Separation of time scales and convergence to the coalescent in structured populations*

Magnus Nordborg† Stephen M. Krone‡

August 31, 2001

Abstract

Many biological phenomena can be thought of and modeled in terms of generalized population structure, with individuals having a number of different possible states. Certain groups of states in such models are often connected by "migration" that occurs on a time scale that is much faster than the coalescent time scale. We show that, when viewed on the coalescent time scale, the structure associated with such groups of states collapses, and is replaced by a kind of averaging over the states. When the entire structure collapses, the standard coalescent is retrieved. The effect of the population structure on the coalescent is then captured by a simple scaling factor that is related to various notions of "effective population size." It is also possible for parts of the structure to collapse, leading to a reduced structured coalescent.

1 Introduction

The stochastic process known as the coalescent (Kingman, 1982a,b) has come to play a central role in modern population genetics (reviewed in: Hudson, 1990; Donnelly and Tavaré, 1995; Nordborg, 2001). The coalescent is a model of random gene genealogies, or lines of descent. It can be viewed as a large-population approximation to gene genealogies in a number of neutral models with finite population size, such as the Wright–Fisher model or the Moran model, that differ in their assumptions about how reproduction occurs. The coalescent is in this sense analogous to the diffusion approximations in classical population genetics, which can be viewed as approximations to the forward-time dynamics of allele frequencies in the same discrete models (see Ewens, 1979, chapter 5). The remarkable fact is that, for a wide range of discrete models, the same diffusion or coalescent approximation is obtained, *as long as time is scaled appropriately.* The differences between the discrete models is thus captured by a scaling factor.

*SK was supported in part by NSF grants DMS-00-72198 and EPS-00-80935.
†Department of Biological Sciences, University of Southern California
‡Department of Mathematics, University of Idaho

The fact that different discrete processes converge to the same approximation as the population size tends to infinity, is of great practical importance; it means that conclusions based on this approximation will be true (modulo the approximation) for all the processes that converge to it, making these conclusions more robust. We will return to the practical implications in Section 4. Our primary purpose in this chapter is to unify and extend recent results (Nordborg and Donnelly, 1997; Nordborg, 1997, 1999; Möhle, 1998; Kaj et al., 2001), that considerably extend the range of phenomena that can be modeled using the coalescent. The basic idea is that:

1. many biological phenomena can be modeled as generalized population structure, with individuals belonging to different states;
2. certain groups of states in such models are often connected by "migration" that occurs on a time scale that is much faster than the coalescent time scale;
3. when viewed on the coalescent time scale, the structure associated with such groups of states collapses, and is replaced by a kind of averaging over the states.

When the entire structure collapses, the standard coalescent is retrieved in the limit as the population size tends to infinity. The effect of the population structure on the coalescent is then captured by a simple scaling factor, which is related to various notions of "effective population size." When there is only only a partial collapsing of states, one obtains a structured coalescent in the limit.

1.1 The coalescent

It is easiest to explain the coalescent starting from a neutral discrete-time Wright–Fisher model. Consider a haploid population of constant size N, in which generation $\tau + 1$ is formed by multinomial sampling from generation τ. Take a sample of size n from the current population (time $\tau = 0$) and trace its ancestry. Backwards in time, each individual can be viewed as picking its parent at random from among the N individuals in the previous generation. Whenever two (or more) individuals pick the same parent, their lineages are said to *coalesce*. The number of distinct lineages, or ancestors, decreases as we go back in time, and ultimately reaches one.

The genealogy of the sampled individuals consists of two elements. First, there is the topology, *i.e.*, who coalesces with whom. The topology is easy to model: Because of neutrality, individuals are equally likely to reproduce; therefore all pairs of lineages must be equally likely to coalesce. The second element is the branch lengths, and this is where the coalescent scaling comes in. If we follow two lineages back in time, it is easy to see that the probability that they remain distinct τ generations into the past is $(1 - 1/N)^\tau$; *i.e.*, the coalescence time is geometrically distributed with parameter $1/N$. If N is large, this can be approximated by an exponential distribution, in the following manner. Rescale time so that one unit of scaled time corresponds to N generations. Then the probability that the two lineages remain distinct for more than t units of scaled time is

$$\Big(1-\frac{1}{N}\Big)^{\lfloor Nt\rfloor}\to e^{-t}, \tag{1}$$

as N goes to infinity (where $\lfloor Nt\rfloor$ is the largest integer less than or equal to Nt). Thus, in the limit, the coalescence time for a pair of lineages is exponentially distributed with mean 1. Now consider k lineages. The probability that none of them coalesce in the previous generation is

$$\prod_{i=0}^{k-1}\frac{N-i}{N}=\prod_{i=1}^{k-1}\Big(1-\frac{i}{N}\Big)=1-\frac{\binom{k}{2}}{N}+O\Big(\frac{1}{N^2}\Big), \tag{2}$$

and the probability that more than two do so is $O(1/N^2)$. Let $T_N(k)$ be the (scaled) time till the first coalescence event, given that there are currently k lineages. By the same argument as above, as $N\to\infty$, $T_N(k)$ converges to a random variable $T(k)$ which is exponentially distributed with mean $2/[(k(k-1)]$. Furthermore, the probability that more than two lineages coalesce in the same generation can be neglected. Thus, under the coalescent approximation, the number of distinct lineages in the ancestry of a sample of (finite) size n decreases in steps of size one back in time; so $T(k)$ is the time from k to $k-1$ lineages (see Figure 1).

The standard coalescent is thus a continuous-time Markov chain with state space $\{1,2,\ldots,n\}$, and infinitesimal generator matrix Q, having components:

$$q_{i,j}=\begin{cases}\binom{i}{2} & \text{if } j=i-1;\\ -\binom{i}{2} & \text{if } j=i;\\ 0 & \text{otherwise.}\end{cases} \tag{3}$$

The structured coalescent is a straightforward extension, in which the population consists of several patches, coalescence is only allowed between lineages in the same patch, and migration occurs at fixed rates (see Notohara, 1990; Herbots, 1997).

In practice, the fact that the random genealogy of the discrete-time Wright–Fisher model (with time measured in units of N generations) "converges" to the coalescent process as "$N\to\infty$" means that, for a given Wright–Fisher model with *fixed* (and not too small) population size, the genealogy of a sample can be approximated by the coalescent provided we translate t units of coalescent time into $\lfloor Nt\rfloor$ generations.

Thus, the correct scaling of time in the Wright–Fisher model is N, the population size. The intuitive reason for this is simply that the rate of coalescence depends on the population size: lineages in a large population "find each other" more slowly than lineages in a small population. However, factors other than the population size can also affect the coalescence rate. For example, the variance in reproductive success should play a role because, if a small number of

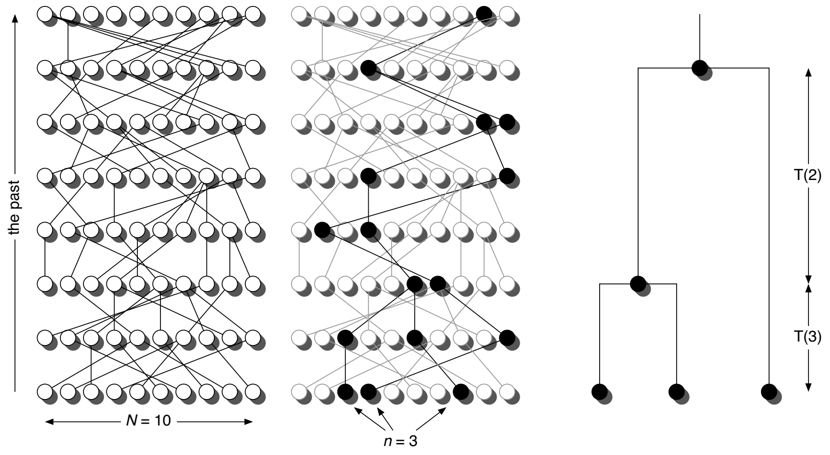

Figure 1: The basic idea behind the coalescent. The left plot shows the genealogy of a small population that is evolving according to the Wright–Fisher model of reproduction. In the center plot, the genealogy of a sample of size $n = 3$ is highlighted. The right plot shows this genealogy in a more concise manner. Because all individuals are reproductively equivalent, the horizontal position of each ancestor carries no meaning. The coalescent models the genealogy of a sample of n individuals as a random bifurcating tree, where the $n-1$ coalescence times $T(n)$, $T(n-1), \ldots, T(2)$ are mutually independent, exponentially distributed random variables. Each pair of lineages coalesces independently at rate 1, so the total rate of coalescence when there are k lineages is "k choose 2."

individuals produce most of the offspring each generation, then the probability that lineages "pick" the same parent must increase, as must surely the rate of coalescence. In the Wright–Fisher model, each individual produces a binomially distributed number of offspring with mean 1 (as it must for the population size to remain constant) and variance $1 - 1/N$. If we consider a generalized model in which the variance of the number of offspring produced by a given individual is $\sigma^2 + O(1/N)$, $0 < \sigma^2 < \infty$, rather than $1 - 1/N$, we find that the standard coalescent is obtained in the limit if time is scaled in units of N/σ^2 generations. Note that $\sigma^2 = 1$ in the standard Wright–Fisher model. The change to the model can be fully accounted for by a simple linear change in the time scale. This is of practical importance, because the variance in reproductive success assumed under the Wright–Fisher model is almost certainly unrealistically small for most organisms. As we will see below, models that are seemingly much more different from the Wright–Fisher model can often be treated in the same way.

The results we are about to describe are applicable to a broad range of models. However, writing them down in full generality would be difficult, and would require a considerable amount of abstract notation. We have therefore chosen to first go through a basic example in full detail (in Section 2 and the Appendix).

This example assumes a haploid model and independently-migrating ancestors. Thereafter, we describe more elaborate examples heuristically (in Section 3).

2 Basic Example

2.1 Structure and scaling

2.1.1 Forward-Time Dynamics

We begin by outlining the general setting and terminology that will be used throughout the paper, and then focus on a haploid "structured Wright–Fisher model" with the nice feature that ancestral lineages migrate independently.

Many models fit into the following hierarchical scheme. The basic idea is that a population can sometimes be partitioned into distinct groups. These groups might coincide with spatial structure, age classes, or some other classification (or a combination of such things). We allow movement between groups via some form of "migration." This might be actual geographical migration, as in the case of spatial structure, or some other kind of movement between groups such as a change in age class. The migration rates will sometimes be quite different between certain groups. For example, if we group individuals according to their spatial location and age class, migration between age classes in the same location might be much faster (perhaps each generation) than migration between different spatial locations (which might happen very infrequently). On a given time scale, migration between some classes may be invisible because it is so rare that it is essentially never seen, whereas migration between other classes may be so rapid that its effects appear only in an average sense. Such differences in scaling will lead to a hierarchical structuring of the population. For reasons that will become clear, we will distinguish between migration rates which are fast or slow relative to the rate of coalescence.

Consider a discrete-time model in which the total population size is fixed at N in each generation. Note that "generation" is not used in the biological sense of "life span": individuals may live several generations and need not be of the same age, and we might have only individuals in a certain class giving birth (see Section 3.2 for such an example). Suppose that each individual in the population can belong to any one of L distinct states and denote the collection of these states by $\mathcal{G}$. Let us suppose for now that the states are labeled so that $\mathcal{G} = \{1, 2, \ldots, L\}$. (We will find it convenient to change this labeling later in cases where there is a significant amount of structure in the population, but the above notation will suffice for now to fix ideas.) We call these states *microstates*; they represent all the information we have (or wish to keep track of) regarding the individuals in the population. We will refer to the set $\mathcal{G}$ as the *individual state space* since, at any given time, an individual in the population will be in exactly one of these states. Each state will be connected to some or all of the other states via some form of migration, which can involve either propagules or adult individuals.

What we have in mind is a collection of states, and migration probabilities that scale like powers of $1/N$. The orders of magnitude in this scaling determine a hierarchy, with clusters of states having the property that migration within each cluster occurs on a much faster time scale than migration between clusters. We will be able to treat such clusters of states as single "*macrostates*" with an averaged or effective rate of migration to other clusters. Thus, some of the structure can collapse in the limit as $N \to \infty$.

To proceed, it is necessary to specify in detail how reproduction and migration occur. In the remainder of this section, we will assume a model in which each individual lives one generation, with reproduction and migration occurring according to a

structured Wright–Fisher model: Each generation, an effectively infinite number of propagules are produced by each individual. These then migrate, after which the adult classes (of the appropriate, finite sizes) are reconstituted by sampling from the post-migration propagules.

Note that this implies that some form of density regulation is in force. However, our results apply as long as it is possible to trace lineages backward in time as described below.

2.1.2 Backward-Time Dynamics

Our goal is to trace the *ancestry* of a sample from a structured population of the above form, and, after suitable scaling, determine the form of the limiting (possibly structured) coalescent. Although we would typically start with a description of the population dynamics forward in time, our interest here is really in following the ancestry of a finite sample of individuals backward in time from the current generation. Since the models we consider are neutral, we need not keep track of which ancestral lineage is in which state; rather, we count only the *number* of ancestors in each state.

Thus, to specify the vector of ancestral lineages in discrete time (*i.e.*, for finite N), we define the backwards *configuration process*, $\mathbf{X}(\tau) = (X_1(\tau), \ldots, X_L(\tau))$ with $X_i(\tau)$ denoting the number of ancestors in state i, τ generations into the past. The configuration process is a discrete-time Markov chain $\{\mathbf{X}(\tau),\ \tau = 0, 1, \ldots\}$ whose state space consists of vectors $\mathbf{x} = (x_1, \ldots, x_L)$ which specify, at any time in the past, the number of ancestral lineages in each state (*i.e.*, microstate). Note that, although the configuration process depends on the population size N, its state space is assumed not to depend on N. The configuration process evolves by (backwards) migration and coalescing of ancestors with the appropriate probabilities as we move back in time one generation at a time. It is this backwards configuration process which will be our starting point in the analysis that follows. Specific examples will be worked out below. Let b_{ij} denote the probability that a lineage "migrates" from state i to state j one generation back in time. For example, if the forward migration probabilities are denoted by m_{ij}, and if the number of individuals in each state is constant, say N_i in state i, then we would have

$$b_{ij} = \frac{N_j m_{ji}}{\sum_k N_k m_{ki}}. \tag{4}$$

Because the number of migrants is infinite in the structured Wright-Fisher model, lineages migrate (backwards) independently of each other. This is not the case in models where the number of migrants is finite. However, as will become clear in Section 3, independent migration is not necessary for our theory to work.

