Fixation probability for a beneficial allele and a mutant strategy in a linear game under weak selection in a finite island model

Véronique Ladret, Sabin Lessard*

Département de mathématiques et de statistique, Université de Montréal, C.P. 6128 Succursale Centre-ville, Montréal, Qué., Canada H3C 3J7

Abstract

The effect of population structure on the probability of fixation of a newly introduced mutant under weak selection is studied using a coalescent approach. Wright's island model in a framework of a finite number of demes is assumed and two selection regimes are considered: a beneficial allele model and a linear game among offspring. A first-order approximation of the fixation probability for a single mutant with respect to the intensity of selection is deduced. The approximation requires the calculation of expected coalescence times, under neutrality, for lineages starting from two or three sampled individuals. The results are obtained in a general setting without assumptions on the number of demes, the deme size or the migration rate, which allows for simultaneous coalescence or migration events in the genealogy of the sampled individuals. Comparisons are made with limit cases as the deme size or the number of demes goes to infinity or the migration rate goes to zero for which a diffusion approximation approach is possible. Conditions for selection to favor a mutant strategy replacing a resident strategy in the context of a linear game in a finite island population are addressed.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Finite island model; Beneficial allele; Fixation probability; Coalescence times; Evolutionary game theory; One-third law

1. Introduction

How does selection affect the probability that a single mutant ultimately fixes in the entire population? This is one of the oldest questions raised by mathematical population geneticists. In the framework of an ideal diploid population, Fisher (1922, 1930) and Haldane (1927) found that the probability of fixation of a new beneficial allele should be approximately given by twice the selective advantage of the heterozygote. Using a diffusion approximation for the frequency of an allele under weak selection in a finite population and allowing for deleterious mutation, arbitrary dominance and arbitrary but large effective population size, Kimura (1957, 1962) obtained a more general formula for the probability of fixation of a mutant given its initial frequency (see also Caballero and Hill, 1992).

In many biological contexts, selection differences are not constant but shaped by the composition of the population.

With interacting individuals, evolutionary game theory appears as a key theoretical framework to make predictions about evolution (Hamilton, 1967; Maynard Smith and Price, 1973). The approach has been applied notably to sex ratio evolution, parental care, dispersal strategies and predator–prey or host–parasite interactions (see, e.g., Maynard Smith, 1982; Godfray, 1995; Abrams and Matsuda, 1997; Hardy, 2002; Taylor et al., 2006). Some linear games involving two strategies, namely the prisoner's dilemma (Axelrod and Hamilton, 1981) and the snowdrift game (Sugden, 1986), have been used to explain the evolution of cooperation and they have been documented for certain viruses (Turner and Chao, 1999). Similarly, the rock–scissors-paper game has served to model the three mating behaviors of the male lizard, *Uta stansburniana* (Sinervo and Lively, 1996) and poisoning strategies used by bacteria (Kerr et al., 2002).There are many other examples of populations engaged in game-like-theory interactions (see, e.g., Nowak and Sigmund, 2004, for a recent review).

Traditionally studied in the framework of infinite populations (Maynard Smith and Price, 1973; Maynard

---

*Corresponding author.

E-mail address: lessards@dms.umontreal.ca (S. Lessard).

Smith, 1974; Taylor and Jonker, 1978; Hofbauer and Sigmund, 1998), linear games have been considered recently in the context of a finite population (Nowak et al., 2004; Wild and Taylor, 2004; Lessard, 2005; Orzack and Hines, 2005; Imhof and Nowak, 2006). In this framework, selection may favor or oppose a mutant strategy replacing a resident strategy. This will be the case if the probability of fixation of the mutant is larger or smaller than it would be under neutrality. Then, a question of interest is the following: does selection oppose any mutant strategy replacing an ESS, an evolutionarily stable strategy in Maynard Smith and Price's (1973) sense? Another one is: what is the condition for a strategy A to replace a strategy B when both are the best replies to themselves? The answer for a panmictic population is that the frequency of A at the unstable equilibrium in the replicator dynamics should be less than $1/3$. This has been termed the one-third law in Nowak et al. (2004).

The effect of population structure on the fixation probability was addressed by Maruyama (1970, 1974). He considered Wright's (1931) island model, except that the population consists of a finite number of demes or colonies, and some other models with conservative migration (which means that the subpopulation sizes are not changed by migration), and he showed that the probability of fixation of additive beneficial alleles remains the same. This result relies on an analogy between such populations and a panmictic population with a rescaled size. It was confirmed by separation-of-time-scales arguments in diffusion approximations in the case where migration is strong (Nagylaki, 1980, 1982, 2000) or low (Slatkin, 1981; Takahata, 1991) and also in the case of a large number of demes (Cherry and Wakeley, 2003; Wakeley, 2003). However, Barton (1993) and more recently Cherry (2003b) showed that, under extinction and recolonization of demes, the probability of fixation of beneficial alleles can be altered by population structure. This is the case because the scaled selection intensity depends on the population structure. Whitlock (2003) studied several geographical models, including the stepping-stone model, under various assumptions as hard versus soft selection, deleterious mutation and arbitrary dominance with respect to the probability of fixation. See also Cherry (2003a) for the case of frequency-dependent selection and Rousset and Billiard (2000) and Roze and Rousset (2003, 2004) for the effect of inbreeding and the role of identity measures between interacting individuals.

In this article we use a coalescent approach suggested in Rousset (2003) and rigorously justified in Lessard and Ladret (2007) to study the probability of fixation of a single mutant allele under the joint effect of population structure and weak selection. A finite island model with D demes of N individuals is assumed and two selection scenarios for viability (fitness) differences between individuals within demes are considered. The first scenario is the well-known beneficial allele model previously studied in the framework of an island model by Maruyama (1970), among others.

Viability differences in the second scenario, called the linear game model and recently studied in the context of a single deme by Nowak et al. (2004), among others, are a result of pairwise interactions between individuals using mixed strategies. The selection differences will be assumed to be small and determined at a single locus and the population to be haploid, but the results will remain valid in the diploid case under additivity assumptions. Generations will be discrete, non-overlapping and migration will take place after reproduction so that a proportion m of individuals in each deme will come from the other demes. Then, there will be selection within each deme (soft selection) followed by random sampling to restore the deme size. In the case of a diploid population, migration is assumed to be gametic. While these models have already been studied in limit cases using diffusion approximations with scaled population parameters (see, e.g., Whitlock, 2003; Lessard, 2005, and references therein), the method presented in this paper allows for arbitrary number of demes D, arbitrary deme size N and arbitrary backward migration rate m. The approximation obtained for the probability of fixation assumes all these parameters fixed and it is valid as long as the intensity of selection is small compared to the inverse of the total population size.

We will show that the derivative of the fixation probability with respect to the intensity of selection evaluated at neutrality can be expressed in terms of expected coalescence times, under neutrality, for samples of two individuals in the case of a beneficial mutant allele and for samples of two and three individuals in the case of a mutant strategy in a linear game. These expected coalescence times can be found exactly using a standard conditioning procedure for Markov chains. Limit values as the number of demes or the deme size goes to infinity or as the migration rate goes to zero, which correspond to different limit models previously studied, are deduced: (1) the structured coalescent (SC) extending Kingman's (1982a, b) coalescent to a subdivided population and in particular to the finite island model in the case of a migration rate m of order $1/N$ such that Nm tends to a positive constant as the deme size, N, goes to infinity (Notohara, 1990; Herbots, 1994; Wilkinson-Herbots, 1998, 2003); (2) the strong- (or high-) migration (SM or HM, respectively) limit in the case of a migration rate m remaining constant (equal to 1, respectively) as the deme size N approaches infinity (Nagylaki, 1980, 2000; Notohara, 1993, 2000; Bahlo and Griffiths, 2001); (3) the low-migration (LM) limit in the case of a migration rate m going to 0 such that Nm tends to 0 as N goes to infinity (Slatkin, 1981; Takahata, 1991; Notohara, 2001); (4) the many-demes (MD) limit in the case of a number of demes D that tends to infinity (Wakeley, 1998, 2003, 2004). In this case, it has been noticed that the deme size N does not have to be large and the migration rate m small (Wakeley and Lessard, 2003; Lessard and Wakeley, 2004).

The results are presented as follows. In Section 2, the finite island model with the two selection regimes

considered, a beneficial allele and a linear game, is described. In Section 3, a first-order approximation for the probability of fixation of a single mutant with respect to the intensity of selection is deduced. In Section 4, it is shown that the derivative of the fixation probability with respect to the intensity of selection evaluated at neutrality, called the fixation coefficient, can be expressed in terms expected coalescence times of lineages for two or three sampled individuals in the neutral setting. In Section 5, systems of linear equations for these expected coalescence times are deduced and their exact solutions are given analytically (see the Appendix for samples of three individuals), illustrated numerically and compared to their approximations in different limit cases: low migration, high or strong migration, structured coalescent, many demes. In Section 6, conclusions on the approximation of the fixation probability in the different settings are drawn. Finally, in Section 7, we discuss differences between results obtained from a diffusion approximation and results obtained from a coalescent approach in the case of a beneficial allele and between conditions for a strategy to replace another in the case of a linear game in an island model versus a panmictic population.

2. Finite island model under two selection regimes

Consider a haploid population subdivided into D demes, $D \geqslant 3$, with N individuals in each deme. Suppose that the population undergoes discrete, non-overlapping generations represented by time steps $t = 0, 1, 2, \ldots$. At the beginning of every generation, each individual in the population produces the same very large number of offspring, which then disperse independently among the demes. Let m_{ij} be the proportion of offspring in deme i that come from deme j and assume these backward migration rates to be constant over time. Moreover, suppose that the backward migration rates to other demes are all equal, that is, $m_{ij} = m/(D-1)$ for all $i \neq j$, where $m = 1 - m_{11}$ is the probability that an offspring in any given deme has been produced in one of the $D-1$ other demes. We assume throughout $0 < m < 1$.

Consider two types of individuals in the population, a wild or resident type B and a mutant type A. If x_i denotes the frequency of A in deme i at the beginning of a given generation and $\tilde{x}_i$ the frequency of A in deme i after migration, then we have

$$\tilde{x}_i = (1-m)x_i + \frac{m}{D-1}\sum_{j \neq i} x_j. \tag{1}$$

Notice that

$$1 - \tilde{x}_i = (1-m)(1-x_i) + \frac{m}{D-1}\sum_{j \neq i}(1-x_j). \tag{2}$$

Migration is followed by selection among offspring within the same deme. This is known as soft selection. The two selection schemes considered are the following.

Beneficial allele: Under this selection scheme, the fitnesses of A and B offspring are $1 + s$ and 1, respectively, where $s > 0$. Thus, the frequency of A among the individuals in deme i after selection is

$$x_i' = \frac{\tilde{x}_i(1+s)}{\tilde{x}_i(1+s) + (1-\tilde{x}_i)}.$$

Throughout the paper, we make the assumption that the selection intensity s is small, which models weak selection. In such a case, a first-order development of x_i' with respect to s leads to

$$x_i' = \tilde{x}_i + s\tilde{x}_i(1-\tilde{x}_i) + o(s). \tag{3}$$

Notice that we get the same first-order development if the population is diploid, migration is gametic and selection occurs after random union of gametes within the demes according to the fitnesses $1 + 2s$, $1 + s$ and 1 for the genotypes AA, AB and BB, respectively, in which case we have

$$x_i' = \frac{\tilde{x}_i^2(1+2s) + \tilde{x}_i(1-\tilde{x}_i)(1+s)}{\tilde{x}_i^2(1+2s) + 2\tilde{x}_i(1-\tilde{x}_i)(1+s) + (1-\tilde{x}_i)^2}$$

in deme i after selection.

