

# Genetic load in subdivided populations: interactions between the migration rate, the size and the number of subpopulations

K Theodorou<sup>1</sup> and D Couvet<sup>2</sup>

<sup>1</sup>Biodiversity Conservation Laboratory, Department of Environmental Studies, University of the Aegean, University Hill, Mytilene 81100, Greece; <sup>2</sup>CRBPO, Muséum National d'Histoire Naturelle, 55, rue Buffon, Paris 75005, France

We assess the relative importance of migration rate, size and number of subpopulations on the genetic load of subdivided populations. Using diffusion approximations, we show that in most cases subdivision has detrimental effects on fitness. Moreover, our results suggest that fitness increases with subpopulation size, so that for the same total population size, genetic load is relatively lower when there are a small number of large subpopulations. Using elasticity analysis, we show that the size of the subpopulations appears to be the parameter that most strongly determines genetic load. Interconnecting subpopulations via migration would also be

of importance for population fitness when subpopulations are small and gene flow is low. Interestingly, the number of subpopulations has minor influence on genetic load except for the case of both very slightly deleterious mutations and small subpopulations. Elasticities decrease as the magnitude of deleterious effects increases. In other words, population structure does not matter for very deleterious alleles, but strongly affects fitness for slightly deleterious alleles. *Heredity* (2006) **96**, 69–78. doi:10.1038/sj.hdy.6800762; published online 23 November 2005

**Keywords:** genetic load; deleterious mutation; subdivided population; conservation

## Introduction

Population subdivision affects both the evolution and the persistence of populations. For instance, subdivision has been shown to have an important effect on the probability of fixation of beneficial and deleterious alleles (Whitlock, 2003), the evolution of mating systems (Ronfort and Couvet, 1995) or the probability of population extinction (Saccheri *et al.*, 1998; Higgins and Lynch, 2001). One of the reasons of this influence is that subdivision changes the way in which the different evolutionary processes (selection, genetic drift, migration) act on allele frequency, compared to a continuous population. As a result, the population's genetic load (ie the decline in fitness due to accumulation of deleterious alleles) can be strongly determined by population structure (Glémin *et al.*, 2003).

Subdivision can vary in several ways, including the size and the number of the subpopulations and the rate of migration between subpopulations. Changes in these parameters can significantly modify the balance between drift and selection within subpopulations and, thus, genetic load: (i) for slightly deleterious and partially recessive alleles, subpopulation size determines both the response to selection and the strength of genetic drift; larger subpopulations should be associated with lower

frequencies of deleterious alleles, (ii) migration between subpopulations restores genetic variability within subpopulations and, therefore, enhances selection, and (iii) the number of subpopulations influences population genetic variance; increasing the number of subpopulations should result in a higher genetic differentiation between subpopulations and, thus, to a higher potential for fitness to be restored by migration. It is also interesting to note that subdivision can have variable effects according to the characteristics of deleterious mutations. For instance, genetic variance within subpopulations could decrease for nearly additive alleles but it can increase for highly recessive alleles (Whitlock, 2002).

Although several theoretical studies examine genetic load in subdivided populations (Couvet, 2002; Whitlock, 2002; Glémin *et al.*, 2003; Roze and Rousset, 2004 to cite only some of the most recent), the issue of the relative importance of the parameters that determine population structure on genetic load has not been explicitly addressed.

The goal of this study is, therefore, to investigate how genetic load in a population depends on the particular form that subdivision takes. Using diffusion approximations, we explore a finite island model (a population consisting of a finite number of identical subpopulations that exchange migrants at random). We examine the relative impact of the parameters that determine population structure in our model (the migration rate, the size and the number of subpopulations) on genetic load. This is carried out by means of elasticity analysis, that is, the proportional change in genetic load resulting from a proportional

Correspondence: K Theodorou, Biodiversity Conservation Laboratory, Department of Environmental Studies, University of the Aegean, University Hill, Mytilene 81100, Greece. E-mail: ktheo@aegean.gr  
Received 10 September 2004; accepted 16 August 2005; published online 23 November 2005

change in each of the parameters. This approach is widely used in population dynamics studies in order to assess the impact of demographic factors, such as mortality or reproductive rate of different age or stage classes, on the growth rate of the population (eg Caswell, 1989; Johnson and Braun, 1999). However, to our knowledge, it is the first time that this approach has been applied to population genetics-related issues.

This analysis allows us to gain insight on how changes in the population structure affect the process of mutation accumulation. Furthermore, our results have implications for the conservation of fragmented populations since they allow us to assess the efficiency of different conservation practices according to population structure (see Soulé and Orians, 2002). Our analysis helps to identify whether the enhancement of migration between subpopulations, the reinforcement of subpopulation size, or the establishment of new subpopulations is the most efficient way to decrease the genetic load of a given population, and, hence, to increase its viability. Our analysis allows us also to refine previous rules of thumbs focusing only on the connectivity between subpopulations, such as that of 'one migrant per population per generation' (OMPG, Mace and Lande, 1991; Mills and Allendorf, 1996).

## Methods

### Basic assumptions

We considered a population of total size  $N_{\text{tot}}$  that is subdivided into  $n$  identical subpopulations of size  $N$  each ( $N_{\text{tot}} = nN$ ); all the subpopulations exchange migrants at random at rate  $m$ . In the literature, this population structure is referred to as the finite island model.

We developed a single locus, two alleles model with  $A$  being the wild-type allele and  $a$  an unconditionally deleterious mutant. Mutation occurs from  $A$  to  $a$  with a rate  $u$  per locus and per generation. Note that in diploid organisms,  $u = U/2L$ , where  $U$  is the mutation rate of the entire genome and  $L$  is the number of loci in the genome. We assumed that reverse mutations occur at a negligible rate. The relative fitness of the genotypes  $AA$ ,  $Aa$ ,  $aa$  is  $1$ ,  $1-hs$ ,  $1-s$  respectively, where  $h$  measures the degree of dominance of the deleterious allele and  $s$  the magnitude of the deleterious effect in homozygotes.

### Diffusion approximations

To study the pattern of genetic load according to population structure, we used diffusion approximations. This method was first developed by Ohta and Kimura (1969, 1971). Although the aim of their paper was to study linkage disequilibrium between two loci, Ohta and Kimura gave a general method for deriving the moments of two or more variables describing gene frequency distributions.

Following the analysis of Ohta and Kimura (1969, 1971), let  $\varphi(x_1, \dots, x_n)$  be the density function of the stationary distribution of the  $n$  independent variables  $x_1, \dots, x_n$ .

Consider a function  $f(x_1, \dots, x_n)$  and let  $E\{f(x_1, \dots, x_n)\} = \int_0^1 f(x_1, \dots, x_n) \varphi(x_1, \dots, x_n) dx_1 \dots dx_n$  be its expectation with respect to the distribution  $\varphi(x_1, \dots, x_n)$ .

