A framework for estimating the fixation time of an advantageous allele in stepping-stone models

M. HARTFIELD
Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK

Introduction

The interaction between adaptive mutation (Haldane, 1924; Fisher, 1930) and population subdivision (Wright, 1951) is an area that has been the subject of an extensive body of population genetics research. Much work has focused on how different aspects of population subdivision affect the fixation probability of an advantageous allele (Patwa & Wahl, 2008), such as extinction and recolonization of demes (Barton, 1993; Whitlock, 2003; Cherry, 2003b, 2004), the impact of selfing and dominance of mutations (Whitlock, 2003; Roze & Rousset, 2003), frequency-dependent selection (Cherry, 2003a; Pannell et al., 2005), and environmental heterogeneity (Lenormand, 2002; Whitlock & Gomulkiewicz, 2005; Vuilleumier et al., 2008). There have also been more recent investigations as to how the emergence of multiple advantageous traits interact with each other in spatial populations and how migration prevents these mutations from interfering with one another (Ralph & Coop, 2010; Martens & Hallatschek, 2011).

One specific area that has gathered interest due to its impact on a wide array of evolutionary phenomena is the fixation time of a favourable allele, as it travels through a series of distinct populations. If the allele has the same selective effect in all subpopulations with additive dominance ($h = 1/2$) and if the deme size is independent of their mean fitness, then the fixation probability of the allele would be the same as in a panmictic population (Maruyama, 1970). However, the time to fixation will increase if the migration rate $m << 1$, which causes an allele to migrate to neighbouring demes in a stepwise fashion.

This slowing effect plays an important role in various evolutionary processes, such as preserving sexuals against asexual invasion (Peck et al., 1999; Salathé et al., 2006), maintaining underdominant chromosomal
inversions (Lande, 1979), altering the dynamics of new species invasion into an existing spatially extended population, if there is hybridization with existing species (Shigesada & Kawasaki, 1997), and determining whether migration rates are high enough to prevent neutral divergence between neighbouring regions (Morjan & Rieseberg, 2004). The slower spread also causes hitch-hiking within demes to affect patterns of linked neutral diversity, which can alter measures of population subdivision such as $F_{ST}$ (Slatkin & Wiehe, 1998; Santiago & Caballero, 2005; Bierne, 2010), and skew estimates of the strength of selective sweeps (Barton, 2000; Kim & Maruki, 2011). Substitution rates at selected loci are also reduced, due to the increased time needed to fix adaptive alleles (Gordo & Campos, 2006).

Fisher (1937) determined that if an allele invaded a spatially continuous population, then it would spread with speed $2\sqrt{sm}$, where $s$ is the selective advantage of the allele. This model is accurate if there is a high rate of migration so that the allele travels in a continuous manner ($m \gg s$) and drift effects in the migration rate are negligible ($Nm \gg 1$). It is not applicable in structured populations with low migration rates between adjacent demes, however, as the allele would not spread as a travelling wave. This was demonstrated by Slatkin (1976), who estimated the mean time taken for a sweep to establish itself in a neighbour, in a two deme system. Using numerical simulations, it was found that such a structured population reduces the speed of the spread of the allele by 14-fold, compared to the result predicted using Fisher’s travelling-wave solution. Slatkin (1981) subsequently used Markov chain methods to estimate an upper limit to fixation time, if migration between demes is weak. Kim & Maruki (2011) adopted a similar method in their analysis of how population subdivision affects heterozygosity at a linked neutral locus, in a haploid population. They determined that the mean ‘delay time’ before an allele is established in a new region is given by (Kim & Maruki, 2011, eqn 5):

$$\frac{1}{s} \log \left[ 1 + \frac{s}{m} \right]$$

if migration is frequent ($4Nm \gg 1$). A similar result was derived by Pálek & Barton (1997) when approximating the spread of a travelling wave through a structured population. However, Slatkin (1976, 1981) and Kim & Maruki (2011) assumed that the mean time needed for an allele to migrate and establish in a new deme (the ‘delay’ time) would be the same for every transfer to a new deme that an allele makes, irrespective of the location of the deme or the manner in which it was connected to its neighbours. Therefore, to calculate the overall time needed for an advantageous allele to fix in a population consisting of more than two demes, the mean delay time is multiplied by the number of transfers that the allele makes to a neighbouring deme before it is present in all populations. The analysis in this paper will show that this assumption is only accurate if migration is very weak ($N_{e}pm \ll 1$ for $N_{D}$, the population size of the deme, as also determined by Slatkin (1981)) and subpopulations are arranged in a one-dimensional formation. Otherwise, migration effects will reduce the delay time in subsequent demes. Slatkin (1976) also assumed that whilst the rate of spread of an allele would be quicker in a two-dimensional populations, due to the greater number of routes that an allele could take in order to spread, the same lag time would apply to each migration event. It will be shown that the lag times alter between different demes in a two-dimensional population, as some demes experience a greater influx of migrants (and efflux of emigrants) than others.

To try and calculate the fixation time in a more general subdivided population, Whitlock (2002, 2003) determined the mean change in allele frequency in a population whose level of subdivision can be measured using Wright’s $F_{ST}$ statistic (Wright, 1951):

$$F_{ST} = \frac{V[x]}{x(1-x)}$$

where $V[x]$ is the variance in frequency of the selected allele between demes and $x$ is the population mean frequency of the allele. Analytic values were obtained for populations where there is either ‘hard’ or ‘soft’ selection. With ‘hard’ selection, the contribution of each deme to the overall population in the next generation is determined by the mean fitness of the individuals within it, and ‘soft’ selection arises when each deme contributes individuals independently of the mean fitness of it (Wallace, 1975; Whitlock, 2002). The terms for the change in mean frequency and variance in frequency were then inserted in the diffusion equations outlined by Kimura & Ohta (1969) to calculate the fixation time. It was shown that this method provided an accurate estimate when applied to an island model and a stepping-stone model with demes arranged in a circle.

