifest local interactions explicitly as density-dependent fecundity. We used a moment closure approach that expresses higher-order moments in terms of lower-order moments, therefore allowing those higher-order moments to vary dynamically as we began to limit migration. We showed that limited migration can affect asexual populations by segregating types. Essentially, limiting migration has no effect on the generation of mutants, but hampers the effective strength of selection (Cherry and Wakeley, 2003) by sheltering alleles from global competition, and so tips the mutation–selection balance in favor of deleterious mutations. Generally, whenever there is both variation in localities and local interaction, migration rate will be a salient factor in determining the frequency of deleterious mutants.

Mutant frequency is sometimes used to estimate mutation rates of microbes. Using such a method, a structured environment may appear mutagenic because a higher frequency of mutants is found. For example, Bjedov et al. (2003) find a disparity in mutant frequencies between liquid and agar bacterial cultures, and attribute it to oxidative stress incurred during colonial growth on agar. This explanation is certainly plausible, but the colony structure itself may contribute to the increased mutant frequency. When going from an unstructured to a structured environment (e.g., a flask to an agar plate), the frequency of deleterious mutants may increase even if the mutation rate is constant.

### 4.1. Adaptive valley crossing

If a single deleterious mutation is complemented by a second mutation that improves the fitness of the organism above the wild type, the frequency (and number) of the original mutant may be relevant for crossing “adaptive valleys”. Recent theoretical studies have elucidated how well-mixed populations cross adaptive valleys, and at what rate (Weissman et al., 2010, 2009). However, Sewall Wright’s shifting balance process is predicated on the idea that, collectively, semi-isolated subpopulations would explore a landscape in a way unavailable to well-mixed populations (Pigliucci, 2008; Wright, 1988, 1932). To cross a valley, a population must first discover a new peak, and then have the peak genotype spread through the population. Increasing migration between separated patches hinders exploration of novel genotypes (Whitlock, 2003), but, once a beneficial genotype is discovered, the increased migration facilitates its spread (Jain et al., 2011; Rozen et al., 2008). In the rugged landscapes that were the focus of Wright’s shifting balance, the rate-limiting step in adaptation may be the discovery of novel genotypes (i.e., finding new peaks) rather than their spread through a population. If this is the case, limited migration may speed the rate of adaptation.

If valley crossing requires multiple “downward” steps, the facilitating effect of limited migration is amplified. Limited

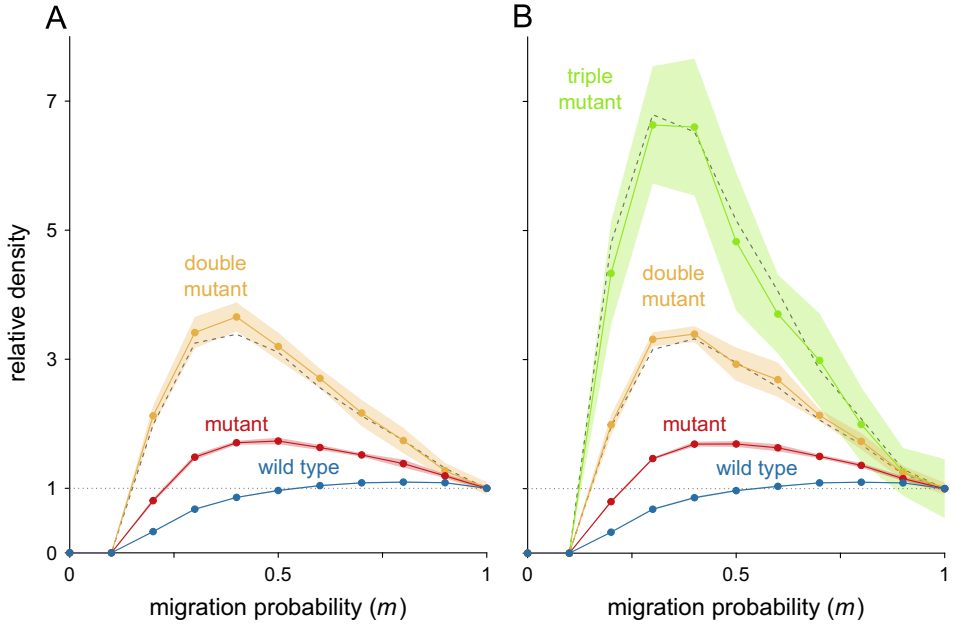


**Fig. 4.** Stochastic simulations with a chain of sequential deleterious mutants in a metapopulation approach. The wild-type genotype yields the first mutant via mutation; the first mutant yields the relatively deleterious double mutant via mutation; and so on. The density of each genotype is shown relative to its density in a well-mixed population (given by the dotted line at unity). The effect of migration probability on relative density in a community with two (A) and three (B) mutants are shown. Points and colored shading represent the mean ± SD of 12 replicate simulations, while dashed lines roughly matching the double-mutant frequency represent the square of the mutant to wild-type ratio, multiplied by the wild-type density, for each $m$. In (B), the cubes of the ratios are also shown, roughly matching the triple-mutant density. Parameter values used in this simulation are: $P = 10^4$, $f = 0.5$, $\beta = 0.2$, $\mu_{W\to M} = \mu_{M\to M2} = \mu_{M2\to M3} = 0.1$, $\mu_{M3\to M2} = \mu_{M2\to M} = \mu_{M\to W} = 0.01$, $\delta_W = 0.05$, $\delta_M = 0.1$, $\delta_{M2} = 0.2$, $\delta_{M3} = 0.4$.

