Available online at www.sciencedirect.com
SCIENCE DIRECT®
Theoretical Population Biology 66 (2004) 83–91
Theoretical Population Biology
http://www.elsevier.com/locate/ytpbi

The many-demes limit for selection and drift in a subdivided population

John Wakeley* and Tsuyoshi Takahashi

Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA

Received 20 October 2003

Available online 15 July 2004

Abstract

A diffusion approximation is obtained for the frequency of a selected allele in a population comprised of many subpopulations or demes. The form of the diffusion is equivalent to that for an unstructured population, except that it occurs on a longer time scale when migration among demes is restricted. This many-demes diffusion limit relies on the collection of demes always being in statistical equilibrium with respect to migration and drift for a given allele frequency in the total population. Selection is assumed to be weak, in inverse proportion to the number of demes, and the results hold for any deme sizes and migration rates greater than zero. The distribution of allele frequencies among demes is also described.

Keywords: Gene frequency; Diffusion; Migration; Selection; Random genetic drift

1. Introduction

The majority of work in population genetics is based on models which do not include geographic structure. However, some account must be made of structure because the members of a species are always distributed over space and in very few cases are individuals able to easily traverse the entire species range. Thus, biological populations are not well mixed, or panmictic, at least not within a single generation. One of the most surprising results of theoretical population genetics is that it takes only a few migrant individuals per generation between demes (subpopulations) to make a subdivided population effectively indistinguishable from a panmictic one (Wright, 1931; Moran, 1959). The reason for this is that the time scale of change in the population (e.g., of allele frequencies) is long, on the order of the population size number of generations, and over such long times a few migrants per generation is enough to mix the population. Still, it is not difficult to find evidence of geographic structure in samples of genetic data (Slatkin, 1985), and so it will often be necessary to model population structure explicitly.

We consider a simple model of subdivision in which there can be selective differences between two alleles at a genetic locus. Our goal is to establish a diffusion approximation for the allele frequency in the population, in particular for the case where the population is subdivided into many demes. This follows some recent work on a similar model which assumed discrete generations and Wright–Fisher (Fisher, 1930; Wright, 1931) reproduction within demes (Wakeley, 2003). Here we assume instead that generations are overlapping and that reproduction occurs according to a Moran (1958a) model. One motivation for studying a Moran model is that the demonstration of the limiting diffusion result is much simpler than in the Wright–Fisher case. The other motivation is to show that the result does not depend on generations being discrete, so that the diffusion approximation may be applied to organisms with a variety of modes of reproduction.

We begin with a brief comparison of results for the panmictic versions of the Moran and Wright–Fisher models, as this is helpful in comparing our results to those in Wakeley (2003). Under the Moran model it is possible to obtain exact expressions for many quantities

*Corresponding author.
E-mail address: wakeley@fas.harvard.edu (J. Wakeley).

0040-5809/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.tpb.2004.04.005

of evolutionary interest which can only be found approximately under the Wright–Fisher model. Results differ between the two models when the population size N is not large, but the same form of a limiting diffusion holds for both models as N tends to infinity. If there are two alleles at some genetic locus, and the mutant allele has a selective advantage s over the wild-type allele, then the diffusion of the mutant allele frequency x depends on the first two moments of the change in x over one time step:

$$E[X(1)-x] = \frac{s}{N}x(1-x) + O(s^2/N), \quad (1)$$

$$E[(X(1)-x)^2] = \frac{2}{N^2}x(1-x) + O(s/N^2), \quad (2)$$

for the Moran model, and

$$E[X(1)-x] = sx(1-x) + O(s^2), \quad (3)$$

$$E[(X(1)-x)^2] = \frac{1}{N}x(1-x) + O(s/N), \quad (4)$$

for the Wright–Fisher model, e.g., see Ewens (1979) and Moran (1962).

Therefore, the natural diffusion time scale for the Wright–Fisher model is in units of N generations, whereas under the Moran model it is $N^2/2$ time steps. Each time step in the Moran model involves a single birth–death event, and it is usual to make comparisons by equating N time steps in the Moran model to a single Wright–Fisher generation. Interestingly, this still leaves a factor of two difference in the rates of genetic drift, with drift being faster under the Moran model. This difference is due to the different distributions of offspring number per individual in the two models (Moran and Watterson, 1959; Feldman, 1966). Measuring time on the natural scale for each model, and letting N go to infinity, the two models converge to the same diffusion process if the scaled selection coefficient is defined to be $\gamma = Ns$ in the Wright–Fisher model and $\gamma = Ns/2$ in the Moran model.

1.1. Diffusion limits for subdivided populations

The typical diffusion approximation in population genetics is, as above, to consider the limiting process as the population size tends to infinity. This has been extended to subdivided populations by considering that the sizes of demes tend to infinity while the number of demes remains fixed. There have been three kinds of diffusion results for such models, which differ in how they treat the migration rates.

