Bulletin of Mathematical Biology (2007) 69: 837–860
DOI 10.1007/s11538-006-9164-y

ORIGINAL ARTICLE

The Role of Life Cycle and Migration in Selection for Variance in Offspring Number

Max Shpak[a,*], Stephen R. Proulx[b]

[a] *Department of Biological Sciences, University of Texas at El Paso, El Paso, TX 79968, USA*
[b] *Department of Ecology and Evolutionary Biology, Iowa State University, Ames, IA 50011, USA*

Received: 5 June 2006 / Accepted: 1 September 2006 / Published online: 15 November 2006

Abstract For two genotypes that have the same mean number of offspring but differ in the variance in offspring number, natural selection will favor the genotype with lower variance. In such cases, the average growth rate is not sufficient as a measure of fitness or as a predictor of fixation probability. However, the effect of variance in offspring number on the fixation probability of mutant strategies has been calculated under several scenarios with the general conclusion that variance in offspring number reduces fitness in proportion to the inverse of the population size [Gillespie, J., Genetics 76:601–606, 1974; Proulx, S.R., Theor. Popul. Biol. 58:33–47, 2000]. This relationship becomes more complicated under a metapopulation scenario where the "effective" population size depends on migration rate, population structure, and life cycle. It is shown that in a life cycle where reproduction and migration (the birth-migration-regulation life cycle, or BMR) occur prior to density regulation within every deme, the fitness of a strategy depends on migration rate. When migration rates are near zero, the fitness of the strategy is determined by the size of individual demes, so that the strategy favored in small populations tends to be fixed. As migration rate increases and approaches panmixis between demes, the fitness of a reproductive strategy approaches what its value would be in a single, panmictic deme with a population size corresponding to the census size of the metapopulation. Interestingly, when the life cycle is characterized by having density regulation in each deme prior to migration (the BRM life cycle) the fixation probability of a strategy is independent of migration rate. These results are found to be qualitatively consistent with the individual-based simulation results in Shpak [Theor. Biosci. 124:65–85, 2005].

Keywords Variance in fitness · Migration · Metapopulation · Hard and soft selection · Bet-hedging

*Corresponding author.
E-mail addresses: mshpak@utep.edu (Max Shpak), proulx@iastate.edu (Stephen R. Proulx).

1. Introduction: A review of previous work

In many evolutionary systems, there are tradeoffs between producing large numbers of offspring and reducing variance in offspring number (and fitness), since wide fluctuations can readily lead to extinction even when mean fitness is high. Reducing variance in offspring number is accomplished by "bet hedging" and spreading reproductive risk (e.g., Seger and Brockmann, 1987; Stearns, 2000). Variance in fitness can be due to intergenerational environmental fluctuations, and it can be shown that in such cases the geometric mean fitness is the best predictor of the probability of ultimate fixation or loss of an allele (Haldane and Jayakar, 1963; Gillespie, 1977). The effects of environmental stochasticity are independent of population size, since the environmental variation effects the viability and fecundity of every individual of a given genotype in the population in the same way. In contrast, intrinsic variance in fitness due to variation in offspring number within a generation involve differences in fecundity between different individuals of the same genotype. The population effects of this sampling variance on the performance of a genotype do scale with population size, so that selection for bet-hedging should be more pronounced in smaller populations (Gillespie, 1974).

Consider two competing alleles in a haploid population. An individual carrying one allele or the complementary copy produces an average of μ_1 and μ_2 offspring in each generation (respectively), with corresponding variance in offspring number per clutch of σ_1^2 and σ_2^2. It has been shown (Gillespie, 1974) that the fixation probabilities cannot be predicted from arithmetic mean fitness alone, i.e., for sufficiently high variance $\sigma_2^2 < \sigma_1^2$, the second allele can have a higher probability of fixation even when $\mu_2 < \mu_1$. The condition for the first strategy having a higher probability of fixation than the second, first derived by Gillespie using the diffusion approximation is

$$\mu_2 - \frac{\sigma_2^2}{n} < \mu_1 - \frac{\sigma_1^2}{n}. \tag{1}$$

This inequality states that a higher mean offspring number (mean fitness) is not necessarily favored by natural selection if the variance is high because of the possibility of the high mean, high variance strategy producing a lower number of offspring than the competitor in any given trial. Quite intuitively, the variance effect is less pronounced in larger populations due to the averaging effects of reduced sample variance for large n. This has important implications for organisms where there is a tradeoff between producing a high average number of offspring and reducing the variance in offspring number because the outcome of selection for one strategy or another will depend on the population size. For example, a semelparous reproductive strategy, all else being equal, has a higher variance in surviving offspring when clutches succeed or fail as a whole than an iteroparous strategy with the same mean number of offspring. If the semelparous strategy also has a somewhat higher mean, whether it or the competing iteroparous

strategy has a higher fixation probability will depend on whether the population is large or small (see also Demetrius and Gundlach, 1999, 2000; Proulx, 2000; Shpak, 2005).

Gillespie calculated the approximate relationship between mean, variance, and "effective" fitness by a change of variables in a diffusion equation and by collecting the coefficients associated with the first derivative of the density function and frequency p. Following Proulx (2000), the relationship between variance in offspring number and selection can also be derived from first principles by calculating the expected change in the number of individuals $x_1(t)$ carrying alleles of the first type, with

$$x_1(t+1) = \sum_{i=1}^{x_1} (\mu_1 + \xi_1[i]), \tag{2}$$

where $\xi_1[i]$ is a normally distributed random variable with mean 0 and variance σ_1^2 (with corresponding equations for the number of individuals with the second allele x_2). For the purposes of the derivations below, it is assumed that the values μ are close to unity, i.e., having the form $\mu = 1 + \zeta$ where $\zeta \ll 1$. (This is done without loss of information or generality in order to avoid issues of rescaling the selection equations in terms of relative rather than absolute fitness, e.g., Gillespie, 1975, Eq. (15).)

Note that the only contribution to variance considered here is that due to differences in offspring number, the variance due to genetic drift (binomial or multinomial sampling of a fixed number of individuals from an offspring pool) is not explicitly considered as contributing to ξ.

The frequency in the next time step will be

$$p(t+1) = \frac{\sum_{i=1}^{x_1} (\mu_1 + \xi_1[i])}{\sum_{i=1}^{x_1} (\mu_1 + \xi_1[i]) + \sum_{i=1}^{x_2} (\mu_2 + \xi_2[i])}, \tag{3}$$

where p is the frequency of the first or "reference" allele in the system.

At this stage, the effects of drift can be included by considering another sampling process that introduces variance. However, so long as this sampling process is random and equiprobable, the expected frequency of allele 1 in the next generation will be given by the equation mentioned earlier (Proulx, 2000). In order to accurately describe selection in terms of first- and second-order terms alone, it will be assumed that $\xi_1[i]$ is small, of the order ϵ. For convenience, the random variable ξ_i is replaced by ϵz_1, where z_1 is a rescaled random variable of order ~ 1, and ϵ is a constant $\ll 1$. Equation (3) then becomes

$$p(t+1) = \frac{x_1\mu_1 + \sum_{i=1}^{x_1} \epsilon z_1[i]}{x_1\mu_1 + \sum_{i=1}^{x_1} \epsilon z_1[i] + x_2\mu_2 + \sum_{i=1}^{x_2} \epsilon z_2[i]}. \tag{4a}$$

The previous expression can be written as a Taylor series expansion (up to second-order terms) as a function of ϵ about $\epsilon_0 = 0$.

$$p(t+1) = \frac{x_1\mu_1}{x_1\mu_1 + x_2\mu_2} + \epsilon\frac{f_1}{g_1} - \epsilon^2\left(\frac{\sum_{k=1}^{x_1} z_1[k]\left(\sum_{k=1}^{x_1} z_1[k] + \sum_{k=1}^{x_2} z_2[k]\right)}{(x_1\mu_1 + x_2\mu_2)^2} - \frac{x_1\mu_1\left(\sum_{k=1}^{x_1} z_1[k] + \sum_{k=1}^{x_2} z_2[k]\right)^2}{(x_1\mu_1 + x_2\mu_2)^3}\right) + O(\epsilon^3)\ldots \tag{4b}$$

where

$$f_1 = (x_1\mu_1 + x_2\mu_2)\sum_{k=1}^{x_1} z_1[k] - x_1\mu_1\left(\sum_{k=1}^{x_1} z_1[k] + \sum_{k=1}^{x_2} z_2[k]\right)$$

$$g_1 = (x_1\mu_1 + x_2\mu_2)^2.$$

Such an expansion can be simplified under the assumption that each reproductive event in a population is independent, so that the covariances $\mathrm{cov}(z_1[k], z_1[j]) = 0$ and $\mathrm{cov}(z_1[k], z_2[j]) = 0$ for all parents k, j. Furthermore, the expectation values of z_1, z_2 are 0, so that the expected means and variances are:

$$E\left[\sum_{k=1}^{x_1} z_1[k]\right] = E\left[\sum_{k=1}^{x_2} z_2[k]\right] = 0$$

$$E\left[\left(\sum_{k=1}^{x_1} z_1[k]\right)^2\right] = x_1 s_1; \qquad E\left[\left(\sum_{k=1}^{x_2} z_2[k]\right)^2\right] = x_2 s_2.$$