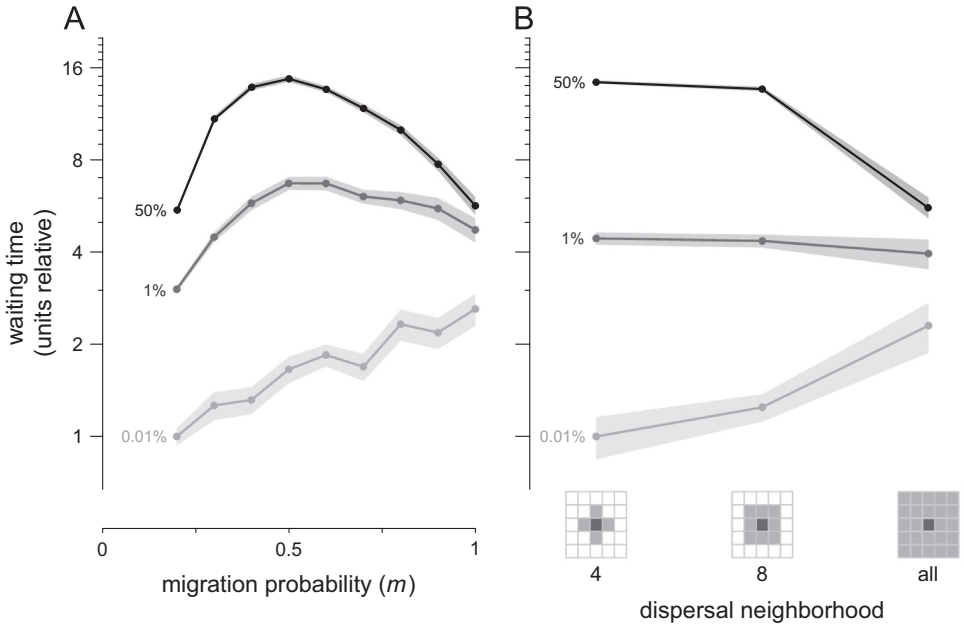


**Fig. 5.** Discovery versus spread in simulated populations. Both metapopulation (A) and lattice (B) simulations were initialized with wild type genotypes only. There are three successively more deleterious mutants comprising a valley between the wild type genotype and a highly beneficial mutant (accessible from the third deleterious mutant). The waiting time before the beneficial mutant reaches a given frequency is shown. As migration becomes more limited, the waiting time to the discovery of the beneficial mutant decreases (as $m$ decreases from 1, the 0.01% profile drops; as the dispersal neighborhood shrinks, the 0.01% profile drops). However, a greater degree of structure inhibits the spread of these beneficial mutants (and thus the time to reach a substantial frequency of the beneficial mutant can increase under initial limitations to migration—see 1% and 50% trajectories). Extremely limited migration decreases total population size, facilitating spread. Fixation does not occur as back mutation is allowed and the numbers of patches are large. Points and shaded regions represent the mean $\pm$ SEM of 24–36 (A) or 16 (B) replicate simulations using parameter values $f=0.5$, $\beta=0.2$, $\mu_{W\to M}=\mu_{M\to M2}=\mu_{M2\to M3}=\mu_{M3\to M4}=0.01$, $\mu_{M4\to M3}=\mu_{M3\to M2}=\mu_{M2\to M}=\mu_{M\to W}=0.01$, $\delta_W=0.05$, $\delta_M=0.1$, $\delta_{M2}=0.2$, $\delta_{M3}=0.4$, $\delta_{M4}=0.025$. For (A), $P=10^4$. For (B), $L=200$.

migration protects not only deleterious single mutants from competition with wild types, but also relatively deleterious double mutants from competition with single mutants (and wild types). When deleterious double mutants are added to our metapopulation simulation, we see the amplified effect of limited migration on the double-mutant frequency (Fig. 4A); when we add triple mutants, the effect of limited migration amplifies further (Fig. 4B). This effect on double and triple mutants is also observed in our lattice-based approach (Sup. Fig. 4).

For sufficiently wide valleys, a population starting with only wild type individuals may discover the peak genotype faster when its migration is limited (Fig. 5). For certain parameter values, a population whose migration is limited may cross even the narrowest valley – one deleterious mutant between two peak genotypes – faster than an unstructured population (Bitbol and Schwab, 2014). Note that when only upward steps are required for adaptation (e.g., a smooth landscape) then the rate-limiting step in adaptation is the spread of beneficial genotypes, and thus limited migration will inhibit adaptation (Kryazhimskiy et al., 2012).

The limited migration of individuals generally slows the spread of advantageous traits. However, it is precisely this dampening of competition that can allow spatially structured populations to safeguard deleterious mutants. By harboring this diversity, it may be possible for structured populations to discover novel genotypes faster, even if the benefit spreads more slowly.

## Acknowledgements

The authors thank Michael Whitlock and Ruth Shaw for their helpful comments, and Jacob Speidel, Brian Connelly, and Joshua Nahum for their coding help. The authors are also grateful to the anonymous reviewers for their many useful suggestions. This work was supported in part by the National Science Foundation Graduate Research Fellowship under grants nos. DGE-0718124 and DGE-1256082, as well as NSF Cooperative Agreement DBI-0939454 and an NSF CAREER Award (DEB0952825).

## Appendix 1. Moment equations

### *Mean density dynamics*

In this appendix, we derive the dynamical equations for our first and second-order moments. We assume our population inhabits a metapopulation of infinite patches (each of which houses a finite population), allowing us to use expectation values for our patch dynamics. We start with the dynamics of the mean genotype abundances. Let $q_i(t)$ be a random variable giving the number of individuals of genotype $i\in\{W,M\}$ within a randomly selected patch at time $t$. If we consider a period of time, $\Delta t$, small enough that the probability of more than one event occuring during that interval is vanishing small, we have the following:

$$q_i(t+\Delta t)=\begin{cases} q_i(t)-1 & \text{with probability } P_i^- \\ q_i(t)+1 & \text{with probability } P_i^+ \end{cases}, \tag{A1.1}$$