The first, due to Wright (1931), assumes that the migration rate m for a deme, which is the fraction of the deme that is replaced by migrants every generation, scales inversely with the size of the deme. The result is that the effect of migration is described by the product Nm. For multiple demes, the diffusion approximation would require as many dimensions as there are demes, e.g., see Pollack (1968), and few results are available. However, this limit underlies many standard results for subdivided populations, even if they were not described explicitly as results of a diffusion process. It is also the basis of the backward-time genealogical model known as the structured coalescent (Notohara, 1990; Herbots, 1994; Wilkinson-Herbots, 1998). For the case of a single deme that receives migrants from a "continental" source population, in which the allele frequency x is constant for all time, Wright (1931) used this large-N, small-m approximation to obtain a stable distribution for the allele frequency within the deme as a balance between immigration and genetic drift. The same result holds for a single deme in Wright's (1931) island model, which assumes an infinite number of demes. In this case, the source of migrants is the entire population which, because it is infinite, can be assumed to have a constant allele frequency.

The second kind of diffusion result assumes that the migration rates between demes are constant while the deme sizes approach infinity. This is the strong-migration limit of Nagylaki (1980), and the result is that the diffusion becomes identical to the panmictic diffusion, but with an effective population size that depends on the pattern of migration. In the case of conservative migration, which means that the emmigration and immigration balance perfectly for every deme, the effective size is equal to the total size of the population, otherwise it is different. The strong-migration limit makes explicit the notion mentioned above that large populations will become well mixed, even by seemingly small amounts of migration. In the strong-migration limit, a large population means $N \to \infty$ for a finite number of demes. The strong-migration limit has also been studied in genealogical, or coalescent, models (Notohara, 1993), with the result of course being convergence to Kingman's (1982) coalescent, but with a possibly different effective size.

The third kind of diffusion result is the low-migration limit studied by Slatkin (1981). Slatkin (1981) assumed that the deme sizes approach infinity and the migration rates approach zero such that the product Nm also approaches zero. The result is that the fixation or loss of an allele within a deme occurs much more quickly than changes in the total population which are mediated by migration. Slatkin (1981) assumed Wright–Fisher reproduction within demes, but showed that dynamics in the total population occurred by the fixation or loss of alleles in one deme at a time, so that the demes become like the individuals in a Moran model of reproduction. Since this model was formulated for a finite number of demes and changes in allele frequency occur in discrete jumps ($\pm$ one deme), this is not a diffusion in the sense of the two limits above. We note also that the general form of Slatkin's (1981) result should not depend on the

sizes of demes being large. Genealogical versions of the low-migration limit are due to Takahata (1991) and Notohara (2001).

We study a fourth kind of diffusion limit for a subdivided population, which was introduced in a forward-time model by Wakeley (2003), and in a genealogical context without selection by Wakeley (1998). Specifically, we seek a diffusion approximation for the allele frequency in the total population which holds in the limit as the number of demes tends to infinity, with time measured appropriately and with proportionately weak selection. Our results are related to the three diffusion approximations described above in that they rely on a separation of time scales between different processes. First, we apply the same theorem, due to Ethier and Nagylaki (1980), that Nagylaki (1980) used to prove the strong-migration limit. Our results are also similar to Wright's (1931) infinite-island model. However, under the infinite-island model the allele frequency among migrants or in the total population is simply assumed to be constant, while we describe how it changes by a diffusion process. This diffusion occurs on a much longer time scale than changes within demes, so our results are akin to those of Slatkin (1981) except that the source of the difference in time scales is that there are many demes rather than the migration rates being small.

We show below that a many-demes diffusion exists and that it does not depend on the allele frequencies within particular demes, but rather on the equilibrium properties of the ensemble of demes. The fast time scale of migration and drift within demes guarantees that the demes will always be close to statistical equilibrium for a given allele frequency in the total population. The diffusion for the allele frequency in the total population is identical in form to the diffusion approximation for the allele frequency in an unstructured population, the only difference being that the time scale of the process is increased by the type of subdivision considered here.

1.2. A model of subdivision with overlapping generations

We consider a population subdivided into D demes, each of which contain N haploid individuals. We assume here that time occurs in discrete steps, and later show convergence to a continuous time process. However, it should also be possible to begin with a continuous time model, as in Moran (1958b), and obtain similar results. In each time step, a single individual is chosen to die and another, possibly the same individual, is chosen to reproduce. We assume that two alleles are segregating at a single locus, and that the fitness, or viability, of individuals carrying the "mutant" allele differs from that of the individuals carrying the "wild-type" allele. Selection and drift occur within each deme, and there is migration among demes.

The order of events in a single unit of time is as follows. A deme is chosen at random, uniformly among all D demes, then an individual within that deme is chosen to die. Viability selection acts within the deme such that individuals with the mutant allele have relative death rate λ_2 and individuals with the wild-type have relative death rate λ_1. We assume that $\lambda_2 = (1 - s_D)\lambda_1$, so that the selective advantage or disadvantage of the mutant compared to the wild-type is $s_D = 1 - \lambda_2/\lambda_1$. Later, we will consider s_D to be small, in proportion to $1/D$, and this is the reason for the subscript. Once an individual is chosen to die, an individual is chosen to reproduce and the offspring of this individual replaces the one who died. With probability $1 - m$ the individual chosen to reproduce comes from the same deme as the individual chosen to die. In this case, each of the N individuals in the deme is equally likely to be the one chosen to reproduce. With probability m the individual chosen to reproduce is randomly sampled from the entire population. In this case, each of the ND individuals in the population is equally likely to be the one chosen to reproduce. In both cases, it is possible that the same individual is chosen to die and to reproduce.

