Chapter 6
Spatial Structure

6.1 Subdivided Populations and the Structured Coalescent

Most models of spatially structured populations have the same basic format. The population is assumed to be *subdivided* into *demes*, which one can think of as 'islands' of population. The demes sit at the vertices of a graph and interaction between the subpopulations in different demes is through migration (or more accurately exchange) of individuals along the edges of the graph. The most elementary example is Wright's island model. This is how he introduced it in (Wright (1943)):

> The simplest model is that in which the total population is assumed to be divided into subgroups, each breeding at random within itself, except for a certain proportion of migrants drawn at random from the whole. Since this situation is likely to be approximated in a group of islands, we shall refer to it as the island model.

This corresponds to taking islands at the vertices of a complete graph. More generally one chooses the graph to caricature the spatial environment in which the population evolves. For example populations evolving in a two-dimensional spatial continuum are often approximated by taking the demes to sit at the vertices of $\mathbb{Z}^2$.

To get a feel for the effect that this will have on the genealogical trees for the population we first take a very simple example. Consider a population that is divided into just two demes with migration between the two. This simple model also 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 6.1 (Wright–Fisher model with migration). A population of size N is structured into two demes, 1 and 2 with population sizes $N_1 = N\omega_1$ in deme 1 and $N_2 = N\omega_2$ in deme 2. Each subpopulation reproduces (independently) according to the neutral Wright–Fisher model except that now, after each reproduction step, a proportion of the population in each deme is exchanged. In other words $\mu_1 N_1$ individuals migrate from deme one to deme two and $\mu_2 N_2$ go the other way. In order to maintain constant population size in each deme, we take $\mu_1 N_1 = \mu_2 N_2$.

We can establish the genealogy of a sample from such a population exactly as in Sect. 5.5. Here things are easier because the population size in each deme is

A. Etheridge, *Some Mathematical Models from Population Genetics*, Lecture Notes in Mathematics 2012, DOI 10.1007/978-3-642-16632-7_6,

constant. Because the individuals in a given deme are indistinguishable from one another, the probability that an individual in deme 1 had a parent in deme 2 is just the proportion of individuals in deme 1 after the migration step that had parents in deme 2, namely $\mu_2 N_2/N_1 = \mu_2\omega_2/\omega_1$. Similarly, the probability that an individual in deme 2 had a parent in deme 1 is $\mu_1\omega_1/\omega_2$. To obtain a diffusion limit we suppose that $\mu_i = \nu_i/N$ 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 $\mathcal{O}(1/N^2)$, in the diffusion timescale we only see coalescences between lineages in the same deme. Our time unit is the *total population* size, as opposed to the population size in one of the demes, so each pair of lineages currently in deme i, coalesces at instantaneous rate $1/\omega_i$. 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

- Ancestral lineages *migrate* from deme one to deme two at rate $\nu_2\omega_2/\omega_1$ and from deme two to deme one at rate $\nu_1\omega_1/\omega_2$.
- Any pair of lineages currently in deme i *coalesces* at instantaneous rate $1/\omega_i$.

Remark 6.2. Notice that the rate of migration of ancestral lineages is weighted by the ratio of the population size in the two demes, just as in Sect. 5.5, so that backwards in time the migration mechanism is biased towards the more populous deme, and, again as in Sect. 5.5, the rate of coalescence within a deme depends on population size there. The analogous result will hold for more general structured populations.

Here we have fixed the total population size in each deme so that different ancestral lineages evolve independently. If we allowed the population size in each deme to fluctuate randomly, then this would no longer be the case. Loosely, knowing that one lineage jumps to a deme tells us that the population size there is probably larger and so other lineages are more likely to jump there too.

Just as we passed to a diffusion approximation from the Wright–Fisher model for a panmictic population, we can also pass to a diffusion approximation for the structured Wright–Fisher model. We assume that the population size in each deme is large enough that the Wright–Fisher diffusion provides a good approximation for the effect of the random resampling due to reproduction. This leads to Kimura's stepping stone model (Kimura (1953)).

Definition 6.3 (Kimura's stepping stone model). We suppose that a population that is distributed across a collection of demes indexed by some set I is also subdivided into two allelic types labelled a and A. The proportion of a-alleles in deme i at time t is denoted by $p_i(t)$. Under Kimura's *stepping stone model*:

$$dp_i = \sum_j m_{ji}(p_j - p_i)dt + \sqrt{\frac{1}{N_e}p_i(1-p_i)}dW_i. \tag{6.1}$$

Here m_{ij} reflects migration between demes and satisfies

$$\sum_{j\neq i} m_{ij} = \sum_{j\neq i} m_{ji} \tag{6.2}$$

(in order to maintain constant population size in each deme). The parameter N_e is the (effective) population size in each deme and the $\{W_i\}_{i\in I}$ are independent standard Brownian motions.

In other words we have a system of interacting Wright–Fisher diffusions. To understand the first term on the right hand side of (6.1), note that type a individuals arrive in deme i at total rate $N_e \sum_j m_{ji} p_j$ and leave at total rate $N_e \sum_j m_{ij} p_i$ and observe that by (6.2)

$$\sum_j m_{ji} p_j - \sum_j m_{ij} p_i = \sum_j m_{ji} p_j - \sum_j m_{ji} p_i = \sum_j m_{ji}(p_j - p_i).$$

Remark 6.4. We can more generally take $N_e(i)$ for the effective population size in deme i, reflecting different population sizes in different demes, but then since we are assuming that the population size in each deme is maintained we must assume that

$$N_e(i) \sum_{j\neq i} m_{ij} = \sum_{j\neq i} N_e(j) m_{ji},$$

and the first term in (6.1) becomes

$$\sum_j \frac{N_e(j)}{N_e(i)} m_{ji}(p_j - p_i).$$

Lemma 6.5. *For a population evolving according to (6.1), the genealogical trees relating a finite sample consisting of n_i individuals from deme i for each $i \in I$ are traced out by the system of coalescing random walks whose evolution is described as follows:*

- *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 $n_i m_{ji}$.*

6.2 Duality

In this section we outline another connection between the stepping stone model and the *structured coalescent* of Lemma 6.5. This is through a powerful technique called the *method of duality*. To illustrate the strengths (and limitations) of the approach, we are going to extend the stepping stone model slightly to incorporate selection.

