Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

Stochastic models of evolution in genetics, ecology and linguistics

R A Blythe[1] **and A J McKane**[2]

[1] SUPA, School of Physics, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, UK
[2] Theory Group, School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK
E-mail: R.A.Blythe@ed.ac.uk and Alan.McKane@manchester.ac.uk

Received 16 March 2007
Accepted 1 May 2007
Published 23 July 2007

Online at stacks.iop.org/JSTAT/2007/P07018
doi:10.1088/1742-5468/2007/07/P07018

Abstract. We give an overview of stochastic models of evolution that have found applications in genetics, ecology and linguistics for an audience of non-specialists, especially statistical physicists. In particular, we focus mostly on neutral models in which no intrinsic advantage is ascribed to a particular type of the variable unit, for example a gene, appearing in the theory. In many cases these models are exactly solvable and furthermore go some way to describing observed features of genetic, ecological and linguistic systems.

Keywords: stochastic particle dynamics (theory), population dynamics (theory), interacting agent models

ArXiv ePrint: cond-mat/0703478

J. Stat. Mech. (2007) P07018

Contents

1. Introduction

The distinguishing feature of statistical mechanics as a theory of many-body physics is that it has probabilistic notions at its core. At equilibrium, the principle of equal *a priori* probabilities is invoked as a means to identify the typical behaviour of a system as a whole. Out of equilibrium, the evolution of a system is usually couched in terms of stochastic dynamics, as a glance at many articles in this journal will confirm. The procedure of stochastic modelling is not, of course, restricted to physics. In particular, stochastic models have played a pivotal role in understanding the dynamics of evolutionary systems and as such one sees many similarities in the approaches and methods that have been used to those employed by statistical physicists. The most well-established discipline for

which this is the case is population genetics, i.e., the study of gene frequencies among a population of reproducing individuals. More recently, models very similar to those that arise in the population genetics arena have also been applied in ecology (where the frequencies of particular species in an ecosystem are of interest) and language change (where now the focus is on frequencies of linguistic variants used by a community of speakers). In this paper, it is our aim to introduce key concepts and results from these fields to practitioners in statistical mechanics in order to stimulate further cross-fertilization between them.

This is not the first overview of the area with this audience in mind, various lecture notes and review articles having appeared over the years such as [1]–[3]. Mostly these works have focused on the question of adaptation via selective advantage. Although we briefly touch on some of these ideas later in this review, our main focus is on *neutral* models in which selection is absent and stochastic effects are most pronounced. Part of our motivation lies in the application these neutral models have found in genetics, ecology and linguistics (see below). However, we also feel that these models are of a fundamental interest from the point of view of non-equilibrium statistical mechanics.

Despite the diverse applications we have mentioned, we shall mostly use the language of population genetics, this being the discipline concerned that has the strongest tradition of quantitative modelling and mathematical analysis. We must therefore begin by introducing relevant terminology that may be unfamiliar to physicists. In the simplified mathematical models we shall be describing, a *gene* is taken to mean a variable unit that codes for a specific trait (such as eye colour), and different variants of the gene (coding in this example for different eye colours) are called *alleles*. Changes in allele frequency can occur as a consequence of three processes. First, there is reproduction, in which offspring organisms acquire copies of alleles from their parents. Secondly, in the copying process, random mutations can occur which may change one allele to another, or create a completely new allele. Finally, for whatever reason, organisms carrying one allele may end up having more offspring than another: this is selection.

The simplest stochastic model of an evolving population dates from the 1930s and was introduced independently by Fisher [4] and Wright [5]. In this model, the number of organisms is taken to be constant from one generation to the next, and each instance of a gene in one generation is an exact copy of one randomly chosen with replacement from the previous generation. Such a process could be realized in a number of ways. Let us consider first of all *haploid* organisms which carry only one instance of each gene and (possibly) reproduce asexually. A new generation could then be formed by each of the organisms producing an infinite number of *gametes* (e.g., seeds) before dying; then, a random fixed-size sample of these gametes survive to become offspring organisms. More often, geneticists focus on *diploid* organisms which carry two copies of each gene and reproduce sexually. Now, each organism produces two infinite sets of gametes, one for each allele carried by the parent. Then, after dying, the next generation is formed by combining the requisite number of randomly chosen gamete pairs. Either way, the mean number of offspring that any individual has in the following generation is one, and so there is no selection: this is a *neutral* model. Furthermore the fact that the allele instances present in one generation are a random sample of those present in the previous generation means that allele frequencies fluctuate through sampling effects alone. These fluctuations are called *genetic drift* and—in the absence of any mutations—over time cause permanent

J. Stat. Mech. (2007) P07018

extinction of one or more alleles: the steady state of such a model is one in which only a single allele remains: this allele is then said to have *fixed* in the population.

A haploid, or randomly mating diploid population evolving in this way is often referred to as an *ideal* population. In reality, individuals do not mate at random, and there is often a preference, or requirement, for mating to occur between or within different *classes* of individuals. Most obviously, in sexually reproducing organisms, females can mate only with males; other factors, such as individuals' geographical locations and age (when one has overlapping generations) may also lead to deviations from ideal behaviour. Nevertheless, in some cases it is found that the predictions of the ideal model are relevant if one replaces the actual size of the population with an effective size, which can be thought of as a number of breeding individuals. We will discuss this procedure in more detail later in this paper.

Neutral theories in population genetics (as well as in the ecological and linguistic applications—see below) seem to have a far greater realm of validity than might naively be expected. The idea that many genetic mutations are effectively neutral, that is do not significantly affect the fitness of the carrier, was pioneered by Kimura [6, 7]. Thus changes in their frequency are due to chance, rather than selection. This is not to deny that natural selection has an important role in the development of morphological and behavioural characteristics, only that random genetic drift has a more significant effect than had hitherto been envisaged. The relationship of the neutral theory to selection was explored further with the introduction of the concept of 'near-neutrality' [8, 9], in which the extent to which genes are affected by drift and selection is a function of the effective size of a breeding population.

In ecology the concept of neutrality is more recent, and is frequently associated with Hubbell, who has developed and popularized the concept during the last decade [10]. The idea of neutrality in ecology can be explained by imagining a set of similar species in a local community, trees in a forest, for example. When a particular tree dies, it is replaced by an 'offspring' of one of the other trees in the forest—picked at random. This is similar to the Wright–Fisher model, except that now individuals (trees) are the fundamental entities, not genes, and they come in different species, rather than as different alleles. As in the genetic case, successive random events will eventually lead to the extinction of one species by chance, and then another, until eventually only one will remain. That this does not happen is postulated to be due to new individuals occasionally being introduced from outside the local community by immigration. The distribution of species abundances calculated from neutral theory gives a remarkably good fit to data [11], however the extreme simplicity of the model has made it very controversial. Some ecologists see it as a good 'zeroth-order approximation' on which to build a more complete description [12], while others believe some other kind of starting point is required [13, 14].

The analogies to a model of language change [15]–[19] are a little more abstract. Here, there is a retained memory of usages of a particular variant of a linguistic variable (such as a vowel sound or grammatical structure); when a variant is reproduced, its retention in a speaker's memory displaces a record of an earlier utterance. Thus a speaker's perception of the frequency with which a particular variant is used—which in the quantitative version of the model which has been developed [20], is taken as the frequency with which that same speaker produces that variant—also exhibits genetic drift-like fluctuations over time. The imitation of linguistic variables according to usage frequencies experienced by individual

J. Stat. Mech. (2007) P07018

speakers falls within a paradigm referred to as the *usage based model* in linguistics [21] (this is to distinguish it from theories in which aspects of the grammar come 'pre-programmed'). Investigations of the model of language change introduced in [20] have only just begun, but it is already clear that the structure of the network of speakers, shaped by both geography and social interactions, is important in determining the time needed for a particular variant to become fixed. In some cases, such as in new dialect formation [22, 23], fixation of linguistic variants has been seen to occur on the timescale of a human generation: it is apparently possible that this may occur entirely through random effects, without the need to invoke any selection mechanism [24].

In the next section of this paper, we present a precise definition of the model of neutral genetic drift in an ideal population, along with the non-ideal generalization of a subdivided population. We also demonstrate concretely the relationship to the ecological and linguistic models just described. Then, in sections 3 and 4 we describe two complementary analytical approaches. The first starts with a master equation and leads to a diffusion approximation for allele frequencies; this type of approximation was popular in the 1950s, largely through Kimura's efforts. Although implicitly used previously, the 1980s saw the formalization of an approach called the coalescent, in which the statistical properties of the ancestry of a present-day population yield additional insights into the population dynamics. Finally, in sections 5 and 6 we review some of the special features of non-ideal and non-neutral evolution respectively, before making general remarks in the conclusion, section 7.

2. Genetic drift: the Wright–Fisher and Moran models

2.1. The Wright–Fisher model

The simplest, and most widely known [25]–[28], model of random genetic drift is the Wright–Fisher model [4, 5]. Suppose we have a population of individuals in generation t who mate randomly to produce the new generation $t+1$. We focus on one particular gene which may be in one of two states (alleles) which we call A and B. The Wright–Fisher model is based on the idea of the gene pool of the t generation individuals being sampled to give the genetic structure of the $t+1$ generation. In reality, as discussed in section 1 individuals are diploid, but for convenience we assume that individuals are haploid, that is, there is one-to-one correspondence between individuals and genes. The process of constructing the new generation from the current one is analogous to sampling coloured balls—representing different alleles (types of gene)—from a bag or urn. The steps in the process are illustrated in figure 1.

Suppose that in generation t there are n' A alleles and $(N-n')$ B alleles. We wish to sample this gene pool N times (with replacement) in order to create the new generation $t+1$. The probability that in N trials of sampling the pool will get n A alleles is simply given by the binomial distribution:

$$\binom{N}{n} p_1^n p_2^{N-n}, \tag{1}$$

where $p_1(n')$ is the probability of picking an A $(=n'/N)$ and $p_2(n')$ is the probability of picking a B $(=(N-n')/N)$. So the probability that there is a transition from the original

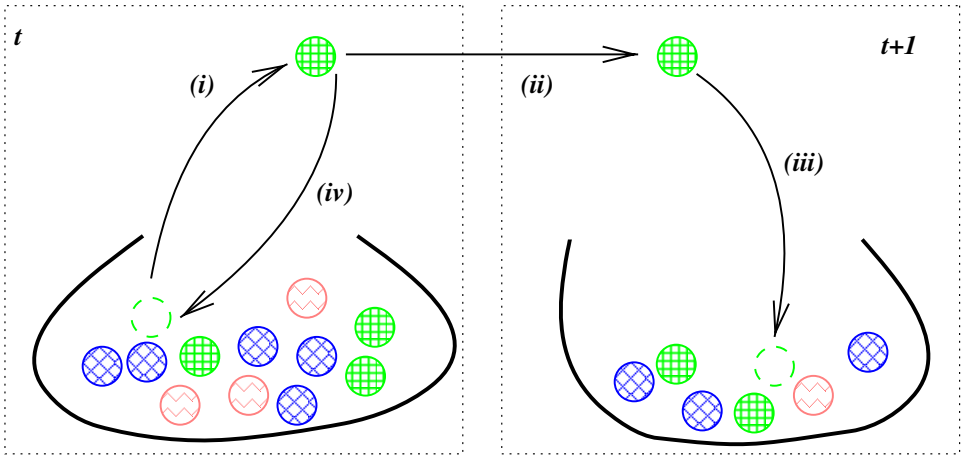

Figure 1. Wright–Fisher 'beanbag' population genetics. The population in generation $t+1$ is constructed from generation t by (i) selecting a gene from the current generation at random; (ii) copying this gene; (iii) placing the copy in the next generation; (iv) returning the original to the parent population. These steps are repeated until generation $t+1$ has the same sized population as generation t.

state (characterized by n' A alleles) to the new state (characterized by n A alleles) is given by

$$p_{n\,n'} = \binom{N}{n} \left(\frac{n'}{N}\right)^n \left(1 - \frac{n'}{N}\right)^{N-n}. \tag{2}$$

It should be noted that many mathematicians will write $p_{n'\,n}$ when they are referring to the transition probability from the state n' to the state n, whereas we follow the usual convention in physics and denote this by $p_{n\,n'}$. This makes much of the matrix algebra more natural—involving post-multiplication of vectors, rather than pre-multiplication.

What we wish to calculate is the probability of finding n A alleles in generation t, given that there were n_0 initially. We denote this probability by $P(n,t|n_0,0)$, which we will frequently abbreviate to $P(n,t)$, taking the initial condition as given. Since the construction of the gene pool at generation $t+1$ only depends on the state of the gene pool in generation t, and not on its state in earlier generations, the process is Markov. Theoretical biologists, like mathematicians, prefer to model such processes in terms of Markov chains. Thus they define firstly, a probability vector

$$\underline{P}(t) = \begin{pmatrix} P(1,t) \\ P(2,t) \\ P(3,t) \\ \vdots \end{pmatrix}$$

and secondly a transition matrix

$$\mathcal{P} = \begin{pmatrix} p_{11} & p_{12} & \cdots \\ p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

The evolution of the system is then given by the matrix equation

$$\underline{P}(t+1) = \mathcal{P}\underline{P}(t) \qquad \text{or} \qquad P(n,t+1) = \sum_{n'} p_{n\,n'} P(n',t). \tag{3}$$

Note that the sum of the components of the probability vector must add up to 1, and the sum of the entries in any one column of the transition matrix must add up to 1 (since $\sum_n P(n,t) = 1$ and $\sum_n p_{n\,n'} = 1$). Matrices whose column entries add up to 1 are known as *stochastic matrices*, although the convention of writing the initial state as the index on the left means that in the literature the statement is usually that the row entries add up to unity. It is now straightforward to formally express the state of the system in generation t in terms of this initial state:

$$\underline{P}(t) = \mathcal{P}\,\underline{P}(t-1) = \mathcal{P}\,\mathcal{P}\,\underline{P}(t-2) = \cdots = \mathcal{P}^t \underline{P}(0). \tag{4}$$

Clearly the large t behaviour will be dominated by the largest (in magnitude) eigenvalues of $\mathcal{P}$ and their corresponding eigenvectors. The eigenvalues of the Wright–Fisher model are known [29], and are given by

$$\lambda_r = \binom{N}{r} \frac{r!}{N^r}, \qquad r = 0, \ldots, N. \tag{5}$$

Some simple deductions may be made from the Wright–Fisher model formulated as a Markov chain. For instance, we may ask what the system looks like after a large number of generations. If it tends to a stationary state, and if we call this stationary probability vector $\underline{P}^*$, it follows that $\mathcal{P}\,\underline{P}^* = \underline{P}^*$. That is, $\underline{P}^*$ is a right eigenvector of $\mathcal{P}$ with eigenvalue 1. Intuitively it seems clear that the final state is either all A (the A allele has become *fixed*) or all B (the B allele is fixed). This would correspond to a stationary state given by

$$\underline{P}^* = \begin{pmatrix} 1-\Pi \\ 0 \\ \vdots \\ 0 \\ \Pi \end{pmatrix}. \tag{6}$$

Here Π is the probability that allele A will eventually become fixed. If we calculate $\mathcal{P}\,\underline{P}^*$ using the above expression, we find that it is indeed equal to $\underline{P}^*$, confirming that $\underline{P}^*$ is of this form.

Another simple deduction is that the expected number of A alleles does not change from one generation to the next:

$$\begin{aligned} \langle n(t+1)\rangle &= \sum_n n P(n,t+1) = \sum_n n \sum_{n'} p_{n\,n'} P(n',t) \\ &= \sum_{n'} n' P(n',t) = \langle n(t)\rangle, \end{aligned} \tag{7}$$

since $\sum_n n p_{n\,n'} = N p_1 = n'$. In particular, $\langle n(t)\rangle = \langle n(0)\rangle = n_0$. We can use this result to determine Π in equation (6). To do this we note from equation (6) that as $t \to \infty$ only

two states are possible: $n = N$ (with probability Π) and $n = 0$ (with probability $1 - \Pi$). Therefore

$$\lim_{t\to\infty} \langle n(t) \rangle = N\Pi + 0(1 - \Pi) \Rightarrow n_0 = N\Pi. \tag{8}$$

This gives the explicit form for $\underline{P}^*$ and also tells us that the probability that eventual fixation of A will occur is n_0/N.

2.2. The Moran model

Statistical physicists are not used to formulating stochastic processes in the way described above for the Wright–Fisher model; they usually use continuous time master equations and define transitions for a single stochastic step rather than for N bundled together [30]. A variant of the Wright–Fisher model, which is closer in spirit to stochastic processes considered by statistical physicists, was introduced by Moran [31]. This model does not have a fixed 'previous generation' from which N genes are sampled (with replacement) to form the next generation. Instead, genes are chosen to die one at a time, and replaced by new ones which are A with probability p_1 and B with probability p_2. These probabilities are calculated from the state of the system prior to the death of the chosen individual: $p_1(n) = n/N$ and $p_2(n) = (N - n)/N$, where n is the number of A alleles just before the combined birth–death event.

The Moran model, unlike the Wright–Fisher model, has overlapping generations: after N sampling events in the Moran model, each gene need not have been replaced, and several might not have survived very long at all. Due to the impossibility of unambiguously identifying a 'generation' in this case, a generation in the Moran model is frequently defined as the period over which a single sampling event takes place. This is a confusing use of the word 'generation', but at least it is well defined. So, in one generation the population of A may go up by one (if a B is chosen to die, and an A is chosen to be born) or down by one (if an A is chosen to die, and a B is chosen to be born). The number of A alleles will remain unchanged if both an A is chosen to die and to be born, and also if a B is chosen to die and to be born. This gives the transition matrix to be

$$p_{n'\,n} = \begin{cases} \left(1 - \dfrac{n}{N}\right) p_1(n), & \text{if } n' = n + 1 \\ \left(\dfrac{n}{N}\right) p_1(n) + \left(1 - \dfrac{n}{N}\right) p_2(n), & \text{if } n' = n \\ \left(\dfrac{n}{N}\right) p_2(n), & \text{if } n' = n - 1 \\ 0, & \text{otherwise,} \end{cases} \tag{9}$$

from which it follows that $\sum_{n'} p_{n'\,n} = 1$.

The explicit expressions for the non-zero elements of the transition matrix of the Moran model are therefore

$$\begin{aligned} p_{n+1\,n} &= \left(1 - \frac{n}{N}\right)\left(\frac{n}{N}\right), \\ p_{n\,n} &= \left(\frac{n}{N}\right)^2 + \left(1 - \frac{n}{N}\right)^2, \\ p_{n-1\,n} &= \left(\frac{n}{N}\right)\left(1 - \frac{n}{N}\right). \end{aligned} \tag{10}$$

The transition matrix (9) for the Moran model is much simpler than for the Wright–Fisher model, being tri-diagonal. Therefore it is not surprising that the eigenvalues of the transition matrix (9) can, as for the Wright–Fisher model, be found exactly [31]:

$$\lambda_r = 1 - \frac{r(r-1)}{N^2}, \qquad r = 0, \ldots, N, \tag{11}$$

but, unlike the Wright–Fisher model, the corresponding eigenvectors are also known [32]. The two largest eigenvalues are both unity; they have left eigenvectors $(1, 1, \ldots, 1)$ and $(0, 1, \ldots, N)$ and corresponding right eigenvectors $(1, 0, 0, \ldots, 0)^{\mathrm{T}}$ and $(0, 0, \ldots, 0, 1)^{\mathrm{T}}$. Clearly the last two column vectors correspond to the loss and fixation of the A allele. The third largest eigenvalue, $\lambda_2 = 1 - (2/N^2)$ has the left eigenvector $(0, N-1, 2(N-2), \ldots, n(N-n), \ldots, 0)$ and right eigenvector $(-(1/2)(N-1), 1, 1, \ldots, 1, -(1/2)(N-1))^{\mathrm{T}}$ [31, 32]. This column vector governs the behaviour, after a large number of generations, of the non-fixed states. In particular, we have, once t is sufficiently large,

$$\underline{P}(t) \approx \begin{pmatrix} 1 - \frac{n_0}{N} \\ 0 \\ \vdots \\ 0 \\ \frac{n_0}{N} \end{pmatrix} + \frac{6n_0(N-n_0)}{N(N^2-1)} \begin{pmatrix} -\frac{(N-1)}{2} \\ 1 \\ \vdots \\ 1 \\ -\frac{(N-1)}{2} \end{pmatrix} \left(1 - \frac{2}{N^2}\right)^t, \tag{12}$$

where n_0 is the initial number of A alleles. We see that the unfixed states are equally populated except at the boundaries, in agreement with graphical representations of the pdf displayed in, e.g., [5, 26], which show it to be essentially flat away from the boundaries after many generations.

2.3. Mutations in the Wright–Fisher and Moran models

So far we have been considering situations where the change in composition of the population is caused by pure genetic drift—randomness with no underlying deterministic behaviour. We now include the effects of mutation: an A allele may mutate with a probability of u to a B allele, and a B allele may mutate with a probability of v to an A allele. These are probabilities per generation, so once again these quantities in the Wright–Fisher model will differ from those in the Moran model by a factor of N.

To include mutation, one chooses an allele from the current population to die, and replaces it with one of a type chosen also with a probability proportional to the current population sizes. This is as in the original models. The mutational stage is inserted at the end of the process, so the chosen successor is allowed to mutate before being introduced into the system to define the new population. In other words, when we pick an allele to be the 'parent' of a 'child' in the next generation, the offspring can mutate with the probabilities u or v. For instance, if an A allele is chosen, there is a probability of $(1-u)$ that the replacement allele is also an A and of u that it is a B. So, with mutation added,

$$p_1(n) = (1-u)\left(\frac{n}{N}\right) + v\left(1 - \frac{n}{N}\right), \tag{13}$$

and

$$p_2(n) = u\left(\frac{n}{N}\right) + (1-v)\left(1 - \frac{n}{N}\right). \tag{14}$$

The transition matrices for the Wright–Fisher and Moran models are again found from equations (1) and (9), but using p_1 and p_2 given by equations (13) and (14). So for instance, the non-zero elements of the transition matrix for the Moran model are

$$\begin{aligned} p_{n+1\,n} &= (1-u)\left(1-\frac{n}{N}\right)\left(\frac{n}{N}\right) + v\left(1-\frac{n}{N}\right)^2, \\ p_{n\,n} &= (1-u)\left(\frac{n}{N}\right)^2 + (1-v)\left(1-\frac{n}{N}\right)^2 + (u+v)\left(\frac{n}{N}\right)\left(1-\frac{n}{N}\right), \\ p_{n-1\,n} &= (1-v)\left(\frac{n}{N}\right)\left(1-\frac{n}{N}\right) + u\left(\frac{n}{N}\right)^2. \end{aligned} \tag{15}$$

Allowing for the possibility of mutation means that the expected number of A alleles is no longer conserved. For the Wright–Fisher model this can be calculated as in equation (7), but now

$$\sum_n n p_{n\,n'} = N p_1 = (1-u)n' + v(N-n'), \tag{16}$$

and so

$$\langle n(t+1)\rangle = (1-u)\langle n(t)\rangle + v\left(N - \langle n(t)\rangle\right). \tag{17}$$

For the Moran model we calculate $\sum_n n\, p_{n\,n'}$ directly from equation (15) and find that

$$\langle n(t+1)\rangle = \langle n(t)\rangle + v\left(1-\frac{\langle n(t)\rangle}{N}\right) - u\left(\frac{\langle n(t)\rangle}{N}\right), \tag{18}$$

which is of the same form as equation (17), apart from the factors of N dividing the mutation probabilities. Equations (17) and (18) give the number of A alleles in an infinite population, where no fluctuations are present. Written in terms of allele frequencies, $a_t = \langle n(t)\rangle/N$ and $b_t = 1 - a_t$, equation (17) is nothing else but the deterministic equation for mutational change, $a_{t+1} = (1-u)a_t + vb_t$, found in many textbooks on population genetics [25]–[27].