Ohta and Kimura (1969) showed that for the function  $f(x_1, \dots, x_n)$  the following diffusion equation holds:

$$\frac{dE\{f\}}{dt} = E \left\{ \frac{1}{2} \sum_{i=1}^n V_{\delta x_i} \frac{\partial^2 f}{\partial x_i^2} + \sum_{\substack{i,j=1 \\ i \neq j}}^n C_{\delta x_i \delta x_j} \frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_{i=1}^n M_{\delta x_i} \frac{\partial f}{\partial x_i} \right\} \quad (1)$$

At equilibrium,  $dE\{f\}/dt = 0$  in the above equation.

In our case,  $x_i$  is the frequency of the deleterious allele in subpopulation  $i$ ;  $M_{\delta x_i}$  corresponds to the mean change in allele frequency within subpopulation  $i$  per generation;  $V_{\delta x_i}$  is the variance in allele frequency within subpopulation  $i$  due to random sampling of gametes, and  $C_{\delta x_i \delta x_j}$  is the covariance of the change in allele frequency due to random sampling in subpopulations  $i$  and  $j$ .

In the case of the finite island model, the terms  $M_{\delta x_i}$ ,  $V_{\delta x_i}$ ,  $C_{\delta x_i \delta x_j}$  of equation (1) can be expressed as follows:

$$M_{\delta x_i} \approx M_{\delta x_i}^{\text{mut}} + M_{\delta x_i}^{\text{sel}} + M_{\delta x_i}^{\text{migr}}$$

That is, we made the assumption that the change in allele frequency,  $x_i$ , is due to the additive action of mutation ( $M_{\delta x_i}^{\text{mut}}$ ), selection ( $M_{\delta x_i}^{\text{sel}}$ ) and migration ( $M_{\delta x_i}^{\text{migr}}$ ). In other words, we neglected interaction terms between these pressures. This assumption will hold as long as changes in allele frequency between two generations are small.

Hence, the mean change in allele frequency is approximately given by the equation:

$$M_{\delta x_i} \approx u(1-x_i) - sx_i(1-x_i)[h + (1-2h)x_i] - mx_i + \frac{m}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^n x_j \quad (2)$$

Further, the variance,  $V_{\delta x_i}$ , is given by the equation:

$$V_{\delta x_i} = \frac{x_i(1-x_i)}{2N} \quad (3)$$

And the covariance of the change in frequencies in two different subpopulations due to random sampling is  $C_{\delta x_i \delta x_j} = 0$ .

Note that Glémin *et al* (2003) used the same method to examine the pattern of genetic load for the finite island model. However, they assumed that the change in allele frequency due to selection is linear ( $M_{\delta x_i}^{\text{sel}} \approx -hsx_i$ ). As the authors pointed out, this approximation holds only when local drift is weak relative to selection, that is, for mutations that are strongly deleterious and not highly recessive and large subpopulations ( $Nhs > 5$ ; Glémin *et al*, 2003).

In order, therefore, to examine the case of slightly deleterious alleles and/or small subpopulations, we dropped the assumption of linear change in gene frequency. However, this leads to an infinite system of equations. To close the system, we assumed that the larger moments may be neglected. As it will be shown later (see 'Validity of diffusion approximations' in the Results), one has to include up to the third moment of gene frequencies in order to accurately describe the process of the accumulation of slightly deleterious mutations.

We need therefore to calculate the expected value of  $x_i$ ,  $x_i^2$ ,  $x_i^3$ ,  $x_i x_j$ ,  $x_i x_j^2$ ,  $x_i x_j x_k$  with  $i, j, k \in [1, n]$ . Owing to the symmetry of the island model, the expected values of the

moments of the allele frequencies will be identical for all demes. Hence, if we discard  $O(x^4)$  terms, we have six unknown parameters:  $E\{x\}$ ,  $E\{x^2\}$ ,  $E\{x^3\}$ ,  $E\{xx'\}$ ,  $E\{x^2x'\}$ ,  $E\{xx'x'\}$ . In order to obtain the expected values of these parameters, we introduce different functions  $f$  into Equation (1) until a  $6 \times 6$  system is built.

(1)  $f = x_i$ . Equation (1) implies:

$$\begin{aligned} \frac{dE\{x\}}{dt} &= E\{M_{\delta x_i}\} \Rightarrow \frac{dE\{x\}}{dt} \\ &= u - (hs + u + m)E\{x_i\} - (1 - 3h)sE\{x_i^2\} \\ &\quad + (1 - 2h)sE\{x_i^3\} + \frac{m}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^n x_j \end{aligned}$$

After simplifications due to the symmetry of the model (ie  $m/(n-1)E\{\sum_{j=1, j \neq i}^n x_j\} = mE\{x\}$ ) we take:

$$\begin{aligned} \frac{dE\{x\}}{dt} &= u - (hs + u)E\{x\} - (1 - 3h)sE\{x^2\} \\ &\quad + (1 - 2h)sE\{x^3\} \end{aligned} \quad (1a)$$

Similarly, (2)  $f = x_i^2$ :

$$\begin{aligned} \frac{dE\{x_i^2\}}{dt} &= E\{2x_iM_{\delta x_i} + V_{\delta x_i}\} \\ &\Rightarrow \frac{dE\{x^2\}}{dt} = \left(2u + \frac{1}{2N}\right)E\{x\} \\ &\quad - 2\left(hs + u + m + \frac{1}{4N}\right)E\{x^2\} \\ &\quad - 2(1 - 3h)sE\{x^3\} + 2mE\{xx'\} \end{aligned} \quad (1b)$$

(3)  $f = x_i x_j$ :

$$\begin{aligned} \frac{dE\{x_i x_j\}}{dt} &= E\{x_i M_{\delta x_j} + x_j M_{\delta x_i}\} \\ &\Rightarrow \frac{dE\{xx'\}}{dt} = 2uE\{x\} - 2\left(hs + u + \frac{m}{n-1}\right)E\{xx'\} \\ &\quad - 2(1 - 3h)sE\{xx'^2\} + \frac{2m}{n-1}E\{x^2\} \end{aligned} \quad (1c)$$

(4)  $f = x_i^3$ :

$$\begin{aligned} \frac{dE\{x_i^3\}}{dt} &= E\{3x_i V_{\delta x_i} + 3x_i^2 M_{\delta x_i}\} \Rightarrow \frac{dE\{x^3\}}{dt} \\ &= 3\left(u + \frac{1}{2N}\right)E\{x^2\} - 3\left(hs + u + m + \frac{1}{2N}\right)E\{x^3\} \\ &\quad + 3mE\{x^2x'\} \end{aligned} \quad (1d)$$

(5)  $f = x_i^2 x_j$ :

$$\begin{aligned} \frac{dE\{x_i^2 x_j\}}{dt} &= E\{x_j V_{\delta x_i} + 2x_i x_j M_{\delta x_i} + x_i^2 M_{\delta x_j}\} \Rightarrow \frac{dE\{x^2x'\}}{dt} \\ &= \left(2u + \frac{1}{2N}\right)E\{xx'\} + uE\{x^2\} + \frac{m}{n-1}E\{x^3\} \\ &\quad - \left(3hs + 3u + \frac{(2n-3)m}{n-1} + \frac{1}{2N}\right)E\{x^2x'\} \\ &\quad + \frac{2(n-2)m}{n-1}E\{xx'x'\} \end{aligned} \quad (1e)$$

