Modelling evolution in a spatial continuum

N H Barton

Institute of Science and Technology
Am Campus I
A-3400 Klosterneuberg
Austria

E-mail: nick.barton@ist-austria.ac.at

A M Etheridge

Department of Statistics
University of Oxford
1 South Parks Road
Oxford OX1 3TG
UK

E-mail: etheridg@stats.ox.ac.uk

A Véber

Centre de Mathématiques Appliquées
École Polytechnique
Route de Saclay
91128 Palaiseau Cedex
France

E-mail: amandine.veber@cmap.polytechnique.fr

Abstract. We survey a class of models for spatially structured populations which we have called spatial Λ-Fleming-Viot processes. They arise from a flexible framework for modelling in which the key innovation is that random genetic drift is driven by a Poisson Point Process of spatial 'events'. We demonstrate how this overcomes some of the obstructions to modelling populations which evolve in two (and higher) dimensional spatial continua, how its predictions match phenomena observed in data, and how it fits with classical models. Finally we outline some directions for future research.

1. Introduction

Almost all naturally occurring populations contain abundant genetic variation. This variation, generated by mutation and recombination, is then modified through the evolutionary processes of genetic drift, the movement of genes from place to place, and natural selection. A fundamental goal of theoretical population genetics is to understand the interactions between these evolutionary processes and their relative importance in shaping the patterns of genetic variation that we see in the world around us.

A natural starting point is a 'forwards in time' model for the way in which frequencies of different genetic types (*allele* frequencies) change with time. Such a model should then be compared to data. This generally takes the form of DNA sequences taken from individuals sampled from the population. These sequences are then used to infer the *genealogical relationships* between the individuals in the sample. For example they may be used to infer the tree describing the shared ancestry of a particular gene from the differences in the DNA sequence corresponding to that gene in different individuals in the sample. Thus, in order to compare the predictions of a mathematical model to data, one also needs a 'backwards in time' description of the genealogical trees relating individuals in a sample from a population evolving under that model. The resulting models are collectively known as *coalescents*.

The outstanding success in this area has been Kingman's coalescent which provides a description of the genealogical trees relating individuals in a sample from a highly idealised population (see §2.1). The resulting coalescent model is parametrised by a single number, the total population size, N. Although the assumptions of the population models of §2.1 are far too stringent to be satisfied by any real population (even in a laboratory), the power of the Kingman coalescent stems from that fact that if one replaces the census population size, N, by an *effective* population size, N_e, then the Kingman coalescent can be applied in an enormous variety of situations. Provided the sample is taken from sufficiently well-separated individuals, this approach even works when the population is spatially structured (see §4.1). Thus the effective population size is somehow capturing the effects of population structure, natural selection, variable population size and so on. However, it is important to note that the effective population size is typically orders of magnitude different from census population size. For example, for humans, while the census population size now stands at 7×10^9, the *effective* population size is $\mathcal{O}(10^4)$ ([1, 2]). We would like to understand how the different processes of evolution are feeding into this number. Moreover, the approach assumes that we are sampling 'uniformly from the whole population' and does not capture any information about spatial patterns of genetic variation. In order to address these issues we must explicitly incorporate spatial structure into our models.

The purpose of this article is to survey a framework for modelling spatially structured populations which was introduced, and for which some preliminary analyses were carried out, in a series of recent papers [3, 4, 5, 6, 7, 8]. The main innovation of this framework is the approach to modelling genetic drift, which, as we shall see, overcomes

obstacles that appear in the classical approach. This leads to a class of measure-valued processes which we believe to be of independent mathematical interest. The framework is very versatile: we can approximate the 'discrete deme' models of spatially structured populations from classical population genetics (§3.6), but it goes beyond that in incorporating the large-scale extinction/recolonisation events that dominate the demographic histories of many species (§3.1). In providing mathematically well-defined and consistent models for the (forwards in time) dynamics of allele frequencies and the (backwards in time) genealogical trees relating individuals in a sample from the population, it offers us a powerful toolbox with which to derive large scale results that do not depend upon the details of the construction.

It is useful to place this work in its historical context and so we begin with a brief description of some of the more widely studied models of population genetics. For more thorough mathematical introductions see, for example, [9, 10, 11] and for more of the biologicial background see [12]. Rather than presenting formal proofs, the core of the article (§3-§5) aims to motivate our approach and explain why results are true. For technical proofs the reader is referred to the original articles. Although, in order to simplify the exposition we have, in several places, considered special cases of our rigorous results, for ease of reference we have aimed as far as possible to use notation consistent with the original work. We shall also describe some more recent work which incorporates natural selection into our framework and explain how this approach is connected to established models. In §6 we briefly mention some related models before, finally, in §7 outlining some directions of future research.

Before embarking on this, let us make some remarks about our assumptions. If we trace the ancestry of a sample of chromosomes from an asexually reproducing population, each chromosome can be identified with a unique parent and so as we trace backwards in time ancestral lineages can only merge (when two chromosomes have a common parent) and not branch. As a result the ancestry is encoded in a tree. If chromosomes in sexually reproducing populations were passed down from parent to offspring as indivisible blocks, then the same would be true. However, as a result of *recombination*, the chromosome that a human mother, for example, passes to her child is not an exact copy of one of her chromosomes, but rather a mosaïc of complementary blocks taken from the pair of chromosomes that she carries. Thus, if we wish to trace the ancestry of a sample from a sexually reproducing population, then for each individual in the sample, we must trace back its parents, grandparents, great grandparents and so on. Eventually, since population sizes are finite, we will see individuals appear at multiple points in the resulting pedigree (even for a sample of size one). Thus the ancestral relationships between individuals will form a complex branching and coalescing web. Nonetheless, if one is interested, not in whole chromosomes, but in sufficiently small blocks of genome, it is reasonable to ignore recombination within the blocks and treat them as indivisible. This greatly simplifies the mathematical analysis. To avoid unnecessary complication, for (most of) the remainder of this article we shall do this. However, in §4.3 we comment briefly on how our approach can be adapted to incorporate

recombination and the implications for statistical inference.

Let us close this section with some terminology. When we deal with populations subject to recombination, two genetic loci are said to be *tightly linked* if they are sufficiently close together that the possibility of recombination between them can be ignored. Two loci are *loosely linked* if recombination events between them are sufficiently frequent that types at the two loci are inherited essentially independently (for example if they are on different chromosomes). Finally, most populations are either *haploid*, meaning that individuals carry one copy of each chromosome, or *diploid*, meaning that chromosomes are carried in pairs.

2. Some history

2.1. Drift

To establish terminology and fix ideas, in this subsection we recall two classical approaches to modelling genetic drift. Our spatial model borrows ideas from both.

The first model was developed independently by R.A. Fisher and S. Wright (see [9] for some history). It considers an (idealised) unstructured population, of constant (large) size, N, in which all individuals are equally fit.

Definition 2.1 (The neutral Wright-Fisher model) *Under the neutral Wright-Fisher model for a haploid population of constant size N, the population evolves in discrete generations. The number of offspring (in generation $t+1$) of each of the individuals in generation t is determined by multinomial sampling with equal weights. In other words, each individual in generation $t+1$ chooses its parent independently at random from those present in generation t.*

It is a simple matter to describe the genealogical trees relating individuals in a sample from a population evolving according to this model. The probability that two individuals share a common parent in the previous generation is $1/N$ and so, for a sample of size two, the time back to their most recent common ancestor (MRCA) has a geometric distribution with parameter $1/N$. Since this has mean N, we immediately see that the appropriate timescale for evolution is units of N generations. In these time units, the time to the MRCA is approximately exponentially distributed with parameter 1. The chance that three or more individuals have a common parent is $\mathcal{O}(1/N^2)$, and, similarly, the chance that two distinct pairs of ancestral lineages merge in a single generation is $\mathcal{O}(1/N^2)$. Thus for large populations, with high probability, all lineages ancestral to a finite sample will merge (pairwise) without us seeing any such events. Putting this together, for a sufficiently large population, the ancestry of a sample of size k is described as follows: the time we must trace back until we see the first merger of ancestral lineages is exponential with rate $\binom{k}{2}$ (the minimum of the exponential one random variables governing the $\binom{k}{2}$ possible pairwise mergers). At that time it is equally likely to be any of the pairs of lineages which merges. There are then $(k-1)$ remaining

lineages and the additional time we must wait until the next merger is exponential with parameter $\binom{k-1}{2}$ and so on. This is an informal description of *Kingman's coalescent*.

In order to obtain the Kingman coalescent, we first measured time in units of N generations and then let $N \to \infty$. To see what this corresponds to in our forwards in time Wright-Fisher model, we suppose that our population is divided into just two types, which we label a and A, and that each offspring inherits the type of its parent. Since, if the proportion of type a alleles in the parental population is p, the absolute number of type a offspring is binomial with parameters (N, p), it is elementary to see that the change Δp in the proportion over a single generation satisfies

$$\mathbb{E}[\Delta p] = 0, \quad \mathbb{E}[(\Delta p)^2] = \frac{1}{N}p(1-p) \quad \text{and } \mathbb{E}[(\Delta p)^k] = \mathcal{O}\left(\frac{1}{N^2}\right) \text{ for } k \geq 3.$$

One generation corresponds to $1/N$ of our new units of time, so the N in these expressions can be interpreted as the length of the infinitesimal time interval over which we are measuring the change in proportion, and we deduce (for example from §15.1 of [13]) that in the limit as $N \to \infty$ the proportion of a-alleles in the population will evolve according to the *Wright-Fisher diffusion*,

$$dp_t = \sqrt{p(1-p)}dW_t,$$

where $\{W_t\}_{t\geq 0}$ is a standard Brownian motion. This is the classical model of *genetic drift*. It is an unfortunate accident of history, that the biological term 'drift' refers to this purely stochastic term, in contrast to the usual mathematical terminology. We see immediately that under the action of drift alone, one of the types in our population will eventually fix (the diffusion will be absorbed at either $p = 0$ or $p = 1$).

We now have two dual models: *forwards in time*, allele frequencies evolve according to the Wright-Fisher diffusion and *backwards in time*, genealogical trees relating individuals in a sample are described by the Kingman coalescent. Before we can compare to data, we must take into account the rescaling of time that has taken place and our dual models become:

Forwards in time:	Backwards in time:
$dp_\tau = \sqrt{\frac{1}{N_e}p_\tau(1-p_\tau)}dW_\tau$	Coalescence at rate $\frac{1}{N_e}\binom{k}{2}$

Notice that in place of the census population size, N, we have substituted an *effective* population size, N_e. As commented in §1, the Kingman coalescent provides an excellent model in an enormous variety of situations, provided we make this substitution and this is reflected in the Wright-Fisher diffusion. The size of N_e is a measure of the strength of the genetic drift - the smaller the effective population size, the more quickly genetic drift will wipe out variability in our population. Of course, without variability we couldn't infer the genealogical trees relating individuals in our sample. The ultimate source of all variation is mutation. In this classical setting, mutations are modelled as falling at a Poisson rate along each branch of the coalescent tree (see, e.g.[11]).

A mathematically convenient consequence of this 'robustness' of the Kingman coalescent and the corresponding Wright-Fisher diffusion, is that we can expect the same limiting models if we replace the Wright-Fisher model by essentially any model that captures the key assumptions of Definition 2.1. A particular instance of this is that we can replace the Wright-Fisher model (in which generations are discrete) by the *Moran model* in which generations overlap.

Definition 2.2 (The neutral Moran model [14]) *A population of size N is said to evolve according to the neutral Moran model if at exponential rate $\binom{N}{2}$ a pair of individuals is chosen uniformly at random from the population, one dies and the other splits in two.*

There are many ways to parametrize this model. Our choice here has the advantage that the genealogy of a sample from the population is given by Kingman's coalescent (there is no need to rescale time). Notice that reproduction is driven by a Poisson process (we shall see this reflected in our spatial model of §3).

The duality that we have described between the Kingman coalescent and the Wright-Fisher diffusion is 'strong' in the sense that the genealogical trees relating individuals in a sample from the Wright-Fisher model converge to the Kingman coalescent and so it is reasonable to refer to the Kingman coalescent as describing the genealogy of a sample from a population evolving according to the limiting diffusion. Donnelly and Kurtz ([15]) exploited the fact that one can embed the Moran model for a population of size N into one for a population of size $N+1$ to construct the Wright-Fisher diffusion (or more generally the Fleming-Viot process, [16, 17], which describes allele frequencies under the limiting model when the type space consists of a possibly infinite number of types) and the Kingman coalescent simultaneously on the *same* probability space. Often one reports a weaker form of the duality, *moment duality*. If $n(t)$ denotes the number of ancestral lineages alive in the Kingman coalescent at time t, then it follows that if $p(t)$ denotes the proportion of a-alleles at time t under the Wright-Fisher diffusion, then

$$\mathbb{E}_{p(0)}[p(t)^{n(0)}] = \mathbb{E}_{n(0)}[p(0)^{n(t)}]. \tag{1}$$

Here (and throughout) we use $\mathbb{E}_x[X(t)]$ to denote the expectation of the random variable $X(t)$ given that $X(0) = x$. The expectation on the left of (1) is with respect to the distribution of the Wright-Fisher diffusion, while that on the right is for the Kingman coalescent. The arrow of time for these two processes points in opposite directions, but we follow the usual convention of denoting time for the Kingman coalescent as a positive quantity. Thus in (1), a sample of $n(0)$ individuals is taken at time t for the Wright-Fisher diffusion (which corresponds to time 0 for the Kingman coalescent) and $n(t)$ is the number of ancestral lineages alive at time t before the sampling time (which is time 0 for the Wright-Fisher diffusion).

To see why (1) should hold, the left hand side is the probability that all individuals in a sample taken from the population at (Wright-Fisher) time t are of type a and, in the absence of mutation, this is the same as the probability that all individuals ancestral

to the sample at time 0 were of type a. The number of individuals alive at time 0 that are ancestral to a sample taken at time t is $n(t)$ and so the right hand side is precisely the probability that we seek. The analogue of (1) will be a key tool for analysis of our more complex models. However, it is important to understand that a moment duality of this form does *not* in general imply the stronger duality. In [18], Taylor gives examples of different individual based models for which the forwards in time processes of allele frequencies all converge to the same limit, but the genealogical processes converge to *different* limits.