This paper aims to extend and complement previous studies by laying out a framework for calculating the fixation time of an advantageous allele in a general structured population, where the allele travels in a stepwise manner between demes. That is, the advantageous allele spreads in a single deme before migrating and establishing in a neighbour at a specific time-point, as opposed to travelling in a continuous manner through space (as in Fisher, 1937). The assumption of a stepwise movement of the allele holds if the migration rate is small ($m < s$ for $s$ the selective advantage of the allele). Models are formulated by considering the total number of demes (and the size of each), how they are connected and the migration rate between each region. An accurate predictor of the fixation time is made by assuming that the allele increases in frequency within each deme deterministically, but the time that such an allele establishes itself in connected neighbours is
influenced in a stochastic manner. A similar mix of deterministic and stochastic equations was used by Karasov et al. (2010) to calculate the fixation time of novel mutations arising at the Ace locus in a panmictic Drosophila population.

The model outlined in this paper can be used to investigate natural systems where $F_{ST}$ might not be the most accurate indicator of how subdivided a population is, and thus in informing how population subdivision will delay the spread of a selected allele. This may arise if selection acting on loci skews observed estimates of population subdivision (Lewontin & Krakauer, 1973), which arises if $s > m$, as assumed in this analysis (estimates of $F_{ST}$ are approximately the same for selected and neutral loci if $s < m$ (Whitlock, 2002)). $F_{ST}$ estimates may also give incomplete information on how the spread of a selected allele is affected by population subdivision, if the selective strength of the allele changes over time. The method described in this paper is flexible enough so that it can be applied to different kinds of stepping-stone model, which is subsequently demonstrated for two types of stepping-stone populations, spread out over one dimension and two dimensions, respectively. An added advantage of this analysis is that it can be used to inform on how migration itself can affect the spread of the allele, by introducing copies of it into neighbouring demes after it has established. This can help determine whether migration rates are sufficiently high between demes to prevent neighbouring regions from diverging (as reviewed in Morjan & Rieseberg, 2004). It also informs on determining when migration is sufficiently high enough in populations consisting of a large number of demes, so that the selected allele moves as a travelling wave. In such cases, Fisher’s solution can then be used to measure fixation time instead.

**First model: selection alone affects allele frequency within a deme**

The first model considers the growth in frequency of a rare allele, which is determined within each deme just by selection acting on it. It is assumed that migration between neighbouring demes can transfer the allele to a new population, but does not affect the allele frequency within each deme. This process continues until the allele fixes in all demes. The mean time taken for an advantageous allele to establish in a neighbouring deme was derived in a similar fashion by Slatkin (1976) and Kim & Maruki (2011) and provides a natural starting point for the present analysis. It is assumed that selection acting on the advantageous allele is strong enough so that it sweeps through each subpopulation in a deterministic manner, and also that migration is frequent but weak compared to selection ($N_{gt} >> 1$ and $m < s$, for $N_g$ the deme population size), so that the allele travels through each deme in a stepwise formation. In spite of these assumptions, it will be shown that these models are generally robust to small $N_{gt}$ values; if $m >> s$, then they will match up to Fisher’s (1937) travelling-wave solution.

Consider a finite haploid population of size $N$, spread equally over $D$ demes, so there are $N_g = N/D$ individuals per deme. After a new generation is created, a proportion $m$ of individuals migrate to a neighbouring deme. At $t = 0$, an individual in a single deme acquires an advantageous mutant, with selection $s$ acting on it (so the fitness of the individual carrying that allele increases from 1 to $1 + s$). It is assumed that the allele is not lost stochastically and proceeds to increase in frequency within that deme. The frequency of the allele at time $t$ is denoted by $p(t)$, which is given by the logistic growth equation (Haldane, 1924):

$$p(t) = \frac{p_0 e^{rt}}{1 - p_0 (1 - e^{rt})}$$

Here, $p_0$ is the initial frequency of the allele, which is set to (Barton, 1994):

$$p_0 = \frac{e^{-\gamma}}{2N_p s}$$

where $\gamma \approx 0.577$ is Euler’s constant. This value is the ‘effective’ initial frequency, which takes into account the accelerated rise in allele frequency if we only consider cases where the allele is not lost stochastically.

At time $t$ in this first deme, the probability that an individual advantageous allele migrates to a neighbour is given by $mp(t)$; the mean number of migrants is therefore equal to $N_g mp(t)$. To ensure that each deme is kept at a constant size, after an individual migrates to a neighbour, an individual in the target deme is then moved back to the focal deme. Therefore, the total proportion of individuals that migrate between demes every generation is equal to $2mp(t)$, of which half of these will move to one of two neighbouring demes. Overall, the total proportion of alleles that migrate to a specific neighbour is equal to $1/2 \times 2mp(t) = mp(t)$. Once the allele transfers, the probability of it then establishing itself in the new population is given by $2s$, for $1 >> s >> 1/N$ (Haldane, 1927). Here, ‘establishment’ of the allele is defined as the arrival of a copy of the allele that will eventually fix in the population, as opposed to a copy that is lost by stochastic drift. Thus, the overall probability that an allele will migrate and establish itself in a neighbouring deme at that generation is $P(t) = 2mp(t)$. Since an allele only has to establish itself once, then it would have failed to do so in previous generations, each time with probability $1 - P(t')$ (for $t' < t$). Therefore, the probability that the first establishment occurs at time $t$, denoted by $Q(t)$, is equal to:

$$Q(t) = \frac{1}{N_p} \left[ \prod_{t'=0}^{t-1} (1 - N_p P(t')) \right] N_p P(t)$$

Note that $P(t)$ is multiplied by $N_p$ to account for the mean number of advantageous alleles that migrate, which
equals $N_d p m(t)$. The calculation of eqn 5 can be greatly speeded up by approximating the product term; this method was similarly used in simplifying eqn 3 of Hartfield & Otto (2011). If each probability $P(t)$ is small, then the product term can be written as:

$$\prod_{t=0}^{t-1} (1 - N_d P(t')) \approx \exp\left(-\int_0^{t-1} N_d P(t') dt\right)$$

(6)

This is a valid approximation since $N_d p m$ is not generally found to be large; Morjan & Rieseberg (2004) notes that most estimates from natural populations lie below 10. Therefore, the compound parameter $N_d p(t) = 2 N_d p m(t)$ is small due to the $sp(t)$ term. By evaluating the integral in eqn 6, the following is obtained:

$$\exp\left(-\int_0^{t-1} N_d P(t') dt\right) = \left[\frac{2e^{-N_d s}}{2e^{N_d s} + e^{(t-1)}} - 1\right]^{2N_c m}$$

(7)

This derivation is outlined in Supporting Information Appendix S1. From $Q(t)$ the mean time until the allele establishes itself in a neighbouring population can then be calculated. We define this time as $MT_1$ (‘mean time 1’):

$$MT_1 = \sum_{t=1}^{\infty} tQ(t)$$

(8)

$MT_1$ is calculated numerically by computing the sum up to a large upper bound, so that it does not increase further.

