The Annals of Applied Probability
1999, Vol. 9, No. 4, 1091–1148

GENEALOGICAL PROCESSES FOR FLEMING–VIOT MODELS WITH SELECTION AND RECOMBINATION

BY PETER DONNELLY[1] AND THOMAS G. KURTZ[2]

University of Oxford and University of Wisconsin—Madison

Infinite population genetic models with general type space incorporating mutation, selection and recombination are considered. The Fleming–Viot measure-valued diffusion is represented in terms of a countably infinite-dimensional process. The complete genealogy of the population at each time can be recovered from the model. Results are given concerning the existence of stationary distributions and ergodicity and absolute continuity of the stationary distribution for a model with selection with respect to the stationary distribution for the corresponding neutral model.

1. Introduction. The Fleming–Viot measure-valued diffusion arises as the large population limit of a wide class of population genetics models. Together with the Dawson–Watanabe process which arises from branching models, it is one of the more well studied measure-valued processes. For a recent review of available results about the Fleming–Viot process, see Ethier and Kurtz (1993) and references therein.

Measure-valued diffusions are often motivated by first considering a class of prelimiting finite-population models. The dynamics in such discrete contexts are easily specified in terms of the behavior of the individuals in the population, and the composition of the population is naturally represented as a measure on the set, E, of possible types. Measure-valued diffusions then arise because the associated discrete measure-valued processes behave sensibly (after appropriate rescaling) in the large population limit. On the other hand, the discrete population models which keep track of the fates of individuals make no sense for infinite-population sizes. Thus, while it might be convenient in applications to think of the measure-valued diffusion as describing the evolution of a hypothetically infinite population, it is difficult to make this precise.

Donnelly and Kurtz (1996, 1999) have recently given a discrete construction of a class of neutral measure-valued population processes. Loosely speaking, the idea is to “bring back the particles.” First, a (one-dimensional) process P describing the total mass of the measure-valued process is constructed. Conditional on P, an E^∞-valued process, X, is described with the property that for any t the collection $(X_1(t), X_2(t), \dots)$ is exchangeable. For fixed t, the

Received June 1998; revised February 1999.
[1]Supported in part by NSF Grant DMS-95-05129, Unverstiy of Chicago Block Fund and UK ESRC Grant B/AF 1255.
[2]Supported in part by NSF Grant DMS-96-26116.
AMS 1991 *subject classifications*. Primary 60J25, 92D10; secondary 60K35, 60J70.
Key words and phrases. Genetic models, recombination, selection, Fleming–Viot process, particle representation, measure-valued diffusion, exchangeability, genealogical processes, coalescent.

components of X, which we call particles, thus "carry" an empirical measure $Z(t)$, the de Finetti measure associated with the exchangeable sequence. As t varies, the process $\tilde{Z} \equiv PZ$ (exists a.s. and) is a version of the appropriate measure-valued process. The particle process X can be thought of as a description of the infinite population described by $\tilde{Z}$, with a careful choice of labels for the individuals.

Aside from inherent interest, such discrete representations considerably simplify the study of the measure-valued processes. One of the reasons is that the discrete construction also carries the "genealogy" of the measure-valued model in an explicit and simple form (much simpler, e.g., than the Dawson–Perkins historical process). Many questions of interest, including sample path properties, ergodicity and the structure of moment measures, are profitably studied in terms of genealogy.

This paper constructs and exploits a discrete representation for Fleming–Viot models with recombination and selection. In the next section, we describe versions of the classical Moran model from population genetics. This is perhaps the simplest of the class of finite-population models in the domain of attraction of the Fleming–Viot diffusion. There are two natural methods for introducing selection depending on whether the type of an individual affects the rate at which the individual dies (so-called viability selection) or the rate at which it reproduces (fertility selection). In each case the Fleming–Viot process arises when the recombination probability and all differences between the fitnesses of distinct individuals are of the order of the inverse of the population size. The key to the discrete representation of the measure-valued diffusion is a particular labelling of the individuals in the Moran model population. With this labelling, the dynamics of individuals still make sense as the population size goes to infinity.

The countable representation is described in Section 3. Informally, the neutral construction appropriate to Fleming–Viot in Donnelly and Kurtz (1999) is augmented with two types of additional events on each level. The first type, "selective events," results in potential discontinuities on a level due to additional births (fertility selection) or deaths (viability selection), which depend on the type on the level and the current value of the empirical measure. The second type of additional event, induced by recombinations, results in discontinuities in which the type on a level is replaced by a random choice according to a Markov kernel, which depends on the current type, and a random choice from the empirical measure.

There are various ways to characterize the E^{∞}-valued particle process X. One is as a solution of an infinite system of ordinary stochastic differential equations; another is via an associated martingale problem. These issues are discussed in Section 4, where we give theorems concerning the uniqueness of solutions to the system of equations (Section 4.1) and to the martingale problem (Section 4.2). Further, provided $X(0)$ is exchangeable, solutions to either the system of stochastic equations or to the martingale problem are exchangeable for any t. It follows from this exchangeability that for any k, the joint distribution of any k-tuple $X_{i_1}(t), \dots, X_{i_k}(t)$ is that of a "sample of size

k" from $Z(t)$. (More formally, $E[f(X_{i_1}(t), \dots, X_{i_k}(t)) \mid Z(t)] = \langle f, Z(t)^k \rangle$, so that moment measures of the measure-valued process are related to the joint distributions of collections of particles.) In fact, this exchangeability holds also for any stopping time with respect to the filtration generated by the empirical measure process. By exchangeability, there is a one-to-one correspondence between stationary distributions for X and those for the measure-valued process (the latter being the de Finetti representing measures for the former), and (see Section 5) the moment measures of the stationary distribution of the measure-valued process are determined by the stationary finite-dimensional distributions of the particle process X.

As they are specified in Section 4, the particle processes do not carry information about genealogy. The problem is that when a selective (or recombination) event involves a choice from the empirical distribution of the particles, the ancestry prior to this choice is not identifiable. Section 6 introduces a particular method for making selections from the empirical measure in such a way that ancestral information prior to such events is known. Informally, the idea is to choose uniformly from all the (neutral) genealogical paths leading back from the time at which the choice is made. Loosely, one can think of the choice at a selective (or recombination) event as choosing a particle "from infinity." The specific mechanism introduced in Section 6 has the property that while we cannot identify "which" particle is chosen at the time of the selective event, we do know who its ancestor was at any time strictly prior to that event. In particular, the method allows us to trace genealogical information back through selective and recombination events.

Sections 7 and 8 involve a detailed study of genealogical structure in the model. In particular, the *ancestral influence graph* is introduced. This has a very simple probabilistic structure and can be thought of as a supragenealogy, in the sense that the true genealogy is contained within it. The joint distribution of a collection of particles at stationarity can be constructed by choosing a type according to the stationary first moment measure of Z and tracing the effects of mutation forward through the appropriate influence graph. Further, if the ancestral influence graph is known from some time t back to s, and the types on the (a.s. finite number of) levels in the graph at s are specified, one can recover the actual genealogy of the population over the period from t back to s. In the absence of selection, the influence graph is just the ancestral recombination graph [e.g., Griffiths and Marjoram (1997)]. In the absence of recombination, it is a generalization to general diploid selection of Krone and Neuhauser's (1997) ancestral selection graph. They used graphical constructions from the particle system world to introduce their ancestral selection graph for haploid selection in a finite-population size Moran model. In addition to its more general setting, our approach actually couples the large population limit of this genealogical process with the Fleming–Viot process. (In fact, more is true in that we simultaneously couple genealogy for all times, rather than simply back from a particular time, as would be typical in the coalescent approach to population genetics models.)

The question of whether and to what extent selection changes the distribution of the ancestral tree in genetics models has been of longstanding interest [e.g., Tavaré (1984), Krone and Neuhauser (1997), Neuhauser and Krone (1997) and references therein]. Krone and Neuhauser's recent work and the results described here (see in particular Section 8.4) at least provide accessible tools for tackling the problem. Neuhauser and Krone (1997) established that, at stationarity, selection does not affect genealogy in the special cases of no mutation or in the limit as the mutation rate goes to infinity. (Note that in both these examples, the models in effect become neutral. If the mutation rate is zero, then at stationarity all members of the population will have the same type and no selective differences will be apparent. In the high mutation rate case, knowing the shape of the genealogical tree gives no information about the type of an individual sampled from along the tree, since the high mutation rate will "randomize" the type to that of an independent draw from the stationary distribution of the mutation process. It follows that information about types provides no information about genealogy, so the genealogy must be the same as in the neutral case.) In contrast, we show that, in a particular nonstationary setting, genealogy under selection is different from genealogy under neutrality.

In Section 8.5, we characterize the distribution of the type on a level immediately after a selective birth event. In the case of genic selection, we show that the expected fitness immediately after such an event is given by the mean fitness of the population plus the variance in fitness across the population, a result somewhat reminiscent of Fisher's "fundamental theorem of natural selection." This observation leads us to conjecture, that apart from the special limiting settings considered by Krone and Neuhauser, selection does change genealogy. To illustrate the intuition, consider genic selection with a finite number of types. The effect of selective events on a level is to (stochastically) increase the fitness of the type on that level. This increases the chances that different ancestors have the same (fitter) type, hence changing (in this case increasing) coalescence rates.

Section 9 concerns ergodicity properties of the processes. We give conditions on the mutation process which ensure uniqueness of stationary distributions, respectively, uniform ergodicity, for the particle process (and hence for the measure-valued process). We also prove that for the model with selection but not recombination, stationary distributions under selection will be mutually absolutely continuous with respect to those under neutrality provided only that the mutation process is strongly connected (by which we mean that for any two initial distributions there exists a time t at which the transient distributions are not mutually singular). This theorem considerably generalizes known results.

If $\{\mathscr{F}_t\}$ is a filtration and U is a process with independent increments, we will say that U is *compatible* with $\{\mathscr{F}_t\}$ if U is $\{\mathscr{F}_t\}$-adapted and for all $t, s \geq 0$, $U(t+s) - U(t)$ is independent of $\mathscr{F}_t$.

Let F be a metric space. A function x: $[0, \infty) \to F$ is *cadlag* if it is right continuous at each $t \geq 0$ and has left limits at each $t > 0$. The space of cadlag, F-valued functions on $[0, \infty)$ will be denoted by $D_F[0, \infty)$.

2. Moran models with selection.

We consider a variation of the classical Moran model in which mutation is assumed to occur continuously in time, rather than at the moment of reproduction. In particular, our model consists of a collection of n particles, each of which has a "type" represented as a point in a complete, separable metric space E, that is, $X^n(t) = (X_1^n(t), \dots, X_n^n(t)) \in E^n$. Between birth and death events, the types of the particles evolve independently according to a Markov process with generator B. We will refer to this process as the *mutation process*. We first consider models with viability selection.

2.1. *Models with viability selection.*

Let $\mathscr{P}(E)$ denote the space of probability measures on E and let $\mathscr{P}^n(E) \subset \mathscr{P}(E)$ be the subset of purely atomic probability measures whose atoms have sizes that are multiples of $1/n$. For $x = (x_1, \dots, x_n) \in E^n$, let $\mu_x = (1/n) \sum_{i=1}^n \delta_{x_i} \in \mathscr{P}^n(E)$ and let β be a nonnegative, bounded, measurable function defined on $E \times \mathscr{P}(E)$. The generator for the particle model is given by

$$
\begin{aligned}
A_0^n f(x) &= \sum_{i=1}^n B_i f(x) \\
&\quad + \frac{1}{2} \sum_{1 \leq i \neq j \leq n} \left(1 + \frac{2}{n} \beta(x_j, \mu_x)\right) \bigl(f(\eta_j(x \mid x_i)) - f(x)\bigr),
\end{aligned} \tag{2.1}
$$

where B_i is just B operating on f as a function of x_i and for $x \in E^n$ and $z \in E$, $\eta_j(x \mid z)$ is the element of E^n obtained from x by replacing x_j by z. Note that the larger $\beta(x_j, \mu_x)$ is, the more likely the jth particle is to "die" and be replaced by a copy (the *offspring*) of a randomly selected particle, so large β reduces the viability of an individual and corresponds to low fitness.

To avoid discussions of the domain of the generator, we assume that B is a bounded operator, that is, for $g \in B(E)$,

$$
Bg(z) = \lambda(z) \int_E (g(y) - g(z)) q(z, dy),
$$

where λ is a nonnegative, bounded, measurable function on E and q is a transition function on E. We will need to consider an unbounded generator later, but the extension of the results in that particular case follows by a simple projection argument. More general unbounded generators can be handled as in Donnelly and Kurtz (1996). If B is bounded, then A_0^n is also a bounded operator, and existence and uniqueness of solutions of the martingale problem for A_0^n are immediate.

For $\mu \in \mathscr{P}^n(E)$, there exists $x \in E^n$ such that $\mu = (1/n)\sum_{i=1}^n \delta_{x_i}$. For $m \le n$, let $\mu^{(m)}$ be the probability measure on E^m defined by

$$\mu^{(m)} = \frac{1}{n(n-1)\cdots(n-m+1)} \sum \delta_{(x_{i_1}, \dots, x_{i_m})},$$

where the sum is over all choices of $1 \le i_1, \dots, i_m \le n$ with $i_k \ne i_l$. Note that for $\mu \in \mathscr{P}^n(E)$, $\mu^{(m)}$ depends only on μ, not on the particular choice of x. For $f \in B(E^n)$, define $F \in B(\mathscr{P}^n(E))$ by

$$F(\mu) = \langle f, \mu^{(n)} \rangle$$

and $\mathbb{A}^n F$ by

$$\mathbb{A}^n F(\mu) = \langle A_0^n f, \mu^{(n)} \rangle.$$

It will be useful to note that if $f \in B(E^m)$, $m \le n$, then

$$\begin{aligned}\mathbb{A}^n F(\mu) &= \sum_{i=1}^m \langle B_i f, \mu^{(m)} \rangle + \frac{1}{2} \sum_{1 \le i \ne j \le m} \langle \Phi_{ij} f - f, \mu^{(m)} \rangle \\ &\quad + \frac{n-m}{n} \sum_{j=1}^m (\langle \beta(\cdot, \mu) \otimes f, \mu^{(m+1)} \rangle - \langle \beta_j(\cdot, \mu) f, \mu^{(m)} \rangle) \\ &\quad + \frac{1}{n} \sum_{1 \le i \ne j \le m} (\langle \beta(\cdot, \mu) \otimes \Phi_{ij} f, \mu^{(m)} \rangle - \langle \beta_j(\cdot, \mu) f, \mu^{(m)} \rangle),\end{aligned} \tag{2.2}$$

where $\Phi_{ij} f$ is the function of $m-1$ variables obtained by setting the ith and jth variables in f equal (note that a function g of $m-1$ variables can be viewed as a function of m variables and $\langle g, \mu^{(m)} \rangle = \langle g, \mu^{(m-1)} \rangle$). The product $h = \beta(\cdot, \mu) \otimes f$ is defined by

$$h(x_1, \dots, x_{m+1}, \mu) = \beta(x_{m+1}, \mu) f(x_1, \dots, x_m)$$

and for $1 \le j \le m$, $h_j = \beta_j(\cdot, \mu) f$ is defined by

$$h_j(x_1, \dots, x_m, \mu) = \beta(x_j, \mu) f(x_1, \dots, x_m).$$

In what follows, for $x \in E^n$, we will refer to x_i as the type of the particle at *level* i. We define a second generator for a Markov process in E^n by

$$\begin{aligned}A^n f(x) &= \sum_{i=1}^n B_i f(x) + \sum_{1 \le i < j \le n} (f(\theta_j(x \mid x_i)) - f(x)) \\ &\quad + \sum_{k=1}^n \sum_{1 \le i < j \le n} \left(\frac{2}{n^2} \beta(x_k, \mu_x) \right) (f(\theta_{jk}(x \mid x_i)) - f(x)).\end{aligned} \tag{2.3}$$

In this formula, for $x \in E^n$ and $z \in E$, $y = \theta_j(x \mid z)$ is the element of E^n satisfying

$$\begin{aligned} y_k &= x_k, & k &\le j-1, \\ y_j &= z, \\ y_k &= x_{k-1}, & k &> j \end{aligned}$$

so $\theta_j(x \mid x_i) = (x_1, \dots, x_{j-1}, x_i, x_j, \dots, x_{n-1})$, with the obvious modification if $j = 1$ or n. In words, we say that a copy of the particle at level i is inserted at level j and the top level particle is killed. Similarly, $y = \theta_{jk}(x \mid x_i)$ is the element of E^n obtained from x by killing the particle at level k and, in the resulting element of E^{n-1}, inserting a copy of the particle with index i at level j, creating the new element on E^n.

The first term on the right-hand side of (2.3) models the changes in the n particles due to mutation. We will refer to the transitions corresponding to the second term as *nonselective birth–death events* and refer to transitions corresponding to the third term as *selective birth–death events*.

Note that for A_0^n, the order in which the particles are numbered is not important. In particular, if $\tilde{X} = (\tilde{X}_1, \dots, \tilde{X}_n)$ is a solution of the martingale problem for A_0^n and $(\gamma_1, \dots, \gamma_n)$ is some permutation of $(1, \dots, n)$, then $(\tilde{X}_{\gamma_1}, \dots, \tilde{X}_{\gamma_n})$ is also a solution. On the other hand, order is critical in solutions of the martingale problem for A^n. For example, in each nonselective birth–death event, the highest numbered particle dies.

Again setting $F(\mu) = \langle f, \mu^{(n)} \rangle$, we have the following relationship between A_0^n and A^n:

$$\mathbb{A}^n F(\mu) = \langle A_0^n f, \mu^{(n)} \rangle = \langle A^n f, \mu^{(n)} \rangle. \tag{2.4}$$

This identity is the basis for the following theorem.

THEOREM 2.1. *Let A_0^n and A^n be given by* (2.1) *and* (2.3), *and let $\tilde{X}$ and X be solutions of the respective martingale problems. Define*

$$\tilde{Z}(t) = \frac{1}{n} \sum_{i=1}^{n} \delta_{\tilde{X}_i(t)}, \qquad Z(t) = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i(t)}.$$

Suppose $\tilde{X}(0)$ and $X(0)$ have the same exchangeable distribution. Then for each $t > 0$, $\tilde{X}(t)$ and $X(t)$ have the same exchangeable distribution,

$$E[f(X_1(t), \dots, X_n(t)) \mid \mathscr{F}_t^Z] = \langle f, Z(t)^{(n)} \rangle, \tag{2.5}$$

and $\tilde{Z}$ and Z have the same distribution on $D_{\mathscr{P}(E)}[0, \infty)$.

REMARK 2.2. (a) The analogous result in the neutral case was proved in Donnelly and Kurtz (1999) using a coupling argument. For the sake of variety, we use a conditioning argument here.

(b) The result also holds with θ_j replaced by η_j in the definition of A^n. The neutral version of that model was studied in Donnelly and Kurtz (1996). We could also take

$$\begin{aligned} A^n f(x) = {} & \sum_{i=1}^{n} B_i f(x) + \sum_{1 \le i < j \le n} (f(\theta_j(x \mid x_i)) - f(x)) \\ & + \sum_{1 \le i \ne j \le n} \left(\frac{1}{n} \beta(x_j, \mu_x) \right) (f(\eta_j(x \mid x_i)) - f(x)). \end{aligned} \tag{2.6}$$

(c) Note that we are only asserting that $X(t)$ and $\tilde{X}(t)$ have the same distribution. The *processes* X and $\tilde{X}$ clearly do not have the same distribution.

PROOF. Let $(\sigma_1, \ldots, \sigma_n)$ be some permuation of $(1, \ldots, n)$. By symmetry, if $\tilde{X}_1, \ldots, \tilde{X}_n$ is a solution of the martingale problem for A_0^n, then $(\tilde{X}_{\sigma_1}, \ldots, \tilde{X}_{\sigma_n})$ is a solution of the martingale problem for A_0^n. It follows that

$$\begin{aligned} &E[f(\tilde{X}_{\sigma_1}(t), \ldots, \tilde{X}_{\sigma_n}(t)) \mid \mathscr{F}_t^{\tilde{Z}}] \\ &\qquad - \int_0^t E[A_0^n f(\tilde{X}_{\sigma_1}(s), \ldots, \tilde{X}_{\sigma_n}(s)) \mid \mathscr{F}_s^{\tilde{Z}}]\, ds \end{aligned} \tag{2.7}$$

is an $\{\mathscr{F}_t^{\tilde{Z}}\}$ martingale. Averaging (2.7) over all permutations, we see that

$$\langle f, \tilde{Z}(t)^{(n)} \rangle - \int_0^t \langle A_0^n f, \tilde{Z}(s)^{(n)} \rangle \, ds$$

is an $\{\mathscr{F}_t^{\tilde{Z}}\}$ martingale. [Note that the conditioning in (2.7) can be dropped since the averaged expressions are measurable with respect to the conditioning σ algebra.] It follows that if $\tilde{X}$ is a solution of the martingale problem for A_0^n, then the corresponding empirical measure $\tilde{Z}$ is a solution of the martingale problem for $\mathbb{A}^n$. By (2.4), it follows that $\tilde{Z}$ is a solution of the martingale problem for $\mathbb{A}^n F(\mu) = \langle A^n f, \mu^{(n)} \rangle$.

We apply Corollary 3.5 of Kurtz (1998a). Let E (in the notation of that corollary) be E^n (in the notation of the present paper), $E_0 = \mathscr{P}_e(E^n)$ (the collection of exchangeable probability measures on E^n), $\gamma(x) = \mu_x^{(n)}$ and $\alpha(\nu, dz) = \nu(dz)$ for $\nu \in E_0 = \mathscr{P}_e(E^n)$. Then C, in the corollary, is $\mathbb{A}^n$ and it follows that there exists a solution X of the martingale problem for A^n such that the corresponding empirical measure process Z has the same distribution as $\tilde{Z}$. The identity (3.6) in Kurtz (1998a) implies (2.5), which in turn gives the exchangeability of $X(t)$.

