Theoretical Population Biology **61**, 31–48 (2002)
doi:10.1006/tpbi.2001.1557, available online at http://www.idealibrary.com on IDEAL®

Neutral Evolution in Spatially Continuous Populations

Nick H. Barton and Frantz Depaulis

Institute of Cell, Animal and Population Biology, University of Edinburgh, King's Building, West Mains Road, EH9 3JT Edinburgh, United Kingdom

and

Alison M. Etheridge

Department of Statistics, University of Oxford

Received September 15, 2000

We introduce a general recursion for the probability of identity in state of two individuals sampled from a population subject to mutation, migration, and random drift in a two-dimensional continuum. The recursion allows for the interactions induced by density-dependent regulation of the population, which are inevitable in a continuous population. We give explicit series expansions for large neighbourhood size and for low mutation rates respectively and investigate the accuracy of the classical Malécot formula for these general models. When neighbourhood size is small, this formula does not give the identity even over large scales. However, for large neighbourhood size, it is an accurate approximation which summarises the local population structure in terms of three quantities: the effective dispersal rate, σ_e; the effective population density, ρ_e; and a local scale, κ, at which local interactions become significant. The results are illustrated by simulations.

1. INTRODUCTION

Many organisms live in two dimensions. Over local scales, they may be restricted to discrete demes, scattered over a continuous habitat, or distributed in some more complex pattern. Regardless of these local details, the population may appear more or less continuous when viewed over a larger scale. To describe the evolution of such populations, we require a simple model which can approximate a wide variety of local structures in terms of a few parameters.

For a population that is large enough that density and allele frequencies evolve deterministically, allele frequencies can be approximated over large spatial and temporal scales by a diffusion equation whose coefficients depend only on the mean and variance of the distance between parent and offspring, and not on finer details of the dispersal distributions (Nagylaki, 1975). In one dimension, this approximation extends to include random genetic drift. One can obtain a diffusion approximation for allele frequency via a simple rescaling and, thus, simultaneously, derive a second diffusion equation for the correlations between allele frequencies at different locations (Nagylaki, 1978a,b). More generally, one obtains a stochastic partial differential equation for the allele frequencies (Shiga, 1988) for which one can calculate correlations of all orders.

In two dimensions, as we shall see in a little more detail in Section 2, this rescaling fails (Nagylaki, 1978a). The equations for the correlations "blow up" on local scales (that is, on scales comparable with the distance moved by a single gene over one generation) and the natural

0040-5809/02 $35.00

analogue of the one-dimensional stochastic partial differential equation for allele frequencies has no solution.

The difficulty arises from the assumption that there is a local population density which is constant. A consequence of this is that if we trace the lineages of a sample from the population *backward* in time, then they evolve as independent random walks (or Brownian motions in the diffusion limit). However, in two dimensions, two independent Brownian motions will, with probability one, never be in the same place at the same time. In other words, our lineages will never intersect. Because random fluctuations arise from the coalescence of ancestral lineages, this implies that the allele frequencies solve a *deterministic* partial differential equation in the diffusion limit.

We should like then a model that incorporates the stochastic effects arising from finite (and fluctuating) local population. Over small scales, the distribution of allele frequencies depends on detailed local structure of the population and these local details are propagated by the spatial motion of lineages to influence patterns over larger scales. However, this leads to a second difficulty. In one or two dimensions, there is no satisfactory model for the population dynamics in a truly continuous habitat. Wright (1943) and Malécot (1948) derived results for a continuous habitat by assuming that genes reproduce and disperse independently of each other, and that they are scattered in a stationary Poisson distribution. However, these assumptions are incompatible: independent reproduction leads to ever increasing clumping (Felsenstein, 1975; Sawyer and Fleischmann, 1978). Some form of "density-dependent regulation" is required to stabilise the population. A drastic solution is to use a stepping stone model, in which the population is forced to occupy discrete demes, each with a fixed population size (Kimura and Weiss, 1964; Malécot, 1948; Maruyama, 1971). However, population genetics and population dynamics cannot be separated if the population is truly continuous: genes that occupy crowded regions must leave fewer descendants, and so there must be local interactions between nearby lineages.

In this paper, we set out an approximation to a population with an arbitrary local structure. This is almost identical to that derived by Wright (1943) and Malécot (1948), but allows for the interactions induced by local density-dependent regulation. Our ultimate aim is to include nonlinear forces such as selection, and to describe the genealogical relation between samples of many genes. However, we deal here only with the very simplest case of the relationship between two neutral genes sampled from a population that has reached a stationary state.

We introduce the problem with a reasonably detailed discussion of the continuous-time stepping stone model in Section 2. We then discuss diffusion approximations to the stepping stone model, including, in Section 3, a stochastic partial differential equation for allele frequencies in populations with large "neighbourhood size" that remains valid in two dimensions. Inter alia we provide an elementary derivation of Malécot's (1948) formula for the probability of identity of two individuals sampled from a population in which local population density is constant. In Section 4 we then set out a general recursion for the identity in state between two genes, subject to mutation, migration, and random drift, and which allows for the local interactions that are inevitable in a continuous population. Following Malécot, in Section 5, we take the Fourier transform of the recursion to reduce the iteration to algebra, while in Section 6, we prepare the ground for using the resulting expression to write down an exact solution to the recursion as a series in inverse powers of the neighbourhood size in Section 7. Finally, in Section 8, we give another explicit series expansion which gives accurate approximations even for small neighbourhood size, provided that the mutation rate is low. These results are illustrated by simulations of a continuous population, and by numerical solutions for exact population models.

2. DISCRETE DEMES

In this section we use the classical stepping stone model as a starting point from which to explore some of the difficulties that the recursion in Section 4 below seeks to address. For simplicity, we restrict our attention to the case in which there is no selection and we track just two alleles, a and A say, at each site. The stepping stone model assumes that the population is divided into a finite or countable collection of discrete demes, indexed by the elements of some set I. The size of the population in each deme is fixed and assumed large enough that we can take a diffusion approximation. Typically I will be $\mathbb{Z}^d$, d being the spatial dimension, but the approximation that we derive for correlations between alleles will not be sensitive to the choice of lattice provided that (when suitably rescaled) the migration of individuals can be approximated by a Brownian motion. We use $p_i(t)$ to denote the proportion of one of the alleles, a say, in the ith deme. For simplicity, we use a continuous-time model; this corresponds to the discrete-time models of Kimura and Weiss (1964) and Malécot (1969) in the limit of large deme size and small migration rates. Note, however, that, by convention, overlapping generation models lead to a

reduction of the effective population size by a factor 2 compared to discrete generation models.

DEFINITION 2.1 (Continuous-Time Stepping Stone Model). The *continuous-time stepping stone model* for the evolution of $\{p_i(t)\}_{i\in I}$ is the following system of stochastic differential equations:

$$\begin{aligned} dp_i(t) = D\sum_{j\in I} m_{ij}(p_j(t)-p_i(t))\,dt - \mu_1 p_i(t)\,dt \\ +\mu_2(1-p_i(t))\,dt \\ +\sqrt{\gamma p_i(t)(1-p_i(t))}\,dB_i(t). \end{aligned} \tag{1}$$

Here, μ_1 and μ_2 are the mutation rates from type a to A and from A to a respectively, $\{B_i\}_{i\in I}$ are independent Brownian motions and the constants $\{m_{ij}\}_{j\in I}$ are the migration rates from deme j to deme i. D scales the rate of migration. We assume that $m_{ij}=m_{ji}$ and that $\sum_j m_{ij}=1$. If time is measured in units of generation time, then γ is the variance of the number of offspring of an individual divided by the population size of a deme. Of course, in general, one can take these parameters to vary from deme to deme, but here we shall assume that they are constant.

Calculations for the stepping stone model are most easily performed using the *duality* between the stepping stone model and the process known as the structured coalescent.

DEFINITION 2.2 (Structured Coalescent). The *structured coalescent* takes its values in $\{(n_i)_{i\in I}: n_i\in\mathbb{N}\}$. The dynamics of the process have three ingredients:

1. *Migration:* The transition

$$\begin{cases} n_i \mapsto n_i - 1 \\ n_j \mapsto n_j + 1 \end{cases}$$

takes place at rate $Dn_i m_{ij}$.

2. *Death at site i:* The transition

$$n_i \mapsto n_i - 1$$

takes place at rate $\mu_2 n_i$.

3. *Coalescence at site i:* The transition

$$n_i \mapsto n_i - 1$$

takes place at rate $\frac{1}{2}\gamma n_i(n_i-1)$.

The duality between the stepping stone model and the structured coalescent is described in the following lemma, which can be found in Shiga (1982). For vectors $\underline{p}=(p_i)_{i\in I}$ and $\underline{n}=(n_i)_{i\in I}$, with $0\leqslant p_i\leqslant 1$ and $n_i\in\mathbb{N}$, we write

$$\underline{p}^{\underline{n}} = \prod_{i\in I} p_i^{n_i} \qquad \text{assuming} \quad \sum_{i\in I} n_i < \infty$$

LEMMA 2.3. *Let $\underline{p}(t)$ denote the solution to the system* (1) *and let $\underline{n}(t)$ denote the structured coalescent. Then for each fixed $t\geqslant 0$,*

$$\mathbb{E}[\underline{p}(t)^{\underline{n}(0)}] = \mathbb{E}\left[\underline{p}(0)^{\underline{n}(t)} \exp\left(-\int_0^t \mu_1 |\underline{n}(s)|\,ds\right)\right], \tag{2}$$

where we have written $|\underline{n}(s)|$ for $\sum_{i\in I} n_i(s)$.

Proof. An application of Itô's Lemma shows that $\underline{p}(s)^{\underline{n}(t-s)}\exp(-\mu_1\int_0^{t-s}|\underline{n}(s)|\,ds)$ is a martingale for $0\leqslant s\leqslant t$. Integrating the expectation over $[0,t]$ yields (2). ∎

Equation (2) gives a relation between the moments of allele frequency at time t, and the coalescent process at time t. We shall exploit this below to find an expression for the correlations between allele frequencies in different demes.

Remark 2.4. Although, as we have stated it, the duality is simply a formal relationship between the distributions of two stochastic processes, the structured coalescent also has a direct interpretation in our stepping stone model. Suppose that we take a sample from the population at time t that consists, for each i, of n_i individuals of type a from the ith deme. The distribution of the genealogy of the sample, as we trace back in time, can be obtained by setting $\mu_2=0$ in our description of the structured coalescent. The "deaths" in the structured coalescent then correspond to mutation events as we trace backwards in time. That is, a "death" of a lineage occurs at the time of its most recent mutation from type A to type a.

