STOCHASTIC MODELS IN BIOLOGICAL SCIENCES
BANACH CENTER PUBLICATIONS, VOLUME 80
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2008

DRIFT, DRAFT AND STRUCTURE: SOME MATHEMATICAL MODELS OF EVOLUTION

ALISON M. ETHERIDGE
Department of Statistics, University of Oxford, UK
E-mail: etheridg@stats.ox.ac.uk

Abstract. Understanding the evolution of individuals which live in a structured and fluctuating environment is of central importance in mathematical population genetics. Here we outline some of the mathematical challenges arising from modelling structured populations, primarily focussing on the interplay between forwards in time models for the evolution of the population and backwards in time models for the genealogical trees relating individuals in a sample from that population. In addition to classical models we describe a special case of a new model introduced in very recent work with Nick Barton. A number of directions for future research are suggested.

1. Introduction. If we are to make inferences from genetic data then we must understand how genes evolve in a structured and fluctuating population. Two types of structure are important. First, individuals live in a particular spatial location and their rate of reproduction depends on the local environment and competition from other individuals living nearby. Second, genes are embedded in different *genetic* backgrounds—because genes are organised on chromosomes and chromosomes in turn are grouped into individuals, different genes do not evolve independently of one another. In particular, the evolution of a gene which is itself selectively neutral in that it does not confer any particular advantage or disadvantage to the organism that carries it can nonetheless be influenced by natural selection acting on other genes in that same organism.

Modelling structured populations presents challenging mathematical problems and our aim here is to outline some of the existing approaches and their drawbacks. Of particular importance is the interplay between forwards in time models for the evolution of a population and backwards in time models for the *ancestry* of a population. We shall mostly be concerned with spatial structure rather than genetic structure, but we shall

2000 *Mathematics Subject Classification*: Primary 60K35, 92D10, 92D15, 92D40; Secondary 60G52, 60G55, 60G57.

Key words and phrases: spatial population models, extinction-recolonisation, coalescents.

Research supported in part by EPSRC grants GR/T19537/01, GR/S41203/01

The paper is in final form and no version of it will be published elsewhere.

argue that some of the mathematics that arises in the study of populations undergoing selection could also throw new light on populations evolving in a spatial continuum and in this context we will describe a new model which overcomes some of the shortcomings of existing models. However, there remain many more questions than answers.

In §2 we describe Kingman's coalescent model for the ancestry of an idealised population and introduce the notion of genetic drift. In a *neutral* theory of evolution in which all individuals in the population are equally fit, it is this genetic drift, which corresponds to random reproduction in the population, that drives the changes in frequency of different genetic types. But in such a theory this basic model predicts that genetic variability in a population is determined by the size of the population and this is simply not consistent with data. What we observe is low variation in polymorphism levels between species in contrast to the high variation in census population sizes. Moreover, a theory of evolution based on genetic drift alone predicts dramatically more genetic variability than observed in nature. Other evolutionary forces must be incorporated into our model. In §3 we consider the effect of natural selection on genealogies. In particular we discuss *genetic hitchhiking* and present a model due to Gillespie in which the effect of genetic drift is supplemented by what he dubs genetic *draft*. Under Gillespie's model the ancestry of the population is described by a coalescent with multiple mergers. We recall some basic results about such coalescents in §4. In §5 we turn to *spatial* structure and outline the stepping stone model and some of its shortcomings. The most obvious difficulty is that it supposes the population to be *subdivided* and does not easily extend to the case where it is not. In §6 we present a new model, very recently introduced in joint work with Nick Barton, which not only applies to populations evolving in a spatial continuum but also allows for the inclusion of certain large-scale extinction-recolonisation events which have dominated the demographic history of many species. Under this model, the ancestry of the population is described by what one can call a *spatial* Λ-coalescent or more generally a spatial Ξ-coalescent. It is more general than other processes of that name and we believe biologically more relevant. Our new model incorporates demography through extinction-recolonisation events, but it treats them as instantaneous and it does not, at this stage, allow for fluctuations in population size or interactions between competing species. We discuss some demographic models in §7 before, finally, in §8 suggesting some directions for future research.

REMARK 1.1 (Some remarks on terminology). We shall use interchangeably the terms *gene* and *genetic locus*. For our purposes it is convenient to think of chromosomes as linear and genes or genetic loci being points on that linear set. The different forms that a gene can take, which we will use to label individuals in the population, are called *alleles*.

2. Genetic drift.

Typically the patterns in genetic data are used to make inferences about the genealogical relationships between individuals in a sample from a population and so it is these genealogical relationships, in other words the family trees that relate individuals in the sample, that we try to model.

The most basic model is the classical Kingman coalescent. This process was introduced by Kingman (1982) and it provides a simple and elegant description of the genealogical relationships between individuals in a *large*, randomly mating population of constant size in which all individuals are equally fit. We're going to suppose for simplicity that we are modelling a *haploid* population which just means that each individual has exactly one parent.

Remark 2.1 (Diploid populations). For diploid populations such as our own, in which chromosomes are carried in pairs, provided one is only interested in one gene one typically treats the chromosomes as if they formed a haploid population of size $2N$, ignoring the pairing into individuals. Because, as we explain in §3, in diploid populations chromosomes are not typically passed down as indivisible blocks from parent to offspring, this device does not work if one is following more than one genetic locus. In that setting the genetic types at different loci on a chromosome may be derived from two different chromosomes in the previous generation and to trace the full ancestry we must follow both parental lineages.

Before we can describe the genealogical trees, we need a model for the way in which the population evolves *forwards* in time. The workhorse of mathematical population genetics is the Wright-Fisher model. In this model the population evolves in discrete generations. The family sizes of individuals in the parental population are determined by multinomial sampling with equal weights on all individuals. An equivalent (and convenient) way to say that is the following:

Definition 2.2 (The neutral Wright-Fisher model). A population of N genes evolves in discrete generations. Generation $(n+1)$ is formed from generation n by choosing N genes at random with replacement. i.e. Each gene in generation $(n+1)$ chooses its parent uniformly at random from those present in generation t.

Suppose now that we take a sample of size k from such a population. It is an easy matter to determine the family trees which relate individuals in the sample. For example, suppose that $k = 2$. The probability that the two individuals share a common parent is $1/N$. Conditional on having distinct parents, the probability that they share a common grandparent is again $1/N$ and so on. Thus the time to the most recent common ancestor of the two individuals has a geometric distribution with success parameter $1/N$. This has mean N and so for large populations it is convenient to measure time in units of N generations. The time to the most recent common ancestor of our sample of size two is then approximately exponentially distributed with parameter 1.

More generally, for a sample of size k, the probability that three or more of them share a common parent is $\mathcal{O}(1/N^2)$ and the chance that two *different* pairs of individuals share common parents is also $\mathcal{O}(1/N^2)$. For large N then the time until we see *any* event as we trace back the ancestral lineages is approximately the minimum of $\binom{k}{2}$ exponentially distributed random variables each with parameter one, in other words an exponentially distributed time with parameter $\binom{k}{2}$. At that time it is equally likely to be any of the pairs of ancestral lineages which *coalesce* into a single lineage. We then wait an exponentially distributed time with parameter $\binom{k-1}{2}$ until the next event—which is a coalescence of two of the surviving lineages, again chosen at random—and so on. For large populations this then provides an approximate description of the genealogy of the sample if we measure time in units of size N.

Notice that because we are modelling a haploid population (so individuals have only one parent) the family tree gets *smaller* as we go backwards in time, unlike the growing trees that describe your family history in which you track two parents of every ancestor (cf. Remark 2.1).

If we label individuals in the population $\{1, \ldots, k\}$ then the process we have just described induces a continuous time Markov chain on the space of equivalence relations on $[k] = \{1, 2, \ldots, k\}$. Individuals with common ancestor at time t before the present form a single equivalence class.

Definition 2.3 (Kingman's coalescent). A $[k]$-*coalescent* is a continuous time Markov chain on $\mathcal{E}_k$, the space of equivalence relations on $[k]$, with transition rates $q_{\xi,\eta}$ $(\xi, \eta \in \mathcal{E}_k)$ given by

$$q_{\xi,\eta} = \begin{cases} 1 & \text{if } \eta \text{ is obtained by coalescing two of the equivalence classes of } \xi, \\ 0 & \text{otherwise.} \end{cases}$$

The *Kingman coalescent* on $\mathbb{N}$ is a process of equivalence relations on $\mathbb{N}$ with the property that, for each k, its restriction to $[k]$ is a k-coalescent.