The terms s_1, s_2 are rescaled variance terms such that $\sigma_i^2 = \epsilon^2 s_i$. Substituting these variance and covariance relations into a Taylor expansion (4b), together with the first moments μ, one obtains an estimated expectation value. By abuse of notation, we will use p in place of $p(t)$ on the right-hand side of the equations to denote the frequency of the type 1 allele at time t. Assuming fixed population size, the final form of the expectation value $E[p(t+1)]$ is calculated by substituting np for x_1 and $n(1-p)$ for $x_2 = n - x_1$ (the factor of n in the numerator and denominator cancels). The mean change allele frequency is just the difference between $E[p(t+1)]$ and p, which is

$$E[\Delta p] = E[p(t+1)|p(t)] - p = \frac{p(1-p)(\mu_1 - \mu_2)}{p\mu_1 + (1-p)\mu_2} - \frac{(1-p)p\left(\mu_2\sigma_1^2 - \mu_1\sigma_2^2\right)}{n\left(p\mu_1 + (1-p)\mu_2\right)^3}. \tag{5}$$

In the expression, the first term represents the rate of change in allele frequency due to differences in mean fitness (the deterministic component of the process,

which is the same as the equation for change in gene frequency due to selection when there is no variance in offspring number), while the second term represents the stochastic contribution of offspring variance on the selection differential.

In the limit of weak selection and mean offspring numbers μ_1 and μ_2 near unity, the aforementioned expression can be further simplified to the $M(p)$ term derived by Gillespie (1974),

$$M(p) = p(1-p)\left((\mu_1 - \mu_2) - \frac{\sigma_1^2 - \sigma_2^2}{n}\right). \tag{6}$$

(note that if μ is not of order unity, then (6) must be rescaled such that $\mu = \alpha(1+\zeta)$, with the first term in parentheses scaling as α^{-1} and the second as α^{-2}, due to the order μ^{-1} term associated with the first term on the right-hand side of (5) and the $\mu/\mu^3 = \mu^{-2}$ associated with the second term on the right-hand side of Eq. (5)).

The previous equation is a good approximation to (5) in cases where directional selection is weak and most of the selection differential is due to variance in offspring number. When directional selection is strong, Eq. (5) provides a better estimate of the expected change in allele frequency than (6). For a discussion of the limitations of these approximations, see Proulx (2000).

For variance terms of order $\epsilon \ll 1$, $M(p)$ is the directional component in the diffusion approximation, i.e., the Kolmogorov forward and backward equations (e.g., Kimura, 1964; Crow and Kimura, 1970; Ewens, 2004). The backward equation, which is used to derive probabilities of fixation, is

$$\frac{\partial \phi(p,t)}{\partial t} = M(p)\frac{\partial \phi(p,t)}{\partial p} + \frac{V(p)}{2}\frac{\partial^2 \phi(p,t)}{\partial p^2} \tag{7a}$$

where $\phi(p,t)$ is the probability density function of the frequency p at time t, and $V(p)$ is a term describing the second moments in the changes of allele frequency, again in the limit where selection coefficients and variance terms are of order $\epsilon \ll 1$.

To compute the probability of ultimate fixation, one calculates the limit of $\phi(p,t)$ as $t \to \infty$. If the limiting distribution is unique, then this is equivalent to calculating the equilibrium value of $\phi(p)$ by setting the time derivative equal to zero. In this limit, (7a) reduces to a second order ordinary differential equation

$$M(p)\frac{\mathrm{d}\phi(p,t)}{\mathrm{d}p} + \frac{V(p)}{2}\frac{\mathrm{d}^2\phi(p,t)}{\mathrm{d}p^2} = 0 \tag{7b}$$

which can be solved in terms of the $M(p)$ and $V(p)$ coefficients alone, as

The second moment in the change of the allele frequency is

$$\begin{aligned} E[(\Delta p)^2] &= E\left[(p(t+1) - p(t))^2 \middle| p(t)\right] \\ &= E\left[p^2(t+1) \middle| p(t)\right] - 2pE[p(t+1)|p(t)] + p^2 \end{aligned} \tag{8}$$

The first quadratic term from the previous equation is calculated by applying the same substitutions of variables and the same assumptions about the scaling of the mean and variance terms

$$p^2(t+1) = \frac{\left(\sum_{k=1}^{x_1}(\mu_1 + \xi_1[k])\right)^2}{\left(\sum_{k=1}^{x_2}(\mu_1 + \xi_1[i]) + \sum_{k=1}^{x_2}(\mu_2 + \xi_2[k])\right)^2} \tag{9}$$

The expectation of the second moment minus $2E[p(t+1)] + p^2$ gives one the second moment in the change of allele frequency:

$$\begin{aligned} E[(\Delta p)^2] = {} & \frac{n^2 p^2 \mu_1^2}{(np\mu_1 + n(1-p)\mu_2)^2} - \frac{p(3p\mu_1 - (1-p)\mu_2)\sigma_1^2}{n(p\mu_1 + (1-p)\mu_2)^3} \\ & + \frac{3p^2\mu_1^2(p\sigma_1^2 + (1-p)\sigma_2^2)}{n(p\mu_1 + (1-p)\mu_2)^4} - 2p\left(\frac{p\mu_1}{p\mu_1 + (1-p)\mu_2}\right. \\ & \left. - \frac{(1-p)p(\mu_2\sigma_1^2 - \mu_1\sigma_2^2)}{n(p\mu_1 + (1-p)\mu_2)^3}\right) + p^2. \end{aligned} \tag{10}$$

Both the variance and directional selection terms contribute to the second moment. However, in the limit where the mean values μ approach unity, the previous expression simplifies to the diffusion term $V(p)$ in Gillespie (1975) and Proulx (2000), which is independent of the directional selection terms μ:

$$V(p) = \frac{p(1-p)}{n}((1-p)\sigma_1^2 + p\sigma_2^2). \tag{11}$$

Note that this diffusion term does not take into consideration an additional source of stochasticity that occurs in finite populations even in the absence of variance in offspring number, the sampling of n gametes in each generation from the total pool (as in the classic Wright–Kimura equations for genetic drift). Gillespie (1975) derived $E[(\Delta p)^2]$ for the case where there is variance due to both gametic sampling and intrinsic differences in offspring number between individuals of the same genotype. In the limiting cases where there is no sampling error, his equations reduce to (13), while in the limiting case of zero variance in offspring number they reduce to the Wright–Kimura equations.

Given the first and second moments in the appropriate limits $M(p)$, $V(p)$, the Kolmogorov backward equation (7) can be solved for the fixation probability of an allele,

$$U(p) = \frac{\int_0^p \exp\left[-\int 2\frac{M(x)}{V(x)}\,\mathrm{d}x\right]\mathrm{d}x}{\int_0^1 \exp\left[-\int 2\frac{M(x)}{V(x)}\,\mathrm{d}x\right]\mathrm{d}x} = \frac{\int_0^p \left((1-x)\sigma_1^2 + x\sigma_2^2\right)^{2\left(\frac{N(\mu_1-\mu_2)}{\sigma_1^2-\sigma_2^2}-1\right)}\mathrm{d}x}{\int_0^1 \left((1-x)\sigma_1^2 + x\sigma_2^2\right)^{2\left(\frac{N(\mu_1-\mu_2)}{\sigma_1^2-\sigma_2^2}-1\right)}\mathrm{d}x}. \tag{12}$$

where p in the integral limits is a random variable representing initial frequency (e.g., Kimura, 1964).

It is also the case that when the coefficients of $M(p)$ or $V(p)$ are large (i.e., μ_1–μ_2 and σ_1^2–σ_2^2 are not $\ll 1$), the assumptions behind the diffusion approximation to the discrete time change in allele frequency are violated, with $M(p)$ and $V(p)$ no longer valid as approximations to the first and second moments of change in allele frequency. For large valued differences in selection or variance terms, estimates of fixation probability from Eq. (12) are not always accurate. Nevertheless, Eq. (13) has been found (Shpak, 2005) to give reasonable approximations to fixation probabilities calculated from individual-based simulations for the same parameters even when both variance and directional selection contributions were fairly large (e.g., order $\sim$0.1 for selection differentials, order $\sim$1 for variances).

