COALESCENCE THEORY FOR A GENERAL CLASS OF STRUCTURED POPULATIONS WITH FAST MIGRATION

O. HÖSSJER,* *Stockholm University*

Abstract

In this paper we study a general class of population genetic models where the total population is divided into a number of subpopulations or types. Migration between subpopulations is fast. Extending the results of Nordborg and Krone (2002) and Sagitov and Jagers (2005), we prove, as the total population size N tends to ∞, weak convergence of the joint ancestry of a given sample of haploid individuals in the Skorokhod topology towards Kingman's coalescent with a constant change of time scale c. Our framework includes age-structured models, geographically structured models, and combinations thereof. We also allow each individual to have offspring in several subpopulations, with general dependency structures between the number of offspring of various types. As a byproduct, explicit expressions for the coalescent effective population size N/c are obtained.

Keywords: Age-structured population; coalescence theory; effective population size; geographical substructure; weak convergence

2010 Mathematics Subject Classification: Primary 92D25
Secondary 60J28; 60F17

1. Introduction

When studying the joint ancestry of a sample of n individuals, the coalescent has emerged as an important tool. It was originally introduced in Kingman (1982a), (1982b) in the context of a haploid Wright–Fisher (WF) model with constant population size N and nonoverlapping generations. When $N \to \infty$ and time backwards is rescaled in units of N, the number of ancestors of the WF model converges to a continuous-time pure-death Markov process with the number of ancestors as states and intensity

$$c_a = \binom{a}{2}$$

of a transition from a to $a - 1$ for $2 \leq a \leq n$, with $a = 1$ an absorbing state. This process, frequently referred to as Kingman's coalescent, is surprisingly robust and turns up as the limit ancestral process as $N \to \infty$ for a number of more complicated population genetic models, as reviewed, e.g. in Möhle (2000). Indeed, for diploid models (see Möhle (1998b), Möhle and Sagitov (2003), and Pollak (2010)), self-fertilization and partial selfing (see Fu (1997), Nordborg and Donnelly (1997), and Möhle (1998a)), geographically structured populations with fast migration (see Nordborg and Krone (2002)), age-structured populations with overlapping generations (see Sagitov and Jagers (2005)), and populations with a rapidly

changing Markov regime population size (see Jagers and Sagitov (2004) and Pollak (2010)), it turns out that the limiting ancestral process is a time-scaled version of Kingman's coalescent with intensity

$$c_a = c \binom{a}{2}$$

of a transition from a to $a - 1$. The constant c quantifies the speed at which ancestral lines merge and, hence, the rate at which genetic drift occurs.

In this paper we study a very general class of structured population genetic models of constant size N, divided into a number of subpopulations or types of individual. These may represent geographical sites, age classes, or combinations thereof. It is assumed that migration between subpopulations occurs at a fast rate, $O(1)$, under the original time scale, as opposed to coalescence events, which occur at a slower rate, $O(N^{-1})$. We prove convergence in the Skorokhod topology to Kingman's coalescent under very general conditions. In particular, we allow for dependency between the number of offspring of various types of each individual. In particular, we give a general formula (3.2) for the coalescence rate c. Our results generalize previous convergence results for fast migration models in Nordborg and Krone (2002) and Sagitov and Jagers (2005).

It is often convenient to approximate a population of size N with a size N_e WF model, where N_e is referred to as the *effective population size*. This is achieved by extracting some property of the population and finding the WF model for which this property is the same. Depending on the studied property, various versions of N_e are derived, such as the variance effective population size, the inbreeding effective population size, the eigenvalue effective size, and other versions; see Ewens (1982, pp. 119–128), (2004), Orrive (1993), and Waples (2002) for reviews. Another version, the *coalescent effective population size*, was introduced in Nordborg and Krone (2002) as

$$N_e = \frac{N}{c}, \tag{1.1}$$

provided the limiting ancestral process is a time-scaled version of Kingman's coalescent. Sjödin *et al.* (2005) argued that (1.1) is the most useful notion of effective population size since the coalescent summarizes the most important (one-locus) aspects of the population history. On the other hand, (1.1) is undefined for models with other limiting ancestral processes, for instance the structured coalescent for geographically structured populations with slow migration (see Notohara (1990), Herbots (1997), Nordborg and Krone (2002), and Sampson (2006)).

An interesting application of our findings is the calculation of the coalescent effective population size (1.1) for models with combined geographical and age structure, or models with age structure where survival and fertility are not independent. This extends the results of, e.g. Sagitov and Jagers (2005); see also Cenik and Wakeley (2010). The approach of Engen *et al.* (2005a), (2005b), on the other hand, can be used to calculate the variance effective population size using forward diffusion approximations.

The paper is organized as follows. In Sections 2 and 3 we present the population genetics model in detail and prove the main convergence result (Theorem 4.1) in Section 4. A number of examples are given in Section 5 and an extension to migrating colonies, where individuals are forced to migrate in groups, is discussed in Section 6. In Section 7 we summarize and discuss possible extensions of our results, and the proof of Theorem 4.1 is given in Appendix A.

2. Forward dynamics

Consider a haploid population of constant size N, divided into L subpopulations (or types) of sizes $N_i = Na_i$, $i = 1, \ldots, L$, where $\sum_{i=1}^{L} a_i = 1$. The population evolves in discrete generations. In each new generation, an exact number $N_k m_{ki} = Na_k m_{ki}$ of all individuals in subpopulation i originate from subpopulation k. Since the subpopulation sizes are assumed to remain constant through time, this puts restrictions

$$\sum_{k=1}^{L} a_k m_{ki} = a_i, \qquad i = 1, \ldots, L, \tag{2.1}$$

on m_{ki} and a_k. In our setting (2.1) typically requires that the largest eigenvalue of $\boldsymbol{M} = (m_{ki})_{k,i=1}^{L}$ is 1, with left eigenvector $(a_1, \ldots, a_L)$; see Caswell (2001, pp. 72–92).

Consider a fixed generation, and suppose that the individuals of subpopulation k are numbered $l = 1, \ldots, N_k$. Let v_{ki}^l denote the number of 'offspring' of individual l of subpopulation k that end up in subpopulation i. This includes children of l that are born in subpopulation i and/or l itself, if it survives to the next generation and either remains in subpopulation $i = k$ or migrates to subpopulation $i \neq k$. In order to guarantee (2.1), we impose

$$\sum_{l=1}^{Na_k} v_{ki}^l = Na_k m_{ki}, \qquad 1 \leq k, i \leq L. \tag{2.2}$$

Let $\boldsymbol{v}_k^l = (v_{k1}^l, \ldots, v_{kL}^l)$ represent the number of offspring of various types of individual l of subpopulation k. Crucial for our subsequent development will be the distributional properties of $\boldsymbol{v}_k^l$. To this end, we assume that

(A1) for each k, $\{\boldsymbol{v}_k^l\}_{l=1}^{N_k}$ are exchangeable random vectors,

(A2) $\{\boldsymbol{v}_k^l\}_{l=1}^{N_k}$ are independent for different k.

Exchangeability is a natural generalization of independence, first introduced in population genetics in Cannings (1974). Because of exchangeability, we will sometimes drop the subscript l and write $v_{ki}^l = v_{ki}$. In particular, (2.2) and (A1) imply that

$$\mathrm{E}(v_{ki}) = m_{ki}.$$

We also introduce the second moment quantities

$$V_{kij} = \begin{cases} \mathrm{E}(v_{ki}(v_{ki} - 1)), & i = j, \\ \mathrm{E}(v_{ki} v_{kj}), & i \neq j. \end{cases}$$

Whereas m_{ki} is independent of N, V_{kij} typically depends on N to some extent, although not asymptotically as N grows.

Figure 1 illustrates our notation for two consecutive generations of a population with two subpopulations.
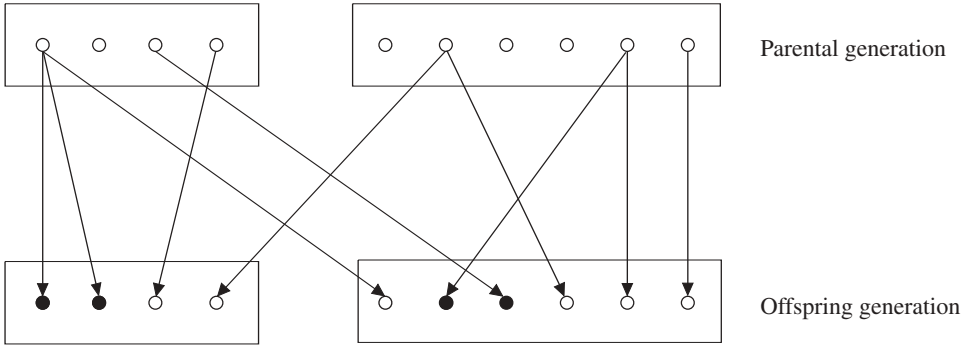
FIGURE 1: Two consecutive generations of a population with $N = 10$ and $L = 2$ subpopulations (the squares) numbered 1 and 2 from left to right. Individuals are marked with circles and offspring with arrows. The subpopulation fractions are $a_1 = 0.4$ and $a_2 = 0.6$, and the migration rates are $m_{11} = \frac{3}{4}$, $m_{12} = \frac{1}{2}$, $m_{21} = \frac{1}{6}$, and $m_{22} = \frac{4}{6}$. The individuals of subpopulation 1 are more productive, since $m_{11} + m_{12} = \frac{5}{4} > 1$ and $m_{21} + m_{22} = \frac{5}{6} < 1$. If individuals are numbered from left to right within each subpopulation, we have, for the parental generation, $\boldsymbol{v}_1^1 = (2, 1)$, $\boldsymbol{v}_1^3 = (0, 1)$, $\boldsymbol{v}_1^4 = (1, 0)$, $\boldsymbol{v}_2^2 = (1, 1)$, $\boldsymbol{v}_2^5 = (0, 2)$, $\boldsymbol{v}_2^6 = (0, 1)$, and $\boldsymbol{v}_k^l = (0, 0)$ for all other k and l. The backward transition matrix $\boldsymbol{B}$ has entries $b_{11} = \frac{3}{4}$, $b_{12} = \frac{1}{4}$, $b_{21} = \frac{1}{3}$, and $b_{22} = \frac{2}{3}$. Since $\boldsymbol{M}$ is not a stochastic matrix, the stationary distribution of the backward Markov chain, $(\gamma_1, \gamma_2) = (\frac{4}{7}, \frac{3}{7})$, differs from (a_1, a_2). The $n = 4$ sampled individuals have filled circles and, hence, the type distribution is $\boldsymbol{x}_{\text{offspring}} = (2, 2)$ in the offspring generation and $\boldsymbol{x}_{\text{parent}} = (2, 1)$ in the parental generation. Note that $|\boldsymbol{x}_{\text{offspring}}| - |\boldsymbol{x}_{\text{parent}}| = 4 - 3 = 1$ due to one coalescence event, the two children in the sample of individual 1 of subpopulation 1.