We shall use the term coalescent for both the process of coalescing ancestral lineages and the induced partition-valued process.

In passing to the large population limit we have taken a *diffusion approximation*. The rôle of the Wright-Fisher model is to identify the *correct* diffusion approximation. We would have obtained the same approximate genealogical trees for a wide variety of finite population models. In particular, provided that the variance of the number of offspring of each individual remains bounded as we pass to the large population limit, (up to a constant time change reflecting this variance) Kingman's coalescent provides an approximation to the genealogical trees for any exchangeable population model. It is this robustness of Kingman's coalescent which makes it such a powerful tool.

To see why one might call this a diffusion approximation, consider the effect of letting population size tend to infinity on the forwards in time population model. For simplicity suppose that the population is subdivided into two genetic types which we label a, A. We continue to suppose that these types are selectively neutral so that all individuals in our population are equally fit. When we pass to the infinite population limit we will not be able to talk about the *numbers* of individuals of each type, but we will be able to talk about the *proportion* of each type at time t (measured in units of size N). Let us write p_t for the proportion of the population at time t which is of type a. Then in a single generation, corresponding to a time interval of length $\delta t = 1/N$, since the number of a individuals in the new generation has the binomial distribution with N trials and success probability p, the change Δp in that proportion satisfies

$$\mathbb{E}[\Delta p] = 0, \quad \mathbb{E}[(\Delta p)^2] = \frac{1}{N^2} Np(1-p) = \delta t p(1-p), \tag{1}$$

and

$$\mathbb{E}[(\Delta p)^k] = \mathcal{O}(\delta t)^2 \text{ for } k \geq 3. \tag{2}$$

As $N \to \infty$ we see that, over an infinitesimal timestep of length δt, Δp is approximately normally distributed with

$$\Delta p \sim N(0, p(1-p)\delta t).$$

In other words, in the limit $\{p_t\}_{t\geq 0}$ follows the diffusion with generator

$$\mathcal{L}f(p) = \frac{1}{2}p(1-p)\frac{\partial^2 f}{\partial p^2},$$

or equivalently

$$dp_t = \sqrt{p_t(1-p_t)}dW_t \tag{3}$$

where $\{W_t\}_{t\geq 0}$ is a standard Brownian motion. Remember that in the passage to the limit we measured time in units of size N and so the appropriate approximation for a population of size N in *real* time units, denoted by τ, is

$$dp_\tau = \sqrt{\frac{1}{N}p_\tau(1-p_\tau)}dW_\tau$$

and backwards in time the coalescence of ancestral lineages is at rate $\frac{1}{N}\binom{k}{2}$. Crucially, *the application of the model requires knowledge of N.*

The stochastic force which drives the Wright-Fisher diffusion is called *genetic drift* and its strength is entirely determined by population size. There is considerable debate about the relative importance of genetic drift—which arises just from random *resampling* (reproduction) in a population—and other evolutionary forces, not least because a theory based on drift alone predicts a strong footprint of population size in data which is simply not present.

A basic observation is that genetic diversity is orders of magnitude lower than expected from census numbers and 'standard' genetic drift, Lewontin (1974), Gillepsie (2001). Something else is going on.

3. Genetic draft. So let's consider some of the other forces which might shape our genealogical trees. The first is *genetic structure*. In particular, we examine the influence of selection at one genetic locus on the genealogy of a sample from a neutral locus on the same chromosome. We shall suppose that the gene at the selected locus occurs in two forms which we label b, B. We start by finding the forwards in time analogue of equation (3)

for the frequency of b-alleles. As usual we start from a Wright-Fisher model.

Definition 3.1 (Wright-Fisher model with genic selection). A population of N genes occurring in two alleles (types) b and B evolves in discrete generations. Generation $(n+1)$ is formed from generation n by drawing N genes at random with replacement. If the *relative fitness* of the two alleles is $1+s:1$ and the frequency of b-alleles in the parental generation is p, then at each draw the probability that a b-allele is selected is

$$\frac{(1+s)p}{(1+s)p+(1-p)}.$$

Now we pass to what is known as the *weak selection* limit, obtained by assuming that s is $\mathcal{O}(\frac{1}{N})$. So writing $s=\frac{\sigma}{N}$ and using x_t to denote the proportion of b-alleles in the population, we perform the analogue of the calculation which led us to (1) to obtain this time

$$\begin{aligned}\mathbb{E}[\Delta x] &= \frac{(1+s)x}{(1+s)x+(1-x)} - x = \frac{sx(1-x)}{1+sx}\\ &= \frac{\sigma}{N}x(1-x) + \mathcal{O}(\frac{1}{N^2})\\ &= \delta t\sigma x(1-x) + \mathcal{O}(\delta t^2)\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[(\Delta x)^2] &= \frac{1}{N^2}N\frac{(1+s)x}{(1+s)x+(1-x)}\left(1-\frac{(1+s)x}{(1+s)x+(1-x)}\right)\\ &= \delta t x(1-x) + \mathcal{O}(\delta t^2).\end{aligned}$$

Once again we have

$$\mathbb{E}[(\Delta x)^k] = \mathcal{O}(\delta t)^2 \quad \text{for } k \geq 3.$$

As $N \to \infty$ we see that over an infinitesimal timestep of length δt,

$$\Delta x \sim N(\sigma x(1-x)\delta t, x(1-x)\delta t).$$

In other words (3) is replaced by

$$dx_t = \sigma x_t(1-x_t)dt + \sqrt{x_t(1-x_t)}dW_t.$$

In *real* time units we have

$$dx_\tau = sx_\tau(1-x_\tau)d\tau + \sqrt{\frac{1}{N}x_\tau(1-x_\tau)}dW_\tau. \tag{4}$$

We now use this equation to investigate the fate of a new selectively advantageous mutation arising in an otherwise neutral population. It will start at frequency $\frac{1}{N}$ in the population. Applying the theory of speed and scale for one-dimensional diffusions (see e.g. Karlin & Taylor 1981) to the diffusion (4), one finds that the probability that the new advantageous mutation will become *fixed* in the population, that is the probability that $\{x_\tau\}_{\tau\geq 0}$ hits one before it hits zero, given that it starts at $\frac{1}{N}$ is

$$P_{\text{fix}} = \frac{1-\exp(-2\sigma/N)}{1-\exp(-2\sigma)} \approx \frac{2s}{1-\exp(-2Ns)}. \tag{5}$$

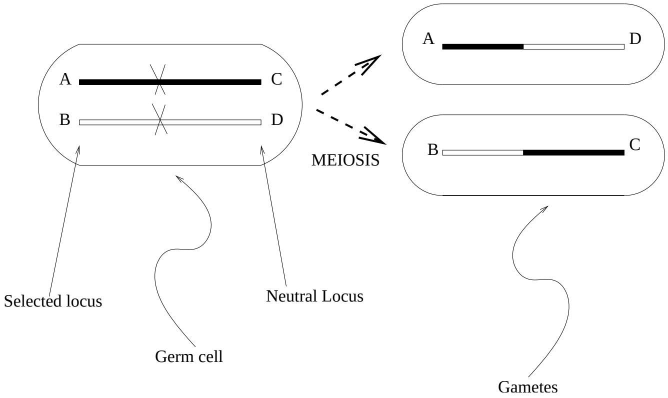
Given that it does so, a Green function calculation which is given in detail in Etheridge et al. (2006), tells us that the expected value of the time T after its first appearance when it achieves fixation, in *real* time units, is

$$\mathbb{E}[T] = 2\frac{\log(Ns)}{s} + \mathcal{O}\left(\frac{1}{s}\right). \tag{6}$$

Moreover, from their Lemma 3.1, the variance $var[T]$ is $\mathcal{O}(1/s^2)$.

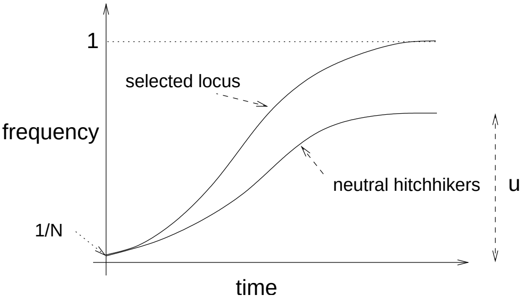
Definition 3.2 (Selective sweep). The process whereby the advantageous mutation increases to fixation is called a *selective sweep*.