Finally, if $\tilde{X}(0)$ is exchangeable, then the exchangeability of $\tilde{X}(t)$ follows from the symmetry of A_0^n, and uniqueness of solutions of the martingale problems implies $\tilde{X}(t)$ and $X(t)$ have the same distribution. □

2.2. *Models with fecundity selection.* We now consider models with fecundity selection. The Moran model now becomes

$$\begin{aligned} A_0^n f(x) = {} & \sum_{i=1}^n B_i f(x) \\ & + \frac{1}{2} \sum_{1 \le i \ne j \le n} \left(1 + \frac{2}{n} \sigma(x_i, \mu_x)\right)(f(\eta_j(x \mid x_i)) - f(x)), \end{aligned} \tag{2.8}$$

where, as before, σ is a nonnegative, bounded, measurable function on $E \times \mathscr{P}(E)$, $\sum_{i=1}^n B_i f$ is the generator for n independent copies of the E-valued Markov process with generator B and $\eta_j(x \mid x_i)$ is the element of E^n obtained from x by replacing x_j by x_i. Note that the only difference between (2.1) and

(2.8) is that in (2.8), β is replaced by σ which depends on the type of the parent rather than on the type of the particle to be killed, that is, selection affects the birth rate rather than the death rate. In particular, large σ gives a larger rate of reproduction and hence higher fitness.

We take the generator of the ordered model to be

$$
\begin{aligned}
A^n f(x) &= \sum_{i=1}^{n} B_i f(x) + \sum_{1\le i<j\le n} (f(\theta_j(x \mid x_i)) - f(x)) \\
&\quad + \sum_{i=1}^{n-1} \sum_{1\le j\le n} \left(\frac{1}{n}\sigma(x_i, \mu_x)\right)(f(\theta_j(x \mid x_i)) - f(x)),
\end{aligned} \tag{2.9}
$$

where θ_j is as above.

For $F(\mu) = \langle f, \mu^{(n)}\rangle$, we again define

$$
\mathbb{A}^n F(\mu) = \langle A_0^n f, \mu^{(n)}\rangle = \langle A^n f, \mu^{(n)}\rangle.
$$

To check that the second equality holds, note that

$$
\begin{aligned}
\langle A^n f, \mu^{(n)}\rangle &= \sum_{i=1}^{n} \langle B_i f, \mu^{(n)}\rangle + \sum_{1\le i<j\le n} (\langle \Phi_{ij} f, \mu^{(n)}\rangle - \langle f, \mu^{(n)}\rangle) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n-1}\sum_{j=1}^{i} (\langle \sigma_{i+1}\Phi_{i+1,j} f, \mu^{(n)}\rangle - \langle \sigma_i f, \mu^{(n)}\rangle) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} (\langle \sigma_i\Phi_{ij} f, \mu^{(n)}\rangle - \langle \sigma_i f, \mu^{(n)}\rangle) \\
&= \sum_{i=1}^{n} \langle B_i f, \mu^{(n)}\rangle + \frac{1}{2}\sum_{1\le i\ne j\le n} (\langle \Phi_{ij} f, \mu^{(n)}\rangle - \langle f, \mu^{(n)}\rangle) \\
&\quad + \frac{1}{n}\sum_{i=2}^{n}\sum_{j=1}^{i-1} (\langle \sigma_i\Phi_{ij} f, \mu^{(n)}\rangle - \langle \sigma_i f, \mu^{(n)}\rangle) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} (\langle \sigma_i\Phi_{ij} f, \mu^{(n)}\rangle - \langle \sigma_i f, \mu^{(n)}\rangle) \\
&= \langle A_0^n f, \mu^{(n)}\rangle.
\end{aligned}
$$

The appearance of $\sigma_{i+1}\Phi_{i+1,j} f$ in the second expression reflects the fact that if an offspring is inserted at a level j below the original level i of the parent, the new level of the parent is $i+1$. Equality of the second terms in the second and third expressions follows from the fact that $\langle \Phi_{ij} f, \mu^{(n)}\rangle = \langle \Phi_{ji} f, \mu^{(n)}\rangle$. If

f depends only on the first m variables ($m \le n$), then

$$\begin{aligned} \mathbb{A}^n F(\mu) = {} & \sum_{i=1}^{m} \langle B_i f, \mu^{(m)} \rangle + \sum_{1 \le i < j \le m} \langle \Phi_{ij} f - f, \mu^{(m)} \rangle \\ & + \frac{n-m}{n} \sum_{j=1}^{m} (\langle \sigma_j(\cdot, \mu) f, \mu^{(m)} \rangle - \langle \sigma(\cdot, \mu) \otimes f, \mu^{(m+1)} \rangle) \\ & + \frac{1}{n} \sum_{1 \le i \ne j \le m} (\langle \sigma_i(\cdot, \mu) \Phi_{ij} f, \mu^{(m)} \rangle - \langle \sigma_i(\cdot, \mu) f, \mu^{(m)} \rangle). \end{aligned} \tag{2.10}$$

The proof of the following theorem is exactly the same as the proof of Theorem 2.1.

THEOREM 2.3. *Let A_0^n and A^n be given by* (2.8) *and* (2.9), *and let $\tilde{X}$ and X be solutions of the respective martingale problems. Define*

$$\tilde{Z}(t) = \frac{1}{n} \sum_{i=1}^{n} \delta_{\tilde{X}_i(t)}, \qquad Z(t) = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i(t)}.$$

Suppose $\tilde{X}(0)$ and $X(0)$ have the same exchangeable distribution. Then for each $t > 0$, $\tilde{X}(t)$ and $X(t)$ have the same exchangeable distribution,

$$E[f(X_1(t), \ldots, X_n(t)) \mid \mathscr{F}_t^Z] = \langle f, Z(t)^{(m)} \rangle,$$

and $\tilde{Z}$ and Z have the same distribution on $D_{\mathscr{P}(E)}[0, \infty)$.

REMARK 2.4. We could replace θ_j by η_j in either the second or third terms in the definition of A^n. We will, in fact, make this substitution in the third term in the model considered in the next section.

2.3. *Models with recombination.* Recombination can be included in the above models. Let $R(x, y, dz)$ be a transition function from $E \times E \to E$ and let $\alpha > 0$. In the model to be considered, recombination events occur at rate α, and conditioned on the occurrence of a recombination, R gives the probability distribution that parents of types x and y have an offspring of type z. [See Ethier and Kurtz (1993) for additional discussion of general recombination.] We also specialize the selection mechanism to the usual diploid model, which is essentially equivalent to taking

$$\sigma(x, \mu) = \int_E \sigma(x, y) \mu(dy).$$

In the case of fecundity selection, the generator of the Moran model becomes

$$
\begin{aligned}
A_0^n f(x) = \sum_{i=1}^n B_i f(x) + \frac{1}{2n} \sum_{1\le i\ne j\ne k\le n} \left(1 + \frac{2}{n}\sigma(x_i, x_k)\right) \\
\times \left(\left(1 - \frac{2\alpha}{n}\right) f(\eta_j(x \mid x_i)) \right. \\
\left. + \frac{2\alpha}{n} \int_E f(\eta_j(x \mid z)) R(x_i, x_k, dz) - f(x)\right) \\
= \sum_{i=1}^n B_i f(x) + \frac{n-2}{2n}\left(1 - \frac{2\alpha}{n}\right) \sum_{1\le i\ne j\le n} (f(\eta_j(x \mid x_i)) - f(x)) \\
+ \frac{1}{n^2}\left(1 - \frac{2\alpha}{n}\right) \sum_{1\le i\ne j\ne k\le n} \sigma(x_i, x_k)(f(\eta_j(x \mid x_i)) - f(x)) \\
+ \frac{\alpha}{n^2} \sum_{1\le i\ne j\ne k\le n} \left(1 + \frac{2}{n}\sigma(x_i, x_k)\right) \\
\times \left(\int_E f(\eta_j(x \mid z)) R(x_i, x_k, dz) - f(x)\right)
\end{aligned} \tag{2.11}
$$

and the corresponding ordered model is

$$
\begin{aligned}
A^n f(x) = \sum_{i=1}^n B_i f(x) + \frac{n-2}{n}\left(1 - \frac{2\alpha}{n}\right) \sum_{1\le i< j\le n} (f(\theta_j(x \mid x_i)) - f(x)) \\
+ \frac{1}{n^2}\left(1 - \frac{2\alpha}{n}\right) \sum_{1\le i\ne k\le n-1} \sum_{j=1}^n \sigma(x_i, x_k)(f(\theta_j(x \mid x_i)) - f(x)) \\
+ \frac{\alpha}{n^2} \sum_{1\le i\ne k\le n-1} \sum_{j=1}^n \left(1 + \frac{2}{n}\sigma(x_i, x_k)\right) \\
\times \left(\int_E f(\theta_j(x \mid z)) R(x_i, x_k, dz) - f(x)\right),
\end{aligned} \tag{2.12}
$$

where θ_j and η_j are as above. The fact that

$$\langle A_0^n f, \mu^{(n)}\rangle = \langle A^n f, \mu^{(n)}\rangle$$

follows as before.

We will actually use a slightly different model, namely,

$$
\begin{aligned}
A_0^n f(x) &= \sum_{i=1}^n B_i f(x) + \frac{1}{2n} \sum_{1\le i\ne j\ne k\le n} \left(1+\frac{2}{n}\sigma(x_i,x_k)\right) \\
&\qquad\qquad \times (f(\eta_j(x\mid x_i)) - f(x)) \\
&\quad + \frac{\alpha}{n}\sum_{1\le i\ne k\le n}\int_E (f(\eta_i(x\mid z)) - f(x))R(x_i,x_k,dz) \\
&= \sum_{i=1}^n B_i f(x) + \frac{n-2}{2n}\sum_{1\le i\ne j\le n}(f(\eta_j(x\mid x_i)) - f(x)) \\
&\quad + \frac{1}{n^2}\sum_{1\le i\ne j\ne k\le n}\sigma(x_i,x_k)(f(\eta_j(x\mid x_i)) - f(x)) \\
&\quad + \frac{\alpha}{n}\sum_{1\le i\ne k\le n}\left(\int_E f(\eta_i(x\mid z))R(x_i,x_k,dz) - f(x)\right),
\end{aligned}
\tag{2.13}
$$

where the corresponding ordered model is (see Remark 2.4 above)

$$
\begin{aligned}
A^n f(x) &= \sum_{i=1}^n B_i f(x) + \frac{n-2}{n}\sum_{1\le i< j\le n}\big(f(\theta_j(x\mid x_i)) - f(x)\big) \\
&\quad + \frac{1}{n^2}\sum_{1\le i\ne j\ne k\le n}\sigma(x_i,x_k)\big(f(\eta_j(x\mid x_i)) - f(x)\big) \\
&\quad + \frac{\alpha}{n}\sum_{1\le i\ne k\le n}\left(\int_E f(\eta_i(x\mid z))R(x_i,x_k,dz) - f(x)\right).
\end{aligned}
\tag{2.14}
$$

This model essentially treats recombination as a type of mutation. We will see that for large n there is little difference between the two models.

The proof of the following theorem is exactly the same as the proof of Theorem 2.1.

THEOREM 2.5. *Let A_0^n and A^n be given by* (2.13) *and* (2.14), *and let $\tilde{X}$ and X be solutions of the respective martingale problems. Define*

$$\tilde{Z}(t) = \frac{1}{n}\sum_{i=1}^n \delta_{\tilde{X}_i(t)}, \qquad Z(t) = \frac{1}{n}\sum_{i=1}^n \delta_{X_i(t)}.$$

Suppose $\tilde{X}(0)$ and $X(0)$ have the same exchangeable distribution. Then for each $t > 0$, $\tilde{X}(t)$ and $X(t)$ have the same exchangeable distribution,

$$E\big[f(X_1(t),\ldots,X_n(t)) \mid \mathscr{F}_t^Z\big] = \langle f, Z(t)^{(m)}\rangle,$$

and $\tilde{Z}$ and Z have the same distribution on $D_{\mathscr{P}(E)}[0,\infty)$.

3. Infinite-population limit. The point of introducing the ordered generators in (2.3), (2.9) and (2.14) is that they are well behaved (that is, converge in an appropriate sense) as the population size $n \to \infty$, whereas the original Moran model generators are not. Letting X^n denote the solution of the martingale problem for A^n (since the dependence on n is now important) and letting $Z_n(t) = (1/n)\sum_{i=1}^n \delta_{X_i^n(t)}$, we can view X^n as determining a process in E^∞ in which components with indices greater than n do not vary. Consequently, we can view A^n as an operator on functions in $B(E^\infty)$. For example, in the case of viability selection, if f depends only on the first m variables, then as $n \to \infty$, the generator $A^n f$ given by (2.3) converges to

$$\begin{aligned} Af(x,\mu) &= \sum_{i=1}^m B_i f(x) + \sum_{1\le i<j\le m} (f(\theta_j(x \mid x_i)) - f(x)) \\ &\quad + \sum_{k=1}^m \beta(x_k,\mu)(f(\psi_k(x)) - f(x)), \end{aligned} \tag{3.1}$$

where for $x \in E^\infty$, $\psi_k(x)$ is the element in E^∞ obtained by eliminating the kth component and shifting the components above level k down one; that is, $y = \psi_k(x)$ satisfies

$$\begin{aligned} y_i &= x_i, \qquad i < k, \\ y_i &= x_{i+1}, \qquad i \ge k. \end{aligned}$$

In particular, if f depends on the first m components of x, then $Af(x,\mu)$ depends on the first $m+1$ components. Similarly, the generator given by (2.6) converges to

$$\begin{aligned} Af(x,\mu) &= \sum_{i=1}^m B_i f(x) + \sum_{1\le i<j\le m} (f(\theta_j(x \mid x_i)) - f(x)) \\ &\quad + \sum_{j=1}^m \int_E \beta(x_j,\mu)(f(\eta_j(x \mid y)) - f(x))\mu(dy). \end{aligned} \tag{3.2}$$

Observing that for large n and $\mu \in \mathscr{P}^n(E)$, $\mu^{(m)}$ is essentially product measure μ^m, for $F(\mu) = \langle f, \mu^m\rangle$, we can identify the limit of $\mathbb{A}^n$ as

$$\begin{aligned} \mathbb{A}F(\mu) &= \sum_{i=1}^m \langle B_i f, \mu^m\rangle + \tfrac{1}{2}\sum_{1\le i\ne j\le m} \langle \Phi_{ij} f - f, \mu^m\rangle \\ &\quad + \sum_{j=1}^m (\langle \beta(\cdot,\mu)\otimes f, \mu^{m+1}\rangle - \langle \beta_j(\cdot,\mu) f, \mu^m\rangle), \end{aligned} \tag{3.3}$$

which satisfies

$$\mathbb{A}F(\mu) = \langle Af(\cdot,\mu), \mu^{m+1}\rangle.$$

To make the meaning of this convergence precise, suppose that for each n, $X^n(0)$ is exchangeable and that $X^n(0) \Rightarrow X(0)$ [$X(0)$ will necessarily be exchangeable]. Then it is easy to check that the sequence of processes (X^n, Z_n) is relatively compact in $D_{E^\infty \times \mathscr{P}(E)}[0,\infty)$ in the sense of convergence in distribution in the Skorohod topology, and, assuming f, $B_i f$ and β are all continuous, any limit point (X, Z) will have the property that

$$f(X(t)) - \int_0^t Af(X(s), Z(s))\,ds \tag{3.4}$$

is an $\{\mathscr{F}_t^{X,Z}\}$ martingale and

$$\begin{aligned} &\langle f, Z^m(t)\rangle - \int_0^t \big\langle Af(\cdot, Z(s)), Z^{m+1}(s)\big\rangle\,ds \\ &\quad = F(Z(t)) - \int_0^t \mathbb{A}F(Z(s))\,ds \end{aligned} \tag{3.5}$$

is an $\{\mathscr{F}_t^{Z}\}$ martingale. [Note that the distinction between the two filtrations is essential. In general, (3.5) will not be an $\{\mathscr{F}_t^{X,Z}\}$ martingale. See Theorem 2.7 of Donnelly and Kurtz (1996).]

By the exchangeability of $X^n(t)$ and the definition of Z_n,

$$E\big[f(X_1^n(t),\dots,X_m^n(t))\big] = E\big[\langle f, Z_n^{(m)}(t)\rangle\big],$$

and passing to the limit, we have

$$E\big[f(X_1(t),\dots,X_m(t))\big] = E\big[\langle f, Z^m(t)\rangle\big].$$

It follows that conditional on $Z(t)$, $X_1(t), X_2(t), \dots$ are iid with distribution $Z(t)$, and hence that $X(t)$ is exchangeable with de Finetti measure Z. In particular,

$$Z(t) = \lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^m \delta_{X_i(t)} \quad \text{a.s.} \tag{3.6}$$

and hence, Z is $\{\mathscr{F}_t^X\}$-adapted and $\mathscr{F}_t^{X,Z} = \mathscr{F}_t^X$.

For fecundity selection, the limiting operators are

$$\begin{aligned} Af(x,\mu) &= \sum_{i=1}^m B_i f(x) + \sum_{1\le i<j\le m}\big(f(\theta_j(x\mid x_i)) - f(x)\big) \\ &\quad + \sum_{1\le j\le m}\int_E \sigma(z,\mu)\big(f(\theta_j(x\mid z)) - f(x)\big)\mu(dz), \end{aligned} \tag{3.7}$$

where, as above, $\theta_j(x\mid z)$ is the element of E^∞ obtained from x by inserting $z\in E$ at the jth level, and

$$\begin{aligned} \mathbb{A}F(\mu) &= \sum_{i=1}^m \langle B_i f, \mu^m\rangle + \sum_{1\le i<j\le m}\langle \Phi_{ij} f - f, \mu^m\rangle \\ &\quad + \sum_{j=1}^m \big(\langle \sigma_j(\cdot,\mu) f, \mu^m\rangle - \langle \sigma(\cdot,\mu)\otimes f, \mu^{m+1}\rangle\big). \end{aligned} \tag{3.8}$$

Note that if $\sigma \le \bar{\sigma}$ and we let $\beta = \bar{\sigma} - \sigma$, then (3.3) and (3.8) are the same. Of course, $\mathbb{A}$ is the generator of the Fleming–Viot process with selection. See, for example, Ethier and Kurtz (1993).

The infinite-population limit for the generator given in (2.12) is

$$\begin{aligned} Af(x,\mu) = {} & \sum_{i=1}^m B_i f(x) + \sum_{1\le i<j\le m} (f(\theta_j(x \mid x_i)) - f(x)) \\ & + \sum_{j=1}^m \int_{E\times E} \sigma(y_1, y_2)(f(\theta_j(x \mid y_1)) - f(x))\mu(dy_1)\mu(dy_2) \\ & + \alpha \sum_{j=1}^m \int_{E\times E} \left(\int_E f(\theta_j(x \mid z)) R(y_1, y_2, dz) - f(x) \right) \\ & \qquad \times \mu(dy_1)\mu(dy_2) \end{aligned} \tag{3.9}$$

while the limit for the generator given in (2.14) is

$$\begin{aligned} Af(x,\mu) = {} & \sum_{i=1}^m B_i f(x) + \sum_{1\le i<j\le m} (f(\theta_j(x \mid x_i)) - f(x)) \\ & + \sum_{j=1}^m \int_{E\times E} \sigma(y_1, y_2)(f(\eta_j(x \mid y_1)) - f(x))\mu(dy_1)\mu(dy_2) \\ & + \alpha \sum_{j=1}^m \int_{E} \left(\int_E f(\eta_j(x \mid z)) R(x_j, y, dz) - f(x) \right) \mu(dy). \end{aligned} \tag{3.10}$$

Letting $H_j f(x_1, \ldots, x_{m+1}) = \int_E f(\eta_j(x \mid z)) R(x_j, x_{m+1}, dz)$, the limit of $\mathbb{A}^n$,

$$\begin{aligned} \mathbb{A}F(\mu) = {} & \sum_{i=1}^m \langle B_i f, \mu^m \rangle + \sum_{1\le i<j\le m} \langle \Phi_{ij} f - f, \mu^m \rangle \\ & + \sum_{j=1}^m (\langle \sigma_j(\cdot,\cdot) f, \mu^{m+1} \rangle - \langle \sigma(\cdot,\cdot) \otimes f, \mu^{m+2} \rangle) \\ & + \alpha \sum_{j=1}^m (\langle H_j f - f, \mu^{m+1} \rangle), \end{aligned} \tag{3.11}$$

is the same for both of these models, partially justifying our claim that the two approaches to the treatment of recombination are asymptotically equivalent.

Note that we are not really claiming to have proved a limit theorem in this section in that we have not yet proved a uniqueness result for the limiting martingale problem and there are technical difficulties to overcome unless one makes continuity assumptions on B, σ, β and R. Uniqueness for the martingale problem for $\mathbb{A}$ given by (3.11) follows by duality arguments [see Ethier and Kurtz (1993)]. In Section 4.1, we will prove uniqueness for a system of stochastic differential equations corresponding to (3.10), and in Section 7 we will prove convergence of the finite-dimensional system corresponding to

(2.14). In Section 4.2, we will consider uniqueness for the martingale problem for the infinite system.

4. Demographic representation of infinite-population models with selection and recombination. In Donnelly and Kurtz (1996, 1999), the "demography" of the infinite particle models was represented in terms of a collection of independent unit Poisson processes $\{L_{ij},\ 1 \le i < j\}$. In the construction of the solution of the martingale problem for the neutral (no selection) model without recombination given by

$$Af(x) = \sum_{i=1}^{m} B_i f(x) + \sum_{1 \le i < j \le m} (f(\theta_j(x \mid x_i)) - f(x)), \tag{4.1}$$

the jump times of L_{ij} determine when level j "looks down" to level i and a copy of the type at level i is inserted at level j. To be precise, suppose that $E \subset \mathbb{R}^d$ (or some more general linear space) and that we can write the mutation process as a solution of a stochastic differential equation

$$X_0(t) = X_0(0) + \int_{U\times[0,\,t]} h(X_0(s-), u) M_0(du \times ds), \tag{4.2}$$

where M_0 is a Poisson random measure on $U \times [0, \infty)$ with mean measure $\nu \times m$ for some measure ν on U and Lebesgue measure m. Any pure jump Markov process in E can be represented in this way. For X_0 satisfying (4.2), the corresponding generator is

$$Bf(x) = \int (f(x + h(x, u)) - f(x))\nu(du).$$

For simplicity, we assume that B is a bounded operator which allows us to assume that $\nu(U) < \infty$.