In this first model, it is assumed that the rise in frequency of the allele in new demes is determined entirely by selection acting on it, and the effect of migration on its frequency within subsequent demes (through the transfer of alleles between demes) is negligible. Therefore in this model, $MT_1$ not only determines the mean time taken for the allele to become established in the neighbouring deme to which the allele first arose, but also other demes thereafter, as assumed by Slatkin (1976) and Kim & Maruki (2011). Once the allele establishes itself in the furthest deme, it no longer has to migrate so it only remains to consider the time needed for it to fix within this last deme. Labelling this time as $MT_2$, this is given by the time needed for the allele to reach a frequency of $1 - p_0$:

$$MT_2 = \frac{1}{s} \log \left[\frac{(1 - p_0)^2}{(1 - p_0^2)}\right]$$

(9)

$$MT_2 = \frac{1}{s} \log \left[\frac{(1 - 2N_d s e^{N_d s})^2}{1 - 2N_d s e^{N_d s}}\right]$$

(10)

Note that Kimura & Ohta (1969) formulated an expression for allele fixation time in a finite panmictic population, using stochastic diffusion equations. However, I use a deterministic equation to calculate $MT_2$ so as to retain consistency with the deterministic formulation of $MT_1$. Also note that this calculation implicitly assumes that once the furthest deme in the chain has reached fixation, then so have all other subpopulations; there are no other demes that are polymorphic at that time. This is a sensible assumption if alleles are strongly selected for, but could be violated for small $N_d s$ values. Despite these caveats, it shall be seen that the following models provide an accurate match to simulation data with these assumptions in place.

Let there be $D'$ demes between the first deme where the allele first appears and the furthest deme from it. Note that $D'$ is usually not equal to the total number of demes present in a population. For example, if there exist $D$ demes arranged in a circular stepping-stone formation, then $D' = D/2$ if $D$ is even or $D' = [(D-1)/2 + 1]$ if $D$ is odd. $D'$ signifies the number of demes an advantageous allele has to traverse before it covers the whole population. In this model, after it first appears, the advantageous allele will migrate $D'-1$ times in order to get to the furthest deme, with the mean time taken for each establishing migration to occur equal to $MT_1$. Then, it has to fix in the furthest deme, which takes $MT_2$ generations on average. Thus, the mean time to fixation over the whole population is equal to $(D'-1)MT_1 + MT_2$ generations. Supporting Information Appendix S2 outlines *Mathematica* 8.0 code (Wolfram Research Inc., 2010) for calculating this value.

This first model can be used to estimate the relative proportion of the total fixation time needed for an adaptive allele to transfer to a new deme (as given by $MT_1$) and for the allele to fix in the final deme ($MT_2$). Specifically, it can be determined when a particular part of the calculation, such as $MT_2$, only contributes a small amount to the overall fixation time (such as $< 5\%$).

Figure S1 plots the total fixation time contributed by $MT_2$ as a function of the number of demes, if the migration rate is low ($N_d p = 0.1$ with $N_d = 2000$). As expected, the contribution to $MT_2$ falls as the number of demes $D'$ increases, and the contribution by $MT_2$ to the overall fixation time increases if the selective strength of the allele $N_d s$ is higher. If $N_d s = 10$ then $MT_2$ contributes $< 5\%$ to the overall fixation time if $D'$ exceeds 20, but if $N_d s = 50$, then $D'$ has to exceed 70 for the contribution made by $MT_2$ to fall below 5%. This demonstrates that if the allele is strongly selected for, $MT_2$ provides a relatively higher contribution to the total fixation time, since most demes remain polymorphic when the adaptive allele reaches the final deme.

**Second model: selection and migration both affect allele frequency within a deme**

It is entirely feasible that whilst the advantageous allele is travelling between the first and furthest deme, migration can affect the frequency of the allele in intermediate demes. This situation can arise, for example, through the introduction of more copies of the allele from the
previous deme or if the frequency of the allele is reduced as individuals leave. Since this process can have a significant effect on the fixation time, the first model is adjusted to take such migration effects into account. The basic derivation assumes a fixed \( s \) and \( m \), but these can be altered if applying the model to a population with differing values between demes.

In the first deme, migration cannot bring in any new alleles from neighbours, so selection alone determines the frequency of the allele in that deme. Thus, the mean time for it to become established in a second deme is \( MT1 \), as before. Similarly, the time to fixation in the furthest deme is kept as \( MT2 \). In intermediate demes (demes 2 to \( D' - 1 \)), once the advantageous allele establishes itself, it is assumed that the allele frequency not only changes due to selection, but also due to migration moving copies of the allele between neighbouring demes. In order to account for these extra effects, a system of differential equations needs to be formulated that model migration affecting the frequency of the allele within demes. These equations can then be used to calculate the delay time before an allele establishes in a new area, in a similar manner to the first model. As it shall be seen, incorporating these effects into the model causes a significant reduction in fixation time, because, even though migration is weak relative to selection, causes a significant reduction in fixation time, because, even though migration is weak relative to selection (\( m \ll s \)), the scaled rate of migration can be significant \( (N_{2p}m = O(1)) \) and thus can affect the frequency of the allele within different demes.