This difference equation for a_t may be easily solved by writing $a_t = v/(u+v) + c_t$, so that $c_{t+1} = (1-u-v)c_t$. The equation for c_t can be solved iteratively to give $c_t = (1-u-v)^t c_0$, which yields

$$\begin{aligned} a_t &= \frac{v}{u+v} + (1-u-v)^t\left[a_0 - \frac{v}{u+v}\right] \\ &= \frac{v}{u+v}\left[1-(1-u-v)^t\right] + a_0\,(1-u-v)^t. \end{aligned} \tag{19}$$

Since $0 < 1-u-v < 1$, $a_t \to v/(u+v)$ as $t \to \infty$.

The addition of mutations also means that neither the A nor the B allele can become fixed—there is always the possibility of an allele being created by mutation, even if there are currently none present in the population. Thus the behaviour of both the Wright–Fisher and Moran models with mutation, after a large number of generations, is more complicated to obtain. Nevertheless, the eigenvalues of the transition matrices are known. For the Wright–Fisher model [29]:

$$\lambda_r = (1-u-v)^r \binom{N}{r}\frac{r!}{N^r}, \qquad r = 0, \ldots, N, \tag{20}$$

and for the Moran model [27]

$$\lambda_r = 1 - (u+v)\frac{r}{N} - (1-u-v)\,\frac{r(r-1)}{N^2}, \qquad r = 0, \ldots, N. \tag{21}$$

The Moran model with mutation can be solved analytically [33, 34]. In addition, approximate methods exist, such as the diffusion approximation discussed in section 3, which describe these models well. Finally, all of the models can be generalized to the cases with two sexes and diploid individuals [32]. These models are more complicated, but do not in our view make the underlying ideas or the connection between different approaches any clearer. For all these developments we refer the reader to the literature on the subject [32, 26, 28, 27]; meanwhile, in section 5 we shall examine how sexual reproduction and diploidy are handled in models that are more easily studied analytically.

2.4. Migration between islands

In the models of genetic drift considered so far, new individuals were produced from their 'parents' randomly. If the organisms had been diploid, we would have said that the mating was random. This randomness assumption may model some situations, for example, if sperm is released into a pond or lake by a group of male fish in order to fertilize eggs laid by females nearby. However, there are many instances where mating is not random. One of the most interesting is when groups of individuals are geographically separated. Then, if there is little migration between the populations, so that they are effectively isolated, stochastic effects mean that they will develop in different ways. For instance, some alleles may be lost in some of the populations, but not in others.

Models which describe the population genetics of subdivided populations which are partly isolated from each other are called 'island models' and were first investigated by Wright [5]. The island populations themselves are referred to as *demes*. In early models, migration between demes was essentially independent of the distance between the islands which contained them, that is, each deme interchanged genes equally with every other deme. More elaborate models soon followed, such as stepping stone models, where the rate of migration between demes is a function of the distance between the islands [35]–[37]. These models can be generalized by defining a migration matrix, g_{ij}, which specifies the probability that, if an individual in deme j has had an offspring, the latter immediately migrates to deme i (so that it has arrived at the start of the next generation).

The inclusion of migration into the models so far considered in this section is in principle straightforward, but in some cases the matrix of transition probabilities becomes cumbersome to write down. There are also potentially several different models which can be constructed depending on the order in which the various processes of birth, mutation and migration take place, how many migrants are permitted into and out of each deme, and so on. The introduction of migration into Moran-type models is, as expected, easier to formulate and seems more natural to statistical physicists. In the usual spirit of this approach, migrations will be considered as single events, for instance, a migration of a gene from deme j to deme i. This is in contrast to many other approaches to stochastic models of migration, including one introduced by Moran [38], where a fixed number of individuals are chosen to migrate from one deme to another. We shall consider the class of models for which migration can be characterized through the matrix g_{ij} defined above: this class includes the simple cases considered by early workers, such as uniform migration

rate between all demes or stepping stone models. The number of demes will be taken to be L and we will assume that they will each consist of N genes. For simplicity we will assume that the mutation rates will be the same on each island. In other words, each island will be an exact copy of the systems considered previously in this section, but coupled together by migration processes.

To illustrate the basic idea, suppose that there are only two islands, not L as assumed above. Since each island contains exactly N genes, only two integers are required to specify the state of the system: n_1 and n_2, where n_i is the number of A alleles in deme i, where $i = 1, 2$. We begin with the basic Moran model of section 2.2 where only the birth/death process is considered. When migration is added, a 'death' is not necessarily followed by a 'birth'—it can be followed by a migration event. The algorithm we use is as follows. First, we randomly pick a 'parent', or gene to be copied, from deme i with probability f_i. Suppose this parent is on island 2. Then with probability g_{12} a migration event takes place, which entails a death on island 1, and the copied gene from island 2 occupying the space that was left behind. With probability $g_{22} = 1 - g_{12}$ we assume that no migration event takes place and a death takes place on island 2, with the copied gene from island 2 replacing it. Note that the total probability per generation that an individual migrates from deme i to deme j per generation is $g_{ji}f_i$.

Since N is fixed for each deme, if B increases by 1 on island i, this is equivalent to A decreasing by 1 on island i. So the four independent transition probabilities are given by

$$\begin{aligned}
p_{(n_1+1,n_2)\,(n_1,n_2)} &= f_1\,(1-g_{21})\left(1-\frac{n_1}{N}\right)\frac{n_1}{N} + f_2 g_{12}\left(1-\frac{n_1}{N}\right)\frac{n_2}{N},\\
p_{(n_1-1,n_2)\,(n_1,n_2)} &= f_1\,(1-g_{21})\left(1-\frac{n_1}{N}\right)\frac{n_1}{N} + f_2 g_{12}\frac{n_1}{N}\left(1-\frac{n_2}{N}\right),\\
p_{(n_1,n_2+1)\,(n_1,n_2)} &= f_2\,(1-g_{12})\left(1-\frac{n_2}{N}\right)\frac{n_2}{N} + f_1 g_{21}\left(1-\frac{n_2}{N}\right)\frac{n_1}{N},\\
p_{(n_1,n_2-1)\,(n_1,n_2)} &= f_2\,(1-g_{12})\left(1-\frac{n_2}{N}\right)\frac{n_2}{N} + f_1 g_{21}\frac{n_2}{N}\left(1-\frac{n_1}{N}\right),
\end{aligned} \tag{22}$$

with

$$p_{(n_1,n_2)\,(n_1,n_2)} = 1 - p_{(n_1+1,n_2)\,(n_1,n_2)} - p_{(n_1-1,n_2)\,(n_1,n_2)} - p_{(n_1,n_2+1)\,(n_1,n_2)} - p_{(n_1,n_2-1)\,(n_1,n_2)}. \tag{23}$$

We can obtain the deterministic equation, valid in the limit $N \to \infty$, describing the effect of migration on the number of A alleles on one island, by using the same method as was used to find equation (18), describing the effects of mutation. In this case we calculate $\sum_{n_1,n_2} n_1\, p_{(n_1,n_2)\,(n_1',n_2')}$ from equations (22) and (23) to find that

$$\langle n_1(t+1)\rangle = \langle n_1(t)\rangle + \frac{f_2 g_{12}}{N}\left[\langle n_2(t)\rangle - \langle n_1(t)\rangle\right], \tag{24}$$

which is of the same form as equation (18) for mutations. The corresponding expression for the expected number of A alleles on island 2, $\langle n_2(t+1)\rangle$ is obtained by exchanging all 1 and 2 indices. Using these expressions one can show that the weighted mean

$$\bar{n}(t) \equiv f_1 g_{21}\langle n_1(t)\rangle + f_2 g_{12}\langle n_2(t)\rangle \tag{25}$$

is conserved, i.e., $\bar{n}(t+1) = \bar{n}(t)$, whilst the difference

$$\delta n(t) \equiv \langle n_1(t)\rangle - \langle n_2(t)\rangle \tag{26}$$

decays geometrically since

$$\delta n(t+1) = \left[1 - \frac{f_1 g_{21} + f_2 g_{12}}{N}\right] \delta n(t). \tag{27}$$

In other words, migration results in a convergence of the mean allele frequencies on the two islands to a common value which in turn eventually reaches a stationary value n^* that can be worked out from equation (25) once the initial values $n_1(0)$ and $n_2(0)$ are known:

$$n^* = \frac{f_1 g_{21} n_1(0) + f_2 g_{12} n_2(0)}{f_1 g_{21} + f_2 g_{12}}. \tag{28}$$

2.5. Mapping between population genetic, ecological and linguistic models

In section 1 we have already discussed the analogies that can be drawn between models of population genetics, individual based models (IBMs) in ecology and models of the evolution of language. The mathematical formalism we have introduced in the current section allows us to make these analogies more concrete, and to define mappings between models in these three distinct areas.

The genetic models which we have described have been of a gene at a single position (locus) on a chromosome, involved only one chromosome, and has not included sex: new genes were 'born' from a single parent. In this very simplified genetics, we can directly identify a gene with an individual and vice versa. In this case different alleles can be thought of as different species. Therefore in the context in which we have been working, the mappings (genes $\leftrightarrow$ individuals) and (alleles $\leftrightarrow$ species) is very natural. The idea of an island is common to both population genetics and population ecology, where it is called a patch, and we have the further correspondence (deme $\leftrightarrow$ local community).

To extend these mappings to the evolutionary model of language change [15, 16, 18, 19] described briefly in section 1, we begin by noting that the central idea of this approach is to make an analogy between 'different ways of saying the same thing'—called *variants* of a word or set of words, and alleles. The specific 'word or set of words' under consideration is called the *lingueme*. Thus the first analogy is (variants $\leftrightarrow$ alleles). In the model [20], each speaker's utterances may contain different variants of a lingueme. In the case of two variants, the speaker's grammar can be modelled as being made up of N *tokens*, of which n are of one variant and $N-n$ of the other. An utterance by speaker i will then produce tokens which modify that speaker's own grammar (death and birth of tokens belonging to speaker i) and which may modify the grammar of speaker j (migration of tokens from speaker i to speaker j). This gives the further analogies (tokens $\leftrightarrow$ genes) and (speakers $\leftrightarrow$ islands). Table 1 summarizes the mappings between these different areas.

While we have so far assumed that there are only two alleles, A and B, in the case of population genetics, the analogies with population ecology and evolutionary linguistics make it clear that we will frequently be interested in having more than two species or variants in our models. Therefore in the next section will generalize the models so that there are M possible types of gene/individual/token.

The interpretation of 'genetic drift', mutation and migration in population ecology and evolving linguistics is interesting. The species being considered in the analogy are not to be thought of as predators and prey, but rather as species of a similar type ('trophically similar species') competing for a common resource, such as space or light. The fact

J. Stat. Mech. (2007) P07018

Table 1. The correspondences between constituents, concepts and processes in population genetics, population ecology and language evolution.

Population genetics	Population ecology	Language evolution
Gene	Individual	Token
Allele	Species	Variant
Island	Patch	Speaker
Deme	Local community	Grammar
Population	Metacommunity	Speech community
Mutation	Speciation	Innovation
Migration	Immigration	Conversation

that when a death of an individual occurs, it is immediately replaced by another (either through a birth or by immigration) means that there is intense competition; as soon as a 'vacancy' occurs, it is immediately filled. This coupling of the death rate to the birth/immigration rate so that the overall population remains fixed, is called the 'zero-sum rule' in biology. Models where trophically similar species (for example, forests of trees) change their population sizes through purely random dynamics analogous to pure genetic drift, are called neutral models [10]. They are, at present, quite controversial [39], mainly because measurable 'physical' quantities such as the species abundance distribution arise as a result of random effects. In the context of neutral theories a mutation corresponds to speciation, where a single individual of a species changes into an individual of a different species. This is, of course, a very coarse-grained description of a very complex process. Immigration has the same interpretation as migration in the genetic case.

In the linguistics application, we think of different linguistic variants competing for use by members of the speech community. Here, the analogue of 'trophically similar species' would be functionally equivalent variants, or 'two ways of saying the same thing'. For true neutrality, one would further require that no speakers ascribe any social value ('prestige') to a variant, which might cause them to reproduce it more (or, indeed, less) often they have heard it used thus far. There are many ways in which new variants might be created, a process referred to as *innovation* [17]. Basically, most of these arise from speakers needing to communicate something that they have not communicated before; other forces behind innovation include articulatory constraints which may involve speakers producing one variant, even if they aim to produce another. This latter case, in particular, could be modelled by mutation rates between variants, as we have described for the genetics application. Migration corresponds to the process whereby tokens from the grammar of one speaker are incorporated into the grammar of another speaker with whom the first is conversing. Two factors contribute to the overall rate at which tokens migrate from one grammar to another: first, there is the frequency with which two speakers converse; secondly, speakers may give higher or lower weights to the utterances of other speakers based, for example, on the listener's perception of how similar the interlocutor is to themselves, or whether they wish to acquire social status by imitating variants associated with a prestigious group. Such sociolinguistic effects are discussed in more detail in [17]. Meanwhile, it has been proposed [23] that when dialects are brought into contact, as happened for example when people from Britain and Ireland migrated to New Zealand, a truly neutral theory (i.e., one in which these social factors are claimed

to be unimportant) should describe observed effects, such as convergence to a relatively homogeneous dialect among a large community of speakers on a timescale of order of a human lifetime. Whether a genetic drift-like process actually allows for such behaviour is another matter of debate and is under current investigation.

In this section we have introduced the simplest stochastic models of population genetics and brought out analogies with similar models in population ecology and evolutionary linguistics. In the next section we will write these models in a form more familiar to statistical physicists, generalize them, and also show that when the number of genes/individuals/tokens for an island/patch/speaker is large, they may be replaced by models based on Fokker–Planck equations, rather than on master equations.

3. Master and Fokker–Planck equations

As we have seen in section 2, the Wright–Fisher model is quite difficult to analyse, especially when mutation and migration are allowed for, and so very early on workers resorted to the *diffusion approximation* [40]–[42]. This is a description of the process, valid when N is large, where (a) the (rational) allele frequencies n/N are replaced by real numbers x, $0 \leq x \leq 1$, and (b) the generational label, t, is rescaled by a factor of $1/N$. Specifically, t now measures units of N generations, so that one generation becomes only $1/N$ units of t. The governing equation is now a Fokker–Planck equation for $P(x, \tau'|x_0, 0)$ which for pure genetic drift has the form

$$\frac{\partial P}{\partial \tau'} = \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[x(1-x)P\right], \tag{29}$$

where $\tau' = t/N$. It is possible to add terms to equation (29) which account for mutations and migrations. In essence, this is a mesoscopic description, whereas the Wright–Fisher equation was a gene based description. In general we will adopt the terminology [30] that a Fokker–Planck equation provides a mesoscopic description of the system; macroscopic equations are not stochastic and do not include the effect of fluctuations. The equation (29) is referred to as a non-linear Fokker–Planck equation because the diffusion coefficient depends on x.

In the last section we also described how the Moran model, while similar to the Wright–Fisher model, was in many ways easier to write down and generalize. Therefore in this section we will formulate the basic stochastic processes that interest us in terms of Moran-type models, rather than Wright–Fisher models. An advantage is that Moran-type models may be formulated as master equations, a step that is not typically followed in the mathematical biology literature. The advantage of formulating a stochastic process in this way, is that there then exist systematic methods to approximate the master equation for large N as a Fokker–Planck equation [30]. It will turn out that the Fokker–Planck equation corresponding to the large N limit of the Moran model has the same form as equation (29), that is, has the same limit as the Wright–Fisher model.

We will begin by discussing the reformulation of the Markov chain based description of population genetics given in section 2 to the master equation description which we will use in the rest of the section.

J. Stat. Mech. (2007) P07018

3.1. Master equations and population genetics

We begin by rewriting the standard dynamical relation for a Markov chain, $P(n,t+1) = \sum_{n'} p_{n\,n'} P(n',t)$ as

$$P(n,t+1) - P(n,t) = \sum_{n'} p_{n\,n'} P(n',t) - \sum_{n'} p_{n'\,n} P(n,t), \tag{30}$$

using $\sum_{n'} p_{n'\,n} = 1$. Now the terms in the sums with $n' = n$ may be omitted—since they cancel between the two terms on the right-hand side—and so only the terms involving an *actual transition*, that is, terms with $n' \neq n$, appear in the equation.

Up until this point, we have been careful not to refer to t as a 'time'—it has simply been an integer labelling generations. However now we interpret t as time divided into intervals Δt taken to be sufficiently small that at most one sampling event occurs during it. Dividing equation (30) by Δt, and omitting the terms with $n' = n$, yields

$$\frac{P(n,t+\Delta t) - P(n,t)}{\Delta t} = \sum_{n' \neq n} \left(\frac{p_{n\,n'}}{\Delta t}\right) P(n',t) - \sum_{n' \neq n} \left(\frac{p_{n'\,n}}{\Delta t}\right) P(n,t). \tag{31}$$

We now adopt a different view of the stochastic process. Instead of assuming that exactly one sampling event happens per generation (including $n \to n$, where no transition actually occurs), we now have sampling events occur at unit rate so that *on average* one event takes place per generation. Once t is sufficiently large, by far the most likely number of sampling events that will have taken place will be equal to t. Therefore one expects the properties of the continuous and discrete time processes to concur at large times. Additionally, we replace the transition probability from state n' to n per generation $p_{n\,n'}$ with the transition rate per unit time $T(n|n')$.

These two quantities are related via the expression

$$p_{n\,n'} = T(n|n')\Delta t + \mathcal{O}(\Delta t)^2, \qquad n \neq n' \tag{32}$$

where the terms of order $(\Delta t)^2$ and higher represent the probability of two or more events happening in time Δt. Substituting (32) into equation (31), and letting $\Delta t \to 0$, gives

$$\frac{\mathrm{d}P(n,t)}{\mathrm{d}t} = \sum_{n' \neq n} T(n|n')\, P(n',t) - \sum_{n' \neq n} T(n'|n)\, P(n,t), \tag{33}$$

which is the usual master equation used by statistical physicists [30, 43, 44]. Typically, the transition rates $T(n|n')$ are assumed given (this specifies the model) and we wish to determine the $P(n,t)$. The initial condition is that the system is in state n_0 at $t = 0$. Boundary conditions may also have to be given.

To illustrate these ideas we will look at three examples, all of which were introduced in section 2. We begin with the Moran model for pure genetic drift, which has non-zero elements of the transition matrix given by equation (10). In the continuous time formulation, the population evolves by two individuals being sampled uniformly with replacement from the population as a Poisson process with the mean time between events taken to be unity. One of the sampled individuals is designated the parent which is copied to create an offspring; the other sampled individual is sacrificed to make way for the new offspring. To obtain the transition rates in the master equation (33), we need not

consider the matrix element $p_{n\,n}$, since this does not involve an actual transition. We find therefore

$$T(n+1|n) = \left(1-\frac{n}{N}\right)\left(\frac{n}{N}\right), \qquad T(n-1|n) = \left(\frac{n}{N}\right)\left(1-\frac{n}{N}\right). \tag{34}$$

We now turn to the Moran model with mutation, which has non-zero elements given by equation (15). We will contrast two ways one might construct a continuous time process to mirror the discrete time process as defined in section 2.3. The most obvious is to follow the prescription given earlier: that is, to have Poisson sampling events as described in the previous paragraph, and to mutate the copy of the parent to allele B with probability u if the chosen parent was an A or with probability v in the opposite direction. The transition rates $T(n+1|n)$ and $T(n-1|n)$ are then given by the expressions for $p_{n+1\,n}$ and $p_{n-1\,n}$ appearing in equation (15), i.e.,

$$\begin{aligned} T(n+1|n) &= (1-u)\left(1-\frac{n}{N}\right)\left(\frac{n}{N}\right) + v\left(1-\frac{n}{N}\right)^2, \\ T(n-1|n) &= (1-v)\left(\frac{n}{N}\right)\left(1-\frac{n}{N}\right) + u\left(\frac{n}{N}\right)^2. \end{aligned} \tag{35}$$

The second approach we shall take differs in that the mutation process is divorced from the death/birth process—the two are considered to take place independently with mutations occurring spontaneously, rather than between a death and a birth. Specifically, for a fraction γ of the time we sample two alleles, as previously described, and we replace one with a copy of the other without mutation. The rest of the time, i.e., with probability $1-\gamma$, we sample only one allele, mutate it and replace it in the population. If it is in state A we mutate it to state B with probability U, and if it is in state B, we mutate it to state A with probability V. This leads to transition rates of the form

$$T(n+1|n) = \gamma\left(1-\frac{n}{N}\right)\left(\frac{n}{N}\right) + (1-\gamma)V\left(1-\frac{n}{N}\right) \tag{36}$$

$$T(n-1|n) = \gamma\left(\frac{n}{N}\right)\left(1-\frac{n}{N}\right) + (1-\gamma)U\left(\frac{n}{N}\right). \tag{37}$$

This is not obviously a Moran-type process. To make contact with the discrete time Moran process, we shall insist that, at any given time, the probability for the *next* event to occur in the continuous time process is equal to that for the discrete time process, given the same starting configuration. To do this, we rewrite $p_{n+1\,n}$ and $p_{n-1\,n}$ in equation (15) as

$$\begin{aligned} p_{n+1\,n} &= (1-u-v)\left(1-\frac{n}{N}\right)\left(\frac{n}{N}\right) + v\left(1-\frac{n}{N}\right) \\ p_{n-1\,n} &= (1-u-v)\left(1-\frac{n}{N}\right)\left(\frac{n}{N}\right) + u\left(\frac{n}{N}\right). \end{aligned} \tag{38}$$

Correspondence is obtained if we put

$$\gamma = 1-u-v, \qquad V = \frac{v}{u+v} \qquad \text{and} \qquad U = \frac{u}{u+v}, \tag{39}$$

that is, if a fraction u of the time, A alleles spontaneously change state to B alleles, and a fraction v of the time, B alleles change to A. Note this leads only to positive transition rates if the mutation probabilities satisfy $u+v<1$. Therefore, not all Moran processes in which birth, death and mutation are intertwined can be represented by a model in which

mutation occurs independently of birth and death. Nevertheless, a model with transition rates of this form was introduced by Maruyama (model I in the appendix of [28]) and it has been remarked [45] that the results obtained by this model and the Moran model seem to be essentially identical. Our derivation of the former starting from the latter shows why this is the case.