The most significant results in the single deme model follow directly from (1), namely, that for nearly equal arithmetic mean numbers of offspring per generation, the genotype that produces the smaller variance in offspring number will be favored by selection. Furthermore, in the case where the higher mean strategy also has a higher variance (i.e., $\mu_1 > \mu_2$ and $\sigma_1^2 > \sigma_2^2$), there will be a critical population size $\hat{n}$ at which the two alleles have equal fitness (and consequently, population sizes $n < \hat{n}$ will favor the low-mean, low-variance strategy, while population sizes $\hat{n} < n$ favor the high-mean, high-variance strategy). This critical population size (Gillespie, 1974) is calculated by setting both sides of Eq. (1) equal:

$$\hat{n} = \frac{\sigma_2^2 - \sigma_1^2}{\mu_1 - \mu_2}. \tag{13}$$

This relationship between population size and clutch size average/variance becomes less straightforward in the context of a metapopulation. If instead of a single, isolated deme, there are multiple demes exchanging migrants with one another, it seems apparent that the "effective" population size from the standpoint of selection on offspring variance will depend on deme structure and migration rate. For instance, in the absence of explicit spatial structure, if there are D demes of n individuals, one would expect that for very low migration rates the effective fitness of a strategy would be approximated by $\mu - (\sigma^2/n)$ (nearly independent demes), while for very high migration rates (approaching complete mixing) the effective fitness would be closer to $\mu - (\sigma^2/nD)$, with values in the denominator between n and Dn for intermediate migration rates. This suggests that for values of μ and σ^2 such that $n < \hat{n}$ and $\hat{n} < nD$, there should also be a critical value of migration rate at which a high-variance, low mean strategy starts to be disfavored by selection.

In Shpak (2005), this problem was investigated using individual-based simulations. It was found that for the metapopulation scenario described earlier, with D demes with n individuals, each of which exchanges a fraction m with the $(D-1)$ other demes, under certain cases there was indeed a critical value of $\hat{m}$ at which a strategy disfavored in a single deme of size $n < \hat{n}$ (i.e., higher mean, high variance) begins to be favored due to the effects of a greater "effective" population size under higher migration.

Significantly, it was found that migration rate only had this effect on selection for variance in a metapopulation if the migration occurred after reproduction but prior to selection. If reproduction and density regulation took place within

individual demes and was followed by migration of post-selection adults, there was no effect and the strategy that was favored in the individual demes in the absence of migration remained so even for high migration rates. The heuristic argument for why this is the case is that in the Birth → Regulation → Migration life cycle, the entire sampling process takes place separately in individual demes, so that migration only accomplishes a "mixing" (and subsequent homogenization) of allele frequencies. In contrast, for the second life cycle (Birth → Migration → Regulation), the pool of offspring that contribute to a given deme are sampled from the entire metapopulation, so that the effective sample variance depends on the contribution across demes as well as within demes in every generation.

In a sense, these differences in outcomes are due to the fact that the different life cycles lead to soft and hard selection, respectively (e.g., Wallace, 1968; Wade, 1985). Under the first (Birth → Regulation → Migration) life cycle, population regulation takes place within each deme prior to migration, so that every deme sends an equal number of migrants to each neighbor in every generation. Since population density regulation occurs within each deme before migration, this corresponds to soft selection in the sense of Wade (1985). In contrast, in the Birth → Migration → Regulation life cycle, migration takes place prior to regulation, which occurs at the global (metapopulation) level. Even though the total population and deme sizes are fixed from one generation to the next, there is no density regulation at the demic level prior to dispersal, which effectively imposes a hard selection regime (e.g., Gardner and West, 2006). This is because demes with higher average fitness contribute more migrants to the entire metapopulation, with density regulation occurring only after the potentially disproportionate representation from the different demes has occurred.

These observed differences can be confirmed by calculating the expected change in allele frequency under different selection regimes. Later, derivations of $E(\Delta p)$ are presented for spatially unstructured metapopulations for the two life cycles. The derivations essentially follow the methods used in deriving Eq. (6), apart from the introduction of factors of $(1-m)$ for "residents" and m for "migrants" for every deme and its $D-1$ neighbors.

2. Expected change in allele frequency in metapopulations

The life cycle of an organism in a metapopulation can be broken down into the processes of migration, reproduction (birth), and density-dependent regulation. In the models discussed later, the deme size (and therefore total metapopulation size) is held constant from one generation to the next, irrespective of the total number of incoming or outgoing migrants or the absolute number of progeny. Reproduction is again described by every genotype in the ith deme producing a random number of progeny equal to $\mu + \xi_i = \mu + \epsilon z$, so that the random variable describing offspring number variance scales as $\epsilon \ll 1$ and $\mu \sim 1$.

A number of additional approximations are made in deriving the expected change in allele frequency in these model systems. For both the Birth → Regulation → Migration (BRM) and Birth → Migration → Regulation (BMR) life cycles, the contribution of migrants in each deme is approximated as a proportion

$m\bar{x}$ (i.e., the average number of alleles of the first type in the metapopulation). This assumes that there is a sample pool of migrants from all demes which then migrate at random (so that migrants from a given deme can in principle return to the parental deme). While this is a reasonable assumption for certain instances of a BMR life cycle (for instance, broadcast spawning marine invertebrates), it is less realistic as an exact model for metapopulations under BRM.

In the BRM life cycle, it is adults (or at least some an age class that follows a juvenile stage at which selection is strongest) that migrate. It is unrealistic to assume a "broadcast and random return" form of migration in this case, so that in reality each deme is only likely to receive migrants from the $D-1$ other demes in the population (or from nearest neighbors, if there is spatial structure). However, treating migration in BRM as a pooled effect is a reasonable approximation when D is very large and/or allele frequencies do not greatly differ across demes. Furthermore, since in the BRM life cycle the contribution of offspring variance to $M(p)$ will be shown to be independent of migration rate, the results pertinent to the question of effective population size and the effective fitness of a reproductive strategy are qualitatively the same for the approximation and a more exact model regardless.

It should also be noted that an BRM life cycle is biologically equivalent to MBR and RMB, since the events occur cyclically and in sequence (with the same reasoning applicable to BMR versus MRB, RBM). In the actual calculations, the only difference between a BRM, MBR, and RMB life cycle is in the census point at which the frequency $p(t)$ is defined. So while the formal expressions may differ for various choices of census point, it should be obvious that the relations between terms (migration, selection coefficients, etc.) in the equations do not change, since the same processes are involved.

Finally, while the calculations for the rates of change of allele frequencies shown later are for haploid genotypes in the name of simplicity, the qualitative results on the effect of population subdivision, migration, and life cycle are essentially the same for diploids. This is because at least for populations in Hardy–Weinberg equilibrium, the equations for $E(\Delta p)$ and $E(\Delta p^2)$ in diploids take the same form as for haploids (as shown in the appendix of Gillespie, 1974) when one substitutes marginal values for the mean and variance of a given allele (i.e., weighted by their frequency of occurrence in homozygotes versus heterozygous states). Since the first and second moments of change in allele frequency for diploids in one deme are the same as equations for haploid selection in a single deme given the right substitution of parameters (i.e., marginal fitness in place of haploid fitness), it follows that the effects of migration and life cycle should also be equivalent in the diploid case because the migration coefficients appear as factors of the same terms in both scenarios.

2.1. Birth → density regulation → migration life cycle

Consider first the life cycle where the sequence of events is Birth → Regulation → Migration (BRM). In the absence of spatial structure, every deme sends a proportion m of its population to each of the remaining $D-1$ demes. The change in the

number of individuals of the first genotype x_i in the ith deme due to migration is approximated as

$$x_i(t+1) = (1-m)x_i + m\bar{x},$$

where $\bar{x} = (1/D)\sum_{j=1}^{D} x_j$.

In exact terms, this corresponds to the case where every deme contributes a fraction m to a common migrant pool that is then divided up between the D demes (so that some of the "migrants" return to the parental population). A more realistic model (particularly for a life cycle where adults migrate) is one where the migrants can only move to other demes. This mode of migration is represented by

$$x_i(t+1) = (1-m)x_i + \frac{m}{D-1}\sum_{j\neq i} x_j$$

(i.e., the ith deme sends fraction m to all of the $D-1$ other demes and receives migrants from the $D-1$ other demes in same ratios). Describing migration in terms of metapopulation mean absolute frequency $\bar{x}$ is an inexact but reasonable estimate if there is minimal difference in allele frequency between demes or if the number of demes is large (so that the estimated mean frequency is not strongly influenced by the removal of one deme frequency from the sample).

The frequency of the first allele in the ith deme after reproduction, density regulation, and migration is

$$\begin{aligned} p_i(t+1) = {} & m\frac{\bar{x}\mu_1 + \epsilon\sum_{k=1}^{\bar{x}} z_1[k]}{\left(\bar{x}\mu_1 + (n-\bar{x})\,\mu_2 + \epsilon\left(\sum_{k=1}^{\bar{x}} z_1[k] + \sum_{k=1}^{n-\bar{x}} z_2[k]\right)\right)} \\ & + (1-m)\frac{\left(x_i\mu_1 + \epsilon\sum_{k=1}^{x_i} z_1[k]\right)}{\left(x_1\mu_1 + (n-x_1)\mu_2 + \epsilon\left(\sum_{k=1}^{x_i} z_1[k] + \sum_{k=1}^{n-x_i} z_2[k]\right)\right)} \end{aligned} \tag{14}$$

Note that normalization occurs at the deme level, corresponding to the "soft selection" mode of density regulation.