Selection has been implemented in two different ways in studies of the Moran model: at death as we do above, and at birth (Moran, 1962; Ewens, 1979). A model of selection at birth can be constructed that would correspond closely to our model of selection at death. In it a deme would be chosen randomly, then an individual within that deme would be chosen to reproduce. Selection would act within the deme such that individuals with the mutant allele would have relative fertility $\lambda_2 = (1 + s_D)\lambda_1$, compared to λ_1 for individuals carrying the wild-type allele. With probability $1 - m$ the offspring of this individual would replace an individual selected at random from within the deme, and with probability m it would replace an individual selected at random from the entire population. Despite the apparent change in sign of s_D from the viability selection model, positive s_D means a selective advantage of the mutant allele in both models. A notable similarity between these two models is that selection acts only within demes. Models in which selection acts across the entire population, i.e. among demes as well as within demes, produce slightly different results. We present detailed results only for the model of selection at death described above, but we note some of these differences below.

2. Theory

This section is organized as follows. After a brief introduction of the state space and the major random variables to be considered, Section 2.1 presents the result

and introduces the theorem used to prove it in the subsequent four sections. Section 2.2 contains the description of the fundamental dynamics, which are changes in the number of mutant alleles within a deme, and the derivation of the equilibrium distribution of a single deme among frequency classes assuming a constant frequency of the mutant allele in the total population. In Section 2.3, it is shown that the entire collection of demes approaches a related equilibrium for a constant allele frequency in the total population. Section 2.4 establishes the (fast) time scale of the approach to equilibrium of demes among frequency classes. Section 2.5 establishes the (slow) time scale of changes in allele frequency in the total population, and this completes the derivation. Section 2.6 considers the question of large deme size.

Following Wakeley (2003), we denote the state of the population using a vector-valued random variable $Z^D(\cdot)$ which records the fraction of demes in each allele-frequency class. Specifically, $Z_i^D(t)$ represents the fraction of demes that contain i copies of the mutant allele at generation t. The superscript is in recognition that the possible configurations of $Z^D(\cdot)$ will depend on the number of demes, and that we will be considering the limiting process as $D\to\infty$. We show below that $Z_i^D(t)$ is always sufficiently close to its equilibrium value $v_i(t)$ that the diffusion approximation for the allele frequency in the total population depends on v rather than on Z. The allele frequency in the total population is given by

$$X^D(t)=\frac{1}{N}\sum_{i=0}^{N} iZ_i^D(t), \tag{5}$$

and the value of this random variable in a particular generation is given by $x(t)$. Below we will also need

$$Y_i^D(t)=Z_i^D(t)-v_i(t), \tag{6}$$

which is the deviation of $Z^D(t)$ from its equilibrium value. The equilibrium $v_i(t)$ is a simple function of N, m, and $X^D(t)$. We will use y and z to denote particular values of the random variables Y^D and Z^D.

2.1. The many-demes diffusion result

Ethier and Nagylaki (1980) proved a general limit theorem for diffusion approximations for Markov processes with two time scales which can be applied to $Z^D(t)$. The theorem applies here because, when D is large and selection is weak, the rate at which allele frequencies within demes approach statistical equilibrium is much faster than the speed of allele-frequency changes in the total population. We have the parameter

$$\gamma=\frac{N}{2}\lim_{D\to\infty} Ds_D, \tag{7}$$

which is the selection coefficient scaled by the total population size, but including the extra factor of two discussed in Section 1.

Adapting Ethier and Nagylaki's (1980) theorem to the present case, we show that over one generation (suppressing t),

$$\varepsilon_D^{-1}E_z[X^D(1)-x]=b(x,y)+o(1), \tag{8}$$

$$\varepsilon_D^{-1}E_z[(X^D(1)-x)^2]=a(x,y)+o(1), \tag{9}$$

$$\varepsilon_D^{-1}E_z[(X^D(1)-x)^4]=o(1), \tag{10}$$

$$\delta_D^{-1}E_z[Y_i^D(1)-y_i]=c_i(x,y)+o(1), \tag{11}$$

$$\delta_D^{-1}\mathrm{Var}_z[Y_i^D(1)]=o(1), \tag{12}$$

in which

$$\varepsilon_D^{-1}=\frac{(ND)^2}{2}\left(1+\frac{1-m}{Nm}\right), \tag{13}$$

$$\delta_D^{-1}=D \tag{14}$$

and

$$a(x,y)=x(1-x)+\sum_{i=0}^{N}\alpha_i y_i, \tag{15}$$

$$b(x,y)=\gamma\frac{N-1}{N}x(1-x)+\sum_{i=0}^{N}\beta_i y_i, \tag{16}$$

$$c_i(x,y)=\sum_{j=0}^{N} y_j P_{ji}^{*}-y_i. \tag{17}$$

In Eqs. (8)–(12) and below, $E_z[\cdot]$ denotes an expectation over $Z^D(t+1)$ for a given $z(t)$. The terms α_i, β_i, and P_{ji}^{*} are functions of N, m, and x, and are detailed below. The first two of these are inconsequential because $y\to 0$ in the limiting diffusion. One further technical requirement, in this case where $\delta_\infty=0$, is that the zero solution, $c(x,0)=0$ of the (deterministic) differential equation

$$\frac{d}{dt}Y(t,x,y)=c(x,Y(t,x,y)),\quad Y(0,x,y)=y \tag{18}$$

is globally asymptotically stable for $x\in(0,1)$.