3. Backward dynamics

Assuming that

(A3) reproduction is independent across generations,

it follows that the ancestral types of an individual form a Markov chain with transition probabilities

$$b_{ik} = \frac{N a_k m_{ki}}{N a_i} = \frac{a_k m_{ki}}{a_i}.$$

By postulating that

(A4) the transition matrix $\boldsymbol{B} = (b_{ik})_{i,k=1}^L$ corresponds to an irreducible aperiodic Markov chain,

we guarantee that the ancestral type history of an individual is a Markov chain with a unique asymptotic distribution; say $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_L)$.

Let $\boldsymbol{X} = (X_1, \ldots, X_L)$ denote the type distribution of a sample of size n, with X_i the number of type-i individuals and $|\boldsymbol{X}| = \sum_{i=1}^L X_i = n$. Assume for a moment that the ancestry of these n individuals form n independent Markov chains with transition matrix $\boldsymbol{B}$. Then the equilibrium distribution is multinomial:

$$\boldsymbol{X} \sim \text{Mult}(n, \gamma_1, \ldots, \gamma_L). \tag{3.1}$$

In practice (3.1) is only an approximation, since (2.2) imposes some dependency between the ancestry of different individuals and there is merging of ancestry due to coalescence events. However, it turns out that this approximation is increasingly accurate as $N \to \infty$. To study the

impact of coalescence we let $N \to \infty$ and assume that

(A5) $\mathrm{E}((\nu_{ki})^3) = o(N)$ as $N \to \infty$ for all $1 \le k, i \le L$,

(A6) $\sigma_{kij} = (\lim_{N \to \infty} V_{kij})/(m_{ki}m_{kj})$ exist for all $1 \le i, j, k \le L$.

Assumption (A6) guarantees that, asymptotically, as $N \to \infty$, coalescence events occur one at a time through pairwise merging of lines. The probability for two individuals of types i and j to coalesce in one generation is $\sigma_{kij}/(Na_k) + o(N^{-1})$, given that both parents are of type k. This implies that coalescence events take place on a slower time scale, $O(N^{-1})$, than migration, which proceeds at a fast rate, $O(1)$; cf. Möhle (1998a).

The one-generation coalescence probability of a sample with type configuration $X = x = (x_1, \ldots, x_L)$ takes place on the slow time scale, and, thus, has the form $H(x)/N + o(N^{-1})$. Following the argument of Nordborg and Krone (2002), we can motivate the form of $H(x)$ by noting that the probability for two type-i individuals to coalesce into a type-k individual is approximately $\binom{x_i}{2}b_{ik}^2\sigma_{kii}/N_k$. Indeed, we can pick the individuals in $\binom{x_i}{2}$ different ways, the probability for both of them to have a type-k parent is approximately b_{ik}^2, and, given this event, the probability of identical parents is approximately σ_{kii}/N_k. Similarly, the probability for two individuals of types $i \ne j$ to coalesce into a type-k individual is, approximately, $x_i x_j b_{ik} b_{jk} \sigma_{kii}/N_k$. Putting things together, by adding probabilities for all i, j, and k, we obtain

$$H(x) = \sum_{k=1}^{L} \frac{1}{a_k} \left(\sum_{i=1}^{L} \binom{x_i}{2} b_{ik}^2 \sigma_{kii} + \sum\sum_{1 \le i < j \le L} x_i x_j b_{ik} b_{jk} \sigma_{kij} \right).$$

Since coalescence events occur at rate $O(N^{-1})$, the ancestral type configuration will typically transverse the state space

$$S_n = \left\{ x = (x_1, \ldots, x_L); \sum_{i=1}^{L} x_i = n \right\}$$

many times and approximately converge to the equilibrium distribution (3.1) before the first coalescence event. Using $\mathrm{E}(X_i X_j) = n(n-1)\gamma_i\gamma_j$ and $\mathrm{E}(X_i(X_i - 1)) = n(n-1)\gamma_i^2$, and rescaling time by a factor N, a coalescence rate

$$c_n = \mathrm{E}(H(X)) = c\binom{n}{2}$$

is obtained for a sample of size n, where

$$c = \sum_{k=1}^{L} \frac{1}{a_k} \sum_{i=1}^{L} \sum_{j=1}^{L} \gamma_i \gamma_j b_{ik} b_{jk} \sigma_{kij} \tag{3.2}$$

is the constant appearing in the denominator of (1.1). We may also rewrite (3.2) in the following way. Assume that I and J are drawn independently from the stationary distribution γ, and that $K|I$ and $K'|J$ are both drawn as the next state of a Markov chain with transition matrix B, currently in state I and J, respectively. Then

$$c = \mathrm{P}(K = K')\mathrm{E}\left(\frac{\sigma_{KIJ}}{a_K} \,\middle|\, K = K' \right)$$

is the product of the probability that the ancestors of I and J belong to the same subpopulation multiplied by the normalized coalescence rate σ_{KIJ}/a_K of I and J given this event.

4. Main result

Let $X_N(\tau)$ denote the ancestral type configuration of a given sample of individuals $\tau = 0$, $1, 2, \ldots$ generations back in time. If $|X_N(0)| = n$, $X_N(\cdot)$ is a time-homogeneous discrete-time Markov chain on the state space

$$S = S_1 \cup \cdots \cup S_n.$$

Changing the time scale by a factor N, we let $A_N(t) = |X_N([Nt])|$ for $t \geq 0$ denote the number of ancestral lines at time t, with $[Nt]$ denoting the largest integer smaller than or equal to Nt. We view it as a random element of $D_E[0, \infty)$, the space of right-continuous functions from $[0, \infty)$ to $E = \{1, \ldots, n\}$. Our main result states that in the limit $N \to \infty$, A_N converges weakly to Kingman's n-coalescent, run on time scale c.

Theorem 4.1. *Assume that (A1)–(A6) hold. Then, as $N \to \infty$,*

$$\{A_N(t);\ t \geq 0\} \xrightarrow{\mathcal{L}} \{A(t);\ t \geq 0\}$$

in terms of weak convergence on $D_E[0, \infty)$ with respect to the Skorokhod topology. Here A is a time-homogeneous continuous-time Markov chain on $D_E[0, \infty)$ with infinitesimal generator matrix $Q = (q_{ab})_{a,b=1}^n$, where $q_{aa} = -c\binom{a}{2}$, $q_{a,a-1} = c\binom{a}{2}$, $a = 2, \ldots, n$, c is given by (3.2), and $q_{ab} = 0$ for all other a, b.

5. Examples

Example 5.1. (*Subpopulations with newborns or adults only.*) Suppose that we have a disjoint decomposition $\{1, \ldots, L\} = \mathcal{A} \cup \mathcal{B}$ of all subpopulations i into those that consist of adults only ($i \in \mathcal{A}$) or newborns only ($i \in \mathcal{B}$). When $i \in \mathcal{B}$, v_{ki}^l represents the number of type-i children of individual l of subpopulation k. When $i \in \mathcal{A}$, v_{ki}^l is an indicator for the event that individual l of subpopulation k migrates from k to i. This implies in particular that

$$\sum_{i \in \mathcal{A}} v_{ki}^l = 1_{\{l \text{ survives to the next generation}\}},$$

$$\sum_{i \in \mathcal{B}} v_{ki}^l = \text{total number of children of } l.$$

In particular, we have

$$V_{kij} = \sigma_{kij} = 0, \qquad i, j \in \mathcal{A},$$

as a simple consequence of the facts that v_{ki}^l and v_{kj}^l are both 0 or 1, and both of them cannot equal 1. Hence, when all subpopulations consist of adults ($\mathcal{A} = \{1, \ldots, L\}$), we have $c = 0$. Of course, this represents a degenerate case when coalescence events never appear.

Example 5.2. (*Geographical structure.*) A population with nonoverlapping generations and L islands corresponds to $\mathcal{B} = \{1, \ldots, L\}$, first treated in Wright (1943). What is traditionally referred to as migration in these models corresponds in our framework to offspring being born at other subpopulations than their parents.