As for the single deme model $p_i(t+1)$ can be rewritten as a Taylor expansion in terms of ϵ up to second order. After collecting terms with the same coefficients and applying the relationships $\mathrm{cov}(z_1[i], z_1[j]) = 0$, $\mathrm{cov}(z_1[i], z_2[j])$, and $(\sum_{k=1}^{x_i} z_i[k])^2 = x_i\sigma^2/\epsilon^2$, together with the substitutions $x_1 = np$, $x_2 = n(1-p)$, $\bar{x} = n\bar{p}$, a simplified expression for $p_i(t+1)$ and thus $E(\Delta p_i) = p_i(t+1) - p_i(t)$ is obtained:

$$\begin{aligned} E(\Delta p_i) = {} & E[p_i(t+1)|p_i(t)] - p_i(t) \\ = {} & \left(\frac{m\bar{p}\mu_1}{\bar{p}\mu_1 + (1-\bar{p})\mu_2} + \frac{(1-m)p_i\mu_1}{p_i\mu_1 + (1-p_i)\mu_2} - p_i\right) \\ & - \left(\frac{m(1-\bar{p})\bar{p}\mu_1\left(\mu_2\sigma_1^2 - \mu_1\sigma_2^2\right)}{n(\bar{p}\mu_1 + (1-\bar{p})\mu_2)^3} + \frac{(1-m)p_i(1-p_i)\left(\mu_2\sigma_1^2 - \mu_1\sigma_2^2\right)}{n(p_i\mu_1 + (1-p_i)\mu_2)^3}\right) \end{aligned} \tag{15}$$

where p_i and $\bar{p}$ are used as shorthand for $p_i(t)$ and $\bar{p}(t)$.

Comparing (15) with the single-deme Equation (6), it can be seen that the terms in the first set of parentheses represent the "deterministic" change in allele frequency due to migration and the differences in allele frequency across demes, and, as in the single deme case, directional selection due to mean differences in fitness between the two alleles. The terms in the second set of parentheses represent selection on variance, which is largely independent of migration in this life cycle except for the change due to differences in relative frequency. When $p_i = \bar{p}$ (allele frequencies equal in all demes), the aforementioned expression reduces to an equation that is identical to (6)–(7), one which is independent of the migration rate m because every $\bar{p} = p_i$ term with a factor of m has a complement p_i factor of $(1 - m)$,

$$M(p_i) = \frac{p_i(1-p_i)(\mu_1 - \mu_2)}{p_i\mu_1 + (1-p_i)\mu_2} - \frac{p_i(1-p_i)\left(\mu_2\sigma_1^2 - \mu_1\sigma_2^2\right)}{n(p_i\mu_1 + (1-p_i)\mu_2)^3}. \tag{16}$$

As a result, it is predicted that the selection dynamics defined by $M(p_i)$ in a metapopulation of D demes of size n under the BRM life cycle are identical to that within a single deme of size n, regardless of migration rate. Migration only contributes to $M(p_i)$ through frequency differences between demes and not because of the sample variance across demes. This can be seen from the fact that in (16) the variance terms scale inversely with n (as for a single deme) independent of migration rate m and the total number of demes D, so that the selection coefficient ("effective fitness") of a strategy remains $\mu - (\sigma^2/n)$. Even when there are differences in allele frequencies across demes, the selection on variance is still of the order σ^2/n, with the only difference between the case of no migration versus migration being that of an allelic (additive) variance factor of $p_i(1-p_i)(\sigma^2/n)$ versus the migration weighted term $(m\bar{p}(1-\bar{p}) + (1-m)p_i(1-p_i))(\sigma^2/n)$.

That migration only contributes to change in allele frequency when $p_i \neq \bar{p}$ can also be seen in substituting the limiting cases of $m = 0$ (no migration) and $m = 1$ (complete mixing) into (16), which give, respectively,

$$M(p_i) = \frac{p_i(1-p_1)(\mu_1 - \mu_2)}{p_i\mu_1 + (1-p_i)\mu_2} - \frac{p_i(1-p_i)\left(\mu_2\sigma_1^2 - \mu_1\sigma_2^2\right)}{n(p_i\mu_1 + (1-p_i)\mu_2)^3} \tag{17a}$$

$$M(p_i) = \frac{\bar{p}(1-\bar{p})(\mu_1 - \mu_2)}{\bar{p}\mu_1 + (1-\bar{p})\mu_2} - \frac{(1-\bar{p})\bar{p}\mu_1\left(\mu_2\sigma_1^2 - \mu_1\sigma_2^2\right)}{n(\bar{p}\mu_1 + (1-\bar{p})\mu_2)^3} \tag{17b}$$

which differ only in the relative roles of p_i, $\bar{p}$ and not in the value of the denominator terms associated with σ^2. The second moment (diffusion) term $V(p_i)$ is also independent of migration rate when allele frequencies are equal (see the first part of the Appendix for derivations). Since both the selection and diffusion terms are equivalent to (7) and (12), the fixation probability of an allele under the BRM life cycle will be approximately same as for a single deme, irrespective of how large the metapopulation is or how high the migration rate (as was found to be the case in the simulations of Shpak (2005)). These considerations lead to the interesting

result that selection on offspring variance can be strong even in populations that appear to be very large and nearly panmictic, provided that selection is localized within small individual demes.

2.2. *Birth → migration → density regulation life cycle*

As was noted in Section 1, if reproduction in every deme is followed by migration prior to culling (as in the BMR life history), the evolutionary dynamics are quite different from the BRM cycle because the sample variance in offspring number will be reduced by contributions from every deme. It is expected that this will be reflected in the form of $M(p_i)$ (and ultimately, the fixation probabilities of an allele) for the BMR life cycle, since the normalization of allele counts (representing soft selection) is done over absolute frequencies across as well as within demes. Here migration is also assumed to occur via the "mixing" used in the previous derivations, such that all demes contribute a certain fraction m of their offspring to a migrant pool which then distributes at random across all demes (allowing for return migration and thus the characterization of the migrant pool as an average $\bar{x}$).

Following the same notation as in the previous derivations, the frequency of the reference allele at the end of the BMR life cycle is

$$p_i(t+1) = \frac{f_2}{g_2} \tag{18}$$

where

$$f_2 = m\left(\bar{x}\mu_1 + \frac{\epsilon\sum_{k=1}^{D\bar{x}} z_1[k]}{D}\right) + (1-m)\left(x_i\mu_1 + \epsilon\sum_{k=1}^{x_1} z_1[k]\right)$$

$$g_2 = m\left(\bar{x}\mu_1 + \frac{\epsilon\sum_{k=1}^{D\bar{x}} z_1[k]}{D}\right) + (1-m)\left(x_i\mu_1 + \epsilon\sum_{k=1}^{D\bar{x}} z_1[k]\right)$$
$$+ m\left((n-\bar{x}) + \mu_2\frac{\epsilon\sum_{k=1}^{D(n-\bar{x})} z_2[k]}{D}\right)$$
$$+ (1-m)\left((n-x_i)\mu_2 + \epsilon\sum_{k=1}^{n-x_i} z_2[k]\right)$$

with the difference that division by the total number of individuals (density regulation) is done at the level of the entire metapopulation after migration. It will be shown that as a consequence, in this life cycle, the number of individuals with the reference allele in the entire metapopulation $D\bar{x}$ contributes to the variance component in every deme.

To calculate the change of p_i in terms of offspring variance for the BMR life cycle, (18) is similarly expanded as a power series up to second order in terms of ϵ. Applying the variance and covariance relations defined in the previous sections

and substituting $np = x_1, n(1-p) = x_2, n\bar{p} = \bar{x}$ as before, $E(\Delta p_i)$ can be rewritten in terms of allele frequencies as:

$$\begin{aligned} E(\Delta p_i) &= E[p_i(t+1)|p_i(t)] - p_i(t) \\ &= \frac{m\bar{p}\mu_1 + (1-m)p_i\mu_1}{m\bar{p}\mu_1 + (1-m)p_i\mu_1 + m(1-\bar{p})\mu_2 + (1-m)(1-p_i)\mu_2} \\ &\quad -\frac{f_3}{g_3} + \frac{f_4}{g_4} \end{aligned} \tag{19}$$

where

$$f_3 = \frac{m^2\bar{p}\sigma_1^2}{D} - (1-m)^2 p_i\sigma_1^2$$

$$g_3 = n\left(m\bar{p}\mu_1 + (1-m)p_i\mu_1 + m(1-\bar{p})\mu_2 + (1-m)(1-p_i)\mu_2\right)^2$$

$$f_4 = (m\bar{p}\mu_1 + (1-m)p_i\mu_1)\left(\frac{m^2 n\bar{p}\sigma_1^2}{D} + (1-m)^2 np_i\sigma_1^2 + \frac{m^2(1-\bar{p})\sigma_2^2}{D} + (1-m)^2(1-p_i)\sigma_2^2\right)$$

$$g_4 = n\left(m\bar{p}\mu_1 + (1-m)p_i\mu_1 + m(1-\bar{p})\mu_2 + (1-m)(1-p_i)\mu_2\right)^3$$

In the case where allele frequencies are equal ($p_i = \bar{p}$) across all demes, the mean change in frequency simplifies to

$$\begin{aligned} E(\Delta p_i) &= \frac{p_i(1-p_i)(\mu_1-\mu_2)}{p_i\mu_1 + (1-p_i)\mu_2} \\ &\quad -\frac{\left((1-m)^2 + \frac{m^2}{D}\right)}{n}\frac{p_i(1-p_i)\left(\mu_2\sigma_1^2 - \mu_1\sigma_2^2\right)}{\left(p_i\mu_1 + (1-p_i)\mu_2\right)^3} \end{aligned} \tag{20}$$

When directional selection is weak and the values of μ_i are near unity, the first moment can be approximated as

$$M(p_i) = p_i(1-p_i)\left(\mu_1 - \mu_2 - \frac{\left((1-m)^2 + \frac{m^2}{D}\right)}{n}\left(\sigma_1^2 - \sigma_2^2\right)\right) \tag{21}$$

It is apparent that the aforementioned expressions are not equivalent to $M(p)$ for a single deme (i.e., Eqs. (5)–(6)). The difference is that rather than scaling as $1/n$, the variance contributions scale as $\{[(1-m)^2] + [m^2/D]\}/n$, which depends on both the migration rate and the number of demes. Even when allele frequencies are equal in all demes, migration has the indirect effect on the selection coefficient of reducing variance by sampling from the pool of offspring throughout the metapopulation.