It is now a straightforward matter to find an expression for the probability of identity of two individuals sampled from demes i and j, say. The first step is to weed out the dependence on the parameters μ_1 and μ_2. To this end we introduce the following notation.

Notation 2.5. Suppose that two (continuous-time) random walks on I, started from i and j at time zero and with transition rates m_{ij}, *coalesce* at rate γ when they are in the same state. We denote by $\tau=\tau(i,j)$ the random time at which they coalesce.

LEMMA 2.6. *Suppose that a population evolving according to the system of Eq.* (1) *has reached stationarity. The probability that two individuals, sampled from demes i and j, respectively, are identical in state, denoted* $\phi(i,j)$*, is given by*

$$\begin{aligned}\phi(i,j) &= \mathbb{E}[p_i(t)\,p_j(t)+(1-p_i(t))(1-p_j(t))]\\ &= \bar{p}^2+(1-\bar{p})^2+2\bar{p}(1-\bar{p})\,\mathbb{E}[e^{-2(\mu_1+\mu_2)\,\tau(i,j)}], \quad (3)\end{aligned}$$

where $\bar{p}$ *is the expected allele frequency when the population is in equilibrium and is given by*

$$\bar{p}=\frac{\mu_2}{\mu_1+\mu_2}.$$

Proof. In order to calculate $\mathbb{E}[p_i(t)\,p_j(t)]$, we apply Eq. (2) with

$$n_k(0)=\begin{cases}1 & k=i,\,j,\,i\neq j\\ 0 & \text{otherwise.}\end{cases}$$

To ensure that the population has reached stationarity, we let $t\to\infty$. Evidently $|\underline{n}(t)|\to 0$ as $t\to\infty$, since $|\underline{n}(0)|=2$ initially. The expectation on the right-hand side of Eq. (2) is now most easily calculated by splitting into two parts, according to whether the first death in $\underline{n}(s)$ takes place before coalescence of the two individuals or not. Notice that, conditional on the coalescence time $\tau(i,j)$, the probability that coalescence happens before the first death is $\exp(-2\mu_2\tau(i,j))$. Also,

$$\mathbb{E}\left[\exp\left(-\int_u^\infty \mu_1\,|\underline{n}(s)|\,ds\right)\middle|\,|\underline{n}(u)|=1\right]=\frac{\mu_2}{\mu_1+\mu_2},$$

since, once $|\underline{n}(u)|=1$, the only possible transition of $\underline{n}$ is a death, which happens at rate μ_2. Combining these observations (and suppressing the dependence of τ on i,j in our notation) gives

$$\begin{aligned}&\mathbb{E}\left[\exp\left(-\int_0^\infty \mu_1\,|\underline{n}(s)|\,ds\right)\right]\\ &\quad=\mathbb{E}\left[e^{-2\mu_1\tau}e^{-2\mu_2\tau}\frac{\mu_2}{\mu_1+\mu_2}\right.\\ &\qquad\left.+(1-e^{-2\mu_2\tau})\int_0^\tau\frac{2\mu_2e^{-2\mu_2 s}}{1-e^{-2\mu_2\tau}}e^{-2\mu_1 s}\,ds\,\frac{\mu_2}{\mu_1+\mu_2}\right]\\ &\quad=\bar{p}^2+\bar{p}(1-\bar{p})\,\mathbb{E}[e^{-2(\mu_1+\mu_2)\,\tau}]. \quad (4)\end{aligned}$$

Performing the same calculation with p replaced by $1-p$ (which amounts to changing the roles of μ_1 and μ_2) and adding the result to (4) completes the proof. ∎

In writing down our stepping stone model, we have made the assumption that when a mutation occurs, it is from one of our two types to the other. We could also have assumed that mutations invariably lead to new types, the so-called *infinitely many alleles* model. This is the situation considered by Malécot (1969). The probability of identity of individuals sampled from i, j for the corresponding model is

$$\mathbb{E}[e^{-2\mu\tau(i,j)}], \quad (5)$$

where μ is now the rate at which new mutations arise. Evidently, for either model, it is key to be able to estimate the quantity (5), that is, the Laplace transform of the coalescence time. Notice that in both the two allele and the infinitely many allele setting, we can now use the probability of identity in state, ϕ, to infer the distribution of the coalescence times, τ (cf. Slatkin, 1991). A natural approach is to try to approximate the random walk followed by the "particles" of the $\underline{n}$-process by a diffusion. In one dimension ($I=\mathbb{Z}$), this can be achieved by a simple rescaling (Shiga, 1988).

LEMMA 2.7 (Rescaling in One Dimension). *Let* $I=\mathbb{Z}$ *and suppose that the coefficients* m_{ij} *in Eq.* (1) *satisfy*

$$D\sum_{j\in\mathbb{Z}} m_{ji}\,|j-i|^2=\sigma^2,$$

for all $i\in\mathbb{Z}$ *and some constant* σ*.* (σ^2 *is the variance of the distance moved in one unit of time.) We rescale the stepping stone model as follows:*

$$\begin{cases} i\mapsto i/\sqrt{N},\\ t\mapsto Nt,\\ \mu_1\mapsto\mu_1/N,\\ \mu_2\mapsto\mu_2/N,\\ \gamma\mapsto\gamma/\sqrt{N}.\end{cases} \quad (6)$$

Then, extending the corresponding stepping stone processes on $\mathbb{Z}/\sqrt{N}$ *to* $\mathbb{R}$*-valued processes by linear interpolation, as* $N\to\infty$*, the processes converge to the solution to the following stochastic partial differential equation (Walsh, 1986),*

$$\begin{aligned}dp=&\left(\frac{\sigma^2}{2}\Delta p-\mu_1 p+\mu_2(1-p)\right)dt\\ &+\sqrt{\gamma p(1-p)}\,W(dx,dt), \quad (7)\end{aligned}$$

where $W(dx,dt)$ *is a Gaussian white noise on* $\mathbb{R}^d\times\mathbb{R}_+$.

The rescaling in (6) is the diffusive scaling on the lattice. The rescaling of the mutation rates counteracts the speeding up of time. The rescaling of γ, which corresponds to increasing the population in each deme by a factor $\sqrt{N}$, is chosen precisely so that the correlations $\mathbb{E}[p(x)\, p(y)]$ converge. The effect on the dual process is that the probability of coalescence by time t is determined by the *local time at the origin* (McKean, 1969) of the difference of two Brownian motions started from x and y, respectively. Because two independent Brownian motions do not collide in dimensions greater than one (with probability one), there is no choice of rescaling for γ that will lead to a nontrivial stochastic limiting equation for the allele frequencies. That is, *simple rescaling fails in dimensions larger than one* (cf. Nagylaki, 1978a).

Even though the diffusive rescaling does not lead to a sensible limiting model, over sufficiently large spatial scales, we can still use Brownian motion as an approximation to a random walk to simplify the calculation of the quantity (5). We suppose that we are working with $I = \mathbb{Z}^d$, so that the separation of the two lineages is itself a random walk. This is not essential to the result, but simplifies the calculation (which is not the central purpose of this paper). Let the initial separation of the lineages be $x \in \mathbb{Z}^d$; then, writing

$$\phi(x) = \mathbb{E}[e^{-2\mu\tau}],$$

we have

$$2\mu\phi(x) - \Delta^{(1)}\phi(x) + \gamma\chi_{\{0\}}(x)\,\phi(x) = \gamma\chi_{\{0\}}(x), \qquad (8)$$

where $\Delta^{(1)}$ denotes the generator of the continuous-time random walk corresponding to the difference of the lineages and $\chi_{\{0\}}(x)$ is the indicator function of $\{0\}$. The continuous approximation is obtained by applying the diffusive rescaling. Since, in the limit, the separation of the two lineages will be the difference between two Brownian motions (which is itself a Brownian motion, but run at twice the speed), the probability of identity is a function of the distance between the two sites, reducing the equation to one variable. Abusing notation by using x also to denote the distance between the points at which the sample is taken, this yields

$$\mu\phi - \sigma^2\left(\phi_{xx} + \frac{d-1}{x}\,\phi_x\right) + \gamma\delta_0\phi = \gamma\delta_0, \qquad (9)$$

where d is the dimension of the space and δ_0 is the Dirac delta function concentrated on the origin. Note that the coefficient in front of (the radial form of) the Laplacian is σ^2 not $\sigma^2/2$, because now we are considering the distance between the two lineages. In one spatial dimension, this is easily solved to yield

$$\phi(x) = \frac{1}{1 + \dfrac{2\sigma\sqrt{2\mu}}{\gamma}}\, e^{-\frac{\sqrt{2\mu}}{\sigma}|x|}. \qquad (10)$$

In two dimensions, we have to get around the now familiar problem that two independent Brownian motions do not meet. As a result of this, the formula we obtain breaks down for separations less than or equal to order $O(\sigma)$.

