Journal of Theoretical Biology 365 (2015) 486–495

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Population structure and the rate of evolution

Xinzhu Wei [a], Lei Zhao [a], Martin Lascoux [a,b,*], David Waxman [a,*]

[a] Centre for Computational Systems Biology, Fudan University, Shanghai 200433, People's Republic of China
[b] Department of Ecology and Genetics, Evolutionary Biology Center, Uppsala University, Uppsala 75236, Sweden

HIGHLIGHTS

- We analysed a two patch model in which the spatial average of selection is zero.
- We examined effects of population size, migration and selection on the rate of evolution.
- For large population sizes, the size appears only in scaled parameters, not on its own.
- We assume that scaled migration and selection strengths are dependent due to ecology/evolution.
- Ecology/evolution affects how the rate of evolution varies with population size.

ARTICLE INFO

Article history:
Received 28 June 2014
Received in revised form
9 October 2014
Accepted 31 October 2014
Available online 11 November 2014

Keywords:
Population subdivision
Migration
Spatially dependent selection
Wright–Fisher model
Diffusion approximation

ABSTRACT

The way population size, population structure (with migration), and spatially dependent selection (where there is no globally optimal allele), combine to affect the substitution rate is poorly understood. Here, we consider a two patch model where mutant alleles are beneficial in one patch and deleterious in the other patch. We assume that the spatial average of selection on mutant alleles is zero. We take each patch to maintain a finite number of N adults each generation, hence random genetic drift can independently occur in each patch. We show that the principal way the population size, N, when large, affects the substitution rate, R_∞, is through its dependence on two composite parameters. These are the scaled migration rate M ($\propto$ population size $\times$ migration rate), and the scaled selection intensity S ($\propto$ population size $\times$ beneficial effect of a mutant). Any relation between S and M that arises for ecological/evolutionary reasons can strongly influence the way the substitution rate, R_∞, depends on the population size, N. In the simplest situation, both M and S are proportional to N, and this is shown to lead to R_∞ increasing with N when S is not large. The behaviour, that R_∞ increases with N, is not inevitable; a more complex relation between S and M can lead to the opposite or other behaviours. In particular, let us assume that dM/dN is positive, as would occur if the migration rate were constant, S is not large, and S depends on M (i.e., $S=S(M)$). We then find that if $S(M)$ satisfies $S(M) > ((1+M)/\sqrt{1+2M})S(0)$ then the substitution rate, R_∞, increases with N, but if $S(M) < ((1+M)/\sqrt{1+2M})S(0)$ then R_∞ decreases with N.

1. Introduction

The relative importance of demography and selection, for a number of issues in modern evolutionary biology including the rate of evolution, has been a topic of active debate. The founders of population genetics held different views on this issue (Fisher, 1930; Haldane, 1932; Wright, 1968). In brief, Wright gave more importance to random genetic drift and population structure in shaping the pattern of genetic diversity than either Fisher or Haldane. Furthermore, following Kimura's development of the neutral theory of molecular evolution (Kimura, 1968), the debate intensified (Akashi et al., 2012).

For the rate of evolution, two aspects of demography have received much attention and continue to do so. These are

(i) the effects of population size (see, e.g., Ohta, 1972; Gillespie, 2001; Akashi et al., 2012; Balloux and Lehmann, 2012; Lanfear et al., 2014);
(ii) the effects of population structure (see, e.g., Gavrilets and Gibson, 2002; Frearn et al., 2013).

In an important paper, Ohta (1972) argued that the speed of evolution is likely to be higher in small populations than in large ones. Her main

* Corresponding authors.
E-mail addresses: martin.lascoux@ebc.uu.se (M. Lascoux), davidwaxman@fudan.edu.cn (D. Waxman).

http://dx.doi.org/10.1016/j.jtbi.2014.10.041
0022-5193/

argument was based on the highly plausible idea that the environment experienced by a small population is likely to be more uniform than the corresponding environment of a large population. A random mutant is therefore more likely to be beneficial in a small population than in a large population – where it would have to be beneficial under a larger set of environmental conditions. In other words, decreasing the population size results in an increased proportion of beneficial new mutations, and further, "the mean selection coefficient of all beneficial mutants is also larger" (Ohta, 1972).

The measure of the speed of evolution which Ohta had in mind was the rate of substitution, which is defined as the mean number of mutations that fix/generation. Mutations stochastically arise in a population, and their subsequent fate is also stochastic: most new mutations are rapidly lost, but a small fraction achieve fixation. The rate of substitution characterizes the flux of new mutations which ultimately achieve fixation and hence lead to long lived changes in a population.

Assuming discrete generations, the rate of substitution can be decomposed into a product of two factors:

(i) the expected number of mutations that arise each generation in the population;
(ii) the probability of fixation of a new mutation.

The expected number of mutations that enter a given region of the genome of a population, each generation, is $2Nu$ where N is the number of adults in the population and u is the mutation rate of the genomic region. Thus if we focus on a single locus, then u is the allelic mutation rate, i.e., the probability that a gene in a gamete contains a genetic change, relative to the parental gene.

The probability of fixation of a new mutation has a mathematical form which depends on the selective effect of the mutation. Neutral mutations have a probability of fixation that equals their initial frequency, namely $1/(2N)$. For this case, the substitution rate is $2Nu \times 1/(2N)$,, i.e., it is simply equal to the mutation rate, u, and hence is independent of the population size (Kimura, 1983). This result seems surprisingly robust (Lanfear et al., 2014). There is a vast literature on the probability of fixation of a new mutation which is subject to selection (see, e.g., Fisher, 1930; Haldane, 1932; Kimura, 1962; and for recent work see, e.g., Waxman, 2011). These results can be used in the determination of the substitution rate, when mutations are selected.

Ohta's model involves a single population. A way to introduce environmental heterogeneity is to explicitly assume a subdivided population that consists of patches that are connected by migration. Since Wright's seminal work on this subject (Wright, 1931), subdivided populations and dispersal have played a major role in evolutionary biology. In particular, subdivided populations lie at the core of models of local adaptation (Savolainen et al., 2013). When these more complex (and more realistic) situations occur, the rate of substitution can show different dependencies on the population size. For example, the rate of substitution in a small population can (sometimes) be higher than that in a large population. This will occur if, for a small population, the probability of fixation of a mutation is sufficiently large that it more than compensates for the lower rate at which mutations enter the population. For example, in Ohta's model, the rate of substitution in a small population will be higher than that in a large population due to the latter's higher level of heterogeneity, and hence reduced effectiveness of selection (a higher proportion of mutations entering a large population are deleterious).

Let us consider the effects of selection and population subdivision on the substitution rate. In the simplest case of additive selection at a single locus in an unstructured randomly mating population, beneficial mutations result in a large population size producing a higher substitution rate than a small population size. By contrast, if the mutations are deleterious, then a large population size will produce a smaller substitution rate than a small population size (see, e.g., Lanfear et al., 2014).

In the case of a structured population with spatially dependent selection matters are more complex. When mutant alleles are *beneficial in one patch*, and the same alleles are *deleterious in another patch*, but the patches are connected by migration and the spatial average of selection on mutants is *zero*, it is unclear under which circumstances the substitution rate will be larger in a large population than in a small population. Another way of saying this is that it is unclear which has more influence on the substitution rate: the beneficial aspect of the alleles in one patch, or their deleterious aspect in the other patch.

In the present work, we investigate the dependence of the rate of substitution on population size, and in particular, the circumstances or conditions where it can be larger in a small subdivided population than a large subdivided population. While this is somewhat motivated by Ohta's hypothesis, we explicitly include

(i) finite population size,
(ii) population structure,
(iii) spatially dependent selection.