$f = x_i x_j x_k$ :

$$\begin{aligned} \frac{dE\{x_i x_j x_k\}}{dt} &= E\{x_j x_k M_{\delta x_i} + x_i x_k M_{\delta x_j} + x_i x_j M_{\delta x_k}\} \\ &\Rightarrow \frac{dE\{xx'x'\}}{dt} = uE\{xx'\} + \frac{2m}{n-1}E\{x^2x'\} \\ &\quad - (hs + u + \frac{2m}{n-1})E\{xx'x'\} \end{aligned} \quad (6)$$

In order to set the initial conditions of the above system of differential equations, we assumed that subpopulations are founded by individuals descended from an infinite population, that is

$$\begin{aligned} E\{x\}_{t=0} &= q_{eq}, \quad E\{x^2\}_{t=0} = q_{eq}^2, \quad E\{xx'\}_{t=0} = q_{eq}^2, \\ E\{x^3\}_{t=0} &= q_{eq}^3, \quad E\{x^2x'\}_{t=0} = q_{eq}^3, \quad E\{xx'x'\}_{t=0} = q_{eq}^3 \end{aligned}$$

where  $q_{eq}$  is given by the equation  $s(1-2h)q_{eq}^2 + sh(1+u)q_{eq} - u = 0$  (Crow and Kimura, 1970). This is an assumption frequently made in previous studies (eg Wang *et al*, 1999; Theodorou and Couvet, 2004). In any case, we found that changing the initial conditions had no significant effect on the patterns observed.

A numerical solution of the above system was obtained using the program Mathematica. However, an analytical solution of the moments of allele frequency can be obtained for the equilibrium state (where all the left-hand parts of the equations equal to zero). The explicit expressions, however, are formidable and not given.

#### Fitness over many loci

In order to extend this model to the entire genome, we considered that loci act multiplicatively and independently. The expected mean fitness over loci within a subpopulation can then be calculated as (Whitlock *et al*, 2000):

$$\bar{w} = (1 - hs)^{n_{het}} (1 - s)^{n_{hom}} \quad (4)$$

where  $n_{het} = LE\{2x(1-x)\} = 2L[E\{x\} - E\{x^2\}]$  is the expected number of heterozygous loci at equilibrium within a subpopulation and  $n_{hom} = LE\{x^2\}$  is the expected number of homozygous loci ( $L$  is the number of loci in the diploid genome).

#### Multilocus simulations

In order to check the validity of our results, we performed extensive multilocus simulations. This permits us to (i) decide the order of gene frequency moments required to accurately describe the accumulation of slightly deleterious alleles, and (ii) assess the influence of linkage disequilibrium, which is thought to be important in small and subdivided populations, on the accuracy of our results.

In each generation, the order of operations in the life cycle of each population was mutation, random mating, selection, and finally migration. The number of new mutations arising in an individual at any generation was assumed to follow a uniform distribution with mean  $U = 2Lu$  (we assume that backward mutations are negligible), and the loci where mutations occurred were chosen at random. The number of loci was set to  $L = 2560$ . The production of an offspring within each subpopulation was simulated as follows. Two parents, chosen at random, produce an offspring. The selective

value of the offspring produced was calculated from the fitness function of equation (4). The selective value of the offspring was compared to a random number from the uniform distribution between 0 and 1. The offspring was retained if its selective value was higher than this number. When an offspring was rejected, two parents were again chosen at random, which meant that an individual could produce offspring with several mates. This process was repeated until  $N$  offspring produced (equal to the size of the subpopulations).

At the end of each generation, zygote migration occurred at a constant rate  $m$ . To decide if an individual migrated, we compared  $m$  to a random number drawn from the uniform distribution between 0 and 1. If  $m$  was greater than the random number then the individual migrated otherwise it remained in the native subpopulation.

For each parameter set, we ran 50 simulations, and we computed the average fitness between generations 4000 and 5000 (a sufficient number of generations to reach equilibrium).

### Elasticity analysis

In order to investigate the relative importance of the parameters that determine population structure ( $n$ ,  $m$ ,  $N$ ) on the genetic load of the population, we calculated the elasticity of the mean fitness with respect to each of these parameters.

In demographic analysis, elasticity is defined as the proportional change in the population growth rate resulting from a proportional change in the coefficients of population growth (such as age or stage-dependent survival and fecundity). As Caswell pointed out (1989, pp. 118–119), together with the calculation of the asymptotic population growth rate, an important part of a demographic analysis is the investigation of how the growth rate would vary in response to changes in the vital rates, that is, identifying the vital rates whose change will most affect population dynamics. In order to examine the elasticity of genetic load with regard to population structure, we applied this approach and simply substituting the growth rate with mean fitness and the demographic factors with the parameters of population structure.

Consequently, elasticity analysis permits us to find the proportional change in genetic load resulting from a proportional change in the values of the parameters under study. We used elasticity, rather than sensitivity, because the parameters under study are measured and managed on different scales (Caswell, 1989); a given change in the migration rate,  $m$ , and the subpopulation size,  $N$ , for instance, represents a much greater proportional change in the former than in the latter parameter. The elasticities are calculated according to Caswell (1989, p 132):

$$e(m) = \frac{m}{\bar{w}} \frac{\partial \bar{w}}{\partial m} \quad (5a)$$

$$e(n) = \frac{n}{\bar{w}} \frac{\partial \bar{w}}{\partial n} \quad (5b)$$

$$e(N) = \frac{N}{\bar{w}} \frac{\partial \bar{w}}{\partial N} \quad (5c)$$

where  $e(m)$ ,  $e(n)$ ,  $e(N)$  are the elasticities of mean fitness with respect to the migration rate, the number and the size of the subpopulations, respectively.

At disequilibrium, the system of equation 1(a)–(e) was solved numerically (and not analytically) and therefore the above equations cannot be used directly to calculate elasticities. An approximation was therefore used according to which

$$e(x) = \frac{x}{\bar{w}} \frac{\Delta \bar{w}}{\Delta x} \quad (5d)$$

with  $x = m$ ,  $n$  or  $N$

### The values of mutation parameters

The issue of the value of the mutation parameters in most organisms is yet unresolved. Most available results come from experiments conducted on *Drosophila* species. Several of these experiments suggest that the majority of mutations are slightly deleterious ( $s = 0.01$ – $0.03$ ) with mutation rates per diploid genome of  $U = 1$  and a mean dominance coefficient of  $h = 0.2$ – $0.4$  (see review in Lynch *et al*, 1999).

However, the validity of these estimates has been recently questioned; new experimental studies on *Drosophila melanogaster* and other organisms suggest that mutation rates are much lower,  $U = 0.01$ – $0.02$ , and the average effect in fitness higher than previously thought,  $s = 0.1$ – $0.2$  (Bataillon, 2000; Caballero *et al*, 2002). Other studies report intermediate mutation rates of  $U \approx 0.1$  (Fry, 2001; Fry and Heinsohn, 2002) while Rodríguez-Ramilo *et al* (2004) found evidence for both sets of mutation parameters (slightly deleterious – high mutation rate and mildly deleterious – low mutation rate).