Then by Theorem 3.3 of Ethier and Nagylaki (1980), as D tends to infinity the above system reduces to a diffusion $x(\cdot)$ with generator

$$\mathscr{L}=\frac{1}{2}a(x,0)\frac{d^2}{dx^2}+b(x,0)\frac{d}{dx}. \tag{19}$$

This is our main result, and is shown to be true in the next four sections. From (15) and (16) it is clear that the form of this diffusion is identical to that for an unstructured population (Ewens, 1979), only with a different time scale and the usual factor of two difference between the Moran model and the Wright–Fisher model. The time scale depends on $(ND)^2/2$, as in the unstructured case (see Eq. (2)), but is lengthened by the factor $1+(1-m)/(Nm)$ which is greater than one if

$m<1$ and is equal to one if $m=1$. There is an additional factor $(N-1)/N$ which decreases the strength of selection, and reflects the fact that there is no within-deme variation for selection to act upon when $N=1$. Note that populations in which selection acts across the entire populations do not have this property, and instead become identical in all respects to an unstructured population when $N=1$ and $m=1$.

2.2. Allele frequencies within a deme

Wright (1931) studied the equilibrium distribution of allele frequencies in a single deme that receives migrants from a continent of infinite size or from the other demes in an infinite island model. What is important in either case, and what allows the frequencies within the deme to reach an equilibrium, is that the allele frequencies among migrants are constant. Many others since have studied distributions of this sort, which are essentially the same as equilibrium distributions under bi-directional mutation. Rannala (1996), for example, recently studied the within-deme equilibrium distribution of allele frequencies in a Moran type model of reproduction with multiple alleles. Our results are closely related to these, in particular those of Moran (1962, pp. 129–132), except that the allele frequencies among migrants are allowed to change according to the diffusion (19) above. However, the time scale of these changes is so much longer than the time scale of migration and drift within demes that it can be treated as constant in establishing the equilibrium distribution.

Consider what happens in the model described above when an individual is chosen to die within a deme that contains i copies of the mutant allele and $N-i$ copies of the wild-type allele. The probability that a mutant individual is the one chosen to die is equal to

$$\frac{\lambda_2 i}{\lambda_2 i+\lambda_1(N-i)}=\frac{(1-s_D)i}{N-s_D i} \tag{20}$$

$$=\frac{i}{N}-s_D\frac{i(N-i)}{N^2}+O(s_D^2), \tag{21}$$

and the probability that a wild-type individual is chosen to die is one minus this. The individual is replaced by a mutant or by a wild-type individual with a probability that depends on the migration rate and the frequency of mutant individuals in the entire population, as well as on i. Overall, the single-step probabilities of transition for a deme which contains i mutants now and in which an individual is chosen to die are given by

$$P_{i,i+1}=\left[\frac{N-i}{N}+s_D\frac{i(N-i)}{N^2}\right]\times\left[(1-m)\frac{i}{N}+mx\right]+O(s_D^2), \tag{22}$$

$$P_{i,i-1}=\left[\frac{i}{N}-s_D\frac{i(N-i)}{N^2}\right]\times\left[(1-m)\frac{N-i}{N}+m(1-x)\right]+O(s_D^2), \tag{23}$$

$$P_{i,i}=1-P_{i,i+1}-P_{i,i-1}, \tag{24}$$

$$P_{i,j}=0 \quad \text{if } |i-j|>1. \tag{25}$$

Thus, for a constant x, the allele-frequency changes within a deme form a Markov process with transition matrix $\mathbf{P}$, which has entries (22)–(25).

Of course, x will change from time step to time step, but in a moment we will see that these changes are slow in comparison to the allele-frequency changes within a deme. For a constant x, the equilibrium allele-frequency distribution v within a deme is the solution to

$$v=v\mathbf{P}, \tag{26}$$

subject to the condition

$$\sum_{i=0}^{N} v_i=1. \tag{27}$$

These equations have a unique solution, which is given by Moran (1962, p. 132) after making the substitutions $\alpha_1=m(1-x)$ and $\alpha_2=mx$ for Moran's mutation parameters. Note that in the limit of large N and small m, with $M=\lim_{N\to\infty} Nm$ and for a constant $y=i/N$, the solution converges on a continuous distribution of allele frequencies within the deme (Moran, 1962) which is identical to the corresponding limit result for Wright–Fisher reproduction within the deme (Wright, 1931).