2.2. Introducing space

Modelling spatially structured populations has a long history. Wright's island model ([19]) was an early attempt to understand the effects of spatial subdivision. In this model, the 'islands' of population are assumed to sit at the vertices of a complete graph and individuals migrate between islands. This has been refined and extended in many ways, but most models have the same basic structure: the population is subdivided into islands or *demes* which are situated at the vertices of a graph. Interaction between subpopulations is through migration of individuals along the edges of the graph.

An important example of models of this class is the Kimura stepping stone model, [20]. Let us index the vertices of the graph by a finite or countable set I and, for $i, j \in I$, write $i \sim j$ if the vertices labelled i and j are neighbours. Just as the Wright-Fisher diffusion was obtained as a limit of the neutral Wright-Fisher model as population size went to infinity, the Kimura stepping stone model can be obtained as a limit of a structured Wright-Fisher model as the size of the subpopulation in each deme tends to infinity. In the prelimiting model, in each generation each subpopulation reproduces according to the neutral Wright-Fisher model, but now, in addition, after reproduction it exchanges a proportion of individuals with its neighbours. More precisely, for each pair (i, j) with $i \sim j$, m_{ij} individuals chosen at random from deme i migrate to deme j and, to preserve population size in each deme, we suppose that $\sum_{j:i\sim j} m_{ij} = \sum_{j:i\sim j} m_{ji}$. We suppose that the effective population size in each deme is N_e and that $\sum_{i:i\sim j} m_{ij} \ll N_e$. (In fact, under this assumption, one obtains the same approximation if, for example, a Poisson number of individuals with mean m_{ij} migrate from deme i to deme j after each reproduction event.) If the population is divided into just two types, a and A, writing p_i for the proportion of a-alleles in deme i, the Kimura stepping stone model takes the form

$$dp_i = \sum_{j:j\sim i} m_{ji}(p_j - p_i)dt + \sqrt{\frac{1}{N_e}p_i(1-p_i)}dW_i, \tag{2}$$

where $\{W_i\}_{i\in I}$ are independent standard Brownian motions. This standard stepping stone model does not consider the effect of spatial variation in population density of migration, although this is easily incorporated (see, for example, [21]).

We can also mimic our previous arguments to understand the genealogical trees relating individuals in a sample from a population whose dynamics are governed by (2).

Just as we never saw mergers of three or more lineages in the Kingman coalescent, so (since $\sum_{i:i\sim j} m_{ij} \ll N_e$) we never see a simultaneous merger and migration in this *structured* coalescent. The genealogical tree relating a finite sample of individuals, consisting of n_i individuals from deme i for each $i \in I$, is traced out by the system of coalescing random walks $\{(n_i(t))_{i\in I}\}_{t\geq 0}$, where

- for each $i \in I$, $n_i \mapsto n_i - 1$ at instantaneous rate $\frac{1}{N_e}\binom{n_i}{2}$
- for each $i, j \in I$ with $i \neq j$, $\begin{cases} n_i \mapsto n_i - 1 \\ n_j \mapsto n_j + 1 \end{cases}$ at instantaneous rate $n_i m_{ji}$.

Once again this implies a weaker moment duality between the stepping stone model and the structured coalescent which we shall see reflected in our model in a spatial continuum. For any vector $\underline{n}(0) = (n_i(0))_{i\in I}$ with non-negative integer components and $\sum_i n_i(0) < \infty$ and any initial condition $\underline{p}(0) = (p_i(0))_{i\in I}$ for (2),

$$\mathbb{E}_{\underline{p}(0)}\left[\prod_{i\in I} p_i(t)^{n_i(0)}\right] = \mathbb{E}_{\underline{n}(0)}\left[\prod_{i\in I} p_i(0)^{n_i(t)}\right]. \tag{3}$$

Note once again that time for the coalescent process runs backwards and so we have adopted the same conventions as in (1).

2.3. Spatial continua: the pain in the torus

In some situations, the stepping stone model is very natural. However, many populations are not subdivided, but instead are distributed across a spatial continuum. For such populations, in order to use the stepping stone model, one must impose an artificial subdivision of space and choose a graph (often $\mathbb{Z}^2$) which caricatures the local geography. One would like instead an analogue of the stepping stone model in a spatial continuum.

The stepping stone model is a system of stochastic ordinary differential equations, indexed by the set I, for which a natural continuum analogue would be the stochastic *partial* differential equation

$$dp_t = \frac{1}{2}\Delta p dt + \sqrt{\frac{1}{N_e} p_t(1-p_t)}W(dt, dx), \tag{4}$$

where W is a space-time white noise. In one space dimension, one can obtain (4) through a diffusive rescaling of the stepping stone model on $\mathbb{Z}$. Moreover, under this rescaling, the system of coalescing random walks which describes the genealogy of a sample from the population (2) converges to a system of Brownian motions which coalesce at a rate depending upon the local time that they spend together. However, in two spatial dimensions equation (4) has no solution. Moreover, if one applies the diffusive rescaling to the stepping stone model on $\mathbb{Z}^2$, the limit obtained is deterministic. The easiest way to see this is to consider what happens when one rescales the genealogical trees. Just as in one dimension, the motion of a single lineage will converge to Brownian motion, but since two Brownian motions in two dimensions will never meet, we don't see any

coalescence and, as we saw in §2.1, the coalescence backwards in time corresponds to the random genetic drift forwards in time.

Although in two spatial dimensions, two Brownian motions will never meet, they *will* come arbitrarily close to one another infinitely often. This suggests another possible approach: model the genealogy of a sample from the population as a system of Brownian motions which coalesce at an instantaneous rate depending upon their separation. We must also specify the location of the parent, for example as the midpoint between the two coalescing lineages. However, such a model lacks *sampling consistency*. To see this, take the genealogical tree relating a sample of three individuals and delete the ancestral lineage corresponding to a randomly chosen individual in the sample. With positive probability, this lineage was involved in the most recent coalescence event in our full tree and at the time of that coalescence event, the other lineage which was involved (which is ancestral to our subsample) will jump. This jump would *not* be seen if we modelled the genealogy of the subsample directly.

Malécot and Wright attempted to model populations living in a spatial continuum in the 1940s ([22, 19]). Their model assumes that individuals are dispersed according to a Poisson Point Process of constant intensity in $\mathbb{R}^2$. The reproduction mechanism mimics that of the Wright-Fisher model: the population evolves in discrete generations and the number of offspring of each individual is Poisson with mean one. This is exactly the distribution of the family size of a single individual in the Wright-Fisher model as $N \to \infty$. The twist is that offspring now have a spatial position which is sampled (independently) from a Gaussian distribution centred on the spatial location of their parent. In addition, the model incorporates mutation: with probability μ an offspring, rather than inheriting the genetic type of its parent, is of a type never before seen in the population.

The work of Malécot and Wright predates the coalescent by forty years. Instead of discussing genealogical trees, they calculated $F(x)$, the *probability of identity in state* of two individuals sampled at separation $x \in \mathbb{R}^2$; that is the probability that the two individuals carry identical copies of an ancestral allele. This is equivalent to calculating the generating function of the number of generations back to their MRCA. Malécot uses a recursion to obtain an approximation for $F(x)$ in terms of K_0, the modified Bessel function of the second kind of order zero:

$$F(x) \approx \frac{1}{\mathcal{N} + \log(l/\kappa)} K_0\left(\frac{\|x\|}{l}\right), \qquad \|x\| > \kappa, \tag{5}$$

where κ is a local scale, $l = \sigma/\sqrt{2\mu}$ is the characteristic length scale, σ^2 is the variance of the Gaussian distribution that determines the location of offspring and $\mathcal{N}$ is Wright's *neighbourhood size*, which, loosely, measures the number of 'potential parents' of each offspring. We shall refer to (5) as the *Wright-Malécot formula*.

However, as first observed by Felsenstein in 1975 ([23]), the assumptions of Wright and Malécot are inconsistent. On the one hand they assume that the population is at stationarity and distributed across the plane as a Poisson point process of constant intensity, but on the other that it evolves according to a branching random walk. The

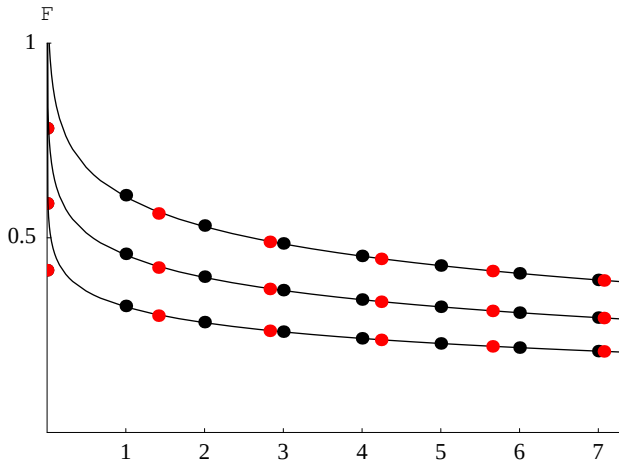


Figure 1. Identity in state in a stepping stone model on $\mathbb{Z}^2$. The horizontal axis is the (Euclidean) separation at which individuals are sampled. The three sets of dots correspond to exact values for three different values of N_e. The curves are obtained from the Wright-Malécot formula.

difficulty is that in (one and) two spatial dimensions, populations whose dynamics are governed by spatial branching processes (with homogeneous branching and dispersal mechanisms) have no stationary distribution. If not extinct, they develop clumps of arbitrarily large density and extent. To overcome this, Felsenstein tried working on a torus instead of the whole plane. However, the total population size is then governed by a Galton-Watson branching process, which will either eventually die out or grow without bound, and so next he tried exogenously specifying the population size. But this does not overcome the clumping. Felsenstein famously dubbed the problem *'the pain in the torus'*. The key point is that to overcome clumping, the population size must be regulated *locally*. There are models based on spatial branching processes which incorporate local regulation of the population size and exhibit stable population dynamics (e.g.[24, 25, 26, 27, 28]), but these are challenging to work with and there seems little hope of recovering explicit descriptions for the corresponding genealogical trees. In [29], we successfully extend the Wright-Malécot formula to a class of population models that incorporate local regulation. However, the usefulness of this result is limited due to the lack of explicit models for which the assumptions of our model can be validated and parameters established.

At this point we have two (unsatisfactory) approaches to modelling populations in a spatial continuum. We can artificially subdivide the population and use a stepping stone model or we can ignore the inconsistencies in the assumptions and use the Wright-Malécot formula. What is striking is just how close these two approaches turn out to be. Figure 1 shows the probability of identity at different separations under the stepping stone model on $\mathbb{Z}^2$ (with $m_{ij} = 1/4$ if $i \sim j$) for three different choices of N_e. The curves are the corresponding predictions of the Wright-Malécot formula (for appropriate choices of parameters). The fit is astonishing, even for a deme-spacing of one.

3. A framework for modelling

3.1. Some biological considerations

We have so far focused on the mathematical shortcomings of the classical models for gene flow. But they also fail to provide an explanation of biological data: in particular, for the wide range of spatial scales over which patterns in allele frequencies persist, for the magnitude of random fluctuations seen even in dense populations, and for correlations across unlinked genetic loci. The Wright-Malecot formula predicts an approximately exponential decay in the probability of identity as a function of separation; in contrast, data reveals that although this is a good approximation over relatively short distances, correlations in allele frequencies persist over much longer spatial scales (e.g. [30]). There is also no convincing explanation for the huge discrepancy between census population size and effective population size that is implied by the moderate levels of genetic diversity seen even in abundant populations ([31]). Finally, the classical theory of isolation by distance implies that patterns at loosely linked loci are independent of each other. Yet, the whole field of phylogeography ([32]) is based on the idea that genetic patterns reflect demographic history - shaped by large-scale population movements - which implies that such patterns are shared across loci. The idea that a population's history can be inferred from its genetics makes no sense under the classical theory. One possible explanation for all three effects is that the demographic history of many species is dominated by large-scale extinction/recolonisation events, which substantially reduce genetic diversity and cause collective movements of genes across large distances.

There is a further consideration when modelling spatially distributed populations. Our approximation of the exact genealogical trees for the Wright-Fisher model, which do (albeit rarely) admit mergers of three or more ancestral lineages, by the Kingman coalescent, in which there are only pairwise mergers, rests on the fact that offspring are selecting parents (independently and uniformly) from a very large pool. In the language of §2.3, *neighbourhood size* is very large. However, in a spatial continuum, neighbourhood size can be small and then pairwise coalescence of ancestral lineages may not dominate.

The framework that we now describe not only overcomes the pain in the torus, but it also allows us to explicitly incorporate both large-scale extinction/recolonisation events and small neighbourhood size. The key innovation is that reproduction and extinction/recolonisation events are both driven by a space-time Poisson process rather than being based upon individuals. Of course, a Poisson process is probably not an appropriate model for some types of large scale events (e.g. glacial maxima), but for others (forest fires, storms etc.) it can be viewed as a reasonable starting point. Moreover, the framework provides a mechanism for modelling populations *conditional* on such events. We shall focus on a particular model that arises from this framework, and briefly mention another variant in §3.3, but it should be emphasised from the outset that the framework itself is very flexible: only the formulation of genetic drift in terms of a Poisson process of events is crucial. The exact form of those events is not important.

3.2. Individual based models

We begin with an individual-based model. For this section we shall suppose that the population evolves in $\mathbb{R}^d$. We can work in any dimension, but from the biological point of view, the most interesting case is $d = 2$. The model is parametrized by a real number $\lambda > 0$ and a measure which we write as

$$\xi(dr, du) = \mu(dr)\nu_r(du) \tag{6}$$

on $(0, \infty) \times (0, 1]$. Here μ is a (possibly infinite) measure on $(0, \infty)$ which will determine, for each r, the rate at which 'events of size r' fall on any given point. For each r, ν_r is a probability measure on $(0, 1]$ which will determine the 'impact' of an event. To avoid trivialities we also assume that $\xi((0, \infty) \times (0, 1]) > 0$. The dynamics are driven by a Poisson point process, Π, on $\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+ \times (0, 1]$ with intensity $dt \otimes dx \otimes \xi(dr, du)$.