Linear game: Under this selection scenario, we assume that the two types of individuals, A and B, are associated with the mixed strategies $\mathbf{p}_A$ and $\mathbf{p}_B$, respectively, these being frequency vectors whose components give the probabilities of using some pure strategies in a contest against an opponent. More precisely, pairwise interactions take place among the offspring in the same deme so that the viabilities (fitnesses) for types A and B in deme i take the forms

$$f_{A,i} = 1 + s\mathbf{p}_A \cdot W\bar{\mathbf{p}}_i$$

and

$$f_{B,i} = 1 + s\mathbf{p}_B \cdot W\bar{\mathbf{p}}_i,$$

respectively, where $s > 0$ measures the selection intensity, W refers to some game matrix, $\mathbf{x} \cdot \mathbf{y}$ denotes the scalar product of two vectors $\mathbf{x}$ and $\mathbf{y}$, and $\bar{\mathbf{p}}_i$ stands for the mean strategy in deme i before selection, that is,

$$\bar{\mathbf{p}}_i = \tilde{x}_i\mathbf{p}_A + (1-\tilde{x}_i)\mathbf{p}_B = \tilde{x}_i(\mathbf{p}_A - \mathbf{p}_B) + \mathbf{p}_B.$$

Following selection, the frequency of A among the individuals in deme i becomes

$$x_i' = \frac{\tilde{x}_i f_{A,i}}{\tilde{x}_i f_{A,i} + (1-\tilde{x}_i)f_{B,i}}.$$

The next generation is obtained by drawing at random N individuals in each deme so that the number of A individuals in deme i follows a binomial distribution of parameters N and x_i'. Then, the frequency of A in deme i has mean x_i' whose first-order development with respect to s is given by

$$x_i' = \tilde{x}_i + s\{\tilde{x}_i(1-\tilde{x}_i)(\mathbf{p}_A - \mathbf{p}_B) \cdot W\bar{\mathbf{p}}_i\} + o(s),$$

or equivalently,

$$x_i' = \tilde{x}_i + s\{\tilde{x}_i^2(1 - \tilde{x}_i)(\mathbf{p}_A - \mathbf{p}_B) \cdot W(\mathbf{p}_A - \mathbf{p}_B)$$
$$+ \tilde{x}_i(1 - \tilde{x}_i)(\mathbf{p}_A - \mathbf{p}_B) \cdot W\mathbf{p}_B\} + o(s), \tag{4}$$

owing to (2). Notice that the same first-order development is obtained in the diploid case if migration is gametic and the fitnesses of the genotypes AA, AB and BB after random union of gametes in deme i are $f_{AA,i} = 1 + 2s\mathbf{p}_A \cdot W\bar{\mathbf{p}}_i$, $f_{AB,i} = 1 + s(\mathbf{p}_A + \mathbf{p}_B) \cdot W\bar{\mathbf{p}}_i$ and $f_{BB,i} = 1 + 2s\mathbf{p}_B \cdot W\bar{\mathbf{p}}_i$, respectively, which lead to the equation

$$x_i' = \frac{\tilde{x}_i^2 f_{AA,i} + \tilde{x}_i(1 - \tilde{x}_i)f_{AB,i}}{\tilde{x}_i^2 f_{AA,i} + 2\tilde{x}_i(1 - \tilde{x}_i)f_{AB,i} + (1 - \tilde{x}_i)^2 f_{BB,i}}.$$

3. Fixation probability under weak selection

In this section we use and extend an approach suggested in Rousset (2003, 2004) to show that the probability of fixation of a single mutant can be expressed to a first order with respect to s as a function of expected coalescence times between pairs of individuals in the case of a beneficial allele, and between pairs and triplets of individuals in the case of a linear game.

Let $X_i(t)$ be the frequency of the mutant type A in deme i at time $t \geqslant 0$. Define the vector $\mathbf{X}(t) = (X_1(t), X_2(t), \ldots, X_D(t))$. Suppose that, at time $t = 0$, there is a single individual of the mutant type A in the whole population and that this individual is in deme 1, that is, $\mathbf{X}(0) = \mathbf{x}_0$, where

$$\mathbf{x}_0 = \left(\frac{1}{N}, 0, \ldots, 0\right).$$

Let $Z(t)$ be the frequency of A in the whole population at time $t \geqslant 0$. Then, we have

$$Z(t) = \frac{1}{D}\sum_{i=1}^{D} X_i(t)$$

and $Z(0) = 1/(ND)$.

Denote by $u(s)$ the probability of ultimate fixation of type A in the whole population when the selection intensity is s. We have to determine $u(0)$ and $u'(0)$, the derivative evaluated at $s = 0$, in order to find the first-order development of $u(s)$ with respect to s, that is,

$$u(s) = u(0) + su'(0) + o(s).$$

The term $u'(0)$ will be called the fixation coefficient. In the following, we shall denote by E_0 (P_0, respectively) the expectation (probability, respectively) in the case of neutrality, that is, when $s = 0$. We will use the notation E (P, respectively) in the case of an unspecified selection intensity s.

In the neutral selection scenario, all individuals in the population will eventually descend from a single random individual in the initial generation as a result of random drift and symmetry. Therefore, the fixation probability of

A when $s = 0$ is

$$u(0) = Z(0) = \frac{1}{ND}.$$

Now, consider $s \geqslant 0$. The sequence $Z(t)$ for $t = 0, 1, \ldots$ is a Markov chain on the states $k/(ND)$ for $k = 0, 1, \ldots, ND$, with initial state $Z(0) = 1/(ND)$ and fixation states $z = 0$ and 1, while all other states are transient. This chain will converge in probability to a variable $Z(\infty)$, which takes the value 1 with probability $u(s)$ and 0 with the complementary probability $1 - u(s)$. Being uniformly bounded the chain will also converge in mean. This justifies to write, as claimed in Rousset (2003, 2004, p. 93), that

$$E[Z(\infty) - Z(0)] = \sum_{t \geqslant 0} E[Z(t + 1) - Z(t)],$$

that is,

$$u(s) - \frac{1}{ND} = \sum_{t \geqslant 0} E[Z(t + 1) - Z(t)]. \tag{5}$$

Then, assuming that $u(s)$ is differentiable with respect to s and that the derivative of the sum in (5) is the sum of the derivatives (see Lessard and Ladret, 2007, for a formal proof under mild regularity conditions on the transition probabilities), we have

$$u'(0) = \sum_{t \geqslant 0} \frac{d}{ds}E[Z(t + 1) - Z(t)]\bigg|_{s=0}.$$

Conditioning on the value $\mathbf{x} = (x_1, x_2, \ldots, x_D)$ taken by $\mathbf{X}(t)$, we get

$$E[Z(t + 1) - Z(t)]$$
$$= \sum_{\mathbf{x}} E[Z(t + 1) - Z(t)|\mathbf{X}(t) = \mathbf{x}]P(\mathbf{X}(t) = \mathbf{x}).$$

It follows that

$$\frac{d}{ds}E[Z(t + 1) - Z(t)]\bigg|_{s=0}$$
$$= \sum_{\mathbf{x}} P_0(\mathbf{X}(t) = \mathbf{x})\frac{d}{ds}E[Z(t + 1) - Z(t)|\mathbf{X}(t) = \mathbf{x}]\bigg|_{s=0}$$
$$+ \sum_{\mathbf{x}} E_0[Z(t + 1) - Z(t)|\mathbf{X}(t) = \mathbf{x}]\frac{d}{ds}P(\mathbf{X}(t) = \mathbf{x})\bigg|_{s=0}.$$

Under neutrality, the frequency of A in the whole population does not change in mean from one generation to the next, that is,

$$E_0[Z(t + 1) - Z(t)|\mathbf{X}(t) = \mathbf{x}] = 0.$$

Thus, we have

$$u'(0) = \sum_{t \geqslant 0}\sum_{\mathbf{x}} P_0(\mathbf{X}(t) = \mathbf{x})$$
$$\times \frac{d}{ds}E[Z(t + 1) - Z(t)|\mathbf{X}(t) = \mathbf{x}]\bigg|_{s=0},$$

where

$$E[Z(t + 1) - Z(t)|\mathbf{X}(t) = \mathbf{x}] = \frac{1}{D}\sum_{i=1}^{D}(x_i' - x_i).$$

We will now consider the two selection scenarios separately.

Beneficial allele: In this case, (3) leads to

$$\frac{d}{ds}E[Z(t+1)-Z(t)|\mathbf{X}(t)=\mathbf{x}] = \frac{1}{D}\sum_{i=1}^{D}\tilde{x}_i(1-\tilde{x}_i),$$

from which

$$u'(0) = E_0(\beta), \tag{6}$$

where

$$\beta = \sum_{t \geqslant 0}\frac{1}{D}\sum_{i=1}^{D}\tilde{X}_i(t)(1-\tilde{X}_i(t)).$$

Using (1), (2) and the identity

$$\sum_{i=1}^{D}\sum_{j\neq i}(1-X_i(t))X_j(t) = \sum_{i=1}^{D}\sum_{j\neq i}X_i(t)(1-X_j(t)),$$

we find that

$$\beta = (1-m)^2\beta_1 + 2m(1-m)\beta_2 + m^2\beta_3, \tag{7}$$

where

$$\beta_1 = \sum_{t \geqslant 0}\frac{1}{D}\sum_{i=1}^{D}X_i(t)(1-X_i(t)),$$

$$\beta_2 = \sum_{t \geqslant 0}\frac{1}{D(D-1)}\sum_{i=1}^{D}\sum_{j\neq i}X_i(t)(1-X_j(t)),$$

$$\beta_3 = \sum_{t \geqslant 0}\frac{1}{D(D-1)^2}\sum_{i=1}^{D}\sum_{j,k\neq i}(1-X_j(t))(1-X_k(t)).$$

Linear game: Owing to (4), this model leads to

$$u'(0) = E_0(\alpha)(\mathbf{p}_A-\mathbf{p}_B)\cdot W(\mathbf{p}_A-\mathbf{p}_B) \\ + E_0(\beta)(\mathbf{p}_A-\mathbf{p}_B)\cdot W\mathbf{p}_B, \tag{8}$$

where β is given in (7) and

$$\alpha = \sum_{t \geqslant 0}\frac{1}{D}\sum_{i=1}^{D}\tilde{X}_i^2(t)(1-\tilde{X}_i(t)).$$

Using (1) and (2), we find that

$$\alpha = (1-m)^3\alpha_1 + m(1-m)^2\alpha_2 + m^2(1-m)\alpha_3 + m^3\alpha_4, \tag{9}$$

where

$$\alpha_1 = \sum_{t \geqslant 0}\frac{1}{D}\sum_{i=1}^{D}X_i^2(t)(1-X_i(t)),$$

$$\alpha_2 = \sum_{t \geqslant 0}\frac{1}{D(D-1)}\sum_{i=1}^{D}\sum_{j\neq i}(2X_i(t)X_j(t)(1-X_i(t)) \\ + X_i^2(t)(1-X_j(t))),$$

$$\alpha_3 = \sum_{t \geqslant 0}\frac{1}{D(D-1)^2}\sum_{i=1}^{D}\sum_{j,k\neq i}(2X_i(t)X_j(t)(1-X_k(t)) \\ + X_j(t)X_k(t)(1-X_i(t))),$$

$$\alpha_4 = \sum_{t \geqslant 0}\frac{1}{D(D-1)^3}\sum_{i=1}^{D}\sum_{j,k,l\neq i}X_j(t)X_l(t)(1-X_k(t)).$$

Hence, in order to evaluate the fixation coefficient $u'(0)$ in the two selection scenarios, we have to find $E_0(\alpha_i)$, for $i = 1,\ldots,4$ and $E_0(\beta_i)$ for $i = 1,2,3$.

4. Fixation coefficient in terms of expected coalescence times

We will use the following notation to describe the configuration of a sample of n individuals in the finite island model (see, e.g., Wakeley, 1998). Let d be the total number of demes that contain sampled individuals and n_i be the number of those demes that contain exactly i sampled individuals. In particular, we have $\sum_{i=1}^{n}n_i = d$ and $\sum_{i=1}^{n}in_i = n$. Then, the sample configuration is represented by the vector $\mathbf{n}_d = (n_1,\ldots,n_d)$ and the set of all such vectors is denoted by S.

We will concentrate on samples of three individuals or less. A sample of one individual has only one possible configuration (1). There are two possible configurations for a sample of two individuals, $(0,1)$ and $(2,0)$, and three possible configurations for a sample of three individuals, $(0,0,1)$, $(1,1,0)$ and $(3,0,0)$. Finally, we denote by $T_\mathbf{n}$ the coalescence time of the genealogy of a sample whose configuration is $\mathbf{n}$, that is, the random variable that counts the number of generations backward in time starting from the n sampled individuals and ending with their most recent common ancestor (MRCA).