Frearn et al. (2013) have considered some specific spatial models, such as star-structured populations and two-dimensional lattices. Their main conclusion is that the time to fixation of new mutations, and thereby the rate of evolution, may be strongly affected by population structure. For instance, 'star graphs' increase the probability of fixation of positively selected mutations while simultaneously decreasing that of disadvantageous ones. However, the expected fixation time is significantly increased. It should be noted that the structure they are referring to is local population structure, and each node in the lattice or star graph is occupied by a single individual. These models are therefore quite different from the ones that are classically used in population genetics, where each node (= deme or subpopulation) is occupied by many individuals that can be exchanged between populations. Gavrilets and Gibson (2002) considered population structure with selection having different directions in different demes. In their work, these authors determined approximations for the probability of fixation of a new mutant allele, within the framework of the diffusion approximation. They proceeded to discuss the implications of their results for the rate of substitution. Their main conclusion was that it is possible for small subdivided populations, with low migration rates, to have higher substitution rates than large subdivided populations, when the intensity of selection over space exceeds a threshold. Building on a few examples and limiting cases they conjectured that this would likely be true for a convex dependence of migration rate on population size.

We begin our analysis, of how the rate of substitution depends on the population size, by first specifying a simple model that we term the *Basic Model*. This model captures many of the essential features of the problem, and allows an exploration of the key issues.

In the Basic Model there are two patches in which each maintain N adults in each generation. When the population size of each patch, N, is finite, random genetic drift can occur. We shall work within the framework of a Wright–Fisher model for the two-patch system. Combining this model with ideas/results from diffusion analysis yields numerical results for a finite population that includes migration and selection. This allows us to analyse the dynamics of the system and to show that, to good accuracy, the rate of substitution, relative to the neutral rate of substitution, depends on only two composite parameters. These are the scaled rates of migration and selection, which we write as M and S, respectively, and are defined by

$$M = 4Nm, \quad S = 4Ns \tag{1}$$

where m is the probability of an individual migrating from one patch to another in one generation, and s (≥ 0) is the selection coefficient

associated with a mutant allele in the patch where it is beneficial (the same allele is deleterious in the other patch, where it has a selection coefficient of $-s$).

The dependence of the substitution rate on the composite parameters M and S indicates that the association between the rate of evolution and the population size, within the Basic Model, is most naturally discussed in terms of M and S rather than in terms of N, m and s separately. In particular, it becomes natural to investigate how any relationship between the scaled strengths of migration and selection, i.e., any relationship between M and S, affects the rate of evolution, rather than the relationship between N and m, as investigated by Gavrilets and Gibson (2002).

2. Basic Model

The Basic Model that we adopt in this work has the following ingredients (cf. Gavrilets and Gibson, 2002):

(i) A population which resides at two patches, termed Patch 1 and Patch 2.
(ii) Symmetric migration between the two patches, which occurs at rate (i.e., probability/generation) of m. We shall assume m is small ($m \ll 1$).
(iii) Spatially dependent selection, which acts multiplicatively (i.e., selection is genic) at all loci, with each locus having two alleles which are denoted by a and A. The a allele (the resident allele) has a relative fitness of 1 in both patches, while the A allele (the mutant allele) is oppositely selected in the two patches: in Patch 1 the A allele is beneficial, and has a relative fitness $1+s$ (with $s>0$), while in Patch 2 the A allele is deleterious and has a relative fitness $1-s$. Thus the relative fitnesses of the AA, Aa and aa genotypes in Patch 1 are, respectively, $(1+s)^2$, $(1+s)$ and 1, while in Patch 2 they are $(1-s)^2$, $(1-s)$ and 1. We shall assume s is small ($s \ll 1$), in which case multiplicative and additive selection are effectively equivalent.[1]
(iv) The absence of mutation during the time the A allele is segregating. Thus once a single A allele has been produced at a locus by mutation in one of the patches, no further mutations occur at the locus while the A allele is segregating.
(v) Number regulation, which maintains the same number of N adults in each patch, each generation.

Number regulation takes place within the context of a life cycle. We adopt the following life cycle for the processes occurring in one generation.

Adults (Generation t)
↓ produce an effectively infinite number of gametes then die
Gametes
↓ undergo random union after some gametes migrate between patches
Zygotes
↓ undergo selection in the patch they reside
Juveniles
↓ undergo number regulation; carried out independently in each patch, leaving N adults in each patch; random genetic drift occurs at this stage in a finite population
Adults (Generation $(t+1)$)

When the population size is effectively infinite, there are no effects of random genetic drift and the behaviour of the population is deterministic. In this case, the level of polymorphism that can be achieved in the population at equilibrium follows from a balance between migration and selection. This has been extensively studied (Levene, 1953; Bulmer, 1972, and for a general presentation, see Felsenstein, 2013).

When the population size is finite, random genetic drift generally operates and fixation and loss are then possible. The probability of fixation of a single mutant A allele has been obtained under different regimes, different approximations, and generalizations of the Basic Model (Tachida and Iizuka, 1991; Gavrilets and Gibson, 2002; Whitlock and Gomulkiewicz, 2005).

More recently, Yeaman and Otto (2011) determined the conditions under which polymorphism is likely to be maintained when population size is finite and there is recurrent mutation. These authors ensured maintenance of both alleles in the total population by having a high rate of bidirectional mutation, and determined the migration rate "below which populations were locally adapted for a substantial fraction of time in each patch".

3. Results

We base our analysis on a Wright–Fisher model for a single locus of a finite population that is located in two patches. In a conventional Wright–Fisher model, with only a single patch, a state of the population is specified by a single number, namely the number of A alleles. In the two-patch case studied here, the corresponding Wright–Fisher model is more complex since a state of the total population is specified by two independent numbers, namely the number of A alleles in Patch 1 and the corresponding number in Patch 2.

Let the rate of substitution, relative to the neutral rate (which is u), be denoted by R. Henceforth we shall refer to R simply as the substitution rate, although technically it is a *scaled* substitution rate. We use $P_{\mathrm{fix}}(x,y)$ to denote the probability of ultimate fixation of the A allele (i.e., it ultimately achieves a frequency of unity in *both* patches) when it starts at frequency x in Patch 1 and frequency y in Patch 2. The average number of mutations that fix/generation from mutations arising in Patch 1 is $2\times N\times u\times P_{\mathrm{fix}}(1/2N,0)$ and there is a similar expression for mutations arising in Patch 2. Hence we have

$$R=\frac{2\times N\times u\times\left[P_{\mathrm{fix}}\left(\frac{1}{2N},0\right)+P_{\mathrm{fix}}\left(0,\frac{1}{2N}\right)\right]}{u}$$
$$=2N\times P_{\mathrm{fix}}\left(\frac{1}{2N},0\right)+2N\times P_{\mathrm{fix}}\left(0,\frac{1}{2N}\right) \quad (2)$$

and in Appendix A we give full details of the Wright–Fisher model for the two patch model. In Appendix B we show that the diffusion approximation of the Wright–Fisher model suggests that the substitution rate, R, is (with corrections of order N^{-1}) a function only of the scaled migration and selection parameters M and S of Eq. (1). This is indeed found to be the behaviour of the Wright–Fisher model in a numerical investigation (see Appendix C). We thus write

$$R=R_\infty(M,S)+O(N^{-1}) \quad (3)$$

[1] Note that selection acts separately in each patch. This means we can independently assign the precise way selection acts on the two alleles in a patch. Thus while the adopted scheme corresponds to the relative fitnesses of the AA, Aa and aa genotypes in Patch 2 of $(1-s)^2$, $(1-s)$ and 1, we could choose different relative fitnesses (while keeping the original assignments of fitnesses of Patch 1) and still obtain the same dynamical behaviour. For example, we could choose the relative fitnesses in Patch 2 of 1, $(1+s)$ and $(1+s)^2$ (neglecting very small terms of order s^2). Thus the Basic Model has more generality than may initially be apparent.

where $R_\infty(M,S)$ is the substitution rate when M and S have fixed values and N has been allowed to become arbitrarily large ($N\to\infty$).