With this mutation process, the solution of the martingale problem for (4.1) can be obtained as the solution $X(t) = (X_1(t), X_2(t), \ldots)$ of the system of stochastic differential equations

$$\begin{aligned} X_j(t) = X_j(0) &+ \int_{U\times[0,\,t]} h(X_j(s-), u) M_j(du \times ds) \\ &+ \sum_{i=1}^{j-1} \int_0^t (X_i(s-) - X_j(s-))\, dL_{ij}(s) \\ &+ \sum_{1 \le i < k < j} \int_0^t (X_{j-1}(s-) - X_j(s-))\, dL_{ik}(s), \end{aligned} \tag{4.3}$$

where the $\{M_j\}$ are independent copies of M_0 and independent of the $\{L_{ij}\}$. (Of course, we always assume that $X(0)$ is independent of the driving Poisson processes.) The second term on the right-hand side determines the evolution of the mutation process on level j, the third term determines when a copy of the type at level i is inserted at level j and the fourth term reflects the insertion of new particles at levels below level j.

Note that when a copy of the particle at level i is inserted at level j, we can interpret the particle at level i as being the parent of the new particle at level j, and consequently, we can trace the ancestry of any particle back to time zero. (If the mutation process has a stationary version, we can construct a version with time starting at $-\infty$ and trace ancestry into the indefinite past.) For each $t \geq 0$ and $k = 1, 2, \ldots$, let $N_k^t(s)$, $0 \leq s \leq t$, be the level at time s of the ancestor of the particle at level k at time t. In terms of the L_{ij}, for $0 \leq s \leq t$,

$$\begin{aligned} N_k^t(s) = k &- \sum_{1 \leq i < j < k} \int_s^t I_{\{N_k^t(u) > j\}}\, dL_{ij}(u) \\ &- \sum_{1 \leq i < j \leq k} \int_s^t (j - i) I_{\{N_k^t(u) = j\}}\, dL_{ij}(u)\,. \end{aligned} \tag{4.4}$$

For $s < t$, the set of values

$$\Gamma(s, t) = \big\{ N_k^t(s) : k = 1, 2, \ldots \big\} \tag{4.5}$$

is finite; that is, only finitely many individuals alive at time s have descendants alive at time t, and for each t, $\Gamma(\cdot, t)$ determines a version of Kingman's (1982) coalescent which models the genealogy of a neutral population. [See the discussions of genealogy in Donnelly and Kurtz (1996, 1999).]

Griffiths and others [see Griffiths and Majoram (1997) for a recent survey] have considered genealogies for models with recombination, introducing an *ancestral recombination graph* as a generalization of the coalescent. Krone and Neuhauser (1997) have given a description of genealogy for a population subject to certain kinds of selection in terms of a corresponding *ancestral selection graph*. In our setting, which includes both selection and recombination, we will refer to the analogous object as the *ancestral influence graph*. We will generalize the system (4.3) to the model with generator (3.10) in such a way that for each t, the ancestral influence graph for the infinite population at time t is embedded in the corresponding demography in a manner analogous to the embedding of Kingman's coalescent in the neutral demography.

Note that if $\bar{\sigma} \geq \sigma$, then we can rewrite A as

$$\begin{aligned} Af(x, \mu) = \sum_{i=1}^m B_i f(x) &+ \sum_{1 \leq i < j \leq m} (f(\theta_j(x \mid x_i)) - f(x)) \\ &+ \sum_{j=1}^m \bar{\sigma} \int_{E \times E} \bigg(\frac{\sigma(y_1, y_2)}{\bar{\sigma}} f(\eta_j(x \mid y_1)) \\ &\qquad + \frac{\bar{\sigma} - \sigma(y_1, y_2)}{\bar{\sigma}} f(x) - f(x) \bigg) \mu(dy_1)\mu(dy_2) \\ &+ \sum_{j=1}^m \alpha \int_E \bigg(\int_E f(\eta_j(x \mid z)) R(x_j, y, dz) - f(x) \bigg) \mu(dy). \end{aligned} \tag{4.6}$$

Written in this form, the jth term in the selection operator has the following interpretation. At the jump times of a Poisson process with intensity $\bar{\sigma}$, a sample of size 2, say U_1 and U_2, is drawn from the empirical measure Z (that is, if the jump time is τ, $E[f(U_1, U_2) \mid Z(\tau)] = \langle f, Z(\tau)^2 \rangle$). Conditioned on U_1 and U_2, with probability $\bar{\sigma}^{-1}\sigma(U_1, U_2)$, the particle at level j is replaced by a particle of type U_1, and with probability $\bar{\sigma}^{-1}(\bar{\sigma} - \sigma(U_1, U_2))$, there is no change, that is, $X(\tau) = X(\tau-)$. To describe the recombination events, let $r: E \times E \times [0,1] \to E$ have the property that if V is uniformly distributed on $[0,1]$, then $r(y_1, y_2, V)$ has distribution $R(y_1, y_2, \cdot)$. At the jump times of a Poisson process with parameter α, a sample U is drawn from the empirical measure Z, a uniform random variable V is generated and the type on the jth level is set equal to $r(X_j(\tau-), U, V)$.

To build these descriptions into a stochastic equation, we must have a way to generate samples from the empirical measure. Let $\rho: \mathscr{P}(E) \times [0,1] \to E$ have the property that if $\mu \in \mathscr{P}(E)$ and V is uniformly distributed on $[0,1]$, then $\rho(\mu, V)$ has distribution μ. [See Blackwell and Dubins (1983) for a construction of such a ρ with nice continuity properties.] We will later give a special method for generating ρ that has other uses in our setting, but for the moment, we leave the choice of ρ arbitrary. For $j = 1, 2, \ldots$, let K_j be a Poisson random measure on $[0,1]^3 \times [0, \infty)$ with mean measure $\bar{\sigma} m^4$, where m^d is d-dimensional Lebesgue measure, and let J_j be a Poisson random measure on $[0,1]^2 \times [0,\infty)$ with mean measure αm^3. Of course $K_j([0,1]^3 \times [0, \cdot])$ is a Poisson process with intensity $\bar{\sigma}$ and $J_j([0,1]^2 \times [0, \cdot])$ is a Poisson process with intensity α. Following the terminology of Krone and Neuhauser, we will refer to the jump times of the K_j as $\bar{\sigma}$-branch points and the jump times of the J_j as α-branch points.

The desired system of equations is given by

$$
\begin{aligned}
X_j(t) = X_j(0) &+ \int_{U \times [0,t]} h(X_j(s-), u) M_j(du \times ds) \\
&+ \sum_{i=1}^{j-1} \int_0^t (X_i(s-) - X_j(s-)) \, dL_{ij}(s) \\
&+ \sum_{1 \le i < k < j} \int_0^t (X_{j-1}(s-) - X_j(s-)) \, dL_{ik}(s) \\
&+ \int_{[0,1]^3 \times [0,t]} \big(\rho(Z(s-), u_1) - X_j(s-)\big) \\
&\quad \times I_{[0, \bar{\sigma}^{-1}\sigma(\rho(Z(s-), u_1), \rho(Z(s-), u_2))]}(u_3) \\
&\quad \times K_j(du_1 \times du_2 \times du_3 \times ds) \\
&+ \int_{[0,1]^2 \times [0,t]} \big(r(X_j(s-), \rho(Z(s-), u_1), u_2) - X_j(s-)\big) \\
&\quad \times J_j(du_1 \times du_2 \times ds).
\end{aligned} \tag{4.7}
$$

To understand the K_j integral term, note that

$$N_{\sigma, j}(t) = \int_{[0,1]^3 \times [0,t]} I_{[0, \bar{\sigma}^{-1}\sigma(\rho(Z(s-), u_1), \rho(Z(s-), u_2))]}(u_3) \times K_j(du_1 \times du_2 \times du_3 \times ds)$$

is a counting process with intensity $\langle \sigma, Z(t)^2 \rangle$, and at each jump time $\tau_{\sigma, j, k}$ of $N_{\sigma, j}$, the type at level j is replaced by a random type with distribution satisfying

$$E[g(X_j(\tau_{\sigma, j, k})) \mid Z(\tau_{\sigma, j, k})] = \frac{\langle \sigma_1 g, Z(\tau_{\sigma, j, k})^2 \rangle}{\langle \sigma, Z(\tau_{\sigma, j, k})^2 \rangle},$$

where $\sigma_1 g(y_1, y_2) = \sigma(y_1, y_2) g(y_1)$. For the more general σ appearing in (3.7) (but with the insertion operator θ_j replaced by η_j), the selection term in (4.7) can be replaced by

$$\int_{[0,1]^2 \times [0,t]} (\rho(Z(s-), u_1) - X_j(s-)) I_{[0, \bar{\sigma}^{-1}\sigma(\rho(Z(s-), u_1), Z(s-))]}(u_2) \times K_j(du_1 \times du_2 \times ds). \tag{4.8}$$

Note that this term gives a different system in the case of diploid σ, reflecting the fact that different systems of stochastic differential equations may correspond to the same martingale problem.

4.1. *Uniqueness for the infinite system.* In this subsection we consider a more general system that combines the selection and recombination terms into a single term. In place of $\{K_j\}$ and $\{J_j\}$ in (4.7), the equations will be driven by Poisson random measures (denoted $\{K_j\}$) on a general space of the form $S \times [0, \infty)$, where for definiteness we will assume that S is a complete, separable metric space ($S = [0,1]^3 \cup [0,1]^2$ in (4.7)). Specifically, the K_j will be independent Poisson random measures on $S \times [0, \infty)$, each with mean measure $\zeta \times m$ for some finite measure ζ on S.

Fix $\mu_0 \in \mathscr{P}(E)$, and define γ: $E^\infty \to \mathscr{P}(E)$ by

$$\gamma(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i},$$

if the limit exists in the weak topology and by $\gamma(x) = \mu_0$ otherwise. The system is

$$\begin{aligned} X_j(t) = X_j(0) &+ \int_{U \times [0,t]} h(X_j(s-), u) M_j(du \times ds) \\ &+ \sum_{i=1}^{j-1} \int_0^t (X_i(s-) - X_j(s-))\, dL_{ij}(s) \\ &+ \sum_{1 \le i < k < j} \int_0^t (X_{j-1}(s-) - X_j(s-))\, dL_{ik}(s) \\ &+ \int_{S \times [0,t]} (q(X_j(s-), Z(s-), u) - X_j(s-)) K_j(du \times ds), \end{aligned} \tag{4.9}$$

where $Z(t) = \gamma(X(t))$ and $X(0)$, $\{M_j\}$, $\{L_{ij}\}$ and $\{K_j\}$ are independent. Applying Itô's formula, we see that any solution of (4.9) is a solution of the martingale problem for

$$\begin{aligned} Af(x,\mu) &= \sum_{i=1}^{m} B_i f(x) + \sum_{1\le i<j\le m} (f(\theta_j(x\mid x_i)) - f(x)) \\ &\quad + \sum_{j=1}^{m} \beta \int_E (f(\eta_j(x\mid z)) - f(x)) Q(x_j,\mu,dz), \end{aligned} \tag{4.10}$$

where $\beta = \zeta(S)$ and Q is the transition function given by

$$Q(x_j,\mu,C) = \frac{1}{\beta}\int_S I_{\{q(x_j,\mu,u)\in C\}}\zeta(du).$$

It will actually be convenient to think of q as a function of x rather than μ and to allow q to depend on an additional stochastic process ξ.

To be specific, let $\{\mathcal{F}_t\}$ be a filtration and assume that $\{K_j\}$, $\{L_{ij}\}$ and $\{M_j\}$ are compatible with $\{\mathcal{F}_t\}$ and $X(0)$ is $\mathcal{F}_0$-measurable. Let ξ be a cadlag $\{\mathcal{F}_t\}$-adapted process with values in a complete, separable metric space F. Assume that

$$q\colon E\times E^\infty \times F\times S\to E. \tag{4.11}$$

The system then becomes

$$\begin{aligned} X_j(t) &= X_j(0) + \int_{U\times[0,t]} h(X_j(s-),u) M_j(du\times ds) \\ &\quad + \sum_{i=1}^{j-1}\int_0^t (X_i(s-) - X_j(s-))\, dL_{ij}(s) \\ &\quad + \sum_{1\le i<k<j}\int_0^t (X_{j-1}(s-) - X_j(s-))\, dL_{ik}(s) \\ &\quad + \int_{S\times[0,t]} (q(X_j(s-), X(s-), \xi(s-), u) - X_j(s-)) K_j(du\times ds). \end{aligned} \tag{4.12}$$

A solution of (4.12) will still be a solution of the martingale problem for (4.10) if

$$\begin{aligned} &Q(X_j(t-), Z(t-), C) \\ &\quad = \frac{1}{\beta}\int_S I_{\{q(X_j(t-),X(t-),\xi(t-),u)\in C\}}\zeta(du), \quad \text{a.s.} \end{aligned} \tag{4.13}$$

The basic assumption on q will be that there exists a constant $D > 0$ such that

$$\begin{aligned} &\int_S I_{\{q(x_j,x,v,u)\ne q(y_j,y,v,u)\}}\zeta(du) \\ &\quad \le D\left(I_{\{x_j\ne y_j\}} + \limsup_{m\to\infty}\frac{1}{m}\sum_{i=1}^{m} I_{\{x_i\ne y_i\}}\right) \quad \text{a.s.} \end{aligned} \tag{4.14}$$

Taking $S = [0,1]^3 \cup [0,1]^2$ and appropriately defining ζ and q, (4.7) is a special case of (4.12).

We can write

$$M_j = \sum_{k=1}^{\infty} \delta_{(u_k^{1,j}, s_k^{1,j})}, \qquad K_j = \sum_{k=1}^{\infty} \delta_{(u_k^{2,j}, s_k^{2,j})}.$$

Let

$$\Lambda = \{(j, s_k^{1,j}) : j, k = 1, 2, \dots\} \cup \{(j, s_k^{2,j}) : j, k = 1, 2, \dots\} \cup \{(j, 0) : j = 1, 2, \dots\}.$$

For $N_k^t(s)$ given by (4.4), that is, the ancestral level in the neutral genealogy, let

$$\tau_k^t = \sup\{s < t : (N_k^t(s), s) \in \Lambda\}$$

and note that $X_k(t) = X_{N_k^t(\tau_k^t)}(\tau_k^t)$. Letting $\Lambda_t = \{(j, s) \in \Lambda : s < t\}$, it follows that Z must be of the form

$$Z(t) = \sum_{(j,s)\in\Lambda_t} a(j, s, t)\delta_{X_j(s)}, \tag{4.15}$$

where

$$a(j, s, t) = \lim_{m\to\infty} \frac{1}{m} \sum_{k=1}^{m} I_{\{(N_k^t(\tau_k^t), \tau_k^t)=(j,s)\}}.$$

Properties of the neutral infinite alleles model imply that, for $t > 0$, this limit exists and $\sum_{(j,s)\in\Lambda_t} a(j, s, t) = 1$ a.s. One consequence of the representation (4.15) is that the system (4.12) makes sense even without an assumption of exchangeability. For any initial condition $X(0)$, under appropriate conditions on q, we can construct a solution iteratively. Let $X_j^0(t) \equiv X_j(0)$ and for $n = 0, 1, \dots,$

$$Z^n(t) = \sum_{(j,s)\in\Lambda_t} a(j, s, t)\delta_{X_j^n(s)} = \lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} \delta_{X_i^n(t)} \tag{4.16}$$

and

$$\begin{aligned} X_j^{n+1}(t) = X_j(0) &+ \int_{U\times[0,t]} h(X_j^{n+1}(s-), u) M_j(du \times ds) \\ &+ \sum_{i=1}^{j-1} \int_0^t (X_i^{n+1}(s-) - X_j^{n+1}(s-))\, dL_{ij}(s) \\ &+ \sum_{1\le i<k<j} \int_0^t (X_{j-1}^{n+1}(s-) - X_j^{n+1}(s-))\, dL_{ik}(s) \\ &+ \int_{S\times[0,t]} (q(X_j^{n+1}(s-), X^n(s-), \xi(s-), u) - X_j^{n+1}(s-)) \\ &\quad \times K_j(du \times ds). \end{aligned} \tag{4.17}$$

THEOREM 4.1. *Suppose that there exists a constant $D > 0$ such that* (4.14) *holds. Then*:

(a) $X = \lim_{n\to\infty} X^n$ *exists and* X *is the unique solution of* (4.12).

(b) *If* $X(0)$ *is exchangeable and* (4.13) *holds, then for each* $f \in B(E^m)$, $m = 1, 2, \ldots,$

$$E\big[f(X_1(t), \ldots, X_m(t)) \mid \mathscr{F}_t^{\gamma(X)}\big] = \langle f, \gamma(X(t))^m\rangle.$$

In particular, for each $t \geq 0$, $X(t)$ *is exchangeable.*

(c) *For arbitrary* $X(0)$, *if* (4.13) *holds, then for each* $f \in B(E^m)$, $m = 1, 2, \ldots,$

$$\lim_{t\to\infty} E[f(X_1(t), \ldots, X_m(t)) \mid \mathscr{F}_t^{\gamma(X)}] - \langle f, \gamma(X(t))^m\rangle = 0.$$

PROOF. Let $\tilde{X}_j^{T,n}(t) = X^n_{N_j^T(t)}(t)$, where again N_j^T gives the ancestral level in the neutral genealogy determined by (4.4). For each $T > 0$ and $j = 1, 2, \ldots,$ define

$$\tilde{M}_j^T(C \times [0, t]) = \sum_{i=1}^{j} \int_{C\times[0,t]} I_{\{N_j^T(s)=i\}} M_i(du \times ds)$$

and define $\tilde{K}_j$ similarly. Since the L_{ij} are independent of the M_j and K_j, it follows that $\tilde{M}_j$ and $\tilde{K}_j$ are Poisson random measures with mean measures $\nu \times m$ and $\zeta \times m$, respectively (but not independent). Then

$$(4.18)\qquad \begin{aligned} \tilde{X}_j^{T,n+1}(t) = X_{N_j^T(0)}(0) &+ \int_{U\times[0,t]} h(\tilde{X}_j^{T,n+1}(s-), u)\tilde{M}_j(du \times ds) \\ &+ \int_{S\times[0,t]} (q(\tilde{X}_j^{T,n+1}(s-), X^n(s-), \xi(s-), u) - \tilde{X}_j^{T,n+1}(s-)) \\ &\qquad \times \tilde{K}_j(du \times ds), \end{aligned}$$

and from this equation, it follows that $\tilde{X}_j^{T,n+1}(s) = \tilde{X}_j^{T,n}(s)$ until the time of the first jump in $\tilde{K}_j(du \times ds)$ at which $X^n(s-) \neq X^{n-1}(s-)$ [or, more precisely, $q(\tilde{X}_j^{T,n}(s-), X^n(s-), \xi(s-), u) \neq q(\tilde{X}_j^{T,n}(s-), X^{n-1}(s-), \xi(s-), u)$]. Consequently, setting $V_j^{T,n+1}(t) = I_{\{\tilde{X}_j^{T,n+1}(t)\neq \tilde{X}_j^{T,n}(t)\}}$ [and hence $1 - V_j^{T,n+1}(t) = I_{\{\tilde{X}_j^{T,n+1}(t)= \tilde{X}_j^{T,n}(t)\}}$] yields

$$\begin{aligned} &V_j^{T,n+1}(t) \\ &\quad \leq \int_{S\times[0,t]} (1 - V_j^{T,n+1}(s-)) \\ &\qquad \times I_{\{q(\tilde{X}_j^{T,n}(s-), X^n(s-), \xi(s-), u)\neq q(\tilde{X}_j^{T,n}(s-), X^{n-1}(s-), \xi(s-), u)\}} \tilde{K}_j(du \times ds) \end{aligned}$$

$$
\begin{aligned}
&= \int_{S\times[0,t]} (1 - V_j^{T,n+1}(s-)) \\
&\quad \times I_{\{q(\tilde{X}_j^{T,n}(s-),\, X^n(s-),\, \xi(s-),\, u)\neq q(\tilde{X}_j^{T,n}(s-),\, X^{n-1}(s-), \xi(s-),\, u)\}} \\
&\quad \times (\tilde{K}_j(du\times ds) - \zeta(du)\,ds) \\
&\quad + \int_{S\times[0,t]} (1 - V_j^{T,n+1}(s-)) \\
&\qquad \times I_{\{q(\tilde{X}_j^{T,n}(s-),\, X^n(s-),\, \xi(s-),\, u)\neq q(\tilde{X}_j^{T,n}(s-),\, X^{n-1}(s-),\, \xi(s-),\, u)\}} \zeta(du)\,ds \\
&\le \text{ martingale } + D\int_0^t R^n(s)\,ds,
\end{aligned}
$$

where

$$
R^n(s) = \sum_{(i,r)\in\Lambda_s} a(i,r,s) I_{\{X_i^n(r)\neq X_i^{n-1}(r)\}} = \lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^m I_{\{X_i^n(s)\neq X_i^{n-1}(s)\}}.
$$

Setting $t = T$ and taking expectations of both sides, we have

$$
\begin{aligned}
P\{X_j^{n+1}(T) \neq X_j^n(T)\} &\le D\int_0^T E[R^n(s)]\,ds \\
&= D\int_0^T \lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^m P\{X_i^n(s)\neq X_i^{n-1}(s)\}\,ds,
\end{aligned}
\tag{4.19}
$$

which in turn implies

$$
E[R^{n+1}(T)] \le D\int_0^T E[R^n(s)]\,ds.
$$

It follows that $\lim_{n\to\infty} E[R^n(T)] = 0$, so by (4.19), $\lim_{n\to\infty} P\{X_j^{n+1}(T) \neq X_j^n(T)\} = 0$ for each j and hence X^n converges. Uniqueness follows by essentially the same estimates.