The WF model represents the simplest reproduction scenario, where the $N_k m_{ki}$ individuals of subpopulation i with parents in subpopulation k pick their parents independently from k, i.e.

$$(v_{ki}^1, \ldots, v_{ki}^{N_k}) \sim \text{Mult}\left(N_k m_{ki};\ \frac{1}{N_k}, \ldots, \frac{1}{N_k} \right),$$
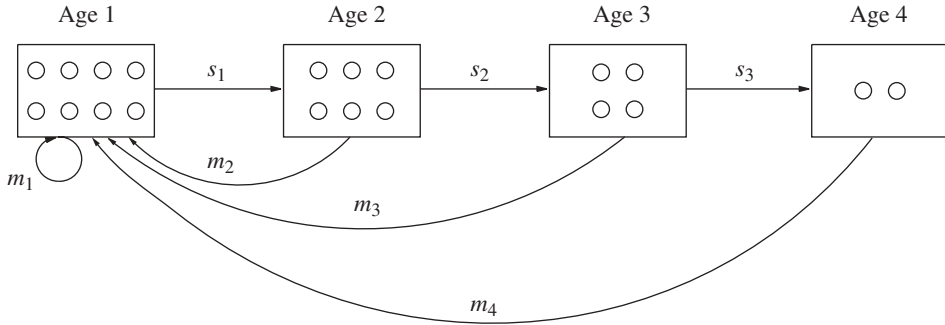
FIGURE 2: Illustration of the age-structured model of Example 5.3 with $L = 4$ age classes and $N = 20$ individuals, marked as circles. The migration rates s_k and m_k are indicated above the arrows. Since $a_1 = 0.4, a_2 = 0.3, a_3 = 0.2$, and $a_1 = 0.1$, it follows from (5.1) that $s_1 = \frac{3}{4}, s_2 = \frac{2}{3}$, and $s_3 = \frac{1}{2}$.

independently for different k, i. From this, it follows that $\lim_{N \to \infty} V_{kij} = m_{ki} m_{kj}$ for any pair i, j of subpopulations, and, hence, $\sigma_{kij} \equiv 1$. This gives

$$c = \sum_{k=1}^{L} \frac{1}{a_k} \sum_{i=1}^{L} \sum_{j=1}^{L} \gamma_i \gamma_j b_{ik} b_{jk} = \sum_{k=1}^{L} \frac{\gamma_k^2}{a_k},$$

as noted in Nordborg and Krone (2002). In the special case when the row sums of $\boldsymbol{M}$ equal 1, i.e. when $\boldsymbol{M}$ is a transition matrix of a 'forward Markov chain', we note that (see Naglyaki (1980))

$$\sum_{i=1}^{L} a_i b_{ik} = \sum_{i=1}^{L} a_i \frac{a_k m_{ki}}{a_i} = a_k \sum_{i=1}^{L} m_{ki} = a_k.$$

In view of the assumed uniqueness of the stationary distribution of $\boldsymbol{B}$, this implies that $\gamma_k = a_k$ and, hence, $c = 1$.

Example 5.3. (*Age-structured models.*) Felsenstein (1971) and Sagitov and Jagers (2005) considered a model where subpopulations represent age classes. This also incorporates the seed-bank model of Kaj *et al.* (2001). In detail, $\{1, \ldots, L\}$ represent age classes of individuals of increasing age, with $\mathcal{B} = \{1\}$ containing all newborns and $\mathcal{A} = \{2, \ldots, L\}$ containing the adult age classes. The migration rate matrix has the form

$$\boldsymbol{M} = \begin{pmatrix} m_1 & s_1 & 0 & 0 & \cdots & 0 \\ m_2 & 0 & s_2 & 0 & \cdots & 0 \\ m_3 & 0 & 0 & s_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ m_{L-1} & 0 & 0 & 0 & \cdots & s_{L-1} \\ m_L & 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

where $m_k = m_{k1}$ represents the expected number of newborn children and $s_k = m_{k,k+1}$ represents the survival probability for individuals of age class k; cf. Figure 2 for an illustration.

From (2.1) we deduce that

$$a_k = a_1 \prod_{i=1}^{k-1} s_i, \qquad a_1 = \sum_{k=1}^{L} a_k m_k, \tag{5.1}$$

which in conjunction with $\sum_{k=1}^{L} a_k = 1$ imposes some restrictions, either on the reproduction parameters m_k or the survival parameters s_k. If (5.1) holds we get nonzero backward transition probabilities $b_{1k} = m_k a_k / a_1 =: p_k$ and $b_{k+1,k} = 1$. The corresponding stationary distribution is $\gamma_k = \sum_{i=k}^{L} p_i / \gamma$, where $\gamma = \sum_{k=1}^{L} k p_k$ is a normalizing constant.

Set

$$V_k = \lim_{N \to \infty} V_{k11} = \lim_{N \to \infty} E(\nu_{k1}(\nu_{k1} - 1)) \quad \text{for } k = 1, \ldots, L$$

and

$$C_k = \lim_{N \to \infty} \text{cov}(\nu_{k1}, \nu_{k,k+1}) \quad \text{for } k = 1, \ldots, L - 1.$$

Then

$$\sigma_{kij} = \begin{cases} 1 + \dfrac{C_k}{m_k s_k} & \text{if } k = 1, \ldots, L - 1 \text{ and } (i, j) = (1, k+1) \text{ or } (k+1, 1), \\ \dfrac{V_k}{m_k^2} & \text{if } k = 1, \ldots, L - 1 \text{ and } i = j = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Inserting this into (3.2), and using the fact that $m_k = a_1 p_k / a_k$, we find that

$$c = 2 \sum_{k=1}^{L-1} \frac{1}{a_k} \gamma_1 b_{1k} \gamma_{k+1} b_{k+1,k} \left(1 + \frac{C_k a_k}{a_1 p_k s_k} \right) + \sum_{k=1}^{K} \frac{1}{a_k} (\gamma_1 b_{1k})^2 \frac{V_k a_k^2}{(a_1 p_k)^2}$$

$$= \frac{2}{\gamma} \sum_{k=1}^{L-1} \frac{p_k \gamma_{k+1}}{a_k} + \frac{2}{\gamma a_1} \sum_{k=1}^{L-1} \frac{C_k \gamma_{k+1}}{s_k} + \frac{1}{\gamma^2 a_1^2} \sum_{k=1}^{L} V_k a_k, \tag{5.2}$$

which is a generalization of the formula with $C_k \equiv 0$ derived in Sagitov and Jagers (2005). If $C_k < 0$ or $C_k > 0$, individuals with many children will respectively have a smaller or larger chance of surviving to the next generation. In the latter case reproduction becomes more unevenly distributed among parents, which naturally causes the coalescence rate c to increase, as illustrated in Figure 3.

Example 5.4. (*Combined age and geographical structure.*) Consider a simple model with two islands and two age groups within each island: newborns and adults, as illustrated in Figure 4. We write $k = (k_1, k_2)$ for a subpopulation, where $k_1 \in \{1, 2\}$ refers to the island and $k_2 \in \{1, 2\}$ refers to the age group. All newborn children survive to the next generation but remain within the same island. Adults have children, all of whom reside within the same island (probability $1 - q$) or move to the other island (probability q). Adults do not survive to the next generation. Let $\eta_{k_1}^l$ denote the number of children of the lth adult of island k_1, and let $I_{k_1}^l$ denote the indicator for the event that the children move to the other island. We thus have

$$\nu_{ki}^l = \begin{cases} 1, & k = (k_1, 1), \ i = (k_1, 2), \\ (1 - I_{k_1}^l) \eta_{k_1}^l, & k = (k_1, 2), \ i = (k_1, 1), \\ I_{k_1}^l \eta_{k_1}^l, & k = (k_1, 2), \ i = (3 - k_1, 1), \\ 0, & \text{otherwise.} \end{cases}$$

It is assumed that $\eta_{k_1}^l$ and $I_{k_1}^l$ are asymptotically independent, with