In the limiting cases of $m = 0$ and $m = 1$, the respective values of $M(p_i)$ are

$$M(p_i) = \frac{p_i(1-p_1)(\mu_1-\mu_2)}{p_i\mu_1 + (1-p_i)\mu_2} - \frac{p_i(1-p_i)\left(\mu_2\sigma_1^2 - \mu_1\sigma_2^2\right)}{n\left(p_i\mu_1 + (1-p_i)\,\mu_2\right)^3} \tag{22a}$$

$$M(p_i) = \frac{\bar{p}(1-\bar{p})(\mu_1-\mu_2)}{\bar{p}\mu_1 + (1-\bar{p})\mu_2} - \frac{(1-\bar{p})\bar{p}\mu_1\left(\mu_2\sigma_1^2 - \mu_1\sigma_2^2\right)}{nD\left(\bar{p}\mu_1 + (1-\bar{p})\mu_2\right)^3}. \tag{22b}$$

For a BMR life cycle then, migration rate determines the effective population size "seen" by selection acting on offspring variance. When $m = 0$, each deme is independent of the others, so that $M(p)$ is the same as (6) for a single deme (i.e., n in the denominator of the variance terms). When $m = 1$ (corresponding to a case of complete mixing), the denominator is nD, the total metapopulation size. Hence, for the BMR life cycle, even when $p_i = \bar{p}$, it can be seen that (22a) and (22b) are obviously not equivalent: one has a variance contribution to fitness that scales with deme size n, the other with metapopulation size nD.

On heuristic grounds, it was argued in Shpak (2005) that the effective population size (denominator of the variance terms in $M(p)$) would scale as $(1-m)n + mDn$, which had the desired explicit dependence on migration rate and limiting properties. The form of Eq. (20) suggests that the relationship is actually more complicated for intermediate values of m. The variance contributions to $M(p_i)$ under the BMR life cycle actually scale as functions of the squares of migration rates, with n in the denominator for"residents" and nD for the "migrants,"

$$\frac{(1-m)^2 p_i\sigma_1^2}{n} + \frac{m^2\bar{p}\sigma_1^2}{nD}. \tag{23}$$

If this is compared to $(p\sigma^2)/n$ for a single deme, then it can be argued that migration in the BMR model induces an effective population size n_e from the standpoint of selection for variance, so that the strength of selection against offspring number variance in a single deme with size n_e would be the same as that in the metapopulation with D demes of size n and migration rate m, i.e.,

$$\frac{p\sigma^2}{n_e} = \frac{(1-m)^2 p_i\sigma_1^2}{n} + \frac{m^2\bar{p}\sigma_1^2}{nD}. \tag{24}$$

Ignoring the effects of frequency differences between demes by setting $\bar{p} = p_i = p$,

$$n_e \approx \frac{nD}{D(1-m)^2 + m^2}, \tag{25}$$

which increases towards nD as m approaches unity, and decreases towards n as m approaches 0.

It is emphasized that "effective population size" is always defined not as an absolute quantity but as a parameter specific to some population-level process. In much of the population genetics literature (e.g., Ewens, 2004), effective population size

is defined in reference to genetic drift and coalescence processes in a population. Here, we use "effective population size" as the size of a single panmictic population that would give the same strength of selection on offspring variance as for a deme in a subdivided metapopulation with migration rate m, as opposed to the probability or expected time of coalescence.

Together, Eqs. (24) and (25) yield the prediction that if $\mu_1 < \mu_2$, $\sigma_1^2 < \sigma_2^2$, and n is small enough so that $\mu_2 - (\sigma_2^2/n) < \mu_1 - (\sigma_1^2/n)$, for sufficiently many demes and a high enough migration rate there will be a critical value m such that n_e in (25) satisfies Eq. (13), so that the strategy with higher effective fitness at zero or low migration is disfavored at high migration rates because the selection against variance is less pronounced in a higher effective population size). We explore this theme further in Section 4.

As far as calculating actual fixation probabilities under BMR is concerned, the derivations of $V(p_i)$ due to offspring number variance in a metapopulation are presented in the Section A.2 of the Appendix. As with $M(p_i)$, the expression is analogous to that of a single deme, but with n_e in place of n as the denominator of the variance terms. Together, these terms provide enough information to estimate fixation probabilities when the differences in mean and variance terms of competing genotypes are sufficiently small.

3. Fixation probabilities: Caveats and comparisons

In the special case where allele frequencies are equal in all demes, the fixation probability of an allele corresponding to a given strategy can be approximated by (14) because the ratios of $M(p_i)/V(p_i)$ will be integrable. If weak selection is assumed, then (22b) and (32) can be used to approximate the drift and the diffusion term ratio, and the equation can be integrated analytically. The results of these integrals are shown in Fig. 1, which calculates the fixation probability of the high variance, high mean strategy ($\mu_1 = 1.1$, $\sigma_1^2 = 10.89$), competing against the low-mean, low-variance strategy (with $\mu_2 = 1.0, \sigma_2^2 = 0.9$) in a metapopulation of $D = 10$ demes of size $n = 50$ for a range of migration rates. The initial frequency is 0.5, so a fixation probability of one half corresponds to effective neutrality.

Plotted alongside the PDE solution are the fixation frequencies sampled over the same range of migration rates, calculated from the individual-based simulations with the same selection and variance parameters as in the analytical predictions. The model BMR life cycle was implemented in the simulation as follows: each high variance individual produces a single clutch of 11 offspring, which all survive or die with probability 0.1. All individuals of the competing strategy produce 10 clutches of a single offspring, each of which survives (individually) with the same probability of 0.1. After reproduction, migration is implemented such that the number of migrants sent from each deme is determined by a Poisson distribution with mean $\lambda = m/(Dn')$, where m is the proportion of migrants and n' the combined number of offspring of both genotypes in a deme. Each migrant's genotype is chosen as a Bernoulli probability equal to its frequency in the offspring pool. The last stage of the life cycle is population regulation, where exactly $n = 50$ individuals from the total offspring pool are chosen after migration (essentially as

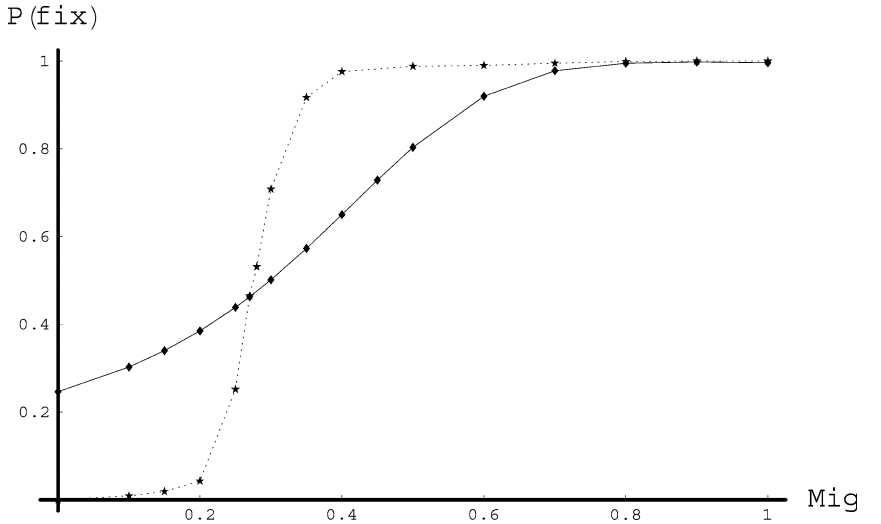

Fig. 1 Fixation probabilities of the high-mean, high-variance strategy as a function of migration rate under the BMR life cycle. Here deme size is held constant at $n = 50$ and there are $D = 10$ demes. The "reference" allele whose fixation probability is calculated corresponds to mean offspring number $\mu_1 = 1.1$ and variance $\sigma_1^2 = 10.89$. The competing strategy has a mean value $\mu_2 = 1.0$ and a variance $\sigma_2^2 = 0.9$. The initial frequencies of both strategies are set to $p = 0.5$ in every deme, so that fixation probabilities greater than unity represent higher than neutral effective fitness, $M(p) > 0$. The *solid line* represents calculated fixation probabilities from Eq. (16), using the weak selection forms of $M(p)$ and $V(p)$, as in Eqs. (21) and (32). The *dashed line* and *points* represent the fixation probabilities calculated for the same parameters in individual-based simulations for a variety of migration rates.

a binomial probability with $n = 50$ and $p =$ frequency of low variance individuals after migration). The simulations are discussed in further detail in Shpak (2005), and the *C* code used is available from the senior author on request.