Let V be an E^∞-valued process adapted to $\{\mathscr{F}_t\}$ for which the empirical measure process

$$
Z^V(t) = \lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^m \delta_{V_i(t)} = \gamma(V(t))
$$

exists, and let Y satisfy

$$
\begin{aligned}
Y_j(t) &= Y_j(0) + \int_{U\times[0,t]} h(Y_j(s-),u) M_j(du\times ds) \\
&\quad + \sum_{i=1}^{j-1}\int_0^t (Y_i(s-) - Y_j(s-))\,dL_{ij}(s) \\
&\quad + \sum_{1\le i<k<j}\int_0^t (Y_{j-1}(s-) - Y_j(s-))\,dL_{ik}(s) \\
&\quad + \int_{S\times[0,t]} (q(Y_j(s-), V(s-), \xi(s-), u) - Y_j(s-)) K_j(du\times ds).
\end{aligned}
\tag{4.20}
$$

If $Y(0)$ is exchangeable and (4.13) holds, then we claim that

$$E[f(Y_1(t),\dots,Y_m(t)) \mid \mathscr{F}_t^{\gamma(Y),\gamma(V)}] = \langle f, \gamma(Y(t))^m\rangle. \tag{4.21}$$

This identity is a consequence of the corresponding identity for the neutral model, that is, the solution of (4.20) without the K_j integral terms. To see that the more general identity holds, first replace K_j by the counting measure obtained by setting

$$K_j^\varepsilon(du\times\{k\varepsilon\}) = K_j(du\times(k\varepsilon,(k+1)\varepsilon])I_{\{K_j(S\times(k\varepsilon,(k+1)\varepsilon])\le 1\}},$$

and let Y^ε denote the solution of the resulting equation. During time intervals of the form $[k\varepsilon,(k+1)\varepsilon)$, the system evolves as the neutral particle system, while at time $k\varepsilon$, the $Y_j^\varepsilon(k\varepsilon)$ are conditionally independent given $\mathscr{G}_{k\varepsilon} = \sigma(V(s), Y^\varepsilon(s) : s < k\varepsilon)$ with

$$P\{Y_1^\varepsilon(k\varepsilon)\in C_1,\dots,Y_m^\varepsilon(k\varepsilon)\in C_m \mid \mathscr{G}_{k\varepsilon}\} = \prod_{i=1}^m Q(Y_i^\varepsilon(k\varepsilon-), Z^V(k\varepsilon-), C_i),$$

and (4.21), with Y replaced by Y^ε, follows by the corresponding result for the neutral particle system and this conditional independence. Letting $\varepsilon\to\infty$, (4.21) follows and, as a consequence, we have

$$E[f(X_1^{n+1}(t),\dots,X_m^{n+1}(t)) \mid \mathscr{F}_t^{\gamma(X^{n+1}),\gamma(X^n)}] = \langle f, \gamma(X^{n+1}(t))^m\rangle. \tag{4.22}$$

Letting $n\to\infty$, we obtain part (b).

If $X(0)$ is not exchangeable, let Y denote the solution of (4.20) with $V = X$ and $Y_j(0) = X_1(0)$ for all j. Then

$$E[f(Y_1(t),\dots,Y_m(t)) \mid \mathscr{F}_t^{\gamma(Y),\gamma(X)}] = \langle f, \gamma(Y(t))^m\rangle. \tag{4.23}$$

Letting $n\to\infty$ in (4.18), we have that $\tilde X_j^T(t) \equiv X_{N_j^T(t)}(t)$ satisfies

$$\begin{aligned}\tilde X_j^T(t) &= X_{N_j^T(0)}(0) + \int_{U\times[0,t]} h(\tilde X_j^T(s-),u)\tilde M_j(du\times ds)\\ &\quad + \int_{S\times[0,t]} (q(\tilde X_j^T(s-), X(s-), \xi(s-), u) - \tilde X_j^T(s-))\tilde K_j(du\times ds),\end{aligned} \tag{4.24}$$

and similarly defining $\tilde Y_j^T(t) \equiv Y_{N_j^T(t)}(t)$, we have

$$\begin{aligned}\tilde Y_j^T(t) &= Y_{N_j^T(0)}(0) + \int_{U\times[0,t]} h(\tilde Y_j^T(s-),u)\tilde M_j(du\times ds)\\ &\quad + \int_{S\times[0,t]} (q(\tilde Y_j^T(s-), X(s-), \xi(s-), u) - \tilde Y_j^T(s-))\tilde K_j(du\times ds).\end{aligned} \tag{4.25}$$

Consequently, setting $C_t = \{N_i^t(0) = 1, \text{ all } i\}$ (that is, the event that the neutral population at time t has a unique common ancestor, necessarily at the bottom level, at time 0), we have $X(t) = Y(t)$ on C_t. Then

$$E\big[I_{C_t} h(X_1(t), \dots, X_m(t)) \mid \mathscr{F}_t^{\gamma(Y),\gamma(X)}\big] = E\big[I_{C_t} h(Y_1(t), \dots, Y_m(t)) \mid \mathscr{F}_t^{\gamma(Y),\gamma(X)}\big],$$

and since $\lim_{t\to\infty} P(C_t) = 1$, part (c) follows. □

4.2. *Corresponding martingale problem.* To be able to apply certain general theorems regarding martingale problems without making continuity assumptions on q [and the selection and recombination mechanisms in (3.10)], we need to show that the generator has a particular representation. For simplicity, we assume that $B: \bar{C}(E) \to \bar{C}(E)$, although that assumption could be relaxed as well. We consider the process with generator (4.10).

Let $F = \mathscr{P}(E)^\infty$ and define $\Psi: x \in E^\infty \to (\psi_1, \psi_2, \dots) \in F$ by

$$\psi_j(x, dz) = Q(x_j, \gamma(x), dz).$$

Let $\mathscr{D}_0$ be a countable subset of $\bar{C}(E)$ that is closed under multiplication and is dense in $\bar{C}(E)$ in the sense of convergence of bounded sequences, uniformly on compact subsets. Let $\mathscr{D}(A^0) = \{\prod_{i=1}^m f_i(x_i) : f_i \in \mathscr{D}_0, m = 1, 2, \dots\}$. Define $A^0 : \mathscr{D}(A^0) \subset \bar{C}(E^\infty) \to \bar{C}(E^\infty \times F)$ by

$$A^0 f(x, \nu) = \sum_{i=1}^m B_i f(x) + \sum_{1 \le i < j \le m} (f(\theta_j(x \mid x_i)) - f(x)) + \sum_{j=1}^m \beta \int_E (f(\eta_j(x \mid z)) - f(x)) \nu_j(dz) \tag{4.26}$$

and observe that for $\mu = \gamma(x)$,

$$\begin{aligned} Af(x, \mu) &= \sum_{i=1}^m B_i f(x) + \sum_{1 \le i < j \le m} (f(\theta_j(x \mid x_i)) - f(x)) \\ &\quad + \sum_{j=1}^m \beta \int_E (f(\eta_j(x \mid z)) - f(x)) Q(x_j, \mu, dz) \\ &= A^0 f(x, \Psi(x)). \end{aligned}$$

It follows that A satisfies the conditions of Theorem 2.7 of Kurtz (1998a), and the results of that paper will be used in several places in the current paper.

Note that under the assumption that B is a bounded operator, any solution of the martingale problem for A with $\mathscr{D}(A) = \mathscr{D}(A^0)$ is a solution of the martingale problem for A with $\mathscr{D}(A) = \bigcup_m B(E^m)$. [First extend A to the linear span of $\mathscr{D}(A^0)$ and then close the operator under bounded pointwise convergence. See Ethier and Kurtz (1986), Proposition 4.3.1.] As in (3.11), for

$f \in B(E^m)$, define $F(\mu) = \langle f, \mu^m \rangle$ and

$$\mathbb{A}F(\mu) = \langle Af(\cdot, \mu), \mu^m \rangle. \tag{4.27}$$

We will also, on occasion, need to incorporate information about the driving processes $\{K_j\}$, $\{L_{ij}\}$ and $\{M_j\}$ into the state description of the process. Let $\mathscr{G}_U \subset \mathscr{B}(U)$ $[\mathscr{G}_S \subset \mathscr{B}(S)]$ be countable and be an algebra (that is, closed under finite unions and complements) that generates the σ algebra $\mathscr{B}(U)$ $[\mathscr{B}(S)]$. For $C \in \mathscr{G}_U$, let $\lambda^U_{C,j}$ have values in $\{-1, 1\}$ and satisfy

$$\lambda^U_{C,j}(t) = \lambda^U_{C,j}(0)(-1)^{M_j(C\times[0,t])}.$$

For $C \in \mathscr{G}_S$, define $\lambda^S_{C,j}$ similarly and

$$\lambda^L_{ij}(t) = \lambda^L_{ij}(0)(-1)^{L_{ij}(t)}.$$

Note, for example, that

$$M_j(C \times [0, t]) = -\tfrac{1}{2}\int_0^t \lambda^U_{C,j}(s-)\, d\lambda^U_{C,j}(s), \tag{4.28}$$

so M_j can be recovered from $\{\lambda^U_{C,j} : C \in \mathscr{G}_U\}$. The generator A_λ for the process $(X, \lambda^L, \lambda^U, \lambda^S)$ can be derived using Itô's formula. If $(X, \lambda^L, \lambda^U, \lambda^S)$ is a solution of the martingale problem for A_λ, and if L_{ij}, M_j and K_j are defined in terms of λ^L, λ^U and λ^S as in (4.28), then X will satisfy (4.9).

If we set $\mathscr{I} = \{(i, j) : 1 \le i < j\}$ and $E_\lambda = \{-1, 1\}^{\mathscr{I}} \times \{-1, 1\}^{\mathscr{G}_U} \times \{-1, 1\}^{\mathscr{G}_S}$, then the state space for $(X, \lambda^L, \lambda^U, \lambda^S)$ is $E^\infty \times E_\lambda$. Let $\Pi_\lambda \in \mathscr{P}(E_\lambda)$ be the infinite product measures with factors $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$. [Note that Π_λ will be a stationary distribution for $(\lambda^L, \lambda^U, \lambda^S)$.] Then for $f \in B(E^m \times E_\lambda)$, considered as a function in $B(E^\infty \times E_\lambda)$, define $\Pi_\lambda f(x_1, \ldots, x_m) = \int_{E_\lambda} f(x_1, \ldots, x_m, v)\Pi_\lambda(dv)$ and note that

$$\Pi_\lambda A_\lambda f(x_1, \ldots, x_m) = A\Pi_\lambda f(x_1, \ldots, x_m, \gamma(x)). \tag{4.29}$$

As a consequence of this identity, we have the following lemma.

LEMMA 4.2. *If X is a solution of the martingale problem for A, then there exists a solution $(\tilde{X}, \lambda^L, \lambda^U, \lambda^S)$ of the martingale problem for A_λ such that X and $\tilde{X}$ have the same distribution. In particular, any solution of the martingale problem for A is a weak solution of* (4.9).

REMARK 4.3. A process X is a *weak solution* of (4.9) if there exists a probability space on which are defined Poisson random measures $\{\hat{M}_j\}$, $\{\hat{L}_{ij}\}$ and $\{\hat{K}_j\}$ with the same joint distributions as $\{M_j\}$, $\{L_{ij}\}$ and $\{K_j\}$ and a process

$\hat{X}$ with the same distribution as X such that

$$\begin{aligned}\hat{X}_j(t) &= \hat{X}_j(0) + \int_{U\times[0,t]} h(\hat{X}_j(s-),u)\hat{M}_j(du\times ds)\\ &\quad + \sum_{i=1}^{j-1}\int_0^t (\hat{X}_i(s-)-\hat{X}_j(s-))\,d\hat{L}_{ij}(s)\\ &\quad + \sum_{1\le i<k<j}\int_0^t (\hat{X}_{j-1}(s-)-\hat{X}_j(s-))\,d\hat{L}_{ik}(s)\\ &\quad + \int_{S\times[0,t]} (q(\hat{X}_j(s-),\hat{Z}(s-),u)-\hat{X}_j(s-))\hat{K}_j(du\times ds),\end{aligned} \tag{4.30}$$

where $\hat{Z}(t) = \gamma(\hat{X}(t))$.

PROOF. The existence of the solution of the martingale problem for A_λ follows from Corollary 3.5 and Theorem 2.7 of Kurtz (1998a). Then, as observed above, any solution of the martingale problem for A_λ satisfies (4.9). See Kurtz (1998b) for further discussion of results of this type. □

We are primarily interested in solutions X of the martingale problem for A for which $(X_1(t), X_2(t), \dots)$ is exchangeable. In particular, exchangeability implies

$$\gamma(X(t)) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i(t)} \quad \text{a.s.}$$

To be precise, for $\nu_0 \in \mathscr{P}_e(E^\infty)$, the collection of exchangeable probability distributions on E^∞, we will say that X is a solution of the *exchangeable martingale problem* for (A, ν_0) if there exists a filtration $\{\mathscr{F}_t\}$ such that for each $f \in \mathscr{D}(A)$,

$$f(X(t)) - \int_0^t Af(X(s))\,ds$$

is an $\{\mathscr{F}_t\}$ martingale and for each $t \ge 0$, $X(t)$ is exchangeable.

Let $\mathscr{H}$ be the collection of functions of the form $f(x_{\kappa_1^1}, \dots, x_{\kappa_m^1}) - f(x_{\kappa_1^2}, \dots, x_{\kappa_m^2})$, where $f \in B(E^m)$ for some $m \ge 1$ and for $i = 1, 2$, $(\kappa_1^i, \dots, \kappa_m^i)$ is a permutation of $(1, \dots, m)$. Let $\mathscr{P}_{\mathscr{H}}(E^\infty) \subset \mathscr{P}(E^\infty)$ be the collection of probability measures $\nu_0 \in \mathscr{P}(E^\infty)$ satisfying $\int_{E^\infty} h\,d\nu_0 = 0$ for each $h \in \mathscr{H}$ and note that $\mathscr{P}_{\mathscr{H}}(E^\infty) = \mathscr{P}_e(E^\infty)$. It follows that the exchangeable martingale problem for (A, ν_0) is just the *restricted martingale problem* for $(A, \mathscr{H}, \nu_0)$ in the sense of Kurtz [(1998a), Section 3].

We assume that F in (4.11) is a linear space and that the process ξ in (4.12) is the unique solution of a stochastic differential equation driven by $\{M_j\}$, $\{L_{ij}\}$ and (possibly) additional independent Poisson random measures

$\{\Xi_\alpha\}$, but that ξ is independent of $\{K_j\}$. We write

$$\begin{aligned}\xi(t) = \xi(0) &+ \sum_j \int_{U\times[0,t]} F_j(\xi(s-),u)M_j(du\times ds)\\ &+ \sum_{ij}\int_0^t G_{ij}(\xi(s-))\,dL_{ij}(s)\\ &+ \sum_\alpha \int_{U_\alpha\times[0,t]} H_\alpha(\xi(s-),u)\Xi_\alpha(du\times ds).\end{aligned}\tag{4.31}$$

[See (6.2) for a particular example.] If X is a solution of (4.9), then, at least formally, (X,ξ) is a solution of the martingale problem for an operator $A^1 = A^0 + B^1$, where

$$B^1 f(x,v) = \sum_j \beta \int_S \big(f(\eta_j(x \mid q(x_j,\gamma(x),u)),v) - f(x,v)\big)\zeta(du),\tag{4.32}$$

and if X is a solution of (4.12), then (X,ξ) is a solution of the martingale problem for $A^2 = A^0 + B^2$, where

$$B^2 f(x,v) = \sum_j \beta \int_S \big(f(\eta_j(x \mid q(x_j,x,v,u)),v) - f(x,v)\big)\zeta(du).\tag{4.33}$$

We will assume that solutions of the stochastic equations are, in fact, solutions of the corresponding martingale problems (an assertion that is usually easy to check using Itô's formula) and, in addition, we will assume that any solution of the martingale problem for A^1 is a weak solution of the system (4.9) and (4.31) and that any solution for A^2 is a weak solution of the system (4.12) and (4.31) (assertions that can usually be proved by the argument used in the proof of Lemma 4.2).

THEOREM 4.4. *Let A be given by* (4.10) *and let q satisfy* (4.11) *and* (4.14). *Let $\Gamma \subset E^\infty \times F$ be the set of (x,v) such that*

$$Q(x_j,\gamma(x),C) = \frac{1}{\beta}\int_S I_{\{q(x_j,x,v,u)\in C\}}\zeta(du), \quad C \in \mathscr{B}(E),$$

for all $j = 1,2,\dots$ and $t \ge 0$. Let $A^1 = A^0 + B^1$ and $A^2 = A^0 + B^2$ be defined on the same domain with $B^1$and B^2 given by (4.32) *and* (4.33), *respectively, and assume that (X,ξ) is a solution of the martingale problem for A^1 if and only if it is a weak solution of the system* (4.9) *and* (4.31) *and that (X,ξ) is a solution of the martingale problem for A^2 if and only if it is a weak solution of the system* (4.12) *and* (4.31). *Suppose that every process (X,ξ) that solves the system* (4.9) *and* (4.31) *or the system* (4.12) *and* (4.31) *satisfies $(X(t),\xi(t)) \in \Gamma$ a.s., $t > 0$. Then*:

(a) *For each $\nu_0 \in \mathscr{P}(E^\infty)$ there exists a unique solution of the martingale problem for (A,ν_0).*

(b) *If* $\nu_0 \in \mathscr{P}_e(E^\infty)$ *and* X *is a solution of the martingale problem for* (A, ν_0), *then* X *is a solution of the exchangeable martingale problem for* (A, ν_0) *and* $Z = \gamma(X)$ *is a solution of the martingale problem for* $\mathbb{A}$ *given by* (4.27).

(c) *If* Z *is a solution of the martingale problem for* $\mathbb{A}$ *and, for* $f \in B(E^m)$, $\langle f, \nu_0^m \rangle = E[\langle f, Z^m(0)\rangle]$, *then there exists a solution* X *of the exchangeable martingale problem for* (A, ν_0) *such that the distribution of* $\gamma(X)$ *is the same as the distribution of* Z.

REMARK 4.5. In Section 6, we will apply this theorem to obtain existence and uniqueness for A given in (3.10). More generally, the methods of Section 6 can be applied to (4.10) with

$$Q(y, \mu, C) = \int_{E^m} h(y, x_1, \dots, x_m, C)\mu(dx_1)\cdots\mu(dx_m).$$

In particular, models of frequency-dependent selection [see, for example, Neuhauser (1998)] with selection parameters of the form

$$\sigma(x, \mu) = \int_{E^m} s(x, x_1, \dots, x_m)\mu(dx_1)\cdots\mu(dx_m)$$

are covered.

PROOF OF THEOREM 4.4. Existence of solutions of the martingale problem for A is immediate since, by Itô's formula, any solution of (4.12) is a solution of the martingale problem. Uniqueness for the martingale problem will follow from uniqueness for the system provided every solution of the martingale problem is a weak solution of the system. It follows from Lemma 4.2 that every solution X of the martingale problem for A is a weak solution of (4.9) and hence can be coupled with a solution ξ of (4.31) to give a solution of the martingale problem for A^1. However, $A^1 f(x, v) = A^2 f(x, v)$ for $(x, v) \in \Gamma$, so by the assumption $(X(t), \xi(t)) \in \Gamma$, $t > 0$, (X, ξ) is a solution of the martingale problem for A^2 and hence is a weak solution of the system (4.12)–(4.31). Uniqueness for (4.12) then gives uniqueness for the martingale problem.

If $\nu_0 \in \mathscr{P}_e(E^\infty)$ and X is a solution of the martingale problem for (A, ν_0), then by part (b) of Theorem 4.1, X is a solution of the exchangeable martingale problem of (A, ν_0). Let $Z = \gamma(X)$. Then for $f \in \mathscr{D}(A)$, it follows that

$$E\big[f(X_1(t), \dots, X_m(t)) \mid \mathscr{F}_t^Z\big] - \int_0^t E\big[Af(X_1(s), \dots, X_m(s), Z(s)) \mid \mathscr{F}_s^Z\big]\, ds \tag{4.34}$$

is an $\{\mathscr{F}_t^Z\}$ martingale and by part (b) of Theorem 4.1, (4.34) is

$$\langle f, Z(t)^m \rangle - \int_0^t \langle Af(\cdot, Z(s)), Z(s)^m \rangle\, ds,$$

which gives part (b).

To prove part (c), apply Corollary 3.7 of Kurtz (1998a) with E in the corollary corresponding to E^∞ in the present paper, Y to the present Z, Z to the present

X, $E_0 = \mathscr{P}(E)$, $C = \mathbb{A}$ and $\alpha(\mu, dx)$, the transition function from $\mathscr{P}(E)$ to E^∞ determined by

$$\int_{E^\infty} f(x)\alpha(\mu, dx) = \langle f, \mu^m \rangle$$

for $f \in B(E^m)$ considered as a subset of $B(E^\infty)$. If Z is a solution of the martingale problem for $\mathbb{A}$, then the corollary gives existence of the desired X. □

4.3. *Systems for viability selection.* In this section we have concentrated on systems with fecundity selection. Similar systems can be developed for models with viability selection. The model given by (3.1) is particularly appealing in the case of genic selection. Taking $\alpha = 0$ for simplicity, the corresponding system is

$$\begin{aligned} X_j(t) = X_j(0) &+ \int_{U\times[0,t]} h(X_j(s-), u) M_j(du \times ds) \\ &+ \sum_{i=1}^{j-1} \int_0^t (X_i(s-) - X_j(s-))\, dL_{ij}(s) \\ &+ \sum_{1\le i<k<j} \int_0^t (X_{j-1}(s-) - X_j(s-))\, dL_{ik}(s) \\ &+ \sum_{k=1}^{j} \int_{[0,1]\times[0,t]} (X_{j+1}(s-) - X_j(s-)) I_{[0, \bar\beta^{-1}\beta(X_k(s-))]}(u) \\ &\times K_k(du \times ds), \end{aligned} \tag{4.35}$$

where $\bar\beta$ is a constant satisfying $0 \le \beta \le \bar\beta$ and the K_k have mean measure $\bar\beta m^2$.

5. Stationarity. Let $\mathscr{P}_e(E^\infty) \subset \mathscr{P}(E^\infty)$ be the collection of exchangeable distributions.

LEMMA 5.1. *Under the conditions of Theorem* 4.4, *if* Π *is a stationary distribution for* A, *then* $\Pi \in \mathscr{P}_e(E^\infty)$.