$$m_{k_1} = E(\eta_{k_1}^l), \qquad V_{k_1} = \lim_{N \to \infty} E(\eta_{k_1}^l (\eta_{k_1}^l - 1)), \qquad q = P(I_{k_1}^l = 1).$$
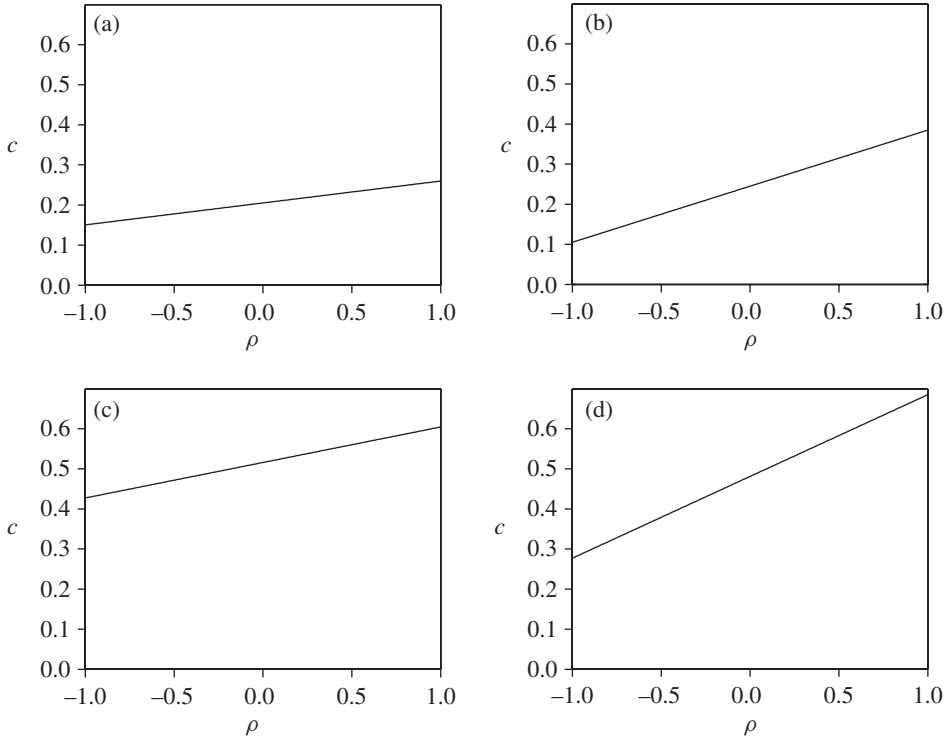
FIGURE 3: Plots of c as a function of the correlation $\rho = \mathrm{Corr}(v_{k1}, v_{k,k+1})$, $k = 1, \ldots, L-1$, for the age-structured model of Example 5.3 with $L = 4$ age classes, survival probabilities $s = (s_1, s_2, s_3)$, offspring sizes with means $\boldsymbol{m} = (m_1, m_2, m_3, m_4)$, and dispersion coefficients $\sigma_k = V_k/m_k^2$, where (a) $\boldsymbol{s} = (1, 1, 0.5)$, $\boldsymbol{m} = 2(1, 1, 1, 1)/7$, $\sigma_k \equiv -\frac{3}{2}$; (b) $\boldsymbol{s} = (1, 1, 0.5)$, $\boldsymbol{m} = 2(0, 0, 1, 1)/3$, $\sigma_k \equiv \frac{1}{2}$; (c) $\boldsymbol{s} = (0.5, 0.5, 0.5)$, $\boldsymbol{m} = 8(1, 1, 1, 1)/15$, $\sigma_k \equiv \frac{1}{8}$; and (d) $\boldsymbol{s} = (0.5, 0.5, 0.5)$, $\boldsymbol{m} = 8(0, 0, 1, 1)/3$, $\sigma_k \equiv \frac{13}{8}$. The proportionality constants of $\boldsymbol{m}$ are chosen to satisfy (5.1) and σ_k is chosen to allow for the largest possible range of correlations. Indeed, $\rho = \pm 1$ requires $v_{k1} = m_k \pm (v_{k,k+1} - s_k)\sqrt{\mathrm{var}(v_{k1})/(s_k(1 - s_k))}$ when $0 < s_k < 1$, with $\mathrm{var}(v_{k1}) = m_k + m_k^2(\sigma_k - 1)$. This amounts to $v_{k1} = 2m_k 1_{\{v_{k,k+1}=(1+\rho)/2\}}$ when $s_k = 0.5$ and $\sigma_k = 2 - 1/m_k$, i.e. individuals have a fixed number $2m_k$ of children if and only if they survive ($\rho = 1$) or if and only if they die ($\rho = -1$), respectively. Although this violates the requirement on v_{k1} to be a nonnegative integer, it still suggests that ρ can be chosen close to ± 1.

This implies that

$$m_{ki} = \begin{cases} 1, & k = (k_1, 1),\ i = (k_1, 2), \\ (1 - q)m_{k_1}, & k = (k_1, 2),\ i = (k_1, 1), \\ qm_{k_1}, & k = (k_1, 2),\ i = (3 - k_1, 1), \\ 0, & \text{otherwise}, \end{cases}$$

and

$$\sigma_{kij} = \begin{cases} \dfrac{V_{k_1}}{(1 - q)m_{k_1}^2}, & k = (k_1, 2),\ i = j = (k_1, 1), \\[2ex] \dfrac{V_{k_1}}{qm_{k_1}^2}, & k = (k_1, 2),\ i = j = (3 - k_1, 1), \\[2ex] 0, & \text{otherwise}. \end{cases}$$
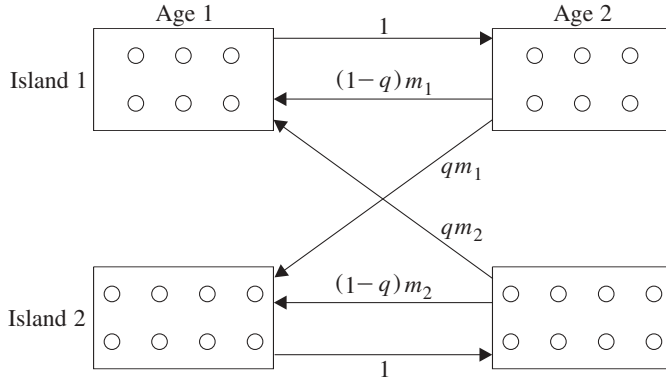
FIGURE 4: Illustration of migration rates of the model in Example 5.4, with two islands, two age classes, and $L = 2 \cdot 2 = 4$ subpopulations. Individuals are depicted as circles, $N = 28$, $a_{11} = a_{12} = \frac{3}{14}$, and $a_{21} = a_{22} = \frac{2}{7}$.

Without loss of generality, we assume that the adults of island 1 are less productive than those of island 2, i.e. $m_1 \leq 1 \leq m_2$. In order for (2.1) to have a solution, it is required that

$$m_2 = \frac{1 - (1 - q)m_1}{1 - q - (1 - 2q)m_1},$$

with island sizes $a_{11} = a_{12} = a_1/2$ and $a_{21} = a_{22} = a_2/2$, where

$$a_1 = \frac{qm_2}{q(m_1 + m_2) + 1 - m_1} = 1 - a_2$$

is the total fraction of individuals in island 1. The backward transition probabilities are

$$b_{ik} = \begin{cases} 1, & i = (i_1, 2), \ k = (i_1, 1), \\ (1 - q)m_{i_1}, & i = (i_1, 1), \ k = (i_1, 2), \\ 1 - (1 - q)m_{i_1}, & i = (i_1, 1), \ k = (1 - i_1, 2), \\ 0, & \text{otherwise,} \end{cases}$$

which has a unique stationary distribution, with $\gamma_{11} = \gamma_{12} = \gamma_1/2$ and $\gamma_{21} = \gamma_{22} = \gamma_2/2$, where

$$\gamma_1 = \frac{1 - (1 - q)m_2}{2 - (1 - q)(m_1 + m_2)} = 1 - \gamma_2$$

is the equilibrium probability that the backward Markov chain is in island 1. Putting things together, we obtain

$$c = \sum_{k_1=1}^{2} \frac{2}{a_{k_1}} \left(\frac{(\gamma_{k_1}(1 - q)m_{k_1}/2)^2 V_{k_1}}{(1 - q)m_{k_1}^2} + \frac{((\gamma_{3-k_1})(1 - (1 - q)m_{3-k_1})/2)^2 V_{k_1}}{qm_{k_1}^2} \right).$$

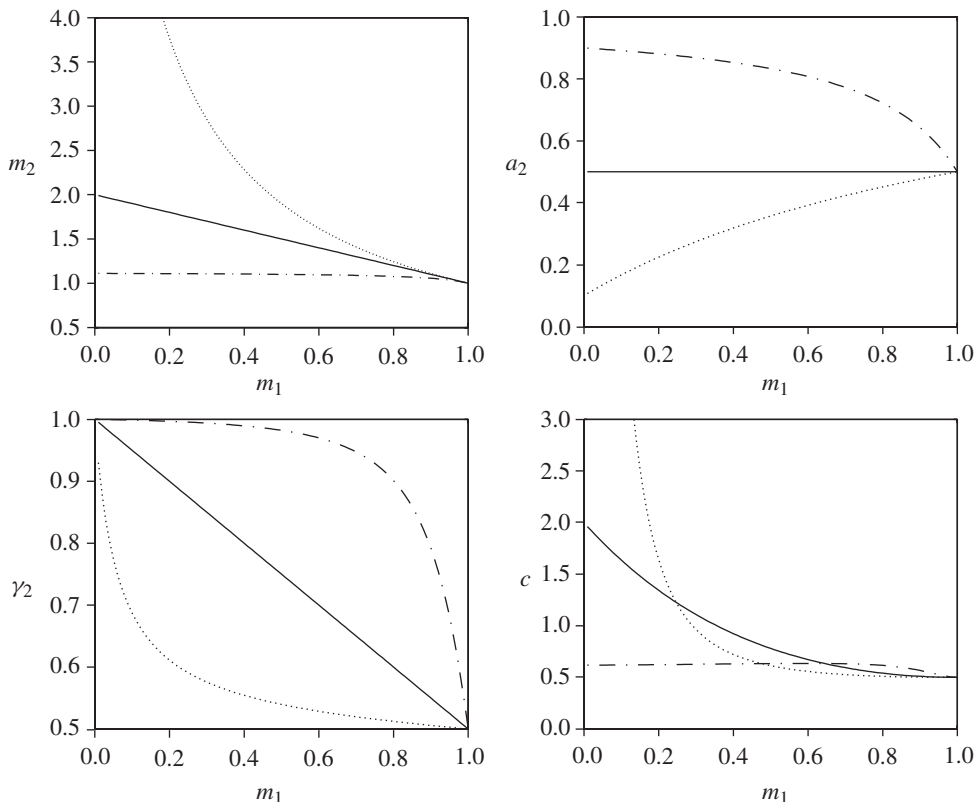Figure 5 illustrates this model for various parameter values.

FIGURE 5: Plots of m_2, a_2, γ_2, and c as a function of m_1 for two islands, each with two age classes (Example 5.4). The migration probability q is either 0.1 (*dash–dot lines*), 0.5 (*solid lines*), or 0.9 (*dotted lines*). WF-type reproduction is assumed ($V_1/m_1^2 = V_2/m_2^2 = 1$). The coalescence rate $c = 0.5$ when individuals of both islands are equally productive ($m_1 = m_2 = 1$), regardless of q. When q is small, the more productive individuals have $m_2 \approx 1$ regardless of m_1. In particular, when q and m_1 are both small, the whole system can be approximated by the age-structured model of Example 5.3 with $L = 2$, where children always survive, adults but not newborns have children, and, hence, $c = 0.5$ according to (5.2). When q is large and m_1 small, there is a small fraction of very productive individuals in subpopulation 2, and a large fraction of individuals with low productivity in island 1. This causes the coalescence rate c to increase.