Definition 6.6 (Kimura's stepping stone model with selection). We suppose that a population that is distributed across a collection of demes indexed by some set I is also subdivided into two allelic types labelled a and A. The proportion of a-alleles in deme i at time t is denoted by $p_i(t)$. Under Kimura's *stepping stone model with selection*

$$dp_i = \sum_j m_{ji}(p_j - p_i)dt + \alpha p_i(1-p_i)dt + \sqrt{\frac{1}{N_e}p_i(1-p_i)}dW_i. \tag{6.3}$$

Here again m_{ij} reflects migration between demes and satisfies

$$\sum_{j\neq i} m_{ij} = \sum_{j\neq i} m_{ji}$$

(in order to maintain constant population size in each deme). The N_e is the (effective) population size in each deme and the $\{W_i\}_{i\in I}$ are independent standard Brownian motions.

The idea of duality is simple. We should like to express the distribution of the process $\underline{p} = (p_i)_{i\in I}$ that we are actually interested in, in our case allele frequencies in different demes, in terms of another (simpler) random variable, $\underline{n}$, that may take values in a completely different state space. The aim is to find a function f for which the following relationship holds:

$$\frac{d}{du}\mathbb{E}\left[f\left(\underline{p}(u),\underline{n}(t-u)\right)\right] = 0, \quad 0 \leq u \leq t, \tag{6.4}$$

so that

$$\mathbb{E}\left[f\left(\underline{p}(t),\underline{n}(0)\right)\right] = \mathbb{E}\left[f\left(\underline{p}(0),\underline{n}(t)\right)\right].$$

If, as the second argument of $f(\underline{p},\underline{n})$ varies, this provides a wide enough class of functions, then this is enough to characterise the distribution of $\underline{p}$. In particular, *existence* of a dual process is often used to prove *uniqueness* (in distribution) of the original process. A good reference is Ethier and Kurtz (1986), see also Etheridge (2000).

It is usually far from evident how to identify a suitable function f, but many models that arise in genetics have *moment duals*. These provide expressions for the moments and mixed moments of the process,

$$\mathbb{E}\Big[\prod_{i\in I} p_i^{n_i}\Big],$$

where $\underline{n} = (n_i)_{i\in I}$ is a vector with non-negative integer entries, a finite number of which are non-zero. In our dual process we are going to think of n_i as representing a number of 'particles' in deme i. The function f is defined by

$$f(\underline{p},\underline{n}) = \underline{p}^{\underline{n}} \equiv \prod_{i\in I} p_i^{n_i}$$

and our aim is to find dynamics for the process $\underline{n}(t)$ that guarantee that (6.4) is satisfied. The first step is to calculate $d\underline{p}^{\underline{n}}$ with $\underline{n}$ held fixed. Writing $\underline{e}_i$ for the vector consisting entirely of 0s except for a 1 in the ith position,

$$d\left(\underline{p}^{\underline{n}}\right) = \sum_i n_i \underline{p}^{\underline{n}-\underline{e}_i} \left[\sum_j m_{ji}\left(p_j - p_i\right) + \alpha p_i\left(1-p_i\right)\right] dt$$
$$+ \sum_i \frac{1}{2N_e} n_i\left(n_i-1\right) \underline{p}^{\underline{n}-2\underline{e}_i} p_i\left(1-p_i\right) dt + \sum_i (\ldots)\, dW_i.$$

Notice that, because we take the expectation in (6.4), we don't care about the exact form of the martingale term. Rearranging,

$$d\left(\underline{p}^{\underline{n}}\right) = \sum_i n_i \sum_j m_{ji}\left(\underline{p}^{\underline{n}+\underline{e}_j-\underline{e}_i} - \underline{p}^{\underline{n}}\right) dt + \sum_i n_i \alpha\left(\underline{p}^{\underline{n}} - \underline{p}^{\underline{n}+\underline{e}_i}\right) dt$$
$$+ \sum_i \frac{1}{2N_e} n_i\left(n_i-1\right)\left(\underline{p}^{\underline{n}-\underline{e}_i} - \underline{p}^{\underline{n}}\right) dt + \sum_i (\ldots)\, dW_i. \tag{6.5}$$

Our task is to identify dynamics for $\underline{n}(t)$ that ensure that (6.4) holds. To do this, we now think of evaluating $d\underline{p}^{\underline{n}}$ with $\underline{p}$ held fixed. Notice that since we evaluate $\underline{n}$ at time $t-u$ in (6.4) we pick up an extra minus sign. To cancel the first term in (6.5), particles should migrate according to the time reversal of the random walk that governed the forwards in time evolution of the individuals in our biological population. To cancel the second term we assume that $\alpha \leq 0$. Note that there is no loss of generality in doing so because if we consider $1-p$ in place of p, that is we look at the proportion of A alleles instead of a alleles, the only effect on (6.6) is to switch the sign of α. If $\alpha < 0$, then the second term will be cancelled by assuming that particles in the dual give birth (split in two) at rate $-\alpha$. Finally, to deal with the last term, we suppose that at instantaneous rate $1/N_e$ each pair of particles currently in deme i coalesces to form a single particle.