Each point (t, x, r, u) of Π is thought of as an 'event' which affects only individuals living within $B_r(x)$, the closed ball of radius r and centre $x \in \mathbb{R}^d$. Frequent 'small' events model ordinary reproduction, whereas infrequent 'large' events mimic the effects of large-scale extinction/recolonisations. More precisely, at a point (t, x, r, u) of Π, if $B_r(x)$ is empty do nothing. Otherwise:

- choose a parent uniformly at random from those individuals present in the ball;
- each individual in $B_r(x)$ (including the parent), independently, dies with probability u;
- throw down new individuals (with the same type as the parent) according to an independent Poisson point process with intensity $u\lambda \mathbf{1}_{B_r(x)}(y)dy$ (where $\mathbf{1}_A$ is the indicator function of the set A).

This mechanism can then be thought of as regulating the reproductive success of individuals. If the ball $B_r(x)$ is crowded, then each individual living there has only a small chance of reproducing. On the other hand if the ball is only sparsely populated they have a significant chance of producing a Poisson number of offspring with mean $\lambda u V_d(r)$ where $V_d(r)$ is the volume of a ball of radius r in $\mathbb{R}^d$.

In order to ensure that this process exists, in [4] we assumed that

$$\int_0^\infty \int_0^1 u r^d \left(1 + r^d\right) \nu_r(du)\mu(dr) < \infty. \tag{7}$$

In fact this condition is stronger than is really required and A.V. and Anton Wakolbinger (personal communication) have shown that the model exists under the weaker condition (12) below.

If we were to allow 'births' in events when there is no potential parent present in the ball, then the population would have a Poisson point process with intensity λdx as stationary distribution. However, we do not. Nonetheless, because neighbourhoods overlap, an empty region can subsequently become recolonised. The question is whether this is enough to prevent the population from dying out. In [4], it is shown that there is a critical value of the parameter λ below which extinction is certain, but above which

the population, started from a Poisson random measure with constant intensity, survives (indeed there is an ergodic stationary distribution).

Ideally, we would now identify the distribution of the genealogy of a sample from such a population. However, it turns out that this is complicated. Because regions can, and do, become empty, knowing that there is an individual in our sample at a point $x \in \mathbb{R}^2$ tells us something about the Poisson process of events that have occurred. In particular, it is not possible to infer the ancestry by simply reversing the Poisson process of events. On the other hand, for sufficiently large λ, one expects the genealogy to be well approximated by the system of coalescing lineages generated by reversing the Poisson process and assuming that the regions are never empty. The methods of [4] can be used to make this precise. Here, instead, in the next subsection we turn to a variant of the model in which explicit calculations are feasible and examine the stability of genealogical trees as the density of individuals changes in that setting.

3.3. A Gaussian replacement mechanism

The individual based model considered above is very special. There is of course no reason to suppose that reproduction or extinction/recolonisation events should affect discs and still less that they should have the same 'impact' on everyone living in a region. An alternative model was studied in [6]. The dynamics are again driven by a Poisson point process of events, but now instead of affecting a compact region, an event has the potential to affect all individuals in the population, but with a probability that decreases (with Gaussian decay) with distance from the 'centre' of the event. More precisely, dynamics are driven by a Poisson process, Π_1, with intensity $\Lambda dt \otimes dx$ for some constant Λ. If $(t, x) \in \Pi_1$ then we first choose a parent by taking a weighted sample from the current population, where an individual at y is given weight

$$v(y, x) = \exp\left(-\frac{\|y - x\|^2}{2\alpha^2\theta^2}\right).$$

Here we will assume that $\alpha, \theta \in \mathbb{R}$ are constant, but they can, more generally, be taken to be random (and not necessarily independent). Each individual is killed with a weighted probability, so that an individual at z, say, is killed with probability

$$u(z, x) = u_0 \exp\left(-\frac{\|z - x\|^2}{2\theta^2}\right)$$

where $0 < u_0 \leq 1$. Offspring, of the same type as the parent, are thrown down according to a Poisson process with intensity $\lambda u(z, x)dz$.

This time, the process has a stationary distribution for all choices of λ, given by a homogeneous Poisson process in the plane with intensity λdx. Moreover, the genealogy *can* be obtained by simply reversing the Poisson process of events. However, analogous to the effect of empty regions in our disc model, if λ is small, it may be that an ancestral lineage sometimes has to make a very long jump to find a parent. In contrast to the disc model, in this setting, provided we assume that the population is at stationarity,

it is possible to establish a closed form expression for the distribution of that jump and our aim is to estimate just how sensitive this is to changes in the population density λ. This is equivalent to understanding how the distribution of the position of the parent picked in an event depends on λ. In particular, we are interested in seeing how rapidly this converges to a limit as $\lambda \to \infty$.

Since the dynamics are spatially homogeneous, it is enough to consider a reproduction event centred on the origin. Let us identify the probability that the parent chosen in the event is in an infinitesimal neighbourhood of the point y. First, there must be a point of the population at y which happens with probability λdy. Second, we must have chosen that point when we took our weighted sample. Now, given that there is an individual at y, the rest of the population is distributed according to an independent Poisson point process with intensity λdx. Thus the probability that we sampled the individual at y is $\mathbb{E}[a/(a+S)]dy$ where

$$a = \exp(-\frac{y^2}{2\alpha^2\theta^2}), \quad S = \sum_{X_i \in \pi} \exp(-\frac{X_i^2}{2\alpha^2\theta^2}),$$

and π is a Poisson point process with intensity λdx. To calculate this probability, we first find the Laplace transform of S.

$$\mathbb{E}[\exp(-\eta S)] = \mathbb{E}\left[\prod_i \exp\left(-\eta e^{-X_i^2/(2\theta^2\alpha^2)}\right)\right] = \exp\left(-\lambda \int_{\mathbb{R}^2} (1-\phi(x))dx\right),$$

where

$$\phi(x) = \exp\left(-\eta e^{-x^2/(2\theta^2\alpha^2)}\right).$$

Transforming to polar coordinates and making the substitution $u = e^{-r^2/(2\theta^2\alpha^2)}$ yields

$$\mathbb{E}[e^{-\eta S}] = \exp\left(-2\theta^2\alpha^2\pi\lambda \int_0^1 \left(\frac{1-e^{-\eta u}}{u}\right) du\right).$$

Now observe that

$$\mathbb{E}\left[\frac{a}{a+S}\right] = \int_0^\infty \mathbb{E}\left[a \exp\left(-\eta(a+S)\right)\right] d\eta,$$

and use that

$$\eta - \frac{\eta^2}{4} \leq \int_0^1 \left(\frac{1-e^{-\eta u}}{u}\right) du \leq \eta,$$

to obtain

$$\mathbb{E}\left[\frac{a}{a+S}\right] = \frac{a}{a+2\theta^2\alpha^2\pi\lambda}\left(1+\mathcal{O}\left(\frac{1}{\lambda}\right)\right) = \frac{a}{2\theta^2\alpha^2\pi\lambda}\left(1+\mathcal{O}\left(\frac{1}{\lambda}\right)\right).$$

From this, we see that the probability of sampling a parent in an infinitesimal neighbourhood of y is

$$\frac{1}{2\theta^2\alpha^2\pi}\exp\left(-\frac{y^2}{2\theta^2\alpha^2}\right)\left(1+\mathcal{O}\left(\frac{1}{\lambda}\right)\right)dy.$$

Even this rather crude calculation guarantees that up to a relative error of order $1/\lambda$ we can approximate the jump of the ancestral lineage by the Gaussian distribution obtained in the limit as $\lambda \to \infty$. Of course, for $\lambda < \infty$, successive jumps will not be independent, but an analogous argument guarantees the rapid decay of correlations as λ increases.

In the disc replacement model of the previous section there is similar stability. As λ increases, the chance of a region that is hit by an event being empty decreases rapidly as λ increases.

3.4. The spatial Λ*-Fleming-Viot model*

Since the distribution of the genealogical trees relating individuals in a sample rapidly stabilises as the density λ of individuals in our population grows, it seems reasonable to consider the model obtained as $\lambda \to \infty$. At first sight this seems analogous to letting neighbourhood size tend to infinity in a structured Wright-Fisher model. However, as we illustrate in §3.6, this is not the right interpretation and in fact our model will retain a signature of finite neighbourhood size. As a result, the genealogical trees relating individuals in a sample from the population will be spatial versions of so-called Λ-coalescents which admit 'multiple mergers', by which we mean that three or more lineages can coalesce in a single event. We describe this in detail in §3.5, but first we formulate the (limiting) forwards in time model for allele frequencies.

The *spatial* Λ*-Fleming-Viot process* is the name given to the process obtained in the limit as $\lambda \to \infty$ in our individual-based model. To understand the form of the limit, let us begin with a *non-spatial* analogue of the individual based model of §3.2. Suppose then that our population initially consists of a Poisson number of individuals with parameter λ. The dynamics are driven by a Poisson point process Π_0 on $(0,\infty) \times (0,1]$ with intensity $dt \otimes F(du)$ for a suitable measure F on $(0,1]$. At a point $(t,u) \in \Pi_0$, provided the population is not already extinct, we choose a parent at random from the population at time $t-$ (that is immediately before the event); each individual, independently, dies with probability u; and a Poisson number of offspring with mean λu, all of the same type as the parent, are born. We measure the population in units of size λ. In these units, at time zero it is of size $1 + \mathcal{O}(1/\sqrt{\lambda})$ and the proportion of the population replaced at a reproduction event is $u + \mathcal{O}(1/\sqrt{\lambda})$. We see that as $\lambda \to \infty$ the population size becomes fixed at 1. The model specifies the distribution of types in the population and so is represented by a probability measure ρ on a (compact) type space K. For $(t,u) \in \Pi_0$, if the distribution of types in the population at time $t-$ is $\rho(t-,\cdot)$, then immediately after the event it is given by

$$\rho(t,\cdot) = (1-u)\rho(t-,\cdot) + u\delta_k,$$

where the parental type, k, is chosen according to $\rho(t-,\cdot)$. For reasons that will emerge in §3.5 we shall call this process a Λ*-Fleming-Viot process*. Notice that in passing to this limit, we are not rescaling time and so our limiting model is driven by the same Poisson process Π_0 of events.

The spatial analogue of the Λ-Fleming-Viot process is our main object of study. Whereas, at each time $t \geq 0$, the Λ-Fleming-Viot process specifies a single probability measure on K (which determines the distribution of types in a panmictic population), our spatial analogue will specify a different probability measure on K at *every* point in space. The interpretation is that if we were to sample an individual from the point x at time t, then its genetic type is determined by a random pick from $\rho(t, x, \cdot)$.

Definition 3.1 (The spatial Λ-Fleming-Viot process) *The* spatial Λ-Fleming-Viot process, $\{\rho(t, x, \cdot), x \in \mathbb{R}^d, t \geq 0\}$, *specifies a probability measure on a compact type space K for every $t \geq 0$ and every $x \in \mathbb{R}^d$. With the notation above, the dynamics of the process are as follows. At every point (t, x, r) of the Poisson point process Π (independently), choose $u \in (0, 1]$ according to the measure $\nu_r(du)$. Select a point z at random from $B(x, r)$ and a type k at random according to $\rho(t-, z, \cdot)$. For all $y \in B(x, r)$,*

$$\rho(t, y, \cdot) = (1 - u)\rho(t-, y, \cdot) + u\delta_k. \tag{8}$$

Sites outside $B(x, r)$ are not affected, that is $\rho(t, y, \cdot) = \rho(t-, y, \cdot)$ for every $y \notin B(x, r)$.

We shall require some conditions on the intensity of the Poisson process Π if such a process is to exist. These will be made precise in §3.5.

3.5. The spatial Λ-coalescent

The key tool for analysing the spatial Λ-Fleming-Viot process is its coalescent dual. Indeed in [5] existence of the process is proved from existence of the dual using powerful results of [33]. To understand that dual process, and to make the connection with earlier work, we first consider the non-spatial context. As we saw in §3.4, in the non-spatial version of our Λ-Fleming-Viot process, the allele frequencies change at times dictated by the points of a Poisson point process. For convenience we think of time being extended to the whole real line so that this Poisson point process is reversible. The dual process of coalescing lineages is then driven by the time-reversed process. Suppose that we sample individuals uniformly at random from the population. Then at a point (t, u) of the reversed Poisson point process, all those ancestral lineages lying in the portion u of the population corresponding to 'offspring' in the forwards in time model will coalesce into a common parent. Since our sample was picked at random, each, independently, has probability u of being among the offspring. Thus, if there are currently n lineages ancestral to our sample, the chance of a particular subset of k of them coalescing is $u^k(1 - u)^{n-k}$ and since these coalescence events were driven by the time reversal of the Poisson point process Π_0, the rate at which we see such an event can be written as

$$\int_0^1 u^k(1 - u)^{n-k} F(du).$$

We recognise these rates as those of a Λ-*coalescent*. These coalescents were introduced independently by Donnelly & Kurtz, Pitman and Sagitov in [34, 35, 36]. They exist provided that $F(du) = \Lambda(du)/u^2$ where Λ is a finite measure on $[0, 1]$. The representation

in terms of jumps of a Poisson point process rests on Λ having no atom at 0. The process still exists with such an atom, which corresponds to adding extra pairwise events (a Kingman component) to the coalescent process, but our spatial extension does not admit such a term. The duality between Λ-coalescents and Λ-Fleming-Viot processes was made explicit by Bertoin & Le Gall in [37].