6. Rapidly varying population size and migrating colonies

Interestingly, our framework has similarities to coalescence theory with a rapidly varying population size (see Jagers and Sagitov (2004)), whose variation backward in time is controlled by an aperiodic and irreducible Markov chain with transition matrix $\boldsymbol{B}$, states $Na_1, \ldots, Na_L$, and equilibrium distribution $\boldsymbol{\gamma}$. We can regard this as *the whole population* moving between subpopulations of sizes Na_i. This requires a slight redefinition of the migration rates (2.1) and the forward reproduction condition (2.2). If the whole population is currently in i, let K denote the subpopulation of the previous generation, i.e. $P(K = k \mid i) = b_{ik}$. We then assume that

$$a_k m_{ki} = a_i, \qquad Na_i = \sum_{l=1}^{Na_k} v_{ki}^l, \quad \text{if } K = k. \qquad (6.1)$$

A sample of size n can only be in one of L configurations, $\boldsymbol{x}_1 = n\boldsymbol{e}_1, \ldots, \boldsymbol{x}_L = n\boldsymbol{e}_L$, where $\boldsymbol{e}_i$ has a 1 in position i and 0s elsewhere. The coalescence probability is $\sigma_{kii}/(Na_k) + o(N^{-1})$ for a pair of individuals of subpopulation i to coalesce, given that it is known that they both originate from subpopulation k, where

$$\sigma_{kii} = \frac{V_{kii}}{m_{ki}^2} = \frac{V_{kii}}{(a_i/a_k)^2}.$$

The coalescence rate for configuration $\boldsymbol{x}_i$ is $H(\boldsymbol{x}_i)N^{-1} + o(N^{-1})$, where

$$H(\boldsymbol{x}_i) = \binom{n}{2} \sum_{k=1}^{L} b_{ik} \frac{\sigma_{kii}}{a_k}.$$

The (quasi) equilibrium distribution of the type configuration $\boldsymbol{X}$ of a sample of size n, before any coalescence events have occurred, is

$$\mathrm{P}(\boldsymbol{X} = \boldsymbol{x}_i) = \gamma_i, \qquad i = 1, \ldots, L.$$

This gives a coalescence rate $c_n = c\binom{n}{2}$, with

$$c = \mathrm{E}(H(\boldsymbol{X})) = \sum_{i=1}^{L} \sum_{k=1}^{L} \gamma_i b_{ik} \frac{V_{kii}}{m_{ki}^2 a_k} \overset{(6.1)}{=} \sum_{i=1}^{L} \sum_{k=1}^{L} \gamma_i b_{ik} V_{kii} \frac{a_k}{a_i^2}, \tag{6.2}$$

in agreement with Jagers and Sagitov (2004).

In principle, we can generalize (6.2) to a colony whose n members are constrained to reside within the same subpopulation, although the whole population may not be forced to that. For instance, the first equality in (6.2) remains valid if the whole colony moves between subpopulations according to $\boldsymbol{B}$, but migration of the whole population is controlled by (2.1)–(2.2).

Further extensions are possible, for instance, a set of $e > 1$ colonies of sizes $n_1, \ldots, n_e$ that move independently between subpopulations, but the members of each colony are restricted to migrate together, within the same subpopulations. However, individuals need not necessarily coalesce within the same colony, only within the same subpopulation.

Yet another possibility is a geographically structured coalescent model with rapidly varying populations size; cf. Sampson (2006). A subpopulation then involves both the geographical site and its size, and the backward Markov chain controls movement between these subpopulations. The backward history of a sample of size n corresponds to a migrating colony constrained not to reside in one single subpopulation, but rather to a group of subpopulations that correspond to restrictions imposed by population sizes.

7. Discussion

We have considered a general class of haploid population genetic models with constant population size and fast migration. The main result is a proof of weak convergence of the ancestral process towards Kingman's coalescent as the total population size $N \to \infty$. An important application of this result is the calculation of coalescent effective population sizes for populations with age and spatial structures, allowing for dependency between the number of offspring of various types of individual.

Several extensions of our work are possible. For instance, Engen *et al.* (2005a) derived the variance effective population size for age-structured diploid models. It would be interesting to obtain expressions for the coalescent effective diploid population size in our context of age and spatial structures. It would also be useful to study migration at a slow rate, $O(N^{-1})$, and extend the structured coalescent results of Nordborg and Krone (2002) for WF-type reproduction to our setting.

Rapidly varying population sizes (see Jagers and Sagitov (2004)) could also be analyzed, as indicated in Section 6. Another possibility is to study scenarios where the population size varies on an intermediate time scale of $O(N)$. In this case the limit process is typically a stochastic time change of Kingman's coalescent (see Kaj and Krone (2003) and Sano *et al.* (2004)).

In order to simplify the proofs, we have required a fixed amount of migration (2.2) between subpopulations. We conjecture that this condition can also be relaxed. For instance, a condition like

$$\sum_{k=1}^{L} \sum_{l=1}^{Na_k} v_{ki}^l = Na_i, \qquad \sum_{l=1}^{Na_k} v_{ki}^l = Na_k m_{ki}(1 + O_p(N^{-1/2})),$$

would incorporate the *nonconservative migration* model of Sampson (2006), where individuals produce an effectively infinite number of propagules that migrate between subpopulations at fixed proportions m_{ki}. The next generation is then formed by sampling randomly, within each subpopulation, from the infinite pool of propagules.

Appendix A. Proof of Theorem 4.1

We will follow the method of proof in Kaj *et al.* (2001), Nordborg and Krone (2002), and Sagitov and Jagers (2005).

Let d_a denote the number of states of S_a, and let $d = d_1 + \cdots + d_n$ be the number of states of S. The states of each S_a are numbered in some fixed but otherwise arbitrary way and the states of S are numbered so that the d_1 states of S_1 are listed first, followed by the d_2 states of S_2, etc. until we finally reach the d_n states of S_n. Let $\mathbf{\Pi}_N$ denote the $d \times d$ transition matrix of the Markov chain $\mathbf{X}_N$. It was shown in Nordborg and Krone (2002), using Theorem 2.12 of Ethier and Kurz (1986, p. 173), that it suffices to establish

$$\lim_{N \to \infty} \sum_{\mathbf{y} \in S_b} \mathbf{\Pi}_N^{[Nt]}(\mathbf{x}, \mathbf{y}) = \sum_{w=0}^{\infty} \frac{t^w}{w!} q_{ab}^{(w)} \tag{A.1}$$

for all $1 \le a, b \le n$ and $\mathbf{x} \in S_a$. Here $\mathbf{\Pi}_N^{\tau}$ is the τ-fold product of $\mathbf{\Pi}_N$ and $q_{ab}^{(w)}$ is the element a, b of $\mathbf{Q}^w$.

We will establish below the existence of $d \times d$ matrices $\mathbf{A}$ and $\mathbf{C}$ whose entries do not involve N, such that

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

with $\lim_{\tau \to \infty} \mathbf{A}^{\tau} = \mathbf{P}$ a $d \times d$ block diagonal matrix $\mathbf{P} = \text{diag}(\mathbf{P}_1, \dots, \mathbf{P}_n)$, where $\mathbf{P}_a$ is a $d_a \times d_a$ matrix having identical rows corresponding to the asymptotic distribution (3.1), i.e.

$$\lim_{\tau \to \infty} \mathbf{A}^{\tau}(\mathbf{x}, \mathbf{y}) = \mathbf{P}(\mathbf{x}, \mathbf{y}) = \frac{a!}{y_1! \cdots y_L!} \prod_{i=1}^{L} \gamma_i^{y_i} \quad \text{for } \mathbf{x}, \mathbf{y} \in S_a. \tag{A.3}$$

We can then invoke a crucial lemma due to Möhle (1998a), i.e.

$$\lim_{N \to \infty} \left(A + \frac{1}{N} C + o\left(\frac{1}{N}\right) \right)^{[Nt]} = P - I + e^{tG}, \tag{A.4}$$

where I is the $d \times d$ identity matrix and $G = PCP$ is the limiting (as $N \to \infty$) infinitesimal generator matrix of X_N run on the time scale N. Because of (A.4), (A.1) will follow if we can show that

$$\sum_{y \in S_a} G^w(x, y) = q_{ab}^{(w)} \tag{A.5}$$

for all $w \geq 0$, $1 \leq b \leq a \leq n$, and $x \in S_a$. Equation (A.5) was proved recursively with respect to w in Nordborg and Krone (2002) for a WF model with birth types only. Scrutinizing their proof, it turns out that (A.5) will follow in our setting as well if we can verify that

$$\sum_{y \in S_{a-1}} C(x, y) = H(x), \qquad 2 \leq a \leq n, \tag{A.6}$$

and

$$C(x, y) = 0 \quad \text{if } x \in S_a, \, y \in S_b, \, 1 \leq b \leq a - 2 \leq n - 2. \tag{A.7}$$

Hence, in order to prove Theorem 1, we need to verify (A.2), (A.3), (A.6), and (A.7) for our model.

Let $x \in S_a$ for some $2 \leq a \leq n$ represent the type configuration of a fixed offspring generation and let $y \in S_b$, $1 \leq b \leq a$, represent that of the corresponding parent generation. It will be convenient to represent x as $u = (u_1, \ldots, u_a)$, assuming that the a individuals in x are numbered in some fixed but arbitrary way, with u_s the type of individual s. Assume that the parent of u_s is of type v_s, and let $v = (v_1, \ldots, v_a)$ represent the types of all parents.