We have recovered a spatial version of the ancestral selection graph.

Lemma 6.7. *Suppose that $\underline{p}(t)$ evolves according to the Kimura stepping stone model with selection of Definition 6.6 with $\alpha < 0$ and that the process $\underline{n}$, taking values in $\mathbb{Z}_+^I$ (that is vectors indexed by I with non-negative integer components) and with $\underline{n}(0)$ having only finitely many non-zero components, evolves as follows:*

- $n_i \mapsto n_i + 1$ *at rate* $-\alpha n_i$
- $\begin{cases} n_i \mapsto n_i - 1 \\ n_j \mapsto n_j + 1 \end{cases}$ *at rate* $n_i m_{ji}$
- $n_i \mapsto n_i - 1$ *at rate* $\frac{1}{2N_e} n_i\left(n_i-1\right)$.

Then we have the duality relationship

$$\mathbb{E}\left[\underline{p}(t)^{\underline{n}(0)}\right] = \mathbb{E}\left[\underline{p}(0)^{\underline{n}(t)}\right].$$

It is easy to explain this result probabilistically. Calculating $\mathbb{E}[\underline{p}(t)^{\underline{n}(0)}]$ is equivalent to asking what is the probability that in a sample consisting of $n_i(0)$ individuals from deme i for each $i \in I$, all individuals are of type a. Just as in the ancestral selection graph of Definition 5.12, the process $\underline{n}(t)$ traces all 'potential' ancestors. The migration and coalescence is what we expect from tracing ancestral lineages of individuals in the sample. The branching of course reflects selection. It is most easily understood in terms of the Moran model with selection of Definition 5.9. The extra 'potential' selective events in the Moran model take place at rate $|\alpha|$. Here (in contrast to Definition 5.9) we are assuming that A has a selective advantage and so if we are to emerge with a type a individual from such a selective event, it must be that both individuals sampled at the event were type a. This happens with probability p^2, hence the branch in the structured coalescent dual – we must check the ancestry of *both* potential parents at such an event.

Remark 6.8. Although the process $\{\underline{n}(t)\}_{t\geq 0}$ has an interpretation in terms of the genealogy of a sample from the population, it is important to remember that the duality relation (6.4) is not enough to guarantee this, c.f. Remark 3.7.

Let's use this duality to try to make some qualitative statements about the long-time behaviour of a population evolving in a two-dimensional habitat. We take $I = \mathbb{Z}^2$ and suppose that migration corresponds to the discrete Laplacian (that is $m_{ij} = 1/4$ if i and j are neighbours and zero otherwise). We consider two separate cases.

First suppose $\alpha < 0$ and to avoid special cases suppose that $0 < p_i(0) < 1$ for all $i \in \mathbb{Z}^2$. Let's calculate

$$\mathbb{E}\left[\underline{p}(t)^{\underline{n}(0)}\right] \qquad \text{as } t \to \infty,$$

for a non-trivial $\underline{n}(0)$. In the dual process of branching and coalescing random walks, branches take place all the time, whereas particles only coalesce when they are in the same site, and the random walk is dispersing them across $\mathbb{Z}^2$, so we expect the number of particles to eventually grow without bound. Irrespective of $\underline{n}(0)$ then,

$$\mathbb{E}\left[\underline{p}(t)^{\underline{n}(0)}\right] = \mathbb{E}\left[\underline{p}(0)^{\underline{n}(t)}\right] \to 0 \qquad \text{as } t \to \infty.$$

Asymptotically, all individuals in our sample will be of type A. This of course makes sense biologically because the type A individuals have a selective advantage.

Now suppose that $\alpha = 0$ so that both alleles are selectively neutral. First we calculate $\mathbb{E}[p_i(t)p_j(t)]$ as $t \to \infty$. To do this, we start the dual process from one particle in site i and one in site j at time zero and see what happens as $t \to \infty$. Now there are no branches any more, just migration and coalescence. The distance between the two particles follows a two-dimensional random walk. Eventually they will come together. When that happens, there is some chance that they will coalesce before they move apart. If they don't coalesce, eventually they will come back together and once again they will have some chance of coalescence. And so on. In finite time they *will* coalesce. Then there will just be a single individual exploring $\mathbb{Z}^2$. The same argument applies for any $\underline{n}(0)$ (with finitely many non-zero components).

Eventually, there will just be a single individual exploring $\mathbb{Z}^2$. Thus

$$\mathbb{E}\left[\underline{p}(t)^{\underline{n}(0)}\right] \to \overline{p} \qquad \text{as } t \to \infty,$$

where $\overline{p}$ is a constant determined by the average initial proportion of a alleles in the population at time zero. How can this happen? Well, only if

$$\underline{p}(t) \xrightarrow{fdd} \begin{cases} \underline{1} \text{ with probability } \overline{p} \\ \underline{0} \text{ with probability } 1-\overline{p} \end{cases} \qquad \text{as } t \to \infty,$$

where $\underline{1}$ is the vector all of whose entries are 1 and $\underline{0}$ is the vector consisting entirely of 0s and the convergence is in the sense of finite dimensional distributions. So even though neither type has a selective advantage, for large times we expect our sample to consist entirely of a or entirely of A alleles. In the non-spatial setting, since the Wright–Fisher diffusion with no selection or mutation is in the natural scale, the probability that the a allele fixes is its initial frequency (see Lemma 3.14). In the spatial setting, which allele we see in our sample is determined by $\overline{p}$.