If chromosomes are passed down as indivisible blocks then during the course of the sweep the genetic type along the whole of the chromosome on which the mutation arises will also become fixed. But in a diploid population such as our own in which chromosomes are carried in *pairs* they are not passed down as indivisible blocks. So (ignoring Y-chromosomes) although you will have inherited one chromosome in the pair from your mother and one from your father, it is not the case that the chromosome passed down by a parent to their offspring is an exact copy of one of the parental chromosomes but instead it is a mosaic of the two chromosomes in the parental pair. This is due to a process called *recombination*. Here, roughly, is how it works. During the reproductive process, each parent produces a large (effectively infinite) number of *germ cells*. These are cells all of the same genotype as the parent. These split into *gametes*. Gametes are cells that contain just one chromosome from each pair. It is the gametes that will fuse at random (one from the father, one from the mother) to produce the offspring. But during the process of splitting into gametes, which is called *meiosis*, the two chromosomes can exchange genetic material. This is illustrated in the cartoon below.



Such an event is called a *recombination* or *crossover* event. In general there could be more than one crossover event between two loci, indeed the number is often modelled as a Poisson process, but for our purposes we just need to know that there is some positive probability that the type at the neutral locus is inherited from a different parental chromosome from that at the selected locus.

As a result of the recombination events, during a selective sweep the frequency at a linked neutral locus may not be swept to fixation. Nonetheless, if the locus is sufficiently close to the selected locus that the probability of recombination is not too high, the frequency of the neutral allele which was fortunate enough to be associated with the selected allele when it arose will receive a boost from the selective sweep. This boost to the neutral locus is known as *genetic hitchhiking*, a term introduced by Maynard Smith & Haigh (1974). Schematically, the picture is something like the following:



We have used u to denote the proportion of neutral alleles at the end of the sweep descended from the individual associated with the favourable mutation when it first arose; that is u is the proportion of 'hitchhikers' in the population. Of course u is a random variable with a distribution depending on the strength of selection and the rate of recombination. Notice in particular that since we know from (6) that the sweep lasts $\mathcal{O}(\log(Ns)/s)$ generations, in order for the hitchhiking effect to be significant we require that the probability r of a recombination event is at most $\mathcal{O}(\frac{s}{\log(Ns)})$ in each generation. (This corresponds to the bound $r/s < 0.1$ given in Fay & Wu 2000.)

Gillespie (2000, 2001) investigated a model in which these strongly selected mutations which give rise to hitchhiking events occur at the points of a Poisson process. He assumes that selection is strong enough that the duration of the sweeps causing the hitchhiking events that affect a given locus is small compared to the time between them so that we can ignore the possibility that a locus will be subject to two simultaneous hitchhiking events. Then, in much the same way as we calculated the first two moments of the change of allele frequency over a single generation of the Wright-Fisher model, Gillespie calculated the first two moments in the change in allele frequency at the neutral locus over the course of a hitchhiking event. Again suppose that there are just two alleles at the neutral locus, a and A say. Suppose that the frequency of a-alleles before the sweep is p and consider the change Δp in frequency at the end of the sweep. Two things can happen:

1. If the a-allele hitchhikes, which happens with probability p, at the end of the sweep the frequency will be $u + p(1 - u)$.
2. If the a-allele does not hitchhike, which happens with probability $1-p$, at the end of the sweep the frequency will be $(1 - u)p$.

The expected change in frequency over the course of the sweep (the expectation is with respect to the u-random variable) is then

$$\mathbb{E}[\Delta p] = \mathbb{E}[p(u + p(1 - u)) + (1 - p)p(1 - u)] - p = 0,$$

while

$$\mathbb{E}[(\Delta p)^2] = \mathbb{E}[p(u - pu)^2 + (1 - p)(-pu)^2] = p(1 - p)\mathbb{E}[u^2].$$

Notice that this looks a lot like genetic drift (equation 1) except that the rôle of $\frac{1}{N}$ is played by $\mathbb{E}[u^2]$—multiplied by the rate of sweeps to which the locus is subjected. This rate of sweeps is called the *substitution rate*.

Now Gillespie argues that $\mathbb{E}[u^2]$ times the substitution rate is much less sensitive to population size than $\frac{1}{N}$ and so this effect would allow us to carry over much of modern population genetics to a setting in which predictions are much less sensitive to population size than a theory based on drift alone.

Remark 3.3. At first sight the claim that substitution rate is rather insensitive to population size is counter-intuitive. If the chance of an advantageous mutation arising on an individual is μ in each generation, then the rate at which advantageous mutations arise across the whole population is $2N\mu$ and in view of equation (5), one is tempted to approximate the rate of substitution as $4N\mu s$ which obviously depends very strongly on population size. However, our calculation of the fixation probability for a new selectively advantageous allele supposed that the mutation arose in an otherwise neutral population, but as the mutation rate increases we can expect to see *overlapping* sweeps. These will interfere with one another and in the absence of recombination we expect (see Cuthbertson et al. 2007) that if we assume what is known as multiplicative fitness, so the relative fitness of an individual carrying k mutations is $(1+s)^k$, then the rate of substitution grows at a rate proportional to $\log N$. In the hitchhiking world there *is* recombination, but we are only interested in a small portion of genome around the neutral locus in which recombination rates are low ($\mathcal{O}(\frac{1}{\log N})$) and so although the rate of adaptation will be increased by recombination, one does not expect it to grow significantly faster than $\mathcal{O}(\log N)$. Moreover, although we have modelled it as continuous, the genome is in reality discrete and so there is a bound on the number of distinct beneficial mutations that can arise in this small portion of genome around the neutral locus.

But Gillespie's *pseudo-hitchhiking model* as this is called differs in another important way from genetic drift. In the Wright-Fisher diffusion $\mathbb{E}[(\Delta p)^3] = \mathcal{O}(\delta t)^2$. Here that is *not* the case. In fact *all* moments of Δp will be of order δt. What this means for the genealogical trees relating individuals in a sample from the population is that whereas for the Kingman coalescent that corresponds to the Wright-Fisher diffusion we only ever saw two ancestral lines merging at a time, now there can be *multiple mergers* of ancestral lines.

In fact as we trace the ancestral lineages for a sample from the neutral locus back through the sweep, all those corresponding to genetic hitchhikers will merge into a single ancestor (corresponding to the founder of the sweep). The probability that any other lineages are involved in coalescence events during the course of the sweep is $\mathcal{O}(\frac{1}{\log N})$ and so to a first approximation we can approximate the genealogy at a neutral locus under the hitchhiking model by a coalescent in which at a coalescence event we see just a single merger. Conditional on u (which is independent at each successive event), for each ancestral lineage (independently) we flip a coin which comes up heads with probability u. All the ancestral lineages with a head merge into a single lineage. In particular *multiple* (by which we mean more than two) lineages can coalesce in a single event. This is a special case of something called a Λ-coalescent. Because in reality the sweep is not instantaneous, with probability of $\mathcal{O}(1/\log N)$ we see coalescence of lineages in the portion $1-u$ of the population whose type at the neutral locus does not derive from that of the originator of the sweep and so a better approximation for the genealogy is given by a coalescent which allows for *simultaneous (multiple) mergers*. Such coalescents are called Ξ-coalescents. A detailed discussion of the genealogy of the pseudo-hitchhiking model can be found in Durrett & Schweinsberg (2005). See also Durrett & Schweinsberg (2004), Schweinsberg & Durrett (2005) and Etheridge et al. (2006) for more on the effect of selective sweeps on genealogies.

4. Coalescents with multiple mergers. The processes known as Λ-coalescents were introduced independently by Pitman (1999) and Sagitov (1999). Like Kingman's coalescent they take their values among partitions of $\mathbb{N}$ and their laws can be prescribed by specifying the restriction to partitions of $[k] = \{1, 2, \ldots, k\}$ for each $k \in \mathbb{N}$.

Definition 4.1 (Λ-coalescents). A Λ-*coalescent* $\{\pi(t)\}_{t\geq 0}$ is a Markov process taking its values among partitions of $\mathbb{N}$ with the property that for each k, the restriction to $[k]$, $\{\pi_k(t)\}_{t\geq 0}$, is also a Markov process and if there are currently n blocks in $\pi_k(t)$ then each transition involving j of the blocks merging into one happens at rate $\beta_{n,j}$ (which is *independent* of k) and these are the only possible transitions.