In the spatial setting, the picture is very similar. Once we extend time to the whole real line, the Poisson point process Π that dictates the dynamics of the model is reversible. At any point $(t, x, r, u) \in \Pi$, each of the lineages ancestral to our sample that is in $B_r(x)$ will (independently), with probability u, jump to the position of the 'parent' of the event, which is uniformly distributed on $B_r(x)$. Crucially, this is true even if there is only a single lineage within the ball (otherwise we would lose sampling consistency, c.f. §2.3). As a result, if we follow a single lineage, then it evolves in a series of jumps with intensity

$$dt \otimes \left(\int_{[|x|/2,\infty)} \int_{(0,1]} \frac{L_r(x,0)}{V_d(r)} u \nu_r(du) \mu(dr) \right) \, dx, \tag{9}$$

where $L_r(x, y)$ is the volume of $B_r(x) \cap B_r(y)$. To see where equation (9) comes from, suppose (without loss of generality) that our lineage is currently at 0. In order for it to jump to x, both 0 and x must lie within $B_r(z)$ for some $(t, z, r, u) \in \Pi$. The volume of admissible centres z is $L_r(x, 0)$. If the lineage is to jump to x, then the point x must have been chosen as the location of the parent, which, since the parental position is uniformly distributed on the ball affected by the event, happens with probability $dx/V_d(r)$ and, finally, the lineage must be among the 'offspring' of the event, which happens with probability u.

In order for this jump intensity to correspond to a well-defined Lévy process, we require that

$$\int (\min(1, |x|^2) \int_{[|x|/2,\infty)} \int_{(0,1]} \frac{L_r(x,0)}{V_d(r)} u \nu_r(du) \mu(dr) dx < \infty. \tag{10}$$

Two lineages currently at separation $y \in \mathbb{R}^d$ will coalesce if they are *both* affected by an event, which will happen at instantaneous rate

$$\int_{[|y|/2,\infty)} L_r(y,0) \int_{(0,1]} u^2 \nu_r(du) \mu(dr). \tag{11}$$

Evidently, if this is bounded, then so too will be the rates of all other possible coalescence events.

One might hope that these two conditions would be enough to guarantee existence of our model. But in order to identify our process we need a stronger condition. We suppose that

$$\Lambda(du) = \int_{(0,\infty)} u r^d \nu_r(du) \mu(dr) \tag{12}$$

defines a finite measure on $[0, 1]$. As we noted in §3.2, although weaker than the condition (7) which was required in [4], this condition also suffices for existence of the prelimiting model of §3.2.

We defer the formal statement of the moment duality between the spatial Λ-Fleming-Viot process and the spatial Λ-coalescent described above until we have specialised to type space $K = \{0, 1\}$ in §4.4 (where the necessary notation is less intimidating).

A rigorous proof of convergence of the individual based model of §3.2 to the limiting spatial Λ-Fleming-Viot process that we have described is the object of joint work of A.M.E. and Tom Kurtz. A corollary of that work is that the genealogies also converge so that (as for the duality between the Wright-Fisher diffusion and the Kingman coalescent) the duality between the forwards in time model of allele frequencies and the spatial Λ-coalescent is not just a weak moment duality.

3.6. Parameters

In order to gain a better understanding of some of the parameters in the model, it is helpful to think about how we would try to approximate Kimura's stepping stone model in this framework. To this end, let us replace $\mathbb{R}^2$ by $\mathbb{Z}^2$ and instead of taking events to be balls of radius r based upon points $x \in \mathbb{R}^2$, suppose that events cover exactly two neighbouring lattice points. More precisely, for each $x = (x_1, x_2) \in \mathbb{Z}^2$, events covering $\{(x_1, x_2), (x_1 + 1, x_2)\}$, $\{(x_1, x_2), (x_1 - 1, x_2)\}$, $\{(x_1, x_2), (x_1, x_2 + 1)\}$, $\{(x_1, x_2), (x_1, x_2 - 1)\}$, each arrive according to independent Poisson processes of rate ρ. When such an event falls, each of the points covered is equally likely to be chosen as the location of the 'parent' of the event. For simplicity, let us suppose that u, the proportion of the population replaced during an event, is fixed. To make the comparison to the stepping stone model, we investigate the behaviour of ancestral lineages under this model. Each will jump to a neighbouring site at rate $m = 4\rho u$. If two lineages are at the same lattice point they coalesce at rate $8\rho u^2$, which we equate to $\frac{1}{N_e}$. We ignore the fact that lineages in neighbouring demes can coalesce in our model, as this argument is meant to be no more than heuristic. Roughly then, using $m = 4\rho u$ to eliminate ρ from the expression for $1/N_e$, the parameter u can be expressed in terms of migration rate and 'local population density' as $u = 1/(2mN_e)$. The quantity mN_e is proportional to Wright's neighbourhood size, which we introduced in §2.3. What this shows is that keeping u macroscopic as we pass to the limit as $\lambda \to \infty$ retains a signature of finite neighbourhood size. Thus our limit should *not* be thought of as analogous to allowing $N \to \infty$ when we pass, for example, from the Wright-Fisher model to a diffusion approximation. This is further reflected in the fact that if more than two lineages are in a region hit by an event then any subset of them can coalesce during the event.

4. Some results in the neutral case

In this section we investigate how our model addresses some of the biological issues raised in §3.1.

4.1. Large-scale events and genetic diversity

The results of this subsection are from [38] and can also be found in detail in [5]. For ease of exposition, we do not present them in their full generality.

In [39], Zähle, Cox & Durrett consider the Kimura stepping stone model on a torus of side L in $\mathbb{Z}^2$. They show, in particular, that if one samples a finite number of individuals uniformly at random from the torus, then as $L \to \infty$, measuring time in units of $\mathcal{O}(L^2 \log L)$, the genealogy of the sample converges to a Kingman coalescent. Let's begin by trying to understand this result. Suppose that we sample two individuals according to the uniform distribution on $(\mathbb{Z} \bmod L)^2$ and write t_0 for the time since their MRCA. We divide t_0 into two phases. The first, of duration T_0, is the period until the lineages are first in the same deme; the second is the additional period until they actually coalesce. The first observation of [39] is that $T_0/(L^2 \log L)$ converges to an exponentially distributed random variable as $L \to \infty$; the second is that the additional time, $t_0 - T_0$, is asymptotically negligible in the timescale $L^2 \log L$. They consider several cases, in some of which the local population density can be very big, but in the scenario we consider here it is $\mathcal{O}(1)$. The extension to larger samples uses the fact (which we already see reflected in the exponential distribution of $T_0/(L^2 \log L)$) that the time $L^2 \log L$ is long enough for a random walk to reach its mixing time on the torus of side L in $\mathbb{Z}^2$ and so at the time when a pair first come into a common deme the positions of the lineages ancestral to the sample are no longer correlated with their starting points. This gives exchangeability: each pair of lineages is equally likely to coalesce. Moreover, when a first pair of lineages comes together, the other lineages are still far apart and so we will not see 'multiple' mergers of lineages.

In order to investigate the reduction in genetic diversity (or equivalently in effective population size) resulting from the large-scale extinction/recolonisation events in the spatial Λ-Fleming-Viot process, it is natural to mimic the approach of [39] and work on a large torus in $\mathbb{R}^2$.

We write $\mathbb{T}(L)$ for the torus of side L in $\mathbb{R}^2$. We shall consider two types of event. Small events will affect uniformly bounded regions. The rate at which we see small events of radius r will be governed by a σ-finite measure $\mu_s(dr)$ on $[0, R_s]$. Large events will affect regions with a radius of $\mathcal{O}(L^\alpha)$ for some $0 < \alpha \leq 1$. The rate at which we see events of radius $L^\alpha r$ will be determined by a σ-finite measure $\mu_B(dr)$ on $[0, R_B]$.

More precisely, the dynamics of our population will be driven by *two* Poisson point processes:

- small events are driven by Π_L^s, a Poisson point process on $\mathbb{R}_+ \times \mathbb{T}(L) \times [0, R_s] \times (0, 1]$ with intensity $dt \otimes dx \otimes \xi_s(dr, du)$ where $\xi_s(dr, du) = \mu_s(dr)\nu_r^s(du)$;
- large events are driven by Π_L^B, a Poisson point process on $\mathbb{R}_+ \times L^{-\alpha}\mathbb{T}(L) \times [0, R_B] \times (0, 1]$ with intensity $\frac{1}{\rho_L} dt \otimes dx \otimes \xi_B(dr, du)$ where $\xi_B(dr, du) = \mu_B(dr)\nu_r^B(du)$.

The reproduction mechanism is as before except that at a point (t, x, r, u) of Π_L^B, a reproduction event takes place in the ball centred at $L^\alpha x$ and of radius $L^\alpha r$. The

parameter ρ_L determines the relative frequency of small and large events and therefore their relative importance in shaping the genealogy of a sample.

Of course, since the sum of two independent Poisson processes is again a Poisson process, this is the same as the model of Definition 3.1, but we have divided the driving Poisson Point Process into two parts in order to disentangle the respective effects of large and small events on the genealogical trees.

It is important to understand that the effect of large scale events here is very different from that of adding long range dispersal in a classical stepping stone model. Rather than a single offspring being born at a very large displacement from its parent, here as a result of a 'big' event offspring of a single parent replace a proportion of the population at *every* point within a large ball. If we wish to approximate a stepping stone model with long range dispersal within this framework, then we must do something analogous to the scalings of §5.3 in which the proportion of the population within the affected region which is replaced during a reproduction event shrinks to zero. We are unaware of the existence in the literature of a proof, in that setting, of an analogue of Theorem 4.1.

Without the large events, the model is very much like the stepping stone model and so by analogy with the results of [39] we expect that on timescales of $\mathcal{O}(L^2 \log L)$ the genealogy of a uniform sample from the torus should look approximately like a Kingman coalescent. Our first result says that for any $\alpha < 1$ the genealogy of a uniform sample will still be close to a Kingman coalescent, but the timescale can depend on both big and small events. To understand that timescale, let us consider just two lineages. As for the stepping stone model, the time to coalescence can be divided into two phases. If ρ_L is not too big, that is large events are sufficiently frequent, then the *first* phase is the time that it takes for the two lineages to come within distance $2R_B L^\alpha$ of one another, so that there is some chance that they will be hit by the same event. The *second* phase is the additional time to coalescence which, if ρ_L is not too big, will be triggered by a large event. If, on the other hand, ρ_L is big, then large events are too infrequent to alter the genealogy and the coalescence will be caused by a small event. The first phase is then the time to come within distance $2R_s$ and the second is the additional time to coalescence. The transition between the two regimes is, as one expects from the results for the stepping stone model, when $\rho_L \propto L^2 \log L$.

To state a more precise result, let σ_s^2 (resp. $\sigma_B^2 L^{2\alpha}/\rho_L$) denote the variance in the displacement of a single ancestral lineage in one time unit due to small (resp. large) events.

Theorem 4.1 (Special case of Theorem 3.3 of [5]) *Define*

$$\omega_L = \begin{cases} \frac{(1-\alpha)\rho_L L^2 \log L}{2\pi\sigma_B^2 L^{2\alpha}} & \text{if } \frac{L^{2\alpha}}{\rho_L} \to \infty, \\ \frac{(1-\alpha)L^2 \log L}{2\pi(\sigma_s^2 + b\sigma_B^2)} & \text{if } \frac{L^{2\alpha}}{\rho_L} \to b \in [0,\infty), \\ \frac{L^2 \log L}{2\pi\sigma_s^2} & \text{if } \frac{L^2 \log L}{\rho_L} \to 0. \end{cases}$$

Then if we measure time in units of ω_L the genealogy of a uniform random sample from

$\mathbb{T}(L)$ *converges in law to the Kingman coalescent.*

In particular, we see that big events can change the effective population size, even though the genealogy is asymptotically determined by a Kingman coalescent. Importantly, in this case coalescence may be due to large scale events, but asymptotically even the large scale events capture at most two lineages at a time.

For $\alpha = 1$, so that big events affect a significant proportion of the species range, the picture is much richer:

Theorem 4.2 (Summary of Theorem 3.7 of [5]) *Suppose that* $\alpha = 1$.

(i) if $\rho_L/L^2 \to b$, *rescaling the torus by* $1/L$, *on timescale* L^2 *the genealogy converges to a* spatial Λ*-coalescent on* $\mathbb{T}(1)$ *in which, between mergers, lineages follow independent Brownian motions;*

(ii) if $\rho_L/(L^2 \log L) \to \beta$, *on timescale* $L^2 \log L$ *the genealogy converges to a (non-spatial)* Λ*-coalescent;*

(iii) if $\rho_L \gg L^2 \log L$ *then on timescale* $L^2 \log L$ *the genealogy converges to a Kingman coalescent.*

In particular, in the first two cases the limit is no longer a Kingman coalescent and we are seeing effects of the large scale events that cannot be captured through a simple timechange of the classical Kingman coalescent.

We refer to the original article for exact expressions for the parameters in the limits and content ourselves with trying to understand why the limits take these particular forms. Once again $L^2 \log L$ arises as the mixing time of a random walk making jumps of size $\mathcal{O}(1)$ over $\mathbb{T}(L)$. In the first scenario, the diffusive rescaling gives rise to the Brownian motion of ancestral lineages in between large events. In this case the mixing time has not been achieved when the first large event is seen, so lineages still have some memory of their starting position at that time and the coalescent retains some spatial structure. Since large events encompass a positive fraction of the torus, we may see multiple mergers of ancestral lineages. In the second case, the mixing time is achieved before we see a large event, and so the limiting coalescent has no spatial structure. For each surviving lineage, when a large event arrives its chance of being in the region affected by the event is just the proportion of the torus covered by the event, irrespective of its starting position and independent of all other surviving lineages. Moreover, in between large events, which are separated by times of $\mathcal{O}(L^2 \log L)$, we may see some pairwise coalescences due to small events, leading to a Kingman component in the limiting coalescent. In the last case, all coalescence is dictated by small events, before the first large event arrives.