6.3 Collapse of Structure

Having established the genealogical trees relating individuals in a sample from a subdivided population one can look for the effect of structure on simple summary statistics of the coalescent trees. Perhaps the best known result is the following.

Lemma 6.9. *Suppose that a population evolves according to Wright's island model with D demes and population size N in each deme. Then the mean coalescence time of two ancestral lineages sampled from within the* same *island is equal to that of two lineages sampled from a panmictic population of size DN* independent *of the rate of migration between islands.*

Remark 6.10. In fact this result can be extended. For a surprisingly wide range of models of subdivided populations, the mean coalescence time of a sample of two lineages from within a single subpopulation will be equal to that of two individuals sampled from a panmictic population of the same total size, irrespective of the detailed pattern of migration. Conditions to guarantee this can be found in Wilkinson-Herbots (2003).

Proof of Lemma 6.9. Let us write T_{11} for the mean time to coalescence of two lineages sampled at random from within the *same* island and T_{12} for the mean time to coalescence of two lineages sampled from *different* islands. Suppose that the rate of migration of each lineage is m. We condition on the first event to hit the two sampled lineages. If they are in the same island then this can be a migration or a coalescence and happens at exponential rate $2m + 1/N$. If they are in different islands then the event is necessarily a migration. It occurs at rate $2m$ and it can leave

the lineages in different islands (with probability $(D-2)/(D-1)$ since only one of the $D-1$ possible targets contains the other lineage) or the same island (with probability $1/(D-1)$). This leads to the linear equations

$$T_{11} = \frac{1}{\frac{1}{N}+2m} + \frac{2m}{\frac{1}{N}+2m}T_{12},$$

$$T_{12} = \frac{1}{2m} + \frac{D-2}{D-1}T_{12} + \frac{1}{D-1}T_{11}.$$

Solving these we obtain

$$T_{11} = ND$$

as required. □

We can also solve for T_{12} to obtain

$$T_{12} = \frac{D-1}{2m} + ND.$$

This quantity, by contrast, does depend on the migration rate, but if $m \to \infty$ then $T_{12} \to ND$ and the mean time to coalescence behaves as for a panmictic population even if we sample from different demes. One can take this further. Bahlo and Griffiths (2001) find an explicit expression for the Laplace transform of the distribution of the time to the most recent common ancestor of a sample of size two and from this show that, as $m \to \infty$, the whole distribution of the time to the MRCA converges to that of a sample of size two from a panmictic population.

It is natural to ask whether this extends to the genealogical tree of a larger sample from the population. The answer, it turns out, is yes. This is part of a much wider phenomenon in which, because migration and coalescence are happening on different timescales, we see a 'collapse' of structure in the structured coalescent. Nordborg and Krone (2002) summarise the situation beautifully. Here we shall just skim the surface. We consider a population that is subdivided into different states. These could be demes as before or, more generally, age classes, genetic types and so on. If 'migration' (which could be through ageing or mutation for example) between some groups of states is happening on a much faster timescale than coalescence, then the structure associated with those groups collapses and each is replaced, through some sort of averaging procedure, by a single 'metastate'. We already saw an effect like this is Sect. 2.3. When the entire structure collapses, we recover the Kingman coalescent with an effective population size, but one can also recover a structured coalescent in the limit. (We shall see something analogous to this in Sect. 6.5.) To illustrate collapse of structure we consider a very simple example.

Example 6.11. Suppose that our population, which evolves in discrete generations, is divided into two demes with sizes $N_1 = a_1 N$ and $N_2 = a_2 N$. In each generation, ancestral lineages migrate between demes with strictly positive probabilities β_{12} and β_{21} and we write (γ_1, γ_2) for the stationary distribution of the corresponding random walk. In contrast to Definition 6.1, we do not suppose that β_{ij} is $\mathcal{O}(1/N)$.

Coalescence within demes is with probability $1/N_i = 1/(Na_i)$ in deme i in each generation. Measuring time in units of N generations, the genealogy of a sample from the population converges to a Kingman coalescent in which if there are currently k ancestral lineages, a pair chosen at random will coalesce at rate $c\binom{k}{2}$ where

$$c = \sum_{i=1}^{2} \frac{1}{a_i} \gamma_i^2.$$

We verify this result only when starting from two ancestral lineages and refer to Nordborg and Krone (2002) for a more general result. In this case we can record the possible states of the process of ancestral lineages as

$$\{(1,0),(0,1),(2,0),(1,1),(0,2)\}.$$

Ignoring terms of $\mathcal{O}(1/N^2)$, the backwards in time transition matrix of the process of ancestral lineages can be written as

$$\Pi_N = A + \frac{1}{N} B$$

where the matrix A corresponds to migration of ancestral lineages and the matrix B to coalescence within demes. The key result is the following Lemma which can be found in Möhle (1998).

Lemma 6.12. *Let $t, K \geq 0$ be fixed and let $(c_N)_{N\in\mathbb{N}}$ be a sequence of positive real numbers with $\lim_{N\to\infty} c_N = 0$. Further let $A = (a_{ij})$ be a matrix with $\|A\| \equiv \max_i \sum_j |a_{ij}| = 1$ such that $P = \lim_{n\to\infty} A^n$ exists. Then*

$$\lim_{N\to\infty} \sup_{\|B\|\leq K} \|(A + c_N B)^{[t/c_N]} - (P + c_N B)^{[t/c_N]}\| = 0.$$

If $(B_N)_{N\in\mathbb{N}}$ is a matrix sequence such that $G = \lim_{N\to\infty} PB_NP$ exists, then

$$\lim_{N\to\infty} (A + c_N B_N)^{[t/c_N]} = P - I + e^{tG} \qquad \forall t > 0.$$

This generalises the familiar identity $\lim_{N\to\infty}(I + A/N)^N = e^A$. An easy consequence of this is the following useful theorem.