In order to characterize the coalescence pattern between u and v, we first note that $y_k \leq |\{s; v_s = k\}|$, with equality for all $k = 1, \ldots, L$ if and only if there are no coalescence events. If $y_k < |\{s; v_s = k\}|$, some parents s with $v_s = k$ will correspond to the same individual. We can identify these sets of 'identical parents' by numbering the type-k individuals of the parent generation as $l = 1, \ldots, N_k$ in some fixed but arbitrary way and then letting $l = (l_1, \ldots, l_a)$ represent the individual numbers of the parents in $v = (v_1, \ldots, v_a)$ within their respective subpopulations, i.e. $1 \leq l_s \leq N_{v_s}$. Then y_k is the number of different l_s among those s with $v_s = k$. The coalescence pattern between u and v is summarized through

$$g = \{g_k^r; k = 1, \ldots, L, \, r = 1, \ldots, y_k\},$$

where $g_k^r = (g_{k1}^r, \ldots, g_{kL}^r)$ and $g_{ki}^r = |\{s; u_s = i, \, v_s = k, \, l_s = l_{s_{kr}}\}|$ is the number of offspring of type i, within the sample, of the rth parent of type k. We assume (without loss of generality) that s_{kr} is selected to be the rth new value of l_s among all $\{s; v_s = k\}$, transversed in increasing order. Thus, g_k^r represents the offspring subpopulation membership distribution within the sample of the rth parent of type k.

Since $y = y(v, g)$, we can write the backward transition probability as

$$\Pi_N(x, y) = \sum_{\substack{v, g \\ y(v,g)=y}} P(v, g \mid u), \tag{A.8}$$

independently of how we choose to represent x by u. In order to find an expression for $P(v, g \mid u)$, note first that

$$P(v, g, l \mid u) = f(x, g) \frac{E(\prod_{k=1}^{L} \prod_{r=1}^{y_k} (v_k^{l_{s_{kr}}}) g_k^r)}{\prod_{i=1}^{L} (N_i)_{x_i}}, \tag{A.9}$$

where

$$f(x, g) = \prod_{i=1}^{L} \frac{x_i!}{(x_i - \tilde{x}_i)! \prod_{k=1}^{L} \prod_{r; |g_k^r| \geq 2} (g_{ki}^r)!}$$

denotes the number of ways to vary which individuals within u should merge according to the coalescence pattern g, with $|g_k^r| = \sum_{i=1}^{L} g_{ki}^r$ and $\tilde{x}_i = \sum_{k=1}^{L} \sum_{r; |g_k^r| \geq 2} g_{ki}^r$ denoting the total number of children of type i that merge. We use the notation $(N_i)_{x_i} = N_i(N_i - 1) \cdots (N_i - x_i + 1)$, so that the denominator of (A.9) refers to the number of ways to assign individuals that are offspring of parents within the sample. The numerator of (A.9) is (the expected value of) the number of these that conform with the coalescence pattern, with $(v_k^l) g_k^r = \prod_{i=1}^{L} (v_{ki}^l) g_k^r$.

Invoking the exchangeability and independence assumptions, (A1)–(A2), we find that

$$P(v, g \mid u) = \sum_l P(v, g, l \mid u)$$

$$= f(x, g) \frac{\prod_{k=1}^{L} (N_k)_{y_k}}{\prod_{i=1}^{L} (N_i)_{x_i}} \prod_{k=1}^{L} E\left(\prod_{r=1}^{y_k} ((v_k^r) g_k^r)\right), \tag{A.10}$$

where the sum is taken over all $\prod_{k=1}^{L} (N_k)_{y_k}$ choices of l that conform with v and g.

In order to approximate (A.10), we write $U_N \sim V_N$ to denote that $U_N / V_N \to 1$ as $N \to \infty$. Since $N_k = Na_k$, it is easy to show that

$$\frac{\prod_{k=1}^{L} (N_k)_{y_k}}{\prod_{i=1}^{L} (N_i)_{x_i}} \sim N^{|y| - |x|} \prod_{k=1}^{L} a_k^{y_k - x_k}. \tag{A.11}$$

Substituting (A.11) into (A.10) we find that

$$P(v, g \mid u) \sim f(x, g) N^{|y| - |x|} \prod_{k=1}^{L} a_k^{y_k - x_k} \prod_{k=1}^{L} E\left(\prod_{r=1}^{y_k} (v_k^r) g_k^r\right). \tag{A.12}$$

We will now simplify (A.12) further in order to approximate $\Pi_N(x, y)$ in three major cases.

Case I: no coalescence events. Assume that $|x| = |y|$. It follows from Lemma A.1 below that

$$\prod_{k=1}^{L} E\left(\prod_{r=1}^{y_k} (v_k^r) g_k^r\right) \sim \prod_{k=1}^{L} \prod_{r=1}^{y_k} E((v_k^r) g_k^r) \sim \prod_{k=1}^{L} \prod_{i=1}^{L} (m_{ki})^{n_{ki}}, \tag{A.13}$$

where

$$n_{ki} = |\{s; u_s = i, v_s = k\}| = \sum_{r=1}^{y_k} g_{ki}^r$$

denotes the migration frequency between subpopulations k and i within the sample. In the last step of (A.13) we used the fact that g_k^r has 1 in a single position, say i, and 0s elsewhere, so that $E((v_k^r) g_k^r) = m_{ki}$.

Combining (A.12) and (A.13) with $f(\boldsymbol{x}, \boldsymbol{g}) = 1$ and $m_{ki} = a_i b_{ik}/a_k$, we obtain

$$P(\boldsymbol{v}, \boldsymbol{g} \mid \boldsymbol{u}) \sim \prod_{k=1}^{L} a_k^{y_k - x_k} \prod_{k,i} \left(\frac{a_i b_{ik}}{a_k} \right)^{n_{ki}} = \prod_{k,i} b_{ik}^{n_{ki}}, \qquad (A.14)$$

where in the last equality we used $x_i = \sum_k n_{ki}$ and $y_k = \sum_i n_{ki}$. Since the transition probability in (A.14) does not involve coalescence events, we drop $\boldsymbol{g}$ in the notation and write $P(\boldsymbol{v} \mid \boldsymbol{u})$. In view of (A.2), (A.8), and (A.14), we thus define

$$A(\boldsymbol{x}, \boldsymbol{y}) = \lim_{N \to \infty} \sum_{\substack{\boldsymbol{v} \\ \boldsymbol{y}(\boldsymbol{v}) = \boldsymbol{y}}} P(\boldsymbol{v} \mid \boldsymbol{u}) = \sum_{\substack{\boldsymbol{v} \\ \boldsymbol{y}(\boldsymbol{v}) = \boldsymbol{y}}} \prod_{s=1}^{L} b_{u_s v_s}$$

when $|\boldsymbol{y}| = |\boldsymbol{x}| = a$, with $y_k(\boldsymbol{v}) = |\{s;\ v_s = k\}|$. When $|\boldsymbol{y}| \neq |\boldsymbol{x}|$, we set $A(\boldsymbol{x}, \boldsymbol{y}) = 0$. Note that A is obtained by the components v_s changing independently according to a Markov chain with transition matrix $\boldsymbol{B}$ and then keeping track of the subpopulation distribution $\boldsymbol{y} = \boldsymbol{y}(\boldsymbol{v})$. Although $\boldsymbol{v} \to \boldsymbol{y}(\boldsymbol{v})$ is noninvertible, it still preserves the Markov property when applied to a Markov chain with transition matrix $\boldsymbol{B}^{\otimes L}$. Therefore,

$$A^{\tau}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\substack{\boldsymbol{v} \\ \boldsymbol{y}(\boldsymbol{v}) = \boldsymbol{y}}} \prod_{s=1}^{L} \boldsymbol{B}_{u_s, v_s}^{\tau}.$$

Since $\boldsymbol{\gamma}$ is the equilibrium distribution of $\boldsymbol{B}$,

$$\lim_{\tau \to \infty} A^{\tau}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\substack{\boldsymbol{v} \\ \boldsymbol{y}(\boldsymbol{v}) = \boldsymbol{y}}} \prod_{s=1}^{L} \gamma_{v_s} = \frac{a!}{\prod_{i=1}^{L} y_i!} \prod_{i=1}^{L} \gamma_i^{y_i},$$

and this proves (A.3).