Defining $\xi_{li}(t) = 1$ if individual l in deme i at time t is of type A and 0 otherwise, the frequency of A in deme i at time t can be expressed as

$$X_i(t) = \frac{1}{N}\sum_{l=1}^{N}\xi_{li}(t),$$

and its kth power as

$$X_i^k(t) = \frac{1}{N^k}\sum_{i_1+\cdots+i_N=k}\binom{k}{i_1,\ldots,i_N}\xi_{1i}^{i_1}(t)\cdots\xi_{Ni}^{i_N}(t).$$

The variables $\xi_{li}(t)$ being exchangeable and satisfying $\xi_{li}^{i_l}(t) = \xi_{li}(t)$ if $i_l \geqslant 1$ and 1 if $i_l = 0$, the expected value of $X_i^k(t)$ under neutrality will be

$$E_0[X_i^k(t)] = \frac{1}{N^k}\sum_{l=1}^{\min(k,N)}\binom{N}{l}l!S_{kl}E_0[\xi_{1i}(t)\ldots\xi_{li}(t)], \tag{10}$$

where S_{kl} is a Stirling number of second kind, that is,

$$S_{kl} = \frac{1}{l!}\sum_{\substack{i_1+\cdots+i_l=k \\ i_1,\ldots,i_l \geqslant 1}}\binom{k}{i_1,\ldots,i_l},$$

and represents the number of ways that k distinct elements can be partitioned into l nonvoid subsets (see, e.g., Abramowitz and Stegun, 1965).

Moreover, we have

$$\frac{1}{D}\sum_{i=1}^{D}E_0[\xi_{1i}(t)\ldots\xi_{li}(t)] = \frac{1}{ND}P_0(T_{\mathbf{e}_l}\leqslant t), \qquad (11)$$

where $\mathbf{e}_l = (0,\ldots,0,1)$ represents a random sample of l individuals in a single deme. Indeed, notice that $E_0[\xi_{1i}(t)\ldots\xi_{li}(t)]$ is the probability for the most recent common ancestor (MRCA) of l individuals chosen at random without replacement in deme i at time t to be of type A. This is the probability for these individuals to have a common ancestor at time 0 multiplied by the probability for this common ancestor to be of type A. The former is the probability for the coalescence time to the MRCA to be less or equal to t, that is, $P_0(T_{\mathbf{e}_l}\leqslant t)$. The latter is $1/N$ if the common ancestor at time 0 is in deme 1, which contains a single individual of type A, and 0 otherwise. Let $A_{l,ij}(t)$ represent the event that l individuals chosen at random without replacement in deme i at time t have a common ancestor in deme j at time 0 given that they have a common ancestor at time 0. Then, we have

$$E_0[\xi_{1i}(t)\ldots\xi_{li}(t)] = \frac{1}{N}P_0(T_{\mathbf{e}_l}\leqslant t)P_0(A_{l,i1}(t)).$$

Eq. (11) follows from the symmetry property

$$P_0(A_{l,i1}(t)) = P_0(A_{l,1i}(t))$$

and the fact that

$$\sum_{i=1}^{D}P_0(A_{l,1i}(t)) = 1.$$

Using the equality $P_0(T_{\mathbf{e}_l}\leqslant t) = 1 - P_0(T_{\mathbf{e}_l} > t)$, we deduce from (10) and (11) that

$$\frac{1}{D}\sum_{i=1}^{D}E_0[X_i^k(t)]$$
$$= \frac{1}{ND} - \frac{1}{DN^{k+1}}\sum_{l=1}^{\min(k,N)}\binom{N}{l}l!S_{kl}P_0(T_{\mathbf{e}_l} > t).$$

Recalling the identities $S_{n,n} = 1$, $S_{n,n-1} = n(n-1)/2$ and $S_{n,2} = 2^{n-1} - 1$, it follows that

$$\frac{1}{D}\sum_{i=1}^{D}E_0[X_i(t)] = \frac{1}{ND},$$

$$\frac{1}{D}\sum_{i=1}^{D}E_0[X_i^2(t)]$$
$$= \frac{1}{ND}\left(1 - \left(1 - \frac{1}{N}\right)P_0(T_{(0,1)} > t)\right),$$

$$\frac{1}{D}\sum_{i=1}^{D}E_0[X_i^3(t)] = \frac{1}{ND}\left(1 - \frac{3}{N}\left(1 - \frac{1}{N}\right)P_0(T_{(0,1)} > t)\right.$$
$$\left. - \left(1 - \frac{2}{N}\right)\left(1 - \frac{1}{N}\right)P_0(T_{(0,0,1)} > t)\right).$$

Similarly, we have

$$\frac{1}{D(D-1)}\sum_{i=1}^{D}\sum_{j\neq i}E_0[X_i(t)X_j(t)]$$
$$= \frac{1}{D(D-1)N^2}\sum_{i=1}^{D}\sum_{j\neq i}\sum_{k,l=1}^{N}E_0[\xi_{ki}(t)\xi_{lj}(t)],$$

which is equal to

$$\frac{1}{ND}(1 - P_0(T_{(2,0)} > t)),$$

the probability for two individuals chosen at random in two different demes chosen at random at time t to be both of type A. The same arguments apply, including the symmetry argument, for three individuals in three different demes to yield

$$\frac{1}{D(D-1)(D-2)}\sum_{i=1}^{D}\sum_{j\neq i}\sum_{k\neq i,j}E_0[X_i(t)X_j(t)X_k(t)]$$
$$= \frac{1}{ND}(1 - P_0(T_{(3,0,0)} > t)).$$

Considering two individuals in the same deme and a third individual in a different deme, we find

$$\frac{1}{D(D-1)}\sum_{i=1}^{D}\sum_{j\neq i}E_0[X_i^2(t)X_j(t)]$$
$$= \frac{1}{D(D-1)N^3}\sum_{i=1}^{D}\sum_{j\neq i}\sum_{k,m=1}^{N}E_0[\xi_{ki}(t)\xi_{mj}(t)]$$
$$+ \frac{1}{D(D-1)N^3}\sum_{i=1}^{D}\sum_{j\neq i}\sum_{k\neq l,m=1}^{N}E_0[\xi_{ki}(t)\xi_{li}(t)\xi_{mj}(t)],$$

which is equal to

$$\frac{1}{ND}\left(1 - \frac{1}{N}P_0(T_{(2,0)} > t) - \left(1 - \frac{1}{N}\right)P_0(T_{(1,1,0)} > t)\right).$$

Also, some algebraic manipulations yield

$$\sum_{i=1}^{D}\sum_{j,l\neq i}E_0[X_j(t)X_l(t)]$$
$$= \frac{D(D-1)(D-2)}{ND}(1 - P_0(T_{(2,0)} > t))$$
$$+ \frac{D(D-1)}{ND}\left(1 - \left(1 - \frac{1}{N}\right)P_0(T_{(0,1)} > t)\right),$$

$$\sum_{i=1}^{D} \sum_{j,k \neq i} E_0[X_i(t)X_j(t)X_k(t)]$$

$$= \frac{D(D-1)(D-2)}{ND}(1 - P_0(T_{(3,0,0)} > t)) + \frac{D(D-1)}{ND}$$

$$\times \left(1 - \frac{1}{N}P_0(T_{(2,0)} > t) - \left(1 - \frac{1}{N}\right)P_0(T_{(1,1,0)} > t)\right),$$

$$\sum_{i=1}^{D} \sum_{j,k,l \neq i} E_0[X_j(t)X_k(t)X_l(t)]$$

$$= \frac{D(D-1)(D-2)(D-3)}{ND}(1 - P_0(T_{(3,0,0)} > t))$$

$$+ \frac{3D(D-1)(D-2)}{ND}\left(1 - \frac{1}{N}P_0(T_{(2,0)} > t)\right.$$

$$- \left(1 - \frac{1}{N}\right)P_0(T_{(1,1,0)} > t)\right)$$

$$+ \frac{D(D-1)}{ND}\left(1 - \frac{3}{N}\left(1 - \frac{1}{N}\right)P_0(T_{(0,1)} > t)\right.$$

$$- \left(1 - \frac{3}{N} + \frac{2}{N^2}\right)P_0(T_{(0,0,1)} > t)\right).$$

Since $\sum_{t \geq 0} P_0(T_{\mathbf{n}} > t) = E_0(T_{\mathbf{n}})$, the above calculations lead to the following expressions:

$$E_0(\beta_1) = \frac{1}{ND}\left(1 - \frac{1}{N}\right)E_0(T_{(0,1)}),$$

$$E_0(\beta_2) = \frac{1}{ND}E_0(T_{(2,0)}),$$

$$E_0(\beta_3) = \frac{1}{ND(D-1)}\left((D-2)E_0(T_{(2,0)})\right.$$

$$+ \left(1 - \frac{1}{N}\right)E_0(T_{(0,1)})\right),$$

$$E_0(\alpha_1) = \frac{1}{ND}\left(1 - \frac{1}{N}\right)\left(\left(1 - \frac{2}{N}\right)E_0(T_{(0,0,1)})\right.$$

$$- \left(1 - \frac{3}{N}\right)E_0(T_{(0,1)})\right),$$

$$E_0(\alpha_2) = \frac{1}{ND}\left(3\left(1 - \frac{1}{N}\right)E_0(T_{(1,1,0)}) - \left(2 - \frac{3}{N}\right)E_0(T_{(2,0)})\right.$$

$$- \left(1 - \frac{1}{N}\right)E_0(T_{(0,1)})\right),$$

$$E_0(\alpha_3) = \frac{1}{ND(D-1)}\left(3(D-2)E_0(T_{(3,0,0)})\right.$$

$$+ 3\left(1 - \frac{1}{N}\right)E_0(T_{(1,1,0)}) - \left(3D - 4 - \frac{3}{N}\right)E_0(T_{(2,0)})$$

$$- \left(1 - \frac{1}{N}\right)E_0(T_{(0,1)})\right),$$

$$E_0(\alpha_4) = \frac{m^3}{(D-1)^2}\frac{1}{ND}\left((D-2)(D-3)E_0(T_{(3,0,0)})\right.$$

$$+ 3(D-2)\left(1 - \frac{1}{N}\right)E_0(T_{(1,1,0)})$$

$$+ \left(1 - \frac{3}{N} + \frac{2}{N^2}\right)E_0(T_{(0,0,1)})$$

$$- (D-2)\left(D - 1 - \frac{3}{N}\right)E_0(T_{(2,0)})$$

$$- \left(1 - \frac{1}{N}\right)\left(D - 1 - \frac{3}{N}\right)E_0(T_{(0,1)})\right).$$

Finally, we get

$$E_0(\alpha) = (1-m)^3\frac{1}{ND}\left(\left(1 - \frac{3}{N} + \frac{2}{N^2}\right)E_0(T_{(0,0,1)})\right.$$

$$- \left(1 - \frac{1}{N}\right)\left(1 - \frac{3}{N}\right)E_0(T_{(0,1)})\right)$$

$$+ m(1-m)^2\frac{1}{ND}\left(3\left(1 - \frac{1}{N}\right)E_0(T_{(1,1,0)})\right.$$

$$- \left(2 - \frac{3}{N}\right)E_0(T_{(2,0)}) - \left(1 - \frac{1}{N}\right)E_0(T_{(0,1)})\right)$$

$$+ \frac{m^2(1-m)}{(D-1)}\frac{1}{ND}\left(3(D-2)E_0(T_{(3,0,0)})\right.$$

$$+ 3\left(1 - \frac{1}{N}\right)E_0(T_{(1,1,0)})$$

$$- \left(3D - 4 - \frac{3}{N}\right)E_0(T_{(2,0)}) - \left(1 - \frac{1}{N}\right)E_0(T_{(0,1)})\right)$$

$$+ \frac{m^3}{(D-1)^2}\frac{1}{ND}\left((D-2)(D-3)E_0(T_{(3,0,0)})\right.$$

$$+ 3(D-2)\left(1 - \frac{1}{N}\right)E_0(T_{(1,1,0)})$$

$$+ \left(1 - \frac{3}{N} + \frac{2}{N^2}\right)E_0(T_{(0,0,1)})$$

$$- (D-2)\left(D - 1 - \frac{3}{N}\right)E_0(T_{(2,0)})$$

$$- \left(1 - \frac{1}{N}\right)\left(D - 1 - \frac{3}{N}\right)E_0(T_{(0,1)})\right), \tag{12}$$

$$E_0(\beta) = (1-m)^2\frac{1}{ND}\left(1 - \frac{1}{N}\right)E_0(T_{(0,1)})$$

$$+ 2m(1-m)\frac{1}{ND}E_0(T_{(2,0)})$$

$$+ \frac{m^2}{(D-1)}\frac{1}{ND}\left((D-2)E_0(T_{(2,0)})\right.$$

$$+ \left(1 - \frac{1}{N}\right)E_0(T_{(0,1)})\right). \tag{13}$$

Therefore, the fixation coefficient for a beneficial allele can be expressed in terms of expected coalescence times under neutrality for samples of two individuals in agreement with Rousset (2004, p. 98) and the fixation coefficient for an allele coding for a mutant strategy in a linear game in terms

of expected coalescence times under neutrality for samples of two and three individuals. In order to go further, we have to find these expected times with respect to the parameters of the population structure. This is done in the next section using a standard conditioning procedure for Markov chains, first in the most general setting without any assumption on the number of demes, the deme size or the migration rate, and then in some special limit cases.