Theorem 6.13 (Möhle 1998). *Let $X_N = \{X_N(r)\}_{r\in\mathbb{N}_0}$ be a sequence of time homogeneous Markov chains on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the same finite state space S and let Π_N denote the transition matrix of X_N. Assume that the following conditions are satisfied.*

1. *$A = \lim_{N\to\infty} \Pi_N$ exists and $\Pi_N \neq A$ for all sufficiently large N.*
2. *$P = \lim_{n\to\infty} A^n$ exists.*
3. *$G = \lim_{N\to\infty} PB_NP$ exists, where $B_N = (\Pi_N - A)/c_N$ and $c_N = \|\Pi_N - A\|$ for all $N \in \mathbb{N}$.*

If the sequence of initial probability measures $\mathbb{P}_{X_N(0)}$ converges weakly to some probability measure μ, then the finite dimensional distributions of the process $\{X_N([t/c_N])\}_{t\geq 0}$ converge to those of a continuous time Markov process $(X_t)_{t\geq 0}$ with initial distribution $X_0 \stackrel{d}{=} \mu$, transition matrix $\Pi(t) = P - I + e^{tG}$, $t > 0$, and infinitesimal generator G.

Remark 6.14. This is a special case of a general class of results in perturbation theory which are discussed, for example, in Ethier and Kurtz (1986), Chap. 1, Sect. 7.

Since P is a projection, that is $P^2 = P$, we have that

$$P - I + e^{tG} = Pe^{tG} = e^{tG}P.$$

(To see this expand e^{tG}, and hence the left hand side, as a series and note from the definition of G that $PG = G = GP$.) This tells us that the limiting process is obtained by first projecting, using P, onto the stationary distribution of the 'fast process' governed by A and then applying the generator G.

In our example,

$$P = \begin{pmatrix} \gamma_1 & \gamma_2 & 0 & 0 & 0 \\ \gamma_1 & \gamma_2 & 0 & 0 & 0 \\ 0 & 0 & \gamma_1^2 & 2\gamma_1\gamma_2 & \gamma_2^2 \\ 0 & 0 & \gamma_1^2 & 2\gamma_1\gamma_2 & \gamma_2^2 \\ 0 & 0 & \gamma_1^2 & 2\gamma_1\gamma_2 & \gamma_2^2 \end{pmatrix},$$

$c_N = \frac{1}{N}$ and

$$B = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \frac{1}{a_1} & 0 & -\frac{1}{a_1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{a_2} & 0 & 0 & -\frac{1}{a_2} \end{pmatrix}.$$

We can then calculate G as

$$PBP = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ c\gamma_1 & c\gamma_2 & -c\gamma_1^2 & -2c\gamma_1\gamma_2 & -c\gamma_2^2 \\ c\gamma_1 & c\gamma_2 & -c\gamma_1^2 & -2c\gamma_1\gamma_2 & -c\gamma_2^2 \\ c\gamma_1 & c\gamma_2 & -c\gamma_1^2 & -2c\gamma_1\gamma_2 & -c\gamma_2^2 \end{pmatrix},$$

with

$$c = \frac{\gamma_1^2}{a_1} + \frac{\gamma_2^2}{a_2}.$$

We can collapse states according to the number of lineages to see that we have recovered exactly the Kingman coalescent (up to the time change by c). The assignment

of demes to lineages is just by independent sampling according to the stationary distribution of the random walk. □

In this example, with a fixed and finite number of demes, the result is not really surprising. On the time scale of the coalescence, at any given instant the random walks have probability about γ_1^2 of both being in deme 1 in which case they have instantaneous coalescence rate $1/a_1$ and they have probability about γ_2^2 of both being in deme 2 in which case they coalesce at instantaneous rate $1/a_2$. When we look at larger numbers of lineages, convergence to the coalescent hinges on the exchangeability of lineages. Ancestral lineages have 'forgotten' all about their starting point by the time we see a coalescence event, and so it is equally likely to be any pair of lineages that coalesce. For a general spatial model, we cannot expect this exchangeability for the ancestral lineages of an arbitrary sample. Lineages sampled close together are more likely to coalesce first. However, if coalescence times are long enough, then the lineages have time to 'mix'. Zähle et al. (2005) consider a stepping stone model on a large two-dimensional torus in $\mathbb{Z}^2$. They show that if individuals are sampled uniformly from the torus, then as the side of the torus tends to infinity the genealogy does indeed converge to a Kingman coalescent (with an appropriate effective population size). We shall describe a close analogue of their result in Sect. 6.5.

Collapse of structure can also be seen in island models with large numbers of demes. This is demonstrated in a series of papers by Wakeley and coworkers (e.g., Wakeley (2001), Wakeley and Aliacar (2001)). In contrast to the setting of Nordborg and Krone (2002), the population size, N, in each deme is assumed to be fixed and finite, but the number, D, of demes grows without bound. While within the same deme each pair of lineages coalesces at rate $1/N$, but migration between demes at a rate of $\mathcal{O}(1)$ sends each lineage to a new deme chosen uniformly at random from the $D-1$ available. If the sample size is much smaller than the number of demes, then the chance of landing on a deme that is already occupied by another ancestral lineage is $\mathcal{O}(1/D)$. For large D the history of a sample can then be divided into two phases. During the first *scattering phase*, which is $\mathcal{O}(1)$ generations long, lineages within the same deme will experience a mixture of coalescence and migration to unoccupied demes, until there is at most one lineage in each deme. Never again during the history of the sample will we see more than two lineages in a single deme. During the second *collecting phase*, which is $\mathcal{O}(D)$ generations long, lineages migrate between demes with the possibility of coalescence only when they are in the same deme. Measuring time in units of D generations, we have a tractable ancestral process in which the scattering phase is instantaneous (corresponding to the projection P of Remark 6.14) and the collecting phase is a Kingman coalescent.