The simplest way to account for migration effects over a large number of demes is to break the problem down, and consider a closed system of equations in which the allele moves between just two linked demes. These two regions are representative of the deme in which the advantageous allele previously resided and the deme in which it has just become established. This system therefore assumes that only one other subpopulation ‘feeds’ advantageous alleles into the current deme; this assumption may be violated if a deme is connected to many neighbours, such as in a population spread out over a two-dimensional torus. The next section demonstrates how migration to and from multiple neighbours can be accounted for.

Define \( p_2(t) \) as the frequency of the advantageous allele in the deme where it has just become established. Time is reset, so \( t = 0 \) is defined as the time when the establishing mutation first appears in the new deme. Furthermore, \( q_2(t) \) is defined as the frequency of the allele in the previous deme, from which the advantageous allele is migrating. Under these assumptions, the following set of differential equations are formed:

\[
\frac{dp_2(t)}{dt} = sp_2(t)(1 - p_2(t)) - mp_2(t) + mq_2(t) \tag{11}
\]

\[
\frac{dq_2(t)}{dt} = qp_1(t)(1 - q_2(t)) + mp_2(t) - mq_2(t) \tag{12}
\]

This system considers the allele growing in frequency within the deme due to selection (as denoted by the \( sp_2(t)(1 - p_2(t)) \) term, along with its equivalent for \( q_2 \); migration introducing the allele from the previous deme to the current deme (denoted by the \( + mp_2(t) \) terms); and migration moving the allele back to the previous deme (denoted by the \( -mq_2(t) \) terms). Note that in order to keep the system of equations closed (so that \( p_2, q_2 \) can reach a maximum frequency of one), the system only consider migration occurring between these two demes alone. In reality, migration can also shift copies of the allele back to other demes or forward to demes where it has yet to establish (such individuals are then lost by stochastic drift). In order to fully account for these migration effects, it would be necessary to set up a system of equations for all demes in the chain, which would be unwieldy. However, it is possible to produce an accurate model even if these effects are not considered, as these have a minimal effect on allele frequencies. This is because the allele would have fixed in previous demes, so migration from the first deme considered (where the allele frequency is denoted by \( p_2 \)) to the one that lies previous to it in the chain would not affect the average gene frequency within the first deme. Similarly, only a tiny fraction of individuals would be lost stochastically due to extra migration from the second deme considered (where the allele frequency is denoted by \( p_2 \)). It will be seen that the adjusted model formed using the above equations still gives an accurate calculation of fixation time.

This system has initial frequency \( p_2(0) = p_0 \) (as defined by eqn 4) and \( q_2(0) = p(\text{MT1}) \) (the frequency of the allele in the previous deme, at the mean time when it establishes itself in the new population). This system can be evaluated numerically (e.g. by using the ‘NDSolve’ function in Mathematica).

Similar calculations as before can be used to find the mean time before the allele establishes itself in a subsequent deme. The probability that an establishing migration event occurs at time \( t \) is \( P_2(t) = 2mp_2(t) \). As with the previous model, if the first establishing migration occurs at time \( t \), then the allele would have failed to establish in previous generations with probability \( (1 - P_2(t)) \). So the probability that the first establishing migration takes place at generation \( t \) is:

\[
Q_2(t) = \prod_{\tau=0}^{t-1} (1 - N_0P_2(\tau)) N_0P_2(t) \tag{13}
\]

Therefore, the mean time for establishment in the next deme is defined as:

\[
\text{MT1a} = \sum_{t=1}^{\infty} tQ_2(t) \tag{14}
\]

In this second model, the allele takes \( \text{MT1} \) generations to leave the first deme and establish itself in the second. It then takes \( \text{MT1a} \) generations, on average, for the allele to...
establish itself in subsequent demes, which occurs \(D' - 2\) times if \(s, m\) do not differ between demes. Finally, the allele fixes within the furthest deme in MT2 generations. Under this model, the mean number of generations needed for the allele to fix would be \(MT1 + (D' - 2)MT1a + MT2\). Supporting Information Appendix S3 gives an example notebook that calculates this time.

**Weak-migration approximation.** In the limit of weak migration relative to selection \((m \ll s)\), it is likely that the allele would be fixed in the preceding deme at the time when it establishes in the focal deme. In this case, by setting \(q_2 = 1\) in eqns 11–12, a single differential equation is produced:

\[
\frac{dp_2(t)}{dt} = sp_2(t)(1 - p_2(t)) + m(1 - p_2(t)) \quad (15)
\]

This can be easily solved:

\[
p_2(t) = 1 - \frac{(m + s)(1 - 2Nsce^t)}{s(1 - 2Nsce^t - e^{m+s/t}(1 - 2Nsme^t))} \quad (16)
\]

This form of \(p_2(t)\) can be used with eqns 13 and 14 to obtain a weak-migration approximation for MT1a. As for model one, it is possible to approximate the product term in eqn 13 since each compound probability \(P_2(t)\) is small:

\[
\exp\left( -\int_0^{t-1} N_d P_2(t')dt' \right) = \frac{2(m + s)e^{\gamma - \frac{mt}{1+\gamma}}N_d}{2Nsce^t + e^{m+s}(1-e^t)(1+2Nsme^t)-1}^{2Nem} \quad (17)
\]

Together with eqn 14, this approximation can be used to produce an analytical formula for the mean fixation time, which is derived in Supporting Information Appendix S4.

**Correction for multiple demes in a two-dimensional population**

The above derivations are good starting models where the spread of an advantageous allele can be described as a series of sequential migrations to connected demes along a linear path. This assumption holds, for example, if demes are arranged in a circular formation, with migration possible between one of its two neighbouring demes (hereafter denoted as the ‘one-dimensional’ case). However, if there exist multiple paths along which the advantageous allele can travel, given the first deme that it has migrated to, then these approximations overestimate the time taken for an advantageous allele to fix. This situation arises if demes are arranged in a grid over a two-dimensional torus, with migration possible from a deme to one of its four neighbours (hereafter denoted as the ‘two-dimensional’ case).