We now turn to the Moran model with migration between two islands, discussed in section 2.4. In this case the state of the system is specified by two integers: $\underline{n} = (n_1, n_2)$. The considerations leading to the master equation (33), where the state is specified by a single variable, generalize in a straightforward way to the case where the state is specified by a set of integers:

$$\frac{\mathrm{d}P(\underline{n},t)}{\mathrm{d}t} = \sum_{\underline{n}'\neq\underline{n}} T(\underline{n}|\underline{n}')\,P(\underline{n}',t) - \sum_{\underline{n}'\neq\underline{n}} T(\underline{n}'|\underline{n})\,P(\underline{n},t), \tag{40}$$

where in the case under consideration here $\underline{n} = (n_1, n_2)$. The transition probabilities are given by equation (22). To construct the corresponding continuous time process, we take the rate at which a parent is chosen in deme j, the offspring of which displaces an individual in deme i, to be

$$G_{ij} = f_j g_{ij}, \tag{41}$$

where we recall that f_j is the probability the parent is in deme j, and g_{ij} the probability that the offspring lands in deme i, conditioned on the parent being in deme j. Since $\sum_j f_j = 1$ and $\sum_i g_{ij} = 1$ for any j, the total rate at which pairs of individuals are sampled is

$$\sum_{i,j} G_{ij} = \sum_j f_j \sum_i g_{ij} = 1 \tag{42}$$

as required, even though the sampling is now non-uniform.

The transition probabilities for this model then follow as

$$\begin{aligned}
T(n_1+1,n_2|n_1,n_2) &= G_{11}\left(1-\frac{n_1}{N}\right)\frac{n_1}{N} + G_{12}\left(1-\frac{n_1}{N}\right)\frac{n_2}{N},\\
T(n_1-1,n_2|n_1,n_2) &= G_{11}\left(1-\frac{n_1}{N}\right)\frac{n_1}{N} + G_{12}\frac{n_1}{N}\left(1-\frac{n_2}{N}\right),\\
T(n_1,n_2+1|n_1,n_2) &= G_{22}\left(1-\frac{n_2}{N}\right)\frac{n_2}{N} + G_{21}\left(1-\frac{n_2}{N}\right)\frac{n_1}{N},\\
T(n_1,n_2-1|n_1,n_2) &= G_{22}\left(1-\frac{n_2}{N}\right)\frac{n_2}{N} + G_{21}\frac{n_2}{N}\left(1-\frac{n_1}{N}\right).
\end{aligned} \tag{43}$$

Note that the diagonal matrix elements G_{ii} give the rates at which an individuals are chosen to reproduce without their offspring subsequently migrating. From the master equation (40) and the transition rates (43), we can obtain the deterministic equation, valid in the limit $N \to \infty$, describing the effect of migration on the number of A alleles on one island, by calculating $\mathrm{d}\langle n_i\rangle/\mathrm{d}t$. Focusing on island 1, we multiply equation (40) by n_1 and sum over all n_1 and n_2. By shifting the quantities being summed over by ± 1 in some of the sums over n_1, one finds that

$$\frac{\mathrm{d}\langle n_1\rangle}{\mathrm{d}t} = \langle T(n_1+1,n_2|n_1,n_2)\rangle - \langle T(n_1-1,n_2|n_1,n_2)\rangle. \tag{44}$$

Substituting for the transitions rates using equation (43) yields

$$\frac{\mathrm{d}\langle n_1\rangle}{\mathrm{d}t} = G_{12}\left(\frac{\langle n_2\rangle}{N} - \frac{\langle n_1\rangle}{N}\right), \tag{45}$$

which is in agreement with equation (24) as one would expect.

3.2. Fokker–Planck equations and the large N limit

Having obtained the transition rates (34), (37) and (43) for pure genetic drift involving two alleles, together with mutation and migration between two islands, we can now investigate the form that the governing equations take in the limit where N is large. It will turn out that these processes can be described by Fokker–Planck equations in this limit. This will allow more complex situations, for example involving L islands or M alleles, to be formulated in a rather straightforward way.

We again begin with the Moran model for pure genetic drift. When formulated as a master equation, it has the form (33) with transition rates (34). Such an equation is of 'the diffusion type', since there is no macroscopic equation and the large N approximation results in a non-linear Fokker–Planck equation [30]. In this case, the large N limit can be obtained in a straightforward fashion by substituting $n = xN$, and expanding out any remaining terms in N in a $1/N$ expansion. We find

$$\begin{aligned}\frac{\partial P}{\partial t} = &\left(x - \frac{1}{N}\right)\left(1 - x + \frac{1}{N}\right)\left\{P(x,t) - \frac{1}{N}\frac{\partial P(x,t)}{\partial x} + \frac{1}{2N^2}\frac{\partial^2 P(x,t)}{\partial x^2}\right\}\\ &+ \left(x + \frac{1}{N}\right)\left(1 - x - \frac{1}{N}\right)\left\{P(x,t) + \frac{1}{N}\frac{\partial P(x,t)}{\partial x} + \frac{1}{2N^2}\frac{\partial^2 P(x,t)}{\partial x^2}\right\}\\ &- 2x(1-x)P(x,t) + \mathcal{O}(1/N^3).\end{aligned} \tag{46}$$

After a little algebra, this reduces to

$$\frac{\partial P}{\partial t} = \frac{1}{N^2}\frac{\partial^2}{\partial x^2}\left[x(1-x)P\right] + \mathcal{O}(1/N^3). \tag{47}$$

So if we define a scaled time $\tau = 2t/N^2$, then taking $N \to \infty$ we obtain

$$\frac{\partial P}{\partial \tau} = \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[x(1-x)P\right], \tag{48}$$

which is the diffusion equation (29). This shows that two different gene based models—the Wright–Fisher and Moran models—give the same mesoscopic Fokker–Planck equation. This is a familiar situation in statistical physics: microscopic models describing the same 'physics', but formulated in slightly different ways, typically will give rise to the same mesoscopic equation. The only difference between the Wright–Fisher and the Moran model is that, as usual, the timescales differ. The redefinition of a generation in the Wright–Fisher model leads to $\tau' = t/N$, whereas the scaling of time in the Moran model gives $\tau = 2t/N^2$. In other words, a generation in the Wright–Fisher model is $N/2$ times as long as in the Moran model in this mesoscopic description.

When the diffusion approximation is applied to the Moran model which includes mutational and migrational effects, the rates of migration and mutation have to be rescaled by a factor of N. If this were not so, the large N limit of the master equation

would not be a Fokker–Planck equation of the diffusion type, but a linear Fokker–Planck equation which would describe fluctuations about the deterministic equation (17) [30, 34]. In the backward-time formulation discussed later (see section 4), we will also see that this scaling corresponds to a regime in which all relevant processes affecting the ancestry of the population occur on the same timescale.

Let us first consider the models with mutation. The scaled mutation rates, with the factor of N mentioned above and a factor of 2 to be explained below, are defined by

$$\mathcal{U} = \frac{uN}{2}, \qquad \mathcal{V} = \frac{vN}{2}. \tag{49}$$

In section 3.1 we discussed two ways of introducing mutation into the Moran model. The first resulting in transition rates given by equation (35) and the second in transition rates given by equation (37), with γ, U and V given by equation (39). There is no need to carry out the $1/N$ expansion for each model in turn, since the transitions rates for both are identical (although one has the restriction $u + v < 1$), and which using the scaled forms (49) read

$$\begin{aligned} T(n+1|n) &= \left(1 - \frac{2\mathcal{U}}{N} - \frac{2\mathcal{V}}{N}\right)\left(1 - \frac{n}{N}\right)\left(\frac{n}{N}\right) + \frac{2\mathcal{V}}{N}\left(1 - \frac{n}{N}\right) \\ T(n-1|n) &= \left(1 - \frac{2\mathcal{U}}{N} - \frac{2\mathcal{V}}{N}\right)\left(\frac{n}{N}\right)\left(1 - \frac{n}{N}\right) + \frac{2\mathcal{U}}{N}\left(\frac{n}{N}\right). \end{aligned} \tag{50}$$

Since the term (47) is already of order $1/N^2$, the terms involving $\mathcal{U}/N$ and $\mathcal{V}/N$ which multiply it give terms of order $1/N^3$, which do not contribute to the Fokker–Planck equation. Therefore the only additional terms due to mutation are the second terms on the right-hand side of the transitions probabilities (50), which are

$$\begin{aligned} &\left\{\frac{2\mathcal{V}}{N}\left(1 - x + \frac{1}{N}\right)\right\}\left\{P(x,t) - \frac{1}{N}\frac{\partial P(x,t)}{\partial x}\right\} + \left\{\frac{2\mathcal{U}}{N}\left(x + \frac{1}{N}\right)\right\} \\ &\qquad \times \left\{P(x,t) + \frac{1}{N}\frac{\partial P(x,t)}{\partial x}\right\} - \left[\frac{2\mathcal{V}}{N}(1-x) + \frac{2\mathcal{U}}{N}x\right] P(x,t) + \mathcal{O}(1/N^3), \end{aligned} \tag{51}$$

which on simplification yields

$$\frac{2}{N^2}\frac{\partial}{\partial x}\left[\{\mathcal{U}x - \mathcal{V}(1-x)\}\, P\right] + \mathcal{O}(1/N^3). \tag{52}$$

Adding this additional term to equation (47), and using rescaled time units $\tau = 2t/N^2$, we find on letting $N \to \infty$ that the additional mutation processes present in either model give rise to the Fokker–Planck equation

$$\frac{\partial P}{\partial \tau} = \frac{\partial}{\partial x}\left[\{\mathcal{U}x - \mathcal{V}(1-x)\}\, P\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[x(1-x)P\right]. \tag{53}$$

The restriction $u+v < 1$, which applied to one of the models, takes the form $\mathcal{U}+\mathcal{V} < N/2$, when expressed in terms of the scaled mutation rates. Since in most applications $\mathcal{U}$ and $\mathcal{V}$ are numerically small, and in any case we are taking the limit $N \to \infty$, this condition will always hold. Furthermore, the Fokker–Planck equation (53) agrees with that found by applying the diffusion approximation to the Wright–Fisher model with mutation [26]. The additional factors of 2 introduced in equation (49) were included to conform with the conventions used in the Wright–Fisher case.

doi:10.1088/1742-5468/2007/07/P07018

J. Stat. Mech. (2007) P07018

The introduction of two islands or speakers has been discussed in section 2.4 and the transition rates given in equation (43). To derive the Fokker–Planck equation from the master equation in this case, we note that the first terms on the right-hand sides of these transition rates are exactly as for pure drift, up to the inclusion of pre-factors G_{11} or G_{22}. Therefore, the contribution from events which do not involve actual migration is

$$\sum_{i=1}^{2} \frac{G_{ii}}{N^2} \frac{\partial^2}{\partial x_i^2} \left[x_i(1-x_i)P\right] + \mathcal{O}(1/N^3). \tag{54}$$

The second terms on the right-hand side of the transition rates in equation (43), for migration from island j to island i are:

$$\begin{aligned} &+G_{ij}\left(1-x_i+\frac{1}{N}\right)x_j\left\{P(\underline{x},t)-\frac{1}{N}\frac{\partial P(\underline{x},t)}{\partial x_i}\right\} \\ &\quad + G_{ij}\left(x_i+\frac{1}{N}\right)(1-x_j)\left\{P(\underline{x},t)+\frac{1}{N}\frac{\partial P(\underline{x},t)}{\partial x_i}\right\} \\ &\quad - G_{ij}\left[(1-x_i)\,x_j + x_i\,(1-x_j)\right]P(\underline{x},t) + \mathcal{O}\left(\frac{1}{N^2}\right) \\ &= G_{ij}\frac{1}{N}\frac{\partial}{\partial x_i}\left[(x_i-x_j)\,P\right] + \mathcal{O}\left(\frac{1}{N^2}\right). \end{aligned} \tag{55}$$

In order for the Fokker–Planck equation to be of the diffusion type, the rates of migration between demes, i.e., G_{ij} for $i \neq j$, have to be scaled by a factor of N, just as the mutation rates were. Let us now parametrize these off-diagonal matrix elements as $G_{ij} = 2\mathcal{G}_{ij}/N$ where $\mathcal{G}_{ij}$ is of order unity. Then, $G_{ii} = f_i + \mathcal{O}(1/N)$, and so the Fokker–Planck equation becomes

$$\begin{aligned} \frac{\partial P(\underline{x},t)}{\partial t} &= \frac{1}{N^2}\sum_{i=1}^{2} f_i \frac{\partial^2}{\partial x_i^2}\left[x_i(1-x_i)P\right] \\ &\quad + \frac{2}{N^2}\left(\mathcal{G}_{12}\frac{\partial}{\partial x_1} - \mathcal{G}_{21}\frac{\partial}{\partial x_2}\right)\left[(x_1-x_2)\,P\right] + \mathcal{O}\left(\frac{1}{N^3}\right). \end{aligned} \tag{56}$$

Introducing $\tau = 2t/N^2$, as before, we can let $N \to \infty$ and obtain the Fokker–Planck equation for migration between two islands:

$$\frac{\partial P}{\partial \tau} = \frac{1}{2}\sum_{i=1}^{2} f_i \frac{\partial^2}{\partial x_i^2}\left[x_i(1-x_i)P\right] + \left(\mathcal{G}_{12}\frac{\partial}{\partial x_1} - \mathcal{G}_{21}\frac{\partial}{\partial x_2}\right)\left[(x_1-x_2)\,P\right]. \tag{57}$$

In this mesoscopic formulation all trace of the individual genes has disappeared, and we are left only with a description in terms of the fractions of genes on island i which are of type A.

3.3. Migration of genes having M alleles between L islands

We have so far, for simplicity, mainly restricted ourselves to considering two islands and two alleles. However the formalism we have discussed carries over with little modification to the cases of an arbitrary number of island and alleles. In order to avoid introducing too many complicating looking expressions in this section, we have relegated some of the details to an appendix.

J. Stat. Mech. (2007) P07018

3.3.1. L islands. We assume that on each island we have a deme labelled by i, with n_i being the number of A alleles and $(N - n_i)$ the number of B alleles. A parent is chosen from island j with probability f_j and subsequently leaves an offspring on island i with probability g_{ij}, where the sum rule $g_{ii} = 1 - \sum_{j\neq i} g_{ji}$ must be satisfied. Then proceeding as in section 2.4, the transition probabilities analogous to those in equation (22) are:

$$p_{(n_1,\ldots,n_i+1,\ldots,n_L)\,(n_i,\ldots,n_L)} = f_i\left(1-\sum_{j\neq i} g_{ji}\right)\left(1-\frac{n_i}{N}\right)\frac{n_i}{N} + \sum_{j\neq i} f_j g_{ij}\left(1-\frac{n_i}{N}\right)\frac{n_j}{N}, \tag{58}$$

and

$$p_{(n_1,\ldots,n_i-1,\ldots,n_L)\,(n_i,\ldots,n_L)} = f_i\left(1-\sum_{j\neq i} g_{ji}\right)\left(1-\frac{n_i}{N}\right)\frac{n_i}{N} + \sum_{j\neq i} f_j g_{ij}\frac{n_i}{N}\left(1-\frac{n_j}{N}\right). \tag{59}$$

To go over to a master equation description, we introduce the set of migration rates via equation (41). The transition rates for the master equation (40) now follow:

$$\begin{aligned} T(n_1,\ldots,n_i+1,\ldots,n_L|n_1,\ldots,n_L) &= \sum_j G_{ij}\left(1-\frac{n_i}{N}\right)\frac{n_j}{N},\\ T(n_1,\ldots,n_i-1,\ldots,n_L|n_1,\ldots,n_L) &= \sum_j G_{ij}\frac{n_i}{N}\left(1-\frac{n_j}{N}\right). \end{aligned} \tag{60}$$

To obtain the Fokker–Planck equation, valid when N is large, is now very simple, since the required calculations are as in the $L = 2$ case. For example, equation (54) is unchanged, except that the sum is now from $i = 1$ to L, and the contribution (55) holds for all pairs i and j with $i \neq j$. Introducing as before $G_{ij} = 2\mathcal{G}_{ij}/N$ for $i \neq j$, rescaling time as $\tau = 2t/N^2$, letting $N \to \infty$ and noting that $G_{ii} \to f_i$ in this limit, we obtain the Fokker–Planck equation for migration between L islands:

$$\frac{\partial P}{\partial \tau} = \frac{1}{2}\sum_{i=1}^{L} f_i \frac{\partial^2}{\partial x_i^2}\left[x_i(1-x_i)P\right] + \sum_{\langle ij\rangle}\left(\mathcal{G}_{ij}\frac{\partial}{\partial x_i} - \mathcal{G}_{ji}\frac{\partial}{\partial x_j}\right)\left[(x_i - x_j)\,P\right]. \tag{61}$$

Here the notation $\langle ij\rangle$ means 'sum over all distinct pairs i and j'.

Fokker–Planck equations may be written as continuity equations [44]; in the case of interest here it takes the form $\partial P/\partial\tau + \sum_i \partial J_i/\partial x_i = 0$, where

$$J_i(\underline{x}, t) = -\frac{1}{2} f_i \frac{\partial}{\partial x_i}\left[x_i(1-x_i)P(\underline{x}, t)\right] - \sum_{j\neq i}\mathcal{G}_{ij}\left[(x_i - x_j)\,P(\underline{x}, t)\right], \tag{62}$$

is the probability current. Multiplying the equation of continuity by x_k and integrating over all x_i $(i = 1, \ldots, L)$ gives

$$\frac{\mathrm{d}\langle x_k\rangle}{\mathrm{d}\tau} = \int \mathrm{d}\underline{x}\, x_k \frac{\partial P}{\partial \tau} = -\sum_i \int \mathrm{d}\underline{x}\, x_k \frac{\partial J_i}{\partial x_i} = \int \mathrm{d}\underline{x}\, J_k, \tag{63}$$

since the probability current vanishes on the boundaries [44]. From equations (62) and (63) we have

$$\frac{\mathrm{d}\langle x_k\rangle}{\mathrm{d}\tau} = -\sum_{j\neq k}\mathcal{G}_{kj}\left(\langle x_k\rangle - \langle x_j\rangle\right). \tag{64}$$

To make contact with previous results for finite N, we reintroduce $n_i = Nx_i$, $G_{ij} = 2\mathcal{G}_{ij}/N$ and $t = N^2\tau/2$, so that equation (64) becomes

$$\frac{\mathrm{d}\langle n_k\rangle}{\mathrm{d}t} = -\sum_{j\neq k} G_{kj}\frac{1}{N}\left(\langle n_k\rangle - \langle n_j\rangle\right). \tag{65}$$

This is in agreement with equation (45) found directly from the master equation in the case $L = 2$.

Let us briefly consider under what conditions the mean of the frequency of A alleles across all islands is conserved. For this we require

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_k \langle n_k\rangle = -\sum_k\sum_{j\neq k}\frac{G_{kj}}{N}\left(\langle n_k\rangle - \langle n_j\rangle\right) = \sum_j \frac{\langle n_j\rangle}{N}\sum_{k\neq j}\left(G_{kj} - G_{jk}\right) = 0. \tag{66}$$

For this to hold for any set of allele frequencies $\langle n_j\rangle$, we require

$$\sum_{k\neq j} G_{kj} = \sum_{k\neq j} G_{jk} \qquad \forall j. \tag{67}$$

Since the total mean A allele frequency is conserved by migration models for which the migration rates satisfy this equality, they are called *conservative*. We will see in section 5 that a number of exact results are known for models with conservative migration.

3.3.2. M alleles. So far in this article we have restricted ourselves to situations where each gene has only two alleles: A and B. There is no reason, other than grounds of simplicity, to do so, and it is possible to generalize the treatment to the situation where there are M possible types of alleles, A_α, where $\alpha = 1,\ldots,M$. If one is considering a single island, the possible states are labelled by the number of A_α alleles denoted by n_α, where $\alpha = 1,\ldots,(M-1)$. Clearly, since $\sum_{\alpha=1}^{M} n_\alpha = N$, $n_M = N - \sum_{\alpha=1}^{(M-1)} n_\alpha$ is not an independent variable. If there are L islands, then there are $L(M-1)$ independent variables: $n_{i\alpha}$ with $i = 1,\ldots,L$ and $\alpha = 1,\ldots,(M-1)$.

We begin by assuming that there is no mutation or migration (that is, $L = 1$). The basic transition probability associated with an A_β allele being replaced by an A_α allele is

$$p_{(n_1,\ldots,n_\alpha+1,\ldots,n_\beta-1,\ldots,n_{M-1})\,(\underline{n})} = \left(\frac{n_\alpha}{N}\right)\left(\frac{n_\beta}{N}\right), \tag{68}$$

where $\alpha \neq \beta$, $\alpha,\beta = 1,\ldots,(M-1)$. Clearly if the transition involves the Mth allele the expression has to be modified, since the number of alleles of type M is not an independent variable. The transition rates which appear in the master equation are obtained as before through consideration of a continuous time process in which pairs of individuals, one of which reproduces and whose offspring displaces the others, are sampled on average once per unit time. This leads to

$$T(n_1,\ldots,n_\alpha+1,\ldots,n_\beta-1,\ldots,n_{M-1}|\underline{n}) = \left(\frac{n_\alpha}{N}\right)\left(\frac{n_\beta}{N}\right), \tag{69}$$

where $\alpha \neq \beta$ and neither is equal to M, and

$$\begin{aligned} T(n_1,\ldots,n_\alpha+1,\ldots,n_{M-1}|\underline{n}) &= T(n_1,\ldots,n_\alpha-1,\ldots,n_{M-1}|\underline{n}) \\ &= \left(\frac{n_\alpha}{N}\right)\left(\frac{N-\sum_{\beta=1}^{M-1} n_\beta}{N}\right), \end{aligned} \tag{70}$$

if either allele $\alpha\,(\neq M)$ increases at the expense of allele M or allele M increases at the expense of allele $\alpha\,(\neq M)$, respectively.