Here we will be considering the limit as D tends to infinity and s_D tends to zero, in which case the effect of selection on the within-deme equilibrium distribution is small. Thus, we are concerned with the matrix $\mathbf{P}^*=\lim_{s_D\to 0}\mathbf{P}$, and we have $\mathbf{P}=\mathbf{P}^*+O(s_D)$. Because $\mathbf{P}^*$, like $\mathbf{P}$, is a continuant, the solution to $v=v\mathbf{P}^*$ can be obtained exactly as

$$v_k=v_0\frac{P^*_{0,1}P^*_{1,2}\cdots P^*_{k-1,k}}{P^*_{1,0}P^*_{2,1}\cdots P^*_{k,k-1}} \tag{28}$$

with v_0 determined by constraint (27); see Ewens (1979, pp. 73–74). Then,

$$v_k=\binom{N}{k}\frac{\Gamma(\frac{Nm}{1-m})\Gamma(\frac{Nmx}{1-m}+k)\Gamma(\frac{Nm(1-x)}{1-m}+N-k)}{\Gamma(\frac{Nm}{1-m}+N)\Gamma(\frac{Nmx}{1-m})\Gamma(\frac{Nm(1-x)}{1-m})}, \tag{29}$$

which differs only by $O(s_D)$ from the solution to Eqs. (26) and (27). Note that distribution (29) can also be expressed as a type of hypergeometric distribution, specifically a Polya distribution; see Eq. (40.13) in Johnson et al. (1997). A Polya distribution was proposed in Wakeley (2003) as an approximation for v in the case of Wright–Fisher reproduction within the deme. Distribution (29) converges to a beta distribution

as N tends to infinity for $y = i/N$ constant and with $M = \lim_{N\to\infty} Nm$ (Moran, 1962).

The mean and the variance of the number of mutant individuals in a deme can be obtained using (26) and (27) or by using the relation $v = v\mathbf{P}^*$ to solve for the moments recursively. They are

$$E_v[K] = Nx, \tag{30}$$

$$\mathrm{Var}_v[K] = \frac{N^2x(1-x)}{Nm+1-m}, \tag{31}$$

and these are required in computing the moments of $X^D(t)$ below. Again these differ only by $O(s_D)$ from the mean and variance that would be obtained using $v = v\mathbf{P}$.

2.3. Allele frequencies among demes

If D demes received migrants from a common source in which the allele frequencies were constant, they would independently approach the equilibrium v, and the distribution of the number of demes containing $i = 0, 1, \ldots, N$ copies of the mutant would be multinomial with parameters D and $v_0, \ldots, v_N$. Let $r_i = Dz_i$, so that $(r_0, \ldots, r_N)$ is a configuration of the demes among frequency classes. Then, in a single generation either a deme moves up a single frequency class, a deme moves down a single frequency class, or no change occurs in the population. We have

$$P\{(\ldots, r_i, r_{i+1}, \ldots) \to (\ldots, r_i - 1, r_{i+1} + 1, \ldots)\} = \frac{r_i}{D} P^*_{i,i+1}, \tag{32}$$

$$P\{(\ldots, r_{i-1}, r_i, \ldots) \to (\ldots, r_{i-1} + 1, r_i - 1, \ldots)\} = \frac{r_i}{D} P^*_{i,i-1}, \tag{33}$$

$$P\{(r_0, \ldots, r_N) \to (r_0, \ldots, r_N)\} = \sum_{i=0}^{N} \frac{r_i}{D} P^*_{i,i}. \tag{34}$$

Eqs. (32)–(34) define an ergodic Markov chain with a finite number of states, which will approach a unique steady state.

Let $p_t(r_0, \ldots, r_N)$ be the probability of configuration $(r_0, \ldots, r_N)$ at time t. The steady state distribution satisfies

$$\begin{aligned} p_\infty(r_0, \ldots, r_N) &= \sum_{i=0}^{N} p_\infty(\ldots, r_{i-1} + 1, r_i - 1, \ldots) \frac{r_{i-1}+1}{D} P^*_{i-1,i} \\ &\quad + \sum_{i=0}^{N} p_\infty(\ldots, r_i - 1, r_{i+1} + 1, \ldots) \frac{r_{i+1}+1}{D} P^*_{i+1,i} \\ &\quad + p_\infty(r_0, \ldots, r_N) \sum_{i=0}^{N} \frac{r_i}{D}(1 - P^*_{i,i-1} - P^*_{i,i+1}). \end{aligned} \tag{35}$$

The solution to (35) is the multinomial distribution

$$p_\infty(r_0, \ldots, r_N) = \frac{D!}{r_0! \cdots r_N!} v_0^{r_0} \cdots v_N^{r_N} \tag{36}$$

which is justified because

$$\sum_{i=0}^{N} \frac{r_i}{D}(P^*_{i,i-1} + P^*_{i,i+1}) = \sum_{i=0}^{N} \frac{r_i}{D}\left(\frac{v_{i-1}}{v_i} P^*_{i-1,i} + \frac{v_{i+1}}{v_i} P^*_{i+1,i}\right). \tag{37}$$

Eq. (37) is shown to be true using the general form of the solution for v_k given in Eq. (28). The methods proposed by Siegert (1949), in the context of theory of gases, can be used to prove convergence to equilibrium (36). Note that Eq. (36) holds for any subsample of K demes if D is replaced with K, and $\sum_{i=0}^{N} r_i = K$, so (36) might be useful in estimating migration rates from samples from a subdivided population.