5. Expected coalescence times

Consider a sample of n individuals described by the vector $\mathbf{n} = (n_1, \ldots, n_d)$. The sample configuration backward in time is modelled as a Markov chain with state space S. Assume that $P_{\mathbf{n}\mathbf{n}'}$ is the probability of transition from state $\mathbf{n}$ to state $\mathbf{n}'$ from one generation to the previous one. The time back to a change in the chain state following coalescence or migration events (notice that several events can occur simultaneously in the general setting) is geometrically distributed with parameter

$$P_{\mathbf{n}} = \sum_{\mathbf{n}' \neq \mathbf{n}} P_{\mathbf{n}\mathbf{n}'}. \tag{14}$$

Moreover the probability that the change is a transition from state $\mathbf{n}$ to state $\mathbf{n}' \neq \mathbf{n}$ is

$$Q_{\mathbf{n}\mathbf{n}'} = P_{\mathbf{n}\mathbf{n}'}/P_{\mathbf{n}}. \tag{15}$$

Thus, conditioning on the first change in the chain state, the expectation of the coalescence time $T_{\mathbf{n}}$, defined as the number of generations backward in time from state $\mathbf{n}$ to state (1), which corresponds to the MRCA, is given by the recurrence equation

$$E_0(T_{\mathbf{n}}) = \frac{1}{P_{\mathbf{n}}} + \sum_{\mathbf{n}' \neq \mathbf{n}} Q_{\mathbf{n}\mathbf{n}'} E_0(T_{\mathbf{n}'}).$$

Measuring time in units of ND generations, the coalescence time from state $\mathbf{n}$ to the MRCA becomes

$$t_{\mathbf{n}} = \frac{T_{\mathbf{n}}}{ND},$$

and its expected value satisfies

$$E_0(t_{\mathbf{n}}) = \frac{1}{\lambda_{\mathbf{n}}} + \sum_{\mathbf{n}' \neq \mathbf{n}} Q_{\mathbf{n}\mathbf{n}'} E_0(t_{\mathbf{n}'}),$$

where

$$\lambda_{\mathbf{n}} = NDP_{\mathbf{n}}. \tag{16}$$

For samples of two individuals, we get the system of equations

$$(\text{I}) \begin{cases} E_0(t_{(0,1)}) = \dfrac{1}{\lambda_{(0,1)}} + Q_{(0,1)(2,0)} E_0(t_{(2,0)}), \\[2mm] E_0(t_{(2,0)}) = \dfrac{1}{\lambda_{(2,0)}} + Q_{(2,0)(0,1)} E_0(t_{(0,1)}), \end{cases}$$

while for samples of three individuals, we have the system

$$(\text{II}) \begin{cases} E_0(t_{(0,0,1)}) = \dfrac{1}{\lambda_{(0,0,1)}} + Q_{(0,0,1)(0,1)} E_0(t_{(0,1)}) \\[1mm] \qquad + Q_{(0,0,1)(2,0)} E_0(t_{(2,0)}) + Q_{(0,0,1)(1,1,0)} E_0(t_{(1,1,0)}) \\[1mm] \qquad + Q_{(0,0,1)(3,0,0)} E_0(t_{(3,0,0)}), \\[2mm] E_0(t_{(1,1,0)}) = \dfrac{1}{\lambda_{(1,1,0)}} + Q_{(1,1,0)(0,1)} E_0(t_{(0,1)}) \\[1mm] \qquad + Q_{(1,1,0)(2,0)} E_0(t_{(2,0)}) + Q_{(1,1,0),(0,0,1)} E_0(t_{(0,0,1)}) \\[1mm] \qquad + Q_{(1,1,0)(3,0,0)} E_0(t_{(3,0,0)}), \\[2mm] E_0(t_{(3,0,0)}) = \dfrac{1}{\lambda_{(3,0,0)}} + Q_{(3,0,0)(0,1)} E_0(t_{(0,1)}) \\[1mm] \qquad + Q_{(3,0,0)(2,0)} E_0(t_{(2,0)}) + Q_{(3,0,0),(0,0,1)} E_0(t_{(0,0,1)}) \\[1mm] \qquad + Q_{(3,0,0)(1,1,0)} E_0(t_{(1,1,0)}). \end{cases}$$

We refer to the Appendix A.1 for the exact expressions of the transition probabilities.

The solution of the system of equations (I) for samples of two individuals is

$$E_0(t_{(0,1)}) = 1 \tag{17}$$

and

$$E_0(t_{(2,0)}) = 1 - \frac{1}{N} + \frac{1}{M(2 - M/N)}, \tag{18}$$

where $M = mND/(D-1)$. This notation will be used throughout.

For samples of three individuals, the solution of the system (II) leads to cumbersome expressions given in the Appendix A.2. Numerical evaluations of $E_0(t_{(0,0,1)})$, $E_0(t_{(1,1,0)})$ and $E_0(t_{(3,0,0)})$ as functions of m and N in the case $D = 5$, $D = 100$ and $D = 1000$ (not shown) indicate that all these expected coalescence times decrease to $4/3$ as m and N increase. Moreover, $E_0(t_{(0,0,1)})$ approaches a finite limit as m and N decrease, while $E_0(t_{(1,1,0)})$ and $E_0(t_{(3,0,0)})$ tend to infinity. We will now focus more closely on limit cases.

5.1. Structured coalescent (SC)

Under this scenario, the migration rate m is of order $1/N$ and simultaneous events of migration or coalescence can be neglected when N is large. In the limit, only one event of migration or coalescence can occur at a time and the system of equations (II) reduces to

$$E_0(t_{(0,0,1)}) = \frac{1}{3(M(D-1)+D)} + \frac{D}{(M(D-1)+D)} E_0(t_{(0,1)})$$
$$\qquad + \frac{M(D-1)}{(M(D-1)+D)} E_0(t_{(0,0,1)}),$$

$$E_0(t_{(1,1,0)}) = \frac{1}{D + 2M(D-1) - M}$$
$$\qquad + \frac{D}{D + 2M(D-1) - M} E_0(t_{(2,0)})$$

$$+ \frac{2M(D-2)}{D+2M(D-1)-M} E_0(t_{(3,0,0)})$$

$$+ \frac{M}{D+2M(D-1)-M} E_0(t_{(0,0,1)}),$$

$$E_0(t_{(3,0,0)}) = \frac{1}{6M} + E_0(t_{(0,0,1)}),$$

where $M = mND/(D-1)$. Letting N go to infinity in (17) and (18), we find $E_0(t_{(0,1)}) = 1$ and $E_0(t_{(2,0)}) = 1 + 1/(2M)$ in agreement with previous authors (see, e.g., Wilkinson-Herbots, 1998, 2003; Bahlo and Griffiths, 2001), and then some algebraic manipulations lead to

$$E_0(t_{(0,0,1)}) = \frac{4}{3} + \frac{1}{6(M+1)} - \frac{1}{6D(M+1)},$$

$$E_0(t_{(1,1,0)}) = \frac{4}{3} + \frac{1}{2M} - \frac{1}{6D(M+1)},$$

$$E_0(t_{(3,0,0)}) = \frac{4}{3}\left(1 + \frac{1}{2M}\right) - \frac{1}{6D(M+1)}.$$

5.2. Many demes (MD)

When the number of demes D goes to infinity under the structured coalescent assumptions, we get

$$E_0(t_{(0,0,1)}) = \frac{4}{3} + \frac{1}{6(M+1)},$$

$$E_0(t_{(1,1,0)}) = \frac{4}{3} + \frac{1}{2M},$$

$$E_0(t_{(3,0,0)}) = \frac{4}{3}\left(1 + \frac{1}{2M}\right),$$

with $E_0(t_{(0,1)}) = 1$ and $E_0(t_{(2,0)}) = 1 + 1/(2M)$, where M takes its limit value, that is, $M = Nm$. Notice that these results can be derived using a separation-of-time-scales argument (Wakeley, 1998; Lessard and Wakeley, 2004).

5.3. Low migration (LM)

Under the low migration scenario, the migration parameter m is of order smaller than $1/N$ so that the product Nm tends to zero as N goes to infinity. In this case, we get the approximations

$$E_0(t_{(0,0,1)}) = \frac{4}{3} + \frac{1}{6}\left(1 - \frac{1}{D}\right), \tag{19}$$

$$E_0(t_{(1,1,0)}) = \frac{1}{2M}, \tag{20}$$

$$E_0(t_{(3,0,0)}) = \frac{2}{3M}, \tag{21}$$

with $E_0(t_{(0,1)}) = 1$ and $E_0(t_{(2,0)}) = 1/(2M)$. These approximations are in agreement with the result of Takahata (1991), who showed more generally that, in the LM limit,

we have

$$E_0(Mt_{\mathbf{n}}) = \left(1 - \frac{1}{d}\right),$$

where d is the number of demes occupied by the sampled individuals. This means that, measuring time in units of $ND/(2M) = (D-1)/(2m)$ generations and letting m go to zero, samples of individuals from the same deme coalesce instantaneously, while samples from d different demes behave like samples of d individuals in a panmictic population of size $(D-1)/(2m)$ as described by Kingman's (1982a, b) coalescent. (See also Slatkin, 1981; Notohara, 2001.)