Most of our analysis considers the mutation set  $U = 1$ ,  $s = 0.02$  and  $h = 0.3$ , since the majority of arguments is in favor of such values (Lynch *et al*, 1999). However, the dependence of our results on the choice of the mutations parameters was also checked.

## Results

### Validity of diffusion approximations

In order to test the validity of diffusion approximations, we proceeded as follows:

First, we compared our results obtained by diffusion approximations with those generated by the transition matrix approach, since the latter method has been used and tested repeatedly (see for instance Ewens, 1979; Lynch *et al*, 1995; Schoen *et al*, 1998; Couvet, 2002; Theodorou and Couvet, 2002). We checked the robustness of our approximations with regard to mutational ( $h$ ,  $s$ ) and demographic parameters ( $N$ ,  $m$ ) for an infinite array of subpopulations. The reader can refer to Couvet (2002) for a complete presentation of the transition matrix approach in the case of the infinite island model.

Our results show that diffusion approximations are in very good agreement with the transition matrix approach (Table 1). The agreement is remarkable when slightly deleterious and not highly recessive mutations are considered (eg  $s = 0.02$ ,  $h = 0.4$ ), which is most convenient since our study focuses exactly on this type of mutation. However, for low migration ( $m \leq 0.01$ ) and for highly recessive and moderately deleterious alleles

( $h = 0.1, s = 0.2$ ), diffusion approximations underestimate fitness.

The inclusion of the third moments of allele frequency in the diffusion equations improves significantly the precision of our approximations in the case of slightly deleterious and highly recessive alleles ( $hs$  low). In all the other cases, similar results are yielded irrespective of whether one includes the second or the third moments of allele frequency.

Second, we performed stochastic multilocus simulations in order to check the influence of the number of subpopulations ( $n$ ) and the effects of linkage disequilibrium on the precision of diffusion approximations. The results of these simulations show that our approximations describe very accurately ( $<5\%$  deviation from the results of the simulations) the accumulation of slightly deleterious alleles as long as the total population is  $N_{\text{tot}} \geq 200$ , (eg  $n \geq 10$  when  $N = 20$  or  $n \geq 4$  when  $N = 50$ ) (Table 2).

In the case of small populations ( $N_{\text{tot}} = 200\text{--}600$  individuals), the inclusion of the third moments of allele frequency improves considerably the agreement between diffusion approximations and simulations for the majority of subpopulation sizes and migration rates checked. This is particularly so for small subpopulations ( $N = 20$ ) and/or low migration rates ( $m \leq 0.025$ ).

Finally, in the case of disequilibrium the agreement between diffusion approximations and multilocus simulations is excellent for all population structures and values of the mutation parameters used in this study ( $<5\%$  deviation for the first 100 generations after subdivision, results not shown).

In conclusion, our approximations seem to yield consistent results for a large range of population

structures and properties of deleterious alleles ( $h, s$ ). Moreover, the extension of the approximations of Glémin *et al* (2003) in order to include higher moments of allele frequency proved to be necessary when one wishes to describe the accumulation of slightly deleterious alleles in small populations.

### The effects of subdivision on fitness

We compared fitness under different population structures with that of a single undivided population of the same total size, both in the short and in the long term (Figure 1a and b).

For deleterious mutations of intermediate levels of dominance ( $h = 0.3$ ), subdivision has negative effects on fitness. Furthermore, a population consisting of a few large subpopulations is associated with higher fitness relative to a population of many small subpopulations. The relative advantage of having large subpopulations decreases for strongly connected subpopulations and highly recessive mutations. Moreover, for low levels of dominance of deleterious alleles ( $h = 0.1$ ), a subdivided population show higher fitness than a single one as long as migration is high ( $m \simeq 0.05$ ).

Note that the above conclusions remain the same whether we examine the first 50 generations or equilibrium (compare Figure 1a and b). Furthermore, for large and strongly interconnected subpopulations, the agreement between the short and the long term is even quantitative suggesting that, in these cases, equilibrium is attained in a small number of generations.

It must be, however, underlined that for highly recessive alleles ( $h = 0.1$ ) fitness is significantly lower after 50 generations of subdivision than at equilibrium.

**Table 1** The fitness of the subdivided population, relative to the fitness of an effectively infinite population, calculated by means of the transition matrix approach (matrix) and diffusion approximations when the first two or three orders of gene frequency are included in the calculations (referred to  $O(x^3)$  and  $O(x^4)$ , respectively)

M	N = 20 s = 0.02									
	h = 0.1					h = 0.4				
	Matrix	$O(x^3)$	$O(x^3)$ -error	$O(x^4)$	$O(x^4)$ -error	Matrix	$O(x^3)$	$O(x^3)$ -error	$O(x^4)$	$O(x^4)$ -error
0.010	0.553	1.367	147.20	0.557	0.72	0.424	0.676	59.43	0.487	14.86
0.025	0.904	1.316	45.58	0.863	-4.54	0.706	0.785	11.19	0.704	-0.28
0.050	1.047	1.256	19.96	1.019	-2.67	0.843	0.863	2.37	0.830	-1.54
	s = 0.2									
0.010	1.122	1.27	13.19	0.974	-13.19	0.899	0.893	-0.67	0.867	-3.56
0.025	1.138	1.239	8.88	1.005	-11.69	0.92	0.906	-1.52	0.888	-3.48
0.050	1.082	1.2	5.17	1.025	-5.27	0.94	0.923	-1.81	0.911	-3.09
	N = 50 s = 0.02									
0.010	0.952	1.308	37.39	0.903	-5.15	0.762	0.831	9.06	0.772	1.31
0.025	1.075	1.230	14.42	1.050	-2.33	0.884	0.899	1.70	0.879	-0.57
0.050	1.091	1.162	6.51	1.083	-0.73	0.938	0.939	0.11	0.932	-0.64
	s = 0.2									
0.010	1.140	1.194	4.74	1.024	-4.91	0.965	0.954	-1.14	0.949	-1.66
0.025	1.128	1.157	2.57	1.050	-4.26	0.971	0.960	-1.13	0.957	-1.44
0.050	1.108	1.119	0.99	1.061	-4.24	0.978	0.968	-1.02	0.965	-1.33

Error designates the % deviation of the results of diffusion approximations from those of the transition matrix approach. The genome mutation rate is set to  $U = 1$ ;  $m$  denotes the rate of migration between subpopulations.