We can now carry out a large N expansion on the master equation with the transition rates (69) and (70), just as we did in equation (46). The algebraic details are a little more messy for general M, and so we have given them in the appendix. Rescaling the time as before, one finds the Fokker–Planck equation

$$\frac{\partial P}{\partial \tau} = \frac{1}{2}\left\{\sum_{\alpha=1}^{M-1}\frac{\partial^2}{\partial x_\alpha^2}\left[x_\alpha\left(1-x_\alpha\right)P\right] - \sum_{\alpha=1}^{M-1}\sum_{\beta\neq\alpha}^{M-1}\frac{\partial^2}{\partial x_\alpha \partial x_\beta}\left[x_\alpha x_\beta P\right]\right\}, \tag{71}$$

in the limit $N\to\infty$. This the usual multi-allelic diffusion equation [46].

If we now add the possibility of mutations occurring, $M(M-1)$ mutation probabilities have to be defined to describe mutations from allele A_δ to allele A_α. We will denote this probability by $u_{\alpha\delta}\,(\alpha\neq\delta)$. To see how this matrix enters the formalism, let us return to the case of two alleles and rewrite equations (13) and (14) as

$$p_\alpha = \sum_{\delta=1}^{2} u_{\alpha\delta}(n_\delta/N), \tag{72}$$

where we have written A and B as A_1 and A_2 respectively, $n_2 = N - n_1$, and where the *mutation matrix* is given by

$$\mathbf{u} = \begin{bmatrix} 1-u & v \\ u & 1-v \end{bmatrix}. \tag{73}$$

Notice that the condition that the columns of the matrix $\mathbf{u}$ add up to 1 ensures that $\sum_\alpha p_\alpha = 1$, since $\sum_\delta n_\delta = N$. For the case of M different alleles $\mathbf{u}$ has entries $u_{\alpha\delta}$ if $\alpha\neq\delta$, and $1-\sum_{\gamma\neq\delta}u_{\gamma\delta}$ for the diagonal entry (δ,δ). Therefore to deduce the form of the transition probabilities in the Moran model with M alleles and mutation, we may follow the same arguments which led to equation (15) in the case of two alleles. For instance, if an A_β allele is replaced by an A_α allele, then

$$p_{(n_1,\ldots,n_\alpha+1,\ldots,n_\beta-1,\ldots,n_{M-1})\,(\underline{n})} = p_\alpha(\underline{n})\left(\frac{n_\beta}{N}\right). \tag{74}$$

This will hold for all α and β as long as we interpret n_M as $1-\sum_{\gamma=1}^{(M-1)} n_\gamma$. The change in the mean value of n_α with time can be calculated in an analogous way to equation (18) and one finds that

$$\langle n_\alpha(t+1)\rangle = \langle n_\alpha(t)\rangle + \sum_{\beta\neq\alpha}\left\{u_{\alpha\beta}\frac{\langle n_\beta(t)\rangle}{N} - u_{\beta\alpha}\frac{\langle n_\alpha(t)\rangle}{N}\right\}. \tag{75}$$

As in section 3.1, we shall follow two approaches to mutation in the continuous time formulation. First, we shall have pairs of individuals sampled on average once per unit time, and the offspring of the parent changed to allele α with probability $u_{\alpha\delta}$ given that the sampled parent carried allele δ. Transition rates for this model are then given by

$$T(n_1\ldots n_\alpha+1\ldots n_\beta-1\ldots n_{M-1}|\underline{n}) = \left(\sum_\delta u_{\alpha\delta}\frac{n_\delta}{N}\right)\frac{n_\beta}{N} \tag{76}$$

for $\alpha \neq \beta$ and it is understood that n_M is always implicitly given by $1 - \sum_{\alpha=1}^{M-1} n_\alpha$. In the appendix we derive the Fokker–Planck equation for this system, valid when N is large. Again, it is necessary to rescale the rates of mutation between different alleles, i.e.,

$$\mathcal{U}_{\alpha\delta} = \frac{N u_{\alpha\delta}}{2} \qquad \alpha \neq \delta. \tag{77}$$

The Fokker–Planck equation reads

$$\frac{\partial P}{\partial \tau} = -\sum_{\alpha=1}^{M-1} \frac{\partial}{\partial x_\alpha} \left[\mathcal{A}_\alpha(\underline{x}) P\right] + \frac{1}{2} \sum_{\alpha=1}^{M-1} \sum_{\beta=1}^{M-1} \frac{\partial^2}{\partial x_\alpha \partial x_\beta} \left[\mathcal{D}_{\alpha\beta}(\underline{x}) P\right], \tag{78}$$

where $\mathcal{A}_\alpha(\underline{x})$ and $\mathcal{D}_{\alpha\beta}(\underline{x})$ are given by

$$\mathcal{A}_\alpha(\underline{x}) = \sum_{\beta=1}^{M} \left(\mathcal{U}_{\alpha\beta} x_\beta - \mathcal{U}_{\beta\alpha} x_\alpha\right), \tag{79}$$

and

$$\mathcal{D}_{\alpha\beta}(\underline{x}) = \begin{cases} x_\alpha(1 - x_\alpha), & \text{if } \alpha = \beta \\ -x_\alpha x_\beta, & \text{if } \alpha \neq \beta. \end{cases} \tag{80}$$

Note that the term in the sum with $\alpha = \beta$ in (79) is zero, so it is not necessary to define the diagonal rescaled mutation matrix elements $\mathcal{U}_{\alpha\alpha}$.

It is worth, briefly, considering the variant of the continuous time mutation model in which mutation events are separated from the birth–death events. As in section 3.1, we shall sample alleles on average once per generation. For a fraction γ of the time, we sample a pair of alleles and replace one of them with a copy of the other. For the remaining fraction $1 - \gamma$ of the time, only one allele is chosen and this is mutated to allele α with probability $U_{\alpha\beta}$, given that it is of type β. The transition rates for this model are

$$T(n_1 \ldots n_\alpha + 1 \ldots n_\beta - 1 \ldots n_{M-1} | \underline{n}) = \gamma \left[\frac{n_\alpha}{N} \frac{n_\beta}{N}\right] + (1 - \gamma) \left[U_{\alpha\beta} \frac{n_\beta}{N}\right]. \tag{81}$$

Compare now with (76) which can be written as

$$T(n_1 \ldots n_\alpha + 1 \ldots n_\beta - 1 \ldots n_{M-1} | \underline{n}) = \left(1 - \sum_{\delta \neq \alpha} u_{\delta\alpha}\right) \frac{n_\alpha}{N} \frac{n_\beta}{N} + \sum_{\delta \neq \alpha} u_{\alpha\delta} \frac{n_\delta}{N} \frac{n_\beta}{N}. \tag{82}$$

Since no allele frequencies n_δ/N with $\delta \neq \alpha, \beta$ appear in (81), the only way the two can be made equal is if we take $u_{\alpha\delta} = u_\alpha$. Then, the sum over n_δ/N gives $1 - n_\alpha/N$, since all allele frequencies add up to unity. This results in

$$T(n_1 \ldots n_\alpha + 1 \ldots n_\beta - 1 \ldots n_{M-1} | \underline{n}) = \left(1 - \sum_{\delta} u_\delta\right) \frac{n_\alpha}{N} \frac{n_\beta}{N} + u_\alpha \frac{n_\beta}{N}. \tag{83}$$

Correspondence with (81) is then obtained as long as

$$\gamma = 1 - \sum_{\delta} u_\delta \qquad \text{and} \qquad U_\alpha = \frac{u_\alpha}{\sum_\delta u_\delta}, \tag{84}$$

where we have taken $U_{\alpha\beta} = U_\alpha$ for all $\beta \neq \alpha$. We see that, in common with the case $M = 2$ discussed in section 3.1, the sum of mutation rates $\sum_\delta u_\delta$ must be less than one in

order for the simplified model (in which mutation is a spontaneous process) to have the same dynamics as the full model in which mutation, birth and death are combined.

We remark that this condition is satisfied in the limit $N \to \infty$ when mutation rates are scaled with N as above. The Fokker–Planck equation (78) results, albeit with a simplified deterministic (potential) term

$$\mathcal{A}_\alpha(\underline{x}) = \mathcal{U}_\alpha - \sum_{\beta=1}^{M} \mathcal{U}_\beta x_\alpha, \tag{85}$$

on account of the mutation rates depending only on the end product of the mutation. Note that this time, the term $\beta = \alpha$ is included in the summation. Even with this simplification, the Fokker–Planck equation is a non-linear partial differential equation in $(M-1)$ variables, and one therefore does not expect it to be readily solved. The stationary solution has long been known to be

$$P^*(x_1, x_2, \ldots, x_M) = \Gamma\left(2\sum_{\alpha=1}^{M} \mathcal{U}_\alpha\right) \prod_{\alpha=1}^{M} \frac{x_\alpha^{2\mathcal{U}_\alpha - 1}}{\Gamma(2\mathcal{U}_\alpha)} \tag{86}$$

in which $\Gamma(u)$ is the usual Gamma function (see, e.g., [47] for a discussion of early derivations of this result). In this steady state, the probability current vanishes everywhere, i.e., detailed balance is satisfied. For a general set of mutation rates, i.e., one for which the mutation rates depend on both initial and final allele states, the stationary solution of the Fokker–Planck equation has not been found. One therefore speculates that detailed balance is not satisfied for these more general models.

Meanwhile, a remarkable feature of the Fokker–Planck equation under the restriction $\mathcal{U}_{\alpha\beta} = \mathcal{U}_\alpha$ is that the full time dependence can be found, not just the steady state. There are at least two ways to achieve this; either one can explicitly construct the set of polynomials orthogonal to the weight function given by (86) to obtain the eigenfunctions of the operator appearing in (78) [48]. Alternatively, one can make the change of variables

$$\xi_\alpha = \frac{x_\alpha}{1 - \sum_{\beta<\alpha} x_\beta}, \qquad \alpha = 1, \ldots, M-1, \tag{87}$$

under which the equation becomes separable [49]. Essentially this is possible because of the simple way that the M-allele problem is 'nested' in the $(M+1)$-allele problem. This is also responsible for most of the simplifications in the calculations of many other quantities of interest, such as mean times to fixation, the probability of a particular sequence of extinctions, and so on [49]. In this approach, it is also learnt that the eigenfunctions of the Fokker–Planck equation are Jacobi polynomials.

3.3.3. L islands and M alleles. The final case to consider is when there are M alleles and L islands. The analysis is a combination of the $M = 2$, general L case considered in section 3.3.2 and the $L = 1$, general M case just considered. So, for example, not including mutation for the moment, the terms in the transition probability which do not involve migration are exactly as in equation (69) with indices i added to the n_α or n_β, as well as an extra factor of $f_i(1 - \sum_{j\neq i} g_{ij})$. Alternatively, the transition probabilities are

exactly as in equations (58) and (59) with indices α and β added:

$$p_{(\ldots,n_{i\alpha}+1,\ldots,n_{i\beta}-1,\ldots)}(\underline{n}) = f_i\left(1-\sum_{j\neq i} g_{ji}\right)\left(\frac{n_{i\beta}}{N}\right)\left(\frac{n_{i\alpha}}{N}\right)+\sum_{j\neq i} f_j g_{ij}\left(\frac{n_{i\beta}}{N}\right)\left(\frac{n_{j\alpha}}{N}\right). \tag{88}$$

Here we have removed an A_β allele from island i and replaced it with an A_α allele, also from island i (first term) or with a migrant from island j (second term).

We assume that mutation occurs at the point that a parent is copied to make the new offspring, no matter whether that offspring remains on its parent's island or emigrates. Furthermore, the probability that a parent with allele β gives birth to an offspring with allele α does not depend on either the source or target islands, i.e., the mutation occurs with probability $u_{\alpha\beta}$ as before. Including this mutation process, the transition probability becomes

$$p_{(\ldots,n_{i\alpha}+1,\ldots,n_{i\beta}-1,\ldots)}(\underline{n}) = f_i\left(1-\sum_{j\neq i} g_{ji}\right)p_\alpha(\underline{n}_i)\left(\frac{n_{i\beta}}{N}\right)+\sum_{j\neq i} f_j g_{ij} p_\alpha(\underline{n}_j)\left(\frac{n_{i\beta}}{N}\right). \tag{89}$$

As before, $\alpha \neq \beta$ and it is understood that n_{iM} should be replaced by $1-\sum_{\gamma=1}^{(M-1)} n_{i\gamma}$. The transition rates follow as before.

Starting from the master equation (40) it is straightforward to obtain the Fokker–Planck equation using these transition rates, the algebraic steps simply being a combination of those which need to be carried out for the cases considered in sections 3.3.1 and 3.3.2. The details are given in the appendix, where it is shown that the resulting Fokker–Planck equation is given by

$$\begin{aligned}\frac{\partial P}{\partial \tau} = &-\sum_{i=1}^{L} f_i \sum_{\alpha=1}^{M-1}\frac{\partial}{\partial x_{i\alpha}}\left[\mathcal{A}_\alpha(\underline{x}_i)P\right]+\sum_{\langle ij\rangle}\sum_{\alpha=1}^{M-1}\left(\mathcal{G}_{ij}\frac{\partial}{\partial x_{i\alpha}}-\mathcal{G}_{ji}\frac{\partial}{\partial x_{j\alpha}}\right)\left[(x_{i\alpha}-x_{j\alpha})\,P\right]\\ &+\frac{1}{2}\sum_{i=1}^{L} f_i\sum_{\alpha=1}^{M-1}\sum_{\beta=1}^{M-1}\frac{\partial^2}{\partial x_{i\alpha}\partial x_{i\beta}}\left[\mathcal{D}_{\alpha\beta}(\underline{x}_i)P\right],\end{aligned} \tag{90}$$

where $\mathcal{A}_\alpha(\underline{x})$ and $\mathcal{D}_{\alpha\beta}(\underline{x})$ are given by equations (79) and (80) respectively.

An equation of a very similar form to equation (90) has been obtained in connection with the model of language evolution previously discussed [20]. In that case the equation was derived through purely mesoscopic reasoning—there were no 'islands' of N tokens, only fractions, x, of different variants in a speaker's grammar. In this case in order to derive the Fokker–Planck equation, the mutation and migration rates also had to be scaled, now not by N, but by λ, the weight given to new tokens. The generational time had also to be scaled, this time by λ^{-2}. Thus the scalings required to derive the Fokker–Planck equation are identical in both cases.

4. Backward-time formulation: coalescent theory

In section 2 we described the Wright–Fisher model as a forward-time process, i.e., one specifies the probability of having a certain number of alleles in a subsequent generation *given* the number present in the current generation. It is perfectly legitimate instead to ask for the probability that a certain number of individuals in the current generation are

all descended from a set of individuals from the *previous* generation. As we describe in this section, one can reconstruct the ancestry of the present-day population given that it has evolved under the dynamics defined in section 2. We shall also show below that this can be used to recover any desired property of the population going forward in time from a known initial condition. Thus these forward- and backward-time approaches are entirely equivalent.

4.1. Reconstructing the ancestry

To describe the backward-time approach in detail, we return to the ideal Wright–Fisher population of N haploid individuals described in section 2. Each of these individuals must have a parent from the previous generation; some of them may share a parent, in which case there are individuals in the previous generation that did not have any offspring. Hence, as one looks back in time the number of ancestors of the present-day population must decrease until a single ancestor of the entire population is found.

Let us assume that at t generations in the past (note that in this section, increasing t means going back further in time) there are n ancestors of the present-day population. The probability that any pair of these ancestors have the same parent is $1/N$. This can be understood from the fact that when going forward in time, a new generation can be constructed by assigning to each of the N offspring a parent chosen at random from one of the N individuals in the previous generation. The probability that out of any pair of offspring, the second receives the same parent as the first is then clearly $1/N$ as claimed. If the number of ancestors is much smaller than the total population size, $n \ll N$, the probability that three or more individuals are siblings, or that there is more than one pair of siblings among the subpopulation of n ancestors, is of order $1/N^2$ and can thus be neglected in the limit $N \to \infty$. As one steps back a single generation, therefore, the number of ancestors decreases by one with probability

$$p_c(n) = \binom{n}{2}\frac{1}{N}, \tag{91}$$

since there are $\binom{n}{2}$ pairs that could find a common parent; otherwise the number of ancestors stays the same.

It is customary to represent this decrease in the number of ancestors in the form of a tree with the vertical direction representing time (going upwards corresponds to going backward in time). Each branch represents an ancestor of the present day population and two branches coalesce when the corresponding ancestors find a common parent—see figure 2. For this reason this backward-time process was called 'the coalescent' when formalized mathematically by Kingman in the 1980s [50], although the underlying ideas had previously been employed by population geneticists prior to this point (some discussion of this earlier work can be found in [51]).

Another way to picture this process, which may be more appealing for physicists, is as a particle reaction system. Let there be a lattice (or, to be more technically correct, a graph) with N sites (nodes), n of which are occupied by a particle. In each time step, each particle either remains where it is, or hops with some probability to one of its neighbours. Should two particles ever occupy the same site simultaneously, they immediately coalesce—see figure 3. One can verify that the backward-time formulation

doi:10.1088/1742-5468/2007/07/P07018

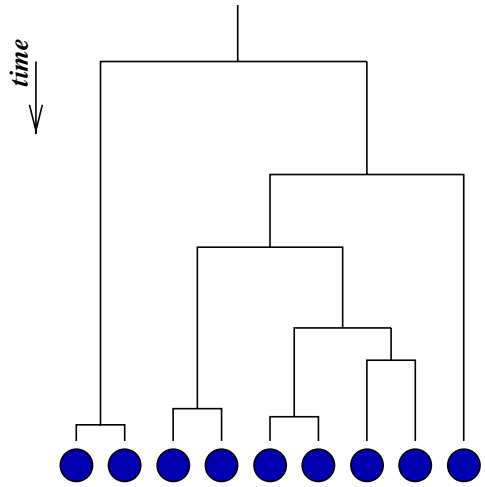

Figure 2. Graphical representation of the coalescent process. Each vertical line corresponds to a lineage leading to the final (present-day) population. Two lineages merge whenever the ancestors they correspond to find a common parent. Since this coalescent rate is proportional to $n(n-1)$, where n is the number of lineages remaining and is much smaller than the total population size N, the expected time between coalescence events considerably lengthens as one goes back in time.

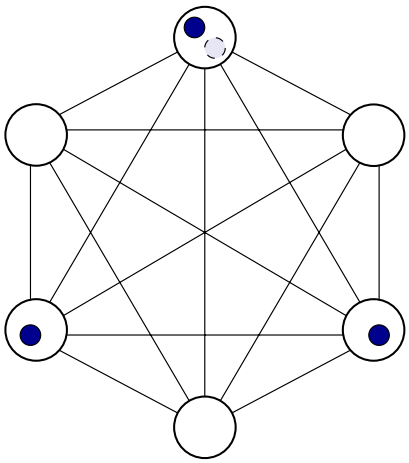

Figure 3. Interpretation of the ancestral dynamics as a reaction–diffusion process on the complete graph. Each particle represents a lineage, and hops with probability $1/(N-1)$ to a neighbouring site in each time step. Should two particles ever coincide on the same site, they react immediately, leaving only a single particle.

of the ideal Wright–Fisher model corresponds to this reaction–diffusion process on the complete graph (i.e., a network in which each site is connected to every other—also sometimes called a mean-field geometry).

From p_c one can calculate statistics of the length of time that elapses between a state with n ancestors to a state with $m < n$ ancestors (as long as both n and m are much

smaller than N). This is because p_c is constant until a coalescence occurs, and hence the time spent in a state with n ancestors has a geometric distribution

$$P_n(t) = [1 - p_c(n)]^{t-1} p_c(n). \tag{92}$$

The mean of this distribution is thus $1/p_c(n)$, and since successive coalescence events are uncorrelated, the time from n to m ancestors is

$$T(n \to m) = \sum_{\ell=m+1}^{n} \frac{1}{p_c(l)} = 2N \sum_{\ell=m+1}^{n} \frac{1}{\ell(\ell-1)} = 2N \left(\frac{1}{m} - \frac{1}{n} \right). \tag{93}$$

This result shows that the ancestry of a present-day population is dominated by an era in which the number of ancestors is small: the mean time in which there are two ancestors $T(2 \to 1)$ is more than half the time $T(m \to 1)$ to a single common ancestor from any number $m \ll N$ ancestors. In fact, it can further be shown that if one starts with a sample comprising the entire population $m = N$, in the limit of an infinite population $N \to \infty$, after any finite rescaled time $\tau = t/N$, the number of ancestors remaining is finite (i.e., a vanishing fraction of the initial population) [52]. This is significant because it means that the condition for only pairwise coalescences with non-zero probability, i.e., that $n \ll N$, is satisfied at any finite rescaled time τ in the past.

This backward-time description of Wright–Fisher population dynamics has a number of practical benefits. Firstly, the data that geneticists have to hand comes from some present-day sample of individuals, and thus it is sensible to condition the outcome of the evolutionary dynamics on this known outcome. Indeed, geneticists are often interested in the information obtained by reconstructing the ancestry of sampled genes, and a number of sophisticated methods based on ensembles of ancestral trees have been developed with this aim in mind [53]. Secondly, it is very much more efficient to simulate the backward-time process: to construct a genealogical tree for n present-day individuals with the desired rate involves generating $n - 1$ waiting times from the appropriate geometric distributions (92). Furthermore, having constructed this tree, one can superimpose various models of mutation (such as that described previously) since selectively neutral mutations (i.e., those that do not cause individuals carrying a particular allele to have more offspring than others) by definition do not affect the population dynamics. Finally, as we will discuss in section 5 below, the backward-time method generalizes more readily to non-ideally mating populations (although one is restricted to selectively neutral alleles).