The correspondence between the simulations and calculated fixation probabilities is imprecise for a number of reasons. First, the estimated $V(p)$ does not incorporate the genetic drift due to gamete sampling that occurs in all finite populations. The model assumes that all demes have the same allele frequency, but this will not be true in general because of the stochastic nature of the process. Interdeme variance in allele frequencies will cause deviations from the predicted mean change in allele frequency, but is not expected to cause differences in the sign of the mean change in allele frequency. This is because most parameters cause the mean change in allele frequency to have the same sign independent of allele frequency (Proulx, 2000). Thus, even with interdeme variance in allele frequency the global (i.e., metapopulation) mean change in allele frequency will have the same sign as the mean change when all demes have the mean allele frequency. Although the mean change in the global allele frequency $M(p)$ does predict which allele will have a greater than neutral fixation probability, the diffusion terms $V(p)$ are required to determine the value of the fixation probability. In many cases, however, the relative magnitudes of mean change in allele frequency are sufficient to make

qualitative predictions about the evolutionary dynamics (Proulx and Day, 2001; Nowak et al., 2004; Wild and Taylor, 2004).

Even for the case of equal allele frequencies in all demes, in order to calculate fixation probabilities from solutions to the Kolmogorov backward equation, the conditions for the diffusion approximation must be met, namely weak selection small variance contributions. This requires that coefficients of first and second moments of allele frequency change are of order $\epsilon \ll 1$, while higher order moments are negligible. When these conditions are violated, solutions to the diffusion equation will not necessarily give a close estimate of fixation probabilities (e.g., Kimura, 1964). Furthermore, the only source of stochasticity in the equations leading to solutions of the form (12) is due to variance in offspring number, as opposed to the additional genetic drift caused by the sampling of n gametes from the total pool. This is what accounts for the higher probability of loss of the high-variance allele when the migration rate (and therefore n_e) is small relative to the predictions of the Kolmogorov backward equations, as can be seen in Fig. 1.

Nevertheless, by comparing analytical results with the simulations, it can be seen that they give a reasonable first-order approximation to the observed fixation probability (particularly for higher migration rates and effective population sizes, when selection on offspring number variance is weak). In addition, even when the diffusion approximation breaks down, if $M(p)$ has a constant sign with respect to p, then the process is a martingale and the fixation probability is bounded by the neutral fixation rate. Thus, the diffusion approximation does correctly predict the critical migration value at which the high-mean, high-variance strategy starts to have a higher fixation probability than its competitor ($m \approx 0.29$, corresponding to an effective population size of just under 100 at which the two strategies are effectively "neutral"). Irrespective of the diffusion term, if $M(p) > 0$, it will have a higher than neutral probability of fixation, and it can be seen that the fixation probability is greater than 50% (corresponding to neutrality, where fixation probability is equal to initial frequency) at precisely the value of m for which the sign of $M(p)$ changes.

4. Discussion

There are a several potentially important consequences of selection on variance of offspring production in metapopulations. That the effective fitness of a strategy where there is variance in clutch size depends on population size was established by Gillespie (1974), so it stands to reason that population processes that lead to differences between census size and "effective" population size with respect to selection on offspring number variance (such as density fluctuations, differences in frequency between the sexes, e.g., Proulx, 2000) can lead to differences in the relative advantage of a strategy with a given mean and variance values.

Migration and population structure are known to cause a discrepancy between census and effective population size from the standpoint of expected coalescence times and genetic drift (e.g., Whitlock and Barton, 1997; Wakeley and Aliacar, 2001; Cherry and Wakeley, 2003), which might suggest that other sources of variance would be influenced by subdivision. Defined in terms of coalescence and drift

processes, subdivision and reduced migration rate lead to higher effective population size (i.e., increase expected time to coalescence or time to fixation or loss by genetic drift), while effective size as defined by intensity of selection on offspring variance decreases with subdivision and reduced migration rates. This difference is due to the fact that the two definitions of effective size involve two different processes—the first the coalescence of a random branching process, which depends on the absolute number of individuals in each deme and their migration rate, irrespective of interaction or life cycle, while in these models the effective population size is defined specifically in terms of the probability that during the processes of selection and density regulation a pair of competing individuals will have come from the same deme. This tendency (referred to as "population viscosity" by Queller, 1992) is measured by the parameter $(1-m)^2$, the probability that two individuals from the same deme will ultimately interact. It is from this probability of interaction that the significance of the sequence of events in a life cycle and whether density regulation occurs at the demic or metapopulation level arises.

In models of kin selection, for example, this measure of viscosity plays a role in that it reflects the probability that individuals will interact with kin versus unrelated individuals (Queller, 1992; Gardner and West, 2006). In the context of kin selection models, Gardner and West have shown that the strength of selection is measured as the sum of a "soft" (intrademic) selection term and a "hard" (interdemic) selection term weighted by $(1-m)^2$, which is similar to the partitioning of the selection differential on offspring number variance in Eq. (21) into within and between deme components. When selection is strictly intrademic, of course (as in the BRM life cycle), then no such partitioning is possible, and the selection coefficients are the same as those in a single deme.

What these results suggest is that the influence of offspring variance on the performance of a given strategy cannot necessarily be ignored even in large populations. Traditionally, the significance of Eqs. (6)–(7) and of evolutionary bet-hedging was considered by many to be of only academic interest, on the grounds that most biological populations were large enough for sample variance to play a minor role. What the results in this paper show is that even in a large metapopulation with extensive mixing, if the life cycle is of the BRM type and the population consists of many small demes, the effect of offspring variance is essentially the same as it would be for a single small deme. Consequently, the impact of offspring variance must be assessed on a case-by-case basis, depending on life cycle and population structure as well as on the fitness and variance parameters.

In other words, if there is a tradeoff between producing a high mean number of offspring and reducing variance (as in the semelparous vs. iteroparous regimes mentioned in Section 1), the outcome of selection will depend not only on the census number in the population but also on life cycle and migration rate. Using the semelparity and iteroparity examples given in the previous section, an organism that reproduces once and produces an average of one clutch $k_1 = 1$ of $\omega_1 = 11$ offspring that survive (as a whole clutch) with a probability of $\pi = 0.1$ and fail with a probability of $1-\pi = 0.9$ has a mean fitness $\mu_1 = k_1\omega_1\pi = 1.1$ and variance $\sigma_1^2 = k_1\omega_1^2\pi(1-\pi) = 10.89$. If it competes against an iteroparous strategy that produces a $k_2 = 10$ clutches of $\omega_2 = 1$ single offspring (with the probability of surviving $\pi = 0.1$), the parameters are $\mu_2 = 1.0$ and $\sigma_2^2 = 0.9$. Eq. (15)

predicts that for a population size $n < 99$, the iteroparous strategy will be favored in spite of having a lower arithmetic mean. If the selection takes place in the context of a metapopulation with $D = 10$ demes with $n = 50$ individuals per deme, the iteroparous strategy will always be favored in BRM life cycle regardless of migration rate.

In a BMR life cycle, there will be a critical value of migration rate at which the higher mean semelparous strategy starts to be favored. This critical value can be approximated as the value of m that gives $n_e = 99.9$ in Eq. (27), with $n = 50$ and $D = 10$. Solving the quadratic equation for m, the root less than unity is $m = 0.289$. This corresponds reasonably well with the individual-based simulations shown in Fig. 1, where the probability of fixing the iteroparous strategy (with initial frequency $p = 0.5$) was near 50% for a migration rate of 0.27. Equation (25) gives a much better estimate of the critical migration rate at which the strategies become effectively neutral than the linear approximation proposed in Shpak (2005).

These results predict broad trends in the evolution of life histories and reproductive strategies in various organisms. Since for the same mean value of offspring number an iteroparous strategy produces a lower variance in surviving progeny than a semelparous strategy, it is likely that semelparity should be less common in organisms with small population sizes, or in highly structured populations where selection on juveniles is local with respect to their place of birth (i.e., with a life cycle of the BRM type). For a large population or a metapopulation of organisms with a BMR life cycle, the selective penalty for high offspring variance is lower given sufficiently high migration and a large number of demes. Therefore, one might expect semelparity to be more common in organisms with a BMR life cycle, particularly if the semelparous strategy can produce a higher mean number of offspring overall.

At an empirical level, this means that one would expect semelparity to be more common among organisms where the majority of incoming migrants to any deme are juveniles so that the most important selection takes place after migration. This is the case with broadcast spawning marine invertebrates, many of which have large, widely distributed metapopulations that exchange migrants via planktonic eggs and larvae, and with plants that disperse seeds over long distances. Semelparity would probably be more prevalent in such organisms than in (for example) birds or large mammals, where most of the migration between demes is by adults that have already been subjected to an entire lifetime of selection.