6.4 Evolution in a Spatial Continuum and the Pain in the Torus

So far we have concentrated on subdivided populations, but, in reality, many biological populations evolve in a spatial continuum. Wright (1943) and Malécot (1948) considered populations evolving in $\mathbb{R}^1$ and $\mathbb{R}^2$. They make similar assumptions.

Malécot, for example, assumes that (I) individuals are distributed randomly with constant expected density everywhere in space; (II) each individual, independently, produces a Poisson number of offspring with mean one; and (III) each offspring migrates independently, with the displacements being drawn from some distribution $m(x)$, for example a normal distribution. However, as Felsenstein (1975) observed, these assumptions are inconsistent. A population evolving according to (II) and (III) violates (I). In fact, if it is distributed over all of $\mathbb{R}^1$ or $\mathbb{R}^2$ it develops larger and larger clumps separated by greater and greater distances. This is not overcome by working on a torus as then the population dies out. Counteracting this, for example by conditioning the total population size to be a constant N does not overcome the problem of clumping. Felsenstein dubs these problems *'the pain in the torus'*.

Backwards in time, both Wright and Malécot assume that the probability that two individuals have a common parent in the previous generation is a function of their separation (determined by convolving two copies of the distribution $m(x)$) and that if they did not have a common parent their parents' positions are determined by independent copies of $m(x)$. Evidently this (backwards in time) description of the genealogy is not consistent with their forwards in time model for the evolution of the population.[1] So can we find consistent forwards and backwards in time models? In view of the success of the stepping stone model it is natural to use that as a starting point and to try to replace the system of interacting stochastic (ordinary) differential equations by a single stochastic *partial* differential equation. In one spatial dimension this can be achieved by applying the *diffusive rescaling* to the stepping stone model (so that the random walk governing migration of individuals converges to Brownian motion). This results in the limiting equation

$$dp = \frac{1}{2}\Delta p dt + \sqrt{\gamma p(1-p)}dW, \tag{6.6}$$

where W is now a space-time white noise. This was proved by Shiga (1988), who also established convergence of the system of coalescing random walks that describe the genealogy in the stepping stone model to a system of Brownian motions that coalesce at a rate determined by the local time that they spend together. This generalises work of Nagylaki (1978; 1978) who derived, under the same rescaling, an equation for the correlations between allele frequencies at different locations. In two dimensions Nagylaki showed that the rescaling fails. The equations for the correlations 'blow up' on scales comparable with the distance moved by a single gene over a single generation. Correspondingly, (6.6) has no solution; the white noise is 'too rough'. (See Walsh (1986) for an introduction to stochastic partial differential equations.) Moreover, if one applies the diffusive rescaling to the stepping stone model then one recovers a deterministic heat equation. It is easy to see why by thinking about the genealogical process of coalescing random walks.

[1] Wright and Malécot thought about probability of identity in allelic state under an infinitely many alleles mutation model rather than genealogies, but as we saw in Sect. 2.4 the two notions are closely related.

Under the diffusive rescaling the random walks converge to Brownian motions, but two independent Brownian motions evolving in $\mathbb{R}^2$ never meet and so we lose the coalescence. The coalescence is what reflects the noise term (which in turn models the randomness of reproduction) and so with no coalescence we cannot expect any noise.

Remark 6.15. If one modifies (6.6) by replacing the white noise W by a suitable 'coloured' noise, obtained for example by convolving W with a function from $L^2(\mathbb{R}^d)$, then the new equation *does* have a solution. Although at first sight this equation looks natural, it is not, as one might hope, what one obtains by taking a rescaling limit of an individual based model in which parents are chosen from a neighbourhood (rather than the same location). Indeed it is not clear how to obtain it as the limit of *any* individual based model.

Alternatively, instead of modifying the forwards in time stepping stone model, one can try to modify the corresponding (backwards in time) structured coalescent. An obvious approach 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). However, this process of coalescence violates the *consistency* of Remark 2.4. To see this, take the tree corresponding to a sample of size k and consider a subtree of size two. Whenever one of these two ancestral lineages is involved in a coalescence event in the full tree it will jump. We would not see this jump if we modelled the tree relating just two individuals directly. Furthermore, there is no corresponding *forwards* in time model for the evolution of the population.

Wright and Malécot assume an infinitely many alleles mutation model in which, in each generation, each offspring (independently) has a new allelic type with some fixed probability. They find an expression for the probability that two individuals, sampled at distance x apart, have the same allelic type. Although based on inconsistent assumptions, the formula provides an astonishingly good fit to the two-dimensional stepping stone model. This can be seen for example in Fig. 1 of Barton et al. (2002). That paper shows that under certain conditions the Wright–Malécot formula can be extended to continuum population models which incorporate local structure. Over all but very small scales, the resulting probability of identity can be written as a function of three parameters: the *effective dispersal rate*, the *neighbourhood size* and the *local scale*. The difficulty is that there is a shortage of explicit models for which the assumptions underlying this result can be verified and the effective parameters established. Moreover, the formula only applies to samples of size two.