**Table 2** The fitness of the subdivided population, relative to the fitness of an effectively infinite population, calculated by means of multilocus stochastic simulations (Simul) and diffusion approximations when the first two or three orders of gene frequency are included in the calculations (referred to  $O(x^3)$  and  $O(x^4)$ , respectively)

$N_{tot}$	$m = 0.01$									
	$N = 20$					$N = 50$				
	Simul	$O(x^3)$	$O(x^3)$ -error	$O(x^4)$	$O(x^4)$ -error	Simul	$O(x^3)$	$O(x^3)$ -error	$O(x^4)$	$O(x^4)$ -error
100	0.213	0.872	309.39	0.490	130.05	0.142	0.946	566.20	0.740	421.13
200	0.492	0.877	78.25	0.520	5.69	0.772	0.959	24.22	0.798	3.37
400	0.524	0.88	67.94	0.545	4.01	0.836	0.965	15.43	0.823	-1.56
600	0.544	0.881	61.95	0.562	3.31	0.848	0.967	14.03	0.831	-2.00
	$m = 0.025$									
100	0.343	0.891	159.77	0.617	79.88	0.522	0.951	82.18	0.771	47.70
200	0.674	0.903	33.98	0.682	1.19	0.853	0.973	14.07	0.862	1.06
400	0.748	0.91	21.66	0.745	-0.40	0.926	0.986	6.48	0.906	-2.16
600	0.764	0.913	19.50	0.757	-0.92	0.941	0.99	5.21	0.919	-2.34
	$m = 0.05$									
100	0.430	0.902	109.77	0.72	36.62	0.538	0.954	77.32	0.783	45.54
200	0.778	0.921	18.38	0.804	3.34	0.882	0.98	11.11	0.891	1.02
400	0.851	0.932	9.52	0.854	0.35	0.961	0.997	3.61	0.944	-2.38
600	0.874	0.937	7.21	0.870	-0.46	0.977	1.003	2.66	0.966	-1.74

Error designates the % deviation of the results of diffusion approximations from those of the simulations. The selection coefficient is set to  $s = 0.02$ , the dominance coefficient to  $h = 0.3$  and the genome mutation rate to  $U = 1$ .  $N_{tot}$  denotes the total population size ( $N_{tot} = nN$ ) and  $m$  denotes the rate of migration between subpopulations.

This is because, with subdivision, homozygosity increases and the previously masked deleterious alleles are now exposed to selection. During purging, fitness is low and it rebounds only after a considerable number of generations; for  $m = 0.025$ , it can take a 100 generations for fitness to recover to levels close to equilibrium (results not shown). A transient genetic load due to highly recessive alleles was also observed by Kirkpatrick and Jarne (2000) for bottlenecked populations.

#### Relative influence of migration rate, the size and the number of subpopulations on genetic load

Figures 2a and b show that there is a clear hierarchy in the influence of the parameters of population structure on genetic load. For a given population structure, a proportional change in the subpopulation size,  $N$ , would cause the highest proportional change in fitness.

Subpopulation-size elasticity,  $e(N)$  maximizes when both migration rate and subpopulation size are small. In weakly interconnected subpopulations the frequency of deleterious mutations is mainly determined by the balance between selection and drift; increasing the size of the subpopulations increases the efficiency of selection while decreasing the strength of drift. This explains the significant decrease of subpopulation-size elasticity for large subpopulations, for example,  $e(N) = 0.760$  and  $0.278$  for  $N = 20$  and  $50$ , respectively ( $m = 0.01$ ,  $n = 10$ ).

Changes in the migration rate would have an important effect on genetic load (ie  $e(m)$  is large) when subpopulation sizes are small. However, the dependence of genetic load on migration rate decreases considerably when migration rate is already enhanced and/or subpopulations are large; large subpopulations can efficiently purge deleterious alleles even for low migration

rates. In any case, the elasticity with respect to migration rate,  $e(m)$ , is always lower than subpopulation-size elasticity,  $e(N)$ .

Finally, our results concerning subpopulation-number elasticity,  $e(n)$ , suggest that for most population structures, the number of subpopulations appears to be a parameter of minor importance in terms of population fitness; fitness would be independent of the number of interconnected subpopulations apart from the case of small subpopulation sizes and enhanced migration (Figure 2a).

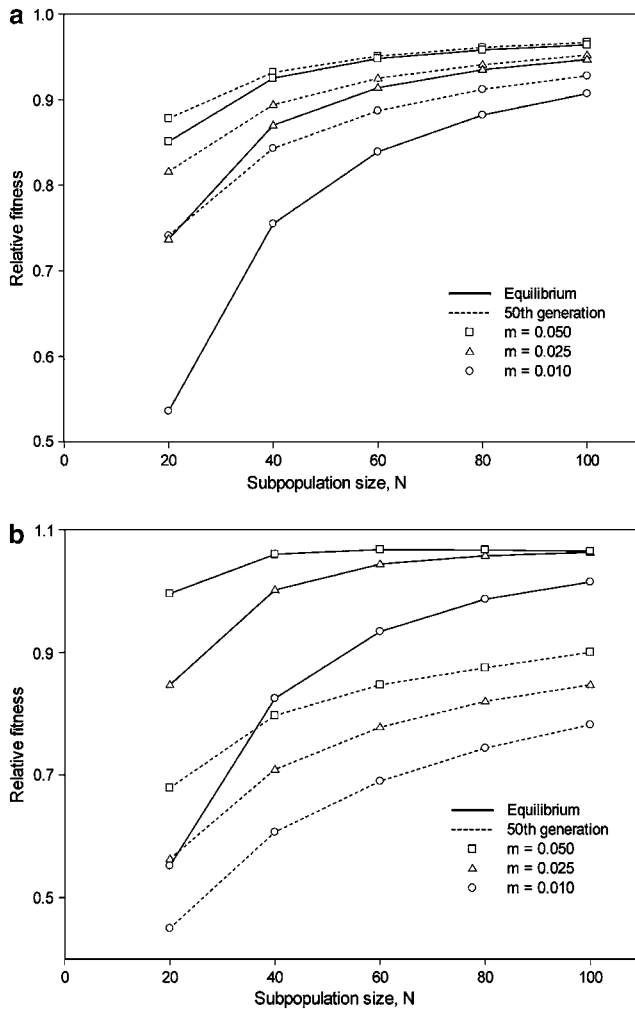
The elasticity analysis concerning the first 50 generations after subdivision, leads to the same conclusions about the relative influence of the migration rate and the size of subpopulations. Although, we observed lower values of elasticity relative to equilibrium when both migration and subpopulation sizes are low, elasticities remain quite large. To give an example in terms of fitness, when  $N$  is divided by a factor of two, the rate of decrease in fitness will be between 5 and 15% per generation (as elasticity varies between 0.1 and 0.3; see Figure 2b).

#### Dependence of the results on the choice of mutation parameters

Although in the section 'Values of mutation parameters' we argued that the set of mutation parameters used so far would be the more plausible, it is of interest to investigate the robustness of our results for other values of the mutation parameters.

Selection coefficient determines strongly the magnitude of elasticity of genetic load with regard to the parameters of population structure ( $N$ ,  $m$ ,  $n$ ).