Generally $\pi(0)$ is taken to be the partition all of whose blocks are singletons. For our purposes the Λ-coalescent describes the ancestry of a population whose individuals are labelled by $\mathbb{N}$. Each block in the partition at time t corresponds to a single ancestor at time t before the present with the elements of the block being the descendants of that ancestor. Whereas for the Kingman coalescent the only transitions are mergers of pairs of blocks, for the Λ-coalescent there can be mergers of three or more lineages. The key point is that the conditions in Definition 4.1 ensure sampling consistency: the coalescent obtained by taking a sample of size k_1 and then sampling $k_2 < k_1$ of the individuals in the sample and looking at the restriction of the process to that subsample has the same distribution as the coalescent obtained by simply starting from a sample of size k_2.

If such a process is to exist, the parameters $\{\beta_{n,j}, 2 \leq j \leq n\}$ cannot be chosen arbitrarily.

Theorem 4.2 (Pitman (1999)). *The Λ-coalescent of Definition 4.1 exists if and only if there is a finite measure Λ on $[0,1]$ for which*

$$\beta_{n,j} = \int_{[0,1]} u^{j-2}(1-u)^{n-j}\Lambda(du). \tag{7}$$

Remark 4.3 (Recovering Kingman's coalescent). Notice that if $\Lambda = \delta_0$, the point mass on zero, then $\beta_{n,j}$ vanishes unless $j = 2$ (corresponding to two blocks, or ancestral lines, merging) in which case it is one so that we recover Kingman's coalescent.

More generally one can consider coalescents with *simultaneous* multiple collisions. Such coalescents were obtained as the limit of the ancestral processes of properly scaled population models (in much the same way as the Kingman coalescent arises as the rescaled ancestral process corresponding to the Wright-Fisher model) by Möhle & Sagitov (2001) who were able to classify all coalescents arising in this way as a limit of populations with exchangeable reproduction mechanisms. Schweinsberg (2000) independently obtained the same class of coalescents (without a passage to the limit) and characterised the possible rates of mergers of ancestral lineages in terms of a single measure Ξ on the infinite simplex

$$\Delta = \{(x_1, x_2, \ldots) : x_1 \geq x_2 \geq \ldots \geq 0, \sum_{i=1}^{\infty} x_i \leq 1\}.$$

The resulting coalescents are now known as Ξ-coalescents.

The Λ-coalescent specifies the genealogy of a sample from a forwards in time population model introduced by Bertoin & Le Gall (2003) that we shall call the Λ-Fleming-Viot process. This is closely related to work of Donnelly & Kurtz (1999). See also Birkner et al. (2005) and Berestycki et al. (2007) for explicit simultaneous constructions of the Λ-Fleming-Viot process and its genealogical trees.

The Λ-Fleming-Viot process takes its values among probability measures on $[0,1]$. We will describe it in terms of its generator, $\mathcal{R}$, acting on functions of the form

$$G(\rho) = \int f(x_1, \ldots, x_n)\rho(dx_n)\ldots\rho(dx_1), \tag{8}$$

where $n \in \mathbb{N}$ and $f : [0,1]^p \to \mathbb{R}$ is measurable and bounded. First we need some notation. If $x = (x_1, \ldots, x_n) \in [0,1]^n$ and $J \subseteq \{1, \ldots, n\}$ we write

$$x_i^J = x_{\min J} \text{ if } i \in J, \text{ and } x_i^J = x_i \text{ if } i \notin J, i = 1, \ldots, n. \tag{9}$$

Definition 4.4 (Λ-Fleming-Viot process). Let Λ be a finite measure on $[0,1]$. The *Λ-Fleming-Viot process* has generator

$$\mathcal{R}G(\rho) = \sum_{J\subseteq\{1,\ldots,n\},|J|\geq 2} \beta^{\Lambda}_{n,|J|} \int (f(x_1^J, \ldots, x_n^J) - f(x_1, \ldots, x_n))\rho(dx_n)\ldots\rho(dx_1), \tag{10}$$

where $\beta^{\Lambda}_{n,j}$ is defined in equation (7). When $\Lambda(\{0\}) = 0$ (so there is no Kingman component), this can also be written

$$\mathcal{R}G(\rho) = \int_{(0,1]}\int_{[0,1]} (G((1-u)\rho + u\delta_k) - G(\rho))\rho(dk)u^{-2}\Lambda(du). \tag{11}$$

The Λ-Fleming-Viot process is most easily understood in the case $\Lambda(\{0\}) = 0$. In that case one can think of it as follows. The population at time t is described by a probability measure $\rho(t)$ on the type space $[0,1]$. Take a Poisson point process on $\mathbb{R}_+ \times (0,1]$ with intensity measure $dt \otimes u^{-2}\Lambda(du)$ which picks times and sizes of jumps for our population process. At a jump time t with corresponding jump size u, a portion u of the population is killed and replaced by offspring of an individual chosen at random from $\rho(t-)$. Thus

$$\rho(t) = (1-u)\rho(t-) + u\delta_x$$

where x is chosen according to $\rho(t-)$. The duality with the Λ-coalescent is evident. To construct the genealogy of a sample from such a population, suppose that there are currently k ancestral lineages. We trace backwards in time until we first encounter a point of our Poisson process. Suppose that the jump size that this specifies is u. Then for each lineage, independently, we flip a coin which shows heads with probability u. All lineages with a head merge into a common ancestor and the process continues.

In Gillespie's pseudohitchhiking model, $dt \otimes \frac{1}{u^2}\Lambda(du)$ is determined by the rate of substitutions and the distribution of the hitchhiking effect encoded by u.

Evidently one can write down analogous population models whose genealogy is determined by the Ξ-coalescents, but we omit that here.

A population evolving according to the Λ-Fleming-Viot process forms a single mating unit, but real populations are spatially structured. For example they may be subdivided into discrete locations or distributed over a two-dimensional continuum. It is to spatial structure that we now turn our attention.

5. Spatial structure.

The standard approach to space is to assume that the population is subdivided into *demes* (which one can think of as 'islands'), each of which is of large constant size. Interaction between demes is through *migration* (or more accurately exchange) of individuals.

To get a feel for the effect that this will have on the genealogical trees for the population we first consider a very simple example. Consider a population that is divided into just two demes with migration between the two. This simple model arises as a model for a single population divided into two genetic types which are in approximate equilibrium in the population, but in which there is mutation between types. The Wright-Fisher model is adapted to this setting as follows:

Definition 5.1 (Wright-Fisher model with migration). A population of size N is structured into two types, P and Q with $N_1 = N\omega_1$ of type P and $N_2 = N\omega_2$ of type Q. Each sub-population reproduces (independently) according to the neutral Wright-Fisher model except that now, after each reproduction step, a proportion of the population in each background is exchanged. In other words $\bar{\mu}_1 N_1$ individuals migrate from background one (P) to background two (Q) and $\bar{\mu}_2 N_2$ go the other way. In order to maintain constant population size in each background, we take $\bar{\mu}_1 N_1 = \bar{\mu}_2 N_2$.

Now let's establish the genealogy of a sample from such a population. First consider what happens to a single ancestral lineage as we trace backwards in time. Because the individuals in a given background are indistinguishable from one another, the probability that an individual of type P had a parent of type Q is just the proportion of individuals in the P-background after the migration step that had parents in the Q-background, namely $\frac{\bar{\mu}_2 N_2}{N_1} = \frac{\bar{\mu}_2 \omega_2}{\omega_1}$. Similarly, the probability that an individual in the Q background had a parent in the P background is $\frac{\bar{\mu}_1 \omega_1}{\omega_2}$. To obtain a diffusion limit we suppose that $\overline{\mu}_i = \frac{\mu_i}{N}$ where $N = N_1 + N_2$ is the total population size and we measure time in units of size N. Since the chance of a migration event and a coalescence event both affecting our ancestral lineages in a single generation is $\mathcal{O}(\frac{1}{N^2})$, in the diffusion timescale we only see coalescences between lineages in the same background. Our time unit is the *total population* size, as opposed to the population size in one of the backgrounds, so each pair of lineages currently in background i, coalesces at instantaneous rate $\frac{1}{\omega_i}$. We are implicitly assuming that $N\omega_i$ is *large* so that we never see multiple mergers. The genealogical trees for this model can then be described by a *structured (Kingman) coalescent*. As we trace backwards in time

1. ancestral lineages *migrate* from background one to background two at rate $\frac{\mu_2 \omega_2}{\omega_1}$ and from background two to background one at rate $\frac{\mu_1 \omega_1}{\omega_2}$,
2. any pair of lineages currently in background i *coalesces* at instantaneous rate $\frac{1}{\omega_i}$.