We shall proceed, neglecting the N^{-1} corrections in Eq. (3), and we shall now give some of the properties of the substitution rate, $R_\infty(M,S)$, which follow from consideration of the Wright–Fisher model combined with inferences from the diffusion approximation.

3.1. Dependence of the substitution rate on M and S

We first note that for any non-zero value of the scaled migration rate, M, the substitution rate satisfies

$$R_\infty(M,0)=1. \tag{4}$$

This follows from Eq. (2) using, e.g., the $s\to 0$ limit of Eq. (8) of Gavrilets and Gibson (2002), which yields $P_{\rm fix}(1/2N,0)=P_{\rm fix}(0,1/2N)=1/4N$. Numerically, we find from the Wright–Fisher model, in the absence of selection and for any population size, that the substitution rate has a value that is indistinguishable from unity. Thus in the absence of selection, the substitution rate coincides with the neutral rate for a single unstructured population, independent of the population size and independent of the migration rate in the two patch problem (cf. Maruyama, 1970).

Next we note that in the Basic Model, selection of the A allele in one patch is equal and opposite to the selection in the other patch, and that apart from this, both patches are equivalent. Given, also, that mutations are equally likely to arise in each patch, the implication is that the rate of substitution, $R_\infty(M,S)$, cannot depend on the sign of S. Making the plausible assumption that the relationship between $R_\infty(M,S)$ and S is smooth, it must be a function of S^2. Thus we assume that for S not large,

$$R_\infty(M,S)\simeq 1+C_2(M)\times S^2 \tag{5}$$

where the coefficient $C_2(M)$, of S^2, is generally a function of the scaled migration rate, M. For values of S smaller than 2 we have numerically found that the approximation in Eq. (5) has an error smaller than 7% for a wide range of M values up to $M=20$ while for S smaller than 1 the error is smaller than 2%. These numerical results give high confidence in the assumptions underlying Eq. (5). In particular, they allow us to rule our dependence of $R_\infty(M,S)$ on the absolute value of S (i.e., $|S|$). Extensions of Eq. (5), so that it applies at larger S, are possible, for example, by including higher order terms in S^2, e.g., $R_\infty(M,S)\approx 1+C_2(M)\times S^2+C_4(M)\times S^4$ and in Fig. 1 we show the dependence of $C_2(M)$ and $C_4(M)$ on M.

In all cases we have looked at, we have found that the coefficient $C_2(M)$ is positive. Positivity suggests that while the effect of the patch where selection is negative is to reduce the substitution rate, this is more than compensated by effect of the opposite, positive selection, in the other patch. Overall, the $\pm s$ scheme of selection of the Basic Model leads to a net increase of the rate of substitution over the neutral case.

In our numerical work, we have found that $C_2(M)$ decreases with M. Such a behaviour is intuitively reasonable. The larger the value of M, the more well-mixed the populations in the two patches are. Hence the selection experienced by a mutant allele becomes closer to an average over both positive and negative selection and becomes increasingly independent of where the allele arose.

We have obtained numerical results for the coefficient of the S^2 in R_∞, namely $C_2(M)$. From a plot of $C_2(M)\times(1+M)^2$ against M we observed a result *very* close to a straight line, with near rational values of the intercept and slope. This suggested that $C_2(M)$ is very close in form to

$$C_2(M)=\frac{1+2M}{12(1+M)^2}. \tag{6}$$

We have investigated Eq. (6) for a range of M from $M=0$ to $M=20$. We find that Eq. (6) is within 10^{-5} of the numerical results. We suspect that Eq. (6) is exact for all M, although we have no mathematical proof of this.

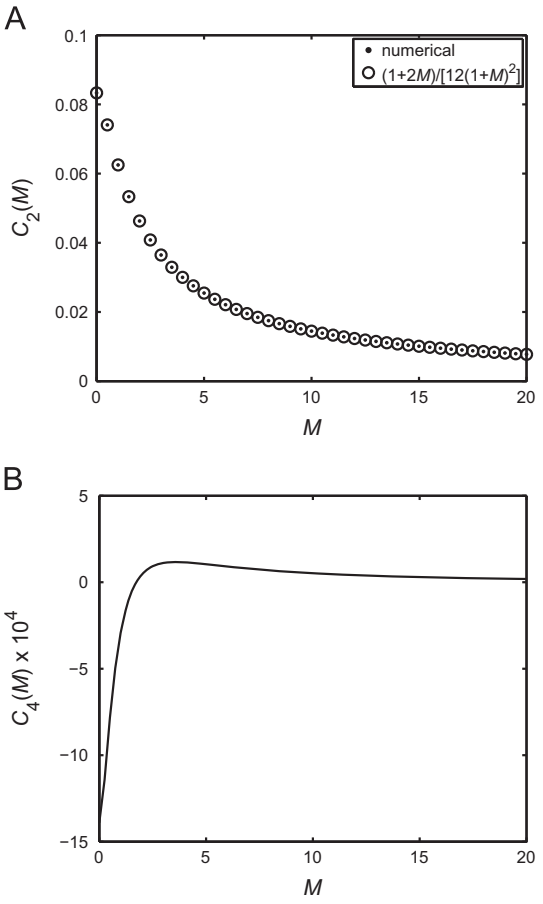

Fig. 1. In Panels A and B we show how the coefficients $C_2(M)$ and $C_4(M)$, in the expansion of the substitution rate $R_\infty(M,S)\simeq 1+C_2(M)\times S^2+C_4(M)\times S^4$, depend on M. In Panel A we also plot the result for $C_2(M)$ given in Eq. (6), which is extremely close to the numerical results for all values of M from 0 to 20 that we have considered. The form of $C_2(M)$ in Eq. (6) behaves as $C_2(M)\simeq 1/12-M^2/12$ for small M ($M\ll 1$) and as $C_2(M)\simeq 1/(6M)$ for large M ($M\gg 1$). From Panel B we note that the largest magnitude of $C_4(M)$ occurs at $M=0$, where $|C_4(0)|\simeq 1.4\times 10^{-3}$. We obtained the figures by first numerically determining $R_\infty(M,S)$ from an extrapolation of the substitution rate of a Wright–Fisher model to $N\to\infty$, when M and S were held fixed. We then repeated the numerical calculation for different S, and fitted a polynomial in S^2 to $R_\infty(M,S)$ to obtain the coefficients $C_2(M)$ and $C_4(M)$ (see Appendix C for details).

In Panel A of Fig. 1 we illustrate how $C_2(M)$ depends on M, while Panel B contains the corresponding plot of $C_4(M)$.

3.2. Substitution rate in the presence of a relationship between M and S

In the region where Eq. (5) applies, i.e., where S not large, we can use this equation to determine a condition on whether the rate of substitution either increases or decreases with the number of adults, N, in each patch, that is, whether dR_∞/dN is positive or negative.

In the very simplest case, where m and s (which appear in $M=4Nm$ and $S=4Ns$) are constants, it is straightforward to show that dR_∞/dN is positive (dR_∞/dN can be written as $(M^2+3M+1)S^2/[6N(1+M)^3]$). Thus in this case the rate of substitution increases with population size.