4.2. Comparison with the Wright-Malécot formula

In the previous section we saw that large-scale extinction/recolonisation events certainly have the potential to reduce effective population size. In §3.1 we also suggested that they could explain correlations in allele frequencies over large spatial scales. In particular,

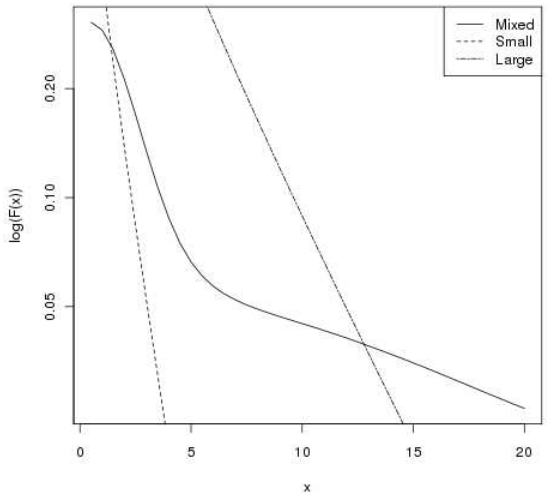


Figure 2. The logarithm of the probability of identity of two individuals as a function of their spatial separation under three different scenarios: just small events, just large events and a mixture of the two. The impact of the large events in this last case is seen by the replacement of one initially approximately exponential rate of decay by a slower rate of decay at larger initial separations.

whereas the Wright-Malécot formula predicts approximate exponential decay in the probability of identity in state of two individuals as a function of their separation, we expect an approximately exponential decay over relatively small scales to be replaced by a slower rate over larger scales. Figure 2 is a simulation of the two-dimensional version of our model by Jerome Kelleher. It shows the logarithm of the probability of identity of two individuals as a function of their separation x under three different scenarios: just small events, just large events, and a mixture of the two. Since we haven't discussed incorporation of mutation in our model, this should be interpreted as the generating function of the time to the MRCA of the two individuals. With just one size of event, we see the approximately exponential decay of the Wright-Malécot formula, one of the characteristics one would hope for from a 'continuum stepping stone model'. When we have a mixture of small and large events we see the rate of decay of identity decrease as spatial separation increases, suggesting that large scale extinction/recolonisation events really do provide one possible explanation of this pattern in observed allele frequencies.

4.3. Introducing Recombination

The results of §4.1 and §4.2 certainly support our claim of §3.1 that large scale extinction/recolonisation events could explain both long range correlations in allele frequencies and moderate levels of genetic diversity. However, we still lack observable measures to distinguish alternative models. For example, Fig. 2 rests on a pairwise

measure and could equally be derived from a stepping stone model with long range dispersal. A possible route to finding such measures exploits the third shortcoming of classical models that we pointed to in §3.1. Whereas under classical models all but tightly linked loci will evolve independently of one another, under our model spatial patterns of different - even unlinked - loci will be correlated.

We focus on the case of just two loci. Due to recombination, one offspring can inherit its types at the two loci from different parents. As a result, as explained in §1, the ancestry of a sample from the population is encoded in a complex branching and coalescing web in which branches arise through recombination and coalescences through shared ancestry.

There are very few mathematical studies of correlations across loci for spatially structured, selectively neutral populations. Wakeley & Lessard ([40]) work in the context of Wright's island model in which demes sit at the vertices of a complete graph. They compute the means, variances and covariances of coalescence times at the two loci, but this does not capture the two-dimensional spatial structure in which we are primarily interested. Zähle et al. ([39]), as an application of the results cited in §4.1, compute the probability of seeing a recombination before a coalescence, but do not study the probability that two recombinant lineages quickly coalesce again, causing correlations to remain strong. As a consequence they do not compute explicitly their measure of 'linkage disequilibrium' which involves the covariance in coalescence times across loci. De & Durrett ([41]) simulate an island model and the stepping stone model, and show that there are significant differences in their genealogies. In particular, the stepping stone structure increases the chance that the genealogies at the two loci are perfectly correlated (as indeed is confirmed by our results below).

To probe correlations in coalescence time across loci in our setting, we must extend our model.

In [8] we extend the spatial Λ-Fleming-Viot process of Definition 3.1 in two ways. The first is to incorporate recombination. The second addresses an obvious criticism of the models described so far: one would expect multiple 'founders' in a large scale extinction/recolonisation event, not just one. In fact, in [8] we allow multiple parents in events at any scale.

Our approach is a simple modification of the framework described in §4.1. Small and large events will be driven by two independent Poisson Point Processes. For any parameter L, we fix a fraction $r_L \in (0,1]$ of recombinants. During a small event (t,x,r,u), a random integer $N_s \geq 1$ is drawn from a fixed distribution, and N_s parents are chosen independently and uniformly at random within the area of the event. We write (a_i, b_i) for the types carried at the two loci by the ith parent. Then for $y \in B_r(x)$, (8) is replaced by

$$\rho^{(L)}(t,y,\cdot) = (1-u)\rho^{(L)}(t-,y,\cdot) + \frac{u(1-r_L)}{N_s}\sum_{i=1}^{N_s}\delta_{(a_i,b_i)} + \frac{ur_L}{N_s(N_s-1)}\sum_{i\neq j}\delta_{(a_i,b_j)}. \quad (13)$$

In words, a fraction $1-u$ of the local population remains unchanged and a fraction

$u(1-r_L)$ (resp., ur_L) is replaced by non-recombinant (resp., recombinant) offspring of the N_s parents. We assume that recolonisation is so rapid after a large scale extinction that the effects of recombination during large scale events can be ignored and so, for large events, (13) is replaced by the corresponding expression with $r_L = 0$. Of course the distribution of the number of parents chosen can differ between large and small events.

To understand correlations in patterns at the two different loci, we investigate the genealogical trees corresponding to these loci. The main result of [8] considers a sample of size two, but is easily extended to larger samples. It states that there is a critical sampling distance

$$D_L^* \approx L^\alpha \sqrt{1 + \frac{\log \rho_L}{r_L \rho_L}},$$

such that if our individuals are sampled at pairwise distances larger than D_L^*, the genealogical processes corresponding to each locus become independent as the torus side L tends to infinity. On the other hand, if our individuals are sampled at pairwise distances less than D_L^*, asymptotically one sees a first phase of complete correlation between the genealogies at different loci, followed by a sudden decorrelation. In other words, asymptotically, coalescence events occurring before a 'decorrelation threshold' (which can be described explicitly) are completely correlated at the two loci, but conditional on being greater than this threshold they are independent.

The key idea is one of 'effective recombination'. Because recombination events result in two ancestral lineages at a small spatial separation, we can expect that many of them will rapidly be followed by a coalescence (due to small events). We declare a recombination event to be *effective* if at least one of the lineages resulting from the event is hit by a large event before such a coalescence. The phase of complete correlation is one in which there are no effective recombinations. The transition to complete decorrelation is when effective recombination kicks in.

These results not only provide a very good understanding of the mechanisms leading to decorrelation, but also a good picture of the local correlations. As a consequence, they suggest appropriate tools for inference of the parameters of local evolution, and to test for the presence of large events impacting the genetic diversity of the population. These two points are the main goals of [42].

As concerns inference, it is clear that the number of parameters involved in the definition of the spatial Λ-Fleming-Viot process is much too large to hope for any exhaustivity. However, recalling the small number of relevant quantities in the Wright-Malécot formula (5) and the comparison presented in §4.2, one can try to recognise the combinations of parameters which, not only matter, but are also observable in data. Following Theorem 4.1, a quantity which one might call *effective population size* is the timescale over which the genealogy of a sample is given by the Kingman coalescent: for large L's,

$$N_e = \omega_L \propto \begin{cases} L^2 \log L & \text{if large events are rare,} \\ \frac{\rho_L}{L^{2\alpha}} L^2 \log L & \text{if large events are frequent.} \end{cases}$$

However, this is only the correct timescale for lineages that are sampled at very large distances from each other, and combines the contribution of both small and large events in the evolution of the population. As we can see from the beginning of the slope in Fig. 2, on much smaller spatial scales, small events may induce the coalescence of close-by lineages before they experience any large event. Consequently, one has to choose an appropriate sampling distance to disentangle the effects of events of different orders of magnitude.

Assuming that some biological considerations give us a good idea of the mean radius R_s of a small event, the strategy developed in [42] consists in sampling a few individuals at distances of the order of 10 or 100 times R_s, and considering their genealogies at many loci. We suppose that, during each reproduction event, each link between two adjacent loci can be broken by recombination with some small probability, while, with a smaller probability still, each site can mutate into a new state, never seen before. Using the analysis carried out in [8], in [42] we derive an approximation of the law of the length of *conserved sequences* between two individuals, i.e. of the number of consecutive loci that are in identical state in the two individuals. We show that this length has a geometric distribution, whose parameter is a function of the probability that two lineages, having just recombined, manage to escape far from each other for a time long enough that one of them may be hit by a mutation (which will then be the end of the conserved sequence). Although this function is rather complex, it can be used as the basis for a maximum likelihood approach to the inference of local evolutionary parameters.

4.4. Patterns of allele frequencies

Our results so far have focussed on the genealogical trees relating individuals sampled from far apart. We now turn to the patterns of allele frequencies generated by our model. We are interested in large scale structures, where the fine details of the construction won't influence the results, but in contrast to Theorem 4.1, not so large that all spatial information is lost.

From now on we revert to our basic model of Definition 3.1. In particular, the underlying geographical space is $\mathbb{R}^d$ rather than the torus. We specialise to just two types of individual which we label $\{0, 1\}$ and then our probability measure on type space can be replaced by

$$w(t, x) = \rho_t(x)(\{1\}),$$

the proportion of 1s at x at time t.

A rather unnatural feature of the model is that once a type is present, provided u is never 1, that type can never be lost: there will always be a trace of its presence. It is natural to ask whether, if a mutation is initially confined to a bounded region, it will have compact range (that is it will only ever be seen in a compact region) or if it will eventually be spread across the whole of $\mathbb{R}^2$. This question was resolved in an important special case in [43]. Saadi specialises to the case $\xi(dr, du) = \delta_r \otimes \delta_u$ for some $r \in (0, \infty)$ and $u \in (0, 1)$. Moreover, for convenience, he takes a slight modification

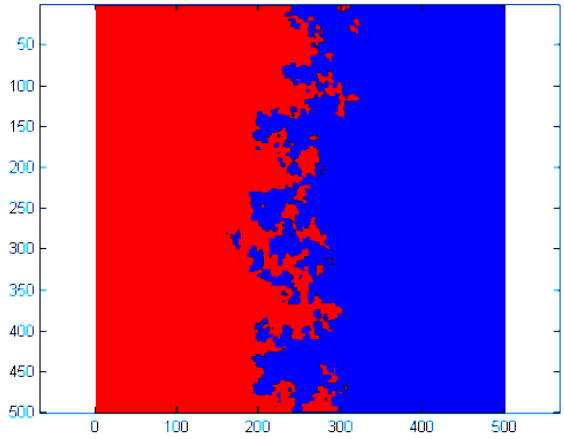


Figure 3. A snapshot of the interface between two types when we take $u = 1$ and a fixed radius of event sizes. The initial condition was a half plane of each type.

of our model in which the parent, instead of being sampled uniformly, is always at the centre of the ball affected by the event. He uses an elegant martingale argument to show that if a mutation is initially confined to a bounded region then its range is, with probability one, bounded. So for events of fixed radii, the range of a mutation is bounded. However, this doesn't provide any information about the shape of the region in which the mutation is eventually represented, nor about the extent to which we can expect to see different genetic types coexisting. In [7], a more detailed analysis of the 'interface' between different types is undertaken. The results are for the spatial Λ-Fleming-Viot process in $\mathbb{R}^d$ with two different types. We start from a half plane of type 1s and a (complementary) half plane of type 0s, that is we take the initial condition $w_0 = \mathbf{1}_{\mathbb{H}}$ where $\mathbb{H} = \{x \in \mathbb{R}^d : x_1 \leq 0\}$ (and x_1 is the first coordinate of x). In one spatial dimension, this reduces to the Heaviside function. Figure 3 shows a snapshot of the population if we take $\xi(dr, du) = \delta_r \otimes \delta_1$. Even with $u = 1$, so that $w(t, x)$ is the indicator function of a random set and there are no points in space at which both types are represented, the interface between the two types is very complex, but it is reasonable to hope that if we 'zoom out' and look over large spatial and temporal scales then some order might emerge.

The key tool in the analysis is moment duality. (Here we modify our notation for the expectation of a random variable and write $\mathbf{E}[X(t)|X(0) = x] = \mathbf{E}_x[X(t)]$.)

Theorem 4.3 (Special case of Theorem 4.2 and §4.2 of [5]) *Sample individuals from locations $x_1, \ldots, x_j$ and write $\xi_t^1, \ldots, \xi_t^{N_t}$ for the locations at time t of lineages evolving according to the spatial Λ-coalescent of §3.5. Then for each $\psi \in C\left((\mathbb{R}^d)^j\right) \cap L^1\left(dx^{\otimes j}\right)$ and every $t \geq 0$,*

$$\int_{(\mathbb{R}^d)^j} \psi(x_1, \ldots, x_j) \mathbb{E}\left[w(t, x_1) \cdots w(t, x_j) \middle| w(0, \cdot) = w_0(\cdot)\right] dx_1 \ldots dx_j$$

$$= \int_{(\mathbb{R}^d)^j} \psi(x_1, \ldots, x_j) \mathbb{E}\left[w_0(\xi_t^1) \cdots w_0(\xi_t^{N_t}) \middle| N_0 = j, \xi_0^1 = x_1, \ldots, \xi_0^j = x_j \right] dx_1 \ldots dx_j.$$

This is a sort of 'weak' moment duality. It implies that

$$\mathbb{E}\left[\prod_{i=1}^{j} w(t, x_i) \middle| w(0, \cdot) = w_0(\cdot) \right] = \mathbb{E}\left[\prod_{i=1}^{N_t} w_0(\xi_t^i) \middle| N_0 = j, \xi_0^1 = x_1, \ldots, \xi_0^j = x_j \right] \tag{14}$$

for Lebesgue almost every $(x_1, \ldots, x_j)$. This can be compared to the corresponding equation (3) for the stepping stone model.