The proof that the zero solution of (18) is globally asymptotically stable does not require such a detailed description of the distribution of demes among frequency classes. Note that (18) is deterministic in that x does not change, and further that the solution $c(x, 0) = 0$ is equivalent to $z = v$. In a single time step, Z_i^D can either increase by $1/D$, decrease by $1/D$, or remain unchanged:

$$Z_i^D(1) = \begin{cases} z_i - \dfrac{1}{D} & \text{with probability } z_i(1 - P_{i,i}), \\ z_i + \dfrac{1}{D} & \text{with probability} \\ & \quad z_{i-1}P_{i-1,i} + z_{i+1}P_{i+1,i}, \\ z_i & \text{otherwise.} \end{cases} \tag{38}$$

Measuring time in units of D generations, with $dt = 1/D$, and letting D go to infinity, so that the $O(1/D)$ parts of $P_{i,j}$ become negligible, we have the differential equation

$$\frac{dz_i}{dt} = z_{i-1}P^*_{i-1,i} + z_{i+1}P^*_{i+1,i} - z_i(P^*_{i,i-1} + P^*_{i,i+1}). \tag{39}$$

The solution of this equation for the vector $z = (z_1, \ldots, z_N)$, with some initial condition $z(0)$, is given by

$$z(t) = z(0)e^{(\mathbf{P}^* - \mathbf{I})t}, \tag{40}$$

in which $\mathbf{I}$ is the $(N+1) \times (N+1)$ identity matrix. The rate matrix $\mathbf{P}^* - \mathbf{I}$ has leading eigenvalue $\lambda_0 = 0$, and all other eigenvalues are less than zero:

$$\lambda_i = -\frac{im}{N} - \frac{i(i-1)(1-m)}{N^2} \tag{41}$$

(Cannings, 1974); see also Ewens (1979, pp. 86–87). Therefore, regardless of initial conditions, $z(t) \to v$ as $t \to \infty$ where v is the solution to

$$0 = v(\mathbf{P}^* - \mathbf{I}), \tag{42}$$

and this is identical to solution (29) of $v = v\mathbf{P}^*$.

2.4. The time scale of drift and migration within demes

Here we establish (11), (12), and (17) which set the (fast) time scale of drift and migration within demes. We have

$$E_z[Y_i^D(1)-y_i]=E_z[Z_i^D(1)-z_i]+E_z[v_i(1)-v_i]. \quad (43)$$

Using (38), the first term on the right above is given by

$$E_z[Z_i^D(1)-z_i] = \frac{1}{D}[z_{i-1}P_{i-1,i}+z_i(P_{i,i}-1)+z_{i+1}P_{i+1,i}] \quad (44)$$

$$= \frac{1}{D}[z_{i-1}P^*_{i-1,i}+z_i(P^*_{i,i}-1)+z_{i+1}P^*_{i+1,i}]+O(s_D/D) \quad (45)$$

since $P_{i,j}-P^*_{i,j}$ is $O(s_D)$.

To compute the second term on the right in (43), we note that the equilibrium v is determined by the value of X^D in each generation. Rewriting Eq. (29), we have

$$v_i(1)=\binom{N}{i}\frac{\Gamma(\frac{Nm}{1-m})}{\Gamma(\frac{N}{1-m})}\prod_{j=0}^{i-1}\left(\frac{NmX^D(1)}{1-m}+j\right) \times\prod_{k=0}^{N-i-1}\left(\frac{Nm(1-X^D(1))}{1-m}+k\right). \quad (46)$$

Then, putting in $X^D(1)=x+\Delta x$ and simplifying, the two products on the right become

$$\prod_{j=0}^{i-1}\prod_{k=0}^{N-i-1}\left[\left(\frac{Nmx}{1-m}+j\right)\left(\frac{Nm(1-x)}{1-m}+k\right)+\Delta x\frac{Nm}{1-m} \times\left(\frac{Nm(1-2x)}{1-m}+k-j\right)-(\Delta x)^2\left(\frac{Nm}{1-m}\right)^2\right]. \quad (47)$$

Expanding this, it is clear that Eq. (46) can be written

$$v_i(1)=v_i+\sum_{j=1}^{2N}C_{i,j,N,m,x}(\Delta x)^j \quad (48)$$

and that the coefficients $C_{i,j,N,m,x}$ are bounded, i.e. do not depend on D.

The moments of $\Delta x=X^D(1)-x$ are the topic of the next section, where it is shown that $E[(X^D(1)-x)^k]=O(s_D/D^k)$ if k is odd, and $E[(X^D(1)-x)^k]=O(1/D^k)$ if k is even. From this and Eq. (48), we have $E[v_i(1)]=v_i+O(s_D/D)$, and we can write the second term on the right in (43) becomes

$$E_z[v_i(1)-v_i]=\frac{1}{D}[v_{i-1}P^*_{i-1,i}+v_i(P^*_{i,i}-1)+v_{i+1}P^*_{i+1,i}]+O(s_D/D). \quad (49)$$

The quantity in brackets in (49) is equal to zero by definition. Putting (45) and (49) into (43) gives

$$E_z[Y_i^D(1)-y_i] = \frac{1}{D}\left(\sum_{j=0}^{N}y_jP^*_{ji}-y_i\right)+O(s_D/D) \quad (50)$$

as needed in Section 2.1.