We will assume a diffusion scaling for the backward migration probabilities. In particular, the probability of moving from state i in the current generation to state j in the previous generation will be of the form $b_{ij} = \beta_{ij}/N^\alpha, i \neq j$, for some $0 \leq \alpha \leq 1$. Of course, $b_{ii} = 1 - \sum_{j\neq i} b_{ij} = 1 - N^{-\alpha} \sum_{j\neq i} \beta_{ij}$. In this case, we will say that b_{ij} has *scaling exponent* α. The β_{ij}'s are constants (*i.e.*, do not depend on N), while the b_{ij}'s are generally functions of N unless $\alpha = 0$. (The above scaling can be made slightly more general by replacing $b_{ij} = \beta_{ij}/N^\alpha$ with $N^\alpha b_{ij} \to \beta_{ij}$, as $N \to \infty$. Such a change will not affect the limiting coalescent process, so we will stick to the simpler setup.) Note that when a migration probability b_{ij} has scaling exponent $\alpha = 0$, this probability will not depend on N. If b_{ij} is a non-zero constant ($\alpha = 0$), then a lineage in state i will migrate to state j in a finite number of generations (independent of N). A typical example is subdivision according to age classes. Here, a lineage might change age class each generation. Migration probabilities with scaling exponent $\alpha > 0$ correspond to migration events which take $O(N^\alpha)$ generations to occur; in particular, such events become more and more rare if we let the population size N increase in the model.

We will assume the following about the backward migration probabilities:

- Migration between states is *irreducible* and *aperiodic*. By irreducibility we mean that the migration Markov chain with jump probabilities given by the b_{ij}'s is irreducible. This just means that it is possible to get from one state to any other state in a finite number of steps. Since the chain has a finite state space, this implies that it will have a unique stationary distribution.

- The migration probabilities are *scaling symmetric* in the sense that the scaling exponent is the same in each direction. In other words, if b_{ij} has scaling exponent α, then so does b_{ji}. Of course, some of the b_{ij}'s can be zero; we just need enough of them nonzero to get irreducibility.

Note that none of these assumptions are necessary. One could for instance imagine models where the noncommunicating states would evolve independently and never come together, or models where some states are transient due to one-way migration. More general cases will be discussed in Section 4.

In the interest of simplicity, we will allow for at most two types of scaling for the migration probabilities. *Slow migration* probabilities have scaling exponent 1, *i.e.*, $b_{ij} = \beta_{ij}/N$; *fast migration* probabilities have scaling exponent α, for some fixed $\alpha \in [0,1)$. If we were to allow several different values of $\alpha < 1$, we would get a more complex hierarchical structure and clustering on several time scales. The methods of this paper would still apply, but there would be several stages of collapsing of states, and the notation would be more involved.

Once we know the backwards migration probabilities, we can form clusters of states that are connected via fast migration. More specifically, we say states i and j are in the same cluster if there is a sequence of states, $i = i_1, i_2, \ldots, i_k = j$, such that $b_{i_u,i_{u+1}}$ is positive and has scaling exponent $\alpha < 1$ for each $u = 1, \ldots, k-1$. These clusters partition the state space into equivalence classes. Distinct clusters can be connected by slow migration only.

As we have seen, slow migration happens on the same time scale as coalescence and hence structure that is associated with slow migration will still be visible in the limiting coalescent. On the other hand, fast migration events happen so quickly compared to coalescence events that, as $N \to \infty$, the number of fast migration events between coalescence events tends to infinity. This gives the fast migration jumps time to reach equilibrium between coalescence events. The effect of this is that, in the limiting coalescent, we will be able to treat each cluster of states connected with fast migration as a single macrostate with an averaged or "effective" rate of migration to other clusters and an effective rate of coalescence for lineages within a given macrostate.

Our models will fall into two classes. In the first, all states are connected by fast migration and the limiting ancestral process is the standard coalescent. In the second, there is both fast and slow migration. There will be clusters of microstates, within which all migration is fast. Movement between clusters, on the other hand, is only via slow migration. If there are $\ell \geq 2$ such clusters then the limiting ancestral process will be a structured coalescent with ℓ patches.

When tracing the ancestry of a sample, it is natural to break up the state space for the configuration process into "levels" determined by the total number of ancestors present in all states combined. While there are r ancestors ($r = 1, \ldots, n$), the configuration process moves among the configurations in *level* r:

$$S_r \equiv \{(x_1, \ldots, x_L) : x_1 + \cdots + x_L = r\}.$$

The initial state $\mathbf{X}(0)$ of the configuration process is a point in S_n corresponding to the sample of size n. The configuration process moves within level n for awhile, this movement corresponding to migration of the ancestors as we move back in time. When a coalescence occurs on level n, we drop down to level $n-1$ if only one coalescence occurs, or down to a lower level if two or more coalescences occur. When the process reaches a new level, it starts moving around among the configurations in that level until another coalescence occurs. This continues until we reach level 1; *i.e.*, until a common ancestor of the sample is found. If we start with a sample of size n, the state space for the configuration process is given by $S = S_1 \cup \cdots \cup S_n$; we will call this the *configuration space*. For any configuration $(x_1, \ldots, x_L) \in S$, we will be able to specify probabilities of jumping to other configurations due to migration and/or coalescence of ancestors. We will show that, as $N \to \infty$, the probability of multiple coalescence events and (if $\alpha > 0$) simultaneous migration/coalescence events go to zero rapidly enough that they do not appear in the coalescent.

A good example to think about is the standard island model. For simplicity, assume we have two demes, 1 and 2, and consider the case $n = 2$. Level 2 in this case consists of the configurations

$$S_2 \equiv \{(2,0), (1,1), (0,2)\},$$

corresponding to having two ancestors in deme 1, one ancestor in each deme, and two ancestors in deme 2, respectively. If the two demes are connected by slow migration, the model converges to a structured coalescent. However, if migration is fast, there is a single macrostate, and the standard coalescent is obtained in the limit (see Figure 2). This latter type of behavior is the subject of Section 2.2.

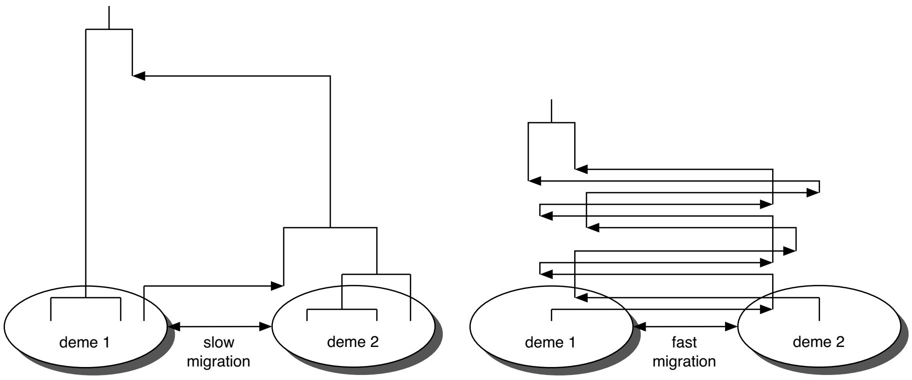

Figure 2: Examples of genealogies in a two-deme model. When migration and coalescence probabilities are of the same order of magnitude, lineages will often coalesce without migrating (see left-hand side). As migration becomes more frequent, lineages will increasingly migrate back and forth between demes a large number of times before coalescing. Coalescence events will then tend to occur in a particular deme with probability proportional to how often that deme is visited, and the original state of the sample no longer matters greatly (see right-hand side). In the limit, the population structure collapses, and the standard coalescent is obtained (described further in Section 2.2).

Now imagine that the demes are further partitioned into two subdemes: 1a and 1b in deme 1, 2a and 2b in deme 2. Assume that migration between subdemes is fast, and migration between demes is slow (see Figure 3). Level 2 in this case consists of the configurations

$$\begin{aligned} S_2 \quad \equiv \quad & \{(2,0;0,0), (1,1;0,0), (0,2;0,0), (1,0;1,0), (1,0;0,1), \\ & (0,1;1,0), (0,1;0,1), (0,0;2,0), (0,0;1,1), (0,0;0,2)\}. \end{aligned}$$

The first three configurations correspond to having two ancestors in deme 1 and none in deme 2; under fast migration between subdemes, these three configurations will collapse to the configuration of macrostates $\langle 2, 0\rangle$ which just records the number of ancestors in each deme. Similarly, the next four configurations in S_2 correspond to having one ancestor in each deme; they will collapse to the configuration $\langle 1, 1\rangle$. The last three configurations will collapse to the $\langle 0, 2\rangle$. This type of behavior is the subject of Section 2.3.

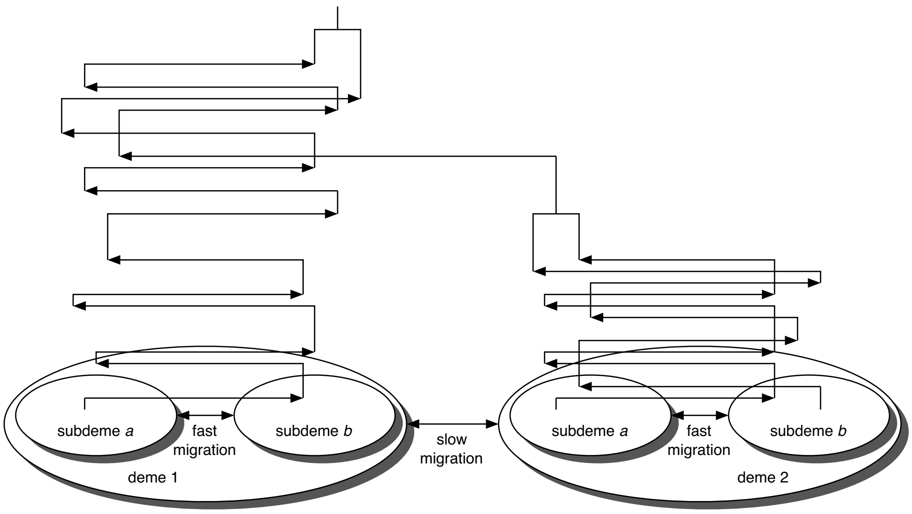

Figure 3: Examples of genealogies in a hierarchical two-deme model. Migration is slow between demes 1 and 2, but fast between the subdemes within each deme. On the coalescent time scale, only migration between the former will be visible. In the limit, a structured coalescent will be obtained, with two macrostates corresponding to the demes. The structure within each deme will collapse (see Section 2.3).

2.2 Full collapse to the standard coalescent

Denote the individual state space by $\mathcal{G} = \{1, \dots, L\}$ and assume that all the microstates are connected by fast migration (and hence collapse to $\ell = 1$ macrostate).

Under the above assumptions, the time between coalescence events will be long enough so that, for large N, the configuration process will reach an equilibrium on each level before the coalescence occurs. These equilibrium distributions on the different levels are studied next.

For a given $r \in \{1, \dots, n\}$, if we consider the configuration process $\mathbf{X}(\tau)$ conditioned to be in level r and to experience no coalescences, we get the *level-r configuration process*, $\mathbf{X}^{(r)}(\tau) = (X_1^{(r)}(\tau), \dots, X_L^{(r)}(\tau))$, $\tau = 0, 1, 2, \dots$, which is governed by backwards migration only. Assume that, for each $r \in \{1, \ldots, n\}$, the level-r configuration process is irreducible and has a unique stationary distribution. To see what this stationary distribution is, let $(\gamma_1, \ldots, \gamma_L)$ be the unique stationary distribution for a single ancestral lineage moving according to the backward migration process determined by the b_{ij}'s. Thus we have

$$\sum_{j=1}^{L} \gamma_j b_{ji} = \gamma_i \tag{5}$$

for each $i \in \{1, \ldots, L\}$. It will be assumed that this stationary distribution is the same for each value of N. In fact, it will be clear in the examples we treat that the level-r configuration process does not depend on N; the dependence on N in the configuration process will come in only through coalescence events, which are ignored in the level-r configuration process.

Now, for each level $r \in \{1, \ldots, n\}$, let $\pi_r = \{\pi_r(\mathbf{x}) : \mathbf{x} \in S_r\}$ be the multinomial distribution

$$\mathrm{Mult}(r; \gamma_1, \ldots, \gamma_L),$$

defined by

$$\pi_r(\mathbf{x}) = \frac{r!}{x_1! \cdots x_L!} \gamma_1^{x_1} \cdots \gamma_L^{x_L}. \tag{6}$$

It is easy to check that this is the stationary distribution for the level-r configuration process. Intuitively, this follows from the fact that the different ancestral lineages move independently from state to state according to the backward migration process.

Let

$$A_N(\tau) \equiv X_1(\tau) + \cdots + X_L(\tau)$$

denote the number of ancestors at step τ for the configuration process with population size N. We stop the process when $A_N(\tau)$ reaches 1, corresponding to the most recent common ancestor of the sample. We want to show that the speeded up process $A_N(\lfloor Nt \rfloor)$ converges to a time-changed version of the standard coalescent, as $N \to \infty$. Thus we will be interested in the behavior of the speeded up configuration process $\mathbf{X}_N(\lfloor Nt \rfloor)$. The collapsing of states alluded to above means that, in the limit, we only need the sum of the components of the configuration process; *i.e.*, we collapse the configuration process. We will describe next how to get the limiting coalescent. This discussion will be kept at a heuristic level in this section so as to aid the understanding of the basic ideas without getting bogged down in technical details. A step-by-step proof of weak convergence will be given in the appendix.