Consider first the modified equation:

$$2\mu\psi - \sigma^2\left(\psi_{xx} + \frac{1}{x}\,\psi_x\right) = \delta_0. \qquad (11)$$

Although the integral makes no sense when $x = 0$, formally, the solution to the two-dimensional diffusion Eq. (11) can be written as

$$\int_0^\infty \frac{1}{4\pi\sigma^2 t}\, e^{-2\mu t} e^{-x^2/(4\sigma^2 t)}\, dt.$$

Now, using Erdelyi (1954, p. 146, Eq. 29), we have

$$\int_0^\infty e^{-pt} t^{\nu-1} e^{-\alpha/(4t)}\, dt = 2\left(\frac{1}{4}\frac{\alpha}{p}\right)^{\nu/2} K_\nu(\sqrt{\alpha p}),$$

$$\mathrm{Re}\,\alpha > 0, \quad \mathrm{Re}\,p > 0.$$

Combining this with linearity, we see that the solution to Eq. (9) in two dimensions is

$$\phi(x) = \frac{\gamma(1-\phi(0))}{2\pi\sigma^2}\, K_0\left(\frac{|x|}{\sigma}\sqrt{2\mu}\right). \qquad (12)$$

We now follow Malécot in our circumvention of the difficulty of two Brownian motions never colliding in two dimensions, and thus of the coalescence time of two lineages being infinite, by declaring that the probability that two lineages currently at separation y coalesce in the previous generation is $\gamma p_{2T}(y, 0)$, where T is the generation time and $p_t(x, y)$ is the *two-dimensional* Brownian transition density (normalised so that the process has variance $\sigma^2 t/2$ at time t). At least for $\mu \ll 1$, if the initial separation of lineages is $x \gg \sigma$, then Eq. (12) provides a good approximation. Now we cannot substitute $x = 0$

and thus solve for $\phi(0)$ directly, but if we assume that ϕ is continuous at the origin, since

$$K_0(z) \sim \log(1/z) \qquad \text{as} \quad z \to 0,$$

substituting $x = \kappa$ and then assuming $\phi(0) \sim \phi(\kappa)$, this leads to

$$1 - \phi(0) = \frac{1}{1 + \frac{\gamma}{2\pi\sigma^2} \log(\sigma/(\kappa\sqrt{2\mu}))}, \tag{13}$$

and

$$\phi(x) = \frac{1}{1 + \frac{\gamma}{2\pi\sigma^2} \log(\sigma/(\kappa\sqrt{2\mu}))} \frac{\gamma}{2\pi\sigma^2} K_0\left(\frac{\sqrt{2\mu}\,|x|}{\sigma}\right), \quad x > \kappa, \tag{14}$$

where κ is a local scale, of order σ, at which the Bessel approximation breaks down. Indeed, Eqs. (12)–(14) correspond to classical asymptotic results of two-dimensional stepping stone theory (Kimura and Weiss, 1964; Malécot, 1969). In general, we shall use $2\rho_e$ to denote the effective population density (in this setting just the population size of a deme). Similarly, σ_e denotes the "effective dispersal rate" (or backward dispersal). In this notation $\gamma = 1/(2\rho_e)$. Defining *Wright's neighbourhood size* by

$$\mathcal{N} \triangleq 4\pi\rho_e\sigma_e^2,$$

this becomes

$$\phi(x) = \frac{1}{\mathcal{N} + \log(\sigma_e/(\kappa\sqrt{2\mu}))} K_0\left(\frac{\sqrt{2\mu}\,|x|}{\sigma_e}\right), \quad x > \kappa. \tag{15}$$

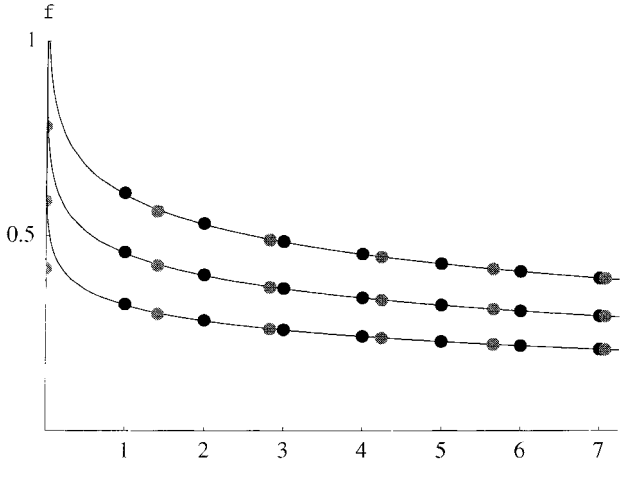

FIG. 1. Accuracy of the Malécot approximation to a stepping stone model in two dimensions. The continuous curves show the approximation to ϕ of Eq. (14), plotted against distance (measured in deme spacings). The mutation rate is $\mu = 10^{-6}$; migration is a symmetric, nearest neighbour random walk with a proportion $m = 0.05$ of individuals being exchanged with each of the four neighbours in each generation. The local scale appropriate for nearest neighbour migration is $\kappa = 1/\sqrt{32}$ deme spacings. Neighbourhood sizes are 2, 5, and 10 (top to bottom). The dark circles show exact results for separations (0, 0), (1, 0), (1, 1), (2, 0), (2, 2), (3, 0), ...

DEFINITION 2.8. We shall refer to Eqs. (10) and (15) as *the Malécot approximation.*

Figure 1 shows that the Malécot approximation of (14) is an extremely close approximation to the two-dimensional stepping stone model, even when it is applied to adjacent demes.

3. A DIFFUSION APPROXIMATION FOR LARGE NEIGHBOURHOOD SIZE

For completeness, in this section, we write down a continuum approximation to the stepping stone model that remains valid in two spatial dimensions. We also check that it provides a good estimate for probability of identity, ϕ, provided that the neighbourhood size is large.

Notice first that the solution to the system of Eqs. (1) can be obtained as the restriction to the lattice points of the solution to

$$dp = (D\Delta^{(1)}p - \mu_1 p + \mu_2(1-p))\,dt + \sqrt{\gamma p(1-p)}\,\hat{W}(dx, dt), \tag{16}$$

where we have used $\Delta^{(1)}$ to denote the infinitesimal generator of the random walk generated by the m_{ij}'s, and $\hat{W}(dx, dt)$ for the convolution of white noise with a constant multiple of an indicator function. The indicator function is chosen to be that of a "cell" about the origin in our lattice (so that the noises experienced at different lattice points are uncorrelated) and the constant is chosen so that the variance of the noise at a lattice point is one.

To approximate Eq. (16), we should like to replace the noise, $\hat{W}(dx, dt)$, by white noise in continuous space, and the discrete Laplacian by the (continuous) Laplacian. However, this leads to the stochastic partial differential equation (7) and, as we remarked in Section 1, this equation has no solution in dimensions larger than one. The

difficulty is the form of the noise term. If instead of a function of p it were a *constant* multiple of white noise, then the equation *would* have a solution, even in two dimensions. Now if the fluctuations in the allele frequencies are small, then we can approximate Eq. (16) by

$$dp = (D\Delta^{(1)}p - \mu_1 p + \mu_2(1-p))\,dt + \sqrt{\gamma\bar{p}(1-\bar{p})}\,\hat{W}(dx, dt), \tag{17}$$

where $\bar{p} = \mu_2/(\mu_1+\mu_2)$ is treated as a constant. Equation (17) can in turn be approximated by the *linear* stochastic partial differential equation

$$dp = \left(\frac{\sigma^2}{2}\Delta p - \mu_1 p + \mu_2(1-p)\right)dt + \sqrt{\gamma\bar{p}(1-\bar{p})}\,W(dx, dt). \tag{18}$$

The remainder of this section is devoted to showing that the probability of identity predicted by Eq. (18) is the first term in the series expansion in inverse powers of neighbourhood size of the true probability of identity.

It is convenient to write

$$\beta^2 \triangleq \gamma\bar{p}(1-\bar{p}),$$

and to consider the corresponding equation for $z \triangleq p - \bar{p}$,

$$dz = \left(\frac{\sigma^2}{2}\Delta z - (\mu_1+\mu_2)\,z\right)dt + \beta\hat{W}(dx, dt). \tag{19}$$

Equation (19) has the explicit solution

$$z(t, x) = e^{-(\mu_1+\mu_2)t}T_t z(0,\cdot)(x) + \int_0^t \beta e^{-(\mu_1+\mu_2)(t-s)}T_{t-s}W(ds,\cdot)(x),$$

where T_t denotes the semigroup with generator $\sigma^2\Delta/2$. This gives

$$\mathbb{E}[z(t, x)\,z(t, y)] = e^{-2(\mu_1+\mu_2)t}T_t z(0,\cdot)(x)\,T_t z(0,\cdot)(y) + \int_0^t \beta^2 e^{-2(\mu_1+\mu_2)s}p_{2s}(x, y)\,ds, \tag{20}$$

where $p_t(x, y)$ is the Brownian transition density (normalised so that the process has variance $\sigma^2 t/2$ at time t), from which it follows that when the population is in equilibrium, the probability of identity in state of two individuals sampled from x and y is

$$\int_0^\infty \left(\beta^2 2\,(\mu_1+\mu_2)\,e^{-2(\mu_1+\mu_2)t}\int_0^t p_{2s}(x, y)\,ds\right)dt, \tag{21}$$

or in the infinitely many alleles setting

$$\int_0^\infty \left(\gamma 2\mu e^{-2\mu t}\int_0^t p_{2s}(x, y)\,ds\right)dt. \tag{22}$$

In other words, we are approximating the quantity (5) by the expression (22). To see how good an approximation this is, first let L_t denote the total amount of time before t that two independent random walks (governed by the m_{ij}'s), started from x and y, respectively, spend in the same deme. Then (provided that x and y are sufficiently separated that Brownian motion is a good approximation to random walk) the expression (22) is approximately

$$\int_0^\infty \gamma 2\mu e^{-2\mu t}\mathbb{E}[L_t]\,dt.$$

We wish to compare this to (5).

Now, the random time of pairwise coalescence τ satisfies

$$\mathbb{P}[\tau > t] = \mathbb{E}[e^{-\gamma L_t}].$$

An integration by parts yields

$$\begin{aligned}\mathbb{E}[e^{-2\mu\tau}] &= \int_0^\infty 2\mu e^{-2\mu t}(1-\mathbb{P}[\tau > t])\,dt \\ &= 1 - \int_0^\infty 2\mu e^{-2\mu t}\mathbb{E}[e^{-\gamma L_t}]\,dt \\ &= \int_0^\infty 2\mu e^{-2\mu t}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n!}\gamma^n\mathbb{E}[L_t^n]\,dt.\end{aligned}$$

Comparing with Eqs. (14) and (15), we see that this power series in γ can be thought of as a power series in inverse powers of the neighbourhood size and thus that, provided neighbourhood size is large, the linear stochastic partial differential equation (18) yields a good approximation to allele frequencies in the stepping stone model. This approximation depends on interaction between individuals within demes becoming negligible as $\gamma \to 0$.

In fact, by replacing white noise by coloured noise (that is the convolution of the white noise with a suitable

L^2 function) in any of our stochastic partial differential equations, we obtain equations that have function-valued solutions in two dimensions and so we can even extend them to incorporate, for example, a selection term. This then suggests a class of continuum models in two dimensions. However, the artificial nature of the stepping stone model (used in the derivation) and the associated assumption of large and known local density represents a major deficiency.

4. THE GENERAL CASE: LOCAL INTERACTIONS

In this section we write down a general model for the probability of identity in state of two individuals sampled from a population evolving in $\mathbb{R}^2$. Unlike the stepping stone models of Section 2, this model incorporates local interactions. In particular, the population density is no longer taken as given, and indeed, is not well defined (see Section 10), so that as we evolve backward in time, the lineages of the two individuals will not follow independent spatial motions. Our model will take the form of a recursion over an arbitrarily chosen time interval Δt. We then state the assumptions that we require in order to iterate the recursion and thus, in later sections, perform the iteration to uncover possible sources of inaccuracy in the classical Malécot formula.