Case II: one coalescence event. Assume that $|\boldsymbol{y}| = |\boldsymbol{x}| - 1$ and that $\boldsymbol{g}$ corresponds to a single coalescence event where two offspring of types i and j are merged into type k. Then

$$f(\boldsymbol{x}, \boldsymbol{g}) = \begin{cases} \binom{x_i}{2}, & i = j, \\ x_i x_j, & i \neq j, \end{cases}$$

and Lemma A.1 below implies that

$$\prod_{\kappa=1}^{L} E\left(\prod_{r=1}^{y_\kappa} (v_\kappa^r)_{g_\kappa^r} \right) \sim \prod_{\kappa=1}^{L} \prod_{r=1}^{y_\kappa} E((v_\kappa^r)_{g_\kappa^r}) \sim \frac{V_{kij}}{m_{ki} m_{kj}} \prod_{\kappa=1}^{L} \prod_{\iota=1}^{L} (m_{\kappa\iota})^{n_{\kappa\iota}}. \qquad (A.16)$$

Combining (A.12), (A.16), and assumption (A6), we thus find that

$$P(\boldsymbol{v}, \boldsymbol{g} \mid \boldsymbol{u}) \sim \frac{f(\boldsymbol{x}, \boldsymbol{g}) \sigma_{kij}}{N} \prod_{\kappa=1}^{L} a_\kappa^{y_\kappa - x_\kappa} \prod_{\kappa, \iota} \left(\frac{a_\iota b_{\iota\kappa}}{a_\kappa} \right)^{n_{\kappa\iota}}$$

$$= \frac{f(\boldsymbol{x}, \boldsymbol{g}) \sigma_{kij}}{N} \prod_{\kappa=1}^{L} a_\kappa^{y_\kappa - \tilde{y}_\kappa} \prod_{\kappa, \iota} b_{\iota\kappa}^{n_{\kappa\iota}}$$

$$= \frac{f(\boldsymbol{x}, \boldsymbol{g}) \sigma_{kij}}{N a_k} \prod_{\kappa, \iota} b_{\iota\kappa}^{n_{\kappa\iota}}, \qquad (A.17)$$

where in the first equality we used

$$x_\iota = \sum_\kappa n_{\kappa\iota}, \qquad \tilde{y}_\kappa := |\{s; v_s = \kappa\}| = \sum_\iota n_{\kappa\iota},$$

and in the second equality we used $\tilde{y}_k = y_k + 1$ and $\tilde{y}_\kappa = y_\kappa$ for $\kappa \neq k$. In view of (A.4), (A.8), and (A.17), we define

$$C(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\substack{\boldsymbol{v}, \boldsymbol{g} \\ \boldsymbol{y}(\boldsymbol{v}, \boldsymbol{g}) = \boldsymbol{y}}} \frac{f(\boldsymbol{x}, \boldsymbol{g})\sigma_{kij}}{a_k} \prod_{\kappa, \iota} b_{\iota\kappa}^{n_{\kappa\iota}} \tag{A.18}$$

when $1 \leq |\boldsymbol{y}| = |\boldsymbol{x}| - 1 \leq n - 1$. Note that k, i, j, and $\{n_{\kappa\iota}\}$ all depend on $\boldsymbol{v}, \boldsymbol{g}$ in (A.18). Given $\boldsymbol{x} \in S_a$, by summing (A.18) over all $\boldsymbol{y} \in S_{a-1}$ we obtain

$$\sum_{\boldsymbol{y} \in S_{a-1}} C(\boldsymbol{x}, \boldsymbol{y}) = \sum_{k=1}^{L} \frac{1}{a_k} \sum_{i=1}^{L} \binom{x_i}{2} \sigma_{kii} \sum_{\boldsymbol{v}} \prod_{\kappa, \iota} b_{\iota\kappa}^{n_{\kappa\iota}}$$

$$+ \sum_{k=1}^{L} \frac{1}{a_k} \sum_{1 \leq i < j \leq L} x_i x_j \sigma_{kij} \sum_{\boldsymbol{v}} \prod_{\kappa, \iota} b_{\iota\kappa}^{n_{\kappa\iota}}$$

$$= \sum_{k=1}^{L} \frac{1}{a_k} \left(\sum_{i=1}^{L} \binom{x_i}{2} b_{ik}^2 \sigma_{kii} + \sum_{1 \leq i < j \leq L} x_i x_j b_{ik} b_{jk} \sigma_{kij} \right)$$

$$= H(\boldsymbol{x}). \tag{A.19}$$

In the first equality of (A.19), when $i = j$, we summed over all $\boldsymbol{v}$ such that $v_{s_1} = v_{s_2} = k$ are kept fixed, with s_1 and s_2 an arbitrarily chosen pair of indexes (out of all $\binom{x_i}{2}$ possible) such that $u_{s_1} = u_{s_2} = i$. We thus have

$$\sum_{\boldsymbol{v}} \prod_{\kappa, \iota} b_{\iota\kappa}^{n_{\kappa\iota}} = \sum_{\boldsymbol{v}} \prod_{s=1}^{L} b_{u_s v_s} = b_{ik}^2, \tag{A.20}$$

where in the last step we interchanged the order of the product and sum and utilized the fact that the row sums of $\boldsymbol{B}$ are 1. Similarly, when $i \neq j$ in (A.19), we sum over all $\boldsymbol{v}$ such that $v_{s_1} = v_{s_2} = k$ are kept fixed, with s_1 and s_2 an arbitrarily chosen pair of indexes (out of all $x_i x_j$ possible) such that $u_{s_1} = i$ and $u_{s_2} = j$. The sum analogous to (A.20) is $b_{ik} b_{jk}$ instead of b_{ik}^2. Summarizing, we see from (A.19) that (A.6) is verified.

Case III: more than one coalescence event. We need to verify (A.7). In view of (A.2), since $\boldsymbol{A}$ is block diagonal, it suffices to prove that

$$\Pi_N(\boldsymbol{x}, \boldsymbol{y}) = o(N^{-1})$$

when $|\boldsymbol{x}| - |\boldsymbol{y}| \geq 2$. However, this is an immediate consequence of (A.8), (A.12), and Lemma A.1, since the number of terms of size $o(N^{-1})$ on the right-hand side of (A.8) is uniformly bounded, independently of N.

Lemma A.1. *Consider y individuals $l = 1, \ldots, y$ from subpopulation k, with coalescence pattern $\boldsymbol{g} = (\boldsymbol{g}_1, \ldots, \boldsymbol{g}_y)$, where $\boldsymbol{g}_l = (g_{l1}, \ldots, g_{lL})$ and g_{li} is the number of offspring of l*

within the sample from subpopulation i. *Let*

$$|\boldsymbol{g}| = \sum_{l=1}^{y} |\boldsymbol{g}_l| = \sum_{l=1}^{y} \sum_{i=1}^{L} g_{li},$$

so that $|\boldsymbol{g}| - y$ *is the number of lines that coalesce. Then, as* $N \to \infty$,

$$\mathrm{E}\left(\prod_{l=1}^{y} (\boldsymbol{v}_k^l)_{\boldsymbol{g}_l}\right) \begin{cases} \sim \prod_{l=1}^{y} \mathrm{E}((\boldsymbol{v}_k^l)_{\boldsymbol{g}_l}) & \text{if } |\boldsymbol{g}| = y, y+1, \\ = o(N^{|\boldsymbol{g}|-y-1}) & \text{if } |\boldsymbol{g}| \geq y+2. \end{cases} \tag{A.21}$$

Proof. We will sketch a proof by establishing (A.21) recursively with respect to $\boldsymbol{g}$. See also Möhle (1998c), Sagitov (1999), and Jagers and Sagitov (2004) for similar calculations. Loosely speaking, we proceed by induction with respect to $(y, \boldsymbol{g})$.

Case 1. When $y = 1$, the result is trivially true when $|\boldsymbol{g}| = 1$, it follows from assumption (A4) for $|\boldsymbol{g}| = 2$, from assumption (A5) and Hölder's inequality for $|\boldsymbol{g}| = 3$, and by repeated use of $v_{ki}^l \leq N_k m_{ki} = O(N)$ (see (2.2)) and Hölder's inequality for $|\boldsymbol{g}| > 3$.

Case 2. Assume that $|\boldsymbol{g}| = y > 1$, i.e. no coalescence events, with $g_{li_l} = 1$ for $l = 1, \ldots, y$, so that i_l is the type of the single descendant of l. Then

$$\mathrm{E}\left(\prod_{l=1}^{y} v_{ki_l}^l\right) = \frac{1}{(N_k)_y} \mathrm{E}\left(\sum_{l_1, \ldots, l_y \text{ different}} \prod_{r=1}^{y} v_{ki_r}^{l_r}\right)$$

$$= \frac{1}{(N_k)_y} \mathrm{E}\left(\sum_{l_1, \ldots, l_y = 1}^{N_k} \prod_{r=1}^{y} v_{ki_r}^{l_r}\right) + o(1),$$

$$\sim \frac{1}{(N_k)_y} \mathrm{E}\left(\prod_{r=1}^{y} \sum_{l_r=1}^{N_k} v_{ki_r}^{l_r}\right) + o(1),$$

$$\sim \frac{1}{(N_k)_y} \prod_{r=1}^{y} N_{ki_r}$$

$$\sim \prod_{r=1}^{y} m_{ki_r},$$

where in the first step we utilized the exchangeability condition (A1), in the second step we obtained a remainder term $o(1)$ by tacitly assuming that (A.21) has already been established for all configurations $\boldsymbol{g}$ with less than y parents, and in the fourth step we used (2.2) and $N_{ki} = N_k m_{ki}$.

Case 3. Assume that $y \geq 2$, $|\boldsymbol{g}_1| = 2$, and $|\boldsymbol{g}_2| = \cdots = |\boldsymbol{g}_L| = 1$. If $g_{1i_1} = 2$ and $g_{li_l} = 1$ for $l = 2, \ldots, y$, we find that

$$\mathrm{E}\left((v_{ki_1}^1)_2 \prod_{l=2}^{y} v_{ki_l}^l\right) = \frac{1}{(N_k - 1)_{y-1}} \mathrm{E}\left(\sum_{l_2, \ldots, l_y \text{ different} > 1} (v_{ki_1}^1)_2 \prod_{r=2}^{y} v_{ki_r}^{l_r}\right)$$

$$= \frac{1}{(N_k - 1)_{y-1}} \mathrm{E}\left(\sum_{l_2, \ldots, l_y = 1}^{N_k} (v_{ki_1}^1)_2 \prod_{r=2}^{y} v_{ki_r}^{l_r}\right) + o(1),$$

$$\sim \frac{1}{(N_k-1)_{y-1}} \, \mathrm{E}\!\left((v_{ki_1}^1)_2 \prod_{r=2}^{y} \sum_{l_r=1}^{N_k} v_{ki_r}^{l_r}\right)$$

$$\sim \frac{1}{(N_k-1)_{y-1}} V_{ki_1 i_1} \prod_{r=2}^{y} N_{ki_r}$$

$$\sim V_{ki_1 i_1} \prod_{r=2}^{y} m_{ki_r},$$

where again in the second step we assumed that (A.21) has already been proved for all $\boldsymbol{g}$ with less than y terms. The case when $g_{1i_1} = g_{1j_1} = 1$ for some $i_1 \neq j_1$ and $g_{li_l} = 1$ for $l = 2, \ldots, y$ is proved analogously.