To get coalescence rates in the limiting ancestral process, we will average coalescence probabilities according to the stationary distribution in each level. To this end, let $h(\mathbf{x}) = h(x_1, \ldots, x_L)$ denote the probability that, when the current configuration is $\mathbf{x} \in S$, a single pair of ancestors coalesces one generation back. Due to the Wright–Fisher reproduction dynamics, the one-step *coalescence probabilities* will have the form

$$h(\mathbf{x}) = N^{-1} H(\mathbf{x}) + O(N^{-2}) \tag{7}$$

for all $\mathbf{x} \in S$. Thus, in the continuous-time coalescent, with one unit of time corresponding to N generations, $H(\mathbf{x})$ will be the coalescence rate when the ancestor process is in configuration $\mathbf{x}$. Since migration probabilities are of order $O(N^{-\alpha})$, the probability of simultaneous migration and coalescence in the same generation is order $O(N^{-1-\alpha})$, which will be negligible in the coalescent limit, unless $\alpha = 0$. Thus, when $\alpha > 0$, the configuration process will remain in a given state $\mathbf{x}$ for $O(N^{\alpha})$ generations. During this time, the coalescence probability remains the same each generation. If migration, say to configuration $\mathbf{y}$, occurs before coalescence, then the coalescence probability changes to $h(\mathbf{y})$, etc. The key is that, when $\alpha < 1$, there will be many (in fact, $O(N^{1-\alpha})$) migration jumps on a given level before a coalescence occurs. This gives the configuration process on a given level time to reach equilibrium between coalescence events.

Two basic ingredients are needed for calculating the effective coalescence rates via the aforementioned averaging. First, the strong law of large numbers for additive functionals of a Markov chain (*cf.* Durrett, 1996, p. 323) yields

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} H\left(\mathbf{X}^{(r)}(i)\right) = \sum_{\mathbf{y}\in S_r} \pi_r(\mathbf{y}) H(\mathbf{y}) \tag{8}$$

with probability one, for any initial distribution on $\mathbf{X}^{(r)}$. Next, we need a slight generalization of the exponential limit in equation (1). If $\{a_N(k) : N \geq 1, k = 1, \ldots, N\}$ is a doubly-indexed array for which $a \equiv \lim_{N\to\infty} \sum_{k=1}^{N} a_N(k)$ is finite, and such that

$$\max_{1\leq k\leq N} a_N(k) \to 0, \text{ as } N \to \infty,$$

then

$$\lim_{N\to\infty} \prod_{k=1}^{N} (1 - a_N(k)) = e^{-a}. \tag{9}$$

A typical example has $a_N(k) = a(k)/N$, where the sequence $a(k)$ is bounded, non-negative, and Cesáro summable.

Given that the configuration process is currently on level r, the probability that there are no coalescences for $\lfloor Nt \rfloor$ generations is

$$\mathbb{P}(\text{no coalescences for } \lfloor Nt \rfloor \text{ generations}) = \prod_{\tau=1}^{\lfloor Nt \rfloor} \left[1 - h(\mathbf{X}^{(r)}(\tau))\right] \to e^{-c_r t}, \tag{10}$$

as $N \to \infty$, where

$$c_r \equiv \lim_{N\to\infty} \sum_{\tau=1}^{N} h(\mathbf{X}^{(r)}(\tau)) = \lim_{N\to\infty} \frac{1}{N} \sum_{\tau=1}^{N} H(\mathbf{X}^{(r)}(\tau)). \tag{11}$$

This suggests that the time it takes (on the coalescent time scale) to get a coalescence when there are r ancestors is $T(r) \sim \text{Exp }(c_r)$. Thus, it is natural to call c_r the *effective coalescence rate* for level r. By (8), we can express this in an intuitively pleasing way as

$$c_r = \sum_{\mathbf{y} \in S_r} \pi_r(\mathbf{y}) H(\mathbf{y}). \tag{12}$$

Intuitively, to get the effective coalescence rate on a given level, one simply averages the coalescence rates for the different configurations according to the stationary distribution for that level.

We will see later that there is a constant c such that $c_r = c\binom{r}{2}$ for each r. Thus c will give the (constant) change in time scale relative to the standard coalescent.

The above limit, which took us from the discrete-time setting to the continuous-time setting, is heuristic and also does not give us the full weak convergence to the coalescent. (Weak convergence is stronger that convergence at a finite number of times, and is needed if one wants to treat events that depend on an infinite number of times.) This will be remedied in the Appendix. We will also show how to compute the value of c_r.

Decomposing the transition matrix. Order the states in S so that level 1 states occur first,..., level n states occur last, the ordering of states within a level being arbitrary but fixed. With this ordering, let $\mathbf{\Pi}_N = (\mathbf{\Pi}_N(\mathbf{x}, \mathbf{y}))_{\mathbf{x},\mathbf{y} \in S}$ be the one-step transition probability matrix for the configuration process $\mathbf{X}(k)$ when the population size is N. To understand the behavior of the speeded up configuration process, we will need to consider the asymptotics of $\mathbf{\Pi}_N^{\lfloor Nt \rfloor}$ as $N \to \infty$. We begin by decomposing Π_N as follows. For $0 \le \alpha \le 1$,

$$\mathbf{\Pi}_N = \mathbf{I} + \frac{1}{N^\alpha}\mathbf{B} + \frac{1}{N}\mathbf{C} + o\left(\frac{1}{N}\right), \tag{13}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{B}$ is a block diagonal matrix of the form

$$\mathbf{B} = \begin{bmatrix} B_{11} & 0 & 0 & \cdots & 0 & 0 \\ 0 & B_{22} & 0 & \cdots & 0 & 0 \\ . & . & . & \cdots & . & . \\ & & & \ddots & & \\ . & . & . & & . & . \\ 0 & 0 & 0 & \cdots & B_{n-1,n-1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & B_{n,n} \end{bmatrix}$$

which arises from backward migration jumps, and $\mathbf{C}$ is a block matrix of the form

$$\mathbf{C} = \begin{bmatrix} -C_{11} & 0 & 0 & \cdots & 0 & 0 & 0 \\ C_{21} & -C_{22} & 0 & \cdots & 0 & 0 & 0 \\ 0 & . & . & \cdots & . & . & . \\ & & & \ddots & & & \\ . & . & . & & . & . & . \\ 0 & 0 & 0 & \cdots & C_{n-1,n-2} & -C_{n-1,n-1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & C_{n,n-1} & -C_{n,n} \end{bmatrix}$$

which arises due to coalescence jumps. In the block matrices $\mathbf{B}$ and $\mathbf{C}$, each 0 denotes a zero matrix of the appropriate size. The above decomposition is easy when the state space of the configuration process does not depend on N, as will always be the case in this contribution.

Example. Here's a simple example to illustrate the structure of the matrix $\mathbf{\Pi}_N$ (see also Nordborg, 1997). Consider a sample of size $n = 2$ from a 2-deme model in which deme 1 has $N_1 = a_1 N$ individuals and deme 2 has $N_2 = a_2 N$. The configuration space is given by

$$S = \{(1,0),(0,1),(2,0),(1,1),(0,2)\},$$

where we have ordered the configurations so that the two level-1 configurations appear first and the three level-2 configurations appear last; other than that, the order is not important. Let the scaling exponent for the backward migration probabilities be α; *i.e.*, the backward migration probabilities have the form $b_{ij} = \beta_{ij}/N^\alpha$. If $0 < \alpha < 1$, with the above ordering of configurations, then

$$\mathbf{\Pi}_N = \begin{bmatrix} 1-\frac{\beta_{12}}{N^\alpha} & \frac{\beta_{12}}{N^\alpha} & 0 & 0 & 0 \\ \frac{\beta_{21}}{N^\alpha} & 1-\frac{\beta_{21}}{N^\alpha} & 0 & 0 & 0 \\ \frac{1}{a_1 N} & 0 & 1-\frac{2\beta_{12}}{N^\alpha}-\frac{1}{a_1 N} & \frac{2\beta_{12}}{N^\alpha} & 0 \\ 0 & 0 & \frac{\beta_{21}}{N^\alpha} & 1-\frac{\beta_{21}+\beta_{12}}{N^\alpha} & \frac{\beta_{12}}{N^\alpha} \\ 0 & \frac{1}{a_2 N} & 0 & \frac{2\beta_{21}}{N^\alpha} & 1-\frac{2\beta_{21}}{N^\alpha}-\frac{1}{a_2 N} \end{bmatrix}.$$

For example, the first row indicates that if the current configuration is $(1,0)$, then in the next generation back in time it will be $(1,0)$ with probability $1 - \frac{\beta_{12}}{N^\alpha}$, and it will be $(0,1)$ with probability $\frac{\beta_{12}}{N^\alpha}$. The other terms in the first row are zero since the configuration process cannot jump from level 1 to level 2. Note that terms of order $o(1/N)$ have been ignored. These ignored terms are multiples of $1/N^2$, $1/N^{1+\alpha}$, etc., corresponding to multiple coalescences, simultaneous coalescence and migration, etc., and will have no effect on the limiting coalescent. In this example and in (13), we have also ignored terms of order $O(1/N^{2\alpha})$ and other powers of $1/N^\alpha$ which arise when two or more ancestral lines try to migrate in the same generation. These will be very rare compared to order $O(1/N^\alpha)$ migration events and will not affect the limit, even though they can occur many times between coalescence events. Recall that we are assuming that all fast-migration probabilities have the same scaling exponent α. The important thing to keep in mind is that order $O(1/N^\alpha)$ events will determine the stationary distribution of the backward configuration process, and order $O(1/N)$ events are the only ones whose details make a mark on the limiting coalescent. This is the reason multiple migration events can be ignored when $0 < \alpha < 1$.

In the above example, the block matrices making up $\mathbf{B}$ and $\mathbf{C}$ are

$$B_{11} = \begin{bmatrix} -\beta_{12} & \beta_{12} \\ \beta_{21} & -\beta_{21} \end{bmatrix}, \quad B_{22} = \begin{bmatrix} -2\beta_{12} & 2\beta_{12} & 0 \\ \beta_{21} & -(\beta_{12}+\beta_{21}) & \beta_{12} \\ 0 & 2\beta_{21} & -2\beta_{21} \end{bmatrix},$$

$$C_{21} = \begin{bmatrix} 1/a_1 & 0 \\ 0 & 0 \\ 0 & 1/a_2 \end{bmatrix}, \qquad -C_{22} = \begin{bmatrix} -1/a_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1/a_2 \end{bmatrix},$$

and C_{11} and C_{12} are zero matrices of sizes 2×2 and 2×3, respectively.

In the special case of $\alpha = 0$, in which migration out of any deme occurs after a finite number of steps (independent of N), the above decomposition simplifies considerably and there typically will be multiple migrations in the same generation. This case is easier and will be treated separately below.

In the other extreme case, $\alpha = 1$, (13) also simplifies since the migration events happen on the same time scale as coalescence events. Here, (13) takes the form

$$\mathbf{\Pi}_N = \mathbf{I} + \frac{1}{N}\mathbf{C} + o\left(\frac{1}{N}\right) \tag{14}$$

where the order $1/N$ term contains elements corresponding to both migration and coalescence. This case, for which there is no collapsing of states, has already been treated by Herbots (1997) and others. The limiting structured coalescent retains all the original structure. Since our objective here is to study models for which there is some collapse, we will not consider the case $\alpha = 1$ further. Our focus will thus be on the first two cases, in which migration is fast compared to coalescence. Both of these cases will lead to a full collapse of microstates and the resulting coalescent will be obtained by a linear time change of the standard coalescent. (We will treat partial collapsing of states in the next subsection.) As we will see, the cases $\alpha = 0$ and $0 < \alpha < 1$ require slightly different treatments.

Calculating the coalescence rate. In general, if we denote the size of level k by $d_k = |S_k|$, then $B_{k,k}$ and $C_{k,k}$ must be $d_k \times d_k$ matrices, and $C_{k,k-1}$ must be a $d_k \times d_{k-1}$ matrix. The overall matrix $\mathbf{\Pi}_N$ is $d \times d$, where

$$d \equiv |S| = \sum_{k=1}^{n} d_k$$

is the size of the configuration space.

We now commence with a careful analysis of the case $\alpha = 0$. The treatment of the case $\alpha > 0$, which is a bit more complicated, will be done later and kept at a more intuitive level.

Case 1 ($\alpha = 0$): In this case, (13) has the form

$$\mathbf{\Pi}_N = \mathbf{A} + \frac{1}{N}\mathbf{C} + O\left(\frac{1}{N^2}\right), \tag{15}$$

where $\mathbf{A} = \mathbf{I} + \mathbf{B}$ is the diagonal block matrix obtained by combining the two constant matrices $\mathbf{I}$ and $\mathbf{B}$. In other words, $\mathbf{A}$ is the constant part of $\mathbf{\Pi}_N$ and has the form

$$\mathbf{A} = \begin{bmatrix} A_{11} & 0 & 0 & \cdots & 0 & 0 \\ 0 & A_{22} & 0 & \cdots & 0 & 0 \\ . & . & . & \cdots & . & . \\ . & . & . & \ddots & . & . \\ 0 & 0 & 0 & \cdots & A_{n-1,n-1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & A_{n,n} \end{bmatrix},$$

where $A_{i,i}$ is the one-step transition matrix for the level-i configuration process (*i.e.*, ignoring coalescence events). As before, the matrix $\mathbf{C}$ contains terms relating to single coalescence events. The $O(N^{-2})$ term includes rare events, like two or more coalescences in a single step, which can be ignored in the limit as $N \to \infty$.