4.2. Relationship between forward- and backward-time dynamics

Clearly, it should be possible to use the backward-time formulation to calculate properties of the forward-time evolution from a fixed initial condition. The way to do this is to note that the process of assigning mutant alleles to individuals in the initial condition is an example of a mutation process that has no effect on the statistics of the line of descent. Note, however, that this is an unusual mutation process in which a large number of mutations occur in parallel at some early time, after which no further mutations take place. Consider now figure 4. It shows the lines of descent from some initial condition: we can take the individuals shown shaded at the earlier time (top) to be mutants, and the remainder to carry the wild-type allele. At the later time, we can select some other set of

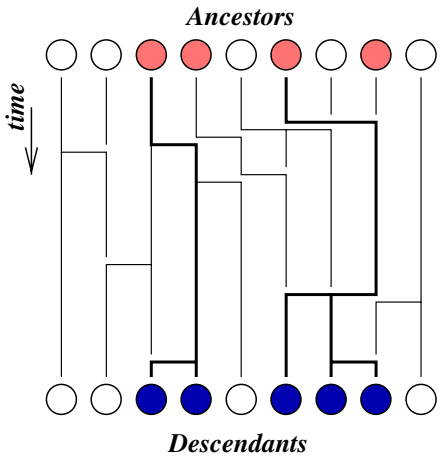

Figure 4. Lines of ascent/descent connecting two sets of individuals, one in the initial condition and one in the final condition. Note here that shading does not refer to the allelic state of any individual, rather whether it is considered to be within or outside the set of interest at the early or late time. The heavy lines show the lines of descent from the early-time set that contribute to the late-time set.

individuals (e.g., that formed by the dark-shaded individuals) and ask for the probability that all are mutants.

Clearly for them all to be mutants, they must all be descended from the mutants in the initial condition (if no mutation events take place in the interim). However, there may also be other mutants present at the later time outside the set of interest (there are two such mutants in figure 4). Therefore, to construct the probability $R(a, d; t)$ that all d sampled individuals at late time are descended from one, more or all of the a mutants at the initial time from $P(b|a; t)$, the probability that there are *precisely* b mutants t generations after the initial condition, one needs to sum the latter, weighted by the probability that the chosen set falls within the late-time mutant set. Given that there are b mutants in the population, there are $\binom{b}{d}$ ways to choose d individuals in such a way that all are mutants. Since there are in total $\binom{N}{d}$ possible ways in which d individuals could have been chosen, we have that

$$R(a, d; t) = \sum_b \frac{\binom{b}{d}}{\binom{N}{d}} P(b|a; t). \tag{94}$$

We can ask the same question using the backward-time language. Let $Q(c|d; t)$ be the probability that the late-time chosen set of d individuals has coalesced into c ancestors t generations previously. For all d late-time individuals all to be mutants, their c ancestors must all be drawn from the initial mutant set. Using the same kind of counting argument as previously one finds that

$$R(a, d; t) = \sum_c \frac{\binom{a}{c}}{\binom{N}{c}} Q(c|d; t). \tag{95}$$

Equating these two expressions gives an implicit duality relation between the forward- and backward-time distributions P and Q; this can be found for example in [54]. Using the

identity

$$\sum_{d=b}^{b'}(-1)^{b+d}\frac{\binom{N}{b}\binom{N-b}{d-b}\binom{b'}{d}}{\binom{N}{d}} = (-1)^b\binom{b'}{b}\sum_{d=b}^{b'}(-1)^d\binom{b'-b}{d-b} = \delta_{b\,b'}, \tag{96}$$

one can write this duality relation to give P explicitly in terms of Q [24]. One finds

$$P(b|a;t) = \sum_{c=1}^{a}\sum_{d=b}^{N}(-1)^{b+d}\frac{\binom{N}{b}\binom{N-b}{d-b}\binom{a}{c}}{\binom{N}{c}}Q(c|d;t). \tag{97}$$

Since it is usually the case that backward-time quantities are more easily obtained than their forward-time counterparts (e.g., by Monte Carlo sampling of the ancestral dynamics, or the equivalent reaction–diffusion process), this formula provides one means for efficient computation of the evolution from a fixed initial condition. For example, if one wants to know the probability that a mutant allele has become fixed by time t, one puts $b = N$ in the foregoing to find

$$P(N|a;t) = \sum_{c=1}^{a}\frac{\binom{a}{c}}{\binom{N}{c}}Q(c|N;t). \tag{98}$$

Actually, in this case $Q(c|N;t)$ is known analytically [52, 51], as indeed is $P(N|a;t)$, at least in the limit $N \to \infty$ and after rescaling allele frequency $x_0 = a/N$ and time $\tau = t/N$ [26]. Nevertheless, this approach has recently found utility in applications to the fixation time in subdivided populations in which neither of these distributions can be calculated exactly (see [24] and section 5 below).

4.3. Scaling with population size in the coalescent formulation

In section 3 we saw that the characteristic timescale of evolution grew linearly with population size for the Wright–Fisher model, and quadratically in the Moran model. One sees these same scaling relations in the coalescent formulation. Recall that above we argued that the probability that two lineages coalesce in one generation undergoing Wright–Fisher evolution is $p_c(2) = 1/N$. Under the Moran model, the probability that a pair of lineages coalesces is equal to the probability that one of them was chosen to be the parent individual, and the other the offspring. This is $p_c(2) = 2/N^2$ or $p_c(2) = 2/N(N-1)$ if the sampling is done with or without replacement, respectively. The factor of 2 appears here because there are two ways in which a (parent, offspring) permutation can be assigned to a pair of individuals.

In both cases, the total rate of coalescence if there are n lineages present is $p_c(n) = \binom{n}{2}p_c(2)$ as long as $n \ll N$. Therefore in the limit $N \to \infty$ the ancestries of both Wright–Fisher and Moran populations have the same dynamics as long as time as rescaled as $\tau = t/N$ in the former and $\tau = 2t/N^2$ in the latter. This is precisely the scaling that was previously used to obtain the same Fokker–Planck equation for the two models.

In fact, very many neutral population dynamical processes have the same ancestral dynamics as the Wright–Fisher and Moran models in the limit $N \to \infty$ under the appropriate rescaling of time. For example, if one has a model in which the distribution of offspring number for each individual is invariant under exchanges of the individuals, and one can further define a timescale on which the probability of ternary coalescences is

J. Stat. Mech. (2007) P07018

vanishingly small compared to binary coalescences, then (subject to other rather formal requirements), the ancestral dynamics converges in the limit $N \to \infty$ to that of the Wright–Fisher process (see, e.g., [55] for the full mathematical details). Population geneticists refer to this property as the *robustness* of the coalescent; physicists will recognize this lack of sensitivity of the late-time behaviour to details of the dynamics in large populations as a kind of universality.

Furthermore, one can also understand the need to rescale migration and mutation rates with the population size N from the backward-time formulation. Consider the Moran model with two alleles as defined in section 2.3. The probability that a particular ancestor was chosen as the offspring in the previous generation is $1/N$, and further that the allele mutated after the parent was sampled is either u or v (depending on the allele in question). For coalescence and mutation probabilities to be of the same order in N, we therefore require u and v both to be of order $1/N$.

5. Non-ideal populations

We have thus far mostly restricted our discussion to ideal populations, i.e., those in which the individual(s) displaced through the production of new offspring are chosen from a uniform distribution. We now turn to the contrasting case of non-ideal populations, where the individuals to displace are chosen non-uniformly and investigate the various ways in which population subdivision has been handled by population geneticists.

Non-ideal populations are most easily handled within the backward-time formulation, and therefore it is most convenient to describe the migration events in terms of the set of probabilities μ_{ij} that any given individual randomly sampled from subpopulation (or deme) i is the offspring of a parent individual from the previous generation in deme j. For the case $i \neq j$, this is given simply as the joint probability for a parent to be in deme j and an offspring in deme i, which in our earlier notation is

$$\mu_{ij} = f_j g_{ij}. \tag{99}$$

On the other hand, the probability that an individual in deme i is not an immigrant is obtained from the fact that $\sum_j \mu_{ij} = 1$, namely,

$$\mu_{ii} = 1 - \sum_{j \neq i} f_j g_{ij}. \tag{100}$$

It is worth at this stage to explore these migration parameters a little further. First, if subpopulation i comprises N_i individuals, the expected number of immigrants per generation is

$$N_i^{(\mathrm{in})} = N_i \sum_{j \neq i} \mu_{ij}. \tag{101}$$

Meanwhile, the expected number of emigrants from deme i (defined as the number of offspring of individuals that end up in a deme $j \neq i$) is

$$N_i^{(\mathrm{out})} = \sum_{j \neq i} N_j \mu_{ji}. \tag{102}$$

Migration is called *conservative* if the expected number of individuals entering and leaving every deme per generation is the same, i.e., if $N_i^{(\mathrm{in})} = N_i^{(\mathrm{out})}\ \forall\, i$, or equivalently, if

$$\sum_{j\neq i} N_i \mu_{ij} = \sum_{j\neq i} N_j \mu_{ji} \qquad \forall\, i. \tag{103}$$

Note that this corresponds with the definition of conservative migration previously given (67) when all deme sizes are equal, $N_i = N$. Now, let us briefly consider the common ancestor of a subdivided population. The single lineage that remains as $t \to \infty$ will hop from between subpopulations at a rate μ_{ij} from deme i to deme j. Therefore, the stationary distribution of the common ancestor (assumed unique) is given by the solution of the set of equations

$$\sum_{j\neq i} Q_j^* \mu_{ji} = \sum_{i\neq j} Q_i^* \mu_{ij}. \tag{104}$$

We see that if migration is conservative, i.e., if (103) holds, the probability that the common ancestor is in subpopulation i in the steady state is proportional to its size,

$$Q_i^* = \frac{N_i}{L\bar{N}}, \tag{105}$$

where L is the number of subpopulations and $\bar{N} = (1/L)\sum_{i=1}^{L} N_i$ is their mean size.

Clearly, one can construct models of varying complexity and intricacy by suitable choices of the migration probabilities μ_{ij} and subpopulation sizes N_i, although some cases are still excluded from this description. Consider the following classical examples.

Symmetric island model. The simplest model of subdivision has L equal sized subpopulations (islands) of equal size $N_i = N$ from which a uniform fraction μ of the individuals on each island are replaced by offspring chosen at random from the other islands: $\mu_{ij} = \mu/(L-1)$ for all pairs $i \neq j$. The version of this model that has an infinite number of islands was introduced by Wright [5, 56]; the finite island model is often attributed to Maruyama [57] and Latter [58].

Randomly mating diploid population with two sexes. Recall that a diploid population is one that carries two instances of each gene which may or may not be the same allele. A diploid population of L individuals can be thought of as a set of L subpopulations, each of size 2. One gene from in each subpopulation is sampled randomly from the L_{m} males in the populations; the other from the $L_{\mathrm{f}} = L - L_{\mathrm{m}}$ females. To represent this model via the migration parameters μ_{ij}, however, it is necessary to have $2L$ subpopulations each of size 1. We then take $\mu_{ij} = 1/2L_{\mathrm{m}}$ if subpopulation i is a gene that is inherited from a male and j is one of the $2L_{\mathrm{m}}$ genes carried by a male in the previous generation; and $\mu_{ij} = 1/2L_{\mathrm{f}}$ is the corresponding quantity for females.

Randomly mating asexual diploid population with hermaphroditism. As an alternative to sexual reproduction, the two genes carried by a diploid individual could be inherited from the same parent (through a process called *selfing*) with probability S, or from two different parents with probability $1-S$ (by *outcrossing*). This model cannot be reproduced using the μ_{ij} parameters because the parental distribution is not independent for the two genes carried by a single individual.

It is also possible for populations to be subdivided by age group, for example, if only individuals within some age class are fertile. Given that considerable complexity is

possible within even the simple general model of population subdivision outlined here, it is unsurprising that geneticists have sought a small number of parameters with which to characterize and differentiate the various specific cases that can be constructed. Two quantities that appear prominently throughout the population genetics literature are *inbreeding coefficients* and *effective population size*. Both of these concepts are commonly said to have been introduced by Wright [59].

5.1. Inbreeding coefficients, fixation indices and F-statistics

Geneticists are often most interested in diploid organisms, and there an important question is whether the two instances of a gene at a specified locus are the same or not. In the absence of mutation, the two genes are the same only if they share a common ancestor (a situation described as being *identical by descent*). Since this is the case if some *inbreeding* has occurred, the probability F that two genes are identical by descent is sometimes called an *inbreeding coefficient*.

Given a sample of some individuals taken from a population, F can be estimated by examining specific *genetic markers*. For example, one can look for different enzymes that can catalyse a particular chemical reaction, or look directly at DNA sequences (see e.g., [60] for a description of various methods in the context of subdivided populations). Of particular interest is the way in which population structure manifests itself in observed identity probabilities: typically, one expects there to be a greater chance for genes sampled from a subpopulation to be identical, than those from the wider (meta)population.

A quantity that encodes this information and that has become ubiquitous in analyses of population structure is denoted F_{ST}, variously called an inbreeding coefficient, fixation index or F-statistic, and was introduced by Wright [59]. A number of subtly different, sometimes ambiguous, definitions of this quantity can be found dotted around the literature—one senses a slight air of frustration arising from this state of affairs in the introduction of [61] for example. Following the approach of [62] (and others), we avoid getting too heavily into the distinctions between different expressions for F_{ST}, and simply adopt a particular definition, namely that provided by Nei [63] which reads

$$F_{\mathrm{ST}} = \frac{F_0 - \bar{F}}{1 - \bar{F}}. \tag{106}$$

Here F_0 is the probability that two genes drawn at random from the same subpopulation are identical by descent, and $\bar{F}$ is the corresponding quantity for a pair of genes sampled randomly from the entire population. Nevertheless this definition leaves open the question of whether these samples should be weighted according to the subpopulation sizes or not; in practice, this depends on how F_{ST} has been determined from an actual sample at hand.

Despite its ubiquity and utility, there are various aspects of this quantity that are confusing at a first encounter. First there is the (rarely elucidated) notation: S and T simply stand for subpopulation and total population respectively. With reference to diploid populations, one can also define analogous quantities F_{IS} and F_{IT} which measure how much more (or less) closely related the pair of genes contained within an individual are than a pair sampled from the population at large.

Next, one often finds a single expression of F_{ST} quoted for a given model of subpopulation, even though identity probabilities in general change with time. Wright's infinite island model (defined above) provides one example of a case where F_{ST} is time

independent. Since the total population size is infinite, the number of individuals across the whole population that carry a particular allele does not change with time. However, if one takes a subsample of this population of size N, there will be variation in the number of alleles of a particular type, and so one finds $F_0 \neq \bar{F}$ which leads to a non-trivial $F_{\rm ST}$. Wright [56] obtained an expression for $F_{\rm ST}$ by taking the case of two alleles with a fraction x_i of the individuals in subpopulation i carrying one of them (say, the wild type). In such a case, one can show that

$$F_0 = \frac{1}{L}\sum_i \left[x_i^2 + (1-x_i)^2\right] = 2\overline{x^2} - 2\bar{x} + 1 \tag{107}$$

$$\bar{F} = \frac{1}{L^2}\sum_{i,j} \left[x_i x_j + (1-x_i)(1-x_j)\right] = 2\bar{x}^2 - 2\bar{x} + 1, \tag{108}$$

where $\overline{f(x)} = (1/L)\sum_i f(x_i)$. Hence,

$$F_{\rm ST} = \frac{\overline{x^2} - \bar{x}^2}{\bar{x}(1-\bar{x})}. \tag{109}$$

We remark that in the limit of infinite subpopulations (where there is no distinction between sampling with and without replacement), one has $0 \leq F_{\rm ST} \leq 1$, the lower and upper bounds relating to extremes of homogeneous and heterogeneous spatial distributions of alleles respectively. These bounds one can see by noting first that the numerator of (109) is positive, since it is a variance, and second that since all $0 \leq x_i \leq 1$ it follows that $0 \leq \bar{x} \leq 1$ and further that $\overline{x^2} \leq \bar{x}$. When subpopulations are finite, it is possible for $F_{\rm ST}$ to be slightly negative as a consequence of the effects of sampling without replacement: having removed one gene from a subpopulation, the probability of finding an identical gene in another subpopulation could be of order $1/N$ *more* likely than in the same subpopulation. Therefore, if one uses (109) as a general definition of $F_{\rm ST}$ rather than (106), as is done for example in [61], the results obtained will not necessarily agree in situations where subpopulations are finite.

Using (109), Wright [56] showed by an explicit calculation of the variance in the allele frequencies that for the island model

$$F_{\rm ST} = \frac{(1-\mu)^2}{N - (N-1)(1-\mu)^2}. \tag{110}$$

A quick way to obtain this result is to focus on a single subpopulation and assume that any immigrants that arrive from the wider population do not have a common ancestor (and are hence unrelated). This assumption can be justified by switching to the coalescent picture outlined in section 4, suitably extended to cater for population subdivision. If two lineages are situated within a single subpopulation, they coalesce at a rate $1/N$, whereas two lineages in different subpopulations will coalesce at a rate proportional to μ/L. If the subpopulation size N is held fixed whilst $L \to \infty$, a separation of timescales occurs and the probability that two lineages in different subpopulations coalesce effectively vanishes. Hence, $\bar{F}$ is effectively zero (since the chance of two randomly chosen individuals are from the same subpopulation is also of order $1/L$), and $F_{\rm ST} = F_0$. Using now the fact that F_0

does not change with time, one can establish that

$$F_0 = (1-\mu)^2 \left[\frac{1}{N} + \left(1 - \frac{1}{N}\right) F_0 \right] \tag{111}$$

since the right-hand side of this equation gives the value of F_0 after one generation of the Wright–Fisher process (as defined in section 2). This comes about because the probability that the members of any pair are both descended from individuals from the subpopulation of interest is $(1-\mu)^2$; the probability that both are offspring of the same parent is $1/N$; and if they are offspring of different parents, those parents were themselves related with probability F_0. Rearranging this equation yields (110). Note that within this picture, F_{ST} can be interpreted as the probability that two lineages in the same subpopulation coalesce before one of them migrates away (looking backward in time).

It is worth looking more closely at how F_{ST} behaves in the various limits that have previously been established. First, if the migration probability μ remains finite as the subpopulation size $N \to \infty$, $F_{\mathrm{ST}} \to 0$ which indicates an absence of spatial structure in gene frequencies—precisely what one would expect in the strong migration limit. On the other hand, if the combination $N\mu$ is finite as $N \to \infty$, one has

$$F_{\mathrm{ST}} \approx \frac{1}{1+2\mu N} \tag{112}$$

to leading order in $1/N$. As a consequence of this relationship, a value of F_{ST} estimated by sampling from a population is often used as a means to obtain the mean *number* of migrants μN arriving in a subpopulation in each generation, this being considered an easier task practically than tracking the movement of individuals between subpopulations. We also remark that (110) is not restricted to the special case where the migration rate between every pair of islands is the same, since it has only been assumed that F_0 is stationary, and that all immigrants into the subpopulation of interest are unrelated. In practice this means one requires an infinite total population that is at equilibrium, or has stationary mean allele frequencies as a consequence of conservative migration. Nevertheless, these are still quite restrictive assumptions and it turns out that estimating μN in this indirect way is fraught with difficulties [64].

There is therefore much interest in exploring how F_{ST} behaves under more general conditions, and in particular under models of population subdivision in which migration is not conservative, or the number of subpopulations is finite. In these models, allele frequencies vary with time, and typically therefore the focus is on the value of F_{ST} that is approached as equilibrium is reached. In a model with a finite total population, the equilibrium state is one of fixation, as previously discussed in section 2. Hence all identity probabilities tend to unity. However, the ratio that appears in (106) approaches a non-trivial finite value, and thus contains some information about population structure that can be related to gene diversity.

A popular way to calculate this limiting value is to exploit a trick due to Slatkin [65]. Let us trace the ancestry of a pair of genes sampled from the present day population. From the discussion of section 4 we know that eventually a common ancestor of these genes will be found; let us suppose that this happens at time t with probability $c(t)$. Let us also introduce a Poisson mutation process that occurs with probability θ per generation. Then, the probability that the two genes are *identical in state* (i.e., share a common ancestor

and have not been subject to any mutations since the time of coalescence) is

$$F(\theta) = \sum_{t=1}^{\infty} c(t)(1-\theta)^{2t}. \tag{113}$$

In the limit $\theta \to 0$, the concepts of identity by descent and identity in state become equivalent, and so asymptotically,

$$F_{\mathrm{ST}} = \lim_{\theta \to 0} \frac{F_0(\theta) - \bar{F}(\theta)}{1 - \bar{F}(\theta)} = 1 - \frac{T_0}{\bar{T}}, \tag{114}$$

where T_0 and $\bar{T}$ are the mean times for two genes randomly sampled from within or between subpopulations respectively to find a common ancestor. In essence, therefore, the task is to calculate the mean first passage time for a pair of random walkers which hop between sites of a network at rates controlled by the migration parameters μ_{ij}.

This is achieved by solving the set of linear equations

$$T_{ij} = 1 + \sum_{k\ell} \mu_{ik}\mu_{j\ell}T_{k\ell}\left(1 - \frac{1}{N_k}\delta_{k,\ell}\right), \tag{115}$$

where T_{ij} is the mean time for two lineages, one in subpopulation i and one in subpopulation j to coalesce. This equation arises because if one has two distinct lineages, one needs to wait at least one generation for a change of state to occur. The joint probability that the lineages in i and j hop to k and ℓ is $\mu_{ik}\mu_{j\ell}$. If they arrive in different subpopulations, we must then further wait a mean time $T_{k\ell}$ for coalescence to occur. On the other hand, if they land in the same place k, there is a probability $1/N_k$ for coalescence to occur immediately (i.e., a mean waiting time of zero), otherwise we need to wait T_{kk} on average. As with F_{ST}, the question arises as to how to form the averages T_0 and $\bar{T}$ appearing in (114). For example, one could take $T_0 = T_{ii}$ for a specific subpopulation of interest. Alternatively, one could average over all subpopulations, possibly weighted according to their size. It is not uncommon to see both size-weighted and unweighted versions of $\bar{T}$ in the literature, presumably because the correct form depends on whether subpopulation sizes are known experimentally or not (some further discussion is given in [62]).

For arbitrary migration rates (115) can be rather cumbersome to solve, even within simple models of population subdivision. Simplifications occur in the limit of slow migration, where $\mu_{ij} = m_{ij}/\bar{N}$ where $\bar{N}$ is the mean subpopulation size and is taken to infinity. Recall from section 2 that this is the regime in which both drift and migration operate on the same timescale in the limit of an infinite population size; meanwhile, in the backward-time formulation, this scaling has migration events occurring on the same timescale as coalescence (see, e.g., [66]). In this limit, terms which have both lineages hopping simultaneously do not enter into (115). Furthermore, it has been shown [67] that, if migration is conservative, the mean time for coalescence of lineages located in the same subpopulation i is independent of i and given by $T_{ii} = L\bar{N}$. This fact considerably simplifies calculation of F_{ST} for conservative models. For example, Slatkin [65] showed that for the symmetric island model with a finite number L of islands undergoing a slow migration process,

$$F_{\mathrm{ST}} = \frac{1}{1 + 2Nm\,(L/(L-1))^2} \approx \frac{1}{1+2Nm}\left[1 + \frac{2}{1+(1/2Nm)}\frac{1}{L} + \mathrm{O}\left(\frac{1}{L^2}\right)\right] \tag{116}$$

which agrees with the result previously obtained by other methods (such as that used by [58]; see also [62]). We observe the finite size corrections to the infinite island result, (110), are of order $1/L$ which perhaps sheds some light on the ubiquity of the expression (110) in analyses of population structure.