We denote by $\phi(y)$ the probability that two individuals, separated by the vector $y \in \mathbb{R}^2$, are identical in state, under the infinitely many alleles model. The basic recursion traces the origin of these individuals at time t, the current time being $t+\Delta t$:

$$\phi_{t+\Delta t} = e^{-2\mu\,\Delta t}\left\{\int \phi_t\Big(y+x-x'\Big)\, g(x, x', y)\, dx\, dx' + \frac{\Delta t}{2\rho_e} h(y)\right\}. \tag{23}$$

Here $g(x, x', y)$ is the probability that individuals currently separated by y are descendants of two distinct individuals at time t whose displacements from the current individuals are x and x', respectively (Fig. 2). A consequence of this definition is of course that the integral of g over all possible displacements is not 1, but $1-h(y)\,\Delta t/(2\rho_e)$.

The quantity $h(y)\,\Delta t/(2\rho_e)$ is the probability that two individuals at separation y are identical by descent in the previous Δt of time. The function h is normalised so that $\int h(y)\,dy = 1$. Roughly speaking, h is the distribution of distances between pairs that coalesce during a time interval. The normalisation provides a definition of the inbreeding effective density, ρ_e. Note that this effective density depends on the arbitrary time slice Δt (Fig. 3). As usual, μ is the mutation rate, assumed very small.

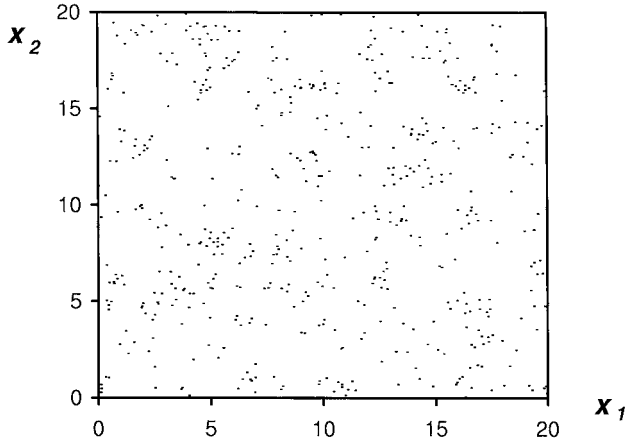

FIG. 2. Zoom of a spatial distribution of simulated individuals at the end of a simulation under the Bolker–Pacala (1997) scheme. Parameters are detailed in text.

Malécot (1969, Chap. IX, Eq. 7) assumed that two genes separated by y had a probability $g_1(x)\, g_1(y-x)\, dx^2$ of both coming from a small region dx in the previous generation, and a probability $1/(2\rho_e\, dx)$ of identity if they did come from this small region (actually he writes δ for $2\rho_e$), where g_1 is a Gaussian density. Thus, his recursion is equivalent to assuming that

$$g(x, x', y) = g_1(x)\, g_1(x')\left(1-\frac{\Delta t}{2\rho_e}\delta_0(y+x-x')\right),$$

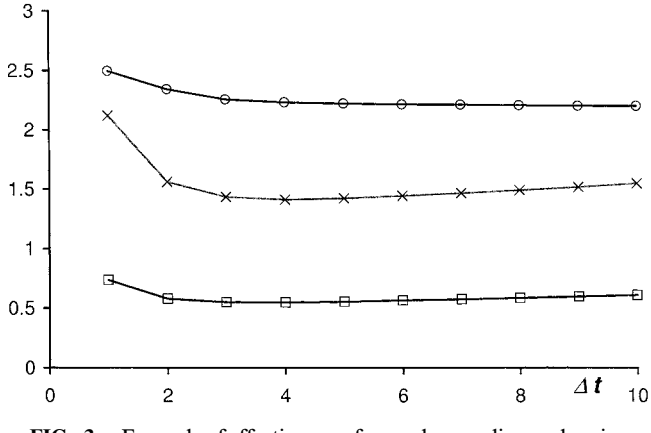

FIG. 3. Example of effective over forward mean dispersal variance (empty circles, for sufficiently large Δt, this tends to a constant, and the corresponding dispersal distribution g_1 to a Gaussian); inbreeding effective density over mean global density (empty squares); and $\mathcal{N}$ (crosses) as a function of the time slice Δt.

where δ_0 is the Dirac delta function at the origin, and

$$h(y) = \int g_1(x)\, g_1(y-x)\, dx.$$

It is important to appreciate that we define the functions g and h, and the recursion (23), for a population which is in a stationary state. This stationary state will in general consist of transient local clusters, whose density influences the reproduction of the genes within them. Local interactions between genes arise not just from the direct influence of one gene on another's reproduction, but from the influence of the common environment on both. The scale over which interactions persist is set by the scale of local density fluctuations, and so can be much longer than the generation time.

Two related classes of assumptions are needed to make progress. First, for sufficiently well-separated genes ($|y|$ large), the chance of coancestry must be negligible ($h(y) \ll 1$), and the movements of the lineages must be independent (i.e., $g(x, x', y)$ must separate into $g_1(x)\, g_1(x')$). Second, for a sufficiently large time step Δt, it must be possible to apply the recursion over successive steps. In general, the movements of lineages in one time step may be correlated with their movements in previous steps—for example, if individuals tend to disperse away from a temporarily crowded cluster. However, provided that we examine the genealogy over sufficiently long times (i.e., choose μ sufficiently small), we can choose Δt sufficiently large for such temporal correlations to become negligible. Note that Δt can be defined for both discrete-time and continuous-time models, and that Eq. (23) includes the stepping stone model as a special case.

The two classes of assumptions are related in that if ancestors Δt in the past tend to be well separated ($|y+x-x'|$ large) then their movements in previous time steps will be independent of each other. It follows from these assumptions that the functions h and g are Gaussian for sufficiently large Δt. However, it is not necessary to assume this in the following.

5. FOURIER TRANSFORMS

Following Malécot (1948), we take the Fourier transform of Eq. (23). We normalise the Fourier transform in two dimensions as follows:

$$\tilde{\phi}(\tilde{x}) = \frac{1}{2\pi}\int e^{ix.\tilde{x}}\phi(x)\, dx, \qquad \phi(x) = \frac{1}{2\pi}\int e^{-ix.\tilde{x}}\tilde{\phi}(\tilde{x})\, d\tilde{x}. \tag{24}$$

We now apply the Fourier transform to Eq. (23). First observe that

$$\frac{1}{2\pi}\int \phi_t(y+x-x')\, g(x, x', y)\, e^{iy.\tilde{y}}\, dx\, dx'\, dy$$

$$= \frac{1}{(2\pi)^5}\int \tilde{\phi}(\tilde{z})\, \tilde{g}(\tilde{a}, \tilde{b}, \tilde{c})\, e^{i\tilde{z}.(y+x-x')} e^{-i\tilde{a}.x} e^{-i\tilde{b}.x'} e^{-i\tilde{c}.y} e^{iy.\tilde{y}}$$
$$dx\, dx'\, dy\, d\tilde{a}\, d\tilde{b}\, d\tilde{c}\, d\tilde{z}$$

$$= \frac{1}{(2\pi)^5}\int \tilde{\phi}(\tilde{z})\, \tilde{g}(\tilde{a}, \tilde{b}, \tilde{c})\, e^{-ix.(\tilde{z}+\tilde{a})} e^{-ix'.(-\tilde{z}+\tilde{b})} e^{-iy.(\tilde{z}+\tilde{c}-\tilde{y})}$$
$$dx\, dx'\, dy\, d\tilde{a}\, d\tilde{b}\, d\tilde{c}\, d\tilde{z}.$$

Now using that

$$(2\pi)^2\, \delta_x = \int e^{ix.\tilde{x}}\, d\tilde{x}, \tag{25}$$

this gives

$$\tilde{\phi}_{t+\Delta t}(\tilde{y}) = e^{-2\mu\,\Delta t}\left\{2\pi\int \tilde{\phi}_t(\tilde{z})\, g(-\tilde{z}, \tilde{z}, \tilde{y}-\tilde{z})\, d\tilde{z} + \frac{\Delta t}{2\rho_e}\tilde{h}(\tilde{y})\right\}.$$

DEFINITION. We now assume that g is of the form

$$g(x, x', y) = g_1(x)\, g_1(x') - \frac{\Delta t}{2\rho_e}\eta(x, x', y).$$

The function g_1 is the density function for the (backward in time) spatial motion of a single individual (typically Brownian motion) at time Δt. Evidently, if the probabilities for the possible origin of our two individuals are to sum to one, we must have that

$$\int \eta(x, x', y)\, dx\, dx' = h(y).$$

This condition, and the conditions that g_1 and h both have integral one, become, in transform space,

$$2\pi\tilde{g}_1(0) = 1,$$
$$2\pi\tilde{h}(0) = 1,$$
$$(2\pi)^2\, \tilde{\eta}(0, 0, \tilde{y}) = \tilde{h}(\tilde{y}).$$

Moreover,

$$\tilde{g}(\tilde{x}, \tilde{x}', \tilde{y}) = \tilde{g}_1(\tilde{x})\, \tilde{g}_1(\tilde{x}')\, 2\pi\delta_{\tilde{y}} - \frac{\Delta t}{2\rho_e}\tilde{\eta}(\tilde{x}, \tilde{x}', \tilde{y}).$$

For Malécot's model,

$$\tilde{h}(\tilde{y}) = 2\pi \tilde{g}_1(\tilde{y})\, \tilde{g}_1(-\tilde{y})$$

$$\tilde{\eta}(\tilde{x}, \tilde{x}', \tilde{y}) = \frac{1}{2\pi}\, \tilde{g}_1(\tilde{x}-\tilde{y})\, \tilde{g}_1(\tilde{x}'+\tilde{y}).$$

The Fourier transform of Eq. (23) now becomes,

$$\tilde{\phi}_{t+\Delta t}(\tilde{y}) = e^{-2\mu\,\Delta t}\Bigg[(2\pi)^2\, \tilde{\phi}_t(\tilde{y})\, \tilde{g}_1(\tilde{y})\, \tilde{g}_1(-\tilde{y}) - 2\pi \int \tilde{\phi}_t(\tilde{z})\, \frac{\Delta t}{2\rho_e}\, \tilde{\eta}(-\tilde{z}, \tilde{z}, \tilde{y}-\tilde{z}) + \frac{\Delta t}{2\rho_e}\, \tilde{h}(\tilde{y})\, d\tilde{z} \Bigg]. \tag{26}$$