We now compute the coalescence rate c_r in level r. We begin by computing the probability of a single coalescence event one generation back in time when the current configuration is $\mathbf{x} = (x_1, \ldots, x_L)$. Breaking this down according to whether the merging ancestors came from the same or different demes, we get

$$\begin{aligned} h(\mathbf{x}) &\equiv \mathbb{P}(\text{coalescence event at next step} \,|\, \mathbf{X} = (x_1, \ldots, x_L)) \\ &= \sum_{k=1}^{L} \frac{1}{a_k N} \left(\sum_{i=1}^{L} \binom{x_i}{2} b_{ik}^2 + \sum_{i=1}^{L} \sum_{j<i} x_i x_j b_{ik} b_{jk} \right) + O(\frac{1}{N^2}). \end{aligned} \tag{16}$$

Multiplying this probability by N and letting $N \to \infty$, gives the coalescence *rate* from configuration $\mathbf{x}$ on the coalescent time scale:

$$H(\mathbf{x}) = \sum_{k=1}^{L} \frac{1}{a_k} \left(\sum_{i=1}^{L} \binom{x_i}{2} b_{ik}^2 + \sum_{i=1}^{L} \sum_{j<i} x_i x_j b_{ik} b_{jk} \right). \tag{17}$$

(This last step is heuristic. It is a good heuristic, however, as we will demonstrate in the Appendix.)

Next, we must weight these rates by the multinomial stationary probabilities for the level-r configuration process given by (6). Thus, if we let $\mathbf{X} =$

$(X_1, \dots, X_L)$ denote a random vector distributed as π_r, the effective coalescence rate in level r is given by

$$\begin{aligned}
c_r &= \sum_{\mathbf{x} \in S_r} \pi_r(x_1, \dots, x_L) H(x_1, \dots, x_L) \equiv \mathbb{E}\left[H(X_1, \dots, X_L)\right] \\
&= \sum_{k=1}^{L} \frac{1}{a_k} \left(\sum_{i=1}^{L} b_{ik}^2 \mathbb{E}\binom{X_i}{2} + \sum_{i=1}^{L} \sum_{j<i} b_{ik} b_{jk} \mathbb{E} X_i X_j \right) \\
&= \sum_{k=1}^{L} \frac{1}{a_k} \left(\sum_{i=1}^{L} b_{ik}^2 \binom{r}{2} \gamma_i^2 + \sum_{i=1}^{L} \sum_{j<i} b_{ik} b_{jk} 2 \binom{r}{2} \gamma_i \gamma_j \right) \\
&= \binom{r}{2} \sum_{k=1}^{L} \frac{1}{a_k} \left(\sum_{i=1}^{L} \sum_{j=1}^{L} \gamma_i b_{ik} \gamma_j b_{jk} \right) \\
&= \binom{r}{2} \sum_{k=1}^{L} \frac{1}{a_k} \gamma_k^2 \\
&\equiv c \binom{r}{2},
\end{aligned} \tag{18}$$

where the second to last equality follows from (5), and

$$c \equiv \sum_{k=1}^{L} \frac{1}{a_k} \gamma_k^2. \tag{19}$$

Note that the pairwise coalescence rate c is independent of r and gives the change in time scale for the coalescent. We refer the reader to the Appendix for the arguments that make the above calculations rigorous, as well as for the proof of weak convergence to the coalescent process.

Case 2 ($0 < \alpha < 1$): We now show how to handle the case $0 < \alpha < 1$. First, we need the stationary distribution $\bar{\gamma} = (\gamma_1, \dots, \gamma_L)$ for the location of a single ancestral lineage. Using the backward migration probabilities, we see that the one-step backward migration transition probability matrix (ignoring coalescences) is an $L \times L$ matrix given by

$$I + \frac{1}{N^\alpha} R$$

where $R_{ii} = -\sum_{j \neq i} \beta_{ij}$, and $R_{ij} = \beta_{ij}$ when $i \neq j$. The stationary distribution satisfies $\bar{\gamma}(I + N^{-\alpha} R) = \bar{\gamma}$, or equivalently

$$\bar{\gamma} R = 0, \tag{20}$$

as well as $\gamma_1 + \cdots + \gamma_L = 1$. Thus, $\bar{\gamma}$ is also the stationary distribution for the continuous-time Markov chain with infinitesimal generator matrix R; note that it does not depend on the specific value of α. As before, once we find $\bar{\gamma}$ satisfying (20), the stationary distribution for the level-r configuration process will be $\pi_r \sim \text{Mult}(r; \gamma_1, \ldots, \gamma_L)$. Since the mean time between migration events is order $O(N^\alpha)$, which is much smaller than the $O(N)$ time it takes between coalescence events, it should be clear that the configuration process will have time to reach equilibrium on each level before another coalescence occurs.

Thus, our approach is basically the same as before. Begin with the one-step coalescence probability from configuration $\mathbf{x} = (x_1, \ldots, x_L)$. Using the scaling $b_{ij} = N^{-\alpha}\beta_{ij}$ for $i \neq j$, and $b_{ii} = 1 - N^{-\alpha}\sum_{j\neq i}\beta_{ij}$, a calculation similar to the one which led to (16) yields

$$\begin{aligned} h(\mathbf{x}) &\equiv \mathbb{P}(\text{coalescence event at next step} \,|\, \mathbf{X} = (x_1, \ldots, x_m)) \\ &= \frac{1}{N}\sum_{k=1}^{L}\frac{1}{a_k}\binom{x_k}{2} + O\left(\frac{1}{N^{1+\alpha}}\right), \end{aligned}$$

where the $O(N^{-(1+\alpha)})$ term includes probabilities of things like simultaneous migration and coalescence.

Multiplying this by N and letting $N \to \infty$, gives the coalescence *rate* from configuration $\mathbf{x}$, on the coalescent time scale:

$$H(\mathbf{x}) = \sum_{k=1}^{L}\frac{1}{a_k}\binom{x_k}{2}. \tag{22}$$

Now weight these rates by the multinomial stationary probabilities for the level-r configuration process with γ determined by (20). Thus, if we let $\mathbf{X} = (X_1, \ldots, X_L) \sim \text{Mult}(r; \gamma_1, \ldots, \gamma_L)$, the effective coalescence rate in level r is given by

$$\begin{aligned} c_r &\equiv \sum_{\mathbf{x}\in S_r}\pi_r(x_1, \ldots, x_L)H(x_1, \ldots, x_L) \\ &= \sum_{k=1}^{L}\frac{1}{a_k}\mathbb{E}\binom{X_k}{2} = \sum_{k=1}^{L}\frac{1}{a_k}\binom{r}{2}\gamma_k^2 \\ &\equiv c\binom{r}{2}, \end{aligned}$$

where c is given by equation (19). Combining cases 1 and 2, we see that the effective coalescence rate, when all states are connected by fast migration with the same scaling exponent $\alpha < 1$, does not depend on the specific value of α.

2.3 Partial collapse and the structured coalescent

Suppose the forward-time model is a structured Wright–Fisher model with two different time scales for migration, fast and slow. This complicates matters a bit, but the basic ideas are still the same. We begin by changing the notation for the individual state space to

$$\mathcal{G} = \{g_{11}, \ldots, g_{1k(1)}, g_{21}, \ldots, g_{2k(2)}, \ldots, g_{\ell 1}, \ldots, g_{\ell k(\ell)}\},$$

where, for $i = 1, \ldots, \ell$, we think of $\{g_{i1}, g_{i2}, \ldots, g_{ik(i)}\}$ as the ith cluster of microstates. The microstates in a cluster are connected by fast migration, and there is no fast migration to any other cluster (only slow migration is allowed between clusters). In the limit as $N \to \infty$, the ith cluster of microstates will collapse to a single macrostate which we denote by g_i. We are assuming here that the number of macrostates ℓ is greater than one. Thus the collection of macrostates is $\{g_1, g_2, \ldots, g_\ell\}$. In this setting, the limiting ancestral process will be a structured coalescent with ℓ patches. Note that the total number of microstates is $L = k(1) + \cdots + k(\ell)$.

The forward-time model would specify things like $m_{(i_1 j_1),(i_2 j_2)}$, the probability that an individual currently in state $g_{i_1 j_1}$ moves to state $g_{i_2 j_2}$ in the next generation. Typically, some of these would scale like $1/N^\alpha$ (fast migration) and some would scale like $1/N$ (slow migration), where $0 \le \alpha < 1$ is fixed.

Let us suppose that the corresponding backward migration probabilities are of the form

$$b_{(i_1 j_1),(i_2 j_2)} = \begin{cases} \dfrac{1}{N}\beta_{(i_1 j_1),(i_2 j_2)}, & \text{if } i_1 \neq i_2 \text{ (slow migration)}, \\[2ex] \dfrac{1}{N^\alpha}\beta_{(i_1 j_1),(i_2 j_2)}, & \text{if } i_1 = i_2 \text{ (fast migration)}. \end{cases} \tag{24}$$

To keep things simple, we again assume that these backwards probabilities are scaling symmetric overall, and that they are irreducible and aperiodic within clusters. If slow migration between clusters has only one irreducible recurrent class, then all lineages will eventually meet. We assume further that the population size in each subdeme is constant; $N_{ij} = a_{ij} N$ individuals in subdeme j of cluster i.

The simplest way to construct the one-step transition matrix $\mathbf{\Pi}_N$ for the backwards configuration process is to first order the configurations so that level 1 comes first, level 2 second, etc. Then, within each level, put together the configurations which have fixed numbers of ancestors in the different clusters. Once this partial ordering is done, one can form the backwards transition matrix as before. It will have the form

$$\mathbf{\Pi}_N = \mathbf{I} + \frac{1}{N^\alpha}\mathbf{B} + \frac{1}{N}\mathbf{C} + o\left(\frac{1}{N}\right), \tag{25}$$

which is similar to (13), except that now the matrix $\mathbf{C}$ contains terms corresponding to slow migration in addition to coalescences. The matrix $\mathbf{B}$ contains all the terms corresponding to fast migration.

The backwards configuration process now looks like

$$\mathbf{X}(\tau) = (X_{11}(\tau), \ldots, X_{1k(1)}(\tau), \ldots, X_{\ell 1}(\tau), \ldots, X_{\ell k(\ell)}(\tau)),$$

with $X_{ij}(\tau)$ denoting the number of ancestors in state g_{ij} at time τ generations in the past. It is necessary to decompose level r of the configuration space according to how many of the r ancestral lineages reside in each of the ℓ clusters. The exact configuration within a cluster will not be important (in the limit) since movement of ancestral lineages within a fast cluster will reach equilibrium before there is a coalescence event or a migration event involving a lineage moving from one cluster to another. Thus, the essential information is contained in a "level vector" $\langle r_1, \ldots, r_\ell \rangle$, in which r_i denotes the level (*i.e.*, the number of lineages) in the ith cluster. Denote configurations by

$$\mathbf{x} = (\overline{x}^1, ..., \overline{x}^\ell),$$

where $\overline{x}^j = (x_1^j, \ldots, x_{k(j)}^j)$, $j = 1, \ldots, \ell$. We will write $\mathbf{x} \sim \langle r_1, \ldots, r_\ell \rangle$ when this configuration is compatible with the level vector in the sense that $r_j = |\overline{x}^j| = x_1^j + \cdots + x_{k(j)}^j$.

To describe the limiting structured coalescent, we now need to specify the coalescence rates within each cluster as well as the migration rates between clusters. We will do this for the case $\alpha = 0$, leaving the treatment of $0 < \alpha < 1$ to the reader. Begin by computing the stationary distribution $\bar{\gamma}^{(i)} = (\gamma_1^{(i)}, \ldots, \gamma_{k(i)}^{(i)})$ for a single lineage moving within cluster i (conditioned on not leaving). This stationary distribution exists because of the assumption of irreducibility and aperiodicity within a cluster. Then, as in the previous subsection, the fact that lineages migrate independently implies that the stationary distribution for the r_i lineages within the ith cluster is multinomial:

$$\pi_{r_i}^{(i)} \sim \mathrm{Mult}(r_i; \gamma_1^{(i)}, \ldots, \gamma_{k(i)}^{(i)}). \tag{26}$$

Since the lineages in different clusters migrate independently, the overall stationary distribution for the configuration process while in level $\langle r_1, \ldots, r_\ell \rangle$ is a product of multinomials:

$$\pi_{\langle r_1, \ldots, r_\ell \rangle} \sim \prod_{i=1}^{\ell} \mathrm{Mult}(r_i; \gamma_1^{(i)}, \ldots, \gamma_{k(i)}^{(i)}). \tag{27}$$

Thus, if we start with configuration $\mathbf{x} \sim \langle r_1, \ldots, r_\ell \rangle$, the configuration process will quickly reach equilibrium (before there is a coalescence or a migration from one cluster to another) and the probability of finding the configuration process in state $\mathbf{y} \sim \langle r_1, \ldots, r_\ell \rangle$ is

$$\pi_{\langle r_1, \ldots, r_\ell \rangle}(\mathbf{y}) = \prod_{i=1}^{\ell} \frac{r_i!}{y_1^i! \cdots y_{k(i)}^i!} (\gamma_1^{(i)})^{y_1^i} \cdots (\gamma_{k(i)}^{(i)})^{y_{k(i)}^i}.$$

The pairwise coalescence rate within cluster i is computed as in (19):

$$c^{(i)} = \sum_{j=1}^{k(i)} \frac{1}{a_{ij}} (\gamma_j^{(i)})^2. \tag{28}$$

To get the migration rates, first note that the probability of exactly one ancestor migrating from cluster i_1 to cluster i_2, given the current configuration is $\mathbf{X} = (\overline{x}^1, ..., \overline{x}^\ell)$, is

$$\begin{aligned} &P\big(\text{migration from cluster } i_1 \text{ to cluster } i_2 | \mathbf{X} = (\overline{x}^1, ..., \overline{x}^\ell)\big) \\ &\qquad = \sum_{j_1=1}^{k(i_1)} x_{j_1}^{i_1} \sum_{j_2=1}^{k(i_2)} b_{(i_1 j_1),(i_2 j_2)} + o\big(\frac{1}{N}\big). \end{aligned}$$

Multiply by N and let $N \to \infty$ to get the corresponding migration *rate* when we are in the current configuration. Finally, weight these rates by the multinomial stationary probabilities given above to get the migration rate (on the coalescent time scale) from cluster i_1 to cluster i_2 when the current level is $\langle r_1, \ldots, r_\ell \rangle$:

$$\mu_{i_1,i_2} = r_{i_1} \sum_{j_1=1}^{k(i_1)} \gamma_{j_1}^{(i_1)} \sum_{j_2=1}^{k(i_2)} \beta_{(i_1 j_1),(i_2 j_2)}. \tag{29}$$

In summary, the rescaled ancestral process for the above model converges to a structured coalescent with ℓ patches. If the current state is given by $\langle r_1, \ldots, r_\ell \rangle$ (*i.e.*, r_i ancestors in patch i), then the transition rates are given by:

$$\langle r_1, \ldots, r_\ell \rangle \to \begin{cases} \langle r_1, \ldots, r_\ell \rangle - \mathbf{e}_i, & \text{at rate } c^{(i)} \binom{r_i}{2}, \\ \langle r_1, \ldots, r_\ell \rangle - \mathbf{e}_{i_1} + \mathbf{e}_{i_2}, & \text{at rate } \mu_{i_1,i_2}, \end{cases} \tag{30}$$

where $\mathbf{e}_i$ denotes the unit vector with a 1 in the ith coordinate, and we interpret $\binom{r_i}{2}$ to be 0 if $r_i < 2$.