5.4. Strong migration and high migration (SM and HM)

In the strong (or high) migration limit, the migration rate m is kept fixed (equal to 1, respectively) as the deme size N approaches infinity. In this case, we have $E_0(t_{(0,1)}) = E_0(t_{(2,0)}) = 1$, in agreement with Nagylaki (2000) who proved that the expected coalescence time for samples of two individuals in the SM limit is the same as the expected coalescence time for samples of two individuals in a panmictic population of size ND with ND generations taken as the unit of time as N goes to infinity (see also Notohara, 2000; Bahlo and Griffiths, 2001). The same remains true for samples of three individuals, that is,

$$E_0(t_{(0,0,1)}) = E_0(t_{(3,0,0)}) = E_0(t_{(1,1,0)}) = 4/3.$$

6. Approximations of the fixation probability

We are now in a position to express $E_0(\alpha)$ and $E_0(\beta)$ in terms of the population parameters. Using the expressions of the expected coalescence times for samples of two and three individuals under neutrality presented in the previous section and the Appendix, tedious but straightforward algebraic manipulations yield

$$E_0(\beta) = 1 - \frac{1}{ND} \tag{22}$$

and

$$E_0(\alpha) = \frac{1}{3} + \frac{\gamma}{\delta}, \tag{23}$$

where

$$\begin{aligned}
\gamma = {} & 7DNm - Nm^4D^5 - 24m^2ND^3 - 10m^2ND^5 + 48mND^3 \\
& + (2/3)Nm^4D^4 - 2Dm - 8D^2 - 30mND^2 + 7m^2ND^2 \\
& + 2D + 3N + 13mD^2 - (14/3)m^2D^2 - 15DN - 27mD^3 \\
& + 9m^2D^5 + 15D^4N - (2/3)m^4D^4 - 7mD^5 - 3D^5N \\
& - (10/3)m^3D^3 - 5m^3D^5 - 30ND^3 + 55m^2D^3/3 \\
& - 68m^2D^4/3 + 25m^3D^4/3 + 23mD^4 + 30ND^2 + m^4D^5 \\
& - 8D^4 + 2D^5 + 12D^3 + 9mND^5 + (10/3)Nm^3D^3 \\
& + 5Nm^3D^5 - 34mND^4 - 25Nm^3D^4/3 + 27m^2ND^4
\end{aligned}$$

and

$$\begin{aligned}
\delta = (&12 - 81DNm + 78Dm + 72D^2 + 207mND^2 \\
&- 4m^4D^3 - 129m^2ND^2 - 20m^3D^2 - 28m^2D \\
&+ 138mD^3 - 18D^4N + 6m^4D^4 + 50m^3D^3 \\
&+ 72ND^3 - 12m - 136m^2D^3 + 54m^2D^4 \\
&- 30m^3D^4 - 42mD^4 - 108ND^2 - 48D \\
&+ 12D^4 - 48D^3 - 18N - 54m^2N^2D^3 \\
&- 18mD^4N^2 - 65Nm^3D^3 + 63mND^4 \\
&+ 54mN^2D^3 + 45Nm^3D^4 + 27m^2D^4N^2 \\
&- 81m^2ND^4 - 54mN^2D^2 - 15m^3D^4N^2 \\
&+ 186m^2ND^3 - 195mND^3 - 9Nm^4D^4 \\
&+ 15m^3N^2D^3 + 24m^2ND + 20Nm^3D^2 \\
&+ 4Nm^4D^3 + 3N^2m^4D^4 + 18mN^2D \\
&+ 27m^2N^2D^2 - 162mD^2 + 110m^2D^2 \\
&+ 6mN + 72DN)D.
\end{aligned}$$

All theoretical results concerning these expected values along with the expected coalescence times for samples of two and three individuals in the different limit cases (SC, SM, HM, LM, MD) and the general setting are summarized in Table 1.

Notice that $E_0(\beta)$ depends only on the total population size, ND. In particular, it does not depend on the migration rate. Moreover, it increases to 1 as N and D tend to infinity and does so very rapidly so that for D and N as small as 5 its value is comprised between 0.96 and 1.

On the other hand, $E_0(\alpha)$ depends on the population structure via the deme size, N, the number of demes, D, as well as the migration rate, m. Its value ranges from $1/2$ to $1/3$. Moreover, $E_0(\alpha)$ remains dependent on the parameters of the model in four of the limit cases considered, either through D (SC, LM), or m and N (SC, MD). Notice also

that $E_0(\alpha)$ varies little with the number of demes, D, and the deme size, N. Fig. 1(a) shows the relative error, defined as the absolute error over the exact value, using the SC limit as an approximation. For $D = 5$ or 100 and N comprised between 5 and 20, the relative error remains smaller than 10% when m ranges from 0.001 and 1, while for the same values of D and any larger value of N it does not exceed 3%. Hence, the number of demes, D, and the deme size, N, do not need to be very large for the SC limit to give a good approximation of $E_0(\alpha)$. As a consequence, $E_0(\alpha)$ can be approximated by its value in the SC limit, which is $1/3 + (1 - 1/D)/(6(M + 1))$, even for small values of D and N. This is also the case for $E_0(\beta)$ with the limit value 1 as shown in Fig. 1(b).

In the MD limit, $E_0(\alpha)$ is equal to $1/3 + 1/(6(M + 1))$ and numerical evaluations (not shown) indicate that this limit provides a good approximation as soon as the number of demes D exceeds 10, for $N \geqslant 20$ and $0.001 \leqslant m \leqslant 1$, the relative error remaining smaller than 4%. On the other hand, using the LM limit $1/2 - 1/(6D)$ as an approximation, the relative error does not exceed 4% as long as M keeps smaller than $1/10$, for $N \geqslant 5$ and $D \geqslant 5$. Finally, in the SM and HM limits, $E_0(\alpha)$ is equal to $1/3$. This approximation holds with a relative error smaller than 10% as soon as $N \geqslant 30$ and $m \geqslant 0.1$, but less than 5% when M exceeds 10 and $N \geqslant 10$, for both $D = 5$ and $D = 100$.

With the expectations of α and β under neutrality in hand, we can obtain expressions for the fixation coefficient, and therefore approximations for the probability of fixation of a newly introduced single mutant under weak selection.

6.1. Beneficial allele

According to (6), the fixation probability $u(s)$ for a single beneficial mutant as a function of the intensity of selection

Table 1

Expected coalescence times for samples of two and three individuals under the most general setting of D demes of size N and their approximations in limit cases when the deme size N is large: LM for low migration (backward migration rate m much smaller than $1/N$), SC for structured coalescent (m of order $1/N$), SM for strong migration (m fixed), HM for high migration ($m = 1$) and MD for many demes (large number of demes D)

	General setting	LM	SC	SM and HM	MD
$E_0(t_{(0,1)})$	1	1	1	1	1
$E_0(t_{(2,0)})$	$1 - \dfrac{1}{N} + \dfrac{1}{M(2 - M/N)}$	$\dfrac{1}{2M}$	$1 + \dfrac{1}{2M}$	1	$1 + \dfrac{1}{2M}$
$E_0(t_{(0,0,1)})$	Appendix A.2	$\dfrac{3}{2} - \dfrac{1}{6D}$	$\dfrac{4}{3} + \left(1 - \dfrac{1}{D}\right)\dfrac{1}{6(M+1)}$	$\dfrac{4}{3}$	$\dfrac{4}{3} + \dfrac{1}{6(M+1)}$
$E_0(t_{(1,1,0)})$	Appendix A.2	$\dfrac{1}{2M}$	$\dfrac{4}{3} + \dfrac{1}{2M} - \dfrac{1}{6D(M+1)}$	$\dfrac{4}{3}$	$\dfrac{4}{3} + \dfrac{1}{2M}$
$E_0(t_{(3,0,0)})$	Appendix A.2	$\dfrac{2}{3M}$	$\dfrac{4}{3} + \dfrac{2}{3M} - \dfrac{1}{6D(M+1)}$	$\dfrac{4}{3}$	$\dfrac{4}{3} + \dfrac{2}{3M}$
$E_0(\beta)$	$1 - \dfrac{1}{ND}$	1	1	1	1
$E_0(\alpha)$	Eq. (23)	$\dfrac{1}{2} - \dfrac{1}{6D}$	$\dfrac{1}{3} + \dfrac{1}{6(M+1)}\left(1 - \dfrac{1}{D}\right)$	$\dfrac{1}{3}$	$\dfrac{1}{3} + \dfrac{1}{6(M+1)}$

Parameter M is defined generally as $M = mND/(D - 1)$, but takes the value Nm in the MD limit.

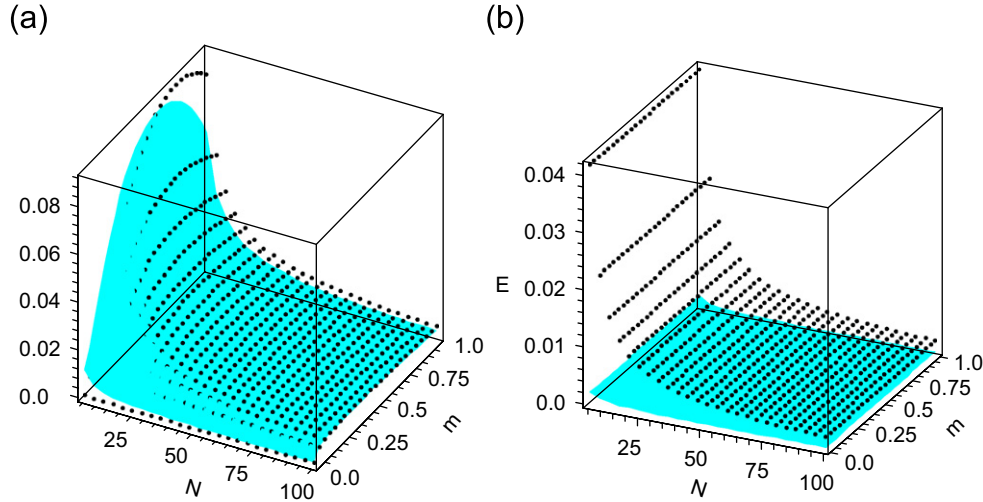(a)   (b)

Fig. 1. Relative error using the SC limit to approximate $E_0(\alpha)$ and $E_0(\beta)$ for $D = 5$ demes (dotted surface) and $D = 100$ demes (solid surface), as functions of the deme size, N, and the migration rate, m. The parameter m ranges from 0.001 to 1, while N ranges from 5 to 100. (a) $E_0(\alpha)$; (b) $E_0(\beta)$.

s has a derivative at $s = 0$ given by $E_0(\beta)$. Therefore, we have

$$u(s) = \frac{1}{ND} + s\left(1 - \frac{1}{ND}\right) + o(s). \tag{24}$$

This approximation under weak selection for D and N fixed is not affected by population subdivision and it is the same for a panmictic population of size ND. This result is in agreement with a diffusion approximation in a framework of a large population (Maruyama, 1970; Whitlock, 2003).

Notice that the intensity of selection s has to be small compared to $1/(ND)$ for the approximation to be valid. Simulation results have shown however that sND does not have to be too small. With $D = 5, N = 10$ and $s = 0.01$, for instance, in which case $sND = 1/2$, the relative error made in approximating $u(s)$, obtained from the frequency of fixation in 10^8 replica, with Eq. (24) did not exceed 5%.

6.2. Linear game

In the context of a linear game, the derivative of the fixation probability with respect to the intensity of selection is given in Eq. (8). Using the expressions (22) and (23) for the expected values $E_0(\beta)$ and $E_0(\alpha)$ in the general setting, we get

$$u(s) = \frac{1}{ND} + s\left\{\left(\frac{1}{3} + \frac{\gamma}{\delta}\right)(\mathbf{p}_A - \mathbf{p}_B) \cdot W(\mathbf{p}_A - \mathbf{p}_B) \right.$$
$$\left. + \left(1 - \frac{1}{ND}\right)(\mathbf{p}_A - \mathbf{p}_B) \cdot W\mathbf{p}_B\right\} + o(s). \tag{25}$$

Using the values of $E_0(\beta)$ and $E_0(\alpha)$ in the different limit cases (see Table 1) and assuming s small compared to $1/(ND)$ when N is large, we have the

approximation

$$u(s) = \frac{1}{ND} + s\left\{\frac{1}{3}(\mathbf{p}_A - \mathbf{p}_B) \cdot W(\mathbf{p}_A - \mathbf{p}_B) \right.$$
$$\left. + (\mathbf{p}_A - \mathbf{p}_B) \cdot W\mathbf{p}_B\right\} + o(s), \tag{26}$$

under strong or high migration (SM or HM),

$$u(s) = \frac{1}{ND} + s\left\{\left(\frac{1}{2} - \frac{1}{6D}\right)(\mathbf{p}_A - \mathbf{p}_B) \cdot W(\mathbf{p}_A - \mathbf{p}_B) \right.$$
$$\left. + (\mathbf{p}_A - \mathbf{p}_B) \cdot W\mathbf{p}_B\right\} + o(s), \tag{27}$$

under low migration (LM),

$$u(s) = \frac{1}{ND} + s\left\{\left(\frac{1}{3} + \frac{1}{6(M+1)}\left(1 - \frac{1}{D}\right)\right)(\mathbf{p}_A - \mathbf{p}_B) \right.$$
$$\left. \cdot W(\mathbf{p}_A - \mathbf{p}_B) + (\mathbf{p}_A - \mathbf{p}_B) \cdot W\mathbf{p}_B\right\} + o(s), \tag{28}$$

under the structured coalescent (SC) assumptions, and

$$u(s) = \frac{1}{ND} + s\left\{\left(\frac{1}{3} + \frac{1}{6(M+1)}\right)(\mathbf{p}_A - \mathbf{p}_B) \cdot W(\mathbf{p}_A - \mathbf{p}_B) \right.$$
$$\left. + (\mathbf{p}_A - \mathbf{p}_B) \cdot W\mathbf{p}_B\right\} + o(s), \tag{29}$$

in the case of many demes (MD). The approximation generally depends on the parameters describing the population structure at neutrality, that is, the deme size, the number of demes and the migration rate. This is the case because the expected value $E_0(\alpha)$ that comes into play in the fixation coefficient as a consequence of the pairwise interactions depends on these parameters. Notice, however, that this is no longer the case in the HM and SM scenarios for which this approximation is the same as the one obtained in a panmictic population of size ND (see Lessard, 2005).