Without loss of generality, assume that the advantageous allele starts in the centre of the grid in a two-dimensional torus population and has to migrate to a deme that lies furthest away from where the allele first arose. For a \(3 \times 3\) array of demes, there are only two possible paths that the allele can take to a specific endpoint (the furthest deme from where the allele first appeared), which cover the shortest possible distance (Fig. 1a). Therefore, the calculations needed to calculate the allele fixation time are equivalent to a one-dimensional model with \(D' = 3\). However, for a \(5 \times 5\) array of demes, there are multiple routes that the advantageous allele can travel along to reach the furthest deme, given the first neighbouring deme that the allele migrates to and establishes in. Figure 1b shows a sample of these routes.

Because of these multiple paths, the second model needs to be altered to consider these differing migration effects. This derivation is altered in two ways. First, the migration coefficient is scaled to reflect the fact that each deme is connected to more than two neighbours. Second, the multiple routes that an adaptive allele can take to fixation is also taken into account. These points...
are addressed in turn; Supporting Information Appendix S5 contains example code for implementing these corrections.

Correcting model two to account for multiple neighbours

With populations arranged in a two-dimensional structure, the migration value used in the models had to be set to half that used when applied to one-dimensional populations. This change represents the variance in migration across an individual axis (for migration between adjacent demes), which is half the overall migration rate. Thus, eqns 11–12 are rewritten as:

\[
\frac{dp_2(t)}{dt} = sp_2(t)(1 - p_2(t)) - \frac{m}{2}p_2(t) + \frac{m}{2}q_2(t) \quad (18)
\]

\[
\frac{dq_2(t)}{dt} = sq_2(t)(1 - q_2(t)) + \frac{m}{2}p_2(t) - \frac{m}{2}q_2(t) \quad (19)
\]

with the probability that an establishing migration event occurring being equal to \( P_2(t) = smp_2(t) \). Similarly, the migration coefficient in Fisher’s travelling-wave solution \( 2\sqrt{sm} \) is scaled by \( 1/\sqrt{2} \), representing the variance in migration over two dimensions. This scaling is discussed further in Appendix 1.

Correcting to account for multiple paths to fixation

Because of the multiple paths that an adaptive allele can take when spreading through the entire population, the second model needs to be altered to take these extra routes into account. Each possible path is considered in turn, and for each deme that lies along it, the number of possible entrance points and exit points are considered in determining how migration affects the frequency of the allele within demes, or the probability of the allele establishing in a neighbour. It will be shown that this adjustment will offer an accurate correction for the population structures considered here, due to the small number of paths considered.

Equations 18–19 are altered to account for the fact that certain demes experience a greater influx of migrants than others, or that there are multiple demes that the advantageous allele can migrate to, whilst travelling to the furthest point. If there exists a deme on the path in which there exist two possible entrance points for the allele, then we consider migration contributing new copies of the mutation into the focal deme from two preceding demes. As an approximation, the usage of \( q_2(t) \) is changed so that in this case it represents the mean frequency of the allele in both these preceding demes, which is equal to the allele frequency in a single deme under the previous model (eqns 18–19). This is a valid simplification to make if the frequency of the selected allele in both subpopulations is approximately equal to each other at the time when it establishes in the focal deme. This assumption is reasonable since the alleles spreads in all directions at equal speed, and the selective advantage of the mutant is the same in all demes in this example. Therefore, the coefficient of migration used in the equations increases by a factor of two, so for a deme experiencing input of adaptive alleles from two neighbours, eqns 11–12 are used to model the increase in allele frequency. Similarly, if there are two possible exit points that an advantageous allele can take in order to reach the same end deme, eqn 13 is calculated with \( P_2(t) = 2smp_2(t) \) instead for that deme, as for a one-dimensional population. For a two-dimensional grid, these are the only changes that need to be made to the original equations, since no more than two demes can feed advantageous alleles into another deme at any time, nor are there more than two possible neighbours for which an allele can then travel to if only considering the shortest possible paths linking the original deme in which the allele arose to a specific corner deme. Whilst there can be three possible exit points for an advantageous allele if spreading through a two-dimensional population, only two of these exits take the allele closer towards a specific final deme that lies furthest from where the allele first arose (Fig. 1). Otherwise, the allele is heading to a different furthest deme or doubling back on itself.

To demonstrate how this correction can be implemented, a \( 5 \times 5 \) grid of demes is used as the simplest possible model for which the adjusted equations can be applied to. However, it should be noted that if this correction was to be applied to a system with a larger number of demes, then the following derivation would have to be altered to take into account extra paths that may not be present in this specific example. Nevertheless, it will be shown that a scaled version of this correction is accurate for populations consisting of a large number of demes (\( D = 100, \text{equivalent to } D' = 10 \)) with high migration rates (\( Npm \geq 1 \)). From Fig. 1b, it can be seen that out of the six possible paths, two of them pass through a deme with one possible entrance and two possible exits, one with one entrance and one exit, and a third with two entrances and one exit. Similarly, there are four paths passing through a deme with one entrance and two exits, a second deme with two entrances and two exits, and a third deme with two entrances and one exit. By averaging over all these possible combinations, a corrected form of eqn 14 is obtained that accounts for the increased speed at which the advantageous allele spreads at. Let \( T_a \) be the mean time taken for the allele to migrate to a neighbour, if present in a deme with one entrance and two exits; \( T_b \) the mean time if a deme has one entrance and one exit; \( T_c \) the mean time if a deme has two entrances and one exit; and \( T_d \) the mean time if a deme has two entrances and two exits. So, for example, \( T_a \) is calculated using eqns 18 and 19 to determine the
frequency of the allele at a specific time, then \( P_s(t) = 2smp_r(t) \) is used to calculate the probability so that it then establishes in a neighbour at time \( t \). By the above reasoning, the mean time taken to migrate in intermediate demes, \( MT_{1a} \), is now:

\[
MT_{1a} = \left( \frac{1}{2} \left( T_a + T_b + T_c \right) + 4 \left( T_a + T_d + T_c \right) \right) / 6
\]  

Note that the above formulation does not take into account paths that wrap around the torus in order to travel to the end deme. However, the results will show that even without considering these paths the corrected calculation is very accurate, as the ratio of different paths with a certain number of entrance and exit points, as given by eqn 21, would remain the same.