Neighbourhood size is the product of the effective dispersal rate (that is the variance of the Gaussian distribution from which an individual's parent is drawn) and the local population density and gives some measure of how many individuals 'interact' in a given generation. Although the Wright–Malécot formula could in principle be extended to larger samples of well-separated genes, if neighbourhood size is

small, multiple coalescences of ancestral lineages could become significant. This observation turned out to be key in writing down a new model which addresses some of the problems identified above.

6.5 The Spatial Λ-Fleming–Viot Process

Recently, in joint work with Nick Barton, we proposed a new framework for modelling populations evolving in a spatial continuum and this will be our final topic. Not only does the proposed framework address some of the issues raised above, including allowing for small neighbourhood size, but it also allows us to explicitly incorporate large-scale extinction-colonisation events into the dynamics of the population. The motivation for this is the basic observation that we made at the beginning of Chap. 5:

> Genetic diversity is orders of magnitude lower than expected from census population size and genetic drift.

While selection certainly plays a rôle in reducing genetic diversity, it is plausible that most of the reduction that we observe relative to the 'null' model of neutral evolution and Kingman's coalescent is due to 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. The new framework provides mathematical models through which to assess the importance of such events relative to some of the other forces that shape genetic diversity.

For simplicity we describe only a particular instance of our approach 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. In many settings it would be natural to take a Poisson number of colonists, say, and then the corresponding coalescent model would be a spatial Ξ-coalescent.

The starting point is an individual based model. The resolution of Felsenstein's 'pain in the torus' is that reproduction events (including the large-scale extinction-recolonisation events) are based on a Poisson point process in space. The rate at which a given region of space is affected by such an event does not grow with local population density and this prevents clumping.

Definition 6.16 (Individual based model). We suppose that the population is initially distributed as a Poisson point process in $\mathbb{R}^d$ (with $d = 2$ being the most interesting case). Let λ be a fixed positive constant, $\mu(dr)$ be a measure on $(0,\infty)$ and, for each $r > 0$, let $\nu_r(du)$ be a probability measure on $[0,1]$ such that

$$\int_0^\infty \int_0^1 u r^d (1+r^d) \nu_r(du) \mu(dr) < \infty. \tag{6.7}$$

Write $\xi(dr,du) = \mu(dr)\nu_r(du)$. The dynamics of the population are as follows:

1. Let Π be a Poisson Point Process on $\mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}_+ \times (0,1]$ with rate $dt \otimes dx \otimes \xi(dr,du)$.
2. If (t,x,r,u) is a point of Π, then at time t throw down a ball $B_r(x)$ of radius r and centre x in $\mathbb{R}^d$.
3. If the ball is empty do nothing. If not:
   a. Choose an individual at random from those in $B_r(x)$;
   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 $u\lambda \mathbf{1}_{B_r(x)} dx$.

Regions of space can, and do, become empty in this model, but, because the neighbourhoods affected by different events overlap, an empty region can subsequently be recolonised. Berestycki et al. (2009) show that there is a critical value of λ above which the process survives and below which it dies out. They also check that under condition (6.7) the process described in Definition 6.16 actually exists.

The difficulty with this model is that it is not easy to write down explicitly the genealogical trees relating individuals in a sample from the population. An ancestor is necessarily in a non-empty patch of space and knowing that a region 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 letting $\lambda \to \infty$ so that there are no empty regions of space. At first sight it looks as though we are thereby losing the possibility of small neighbourhood size, but in fact this is not so: by retaining the same reproduction mechanism, in which each individual hit by a reproduction event has probability u of being killed, we retain the signature of finite neighbourhood size. In particular, we can still see multiple coalescences of ancestral lineages.

Remark 6.17. An alternative model of this type, considered in Barton et al. (2010), has a slightly modified reproduction mechanism. It is again based on a spatial Poisson process, but now if an event is centred on the point x, then an individual at y is killed with probability $u(x,y)$, where $u(x,y)$ is a Gaussian kernel centred on x say. A parent is selected by taking a weighted sample from the population immediately before the event, in which individuals are weighted according to their distance from x according to a (possibly different) Gaussian distribution. Offspring, of the same type as the parent, are distributed according to a Poisson point process with intensity $\lambda u(x,y)$. The resulting population model has a Poisson distribution with intensity λ as its stationary distribution.

Let us now describe the 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 continue to take $E = \mathbb{R}^d$. To each point $x \in E$ and at each time t, the limiting process assigns a probability

measure, $\rho(t,x)$ on K. The idea is that the type of an individual sampled from the point x at time t has distribution $\rho(t,x)$. The reproduction mechanism mirrors that for our discrete time model.

Definition 6.18 (Spatial Λ-Fleming–Viot process). The *spatial Λ-Fleming–Viot process*, denoted $\{\rho(t,x,\cdot), x \in \mathbb{R}^d, t \geq 0\}$, specifies a probability measure on the type space K for every $t \geq 0$ and every $x \in \mathbb{R}^d$. With the notation of Definition 6.16, the dynamics of the process are as follows. At every point (t,x,r,u) of the Poisson point process Π we 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 affecting a proportion $y \in A \subseteq [0,1]$ of individuals at x at rate

$$\int_{(0,\infty)} \int_A C_d r^d \nu_r(du)\mu(dr),$$

where C_d is the volume of the unit ball in $\mathbb{R}^d$. By analogy with the Λ-Fleming–Viot process, we should like

$$\tilde{\Lambda}(du) = \int_{(0,\infty)} u^2 r^d \nu_r(du)\mu(dr) \tag{6.8}$$

to define a finite measure on $[0,1]$. In fact, we require a bit more:

$$\Lambda(du) = \int_{(0,\infty)} u r^d \nu_r(du)\mu(dr) \tag{6.9}$$

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

Remark 6.19. Recall from Remark 5.24 that the existence of Λ-coalescents for which the analogue of (6.8) is satisfied, but not that of (6.9), relies on some cancellation of positive and negative jumps. Our need for the stronger condition in the spatial setting reflects the fact that the existence of overlapping neighbourhoods destroys that cancellation.