(For this and other simple extensions of Kingman's coalescent see Hudson 1990.) The key thing to note is that the rate of migration of ancestral lineages is weighted by the ratio of the population size in the two demes, so that backwards in time the migration mechanism is biased towards the more populous deme, and the rate of coalescence within a deme depends on population size there.

In more general spatial contexts, the standard approach is to generalise the simple two-deme model above to a collection of demes, indexed by some set I. The proportion of a-alleles in deme i at time t is denoted by $p_i(t)$ and is governed by Kimura's *stepping* *stone model* (Kimura 1953):

$$dp_i = \sum_j m_{ij}(p_j - p_i)dt + \sqrt{\gamma p_i(1-p_i)}dW_i. \qquad (12)$$

Here $m_{ij} = m_{ji}$ reflects migration between demes and the coefficient γ reflects the population size in each deme. The $\{W_i\}_{i\in I}$ are independent standard Brownian motions. In other words we have a system of interacting Wright-Fisher diffusions.

Remark 5.2. We can more generally take γ to depend on i, reflecting different population sizes in different demes, but then since we are assuming that the population size in different demes is maintained we must also modify the migration mechanism to reflect the condition $\overline{\mu}_1 N_1 = \overline{\mu}_2 N_2$ in Definition 5.1.

The genealogical trees relating a sample from a population evolving according to Kimura's model are again determined by a structured coalescent which can most easily be described in terms of coalescing random walks. Ancestral lineages follow independent random walks (with rates determined by the coefficients $\{m_{ji}\}$) between the demes as we trace backwards in time but while in the same deme each pair of lineages coalesces at rate one. This result is essentially due to Shiga (1982), but the duality between coalescing random walks and the stepping stone model that he establishes is not as strong as we are claiming here. It is the more recent ideas of Donnelly & Kurtz (1999) that allow one to simultaneously construct the stepping stone model and its genealogical trees.

The stepping stone model is extremely popular, not least because it is characterised in terms of easily interpreted parameters and it is relatively easy to simulate the genealogies. However, for many biological populations it is unnatural to think of them as subdivided and for many years people have been looking for an analogue of the stepping stone model in a two-dimensional continuum.

In one spatial dimension one can apply the *diffusive rescaling* (so that the random walk governing migration of individuals converges to Brownian motion) to the stepping stone model to obtain a stochastic *partial* differential equation

$$dp = \frac{1}{2}\Delta p dt + \sqrt{\gamma p(1-p)}dW, \qquad (13)$$

where W is now a space-time white noise (cf. Nagylaki 1978). But in two dimensions this equation has no solution; the white noise is 'too rough'. (See Walsh 1986 for an introduction to stochastic partial differential equations.) But anyway, in two dimensions, equation (13) is not what comes out of a diffusive rescaling. To see what *does* come out of the diffusive rescaling it is easiest to think about the structured coalescent. The coalescing random walks should converge to coalescing Brownian motions (and indeed in one dimension they do) but in two dimensions, two independent Brownian motions will never meet and so they never coalesce. We are just left with the Laplacian term (corresponding to migration in the population), the genetic drift disappears.

One approach to overcoming this difficulty is is to assume that the genealogical trees can be constructed from Brownian motions which coalesce at an instantaneous rate given by a function of their separation. The position of the common ancestor is generally taken to be the midpoint between the two lineages immediately before the coalescence event (although other distributions are of course possible). Sadly this rather natural model has a number of deficiencies. For example, suppose that under this model one takes the genealogical tree relating a sample of size k and examines the subtree obtained by looking at a subsample of size two. The distribution of that subtree will *not* be the same as the distribution of the genealogical tree corresponding to a sample of size two. The reason is that whenever one of the two ancestral lineages is involved in a coalescence event in the full tree it will jump. Furthermore there is no corresponding *forwards* in time model for the evolution of the population.

One can of course argue that even a population evolving in a continuum can be approximated by a subdivided one provided the spatial subdivision is sufficiently fine, but this raises another important issue. A key assumption in the derivation of the stepping stone model from individual based models is that population size in each deme is large. However, in real biological populations the number of individuals living in a local neighbourhood will often be rather small. The structured coalescent, just as Kingman's coalescent, allows for only pairwise mergers of ancestral lineages, but if several individuals are sampled from one location and neighbourhood size is small then multiple coalescences (by which as usual we mean merging of at least three lineages) will become significant.

6. A new model. Very recently, Barton & Etheridge (2007) have introduced a new model for evolution in a spatial continuum. A significant advantage of the new model is that it allows us to explicitly incorporate certain large scale fluctuations in the population, but it also reflects finite local population density by allowing for multiple coalescences. Recall the basic observation that Gillespie was trying to address with the pseudohitchhiking model—genetic diversity is orders of magnitude lower than expected from census population size and genetic drift. One alternative explanation to the pseudo-hitchhiking effect is that real populations also experience *large scale* fluctuations in which the movement and reproductive success of many individuals are correlated. For example climate change has caused extreme extinction and recolonisation events that dominate the demographic history of humans and other species. It is plausible that such fluctuations are the cause of most of the observed drift, but we need adequate models before we can assess their importance relative to other possible causes.

For simplicitly we describe only a special version of our model which can be thought of as a spatial Λ-Fleming-Viot process with genealogical trees determined by a corresponding spatial Λ-coalescent. In this setting, after an extinction event a region is recolonised by the descendants of a single individual. More generally it would be natural to take a Poisson number of colonists and then the corresponding model would be a spatial Ξ-coalescent.

First we describe a prelimiting model (which would be an interesting object of study in its own right). Suppose a population is initially distributed as a Poisson point process with intensity λ say. Let $\mu(dr)$ be a measure (not necessarily finite, but not quite arbitrary as we shall see below) on $(0, \infty)$ and for each $r \in (0, \infty)$ let $\nu_r(du)$ be a probability measure on $[0, 1]$. The dynamics of our prelimiting model are described as follows:

1. Let Π be a Poisson Point Process on $\mathbb{R}_+ \times \mathbb{R}^2 \times \mathbb{R}_+$ with rate $dt \otimes dx \otimes \mu(dr)$.
2. If (t, x, r) is a point of Π, then at time t throw down a ball $B_r(x)$ of radius r and centre x in $\mathbb{R}^2$.
3. If the ball is empty do nothing. If not:
   (a) Choose an individual at random from those in $B_r(x)$ and select $u \in [0, 1]$ at random according to ν_r.
   (b) For each individual in $B_r(x)$, independently flip a coin which shows heads with probability u and kill all those individuals with a head.
   (c) Throw down individuals with the same type as the selected individual (who may now be dead) according to an independent Poisson Point Process with intensity $u\lambda dx$ for $x \in B_r(x)$.

We shall refer to such events as reproduction events, but in some applications we have in mind they mimic rapid large-scale extinction-recolonisation events.

Of course any reproductive event has positive probability of leaving no individuals in a given region of space, but because neighbourhoods overlap, an empty neighbourhood can subsequently become recolonised and one can check (by a comparison with oriented percolation) that at least for sufficiently large λ the population will survive for all time. The difficulty is that it is not easy to write down explicitly the genealogical trees relating individuals in a sample from this population. An ancestor is necessarily in a non-empty patch of space and knowing that it is non-empty gives information about the rate at which it is hit by reproduction events as one traces back in time, but it is hard to find explicit expressions for this effect. We overcome this difficulty by considering a model in which the population density is infinite, but we retain some of the signature of a finite population by retaining the reproduction mechanism so that a non-trivial proportion of individuals in a neighbourhood are descended from a common ancestor. In particular, this will result in multiple coalescences of ancestral lineages.