We now make the assumption that M and S are not independent parameters, but for ecological/evolutionary reasons are *dependent* beyond simply $M/S=$ constant, as would follow when m and s are constants. For instance dispersal may be favored if it

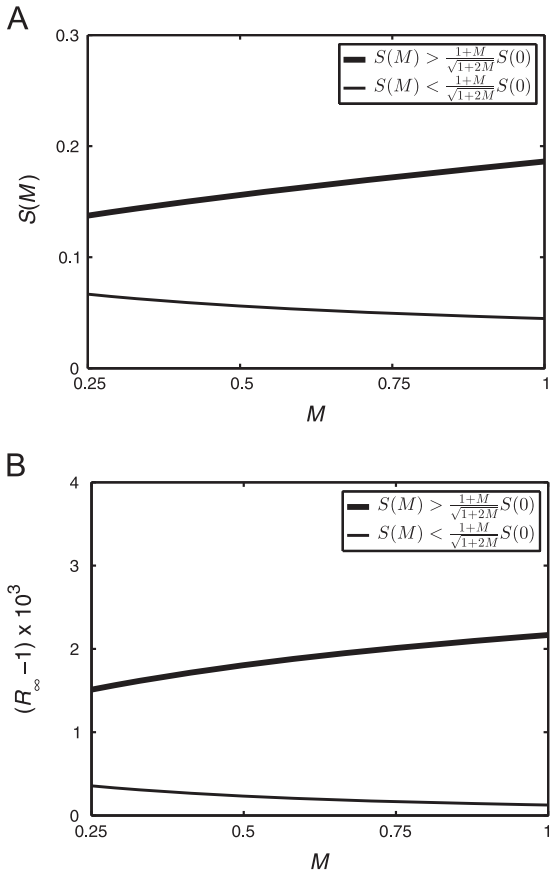

Fig. 2. In this figure we have adopted two different forms for the scaled selection intensity, S, as a function of the scaled migration rate, M, when Eq. (5) applies. The two forms are $S(M)=\frac{1}{10}(1+M)/\sqrt{1+2M}\pm\frac{1}{10}\sqrt{M/2}$. An M dependence of S of the form $S(M)=$ constant $\times(1+M)/\sqrt{1+2M}$ is marginal in the sense it leads to an $R_\infty(M,S(M))$ that is independent of N – because $d(C_2S^2)/dM$ vanishes (see Eq. (7)). We have deviated from this marginal form in a way to obtain $d(C_2S^2)/dM$ being either positive or negative. In Panel A we plot the two different forms of $S(M)$ against M. In Panel B, we assume that the migration rate, m, has a constant value (independent of N), so M ($\propto N$) can be used as a proxy for N. We show that the two different forms of $S(M)$ lead to a substitution rate, R_∞, that either increases or decreases with M, and hence N.

reduces competition between relatives and/or inbreeding depression (Clobert et al., 2001 and references therein). Or we could imagine a mutation which, when M increases, is under higher selection intensity in both patches.

In the present work we assume that S is determined by M (i.e., $S=S(M)$). Thus explicitly, the rate of substitution is now $R_\infty(M,S(M))$, which in this section we shall simply write as R_∞. We have that

$$\frac{dR_\infty}{dN}=\frac{dM}{dN}\frac{d\left(C_2S^2\right)}{dM} \tag{7}$$

and in the present work we assume that dM/dN is positive, as would follow in the particular case where the migration rate, m, is a constant. It then follows from Eq. (7) that R_∞ increases with N if $d\left(C_2S^2\right)/dM$ is positive, and R_∞ decreases with N if $d(C_2S^2)/dM$ is negative. These correspond to the inequalities $d(C_2S^2)/dM\gtrless 0$ from which we obtain $S(M)\gtrless((1+M)/\sqrt{1+2M})S(0)$ where $S(0)$, assumed positive, is the value of $S(M)$ at $M=0$. In this way we obtain the following behaviour of $R_\infty\equiv R_\infty(M,S(M))$:

R_∞ increases with N if $S(M)>\dfrac{1+M}{\sqrt{1+2M}}S(0)$. (8)

R_∞ decreases with N if $S(M)<\dfrac{1+M}{\sqrt{1+2M}}S(0)$. (9)

Eqs. (8) and (9) do not specify the forms of $S(M)$ that lead to R_∞ increasing or decreasing with N, only a property of $S(M)$ that leads to these behaviours. In Fig. 2A we give two curves of $S(M)$ against M, corresponding to two different forms of the dependence of S on M. One form has $S(M)>((1+M)/\sqrt{1+2M})S(0)$ and leads to dR_∞/dN positive, while the other has $S(M)<(1+M)S(0)/\sqrt{1+2M}$ and hence a negative dR_∞/dN. In Fig. 2B the behaviour of the rate of substitution, R_∞ (as given in Eq. (5)), is shown when the two different forms of $S(M)$ of Fig. 2A are adopted.

4. Discussion

Apart from Ohta (1972) and Frearn et al. (2013), who considered structure within a single population, and Gavrilets and Gibson (2002), who studied two demes connected by migration, the effect of population structure on the rate of evolution has received very limited attention. In the present study we derived an explicit formula relating the scaled substitution rate to the scaled rate of migration, $M=4Nm$, and the scaled intensity of selection, $S=4Ns$, for a simple two-deme model with selection. This result appears as a natural extension of results obtained previously in equilibrium studies of two-deme systems (e.g., Yeaman and Otto, 2011) where the maintenance of polymorphism also depends on the balance between selection and migration. Thus our results show that these two parameters are not only involved in the attainment and maintenance of transient polymorphisms but also play a role in the long term evolution of populations and species. We note, however, that in Yeaman and Otto's (2011) study, polymorphism is facilitated by a particularly high rate of bidirectional mutations ($\mu=10^{-4}$) and hence that the equilibrium may actually depend on the three parameters M, S and $4N\mu$ rather than simply on scaled migration and selection parameters.

In our model, the selection pressure has the opposite sign in the two demes. One might expect that the effect of selection would cancel out and that a substitution rate that is close to the neutral rate would follow. However, the substitution rate is *larger* than in the neutral case, and can be appreciably so. The substitution rate has the property that it *increases* when M is *decreased*, so a set of locally adapted populations, with low migration among them, will evolve faster than a single randomly mating population. A similar positive effect of population structure on the rate of evolution, albeit in a rather different type of model, was observed previously by Frearn et al. (2013).

We note that for the Basic Model, Eq. (5) gives a form for the rate of substitution, when S is not large, that we can write as

$$R_\infty(M,S)\simeq 1+C_2(M)\times \mathrm{Var}(S). \tag{10}$$

Here $\mathrm{Var}(S)$ is the variance of selection coefficients across different patches. It is tempting to speculate that when there are a large number of patches, with differing selection strengths, but whose mean is zero, that the rate of substitution will increase with the variance of selection coefficients across different patches, i.e., with $\mathrm{Var}(S)$.

Additionally, it is of interest to see what contribution Patch 1 (where the mutant allele is beneficial) makes to the rate of substitution. We write this contribution as $R_\infty^{beneficial}(N,S)$ and using the methods of the present work we find

$$R_\infty^{beneficial}(M,S)\simeq\tfrac{1}{2}+C_1(M)S+\tfrac{1}{2}C_2(M)S^2 \tag{11}$$

where

$$C_1(M)=\frac{1}{4}\frac{1}{1+M}. \tag{12}$$

Table 1
The fractional contribution that Patch 1 makes to the total substitution rate.

M	S		
	0	1	10
10^{-1}	**0.50**	**0.71**	**0.99**
1	**0.50**	**0.62**	**0.92**
10	**0.50**	**0.52**	**0.68**

Table 2
Large N limit of the scaled mean time, $\tau_{beneficial}$, that a mutant takes to fix when it arises in Patch 1.