3 Further examples

In this section, we apply the theoretical ideas of the last section to some rather different cases of population structure.

3.1 Diploidy

The standard coalescent is developed in a haploid framework, and thus ignores any complications of diploidy; it is usually assumed that these can be taken into account by simply changing the population size from N to $2N$. It will become clear from what follows, why, and in what sense, this simplification is be justified. We begin by looking at hermaphroditism, and then turn to separate sexes.

3.1.1 Partially selfing hermaphrodites

The key to modeling diploid populations is the realization that a diploid population of size N can be thought of as a haploid population of size $2N$, divided into N patches of size 2 (*cf.* Nordborg and Donnelly, 1997; Möhle, 1998).

In the case of partial selfing, whenever a pair of lineages "coalesces" into the same individual, there are two possibilities: either the two lineages pick the same of the 2 available (haploid) parents, or they pick different ones. The former event, which occurs with probability $1/2$, results in a real coalescence, whereas the latter event, which also occurs with probability $1/2$, simply results in two distinct lineages temporarily occupying the same individual. Let S be the probability that a fertilization occurs through *selfing* (self-fertilization), and $1 - S$ the probability that it occurs through *outcrossing* (mating with another individual). If the individual harboring two distinct lineages was produced through selfing, then the two lineages must have come from the same individual in the previous generation, and again pick different parents with probability $1/2$ or coalesce with probability $1/2$. If the individual was produced through outcrossing, the two lineages revert to occupying distinct individuals.

At first glance, it is not clear how to cast the selfing model in the above migration framework. However, after reformulating the model as an equivalent urn scheme, the connection will become clear (see Möhle, 1998, for a similar approach). It should be noted that it is possible to recast the models of the previous section in terms of similar urn schemes. We leave this as a simple exercise for the interested reader (see also Kaj et al., 2001, for another example).

The Ancestral Urn. Consider N boxes (the individuals), each divided into 2 slots (the gametes). Each of the $2N$ slots can contain at most one ball (the ancestral lineages). As we move back one generation, the balls will be relocated to form a new configuration of balls in the boxes, with balls landing in the same slot coalescing to a single ball. The way the balls move will be determined by whether or not the corresponding individuals are selfing. To keep track of this, imagine having two demes, A and B, where A corresponds to individuals produced through selfing and B corresponds to individuals produced through outcrossing. Each box will be assigned to one of these demes, the assignment being randomly made each generation. The probability that a box is assigned to deme A (resp., deme B) is S (resp., $1 - S$), and these choices are made independently for different boxes. To keep track of the relevant information, say a box has type 1 if it contains one ball and type 2 if it contains two balls. For $i = 1, 2$, write x_i^A for the number of boxes in deme A with type i, and x_i^B for the number of boxes in deme B with type i. The configuration $\mathbf{x} = (x_1^A, x_2^A; x_1^B, x_2^B)$ contains all the information we need. Note that the total number of balls (ancestors) is given by $x_1^A + x_1^B + 2(x_2^A + x_2^B)$. Given the current configuration of the urn, the configuration in the next generation back in time is obtained from a sequence of 3 steps:

1. The pair of balls in each 2-box in deme A (independently of the other pairs) relocates to one of the N boxes in the previous generation, each box chosen with probability $1/N$. The two balls in this box independently choose a slot to occupy, each slot being chosen with probability $1/2$. If two balls land in the same slot, they coalesce. (For example, a 2-box in deme A will become a 1-box if both balls pick the same slot within the

new box; this happens with probability 1/2.)

2. Next, all the other balls relocate by independently choosing new slots from among all $2N$ slots, with probability $1/(2N)$ per slot. If a ball lands in a slot containing another ball (either from step 1 or step 2), the balls coalesce.

3. After all the balls have been relocated (perhaps with some coalescences), we independently re-assign the boxes to demes A and B with probabilities S and $1-S$, respectively.

After the above three steps, we get a new configuration $(x_1^A, x_2^A; x_1^B, x_2^B)$ for the urn. Note that coalescences can only occur in steps 1 and 2. Step 3 serves to determine the relocation strategy for the balls in the next generation. If steps 1 and 2 yield N boxes with i having type 1 and j having type 2, then the assignment in step 3 will give us the configuration $(x_1^A, x_2^A; x_1^B, x_2^B) = (i_A, j_A; i - i_A, j - j_A)$ with probability

$$\binom{i}{i_A} S^{i_A}(1-S)^{i-i_A} \cdot \binom{j}{j_A} S^{j_A}(1-S)^{j-j_A}.$$

It should be clear that this urn model is equivalent to the ancestral description for the above model of partial selfing. To relate this to our framework, let the configuration process

$$\mathbf{X}(\tau) = (X_1^A, X_2^A; X_1^B, X_2^B), \quad \tau = 0, 1, 2, \ldots,$$

evolve according to the above rules for the ancestral urn. This gives a discrete-time Markov chain analogous to the configuration processes we dealt with before. Note that the components change on a fast time scale (every generation, in fact), so in the limit as $N \to \infty$, the state space for the ancestral (urn) process collapses to record simply the total number of ancestors $x_1^A + x_1^B + 2(x_2^A + x_2^B)$; *i.e.*, we get full collapse to the standard coalescent with the appropriate time change. Thus, to complete the picture, we need to compute the 1-step transition matrix $\mathbf{\Pi}_N$ for the ancestral urn process, and then proceed as in the previous section.

However, we can simplify the calculations by first reducing the state space. It turns out that all we really need to keep track of at each step is the total number of balls, n, and the number of 2-boxes, x, in deme A. Thus, the states we will record at each step are of the form (n, x). In fact, it is easy to check that knowing this state for a given generation allows us to construct the complete configuration $(x_1^A, x_2^A; x_1^B, x_2^B)$ one generation back by simply using the rules of the urn model. Note that the only states that we will typically see are of the form $(n, 0)$ or $(n, 1)$; all values of $x \geq 2$ occur with probability of order $O(1/N^2)$ and can be ignored. This will become clearer in the calculations below.

If the current state is of the form $(n, 1)$, for some n, then within a few steps, either a coalescence will occur or the two balls in the deme A 2-box will split apart and the state will be $(n, 0)$. In this latter state, the coalescence probability is $O(1/N)$. Thus, we will refer to states with $x = 1$ as being in a "hot" phase and states with $x = 0$ as being in a "cool" phase.

To see how long the process stays in a cool phase, suppose the current state is $(n, 0)$ and compute the probability that no 2-boxes are formed in the next step:

$$\begin{aligned} \mathbb{P}(\text{no two balls land in same box}) &= (1 - \frac{1}{N})(1 - \frac{2}{N}) \cdots (1 - \frac{n-1}{N}) \\ &= 1 - \frac{\binom{n}{2}}{N} + O(\frac{1}{N^2}). \end{aligned} \tag{31}$$

Note that if two balls do land in the same box, they create a 2-box with probability $1/2$ and they coalesce with probability $1/2$. Thus, if the current state is $(n, 0)$,

$$\mathbb{P}(\text{a 2-box is created}) = \mathbb{P}(\text{two balls coalesce}) = \frac{1}{2}\frac{\binom{n}{2}}{N} + O(\frac{1}{N^2}).$$

All other box occupancies have probability $O(1/N^2)$. If we write T_p for the number of steps until a 2-box is created, then as $N \to \infty$,

$$\mathbb{P}(T_p > \lfloor 2Nt \rfloor) = \left(1 - \frac{1}{2}\frac{\binom{n}{2}}{N} + O(\frac{1}{N^2})\right)^{\lfloor 2Nt \rfloor} \to e^{-t\binom{n}{2}}.$$

So, on the coalescent time scale (measuring time in units of $2N$ generations), $T_p \sim \mathrm{Exp}\big(\binom{n}{2}\big)$. When a 2-box is created, it will be assigned to deme A with probability S. So, ignoring coalescences, if we are currently in a cool phase, we move to the hot phase at rate $S\binom{n}{2}$. Thus, the time spent in the cool phase (again, on the coalescent time scale) is $T_c \sim \mathrm{Exp}\big(S\binom{n}{2}\big)$.

Next, we determine what happens in the state $(n, 1)$. Let K_h denote the number of generations spent in the hot phase (ignoring coalescences) before we revert to state $(n, 0)$. Then K_h is a geometric random variable:

$$\mathbb{P}(K_h = k) = S^{k-1}S.$$

At each step in the hot phase, we get a coalescence due to selfing with probability $1/2$. Thus,

$$\begin{aligned} \mathbb{P}(\text{no coalescence}) &= \sum_{k=1}^{\infty} \mathbb{P}(\text{no coalescence}|K_h = k)\mathbb{P}(K_h = k) \\ &= \sum_{k=1}^{\infty} (\frac{1}{2})^k S^{k-1}S = \frac{1-S}{2-S} \end{aligned} \tag{32}$$

and hence $\mathbb{P}(\text{coalescence in hot phase}) = \frac{1}{2-S}$. Note that the coalescence above refers to coalescence via selfing. Also note that, to get the probability of two balls in the same box coalescing before they split, we would replace the S^{k-1}

above with S^k since in this case we would not be assuming that the 2-box is already in A. This yields

$$\mathbb{P}(\text{coalescence for a 2-box}) = \frac{S}{2-S}.$$

Next, we need coalescence rates due to relocation. If we are in the cool phase $(n, 0)$, all n balls jump independently and uniformly over the $2N$ slots. So in the next step

$$\begin{aligned}\mathbb{P}(\text{no coalescence}) &= (1-\frac{1}{2N})(1-\frac{2}{2N})\cdots(1-\frac{n-1}{2N}) \\ &= 1-\frac{\binom{n}{2}}{2N}+O(\frac{1}{N^2}).\end{aligned} \tag{33}$$

So letting $T_c(n)$ denote the number of steps until coalescence in the cool phase,

$$\mathbb{P}(T_c(n) > \lfloor 2Nt \rfloor) = \left(1-\frac{\binom{n}{2}}{2N}+O(\frac{1}{N^2})\right)^{\lfloor 2Nt \rfloor} \to e^{-t\binom{n}{2}}.$$

So on the coalescent time scale, $T_c(n) \sim \mathrm{Exp}\big(\binom{n}{2}\big)$. The hot phase $(n, 1)$ does not last long enough for such order $O(1/N)$ coalescence events to occur (in the limit as $N \to \infty$) and so we can ignore such events.

Finally, putting together the coalescence rates in the two phases, we see that, in the limiting coalescent, the total coalescence rate when there are n balls (ancestors) is

$$\binom{n}{2}\left(1+\frac{S}{2-S}\right),$$

in agreement with Nordborg and Donnelly (1997) and Möhle (1998).

3.1.2 Males and females

Next consider a diploid population that consists of N_m breeding males and N_f breeding females, so that $N = N_m + N_f$. Both N_m and N_f are assumed to be constant. The discussion will be limited to *autosomal* genes, i. e., genes that are not sex-linked. With respect to the genealogy of such genes, the total population can thought of as a haploid population of size $2N$, divided into two patches of size $2N_m$ and $2N_f$, respectively, each of which is further divided into patches of size 2, as in the previous section. Clearly, a lineage currently in a male has probability 1/2 of coming from a male in the previous generation, and probability 1/2 of coming from a female. Within a sex, all individuals are equally likely to be chosen. Define $a_m = N_m/N$ and $a_f = N_f/N$, and assume that reproduction occurs through random union of infinitely many gametes. As in Section 2, this ensures that lineages migrate independently backward in time (with the exception noted below). The model is thus more appropriate for dioecious plants or marine organisms than cows (say). Note that these assumptions are not necessary, but make the analysis much simpler.

The model turns out to be very similar to the one in Section 3.1.1, and there is an equivalent urn model. Again, it turns out that all we need to keep track of is the total number of balls, n, and the number of 2-boxes, x. Whether a ball is currently in a male or in a female is irrelevant, because, going backward in time, irrespective of whether the current box is male or female:

- each ball in a 1-box picks a random male box with probability $1/2$, and a random female box with probability $1/2$, independently of all other balls;
- each pair of balls in a 2-box picks a random male box *and* a random female box with probability 1.