Let us now describe this limiting model a little more precisely. We suppose that each individual in our population has a *type* taken from a set K (for example $K = [0, 1]$) and a *location* in a space E. For illustration, here we take $E = \mathbb{R}^2$. The local population density is taken to be constant in space and time and we shall write $\rho(t, x, \cdot)$, or sometimes for brevity ρ_x, for the probability measure on K which describes the frequencies of different types among those individuals residing at the point x. The reproduction mechanism mirrors that for our discrete time model:

Definition 6.1 (Spatial Λ-Fleming-Viot process). The *spatial Λ-Fleming-Viot process* $\{\rho(t, x, \cdot), x \in \mathbb{R}^2, t \geq 0\}$ specifies a probability measure on the type space K for every $t \geq 0$ and every $x \in \mathbb{R}^2$. With the notation above, the dynamics of the process are as follows. At every point (t, x, r) of the Poisson point process Π we choose $u \in [0, 1]$ independently according to the measure $\nu_r(du)$. We also select a point z at random from $B_r(x)$ and a type k at random according to $\rho(t-, z, \cdot)$. For all $y \in B_r(x)$,

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

Of course we must impose restrictions on the intensity measure if our process is to exist. To see what these should be, consider first the evolution of the probability measure $\rho(t,x,\cdot)$ defining the distribution of types at the point x. This measure experiences a jump of size $y \in A \subseteq [0,1]$ at rate

$$\int_{(0,\infty)} \int_A \pi r^2 \nu_r(du)\mu(dr).$$

By analogy with the Λ-Fleming-Viot process, we should like

$$\tilde{\Lambda}(du) = \int_{(0,\infty)} u^2 r^2 \nu_r(du)\mu(dr) \tag{14}$$

to define a finite measure on $[0,1]$. In fact, (weak) convergence of our individual based model to this limiting model as $\lambda \to \infty$ requires a bit more:

$$\tilde{\Lambda}(du) = \int_{(0,\infty)} u r^2 \nu_r(du)\mu(dr) \tag{15}$$

must define a finite measure on $[0,1]$.

Of course it is not enough to consider a single point. It has to be possible to 'fit together' the type distributions at different sites in a consistent way and the simplest way to guarantee that we can do this is to ensure the existence of a nice dual process describing, for each $n \in \mathbb{N}$, the distribution of lineages ancestral to a sample of size n from the population. Suppose then that a population evolves according to this model and consider the (backwards in time) dynamics of a *single* ancestral lineage. It evolves in a series of jumps with intensity

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

on $\mathbb{R}_+ \times \mathbb{R}^2$ where $L_r(x)$ is the area of $B_r(0) \cap B_r(x)$. If we want this to give a well-defined Lévy process, then we require

$$\int_{\mathbb{R}^2} (1 \wedge |x|^2) \left(\int_{(|x|/2,\infty)} \int_{[0,1]} \frac{L_r(x)}{\pi r^2}\, u\, \nu_r(du)\mu(dr) \right) dx < \infty. \tag{16}$$

Consider now lineages currently at separation $y \in \mathbb{R}^2$. They will coalesce if they are *both* involved in a replacement event which happens at instantaneous rate

$$\int_{(|y|/2,\infty)} L_r(y) \left(\int_{[0,1]} u^2 \nu_r(du) \right) \mu(dr). \tag{17}$$

Of course if two ancestral lineages do coalesce, then their common parent is located at a point selected at random from the ball involved in the reproduction event. Conceptually, this is readily extended to multiple lineages (where we will see multiple mergers). Notice that conditional on not having coalesced, the locations of ancestral lineages are *not* independent of one another. This is entirely analogous to the dependence between ancestral lineages in the coalescent for a continuous (finite) linear population suggested by Wilkins & Wakeley (2002) (see Wilkins 2004 for a two-dimensional analogue).

Only very preliminary calculations have been carried out for this model, but some elementary facts are known. For example, since population density is effectively constant, the path followed by a single lineage is the same forwards and backwards in time. An elementary calculation also allows us to find the distribution of the time, $\tau(x)$, back to the most recent common ancestor of two individuals at separation x in the present day population. Fix $\theta > 0$ and write $\phi(x) = \mathbb{E}[\exp(-\theta\tau(x))]$, then

$$\begin{aligned} 0 = (1-\phi(x)) &\int_{[0,\infty)}\int_{[0,1]} u^2 L_r(x)\nu_r(du)\mu(dr) - \theta\phi(x) \\ &+ \int_{\mathbb{R}^2}\int_{[0,\infty)}\int_{[0,1]} 2\left(u\frac{L_r(y)}{\pi r^2} - u^2\frac{L_r(x,x+y)}{\pi r^2}\right)(\phi(x+y)-\phi(x))\nu_r(du)\mu(dr)dy, \end{aligned}$$

where, as before, $L_r(x)$ is the area of the intersection of the ball of radius r centred on x and the ball of radius r centred on the origin and now $L_r(x, x+y)$ is the volume of the intersection of the three balls of radius r centred on the origin, x and $x+y$ respectively. Although unwieldy to deal with analytically, this quantity can be estimated numerically.

Remark 6.2 (Generalisations). For concreteness we have presented only a simple form of the model. There are many generalisations.

1. Replace $\mathbb{R}^2$ by an arbitrary Polish space.
2. Choose a Poisson number of parents at each reproduction event instead of just one.
3. Choose the distribution of the parents at each reproduction event non-uniformly. For example, recolonisers after a large-scale extinction event may be more likely to come from the boundaries of the region.
4. Impose spatial motion of individuals not linked directly to the reproduction events.
5. Instead of replacing a portion u of individuals from a ball centred on x, replace individuals sampled according to some distribution (e.g. Gaussian) centred on x.
6. ... and many more.

In almost all existing models for spatial evolution, reproduction events affect the allele frequencies in one of a number of discrete neighbourhoods (demes). In our model, neighbourhoods are allowed to overlap. In fact, by replacing $\mathbb{R}^2$ by a discrete model and choosing the intensity of jumps appropriately, we recover the classical stepping stone and island models as special cases. Similarly, we can obtain the spatial version of the Λ-Fleming-Viot process and the corresponding spatial Λ-coalescent studied by Limic & Sturm (2006).

7. Demography. A major shortcoming of all the classical models which we have discussed so far is that they ignore the *demography* of the population. Recall that the key parameter in our theory of genetic drift was population size. This we took to be large and fixed. In our continuum stepping stone model too we have effectively taken local population density to be constant. Moreover, we have assumed that our population evolves in isolation, competing only with individuals of the same species. But real populations are not like this. The size of a population fluctuates both in time and space and different populations are in competition for the same resources.

It is widely believed that if one views populations over sufficiently large spatial and temporal scales, then there should be some averaging effect which would allow one to use classical population genetic models with constant population density but with *effective* parameters replacing the real population parameters. Various conditions are established under which this really holds in Barton et al. (2002), but the usefulness of that result is limited due to a lack of explicit models for which the assumptions can be validated and the effective parameters calculated.

Natural populations interact with one another and with their environment in complex ways and we cannot hope to capture all of them in a single tractable mathematical model. Instead we try to isolate the effects of different aspects of these interactions through the study of simple 'toy' models. In this section we present several such models, all of which are compromises between fully spatial models and interacting particle systems. Many more details and some more general models can be found in Etheridge (2004) and Blath et al. (2007a,b).

In modelling a biological population it is natural to start from a branching process. However, in such models a finite population will either die out or grow without bound. In high dimensions, one can nonetheless obtain models in which population density is finite by allowing individuals to diffuse across an infinite domain, but in one and two dimensions (which is for many biological populations the most relevant), if one assumes that individuals diffuse either according to Brownian motion or a symmetric random walk then the spatial branching process predicts that if not extinct, at large times the population will develop clumps of arbitrarily large density and extent.

In a natural population one expects such clumping to be inhibited by limits on local resources. This effect can be introduced into our models by supposing that individuals living in crowded regions have lower reproductive success than those living in sparsely populated regions. Inspired by Bolker & Pacala (1997), the following class of models (and a continuum analogue) were introduced in Etheridge (2004):

$$dX_i(t) = \sum_j m_{ij}(X_j(t) - X_i(t))dt + \alpha(M - \sum_j \lambda_{ij} X_j(t))X_i(t)dt + \sqrt{\gamma X_i(t)}dB_i(t).$$

Here $X_i(t)$ denotes the size of the population in deme i at time t and $\{\{B_i(t)\}_{t\geq 0}, i \in \mathbb{Z}^d\}$ is a family of independent Brownian motions. The parameters $\{m_{ij}\}$ reflect migration of individuals just as in the stepping stone model. The noise is Feller's noise, corresponding to finite variance continuous state branching. The parameters $\{\lambda_{ij}\}$, which are positive, measure 'neighbourhood size'. It is shown in Etheridge (2004) that provided $m_{ij} > c\lambda_{ij}$ then for sufficiently large M the process will survive. This condition on the relative strengths of migration and competition was arrived at independently by Law et al. (2002). For the case when the competition is only within site, that is $\lambda_{ij} = 0$ for $i \neq j$, Hutzenthaler & Wakolbinger (2007) prove an ergodic theorem and also (under the same condition) show that there is a bound on the value of M below which the population will always go extinct.