where

$$P_i^- = \delta_i q_i(t)\Delta t, \tag{A1.2}$$

and

$$\begin{aligned} P_i^+ = {} & (1-m_i)(1-\mu_{i\to j})F_i\Big(q_i(t),q_j(t)\Big)q_i(t)\Delta t \\ & +(1-m_j)\mu_{j\to i}F_j\Big(q_j(t),q_i(t)\Big)q_j(t)\Delta t \\ & +m_i(1-\mu_{i\to j})E\Big[F_i\Big(q_i(t),q_j(t)\Big)q_i(t)\Big]\Delta t \\ & +m_j\mu_{j\to i}E\Big[F_j\Big(q_j(t),q_i(t)\Big)q_j(t)\Big]\Delta t. \end{aligned} \tag{A1.3}$$

where $E$ is the expectation value over all patches. Thus, the expected change in $q_i$ over our small interval of time is given by:

$$E[\Delta q_i] = E[P_i^+ - P_i^-]. \tag{A1.4}$$

For typographical convenience, we drop the explicit time dependence in our notation for the terms and equations that follow. We use the following notations

$$N_i = E[q_i],$$

$$\sigma_i^2 = \mathrm{var}[q_i],$$

$$C = \mathrm{cov}\left[q_i, q_j\right] = \mathrm{cov}\left[q_j, q_i\right],$$

and we have the following relations

$$E[q_i^2] = N_i^2 + \sigma_i^2, \tag{A1.5}$$

$$E\left[q_i q_j\right] = N_i N_j + C. \tag{A1.6}$$

Using (A1.5), (A1.6), and our per capita birth rate of $F_i(n_i, n_j) = f_i - \beta_i(n_i + \alpha_{ij} n_j)$ we can rewrite (A1.4) as follows:

$$\frac{E[\Delta q_i]}{\Delta t} = -\delta_i N_i + (1-\mu_{i\to j})\left\{f_i N_i - \beta_i\left(\left(N_i^2 + \sigma_i^2\right) + \alpha_{ij}(N_i N_j + C)\right)\right\} + \mu_{j\to i}\left\{f_j N_j - \beta_j\left(\left(N_j^2 + \sigma_j^2\right) + \alpha_{ji}(N_i N_j + C)\right)\right\}. \tag{A1.7}$$

Taking the limit $\Delta t \to 0$, and factoring $N_i$ from the second term and $N_j$ from the third term, we have

$$\frac{dN_i}{dt} = -\delta_i N_i + (1-\mu_{i\to j})\left\{f_i - \beta_i\left(\left(N_i + \frac{\sigma_i^2}{N_i}\right) + \alpha_{ij}\left(N_j + \frac{C}{N_i}\right)\right)\right\} N_i + \mu_{j\to i}\left\{f_j - \beta_j\left(\left(N_j + \frac{\sigma_j^2}{N_j}\right) + \alpha_{ji}\left(N_i + \frac{C}{N_j}\right)\right)\right\} N_j. \tag{A1.8}$$

The terms $N_i + \sigma_i^2/N_i$ and $N_j + C/N_i$ (and the two other similar terms) are more approachable if we allow $N_{j|i}$ to represent the expected number of individuals of genotype $j$ in the patch of a randomly chosen individual of genotype $i$ (rather than a randomly chosen patch). Eq. (A1.6) can now be rewritten as $N_i N_{j|i} = E[q_i q_j] = N_i N_j + C$, and therefore $N_{j|i} = N_j + C/N_i$. Similarly, Eq. (A1.5) yields $N_{i|i} = N_i + \sigma_i^2/N_i$. Using this notation, (A1.8) can be simplified to

$$\frac{dN_i}{dt} = -\delta_i N_i + (1-\mu_{i\to j})\, F_i\left(N_{i|i}, N_{j|i}\right) N_i + \mu_{j\to i}\, F_j\left(N_{j|j}, N_{i|j}\right) N_j. \tag{A1.9}$$

From these equations we see that change in the first order moment (the expected density of genotype $i$) depends on second order moments (the variances and covariance of genotype densities). Thus, we now derive the dynamical equations for the change in these second order moments.

*Variance dynamics*

Again, we consider a very small interval of time, $\Delta t$. The following holds:

$$\Delta q_i^2(t) = q_i^2(t+\Delta t) - q_i^2(t). \tag{A1.10}$$

Using (A1.1), (A1.2), (A1.3) and (A1.10), and again dropping the explicit time dependence in our notation, we see that

$$\Delta q_i^2 = (q_i - 1)^2 - q_i^2 = -2q_i + 1 \text{ with probability } P_i^-, \text{ and} \tag{A1.11}$$

$$\Delta q_i^2 = (q_i + 1)^2 - q_i^2 = 2q_i + 1 \text{ with probability } P_i^+. \tag{A1.12}$$

Thus, the expected change in $q_i^2$ is given by:

$$E[\Delta q_i^2] = E[(2q_i + 1)P_i^+ + (-2q_i + 1)P_i^-] \tag{A1.13}$$

We have the following relations:

$$E[q_i^3] = T_{iii} + 3N_i\sigma_i^2 + N_i^3, \tag{A1.14}$$

$$E\left[q_i^2 q_j\right] = T_{iij} + 2N_i C + N_j\sigma_i^2 + N_i^2 N_j, \tag{A1.15}$$

where $T_{iii}$ and $T_{iij}$ are the central third-order moments. Because of (A1.5), we also have

$$\frac{E[\Delta q_i^2]}{\Delta t} = \frac{\Delta N_i^2}{\Delta t} + \frac{\Delta \sigma_i^2}{\Delta t} \tag{A1.16}$$

Using (A1.5), (A1.6), (A1.14), (A1.15) and (A1.16), taking the limit $\Delta t \to 0$ and using the chain rule (i.e., $\frac{dN_i^2}{dt} = 2N_i\frac{dN_i}{dt}$), we have

$$\frac{d\sigma_i^2}{dt} = \frac{dN_i}{dt} + 2(N_i - \sigma_i^2)\delta_i + 2(1-m_i)(1-\mu_{i\to j})\left\{f_i\sigma_i^2 - \beta_i\left(T_{iii} + 2N_i\sigma_i^2 + \alpha_{ij}(T_{iij} + N_i C + N_j\sigma_i^2)\right)\right\} + 2(1-m_j)\mu_{j\to i}\left\{f_j C - \beta_j\left(T_{jji} + 2N_j C + \alpha_{ji}(T_{iij} + N_i C + N_j\sigma_i^2)\right)\right\}. \tag{A1.17}$$

If we describe the third order moments exactly, we will find ourselves needing to describe fourth order moments, which will in turn require fifth order moments, and so on. Here we use our moment closure technique.