This has equilibrium solution

$$\tilde{\phi}(\tilde{y}) = e^{-2\mu\,\Delta t}\, \frac{\Delta t}{2\rho_e} \left(\frac{\tilde{h}(\tilde{y}) - 2\pi \int \tilde{\phi}(\tilde{z})\, \tilde{\eta}(-\tilde{z}, \tilde{z}, \tilde{y}-\tilde{z})\, d\tilde{z}}{1-(2\pi)^2\, e^{-2\mu\,\Delta t} \tilde{g}_1(\tilde{y})\, \tilde{g}_1(-\tilde{y})} \right)$$

$$= \frac{\tilde{\Omega}(\mu, \tilde{y})}{\mathcal{N}}\, (2\pi\tilde{h}(\tilde{y}) - \tilde{\phi}_e(\tilde{y})). \tag{27}$$

Here, as in Section 3, $\mathcal{N} = 4\pi\rho_e\sigma_e^2$ (where σ_e^2 is the variance of the dispersal distribution g_1) is Wright's (1943) neighbourhood size and is the dimensionless quantity that determines the relative rates of genetic drift and gene flow in two dimensions. We have made the definitions

$$\tilde{\Omega}(\mu, \tilde{y}) = \frac{e^{-2\mu\,\Delta t}\sigma^2\, \Delta t}{1-e^{-2\mu\,\Delta t}(2\pi)^2\, \tilde{g}_1(\tilde{y})\, \tilde{g}_1(-\tilde{y})},$$

and

$$\tilde{\phi}_e(\tilde{y}) = (2\pi)^2 \int \tilde{\phi}(\tilde{z})\, \tilde{\eta}(-\tilde{z}, \tilde{z}, \tilde{y}-\tilde{z})\, d\tilde{z}.$$

We shall take a closer look at $\tilde{\Omega}$ in the next section; this function arises in the case of independent evolution of lineages. The quantity ϕ_e is an "effective probability of identity" that corresponds to an average of the probability of identity over local scales, and has the effect of reducing the magnitude of local fluctuations.

Note that for Malécot's model, this immediately simplifies to give

$$\tilde{\phi}_e(\tilde{y}) = 2\pi\tilde{h}(\tilde{y})\, \phi(0)$$

(where we have used that $\int \tilde{\phi}(\tilde{z})\, d\tilde{z} = 2\pi\phi(0)$). Substituting in Eq. (27) this gives, for the Malécot model,

$$\tilde{\phi}(\tilde{y}) = \frac{\tilde{\Omega}(\mu, \tilde{y})}{\mathcal{N}}\, 2\pi\tilde{h}(1-\phi(0))$$

$$= \frac{\tilde{\Omega}(\mu, \tilde{y})}{\mathcal{N}}\, (2\pi)^2\, \tilde{g}_1(\tilde{y})\, \tilde{g}_1(-\tilde{y})(1-\phi(0)). \tag{28}$$

This can be inverted to give Eq. (15) for small μ.

6. THE ROLE OF $\tilde{\Omega}$

In order to understand the series expansion of Section 8, it is useful to understand the object $\tilde{\Omega}$ in a little more detail. It should be evident from Eq. (28) that it arises from the recursion (23) in the case when lineages evolve independently of one another backward in time. This short section makes this connection explicit.

Consider the following equation:

$$F(y) = e^{-2\mu\,\Delta t} \left(f(y) + \int g_1(x)\, g_1(x')\, F(y+x'-x)\, dx\, dx' \right). \tag{29}$$

As before, we can take the Fourier transform of this equation. This yields

$$\tilde{F}(\tilde{y}) = \frac{1}{2\pi} \int e^{iy\cdot\tilde{y}} F(y)\, dy$$

$$= e^{-2\mu\,\Delta t} \left(\tilde{f}(\tilde{y}) + \frac{1}{2\pi} \int g_1(x)\, g_1(x')\, F(y+x'-x)\, e^{iy\cdot\tilde{y}}\, dx\, dx'\, dy \right)$$

$$= e^{-2\mu\,\Delta t} \left(\tilde{f}(\tilde{y}) + \frac{1}{(2\pi)^4} \int \tilde{g}_1(\tilde{a})\, \tilde{g}_1(\tilde{b})\, \tilde{F}(\tilde{c})\, e^{iy\cdot\tilde{y}} e^{-ix\cdot\tilde{a}} e^{-ix'\cdot\tilde{b}} e^{-i(y+x'-x)\cdot\tilde{c}}\, dx\, dx' dy\, d\tilde{a}\, d\tilde{b}\, d\tilde{c} \right)$$

$$= e^{-2\mu\,\Delta t} (\tilde{f}(\tilde{y}) + (2\pi)^2\, \tilde{g}_1(-\tilde{y})\, \tilde{g}_1(\tilde{y})\, \tilde{F}(\tilde{y})),$$

where, in the last line, we have once again used Eq. (25). Rearranging, this gives

$$\tilde{F}(\tilde{y}) = \frac{1}{\sigma^2\, \Delta t}\, \tilde{\Omega}(\mu, \tilde{y})\, \tilde{f}(\tilde{y}).$$

Malécot's formula corresponds to taking $f(y)$ in Eq. (29) to be

$$f(y) = \frac{(1-F(0))\, \Delta t}{2\rho_e} \int g_1(y-z)\, g_1(z)\, dz.$$

Equation (29) is then precisely the equilibrium version of the equation studied by Malécot (1969, Eq. 3.3.1). (Malécot also takes $\Delta t = 1$ and denotes $2\rho_e$ by δ.)

7. AN EXAMPLE

In this section, we use a simulated example to illustrate the relation between the basic recursion Eq. (23) and Malécot's approximation. We use Bolker and Pacala's (1997) model of density-dependent regulation in continuous space and time. This model was originally used to describe population dynamics, but its extension to describe genealogical structure is straightforward. Individuals are distributed on a 60×60 torus and follow a birth and death process with density dependence according to the following parameters, all expressed per arbitrary time unit: forward dispersal in the absence of density dependence $\sigma = 0.4$; fecundity is constant at $f = 2$; mortality is $m + \alpha\rho$ ($m = 1$, $\alpha = 0.6$). The density perceived by an individual, ρ, is a sum of contributions from each neighbour, with weight given by a Gaussian bivariate function of distance from the neighbour, with variance parameter $\upsilon = 0.4$; $\sqrt{\upsilon}$ defines the scale over which density-dependent regulation operates. At each birth or death, the density perceived by each individual is calculated, giving the birth and death rates for that individual. The location and time of the next event is then chosen appropriately; positions of any offspring are updated; and perceived densities are recalculated for the next step. Figure 2 shows a randomly chosen realisation of this model. This distribution shows substantial clumping: the value of $\bar{c}$, the average spatial covariance weighted by the density dependence kernel U (Bolker and Pacala, 1997), is 0.782 ($P < 0.001$ when tested against a simulated uniform distribution, showing $\bar{c} = 0.000 \pm 0.025$).

Figure 3 shows how the parameters ρ_e, σ_e^2 (the variance of the conditional backward migration function $g_1(x)$ for one individual) and $\mathcal{N} = 4\pi\rho_e\sigma_e^2$ vary with the choice of time slice. The variance of the displacements of genes increase linearly with Δt, so that the effective rate of dispersal tends to a constant for large Δt (upper curve in Fig. 3). The effective rate of dispersal, which is measured by tracing lineages backward in time on Δt, is substantially greater than the rate of forward dispersal. Conversely, the inbreeding effective population density, ρ_e, which is measured from the probability of identity by descent within Δt, is substantially lower than the census density (lower curve in Fig. 3). This is a consequence of the variance in fitness between individuals which is induced by variation in population density, and which is correlated over time. Although ρ_e in principle may vary with Δt, the dependence in this example is very weak. The neighbourhood size, which is the product of effective dispersal and density, is increased by about 50% by the effects of clumping.

Figure 4 compares the observed identity (upper dots) with various approximations. The approximations are calculated using an arbitrary time slice $\Delta t = 5$, for which $\sigma_e^2 = 0.355$ and $\rho_e = 0.636$; $\mathcal{N} = 1.42$. The upper curve shows Malécot's approximation (15), with local scale fitted as $\kappa = 2.9$. The lower curve shows the solution to the basic recursion Eq. (23), based on the observed bivariate distribution of distances between genes at the beginning and end of the time slice (i.e., of $|y + x - x'|$ and $|y|$). This fits closely to the observed identity, and to Malécot's approximation, for well-separated genes ($|y| > 2$, say). However, the basic recursion fails for nearby genes. As discussed above, this is because the average identity between ancestral lineages a given distance apart depends on the local population density, and hence is correlated with whether they produce offspring or not. The bottom of Fig. 4 shows the same graphs, but on a log scale; this shows the agreement over large distances more clearly.