This model is readily extended to competing species. Following Bolker & Pacala (1999) and Murrell & Law (2003), Blath et al. (2007a,b) assume the following strategies for survival for individuals in the population:

1. to colonise relatively unpopulated areas quickly,
2. to quickly exploit resources in those areas,
3. to tolerate local competition.

Suppose that there are just two different populations. Each can adopt a different combination of strategies for survival. We write $\{X(t)\}_{t\geq 0}=\{X_i(t), i\in\mathbb{Z}^d\}_{t\geq 0}$ and $\{Y(t)\}_{t\geq 0}=\{Y_i(t), i\in\mathbb{Z}^d\}_{t\geq 0}$ for our two populations, then in the model of Blath et al. (2007a,b) the populations evolve according to the following system of stochastic differential equations:

$$dX_i(t)=\sum_{j\in\mathbb{Z}^d} m_{ij}(X_j(t)-X_i(t))dt+\alpha\Big(M-\sum_{j\in\mathbb{Z}^d}\lambda_{ij}X_j(t)-\sum_{j\in\mathbb{Z}^d}\gamma_{ij}Y_j(t)\Big)X_i(t)dt \\ +\sqrt{\sigma X_i(t)}dB_i(t), \qquad (18)$$

$$dY_i(t)=\sum_{j\in\mathbb{Z}^d} m'_{ij}(Y_j(t)-Y_i(t))dt+\alpha'\Big(M'-\sum_{j\in\mathbb{Z}^d}\lambda'_{ij}Y_j(t)-\sum_{j\in\mathbb{Z}^d}\gamma'_{ij}X_j(t)\Big)Y_i(t)dt \\ +\sqrt{\sigma Y_i(t)}dB'_i(t), \qquad (19)$$

where $\{\{B_i(t)\}_{t\geq 0},\{B'_i(t)\}_{t\geq 0}, i\in\mathbb{Z}^d\}$ is a family of independent Brownian motions. The (bounded non-negative) parameters m_{ij}, m'_{ij}, λ_{ij}, λ'_{ij}, γ_{ij}, and γ'_{ij} are all supposed to be functions of $\|i-j\|$ alone (where $\|\cdot\|$ denotes the lattice distance on $\mathbb{Z}^d$) and to vanish for $\|i-j\|>R$ for some $R<\infty$.

For the X-population, the first two strategies for survival listed above correspond to taking large m_{ij} and large αM, while the third corresponds to taking small λ_{ij} (conspecific competition) and γ_{ij} (interspecific competition). By varying M we can also model how efficiently the species uses the available resources: a species that can tolerate lower resource levels will have a higher value of M.

Although it is possible to show that there are parameter ranges for which there is long term coexistence of the two populations, the model is difficult to study and it is instructive to pass to a simpler one.

Suppose now that the neighbourhood over which each individual competes is just the site in which it lives so that the only interaction between different points in $\mathbb{Z}^d$ is through migration and assume that the competition parameters λ_{ii},γ_{ii} etc. are constant (and so we may omit the subscripts on them). In addition we suppose that the migration mechanism for the two populations is the same. Write

$$N_i(t)=X_i(t)+Y_i(t),\quad\text{and}\quad p_i(t)=\frac{X_i(t)}{N_i(t)}.$$

Then an application of Itô's formula (and some rearrangement) gives

$$dp_i(t)=\sum_j\frac{N_j(t)}{N_i(t)}m_{ij}(p_j(t)-p_i(t))dt+\alpha p_i(1-p_i)\big[M-\lambda N_i(t)p_i-\gamma N_i(t)(1-p_i)\big]dt \\ -\alpha' p(1-p)\big[M'-\lambda' N_i(t)(1-p_i)-\gamma' N_i(t)p_i\big]dt+\sqrt{\frac{1}{N}p_i(t)(1-p_i(t))}\,dW_i(t),$$

where $\{\{W_i(t)\}_{t\geq 0}, i\in\mathbb{Z}^d\}$ is a family of independent Brownian motions. And conditioning on $N_i(t)\equiv N$ we obtain

$$dp_i(t)=\sum_j m_{ij}(p_j(t)-p_i(t))dt+sp_i(t)(1-p_i(t))(1-\mu p_i(t))dt \\ +\sqrt{\frac{1}{N}p_i(t)(1-p_i(t))}\,dW_i(t), \qquad (20)$$

where

$$s = \alpha M - \alpha' M' + N(\alpha'\lambda' - \alpha\gamma)$$

and

$$\mu = \frac{(\alpha'\lambda' - \alpha\gamma)N + (\alpha\lambda - \alpha'\gamma')N}{\alpha M - \alpha' M' + (\alpha'\lambda' - \alpha\gamma)N}.$$

Equation (20) also has an interpretation from the genetics literature. There, $p_i(t)$ would be thought of as the number of individuals (chromosomes in this context) in the population of type a. These chromosomes are subject to selection, but not the genic selection that we described before which makes it advantageous to carry one of the two alleles, but *balancing selection*. There are some genes for which it is beneficial for an individual to carry one allele of each type (heterozygous advantage). To model this in our Wright-Fisher model with selection, we replace the constant parameter s in (4) by $s_1(1-p) - s_2 p$ to reflect the idea that with probability $1-p$ an a allele will be paired with an A individual in which case it confers the advantage s_1 and with probability p is paired with another a individual in which case it confers the *dis*advantage s_2. If $s_1, s_2 < 0$ then we obtain selection in favour of homozygosity.

If $\mu < 1$, then in each site i there is selection in favour of either the X-type or the Y-type according to whether $s > 0$ or $s < 0$. If $\mu > 1$, in each site i we have selection in favour of *heterozygosity* if $s > 0$ and selection in favour of *homozygosity* if $s < 0$. In the 'neutral' case, $s = 0$, we recover Kimura's stepping stone model.

In the special case when the two populations evolve symmetrically, equation (20) takes the simple form

$$dp_i(t) = \sum_j m_{ij}(p_j(t) - p_i(t))dt + sp_i(t)(1-p_i(t))(1-2p_i(t))dt + \sqrt{\frac{1}{N}p_i(t)(1-p_i(t))}\,dW_i(t). \qquad (21)$$

For general s there is no convenient coalescent dual, but by first transforming the equations one can find an alternative duality with a system of *branching annihilating random walks*.

Definition 7.1 (Branching annihilating random walk). The Markov process $\{n_i(t), i \in \mathbb{Z}^d\}_{t\geq 0}$ with values $n_i(t) \in \mathbb{Z}_+$ and dynamics described by

$$\begin{array}{lll} \left\{\begin{array}{l} n_i \mapsto n_i - 1, \\ n_j \mapsto n_j + 1 \end{array}\right. & \text{at rate } n_i m_{ij} & \text{(migration)} \\ n_i \mapsto n_i + m & \text{at rate } s n_i & \text{(branching)} \\ n_i \mapsto n_i - 2 & \text{at rate } \frac{1}{2}n_i(n_i - 1) & \text{(annihilation)} \end{array}$$

is called a *branching annihilating random walk* with *offspring number m* and *branching rate s*.

The transformation that we require is simple. Let $x_i(t) = 1 - 2p_i(t)$. Then

$$dx_i(t) = \sum_j m_{ij}(x_j(t) - x_i(t))dt + \frac{1}{2}s(x_i^3(t) - x_i(t))dt - \sqrt{(1 - x_i^2(t))}dW_i(t). \qquad (22)$$

LEMMA 7.2. *The system* (22) *is dual to branching annihilating random walk with branching rate* $s/2$ *and offspring number two, denoted* $\{n_i(t), i \in \mathbb{Z}^d\}_{t\geq 0}$*, through the duality relationship*

$$\mathbb{E}[\underline{x}(t)^{\underline{n}(0)}] = \mathbb{E}[\underline{x}(0)^{\underline{n}(t)}],$$

where

$$\underline{x}^{\underline{n}} \equiv \prod_{i\in\mathbb{Z}^d} x_i^{n_i}.$$

Cardy and Täuber (1996), (1998) consider the branching annihilating random walk model of Definition 7.1. Their results are not rigorous, but if true, the implications for the system (21) are as follows. For $\mu = 2$ and $d = 1$ there is a critical value $s_0 > 0$ such that the populations will both persist for all time with positive probability if and only if $s > s_0$. In $d = 2$, there is positive probability that both populations will persist for all time if and only if $s > 0$. For $d \geq 3$ this probability is positive if and only if $s \geq 0$. Roughly speaking, for $d \geq 2$, if there is a homozygous advantage, then the population will initially form homogenic clusters, but ultimately it will be the interactions at the cluster boundaries that dominate and one type will go extinct. In the heterozygous advantage case, there will be long term coexistence of species. In one dimension, the heterozygous advantage must be 'sufficiently strong' if we are to see coexistence.