**Simulation methods**

In order to test the accuracy of these models, the analytical results were compared to values obtained from stochastic simulations coded in C, which track the spread of an advantageous allele through different types of subdivided population. Simulations start with \( N \) haploid individuals divided equally over \( D \) demes, with \( N_D = N/D \) individuals present per deme. Both one-dimensional and two-dimensional structures are simulated.

A new generation is created according to a Wright–Fisher sampling scheme (Fisher, 1930; Wright, 1931). Within each deme, a parent is randomly selected with probability proportional to its fitness and then cloned to produce an offspring. This is repeated \( N_D \) times so that the whole deme is regenerated, which is then repeated for all demes. Individuals then migrate to neighbouring demes. The number of migrants is chosen from a Poisson distribution with mean \( N_{ps} \). \( m \) is the same between each pair of neighbouring demes. For each deme, a migrating individual is chosen at random, then moved to a randomly chosen neighbour. An individual from the neighbour is then moved back to the focal deme, so that \( N_D \) is kept constant.

Initially, the advantageous allele is introduced into a single, randomly selected individual in the first deme. The allele increases the fitness of the individual from 1 to \( 1 + s \); \( s \) is the same in all demes that the allele resides in. The population then undergoes subsequent selection followed by migration until the mutant is fixed or lost in all demes. If it is fixed, it is noted how many generations it took. This is repeated until the allele fixes 1000 times, so that the mean fixation time with a 95% confidence interval is produced.

**Model vs. simulation results**

To test these models, simulations were run with \( N_{ps} \) varying between 0.1 and 5, and \( N_{ps} \) initially varying between 10 and 50. The first results shown here compare the accuracy of the models for one-dimensional structure with five demes (so the maximal distance \( D' = 3 \) or 11 demes \( D' = 6 \), as well as two-dimensional structure with a grid of either 3 \( \times \) 3 demes \( D' = 3 \), or with a 5 \( \times \) 5 grid \( D' = 5 \). The weak-migration approximation (eqn 14 using eqn 17) is presented here for one-dimensional populations only. For comparison, simulation results are compared to the fixation time predicted using Fisher’s (1937) travelling wave model, \( t = D' / (2\sqrt{s}m) \). Whilst plots are only shown here for \( N_D = 2000 \) (except for the cases with a large number of demes), the behaviour outlined below is qualitatively similar for \( N_D = 500 \) and 1000.

For one-dimensional model results, the second model result is very accurate for nearly all \( N_{ps} \) cases for \( D' = 3 \) (Fig. S2) and \( D' = 6 \) (Fig. 2a,b; see also Fig. S3a,b for \( N_{ps} = 0.5 \) and 1 results). The exception lies for \( N_{ps} = 10 \), where the first model more accurately matches up with the simulation results for \( D' = 3 \), and \( D' = 6 \) with \( N_{ps} = 0.1 \) (Fig. 2a). If \( N_{ps} = 0.1 \), the weak-migration approximation slightly underestimates simulation results but is generally accurate (Fig. 2a). As expected, this approximation greatly underestimates fixation time if \( N_{ps} = 2 \) (Fig. 2b). Fisher’s approximation always underestimates the fixation time, especially for \( N_{ps} = 10 \), as the allele does not continuously spread through space.

It was also tested whether these models were still accurate if the overall population consists of a large number of demes. Figure 2c,d shows that with a one-dimensional structure, the second model is accurate if there are 101 demes \( (D' = 51) \) (see also Fig. S3c,d for \( N_{ps} = 0.5 \) and 1 results). As \( N_{ps} \) increases, Fisher’s approximation starts overlapping with simulation results, suggesting that the fixation time of the allele can be modelled as a travelling wave in continuous space for these particular parameters. As with \( D' = 6 \), the weak-selection approximation is accurate but slightly underestimates the fixation time if \( N_{ps} = 0.1 \).

The models are also accurate when applied to a population spread over a two-dimensional torus. For \( D' = 3 \) (Fig. S4), simulation data closely matches with the predictions of model two, with the exception of \( N_{ps} = 10 \) where both models underestimate simulation results. If \( D' = 5 \) (Fig. 3a,b; see also Fig. S5a,b for \( N_{ps} = 0.5 \) and 1 results), then both models initially overestimate the simulation result, with the exception of \( N_{ps} = 2 \) for \( N_{ps} = 10 \). However, once corrected to account for multiple paths (as outlined in the previous section), model two then matches up accurately with simulations for \( N_{ps} \) between 20 and 50, and \( N_{ps} = 10 \) for \( N_{ps} = 0.1 \). These results also verify the fact that one only needs to consider up to two possible exit points per deme to obtain accurate estimates of fixation time for the corrected version of model two.

For a two-dimensional population with 100 demes \( (D' = 10) \), the corrected form of model two with
MT1a (as given by eqns 20 and 21) had to be scaled by $8/3$, so all the coefficients in eqn 21 summed to 8, which is the number of intermediate demes ($D_0^2$). After this change is made, the corrected form of model two is accurate for $N_Dm = \frac{1}{2}$ (Fig. 3d) and $N_Dm = \frac{1}{4}$ (Fig. S5d), but significantly overestimates simulation results for smaller migration rates and $N_Ds \leq 30$ (Fig. 3c; see also Fig. S5c). This discrepancy probably arises due to the presence of more paths that an advantageous allele can take whilst fixing compared with populations consisting of fewer demes, which are not accounted for in the original derivation.