PROOF. Under the conditions of Theorem 4.4, every solution of the martingale problem for A is a weak solution of (4.12). Consequently, the lemma follows from part (c) of Theorem 4.1. □

Note that there is a one-to-one correspondence between $\Pi \in \mathscr{P}_e(E^\infty)$ and $\tilde\Pi \in \mathscr{P}(\mathscr{P}(E))$, where we take $\tilde\Pi$ to be the distribution of the de Finetti measure corresponding to Π. By the definition of $\tilde\Pi$ and $\mathbb{A}$, for every $f \in B(E^m)$ and $F(\mu) = \langle f, \mu^m \rangle$,

$$\int_{E^\infty} Af\, d\Pi = \int_{\mathscr{P}(E)} \langle Af, \mu^{m+2} \rangle \tilde\Pi(d\mu) = \int_{\mathscr{P}(E)} \mathbb{A}F(\mu)\tilde\Pi(d\mu).$$

THEOREM 5.2. *Suppose that* $\Pi \in \mathscr{P}_e(E^\infty)$ *and that* $\tilde{\Pi} \in \mathscr{P}(\mathscr{P}(E))$ *is the distribution of the corresponding de Finetti measure. If for* $f \in B(E^m)$ *and* $F(\mu) = \langle f, \mu^m \rangle$,

$$0 = \int_{E^\infty} Af \, d\Pi = \int_{\mathscr{P}(E)} \langle Af, \mu^{m+2} \rangle \tilde{\Pi}(d\mu) = \int_{\mathscr{P}(E)} \mathbb{A}F(\mu)\tilde{\Pi}(d\mu) \tag{5.1}$$

for A *given by* (4.10), *then* Π *is a stationary distribution for* A *and* $\tilde{\Pi}$ *is a stationary distribution for* $\mathbb{A}$.

REMARK 5.3. Since any measure in $\mathscr{P}(\mathscr{P}(E))$ corresponds to an exchangeable distribution in $\mathscr{P}(E^\infty)$, it follows that there is a one-to-one correspondence between exchangeable stationary distributions for A and stationary distributions for the Fleming–Viot process generated by $\mathbb{A}$. If A satisfies the conditions of Theorem 4.4, then by Lemma 5.1, every stationary distribution for A is exchangeable, so there is a one-to-one correspondence between stationary distributions for A and stationary distributions for $\mathbb{A}$. In particular, if $\Pi_Z \in \mathscr{P}(\mathscr{P}(E))$ is the stationary distribution for the Fleming–Viot process, Z, then the corresponding stationary distribution $\Pi_X \in \mathscr{P}(E^\infty)$ for X is characterized by

$$\int_{E^m} f(x_1, \dots, x_m)\Pi_X^m(dx_1 \times \cdots \times dx_m) = \int_{\mathscr{P}(E)} \langle f, \mu^m \rangle \Pi_Z(d\mu),$$

where $\Pi_X^m \in \mathscr{P}(E^m)$ is the marginal on E^m of Π_X.

PROOF OF THEOREM 5.2. Using the representation of A given in (4.26), the theorem follows from Theorem 3.1 of Kurtz and Stockbridge (1996) for E locally compact and from Theorem 2.7 of Kurtz (1998a) for general complete, separable E. (The usual form of Echeverria's theorem need not apply since we do not want to assume that σ is continuous. For example, in a simple model of heterozygote advantage, $\sigma(x, y) = \sigma_0 I_{\{x \neq y\}}$ for some $\sigma_0 > 0$.) □

Suppose that a stationary distribution Π_X exists. (We will discuss existence in Section 9.) Then with Π_λ as in Section 4.2, it follows from (4.29) that $\Pi_X \times \Pi_\lambda$ will be a stationary distribution for A_λ. [Theorem 2.7 of Kurtz (1998a) again applies.] Let $(X, \lambda^L, \lambda^U, \lambda^S)$ be a stationary solution with marginal distribution $\Pi_X \times \Pi_\lambda$. We can assume that the process is defined on the doubly infinite time interval $(-\infty, \infty)$. If, as in (4.28), we define

$$M_j(C \times (r, t]) = -\tfrac{1}{2} \int_r^t \lambda_{C,j}^U(s-) \, d\lambda_{C,j}^U(s),$$

$$K_j(C \times (r, t]) = -\tfrac{1}{2} \int_r^t \lambda_{C,j}^S(s-) \, d\lambda_{C,j}^S(s)$$

and

$$L_{ij}((r, t]) = -\tfrac{1}{2} \int_r^t \lambda_{ij}^L(s-) \, d\lambda_{ij}^L(s),$$

then X will satisfy

$$
\begin{aligned}
X_j(t) = X_j(r) &+ \int_{U\times(r,t]} h(X_j(s-),u)M_j(du\times ds) \\
&+ \sum_{i=1}^{j-1}\int_{(r,t]}(X_i(s-)-X_j(s-))L_{ij}(ds) \\
&+ \sum_{1\le i<k<j}\int_{(r,t]}(X_{j-1}(s-)-X_j(s-))L_{ik}(ds) \\
&+ \int_{S\times(r,t]}(q(X_j(s-),Z(s-),u)-X_j(s-))K_j(du\times ds)
\end{aligned}
\tag{5.2}
$$

for $-\infty < r < t < \infty$.

6. Incorporating genealogy. It is not clear how to define $\rho(\mu, u)$ so that (4.7) satisfies the conditions of Theorem 4.1 and hence has a unique solution. Even if it were, it would not be possible to recover the genealogy from the solution of this equation, since, in particular, the ancestry of particles substituted by the selection term is not identifiable. We need to be able to generate samples from the empirical measure Z in a way that satisfies the conditions of Theorem 4.1 and that allows us to identify the ancestors of the individual sampled. We accomplish this goal by introducing a family of "genetic markers" which allow us to trace the ancestry of any observed particle.

6.1. *The marker process.* Each particle will be assigned a marker in the space $E_0 = [0,1]^\infty$. The mutation process for the markers (the *marker process* to distinguish it from the original mutation process) will have generator

$$
B^M f(x) = \sum_{k=1}^{\infty}\int_0^1 (f(\eta_k(x\mid z)) - f(x))\,dz,
\tag{6.1}
$$

that is, each component of the marker process evolves independently as a unit rate jump process such that at each jump time, the new value is uniformly distributed on $[0,1]$ independently of the previous value. We define a neutral particle process in E_0^∞ with mutation operator B^M as a solution of

$$
\begin{aligned}
\xi_j(t) = \xi_j(0) &+ \sum_l \int_{[0,1]\times[0,t]}(u-\xi_{jl}(s-))e_l\Xi_{jl}(du\times ds) \\
&+ \sum_{i=1}^{j-1}(\xi_i(s-)-\xi_j(s-))\,dL_{ij}(s) \\
&+ \sum_{1\le i<k<j}(\xi_{j-1}(s-)-\xi_j(s-))\,dL_{ik}(s),
\end{aligned}
\tag{6.2}
$$

where the Ξ_{jl} are independent Poisson random measures on $[0,1]\times[0,\infty)$ with Lebesgue mean measure and e_l is the element in E_0 whose lth component is 1 and all other components are 0.

The system of equations (4.7) for the particle model with selection and recombination becomes

$$
\begin{aligned}
X_j(t) = X_j(0) &+ \int_{U\times[0,t]} h(X_j(s-), u) M_j(du \times ds) \\
&+ \sum_{i=1}^{j-1} (X_i(s-) - X_j(s-))\, dL_{ij}(s) \\
&+ \sum_{1\le i<k<j} (X_{j-1}(s-) - X_j(s-))\, dL_{ik}(s) \\
&+ \int_{[0,1]^3\times[0,t]} (\rho_X(s-, u_1) - X_j(s-)) \\
&\qquad \times I_{[0,\bar{\sigma}^{-1}\sigma(\rho_X(s-,u_1),\rho_X(s-,u_2))]}(u_3) \\
&\qquad \times K_j(du_1 \times du_2 \times du_3 \times ds) \\
&+ \int_{[0,1]^2\times[0,t]} (r(X_j(s-), \rho_X(s-,u_1), u_2) - X_j(s-)) \\
&\qquad \times J_j(du_1 \times du_2 \times ds),
\end{aligned}
\tag{6.3}
$$

where $\rho_X(s-, u)$ will be defined to be a function of $\xi(s-)$ and $X(s-)$. In particular, we must define ρ_X so that if V is uniformly distributed on [0, 1], then the conditional distribution of $\rho_X(s-, V)$ given $\mathscr{F}_s^X$ is $Z(s-)$ [$= Z(s)$ since Z is continuous].

To define ρ_X, we linearly order E_0 lexicographically, that is, for $y, \tilde{y} \in E_0$, $y < \tilde{y}$ if there exists $k \ge 1$ such that $y_i = \tilde{y}_i$ for $i < k$ and $y_k < \tilde{y}_k$. With this ordering, for $0 \le p < 1$, we can define the pth percentile of $(\xi_1(t), \dots, \xi_n(t))$ to be the point $\xi_{k_n(p,t)}(t)$ such that $k_n(p, t) \le n$ and

$$\#\{k \le n : \xi_k(t) \le \xi_{k_n(p,t)}\} \ge [np] + 1$$

and

$$\#\{k \le n : \xi_k(t) \ge \xi_{k_n(p,t)}\} \ge n - [np].$$

Since infinitely many components of ξ_i change during any postive time interval and the changes are independent on different levels, with probability 1, $\xi_i(t) \ne \xi_j(t)$ for $i \ne j$, and hence $k_n(p, t)$ is uniquely determined and $\{p : k_n(p, t) = k\}$ is an interval of length n^{-1}. Consequently, if V is uniformly distributed on [0, 1] and independent of the σ algebra $\sigma(J, K, L, M, \Xi)$, then

$$P\{k_n(V, t) = k \mid \sigma(J, K, L, M, \Xi)\} = \frac{1}{n}, \qquad k = 1, \dots, n,$$

so that $k_n(V, t)$ is independent of $\sigma(J, K, L, M, \Xi)$.

LEMMA 6.1. *Let V be independent of ξ and uniformly distributed on* [0, 1]. *Then $\xi_\infty(V, t) \equiv \lim_{n\to\infty} \xi_{k_n(V,t)}(t)$ exists a.s. in the sense that, with probability* 1, *for each l there exists $n_l(V, t)$ such that $n \ge n_l(V, t)$ implies*

$$\xi_{k_n(V,t),l} = \xi_{\infty,l}(V, t).$$

In addition, with N_j^t *given by* (4.4),

$$N^t(V, s) \equiv \lim_{n\to\infty} N^t_{k_n(V,t)}(s) \tag{6.4}$$

exists a.s. for every $s < t$ *and determines a path back through the neutral genealogy, "starting from infinity";*

$$\rho_X(V, t) \equiv \lim_{s\to t-} X_{N^t(V,s)}(s) \tag{6.5}$$

exists a.s. and

$$\lim_{n\to\infty} P\{X_{k_n(V,t)}(t) = \rho_X(V, t)\} = 1. \tag{6.6}$$

REMARK 6.2. (a) By (6.4), $N^t(V, \cdot)$ gives a randomly selected path starting at infinity, back through the neutral genealogy to time zero. The distribution of $N^t(V, \cdot)$ is uniform over all such paths in the sense that it is the limit of random paths selected uniformly over the paths back from the first n levels at time t.

(b) Since $k_n(V, t)$ is independent of $\mathscr{F}_t^X$ and is uniformly distributed over $\{1, \ldots, n\}$, (6.6) justifies describing $\rho_X(V, t)$ as a sample from the distribution $Z(t)$. If V_1 and V_2 are independent and uniformly distributed over $[0, 1]$, then $k_n(V_1, t)$ and $k_n(V_2, t)$ are independent and uniformly distributed over $\{1, \ldots, n\}$. It then follows that

$$\begin{aligned} E[f(\rho_X(V_1, t), \rho_X(V_2, t)) \mid \mathscr{F}_t^X] &= \lim_{n\to\infty} E[f(X_{k_n(V_1,t)}(t), X_{k_n(V_2,t)}(t)) \mid \mathscr{F}_t^X] \\ &= \lim_{n\to\infty} \frac{1}{n^2} \sum_{1\le i,\, j\le n} f(X_i(t), X_j(t)) \\ &= \langle f, Z(t)^2\rangle, \end{aligned}$$

that is, we can generate independent samples from $Z(t)$ by using independent uniform random variables.

(c) Let

$$\begin{aligned} &q_1(X_j(s-), X(s-), \xi(s-), u_1, u_2, u_3) \\ &\quad = X_j(s-) + (\rho_X(s-, u_1) - X_j(s-)) I_{[0,\, \bar{\sigma}^{-1}\sigma(\rho_X(s-,u_1),\, \rho_X(s-,u_2))]}(u_3) \end{aligned}$$

and

$$q_2(X_j(s-), X(s-), \xi(s-), u_1, u_2) = r(X_j(s-), \rho_X(s-, u_1), u_2),$$

and set $q = q_1 I_{[0,1]^3} + q_2 I_{[0,1]}$. Using (6.6),

$$\begin{aligned}
&\int_{[0,1]^3} I_{\{q_1(X_j(s-), X(s-), \xi(s-), u_1, u_2, u_3)\in C\}}\, du_1\, du_2\, du_3 \\
&\quad = \int_{[0,1]^2} \biggl(\frac{\bar\sigma - \sigma(\rho_X(s-, u_1), \rho_X(s-, u_2))}{\bar\sigma} I_{\{X_j(s-)\in C\}} \\
&\qquad\qquad + \frac{\sigma(\rho_X(s-, u_1), \rho_X(s-, u_2))}{\bar\sigma} I_{\{\rho_X(s-, u_1)\in C\}} \biggr)\, du_1\, du_2 \\
&\quad = \int_{E^2} \biggl(\frac{\bar\sigma - \sigma(x, y)}{\bar\sigma} I_{\{X_j(s-)\in C\}} + \frac{\sigma(x, y)}{\bar\sigma} I_{\{x\in C\}} \biggr) Z(s-, dx) Z(s-, dy)
\end{aligned}$$

and

$$\begin{aligned}
&\int_{[0,1]^2} I_{\{q_2(X_j(s-), X(s-), \xi(s-), u_1, u_2)\in C\}}\, du_1\, du_2 \\
&\quad = \int_{[0,1]^2} I_{\{r(X_j(s-), \rho_X(s-, u_1), u_2)\in C\}}\, du_1\, du_2 \\
&\quad = \int_{E\times[0,1]} I_{\{r(X_j(s-), x, u_2)\in C\}} Z(s-, dx)\, du_2 \\
&\quad = \int_E R(X_j(s-), x, C) Z(s-, dx),
\end{aligned}$$

so (4.13) holds, and any solution of (6.3) is a solution of the martingale problem for (4.6).

PROOF OF LEMMA 6.1. If we restrict the marker process to its first m components, then the corresponding generator

$$\sum_{i=1}^{m} \int_0^1 (f(\eta_i(x \mid z)) - f(x))\, dz$$

is bounded and hence

$$Z_\xi^m(t) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{(\xi_{i1}(t), \dots, \xi_{im}(t))}$$

is purely atomic. [See Ethier and Kurtz (1993), Theorem 7.2 or Donnelly and Kurtz (1996), Theorem 5.2.] Since $\xi_1, \xi_2, \dots$ are conditionally iid given $Z_\xi(t)$, it follows that with probability 1, for n sufficiently large, $(\xi_{k_n(V,t),1}, \dots, \xi_{k_n(V,t),m})$ will fix on the location of one of the atoms of $Z_\xi^m(t)$. Since m is arbitrary, it follows that $\xi(V, t)$ exists a.s.

For each l, the lth component of $\xi(V, t)$, $\xi_l(V, t)$, is the value of a mutation that took place on some level $k_l(V)$ at some time $s_l(V)$ or was a value present in the population at time zero. If $\xi_{k_n(V,t),l}(t) = \xi_l(V, t)$ and $s_l(V) > 0$, it follows that $N_{k_n(V,t)}^t(s_l) = k_l(V)$ and that $N_{k_n(V,t)}^t(s) = N_{k_l(V)}^{s_l(V)}(s)$ for all $0 \le s < s_l(V)$. To prove (6.4), we need to show that

$$\inf_l (t - s_l(V)) = 0 \quad \text{a.s.}$$

If we let $s_l^n(V)$ be the time of the mutation producing $\xi_{k_n(V,t),l}(t)$, the fact that $k_n(V,t)$ is independent of L and Ξ implies that $t-s_1^n(V), t-s_2^n(V), \ldots$ are iid with distribution $P\{t-s_k^n(V)>r\}=e^{-r}I_{[0,t)}(r)$. Since $\lim_{n\to\infty} s_l^n(V)=s_l(V)$ a.s., it follows that $t-s_1(V), t-s_2(V), \ldots$ are independent with the same distribution, and hence $\inf_l(t-s_l(V))=0$ a.s. and (6.4) follows.

As in the proof of Theorem 4.1, define $\tilde{X}_v^T(s)=X_{N^T(v,s)}(s)$, $0\le s<T$, which is well defined for almost every v. For each $T>0$ and $j=1,2,\ldots$, define

$$\tilde{M}_v^T(C\times[0,t])=\sum_{i=1}^{\infty}\int_{C\times[0,t]} I_{\{N^T(v,s)=i\}}M_i(du\times ds)$$

and define $\tilde{K}_v^T$, $\tilde{J}_v^T$ and $\{\tilde{\Xi}_{vl}^T\}$ similarly. As before, since the M_i are independent of the L_{ij} and hence of N_v^T, $\tilde{M}_v^T$ will be a Poisson random measure on $U\times[0,T]$ with mean measure $\nu\times m$ and, similarly, $\tilde{K}_v^T$, $\tilde{J}_v^T$ and $\{\tilde{\Xi}_{vl}^T\}$ will be Poisson random measures with the same distributions as K_j, J_j and $\{\Xi_{jl}\}$, respectively (but not independent). Let

$$\tau_v^T=\sup\bigl\{s<T:\tilde{M}_v^T(U\times(s,T])+\tilde{K}_v^T([0,1]^3\times(s,T])+\tilde{J}_v^T([0,1]^2\times(s,T])>0\bigr\}$$

and note that $\tau_v^T<T$ a.s. and for $\tau_v^T\le s<T$, $\tilde{X}_v^T(s)=\tilde{X}_v^T(\tau_v^T)$, verifying the existence of the limit in (6.5).

To verify (6.6), let $\gamma_v^{T,n}=\sup\{s<T:N^T(v,s)=N^T_{k_n(v,T)}(s)\}$ and note that $\lim_{n\to\infty}\gamma_v^{T,n}=T$. Then

$$\begin{aligned}P\{X_{k_n(V,T)}(T)\ne\rho_X(V,t)\}&\le P\{\gamma_V^{T,n}<\tau_V^T\}\\&\quad+P\{X_{k_n(V,T)}(T)\ne X_{N^T_{k_n(V,T)}(\gamma_V^{T,n})}(\gamma_V^{T,n})\}\\&\le P\{\gamma_V^{T,n}<\tau_V^T\}+E[1-e^{-\beta(T-\gamma_V^{T,n})}],\end{aligned}$$

where $\beta=\nu(U)+\bar{\sigma}+\alpha$, and the limit follows. □

THEOREM 6.3. *For each initial condition $X(0)$, there exists a unique solution of* (6.3).

PROOF. We apply Theorem 4.1 with $S=[0,1]^3\times[0,1]^2$ and

$$\begin{aligned}q(x_j,x,v,u)&=I_{[0,1]^3}(u)I_{[0,\bar{\sigma}^{-1}\sigma(\rho_x(s,u_1),\rho_x(s,u_2))]}(u_3)\rho_x(s,u_1)\\&\quad+I_{[0,1]^2}(u)r(x_j,\rho_x(s,u_1),u_2),\end{aligned}$$

where $u=(u_1,u_2,u_3)$ if $u\in[0,1]^3$ and $u=(u_1,u_2)$ if $u\in[0,1]^2$. Noting that

$$I_{\{q(x_j,x,v,u)\ne q(y_j,y,v,u)\}}\le I_{\{x_j\ne y_j\}}+I_{\{\rho_x(s,u_1)\ne\rho_y(s,u_1)\}}+I_{\{\rho_x(s,u_2)\ne\rho_y(s,u_2)\}},$$

we see that (4.14) holds with

$$D = 2(\bar{\sigma} + \alpha).$$

Existence of a unique solution then follows from Theorem 4.1. □

COROLLARY 6.4. *Let A be given by* (3.10) *with* $0 \le \sigma \le \bar{\sigma}$. *Then for each* $\nu_0 \in \mathscr{P}(E^\infty)$, *there exists a unique solution of the martingale problem* (A, ν_0).

6.2. *The ancestral type processes.* Suppose X is a solution of (6.3) with ρ given by (6.5). For $j = 1, 2, \dots$ and $v \in [0,1]$, define $\tilde{X}_j^T(t) = X_{N_j^T(t)}(t)$ and $\tilde{X}_v^T(t) = X_{N^T(v,t)}(t)$ (which by Lemma 6.1 will be defined for almost every v). We want to derive the stochastic equations satisfied by $\tilde{X}_j^T$ and $\tilde{X}_v^T$. As in the proof of Lemma 6.1, for each $T > 0$ and $j = 1, 2, \dots$, define

$$\tilde{M}_j^T(C \times [0,t]) = \sum_{i=1}^\infty \int_{C\times[0,t]} I_{\{N_j^T(s)=i\}} M_i(du \times ds)$$

and define $\tilde{K}_j^T$, $\tilde{J}_j^T$ and $\{\tilde{\Xi}_{jl}^T\}$ similarly. Again by the independence of the M_i and the L_{ij}, $\tilde{M}_j^T$ will be a Poisson random measure on $U \times [0,T]$ with mean measure $\nu \times m$ and, similarly, $\tilde{K}_j^T$, $\tilde{J}_j^T$ and $\{\tilde{\Xi}_{jl}^T\}$ will be Poisson random measures with the same distributions as K_j, J_j and $\{\Xi_{jl}\}$, respectively. For $v \in [0,1]$, define

$$\tilde{M}_v^T(C \times [0,t]) = \sum_{i=1}^\infty \int_{C\times[0,t]} I_{\{N^T(v,s)=i\}} M_i(du \times ds),$$

as well as $\tilde{K}_v^T$, $\tilde{J}_v^T$ and $\{\tilde{\Xi}_{vl}^T\}$.