*Closing the moments*

When migration is absolute (i.e., $m_i = m_j = 1$), the random variables $q_i$ and $q_j$ are independently Poisson distributed among the patches with means equal to $N_i$ and $N_j$, respectively (see Neuhauser, 2002). For any independent Poisson-distributed random variables, their third order moments can be described exactly in terms of lower-order moments; the homogeneous third central moment is the corresponding first-order moment, while all mixed third central moments are zero:

$$T_{iii} = N_i$$

$$T_{iij} = 0$$

By using these substitutions as approximations when $m_i \approx m_j \approx 1$, we obviate the need to describe higher order moments. This moment closure technique is exact when $m_i = m_j = 1$, and approximate when $m_i \approx m_j \approx 1$.

Substituting our approximations for the third central moments into Eq. (A1.17), we have

$$\frac{d\sigma_i^2}{dt} = \frac{dN_i}{dt} + 2(N_i - \sigma_i^2)\delta_i + 2(1-m_i)(1-\mu_{i\to j})\left\{f_i\sigma_i^2 - \beta_i\left(N_i + 2N_i\sigma_i^2 + \alpha_{ij}(N_i C + N_j\sigma_i^2)\right)\right\} + 2(1-m_j)\mu_{j\to i}\left\{f_j C - \beta_j\left(2N_j C + \alpha_{ji}(N_i C + N_j\sigma_i^2)\right)\right\}. \tag{A1.18}$$

*Covariance dynamics*

Again, we consider a very small interval of time, $\Delta t$. The following holds:

$$\Delta\left(q_i(t)q_j(t)\right) = q_i(t+\Delta t)q_j(t+\Delta t) - q_i(t)q_j(t) \tag{A1.19}$$

Using, (A1.1), (A1.2), (A1.3) and (A1.16), and again dropping the explicit time dependence in our notations, we see that

$$\Delta\left(q_iq_j\right)=(q_i-1)q_j-q_iq_j=-q_j \text{ with probability } P_i^-, \text{ and} \quad \text{(A1.20)}$$

$$\Delta\left(q_iq_j\right)=(q_i+1)q_j-q_iq_j=q_j \text{ with probability } P_i^+. \quad \text{(A1.21)}$$

Thus, the expected change in the quantity $q_iq_j$ is:

$$E\left[\Delta\left(q_iq_j\right)\right]=E\left[-q_jP_i^- -q_iP_j^- +q_jP_i^+ +q_iP_j^+\right]. \quad \text{(A1.22)}$$

From (A1.6), we have the following relation:

$$\frac{E[\Delta\left(q_iq_j\right)]}{\Delta t}=\frac{\Delta(N_iN_j)}{\Delta t}+\frac{\Delta C}{\Delta t}. \quad \text{(A1.23)}$$

Using (A1.5), (A1.6), (A1.14), (A1.15) and (A1.23), taking the limit $\Delta t\to 0$, and using the product rule (i.e., $\frac{\mathrm{d}(N_iN_j)}{\mathrm{d}t}=N_i\frac{\mathrm{d}N_j}{\mathrm{d}t}+N_j\frac{\mathrm{d}N_i}{\mathrm{d}t}$), we have

$$\begin{aligned}\frac{dC}{dt}=&-(\delta_i+\delta_j)C+(1-m_i)(1-\mu_{i\to j})\{f_iC-\beta_i((T_{iij}+2N_iC)\\&+\alpha_{ij}\left(T_{jji}+N_jC+N_i\sigma_j^2\right)\Big)\}+(1-m_j)\mu_{j\to i}\Big\{f_j\sigma_j^2-\beta_j\Big(\left(T_{jjj}+2N_j\sigma_j^2\right)\\&+\alpha_{ji}\left(T_{jji}+N_jC+N_i\sigma_j^2\right)\Big)\}+(1-m_j)(1-\mu_{j\to i})\\&\Big\{f_jC-\beta_j((T_{jji}+2N_jC)+\alpha_{ji}(T_{iij}+N_iC+N_j\sigma_i^2))\Big\}\\&+(1-m_i)\mu_{i\to j}\{f_i\sigma_i^2-\beta_i((T_{iii}+2N_i\sigma_i^2)\\&+\alpha_{ij}(T_{iij}+N_iC+N_j\sigma_i^2))\}.\end{aligned}$$

Substituting our approximations for the third central moments yields

$$\begin{aligned}\frac{dC}{dt}=&-(\delta_i+\delta_j)C+(1-m_i)(1-\mu_{i\to j})\Big\{f_iC-\beta_i\Big(2N_iC+\alpha_{ij}\left(N_jC+N_i\sigma_j^2\right)\Big)\Big\}\\&+(1-m_j)\mu_{j\to i}\Big\{f_j\sigma_j^2-\beta_j\Big(\left(N_j+2N_j\sigma_j^2\right)+\alpha_{ji}\left(N_jC+N_i\sigma_j^2\right)\Big)\Big\}\\&+(1-m_j)(1-\mu_{j\to i})\Big\{f_jC-\beta_j\Big(2N_jC+\alpha_{ji}(N_iC+N_j\sigma_i^2)\Big)\Big\}\\&+(1-m_i)\mu_{i\to j}\Big\{f_i\sigma_i^2-\beta_i\Big((N_i+2N_i\sigma_i^2)+\alpha_{ij}(N_iC+N_j\sigma_i^2)\Big)\Big\}.\end{aligned} \quad \text{(A1.24)}$$

With Eqs. (A1.8), (A1.18) and (A1.24), we have a closed system of five differential equations describing the dynamics of $N_W$, $N_M$, $\sigma_W^2$, $\sigma_M^2$ and $C$.