8. AN EXPANSION IN $1/\mathcal{N}$

In this section we return to our general recursion (23) and obtain an exact solution in the form of a series expansion in powers of $1/\mathcal{N}$. The idea is to write

$$\tilde{\phi} = \sum_{k=1}^{\infty} \tilde{\phi}_k, \qquad \tilde{\phi}_e = \sum_{k=1}^{\infty} \tilde{\phi}_{e,k},$$

with $\tilde{\phi}_k$, $\tilde{\phi}_{e,k}$ of order $1/(\mathcal{N})^k$. Substituting in Eq. (27) we obtain

$$\sum_{k=1}^{\infty} \tilde{\phi}_k(\tilde{y}) = \frac{\tilde{\Omega}(\mu, \tilde{y})}{\mathcal{N}} \left(2\pi\tilde{h}(\tilde{y}) - \sum_{k=1}^{\infty} \tilde{\phi}_{e,k}(\tilde{y}) \right).$$

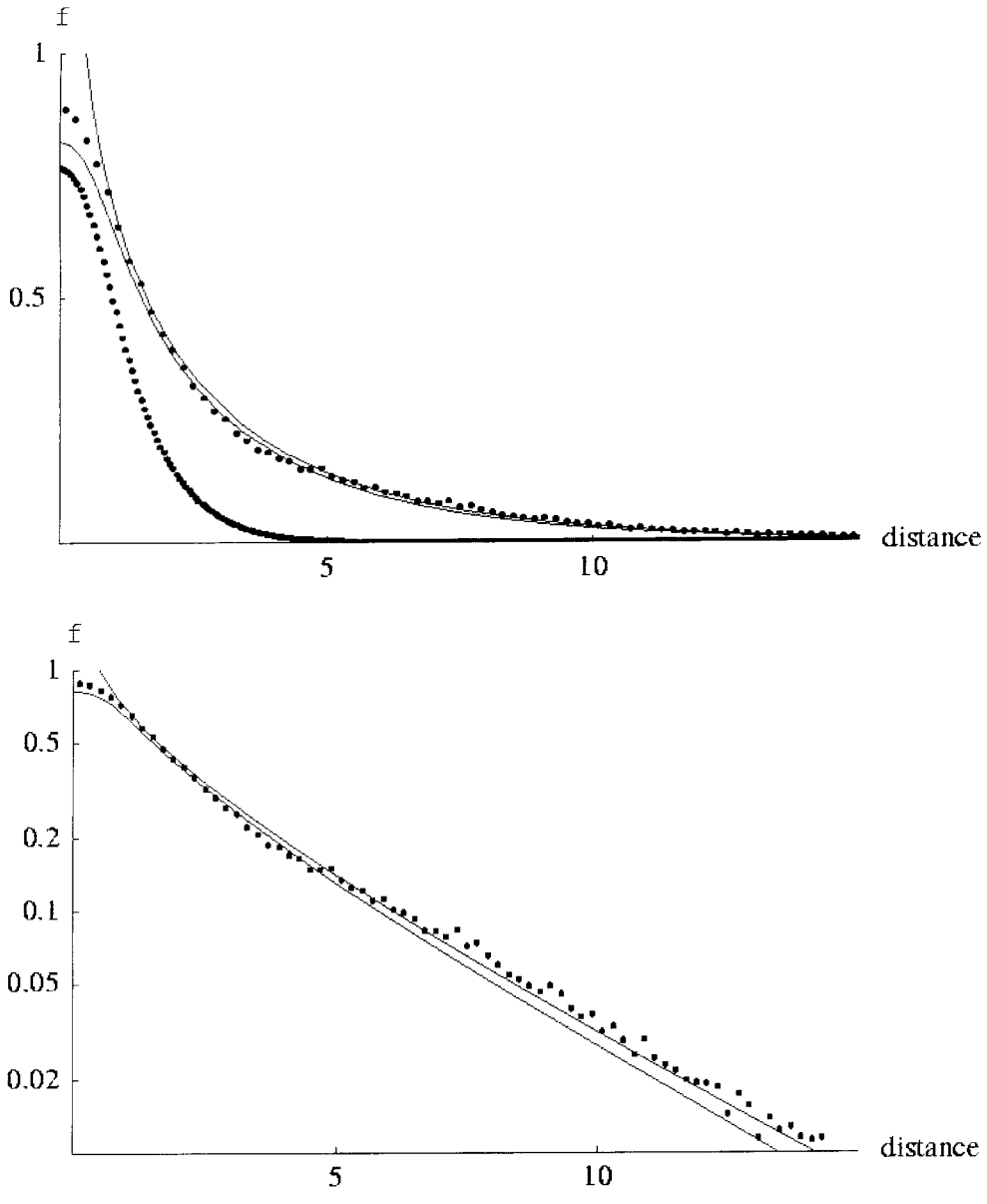

FIG. 4. (Top) Comparison between simulated identities (upper dots), values predicted from the Malécot approximation Eq. (14) with κ adjusted for best fit (upper curve), and solutions of recursion (23) based on the simulated h and g distributions (lower curve, $\mu = 10^{-2}$, $\Delta t = 5$, other parameters as in Fig. 2). The simulated h distribution is also shown (lower dots). (Bottom) Same figure on log scale (showing only identities).

This gives

$$\tilde{\phi}_1(\tilde{y}) = \frac{\tilde{\Omega}(\mu, \tilde{y})}{\mathcal{N}} 2\pi\tilde{h}(\tilde{y}),$$

and if, for ease of notation, we write

$$\tilde{\phi}_{e,0}(\tilde{y}) = -2\pi\tilde{h}(\tilde{y}),$$

then in general

$$\tilde{\phi}_k(\tilde{y}) = -\frac{\tilde{\Omega}(\mu, \tilde{y})}{\mathcal{N}} \tilde{\phi}_{e,k-1}(\tilde{y}),$$

and

$$\tilde{\phi}_{e,k}(\tilde{y}) = (2\pi)^2 \int \tilde{\phi}_k(\tilde{z})\, \tilde{\eta}(-\tilde{z}, \tilde{z}, \tilde{y}-\tilde{z})\, d\tilde{z}$$

$$= -(2\pi)^2 \int \frac{\tilde{\Omega}(\mu, \tilde{z})}{\mathcal{N}} \tilde{\phi}_{e,k-1}(\tilde{z})\, \tilde{\eta}(-\tilde{z}, \tilde{z}, \tilde{y}-\tilde{z})\, d\tilde{z}.$$

Using the observation of Section 6 we can identify the (untransformed) terms in this expansion in terms of solutions to Eq. (29) that arises in the study of independent evolution of lineages. For ease of reference we rerecord the basic Eq. (23) in equilibrium form, and with our special form of the function g:

$$\phi(y) = e^{-2\mu\,\Delta t}\left\{\frac{\Delta t}{2\rho_e}h(y) + \int \phi(y+x-x')\,g_1(x)\,g_1(x')\,dx\,dx' - \int \phi(y+x-x')\,\eta(x,x',y)\,dx\,dx'\right\}. \quad (30)$$

Notice then that ϕ_1 is a solution to (29) with $f(y) = (\Delta t/2\rho_e)\,h(y)$. This solution will dominate the true solution to (30), and the role of $\phi_{e,1}$ is to try to compensate for that. Specifically,

$$\phi_{e,1}(y) = e^{-2\mu\,\Delta t}\int \phi_1(y+x-x')\,\eta(x,x',y)\,dx\,dx'.$$

The next term in the series, ϕ_2, solves (29) with $f(y) = -\phi_{e,1}(y)$. With this definition, $\phi_1+\phi_2$ satisfies (30), but for an extra term that we denote by $\phi_{e,2}$ on the right-hand side. We then solve (29) with $f(y) = -\phi_{e,2}$, and so on. In general,

$$\phi_{e,k}(y) = e^{-2\mu\,\Delta t}\int \phi_k(y+x-x')\,\eta(x,x',y)\,dx\,dx',$$

and ϕ_k solves (29) with f replaced by $-\phi_{e,k-1}$.

We now examine the behaviour of the series expansion over large scales (i.e., for small $\tilde{y}$). Since we have assumed that h and η are localised, one expects that over sufficiently large spatial scales, we can approximate their effect by that of constant multiples of delta functions and thus recover the classical Malécot formula.

In the limit of small μ, all the dependence on $\tilde{y}$ is contained within $\tilde{\Omega}(\mu,\tilde{y})$ and $\tilde{\phi}_{e,k}(\tilde{y}) = \tilde{\phi}_{e,k}(0)+O(\mu)$. Hence,

$$\tilde{\phi}(\tilde{y}) = \sum_{k=1}^{\infty}\tilde{\phi}_k(\tilde{y}) = \frac{\tilde{\Omega}(\mu,\tilde{y})}{\mathcal{N}}\left(1+\sum_{k=1}^{\infty}\tilde{\phi}_{e,k}(0)\right), \quad (31)$$

where

$$\tilde{\phi}_{e,k}(0) = -(2\pi)^3\int \frac{\tilde{\Omega}(\mu,\tilde{z})}{\mathcal{N}}\,\tilde{\phi}_{e,k-1}(\tilde{z})\,\tilde{\eta}(-\tilde{z},\tilde{z},-\tilde{z})\,d\tilde{z}. \quad (32)$$

Thus, the large scale behaviour of the identity is proportional to $\tilde{\Omega}(\mu,\tilde{y})$ and, hence, to the classical Malécot formula. It is deflated by a complicated function of $\mathcal{N}$ and $\log(1/\mu)$, $1+\sum_{k=1}^{\infty}\tilde{\phi}_{e,k}(0)$. This should be compared to the Malécot formula, which is deflated by

$$(1-\phi(0)) = \frac{\mathcal{N}}{\mathcal{N}+\log(1/\sqrt{2\mu})} = 1+\sum_{k=1}^{\infty}(-1)^k\frac{\log(1/\sqrt{2\mu})^k}{\mathcal{N}^k}.$$

The lowest order terms in our expansion should be little affected by the "finite range" interaction introduced by h and η. However, the kth-order term will involve a kth-order convolution, so that the effect of the interaction rapidly ceases to be localised. Thus, for small neighbourhood sizes, even over large scales the identity cannot be summarised precisely by a single factor of the form

$$\frac{\mathcal{N}}{\mathcal{N}+\log(\sigma_e/\kappa\sqrt{2\mu})},$$

valid for a range of μ and $\mathcal{N}$.

If the function η is positive, then the series is alternating, and so if the terms (eventually) tend to zero, then convergence at least is guaranteed.

The first term in this series will be of order $\log(1/\mu)$, and so convergence will be slow if $\log(1/\mu)$ is large, and $\mathcal{N}$ small. In the following section, we derive an alternative expansion appropriate for this case.

9. EXPANSION OF THE SOLUTION FOR SMALL μ

In this section we present a slightly different approach to iteration of the identity (23) that, in good situations, leads to a series in ascending powers of $1/\log(1/\mu)$. In situations where the neighbourhood size is small, but so too is μ, this might lead to a good approximation with comparatively few terms. The idea is to combine the fact that we expect the solution to be close to that for independent spatial motions (essentially obtained by Malécot) and the basic sense of the iteration in the (untransformed) setting of the previous section.

Consider first the recursion

$$\phi(y) = e^{-2\mu\,\Delta t}\left\{\psi(y)+\int \phi(y+x-x')\,g_1(x)\,g_1(x')\,dx\,dx'\left(1-\frac{\Delta t}{2\rho_e}h(y)\right)\right\}, \quad (33)$$

with $\psi(y) = h(y)\,\Delta t/(2\rho_e)$. It is shown how to solve equations of this form in the Appendix. As usual, we can think of ϕ as the moment generating function of the time to coalescence of the two lineages where the probability of coalescence in a time interval of length Δt, if the separation is currently y, is $h(y)\,\Delta t/(2\rho_e)$, and if two lineages do not coalesce then they follow independent Brownian motions. If μ is small, then we may approximate this by the solution to a differential equation. If, in addition, h is Gaussian (this strong assumption is convenient but not necessary), then the solution is well approximated by the Malécot formula in two dimensions. We write

$$\phi_M(y) = \frac{1}{\mathcal{N} + \log(\sigma/\sqrt{2\mu})} K_0\left(\frac{\sqrt{2\mu}\,|y|}{\sigma}\right), \qquad y \gg \sqrt{\mu}.$$

Assumption. From now on, we assume that h is Gaussian.