Next, it was investigated how the accuracy of each model changed with different values of the migration rate, $N_Dm$. Figure 4 plots the fixation time of an advantageous allele as a function of the migration rate $N_Dm$, in populations consisting of a small number of demes ($D = 11$ for one-dimensional models, and $D = 25$ for two-dimensional populations). This was investigated with two different values of $N_Ds$ (10 and 50). In one-dimensional models (Fig. 4a,b), model two provides a very good match to simulation data for all $N_Dm$ values, with the corrected version of model two providing the most accurate match in two-dimensional populations (Fig. 4c,d). The exception is if $N_Ds = 0.1$ and $N_Dm = 0.1$ in two-dimensional populations if $N_Ds = 50$, where all models overestimate the actual fixation time. As expected, the weak-migration approximation is only accurate for $N_{Dm} \approx 0.1$ in one-dimensional populations (Fig. 4a,b). It is also observed that Fisher’s approximation starts to match up with simulation results if the migration rate is low ($N_{Dm} \leq 0.5$), and the allele is strongly selected for ($N_{Ds} = 50$). Otherwise, the analytical models presented in this paper provide a better match with simulation data. The same behaviour is also observed if there are a large number of demes ($D = 101$ for one-dimensional models, and $D = 100$ for two-dimensional populations; Fig. S6). It was also determined that the second model provides a good match with simulation data for $N_{Dm} = 5$ (see Fig. S7 for plots using different values of $N_{Dm}$), although there is no single accurate model for $N_{Ds} = 10$.

One implicit assumption of the analysis is that the overall strength of selection is large, so each allele

Fig. 2 Fixation time of an advantageous allele where the population is divided over a one-dimension structure with 11 demes ($D' = 6$, a and b) and with 101 demes ($D' = 51$, c and d). Results are plotted for the first model (black dashed line), second model (dark grey diamonds), weak-migration approximation (dark grey dashed line), simulation results (black crosses, standard errors lie within the markers) and Fisher’s approximation (gray dash-dotted line). $N_D = 2000$ (a and b) or $N_D = 500$ (c and d), and $N_{Dm} = 0.1$ (a and c) or $N_{Dm} = 2$ (b and d). Note that second model results are plotted as points because the numerical evaluation of eqns 11–12 makes it impractical to plot results as a curve.
increases in frequency within each deme in a deterministic manner. To test how robust these models are for weak selection, Fig. 5 shows how they compare against simulations with $N_D = 100$ for the overall population size (so $N_D s = 1$), where there is a large stochastic component determining the frequency of the allele in each deme. For $N_D m = 0.1–2$, the first model matches up well with simulation data, with model two slightly underestimating the fixation time. For $N_D m = 5$, both models slightly underestimate the simulation fixation time, and Fisher’s travelling-wave model matches up best instead. Here, migration is more stronger than selection so the allele spreads in a continuous manner.

**Discussion**

This paper shows how a mixture of deterministic models measuring the increase in allele frequency within demes, combined with a stochastic analysis of the mean time needed for the allele to establish in a new area, can be combined in order to produce an analytical estimate of the fixation time of advantageous allele in a subdivided population consisting of multiple demes. It is shown that the second model outlined here, which accounted for migration altering frequencies of the advantageous allele in intermediate demes, provides a very good estimate of the fixation time for nearly all cases. This second model needed to be corrected if applied to a two-dimensional structure with a large number of demes, due to the fact that an advantageous allele can take multiple routes between the deme that it first migrated to and to the most distant deme, with intermediate populations having different migration properties. This correction, when applied to the second model, makes it match up very accurately with simulation data. However, although the second model is generally more accurate for $N_D s = 10$, it can be inaccurate if the migration rate is weak and the allele is strongly selected for (Fig. 4d). This implies that for low migration rates, there are extra stochastic effects present that the model does not take into account. These stochastic effects might significantly affect the fixation time if each deme is connected to a large number of neighbours, since this will increase the

---

**Fig. 3** Fixation time of an advantageous allele where the population is divided over a two-dimension structure with 25 demes ($D' = 5$, a and b), and 100 demes ($D' = 10$, c and d). As well as plotting the simulation data and two model results, the corrected version of model 2 that accounts for the different ways in which an advantageous allele can reach a target deme is also shown (light grey circles). $N_D = 2000$ (a and b) or $N_D = 500$ (c and d), and $N_D m = 0.1$ (a and c) or $N_D m = 2$ (b and d).
probability that the advantageous allele successfully establishes in a specific deme. Therefore, a full stochastic treatment should be investigated as part of future work in order to produce a more complete model. Model two also appear to be robust if there are a large number of demes (Figs 2 and 3c,d), although results can be inaccurate in a two-dimensional model with weak selection and migration (Fig. 3c).

By analysing the model, a few key properties of mutant fixation become apparent. An advantageous allele can fix in a two-dimensional structure more quickly than in a one-dimensional model with the same number of demes. This is for two main reasons; first, it is clear that in the two-dimensional case, each deme is connected to more neighbours compared to a deme in a one-dimensional population, so a selected allele can spread through the entire population more quickly. This is reflected by the effective number of demes, $D_0$, that an allele has to travel across, being greatly lower in two-dimensional structures ($D_0 = O(\sqrt{D})$, as opposed to $D_0 = O(D)$ in one-dimensional cases). A more original conclusion is that in two-dimensional structures, the fixation time of an advantageous allele is greatly decreased due to the different paths that it can take, as opposed to assuming that the influx and efflux of the advantageous allele are the same between all demes. This means that some demes experience a greater input of migrants than others, so the allele will increase in frequency faster within these subpopulations. Therefore, the allele will spread faster overall. This conclusion is reflected in the correction applied to model two, which is needed in order to produce an accurate approximation for a large number of demes.

Generally, this analysis has shown that the fixation time of an allele in a subdivided population is reduced by migration effects introducing more copies of an allele after it has established itself in a new deme. This behaviour may alter previously investigated effects of population subdivision, such as how levels of heterozygosity at linked neutral sites are changed or whether there exists adequate gene flow between demes to prevent the populations from diverging. Kim & Maruki (2011), for example, showed how the level of heterozygosity at a linked locus is greatly reduced in demes that lie nearest to where the sweep originated, reflecting how population subdivision delays the fixation of a novel
advantageous allele, thus allowing more recombination to occur (Barton, 2000). This analysis suggests that since migration increases the speed at which the allele fixes in populations consisting of multiple demes, heterozygosity levels would not be broken down to a greater extent, compared to models where such migration effects were not considered. Future work should aim to implement the findings of this analysis into models of genetic hitchhiking, to accurately quantify how heterozygosity would be broken down in stepping-stone populations.