We consider two cases. In the first we allow only small events, whereas in the second we also include large events.

CASE A. For the first result we take $\xi(dr, du) = \delta_r \otimes \delta_u$ for fixed $r \in (0, \infty)$ and $u \in (0, 1]$. A single ancestral lineage will evolve according to a random walk with bounded jumps. This suggests that, if we are looking for large-scale patterns, we should apply a diffusive rescaling, under which the motion of an ancestral lineage will converge to Brownian motion. We therefore set

$$w_t^n(x) = w(nt, \sqrt{n}x) \tag{15}$$

and start from $w_0 = \mathbf{1}_{\mathbb{H}}$.

Theorem 4.4 (Theorem 1.1 of [7]) *There exists a random space-time field $\{w_t^{(2)}(x), x \in \mathbb{R}^d, t \geq 0\}$ such that $w^n \to w^{(2)}$ as $n \to \infty$ in the sense of weak convergence of 'almost all' finite dimensional distributions. Equivalently $\rho^n \to \rho^{(2)}$ in distribution where $\rho^{(2)}$ has density $w^{(2)}$.*

Furthermore, writing $\{B_t^{(2)}\}_{t \geq 0}$ for a standard Brownian motion, there is a $\sigma^2 > 0$ such that if $p_2(t, x) = \mathbb{P}_x[B_{u\sigma^2 t}^{(2)} \in \mathbb{H}]$, then

(i) in $d = 1$, $w^{(2)}$ is a random field of correlated Bernoulli random variables such that $\mathbb{E}[w_t^{(2)}(x)] = p_2(t, x)$ almost everywhere;

(ii) in $d \geq 2$, $w^{(2)}$ is deterministic *and $w_t^{(2)}(x) = p_2(t, x)$ almost everywhere.*

The correlations between the Bernoulli random variables are given by an analogue of (14). There are two important observations:

(i) in one dimension there is no coexistence of types: at a given site $w_t^{(2)}(x) \in \{0, 1\}$;

(ii) the 'speed' of evolution is proportional to the parameter u.

Theorem 4.4 tells us that in $d = 1$, under the diffusive rescaling (15), the limit is the indicator function of a random set. Indeed $w_t^{(2)}$ is equal in distribution to $\mathbf{1}_{x \leq B_{u\sigma^2 t}^{(2)}}$. To see why, note that the dual process of coalescing lineages converges to a system of coalescing Brownian motions which coalesce instantaneously on meeting (c.f. §4.1). In one dimension, the Brownian motions can never cross and so it is impossible for a lineage of type 0 to start to the left of a lineage of type 1. In two dimensions, just as in the diffusive limit of the Kimura stepping stone model, since two independent Brownian motions will never meet, the system of allele frequencies reduces to heat flow.

With this very special choice of $\xi(dr, du)$, the model is very similar to the voter model. In [44], Cox, Durrett & Perkins showed that under a suitable rescaling, in $d \geq 2$ the voter model converges to superBrownian motion. The difference in their setting is that they take a *sparse* initial condition.

CASE B. We now investigate what happens when we introduce some large scale events. Since we are hoping for a nice scaling limit, we choose the rate of those events in such a way that the motion of a single ancestral lineage will scale to a symmetric stable process. One way to do this is to keep $u \in (0, 1]$ fixed and take

$$\mu(dr) = \frac{1}{r^{d+\alpha+1}} \mathbf{1}_{r \geq 1} dr \tag{16}$$

for $\alpha \in (1, 2)$ a fixed parameter (and d the spatial dimension). We then define

$$w_t^n(x) = w(nt, n^{1/\alpha} x).$$

Theorem 4.5 (Theorem 1.5 of [7]) *There exists a random space-time field* $\{w_t^{(\alpha)}(x), x \in \mathbb{R}^d, t \geq 0\}$ *such that* $w^n \to w^{(\alpha)}$ *as* $n \to \infty$ *in the sense of weak convergence of 'almost all' finite dimensional distributions. Equivalently* $\rho^n \to \rho^{(\alpha)}$ *in distribution where* $\rho^{(\alpha)}$ *has density* $w^{(\alpha)}$.

Furthermore, there is a symmetric α*-stable process* $B^{(\alpha)}$ *such that, for all* $t \geq 0$, $w^{(\alpha)}$ *is a random field of correlated Bernoulli random variables such that*

$$\mathbb{E}[w_t^{(\alpha)}(x)] = p_\alpha(t, x) := \mathbb{P}[B_{ut}^{(\alpha)} \in \mathbb{H}] \quad \textit{almost everywhere.}$$

Notice that once again the 'speed' is proportional to u. This time, however, the limit is stochastic in *all* dimensions. At first sight this is surprising. In Case A above (which we shall refer to as $\alpha = 2$), the limit in $d \geq 2$ was deterministic since the genealogy of a sample was asymptotically determined by independent Brownian motions which never meet. It is also the case that independent symmetric α-stable processes with index $\alpha \in (1, 2)$ will not meet in $d \geq 2$. However, the genealogy does not reduce to *independent* processes. As a result, the patterns of allele frequencies that we observe are quite different from those we would see in a stepping stone model with long range dispersal. Here we see local fixation of types, with strong geographic correlations. There we would see the analogue of Theorem 4.4(ii) with p_2 replaced by the corresponding expression for a symmetric stable process.

Let us try to understand heuristically why any pair of lineages will coalesce in a finite time even in $d \geq 2$.

Suppose that two lineages are at separation x. Take some $k > 1$. For any $r > kx$ there is a region with volume proportional to r^d such that any event of radius r whose centre lies in the region will also contain both lineages. Thus the total rate at which the lineages are both 'caught' by a big event is at least

$$\text{Const.} \int_{kx}^{\infty} r^d \mu(dr) = \text{Const.} \int_{kx}^{\infty} \frac{r^d}{r^{d+\alpha+1}} dr \propto \frac{1}{x^\alpha}.$$

If the lineages were evolving independently, then their separation at time t would satisfy $x(t) \sim t^{1/\alpha}$. Of course they are not evolving independently, but censoring events that are big enough to capture both of them should, if anything, slow down the rate at which they move apart and so one estimates that they will be captured by a big event at rate at least $C/(x(t))^\alpha \sim C/t$ and since $\int^\infty 1/tdt = \infty$ they will coalesce in finite time almost surely. Of course this is far from a proof, but in [7] it is replaced by a rigorous argument.

Unlike Case A, in $d = 1$ lineages *can* now jump over one another. The limiting object will still look like the indicator function of a random set, but the structure of that set will be complex. Fig. 4 shows some simulations. Essentially what happens is that from time to time a large event falls on a region in which we see both 0s and 1s, resulting in an ephemeral state in which there are non-trivial proportions of both types throughout the affected region, but that region is rapidly resolved by the much more frequent small events into regions of no coexistence. In the limit we only ever see the result of this process of resolution. Thus, even when $u < 1$, the limit is the indicator function of a random set. Our methods, which are based entirely on the moment duality of Theorem 4.3, are not strong enough to capture any information about the dynamics of the limit and nor do we have any more detailed results on the structure of the random set at a fixed time. Fig. 5 shows a simulation in two dimensions. Although, at first sight, of more mathematical than biological interest, these results show that predictions of our model can differ greatly from those of classical stepping stone models and, in particular, point to the fact that the presence of large scale events can be expected to result in very complex spatial patterns of allele frequencies.

When $\alpha = 2$ and $d = 1$, the rescaling (15) resulted in a Heaviside limit. We have several times made the analogy between this special form of the model and the stepping stone model and as we said in §2.3, under the diffusive scaling, in one dimension the stepping stone model rescales to the stochastic p.d.e.

$$dp = \frac{1}{2}\Delta p dt + \sqrt{\gamma p(1-p)}W(dt, dx). \tag{17}$$

The difference here is that we held the parameter u fixed. Recall from §3.6 that u should be thought of as inversely proportional to neighbourhood size and from §2.1 that the form of the genetic drift in the stepping stone model arises from very large neighbourhood size. This suggests that if we were to allow u to tend to zero as $n \to \infty$ (and so neighbourhood size to tend to infinity) we could obtain the stochastic p.d.e. (17) as a rescaling limit of our model. This will be a special case of the results of §5.3, but there we should like to work in a more general setting which incorporates another evolutionary force, natural selection.

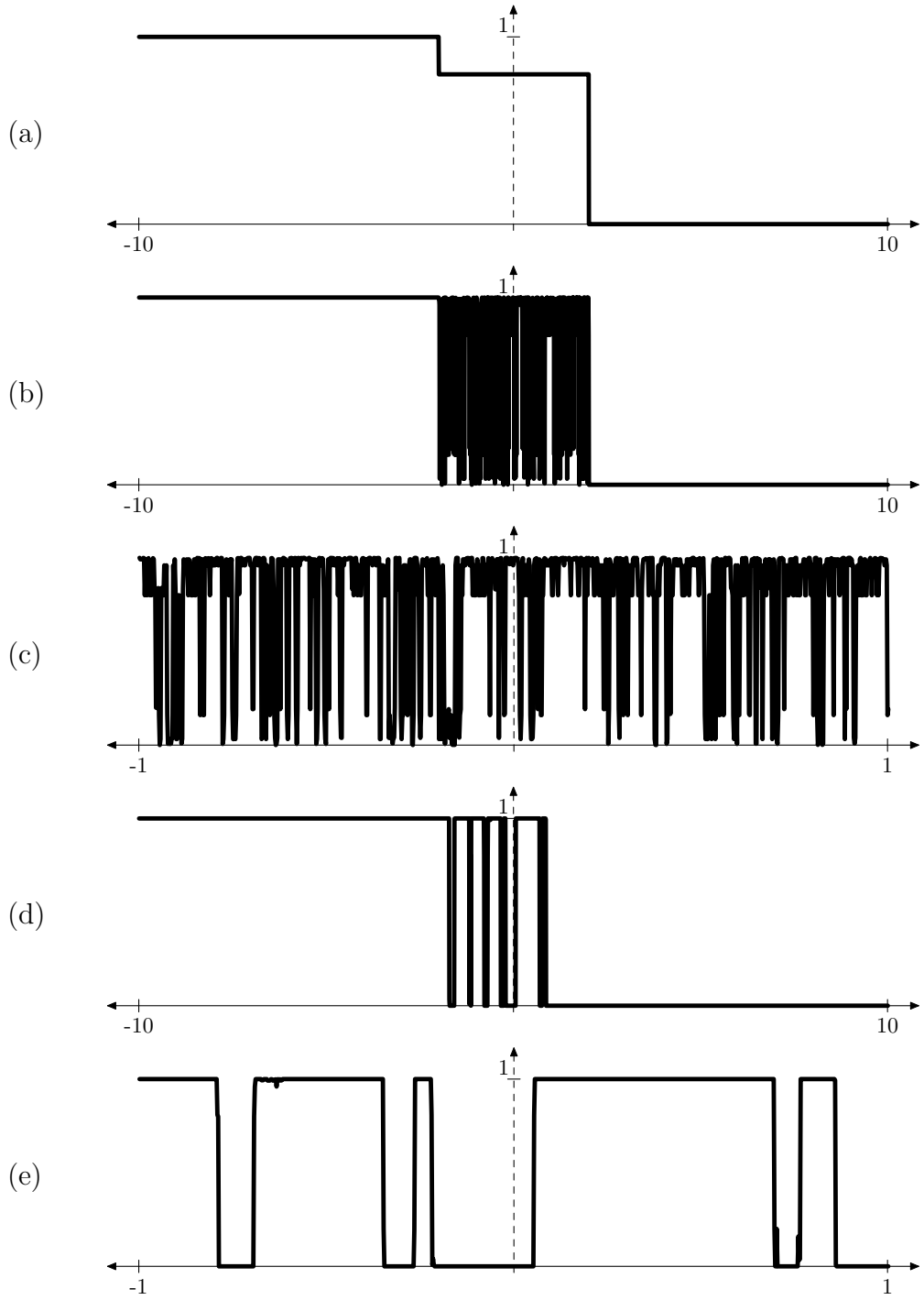
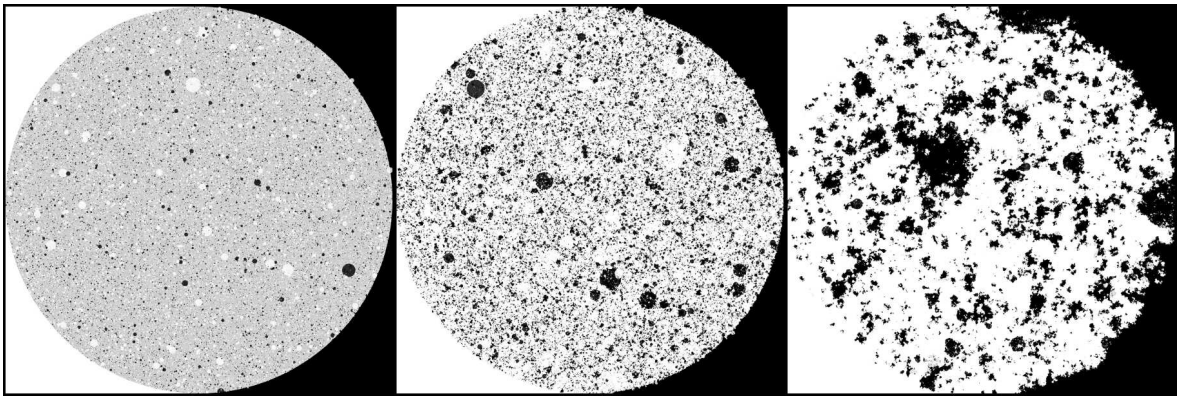


Figure 4. Case B of §4.4 in $d = 1$ on a line of length 20. (a) initial conditions; (b) After 100 events: full range; (c) After 100 events: zooming in; (d) After 10^6 events: full range; (e) after 10^6 events: zooming in. Parameters: $u = 0.8$, $n = 10^4$ and $\alpha = 1.3$.