The tradeoffs involved in determining reproductive strategy are often more complicated than a balance between mean and variance (for example, a balance between survivorship or fecundity at different stages of the life cycle, as discussed by Charnov and Schaffer, 1973), but the effects of offspring variance on the fitness of a genotype are potentially strong enough for this to be an important factor in many animal and plant populations. It would certainly be of interest to find semelparous taxa in nature that have closely related iteroparous congeners, so that the predictions made above about the life cycle and population structure for these competing strategies could be directly tested.

Finally, it is worth mentioning that the BRM and BMR life cycles are in a sense extreme idealizations, in which it is assumed that population regulation occurs largely during one stage of the life cycle and migration at another. In fact

(particularly in organisms where there is overlap between generations), all stages of the life cycle might be subject to culling and to migration alike, though to different degrees. A complete theory of the evolution of offspring variance in metapopulations should take explicit age structure into consideration and allow for different degrees of migration and culling at different stages, with BRM and BMR as two (probably useful and fairly common) limiting cases in nature. Some work on selection on variance terms in age-structured populations has been done (e.g., Demetrius and Gundlach, 1999), and it would be of great interest to extend these matrix models to situations involving multiple demes and migration.

5. Appendix: Second moments of expected change in allele frequency in metapopulations

The derivation of $E(\Delta p^2)$ and the diffusion coefficient $V(p)$ in a single deme is shown in Eqs. (8)–(12). The same approximations and assumptions about migration used to derive $E[p(t+1)]$ in a metapopulation under different life cycles is used to calculate the second moments:

5.1. Birth → regulation → migration life cycle

$$p_i^2(t+1) = \left(\frac{m\left(\bar{x}\mu_1 + \epsilon \sum_{k=1}^{\bar{x}} z_1[k]\right)}{\epsilon\left(\sum_{k=1}^{\bar{x}} z_1[k] + \sum_{k=1}^{n-\bar{x}} z_2[k]\right) + \bar{x}\mu_1 + (n-\bar{x})\,\mu_2} + \frac{(1-m)\left(x_i\mu_1 + \epsilon \sum_{k=1}^{x_i} z_1[k]\right)}{\epsilon\left(\sum_{k=1}^{x_i} z_1[k] + \sum_{k=1}^{n-x_i} z_2[k]\right) + x_i\mu_1 + (n-x_i)\,\mu_2} \right)^2 \tag{26}$$

writing a Taylor expansion in terms of ϵ, applying the variance and covariance relations on the sums of z (reducing the expression to functions of the mean and variance values), and substituting np_i, $n\bar{p}$ for x_i and $\bar{x}$, one can obtain the expected value of $p_i^2(t+1)$. The expression for $E[\Delta p_i^2]$ is simply the aforementioned quantity after appropriate substitutions plus (p_i^2-2p_i $E[p_i(t+1)|p_i(t)]$), where $E[p_i(t+1)|p_i(t)]$ consists of the terms in Eq. (15), i.e.,

$$\begin{aligned} E\left[\Delta p_i^2\right] = & \left(\frac{m\bar{p}\mu_1}{\bar{p}\mu_1 + (1-\bar{p})\mu_2} + \frac{(1-m)p_i\mu_1}{p_i\mu_1 + (1-p_i)\mu_2} \right)^2 \\ & + \frac{m^2n^3\bar{p}^3\mu_1^2\sigma_1^2}{\left(n\bar{p}\mu_1 + n(1-\bar{p})\mu_2\right)^4} - \frac{2m^2n^2\bar{p}^2\mu_1\sigma_1^2}{\left(n\bar{p}\mu_1 + n(1-\bar{p})\mu_2\right)^3} \\ & + \frac{m^2n\bar{p}\sigma_1^2}{\left(n\bar{p}\mu_1 + n\mu_2 - n\bar{p}\mu_2\right)^2} + \frac{(1-m)^2p_i^3\mu_1^2\sigma_1^2}{n\left(p_i\mu_1 + (1-p_i)\mu_2\right)^4} \\ & - \frac{2(1-m)^2p_i^2\mu_1\sigma_1^2}{n\left(p_i\mu_1 + (1-p_i)\mu_2\right)^3} + \frac{(1-m)^2p_i\sigma_1^2}{n\left(p_i\mu_1 + (1-p_i)\mu_2\right)^2} \end{aligned}$$

$$+\frac{(1-m)^2\bar{p}^2(1-\bar{p})\mu_1^2\sigma_2^2}{n(p_i\mu_1+(1-p_i)\mu_2)^4}+\frac{(1-m)^2p_i^2(1-p_i)\mu_1^2\sigma_2^2}{n(p_i\mu_1+(1-p_i)\mu_2)^4}$$

$$+2\left(\frac{m\bar{p}\mu_1}{\bar{p}\mu_1+(1-\bar{p})\mu_2}+\frac{(1-m)p_i\mu_1}{p_i\mu_1+(1-p_i)\mu_2}\right)$$

$$+\left(m\left(\frac{\bar{p}\mu_1\left(\bar{p}\sigma_1^2+(1-\bar{p})\sigma_2^2\right)}{n(\bar{p}\mu_1+(1-\bar{p})\mu_2)^3}-\frac{\bar{p}\sigma_1^2}{n(\bar{p}\mu_1+(1-\bar{p})\mu_2)^2}\right)\right.$$

$$\left.+(1-m)\left(\frac{p_i\mu_1\left(p_i\sigma_1^2+(1-p_i)\sigma_2^2\right)}{n(p_i\mu_1+(1-p_i)\mu_2)^3}-\frac{p_i\sigma_1^2}{(p_i\mu_1+(1-p_i)\mu_2)^2}\right)\right)$$

$$-2p_i\left(\left(\frac{m\bar{p}\mu_1}{\bar{p}\mu_1+(1-\bar{p})\mu_2}+\frac{(1-m)p_i\mu_1}{p_i\mu_1+(1-p_i)\mu_2}\right)\right.$$

$$\left.+\left(\frac{m(1-\bar{p})\bar{p}\mu_1\left(\mu_2\sigma_1^2-\mu_1\sigma_2^2\right)}{n(\bar{p}\mu_1+(1-\bar{p})\mu_2)^3}+\frac{(1-m)p_i(1-p_i)\left(\mu_2\sigma_1^2-\mu_1\sigma_2^2\right)}{n(p_i\mu_1+(1-p_i)\mu_2)^3}\right)\right)+p_i^2. \tag{27}$$

As was the case for the first moments, there is no contribution of offspring variance to $E[\Delta p_i^2]$ in the BRM life cycle except for the weighted difference of allele frequencies between the metapopulation mean and the census deme. In the absence of allele frequency differences, $E[\Delta p_i^2]$) reduces to the following expression,

$$E\left[\Delta p_i^2\right]=\left(\frac{p_i\mu_1}{p_i\mu_1+(1-p_i)\mu_2}\right)^2+\frac{(1-m)^2p_i^3\mu_1^2\sigma_1^2}{n(p_i\mu_1+(1-p_i)\mu_2)^4}$$

$$-\frac{2(1-m)^2p_i^2\mu_1\sigma_1^2}{n(p_i\mu_1+(1-p_i)\mu_2)^3}+\frac{(1-m)^3p_i\sigma_1^2}{n(p_i\mu_1+(1-p_i)\mu_2)^2}$$

$$+\frac{(1-m)^2p_i^2(1-p_i)\mu_1^2\sigma_2^2}{n(p_i\mu_1+(1-p_i)\mu_2)^4}+2\left(\frac{p_i\mu_1}{p_i\mu_1+(1-p_i)\mu_2}\right)$$

$$\times\left(\frac{p_i\mu_1\left(p_i\sigma_1^2+(1-p_i)\sigma_2^2\right)}{n(p_i\mu_1+(1-p_i)\mu_2)^3}-\frac{p_i\sigma_1^2}{n(p_i\mu_1+(1-p_i)\mu_2)^2}\right)$$

$$-2p_i\left(\frac{p_i\mu_1}{p_i\mu_1+(1-p_i)\mu_2}+\frac{(1-p_i)p_i\left(\mu_2\sigma_1^2-\mu_1\sigma_2^2\right)}{n(p_i\mu_1+(1-p_i)\mu_2)^3}\right)+p_i^2. \tag{28}$$

Furthermore, in the limits of complete mixing ($m=1$) and no migration ($m=0$), (28) converges to an equation identical to $E[\Delta p_i^2]$ in a single deme (12) apart from the fact that in the former case $p=\bar{p}$, while in the latter $p=p_i$ (as was the case for the first moment).

For weak selection and low offspring variance (i.e., μ near unity and $\sigma^2 \ll 1$), $V(p)$ can be approximated by the same form as (11) for a single deme.