M	S		
	0	1	10
10^{-1}	**3.5**	**5.1**	**4313.2**
1	**1.2**	**1.3**	**100.3**
10	**1.0**	**1.0**	**1.9**

Table 3
Large N limit of the scaled mean time, $\tau_{deleterious}$, that a mutant takes to fix when it arises in Patch 2.

M	S		
	0	1	10
10^{-1}	**3.5**	**3.0**	**4299.1**
1	**1.2**	**1.3**	**100.3**
10	**1.0**	**1.0**	**2.0**

Eq. (12) has the same status as Eq. (6): we have strong numerical evidence that it is exact, but no proof.

Note that Eq. (11) relies on the scaled selection strengths in the two patches being S and $-S$, thus Eq. (11) is not just a property of Patch 1. Without additional analysis we cannot say what form $R_\infty^{beneficial}$ takes when the selection strengths in the two patches are S_1 and S_2 and $S_2 \neq -S_1$.

To gain further intuition into the importance of the contribution that Patch 1 makes to the total substitution rate, we shall determine the ratio $R_\infty^{beneficial}(M,S)/R_\infty(M,S)$. While it is possible, as the above, to expand $R_\infty(M,S)$ or $R_\infty^{beneficial}(M,S)$ in S, when S is not large, we shall numerically determine the values of the ratio, so we can then allow S to be appreciable. Using the results and methods of the appendices, we have numerically determined $R_\infty^{beneficial}(M,S)/R_\infty(M,S)$ and in Table 1 we give some representative values. From Table 1, we see that the largest contribution of Patch 1, to the total substitution rate, occurs at the lowest value of the scaled migration rate, M, and the largest value of the scaled selection intensity, S. By contrast, scaled migration rates that are large lead to a much weaker dependence on the scaled selection intensity.

Beyond the substitution rate, the typical time it takes a new mutation to fix is another important quantity. Attention has been drawn to this time in a different context (Frearn et al., 2013; Antal and Scheuring, 2006; Wu et al., 2012). The calculations in the present work assume that mutations segregate separately, and matters become much more complicated if this is not the case. The mean time-interval between mutations is the reciprocal of the mutation rate, u^{-1}, and for validity of the results presented this should be larger than the mean time a new mutation takes to fix. However, the fixation time depends on which patch the mutation arises in – the patch where it is beneficial or the one where it is deleterious. Using the methods of the present work we have determined the large N limit of the quantities:

$$\tau_{beneficial} = \frac{E[T_{\text{fix,beneficial}}]}{8N}, \quad \tau_{deleterious} = \frac{E[T_{\text{fix,deleterious}}]}{8N} \tag{13}$$

where $E[\ldots]$ denotes an expected value, conditional on fixation ultimately occurring; $T_{\text{fix,beneficial}}$ and $T_{\text{fix,deleterious}}$ are the random times a single copy of a mutation takes to fix, when arising in Patch 1 and Patch 2, respectively; the denominator $8N$ is (approximately) the mean time a neutral allele would take to fix in a single population whose size equals the combined population size of both patches (i.e., $4 \times 2N$). In Tables 2 and 3 we give some representative values of $\tau_{beneficial}$ and $\tau_{deleterious}$.

The *relevant* mean scaled fixation time is to further average the results in Tables 2 and 3 over the probabilities that a mutation arises and fixes in Patch 1 or Patch 2, which can be obtained from Table 1, and we write this average as $E[T_{\text{fix}}]$. However, the similarities of the results in Tables 2 and 3 make evident what the results will be. In particular, it is clear that strong scaled selection ($S=10$) combined with weak scaled migration ($M=0.1$) leads to a mean scaled fixation time of $E[T_{\text{fix}}] \sim 4300 \times 8N$. For u an allelic mutation rate, with a value of 10^{-6}, the result for $E[T_{\text{fix}}]$ would only be smaller than u^{-1} in a small population, i.e., one with $N \lesssim 30$. For strong scaled selection ($S=10$) and intermediate scaled migration ($M=1$), we have $E[T_{\text{fix}}] \sim 100 \times 8N$ and this is smaller than u^{-1} when $N \lesssim 1250$. The remaining results are much less constrained and can hold for populations of size $N \sim 10^4$ or 10^5.

More generally, the above estimates of the largest population sizes that we can consider, without the mutations interfering, would need to be modified if the *effective population size*, N_e, is significantly smaller than the census size, N (by using a Wright–Fisher model, we have assumed $N_e = N$). This would require a more complex analysis than the one presented here.

In other work on this topic, Gavrilets and Gibson (2002) (as well as Lanfear et al., 2014) focused on the relationship between the substitution rate and the population size – or, more precisely, in the Lanfear et al. case, on the relationship between the substitution rate and the effective population size, N_e. As we have indicated in the Introduction, interest in this relationship traces back to Ohta's (1972) article that indicated that, in some cases, evolution could be more rapid in small populations than in larger ones. More specifically, Gavrilets and Gibson (2002) discussed the effect of the relationship between population size and migration rate on the rate of evolution. In particular they considered two limiting cases: high population size–high migration and low population size–low migration. They noted that if there is a convex relationship between migration and population size then it would be possible to observe higher substitution rates in small populations than in large ones. Part of the biological rationale put forward by Gavrilets and Gibson (2002) was that "in natural populations, population sizes are usually positively correlated with migration rates". To support this claim they cite work by Gaston (1994, 1996, 1998). However, based on these references and more recent studies (Roland et al., 2000; Altwegg et al., 2003) the relation between population size and migration rate does not appear to be always positive, and may, altogether be weak (Gaston, 1996). Because of this and given our findings that the substitution rate depends principally on N via its appearance in scaled parameters ($M=4Nm$, and $S=4Ns$), we examined that the effects of any relationship between scaled migration (M) and scaled selection intensity (S) have on the substitution rate. The dependence of the substitution rate on N was then viewed as being largely a consequence of the M–S relationship. This approach led to the conclusion that there indeed exist parameter regimes where the substitution rate increases with N (including the simplest case where m and s are constants, so $S \propto M$) and other regimes where the substitution rate decreases with N. Hence, while a small

population size can lead to higher rate of substitution than a large population size, this is far from the rule.

Generally, given that genetic drift acts through multiple agencies, it is reasonable to speculate that the focus on the population size (or effective population size) as a primary determinant of key quantities, such as the rate of evolution, may be unwarranted and it might be more fruitful to simply focus on naturally occurring composite parameters involving the (effective) population size such as $\theta = 4N\mu$, $M = 4Nm$ and $S = 4Ns$. Furthermore, these types of composite parameters are often the only parameters that can be estimated from DNA polymorphism data.

Appendix A. Wright–Fisher model for the Basic Model

In this appendix, we give the description of the Wright–Fisher model for the Basic Model outlined in the main text.

The processes taking place in a generation are as given in the life cycle in the main text.

To begin, we use $X(t)$ to denote the frequency of the A allele in adults in Patch 1, in generation t, and $Y(t)$ the corresponding frequency of the A allele in Patch 2.

A.1. Infinite population dynamics

The dynamics in a very large (effectively infinite) population is as follows:

Each adult produces an effectively infinite number of gametes before dying. Thus $X(t)$ and $Y(t)$ are also the frequencies of the A allele in gametes in each patch. With the probability of an individual gamete migrating from its patch of m, the frequencies of the A allele in gametes in generation t, after migration, are

$$\begin{aligned} X^*(t) &= (1-m)X(t) + mY(t) \\ Y^*(t) &= (1-m)Y(t) + mX(t). \end{aligned} \tag{14}$$

The gametes then undergo random union in the patch in which they reside, hence the frequencies of the AA, Aa and aa genotype zygotes are, respectively, $[X^*(t)]^2$, $2X^*(t)[1-X^*(t)]$ and $[1-X^*(t)]^2$ in Patch 1 and $[Y^*(t)]^2$, $2Y^*(t)[1-Y^*(t)]$ and $[1-Y^*(t)]^2$ in Patch 2.