(a) (b) (c)

Figure 5. Case B of §4.4 in two dimensions after (a) 10^5; (b) 10^6; and (c) 10^7 events. We have a square range of edge 8, and the initial patch is a circle of radius 4 with frequency 0.8 (white is frequency 1, black is 0). The model parameters are: $u = 0.8$, $\alpha = 1.3$ and $n = 10^3$.

5. Introducing selection

5.1. Selection in the Wright-Fisher and Moran models

There are many ways of introducing selection into a model for allele frequencies and indeed many different forms of selection that one might wish to model. We concentrate on the simplest type of directional selection.

For the Wright-Fisher model, one would typically weight the choice of parental type. Thus, if at generation t a proportion p of the population is of type a, then the probability that the parent is of type a is $p/(1 + s(1 - p))$ and the probability that it is type A is $(1 + s)(1 - p)/(1 + s(1 - p))$. For $s > 0$, this gives an advantage to type A individuals. If we wish to arrive at a diffusion limit, we take $s \propto \frac{1}{N}$ and then in the limit as $N \to \infty$, on the diffusion timescale, we arrive at

$$dp = -\sigma p(1 - p)dt + \sqrt{p(1 - p)}dW_t \tag{18}$$

as a model for allele frequencies.

It is straightforward to find a moment dual for this system. First apply Itô's formula to p_t^n for n fixed to obtain

$$d(p_t^n) = n\sigma \left(p_t^{n+1} - p_t^n\right) dt + \binom{n}{2} \left(p_t^{n-1} - p_t^n\right) dt + dM_t$$

where M_t is a martingale. Now choose $\{n(t)\}_{t\geq 0}$ to be the birth-death process with rates

$$\begin{cases} n \mapsto n + 1 & \text{at rate } n\sigma, \\ n \mapsto n - 1 & \text{at rate } \binom{n}{2}. \end{cases}$$

Then it is elementary to check that

$$\frac{d}{ds}\mathbb{E}\left[p(s)^{n(t-s)}\right] = 0$$

so that

$$\mathbb{E}\left[p(t)^{n(0)}\right] = \mathbb{E}\left[p(0)^{n(t)}\right], \tag{19}$$

which should be compared to equation (1).

Another way to obtain the same limiting model for allele frequencies is via a Moran model. Recall that in the neutral Moran model, at rate $\binom{N}{2}$ a random pair is selected from the population, one dies and the other reproduces. To mimic the effect of the selection in the Wright-Fisher model above, we bias the choice of parent. Thus if the pair picked consists of one type a and one type A individual, then with probability $(1+s)/2$ it is the type A that reproduces. It is sometimes convenient to think of there being two types of event: neutral events, which occur at rate $(1-s)\binom{N}{2}$, and 'potential selective events' which happen at rate $s\binom{N}{2}$. At a potential selective event, if the pair of individuals chosen consists of one a and one A then with probability *one* it is the A that reproduces. This is simply a reformulation and does not change the process. Once again taking $s \propto 1/N$ and letting $N \to \infty$ we obtain equation (18) as the limiting model for allele frequencies.

The moment duality (19) is easy to understand in the Moran context. If we take a sample of size n from the population at time t, then the left hand side of (19) is the probability that they are all of type a. Neutral events correspond to coalescence of ancestral lineages as in §2.1, but when one of our lineages is hit by a potential selective event, which happens at rate σ, in order for that lineage (which is an offspring of the event) to be type a, it must be that *both* the potential parents sampled in the event were of type a. To confirm this, we must now trace back *two* ancestral lineages, hence the birth in the moment dual.

The branching and coalescing structure swept out as we trace ancestry in this way is known as the *ancestral selection graph*. It was introduced in work of Krone & Neuhauser ([45, 46]). They included mutation between the two types a and A in the population, thus ensuring the existence of a stationary distribution for the process of allele frequencies. They were then able to recover the true genealogical tree of a random sample from the population (at stationarity) using the ancestral selection graph. Our argument only provides a weaker result.

5.2. A simple model

There are also many ways to introduce selection into our spatial model for allele frequencies. Any parameter of the model could depend on genotype. Moreover, parameters might depend both on the genotype of any individual involved (e.g. the parent) and on the local allele frequency. However, here we focus on an analogue of the directional selection of §5.1.

A natural model mimics our approach in the Wright-Fisher model: rather than selecting a parent uniformly at random from those present in the region affected by an event, we choose in a weighted way. Thus if the proportion of a-alleles in the region immediately before the event is $\bar{w}$, then the probability that the parent chosen is of type

a will be $\bar{w}/(1+s(1-\bar{w}))$. Notice that this does not require s to be small, and so can, in particular, be used to model strong selection. Moreover, there is no more computational effort in modelling strong selection than in modelling weak selection. However, since our aim at this stage is to make contact with established models, we focus on weak selection. In this case, we rewrite the probability of choosing a type a parent as

$$\frac{\bar{w}}{1+s(1-\bar{w})} = \bar{w} - s\bar{w}(1-\bar{w}) + \mathcal{O}(s^2) = \bar{w}(1-s) + s\bar{w}^2 + \mathcal{O}(s^2) \quad (20)$$

and that of choosing a type A parent as

$$\frac{(1+s)(1-\bar{w})}{1+s(1-\bar{w})} = (1-s)(1-\bar{w}) + s(1-\bar{w})^2 + 2s\bar{w}(1-\bar{w}) + \mathcal{O}(s^2).$$

Up to an error of $\mathcal{O}(s^2)$, we can then decompose our Poisson process of events into two types. Neutral events, in which the parent is simply chosen at random, are driven by a Poisson process with intensity $(1-s)dt \otimes dx \otimes \xi(dr, du)$. Additional 'potential selective events' are driven by a second Poisson point process with intensity $sdt \otimes dx \otimes \xi(dr, du)$. At such an event *two* potential parents are chosen. If at least one of them is of type A, then we will have a type A parent, otherwise the parent is type a. This of course mirrors what we saw in the Moran model of §5.1 and, in much the same way as there, we can identify a branching and coalescing (weak) moment dual. We can develop the power series in (20) to higher order in s to achieve greater accuracy, but truncating at the term corresponding to s^k will result in a moment dual in which lineages branch into $k+1$.

5.3. One dimension - recovering a stochastic p.d.e.

There is a huge body of work on the fate of a selectively advantageous (or disadvantageous) allele in a spatially structured population, but the difficulties that we encountered with the pain in the torus are reproduced here. In two (or more) spatial dimensions, classical models either treat the population as subdivided or they ignore genetic drift and are deterministic. In one spatial dimension, most work focuses on either deterministic or stochastic versions of the Fisher-KPP equation. This takes the form

$$dp = \frac{1}{2}\Delta p dt - \sigma p(1-p)dt + \epsilon\sqrt{p(1-p)}W(dt, dx). \quad (21)$$

When $\epsilon = 0$ this equation is extremely well understood. It was already studied by Fisher as a model for the way in which a favoured mutation spreads through a population. The equation with noise admits a stochastic travelling wave solution, but the effect of the noise is to slow down the spread of the mutation and there has been much recent interest in understanding exactly how ϵ influences the wavespeed ([47, 48]).

Our aim in this section is to explain how one can obtain (21) as a scaling limit from our model. Details of the calculation can be found in [49]. Here we just explain how to find the correct scaling. The purpose of this is to identify the regime in which our model will behave in the same way as classical stepping stone models and therefore the

appropriate parameter regimes to consider if we wish to emulate the classical approach. Different scalings will, of course, lead to different limits.

Suppose that the events in our model all have the same radius r, say. As intimated at the end of §4.4, in order to obtain a stochastic p.d.e. in the limit, we expect to have to let the parameter u tend to zero as part of our rescaling and so we set $u_n = 1/n^\gamma$, where γ is to be chosen. Consider the rescaled process

$$w(nt, n^\beta x)$$

where β is also to be determined. From §4.4 the 'speed' of an ancestral lineage scales with u, thus in order that the combined effect of these scalings is that ancestral lineages, individually, follow Brownian motions (which is what we expect if we are to see the Laplacian in (21)), we need that $nu_n = n^{2\beta}$. In other words

$$2\beta + \gamma = 1.$$

We also know that in the Moran model, the ratio of the rate of reproductive events to potential selective events must be proportional to n if we are to obtain a diffusion limit. If we set $s_n = \sigma/n^\delta$, the analogous condition here is $nu_n \propto n^\delta$ (the left hand side being the effective timescale). Putting this together, we see that we should define

$$w_t^n(x) = w(nt, n^{1/3}x), \quad \text{with } u_n = \frac{u}{n^{1/3}}, \quad s_n = \frac{\sigma}{n^{2/3}}.$$

It is proved in [49] that this does indeed yield (21) in the limit as $n \to \infty$.

One can perform analogous calculations with events of fixed radius replaced by events governed by the measure μ of (16). If one uses the same parameter s irrespective of the size of the event, then one obtains the stochastic p.d.e.

$$dp = \Delta_\alpha p dt - \sigma p(1-p)dt + \epsilon\sqrt{p(1-p)}W(dt, dx)$$

in the rescaling limit, where Δ_α is the infinitesimal generator of a symmetric stable process of index α. Notice, in particular, that the large scale events don't affect the term due to selection or, unlike the parameter regime of §4.4, the form of the noise. This reflects the fact that we must take $u_n \to 0$ in order to mimic the stepping stone model.

Since the stochastic Fisher KPP equation has no solution in dimension two (or more), the corresponding rescaling can only yield a solution to the deterministic equation in higher dimensions. However, for finite n our model makes perfectly good sense in *any* dimension and provides a natural framework in which to study selection in a spatially distributed population. The scaling of parameters above guides us if we wish to replicate classical results. Figure 6 shows a snapshot of a population evolving according to this model for finite n. An alternative approach to using our model is of course to use the Kimura stepping stone model (2) with an additional term $sp_i(1-p_i)dt$ on the right hand side to model selection. Notice that in contrast to that setting, in Fig. 6 the range of

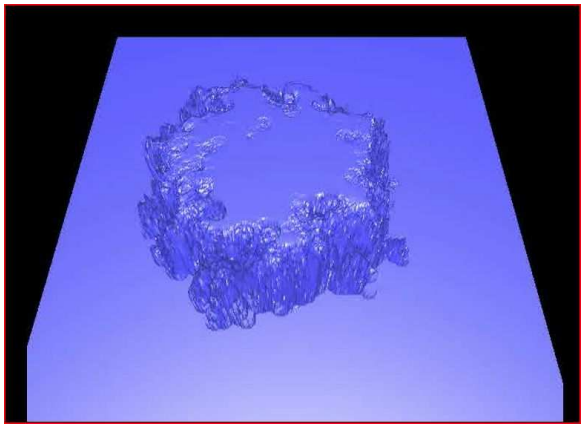


Figure 6. A snapshot of the wave of advance of a selectively advantageous mutation under the model of §5.2.

the compact allele is compact at all times. In the stepping stone model it is present at all sites of $\mathbb{Z}^2$ at any positive time. This is an artefact of the discrete deme structure, the same is not true of solutions to (21). In order to obtain a picture that more closely resembles what one sees for the stepping stone model, we must take a smaller value of u. Although much of the work on equation (21) has been devoted to understanding the behaviour for small values of the parameter ϵ, corresponding to very weak genetic drift, there has been some interest in understanding the strong noise limit ([50]). In order to understand strong genetic drift in our framework we simply keep u macroscopic. In that setting, if we take a strong selection limit we obtain a growth model which can be compared to that of, for example, [51].

6. Related models

In this brief section, let us mention some closely related models that have been studied by others.

There is a huge literature on λ-coalescents (without space). We refer to [52] for an excellent review. Although we have used the term *spatial* Λ*-coalescent* for the process of coalescing ancestral lineages of §3.5, that name was already used by Limic and Sturm [53] for a somewhat different process. Limic and Sturm suppose that the population is subdivided into discrete demes. Ancestral lineages migrate between demes and whenever there is more than one lineage in a deme they are allowed to coalesce according to a Λ-coalescent. This is a natural extension of the structured coalescent dual to Kimura's stepping stone model which we described in §2.2. They also follow [39] in considering the effect of sampling at random from a large torus. They work in dimensions $d \geq 3$ (althought the same results would hold in $d = 2$) and, as one expects, recover a Kingman coalescent with an effective population size. However, since they only allow coalescence within demes, they cannot recover the multiple merger coalescents that we obtain in Theorem 4.2. The time to coalescence in their setting is dominated by the time taken for

two lineages to meet in the same deme. This dictates the timescale of their coalescent and, on this timescale, the chance of ever seeing three or more lineages in the same deme is negligible. This also means that they don't see the reduction in effective population size that results from our large events.

There have been many attempts to construct continuum stepping stone models. The approach of Evans ([33]) that we used to prove existence of our model was originally devised to construct stepping stone models in continuous geographical space. The idea is to first construct a coalescent process, in which ancestral lineages evolve independently until they meet at which point they coalesce, and then show the existence of a forwards in time model which has this coalescent as its dual. In [54], Liang uses the same technology to construct two different 'continuum stepping stone models' for which the geographical space is a circle and coalescence of ancestral lineages is no longer instantaneous, but again occurs only when lineages coincide. As far as we know, ours is the first satisfactory approach in dimensions bigger than one, where, under classical models of dispersal, independently evolving ancestral lineages fail to meet.

There are a variety of models that incorporate extinction/recolonisation events. For example in [55], Kang et al. consider a stepping stone model with extinction/recolonisation events. However, in their setting such events only affect one deme at a time and consequently the long range correlations in allele frequencies that we see in the presence of large events will not be present. Taylor & Véber ([56] consider island models in which extinction events affect multiple demes, but their underlying geographical space is much simpler than that considered here.