This procedure for calculating the fixation index F_{ST} has been followed for a range of models of population subdivision, such as a hierarchical island model [68] (which has one finite island model nested inside another), a model with two subpopulations of different sizes [62] and a continental island model, where migration does not take place directly between the islands but only via a central continent [62]. These models are tractable because of some symmetry or regularity in the migration structure that can be exploited. It is also possible to make progress in cases where the number of subpopulations tends to infinity [69]–[71]. However, we will not discuss these models further here.

5.2. Effective population size

As we described in the previous section, the motivation behind studying F_{ST} is that it is a statistic of a population that can be estimated by examining appropriate data. A second characteristic of biological populations that has also received considerable attention is its *effective size*, N_{e}. Like F_{ST}, this notion can also be attributed to Wright [5], although it is also understood that Crow and co-workers also made significant contributions to its development (see, e.g., the classic text [26]). The basic idea here is that an ideal population is entirely characterized by the number of individuals N that are randomly mating. In a real population, one suspects that fluctuations (like those associated with genetic drift) might be due to some special 'breeding' individuals that form a subset of the full population.

There are as many ways to define N_{e} as there are properties of a population that one might be interested in; furthermore, these different definitions will not generally give the same value for a given model of population subdivision. Nevertheless, for the concept to be useful, one hopes that the effective size determined by examining one quantity is indicative of how the population as a whole will behave. It is not appropriate here to get too involved in all the various definitions and their pros and cons; instead we will highlight two particularly prominent definitions and give a flavour of how they depend on certain aspects of the population structure. For more extension discussions we refer the reader to [72, 73].

One of the most natural definitions of effective size arises from the forward-time prescriptions of the population dynamics given in section 2. Recall that under the Wright–Fisher dynamics, if the allele frequency in one generation is known to be x', in the next it is a random variable x that has variance $x'(1-x')/N$. This connection between the variance of an allele frequency and population size can be exploited to define an effective population size in a more general context. Let $\bar{x}'$ be some measure of an allele frequency across a subdivided population, e.g., the arithmetic mean $\bar{x}' = (1/L)\sum_{i=1}^{L} x'_i$, and $\mathrm{Var}(\bar{x})$ the variance in this quantity after one generation of the underlying population dynamics. An effective size N_{e} of this subdivided population can then be given as

$$N_{\mathrm{e}} = \frac{\bar{x}'(1-\bar{x}')}{\mathrm{Var}(\bar{x})}. \tag{117}$$

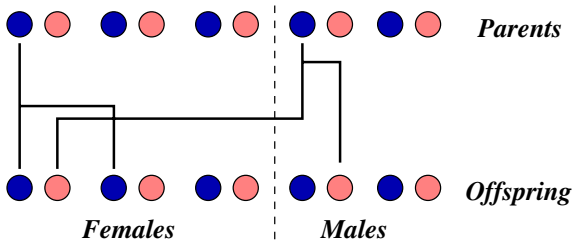

Figure 5. Sexually reproducing diploid population. Each individual carries one gene inherited from a female (shown solid) and one from a female (shown shaded). Two ways in which two distinct genes may coalesce are shown. In both cases, the two offspring must have been inherited from an individual of the same sex.

To distinguish it from other effective sizes than can be defined, this one is called the *variance effective size*. Note that in this expression the variance is over the distribution of allele frequencies $\bar{x}$ generated by the population dynamics, and *not* the variance over subpopulations appearing in (109). A drawback of this definition of N_e is that it is likely to depend on the frequency $\bar{x}$, and possibly other statistical quantities—such as the variance of allele frequencies over subpopulations. Therefore, if one seeks a description of the dynamics that is closed in a small set of random variables, it will typically have to be an approximate description (see, e.g., [74]–[76]).

A definition of effective population size that avoids these difficulties makes use of the fact that when tracing lineages of an ideally mating population backward in time, pairs coalesce at a rate $1/N$. The mean rate of coalescence between pairs of lineages in a subdivided population can then be taken to define the reciprocal of an effective population size. As with the fixation index F_{ST}, this rate will typically vary with time; in those cases, one can adopt the coalescence rate that is approached as one looks infinitely far back in time. The effective size that results is then often referred to as an *asymptotic effective size*. In fact, the same limit has been shown to be approached by a range of different definitions of effective size [77, 73]. This fact can be traced to all definitions being governed asymptotically by the decay rate of the longest-lived non-stationary eigenstate of the underlying population dynamics. Furthermore, this feature makes the asymptotic effective population size an appealing quantity to devote some effort to calculating.

As with F_{ST}, there are a few cases where the mean coalescence rate does not change with time and can be easily be calculated exactly. The first two such cases were introduced at the start of this section.

Randomly mating diploid population with two sexes. If there are two distinct lineages in a diploid population, they must both be inherited from the same parent (and furthermore be the same gene from that parent) if they are to coalesce—see figure 5. Taking two genes from the population at random (with replacement for simplicity), the probability that both were inherited from a male, or both from a female, is $\frac{1}{4}$. Within these classes, the same gene is chosen with probability $1/2L_m$ and $1/2L_f$ for male and female parents respectively. Hence one finds a total coalescence probability per generation of

$$\frac{1}{4}\left(\frac{1}{2L_m}+\frac{1}{2L_f}\right)=\frac{L_m+L_f}{8L_mL_f}=\frac{1}{N_e}, \tag{118}$$

where we have identified the coalescence rate with an inverse population size as described

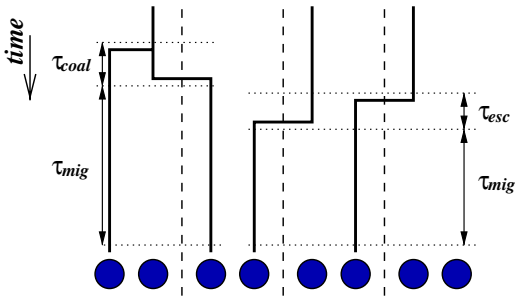

Figure 6. Separation of timescales in a hermaphrodite diploid population. The typical timescale $\tau_{\rm mig}$ for two lineages to migrate into the same individual is very much longer than the typical time to coalescence $\tau_{\rm coal}$ due to selfing, or to escape to different individuals $\tau_{\rm esc}$ due to outcrossing.

above. This result was first given by Wright [5]. Notice that $N_{\rm e} \leq 2(L_{\rm m} + L_{\rm f})$, that is, if there is any imbalance in the number of males and females, the effective number of gene instances in the population is less than their actual number.

Randomly mating asexual diploid population with hermaphroditism. In this model, the effective size can be calculated as a consequence of a separation of timescales. A classical calculation can be found in [78]; here we follow the coalescent based approach introduced in [79], also outlined in [66]. As we have already remarked, this model is subtly different to that described in section 2, in that the parents of two lineages are not chosen independently. Instead, should two lineages find themselves within the same individual there is a selfing probability S that both are descended from genes contained within a single parent individual, and $1-S$ that they are each descended from different individuals. At the time of arrival into a single individual, these two lineages have a probability $1/2$ of immediately coalescing; otherwise, going back a further generation there is a probability $S/2$ that they are both still in the same individual, uncoalesced, or $S/2$ of coalescing, or $1-S$ for the two lineages to return to a state at which they are located in different individuals. Using the properties of the geometric distribution, it can be shown that two lineages arriving at the same individual have a total coalescence probability of $1/(2-S)$, and the mean time to do so is $S/(2-S)$; on the other hand, escape occurs with probability $(1-S)/(2-S)$ and takes on average $2/(2-S)$ generations [79, 66]. Meanwhile, when two lineages are in different individuals, the mean time for them to arrive in the same individual is L, the number of individuals (since the arrival probability per generation is $1/L$). Thus, once L is taken sufficiently large at any fixed S, the rate of coalescence upon arrival is very much faster than the rate of arrival—see figure 6. Then, one can approximate the coalescence rate as the product of the arrival rate and the subsequent probability of coalescence, i.e.,

$$\frac{1}{N_{\rm e}} = \frac{1}{L(2-S)}. \tag{119}$$

Again, the effective (haploid) population size $N_{\rm e} = L(2-S)$ is less than twice the actual population size, unless there is no selfing ($S = 0$), when the effective and actual sizes are equal.

Population subdivision in the strong migration limit. The coupling of a large ratio of coalescence to migration rates and small subpopulation sizes is crucial in admitting the computation of effective size of a hermaphrodite population. In the opposite limit, where migration is very fast compared to coalescence and subpopulation sizes are large, a simple expression valid for general models of population subdivision can also be obtained. The key is to realize that between coalescence events, all lineages are independently distributed according to the stationary distribution generated by the backward-time migration dynamics. That is, we assume that the probability for two lineages to be present in subpopulation i is $(Q_i^*)^2$ where Q_i^* is the stationary distribution for the single common ancestor of the whole population given by the migration balance equations (104). Given that the rate of coalescence for two lineages in subpopulation i is $1/N_i$ it then follows that under this approximation the mean coalescence rate is

$$\frac{1}{N_\mathrm{e}} = \sum_{i=1}^{L} \frac{(Q_i^*)^2}{N_i}. \tag{120}$$

This formula has been shown to be exact (to leading order in $1/N$) if the migration probabilities μ_{ij} obey

$$\lim_{\bar{N}\to\infty} \bar{N}\mu_{ij} = \infty, \tag{121}$$

where $\bar{N}$ is the mean subpopulation size [80]. This equation defines a strong migration limit, within which it is seen that the effective population size is always less than or equal to the total population size $L\bar{N}$. Equality is reached only if the stationary probabilities Q_i^* are in proportion to the subpopulation size, i.e., if $Q_i^* = N_i/(L\bar{N})$. If the stationary distribution is unique, this is the case only if migration is conservative. Since ancestral lineages are well mixed, one does not see any variation between subpopulations, and so the approximation of the entire population as an ideal population with a possibly reduced effective size would appear to be a good one.

For more general models of subdivision, where migration probabilities are typically of order $1/\bar{N}$, fewer exact results are available. Nevertheless, a formula for the effective population size in terms of the asymptotic changes in identity probabilities has been given [73] as

$$\frac{1}{N_\mathrm{e}} = \lim_{t\to\infty} \frac{\sum_{i=1}^{L}(1/N_i)(Q_i^*)^2(1-F_{ii}(t))}{\sum_{i=1}^{L}\sum_{j=1}^{L} Q_i^*Q_j^*(1-F_{ij}(t))}, \tag{122}$$

where $F_{ij}(t)$ is the probability that two individuals, one sampled from subpopulation i and one from subpopulation j, are identical by descent after t generations, as defined in the previous subsection.

This formula takes a particularly simple form in models when migration is conservative. Then, as shown at the start of this section, the probability Q_i^* that the single remaining ancestor is in deme i in the steady state is proportional to N_i, the size of deme i. Then, reduces to

$$\frac{1}{N_\mathrm{e}} = \frac{1}{LN}\frac{1-F_0}{1-\bar{F}}, \tag{123}$$

where F_0 and $\bar{F}$ here are *deme-size-weighted* probabilities of identity for pairs of genes sampled from within demes and from the whole population respectively, that is

$$F_0 = \frac{1}{L}\sum_{i=1}^{L}\frac{N_i}{\bar{N}}F_{ii} \tag{124}$$

$$\bar{F} = \frac{1}{L^2}\sum_{i=1}^{L}\sum_{j=1}^{L}\frac{N_i}{\bar{N}}\frac{N_j}{\bar{N}}F_{ij}. \tag{125}$$

Defining a size-weighted F_{ST} through (106) and the formulae for F_0 and $\bar{F}$ given here, we find that for conservative models N_{e} and F_{ST} can be related through

$$N_{\mathrm{e}} = \frac{L\bar{N}}{1 - F_{\mathrm{ST}}}. \tag{126}$$

When the assumption of conservative migration is relaxed, one does not find a simple relation between effective population size and inbreeding coefficients. Furthermore, it is not clear that a single parameter (such as N_{e} or F_{ST}) should satisfactorily characterize all aspects of the evolutionary dynamics. For example, when migration is weak, it has been argued that the ancestral dynamics can only be fully expressed in terms of the 'structured coalescent' [81] that takes into account the numbers of lineages present in a particular subpopulation at finite times rather than the weighted late-time average implied by (122) [82, 83]. In such cases, it is therefore necessary to examine specific models. For example, the variation of fixation times in structured populations undergoing weak migration has been investigated in an exact approach based on the duality relation given in section 4 for an ideal population [24]. This allows for a comparison with the fixation time that would be seen in an ideal population with an effective size given by (122). In a range of models considered, it was found that the effective size gives an extremely good prediction for the fixation time, except in a particular case where special properties of the network structure admitted fixation to occur in a finite time in an infinite population. This finite time effect has also been seen in voter models on heterogeneous (scale-free) graphs [84, 85], which in fact is a special case of the general model of population subdivision outlined here with all subpopulation sizes $N_i = 1$.

6. Selection

So far we have considered only neutral models of evolution, that is, those for which there is no preference for a particular allele. Despite being apparently a reasonable model for some aspects of genetic, ecological or linguistic behaviour (as we have previously discussed) geneticists in particular have been interested in the fate of alleles that are selected for or against.

The relationship between the genetic make-up of an individual and its survival is of course very complicated. However, one can explore the effects of selection by simply introducing parameters that determine how many offspring an individual carrying a particular allele (or combination of alleles when diploid organisms are being considered) has on average. In this section we offer a small taste of some evolutionary models that encompass selection; for lengthier excursions into this area we direct readers in the first instance to articles written with an audience of physicists in mind [1]–[3].

Let us return to the model that has a randomly mating haploid population of N individuals with two alleles, denoted A and B. In the neutral model, the parent of each individual in an offspring generation is chosen uniformly from all possible parents. When selection is active, it is supposed that an individual with allele A (B) is chosen to be a parent with a weight w_A (w_B). That is, if there are n' A alleles in the parent generation, each offspring has a probability

$$\frac{n'w_A}{n'w_A + (N-n')w_B}$$

of carrying allele A (at least, if no mutations occur). Since each individual is assigned a parent independently, we have that the probability for there to be n A alleles in the next generation is

$$p_{n\,n'} = \binom{N}{n}\left(\frac{n'w_A}{n'w_A + (N-n')w_B}\right)^n \left(\frac{(N-n')w_B}{n'w_A + (N-n')w_B}\right)^{N-n} \tag{127}$$

$$= \frac{1}{(\bar{w}')^N}\binom{N}{n}\left(\left[\frac{n'}{N}\right]w_A\right)^n \left(\left[1-\frac{n'}{N}\right]w_B\right)^{N-n}, \tag{128}$$

where in the second line we have simplified the notation by introducing the mean *fitness* of a population with n' A alleles

$$\bar{w}' = \frac{1}{N}\left(n'w_A + [N-n']w_B\right). \tag{129}$$

Note that when the two alleles have equal fitness, i.e., $w_A = w_B$, (128) reduces to (2) for a neutral population.

It is usual to take a pre-existing 'wild-type' allele (we'll take this to be B) to have fitness $w_B = 1$, and the 'mutant' (A) to have fitness $w_A = 1+s$. Then, the mean change in the number of mutants in one generation, given that there are $n(t)$ mutants in generation t, is

$$\frac{\langle n(t+1)\rangle - n(t)}{N} = s\frac{n(t)}{N}\left(1-\frac{n(t)}{N}\right) + \mathrm{O}(s^2). \tag{130}$$

Denoting the frequency of mutant A alleles as x, the Fokker–Planck equation can be shown (e.g., via a Kramers–Moyal expansion or a large-N expansion) to be

$$\frac{\partial}{\partial t}P(x,t) = -s\frac{\partial}{\partial x}x(1-x)P(x,t) + \frac{1}{2N}\frac{\partial^2}{\partial x^2}x(1-x)P(x,t) \tag{131}$$

when the relative fitness of the mutant s is small. In this instance, small means $s \ll 1$, and not relative to $1/N$. Hence, if at some fixed s one has the (effective) population size $N \gg 1/s$, the effects of drift can typically be neglected. Then the mean allele frequency satisfies the deterministic equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = sx(1-x) \tag{132}$$

and one finds a logistic growth in the number of mutants

$$x(t) = \frac{x(0)}{x(0) + [1-x(0)]\mathrm{e}^{-st}}. \tag{133}$$

Typically it is assumed that the stochastic effects of drift are important only when one of the allele frequencies is small, and so the logistic growth is taken to be representative of the change in frequency of a beneficial mutation in a randomly mating population once its frequency has reached some threshold (see e.g., [86]). This period of rapid growth is often referred to as a *selective sweep*. Of course, a mutant allele is present only in small numbers when it first appears, and here one can work out the probability that a mutant allele will be lost due to genetic drift. To do this one needs to solve the stationary *backward* Fokker–Planck equation corresponding to (131)

$$0 = x(1-x)\left[s\frac{\mathrm{d}}{\mathrm{d}x}Q(x) + \frac{1}{2N}\frac{\mathrm{d}^2}{\mathrm{d}x^2}Q(x)\right] \tag{134}$$

subject to the boundary conditions $Q(0) = 0$ and $Q(1) = 1$, since $Q(x)$ is the probability that a mutant allele becomes fixed given that its initial frequency is x. Integrating once gives $Q'(x) \propto \mathrm{e}^{-2Nsx}$ and again gives $Q(x) = A\mathrm{e}^{-2Nsx} + B$ where A and B are to be fixed by the boundary conditions. This yields [87]

$$Q(x) = \frac{1-\mathrm{e}^{-2Nsx}}{1-\mathrm{e}^{-2Ns}}. \tag{135}$$

In particular, if there is initially only a single mutant, $x = 1/N$, in the limit of an infinite population one has

$$\lim_{N\to\infty} Q(1/N) = \begin{cases} 2s & 0 < s \ll 1 \\ 0 & s \leq 0, \end{cases} \tag{136}$$

where we recall that to obtain (131) it was assumed that s is small. From the backward Fokker–Planck equation (134) one can also find the mean number of generations τ until fixation of a selectively advantageous allele. It is given approximately as

$$\tau = \frac{2\ln N}{s}, \tag{137}$$

when the combination Ns is large (e.g., s some fixed value and $N \to \infty$) [27].

One may ask how selection affects the backward-time formulation of the population dynamics that is couched in terms of genealogies. It turns out that the complexity of the calculations increases considerably, because as one goes back in time one needs to keep track of the number of alleles of each type present in the population. One can sometimes find situations in which there is an equilibrium in these frequencies, for example, when a selectively disadvantageous allele ($s < 0$) is maintained in a population due to recurrent mutations generating it [88]. When there is a selective sweep, it can be shown that the relevant genealogies are those that have *multiple* lineages coalescing simultaneously [89]—compare with the case of neutral evolution when the probability of a triple merger is suppressed by a factor of $1/N$ compared to that of a pairwise coalescence. This is consistent with the fact that fixation is achieved on a timescale logarithmic in N (137): recall that in a state of fixation, *all* N individuals share a common ancestor. One way to formulate the genealogical process with selection is through the 'Λ-coalescent' [90]; meanwhile, certain aspects have recently been elucidated by considering the properties of noisy travelling waves [91].

Given the discussion of section 5, one may also be interested in determining how effects of selection and population subdivision combine. In a number of ways the situation is

J. Stat. Mech. (2007) P07018

rather similar to the neutral case, at least if the selective advantage s (or disadvantage, if negative) is not spatially dependent. For example, when migration is strong one expects fixation probabilities and times to be given by (135) and (137) but with N being an effective population size of the order of the total population size [92]. In the slow migration limit, the mean time for a lineage to hop between subpopulations $\sim N$ is much longer than the duration of a selective sweep $\sim \ln N$ and so typically each subpopulation is taken to be fixed in either the wild-type or mutant state. One thus calculates fixation probabilities and times by having wild-type subpopulations invaded by their mutant neighbours (and vice versa) on the migration timescale, and a flip from the wild-type to mutant state occurring with the probability given by $Q(1/N)$, and in the other direction with the same expression but with s replaced by $-s$. When conservative migration is in force, it can be shown that the probability a mutant allele fixes is independent of the location of the initial mutation [57, 92]: a similar situation occurred in the neutral case (although the fixation probability is different). To see deviations from this behaviour, one needs either to introduce additional processes (such as extinction and recolonization of subpopulations [93]) or relax the assumption of conservative migration [94]. In the latter case one finds that the network structure connecting the subpopulations strongly influences whether selection or drift is the dominant process. For example, star-like structures amplify the effects of selection and can be constructed such that an advantageous mutation appearing almost anywhere is guaranteed to fix. On the other hand, it is also possible to contrive networks in which fixation occurs whenever the mutation occurs within a particular subpopulation, and never if it appears elsewhere. This latter type of behaviour does not, in fact, depend on the mutation having a selective advantage: even neutral mutations appearing in 'well-connected' parts of a system can fix with high probability when migration is a non-conservative process [24].

Shifting focus from networks to continuous space, one can write down an equation for the advance of an advantageous mutation. Recall that in the absence of drift, one has the deterministic equation (132) for the mutant gene frequency x in a subpopulation. With isotropic migration in one dimension with a coordinate u one would augment this equation with a diffusion term

$$\frac{\partial}{\partial t}x(u,t) = D\frac{\partial^2}{\partial u^2}x(u,t) + sx(u)[1-x(u)]. \tag{138}$$

This is known as the Fisher or KPP (Kolmogorov–Petrovsky–Piskunov) equation [95, 96] and admits travelling wave solutions of the form $x(u,t) = f(u-vt)$ where v is the wave velocity. If one anticipates that the leading edge of the wave has an exponential decay $f(\xi) = \mathrm{e}^{-\lambda\xi}$, one finds a range of velocities $v = D\lambda + s/\lambda \geq 2\sqrt{s/D}$ are possible. It turns out that in such an equation, if the interface between the stable and unstable phases (here, regions of high and low mutant frequencies) is sufficiently sharp, the front velocity selected is the smallest allowed [97]. If genetic drift is reintroduced, e.g., by adding a white noise to (138) with zero mean and variance $x(1-x)/N$ (cf (131), but note the usual problems that arise from introducing multiplicative noise in such an ad hoc way), one expects strong fluctuations in the leading edge of the travelling wave as a consequence of the possibility that a newly introduced mutant may go extinct. This leads to a more diffuse profile than in the deterministic case, along with a decrease in the front velocity. It turns out that the noise-free limit is accessed extremely slowly as N is increased: the

doi:10.1088/1742-5468/2007/07/P07018

diffusion constant vanishes as $(\ln N)^{-3}$ and the difference between the front velocity and its asymptotic limit as $(\ln N)^{-2}$ [98].