For nearly neutral to slightly deleterious alleles, elasticities are large suggesting that changes in popula-

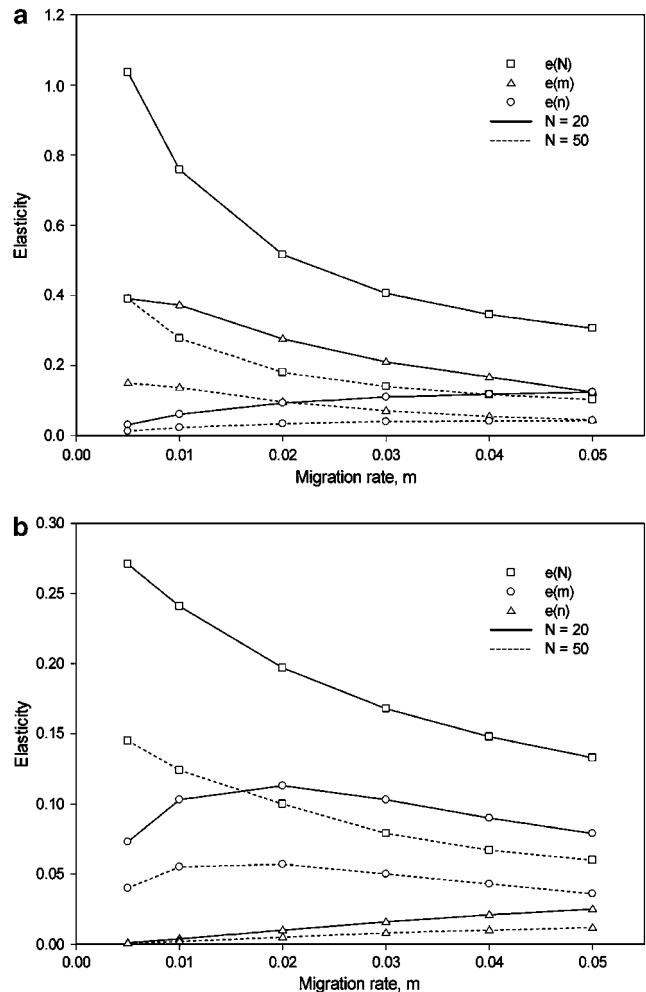


**Figure 1** The fitness of the subdivided population *versus* the subpopulation size for different numbers of migrants. Fitness of the subdivided population is scaled by the fitness of a single population of the same total size with the subdivided one. The total population size is set to  $N_{\text{tot}}=1000$ . The values of the mutation parameters are set to  $s=0.02$ ,  $U=1$ . Results are shown for two dominance coefficients (a)  $h=0.3$  and (b)  $h=0.1$ .

tion structure would have a strong influence on genetic load (Figure 3). Even the elasticity with regard to the number of subpopulations takes high values ( $e(n)=0.325$  when  $s=0.005$  and  $N=50$ ). This means that, for very slightly deleterious alleles, variation in the number of subpopulations would influence genetic fitness.

In contrast, genetic load due to more detrimental mutations ( $s>0.1$ ) depends only slightly on population structure. Selection acts efficiently against these alleles, which segregate in low frequencies for most population structures.

Changes in population structure would have a more important effect on genetic load for nearly additive than for highly recessive alleles (Figure 3). Moreover, the slightly negative elasticities obtained for highly recessive and mildly deleterious mutations ( $h=0.1$ ,  $s\gtrsim 0.1$ ) suggest that, in this case, increasing the size or the number of subpopulations and/or the migration rate could have negative effects on genetic fitness.



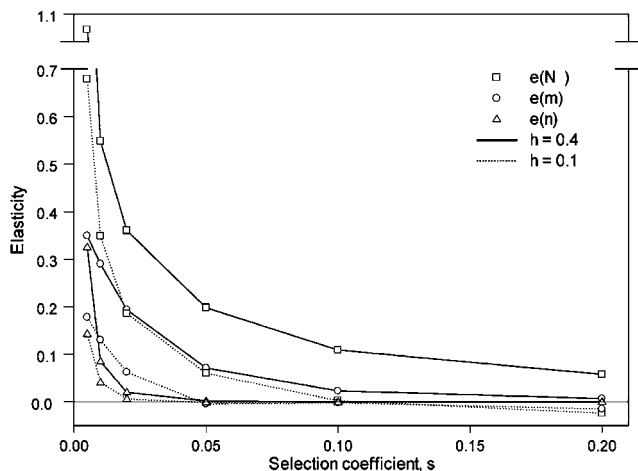
**Figure 2** Elasticity of fitness with regard to subpopulation size,  $e(N)$ , migration rate,  $e(m)$ , and the number of subpopulations,  $e(n)$  (a) at equilibrium and (b) 50 generations after subdivision has started. To calculate elasticity at disequilibrium, the proportional increment in the parameters of population structure is set to 50% (see ‘Methods’ for details). The number of subpopulations is set to  $n=10$ . The selection coefficient is set to  $s=0.02$ , the dominance coefficient is set to  $h=0.3$  and the genome mutation rate to  $U=1$ .

Most importantly, the order of elasticities does not change with the value of the mutation parameters. In other words, the parameter that determines primarily the fitness of the subdivided population remains subpopulation size, whatever the values of  $h$  and  $s$  are.

#### Rules of thumbs under the finite island model

A crucial issue for the conservation of subdivided populations is to find which combination of the parameters of population structure minimizes genetic load.

We addressed this question by calculating the minimum values of the number of migrants, the size and the number of subpopulations for which the reduction in fitness of the subdivided population is less than 5% relative to the fitness of an infinitely large population at mutation–selection equilibrium. This 5% threshold is somewhat arbitrary and it should depend on the vital rates of the studied species. However, in populations of



**Figure 3** Elasticity of fitness versus the selection coefficient,  $s$ , for two values of the dominance coefficient ( $h=0.1$  and  $0.4$ ). The parameters that determine population structure are set to  $N=50$ ,  $m=0.01$  and  $n=10$ .

endangered species, which frequently exhibit low growth rates, a change of 5% in viability could substantially alter their probability of extinction (Johnson and Braun, 1999).

Results are shown for the number of migrants,  $Nm$ , rather than for the migration rate,  $m$ , for three reasons: (i) the level of inbreeding within populations and the genetic differentiation between subpopulations would be determined by the product  $Nm$  rather than by  $m$  alone (see, for instance, Cockerham and Weir, 1987) (ii) the number of migrants is measurable in real populations via the parameter  $F_{ST}$ , and (iii) previous rules of thumb with regard to the viability of fragmented populations were expressed in terms of number of migrants (Mills and Allendorf, 1996; Couvet, 2002).

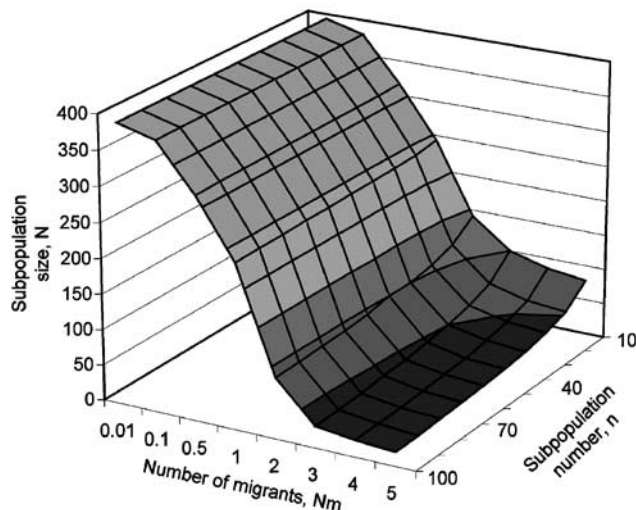
It is important to notice that the minimum number of migrants that ensures low genetic load depends strongly on the size of subpopulations and vice versa (Figure 4).