Using relation (3.12) in Ethier and Nagylaki (1980), and an argument similar to that above,

$$\mathrm{Var}_z[Y_i^D(1)]=\mathrm{Var}_z[Z_i^D(1)-v_i(1)] \quad (51)$$

$$\leqslant 2\mathrm{Var}_z[Z_i^D(1)]+2\mathrm{Var}_z[v_i(1)] \quad (52)$$

$$=2\mathrm{Var}_z[Z_i^D(1)-z_i]+2\mathrm{Var}_z[v_i(1)-v_i] \quad (53)$$

$$\leqslant 2E_z[(Z_i^D(1)-z_i)^2]+2E_z[(v_i(1)-v_i)^2] \quad (54)$$

$$=O(1/D^2)+O(s_D/D). \quad (55)$$

The first term on the right in (54) is $O(1/D^2)$ directly from (38). The second term on the right in (54) is $O(s_D/D)$ since $E[X^D(1)-x]$ is the leading term in $E[(v_i(1)-v_i)^2]$ and using the moments of $X^D(1)-x$ obtained in the next section.

2.5. Allele-frequency changes in the total population

Demonstrating that (8)–(10) are true is easier here than in the case of Wright–Fisher reproduction (Wakeley, 2003). Here, in a single time step the frequency of the mutant allele can either increase by $1/(ND)$, decrease by $1/(ND)$, or remain unchanged:

$$X^D(1)=\begin{cases}x-\dfrac{1}{ND} & \text{with probability } \sum_{i=0}^{N}z_iP_{i,i-1},\\ x+\dfrac{1}{ND} & \text{with probability } \sum_{i=0}^{N}z_iP_{i,i+1},\\ x & \text{otherwise.}\end{cases} \quad (56)$$

The probabilities in (56) are simply the probabilities that the individual chosen to die is in a deme containing i copies of the mutant times the probability that i is either decremented, incremented, or unchanged when that individual is replaced by an offspring individual. Therefore, the moments of the change in allele frequency in the total population are given by

$$E_z[(X^D(1)-x)^k] = \begin{cases}\dfrac{1}{(ND)^k}\sum_{i=0}^{N}z_i(P_{i,i+1}-P_{i,i-1}) & \text{if } k \text{ is odd},\\ \dfrac{1}{(ND)^k}\sum_{i=0}^{N}z_i(P_{i,i+1}+P_{i,i-1}) & \text{if } k \text{ is even},\end{cases} \quad (57)$$

which clearly satisfies the requirement of the diffusion that the higher moments are vanishing.

The time scale of the diffusion and the coefficients $a(x,y)$ and $b(x,y)$ are obtained by putting (22) and (23) in (57) and simplifying. We have

$$\sum_{i=0}^{N} z_i(P_{i,i+1}-P_{i,i-1}) = \sum_{i=0}^{N} z_i\left[m\left(x-\frac{i}{N}\right)+s_D\frac{i(N-i)}{N^2}+O(s_D^2)\right] \quad (58)$$

$$= \sum_{i=0}^{N} (v_i+y_i)\left[m\left(x-\frac{i}{N}\right)+s_D\frac{i(N-i)}{N^2}\right]+O(s_D^2) \quad (59)$$

$$= s_D x(1-x)\frac{(N-1)m}{Nm+1-m}+\sum_{i=0}^{N}\beta_i y_i+O(s_D^2) \quad (60)$$

and

$$\sum_{i=0}^{N} z_i(P_{i,i+1}+P_{i,i-1}) = \sum_{i=0}^{N} z_i\left[m\left(x-\frac{i}{N}\right)\left(1-\frac{2i}{N}\right)+\frac{2i(N-i)}{N^2}\right]+O(s_D) \quad (61)$$

$$= 2x(1-x)\frac{Nm}{Nm+1-m}+\sum_{i=0}^{N}\alpha_i y_i+O(s_D), \quad (62)$$

which makes use of moments (30) and (31) of the distribution v. The terms β_i and α_i are defined to be proportional to the bracketed terms in (59) and (60), respectively. A little algebra gives $a(x,y)$, $b(x,y)$, and ε_D presented in Section 2.1. This completes the application of Ethier and Nagylaki's (1980) theorem, and shows that the diffusion operator (19) governs the change in allele frequencies in the total population, while v of Eq. (28) describes the distribution of demes among frequency classes.

2.6. The limit of large deme size

The essential structure of the diffusion process described above continues to hold in the limit as $N\to\infty$, with the migration equal to $M=\lim_{N\to\infty} Nm$. It is necessary also to define the selection parameter to be equal to $\gamma=\lim_{N\to\infty}\lim_{D\to\infty} NDs_{ND}/2$. Note that the diffusion time scale, in which time is measured in units of ε_D^{-1} generations, already depends linearly on N, so taking the limit $N\to\infty$ simply further lengthens the time scale. The within-deme quantities $P_{i,j}$ and v also converge to continuous limits. This uniform convergence in N was found recently in many-demes genealogical, or coalescent, models (Lessard and Wakeley, 2004).