Case 4. Assume that $y \geq 2$, $|\boldsymbol{g}_1| = 3$, and $|\boldsymbol{g}_2| = \cdots = |\boldsymbol{g}_L| = 1$. The proof of (A.21) is similar to that of case 3.

Case 5. Suppose that $y = 2$, with $|\boldsymbol{g}_1| = |\boldsymbol{g}_2| = 2$ and $g_{1i_1} = g_{1i_2} = g_{2j_1} = g_{2j_2} = 1$ for some $i_1 \neq i_2$ and $j_1 \neq j_2$. We use the exchangeability condition (A1) to deduce that

$$
\begin{aligned}
N_{kj_1} N_{kj_2} \, \mathrm{E}(v_{ki_1}^1 v_{ki_2}^1) &= \mathrm{E}\!\left(\sum_{l_1=1}^{N_k} \sum_{l_2=1}^{N_k} v_{ki_1}^1 v_{ki_2}^1 v_{kj_1}^{l_1} v_{kj_2}^{l_2}\right) \\
&= (N_k-1) \, \mathrm{E}(v_{ki_1}^1 v_{ki_2}^1 v_{kj_1}^2 v_{kj_2}^2) + \mathrm{E}(v_{ki_1}^1 v_{ki_2}^1 v_{kj_1}^1 v_{kj_2}^1) \\
&\quad + (N_k-1)(N_k-2) \, \mathrm{E}(v_{ki_1}^1 v_{ki_2}^1 v_{kj_1}^2 v_{kj_2}^3) \\
&\quad + (N_k-1) \, \mathrm{E}(v_{ki_1}^1 v_{ki_2}^1 v_{kj_1}^1 v_{kj_2}^2) + (N_k-1) \, \mathrm{E}(v_{ki_1}^1 v_{ki_2}^1 v_{kj_1}^2 v_{kj_2}^1),
\end{aligned}
$$

which is equivalent to

$$
\begin{aligned}
\mathrm{E}(v_{ki_1}^1 v_{ki_2}^1 v_{kj_1}^2 v_{kj_2}^2) &= (N_k-1)^{-1} N_{kj_1} N_{kj_2} \, \mathrm{E}(v_{ki_1}^1 v_{ki_2}^1) - (N_k-1)^{-1} \, \mathrm{E}(v_{ki_1}^1 v_{ki_2}^1 v_{kj_1}^1 v_{kj_2}^1) \\
&\quad - (N_k-2) \, \mathrm{E}(v_{ki_1}^1 v_{ki_2}^1 v_{kj_1}^2 v_{kj_2}^3) - \mathrm{E}(v_{ki_1}^1 v_{ki_2}^1 v_{kj_1}^1 v_{kj_2}^2) \\
&\quad - \mathrm{E}(v_{ki_1}^1 v_{ki_2}^1 v_{kj_1}^2 v_{kj_2}^1) \\
&= N_k m_{kj_1} m_{kj_2} V_{ki_1 i_2} - N_k m_{kj_1} m_{kj_2} V_{ki_1 i_2} + o(N) \\
&= o(N),
\end{aligned}
$$

which proves (A.21), since $|\boldsymbol{g}| - y = 4 - 2 = 2$. In the last step we utilized the fact that (A.21) has already been proved for the first two terms (case 1), the third term (case 3), and the last two terms (case 4), on the right-hand side. The case when $i_1 = i_2$ and/or $j_1 = j_2$ is proved analogously.

Case 6. Suppose that $y = 2$, with $|\boldsymbol{g}_1| = 3$ and $|\boldsymbol{g}_2| = 2$. For instance, if $g_{1i_1} = 2$ and $g_{1i_2} = g_{2j_1} = g_{2j_2} = 1$, we derive

$$\mathrm{E}((v_{ki_1}^1)_2 v_{ki_2}^1 v_{kj_1}^2 v_{kj_2}^2) \leq N_{ki_1} \, \mathrm{E}(v_{ki_1}^1 v_{ki_2}^1 v_{kj_1}^2 v_{kj_2}^2) = o(N^2),$$

from case 5 and the fact that $v_{ki}^1 - 1 \leq v_{ki}^1 \leq N_{ki}$, which is a consequence of (2.2). This proves (A.21) since $|\boldsymbol{g}| - y = 5 - 2 = 3$. The proof for other choices of $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ with $|\boldsymbol{g}_1| = 3$ and $|\boldsymbol{g}_2| = 2$ is analogous.

Proceeding in a similar way, we can establish (A.21) recursively for all possible $\boldsymbol{g}$.

Acknowledgements

References

CANNINGS, C. (1974). The latent roots of certain Markov chains arising in genetics: a new approach. I. Haploid models. *Adv. Appl. Prob.* **6,** 260–290.

CASWELL, H. (2001). *Matrix Population Models*, 2nd edn. Sinauer, Sunderland, MA.

CENIK, C. AND WAKELEY, J. (2010). Pacific salmon and the coalescent effective population size. *PLoS One*, **5,** e13019.

ENGEN, S., LANDE, R. AND SAETHER, B.-E. (2005a). Effective size of a fluctuating age-structured population. *Genetics* **170,** 941–954.

ENGEN, S., LANDE, R., SAETHER, B.-E. AND WEIMERSKIRCH, H. (2005b). Extinction in relation to demographic and environmental stochasticity in age-structured models. *Math. Biosci.* **195,** 210–227.

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

EWENS, W. J. (1982). On the concept of effective population size. *Theoret. Pop. Biol.* **21,** 373–378.

EWENS, W. J. (2004). *Mathematical Population Genetics. I. Theoretical Introduction*, 2nd edn. Springer, New York.

FELSENSTEIN, J. (1971). Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics* **68,** 581–597.

FU, Y. X. (1997). Coalescent theory for a partially selfing population. *Genetics* **146,** 1489–1499.

HERBOTS, H. M. (1997). The structured coalescent. In *Progress in Population Genetics and Human Evolution* (Minneapolis, MN, 1994; IMA Vol. Math. Appl. **87**), eds. P. Donnelly and S. Tavaré. Springer, New York, pp. 231–255.

JAGERS, P. AND SAGITOV, S. (2004). Convergence to the coalescent in populations of substantially varying size. *J. Appl. Prob.* **41,** 368–378.

KAJ, I. AND KRONE, S. M. (2003). The coalescent process in a population with stochastically varying size. *J. Appl. Prob.* **40,** 33–48.

KAJ, I., KRONE, S. M. AND LASCOUX, M. (2001). Coalescent theory for seed bank models. *J. Appl. Prob.* **38,** 285–300.

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

KINGMAN, J. F. C. (1982b). On the genealogy of large populations. In *Essays in Statistical Science* (J. Appl. Prob. Spec. Vol. **19A**), eds J. Gani and E. J. Hannan, Applied Probability Trust, Sheffield, pp. 27–43.

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

MÖHLE, M. (1998b). Coalescent results for two-sex population models. *Adv. Appl. Prob.* **30,** 513–520.

MÖHLE, M. (1998c). Robustness results for the coalescent. *J. Appl. Prob.* **35,** 438–447.

MÖHLE, M. (2000). Ancestral processes in population genetics - the coalescent. *J. Theoret. Biol.* **204,** 629–638.

MÖHLE, M. AND SAGITOV, S. (2003). Coalescent patterns in diploid exchangeable population models. *J. Math. Biol.* **47,** 337–352.

NAGLYAKI, T. (1980). The strong-migration limit in geographically structured populations. *J. Math. Biol.* **9,** 101–114.

NORDBORG, M. AND DONNELLY, P. (1997). The coalescent process with selfing. *Genetics* **146,** 1185–1195.

NORDBORG, M. AND KRONE, S. M. (2002). Separation of time scales and convergence to the coalescent in structured populations. In *Modern Development in Theoretical Population Genetics*, eds M. Slatkin and M. Veuille, Oxford University Press, pp. 194–232.

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

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

POLLACK, E. (2010). Coalescent theory for a monoecious random mating population with a varying size. *J. Appl. Prob.* **47,** 41–57.

SAGITOV, S. (1999). The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Prob.* **36,** 1116–1125.

SAGITOV, S. AND JAGERS, P. (2005). The coalescent effective size of age-structured populations. *Ann. Appl. Prob.* **15,** 1778–1797.

SAMPSON, K. Y. (2006). Structured coalescent with nonconservative migration. *J. Appl. Prob.* **43,** 351–362.

SANO, A., SHIMIZU, A. AND IIZUKA, M. (2004). Coalescent process with fluctuating population size and its effective size. *Theoret. Pop. Biol.* **65,** 39–48.

SJÖDIN, P. *et al.* (2005). On the meaning and existence of an effective population size. *Genetics* **169,** 1061–1070.

WAPLES, R. S. (2002). Definition and estimation of effective population size in the conservation of endangered species. In *Population Viability Analysis*, eds S. R. Beissinger and D. R. McCullogh, University of Chicago Press, pp. 147–168.

WRIGHT, S. (1943). Isolation by distance. *Genetics* **28,** 114–138.