One can also consider models with spatially varying fitnesses. A simple example would be if a mutant had fitness advantage $+s$ at position $u > 0$ and disadvantage $-s$ at $u < 0$ (s here is taken to be a positive number) [99]. At $u \to \infty$ one anticipates that the mutant allele would be fixed ($x = 1$), whilst at $u \to -\infty$ only the wild type would be found ($x = 0$). Using (138) the steady state can be found by solving the non-linear equation

$$D\frac{\mathrm{d}^2}{\mathrm{d}u^2}x(u) = \pm sx(u)[1 - x(u)], \tag{139}$$

where the positive sign is taken for $u < 0$ and the negative sign for $u > 0$. The symmetry of the problem implies that $x(u) = 1 - x(-u)$, and in particular that $x(0) = \frac{1}{2}$. However, one finds the step in the fitness landscape at $u = 0$ induces a discontinuity in the gradient of $x(u)$ at that point. To see this, one must solve this differential equation which is achieved by multiplying both sides by $\mathrm{d}x/\mathrm{d}u$ and integrating twice. This procedure leads to [99]

$$x(u) = \begin{cases} \dfrac{3}{2}\left[1 - \tanh^2\left(\dfrac{1}{2}\sqrt{\dfrac{s}{D}}u - \kappa\right)\right] & u < 0 \\ \dfrac{3}{2}\tanh^2\left(\dfrac{1}{2}\sqrt{\dfrac{s}{D}}u + \kappa\right) - \dfrac{1}{2} & u > 0, \end{cases} \tag{140}$$

where

$$\kappa = \frac{1}{2}\ln\left(\frac{\sqrt{3}+\sqrt{2}}{\sqrt{3}-\sqrt{2}}\right) = 1.146\,216\ldots. \tag{141}$$

This function is plotted in figure 7. The spatial variation of allele frequencies due to a balance between migration and a changing fitness landscape has been called a *cline* [100].

Finally one can consider variation in fitness not in real space, but in the space of genotypes (i.e., possible allele combinations). Let us return to a randomly mating diploid population, in which a fraction x_i of all genes are of allele A_i, $i = 1, 2, \ldots, n$. It is fairly straightforward to show [26] that, in the absence of drift, the change in allele frequency per generation is

$$\Delta x_i = \frac{x_i(1 - x_i)}{2}\frac{\partial}{\partial x_i}\ln\bar{w}, \tag{142}$$

where the mean fitness of the population is

$$\bar{w} = \sum_{i,j} x_i x_j w_{ij} \tag{143}$$

and w_{ij} is the fitness of an individual carrying the pair of alleles (A_i, A_j). One can therefore view the function $f(x_1, x_2, \ldots, x_n) = -\ln\bar{w}$ as a kind of free energy defined over the space of allele frequencies which the evolutionary dynamics seeks to minimize. If we now extend the discussion to multiple gene *loci* (i.e., genes coding for different traits) with interactions between them, it is possible for an extremely rugged fitness landscape in the space of all possible allele frequencies to emerge [101]. One can then find a transition between two

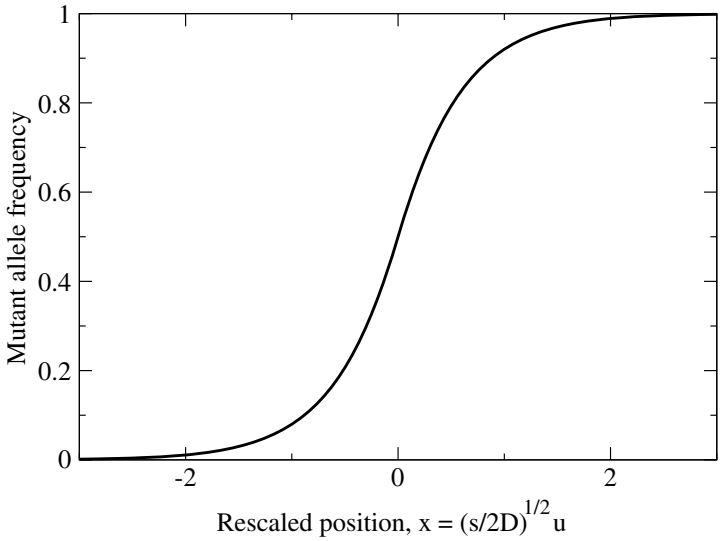

Figure 7. Variation of mutant allele frequency with position in a steady state where selection is balanced by migration. The mutant allele has a selective advantage s at positions $u > 0$, a disadvantage $-s$ at positions $u < 0$. Migration is characterized by a diffusion constant D. Although hard to discern on the plot, there is a discontinuity in the gradient of the allele frequencies at $u = 0$.

distinct regimes induced by a change in mutation rate: if the total number of mutants appearing per generation across the whole population (which is governed by the population size N multiplied by the mutation rate u) is small, fixation of a selectively advantageous allele is likely to occur before the onset of the next mutation (see e.g., [102, 103]). Thus, all individuals in the population are then likely to have the same genotype, and one which tends towards a local fitness maximum. On the other hand, if the number of mutants appearing per generation is large, one is likely to see many different genotypes in the population. In particular, one can find a distribution of genotypes centred on a fitness peak, thus forming a *quasi-species* [104, 102].

7. Conclusion

This paper has been a review of the ideas and formalism used to model stochastic processes in fields that statistical physicists are not typically acquainted with, specifically population genetics, ecology and linguistics. As a consequence, some parts of the discussion will seem familiar, other parts will not. We have tried, and we hope that we have succeeded, to explain the background ideas and motivation, since this will be the greatest obstacle to understanding among a readership of statistical physicists. On the other hand the degree of mathematical sophistication that has been assumed is greater than would be typical outside physics or mathematical biology. This makes this review quite different to others in the same area, and while we expect the readership to be mainly statistical physicists, we hope that some of those working particularly in ecology and linguistics will find our approach to their subject interesting and stimulating.

doi:10.1088/1742-5468/2007/07/P07018

J. Stat. Mech. (2007) P07018

In our discussions of the mathematical models, we have mostly used the language of population genetics, but through of the mappings discussed in section 2.5 the results obtained are more widely relevant. As the evolutionary paradigm becomes even more widely applied, there may be other areas in which analogies can be drawn. It is interesting how neutral processes turn out to have greater importance in all three areas we discussed; at the very least neutral theories can be thought of as null models, against which data and other models can be compared. Most textbooks in population genetics begin their discussion of genetic drift with the Wright–Fisher model, although for physicists the use of non-overlapping generations and a 'time' measured in number of generations will not appear so natural. The Moran model, which has exactly the same limit when the number of genes become large, is far more familiar, resembling a birth/death processes where a death is immediately followed by a birth. In addition, the continuous time limit may easily be taken, leading to a master equation of a kind well known in statistical physics.

We spent some time explaining the relationship between the discrete time approach, based on transition probabilities, and the continuous time master equation approach, based on processes occurring independently at a given rate. The use of master equations is apparently rare in the population genetics literature, which is one of the reasons why we have gone into this approach in such detail. Furthermore, this latter formalism usually turns out to be more efficient. For example, in his book Hubbell [10] formulated his model by analogy with the corresponding discrete time genetic models, and had to perform simulations to generate the stationary probability distribution. However, if the model is formulated as a master equation, the stationary probability distribution can be obtained analytically [105]. We also outlined two ways in which mutation could be introduced. In the first, birth, death and mutation events are combined; this was the original approach of Moran. In the second, the process of mutation is independent of that of birth and death. In the case where there are only two alleles, these two situations can be mapped on to each other, apart from a very mild restriction on the size of the mutation probabilities. However, when there are more than two alleles, these models are only equivalent if the mutation matrix $u_{\alpha\delta}$ is independent of the original state δ.

Another aspect of the paper, which to our knowledge has not been elucidated elsewhere in the context of these problems, is the emphasis we have put on starting with individual (or gene or token) based models and deriving the corresponding mesoscopic model, where the limit of the number of individuals, N, tends to infinity. Simulations are frequently carried out using individuals, and moreover, as is well known, an infinity of individual based models will give the same mesoscopic model in the large N limit. For example, the language evolution model which we have alluded to, was first discussed in a mesoscopic context [20], but this analysis was restricted to quite simple situations, for instance where all speakers spoke equally with each other. When more complex situations are analysed, such as allowing for a non-trivial network topology for links between speakers [24], the individual based picture becomes far more useful. In particular, the backwards-time or coalescent approach is a far more efficient way of performing simulations. In some cases, such as understanding trends in how fixation time changes with network structure, which require simulations of quite large networks, the coalescent formulation is indispensable.

doi:10.1088/1742-5468/2007/07/P07018

J. Stat. Mech. (2007) P07018

In genetics, ecology or language, just as in physics, reality cannot be described by ideal models; there will be a multitude of ways in which real systems deviate from the ideal models created by scientists when they first enter a field. One of the methods that has been devised by population geneticists to deal with this will be very familiar to physicists. This is to characterize a non-ideal system by a few parameters, which will hopefully, if chosen correctly, capture the essence of the system. It may be that a simple model can then be utilized, but with these parameters built in. An example is the effective population size, N_e, discussed in section 5.2, which reflects how the non-ideal nature of the system changes the effective value of N: the effective size of the real population being the number of individuals in the ideal population which gives the same magnitude for the quantity of interest. The effective population size and inbreeding coefficient, discussed in section 5.1, are extremely widely used, but unfortunately not in a very systematic way. We have attempted to illustrate their use, and to draw attention to some of the confusion which exists in the literature, but it is still the case that those employing these useful concepts will need to be vigilant when reading the literature on them.

This last observation opens up some questions that may be worth of further study. First one can ask whether inbreeding coefficients and the effective population size really adequately characterize the evolution of a non-ideal population. Are there other approximation schemes that might allow deviation from ideal behaviour to be described in a systematic and controlled way? Secondly, the backward-time description of neutral evolution will be recognized by non-equilibrium statistical physicists as the reaction–diffusion process $A + A \to A$. Although exact solutions for this process exist for certain mathematically convenient geometries (such as continuous real space [106] and, as we have seen, the complete graph) there is interest in extending the analysis to heterogeneous graphs, which thus far have been treated only using approximate methods [84, 85, 24]. Such solutions would be useful as they would obtain more detail about the evolution of the system, for example, the manner in which an innovation propagates through the system *en route* to fixation. There are many other questions which would benefit from a systematic study of the type that statistical physicists are well placed to carry out. We hope that this paper has stimulated some readers to find out more about these fascinating applications of stochastic methods and to contribute to the development of the field.

Acknowledgments

We wish to thank David Alonso and Alan Bray for useful discussions. RAB holds a Personal Research Fellowship awarded by the Royal Society of Edinburgh.

Appendix. The large N expansion of the master equation for M alleles

In this appendix we give details of carrying out the large N expansion beginning from the master equations involving $M > 2$ alleles, considered in sections 3.3.2 and 3.3.3.

We begin by expanding out the terms containing the transitions rates (70)—which are those involving the type M allele—letting $x_\alpha = n_\alpha/N$, and expanding all other terms in powers of $1/N$. This is the procedure used to obtain equation (46), and the analogous

expression in this case

$$
\begin{aligned}
= \sum_{\alpha=1}^{M-1} \Bigg[& \left(x_\alpha - \frac{1}{N}\right)\left(1 - \sum_{\beta=1}^{M-1} x_\beta + \frac{1}{N}\right)\left\{P - \frac{1}{N}\frac{\partial P}{\partial x_\alpha} + \frac{1}{2N^2}\frac{\partial^2 P}{\partial x_\alpha^2}\right\} \\
& - x_\alpha\left(1 - \sum_{\beta=1}^{M-1} x_\beta\right) P\Bigg] + \sum_{\alpha=1}^{M-1}\Bigg[\left(x_\alpha + \frac{1}{N}\right)\left(1 - \sum_{\beta=1}^{M-1} x_\beta - \frac{1}{N}\right) \\
& \times \left\{P + \frac{1}{N}\frac{\partial P}{\partial x_\alpha} + \frac{1}{2N^2}\frac{\partial^2 P}{\partial x_\alpha^2}\right\} - x_\alpha\left(1 - \sum_{\beta=1}^{M-1} x_\beta\right) P\Bigg] \\
= & \frac{1}{N^2}\sum_{\alpha=1}^{M-1}\Bigg[x_\alpha\left(1 - \sum_{\beta=1}^{M-1} x_\beta\right)\frac{\partial^2 P}{\partial x_\alpha^2} - 2x_\alpha\frac{\partial P}{\partial x_\alpha} \\
& + 2\left(1 - \sum_{\beta=1}^{M-1} x_\beta\right)\frac{\partial P}{\partial x_\alpha} - 2P\Bigg], \qquad \text{(A.1)}
\end{aligned}
$$

neglecting terms of order $1/N^3$ and higher. Carrying out the same procedure for the terms containing the transition rates (69) gives an expression

$$
\begin{aligned}
= \sum_{\alpha=1}^{M-1}\sum_{\beta\neq\alpha}^{M-1} \Bigg[& \left(x_\alpha - \frac{1}{N}\right)\left(x_\beta + \frac{1}{N}\right)\left\{P - \frac{1}{N}\frac{\partial P}{\partial x_\alpha} + \frac{1}{N}\frac{\partial P}{\partial x_\beta}\right. \\
& \left. + \frac{1}{2N^2}\frac{\partial^2 P}{\partial x_\alpha^2} + \frac{1}{2N^2}\frac{\partial^2 P}{\partial x_\beta^2} - \frac{1}{N^2}\frac{\partial^2 P}{\partial x_\alpha \partial x_\beta}\right\} - x_\alpha x_\beta P\Bigg] \\
= & \frac{1}{N^2}\sum_{\alpha=1}^{M-1}\sum_{\beta\neq\alpha}^{M-1}\left[x_\alpha x_\beta\frac{\partial^2 P}{\partial x_\alpha^2} - x_\alpha x_\beta\frac{\partial^2 P}{\partial x_\alpha \partial x_\beta} + 2x_\alpha\frac{\partial P}{\partial x_\beta} - 2x_\alpha\frac{\partial P}{\partial x_\alpha} - P\right], \qquad \text{(A.2)}
\end{aligned}
$$

again neglecting terms of order $1/N^3$ and higher. Adding equations (A.1) and (A.2) gives for the master equation

$$
\begin{aligned}
\frac{\partial P}{\partial t} = \frac{1}{N^2}\Bigg[& -M(M-1)P + 2\sum_{\alpha=1}^{M-1}\left\{(1 - Mx_\alpha)\frac{\partial P}{\partial x_\alpha}\right\} \\
& + \sum_{\alpha=1}^{M-1} x_\alpha(1 - x_\alpha)\frac{\partial^2 P}{\partial x_\alpha^2} - \sum_{\alpha=1}^{M-1}\sum_{\beta\neq\alpha}^{M-1} x_\alpha x_\beta\frac{\partial^2 P}{\partial x_\alpha \partial x_\beta}\Bigg] + \mathcal{O}\left(\frac{1}{N^3}\right) \\
= & \frac{1}{N^2}\left\{\sum_{\alpha=1}^{M-1}\frac{\partial^2}{\partial x_\alpha^2}\left[x_\alpha(1 - x_\alpha)P\right] - \sum_{\alpha=1}^{M-1}\sum_{\beta\neq\alpha}^{M-1}\frac{\partial^2}{\partial x_\alpha \partial x_\beta}\left[x_\alpha x_\beta P\right]\right\}, \qquad \text{(A.3)}
\end{aligned}
$$

up to terms of order $1/N^3$ and higher. As in previous cases, if we define a new time $\tau = 2t/N^2$, and take $N \to \infty$, then we find the result equation (71).

In the main text we discussed two different ways of including mutations. In the first type of model, mutation rates were given by equation (76). This equation may be written as

$$
T(n_1 \ldots n_\alpha + 1 \ldots n_\beta - 1 \ldots n_{M-1}|\underline{n}) = (1 - \rho_\alpha)\frac{n_\alpha}{N}\frac{n_\beta}{N} + \sigma_\alpha(\underline{n}/N)\frac{n_\beta}{N}, \qquad \text{(A.4)}
$$

where

$$\rho_\alpha = \sum_{\gamma\neq\alpha} u_{\gamma\alpha} \qquad \text{and} \qquad \sigma_\alpha(\underline{x}) = \sum_{\delta\neq\alpha} u_{\alpha\delta} x_\delta. \tag{A.5}$$

Suppose that we first look at the case where $\alpha, \beta = 1, \ldots, M-1$ in the transition rate (A.4). This gives the following contribution to the right-hand side of the Fokker–Planck equation:

$$\sum_{\alpha=1}^{M-1}\sum_{\beta\neq\alpha}^{M-1}\left[(1-\rho_\alpha)\left(x_\alpha - \frac{1}{N}\right) + \sigma_\alpha(\underline{x})\right]\left[x_\beta + \frac{1}{N}\right]\left(P + \frac{1}{N}\frac{\partial P}{\partial x_\beta} - \frac{1}{N}\frac{\partial P}{\partial x_\alpha}\right)$$
$$- \sum_{\alpha=1}^{M-1}\sum_{\beta\neq\alpha}^{M-1}\left[(1-\rho_\alpha)\,x_\alpha + \sigma_\alpha(\underline{x})\right]x_\beta P + \mathcal{O}\left(\frac{1}{N^2}\right).$$

The terms involving only the mutation rates are

$$\frac{1}{N}\sum_{\alpha=1}^{M-1}\sum_{\beta\neq\alpha}^{M-1}\left[-\rho_\alpha x_\alpha + \rho_\alpha x_\beta + \sigma_\alpha(\underline{x})\right]P$$
$$+ \frac{1}{N}\sum_{\alpha=1}^{M-1}\sum_{\beta\neq\alpha}^{M-1}\left[-\rho_\alpha x_\alpha x_\beta + \sigma_\alpha(\underline{x})x_\beta\right]\left\{\frac{\partial P}{\partial x_\beta} - \frac{\partial P}{\partial x_\alpha}\right\} + \mathcal{O}\left(\frac{1}{N^2}\right). \tag{A.6}$$

The analogous expression when $\alpha = 1, \ldots, M-1$ and $\beta = M$ is:

$$\sum_{\alpha=1}^{M-1}\left[(1-\rho_\alpha)\left(x_\alpha - \frac{1}{N}\right) + \sigma_\alpha(\underline{x})\right]\left[1 - \sum_{\gamma=1}^{M-1} x_\gamma + \frac{1}{N}\right]\left(P - \frac{1}{N}\frac{\partial P}{\partial x_\alpha}\right)$$
$$- \sum_{\alpha=1}^{M-1}\left[(1-\rho_\alpha)\,x_\alpha + \sigma_\alpha(\underline{x})\right]\left[1 - \sum_{\gamma=1}^{M-1} x_\gamma\right]P + \mathcal{O}\left(\frac{1}{N^2}\right),$$

and when $\beta = 1, \ldots, M-1$ and $\alpha = M$ is:

$$\sum_{\beta=1}^{M-1}\left[(1-\rho_M)\left(1 - \sum_{\gamma=1}^{M-1} x_\gamma - \frac{1}{N}\right) + \sigma_M(\underline{x})\right]\left[x_\beta + \frac{1}{N}\right]\left(P + \frac{1}{N}\frac{\partial P}{\partial x_\beta}\right)$$
$$- \sum_{\beta=1}^{M-1}\left\{(1-\rho_M)\left(1 - \sum_{\gamma=1}^{M-1} x_\gamma\right) + \sigma_M(\underline{x})\right\}x_\beta P + \mathcal{O}\left(\frac{1}{N^2}\right),$$

which lead to the analogous expressions to equation (A.6):

$$\frac{1}{N}\sum_{\alpha=1}^{M-1}\left[-\rho_\alpha x_\alpha + \rho_\alpha\left(1 - \sum_{\gamma=1}^{M-1} x_\gamma\right) + \sigma_\alpha(\underline{x})\right]P$$
$$- \frac{1}{N}\sum_{\alpha=1}^{M-1}\left[-\rho_\alpha x_\alpha\left(1 - \sum_{\gamma=1}^{M-1} x_\gamma\right) + \sigma_\alpha(\underline{x})\left(1 - \sum_{\gamma=1}^{M-1} x_\gamma\right)\right]\frac{\partial P}{\partial x_\alpha} + \mathcal{O}\left(\frac{1}{N^2}\right), \tag{A.7}$$

and

$$\frac{1}{N}\sum_{\beta=1}^{M-1}\left[\rho_M x_\beta - \rho_M\left(1-\sum_{\gamma=1}^{M-1}x_\gamma\right)+\sigma_M(\underline{x})\right]P + \frac{1}{N}\sum_{\beta=1}^{M-1}\left[-\rho_M\left(1-\sum_{\gamma=1}^{M-1}x_\gamma\right)+\sigma_\alpha(\underline{x})\right]x_\beta\frac{\partial P}{\partial x_\beta}+\mathcal{O}\left(\frac{1}{N^2}\right). \tag{A.8}$$

From the expressions (A.6)–(A.8), the term involving mutations in the Fokker–Planck equation can be found. For example, let us focus on the mutation rates involving A_α and A_β with $\alpha,\beta=1,\ldots,M-1$ and $\alpha\neq\beta$. After some straightforward algebra, the terms from equations (A.6) and (A.7) which involve P, but not derivatives of P are

$$\frac{1}{N}\sum_{\alpha=1}^{M-1}\left[\rho_\alpha\left(1-Mx_\alpha\right)+M\sigma_\alpha(\underline{x})\right]P=\frac{1}{N}\sum_{\alpha=1}^{M-1}\sum_{\gamma\neq\alpha}u_{\gamma\alpha}P, \tag{A.9}$$

using equation (A.5). The terms from equations (A.6) and (A.7) which involve derivatives of P are more complicated, but eventually yield

$$\frac{1}{N}\sum_{\alpha=1}^{M-1}\left[\rho_\alpha x_\alpha-\sigma_\alpha(\underline{x})\right]\frac{\partial P}{\partial x_\alpha}=\frac{1}{N}\sum_{\alpha=1}^{M-1}\left[\sum_{\gamma\neq\alpha}u_{\gamma\alpha}x_\alpha-\sum_{\delta\neq\alpha}u_{\alpha\delta}x_\delta\right]\frac{\partial P}{\partial x_\alpha}. \tag{A.10}$$

We may combine equations (A.9) and (A.10) to give

$$-\frac{1}{N}\sum_{\alpha=1}^{M-1}\frac{\partial}{\partial x_\alpha}\left[\sum_{\delta\neq\alpha}u_{\alpha\delta}x_\delta-\sum_{\gamma\neq\alpha}u_{\gamma\alpha}x_\alpha\right]P+\mathcal{O}\left(\frac{1}{N^2}\right). \tag{A.11}$$

Introducing the scaled mutation rates defined in equation (77), this expression equals

$$-\frac{2}{N^2}\sum_{\alpha=1}^{M-1}\frac{\partial}{\partial x_\alpha}\left[\sum_{\delta\neq\alpha}\mathcal{U}_{\alpha\delta}x_\delta-\sum_{\gamma\neq\alpha}\mathcal{U}_{\gamma\alpha}x_\alpha\right]P+\mathcal{O}\left(\frac{1}{N^3}\right). \tag{A.12}$$

Using the rescaled time $\tau=2t/N^2$ and letting $N\to\infty$, we see that we obtain the mutation terms displayed in equations (78) and equations (79), at least in the sector where $\alpha,\beta=1,\ldots,M-1$. In a similar way, the mutation terms where the allele A_M is involved may be derived from equations (A.7) and (A.8). One again finds the result given by equations (78) and (79).