3. Discussion

We have shown that when the number of demes tends to infinity, the diffusion approximation for an unstructured population applies to the allele frequency in a subdivided population with island model migration among demes. We have assumed a Moran model of reproduction, so this work builds on that in Wakeley (2003), where Wright–Fisher reproduction was assumed. Under both models of reproduction, the diffusion has a time scale, or an effective population size $N_e=\varepsilon_D^{-1}$, which is equal that of the diffusion for an unstructured population times a factor which depends on the migration rate and the deme size. This many-demes diffusion result holds for any deme size $N\geqslant 1$ and for any non-zero migration rate $0<m\leqslant 1$, but it also holds in the limit as the deme size tends to infinity, with the usual migration parameter $M=\lim_{N\to\infty} Nm$. Thus, all the many well-known results of the standard unstructured diffusion model, for example in Chapter 5 of Ewens (1979), should be valid in the case where the population is subdivided into many demes.

It is not simply this change in N_e that distinguishes a subdivided population with many demes from a panmictic one. As the allele frequency in the total population changes by drift and selection on this long time scale, the collection of demes remains at the equilibrium v through the action of migration and within-deme drift. While the fast time scale of migration and within-deme drift guarantees that the demes track the shifting equilibrium v closely, it is important to keep in mind that the within-deme process is discrete as long as N is finite. More precisely, the series of mutant allele counts within a single deme forms a time-inhomogeneous Markov chain with transition matrix $\mathbf{P}(t)$ at time t. A single deme will spend much of its time fixed for one or the other allele if m is small, but will have a mutant allele frequency close to x much of the time if m is large. The variance in mutant allele counts among demes at a time when the frequency of the mutant allele in the total population is equal to x is given by Eq. (31). Finally, note that the number of time steps between potential changes in allele count within a deme is geometrically distributed with mean D because each deme has a chance $1/D$ at each time step of being the one in which an individual dies and is replaced.

Simulations, such as those of Cherry and Wakeley (2003), support the use of this approximation over a broad range of D, N, m, and s_D. Similarly, several authors have recently computed the first two moments of the change in allele frequency—$\varepsilon_D a(x,0)$ and $\varepsilon_D b(x,0)$ in the present notation—for a variety of models of subdivision, and then used simulations to assess the accuracy of the implied diffusion approximation. Cherry (2003a, b) studied the effect of dominance or frequency dependence of selective differences and the

effects of extinction and recolonization of demes on fixation probabilities and fixation times. Whitlock (2003) studied these quantities in similar models to those of Cherry (2003a, b) and also showed that the same approach makes useful predictions in the case of the stepping-stone model of migration (Kimura and Weiss, 1964). Roze and Rousset (2003) study partial selfing in addition to the above processes, and present a novel method of calculating $\varepsilon_D a(x,0)$ and $\varepsilon_D b(x,0)$ based on the partial derivatives of a fitness function and on probabilities of genetic identity. Roze and Rousset (2003) recognize that the diffusion relies on a separation of time scales and that it does not depend on the deme size N tending to infinity. The simulations presented in these works for a variety of different population models suggest that the many-demes limit, with its separation of two time scales, should hold in many different situations.

Acknowledgments

We thank Paul Slade and two anonymous reviewers for helpful comments. This work was supported by a Career Award (DEB-0133760) to J.W. from the National Science Foundation.

References

Cannings, C., 1974. The latent roots of certain Markov chains arising in genetics: a new approach. I. Haploid models. Adv. Appl. Prob. 6, 260–290.

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.

Ethier, S.N., Nagylaki, T., 1980. Diffusion approximations of Markov chains with two timescales and applications to population genetics. Adv. Appl. Prob. 12, 14–49.

Ewens, W.J., 1979. Mathematical Population Genetics. Springer, Berlin.

Feldman, M.W., 1966. On the offspring number distribution in a genetic population. J. Appl. Prob. 3, 129–141.

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

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

Johnson, N.L., Kotz, S., Balakrishnan, N., 1997. Discrete Multivariate Distributions. Wiley, New York.

Kimura, M., Weiss, G.H., 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49, 561–576.

Kingman, J.F.C., 1982. The coalescent. Stochastic Process. Appl. 13, 235–248.

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.

Moran, P.A.P., 1958a. Random processes in genetics. Proc. Camb. Phil. Soc. 54, 60–71.

Moran, P.A.P., 1958b. The effect of selection in a haploid genetic population. Proc. Camb. Phil. Soc. 54, 463–467.

Moran, P.A.P., 1959. The theory of some genetical effects of population subdivision. Austr. J. Biol. Sci. 12, 109–116.

Moran, P.A.P., 1962. Statistical Processes of Evolutionary Theory. Clarendon Press, Oxford.

Moran, P.A.P., Watterson, G.A., 1959. The genetic effects of family structure in natural populations. Austr. J. Biol. Sci. 12, 1–15.

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

Notohara, M., 1990. The coalescent and the genealogical process in geographically structured population. 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., 2001. The structured coalescent process with weak migration. J. Appl. Prob. 38, 1–17.

Pollack, E., 1968. On random genetic drift in a subdivided population. J. Appl. Prob. 5, 314–333.

Rannala, B., 1996. The sampling theory of neutral alleles in an island population of fluctuating size. Theor. Popul. Biol. 50, 91–104.

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.

Siegert, A.J.F., 1949. On the approach to statistical equilibrium. Phys. Rev. 76, 1708–1714.

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

Slatkin, M., 1985. Gene flow in natural populations. Ann. Rev. Ecol. Syst. 16, 393–430.

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

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.

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

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

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