Since our spatial Λ-coalescent consists of a system of dependent coalescing jump processes, exact calculations of quantities of interest are only really tractable when one takes suitable rescaling limits as we have done here. On the other hand, it is reasonable to hope that if we replace geographical space with a tree-like structure, then exact calculations will be possible without such a limiting procedure. A natural candidate is the hierarchical group which is often used to mimic higher dimensional spaces. The spatial Λ-coalescent on the hierarchical group is the subject of [57]. In [58], Freeman also works on a geographical space with a hierarchical structure and considers the corresponding spatial Λ-coalescent. The main novelty in his construction is that one can allow 'individuals' to be hit by events at an infinite rate. The space is sufficiently simple that interesting questions can be addressed through branching processes in varying environments, but sufficiently complex that one sees phenomena not present in the non-spatial setting.

There is a huge amount of work, especially in the physics literature, on spatial waves of invasion that goes well beyond what we have cited here. We refer to [59] for a review. For a more general survey of applications of methods of statistical physics in evolutionary biology we refer to [60].

7. Directions for future research

Our work fulfils two distinct rôles. First, it provides a convenient model of movement and reproduction in a truly continuous population. It can be simulated efficiently forwards in time: on a desktop computer, it is feasible to follow 10^8 individuals, or (in the limit of infinite density) up to 10^3 genotypes. If there is no selection, lineages can be simulated back through time in a coalescent process that corresponds precisely to the forwards model. Second, we can include events over a range of scales, allowing both for individual reproduction, and for large events that affect whole regions; such large-scale processes are essential to describe real populations. It is the relative rôle of large and small-scale events that gives greatest scope for future research: in particular, how to make inferences from genetic data, and in understanding selection over different scales.

A substantial community is devoted to inferring population structure and demographic history from genetic data, for its own interest, for practical population management, and to provide a null model against which alleles that affect quantitative traits can be detected. There are a plethora of methods, ranging from qualitative inference from genealogies ('phylogeography') to quantitative tests, which may be essentially descriptive (e.g. spatial autocorrelation; [30]) or based on a specific model. In the latter case, we can distinguish two classes of model: individual reproduction, with different loci evolving more or less independently (as in Wright and Malécot), or models where distinct populations diverge and mix (e.g.[61, 62]). Recently, models of isolation-with-migration that combine elements of both have received attention ([63]). Our model provides a common framework that has the potential to bring together these different approaches: it includes both local gene flow and large-scale population movement, but without imposing artificial boundaries. One focus of our research is to provide a coherent understanding of the diverse methods in use at present.

It will clearly not be possible to infer all the parameters of our model, even with the abundant genomic data that are now available. We therefore have two specific aims in our work on inference: first, to find simple parameter combinations that capture the main features of the process, and second, to find robust ways to reject null hypotheses: for example, that there is only local reproduction, or that selection does not act on particular loci. Here, we may also be able to exploit the fact that with small-scale events, lineages either coalesce locally in the recent past, or escape into the distant past - a generalisation of Wakeley's 'collecting' and 'scattering' phases (e.g. [64, 65]), first identified for the island model.

If any parameter of our model depends on individual genotype, or on genotype frequencies in the local population, then selection will act. There are several fundamental questions here. What is the relative importance of selection on different parameters (i.e., on death rate, or probability of being chosen as a parent) and at different scales? Is there any necessary relation between the classes of event that are responsible for most coalescence and random drift, and the classes on which selection is most effective? (For example, local reproduction might have negligible effect on

random drift, and yet cause the most effective selection.) Which processes most affect the 'tension zones' that bound diverging populations?

The framework can also be used to investigate the interaction between random drift, gene flow and selection, since it provides a mathematically well-defined model of fine-scale reproduction. In particular, we are interested in the spread of a favourable allele through a spatially continuous population, which is dominated by random fluctuations at the leading edge ([66, 67]). This process distorts variation at linked neutral sites. Can such spatially extended sweeps be distinguished, either singly, or through their cumulative effect? Is the signature of spatial hitch-hiking similar to that of large-scale extinction events?

An abundance of DNA sequence data is now becoming available for a wide variety of species - almost all of which are spread over a two-dimensional range. We believe that our framework for modelling reproduction over a range of spatial scales will help us to understand what such data has to tell us about the evolution of species in nature.

Acknowledgments

We should like to thank Jerome Kelleher for the simulations of Figs 2 and 4–6 and Habib Saadi for the picture in Fig. 3. Steve Evans helped with the calculation of §3.3. The work of AME was supported in part by EPSRC grants EP/E065945/1 and EP/I01361X/1, N.B. was supported by the European Research Council grant 250152 and A.V. was supported by the chaire Modélisation Mathématique et Biodiversité of Veolia Environment - École Polytechnique - Museum National d'Histoire Naturelle - Fondation X and by the ANR project MANEGE (ANR-09-BLAN-0215).

References

[1] N Takahata. Allelic genealogy and human evolution. *Mol. Biol. Evol.*, 10(1):2–22, 1993.

[2] H Li and R Durbin. Inference of human population history from individual whole-genome sequences. *Nature*, 475:493–496, 2011.

[3] A M Etheridge. Drift, draft and structure: some mathematical models of evolution. *Banach Center Publ.*, 80:121–144, 2008.

[4] N Berestycki, A M Etheridge, and M Hutzenthaler. Survival, extinction and ergodicity in a spatially continuous population model. *Markov Processes and Related Fields*, 15:265–288, 2009.

[5] N H Barton, A M Etheridge, and A Véber. A new model for evolution in a spatial continuum. *Electron. J. Probab.*, 15:162–216, 2010.

[6] N H Barton, J Kelleher, and A M Etheridge. A new model for extinction and recolonization in two dimensions: quantifying phylogeography. *Evolution*, 64(9):2701–2715, 2010.

[7] N Berestycki, A M Etheridge, and A Véber. Large-scale behaviour of the spatial Λ-Fleming-Viot process. *Ann. Inst. H. Poincaré (to appear)*, arXiv:1107.4254v2[math.PR], 2012.

[8] A M Etheridge and A Véber. The spatial Lambda-Fleming-Viot process on a large torus: genealogies in the presence of recombination. *Ann. Appl. Probab. (to appear)*, arXiv:1106.4050v2[math.PR], 2012.

[9] W J Ewens. *Mathematical Population Genetics I. Theoretical introduction.* Springer-Verlag, 2004.

[10] R Durrett. *Probability models for DNA sequence evolution. 2nd Edition.* Springer-Verlag, 2008.

[11] A M Etheridge. Some mathematical models from population genetics. In *École d'été de probabilités de Saint Flour XXXXIX-2009*, volume 2012. Springer-Verlag, 2011.

[12] N H Barton, D E G Briggs, J A Eisen, D B Goldstein, and N H Patel. *Evolution*. Cold Spring Harbour Press, 2007.

[13] S Karlin and H M Taylor. *A second course in stochastic processes*. Academic Press, 1981.

[14] P A P Moran. Random processes in genetics. *Proc. Camb. Phil. Soc.*, 54:60–71, 1958.

[15] P J Donnelly and T G Kurtz. A countable representation of the Fleming-Viot measure-valued diffusion. *Ann. Probab.*, 24:698–742, 1996.

[16] W H Fleming and M Viot. Some measure-valued Markov processes in population genetics. *Indiana Univ. Math. J.*, 28:817–843, 1979.

[17] S N Ethier and T G Kurtz. Fleming-Viot processes in population genetics. *SIAM J. Control Optim.*, 31:345–386, 1993.

[18] J E Taylor. The genealogical consequences of fecundity variance polymorphism. *Genetics*, 182(3):813–837, 2009.

[19] S Wright. Isolation by distance. *Genetics*, 28:114–138, 1943.

[20] M Kimura. Stepping stone model of population. *Ann. Rep. Nat. Inst. Genetics Japan*, 3:62–63, 1953.

[21] M C Whitlock and N H Barton. The effective size of a subdivided population. *Genetics*, 146:427–441, 1997.

[22] G Malécot. *Les Mathématiques de l'hérédité*. Masson et Cie, Paris, 1948.

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

[24] B M Bolker and S W Pacala. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theor. Pop. Biol.*, 52(3):179–197, 1997.

[25] R Law, D J Murrell, and U Dieckmann. Population growth in space and time: spatial logistic equations. *Ecology*, 84(2):252–262, 2003.

[26] A M Etheridge. Survival and extinction in a locally regulated population. *Ann. Appl. Probab.*, 14(1):188–214, 2004.

[27] M Hutzenthaler and A Wakolbinger. Ergodic behavior of locally regulated branching populations. *Ann. Appl. Probab.*, 17(2):474–501, 2007.

[28] M Birkner and A Depperschmidt. Survival and complete convergence for a spatial branching system with local regulation. *Ann. Appl. Probab.*, 17:1777–1807, 2007.

[29] N H Barton, F Depaulis, and A M Etheridge. Neutral evolution in spatially continuous populations. *Theor. Pop. Biol.*, 61:31–48, 2002.

[30] B K Epperson. *Geographical genetics*. Princeton University Press, Princeton, N.J., 2003.

[31] R C Lewontin. *The genetic basis of evolutionary change*. Columbia University Press, New York., 1974.

[32] J C Avise. *Molecular markers, natural history and evolution*. Sinauer Press, Sunderland, Massechussetts, 2004.

[33] S N Evans. Coalescing markov labelled partitions and a continuous sites genetics model with infinitely many types. *Ann. Inst. Henri Poincaré*, 33:339–358, 1997.

[34] P J Donnelly and T G Kurtz. Particle representations for measure-valued population models. *Ann. Probab.*, 27:166–205, 1999.

[35] J Pitman. Coalescents with multiple collisions. *Ann. Probab.*, 27:1870–1902, 1999.

[36] S Sagitov. The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.*, 26:1116–1125, 1999.

[37] J Bertoin and J-F Le Gall. Stochastic flows associated to a coalescent process. *Prob. Theor. Rel. Fields*, 126:261–288, 2003.

[38] Amandine Véber. *Limit theorems for some spatial branching and coalescent processes*. Ph D thesis, Université Paris-Sud, 2009.

[39] I Zähle, J T Cox, and R Durrett. The stepping stone model ii: genealogies and the infinite sites

model. *Ann. Appl. Probab.*, 15:671–699, 2005.

[40] J Wakeley and S Lessard. Theory of the effects of population structure and sampling on patterns of linkage disequilibrium applied to genomic data from humans. *Genetics*, 164:1043–1053, 2003.

[41] A De and R Durrett. Stepping stone spatial structure causes slow decay of linkage disequilibrium and shifts the site frequency spectrum. *Genetics*, 176:969–981, 2007.

[42] N H Barton, A M Etheridge, J Kelleher, and A Véber. Inference for the spatial Λ-Fleming-Viot process. *In preparation*, 2012.

[43] H Saadi. *Λ-Fleming-Viot processes and their spatial extensions*. DPhil thesis, University of Oxford, 2011.

[44] J T Cox, R Durrett, and E A Perkins. Rescaled voter models converge to super-brownian motion. *Ann. Probab.*, 28:185–234, 2000.

[45] S M Krone and C Neuhauser. Ancestral processes with selection. *Theor. Pop. Biol.*, 51:210–237, 1997.

[46] C Neuhauser and S M Krone. Genealogies of samples in models with selection. *Genetics*, 145:519–534, 1997.

[47] E Brunet and B Derrida. Effect of microscopic noise on front propagation. *J Statist. Phys.*, 103(1-2):269–282, 2001.

[48] C Mueller, L Mytnik, and J Quastel. Effect of noise on front propagation in reaction-diffusion equations of KPP type. *Inv. Math.*, 184(2):405–453, 2011.

[49] A M Etheridge, A Véber, and F Yu. Rescaling limits of the spatial Lambda-Fleming-Viot process with selection. *In preparation*, 2012.

[50] O Hallatschek and K Korolev. Fisher waves in the strong noise limit. *Phys Rev Lett*, 103:108103, 2009.

[51] O Hallatschek and D Nelson. Life at the front of an expanding population. *Evolution*, 64:193–206, 2010.

[52] N Berestycki. Recent progress in coaelscent theory. *Ensaios Matemáticos*, 16, 2009.

[53] V Limic and A Sturm. The spatial Lambda-coalescent. *Electron. J. Probab.*, 11(15):363–393, 2006.

[54] R H Liang. *Two continuum-sites stepping stone models in population genetics with delayed coalescence*. Ph D thesis, UC Berkeley, 2009.

[55] H-C Kang, S M Krone, and C Neuhauser. Stepping stone models with extinction and recolonization. *Ann. Appl. Probab.*, 5:1025–1060, 1995.

[56] J Taylor and A Véber. Coalescent processes in subdivided populations subject to recurrent mass extinctions. *Electron. J. Probab.*, 14:242–288, 2009.

[57] A Greven, F den Hollander, S Kliem, and A Klimovsky. *In preparation*, 2012.

[58] N Freeman. Phas transitions in a spatial coalescent. *arXiv:1109.4363*, 2012.

[59] W van Saarloos. Front propagation into unstable states. *Phys. Rep.*, 386:29–222, 2003.

[60] H de Vladar and N H Barton. The contribution of statistical physics to evolutionary biology. *Trends Ecol Evol*, 26(8):424–432, 2011.

[61] L L Cavalli-Sforza and A W F Edwards. Phylogenetic analysis: models and estimation procedures. *Am. J. Hum. Genet.*, 19:233–257, 1967.

[62] J K Pritchard, M Stephens, and P J Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155:945–959, 2000.

[63] C Pinho and J Hey. Divergence with gene flow: models and data. *Annual Review of Ecology, Evolution and Systematics*, 41:215–230, 2010.

[64] J Wakeley. The coalescent in an island model of population subdivision with variation among demes. *Theor. Pop. Biol.*, 59:133–144, 2001.

[65] J Wakeley and N Aliacar. Gene genealogies in a metapopulation. *Genetics*, 159:893–905, 2001.

[66] R A Fisher. The wave of advance of advantageous genes. *Ann. Eugenics*, 7:355–369, 1937.

[67] O Hallatschek. The noisy edge of traveling waves. *Proc. Natl. Acad. Sci.*, 108(5):1783–1787, 2011.