5.2. Birth → migration → regulation life cycle

The second moment of $p_i(t+1)$ is calculated from the expectation of:

$$E[p_i^2(t+1)|p_i(t)] = \frac{f_5}{g_5} \tag{29}$$

where

$$f_5 = \left(m\left(\frac{\epsilon \sum_{k=1}^{D\bar{x}} z_1[k]}{D} + \bar{x}\mu_1 \right) + (1-m)\left(\epsilon \sum_{k=1}^{x_i} z_1[k] + x_i\mu_1 \right) \right)^2$$

$$g_5 = \left(m\left(\frac{\epsilon \sum_{k=1}^{D\bar{x}} z_1[k]}{D} + \bar{x}\mu_1 \right) + (1-m)\left(\epsilon \sum_{k=1}^{x_i} z_1[k] + x_i\mu_1 \right) \right.$$
$$\left. + m\left(\frac{\epsilon \sum_{k=1}^{D(n-\bar{x})} z_2[k]}{D} + (n-\bar{x})\,\mu_2 \right) + (1-m)\left(\epsilon \sum_{k=1}^{n-x_i} z_2[k] + (n-x_i)\mu_2 \right) \right)^2.$$

Up to the quadratic term in a series expansion about $\epsilon = 0$, the previous expression (after substituting mean and variance values) is:

$$E\big[p_i^2(t+1)|p_i(t)\big] = \frac{f_6}{g_6} - \frac{f_7}{g_7} + \frac{f_8}{g_8} + \frac{f_9}{g_9} \tag{30}$$

where

$$f_6 = (m\bar{p}\mu_1 + (1-m)p_i\mu_1)^2$$
$$g_6 = (m\bar{p}\mu_1 + (1-m)p_i\mu_1 + m(1-\bar{p})\,\mu_2 + (1-m)\,(1-p_i)\,\mu_2)^2$$
$$f_7 = 4\,(m\bar{p}\mu_1 + (1-m)p_i\mu_1)\left(\frac{m^2\bar{p}\sigma_1^2}{D} + (1-m)^2 p_i\sigma_1^2 \right)$$
$$g_7 = n\,(m\bar{p}\mu_1 + (1-m)p_i\mu_1 + m(1-\bar{p})\mu_2 + (1-m)(1-p_i)\mu_2)^3$$
$$f_8 = \frac{m^2\bar{p}\sigma_1^2}{D} + (1-m)^2 p_i\sigma_1^2$$
$$g_8 = n\,(m\bar{p}\mu_1 + (1-m)p_i\mu_1 + m(1-\bar{p})\,\mu_2 + (1-m)\,(1-p_i)\,\mu_2)^2$$
$$f_9 = 3(m\bar{p}\mu_1 + (1-m)p_i\mu_1)^2\left(\frac{m^2\bar{p}\sigma_1^2}{D} \right.$$
$$\left. + (1-m)^2 p_i\sigma_1^2 + \frac{m^2(1-\bar{p})\sigma_2^2}{D} + D(1-m)^2(1-p_i)\sigma_2^2 \right)$$
$$g_9 = n(m\bar{p}\mu_1 + (1-m)p_i\mu_1 + m(1-\bar{p})\mu_2 + (1-m)(1-p_i)\mu_2)^4.$$

Setting all $p_i = \bar{p}$, using the relation $E[\Delta p_i^2] = (E[p_i^2(t+1)|p_i(t)] - 2p_i E[p_i(t+1)|p_i(t)])$, one derives

$$
\begin{aligned}
E\left[\Delta p_i^2\right] = & \left(\frac{p_i\mu_1}{p_i\mu_1 + (1-p_i)\,\mu_2}\right)^2 \\
& - \frac{\left(4\left((1-m)p_i\mu_1 + mp_i\mu_1\right)\left(\left((1-m)^2 + \frac{m^2}{D}\right)p_i\sigma_1^2\right)\right)}{n\left(p_i\mu_1 + (1-p_i)\mu_2\right)^3} \\
& + \frac{\left(\left((1-m)^2 + \frac{m^2}{D}\right)p_i\sigma_1^2\right)}{n\left(p_i\mu_1 + (1-p_i)\mu_2\right)^2} \\
& + \frac{\left(3p_i\mu_1^2\left(\left((1-m)^2 + \frac{m^2}{D}\right)p_i\sigma_1^2 + \left((1-m)^2 + \frac{m^2}{D}\right)(1-p_i)\,\sigma_2^2\right)\right)}{n\left(p_i\mu_1 + (1-p_i)\,\mu_2\right)^4} \\
& - 2p_i\left(\frac{p_i\mu_1}{p_i\mu_1 + (1-p_i)\,\mu_2} - \frac{\left((1-m)^2 + \frac{m^2}{D}\right)}{n}\frac{p_i(1-p_i)\left(\mu_2\sigma_1^2 - \mu_1\sigma_2^2\right)}{\left(p_i\mu_1 + (1-p_i)\,\mu_2\right)^3}\right) \\
& + p_i^2
\end{aligned}
\tag{31}
$$

when the selection coefficients are near unity, one can approximate $E[\Delta p_i^2]$ as the diffusion coefficient, $V(p_i)$,

$$
V(p_i) = \frac{p_i(1-p_i)}{n}\left((1-m)^2 + \frac{m^2}{D}\right)\left((1-p_i)\sigma_1^2 + p_i\sigma_2^2\right). \tag{32}
$$

Since there are closed form solutions for $V(p_i)$ in a metapopulation, one can in principle calculate probabilities of fixation and loss from the Kolmogorov backward Eq. (7) using the integration in (12) for the growth rate and variance parameters. The diffusion term (32) allows (12) to be evaluated exactly when allele frequencies are equal across all demes. For more general models, solutions can probably be derived only with numerical integration.

Acknowledgements

Max Shpak wishes to thank Sergey Gavrilets for his advice and support (via NSF grant DEB-0111613 and NIH grant GM56693) during the writing of an early draft of this paper while at the University of Tennessee (Knoxville), and for startup funding support at the University of Texas (El Paso) during the subsequent revision of the manuscript and derivations. We also thank David Waxman, Lloyd Demetrius, Warren Ewens, and three anonymous reviewers for their assistance, comments, and criticism.

References

Charnov, E.L., Schaffer, W.M., 1973. Life history consequences of natural selection: Cole's result revisited. Am. Nat. 107, 791–793.
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. Harper and Row, New York.
Demetrius, L., Gundlach, M., 1999. Evolutionary dynamics in random environments. In: Crauel, H., Gundlach, M. (Eds.), Stochastic Dynamics. Springer-Verlag, New York, pp. 371–394.
Demetrius, L., Gundlach, M., 2000. Game theory and evolution: Finite size and absolute fitness measures. Math. Biosci. 168, 9–38.
Ewens, W.J., 2004. Mathematical Population Genetics. Springer-Verlag, New York.
Gardner, A., West, S.A., 2006. Demography, altruism, and the benefits of budding. J. Evol. Biol. 19:1707–1716.
Gillespie, J., 1974. Natural selection for within-generation variance in offspring number. I. Genetics 76, 601–606.
Gillespie, J., 1975. Natural selection for within-generation variance in offspring number. II. Genetics 81, 403–413.
Gillespie, J., 1977. Natural selection for variance in offspring numbers: A new evolutionary principle. Am. Nat. 111, 1010–1014.
Haldane, J.B.S., Jayakar, S.D., 1963. Polymorphism due to selection of varying directions. J. Genet. 58, 237–242.
Karlin, S., Taylor, H.M., 1975. A First Course in Stochastic Proccesses. Academic Press, New York, NY.
Kimura, M., 1964. The diffusion model in population genetics. J. Appl. Probab. 1, 177–232.
Nowak, M.A., Sasaki, A., Taylor, C., 2004. Emergence of cooperation and evolutionary stability in finite populations. Nature 428, 646–650.
Proulx, S.R., 2000. The ESS under spatial variation with applications to sex allocation. Theor. Populat. Biol. 58, 33–47.
Proulx, S.R., Day, T., 2001. What can invasion analyses tell us about evolution under stochasticity in finite populations? Select. Mol. Genes Memes 2, 2–15.
Queller, D.C., 1992. Does population viscosity promote kin selection? Trends Ecol. Evol. 7, 322–324.
Seger, J., Brockmann, H.J., 1987. What is bet-hedging? In: Harvey, P.H., Partridge, L. (Eds.), Oxford Surveys in Evolutionary Biology, vol. 4. Oxford University Press, Oxford, pp. 182–211.
Shpak, M., 2005. Evolution of variance in offspring number: The effect of population size and migration. Theor. Biosci. 124, 65–85.
Stearns, S.C., 2000. Daniel Bernoulli: Evolution and economics under risk. J. Biosci. 25, 221–228.
Wade, M.J., 1985. Soft selection, hard selection, kin selection, and group selection. Am. Nat. 125, 61–73.
Wakeley, J., Aliacar, N., 2001. Gene genealogies in a metapopulation. Genetics 159, 893–905.
Wallace, B., 1968. Topics in population genetics. W.W. Norton, New York, NY.
Whitlock, M.C., Barton, N.H., 1997. The effective size of subdivided population. Genetics 146, 427–441.
Wild, G., Taylor, P.D., 2004. Fitness and evolutionary stability in game theoretic models of finite populations. Proc. R. Soc. Lond. Ser. B 2345–2349.