Secondly, this analysis can tell us more on whether gene flow is too low in populations in order to prevent them from diverging, as discussed by Ehrlich & Raven (1969). In a review paper outlining existing data on migration rates, Morjan & Rieseberg (2004) found levels of gene flow to be higher than previously thought, but concluded that ‘there are many species...that lack sufficient gene flow to prevent divergence’. This analysis demonstrates how even in populations with low migration levels, copies of new alleles can be transferred to new subpopulations by migration, thus increasing levels of gene flow between demes. The increase in fixation time can be substantial if subpopulations are closely connected, as in a two-dimensional model (Fig. 3).

This analysis has also highlighted the need to investigate the manner in which populations are connected in natural systems, in order to understand how migration affects the spread of advantageous alleles. The degree of connectivity and type of population structure can have drastic effects on allele fixation time, so information on the manner in which communities are structured would also need to be estimated from field studies, in order to obtain an accurate estimate of fixation time.

Overall, this study highlights how even a modest amount of migration can affect the transfer of alleles into new demes and decrease the fixation time of a selective sweep in structured populations. Future studies of hitchhiking, estimating the probability that neighbouring areas diverge, and other processes affected by population subdivision should take this finding into account, in order to accurately determine the impact migration has on these.

Acknowledgments

I would like to thank Peter Keightley, Nick Barton, Bill Hill, Mike Whitlock and anonymous reviewers for providing helpful comments on the manuscript. I am funded by a Biotechnology and Biological Sciences Research Council (BBSRC) studentship.

References


**Appendix 1**

**Derivation of the migration coefficients used in models**

In order to apply the models to different types of structured populations, the migration rate used in the models is not necessarily the same as the total migration rate between demes. Rather, the migration coefficient used in both models and Fisher’s travelling-wave solution should be the variance in migration over demes, given the directions in which an individual can migrate (Fisher, 1937).

For a one-dimensional model, denote a migration from a focal deme to the population to the left by $-1$ and a migration to the right by $+1$. Since it is equally likely that a migrant will move in either direction, the mean migration coefficient is $1/2 \cdot (-1) + 1/2 \cdot 1 = 0$. Therefore, the variance is $E[X^2] - E[X]^2 = (1/2 \cdot 1 + 1/2 \cdot 1 - 0) = 1$. So in one-dimensional populations, the migration rate used in both models and Fisher’s solution is the same as the overall migration rate between demes.

This is easily extended to a two-dimensional torus model. Denote a migration from a focal deme up to the population above it by the two-dimensional value $(0,1)$, a migration to the right by $(1,0)$, and the negative of these values for migrations down and left, respectively.

$E[X^2] - E[X]^2 = (1/2 \cdot 1 + 1/2 \cdot 1 - 0) = 1$. So in one-dimensional populations, the migration rate used in both models and Fisher’s solution is the same as the overall migration rate between demes.

In order to apply the models to different types of structured populations, the migration rate used in the models is not necessarily the same as the total migration rate between demes. Rather, the migration coefficient used in both models and Fisher’s travelling-wave solution should be the variance in migration over demes, given the directions in which an individual can migrate (Fisher, 1937).

For a one-dimensional model, denote a migration from a focal deme to the population to the left by $-1$ and a migration to the right by $+1$. Since it is equally likely that a migrant will move in either direction, the mean migration coefficient is $1/2 \cdot (-1) + 1/2 \cdot 1 = 0$. Therefore, the variance is $E[X^2] - E[X]^2 = (1/2 \cdot 1 + 1/2 \cdot 1 - 0) = 1$. So in one-dimensional populations, the migration rate used in both models and Fisher’s solution is the same as the overall migration rate between demes.

This is easily extended to a two-dimensional torus model. Denote a migration from a focal deme up to the population above it by the two-dimensional value $(0,1)$, a migration to the right by $(1,0)$, and the negative of these values for migrations down and left, respectively. Again, since there is equal probability of a migration in either of these directions, the mean value is $(0,0)$ and the variance is $(1/2,1/2)$. Hence, when applying model results to two-dimensional populations, the migration rate used is half that in simulations, as this is the variance between two adjacent demes (across the up-down or left-right axes). For Fisher’s solution, the migration rate is scaled by the magnitude of this vector to represent diffusion across both dimensions. This is equal to $1/\sqrt{2}$.

**Supporting information**

Additional Supporting Information may be found in the online version of this article:

**Appendix S1–S5** Derivations as a Mathematica notebook (reader available from http://www.wolfram.com/products/player/).

**Figure S1** The proportion of time contributed by MT2 to model one, as a function of the effective number of demes $D'$ (red line).

**Figure S2** Fixation time of an advantageous allele where the population is divided over a one-dimensional structure with five demes ($D' = 5$).

**Figure S3** Fixation time of an advantageous allele where the population is divided over a one-dimensional struc-
ture with 11 demes ($D' = 6$, a and b), and with 101 demes ($D' = 51$, c and d).

**Figure S4** Fixation time of an advantageous allele where the population is divided over a two-dimensional structure with nine demes (so $D' = 3$).

**Figure S5** Fixation time of an advantageous allele where the population is divided over a two-dimensional structure with 25 demes ($D' = 5$, a and b), and 100 demes ($D' = 10$, c and d).

**Figure S6** Fixation time of an advantageous allele as a function of the migration rate $N_Dm$, where the population is divided over a one-dimensional structure with 101 demes ($D' = 51$, a and b), or a two-dimensional torus with 100 demes ($D' = 10$, c and d).

**Figure S7** Fixation time of an advantageous allele for $N_D = 2000$ and $N_Dm = 5$.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Data deposited at Dryad: doi: 10.5061/dryad.407j1

Received 30 January 2012; accepted 14 May 2012