## Appendix 2. Equilibrium densities

At equilibrium, $dN_W/dt=dN_M/dt=0$. In this appendix, we assume $m_W=m_M=m$, $f_W=f_M=f$, $\beta_W=\beta_M=\beta$, $\alpha_{WM}=\alpha_{MW}=1$, $\mu_{W\to M}=\mu$ and $\mu_{M\to W}=0$. Using these assumptions and Eq. (A1.8), the equilibrium value $\hat{N}_W$ must satisfy the following:

$$0=-\delta_W\hat{N}_W+(1-\mu)\Big\{f\hat{N}_W-\beta\Big(\hat{N}_W^2+\hat{\sigma}_W^2+\hat{N}_W\hat{N}_M+\hat{C}\Big)\Big\}. \quad \text{(A2.1)}$$

If we assume that $m=1$, then $q_i$ and $q_j$ are independently Poisson distributed, and therefore:

$$\hat{\sigma}_W^2=\hat{N}_W, \quad \text{(A2.2)}$$

$$\hat{C}=0. \quad \text{(A2.3)}$$

Using (A2.2) and (A2.3), the non-zero equilibrium in (A2.1) is

$$\hat{N}_W=\frac{(1-\mu)(f-\beta)-\delta_W}{(1-\mu)\beta}-\hat{N}_M. \quad \text{(A2.4)}$$

We denote the total density at equilibrium $\hat{T}=\hat{N}_W+\hat{N}_M$. So, we have

$$\hat{T}=\frac{(1-\mu)(f-\beta)-\delta_W}{(1-\mu)\beta}, \quad \text{(A2.5)}$$

and

$$\hat{N}_W=\hat{T}-\hat{N}_M. \quad \text{(A2.6)}$$

Now we turn to the equilibrial density of the mutant genotype, $\hat{N}_M$, again using (A1.8):

$$\begin{aligned}0=&-\delta_M\hat{N}_M+\Big\{f\hat{N}_M-\beta\Big(\hat{N}_M^2+\hat{\sigma}_M^2+\hat{N}_M\hat{N}_W+\hat{C}\Big)\Big\}\\&+\mu\Big\{f\hat{N}_W-\beta\Big(\hat{N}_W^2+\hat{\sigma}_W^2+\hat{N}_M\hat{N}_W+\hat{C}\Big)\Big\}.\end{aligned} \quad \text{(A2.7)}$$

If we are assuming $m=1$, the resulting Poisson distribution yields

$$\hat{\sigma}_M^2=\hat{N}_M, \quad \text{(A2.8)}$$

Using (A2.3), (A2.6), and (A2.8), the non-zero mutant equilibrium in (A2.7) is

$$\hat{N}_M=\frac{-\mu\hat{T}\Big(f-\beta(\hat{T}+1)\Big)}{-\delta_M+(1-\mu)\Big(f-\beta(\hat{T}+1)\Big)}. \quad \text{(A2.9)}$$

After substituting, using (A2.5), and simplifying, Eqs. (A2.4) and (A2.9) simplify to the following:

$$\hat{N}_W=\frac{\{(1-\mu)\delta_M-\delta_W\}\{(f-\beta)(1-\mu)-\delta_W\}}{\beta(\delta_M-\delta_W)(1-\mu)^2}, \quad \text{(A2.10)}$$

$$\hat{N}_M=\frac{\mu\delta_W\{(f-\beta)(1-\mu)-\delta_W\}}{\beta(\delta_M-\delta_W)(1-\mu)^2}. \quad \text{(A2.11)}$$

In order for $\hat{N}_W$ and $\hat{N}_M$ to be positive, we must have the following two conditions:

$$(1-\mu)(f-\beta)>\delta_W, \quad \text{(A2.12)}$$

$$(1-\mu)\delta_M>\delta_W. \quad \text{(A2.13)}$$

Note that (A2.13) is more stringent than the already assumed $\delta_M>\delta_W$. In all of what follows, we will assume conditions (A2.12) and (A2.13), except where explicitly mentioned. When $\mu=0$, Eqs. (A2.10) and (A2.11) simplify to:

$$\hat{N}_W=\frac{f-\beta-\delta_W}{\beta},\ \hat{N}_M=0, \quad \text{(A2.14)}$$

Which gives a positive density of the wild type (by condition (A2.12)) and no mutant density. When $(1-\mu)\delta_M=\delta_W$ (i.e., right where Eq. (A2.13) starts to be violated), Eqs. (A2.10) and (A2.11) simplify to:

$$\hat{N}_W=0,\ \hat{N}_M=\frac{f-\beta-\delta_M}{\beta}, \quad \text{(A2.15)}$$

Which gives a positive density of the mutant (by condition (A2.12), replacing $\delta_W$ with $(1-\mu)\delta_M$) and no wild-type density. Equilibria in (A2.14) and (A2.15) agree with single species equilibria from ecological models (Neuhauser, 2002; Pacala and Levin, 1997).

We let the fraction of mutants in the population be given by $\Phi(t)$, where

$$\Phi(t)=\frac{N_M(t)}{N_W(t)+N_M(t)}. \quad \text{(A2.16)}$$

Using Eqs. (A2.10) and (A2.11), the mutation–selection balance under full migration is:

$$\hat{\Phi}=\frac{\mu\delta_W}{(1-\mu)(\delta_M-\delta_W)}. \quad \text{(A2.17)}$$

Note that if $\mu=0$, then $\hat{\Phi}=0$. That is, when there is no supply of new mutants through mutation, selection "wins" and no mutants remain at equilibrium; this corresponds to the special case of (A2.14). If $\mu=(\delta_M-\delta_W)/\delta_M$, then $\hat{\Phi}=1$. That is, as $\mu\to(\delta_M-\delta_W)/\delta_M$, mutation "wins" by overwhelming selection and only mutants remain at equilibrium; this corresponds to the special case of (A2.15).