Note that

$$\rho_X(v,t) = \tilde{X}_v^t(t-),$$

so for $0 \le t \le T$,

$$\begin{aligned}\tilde{X}_j^T(t) = {} & X_{N_j^T(0)}(0) + \int_{U\times[0,t]} h\big(\tilde{X}_j^T(s-), u\big)\tilde{M}_j^T(du \times ds)\\ & + \int_{[0,1]^3\times[0,t]} \big(\tilde{X}_{u_1}^s(s-) - \tilde{X}_j^T(s-)\big)\\ & \qquad \times I_{[0,\bar{\sigma}^{-1}\sigma(\tilde{X}_{u_1}^s(s-),\, \tilde{X}_{u_2}^s(s-))]}(u_3)\\ & \qquad \times \tilde{K}_j^T(du_1 \times du_2 \times du_3 \times ds)\\ & + \int_{[0,1]^2\times[0,t]} (r(\tilde{X}_j^T(s-), \tilde{X}_{u_1}^s(s-), u_2) - \tilde{X}_j^T(s-))\\ & \qquad \times \tilde{J}_j^T(du_1 \times du_2 \times ds)\end{aligned} \tag{6.7}$$

and for $0 \le t < T$,

$$\begin{aligned}\tilde{X}_v^T(t) = X_{N^T(v,0)}(0) &+ \int_{U\times[0,t]} h\big(\tilde{X}_v^T(s-), u\big)\tilde{M}_v^T(du \times ds)\\ &+ \int_{[0,1]^3\times[0,t]} \big(\tilde{X}_{u_1}^s(s-) - \tilde{X}_v^T(s-)\big)\\ &\quad\times I_{[0,\bar{\sigma}^{-1}\sigma(\tilde{X}_{u_1}^s(s-), \tilde{X}_{u_2}^s(s-))]}(u_3)\\ &\quad\times \tilde{K}_v^T\big(du_1 \times du_2 \times du_3 \times ds\big)\\ &+ \int_{[0,1]^2\times[0,t]} \big(r(\tilde{X}_v^T(s-), \tilde{X}_{u_1}^s(s-), u_2) - \tilde{X}_v^T(s-)\big)\\ &\quad\times \tilde{J}_v^T(du_1 \times du_2 \times ds).\end{aligned} \tag{6.8}$$

Since

$$\tilde{X}_v^T(T) \equiv \lim_{t\to T-} \tilde{X}_v^T(t) = \rho_X(v, T) \tag{6.9}$$

exists, it follows that the equations are satisfied on the closed interval $[0, T]$.

If $\tilde{K}_v^T$ and $\tilde{J}_v^T$ (or $\tilde{K}_j^T$ and $\tilde{J}_j^T$) are both zero on the interval $[0, T]$, then $\tilde{X}_v^T$ is just a version of the mutation process on the time interval $[0, T]$; that is, it is the unique solution of the stochastic differential equation driven by $\tilde{M}_v^T$. Otherwise, to determine $\tilde{X}_v^T$, we must know the values of $2\tilde{K}_j^T([0,1]^3 \times [0,T]) + \tilde{J}_j^T([0,1]^2 \times [0,T])$ processes $\tilde{X}_u^s$ for values of (u, s) in the support of the counting measure

$$\begin{aligned}I_v^T = &\int_{[0,1]^3\times[0,T]} (\delta_{(u_1,s)} + \delta_{(u_2,s)})\tilde{K}_v^T(du_1 \times du_2 \times du_3 \times ds)\\ &+ \int_{[0,1]^2\times[0,T]} \delta_{(u_1,s)}\tilde{J}_v^T(du_1 \times du_2 \times ds),\end{aligned} \tag{6.10}$$

where $\delta_{(u,s)}$ is the measure on $[0,1] \times [0,\infty)$ that puts unit mass at the point (u, s). [For example, if (u_1, u_2, u_3, s) is a point in $\tilde{K}_v^T$, we must know $\tilde{X}_{u_1}^s(s-)$ and $\tilde{X}_{u_2}^s(s-)$ in order to evaluate $\tilde{X}_v^T(s)$.] To determine each $\tilde{X}_u^s$, we must in turn determine the solutions corresponding to points in the support of I_u^s, so to determine $\tilde{X}_v^T$ we must determine $\tilde{X}_u^s$ for (u, s) in the support of

$$I_v^{T,2} = \delta_{(v,T)} + \int_{[0,1]\times[0,T]} (\delta_{(u,s)} + I_u^s) I_v^T(du \times ds).$$

Iterating this relationship, we must determine $\tilde{X}_u^s$ for all (u, s) in the support of

$$I_v^{T,n+1} = \delta_{(v,T)} + \int_{[0,1]\times[0,T]} I_u^{s,n} I_v^T(du \times ds).$$

Let $|I_v^{T,n}|$ denote $I_v^{T,n}([0,1] \times [0,T])$. In general, $I_v^{T,n}$ will have points of multiplicity greater than 1; however, $\lim_{n\to\infty} |I_v^{T,n}|$ will be an upper bound on the number of solutions $\tilde{X}_u^s$ needed to determine $\tilde{X}_v^T$. Since the martingale properties of I_v^T imply

$$E[|I_v^{T,n+1}|] = 1 + \int_0^T \int_0^1 E[|I_u^{s,n}|](2\bar{\sigma} + \alpha)\,du\,ds$$

and $E[|I_u^{s,n}|]$ does not depend on u and is increasing in n, we have

$$\lim_{n\to\infty} E[|I_v^{T,n}|] = e^{(2\bar{\sigma}+\alpha)T}.$$

Consequently, $\lim_{n\to\infty} |I_v^{T,n}| < \infty$ a.s. and the support of

$$\tilde{I}_v^T = \lim_{n\to\infty} I_v^{T,n} \tag{6.11}$$

is finite.

Let

$$H_v^T = \operatorname{supp}(\tilde{I}_v^T) \tag{6.12}$$

To determine $\tilde{X}_v^T$, we only need to consider a finite system of equations for $\{\tilde{X}_u^s : (u,s) \in H_v^T\}$.

7. Finite-dimensional approximation. Fleming–Viot processes arise naturally as limits of finite-population models. In this section, we show how this convergence can be obtained using the stochastic equations developed above. In particular, consider the finite-dimensional system $1 \le j \le n$:

$$\begin{aligned}
X_j^n(t) &= X_j(0) + \int_{U\times[0,t]} h\big(X_j^n(s-), u\big) M_j(du \times ds) \\
&\quad + \sum_{i=1}^{j-1} \big(X_i^n(s-) - X_j^n(s-)\big)\,dL_{ij}(s) \\
&\quad + \sum_{1\le i<k<j} \big(X_{j-1}^n(s-) - X_j^n(s-)\big)\,dL_{ik}(s) \\
&\quad + \int_{[0,1]^3\times[0,t]} I_{\{k_n(u_1,s)\ne k_n(u_2,s)\ne j\}} \\
&\qquad \times \big(X_{k_n(u_1,s)}^n(s-) - X_j^n(s-)\big) \\
&\qquad \times I_{[0,\bar{\sigma}^{-1}\sigma(X_{k_n(u_1,s)}^n(s-), X_{k_n(u_2,s)}^n(s-))]}(u_3) \\
&\qquad \times K_j(du_1 \times du_2 \times du_3 \times ds) \\
&\quad + \int_{[0,1]^2\times[0,t]} \big(r(X_j^n(s-), X_{k_n(u_1,s)}^n(s-), u_2) - X_j^n(s-)\big) \\
&\qquad \times J_j(du_1 \times du_2 \times ds).
\end{aligned} \tag{7.1}$$

Then X^n is a solution of the martingale problem for

$$
\begin{aligned}
A^n f(x) = {} & \sum_{i=1}^n B_i f(x) + \sum_{1\le i<j\le n} (f(\theta_j(x \mid x_i)) - f(x)) \\
& + \frac{1}{n^2} \sum_{1\le i\ne j\ne k\le n} \sigma(x_i, x_k)(f(\eta_j(x \mid x_i)) - f(x)) \\
& + \frac{\alpha}{n} \sum_{1\le i\ne k\le n} \left(\int_E f(\eta_i(x \mid z)) R(x_i, x_k, dz) - f(x) \right)
\end{aligned} \tag{7.2}
$$

[which is just (2.14) without the factor $(n-2)/n$ multiplying the second term]. For the analogue of (6.7) and (6.8), let $\tilde{X}_j^{T,n}(t) = X^n_{N_j^T(t)}(t)$, $j = 1, \dots, n$, and $\tilde{X}_v^{T,n}(t) = X^n_{N^T_{k_n(v,T)}(t)}(t)$. Then for $0 \le t \le T$,

$$
\begin{aligned}
\tilde{X}_j^{T,n}(t) = {} & X_{N_j^T(0)}(0) + \int_{U\times[0,t]} h(\tilde{X}_j^{T,n}(s-), u) \tilde{M}_j^T(du \times ds) \\
& + \int_{[0,1]^3\times[0,t]} (\tilde{X}_{u_1}^{s,n}(s-) - \tilde{X}_j^{T,n}(s-)) I_{[0,\bar{\sigma}^{-1}\sigma(\tilde{X}_{u_1}^{s,n}(s-), \tilde{X}_{u_2}^{s,n}(s-))]}(u_3) \\
& \times \tilde{K}_j^T(du_1 \times du_2 \times du_3 \times ds) \\
& + \int_{[0,1]^2\times[0,t]} (r(\tilde{X}_j^{T,n}(s-), \tilde{X}_{u_1}^{s,n}(s-), u_2) - \tilde{X}_j^{T,n}(s-)) \\
& \times \tilde{J}_j^T(du_1 \times du_2 \times ds)
\end{aligned} \tag{7.3}
$$

and for $0 \le t \le T$,

$$
\begin{aligned}
\tilde{X}_v^{T,n}(t) = {} & X_{N^T_{k_n(v,T)}(0)}(0) + \int_{U\times[0,t]} h(\tilde{X}_v^{T,n}(s-), u) \tilde{M}^T_{k_n(v,T)}(du \times ds) \\
& + \int_{[0,1]^3\times[0,t]} (\tilde{X}_{u_1}^{s,n}(s-) - \tilde{X}_v^{T,n}(s-)) I_{[0,\bar{\sigma}^{-1}\sigma(\tilde{X}_{u_1}^{s,n}(s-), \tilde{X}_{u_2}^{s,n}(s-))]}(u_3) \\
& \times \tilde{K}^T_{k_n(v,T)}(du_1 \times du_2 \times du_3 \times ds) \\
& + \int_{[0,1]^2\times[0,t]} (r(\tilde{X}_v^{T,n}(s-), \tilde{X}_{u_1}^{s,n}(s-), u_2) - \tilde{X}_v^{T,n}(s-)) \\
& \times \tilde{J}^T_{k_n(v,T)}(du_1 \times du_2 \times ds).
\end{aligned} \tag{7.4}
$$

THEOREM 7.1. *Suppose X^n satisfies* (7.1).

(a) *For each $T > 0$, there exists a set $\mathcal{N}_T \subset [0, T]$ with Lebesgue measure $m(\mathcal{N}_T) = 0$ such that for $v \in [0, T] - \mathcal{N}_T$,*

$$
\lim_{n\to\infty} P\{\tilde{X}_v^{T,n}(t) = \tilde{X}_v^T(t),\ 1 \le t \le T\} = 1, \tag{7.5}
$$

and similarly, with v replaced by j.

(b) *For each* $T > 0$ *and* $j = 1, 2, \dots,$

$$\lim_{n\to\infty} P\{X_j^n(t) = X_j(t),\ 1 \le t \le T\} = 1. \tag{7.6}$$

(c) *If* $X(0)$ *is exchangeable, then for each* $t > 0$, $X(t)$ *is exchangeable, and letting* $Z(t)$ *be the de Finetti measure for* $X(t)$ *and*

$$Z^n(t) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i^n(t)}$$

for $t \ge 0$, $m = 1, 2, \dots$, *and each* $f \in B(E^m)$,

$$\lim_{n\to\infty} E[|\langle f, Z(t)^m\rangle - \langle f, Z_n(t)^m\rangle|] = 0. \tag{7.7}$$

PROOF. Let $\gamma^{T,n}(v) = \max\{t \le T : N_{k_n(v,T)}(t) = N^T(v,t)\}$. Recall that $N_{k_n(v,T)}(t) = N^T(v,t)$ implies $N_{k_n(v,T)}(s) = N^T(v,s)$ for $s \le t$. Since $\gamma^{T,n}(v) \to T$ for almost every v, it follows that

$$\begin{aligned} P\{\tilde{M}_v^T \ne \tilde{M}_{k_n(v,T)}^T\} &\le P\{\tilde{M}_v^T(U \times [\gamma^{T,n}(v), T]) > 0\} \\ &\quad + P\{\tilde{M}_{k_n(v,T)}^T(U \times [\gamma^{T,n}(v), T]) > 0\} \\ &= 2E[1 - e^{-\nu(U)(T-\gamma^{T,n}(v))}], \end{aligned}$$

which converges to zero. Similar inequalities hold for K and J, and by the finiteness of H_v^T defined in (6.12), (7.5) follows.

Since $X_{k_n(u,s)}^n(s-) = \tilde{X}_u^{s,n}(s-)$ by definition and for n large $\tilde{X}_u^{s,n}(s-) = \tilde{X}_u^s(s-)$ with high probability, it follows that X^n satisfies (7.6) with X the solution of

$$\begin{aligned} X_j(t) &= X_j(0) + \int_{U\times[0,t]} h(X_j(s-), u) M_j(du \times ds) \\ &\quad + \sum_{i=1}^{j-1} \big(X_i(s-) - X_j(s-)\big)\, dL_{ij}(s) \\ &\quad + \sum_{1\le i<k<j} \big(X_{j-1}(s-) - X_j(s-)\big)\, dL_{ik}(s) \\ &\quad + \int_{[0,1]^3\times[0,t]} (\tilde{X}_{u_1}^s(s-) - X_j(s-)) \\ &\qquad \times I_{[0,\bar{\sigma}^{-1}\sigma(\tilde{X}_{u_1}^s(s-), \tilde{X}_{u_2}^s(s-))]}(u_3) \\ &\qquad \times K_j(du_1 \times du_2 \times du_3 \times ds) \\ &\quad + \int_{[0,1]^2\times[0,t]} \big(r(X_j(s-), \tilde{X}_{u_1}^s(s-), u_2) - X_j(s-)\big) \\ &\qquad \times J_j(du_1 \times du_2 \times ds). \end{aligned} \tag{7.8}$$

However, $\tilde{X}_u^{s,n}(t) = X_{N_{k_n(u,s)}^s(t)}^n(t) \to X_{N_u^s(t)}(t)$ for almost every u. Consequently, since $\tilde{X}_u^s(t) = X_{N_u^s(t)}(t)$, by (6.5), X is a solution of (6.3).

If $X(0)$ is exchangeable, then the exchangeability of $\{(\tilde{M}_j^t, \tilde{K}_j^t, \tilde{J}_j^t)\}$ implies the exchangeability of

$$\{(\tilde{X}_j^t, \tilde{X}_j^{t,n}),\ j = 1, \dots, n\}.$$

Consequently,

$$P\{X_j(t) = X_j^n(t)\} = P\{\tilde{X}_j^t(t) = \tilde{X}_j^{t,n}(t)\}$$

does not depend on j, $1 \le j \le n$, and setting $\hat{Z}_n(t) = (1/n)\sum_{j=1}^n \delta_{X_j(t)}$, (7.6) implies

$$\lim_{n\to\infty} E[|\langle f, \hat{Z}_n(t)^m\rangle - \langle f, Z_n(t)^m\rangle|] = 0.$$

However, the exchangeability of $X(t)$ implies

$$\lim_{n\to\infty} E[|\langle f, \hat{Z}_n(t)^m\rangle - \langle f, Z(t)^m\rangle|] = 0$$

and (7.7) follows. □

The following corollaries can also be obtained as applications of Theorems 4.1 and 4.4.

COROLLARY 7.2. *If X satisfies* (6.3), *then X is a solution of the martingale problem for A given by* (3.10).

PROOF. The corollary follows from (7.6), (7.7) and the fact that X^n is a solution of the martingale problem for A^n. In particular, for $f(x) = f(x_1, \dots, x_m)$ for $f \in B(E^m)$, the martingale

$$f(X^n(t)) - \int_0^t A^n f(X(s))\,ds$$

converges in L_1 to $f(X(t)) - \int_0^t Af(X(s))\,ds$. □

COROLLARY 7.3. *If $X(0)$ is exchangeable, then for each $t > 0$, $X(t)$ is exchangeable and for $m = 1, 2, \dots,$*

$$E[f(X_1(t), \dots, X_m(t)) \mid \mathcal{F}_t^Z] = \langle f, Z(t)^m\rangle. \tag{7.9}$$

PROOF. By Theorem 2.5,

$$E\left[f(X_1^n(t), \dots, X_m^n(t)) \prod_{j=1}^k \langle h_j, Z_n(t_j)^{m_j}\rangle\right]$$
$$= E\left[\langle f, Z_n(t)^{(m)}\rangle \prod_{j=1}^k \langle h_j, Z_n(t_j)^{m_j}\rangle\right],$$

which, passing to the limit, becomes

$$E\left[f(X_1(t),\dots,X_m(t))\prod_{j=1}^k \langle h_j, Z(t_j)^{m_j}\rangle\right]$$
$$= E\left[\langle f, Z(t)^m\rangle \prod_{j=1}^k \langle h_j, Z(t_j)^{m_j}\rangle\right]$$

and (7.9) follows. □

8. The genealogical tree. We now assume that the J_j, K_j, M_j and Ξ_{jl} and the Poisson processes L_{ij} are defined for the doubly infinite time interval. Assume that X is defined for $t_0 \le t < \infty$ or, as in Section 5, for $-\infty < t < \infty$, and that for each such t, $X(t)$ is independent of

$$\mathscr{G}^t = \sigma\{J_j(\cdot\times[r,s]), K_j(\cdot\times[r,s]), L_{ij}[r,s], M_j(\cdot\times[r,s]),\\ \Xi_{ij}(\cdot\times[r,s]) : t \le r < s, i, j\}.$$

[By uniqueness, this independence will hold for all $t \ge t_0$ if $X(t_0)$ is independent of $\mathscr{G}^{t_0}$.] Let τ be a random variable satisfying $\{\tau \ge t\} \in \mathscr{G}^t$ for all t. Note that $\{\mathscr{G}^t\}$ can be thought of as a filtration with time running backward and τ is then a $\{\mathscr{G}^t\}$-stopping time. Following the usual development of stopping times, we define

$$\mathscr{G}^\tau = \{A \in \mathscr{F} : A \cap \{\tau \ge t\} \in \mathscr{G}^t \text{ all } t\}.$$

Let $\mathscr{F}_{\tau-} = \sigma(X(\tau - s) : s > 0)$. The independent increments properties of J_j, K_j, L_{ij}, M_j and Ξ_{ij} imply the following lemma.

LEMMA 8.1. *Let τ, $\mathscr{G}^\tau$ and $\mathscr{F}_{\tau-}$ be as above. Then $\mathscr{F}_{\tau-}$ is independent of $\mathscr{G}^\tau$. In particular, $X(\tau-)$ is independent of $\mathscr{G}^\tau$ and, if X is stationary with marginal distribution Π_X, the distribution of $X(\tau-)$ is Π_X.*

Throughout this section, $\Pi_X \in \mathscr{P}_e(E^\infty)$ will be a stationary distribution for X, $\Pi_Z \in \mathscr{P}(\mathscr{P}(E))$ will be the corresponding stationary distribution for the Fleming–Viot process and $\pi \in \mathscr{P}(E)$ will be the distribution of an individual sampled from the stationary distribution, that is,

$$\pi(\Gamma) = P\{X_i(t) \in \Gamma\} = \int_{\mathscr{P}(E)} \mu(\Gamma)\Pi_Z(d\mu). \tag{8.1}$$

8.1. *The ancestral influence graph.* Let $\tilde{J}_v^T$, $\tilde{K}_v^T, \cdots$ be defined as before and, as in Section 6.2, we now set

$$I_v^T = \int_{[0,1]^3\times(-\infty,T]} (\delta_{(u_1,s)} + \delta_{(u_2,s)})\tilde{K}_v^T(du_1\times du_2\times du_3\times ds)\\ + \int_{[0,1]^2\times(-\infty,T]} \delta_{(u_1,s)}\tilde{J}_v^T(du_1\times du_2\times ds) \tag{8.2}$$

and similarly define

$$I_j^T = \int_{[0,1]^3\times(-\infty,T]} (\delta_{(u_1,s)} + \delta_{(u_2,s)})\tilde{K}_j^T(du_1 \times du_2 \times du_3 \times ds) + \int_{[0,1]^2\times(-\infty,T]} \delta_{(u_1,s)}\tilde{J}_j^T(du_1 \times du_2 \times ds).$$

Define $I_v^{T,1} = \delta_{(v,T)} + I_v^T$ and

$$I_v^{T,n+1} = \delta_{(v,T)} + \int_{[0,1]\times(-\infty,T]} I_u^{s,n} I_v^T(du \times ds),$$

and let $\tilde{I}_v^T = \lim_{n\to\infty} I_v^{T,n}$. Note that the sequence of measures is increasing and, as in Section 6.2, $\sup_n E[I_v^{T,n}([0,1]\times[t,T])] < \infty$ for all $-\infty < t < T$, so the limit exists. Let

$$\tilde{I}_j^T = \int_{[0,1]\times[0,T]} \tilde{I}_u^s I_j^T(du \times ds).$$

Then $H_j^T = \operatorname{supp}(\tilde{I}_j^T)$ (the support of the measure $\tilde{I}_j^T$) gives the index set for the collection of processes $\tilde{X}_u^s$ that can "influence" the value of $X_j(T) = \tilde{X}_j^T(T)$. For $-\infty < t \le T$, let

$$\Gamma_j^T(t) = \{N_u^s(t) : (u,s) \in H_j^T,\ s > t\} \cup \{N_j^T(t)\}.$$

$\Gamma_j^T(t)$ is the set of levels such that $X_k(t)$, $k \in \Gamma_j^T(t)$, can influence the value of $X_j(T)$. Reversing time, let $Q_j^T(t) = |\Gamma_j^T(T-t)|$. Then

$$\begin{aligned} Q_j^T(t) = 1 + \sum_{i=1}^{\infty}\Big(2\int_{T-t}^{T} I_{\Gamma_j^T(s)}(i) K_i([0,1]^3 \times ds) \\ + \int_{T-t}^{T} I_{\Gamma_j^T(s)}(i) J_i([0,1]^2 \times ds)\Big) \\ - \sum_{1\le i<k} \int_{T-t}^{T} I_{\Gamma_j^T(s)}(i) I_{\Gamma_j^T(s)}(k)\, dL_{ik}(s). \end{aligned}$$

It follows that Q_j^T is a Markov chain adapted to the filtration $\{\mathscr{G}^{T-t}, t \ge 0\}$ with transition intensities $q_{k,k+1} = \alpha k$, $q_{k,k+2} = \bar{\sigma} k$ and $q_{k,k-1} = \binom{k}{2}$ and $Q_j^T(0) = 1$. More generally, if we define $\hat{\Gamma}_m^T(t) = \cup_{j=1}^m \Gamma_j^T(t)$ and $\hat{Q}_m^T(t) = |\hat{\Gamma}_m^T(T-t)|$, then $\hat{Q}_m^T$ is a Markov chain with the same transition intensities and $\hat{Q}_m^T(0) = m$. Using the fact that $E[(\hat{Q}_m^T(t))^2] \ge E[\hat{Q}_m^T(t)]^2$, we have

$$\begin{aligned} E[\hat{Q}_m^T(t)] &= m + \int_0^t \left((2\bar{\sigma}+\alpha)E[\hat{Q}_m^T(s)] - \tfrac{1}{2}E[\hat{Q}_m^T(s)(\hat{Q}_m^T(s)-1)]\right) ds \\ &\le m + \int_0^t \left((2\bar{\sigma}+\alpha+\tfrac{1}{2})E[\hat{Q}_m^T(s)] - \tfrac{1}{2}E[\hat{Q}_m^T(s)]^2\right) ds. \end{aligned}$$

It follows that $\lim_{m\to\infty} E[\hat{Q}_m^T(t)] < \infty$ for all $t > 0$. Consequently, for $t < T$,

$$\hat{\Gamma}^T(t) = \bigcup_{j=1}^{\infty} \Gamma_j^T(t) \tag{8.3}$$

is a finite set with probability 1 and, hence, ∞ is an entrance boundary for the Markov chain $\hat{Q}_\infty^T(t) = |\hat{\Gamma}^T(T-t)|$. Since the Markov chain is irreducible and positive recurrent,

$$\tau_{UA}^T = \sup\{t < T : |\hat{\Gamma}^T(t)| = 1\} > -\infty. \tag{8.4}$$

Following Krone and Neuhauser (1997), we will refer to τ_{UA}^T as the *time of the ultimate ancestor* of the population at time T and refer to the graph determined by $\{(t, j) : j \in \hat{\Gamma}^T(t), \tau_{UA}^T \le t \le T\}$ as the *ancestral influence graph*. Let k_{UA}^T be the unique level in $\hat{\Gamma}^T(\tau_{UA}^T-)$ and observe that $\{\tau_{UA}^T \ge t\} \in \mathscr{G}^t$ and k_{UA}^T is $\mathscr{G}^{\tau_{UA}^T}$-measurable.