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

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

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

$$\int_{\mathbb{R}^d} (1 \wedge |x|^2) \left(\int_{(|x|/2,\infty)} \int_{[0,1]} \frac{L_r(x)}{C_d r^d} u \nu_r(du) \mu(dr) \right) dx < \infty. \tag{6.10}$$

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

$$\int_{(|y|/2,\infty)} L_r(y) \left(\int_{[0,1]} u^2 \nu_r(du) \right) \mu(dr). \tag{6.11}$$

Under condition (6.9), the expressions in (6.10) and (6.11) are both automatically finite. 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 and Wakeley (2002) (see Wilkins (2004) for a two-dimensional analogue).

Remark 6.20 (Spatial Λ-coalescent). Evidently the dual process of ancestral lineages is a spatial version of the Λ-coalescent. However, we emphasise that it differs from that studied by Limic and Sturm (2006).

Recall from Sect. 6.3 that the work of Zähle et al. (2005) shows that it makes sense to define a coalescent effective population size (see Remark 2.8) for a uniform sample from a population evolving according to the stepping stone model on a large torus in $\mathbb{Z}^2$. It is natural to ask whether an analogous result holds here and, if so, what the effect of large scale extinction-recolonisation events is on that effective population size. This question is addressed by Barton et al. (2010) and we finish with a description of their result.

Let $\mathbb{T}(L)$ denote the torus of side L in $\mathbb{R}^2$. Suppose that there are two types of event:

1. *Small* events affecting bounded regions;
2. *large* events affecting regions of diameter $\mathcal{O}(L^\alpha)$, for some $0 \leq \alpha \leq 1$.

The idea is that small events reflect 'ordinary' reproduction, whereas large events model large-scale extinction-recolonisation events. We assume that each ancestral lineage is hit by a small event at rate $\mathcal{O}(1)$, but by a large event at rate $\mathcal{O}(1/\rho(L))$ where $\rho(L) \to \infty$ as $L \to \infty$. We then sample uniformly at random from the whole of $\mathbb{T}(L)$ and ask what happens to the genealogy as $L \to \infty$?

A precise statement can be found in the paper, but here is an outline of the result.

Theorem 6.21 (Barton et al. 2010).

1. *Suppose that* $\alpha < 1$. *On a suitable timescale the genealogy converges to a Kingman coalescent (with an effective parameter). Depending on* $\rho(L)$, *the effective population size that determines the timescale can depend on both large and small scale events.*
2. *Suppose that* $\alpha = 1$. *There are three cases:*
   a. $\rho(L) \approx L^2$. *On timescale* $\rho(L)$, *the coalescent converges to a spatial* Λ*-coalescent in which lineages follow independent Brownian motions in between coalescence events.*
   b. $\rho(L) \approx L^2 \log L$. *On timescale* $\rho(L)$, *the coalescent converges to a (non-spatial)* Λ*-coalescent in which multiple mergers are due to large events and there can also be a Kingman component reflecting coalescence due to small events.*
   c. $\rho(L) \gg L^2 \log L$. *On a timescale* $L^2 \log L$, *the coalescent converges to the Kingman coalescent.*

If there are no large events, then in many ways the model looks like a two-dimensional stepping stone model and so, in view of the results of Zähle et al. (2005), it is no surprise to recover the Kingman coalescent. From a biological perspective, what is interesting is the large effect that even rare extinction-recolonisation events can have on the effective population size.

Since for $\alpha = 1$ large scale events cover a non-negligible fraction of the torus, a mathematically much richer picture emerges. If they happen too frequently, then they can affect multiple lineages while the location of those lineages is still correlated with their starting points. If $\rho(L) \approx L^2 \log L$, the positions of ancestral lineages have homogenised over the torus by the time a large event arrives, but lineages have not necessarily yet all coalesced due to small events. Finally, if large events are too rare, then lineages have all coalesced due to small events before we see a large event and so their effect is lost.

6.6 More General Models

One of the attractions of the approach to modelling outlined above is its flexibility. Although we have presented only the simplest form of the spatial Λ-Fleming–Viot process, it can readily be modified to incorporate more realistic biological assumptions. For example, it would be natural to allow for multiple founders after an extinction-recolonisation event and there is no reason to suppose that they are chosen uniformly from the region affected by the event. Equally, we can incorporate selection, recombination, spatial motion of individuals not linked to reproduction and so on.

From a mathematical perspective, even the simplest model reveals a rich structure. For example, by considering a population subdivided into two types, a and A say, with a sufficiently 'sparse', just as for the voter model, if events affect only

balls of bounded radius then other than in one spatial dimension one can recover a cluster of superBrownian motion as a rescaling limit for the distribution of a-alleles. By incorporating selection and rescaling suitably, one can obtain the (deterministic) Fisher-KPP equation as a limiting description of allele frequencies. In one dimension one can also recover the stochastic partial differential equation

$$dp = \frac{1}{2}\Delta p dt + sp(1-p)dt + \sqrt{\varepsilon p(1-p)}dW, \tag{6.12}$$

where W is space-time white noise. This equation is the focus of a great deal of current research, but in higher dimensions, which are more biologically relevant, it has no solution. By basing reproduction on regions instead of individuals, we have a natural alternative to (6.12), which makes sense in arbitrary spatial dimensions, from which (6.12) can be recovered as a limit in one spatial dimension, and which arises in a natural way as a limit of an individual based model.