Now write our basic recursion as

$$\begin{aligned}\phi(y) = e^{-2\mu\,\Delta t}\Bigg\{ &\frac{\Delta t}{2\rho_e} h(y) + \int \phi(y+x-x')\, g_1(x)\, g_1(x') \\ &dx\,dx' \left(1 - \frac{\Delta t}{2\rho_e} h(y)\right) \\ &+ \int \phi(y+x-x')\, g_1(x)\, g_1(x')\, dx\,dx' \frac{\Delta t}{2\rho_e} h(y) \\ &- \int \phi(y+x-x') \frac{\Delta t}{2\rho_e} \eta(x, x', y)\, dx\,dx' \Bigg\}. \quad (34)\end{aligned}$$

Remark. Before proceeding with the iteration, we observe that Malécot's formula is exact if the last line above vanishes when we substitute $\phi = \phi_M$. That is,

$$\begin{aligned}&\int \phi_M(y+x-x')\, g_1(x)\, g_1(x')\, dx\,dx' \frac{\Delta t}{2\rho_e} h(y) \\ &- \int \phi_M(y+x-x') \frac{\Delta t}{2\rho_e} \eta(x, x', y)\, dx\,dx' = 0.\end{aligned}$$

It is slightly more illuminating to write this as

$$\int \phi_M(y+x-x')[g_1(x)\, g_1(x') - g_0(x, x', y)]\, dx\,dx' = 0. \quad (35)$$

Here we have used g_0 to denote the transition density for the separation of lineages ($g_0 = g/(1 - h\,\Delta t/2\rho_e)$), and we are assuming that the probability of coalescence of two lineages y currently at separation y in the next Δt is always strictly less than one.

We now turn to a series expansion that, under an assumption on g_0 that we make explicit below, allows us to find the correction to Malécot's formula in an elementary way.

Equation (34) suggests that we can approximate the solution to our basic recursion by first solving for the case of independent lineages. The corresponding solution, ϕ_M, will not solve (34). There is an additional term:

$$\begin{aligned}\psi_{e,0}(y) = &\int \phi_M(y+x-x')\, g_1(x)\, g_1(x')\, dx\,dx' \frac{\Delta t}{2\rho_e} h(y) \\ &- \int \phi_M(y+x-x') \frac{\Delta t}{2\rho_e} \eta(x, x', y)\, dx\,dx'. \quad (36)\end{aligned}$$

Using the notation of Eq. (35), we write this as

$$\begin{aligned}\psi_{e,0}(y) = \Bigg\{ &-\int \phi_M(y+x-x')\, g_1(x)\, g_1(x')\, dx\,dx' \\ &+ \int \phi_M(y+x-x')\, g_0(x, x', y)\, dx\,dx' \Bigg\} \\ &\left(1 - \frac{\Delta t}{2\rho_e} h(y)\right). \quad (37)\end{aligned}$$

If we write X_1 for the separation of two independent Brownian motions at time Δt, and X_2 for the separation of two lineages evolving according to g_0, then substituting for ϕ_M, provided that both X_1 and X_2 are (with probability at least 1- $1/\log(1/\sqrt{\mu})$) larger than our local scale κ at which the Bessel approximation breaks down, then even for small neighbourhood size, $\psi_{e,0}(y)$ is bounded above by

$$\begin{aligned}&\left(1 - \frac{\Delta t}{2\rho_e}\right) \frac{1}{\log(1/\sqrt{\mu})} \mathbb{E}_y\left[K_0\left(\frac{\sqrt{2\mu}}{\sigma}|X_2|\right)\right. \\ &\left. - K_0\left(\frac{\sqrt{2\mu}}{\sigma}|X_1|\right)\right]. \quad (38)\end{aligned}$$

(The subscript y reflects the initial separation of the lineages.) Since X_2 is the separation of the lineages conditional on them not having coalesced, if neighbourhood size is small, then we expect the probability that X_2 is small to be dominated by that for X_1. Notice that if X_1

and X_2 are, with high probability, also not too large, then the expectation is approximately

$$\mathbb{E}_y\left[\log\left(\frac{|X_1|}{|X_2|}\right)\right],$$

so that boundedness of this expectation implies $\psi_{e,0}$ is of order $1/\log(1/\sqrt{\mu})$. If we assume that, for large separations, the evolution of lineages is approximately independent, then $\psi_{e,0}(y)$ will tend to zero as $|y| \to \infty$.

We now solve Eq. (33) with $\psi = \psi_{e,0}$. The solution is (to a good approximation)

$$\phi_1(y) = \int \psi_{e,0}(z)\, dz\, \phi_0(y).$$

If the integral is of order $1/\log(1/\sqrt{\mu})$, then $\phi_M - \phi_1$ will solve Eq. (34), but for an error of order $O(\log(1/\mu)^{-2})$. Proceeding in this way leads to a series expansion in powers of $1/\log(1/\sqrt{\mu})$.

To summarise this section, under our standing assumption that Δt is large enough that we can iterate Eq. (23), ignoring correlations over successive time intervals, and under the additional assumption that h is Gaussian, then provided that the expectation in Eq. (38) is bounded, the probability of identity is

$$\phi_M(y)\left(1 - \int \psi_{e,0}(z)\, dz\right),$$

up to an error of order $1/(\log(1/\mu))^2$. If the quantity $\psi_{e,0}(y)$ is large, then the Malécot approximation, ϕ_M, may be very misleading.

10. DISCUSSION

Malécot (1948, 1969) introduced a simple approximation to the relationship between two neutral genes evolving in a continuous two-dimensional population, Eq. (15). This approximation suffers from two difficulties. First, the simplest route to (15) is to assume that the population can be described by a randomly varying allele frequency; the movement of genes is approximated by diffusion at a rate σ^2, and random genetic drift by an uncorrelated white noise, inversely proportional to the density of genes, 2ρ. This naive approximation breaks down in two dimensions, since local fluctuations diverge. The second difficulty, which arises in both one and two dimensions, is that in general, the movement (and possibly the reproduction) of nearby genes is correlated (Kingman, 1977). The assumption made by Malécot, and by almost all subsequent population genetic models of spatially extended populations, is that nearby genes move independently of each other, and that the population is in a stationary state. These assumptions can only hold if individuals are confined to discrete demes, each absolutely regulated in size.

In Section 2, we deal with the first difficulty by showing that for discrete stepping stone models, the Malécot approximation is valid over all but local scales. More precisely, the identity in state between two genes separated by $|x|$ converges to

$$\frac{(1-\phi(0))}{\mathcal{N}} K_0\left(\frac{\sqrt{(2\mu)}\,|x|}{\sigma}\right) \quad \text{for} \quad |x| \gg \sigma, \quad \mu \ll 1.$$

The second difficulty raises substantially greater problems. We present a general recursion, (23), which depends on the probabilities that two genes either derived from two separate genes at time Δt in the past, or are identical by descent from one gene during this arbitrary interval. This recursion is valid provided that the population has reached a stationary state, and provided that the time interval is long enough for movements in successive intervals to be uncorrelated. With the further assumption that well-separated genes move independently, we derive Malécot's approximation in the limit of small mutation rates and large separations. Now, the identity in state converges to

$$\frac{(1-\phi_e(0))}{\mathcal{N}} K_0\left(\frac{\sqrt{(2\mu)}\,|x|}{\sigma_e}\right).$$

Thus, the rate of generation of identity within local regions is deflated by a factor $(1-\phi_e(0))$, where $\phi_e(0)$ is somewhat smaller than $\phi(0)$. The value of this effective local identity, $\phi_e(0)$, and hence the magnitude of identities over larger scales, depends on the complexities of local interactions. We show how it can be calculated explicitly using two alternative series expansions, and compare these calculations with simulations.

In general, the identity over all but the smallest scales can be written as a function (15) of three parameters: the effective dispersal rate, σ_e; the neighbourhood size, $\mathcal{N} = 4\pi\rho_e\sigma_e^2$, where ρ_e is an effective population density; and the local scale, κ, defined through

$$\log\left(\frac{\sigma_e}{\kappa\sqrt{2\mu}}\right) = \frac{\mathcal{N}\phi_e(0)}{(1-\phi_e(0))}.$$

The effective dispersal rate, σ_e, depends on the rate of diffusion of lineages traced backward in time. When

local population densities fluctuate, this can be considerably greater than the variance of distance between parent and offspring, traced forward in time. This is because lineages that disperse out of crowded areas may have much greater long-term success than those that do not. The effective population density, ρ_e, is defined as the inverse of the probability that nearby genes become identical by descent in the previous time Δt, integrated over all separations. This may be considerably lower than the census population density when local densities fluctuate. The local scale, κ, can be thought of as the distance over which the Malécot approximation breaks down, and local interactions become significant.

The value of the Malécot approximation is that it allows the large-scale evolution of arbitrary two-dimensional populations to be described by just three parameters. Conversely, these parameters can be estimated from samples of genes, without knowledge of the detailed local population structure. The covariance between allele frequencies in spatially separated samples is given by (15), while the variance of allele frequencies in principle depends on local complications. However, this variance will, for all practical purposes, be proportional to $\phi_e(0)$, which is determined by the local scale, κ. Moreover, the single value of κ allows calculation of $\phi_e(0)$ for all μ.

Our general recursion (23) is defined in terms of an arbitrary time slice, Δt. In two dimensions, this raises a subtle complication. The probability of identity within time Δt, and hence the effective population density, depends logarithmically on Δt (Fig. 3). (In one dimension, there is no such dependence for $\mu \Delta t \ll 1$.) Similarly, the effective local identity, $\phi_e(0)$, and hence the local scale κ, depends on the choice of time slice. However, any measurable properties defined independent of Δt, such as the identity over large scales, must be independent of this choice. Since, from (12), the identity over large scales is proportional to $(1-\phi_e(0))/\mathcal{N}$, the dependence of numerator and denominator on Δt must cancel. Note, however, that the dependence of both on Δt is negligible for large $\mathcal{N}$ when $\phi_e(0)$ becomes small (Fig. 3).