τ_{UA}^T is not necessarily the time of the most recent common ancestor, since $\hat{\Gamma}^T(t)$ may include levels of particles that do not have descendants at time T. For example, particles that play a role at a $\bar{\sigma}$-branch point may not leave a descendant. It is the case, however, that if we specify the type of $X_{k_{UA}^T}(\tau_{UA}^T)$, then $X(T)$ is uniquely determined. The following result is a consequence of Lemma 8.1.

THEOREM 8.2. *Suppose X is stationary with marginal distribution $\Pi_X \in \mathscr{P}_e(E^\infty)$. Then $X_{k_{UA}^T}(\tau_{UA}^T) = X_{k_{UA}^T}(\tau_{UA}^T-)$ has distribution π given by* (8.1).

8.2. *Core of the ancestral influence graph.* Let $\hat{\Gamma}^T$ be given by (8.3). Note that for $t < T_1 < T_2$, $\hat{\Gamma}^{T_1}(t) \supset \hat{\Gamma}^{T_2}(t)$, and define

$$\hat{\Gamma}(t) = \bigcap_{T>t} \hat{\Gamma}^T(t) = \lim_{T\to\infty} \hat{\Gamma}^T(t),$$

where the limit will exist in the sense that there exists $T_0 > t$ such that $\hat{\Gamma}^T(t) = \hat{\Gamma}^{T_0}(t)$ for all $T > T_0$. Note that $\hat{\Gamma}(t)$ is the subset of levels that are in the ancestral influence graph for every $T > t$. Consequently, we will call $\hat{\Gamma}$ the *core* of the ancestral influence graph. It is clear that run backward in time, $\hat{\Gamma}$ is a stationary, Markov process whose state space is the collection of finite subsets of integers. In particular, $Q(t) = |\hat{\Gamma}(-t)|$ is a stationary version of the Markov chain with transition intensities $q_{k,k+1} = \alpha k$, $q_{k,k+2} = \bar{\sigma} k$ and $q_{k,k-1} = \binom{k}{2}$. Run forward in time, $\hat{\Gamma}$ will, of course, also be a Markov process, and $Q^*(t) = |\hat{\Gamma}(t)|$ will be the stationary time reversal of Q. In particular, if π^Q is the stationary distribution for Q (and, hence, also for Q^*), the transition intensities for Q^* are given by

$$q^*_{k,\,k-1} = \frac{\pi^Q_{k-1}}{\pi^Q_k}\alpha(k-1),$$

$$q^*_{k,\,k-2} = \frac{\pi^Q_{k-2}}{\pi^Q_k}\bar{\sigma}(k-2),$$

$$q^*_{k,\,k+1} = \frac{\pi^Q_{k+1}}{\pi^Q_k}\binom{k+1}{2}.$$

Now assume that X is stationary on $-\infty < t < \infty$. By part (c) of Theorem 4.1, $X(t)$ is exchangeable and, more generally, for each T, $\{\tilde{X}^T_j\}$ is exchangeable. Exchangeability implies that $\hat{Z}(t) = \sum_{k\in\hat{\Gamma}(t)} \delta_{X_k(t)}$ is a Markov process whose state space is the space $\mathscr{I}_f(E)$ of finite, integer-valued measures on E. For functions of the form

$$F\left(\sum_{i=1}^m \delta_{x_i}\right) = \prod_{i=1}^m f(x_i),$$

$\|f\| < 1$, the generator for $\hat{Z}$ is

$$\begin{aligned}
&\mathscr{A}F\left(\sum_{i=1}^m \delta_{x_i}\right)\\
&= \prod_{i=1}^m f(x_i)\Bigg(\sum_{j=1}^m \frac{Bf(x_j)}{f(x_j)} + \frac{q^*_{m,\,m+1}}{m}\sum_{j=1}^m (f(x_j)-1) + \frac{q^*_{m,\,m-2}}{m(m-1)(m-2)}\\
&\qquad\times \sum_{1\le i_1\ne i_2\ne i_3\le m}\Bigg[\frac{\sigma(x_{i_1}, x_{i_3})}{\bar{\sigma}}\left(\frac{1}{f(x_{i_2})f(x_{i_3})} - 1\right)\\
&\qquad\qquad + \left(1 - \frac{\sigma(x_{i_1}, x_{i_3})}{\bar{\sigma}}\right)\left(\frac{1}{f(x_{i_1})f(x_{i_3})} - 1\right)\Bigg]\\
&\qquad + \frac{q^*_{m,\,m-1}}{m(m-1)}\sum_{1\le i_1\ne i_2\le m}\left[\frac{\int f(z)R(x_{i_1}, x_{i_2}, dz)}{f(x_{i_1})f(x_{i_2})} - 1\right]\Bigg)\\
&= \prod_{i=1}^m f(x_i)\Bigg(\sum_{j=1}^m \frac{Bf(x_j)}{f(x_j)} + \frac{q^*_{m,\,m+1}}{m}\sum_{j=1}^m (f(x_j)-1) + \frac{q^*_{m,\,m-2}}{m(m-1)}\\
&\qquad\times \sum_{1\le i_2\ne i_3\le m}\left[1 + \frac{1}{m-2}\sum_{i_1\ne i_2,\, i_3}\frac{\sigma(x_{i_1}, x_{i_3}) - \sigma(x_{i_2}, x_{i_3})}{\bar{\sigma}}\right]
\end{aligned}$$

$$\times\left(\frac{1}{f(x_{i_2})f(x_{i_3})}-1\right)$$
$$+\frac{q^*_{m,m-1}}{m(m-1)}\sum_{1\le i_1\ne i_2\le m}\left[\frac{\int f(z)R(x_{i_1},x_{i_2},dz)}{f(x_{i_1})f(x_{i_2})}-1\right]\Bigg),$$

where the second equality follows by relabelling.

Let $\sigma_1(x, y)$ and $\sigma_2(x, y)$ satisfy $\varepsilon \le \sigma_i(x, y) \le \bar{\sigma} - \varepsilon$, for some $\varepsilon > 0$, and let Ψ_1 and Ψ_2 be core processes corresponding to σ_1 and σ_2 with the same initial distribution. Then the distributions P_1 and P_2 of Ψ_1 and Ψ_2 restricted to a bounded time interval $[0, T]$ are mutually absolutely continuous. To describe the Radon–Nikodym derivative, let $0 < \tau_1 < \tau_2 < \cdots$ be the times at which Q^* jumps by -2 and let $i_1(\tau_i)$ and $i_2(\tau_i)$ be the indices of the particles that are eliminated. Then

$$\begin{aligned}\frac{dP_1}{dP_2}\bigg|_{\mathscr{F}_T} = \prod_{0<\tau_i\le T}&\bigg\{1+\frac{1}{Q^*(\tau_i-)-2}\sum_{j\ne i_1(\tau_i),\, i_2(\tau_i)}\frac{1}{\bar{\sigma}}\big(\sigma_1(X_j(\tau_i-), X_{i_1(\tau_i)}(\tau_i-))\\ &\qquad - \sigma_1(X_{i_2(\tau_i)}(\tau_i-), X_{i_1(\tau_i)}(\tau_i, -))\big)\bigg\}\\ &\times\bigg\{1+\frac{1}{Q^*(\tau_i-)-2}\sum_{j\ne i_1(\tau_i),\, i_2(\tau_i)}\frac{1}{\bar{\sigma}}\big(\sigma_2(X_j(\tau_i-), X_{i_1(\tau_i)}(\tau_i-))\\ &\qquad - \sigma_2\big(X_{i_2(\tau_i)}(\tau_i-), X_{i_1(\tau_i)}(\tau_i-)\big)\big)\bigg\}^{-1}.\end{aligned}$$

If σ_2 is a constant, $\sigma_2(x, y) \equiv \sigma_0$, then even though the system has "extra births," the model is neutral. [In particular, the third term on the right-hand side of (3.11) vanishes.] Assuming that there is no recombination (that is, $\alpha = 0$), then the type process along the ancestral line of any individual in the core process Ψ_2 at time T is just the mutation process. It follows that the distribution of the type process along the ancestral line of an individual in the core process Ψ_1 at time T is absolutely continuous with respect to the distribution of the mutation process.

8.3. *Level of true ancestor.* Fix T and j. We would like to identify the level $\hat{N}^T_j(t)$ of the true ancestor at time $t < T$ of $X_j(T)$. Of course, in the presence of recombination, the definition of $\hat{N}^T_j$ is ambiguous. For simplicity, in tracing $\hat{N}^T_j$ backward in time, we will not change $\hat{N}^T_j$ at recombination times. If R is symmetric in the sense that $R(x_1, x_2, dz) = R(x_2, x_1, dz)$, this convention is equivalent to selecting the "true" ancestor randomly from the two individuals contributing to a recombination.

Then the desired level will satisfy

$$\hat{N}_j^T(t) = N_j^T(t) + \int_{[0,1]^3\times(t,T]} (N^s(u_1,t) - N^s_{\hat{N}_j^T(s)}(t)) I_{[0,\bar{\sigma}^{-1}\sigma(\rho_X(s-,u_1),\rho_X(s-,u_2))]}(u_3) \times \hat{K}_j(du_1\times du_2\times du_3\times ds),$$

where $\hat{K}_j^T$ satisfies

$$\hat{K}_j^T(C\times(t,T]) = \sum_{i=1}^{\infty}\int_{C\times(t,T]} I_{\{\hat{N}_j^T(s)=i\}} K_i(du_1\times du_2\times du_3\times ds).$$

Note that the above equation essentially takes $N_j^T(t)$ as a "first guess" and then corrects this guess at each $\bar{\sigma}$-branch point.

8.4. *Effect of selection on genealogy.* We define the *agreement* of $\xi_i(t)$ and $\xi_j(s)$ by

$$a(\xi_i(t),\xi_j(s)) = \lim_{m\to\infty}\frac{1}{m}\sum_{l=1}^{m} I_{\{\xi_{il}(t)=\xi_{jl}(s)\}}.$$

Note that if $\xi_{il}(t) = \xi_{jl}(s)$, then either the mutation that produced this value is in the ancestral line of both the ith level at time t and the jth level at time s or this marker component was in the population at time zero. Let

$$\tau(i,t,j,s) = 0\vee\sup\{u : N_i^t(u) = N_j^s(u)\}.$$

Then, by the law of large numbers,

$$a(\xi_i(t),\xi_j(s)) = e^{-(t+s-2\tau(i,t,j,s))} a(\xi_{N_i^t(0)}(0),\xi_{N_j^s(0)}(0)). \tag{8.5}$$

In particular, it follows that $\xi_1(t),\dots,\xi_n(t)$ contains complete information about the neutral genealogy of the first n levels at time t. Similarly, if we define

$$\tau(V,t,j,s) = 0\vee\sup\{u : N^t(V,u) = N_j^s(u)\},$$

then

$$a(\xi(V,t),\xi_j(s)) = e^{-(t+s-2\tau(V,t,j,s))} a(\xi_{N_V^t(0)}(0),\xi_{N_j^s(0)}(0)).$$

We can introduce a similar structure into the model with selection and recombination by coupling X satisfying (6.3) with an E_0^∞-valued marker process

Y giving a system for $((X_1, Y_1), (X_2, Y_2), \ldots)$:

$$
\begin{aligned}
X_j(t) = X_j(0) &+ \int_{U\times[0,t]} h(X_j(s-), u) M_j(du \times ds) \\
&+ \sum_{i=1}^{j-1} (X_i(s-) - X_j(s-))\, dL_{ij}(s) \\
&+ \sum_{1\le i<k<j} (X_{j-1}(s-) - X_j(s-))\, dL_{ik}(s) \\
&+ \int_{[0,1]^3\times[0,t]} (\rho_X(s-, u_1) - X_j(s-)) \\
&\quad \times I_{[0,\bar\sigma^{-1}\sigma(\rho_X(s-,u_1),\rho_X(s-,u_2))]}(u_3) \\
&\quad \times K_j(du_1 \times du_2 \times du_3 \times ds) \\
&+ \int_{[0,1]^2\times[0,t]} (r(X_j(s-), \rho_X(s-, u_1), u_2) \\
&\quad - X_j(s-)) J_j(du_1 \times du_2 \times ds),
\end{aligned}
\tag{8.6}
$$

where $\rho_X(s-, u) = \lim_{r\to s-} X_{N^s(u,r)}(r)$, and

$$
\begin{aligned}
Y_j(t) = Y_j(0) &+ \sum_l \int_{[0,1]\times[0,t]} (u - Y_{jl}(s-)) e_l \Xi_{jl}(du \times ds) \\
&+ \sum_{i=1}^{j-1} (Y_i(s-) - Y_j(s-))\, dL_{ij}(s) \\
&+ \sum_{1\le i<k<j} (Y_{j-1}(s-) - Y_j(s-))\, dL_{ik}(s) \\
&+ \int_{[0,1]^3\times[0,t]} (\rho_Y(s-, u_1) - Y_j(s-)) I_{[0,\bar\sigma^{-1}\sigma(\rho_X(s-,u_1),\rho_X(s-,u_2))]}(u_3) \\
&\quad \times K_j(du_1 \times du_2 \times du_3 \times ds),
\end{aligned}
\tag{8.7}
$$

where $\rho_Y(s-, u) = \lim_{r\to s-} Y_{N^s(u,r)}(r)$. Existence and uniqueness for this system is a consequence of the previous result. As in the case for ξ, if we define

$$\hat\tau(i, t, j, s) = 0 \vee \sup\{u : \hat N_i^t(u) = \hat N_j^s(u)\},$$

then, as in (8.5),

$$a(Y_i(t), Y_j(s)) = e^{-(t+s-2\hat\tau(i,t,j,s))} a(Y_{\hat N_i^t(0)}(0), Y_{\hat N_j^s(0)}(0)).$$

Note that at time t, $\zeta(i, j, t) = t - \hat\tau(i, t, j, t)$ is the time since the most recent common ancestor of the individuals at the ith and jth levels and that $\{\zeta(i, j, t), 1 \le i < j\}$ determines the structure of the genealogical tree for the population at time t. It is also clear that conditioned on $\{\zeta(i, j, t), 1 \le i < j\}$, $Y(t)$ is independent of $X(t)$.

The question of whether or not selection has an impact on the distribution of the genealogical tree has been a recent issue. (See the discussion in the

Introduction.) Suppose f is a function only of the marker process, say $f(y) = f(y_1, \ldots, y_m)$. Then, taking $\alpha = 0$ for simplicity and assuming $(X(0), Y(0))$ is exchangeable,

$$\begin{aligned} E[f(Y(t))] = E[f(Y(0))] & \\ &+ E\left[\int_0^t \left(\sum_{i=1}^m B_i^M f(Y(s)) \right.\right. \\ &\left.\left.\qquad + \sum_{1\le i<j\le m} (f(\theta_j(Y(s) \mid Y_i(s))) - f(Y(s))\right) ds\right] \\ &+ \int_0^t E\left[\sum_{i=1}^m (\sigma(X_i(s), X_{m+1}(s)) \right. \\ &\qquad \left. - \sigma(X_{m+2}(s), X_{m+1}(s)))f(Y_1(s), \ldots, Y_m(s))\right] ds, \end{aligned} \tag{8.8}$$

where the selection term has been manipulated using exchangeability and the fact that

$$E[f((X_1(t), Y_1(t)), \ldots, (X_m(t), Y_m(t)))] = E[\langle f, \hat{Z}(t)^m\rangle].$$

Here $\hat{Z}(t)$ denotes the de Finetti measure for $(X(t), Y(t))$.

If selection has no effect on the genealogy or, specifically, if the presence of selection does not affect the one-dimensional distributions of Y, then the left-hand side of (8.8) equals the first two terms on the right and, differentiating by t,

$$E\left[\sum_{i=1}^m (\sigma(X_i(t), X_{m+1}(t)) - \sigma(X_{m+2}(t), X_{m+1}(t)))\, f(Y_1(t), \ldots, Y_m(t))\right] = 0.$$

Since, by exchangeability, each of the summands has the same distribution, we have

$$E\big[(\sigma(X_1(t), X_{m+1}(t)) - \sigma(X_{m+2}(t), X_{m+1}(t)))f(Y_1(t), \ldots, Y_m(t))\big] = 0. \tag{8.9}$$

The identity (8.9) is equivalent to the following: Given a sample of size $m+2$ from the population, let $\mathscr{G}_t^m$ be the σ algebra corresponding to complete information about the genealogy of the first m individuals in the sample. Then

$$E\big[\sigma(X_1(t), X_{m+1}(t)) \mid \mathscr{G}_t^m\big] = E\big[\sigma(X_{m+2}(t), X_{m+1}(t)) \mid \mathscr{G}_t^m\big]. \tag{8.10}$$

If we assume (X, Y) is stationary, then (8.9) and (8.10) hold under the less encompassing assumption that the *stationary* distribution of the genealogy in a model with selection is the same as the neutral genealogy (i.e., the tree given by Kingman's coalescent).

At least in extreme examples, it is easy to check that (8.10) is not valid. Suppose, for example, that there is no mutation, $P\{X_i(0) = X_j(0)\} = 0$ for all

$i \neq j$ and $\sigma(x, x') = I_{\{x=x'\}}$. Then $\sigma(X_i(t), X_j(t)) = I_{\{\zeta(i,j,t)\leq t\}}$, and taking $m = 2$ and $G = \{\zeta(1,2,t) > t\} \in \mathscr{G}_t^2$, (8.10) implies

$$P\{\zeta(1,3,t) \leq t \mid \zeta(1,2,t) > t\} = P\{\zeta(3,4,t) \leq t \mid \zeta(1,2,t) > t\}. \tag{8.11}$$

However, assuming the genealogy with selection is the same as the genealogy without, the conditional probabilities in (8.11) can be calculated using properties of the $\{L_{ij}\}$. Let $\tau_{ij} = \inf\{s : L_{ij}(t-s) < L_{ij}(t)\}$. Then, since the τ_{ij} are independent exponentially distributed random variables,

$$\begin{aligned} p_{13}(t) &\equiv P\{\zeta(1,3,t) \leq t \mid \zeta(1,2,t) > t\} \\ &= P\{\tau_{13} \leq t,\ \tau_{13} < \tau_{23} \mid \tau_{12} > t\} \\ &= \tfrac{1}{2}(1 - e^{-2t}) \end{aligned}$$

and setting $\tau = \min\{\tau_{13}, \tau_{23}, \tau_{14}, \tau_{24}, \tau_{34}\}$,

$$\begin{aligned} &P\{\zeta(3,4,t) \leq t \mid \zeta(1,2,t) > t\} \\ &\quad = \int_0^t P\{N_3^t(0) = N_4^t(0) \mid \tau = s,\ \tau_{12} > t\} 5e^{-5s} ds \\ &\quad = \int_0^t \Big(P\{\tau = \tau_{34} \mid \tau = s, \tau_{12} > t\} \\ &\qquad + P\{N_3^t(0) = N_4^t(0),\ \tau \neq \tau_{34} \mid \tau = s,\ \tau_{12} > t\}\Big) 5^{-5s}\, ds \\ &\quad = \int_0^t \big(\tfrac{1}{5} + \tfrac{4}{5} p_{13}(t-s)\big) 5e^{-5s} ds \\ &\quad = \tfrac{3}{5} - \tfrac{2}{3}e^{-2t} + \tfrac{1}{15}e^{-5t}. \end{aligned}$$

A comparison of the results of these two calculations demonstrates that (8.11) fails.