One expects the same result to be true when $\mu \neq 2$. Notice then that we have selection in favour of heterozygosity precisely when

$$(\alpha\lambda_{ii} - \alpha'\gamma'_{ii})N > \alpha M - \alpha' M', \quad \text{and} \quad (\alpha'\lambda'_{ii} - \alpha\gamma_{ii})N > \alpha' M' - \alpha M.$$

Comparing the quantities $\alpha'\lambda'_{ii} - \alpha\gamma_{ii}$ and $\alpha\lambda_{ii} - \alpha'\gamma'_{ii}$ tells us about the relative effectiveness of the X and Y populations as competitors. If the first is smaller, then the X-population is a less effective competitor. However, provided that $\alpha M > \alpha' M'$, we can even allow it to be negative and we expect to have positive probability of survival for the X-population. This reflects a competition-colonisation tradeoff.

8. Where next? The primary motivation underlying all the work described above is that a theory of evolution based on genetic drift alone predicts dramatically more genetic variability than we observe in data. If we are to correctly interpret that data then we must find ways to distinguish the different causes of reduced genetic diversity.

The new model described in §6 presents a promising platform in which to do this, at least for some evolutionary scenarios, but so far it has barely been explored. The first major issue to resolve is that we must establish the class of individual based models that can be approximated in this unifying framework. Second we must find statistics of the model that distinguish for example the effects of large scale extinction recolonisation events from those of natural selection. What is the relative importance of the (rare) large scale events and the (frequent) small scale events and can we find a footprint of both effects in data?

Demographic models too are a rich source of open problems. The key issue is to find a mathematically tractable model which nonetheless retains sufficient flexibility to incorporate the main strategies for survival that a species can adopt. Even our toy models have proved surprisingly resistant to mathematical analysis. For example, in spite of considerable efforts, although through studying the system (20) directly we have been able to show that for *sufficiently large s* there is positive probability that branching annihilating random walk with offspring number two will survive for all time, we have *not* been able to reproduce the full results of Cardy and Täuber.

Finally let us mention that very few genetic models incorporate demography in a sophisticated way. For example, one might use a model for population size expansion to model a time-varying coalescence rate in Kingman's coalescent, but rather few spatial models incorporate fluctuations in the local population size. However, as explained in detail in Etheridge (2006), even small fluctuations in local population size can have a large effect on the genealogical trees relating individuals in a sample from such a population. So although ecological and evolutionary models are traditionally treated separately, one of the most important challenges that faces us is to find ways to combine demography and genetics into a single tractable model.

Acknowledgements. The author would like to thank the anonymous referee for a detailed and helpful report which led to many clarifications.

References

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

[2] N. H. Barton and A. M. Etheridge, *A continuum stepping stone model*, manuscript, 2007.

[3] J. Berestycki, N. Berestycki, and V. Limic, *On small time asymptotics of Λ-coalescents*, preprint, 2007.

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

[5] M. Birkner, J. Blath, M. Capaldo, A. M. Etheridge, M. M.öhle, J. Schweinsberg, and A. Wakolbinger, *Alpha-stable branching and Beta-coalescents*, Elect. J. Probab. 10 (2005), 303–325.

[6] J. Blath, A. M Etheridge, and M. E Meredith, *Coexistence in locally regulated competing populations and survival of branching annihilating random walk*, Ann. Appl. Probab. 17 (2007), 1474–1507.

[7] J. Blath, A. M Meredith, and M. E Meredith, *Coexistence in locally regulated competing populations and survival of branching annihilating random walk (full version)*, Technical University of Berlin Preprint, 2007.

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

[9] B. M. Bolker and S. W. Pacala, *Spatial moment equations for plant competition: Understanding spatial strategies and the advantages of short dispersal*, American Naturalist 153 (1999), 575–602.

[10] J. L. Cardy and U. C. Täuber, *Theory of branching and annihilating random walks*, Phys. Rev. Lett. 77 (1996), 4780–4783.

[11] J. L. Cardy and U. C. Täuber, *Field theory of branching and annihilating random walks*, J. Stat. Phys. 90 (1998), 1–56.

[12] C. Cuthbertson, A. M. Etheridge, and F. Yu, *Asymptotic behaviour of the rate of adaptation*, preprint, 2007.

[13] P. J. Donnelly and T. G. Kurtz, *Particle representations for measure-valued population models*, Ann. Prob. 27 (1999), 166–205.

[14] R. Durrett and J. Schweinsberg, *Approximating Selective Sweeps*, Theoretical Population Biology 66 (2004), 129–138.

[15] R. Durrett and J. Schweinsberg, *A coalescent model for the effect of advantageous mutations on the genealogy of a population*, Stoch. Proc. Appl. 115 (2005), 1628–1657.

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

[17] A. M. Etheridge, *Evolution in fluctuating populations*, in: A. Bovier, F. Dunlop, A. van Enter, F. den Hollander, and J. Dalibard (eds.), Mathematical Statistical Physics, Lecture notes of the Les Houches Summer School 2005, Elsevier, 2006.

[18] A. M. Etheridge, P. Pfaffelhuber, and A. Wakolbinger, *An approximate sampling formula under genetic hitchhiking*, Ann. Appl. Probab. 16 (2007), 685–729.

[19] J. C. Fay and C.-I. Wu, *Hitchhiking under positive Darwinian selection*, Genetics 155 (2000), 1405–1413.

[20] J. Gillespie, *Genetic drift in an infinite population: the pseudohitchhiking model*, Genetics 155 (2000), 909–919.

[21] J. Gillespie, *Is the population size of a species relavant to its evolution?*, Evolution 55 (2001), 2161–2169.

[22] R. Hudson, *Gene genealogies and the coalescent process*, Oxford Surveys in Evolutionary Biology 7 (1990), 1–44.

[23] M. Hutzenthaler and A. Wakolbinger, *Ergodic behaviour of locally regulated populations*, Ann. Appl. Probab. 17 (2007), 474–501.

[24] S. Karlin and H. M. Taylor, *A Second Course on Stochastic Processes*, Academic Press, 1981.

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

[26] J. F. C. Kingman, *The coalescent*, Stoch. Proc. Appl. 13 (1982), 235–248.

[27] R. Law, D. J. Murrell, and U. Dieckmann, *On population growth in space and time: spatial logistic equations*, Ecology, 2002.

[28] R. C. Lewontin, *The Genetic Basis of Evolutionary Change*, Columbia University Press, New York, 1974.

[29] T. Shiga, *Continuous time multiallelic stepping stone models in population genetics*, J. Math. Kyoto Univ. 22 (1982), 1–40.

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

[31] M. Möhle and S. Sagitov, *A classification of coalescent processes for haploid exchangeable models*, Ann. Probab. 29 (2001), 1547–1562.

[32] D. J. Murrell and R. Law, *Heteromyopia and the spatial coexistence of similar competitors*, Ecology Letters 6 (2003), 48–59.

[33] T. Nagylaki, *A diffusion model for geographically structured populations*, J. Math. Biol. 6 (1978), 375–382.

[34] J. Pitman, *Coalescents with multiple collisions*, Ann. Probab. 27 (1999), 1870–1902.

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

[36] J. Schweinsberg, *Coalescents with simultaneous multiple collisions*, Electron. J. Probab. 5 (2000), 1–50.

[37] J. Schweinsberg and R. Durrett, *Random partitions approximating the coalescence of lineages during a selective sweep*, Ann. Appl. Probab. 15 (2005), 1591–1651.

[38] J. Maynard Smith and J. Haigh, *The hitch-hiking effect of a favourable allele*, Genet. Res. 23 (1974), 23–35.

[39] J. B. Walsh, *An introduction to stochastic partial differential equations*, in: École d'été de probabilités de Saint Flour, Lecture Notes in Mathematics 1180, 1986, 265–439.

[40] J. F. Wilkins, *A separation of timescales approach to the coalescent in a continuous population*, Genetics 168 (2004), 2227–2244.

[41] J. F. Wilkins and J. Wakeley, *The coalescent in a continuous, finite, linear population*, Genetics 161 (2002), 873–888.