Precisely as in the previous section, the only states we will typically see are of the form $(n, 0)$ ("cool" phase) and $(n, 1)$ ("hot" phase). Note that the two balls in a 2-box split apart in one generation: the hot phase is thus instantaneous on the coalescent time scale. To see how long the process stays in the cool phase, suppose the current state is $(n, 0)$ and compute the probability that no 2-boxes are formed in the previous generation. Conditioning on the number, K, of balls that land in males, we obtain

$$\begin{aligned}
&\mathbb{P}(\text{no two balls land in the same box})\\
&\qquad = \mathbb{E}[\mathbb{P}(\text{no two balls land in the same box}|K)]\\
&\qquad = \sum_{k=0}^{n}\binom{n}{k}2^{-n}\Big(1-\frac{\binom{k}{2}}{a_m N}+O\big(\frac{1}{N^2}\big)\Big)\Big(1-\frac{\binom{n-k}{2}}{a_f N}+O\big(\frac{1}{N^2}\big)\Big)\\
&\qquad = 2^{-n}\sum_{k=0}^{n}\binom{n}{k}\Big(1-\frac{\binom{k}{2}}{a_m N}-\frac{\binom{n-k}{2}}{a_f N}\Big)+O\big(\frac{1}{N^2}\big)\\
&\qquad = 1-\frac{\binom{n}{2}}{4a_m a_f N}+O\big(\frac{1}{N^2}\big),
\end{aligned} \tag{34}$$

where the final equality follows (after some algebra) after expanding the sum and equating terms with terms in the expressions for the mean and variance of a binomially distributed random variable.

If two balls do land in the same box, they create a 2-box with probability $1/2$ and they coalesce with probability $1/2$. Thus, if the current state is $(n, 0)$,

$$\mathbb{P}(\text{a 2-box is created}) = \mathbb{P}(\text{two balls coalesce}) = \frac{\binom{n}{2}}{8a_m a_f N}+O\big(\frac{1}{N^2}\big).$$

All other box occupancies have probability $O(1/N^2)$. Thus, on the coalescent time scale, if the current configuration is $(n, 0)$, coalescence occurs at rate $\binom{n}{2}(4a_m a_f)^{-1}$, and 2-box is created at the same rate $\binom{n}{2}(4a_m a_f)^{-1}$. When coalescence occurs, the process jumps to $(n-1, 0)$; when a 2-box is created, it jumps to $(n, 1)$.

Next we determine what happens in the hot phase $(n, 1)$. Consider the probability that no 2-boxes exist in the previous generation. Because the two balls in the same individual *must* choose different sexes in the previous generation, equation (34) does not apply; however, it is easy to see that the probability is of the form

$$\mathbb{P}(\text{no two balls land in the same box}) = 1 - O\big(\frac{1}{N}\big).$$

This expression should be compared with the corresponding probability for the hot phase in the selfing model, give by equation (32). The present model differs from the selfing model in that the coalescence probability in the hot phase is still $O(1/N)$. This means that the hot phase has no effect on the coalescent time scale: precisely as in the selfing model, an infinitesimally small amount of time is spent in this phase; however, unlike in the selfing model, the probability of coalescence is still infinitesimally small.

Thus, coalescence rate when there are n ancestors is

$$\binom{n}{2}\frac{1}{4a_m a_f} = \binom{n}{2}\frac{1}{4a_m(1-a_m)}, \tag{35}$$

so that the coalescence rate is the usual one of $n(n-1)/2$ when the sex ratio is even.

3.2 Stage structure

The techniques described in this chapter are eminently suited for modeling complex life-histories, including various forms of age-structure (Rousset, 1999; Charlesworth, 2001). The general idea of modeling such situations as population structure is well-established (*e.g.*, Orive, 1993, and references therein). In many (or, probably, most) cases, these models converge to the coalescent as described here (see Kaj et al., 2001, for another example). We will illustrate this using a paper chosen more or less at random from the recent literature. Yonezawa et al. (2000) formulated a demographic model for the perennial herb *Fritillaria camtschatcensis*, and estimated its parameters. The main features of their model are the following (see Figure 4). Individual plants belong to one of three stages: single-leaf non-flowering (stage 1), multi-leaf non-flowering (stage 2), and multi-leaf flowering (stage 3). The fraction of the population that is in stage i is denoted a_i, and is assumed to be constant over time. The probability that a stage-i individual will be a stage-j individual next year is m_{ij}. Individuals in all stages reproduce clonally (sexual reproduction is assumed to be negligible). Newly created individuals are always in stage 1. The number of new individuals produced per stage-i individual is f_i. Finally, it is assumed that a fraction d of stage-1 individuals (only) die. Since the a_i are constant, this implies that "migration" with respect to stages 2 and 3 must be conservative, *i.e.*, approximately as many individuals leave and enter each of these stages each year. The estimated parameters support this conclusion. It is less obvious what happens in stage 1. One might think that a large excess of newborn individuals were created each year, and that the population was kept constant through some form of density-dependent mortality, but the estimated parameters suggest that this is not the case. Instead it seems that the number of clonal offspring produced is just sufficient to compensate for the number of dead, *i.e.*, $a_1 d = \sum_i a_i f_i$. Migration (*i.e.* transfer by growth) in and out of stage 1 appears to be conservative to a good approximation.

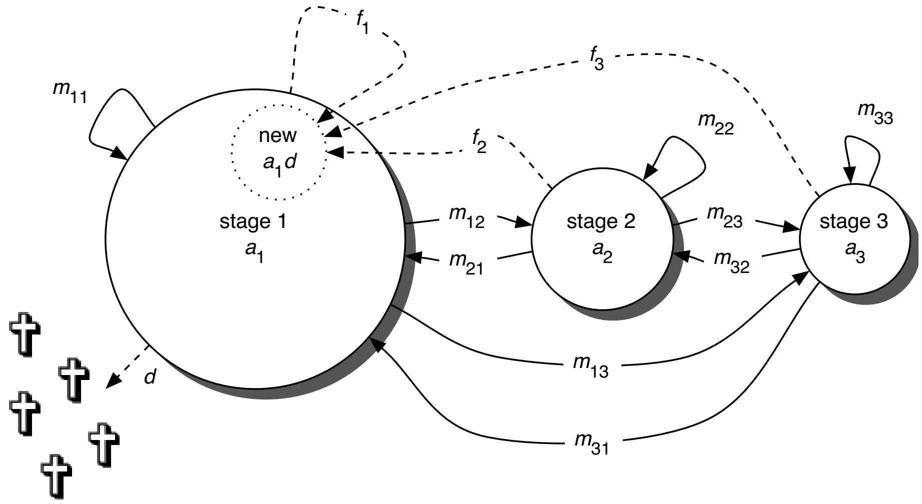

Figure 4: The *F. camtschatcensis* model of Yonezawa et al. (2000). For details and biological background, please see the original paper (note that their notation has been changed to match ours).

In order to turn the description of Yonezawa et al. (2000) into a well-specified population genetics model, it is necessary to make further assumptions about the details of the life cycle. We will assume the following. Each year consists of two phases. First, all existing individuals produce clonal propagules. Stage-i individuals produce a constant number of propagules, $Na_i f_i$; however, the individuals that produce them are chosen randomly. Second, as a result of reproduction and growth, fixed numbers of randomly chosen individuals change stages or die. The number of stage-i individuals that become stage-j individuals is $Na_i m_{ij}$, and the number of stage-1 individuals that die is $Na_1 d$. Note that

$$a_i \sum_{j \neq i} m_{ij} = \sum_{j \neq i} a_j m_{ji}, \forall i \in 1, 2, 3, \tag{36}$$

and

$$a_1 d = \sum_i a_i f_i, \tag{37}$$

so that the numbers of individuals in each stage stay constant. These assumptions avoid certain technical problems associated with modeling density regulation, and are also in agreement with the estimated parameters.

Our purpose is to derive the coalescent in this setting. Before we begin, a slight complication must be addressed. *F. camtschatcensis* reproduces clonally (via bulblets). From the point of view of tracing the genealogy of genes, this means that the two (or more if *F. camtschatcensis* is polyploid) homologous copies of a gene within an individual cannot coalesce — that is, they cannot coalesce until they have been traced back into a sexually reproducing ancestor. This is precisely the property of asexual species that was recently used to argue that bdelloid rotifers have evolved without sexual reproduction (Welch and Meselson, 2000). In the present case, it is possible that *F. camtschatcensis* reproduces sexually at an extremely low rate. If this were the case, one could consider introducing sexual reproduction that occurred on a coalescent time scale, or perhaps even slower (in which case it could be ignored from the point of view of the coalescent). We leave this as an exercise for the reader, and treat *F. camtschatcensis* as if it were haploid. That is, we follow lineages until they coalesce into the same individual (which they will then never leave) and ignore the issue of when (or, perhaps, whether) the copies within individuals coalesce.

Imagine, therefore, a population of N haploid individuals that evolves as described above. We will describe an ancestral urn model for this setting. Because coalescences necessarily involve at least one newborn individual, it is convenient to consider newborn individuals separately. Consider the population just after reproduction has occurred, but before any individual has died or grown into a different stage. Write x'_i for the number of lineages among the newborn in stage i, $i = 1, 2, 3$, and write x_i for the number of lineages among the adults. Given the configuration $\mathbf{x} = (x_1, x_2, x_3, x'_1, x'_2, x'_3)$, the configuration in the previous year is obtained as follows:

1. Newborn lineages are assigned randomly to parents in their stage. Coalescence occurs whenever two newborn lineages are assigned to the same slot (parent), *or* whenever a newborn lineage is assigned to a slot (parent) already occupied by an adult lineage. It is easy to show that the resulting coalescence probabilities in stage i are of the form

$$\frac{\binom{x_i+x'_i}{2} - \binom{x_i}{2}}{Na_i} + O\Big(\frac{1}{N^2}\Big). \tag{38}$$

   Note that all lineages are adult after this step, *i.e.*, the configuration is of the form $(x_1 + x'_1, x_2 + x'_2, x_3 + x'_3, 0, 0, 0)$ minus a random number of lineages in each of the adult classes that disappeared due to coalescence.

2. Each of the remaining lineages then "migrates," independently of all other lineages. Because all newborn individuals end up in stage 1, a stage-1 individual could have been in either of the six states of the configuration process in the previous year. The probability that it was born in stage j, $j = 1, 2, 3$, is

$$\frac{a_j f_j}{\sum_k a_k(f_k + m_{k1})} = \frac{a_j f_j}{a_1},$$

where the equality follows from equations (36)–(37), the assumption that migration and replication preserves population sizes. Similarly, the probability that it grew from stage j, $j = 1, 2, 3$, is

$$\frac{a_j m_{j1}}{\sum_k a_k(f_k + m_{k1})} = \frac{a_j m_{j1}}{a_1}.$$

Stages 2 or 3 are simpler, because individuals in these stages must have grown into them. The probability that an individual currently in stage i, $i = 2, 3$, was in stage j, $j = 1, 2, 3$, the previous year is simply $a_j m_{ji}/a_i$.

The level-r configuration process is a Markov chain where each lineage migrates back in time according to the transition matrix

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} m_{11} & \frac{a_2 m_{21}}{a_1} & \frac{a_3 m_{31}}{a_1} & f_1 & \frac{a_2 f_2}{a_1} & \frac{a_3 f_3}{a_1} \\ \frac{a_1 m_{12}}{a_2} & m_{22} & \frac{a_3 m_{32}}{a_2} & 0 & 0 & 0 \\ \frac{a_1 m_{13}}{a_3} & \frac{a_2 m_{23}}{a_3} & m_{33} & 0 & 0 & 0 \end{bmatrix}.$$

Because a finite number of adult individuals migrate, lineages do not migrate independently of each other. However, this dependence can be shown not to matter in the limit. The stationary distribution of the migration Markov chain is of the form

$$\big(q_1 - f_1 q_1, q_2 - \frac{a_2 f_2}{a_1} q_1, q_3 - \frac{a_3 f_3}{a_1} q_1, f_1 q_1, \frac{a_2 f_2}{a_1} q_1, \frac{a_3 f_3}{a_1} q_1\big), \tag{39}$$

where (q_1, q_2, q_3) is the stationary distribution corresponding to the transition matrix

$$\begin{bmatrix} m_{11} + f_1 & \frac{a_2(m_{21}+f_2)}{a_1} & \frac{a_3(m_{31}+f_3)}{a_1} \\ \frac{a_1 m_{12}}{a_2} & m_{22} & \frac{a_3 m_{32}}{a_2} \\ \frac{a_1 m_{13}}{a_3} & \frac{a_2 m_{23}}{a_3} & m_{33} \end{bmatrix},$$

which can be found using standard methods.

The coalescence rate, c, is calculated as before. From the case $n = 2$, we find that

$$c = \frac{q_1}{a_1} \sum_k f_i \big(2q_i - \frac{a_i f_i}{a_1} q_1\big). \tag{40}$$

4 Conclusion

We have shown that population structure often results in a simple linear change in the time scale of the coalescent. Especially when coupled with the fact that many different biological phenomena can be thought of in terms of population structure, this becomes a powerful modeling tool. In this section, we will first discuss how our results relate to classical concepts of an "effective population size," and, second, discuss the generality of our results, as well as possible future extensions.