## Appendix 3. The effect of structure on mutation–selection balance

In order to explore the role of structure on the mutant frequency, we look at

$$\frac{\partial\Phi}{\partial m}=\frac{\frac{\partial N_M}{\partial m}N_W-\frac{\partial N_W}{\partial m}N_M}{(N_W+N_M)^2}. \tag{A3.1}$$

Here we will evaluate $\partial\Phi/\partial m\big|_{m=1}$. In order to do so, we must find $\partial N_W/\partial m\big|_{m=1}$ and $\partial N_M/\partial m\big|_{m=1}$, which we abbreviate with $\partial N_W/\partial m\big|_1$ and $\partial N_M/\partial m\big|_1$. To do this we differentiate (A1.8) with respect to $m$ and evaluate at the $m=1$ equilibrium. We start with Eq. (A1.8) where $i=W$.

$$0=\left\{(1-\mu)\left[f-\beta(2\hat{N}_W+\hat{N}_M)\right]-\delta_W\right\}\frac{\partial N_W}{\partial m}\bigg|_1-\left\{\beta(1-\mu)\hat{N}_W\right\}\frac{\partial N_M}{\partial m}\bigg|_1$$
$$-\left\{\beta(1-\mu)\right\}\frac{\partial\sigma_W^2}{\partial m}\bigg|_1-\left\{\beta(1-\mu)\right\}\frac{\partial C}{\partial m}\bigg|_1. \tag{A3.2}$$

Again, we see that we will need to consider partial derivatives of higher-order moments with respect to $m$ to solve (A3.1). By differentiating Eq. (A1.8) with $i=M$, (A1.18) with $i=W$, (A1.18) with $i=M$, and (A1.24), all with respect to $m$ and making the appropriate substitutions for when $m=1$, we obtain other equalities involving partial derivatives (similar to (A3.2)). This leads to the following linear system:

$$\mathbf{A}\vec{\partial}_1=\vec{c}, \tag{A3.3}$$

where

$$\mathbf{A}=\begin{bmatrix}(1-\mu)\left\{f-\beta(2\hat{N}_W+\hat{N}_M)\right\}-\delta_W & -\beta(1-\mu)\hat{N}_W & -\beta(1-\mu) & 0 & -\beta(1-\mu)\\ -\beta(1+\mu)\hat{N}_M+\mu(f-2\beta\hat{N}_W) & f-2\beta\hat{N}_M-\beta(1+\mu)\hat{N}_W-\delta_M & -\beta\mu & -\beta & -\beta(1+\mu)\\ \delta_W & 0 & -\delta_W & 0 & 0\\ 0 & \delta_M & 0 & -\delta_M & 0\\ 0 & 0 & 0 & 0 & -\delta_W-\delta_M\end{bmatrix},$$

$$\vec{\partial}_1=\begin{bmatrix}\frac{\partial N_W}{\partial m}\big|_1\\ \frac{\partial N_M}{\partial m}\big|_1\\ \frac{\partial\sigma_W^2}{\partial m}\big|_1\\ \frac{\partial\sigma_M^2}{\partial m}\big|_1\\ \frac{\partial C}{\partial m}\big|_1\end{bmatrix},\ \text{and}\ \vec{c}=\begin{bmatrix}0\\ 0\\ \hat{N}_W(1-\mu)\left\{f-\beta\left(1+2\hat{N}_W+\hat{N}_M\right)\right\}\\ \hat{N}_M\left\{f-\beta\left(1+2\hat{N}_M+(1+\mu)\hat{N}_W\right)\right\}\\ \hat{N}_W\left\{\mu\left(f-\beta\left(1+2\hat{N}_W\right)\right)-2\beta\hat{N}_M\right\}\end{bmatrix}.$$

Solving system (A3.3) and using $\partial N_W/\partial m\big|_1$ and $\partial N_M/\partial m\big|_1$ for Eq. (A3.1) gives the following:

$$\frac{\partial\Phi}{\partial m}\bigg|_{m=1}=-\frac{\beta\mu}{(1-\mu)^2\delta_M\left\{(\delta_M/\delta_W)^2-1\right\}}. \tag{A3.4}$$

We abbreviate $\partial\Phi/\partial m\big|_{m=1}$ as $\partial_m\Phi$. From Eq. (A3.4), it is not difficult to show that $\partial|\partial_m\Phi|/\partial\beta>0$, $\partial|\partial_m\Phi|/\partial\mu>0$, $\partial|\partial_m\Phi|/\partial\delta_M<0$, and $\partial|\partial_m\Phi|/\partial\delta_W>0$. That is, as the competition coefficient $\beta$, the mutation rate $\mu$, or the death rate of the wild type genotype increase, the addition of structure to an unstructured system leads to a greater increase in the mutant class frequency. As the death rate of the mutant is increased, the addition of structure to an unstructured system leads to a smaller increase in the mutant class frequency.

## Appendix 4. Gillespie algorithm

Our simulation is based on a Gillespie algorithm (Gillespie, 1977) that we coded in the Python 2.7 scripting language. The Gillespie algorithm simulates a possible trajectory of a continuous time stochastic system.

In our system of $P$ connected patches, patches must be initialized before simulating evolution. Unless otherwise indicated, we seeded our patches with wild-type and mutant individuals by repeatedly drawing from independent Poisson distributions whose parameters are the full migration equilibria $\hat{N}_W$ and $\hat{N}_M$ from (A2.10) and (A2.11), respectively. The initial population defines update zero, for which the time variable $t$ is also zero. As the populations are seeded from their corresponding $m=1$ equilibrium distributions, structure is introduced as any limited migration simulation begins.