For example, large subpopulations ( $N \geq 100$ ) are necessary to retain population fitness when the number of migrants is low to moderate ( $Nm \leq 2$ ), while fitness is high even for small subpopulations (a few dozens of individuals) when migration is enhanced.

For large subpopulations and moderate migration, the number of subpopulations has not any effect on the threshold values of  $N$  and  $Nm$ . However, when migration is enhanced, the number of interconnected subpopulations decreases, although moderately, the threshold subpopulation size. For example, when  $Nm=5$ , the threshold subpopulation size is  $N=78$  and 20 for subpopulation numbers  $n=10$  and 100, respectively.

## Discussion

The goal of this study was to investigate how population structure determines the degree of mutation accumulation and, consequently, the fitness of a subdivided population. In order to explore this issue, we modified the model of Glémin *et al* (2003). This allowed us to account for cases where genetic drift plays an important role in shaping the frequency of deleterious alleles, that



**Figure 4** Population structure associated with low genetic load. The threshold values of the subpopulation size,  $N$ , and the number of migrants,  $Nm$ , for different numbers of interconnected subpopulations,  $n$ . The threshold values are defined such as the reduction in fitness of the subdivided population to be less than 5% relative to the fitness of a single infinite population. The values of the mutation parameters are as in Figure 2.

is, slightly deleterious alleles and small subpopulation sizes.

We conducted an elasticity analysis in order to identify how changes in the parameters of population structure (ie the migration rate, the number and the size of subpopulations), affect the fitness of subdivided populations. Elasticity analysis is widely used in the field of population dynamics, in order to assess the relative importance of changes in demographic factors on population growth rate (see for instance Caswell, 1989; Silvertown *et al*, 1996; Beissinger and Westphal, 1998).

Elasticity analysis can be of great interest for the genetic management of subdivided populations as it can be used to assess the effectiveness of different management options, through a quantification of the expected outcome. Frankham *et al* (2002) laid out the range of options, which are: (i) increase subpopulation size by enhancing the area and/or the quality of local habitats, (ii) increase migration rate between subpopulations, and (iii) increase the number of interconnected subpopulations by re-establishing extinct populations and creating new habitat corridors.

### Few large or many small populations?

Our analysis shows that subdivision has, in most cases, detrimental effects on fitness. The negative genetic consequences of subdivision were also pointed out by Higgins and Lynch (2001), who studied the interaction of genetic, demographic and environmental factors on the persistence of subdivided populations.

Conversely, Whitlock (2002) argued that, in the case of highly recessive alleles, subdivision could result to an increase in fitness relative to a single population. Our results concerning equilibrium support this view. However, it must be underlined that, fitness can decrease significantly during purging (Figure 1b). Hence, the



conclusion about the positive effects of subdivision on fitness must be moderated, since population persistence can be seriously impaired in the first generations after subdivision.

Moreover, our results point out that to overcome the detrimental effects of subdivision on fitness, large subpopulations must be preserved. This is also true for highly recessive alleles, especially at disequilibrium.

Overall, a general rule can be derived: for the same total population size, a small number of large subpopulations would be associated with higher fitness, compared to numerous but small subpopulations.

Our approach makes a number of simplifying assumptions, for example, it does not account for fluctuations in local and total population size due to demographic and environmental stochasticity. The inclusion of demographic stochasticity should reinforce our conclusions since in larger subpopulations the influence of this source of extinction would be minimized. Moreover, although the importance of the number of subpopulations on persistence increases with the magnitude of environmental fluctuations, the critical number of subpopulations should remain rather low (eg  $n=5-10$ ; Bascombe *et al*, 2002). However, a complete analysis of the joint influence of genetic, demographic and environmental factors on the optimal population structure is still pending.

#### Relative influence of the parameters of population structure

Elasticity analysis revealed that subpopulation size is the most influential parameter in terms of fitness. This is particularly so for the population structures with the more detrimental effects on fitness, that is, small and weakly interconnected subpopulations.

Moreover, we expect the conclusion, that changes in subpopulation size affect more fitness than changes in migration rates, to be particularly robust. The reason is that our model considers that exchanges between subpopulations happen at random. However, in most cases, dispersal is limited and migration should be more probable between neighbouring subpopulations. It has been shown that if migration happens between neighbouring subpopulations, the positive effects of migration on genetic load are less important than under the island model (Higgins and Lynch, 2001). Therefore, our model would rather overestimate the positive effects of migration in real conditions.

In most cases studied here, changing the number of subpopulations has a rather minor effect on genetic load: subpopulation-number elasticity,  $e(n)$ , takes low values ( $e(n) \rightarrow 0$ ) for most population structures. This is most convenient on conservation grounds since it suggests that interconnecting a small number of subpopulations would suffice to restore the greater part of genetic fitness. Furthermore, this result has an important methodological consequence: with regard to the accumulation of deleterious alleles, the assumption of an infinite array of subpopulations is a good approximation even for populations of small total size.

Caution must be paid, however, to the exceptions of this statement: the number of subpopulations can have a significant effect on genetic load for small subpopulations and very slightly deleterious mutations. In such a

case, an increase in the number of interconnected subpopulations could lower genetic load since it maintains the genetic variability of an otherwise highly inbred set of subpopulations. Obviously, in order for the number of subpopulations to have an effect on fitness, gene flow between subpopulations should not be strongly restricted.

#### The case of disequilibrium

The relative influence of the parameters of population structure remains the same whether we examined the first generations after subdivision or equilibrium. The conclusions about the relative effectiveness of the different conservation options are, therefore, valid regardless of the number of generations that the population experiences subdivision.

Moreover, one can notice that changes in subpopulation size and migration rate can have important impacts on the rate of fitness decline in the short term, as witnessed by the rather high values of elasticities.

However, for small and weakly interconnected subpopulations elasticities are significantly smaller in the first generations than at equilibrium. For such population structures, the frequency of deleterious alleles and, consequently, the reduction in fitness, is more important in later generations (compare, for instance, fitness in the first 50 generations with that at equilibrium for  $N=20$  and  $m=0.01$  in Figure 1a). The same proportional increase in the subpopulation size, migration rate or the number of subpopulations (although the influence of the latter is minor) has a higher potential for heterosis and therefore results to a higher proportional recovery of fitness in later generations than in the first generations after subdivision.

#### Reappraisal of previous rules of thumb in conservation

The issue of minimum values of population size or migration rates that would ensure the persistence of fragmented populations has motivated several theoretical and experimental studies. In particular, the optimal level of connectivity between subpopulations has become a central issue in conservation and a rule of thumb of one migrant per generation per subpopulation has been evoked and tested repeatedly (Mace and Lande, 1991; Mills and Allendorf, 1996; Bryant *et al*, 1999; Newman and Tallmon, 2001). However, this rule was determined without any reference to the value of the other parameters of population structure. The question is, therefore, what conclusions can we draw if the three parameters are considered together?