8.5. *Type distribution at a branch point.* The following example demonstrates that Lemma 8.1 does not, in general, hold with $X(\tau-)$ replaced by $X(\tau)$.

PROPOSITION 8.3. *Let X be stationary and satisfy the conditions of Lemma* 8.1, *and let Π_Z be the stationary distribution of the corresponding Fleming–Viot process. Fix t and j, and let $\tau = \sup\{s < t : K_j([0,1]^3 \times [s,t)) > 0\}$. Then*

$$\begin{aligned} &E[f(X_j(\tau))] \\ &\quad = E\bigg[f(X_1(0))\bigg(1 - \frac{\sigma(X_2(0), X_3(0))}{\bar\sigma}\bigg) + f(X_2(0))\frac{\sigma(X_2(0), X_3(0))}{\bar\sigma}\bigg] \\ &\quad = \int_{\mathscr{P}(E)} \int_{E^3} \bigg(f(x_1)\bigg(1 - \frac{\sigma(x_2, x_3)}{\bar\sigma}\bigg) + f(x_2)\frac{\sigma(x_2, x_3)}{\bar\sigma}\bigg) \\ &\qquad \times \mu(dx_1)\mu(dx_2)\mu(dx_3)\Pi_Z(d\mu) \end{aligned} \tag{8.12}$$

and

$$
\begin{aligned}
E\Big[\int_E \sigma(X_j(\tau), z)Z(\tau, dz)\Big]
&= E\bigg[\sigma(X_1(0), X_4(0))\Big(1 - \frac{\sigma(X_2(0), X_3(0))}{\bar\sigma}\Big) \\
&\qquad + \sigma(X_2(0), X_4(0))\frac{\sigma(X_2(0), X_3(0))}{\bar\sigma}\bigg] \\
&= \int_{\mathscr{P}(E)}\int_{E^3}\Big(\sigma(x_1, x_4)\Big(1 - \frac{\sigma(x_2, x_3)}{\bar\sigma}\Big) + \sigma(x_2, x_4)\frac{\sigma(x_2, x_3)}{\bar\sigma}\Big) \\
&\qquad \times\mu(dx_1)\mu(dx_2)\mu(dx_3)\mu(dx_4)\Pi_Z(d\mu) \\
&= \int_{\mathscr{P}(E)}\langle\sigma, \mu^2\rangle\Pi_Z(d\mu) \\
&\quad + \frac{1}{\bar\sigma}\int_{\mathscr{P}(E)}\Big(\int_E\langle\sigma(x_2, \cdot), \mu\rangle^2\mu(dx_2) - \langle\sigma, \mu^2\rangle\Big)\Pi_Z(d\mu) \\
&\geq \int_{\mathscr{P}(E)}\langle\sigma, \mu^2\rangle\Pi_Z(d\mu) = E\big[\sigma(X_1(0), X_2(0))\big] \\
&= E\Big[\int_E \sigma(X_k(\tau), z)Z(\tau, dz)\Big]
\end{aligned}
\tag{8.13}
$$

for $k \neq j$.

In the case of genic selection,

$$E[\sigma(X_j(\tau))] = \int_E \sigma\, d\pi + \int_{\mathscr{P}(E)}(\langle\sigma^2, \mu\rangle - \langle\sigma, \mu\rangle^2)\Pi_Z(d\mu).$$

REMARK 8.4. The inequality in (8.13) makes explicit the intuitive idea that the particle produced at a $\bar\sigma$-branch point should be "more fit" than a particle selected randomly from the population at that time.

PROOF OF PROPOSITION 8.3. By Lemma 8.1, $X(\tau-)$ has distribution Π_X. Note that for $k \neq j$, $X_k(\tau) = X_k(\tau-)$ a.s. If (u_1, u_2, u_3, τ) is the point in K_j at time τ, then

$$f(X_j(\tau)) = f(X_j(\tau-))I_{(\bar\sigma^{-1}\sigma(V, W), 1]}(u_3) + f(V)I_{[0, \bar\sigma^{-1}\sigma(V, W)]}(u_3),$$

where $V = \tilde X^\tau_{u_1}(\tau-)$ and $W = \tilde X^\tau_{u_2}(\tau-)$. Consequently, (8.12) follows from the fact that $X_j(\tau-), V, W$ is a sample of size 3 from $Z(\tau-) = Z(\tau)$, and the distribution of $Z(\tau-)$ is Π_Z. □

9. Ergodicity. Let $\nu_1, \nu_2 \in \mathscr{P}(S)$ and let f_i be the Radon–Nikodym derivative of ν_i with respect to $\nu_1 + \nu_2$. The the *total variation* distance $\|\nu_1 - \nu_2\|$ is given by

$$\int_S |f_1(s) - f_2(s)|(\nu_1(ds) + \nu_2(ds)).$$

Define the *overlap* by

$$O(\nu_1, \nu_2) = \int_S f_1(s) \wedge f_2(s)(\nu_1(ds) + \nu_2(ds)).$$

Let $\nu_0, \mu_1, \mu_2 \in \mathscr{P}(S)$ be given by

$$\nu_0(C) = \frac{1}{O(\nu_1, \nu_2)} \int_C f_1(s) \wedge f_2(s)(\nu_1(ds) + \nu_2(ds))$$

and

$$\mu_i(C) = \frac{1}{1 - O(\nu_1, \nu_2)} \int_C (f_i(s) - f_1(s) \wedge f_2(s))(\nu_1(ds) + \nu_2(ds)).$$

Then μ_1 and μ_2 are mutually singular and

$$\nu_i = O(\nu_1, \nu_2)\nu_0 + (1 - O(\nu_1, \nu_2))\mu_i. \tag{9.1}$$

It follows that

$$\|\nu_1 - \nu_2\| = 2 - 2O(\nu_1, \nu_2)$$

and

$$O(\nu_1, \nu_2) = \inf_{G \in \mathscr{B}(S)} (\nu_1(G) + \nu_2(G^c)). \tag{9.2}$$

The infimum is achieved by G satisfying $\mu_1(G) = 0$ and $\mu_2(G^c) = 0$.

Let C be the generator of a Markov process with state space S and for $\nu \in \mathscr{P}(S)$, let $P_\nu(t)$ denote the one-dimensional distributions for the corresponding process with initial distribution ν. We will say that C is *strongly connected* if and only if for each $\nu_1, \nu_2 \in \mathscr{P}(S)$, there exists a $t > 0$ such that $P_{\nu_1}(t)$ and $P_{\nu_2}(t)$ are not mutually singular, that is, $O(P_{\nu_1}(t), P_{\nu_2}(t)) > 0$. Note that the overlap will be a nondecreasing function of t.

A Markov process is *uniformly ergodic* if

$$\lim_{t \to \infty} \sup_{\nu_1, \nu_2 \in \mathscr{P}(S)} \|P_{\nu_1}(t) - P_{\nu_2}(t)\| = 0. \tag{9.3}$$

Note that a uniformly ergodic Markov process has a unique stationary distribution since, for t sufficiently large, the mapping $\nu \to P_\nu(t)$ will be a (strict) contraction. [See Meyn and Tweedie (1993), Chapter III, for further discussion.]

LEMMA 9.1. *The Markov process with generator C is uniformly ergodic if and only if there exists $t > 0$ such that*

$$\delta(t) = \inf_{\nu_1, \nu_2 \in \mathscr{P}(S)} O(P_{\nu_1}(t), P_{\nu_2}(t)) > 0.$$

PROOF. Suppose $\delta(t) > 0$. Then by (9.1),

$$\begin{aligned}\|P_{\nu_1}(t) - P_{\nu_2}(t)\| &= (1 - O(\nu_1, \nu_2))\|P_{\mu_1}(t) - P_{\mu_2}(t)\| \\ &\le (1 - O(\nu_1, \nu_2))2(1 - \delta(t)) \\ &= (1 - \delta(t))\|\nu_1 - \nu_2\|\end{aligned}$$

and the lemma follows. □

9.1. *Uniqueness of stationary distribution.* Assume that the demography is defined for all $-\infty < t < \infty$. As in (8.4), let

$$\tau_{UA}^T = \sup\{t < T : |\hat{\Gamma}^T(t)| = 1\}$$

and also define

$$\gamma_{UA}^T = \sup\{t < \tau_{UA}^T : |\hat{\Gamma}^T(t)| > 1\}.$$

Note that $\tau_{UA}^T - \gamma_{UA}^T$ is exponentially distributed with parameter $\alpha + \bar{\sigma}$ and that

$$P\left\{\tau_{UA}^T > \frac{T}{2},\ \gamma_{UA}^T < 0\right\} > 0.$$

THEOREM 9.2. *If the mutation process is strongly connected, then the particle process is strongly connected.*

PROOF. Let $P_\nu(t)$ denote the one-dimensional distributions for the mutation process with initial distribution $\nu \in \mathscr{P}(E)$ and let $\hat{P}_{\hat{\nu}}(t)$ be the one-dimensional distributions of the particle process with $\hat{\nu} \in \mathscr{P}(E^\infty)$.

For $\gamma_{UA}^T \vee 0 \le s < \tau_{UA}^T \vee 0$, let $\kappa(s)$ denote the unique level in $\hat{\Gamma}^T(s)$. Let $t_0 = \gamma_{UA}^T \vee 0$. Conditioned on $X_{\kappa(t_0)}$, $X_{\kappa(s)}(s)$ is a version of the mutation process independent of the demography. Let $\hat{\nu}_1, \hat{\nu}_2 \in \mathscr{P}(E^\infty)$ and let $\nu_i \in \mathscr{P}(E)$ be given by

$$\langle f, \nu_i\rangle = E\left[\int f(x_{\kappa(0)})\hat{\nu}_i(dx) \,|\, \tau_{UA}^T > \frac{T}{2},\ \gamma_{UA}^T < 0\right].$$

Then

$$O(\hat{P}_{\hat{\nu}_1}(T), \hat{P}_{\hat{\nu}_2}(T)) \ge P\left\{\tau_{UA}^T > \frac{T}{2},\ \gamma_{UA}^T < 0\right\} O\left(P_{\nu_1}\left(\frac{T}{2}\right), P_{\nu_2}\left(\frac{T}{2}\right)\right). \tag{9.4}$$

Since the probability on the right is positive for every $T > 0$, the theorem follows. □

The following corollary is essentially Theorem 5.3 of Ethier and Kurtz (1999).

COROLLARY 9.3. *If the mutation process B is strongly connected, then there is at most one stationary distibution for A and hence for $\mathbb{A}$.*

PROOF. If there were more than one stationary distribution for A, then there would be two mutally singular stationary distributions. However, strong connectedness for B implies strong connectedness for A, and hence one cannot have two mutually singular stationary distributions for A. □

9.2. *Existence of stationary distributions.* Recall that uniform ergodicity implies both existence and uniqueness of a stationary distribution.

THEOREM 9.4. *If the mutation process is uniformly ergodic, then the particle process is uniformly ergodic.*

REMARK 9.5. At least with the general form of recombination considered here, the particle process may be uniformly ergodic without the mutation process being uniformly ergodic. If $\alpha = 0$, then the mutual absolute continuity of the distribution of the Fleming–Viot processes with and without selection implies that for $f \geq 0$,

$$E[f(X_1(t))] \leq C(t)\langle P_\nu(t), f\rangle,$$

where ν is the distribution of $X_1(0)$ and $C(t)$ is a bound on the Radon–Nikodym derivative. [See Ethier and Kurtz (1993), Theorem 3.3.] Consequently, by (9.2)

$$\inf_{\hat{\nu}_1,\hat{\nu}_2\in\mathscr{P}(E^\infty)} O(\hat{P}_{\hat{\nu}_1}(t), \hat{P}_{\hat{\nu}_2}(t)) \leq C(t) \inf_{\nu_1,\nu_2\in\mathscr{P}(E)} O(P_{\nu_1}(t), P_{\nu_2}(t)).$$

PROOF OF THEOREM 9.4. The result follows from Lemma 9.1 and (9.4). □

COROLLARY 9.6. *If the mutation process is uniformly ergodic, then there exist unique stationary distributions for A and $\mathbb{A}$.*

9.3. *Mutual absolute continuity of selective and neutral stationary distributions.*

THEOREM 9.7. *Suppose that $\alpha = 0$ and that the mutation process is strongly connected. If there exists a (necessarily unique) stationary distribution Π for $\mathbb{A}$ and a stationary distribution Π_0 for the neutral ($\sigma = 0$) Fleming–Viot process, then Π and Π_0 are mutually absolutely continuous.*

REMARK 9.8. For the general form of recombination being considered here, adding a recombination term to a model that has a stationary distribution may give a model that does not have a stationary distribution. Even if the new model has a stationary distribution, it need not be absolutely continuous with respect to the original stationary distribution. On the other hand, if the original model does not have a stationary distribution, adding a recombination term may produce a model with a stationary distribution. The observation made in (2.16) and (2.17) of Ethier and Kurtz (1999) gives a method of constructing examples of all these situations.

PROOF OF THEOREM 9.7. Let Z denote the Fleming–Viot process with generator $\mathbb{A}$ and let Z_0 be the corresponding neutral Fleming–Viot process. If $Z(0)$ and $Z_0(0)$ have the same distribution, then for each $t > 0$, the distributions of Z and Z_0 on $D_{\mathscr{P}(E)}[0, t]$ are mutually absolutely continuous. [See Ethier and Kurtz (1993), Theorem 3.3.] The strong connectedness of the mutation process implies the strong connectedness of the Fleming–Viot processes, and the theorem follows by Theorem A.1. □

9.4. *Simulation of the stationary distribution.* Krone and Neuhauser (1997) pointed out that the construction of the ancestral selection graph (or, similarly, the ancestral influence graph) provides a method of simulating a sample of a given size from the stationary distribution of the Fleming–Viot process. In particular, to simulate a sample of size m, first simulate $\hat{\Gamma}_m^T(t)$ from $t = T$ back to

$$t = \tau_{UA}^{T,m} = \sup\{t < T : |\hat{\Gamma}_m^T(t)| = 1\}.$$

Then specifying a value for $X_{k_{UA}^{T,m}}(\tau_{UA}^{T,m})$, the mutation process can be simulated forward in time along the paths through the graph determined by $\hat{\Gamma}_m^T$. It is only necessary to simulate a graph that is equivalent to $\hat{\Gamma}_m^T$ in the sense that it has the same branch points and coalescences among edges. It is not necessary to simulate level changes that do not involve branching or coalescence. Note that this simulation of $\hat{\Gamma}_m^T$ requires only simulating exponentially distributed random variables and uniform draws from finite sets.

To be correct, the specified value of $X_{k_{UA}^{T,m}}(\tau_{UA}^{T,m})$ should have distribution π, which, unfortunately, will in general be unknown. One approach to this problem is to simulate $\hat{\Gamma}_m^T$ back to time $\tau_{UA}^{T,m} - t_0$ for some large t_0, and then to simulate the type processes forward in time starting from the levels in $\hat{\Gamma}_m^T(\tau_{UA}^{T,m} - t_0)$.

The description of the core of the ancestral influence graph in Section 8.2 suggests another approach. Simulate the core $(X_1(t), \dots, X_{Q^*(t)}(t))$ forward in time for a long period $[0, t_0]$. Starting with m levels, simulate the ancestral influence graph backward in time until a fixed number of levels k is reached (or back to a fixed time and a random number of levels N). Let $\tau = \sup\{s < t_0 : Q^*(s) = k\}$ or $\sup\{s < t_0 : Q^*(s) = N\}$. Initialize the genealogy with the types $(X_1(\tau), \dots, X_{Q^*(\tau)}(\tau))$ and simulate the types forward through the genealogy.

APPENDIX

A.1. Absolute continuity of stationary distributions. For $i = 0, 1$, let A^i be the generator of a Markov process with state space E, and for $\nu \in \mathscr{P}(E)$, let $P^i_\nu(t,\Gamma)$ denote the one-dimensional distributions of the process with initial distribution ν. Call A^i *strongly connected* if for each $\nu,\mu\in\mathscr{P}(E)$ there exists $t>0$ such that $P^i_\nu(t,\cdot)$ and $P^i_\mu(t,\cdot)$ are not mutually singular.

THEOREM A.1. *Let A_0 and A_1 be generators of Markov processes with state space E such that for each $\nu\in\mathscr{P}(E)$ and each $t>0$, $P^0_\nu(t,\cdot)$ and $P^1_\nu(t,\cdot)$ are mutually absolutely continuous. Suppose A_1 is strongly connected. Then A_0 is strongly connected and if π_0 and π_1 are stationary distributions for A_0 and A_1 respectively, then π_0 and π_1 are mutually absolutely continuous.*

PROOF. Let $f_\nu(t,x)$ satisfy $P^0_\nu(t,dx)=f_\nu(t,x)P^1_\nu(t,dx)$. For $\nu,\mu\in\mathscr{P}(E)$, the strong connectedness of A_1 implies that there exists a $t>0,\varepsilon>0$ and $\gamma\in\mathscr{P}(E)$ such that

$$P^1_\nu(t,\Gamma)\wedge P^1_\mu(t,\Gamma)\ge\varepsilon\gamma(\Gamma),\qquad \Gamma\in\mathscr{B}(E).$$

It follows that

$$P^0_\nu(t,\Gamma)\wedge P^0_\mu(t,\Gamma)\ge\varepsilon\int_\Gamma f_\nu(t,x)\wedge f_\mu(t,x)\gamma(dx),\qquad \Gamma\in\mathscr{B}(E),$$

which, by the strict positivity of f_ν and f_μ implies strong connectedness for A_1.

Suppose $C\in\mathscr{B}(E)$ satisfies $\pi_0(C)=0$. Then, since f_{π_0} is strictly positive, it follows that $P^1_{\pi_0}(t,C)=0$ for all $t>0$. There exists a $t>0$, $\varepsilon>0$ and $\gamma\in\mathscr{P}(E)$ such that

$$P^1_{\pi_0}(T,\Gamma)\wedge P^1_{\pi_1}(T,\Gamma)=P^1_{\pi_0}(T,\Gamma)\wedge\pi_1(\Gamma)\ge\varepsilon\gamma(\Gamma).\tag{A.1}$$

Since (A.1) implies $\gamma\ll\pi_1$, there exists $h\ge 0$ such that $\varepsilon\gamma(\Gamma)=\int_\Gamma h(x)\pi_1(dx)$, and the Markov property in turn implies

$$\begin{aligned}0&=E^1_{\pi_0}\left[\int_T^{T+K}I_C(X(t))\,dt\right]\\&\ge E^1_{\pi_1}\left[h(X(0))\int_0^K I_C(X(t))\,dt\right].\end{aligned}$$

Strong connectedness implies uniqueness of stationary distributions, which in turn implies ergodicity of the corresponding stationary process. Dividing the right-hand side of the above inequality by K and letting $K\to\infty$, we obtain

$$0=E^1_{\pi_1}[h(X(0))]\pi_1(C)=\varepsilon\pi_1(C)$$

and, hence, obtain $\pi_1\ll\pi_0$. Since we have verified strong connectedness for A_0, we can interchange the roles of A_0 and A_1 and show that $\pi_0\ll\pi_1$.

Acknowledgment. Part of the work presented here was done in January 1994 while the authors were in residence at the Institute for Mathematics and its Applications.

REFERENCES

BLACKWELL, D. and DUBINS, L. E. (1983). An extension of Skorohod's almost sure representation theorem. *Proc. Amer. Math. Soc.* **89** 691–692.

DONNELLY, P. and KURTZ, T. G. (1996). A countable representation of the Fleming–Viot measure-valued diffusion. *Ann. Probab.* **24** 698–742.

DONNELLY, P. and KURTZ, T. G. (1999). Particle representations for measure-valued population models. *Ann. Probab.* **27** 166–205.

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence.* Wiley, New York.

ETHIER, S. N. and KURTZ, T. G. (1993). Measure-valued diffusion processes in population genetics. *SIAM J. Control Optim.* **31** 345–386.

ETHIER, S. N. and KURTZ, T. G. (1998). Coupling and ergodic theorems for Fleming–Viot processes. *Ann. Probab.* **26** 533–561.

GRIFFITHS, R. C. and MAJORAM, P. (1997). An ancestral recombination graph. *In Progress in Population Genetics and Human Evolution* (P. Donnelly and S. Tavaré, eds.) **87**. Springer, New York.

KINGMAN, J. F. C. (1982). The coalescent. *Stochastic Process. Appl.* **13** 235–248.

KRONE, S. M. and NEUHAUSER, C. (1997). Ancestral processes with selection. *Theoret. Population Biol.* **51** 210–237.

KURTZ, T. G. (1998a). Martingale problems for conditional distributions of Markov processes. *Electronic Journal of Probability* **3**. To appear.

KURTZ, T. G. (1998b). Equivalence of martingale problems and stochastic differential equations. Unpublished manuscript.

KURTZ, T. G. and STOCKBRIDGE, R. H. (1998). Existence of Markov controls and characterization of optimal Markov controls. *SIAM J. Control Optim.* **36** 609–653.

MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability.* Springer, London.

NEUHAUSER, C. (1998). The ancestral graph and gene genealogy under frequency-dependent selection. Preprint.

NEUHAUSER, C. and KRONE, S. M. (1997). The genealogy of samples in models with selection. *Genetics* **154** 519–534.

TAVARÉ, S. (1984). Line-of-descent and genealogical processes, and their applications in genetics models. *Theoret. Population Biol.* **26** 119–164.

DEPARTMENT OF STATISTICS
UNIVERSITY OF OXFORD
1 SOUTH PARKS ROAD
OXFORD OX1 3TG
UNITED KINGDOM
E-MAIL: donnelly@stats.ox.au.uk

DEPARTMENTS OF MATHEMATICS AND STATISTICS
UNIVERSITY OF WISCONSIN—MADISON
480 LINCOLN DRIVE
MADISON, WISCONSIN 53706-1388
E-MAIL: kurtz@math.wisc.edu