Our recursion can only apply if movements in successive time slices are independent of each other. This requirement is more restrictive than it seems, since movements can become correlated indirectly, through fluctuations in the surrounding population, as well as directly, through the immediate effect of one individual on its neighbour's reproduction and dispersal. If a region becomes by chance unusually crowded, then all the individuals within that region will have lower fitness, and may disperse at a different rate. Such correlated effects will last as long as such clumped densities persist, and may extend over large temporal and spatial scales. In particular, we expect such large-scale correlations when the parameters are close to the critical values at which the population goes extinct.

In this paper, we have dealt with only the very simplest problem: the pairwise relationship between two neutral genes within a stationary population. Our calculations of probability of identity in an allelic state extend immediately to give the distribution of pairwise coalescence times: ϕ, considered a function of 2μ, is the Laplace transform of this distribution. The relation with the variance and covariance of allele frequencies in samples is not quite so straightforward, since these depend on how the samples are taken. One might take a series of samples, each consisting of the entire population within some region. These samples would vary in size, and the covariance between them would in general depend on their size. Thus, the estimated covariance in allele frequency would depend on what weight is given to large and small samples (i.e., to dense and sparse regions). However, we expect this effect to be negligible for large neighbourhood size.

The recursion used here extends in an obvious way to three or more genes, and so gives a method for approximating the distribution of genealogies amongst samples of well-separated genes (see Barton and Wilson, 1996, Eq. 3.14). However, additional assumptions need to be made if such genealogies are to be dominated by pairwise coalescence, at a rate described by the effective population density: in principle, multiple coalescence might be significant. If several genes are sampled from one location, and if neighbourhood size is small, then there is an appreciable chance that several will coalesce before diffusing out to become well-separated ancestral lineages. If we define local structure in terms of a time slice Δt, we cannot in general recover a process of pairwise coalescence just as we cannot in a structured coalescent with small deme size. The local genealogical structure will depend on details of the model, and cannot be approximated using our approach.

Extending the recursion (23) to multiple genes soon becomes impracticable, even for numerical calculations. Two more tractable approaches come to mind. First, one would like an algorithm for simulating coalescing lineages backward in time. By analogy with Wright (1943) and Malécot (1948), one could assume that lineages diffuse backward in time independently, and coalesce with probability $1/(2\rho_e \delta A)$ when they approach to within some small region of area δA of each other. This yields the Malécot approximation (15) and so is consistent with our derivations here. However, a rigorous justification for this procedure is needed, and in

any case, it will only apply when $\mathcal{N}$ is large enough that a sufficiently small δA can be found without $1/2\rho_e\ \delta A > 1$.

An alternative approach is to write down a stochastic partial differential equation for allele frequency. The time derivative of the allele frequency will be given as the sum of terms representing diffusion, mutation, and random noise. If we are to make sense of the equation in two dimensions, then the noise should be "coloured" so as to capture the local scale κ. (Over large spatial scales, there will be no correlations in the noise.) Of course, in a spatially continuous population, no tangible allele frequency exists. At any point, there is almost certainly no gene. Nevertheless, such an equation can be used to model allele frequencies in a small region (whose size will be dictated by κ) centred on each point (the overlap of regions around different points being reflected in the coloured noise). However, only in very special cases will there be closed form expressions for the moments of the allele frequencies. Only if the neighbourhood size is sufficiently large can we hope to approximate the equation by one in which the noise term is sufficiently simple to lead to an elementary calculation.

Our approach assumes that the population dynamics are in a stationary state. If the dynamics change slowly, with changing environmental conditions or with changing frequencies of alleles that influence fitness, then our intuition is that the same recursion applies: local fluctuations should reach a quasi-equilibrium over time scales Δt. Extending the Malécot approximation to include selection poses greater difficulties. If selection is strong, then allelic states have an immediate influence on the population dynamics, and the problem appears hopelessly complicated. However, if selection is weak, we can assume that its influence on the numbers of individuals is negligible, and that the distribution of allelic states changes slowly. We can then hope to recover a stochastic partial differential equation of the form indicated above, but now with an extra term representing selection and with the same caveat that unless neighbourhood size is large, calculations of moments may be intractable.

A stochastic model of spatially continuous populations which includes selection as well as migration and random drift would be valuable for understanding the evolution of spatially extended populations. For example, the diffusion approximation which has been so successful in analysing clines in one dimension cannot at present be applied to include random drift in two dimensions (Nagylaki, 1978a,b); yet, almost all data come from two-dimensional populations. Including selection in the framework outlined here would, we believe, also be valuable in clarifying our understanding of group and kin selection (cf. Nunney, 1989), and of the relation between ecological and population genetics processes.

APPENDIX

Solving Equation (33)

In this section we describe an approach to approximating the solution to Eq. (33). We continue to assume that h is a Gaussian density. For easy reference, let us rerecord the equation

$$\phi(y) = e^{-2\mu\,\Delta t}\left\{\psi(y) + \int \phi(y+x-x')\, g_1(x)\, g_1(x')\, dx\, dx' \left(1 - \frac{\Delta t}{2\rho_e} h(y)\right)\right\}.$$

It is convenient to rewrite this as

$$\phi(y) = e^{-2\mu\,\Delta t}\left\{f(y)\,\frac{\Delta t}{2\rho_e} h(y) + \int \phi(y+x-x')\, g_1(x)\, g_1(x')\, dx\, dx' \left(1 - \frac{\Delta t}{2\rho_e} h(y)\right)\right\}. \tag{A1}$$

If we iterate this equation, we see that it has solution

$$\phi(y) = \mathbb{E}_y[f(X_{T-1})\, e^{-2\mu T}], \tag{A2}$$

where T is the time to coalescence of two lineages with initial separation y (measured in units of Δt). Now, as usual, we approximate the distribution of $T\Delta t$ by replacing the random walk by Brownian motion, leading in the case of $f = 1$ to Bessel's equation.

For completeness, we record a rather general situation. Suppose that

$$w(x) = \mathbb{E}_x[l(X_T)\, e^{-2\mu T}],$$

where

$$\mathbb{P}_x[T > t] = \mathbb{E}_x\left[\exp\left(-\int_0^t k(X_s)\, ds\right)\right].$$

Then

$$w(x) = \int_0^\infty \mathbb{E}_x\left[l(X_t)\, k(X_t) \exp\left(-\int_0^t k(X_s)\, ds\right)\right] e^{-2\mu t}\, dt,$$

and an integration by parts reveals that

$$2\mu w - Aw + kw = lk. \tag{A3}$$

(As usual, A is the infinitesimal generator of X.)

Now, we know how to solve this equation for $k = \delta_0$. (The solution is a modified Bessel function of type 0.) Moreover, this choice of k gives a good approximation for the distribution of the coalescence time of the two lineages whenever h is Gaussian. We should like to take this value of k again, but in order to ensure a good approximation to the solution (A2) to Eq. (A1), we must find a means of approximating $f(X_{T-1})$. Now, given that we coalesce at time T, our distribution at time $T-1$ is given by $h(y)$. Thus, if we set $l(0) = \int_{\mathbb{R}^2} h(z)\, f(z)\, dz$ and $k = \delta_0$, the solution to Eq. (A3) will, for small μ, be a good approximation to the expression (A2).

ACKNOWLEDGMENTS

This work was supported by grants from the EPSRC (GR/L10048 and an advanced fellowship for A.M.E.) and NERC (GR3/11635) and by the Darwin Trust of Edinburgh. We thank Anja Sturm for her assistance with the project and anonymous reviewers for helpful comments. This paper is dedicated to Charlotte, A.M.E.'s daughter born during the gestation of the manuscript.

REFERENCES

Barton, N. H., and Wilson, I. 1996. Genealogies and geography, *in* "New Uses for New Phylogenies" (P. H. Harvey, A. J. Leigh Brown, J. Maynard Smith, and S. Nee, Eds.), Oxford Univ. Press, Oxford.

Bolker, B., and Pacala, S. W. 1997. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems, *Theor. Popul. Biol.* **52**, 179–197.

Erdelyi, A. 1954. Tables of integral transforms, "Bateman Manuscript Project," California Institute of Technology.

Felsenstein, J. 1975. A pain in the torus: Some difficulties with the model of isolation by distance, *Amer. Nat.* **109**, 359–368.

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

Kingman, J. F. C. 1977. Remarks on the spatial distribution of a reproducing population, *J. Appl. Probab.* **14**, 577–583.

Malécot, G. 1948. "Les Mathématiques de l'Hérédité," Masson et Cie, Paris.

Malécot, G. 1969. "The Mathematics of Heredity," Freeman, San Francisco.

Maruyama, T. 1971. Analysis of population structure. II. Two dimensional stepping stone models of finite length and other geographically structured populations, *Ann. Hum. Genet.* **35**, 179–196.

McKean, H. P. 1969. "Stochastic Integrals," Academic Press, New York.

Nagylaki, T. 1975. Conditions for the existence of clines, *Genetics* **80**, 595–615.

Nagylaki, T. 1978a. A diffusion model for geographically structured populations, *J. Math. Biol.* **6**, 375–382.

Nagylaki, T. 1978b. Random genetic drift in a cline, *Proc. Natl. Acad. Sci. USA* **75**, 423–426.

Nunney, L. 1989. The maintenance of sex by group selection, *Evolution* **43**, 245–257.

Sawyer, S., and Fleischmann, J. 1978. The maximal geographical range of a mutant allele considered as a subtype of a Brownian branching random field, *Proc. Natl. Acad. Sci. USA* **76**, 872–875.

Shiga, T. 1982. Continuous time multi-allelic stepping stone models in population genetics, *J. Math. Kyoto Univ.* **22**, 1–40.

Shiga, T. 1988. Stepping stone models in population genetics and population dynamics, *in* "Stochastic Processes in Physics and Engineering" (S. Albeverio *et al.*, Eds.), pp. 345–355, Reidel, Dordrecht.

Slatkin, M. 1991. Inbreeding coefficients and coalescence times, *Genet. Res. (Camb.)* **58**, 167–175.

Walsh, J. B. 1986. "An Introduction to Stochastic Partial Differential Equations," Ecole d'été de probabilités de St Flour, Lecture Notes in Mathematics, Vol. 1180, Springer-Verlag, Berlin.

Wright, S. 1943. Isolation by distance, *Genetics* **28**, 114–138.