Selection, of a multiplicative nature, next occurs, and in Patch 1 the relative fitnesses of the AA, Aa and aa genotype juveniles are $(1+s)^2$, $(1+s)$ and 1, respectively, while in Patch 2 they are $(1-s)^2$, $(1-s)$ and 1, respectively. Thus the frequencies of the different genotype juveniles are

$$\left.\begin{array}{ll} AA & \dfrac{(1+s)^2[X^*(t)]^2}{[1+sX^*(t)]^2} \\ Aa & \dfrac{2(1+s)X^*(t)[1-X^*(t)]}{[1+sX^*(t)]^2} \\ aa & \dfrac{[1-X^*(t)]^2}{[1+sX^*(t)]^2} \end{array}\right\} \text{Patch 1}$$

$$\left.\begin{array}{ll} AA & \dfrac{(1-s)^2[Y^*(t)]^2}{[1-sY^*(t)]^2} \\ Aa & \dfrac{2(1-s)Y^*(t)[1-Y^*(t)]}{[1-sY^*(t)]^2} \\ aa & \dfrac{[1-Y^*(t)]^2}{[1-sY^*(t)]^2} \end{array}\right\} \text{Patch 2}$$

In an infinite population, the juveniles simply progress to adults, and the frequencies of the A allele in adults in generation $t+1$ are $X(t+1) = (1+s)X^*(t)/(1+sX^*(t))$ for Patch 1 and $Y(t+1) = (1-s)Y^*(t)/(1-sY^*(t))$ for Patch 2. With the neglect of very small terms of order s^2 or ms we can approximate the equations for $X(t)$ and $Y(t)$ by

$$\begin{aligned} X(t+1) &= D_1(X(t), Y(t)) \\ Y(t+1) &= D_2(X(t), Y(t)) \end{aligned} \tag{15}$$

where $D_1(x,y)$ and $D_2(x,y)$ are the functions:

$$\begin{aligned} D_1(x,y) &= x + sx(1-x) + m(y-x) \\ D_2(x,y) &= y - sy(1-y) + m(x-y). \end{aligned} \tag{16}$$

We note that the life cycle in the main text assumes that first some gametes migrate, that they then undergo random union to form zygotes, and following this the zygotes undergo selection. There are biologically plausible alternatives to this sequence of events. For example, assuming selection occurs in the diploid stage of the life cycle, we have the following two alternatives: (i) Zygotes are formed from the random union of gametes. The zygotes then undergo selection and following this, some of the juveniles (i.e., post-selection individuals) migrate. (ii) Zygotes are formed from the random union of gametes. Some of the zygotes then migrate and following this they undergo selection. These alternatives, and others where selection acts on gametes, lead to results for $D_1(x,y)$ and $D_2(x,y)$ which differ from the forms given in Eq. (16) by very small terms, of the order s^2 or ms, which are terms that have already been assumed to be negligible. The intuition is that as long as selection and migration are weak ($s \ll 1$, $m \ll 1$), the changes they cause in allele frequencies over one generation do not interfere with one another to leading order in s and m. Thus importantly, the infinite population dynamics, when selection and migration are weak, is robust to the precise order and stage of migration and selection in the life cycle.

A.2. Finite population dynamics

In a finite population we assume that population thinning independently occurs, at the juvenile stage, in each patch. The thinning takes the form of randomly sampling the population of a patch, so the numbers of individuals in the patch are reduced to N. With small errors (Nagylaki, 1992, P252; Waxman, 2009), the sampling can be treated as sampling with replacement of the allele frequencies, and this leads to the Wright–Fisher model for the problem. This can be written as

$$\begin{aligned} X(t+1) &= \frac{\mathrm{Bin}(2N, D_1(X(t), Y(t)))}{2N} \\ Y(t+1) &= \frac{\mathrm{Bin}(2N, D_2(X(t), Y(t)))}{2N} \end{aligned} \tag{17}$$

where each instance of a quantity of the form $\mathrm{Bin}(n,p)$ represents a random number that is independently drawn from a binomial distribution with parameters n (representing the number of trials) and p (representing the probability of success on each trial).

Eq. (17) is equivalent to a Markov chain with state labels that have two elements: the first element specifies the state (i.e. frequency of the A allele) of Patch 1, while the second specifies the state of Patch 2. Thus the transition matrix has elements $\mathcal{W}(b,d),(a,c)$, whose value is the probability that the population makes a transition, from one generation to the next, from an A allele frequency that is initially $a/(2N)$ in Patch 1 and $c/(2N)$ in Patch 2, while finally the frequency is $b/(2N)$ in Patch 1, and $d/(2N)$ in Patch 2 (here a, b, c and d can take the values 0, 1,…,$2N$).

We shall use

$$x_n = \frac{n}{2N} \tag{18}$$

which, with $n = 0, 1, \ldots, 2N$, are the allowed frequencies of the A allele in adults in each patch. It then follows from Eq. (17) that the transition matrix elements are given by

$$\mathcal{W}(b,d),(a,c) = \binom{2N}{b} [D_1(x_a, x_c)]^b [1 - D_1(x_a, x_c)]^{2N-b}$$

$$\times \binom{2N}{d} [D_2(x_a, x_c)]^d [1 - D_2(x_a, x_c)]^{2N-d}. \quad (19)$$

Note that if $F_t(a,c)$ represents the probability distribution of the A allele in generation t, then its form in generation $t+1$ is given by

$$F_{t+1}(b,d) = \sum_{a,c=0}^{2N} \mathcal{W}(b,d),(a,c) F_t(a,c).$$

A.3. Alternative labelling of states

There is an alternative way of labelling the states of the two patch Markov chain, where each state of the two patch system is labelled by a single integer, rather than a pair of indices (see, e.g., Waxman, 2009). This allows the problem to be formulated in terms of ordinary matrices and vectors, and hence allows it to be analysed using standard linear algebra software.

Using

$$r = 2N+1 \quad (20)$$

we can convert $\mathcal{W}(b,d),(a,c)$ to an element $W_{f,i}$ of an ordinary matrix $\mathbf{W}$ of size $r^2 \times r^2$ where

$$\begin{aligned} f &= rb+d \\ i &= ra+c \end{aligned} \quad (21)$$

with i and f running from 0 to r^2-1. This transformation can be inverted: $b = \mathrm{int}(f/r)$, $d = f - rb$, $a = \mathrm{int}(i/r)$ and $c = i - ra$ where $\mathrm{int}(k)$ denotes the integer part of k.

Once the problem is reduced to ordinary matrix form, the transformed transition matrix $\mathbf{W}$ has the form

$$\mathbf{W} = \begin{pmatrix} W_{0,0} & W_{0,1} & \cdots \\ W_{1,0} & \cdots & \\ \vdots & & W_{r^2-1,r^2-1} \end{pmatrix} = \begin{pmatrix} 1 & \boldsymbol{\alpha}^T & 0 \\ \mathbf{0} & \mathbf{w} & \mathbf{0} \\ 0 & \boldsymbol{\beta}^T & 1 \end{pmatrix} \quad (22)$$

where a T superscript denotes a matrix transpose, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are column vectors of length r^2-2, $\mathbf{0}$ is a column vector of zeros of length r^2-2 and $\mathbf{w}$ is an $(r^2-2)\times(r^2-2)$ matrix.