4.1 Scaling constants and effective population sizes

So-called effective population sizes have a long history in population genetics (reviewed in: Ewens, 1979; Orive, 1993). The basic idea is to compare the behavior of a given model, with respect to some property, to an "ideal" model (usually the Wright–Fisher model). For example:

variance N_e: Let the frequency of an allele in generation τ be $X(\tau)$. The variance of $X(\tau+1)$, given $X(\tau)=x$, is $x(1-x)/(2N)$ in a diploid Wright–Fisher model. If the same variance in another model is σ^2/N, then

$$N_e^v \equiv \frac{x(1-x)}{2\sigma^2}N.$$

inbreeding N_e: The probability that two randomly chosen gene copies are the descendents of the same gene copy in the previous generation (*i.e.*, the probability that they are identical by descent ["ibd"]) is $1/(2N)$ in a diploid Wright–Fisher model. If the same probability in another model is $\mathbb{P}(\text{ibd})$ then

$$N_e^i \equiv \frac{1}{2\mathbb{P}(\text{ibd})}.$$

By analogy, when structured models converge to the coalescent as described in Section 2.2, it is natural to define:

coalescent N_e: The coalescent scaling in the diploid Wright–Fisher model is $2N$. If the scaling in another model is Nc, then

$$N_e^c \equiv \frac{Nc}{2}.$$

Why do we need yet another definition of N_e? The coalescent N_e is actually conceptually somewhat different from the classical definitions. The reason for this is simple: whereas the classical effective population sizes can be calculated for almost any model, not every model will converge to the coalescent. To say that a particular model has a particular coalescent N_e is to say that: first, it converges in the limit to the coalescent, and, second, it does so with a particular scaling. This is strong claim, because it means that the model will be completely indistinguishable (modulo the limit approximation) from other models that converge to the coalescent. An example will clarify what we mean. It has long been recognized that there is a relationship between inbreeding coefficients and pairwise coalescence times. $\mathbb{P}(\text{ibd})$ is also the per-generation coalescence probability, and an alternative definition of N_e^i utilizes the expected coalescence time (in generations) for a pair of genes (Slatkin, 1991; Orive, 1993). Does this mean that the inbreeding and coalescence effective population sizes will always be the same? Often, but not always: when N_e^c exists, it will be the same as N_e^i. Consider the two-deme example used in Section 2 (see also Figure 2). Irrespective of how the backward migration probabilities scale, it is possible to calculate N_e^i (for examples of this in more complicated settings, see: Nei and Takahata, 1993; Whitlock and Barton, 1997). However, unless migration is fast, the model does not converge to the coalescent, and N_e^c does not exist. We believe that this is a desirable property of an effective population size, at least if the purpose of defining an effective population size is to investigate how genetic drift operates, because a subdivided population *cannot* in general be seen as a random-mating population with a different effective population size. Even though N_e^i can be calculated, all it really tells us is that the expected coalescence time for a pair of genes is the same as in a standard model with size N_e^i — the expected coalescence time for three genes, or the variance of the coalescence time for a pair of genes will not behave similarly. The same is in general true for varying population size (an N_e^c exists only in special cases). This is of crucial importance for data analysis.

In contrast, the coalescent effective population size, when it exists, tells us everything about neutral evolution. However, it is very important to note that it does not predict the behavior of non-neutral models. For example, effective population sizes are sometimes used by conservation geneticists to make predictions about the sustainability of populations (*e.g.*, Yonezawa et al., 2000). This is not warranted. Even if neutral evolution in a structured model can be shown to operate precisely as in a standard model of some size, it does not follow that non-neutral evolution behaves analogously (whether it does or not will certainly depend on the strength of selection, *i.e.*, the scaling of the selection coefficients). To understand the dynamics of loci under selection, it is necessary to study the behavior of a model that incorporates selection.

Indeed, a strong reason for not defining any form of effective population size is that they are frequently interpreted to mean more than they do. Another reason is that the term is inevitably associated with real population sizes (this has in particular plagued arguments about human evolution) even though there is no direct relationship.

4.2 Prospects

Our intention in writing this chapter has been to convey an intuition for how genealogies in structured models may converge to the coalescent. For simplicity, we have not attempted to cover every possible case when formulating the theory (nor have we always been completely rigorous). It is clear that our results can be extended to cases involving several different time scales, migration rates that are not scaling symmetric, and so on, but the notation required to do this rigorously is rather off-putting.

It seems highly plausible that the theory presented here also carries over to the ancestral recombination graph (see Nordborg, 2000), however, no formal proof of this exists that we are aware of. The main difficulty is that the state space with recombination is no longer finite. Nevertheless, we believe that the result of Nordborg (2000) applies generally, and conjecture (informally) that models that converge to the coalescent as described in Section 2.2, will also converge to the ancestral recombination graph. The appropriate scaling for the coalescence rate will be c, and the appropriate scaling for the recombination rate will be the standard one multiplied by the increased $\mathbb{P}(\text{ibd})$ *within individuals.* Similarly, our theory should apply to the ancestral selection graph (Krone and Neuhauser, 1997; Neuhauser and Krone, 1997).

5 Appendix: Proof of weak convergence to the coalescent in the case of full collapse

We will now give a detailed proof of weak convergence to the coalescent for Case 1 ($\alpha = 0$). We follow the approach in Kaj et al. (2001).

Projection theorem. Since we want convergence to a relatively simple type of process (continuous-time pure death process), weak convergence (which is stronger than convergence of finite-dimensional distributions) should not be as complicated as for more general processes. Indeed, Theorem 2.12 on p. 173 of Ethier and Kurtz (1986) will allow us to get weak convergence to the coalescent by showing certain "collapsed semigroups" converge. We will refer to their result as the *"Projection Theorem"* for reasons that will become clear, although it does not appear in Ethier and Kurtz (1986) under that name. (In fact, what they call the projection theorem is something quite different.) As we will, this is fairly straightforward and is quite natural when the state space collapses in the sense discussed above. In their theorem, we let

- $E = \{1, ..., n\}$ be the "ancestor space,"
- $E_N = S$ for all finite N, where $S = S_1 \cup \cdots \cup S_n$ is configuration space,
- $\eta_N : S \to E$ is defined by $\eta_N(\mathbf{x}) = |\mathbf{x}| \equiv x_1 + \cdots + x_L$ for any configuration $\mathbf{x} = (x_1, ..., x_L)$. (So η_N is just the projection which maps a configuration onto the number of ancestors in that configuration.)
- $\mathbf{X}_N(k), k = 0, 1, ...$ is the discrete-time configuration process (when the population size is N),
- $A_N([Nt]) = \eta_N(\mathbf{X}_N([Nt])) = |\mathbf{X}_N([Nt])|$ is the number of ancestors in the speeded up configuration process.

Order the states in S so that level 1 states occur first,..., level n states occur last, the ordering of states within a level being arbitrary but fixed. With this ordering, let $\mathbf{\Pi}_N = (\mathbf{\Pi}_N(\mathbf{x}, \mathbf{y}))_{\mathbf{x},\mathbf{y} \in S}$ be the 1-step transition probability matrix for the configuration process $\mathbf{X}_N(k)$ when the population size is N. Finally, for any function f mapping $\{1,...,n\} \to \mathbb{R}$, set $T(t)f(i) = \mathbb{E}^i f(A(t))$, where $A(t)$ is the coalescent process defined in the introduction by the generator Q, and the superscript i on the expectation refers to the initial state.

The Projection Theorem applied to this setting says that weak convergence of $A_N([N\cdot])$ to $A(\cdot)$ will follow if we can show, for each function $f : E \to \mathbb{R}$ and each configuration $\mathbf{x}$,

$$\sum_{\mathbf{y}\in S} \mathbf{\Pi}_N^{[Nt]}(\mathbf{x},\mathbf{y}) f(|\mathbf{y}|) \to T(t)f(|\mathbf{x}|), \text{ as } N \to \infty. \tag{41}$$

Here, $\mathbf{\Pi}_N^k(\mathbf{x},\mathbf{y})$ is the $(\mathbf{x},\mathbf{y})$ term in the kth power $\mathbf{\Pi}_N^k$ of the transition matrix for the configuration process.

The expression in (41) can be simplified somewhat in the present setting. For the term on the right, we can use the relation $T(t) = e^{tQ}$ to write, for any $i \in \{1,\ldots,n\}$,

$$T(t)f(i) = \sum_{j=1}^{n} f(j) \sum_{k=0}^{\infty} \frac{t^k}{k!} q_{ij}^{(k)},$$

where $Q = (q_{ij})$ is the $n \times n$ generator matrix for the desired coalescent and $q_{ij}^{(k)}$ is the (i,j) element of Q^k. The term appearing on the left in (41) can be written as

$$\sum_{\mathbf{y}\in S} \mathbf{\Pi}_N^{[Nt]}(\mathbf{x},\mathbf{y}) f(|\mathbf{y}|) = \sum_{j=1}^{n} f(j) \sum_{\mathbf{y}\in S_j} \mathbf{\Pi}_N^{[Nt]}(\mathbf{x},\mathbf{y}).$$

Since the outer sum in each expression involves only a finite number of terms, it is enough to prove for each $i \in \{1,...,n\}$ and $\mathbf{x} \in S_i$, and each $j \in \{1,...,n\}$ that

$$\sum_{\mathbf{y}\in S_j} \mathbf{\Pi}_N^{[Nt]}(\mathbf{x},\mathbf{y}) \to \sum_{k=0}^{\infty} \frac{t^k}{k!} q_{ij}^{(k)}, \text{ as } N \to \infty. \tag{42}$$

Thus, the above convergence criterion simplifies to the following.

Theorem 1 *Suppose, for each $i \in \{1,...,n\}$ and $\mathbf{x} \in S_i$, and each $j \in \{1,...,n\}$ that*

$$\sum_{\mathbf{y}\in S_j} \mathbf{\Pi}_N^{[Nt]}(\mathbf{x},\mathbf{y}) \to \sum_{k=0}^{\infty} \frac{t^k}{k!} q_{ij}^{(k)}, \tag{43}$$

as $N \to \infty$. Then the rescaled ancestral process $\{A_N([Nt]) : t \geq 0\}$ converges weakly to the coalescent process $\{A(t) : t \geq 0\}$ defined by equation (3).

A moment's thought shows that this is quite reasonable, given our advertised goal. First of all, note that the sum on the right-hand side is just the (i,j)th element of the (semigroup) matrix e^{tQ}. Thus, this quantity gives the probability that there are j ancestors in the continuous-time coalescent at time t, given that there were i ancestors $(i \geq j)$ at time 0. The quantity on the left-hand side is the probability that the discrete-time configuration process finds itself in level j after $\lfloor Nt \rfloor$ generations, given that it was in some state x in level i in generation

0. Fast migration between coalescence events means that the particular starting configuration $\mathbf{x} \in S_i$ is not important in the limit, and this is reflected in the fact that the right-hand side depends only on i and j. The sum over S_j on the left-hand side is indicative of the anticipated collapsing of states on a given level and the emergence of a single effective coalescence rate arising from an average of coalescence probabilities in that level.

The following result from Möhle (1998) is ideally suited to Case 1 (as well as some slightly more general models) when we want to compute the limit in (43). Recall the structure of Π_N given in equation (15).

Lemma 1 (Möhle's Lemma)

$$\lim_{N\to\infty}\left(\mathbf{A}+\frac{1}{N}\mathbf{C}+o\left(\frac{1}{N}\right)\right)^{[Nt]} = \mathbf{P}-\mathbf{I}+e^{t\mathbf{G}}, \tag{44}$$

where $\mathbf{P}$ *is the matrix* $\mathbf{P} \equiv \lim_{k\to\infty}\mathbf{A}^k$, $\mathbf{I}$ *is the identity matrix, and* $\mathbf{G} \equiv \mathbf{PCP}$.

Here, all matrices are assumed to be of the same size, $d \times d$, where d is the size of the configuration space (*cf.* Section 2.2). It is easy to see that $\mathbf{P}$ will be a block diagonal matrix whose ith block $P_{i,i}$ consists of identical rows given by the stationary distribution π_i of the level-i configuration process. It also follows from the definition of $\mathbf{P}$ that it is a projection matrix (*i.e.*, $\mathbf{P}^2 = \mathbf{P}$). After a suitable collapsing of states, we will see that $\mathbf{G}$ leads to the generator of the coalescent process.

Combining Möhle's Lemma with Theorem 1, we see that the desired limit is

$$\lim_{N\to\infty}\sum_{\mathbf{y}\in S_j}\mathbf{\Pi}_N^{[Nt]}(\mathbf{x},\mathbf{y}) = \sum_{\mathbf{y}\in S_j}(\mathbf{P}-\mathbf{I}+e^{t\mathbf{G}})(\mathbf{x},\mathbf{y}) = \sum_{\mathbf{y}\in S_j}(e^{t\mathbf{G}})(\mathbf{x},\mathbf{y}).$$

(Of course, the only time these quantities are nonzero is when $i \geq j$.) Thus, convergence to the coalescent will follow if we can show that the last sum is equal to $(e^{tQ})_{ij}$ for all $\mathbf{x} \in S_i$. By definition,

$$\sum_{\mathbf{y}\in S_j}(e^{t\mathbf{G}})(\mathbf{x},\mathbf{y}) = \sum_{k=0}^{\infty}\frac{t^k}{k!}\sum_{\mathbf{y}\in S_j}\mathbf{G}^k(\mathbf{x},\mathbf{y}).$$

Furthermore, for any $\mathbf{x} \in S_i$,

$$\sum_{\mathbf{y}\in S_j}\mathbf{G}^k(\mathbf{x},\mathbf{y}) = \sum_{\mathbf{y}\in S_j}(\mathbf{G}^k)_{i,j}(\mathbf{x},\mathbf{y}),$$

where the quantity on the right hand side does not depend on the particular $\mathbf{x}$ but only on $|\mathbf{x}| = i$.