7. Discussion

The probability of fixation of a mutant allele in a population following an island model can be studied using a diffusion approximation if the total population size is large enough (see, e.g., Nagylaki, 1980; Slatkin, 1981). In the case of a large number D of demes containing a fixed number N of haploid individuals with an expected proportion m of individuals coming from all other demes each generation, for instance, Wakeley (2003) (see also Wakeley and Takahashi, 2004, for a selection model of the Moran type, and Cherry, 2003a, for the case of local frequency-dependent selection) showed that the frequency of a beneficial mutant allele A which has fitness $1 + s$ compared to 1 for a wild-type allele B, with s being of order $1/(ND)$, can be approximated by a diffusion process whose drift function is $m(x) = sNDx(1 - x)$ and diffusion function is $v(x) = x(1 - x)$ if time is measured in units of $ND/(1 - F)$ generations, which represents the effective population size, where F is a fixation coefficient under neutrality defined as

$$F = \frac{(1 - m)^2}{Nm(2 - m) + (1 - m)^2}.$$

This is also the case in a diploid population with N individuals in each deme if N is replaced with $2N$, if selection is additive such that the fitnesses of AA, AB and BB are $1 + 2s$, $1 + s$ and 1, respectively, and if migration is gametic. Then, the probability of ultimate fixation of A as a function of s given that it is represented once initially is approximately (see, e.g., Ewens, 2004, and references therein)

$$u(s) = \frac{\int_0^{1/(ND)} \psi(y)\, dy}{\int_0^1 \psi(y)\, dy},$$

where

$$\psi(y) = \exp\left(-2 \int_0^y \frac{m(x)}{v(x)}\, dx\right).$$

This gives

$$u(s) = \frac{1 - \exp(-2s)}{1 - \exp(-2sND)}. \tag{30}$$

If $sND \gg 1$, then the denominator in (30) is close to 1 and a Taylor expansion of the numerator yields

$$u(s) = 2s + o(s). \tag{31}$$

In the case of an additive beneficial allele in a diploid population, this approximation for the probability of fixation is in agreement with Maruyama (1970) who extended a well-known result for a large panmictic population (Haldane, 1927; Fisher, 1922, 1930; see also Kimura, 1957, 1962; Caballero and Hill, 1992, for more general assumptions) to the case of an island model and other cases with conservative migration. See Takahata (1991) and Nagylaki (2000) for limit cases of low migration

and strong migration, respectively, and Whitlock (2003) for more general population structures and genetic assumptions.

On the other hand, if $sND \ll 1$, then a Taylor expansion of both the numerator and the denominator in (30) up to terms of order s^2 yields the approximation

$$u(s) = \frac{1}{ND} + s\left(1 - \frac{1}{ND}\right) + o(s), \tag{32}$$

in agreement with (24). This approximation is valid only when selection is weaker than drift measured by the inverse of the population size, $1/(ND)$. We have deduced this approximation directly without having recourse to a diffusion approximation and we have shown that it is valid not only in a large population but also in a small population for which the assumptions for a diffusion approximation do not hold. This has been made possible by using a coalescent approach suggested by Rousset (2003, 2004). The analysis however relies on the symmetry of the island model and cannot be extended easily to asymmetric models. Let us also stress that letting ND go to infinity in the approximation (32) for the case of selection weaker than drift does not yield the approximation (31) for the case of selection stronger than drift.

In the context of a linear game in a single deme of large size N with a selection intensity s of order $1/N$ such that $sN \ll 1$, a diffusion approximation yields (Lessard, 2005)

$$u(s) = \frac{1}{N} + s\left\{ (\mathbf{p}_A - \mathbf{p}_B) \cdot W\mathbf{p}_B \right.$$
$$\left. + \left(\frac{1}{3}\right)(\mathbf{p}_A - \mathbf{p}_B) \cdot W(\mathbf{p}_A - \mathbf{p}_B) \right\} + O(s/N),$$

for the probability of fixation of a single mutant A. A more precise approximation for any fixed N and s small enough using the same approach as the one presented in this paper, that is, the expected value of the frequency of A and coalescent theory, is (Lessard and Ladret, 2007)

$$u(s) = \frac{1}{N} + s\left(1 - \frac{1}{N}\right)\left\{ (\mathbf{p}_A - \mathbf{p}_B) \cdot W\mathbf{p}_B \right.$$
$$\left. + \left(\frac{N}{3N - 2}\right)(\mathbf{p}_A - \mathbf{p}_B) \cdot W(\mathbf{p}_A - \mathbf{p}_B) \right\} + o(s).$$

See also Imhof and Nowak (2006) for an alternative approach based on a first-step analysis assuming that the fixation probability given any initial frequency is a smooth function of s. In Lessard and Ladret (2007), it is shown that this is the case if the transition probabilities for the frequency of A from one generation to the next are smooth functions of s as in the model at hand. Notice that the fixation probability can be found explicitly, and approximated for s small enough, in the case of a Moran model (see Nowak et al., 2004; Lessard, 2005). A general branching process approach based on stochastic calculus has also been proposed recently (Lambert, 2006).

In the case of D demes, $D \geqslant 3$, with a fixed deme size N, we have shown (see Eq. (25)) that the coefficient of

$s(\mathbf{p}_A - \mathbf{p}_B) \cdot W\mathbf{p}_B$, given by $E_0(\beta) = (1 - (1/ND))$, depends only on the total population size, ND, while the coefficient of $s(\mathbf{p}_A - \mathbf{p}_B) \cdot W(\mathbf{p}_A - \mathbf{p}_B)$, given by $E_0(\alpha)$, generally depends on all the parameters of the population structure, namely, the number of demes, the deme size and the migration rate. In the limit as N goes to infinity, it reduces to $1/3$ when Nm tends infinity (SM or HM), $1/2 - 1/(6D)$ when Nm tends to 0 (LM), $1/3 + (1 - 1/D)/(6(M + 1))$ when $mND/(D - 1)$ tends to M (SC) and $1/3 + 1/(6(M + 1))$ when Nm tends to M and D tends to infinity (MD). Only in the strong or high migration scenario does this coefficient reduce to $1/3$.

As in Nowak et al. (2004), selection favors A replacing B if the fixation probability for a single mutant A is larger than the fixation probability for a single neutral mutant, that is, $u(s) > 1/(ND)$. On the contrary, selection opposes A replacing B if $u(s) < 1/(ND)$. Since the coefficients $E_0(\alpha)$ and $E_0(\beta)$ are positive we find, as in Lessard (2005), that selection opposes A replacing B for selection weak enough and for $\mathbf{p}_A$ different from, but close enough to, $\mathbf{p}_B$ if and only if either

(i) $(\mathbf{p}_A - \mathbf{p}_B) \cdot W\mathbf{p}_B \leqslant 0$,

or, in case of equality in (i),

(ii) $(\mathbf{p}_A - \mathbf{p}_B) \cdot W(\mathbf{p}_A - \mathbf{p}_B) < 0$.

If this condition holds for every $\mathbf{p}_A \neq \mathbf{p}_B$ close enough to $\mathbf{p}_B$, then it holds for every $\mathbf{p}_A \neq \mathbf{p}_B$. This means that $\mathbf{p}_B$ is an evolutionarily stable strategy (ESS) for the game matrix W (Maynard Smith and Price, 1973; Maynard Smith, 1974). This does not mean that $u(s) < 1/(ND)$ for selection weak enough and for all $\mathbf{p}_A \neq \mathbf{p}_B$ if $\mathbf{p}_B$ is an ESS. It will be the case if $\mathbf{p}_B$ has all positive components, which ensures an equality in (i) for all $\mathbf{p}_A \neq \mathbf{p}_B$. But, if $\mathbf{p}_B$ has some null components, then a strict inequality in (i) is possible for $\mathbf{p}_A \neq \mathbf{p}_B$ that has at least one positive component corresponding to a null component of $\mathbf{p}_B$. Then, if the inequality (ii) is reversed, $\mathbf{p}_A$ has to be close enough to $\mathbf{p}_B$ to ensure that

$$E_0(\alpha)(\mathbf{p}_A - \mathbf{p}_B) \cdot W(\mathbf{p}_A - \mathbf{p}_B)$$
$$+ E_0(\beta)(\mathbf{p}_A - \mathbf{p}_B) \cdot W\mathbf{p}_B < 0. \tag{33}$$

How close $\mathbf{p}_A$ has to be to $\mathbf{p}_B$ depends on the coefficients $E_0(\alpha)$ and $E_0(\beta)$.

In the case of two pure strategies $\mathbf{p}_A = (1, 0)$ and $\mathbf{p}_B = (0, 1)$ for a 2×2 game matrix

$$W = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

with $a > c$ and $d > b$, which means that $\mathbf{p}_A$ and $\mathbf{p}_B$ are the best replies to themselves, selection favors A replacing B for selection weak enough if the inequality (33) is reversed, which is equivalent to

$$\frac{d - b}{a - b - c + d} < \frac{E_0(\alpha)}{E_0(\beta)}.$$

The left-hand member of this inequality corresponds to the unstable equilibrium frequency of $\mathbf{p}_A$ for the replicator dynamics in an infinite population, while the right-hand member reduces to $1/3$ in the strong or high migration limit as in a large panmictic population. This corresponds to the one-third law proposed by Nowak et al. (2004). In general, in a population subdivided into D demes according to the finite island model, the ratio $E_0(\alpha)/E_0(\beta)$ is different from $1/3$ but close to $1/3$ unless the migration rate is very small, in which case it gets closer to $1/2$. In this case, the condition for A replacing B becomes less stringent.

The method based on expected coalescence times presented in this paper can be applied under other assumptions on the life cycle or the timing of the different factors as selection preceding dispersal instead of following dispersal. It can also be extended to population structures with slightly more general demographic assumptions, namely subpopulations of equal size with symmetric migration rates as occurs in the circular stepping-stone model and other spatially homogeneous models (see, e.g., Rousset, 2004). Besides weak selection, the main property used is that the probability under neutrality for a common ancestor of individuals chosen at random in subpopulation i to be in subpopulation j is the same if i and j are permuted. In more general settings than the finite island model, however, expected coalescence times are more difficult to compute.