The form of the mutation terms in the Fokker–Planck equation in the second scheme we investigated, defined by the transition rates (83), follows from the results obtained above for the first scheme. This is because if one takes $u_{\alpha\beta}=u_\alpha$ for all $\beta\neq\alpha$, then the transition rates of the two approaches are identical, although in the second case there is the restriction that $\sum_{\delta=1}^{M}u_\delta<1$. We therefore find the same Fokker–Planck equation (78), but now with equation (79) taking the form

$$\mathcal{A}_\alpha(\underline{x})=\sum_{\beta=1}^{M}\left(\mathcal{U}_\alpha x_\beta-\mathcal{U}_\beta x_\alpha\right)=\mathcal{U}_\alpha-\sum_{\beta=1}^{M}\mathcal{U}_\beta x_\alpha, \tag{A.13}$$

which is equation (85). When written out in terms of the scaled mutation rates, the condition on the u_δ in the second case reads

$$\sum_{\delta=1}^{M} \mathcal{U} < \frac{N}{2}, \tag{A.14}$$

which, since the rescaled rates are numerically small and we have taken the limit $N \to \infty$, is always satisfied.

Finally, we consider the most general case where there are L islands and M alleles. As we remarked in the main text, this involves only minor modifications of calculations which have been carried out previously. This is because, since both the mutation and migration rates are scaled by a factor of N, any product of these rates will give a term of order $1/N^3$, and so will not contribute to the Fokker–Planck equation. Therefore we may consider mutation and migration separately, and simply need to insert appropriate indices on the contributions already calculated.

First let, us consider equation (88), which does not include mutation. If we begin by ignoring the migration terms g_{ij} with $i \neq j$, then we see that the transition rates will be of the form given in equations (69) and (70), but multiplied by f_i and with an index i included. Thus the result of a $1/N$ expansion is as in equation (A.3), but with an index i and f_i included:

$$\frac{f_i}{N^2}\left\{\sum_{\alpha=1}^{M-1}\frac{\partial^2}{\partial x_{i\alpha}^2}\left[x_{i\alpha}\left(1-x_{i\alpha}\right)P\right]-\sum_{\alpha=1}^{M-1}\sum_{\beta\neq\alpha}^{M-1}\frac{\partial^2}{\partial x_{i\alpha}\partial x_{i\beta}}\left[x_{i\alpha}x_{i\beta}P\right]\right\}, \tag{A.15}$$

up to terms of order $1/N^3$ and higher. This has to be summed over all i.

Returning to equation (88), the transition rates which involve migration can be found by analogy with the discussion in section 3.3.1, and in particular equation (60), with α and β indices for the allele labels. These appear as in equations (A.1) and (A.2). The result is:

$$\begin{aligned}
G_{ij}\sum_{\alpha=1}^{M-1}\sum_{\beta\neq\alpha}^{M-1}&\left[\left(x_{i\beta}+\frac{1}{N}\right)x_{j\alpha}\left\{P-\frac{1}{N}\frac{\partial P}{\partial x_{i\alpha}}+\frac{1}{N}\frac{\partial P}{\partial x_{i\beta}}\right\}-x_{i\beta}x_{j\alpha}P\right]\\
&+G_{ij}\sum_{\alpha=1}^{M-1}\left[\left(1-\sum_{\gamma=1}^{M-1}x_{i\gamma}+\frac{1}{N}\right)x_{j\alpha}\left\{P-\frac{1}{N}\frac{\partial P}{\partial x_{i\alpha}}\right\}\right.\\
&\left.-\left(1-\sum_{\gamma=1}^{M-1}x_{i\gamma}\right)x_{j\alpha}P\right]+G_{ij}\sum_{\beta=1}^{M-1}\left[\left(x_{i\beta}+\frac{1}{N}\right)\left(1-\sum_{\gamma=1}^{M-1}x_{j\gamma}\right)\right.\\
&\left.\times\left\{P+\frac{1}{N}\frac{\partial P}{\partial x_{i\beta}}\right\}-x_{i\beta}\left(1-\sum_{\gamma=1}^{M-1}x_{j\gamma}\right)P\right]+\mathcal{O}\left(\frac{1}{N^2}\right)\\
&=G_{ij}\frac{1}{N}\left[(M-1)P+\sum_{\alpha=1}^{M-1}\left(x_{i\alpha}-x_{j\alpha}\right)\frac{\partial P}{\partial x_{i\alpha}}\right]+\mathcal{O}\left(\frac{1}{N^2}\right)\\
&=G_{ij}\frac{1}{N}\sum_{\alpha=1}^{M-1}\frac{\partial}{\partial x_{i\alpha}}\left[\left(x_{i\alpha}-x_{j\alpha}\right)P\right]+\mathcal{O}\left(\frac{1}{N^2}\right).
\end{aligned} \tag{A.16}$$

Scaling the migration rate by introducing $\mathcal{G}_{ij} = NG_{ij}/2$, adding an analogous term which represents migration from island i to island j, and summing over all pairs of islands $\langle ij \rangle$ gives

$$\frac{2}{N^2} \sum_{\langle ij \rangle} \sum_{\alpha=1}^{M-1} \left(\mathcal{G}_{ij} \frac{\partial}{\partial x_{i\alpha}} - \mathcal{G}_{ji} \frac{\partial}{\partial x_{j\alpha}} \right) \left[(x_{i\alpha} - x_{j\alpha})\, P \right] + \mathcal{O}\left(\frac{1}{N^3} \right). \tag{A.17}$$

The third and final term is found from equation (89) ignoring the migration terms g_{ij} with $i \neq j$. This is as in equation (76), but multiplied by f_i and with an index i added to the n_β and n_δ. From the discussion leading to equation (A.12) we find that the contribution for the first type of mutation is

$$-\frac{2}{N^2} f_i \sum_{\alpha=1}^{M-1} \frac{\partial}{\partial x_{i\alpha}} \left[\mathcal{A}_\alpha(\underline{x}_i) P \right] + \mathcal{O}\left(\frac{1}{N^3} \right). \tag{A.18}$$

This has to be summed over all i. The second type of mutation gives the same result, but with $\mathcal{A}_\alpha(\underline{x}_i)$ defined as in equation (85).

Putting equations (A.15), (A.17) and (A.18) together gives equation (90), a general Fokker–Planck equation describing a population of genes, of M possible types, subdivided between L different islands.

References

[1] Peliti L, *Introduction to the statistical theory of Darwinian evolution*, 1997 *Preprint* cond-mat/9712027
[2] Baake E and Gabriel W, *Biological evolution through mutation, selection, and drift: an introductory review*, 2000 *Annual Reviews of Computational Physics* vol 7, ed D Stauffer (Singapore: World Scientific) pp 203–64
[3] Drossel B, *Biological evolution and statistical physics*, 2001 *Adv. Phys.* **50** 209
[4] Fisher R A, 1930 *The Genetical Theory of Natural Selection* (Oxford: Clarendon)
[5] Wright S, *Evolution in Mendelian populations*, 1931 *Genetics* **16** 97
[6] Kimura M, *Evolutionary rate at the molecular level*, 1968 *Nature* **217** 624
[7] Kimura M, 1983 *The Neutral Theory of Molecular Evolution* (Cambridge: Cambridge University Press)
[8] Ohta T, *The nearly neutral theory of molecular evolution*, 1992 *Ann. Rev. Ecol. Syst.* **23** 263
[9] Ohta T, *Near-neutrality in evolution of genes and gene-regulation*, 2002 *Proc. Natl Acad. Sci. USA* **99** 16134
[10] Hubbell S P, 2001 *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton, NJ: Princeton University Press)
[11] Condit R *et al*, *Beta-diversity in tropical forest trees*, 2002 *Science* **295** 666
[12] Harte J, *Tail of death and resurrection*, 2003 *Nature* **424** 1006
[13] McGill B J, *A test of the unified neutral theory of biodiversity*, 2003 *Nature* **422** 881
[14] McGill B J, *Empirical evaluation of neutral theory*, 2006 *Ecology* **87** 1411
[15] Hull D L, 1988 *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science* (Chicago: University of Chicago Press)
[16] Hull D L, 2001 *Science and Selection: Essays on Biological Evolution and the Philosophy of Science* (Cambridge: Cambridge University Press)
[17] Croft W A, 2000 *Explaining Language Change: an Evolutionary Approach* (Harlow: Longman)
[18] Croft W, *The Darwinization of linguistics*, 2002 *Selection* **3** 75
[19] Croft W, *The relevance of an evolutionary model to historical linguistics*, 2006 *Competing Models of Linguistic Change: Evolution and Beyond* ed O N Thomsen (Amsterdam: John Benjamins) pp 91–132
[20] Baxter G J, Blythe R A, Croft W and McKane A J, *Utterance selection model of language change*, 2006 *Phys. Rev.* E **73** 046118
[21] Tomasello M, 2003 *Constructing a Language: a Usage-based Theory of Language Acquisition* (Cambridge, MA: Harvard University Press)
[22] Gordon E, Campbell L, Hey J, MacLagan M, Sudbury A and Trudgill P, 2004 *New Zealand English: Its Origins and Evolution* (Cambridge: Cambridge University Press)

J. Stat. Mech. (2007) P07018

[23] Trudgill P, 2004 *New-Dialect Formation: The Inevitability of Colonial Englishes* (Edinburgh: Edinburgh University Press)
[24] Blythe R A, *The propagation of a cultural or biological trait by neutral genetic drift in a subdivided population*, 2007 *Theor. Popul. Biol.* **71** 454
[25] Hartl D L and Clark A G, 1989 *Principles of Population Genetics* 2nd edn (Sunderland, MA: Sinauer Associates)
[26] Crow J F and Kimura M, 1970 *An Introduction to Population Genetics Theory* (New York: Harper and Row)
[27] Ewens W J, 2004 *Mathematical Population Genetics* 2nd edn (Berlin: Springer)
[28] Maruyama T, 1977 *Stochastic Problems in Population Genetics* (Berlin: Springer)
[29] Feller W, *Diffusion processes in genetics*, 1951 *Proc. 2nd Berkeley Symp. on Mathematical Statistics and Probability* ed J Neyman (Berkeley, CA: University of California Press)
[30] van Kampen N G, 1992 *Stochastic Processes in Physics and Chemistry* 2nd edn (Amsterdam: North-Holland)
[31] Moran P A P, *Random processes in genetics*, 1958 *Proc. Camb. Phil. Soc.* **54** 60
[32] Moran P A P, 1962 *The Statistical Processes of Evolutionary Theory* (Oxford: Clarendon)
[33] Karlin S and McGregor J, *On a genetics model of Moran*, 1962 *Proc. Camb. Phil. Soc.* **58** 299
[34] Karlin S and McGregor J, *On some stochastic models in genetics*, 1964 *Stochastic Problems in Medicine and Biology* ed J Gurland (Madison, WI: University of Wisconsin Press) pp 245–71
[35] Kimura M and Weiss G H, *The stepping stone model of population structure and the decrease of genetic correlation with distance*, 1964 *Genetics* **49** 461
[36] Malécot G, *Quelques schemas probabilistes sur la variabilité des populations naturelles*, 1950 *Ann. Univ. Lyon Sci.* A **13** 37
[37] Malécot G, *Migration et parenté génétique moyenne*, 1962 *Entretiens de Monaco en Sciences Humaines* 205
[38] Moran P A P, *The theory of some genetical effects of population subdivision*, 1959 *Aust. J. Biol. Sci.* **12** 109
[39] Alonso D, Etienne R S and McKane A J, *The merits of neutral theory*, 2006 *Trends Ecol. Evol.* **21** 451
[40] Fisher R A, *On the dominance ratio*, 1922 *Proc. R. Soc. Edinburgh* **42** 321
[41] Wright S, *The differential equation of the distribution of gene frequencies*, 1945 *Proc. Nat. Acad. Sci.* **31** 382
[42] Kimura M, *Stochastic processes and distribution of gene frequencies under natural selection*, 1955 *Cold Spring Harbour Symp. Quant. Biol.* **20** 33
[43] Gardiner C W, 2004 *Handbook of Stochastic Processes* (Berlin: Springer)
[44] Risken H, 1989 *The Fokker–Planck Equation* (Berlin: Springer)
[45] Aalto E, *The Moran model and validity of the diffusion approximation in population genetics*, 1989 *J. Theor. Biol.* **140** 317
[46] Crow J F and Kimura M, *Some genetic problems in natural populations*, 1956 *Proc. 3rd Berkeley Symp. on Mathematical Statistics and Probability* ed J Neyman (Berkeley, CA: University of California Press) pp 1–22
[47] Watterson G A, *Heterosis or neutrality?*, 1977 *Genetics* **85** 789
[48] Griffiths R C, *A transition density expansion for a multi-allele diffusion model*, 1979 *Adv. Appl. Probab.* **11** 310
[49] Baxter G J, Blythe R A and McKane A J, *Exact solution of the multi-allelic diffusion model*, 2007 *Math. Biosci.* at press
[50] Kingman J F C, *The genealogy of large populations*, 1982 *J. Appl. Probab.* **19: Essays** 27
[51] Tavaré S, *Line-of-descent and genealogical processes, and their applications in population genetics models*, 1984 *Theor. Popul. Biol.* **26** 119
[52] Griffiths R C, *Asymptotic line-of-descent distributions*, 1984 *J. Math. Biol.* **17** 37
[53] Holder M and Lewis P O, *Phylogeny estimation: traditional and Bayesian approaches*, 2003 *Nat. Rev. Genet.* **4** 275
[54] Möhle M, *Forward and backward diffusion approximations for haploid exchangeable population models*, 2001 *Stoch. Process. Appl.* **95** 133
[55] Möhle M, *Ancestral processes in population genetics—the coalescent*, 2000 *J. Math. Biol.* **204** 629
[56] Wright S, *Isolation by distance*, 1943 *Genetics* **28** 114
[57] Maruyama T, *On the fixation probability of mutant genes in a subdivided population*, 1970 *Genet. Res. Camb.* **15** 221
[58] Latter B D H, *The island model of population differentiation: a general solution*, 1973 *Genetics* **73** 147

J. Stat. Mech. (2007) P07018

[59] Wright S, *The genetical structure of populations*, 1951 *Ann. Eugenics* **15** 323
[60] Neigel J E, *A comparison of alternative strategies for estimating gene flow from genetic markers*, 1997 *Annu. Rev. Ecol. Syst.* **28** 105
[61] Nagylaki T, *Fixation indices in subdivided populations*, 1998 *Genetics* **148** 1325
[62] Wilkinson-Herbots H M, *Genealogy and subpopulation differentiation under various models of population structure*, 1998 *J. Math. Biol.* **37** 535
[63] Nei M, *Analysis of gene diversity in subdivided populations*, 1973 *Proc. Nat. Acad. Sci.* **70** 3321
[64] Whitlock M C and McCauley D E, *Indirect measures of gene flow and migration:* $F_{ST} \neq 1/(4Nm+1)$, 1999 *Heredity* **82** 117
[65] Slatkin M, *Inbreeding coefficients and coalescence times*, 1991 *Genet. Res. Camb.* **58** 167
[66] Nordborg M, *Coalescent theory*, 2003 *Handbook of Statistical Genetics* 2nd edn, ed D J Balding, M Bishop and C Cannings (New York: Wiley)
[67] Strobeck C, *Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision*, 1987 *Genetics* **117** 149
[68] Slatkin M and Voelm L, F_{ST} *in a hierarchical island model*, 1991 *Genetics* **127** 627
[69] Wakeley J and Aliacar N, *Gene genealogies in a metapopulation*, 2001 *Genetics* **159** 893
[70] Wakeley J and Takahashi T, *The many-demes limit for selection and drift in a subdivided population*, 2004 *Theor. Popul. Biol.* **66** 83
[71] Wakeley J and Lessard S, *Corridors for migration between large subdivided populations, and the structured coalescent*, 2006 *Theor. Popul. Biol.* **70** 412
[72] Wang J and Caballero A, *Developments in predicting the effective size of subdivided populations*, 1999 *Heredity* **82** 212
[73] Rousset F, 2004 *Genetic Structure and Selection in Subdivided Populations* (Oxford: Princeton University Press)
[74] Roze D and Rousset R, *Selection and drift in subdivided populations: a straightforward method for deriving diffusion approximations and applications involving dominance, selfing and local extinctions*, 2003 *Genetics* **165** 2153
[75] Rousset F, *Effective size in simple metapopulation models*, 2003 *Heredity* **91** 107
[76] Cherry J L and Wakeley J, *A diffusion approximation for selection and drift in a subdivided population*, 2003 *Genetics* **163** 421
[77] Whitlock M C and Barton N H, *The effective size of a subdivided population*, 1997 *Genetics* **146** 427
[78] Pollak E, *On the theory of partially inbreeding finite populations. I. Partial selfing*, 1987 *Genetics* **117** 353
[79] Nordborg M and Donnelly P, *The coalescent process with selfing*, 1997 *Genetics* **146** 1185
[80] Nagylaki T, *The strong migration limit in geographically structured populations*, 1980 *J. Math. Biol.* **9** 101
[81] Notohara M, *The coalescent and the genealogical process in geographically structured population*, 1990 *J. Math. Biol.* **29** 59
[82] Nordborg M and Krone S M, *Separation of time scales and convergence to the coalescent in structured populations*, 2002 *Modern Developments in Theoretical Population Genetics: The Legacy of Gustave Malécot* ed M Slatkin and M Veuille (Oxford: Oxford University Press) pp 194–232
[83] Sjödin P, Kaj I, Krone S, Lascoux M and Nordborg M, *On the meaning and existence of an effective population size*, 2005 *Genetics* **169** 1061
[84] Sood V and Redner S, *Voter model on heterogeneous graphs*, 2005 *Phys. Rev. Lett.* **94** 178701
[85] Antal T, Redner S and Sood V, *Evolutionary dynamics on degree-heterogeneous graphs*, 2006 *Phys. Rev. Lett.* **96** 188104
[86] Durrett R and Schweinsberg J, *Approximating selective sweeps*, 2004 *Theor. Popul. Biol.* **66** 129
[87] Kimura M, *On the probability of fixation of mutant genes in a population*, 1962 *Genetics* **47** 713
[88] Kaplan N L, Darden T and Hudson R R, *The coalescent process in models with selection*, 1988 *Genetics* **120** 819
[89] Durrett R and Schweinsberg J, *A coalescent model for the effect of advantageous mutations on the genealogy of a population*, 2005 *Stoch. Process. Appl.* **115** 1628
[90] Pitman J, *Coalescents with multiple collisions*, 1999 *Ann. Probab.* **27** 1870
[91] Brunet E, Derrida B, Mueller A H and Munier S, *Noisy traveling waves: effect of selection on genealogies*, 2006 *Europhys. Lett.* **76** 1
[92] Slatkin M, *Fixation probabilities and fixation times in a subdivided population*, 1981 *Evolution* **35** 477
[93] Barton N H, *The probability of fixation of a favoured allele in a subdivided population*, 1993 *Genet. Res. Camb.* **62** 149
[94] Lieberman E, Hauert C and Nowak M A, *Evolutionary dynamics on graphs*, 2005 *Nature* **433** 312
[95] Fisher R A, *The wave of advance of advantageous genes*, 1937 *Ann. Eugenics* **7** 335

J. Stat. Mech. (2007) P07018

[96] Kolmogorov A N, Petrovsky I G and Piskunov N S, *A study of the diffusion equation with increase in the amount of substance, and its application to a biological problem*, 1937 *Bull. Moscow Univ., Math. Mech.* **1** 1
[97] van Saarloos W, *Front propagation into unstable states*, 2003 *Phys. Rep.* **386** 29
[98] Brunet E and Derrida B, *Effect of microscopic noise on front propagation*, 2001 *J. Stat. Phys.* **103** 269
[99] Haldane J B S, *The theory of a cline*, 1948 *J. Genet.* **48** 277
[100] Huxley J S, *Clines: an auxiliary taxonomic principle*, 1938 *Nature* **142** 219
[101] Whitlock M C, Phillips P C, Moore F B G and Tonsor S J, *Multiple fitness peaks and epistasis*, 1995 *Annu. Rev. Ecol. Syst.* **26** 601
[102] Nowak M A, *What is a quasispecies?*, 1992 *Trends Ecol. Evol.* **7** 118
[103] Jain K and Krug J, *Deterministic and stochastic regimes of asexual evolution on rugged fitness landscapes*, 2007 *Genetics* **175** 1275
[104] Eigen M and Schuster P, *Emergence of the hypercycle*, 1977 *Naturwissenschaften* **64** 545
[105] McKane A J, Alonso D and Solé R V, *Analytic solution of Hubbell's model of local community dynamics*, 2004 *Theor. Popul. Biol.* **65** 67
[106] Peliti L, *Renormalisation of fluctuation effects in the $A + A \to A$ reaction*, 1986 *J. Phys. A: Math. Gen.* **19** L365

J. Stat. Mech. (2007) P07018