Evolution of the population occurs over "update" steps. First, for update $u$ each patch $p$ receives four "weights", corresponding to the four possible events in that patch: a wild-type birth, a mutant birth, a wild-type death, and a mutant death. Each event's weight is proportional to its rate. If we let the number of genotypes $W$ and $M$ in a patch $p$ at update step $u$ be given by $n_W(u,p)$ and $n_M(u,p)$, respectively, then the weights are defined as follows:

$$k_1(u,p)=fn_W(u,p)-\beta n_W(u,p)(n_W(u,p)+n_M(u,p)), \tag{A4.1}$$

$$k_2(u,p)=fn_M(u,p)-\beta n_M(u,p)(n_W(u,p)+n_M(u,p)), \tag{A4.2}$$

$$k_3(u,p)=\delta_W n_W(u,p), \tag{A4.3}$$

$$k_4(u,p)=\delta_M n_M(u,p). \tag{A4.4}$$

The event that is attempted at update step $u$ is either a death, a birth with migration, or a birth without migration, and the decision is made stochastically using the following weights:

$$K_{death}(u)=\sum_{p=1}^{P}k_3(u,p)+\sum_{p=1}^{P}k_4(u,p), \tag{A4.5}$$

$$K_{birth\_mig}(u)=m\left\{\sum_{p=1}^{P}k_1(u,p)+\sum_{p=1}^{P}k_2(u,p)\right\}, \tag{A4.6}$$

$$K_{birth\_natal}(u) = (1-m)\left\{\sum_{p=1}^{P}|k_1(u,p)| + \sum_{p=1}^{P}|k_2(u,p)|\right\}. \quad \text{(A4.7)}$$

Time increment $\Delta t$ is drawn from an exponential distribution whose rate is equal to $K_{death}(u)+K_{birth\_mig}(u)+K_{birth\_natal}(u)$, and time parameter $t$ is incremented to $t+\Delta t$. For Fig. 2 and Supplementary Fig. 1, the time increment's exponential distribution rate is equal to $k_1+k_2+k_3+k_4$, which does not noticeably affect the resulting time steps.

The $K_{birth\_natal}(u)$ terms are summed using absolute values due to potentially negative intra-patch birth rates. Since birth rates decrease linearly with density, $k_1(u,p)$ and $k_2(u,p)$ may be negative occasionally in particularly crowded patches. For migrating events, the negative births reduce the mean birth rate, which was never negative for the conditions of our simulations. For non-migrating births, a negative birth event decrements, rather than increments, the chosen genotype population in a patch. Thus, for any patch $r$, if a non-migrating birth event is chosen:

$$n_W(u+1,r) = n_W(u,r) + \mathrm{sgn}(k_1(u,r)) \text{ with probability } (1-\mu)\frac{|k_1(u,r)|}{K_{birth\_natal(u)}}, \quad \text{(A4.8)}$$

$$n_M(u+1,r) = n_M(u,r) + \mathrm{sgn}(k_2(u,r)) \text{ with probability } \frac{|k_2(u,r)|}{K_{birth\_natal}(u)} + \mu\frac{|k_1(u,r)|}{K_{birth\_natal}(u)}, \quad \text{(A4.9)}$$

If a migrating birth event is chosen (i.e., a migrant will 'land' on a random patch):

$$n_W(u+1,r) = n_W(u,r)+1 \text{ with probability } \frac{1-\mu}{P}\sum_{p=1}^{P}\frac{k_1(u,p)}{K_{birth\_mig}(u)}, \quad \text{(A4.10)}$$

$$n_M(u+1,r) = n_M(u,r)+1 \text{ with probability } \frac{1}{P}\left\{\sum_{p=1}^{P}\frac{k_2(u,p)}{K_{birth\_mig}(u)} + \mu\sum_{p=1}^{P}\frac{k_1(u,p)}{K_{birth\_mig}(u)}\right\}, \quad \text{(A4.11)}$$

If a death event is chosen:

$$n_W(u+1,r) = n_W(u,r)-1 \text{ with probability } \frac{k_3(u,r)}{K_{death}(u)}, \quad \text{(A4.12)}$$

$$n_M(u+1,r) = n_M(u,r)-1 \text{ with probability } \frac{k_4(u,r)}{K_{death}(u)} \quad \text{(A4.13)}$$

Simulations were run for 200,000 updates (for $1 \geq m \geq 0.95$), 500,000 updates (for $0.95 > m \geq 0.7$), or 1000,000 updates (for $0.7 > m > 0$). Data was recorded every 5000 updates, and equilibrial values represent a simulation's average state over its final 10,000 updates.

| Term | Explanation | Assumptions |
|---|---|---|
| $W$, $M$ | wild type, mutant genotype labels | |
| $i$, $j$ | arbitrary genotype indices | |
| $n_i$ | number of genotype $i$ in a patch | |
| $f_i$ | intrinsic growth rate of genotype $i$ | $f_W = f_M = f$ |
| $\beta_i$ | competition parameter of genotype $i$ | $\beta_W = \beta_M = \beta$ |
| $\alpha_{ij}$ | competitive effect of genotype $j$ on genotype $i$ | $\alpha_{WM} = \alpha_{MW} = 1$ |
| $\delta_i$ | death rate of genotype $i$ | $\delta_M > \delta_W > 0$ |
| $\mu_{i\to j}$ | mutation probability from genotype $i$ to $j$ | $\mu_{W\to M} = \mu$, $\mu_{M\to W} = 0$ |
| $m_i$ | migration probability of genotype $i$ | $m_W = m_M = m$ |
| $N_i$ | expected number of genotype $i$ over all patches | |
| $N_{i\|j}$ | expected number of genotype $i$ in a patch of a randomly chosen genotype $j$ | |
| $\sigma_i^2$ | variance of genotype $i$ over all patches | |
| $C$ | covariance between genotypes $i$ and $j$ over all patches | |
| $\Phi$ | fraction of mutants at mutation–selection balance | |

## Appendix B. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jtbi.2015.05.003.

## References


Bitbol, A.-F., Schwab, D.J., 2014. Quantifying the role of population subdivision in evolution on rugged fitness landscapes. PLoS Comput. Biol. 10, e1003778. http://dx.doi.org/10.1371/journal.pcbi.1003778.