Appendix A

A.1. Transition probabilities

$$P_{(0,1)(1)} = (1 - m)^2 \frac{1}{N} + \frac{m^2}{D - 1} \frac{1}{N},$$

$$P_{(0,1)(2,0)} = 2m(1 - m) + m^2 \left(1 - \frac{1}{D - 1}\right),$$

$$P_{(2,0)(1)} = m\left(1 - \frac{1}{D - 1}\right) \frac{m}{D - 1} \frac{1}{N} + 2(1 - m)\frac{m}{D - 1}\frac{1}{N},$$

$$P_{(2,0)(0,1)} = 2(1-m)\frac{m}{D-1}\left(1-\frac{1}{N}\right) + m\left(1-\frac{1}{D-1}\right)\frac{m}{D-1}\left(1-\frac{1}{N}\right),$$

$$P_{(0,0,1)(1)} = (1-m)^3\frac{1}{N^2} + m\left(\frac{m}{D-1}\right)^2\frac{1}{N^2},$$

$$P_{(0,0,1)(0,1)} = 3(1-m)^3\frac{1}{N}\left(1-\frac{1}{N}\right) + 3m\left(\frac{m}{D-1}\right)^2\frac{1}{N}\left(1-\frac{1}{N}\right),$$

$$P_{(0,0,1)(2,0)} = 3m(1-m)^2\frac{1}{N} + 3m\frac{m}{D-1}(1-m)\frac{1}{N} + 3m^2\left(1-\frac{1}{D-1}\right)\frac{m}{D-1}\frac{1}{N},$$

$$P_{(0,0,1)(1,1,0)} = 3m(1-m)^2\left(1-\frac{1}{N}\right) + 3(1-m)m\frac{m}{D-1}\left(1-\frac{1}{N}\right) + 3m^2\left(1-\frac{1}{D-1}\right)\frac{m}{D-1}\left(1-\frac{1}{N}\right),$$

$$P_{(0,0,1)(3,0,0)} = 3m^2(1-m)\left(1-\frac{1}{D-1}\right) + m^3\left(1-\frac{1}{D-1}\right)\left(1-\frac{2}{D-1}\right),$$

$$P_{(1,1,0)(1)} = (1-m)^2\frac{m}{D-1}\frac{1}{N^2} + (1-m)\left(\frac{m}{D-1}\right)^2\frac{1}{N^2} + m\left(1-\frac{1}{D-1}\right)\left(\frac{m}{D-1}\right)^2\frac{1}{N^2},$$

$$P_{(1,1,0)(0,1)} = 3(1-m)^2\frac{m}{D-1}\frac{1}{N}\left(1-\frac{1}{N}\right) + 3(1-m)\left(\frac{m}{D-1}\right)^2\frac{1}{N}\left(1-\frac{1}{N}\right) + 3m\left(1-\frac{1}{D-1}\right)\left(\frac{m}{D-1}\right)^2\frac{1}{N}\left(1-\frac{1}{N}\right),$$

$$\begin{aligned}
P_{(1,1,0)(2,0)} =\ & (1-m)^3\frac{1}{N} + (1-m)^2\frac{1}{N}m\left(1-\frac{1}{D-1}\right) + 2(1-m)^2\frac{m}{D-1}\frac{1}{N} \\
& + (1-m)\left(1-\frac{1}{D-1}\right)m\frac{m}{D-1}\frac{1}{N} + 2m\frac{m}{D-1}(1-m)\left(1-\frac{1}{D-1}\right)\frac{1}{N} + 2(1-m)m\left(1-\frac{1}{D-1}\right)\frac{m}{D-1}\frac{1}{N} \\
& + 2\frac{m}{D-1}(1-m)m\frac{1}{N} + 2m\left(1-\frac{1}{D-1}\right)\frac{m}{D-1}m\left(1-\frac{1}{D-1}\right)\frac{1}{N} \\
& + m\left(\frac{m}{D-1}\right)^2\frac{1}{N} + m^2\frac{m}{D-1}\left(1-\frac{1}{D-1}\right)^2\frac{1}{N},
\end{aligned}$$

$$\begin{aligned}
P_{(1,1,0)(0,0,1)} =\ & (1-m)^2\frac{m}{D-1}\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right) + (1-m)\left(\frac{m}{D-1}\right)^2\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right) \\
& + m\left(1-\frac{1}{D-1}\right)\left(\frac{m}{D-1}\right)^2\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right),
\end{aligned}$$

$$\begin{aligned}
P_{(1,1,0)(3,0,0)} =\ & 2(1-m)^2m\left(1-\frac{1}{D-1}\right) + (1-m)m^2\left(1-\frac{1}{D-1}\right)\left(1-\frac{2}{D-1}\right) \\
& + 2(1-m)m\left(1-\frac{1}{D-1}\right)m\left(1-\frac{1}{D-1}\right) + \frac{m^3}{D-1}\left(1-\frac{1}{D-1}\right) + m^3\left(1-\frac{1}{D-1}\right)^2\left(1-\frac{2}{D-1}\right),
\end{aligned}$$

$$P_{(3,0,0)(1)} = 3(1-m)\left(\frac{m}{D-1}\right)^2\frac{1}{N^2} + m\left(1-\frac{2}{D-1}\right)\left(\frac{m}{D-1}\right)^2\frac{1}{N^2},$$

$$P_{(3,0,0)(0,1)} = 3(1-m)\left(\frac{m}{D-1}\right)^2\frac{3}{N}\left(1-\frac{1}{N}\right) + m\left(\frac{m}{D-1}\right)^2\left(1-\frac{2}{D-1}\right)\frac{3}{N}\left(1-\frac{1}{N}\right),$$

$$\begin{aligned}
P_{(3,0,0)(2,0)} =\ & \left(6(1-m)^2\frac{m}{D-1} + 6m\frac{m}{D-1}(1-m)\left(1-\frac{1}{D-1}\right) + 3m\frac{m}{D-1}\right. \\
& \left.\times(1-m)\left(1-\frac{2}{D-1}\right) + 3\frac{m^3}{D-1}\left(1-\frac{2}{D-1}\right)\left(1-\frac{1}{D-1}\right) + 3\frac{m^3}{(D-1)^2}\right)\frac{1}{N},
\end{aligned}$$

$$P_{(3,0,0)(0,0,1)} = 3(1-m)\left(\frac{m}{D-1}\right)^2\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right) + m\left(\frac{m}{D-1}\right)^2\left(1-\frac{2}{D-1}\right)\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right),$$

$$P_{(3,0,0)(1,1,0)} = \left(6(1-m)^2\frac{m}{D-1} + 6m\frac{m}{D-1}(1-m)\left(1-\frac{1}{D-1}\right) + 3m\frac{m}{D-1}\right.$$
$$\left. \times(1-m)\left(1-\frac{2}{D-1}\right) + 3\frac{m^3}{D-1}\left(1-\frac{2}{D-1}\right)\left(1-\frac{1}{D-1}\right) + 3\frac{m^3}{(D-1)^2}\right)\left(1-\frac{1}{N}\right).$$

A.2. Expected coalescence times for samples of three individuals

From the previous exact transition probabilities (Appendix A.1), together with (15) and (16), we get the exact expressions for $Q_{\mathbf{n},\mathbf{n'}}$ and $\lambda_{\mathbf{n}}$ for all vectors $\mathbf{n}, \mathbf{n'}$ corresponding to samples of individuals of size two and three in the general case (without neglecting simultaneous coalescence and/or migration events). Using Maple, we can solve the systems of linear equations (I) and (II) and derive the exact expected coalescence times for samples of two and three individuals. We find:

$$E_0(t_{(0,0,1)}) = \frac{4}{3} + \frac{A_1}{B_1},$$

where

$$A_1 = (-m^4 D^5 + 5m^3 D^5 - 10m^2 D^5 + 9mD^5 - 3D^5 + (2/3)m^4 D^4 - (25/3)m^3 D^4 + 27m^2 D^4 - 34mD^4$$
$$+ 15D^4 + (10/3)m^3 D^3 - 24m^2 D^3 + 48mD^3 - 30D^3 + 7m^2 D^2 - 30mD^2 + 30D^2 + 7mD - 15D + 3)N$$
$$+ (1/3)D(6 + 69mD^3 - 15m^3 D^4 - 21mD^4 - 10m^3 D^2 - 6m + 39mD - 14m^2 D + 27m^2 D^4 - 24D - 2m^4 D^3$$
$$+ 25m^3 D^3 - 68m^2 D^3 - 81mD^2 + 55m^2 D^2 + 3m^4 D^4 - 24D^3 + 36D^2 + 6D^4),$$

$$B_1 = D(-15m^3 D^4 N^2 + 15m^3 D^3 N^2 - 54mD^2 N^2 - 54m^2 D^3 N^2 + 27m^2 D^2 N^2$$
$$+ 3m^4 D^4 N^2 + 27m^2 D^4 N^2 - 18mD^4 N^2 + 18mDN^2 + 54mD^3 N^2 + 20m^3 D^2 N$$
$$+ 207mD^2 N + 63mD^4 N - 65m^3 D^3 N - 9m^4 D^4 N + 4m^4 D^3 N - 18N - 129m^2 D^2 N - 108D^2 N$$
$$- 81mDN - 18D^4 N + 24m^2 DN - 195mD^3 N + 186m^2 D^3 N + 72D^3 N - 81m^2 D^4 N + 6mN + 72DN$$
$$+ 45m^3 D^4 N + 50m^3 D^3 + 72D^2 - 20m^3 D^2 + 138mD^3 + 110m^2 D^2 - 12m - 30m^3 D^4 + 54m^2 D^4$$
$$- 48D^3 - 42mD^4 + 12 - 4m^4 D^3 - 48D + 78mD - 28m^2 D - 162mD^2 + 12D^4 - 136m^2 D^3 + 6m^4 D^4).$$

$$E_0(t_{(1,1,0)}) = \frac{4}{3} + \frac{A_2}{B_2},$$

where

$$A_2 = -12 - 358m^2 D^2 - 444m^3 D^3 + 88m^3 D^2 - 780mD^3 + 912m^2 D^3 + 240D^3 - 180D^2$$
$$+ 66D^6 m - 120m^4 D^6 - 6m^6 D^6 - 150D^6 m^2 + 4m^6 D^5 + 28m^5 D^4 - 12D^6 + 42m^5 D^6$$
$$- 70m^5 D^5 + 180D^6 m^3 + 72D + 12m + 312m^4 D^5 + (18 + 425m^2 D^2 + 550m^3 D^3$$
$$- 280m^3 D^2/3 + 1090mD^3 - 1200m^2 D^3 - 360D^3 + 270D^2 - 99D^6 m + 180m^4 D^6 + 9m^6 D^6$$
$$+ 225D^6 m^2 - (14/3)m^6 D^5 - 98m^5 D^4/3 + 18D^6 - 63m^5 D^6 + 287m^5 D^5/3 - 270D^6 m^3 - 108D$$
$$- 10m - 1324m^4 D^5/3 + 270D^4 - 108D^5 - 950m^2 D^5 + 1028m^4 D^4/3 - 1090m^3 D^4 + 1550m^2 D^4 + 505mD^5$$
$$- 1040mD^4 + 2710m^3 D^5/3 - 244m^4 D^3/3 + 149mD - 50m^2 D - 595mD^2)N - 180D^4 + 72D^5$$
$$+ 652m^2 D^5 - 264m^4 D^4 + 804m^3 D^4 - 1108m^2 D^4 - 342mD^5 + 720mD^4 - 628m^3 D^5 + 72m^4 D^3$$
$$- 126mD + 52m^2 D + 450mD^2 - (1/3)(-18 - 63m^2 D^2 - 41m^3 D^3 - 1062mD^3 + 450m^2 D^3$$
$$+ 720D^3 - 450D^2 + 189D^6 m - 63m^4 D^6 - 261D^6 m^2 - 54D^6 + 9m^5 D^6 - 2m^5 D^5$$
$$+ 180D^6 m^3 + 144D + 77m^4 D^5 - 630D^4 + 288D^5 + 846m^2 D^5 - 14m^4 D^4 + 262m^3 D^4 - 972m^2 D^4$$
$$- 807mD^5 + 1338mD^4 - 401m^3 D^5 - 51mD + 393mD^2)mN^2,$$

$$B_2 = Nm(-2D + 2 + mD)D(-15m^3D^4N^2 + 15m^3D^3N^2 - 54mD^2N^2 - 54m^2D^3N^2$$
$$+ 27m^2D^2N^2 + 3m^4D^4N^2 + 27m^2D^4N^2 - 18mD^4N^2 + 18mDN^2 + 54mD^3N^2 + 20m^3D^2N$$
$$+ 207mD^2N + 63mD^4N - 65m^3D^3N - 9m^4D^4N + 4m^4D^3N - 18N - 129m^2D^2N - 108D^2N$$
$$- 81mDN - 18D^4N + 24m^2DN - 195mD^3N + 186m^2D^3N + 72D^3N - 81m^2D^4N + 6mN$$
$$+ 72DN + 45m^3D^4N + 50m^3D^3 + 72D^2 - 20m^3D^2 + 138mD^3 + 110m^2D^2 - 12m - 30m^3D^4$$
$$+ 54m^2D^4 - 48D^3 - 42mD^4 + 12 - 4m^4D^3 - 48D + 78mD - 28m^2D - 162mD^2 + 12D^4 - 136m^2D^3 + 6m^4D^4).$$