Note that a final state of the population, where the A allele is globally fixed (i.e., at a frequency of 1 in each patch), corresponds to a value of the index f of r^2-1 (using Eq. (21) with $b = d = 2N$). Similarly, an initial state of the population where there is a single copy of the A allele in Patch 1 corresponds to a value of the index i of r (using Eq. (21) with $a=1$ and $c=0$), while a single copy of the A allele in Patch 2 corresponds to $i=1$.

Using standard results for a Markov chain (see, e.g., Waxman, 2009), the probability of fixation of the A allele when it starts from a single copy in Patch 1 and no copies in Patch 2 is

$$P_{\mathrm{fix}}\left(\frac{1}{2N}, 0\right) = [\boldsymbol{\beta}^T (\mathbf{I} - \mathbf{w})^{-1}]_r \quad (23)$$

while the probability of fixation when it starts from no copies in Patch 1 and a single copy in Patch 2 is

$$P_{\mathrm{fix}}\left(0, \frac{1}{2N}\right) = [\boldsymbol{\beta}^T (\mathbf{I} - \mathbf{w})^{-1}]_1. \quad (24)$$

The corresponding mean time of fixation of the A allele, given fixation ultimately occurs, when it starts from a single copy in Patch 1 and no copies in Patch 2, is (see, e.g., Waxman, 2009)

$$E[T_{\mathrm{fix,beneficial}}] = \frac{[\boldsymbol{\beta}^T (\mathbf{I} - \mathbf{w})^{-2}]_r}{[\boldsymbol{\beta}^T (\mathbf{I} - \mathbf{w})^{-1}]_r} \quad (25)$$

and when it starts from a single copy in Patch 2 and no copies in Patch 1, it is

$$E[T_{\mathrm{fix,deleterious}}] = \frac{[\boldsymbol{\beta}^T (\mathbf{I} - \mathbf{w})^{-2}]_1}{[\boldsymbol{\beta}^T (\mathbf{I} - \mathbf{w})^{-1}]_1}. \quad (26)$$

Note that with the indexing scheme adopted, the vector $\boldsymbol{\beta}^T (\mathbf{I} - \mathbf{w})^{-1}$ has an index which starts at 1. Because of this, the fixation probabilities can be directly determined from Eqs. (23) and (24) with no offsets required for the matrix indices.

A.4. Rate of substitution

The average number of mutations that fix/generation, when the mutations originally arose in Patch 1, is $2 \times N \times u \times P_{\mathrm{fix}}(1/2N, 0)$, where u is the allelic mutation rate. The corresponding number of mutation for Patch 2 is $2 \times N \times u \times P_{\mathrm{fix}}(0, 1/2N)$. Hence the rate of substitution, in units of the allelic mutation rate, u, is

$$\begin{aligned} R &= 2N \times P_{\mathrm{fix}}\left(\frac{1}{2N}, 0\right) + 2N \times P_{\mathrm{fix}}\left(0, \frac{1}{2N}\right) \\ &= 2N \times [\boldsymbol{\beta}^T (\mathbf{I} - \mathbf{w})^{-1}]_r + 2N \times [\boldsymbol{\beta}^T (\mathbf{I} - \mathbf{w})^{-1}]_1. \end{aligned} \quad (27)$$

Appendix B. Properties of the rate of substitution

In this appendix, we employ ideas from a diffusion approximation of the problem to infer the dependence of the rate of substitution R on parameters in the model. We show that when terms of order N^{-1} can be neglected, the rate of substitution depends only on the scaled strengths of migration and selection, namely M and S.

B.1. Diffusion approximation

Under the diffusion approximation, we shall use $p_{\mathrm{fix}}(x,y)$ to denote the probability of fixation, when the A allele frequency has an initial value of x in Patch 1 and y in Patch 2. For the Basic Model, this probability obeys the backward diffusion equation (cf. Gavrilets and Gibson, 2002):

$$\begin{aligned} &\left(\frac{x(1-x)}{4N_e}\frac{\partial^2}{\partial x^2} + \frac{y(1-y)}{4N_e}\frac{\partial^2}{\partial y^2}\right) p_{\mathrm{fix}}(x,y) \\ &\quad + [sx(1-x) + m(y-x)]\frac{\partial}{\partial x} p_{\mathrm{fix}}(x,y) \\ &\quad + [-sy(1-y) + m(x-y)]\frac{\partial}{\partial y} p_{\mathrm{fix}}(x,y) = 0 \end{aligned} \quad (28)$$

where N_e denotes the effective population size, and $p_{\mathrm{fix}}(x,y)$ is subject to the boundary conditions $p_{\mathrm{fix}}(0,0) = 0$ and $p_{\mathrm{fix}}(1,1) = 1$.

Here we shall not make any distinction between the effective population size and the census population size, and hence take $N_e = N$.

To proceed, we note that on multiplying Eq. (28) by $4N$ and defining

$$M = 4Nm, \quad S = 4Ns \quad (29)$$

that the equation reduces to

$$\begin{aligned} &\left[x(1-x)\frac{\partial^2}{\partial x^2} + y(1-y)\frac{\partial^2}{\partial y^2}\right] p_{\mathrm{fix}}(x,y) \\ &\quad + [Sx(1-x) + M(y-x)]\frac{\partial}{\partial x} p_{\mathrm{fix}}(x,y) \\ &\quad + [-Sy(1-y) + M(x-y)]\frac{\partial}{\partial y} p_{\mathrm{fix}}(x,y) = 0. \end{aligned} \quad (30)$$

Given that the boundary conditions are independent of N, m and s we can infer that in addition to x and y, $p_{\mathrm{fix}}(x,y)$ also depends on the parameters M and S but nothing else. Thus, under a diffusion approximation, we shall sometimes write the fixation probability as $p_{\mathrm{fix}}(x,y;M,S)$ when we wish to be fully explicit.

Consider the rate of substitution relative to the neutral rate, which we denote by R. Under the diffusion approximation, we

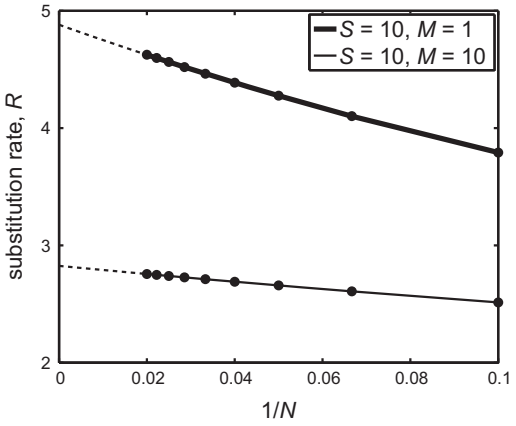

Fig. C1. This figure shows how the substitution rate, R, of the Wright–Fisher model varies with the reciprocal of the population size, N^{-1}. The scaled strengths of migration and selection, $M=4Nm$ and $S=4Ns$, were held fixed for the calculation of both curves of the figure, while N was allowed to range from 10 to 50. The dots represent numerical results from the Wright–Fisher model, while the solid lines are the best curves through these points of the form $R=R_\infty-g/N+h/N^2$. Extrapolating the curves to $1/N=0$, as indicated by the dashed lines in the figure, allows estimation of the substitution rate at $N=\infty$, namely R_∞. For $S=10$ and $M=1$ we obtained $R_\infty\approx 4.88$, $g\approx 13.30$ and $h\approx 24.02$, while for $S=10$ and $M=10$ we obtained $R_\infty\approx 2.82$, $g\approx 3.54$ and $h\approx 4.32$.

have

$$R=2N\times p_{\rm fix}\left(\frac{1}{2N},0;M,S\right)+2N\times p_{\rm fix}\left(0,\frac{1}{2N};M,S\right) \tag{31}$$

where the two terms are the contributions to the substitution rate from mutations that originate in different patches.