So, to get convergence to the coalescent in Case 1, we need to show

$$\sum_{k=0}^{\infty}\frac{t^k}{k!}\sum_{\mathbf{y}\in S_j}(\mathbf{G}^k)_{i,j}(\mathbf{x},\mathbf{y}) = \sum_{k=0}^{\infty}\frac{t^k}{k!}q_{ij}^{(k)}.$$

Equating coefficients of t^k, we see that the convergence criterion in Theorem 1 is equivalent to showing

$$\sum_{\mathbf{y}\in S_j} (\mathbf{G}^k)_{i,j}(\mathbf{x},\mathbf{y}) = q_{ij}^{(k)}, \quad \text{for all } k \geq 0, 1 \leq j \leq i \leq n, \text{ and } \mathbf{x} \in S_i. \tag{45}$$

Structure of **P** *and* **G**. Next, let us specify the structure of **P** and **G**. We know from (6) that the rows of the resulting block matrices $P_{i,i} = \lim_{k\to\infty} A_{i,i}^k$ are all the same and, moreover, that these rows are given by the probabilities for a $\mathrm{Mult}(i;\gamma_1,\ldots,\gamma_L)$ distribution. More specifically, if $y_1^{(i)}, y_2^{(i)}, \ldots, y_{d_i}^{(i)}$ are the ordered elements of S_i, then each row of $P_{i,i}$ is given by the vector

$$\mathbf{M}_i \equiv \left(\pi_i(y_1^{(i)}), \pi_i(y_2^{(i)}), \ldots, \pi_i(y_{d_i}^{(i)})\right) \tag{46}$$

where, for $\mathbf{y} = (y_1, y_2, \ldots, y_L) \in S_i$, we define the multinomial probability

$$\pi_i(\mathbf{y}) \equiv \frac{i!}{y_1! \cdots y_L!}\gamma_1^{y_1}\cdots\gamma_L^{y_L}.$$

If $\mathbf{x} \in S_i$, then the corresponding row in the matrix **C** has non-zero elements only within the sub-matrix $C_{i,i-1}$, with columns corresponding to configurations $\mathbf{y} \in S_{i-1}$, or within $C_{i,i}$, whose columns correspond to configurations $\mathbf{y} \in S_i$. By listing all possible coalescence events starting from a configuration $\mathbf{x} \in S_i$, it follows that each row sum in $C_{i,i-1}$ adds up to the quantity $H(\mathbf{x})$ defined in (7):

$$\sum_{\mathbf{y}\in S_{i-1}} C_{i,i-1}(\mathbf{x},\mathbf{y}) = H(\mathbf{x}). \tag{47}$$

The nonzero columns of $C_{i,i}$ and $C_{i,i-1}$ consist of the same vectors, even though they are not typically in the same locations within these matrices. Hence

$$\sum_{\mathbf{y}\in S_i} C_{i,i}(\mathbf{x},\mathbf{y}) = H(\mathbf{x}). \tag{48}$$

If these steps are not clear, the reader is urged to work through them in the example given in Section 2.2.

Multiplying the block matrices in the definition of **G**, we see that

$$\mathbf{G} = \begin{bmatrix} -G_{11} & 0 & 0 & \cdots & 0 & 0 \\ G_{21} & -G_{22} & 0 & \cdots & 0 & 0 \\ 0 & . & . & \cdots & . & . \\ . & . & \ddots & . & . & . \\ 0 & 0 & \cdots & G_{n-1,n-2} & -G_{n-1,n-1} & 0 \\ 0 & 0 & \cdots & 0 & G_{n,n-1} & -G_{n,n} \end{bmatrix},$$

where

$$\begin{aligned} G_{i,i} &= P_{i,i}C_{i,i}P_{i,i} \\ G_{i,i-1} &= P_{i,i}C_{i,i-1}P_{i-1,i-1}. \end{aligned}$$

These matrices can be simplified because of the special form of the factors.

By (48), (12), and the fact that all rows in $P_{i,i}$ are equal and given by $\mathbf{M}_i$ in (46), we obtain for $\mathbf{x}, \mathbf{y} \in S_i$,

$$\begin{aligned} G_{i,i}(\mathbf{x},\mathbf{y}) &= \sum_{\mathbf{u}\in S_i}\sum_{\mathbf{v}\in S_i} P_{i,i}(\mathbf{x},\mathbf{u})C_{i,i}(\mathbf{u},\mathbf{v})P_{i,i}(\mathbf{v},\mathbf{y}) \\ &= \pi_i(\mathbf{y})\sum_{\mathbf{u}\in S_i}\pi_i(\mathbf{u})H(\mathbf{u}) \\ &= c_i\pi_i(\mathbf{y}), \end{aligned}$$

where c_i is the coalescence rate in level i. Since $P_{i,i}(\mathbf{x},\mathbf{y}) = \pi_i(\mathbf{y})$ for all $\mathbf{x} \in S_i$, we conclude that

$$G_{i,i} = c_i\, P_{i,i}. \tag{49}$$

Similarly,

$$G_{i,i-1} = c_i\, \tilde{P}_{i,i-1}, \tag{50}$$

where $\tilde{P}_{i,i-1}$ is the $d_i \times d_{i-1}$ matrix with identical rows given by the probability vector $\mathbf{M}_{i-1}$ appearing in $P_{i-1,i-1}$.

The next step in establishing the convergence criterion (45) is to compute powers of $\mathbf{G}$. Write $(\mathbf{G}^k)_{i,j}$ for the (i,j) block in the matrix $\mathbf{G}^k$ and $G^k_{i,j}$ for the matrix $G_{i,j}$ raised to the power k. The matrix powers $\mathbf{G}^k$ are lower triangular block matrices given recursively by

$$\begin{aligned} (\mathbf{G}^k)_{i,i} &= (-1)^k G^k_{i,i} = (-1)^k c_i^k P^k_{i,i} = (-1)^k c_i^k P_{i,i}\,, & (51) \\ (\mathbf{G}^k)_{i,j} &= G_{i,i-1}(\mathbf{G}^{k-1})_{i-1,j} - G_{i,i}(\mathbf{G}^{k-1})_{i,j}\,, \quad j = 1, \ldots i-1, & (52) \end{aligned}$$

and, of course, $(\mathbf{G}^k)_{i,j} = 0$ when $j > i$. Note that we have used the fact that $P_{i,i}$ is a projection and hence $P^k_{i,i} = P_{i,i}$.

Remark. To assist the reader in the matrix calculations, we will use the following notational convention. When a matrix D corresponds to transitions in configuration space S, we write $D(\mathbf{x},\mathbf{y})$ for the $(\mathbf{x},\mathbf{y})$ element (with the aforementioned ordering of states). When a matrix Q corresponds to transitions in the "collapsed space" $E = \{1, ..., n\}$, we write $q_{i,j}$ for the (i,j) element. Finally, the notation $D_{i,j}(\mathbf{x},\mathbf{y})$ will represent the $(\mathbf{x},\mathbf{y})$ element of the sub-matrix $D_{i,j}$ in a larger block matrix. As a general rule, we will use boldface letters to denote the large matrices corresponding to the full set of states in S; the submatrices making up these larger matrices will be in regular type.

For $j = i$, by (49) and (51),

$$\sum_{\mathbf{y}\in S_i} (\mathbf{G}^k)_{i,i}(\mathbf{x},\mathbf{y}) = (-1)^k c_i^k \sum_{\mathbf{y}\in S_i} P_{i,i}(\mathbf{x},\mathbf{y}) = (-1)^k c_i^k. \tag{53}$$

For $j = 1, \ldots, i-1$, by (50) and (52),

$$\begin{aligned}
\sum_{\mathbf{y}\in S_j} (\mathbf{G}^k)_{i,j}(\mathbf{x},\mathbf{y}) &= \sum_{\mathbf{y}\in S_j} (G_{i,i-1}(\mathbf{G}^{k-1})_{i-1,j} - G_{i,i}(\mathbf{G}^{k-1})_{i,j})(\mathbf{x},\mathbf{y}) \\
&= c_i \sum_{\mathbf{y}\in S_j} \left(\tilde{P}_{i,i-1}(\mathbf{G}^{k-1})_{i-1,j} - P_{i,i}(\mathbf{G}^{k-1})_{i,j}\right)(\mathbf{x},\mathbf{y}) \\
&= c_i \sum_{\mathbf{y}\in S_j} (\mathbf{G}^{k-1})_{i-1,j}(\mathbf{x},\mathbf{y}) - c_i \sum_{\mathbf{y}\in S_j} (\mathbf{G}^{k-1})_{i,j}(\mathbf{x},\mathbf{y}).
\end{aligned} \tag{54}$$

The last line follows from the fact that any given block $(\mathbf{G}^{k-1})_{i,j}$ of $\mathbf{G}^{k-1}$ will have identical rows. For example, it is easy to check that

$$(\mathbf{G}^2)_{i,i} = c_i^2 P_{i,i},$$

$$(\mathbf{G}^2)_{i,i-1} = -c_i^2 P_{i,i}\tilde{P}_{i,i-1} - c_i c_{i-1}\tilde{P}_{i,i-1}P_{i-1,i-1},$$

and

$$(\mathbf{G}^2)_{i,i-2} = c_i c_{i-1}\tilde{P}_{i,i-1}\tilde{P}_{i-1,i-2},$$

and clearly each of these matrices consists of identical rows. Note also that, to avoid notationally cumbersome resizing of the matrices in the first sum of the right hand side of (54), we take $\sum_{\mathbf{y}\in S_j}(\mathbf{G}^{k-1})_{i-1,j}(\mathbf{x},\mathbf{y})$ to mean the common row sum in the $d_{i-1}\times d_j$ matrix $(\mathbf{G}^{k-1})_{i-1,j}$, which is the same for any choice of $\mathbf{x}$ in S_i.

Recursion. The last step in checking condition (45) is to show that both sides satisfy the same recursion equation. Equations (53) and (54) show that, for each k, the quantities

$$f_k(i,j) \equiv \sum_{\mathbf{y}\in S_j} \mathbf{G}^k(\mathbf{x},\mathbf{y}), \quad \mathbf{x}\in S_i$$

are well defined and satisfy the recursive system of equations

$$\begin{aligned}
f_k(i,i) &= (-1)^k c_i^k, \qquad k \geq 1 \\
f_k(i,j) &= c_i f_{k-1}(i-1,j) - c_i f_{k-1}(i,j), \quad i \geq 2,\ 1 \leq j \leq i-1, \quad k \geq 2.
\end{aligned}$$

This is the same linear system of equations as that satisfied by the elements $q_{i,j}^{(k)}$ of the matrix Q^k. By uniqueness of the solution of this system, we may therefore make the identification

$$f_k(i,j) = q_{i,j}^{(k)},$$

and hence (45) holds.

References

Charlesworth, B. (2001). The effect of life-history and mode of inheritance on neutral genetic variability. *Genet. Res., Camb.*, 77:153–166.

Donnelly, P. and Tavaré, S. (1995). Coalescents and genealogical structure under neutrality. *Annu. Rev. Genet.*, 29:401–421.

Durrett, R. (1996). *Probability: Theory and Examples.* Wadsworth Publishing Company, Belmont, California, 2nd edition.

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

Ewens, W. J. (1979). *Mathematical Population Genetics.* Springer-Verlag, Berlin.

Herbots, H. M. (1997). The structured coalescent. In Donnelly, P. and Tavaré, S., editors, *Progress in Population Genetics and Human Evolution*, pages 231–255. Springer-Verlag, New York.

Hudson, R. R. (1990). Gene genealogies and the coalescent process. In Futuyma, D. and Antonovics, J., editors, *Oxford Surveys in Evolutionary Biology*, volume 7, pages 1–43. Oxford University Press, Oxford.

Kaj, I., Krone, S. M., and Lascoux, M. (2001). Coalescent theory for seed bank models. *J. Appl. Prob.*, 38:285–301.

Kingman, J. F. C. (1982a). The coalescent. *Stochast. Proc. Appl.*, 13:235–248.

Kingman, J. F. C. (1982b). On the genealogy of large populations. In Gani, J. and Hannan, E. J., editors, *Essays in Statistical Science: Papers in Honour of P. A. P. Moran*, pages 27–43. Applied Probability Trust, Sheffield. J. Appl. Prob., special volume 19A.

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

Möhle, M. (1998). A convergence theorem for Markov chains arising in population genetics and the coalescent with selfing. *Adv. Appl. Prob.*, 30:493–512.

Nei, M. and Takahata, N. (1993). Effective population size, genetic diversity, and coalescence time in subdivided populations. *J. Mol. Evol.*, 37:240–244.

Neuhauser, C. and Krone, S. M. (1997). The genealogy of samples in models with selection. *Genetics*, 145:519–534.

Nordborg, M. (1997). Structured coalescent processes on different time scales. *Genetics*, 146:1501–1514.

Nordborg, M. (1999). The coalescent with partial selfing and balancing selection: An application of structured coalescent processes. In Seillier-Moiseiwitsch, F., editor, *Statistics in Molecular Biology and Genetics*, volume 33 of *IMS Lecture Notes-Monograph Series*, pages 56–76. Institute of Mathematical Statistics, Hayward, California.

Nordborg, M. (2000). Linkage disequilibrium, gene trees, and selfing: An ancestral recombination graph with partial self-fertilization. *Genetics*, 154:923–929.

Nordborg, M. (2001). Coalescent theory. In Balding, D. J., Bishop, M. J., and Cannings, C., editors, *Handbook of Statistical Genetics*, pages 179–212. John Wiley & Sons, Inc., Chichester, U. K.

Nordborg, M. and Donnelly, P. (1997). The coalescent process with selfing. *Genetics*, 146:1185–1195.

Notohara, M. (1990). The coalescent and the genealogical process in geographically structured populations. *J. Math. Biol.*, 29:59–75.

Orive, M. E. (1993). Effective population size in organisms with complex life-histories. *Theor. Popul. Biol.*, 44:316–340.

Rousset, F. (1999). Genetic differentiation in populations with different classes of individuals. *Theor. Pop. Biol.*, 55:297–308.

Slatkin, M. (1991). Inbreeding coefficients and coalescence times. *Genet. Res., Camb.*, 58:167–175.

Welch, D. M. and Meselson, M. (2000). Evidence for the evolution of bdelloid rotifers without sexual reproduction of genetic exchange. *Science*, 288:1211–1215.

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

Yonezawa, K., Kinoshita, E., Watano, Y., and Zentoh, H. (2000). Formulation and estimation of the effective size of stage-structured populations in *Fritillaria camtschatcensis*. *Evolution*, 54:2007–2013.