Bjedov, I., Tenaillon, O., Gérard, B., Souza, V., Denamur, E., Radman, M., Taddei, F., Matic, I., 2003. Stress-induced mutagenesis in bacteria. Science 300, 1404–-1409. http://dx.doi.org/10.1126/science.1082240.

Bolker, B., Pacala, S.W., 1997. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. Theor. Popul. Biol. 52, 179–197.

Cherry, J.L., Wakeley, J., 2003. A diffusion approximation for selection and drift in a subdivided population. Genetics 163, 421–428.

Crow, J.F., Kimura, M., 1970. An introduction to Population Genetics Theory.

Débarre, F., Lion, S., Baalen, Gandon, S., 2012. Evolution of host life-history traits in a spatially structured host-parasite system. Am. Nat. 179, 52–63. http://dx.doi.org/10.1086/663199.

Durrett, R., Levin, S., 1997. Allelopathy in spatially distributed populations. J. Theor. Biol. 185, 165–171.

Evans, K.M., Chepurnov, V.A., Sluiman, H.J., Thomas, S.J., Spears, B.M., Mann, D.G., 2009. Highly differentiated populations of the freshwater diatom Sellaphora capitata suggest limited dispersal and opportunities for allopatric speciation. Protist 160, 386–396. http://dx.doi.org/10.1016/j.protis.2009.02.001.

Eyre-Walker, A., Keightley, P.D., 2007. The distribution of fitness effects of new mutations. Nat. Rev. Genet. 8, 610–618. http://dx.doi.org/10.1038/nrg2146.

Fisher, R.A., 1930. The distribution of gene ratios for rare mutations. Proc. R. Soc. (Edinburgh) 50, 205–220.

Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81, 2340–2361. http://dx.doi.org/10.1021/j100540a008.

Glémin, S., Ronfort, J., Bataillon, T., 2003. Patterns of inbreeding depression and architecture of the load in subdivided populations. Genetics 165, 2193–2212.

Haegeman, B., Loreau, M., 2011. A mathematical synthesis of niche and neutral theories in community ecology. J. Theor. Biol. 269, 150–165.

Haldane, J.B.S., 1927. A mathematical theory of natural and artificial selection, Part V: Selection and mutation. Math. Proc. Cambridge Philos. Soc. 23, 838–844. http://dx.doi.org/10.1017/S0305004100015644.

Hauert, C., Doebeli, M., 2004. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. Nature 428, 643–646.

Howells, E.J., Willis, B.L., Bay, L.K., van Oppen, M.J.H., 2013. Spatial and temporal genetic structure of Symbiodinium populations within a common reef-building coral on the Great Barrier Reef. Mol. Ecol. 22, 3693–3708. http://dx.doi.org/10.1111/mec.12342.

Jain, K., Krug, J., Park, S.-C., 2011. Evolutionary advantage of small populations on complex fitness landscapes. Evolution 65, 1945–1955. http://dx.doi.org/10.1111/j.1558-5646.2011.01280.x.

Kerr, B., Riley, M.A., Feldman, M.W., Bohannan, B.J.M., 2002. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. Nature 418, 171–174.

Kryazhimskiy, S., Rice, D.P., Desai, M.M., 2012. Population subdivision and adaptation in asexual populations of *Saccharomyces cerevisiae*. Evolution 66, 1931–-1941. http://dx.doi.org/10.1111/j.1558-5646.2012.01569.x.

Martin, P.H., Canham, C.D., 2010. Dispersal and recruitment limitation in native versus exotic tree species: life-history strategies and Janzen–Connell effects. Oikos 119, 807–824. http://dx.doi.org/10.1111/j.1600-0706.2009.17941.x.

Neuhauser, C., 2002. Effects of local interactions and local migration on stability. Theor. Popul. Biol. 62, 297–308.

Pacala, S.W., Levin, S.A., 1997. Biologically generated spatial pattern and the coexistence of competing species, Spatial Ecology. Princeton University Press, Princeton, New Jersey, USA, pp. 204–232.

Pigliucci, M., 2008. Sewall Wright's adaptive landscapes: 1932 vs. 1988. Biol. Philos. 23, 591–603.

Roze, D., Rousset, F., 2004. Joint effects of self-fertilization and population structure on mutation load, inbreeding depression and heterosis. Genetics 167, 1001–-1015. http://dx.doi.org/10.1534/genetics.103.025148.

Rozen, D.E., Habets, M.G.J.L., Handel, A., de Visser, J.A.G.M., 2008. Heterogeneous adaptive trajectories of small populations on complex fitness landscapes. PLoS One 3, e1715. http://dx.doi.org/10.1371/journal.pone.0001715.
Vanpeteghem, D., Haegeman, B., 2010. An analytical approach to spatio-temporal dynamics of neutral community models. J. Math. Biol. 61, 323–357.
Weissman, D.B., Desai, M.M., Fisher, D.S., Feldman, M.W., 2009. The rate at which asexual populations cross fitness valleys. Theor. Popul. Biol. 75, 286–300.
Weissman, D.B., Feldman, M.W., Fisher, D.S., 2010. The rate of fitness-valley crossing in sexual populations. Genetics 186, 1389–1410.
Whitlock, M.C., 2002. Selection, load and inbreeding depression in a large metapopulation. Genetics 160, 1191–1202.
Whitlock, M.C., 2003. Fixation probability and time in subdivided populations. Genetics 164, 767–779.
Wright, S., 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: Proceedings of the Sixth International Congress on Genetics. pp. 356–366.
Wright, S., 1988. Surfaces of selective value revisited. Am. Nat. 131, 115–123.