On the plausible assumption that $p_{\rm fix}(x,y;M,S)$ is smooth in the vicinity of $(x,y)=(0,0)$ we have

$$p_{\rm fix}\left(\frac{1}{2N},0;M,S\right)=p_{\rm fix}(0,0;M,S)+\frac{1}{2N}\frac{\partial}{\partial x}p_{\rm fix}(x,y;M,S)\Big|_{x=0,y=0}+O(N^{-2})$$
$$=\frac{1}{2N}\frac{\partial}{\partial x}p_{\rm fix}(x,y;M,S)\Big|_{x=0,y=0}+O(N^{-2}). \tag{32}$$

Using this result, and the similar result for $p_{\rm fix}(0,1/2N)$, Eq. (31) becomes

$$R=\frac{\partial}{\partial x}p_{\rm fix}(x,y;M,S)\Big|_{x=0,y=0}+\frac{\partial}{\partial y}p_{\rm fix}(x,y;M,S)\Big|_{x=0,y=0}+O(N^{-1}) \tag{33}$$

and we write this equation as

$$R=R_\infty(M,S)+O(N^{-1}). \tag{34}$$

Hence under the diffusion approximation, the rate of substitution, when terms of order N^{-1} are neglected, depends on the scaled strengths of migration and selection, namely M and S, and no other parameters.

Appendix C. Determination of numerical results

In this appendix, we give details of the way numerical results for this paper were obtained:

(1) *Numerical evidence that $R=R_\infty(M,S)+$ terms which vanish as $N\to\infty$*: We have evaluated the result predicted for the rate of substitution from the Wright–Fisher model, Eq. (27), when the scaled parameters $M=4Nm$ and $S=4Ns$ have *fixed values*, and N is allowed to take progressively larger values from $N=10$ to $N=50$. We have found that for M and S smaller than 10 the resulting value of R is extremely well fitted by a quadratic function of N^{-1} of the form

$$R=R_\infty-\frac{g}{N}+\frac{h}{N^2} \tag{35}$$

where g and h are positive. Such a fit allows us to make an accurate estimate of the rate of substitution when $N\to\infty$, which we write as R_∞. In Fig. C1 we give illustrative plots of how the rate of substitution, R, depends on N^{-1}.

(2) *Dependence of $R_\infty(M,S)$ on M and S*: For a given value of M, and a set of values of S, we determined the values of $R_\infty(M,S)$ via the procedure in Part 1 of this appendix. We fitted a polynomial to $R_\infty(M,S)$ of the form $R_\infty(M,S)=C_0+C_2S^2+C_4S^4+C_6S^6+C_8S^8$. The values of C_0, C_2 and C_4 remained, to good accuracy, stable when the order of the polynomial in S^2 was changed from being quartic to being either cubic or quintic. We recorded the values of C_0, C_2 and C_4 that were obtained, and repeated the above process for different values of M. This led to Fig. 1 of the main text.

Theoretically, the value of C_0 has the value unity, independent of the value of M (see main text), and we found that to high accuracy the numerical results yielded this. Deviations of C_0 from unity were typically smaller than 10^{-4}, sometimes much smaller, depending on the set of S values adopted. For example, for the set of S values $\{0.0.25, 0.5, 0.75, \ldots 2\}$, deviations of C_0 from unity were smaller than 10^{-7}). This level of accuracy, and the close agreement of C_2 with the functional form given in Eq. (6) of the main text, gives us good confidence in the quality of the numerical results presented.

References

Akashi, H., Osada, N., Ohta, T., 2012. Weak selection and protein evolution. Genetics 192, 15–31.

Altwegg, R., Roulin, A., Kestenholz, M., Jenni, L., 2003. Variation and covariation in survival, dispersal, and population size in barn owls Tyto alba. J. Anim. Ecol. 72, 391–399.

Antal, T., Scheuring, I., 2006. Bull. Math. Biol. 68, 1923–1944.

Balloux, F., Lehmann, L., 2012. Substitution rates at neutral genes depend on population size under fluctuating demography and overlapping generations. Evolution 66, 605–611.

Bulmer, M.G., 1972. Multiple niche polymorphism. Am. Nat. 106, 254–257.

Clobert, J., Danchin, E., Dhondt, A.A., Nichols, J.D., 2001. Dispersal. Oxford University Press, Oxford, UK.

Felsenstein, J., 2013. Theoretical Evolutionary Genetics. Freely available at ⟨http://evolution.gs.washington.edu/pgbook/pgbook.html⟩.

Fisher, R.A., 1930. The Genetical Theory of Natural Selection. Dover Publication Inc., New York.

Frearn, M., Rainey, P.B., Traulsen, A., 2013. The effect of population structure on the rate of evolution. Proc. R. Soc. B 280, 1471–2954.

Gaston, K.J., 1994. Rarity. Chapman and Hall, London.

Gaston, K.J., 1996. Species-range size distributions: patterns, mechanisms and implications. Trends Ecol. Evol. 11, 197–201.

Gaston, K.J., 1998. Species-range size distributions: products of speciation, extinction and transformation. Philos. Trans. R. Soc. Lond. B 353, 219–230.

Gavrilets, S., Gibson, N., 2002. Fixation probabilities in a spatially heterogeneous environment. Popul. Ecol. 44, 51–58.

Gillespie, J., 2001. Is the population size of a species relevant to its evolution? Evolution 55, 2161–2169.

Haldane, J.B.S., 1932. The Causes of Evolution. Longmans, London.

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

Kimura, M., 1968. Evolutionary rate at the molecular level. Nature 217, 624–626.

Kimura, M., 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, UK.

Lanfear, R., Kokko, H., Eyre-Walker, A., 2014. Population size and the rate of evolution. Trends Ecol. Evol. 29, 33–41.

Levene, H., 1953. Genetic equilibrium when more than one ecological niche is available. Am. Nat. 87, 311–313.

Maruyama, T., 1970. On the fixation probability of mutant genes in a subdivided population. Genet. Res. 15, 221–225.

Nagylaki, T., 1992. Introduction to Theoretical Population Genetics. Springer, Berlin.

Ohta, T., 1972. Population size and the rate of evolution. J. Mol. Evol. 1, 305–314.

Roland, J., Keyghobadi, N., Fownes, S., 2000. Alpine Parnassius butterfly dispersal: effects of landscape and population size. Ecology 81, 1642–1653.

Savolainen, O., Lascoux, M., Merilä, J., 2013. Ecological genomics of local adaptation. Nat. Rev. Genet. 14, 807–820.

Tachida, H., Iizuka, M., 1991. Fixation probability in spatially changing environments. Genet. Res. Camb. 58, 243–251.

Waxman, D., 2009. Fixation at a locus with multiple alleles: structure and solution of the Wright Fisher model. J. Theor. Biol. 257, 245–251.
Waxman, D., 2011. A unified treatment of the probability of fixation when population size and the strength of selection change over time. Genetics 188, 907–913.
Whitlock, M.C., Gomulkiewicz, R., 2005. Probability of fixation in a heterogeneous environment. Genetics 171, 1407–1417.
Wright, S., 1931. Evolution in Mendelian populations. Genetics 16, 97–159.
Wright, S., 1968. Evolution and the Genetics of Populations. A Treatise in Four Volumes. The University of Chicago Press, Chicago.
Wu, B., Gokhale, C.S., Wang, L., Traulsen, A., 2012. J. Math. Biol. 64, 803–827.
Yeaman, S., Otto, S.P., 2011. Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. Evolution 65, 2123–2129.