$$E_0(t_{(3,0,0)}) = \frac{4}{3} + \frac{A_3}{B_3},$$

where

$$A_3 = -16 - 512m^2D^2 - 636m^3D^3 + 132m^3D^2 - 1060mD^3 + 1268m^2D^3 + 320D^3 - 240D^2$$
$$+ 88D^6m - 160m^4D^6 - 8m^6D^6 - 200D^6m^2 + 6m^6D^5 + 42m^5D^4 - 16D^6 + 56m^5D^6$$
$$- 98m^5D^5 + 240D^6m^3 + 96D + 18m + 428m^4D^5 - 240D^4 + 96D^5 + 878m^2D^5 - 376m^4D^4$$
$$+ 1116m^3D^4 - 1512m^2D^4 - 458mD^5 + 970mD^4 - 852m^3D^5 + 108m^4D^3 - 178mD + 78m^2D$$
$$+ 620mD^2 - (1/3)(-18 - 63m^2D^2 - 41m^3D^3 - 1314mD^3 + 537m^2D^3 + 900D^3$$
$$- 540D^2 + 252D^6m - 84m^4D^6 - 348D^6m^2 - 72D^6 + 12m^5D^6 - 2m^5D^5 + 240D^6m^3$$
$$+ 162D + 98m^4D^5 - 810D^4 + 378D^5 + 1107m^2D^5 - 14m^4D^4 + 322m^3D^4 - 1233m^2D^4 - 1059mD^5$$
$$+ 1716mD^4 - 521m^3D^5 - 51mD + 456mD^2)mN^2 + (24 + 592m^2D^2 + 763m^3D^3 - 403m^3D^2/3$$
$$+ 1470mD^3 - 1638m^2D^3 - 480D^3 + 360D^2 - 132D^6m + 240m^4D^6 + 12m^6D^6 + 300D^6m^2$$
$$- 20m^6D^5/3 - 140m^5D^4/3 + 24D^6 - 84m^5D^6 + 392m^5D^5/3 - 360D^6m^3 - 144D - 15m$$
$$- 1789m^4D^5/3 + 360D^4 - 144D^5 - 1273m^2D^5 + 1418m^4D^4/3 - 1483m^3D^4 + 2092m^2D^4 + 675mD^5$$
$$- 1395mD^4 + 3643m^3D^5/3 - 349m^4D^3/3 + 207mD - 73m^2D - 810mD^2)N,$$

$$B_3 = Nm(-2D + 2 + mD)D(-15m^3D^4N^2 + 15m^3D^3N^2 - 54mD^2N^2 - 54m^2D^3N^2$$
$$+ 27m^2D^2N^2 + 3m^4D^4N^2 + 27m^2D^4N^2 - 18mD^4N^2 + 18mDN^2 + 54mD^3N^2 + 20m^3D^2N$$
$$+ 207mD^2N + 63mD^4N - 65m^3D^3N - 9m^4D^4N + 4m^4D^3N - 18N - 129m^2D^2N - 108D^2N$$
$$- 81mDN - 18D^4N + 24m^2DN - 195mD^3N + 186m^2D^3N + 72D^3N - 81m^2D^4N + 6mN$$
$$+ 72DN + 45m^3D^4N + 50m^3D^3 + 72D^2 - 20m^3D^2 + 138mD^3 + 110m^2D^2 - 12m - 30m^3D^4$$
$$+ 54m^2D^4 - 48D^3 - 42mD^4 + 12 - 4m^4D^3 - 48D + 78mD - 28m^2D - 162mD^2 + 12D^4 - 136m^2D^3 + 6m^4D^4).$$

References

Abramowitz, M., Stegun, I.A., 1965. Handbook of Mathematical Functions. Dover, New York.

Abrams, P.A., Matsuda, H., 1997. Fitness minimization and dynamic instability as a consequence of predator-prey coevolution. Evol. Ecol. 11, 1–20 (reprinted with corrections from 1996, Evol. Ecol. 10, 167–186).

Axelrod, R., Hamilton, W.D., 1981. The evolution of cooperation. Science 211, 1390–1396.

Bahlo, M., Griffiths, R.C., 2001. Coalescence time for two genes from a subdivided population. J. Math. Biol. 43, 397–410.

Barton, N.H., 1993. The probability of fixation of a favoured allele in a subdivided population. Genet. Res. 62, 149–158.

Caballero, A., Hill, W.G., 1992. Effects of partial inbreeding on fixation rates and variation of mutant genes. Genetics 131, 493–507.

Cherry, J.L., 2003a. Selection in a subdivided population with dominance or local frequency dependence. Genetics 163, 1511–1518.

Cherry, J.L., 2003b. Selection in a subdivided population with local extinction and recolonization. Genetics 164, 789–795.

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

Ewens, W.J., 2004. Mathematical Population Genetics 1. Theoretical Introduction, Second ed. Springer, New York.

Fisher, R.A., 1922. On the dominance ratio. Proc. R. Soc. Edinb. 42, 321–341.

Fisher, R.A., 1930. The Genetical Theory of Natural Selection. Oxford University Press, Oxford.

Godfray, H.C.J., 1995. Evolutionary theory of parent-offspring conflict. Nature 376, 133–138.

Haldane, J.B.S., 1927. A mathematical theory of natural and artificial selection. V. Selection and mutation. Proc. Camb. Philos. Soc. 23, 838–844.

Hamilton, W.D., 1967. Extraordinary sex ratios. Science 156, 477–488.

Hardy, I.C.W., 2002. Mother Knows Best: Sons or Daughters? A Review of Sex Ratios. Concepts and Methods. Cambridge University Press, Cambridge.

Herbots, H.M., 1994. Stochastic models in population genetics: genealogy and genetic differentiation in structured populations. Ph.D. Thesis, University of London, London.

Hofbauer, J., Sigmund, K., 1998. Evolutionary Games and Population Dynamics. Cambridge University Press, Cambridge.

Imhof, L.A., Nowak, M.A., 2006. Evolutionary game dynamics in a Wright–Fisher process. J. Math. Biol. 52, 667–681.

Kerr, B., Riley, M., Feldman, M., Bohannan, B., 2002. Local dispersal and interaction promote coexistence in a real life game of rock-paper-scissors. Nature 418, 171–174.

Kimura, M., 1957. Some problems of stochastic processes in genetics. Ann. Math. 28, 882–901.

Kimura, M., 1962. On the probability of fixation of mutant genes in a population. Genetics 47, 713–719.

Kingman, J.F.C., 1982a. The coalescent. Stochast. Proc. Appl. 13, 235–248.

Kingman, J.F.C., 1982b. On the genealogy of large populations. J. Appl. Prob. 19A, 27–43.

Lambert, A., 2006. Probability of fixation under weak selection: a branching process unifying approach. Theor. Popul. Biol. 69, 419–441.

Lessard, S., 2005. Long-term stability from fixation probabilities in finite populations: new perspectives for ESS theory. Theor. Popul. Biol. 68, 19–27.

Lessard, S., Ladret, V., 2007. The probability of fixation of a single mutant in an exchangeable selection model. J. Math. Biol. 54, 721–744.

Lessard, S., Wakeley, J., 2004. The two-locus ancestral graph in a subdivided population: convergence as the number of demes grows in the island model. J. Math. Biol. 48, 275–292.

Maruyama, T., 1970. Effective number of alleles in a subdivided population. Theor. Popul. Biol. 1, 273–306.

Maruyama, T., 1974. A simple proof that certain quantities are independent of the geographical structure of the population. Theor. Popul. Biol. 5, 148–154.

Maynard Smith, J., 1974. The theory of games and the evolution of animal conflicts. J. Theor. Biol. 47, 209–221.

Maynard Smith, J., 1982. Evolution and the Theory of Games. Cambridge University Press, Cambridge.

Maynard Smith, J., Price, G.R., 1973. The logic of animal conflict. Nature 246, 15–18.

Nagylaki, T., 1980. The strong-migration limit in geographically structured populations. J. Math. Biol. 9, 101–114.

Nagylaki, T., 1982. Geographical invariance in population genetics. J. Theor. Biol. 99, 159–172.

Nagylaki, T., 2000. Geographical invariance and the strong-migration limit in subdivided populations. J. Math. Biol. 41, 123–142.

Notohara, M., 1990. The coalescent and the genalogical process in geographically structured populations. J. Math. Biol. 29, 59–75.

Notohara, M., 1993. The strong-migration limit for the genealogical process in geographically structured populations. J. Math. Biol. 31, 115–122.

Notohara, M., 2000. A perturbation method for the structured coalescent with strong migration. J. Appl. Prob. 37, 148–167.

Notohara, M., 2001. The structured coalescent process with weak migration. J. Appl. Prob. 38, 1–17.

Nowak, M.A., Sigmund, K., 2004. Evolutionary dynamics of biological games. Science 303, 793–799.

Nowak, M.A., Sasaki, A., Taylor, C., Fudenberg, D., 2004. Emergence of cooperation and evolutionary stability in finite populations. Nature 428, 646–650.

Orzack, S.H., Hines, W.G.S., 2005. The evolution of strategy variation: will an ESS evolve? Evolution 59, 1183–1193.

Rousset, F., 2003. A minimal derivation of convergence stability measures. J. Theor. Biol. 221, 665–668.

Rousset, F., 2004. Genetic Structure and Selection in Subdivided Populations. Princeton University Press, Princeton.

Rousset, F., Billiard, S., 2000. A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. J. Evol. Biol. 13, 814–825.

Roze, D., Rousset, F., 2003. Selection and drift in subdivided populations: a straightforward method for deriving diffusion approximations and examples involving dominance, selfing, and local extinctions. Genetics 165, 2153–2166.

Roze, D., Rousset, F., 2004. The robustness of Hamilton's rule with inbreeding and dominance: kin selection and fixation probabilities under partial sib mating. Am. Nat. 164, 214–231.

Sinervo, B., Lively, C.M., 1996. The rock-paper-scissors game and the evolution of alternative male reproductive strategies. Nature 380, 240–243.

Slatkin, M., 1981. Fixation probabilities and fixation times in a subdivided population. Evolution 35, 477–488.

Sugden, R., 1986. The Economics of Rights, Co-operation and Welfare. Blackwell, Oxford.

Takahata, N., 1991. Genealogy of neutral genes and spreading of selected mutations in a geographically structured population. Genetics 129, 585–595.

Taylor, P.D., Jonker, L.B., 1978. Evolutionarily stable strategies and game dynamics. Math. Biosci. 40, 145–156.

Taylor, P.D., Day, T., Nagy, D., Wild, G., André, J.B., Gardner, A., 2006. The evolutionary consequences of plasticity in host-pathogen interactions. Theor. Popul. Biol. 69, 323–331.

Turner, P.E., Chao, L., 1999. Prisoner's dilemna in an RNA virus. Nature 398, 441–443.

Wakeley, J., 1998. Segregating sites in Wright's island model. Theor. Popul. Biol. 53, 166–175.

Wakeley, J., 2003. Polymorphism and divergence for island model species. Genetics 163, 411–420.

Wakeley, J., 2004. Recent trends in population genetics: more data! more math! simple models? J. Hered. 95, 397–405.

Wakeley, J., Lessard, S., 2003. Theory of the effects of population structure and sampling on patterns of linkage disequilibrium applied to genomic data from humans. Genetics 164, 1043–1053.

Wakeley, J., Takahashi, T., 2004. The many-demes limit for selection and drift in a subdivided population. Theor. Popul. Biol. 66, 83–91.

Whitlock, M.C., 2003. Fixation probability and time in subdivided populations. Genetics 164, 767–779.

Wild, G., Taylor, P.D., 2004. Fitness and evolutionary stability in game theoretic models of finite populations. Proc. R. Soc. B 271, 2345–2349.

Wilkinson-Herbots, H., 1998. Genealogy and subpopulation in differentiation under various models of population structure. J. Math. Biol. 37, 535–585.

Wilkinson-Herbots, H.M., 2003. Coalescence times and FST values in subdivided populations with symmetric structure. Adv. Appl. Prob. 35, 665–690.

Wright, S., 1931. Evolution in Mendelian populations. Genetics 16, 97–159.