Our results show a large inter-dependence between population size and migration rate (Figure 4). For low subpopulation sizes ( $N \lesssim 50$ ), more than two migrants per generations are needed to ensure high fitness, while low migration rates are sufficient (less than one migrant per generation) for large subpopulations.

It must be also noticed that the benefits from migration do not depend only on the number of migrants; the pattern of migration (eg symmetry in the exchange of migrants between subpopulations or isolation by distance) could be equally if not important in terms of fitness (Higgins and Lynch, 2001; Bouchy *et al*, 2005).

Thus, any rule of thumb with regard to connectivity or estimations of minimum viable populations could lead to

erroneous conclusions as long as they ignore the configuration of the subdivided population.

## Acknowledgements

We thank D Roze for useful comments and S Glémin and D Roze for providing us their manuscripts before publication.

## References

- Bascombe J, Possingham H, Roughgarden J (2002). Patchy populations in stochastic environments: critical number of patches for persistence. *Am Nat* **159**: 128–137.
- Bataillon T (2000). Estimation of spontaneous genome-wide mutation rate parameters: whither beneficial mutations? *Heredity* **84**: 497–501.
- Beissinger SR, Westphal MI (1998). On the use of demographic models of population viability in endangered species management. *J Wildlife Manage* **62**: 821–841.
- Bouchy P, Theodorou K, Couvet D (2005). Metapopulation viability: influence of migration. *Conserv Genet* **6**: 75–85.
- Bryant EH, Backus VL, Clark ME, Reed DH (1999). Experimental tests of captive breeding for endangered species. *Conserv Biol* **13**: 1487–1496.
- Caballero A, Cusi E, García C, García-Dorado A (2002). Accumulation of deleterious mutations: additional *Drosophila melanogaster* estimates and a simulation of the effects of selection. *Evolution* **56**: 1150–1159.
- Caswell H (1989). *Matrix Population Models*. Sinauer: Sunderland, MA.
- Cockerham CC, Weir BS (1987). Correlations, descent measures: drift with migration and mutation. *Proc Natl Acad Sci USA* **84**: 8512–8514.
- Couvet D (2002). Deleterious effects of restricted gene flow in fragmented populations. *Conserv Biol* **16**: 369–376.
- Crow JF, Kimura M (1970). *An Introduction to Population Genetics*. Alpha Editions: Minneapolis.
- Ewens WJ (1979). *Mathematical Population Genetics*, 1st edn. Springer: New York.
- Frankham R, Ballou JD, Briscoe DA (2002). *Introduction to Conservation Genetics*. Cambridge University Press: Cambridge.
- Fry JD (2001). Rapid mutational declines in *Drosophila*. *Genet Res* **77**: 53–60.
- Fry JD, Heinsohn SN (2002). Environment dependence of mutational parameters for viability in *Drosophila melanogaster*. *Genetics* **161**: 1155–1167.
- Glémin S, Ronfort J, Bataillon T (2003). Patterns of inbreeding depression and architecture of the load in subdivided populations. *Genetics* **165**: 2193–2212.
- Higgins K, Lynch M (2001). Metapopulation extinction caused by mutation accumulation. *Proc Natl Acad Sci USA* **98**: 2928–2933.
- Johnson KH, Braun CE (1999). Viability and conservation of an exploited sage grouse population. *Conserv Biol* **13**: 77–84.
- Kirkpatrick M, Jarne P (2000). The effect of a bottleneck on inbreeding depression and the genetic load. *Am Nat* **155**: 154–167.
- Lynch M, Blanchard J, Houle D, Kibota T, Schultz S, Vassilieva L *et al* (1999). Perspective: spontaneous deleterious mutation. *Evolution* **53**: 645–663.
- Lynch M, Conery J, Bürger R (1995). Mutation accumulation and the extinction of small populations. *Am Nat* **146**: 489–518.
- Mace GM, Lande R (1991). Assessing extinction threats: toward a reevaluation of the IUCN threatened species categories. *Conserv Biol* **5**: 148–157.
- Mills LS, Allendorf FW (1996). The one-migrant-per-generation rule in conservation and management. *Conserv Biol* **10**: 1509–1518.
- Newman D, Tallmon DA (2001). Experimental evidence for beneficial fitness effects of gene flow in recently isolated populations. *Conserv Biol* **15**: 1054–1063.
- Ohta T, Kimura M (1969). Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. *Genetics* **63**: 220–238.
- Ohta T, Kimura M (1971). Linkage disequilibrium between two segregating nucleotide sites under the steady flux of mutations in a finite population. *Genetics* **68**: 571–580.
- Rodríguez-Ramilo ST, Pérez-Figueroa A, Fernández B, Fernández J, Caballero A (2004). Mutation–selection balance accounting for genetic variation for viability in *Drosophila melanogaster* as deduced from an inbreeding and artificial selection experiment. *J Evol Biol* **17**: 528–541.
- Ronfort J, Couvet D (1995). A stochastic model of selection on selfing rates in structural populations. *Genet Res* **65**: 209–222.
- Roze D, Rousset F (2004). Joint effects of self-fertilization and population structure on mutation load, inbreeding depression and heterosis. *Genetics* **167**: 1001–1015.
- Saccheri I, Kuussaari M, Kankare M, Vikman P, Fortelius W, Hanski I (1998). Inbreeding and extinction in a butterfly metapopulation. *Nature* **392**: 491–494.
- Schoen D, David J, Bataillon T (1998). Deleterious mutation accumulation and the regeneration of genetic resources. *Proc Natl Acad Sci USA* **95**: 394–399.
- Silvertown J, Franco JM, Menges E (1996). Interpretation of elasticity matrices as an aid to the management of plant populations for conservation. *Conserv Biol* **10**: 591–597.
- Soulé ME, Orians GH (2002). *Conservation Biology. Research Priorities for the Next Decade*. Island Press: Washington.
- Theodorou K, Couvet D (2002). Inbreeding depression and heterosis in a subdivided population: influence of the mating system. *Genet Res* **80**: 107–116.
- Theodorou K, Couvet D (2004). Introduction of captive breeders to the wild: harmful or beneficial? *Conserv Genet* **5**: 1–12.
- Wang J, Hill WG, Charlesworth D, Charlesworth B (1999). Dynamics of inbreeding depression due to deleterious mutations in small populations: mutation parameters and inbreeding rate. *Genet Res* **74**: 165–178.
- Whitlock MC (2002). Selection, load and inbreeding depression in a large metapopulation. *Genetics* **160**: 1191–1202.
- Whitlock MC (2003). Fixation probability and time in subdivided populations. *Genetics* **164**: 767–779.
- Whitlock MC, Ingvarsson PK, Hatfield T (2000). Local drift load and the heterosis of interconnected populations. *Heredity* **84**: 452–457.