THEORETICAL POPULATION BIOLOGY **50**, 91–104 (1996)
ARTICLE NO. 0024

The Sampling Theory of Neutral Alleles in an Island Population of Fluctuating Size

BRUCE RANNALA*

Department of Biology, Yale University, Box 208104, New Haven, Connecticut 06520-8104

Received July 24, 1995

The sampling distribution of neutral alleles under a stochastic birth, death, and immigration (BDI) process, proposed as a model of fluctuating island populations, is studied by analytical methods. A new result is presented for the distribution of allele types in a sample of N individuals from an island when the allele frequencies among immigrants are constant. The sampling distribution of allele types depends on the sample size N, the array of allele frequencies among immigrants $\mathbf{p}$, and the parameter $\theta = \phi/\lambda$, where ϕ is the immigration rate and λ is the individual birth rate. The sampling distribution of alleles does not depend on time or population size, and no "genetic equilibrium" assumption is therefore needed to apply the model to natural populations. The moments of the sampling distribution of allele types are used to calculate the expectation and variance of a sample identity by descent estimate (f_0^N) within islands under the BDI model.

INTRODUCTION

The demographic models that have been applied in classical studies of the genetic structure of subdivided populations (Wright, 1931, 1940, 1943; Kimura, 1953; Kimura and Weiss, 1964; Maruyama, 1970) often provided a weak description of natural populations because they require the unrealistic assumption that population size remains constant over time and space. An important biological question concerns the effect of local population size fluctuations on the genetic variability within, and the genetic differentiation between, semi-isolated populations (Lande, 1992; Whitlock, 1992). This paper further studies the mathematical properties of the continuous-generation island model of population genetics first proposed by Rannala and Hartigan (1995) in which semi-isolated island populations are allowed to fluctuate randomly in size.

Rannala and Hartigan (1995) considered the genetic structure of a haploid island population with continuous generations and a fluctuating

* Current address: Department of Integrative Biology, University of California, Berkeley, CA 94720. E-mail: bruce@mws4.biol.berkeley.edu. Fax: (510)643-6264.

0040-5809/96 $18.00

size determined by a stochastic birth, death, and immigration (BDI) process. The island receives immigrants exclusively from a mainland population with constant allele frequencies. It was shown, in this case, that the expectation of identity by descent (IBD—Malécot, 1948) for random pairs of alleles within an island population and the expected differentiation of allele frequencies among replicate island populations do not depend on the death rate, time, or local population size.

In this paper, the distribution of alleles on islands under a haploid BDI process is studied by analytical methods and it is shown that the limiting distribution of allele frequencies on each island has a Dirichlet density and is identical to the distribution obtained using a diffusion approximation for Wright's island model at equilibrium (with an appropriate change of parameters). For a sample or a population of N individuals under the BDI model, the distribution of the sizes of families of unique allele types is shown to be the compound multinomial Dirichlet, a distribution whose properties are well-known. Surprisingly, the sampling distribution of alleles does not depend on time; the populations generated by a BDI process immediately achieve a steady-state sampling distribution and therefore no "equilibrium" assumption is needed to apply the model. The expectation and variance, over replicate islands, of f_0 the probability of IBD for random pairs of alleles are derived from the statistical moments of this distribution.

1. Review of the Island Model

Wright (1940) considered the distribution of allele frequencies at genetic equilibrium for a single genetic locus with two selectively neutral alleles in a diploid island population of constant size N with discrete generations. A fraction m of the individuals on the island are replaced, in each generation, by immigrants from a large mainland population with alleles A_1 and A_2 in constant frequencies $p:(1-p)$. The expected allele frequencies on the island at genetic equilibrium are $p:(1-p)$ but the higher moments of the distribution of allele frequency, over replicate islands, depend additionally on N and m.

Wright (1940) obtained an approximate solution for the steady-state distribution of allele frequency under this model. The distribution of the frequency α of A_1 among islands for large N and $m \ll 1$ is (from Wright, 1940)

$$\phi(\alpha) = \frac{\Gamma(\theta)}{\Gamma(\theta\alpha)\,\Gamma(\theta[1-\alpha])}\,\alpha^{\theta p-1}(1-\alpha)^{\theta(1-p)-1}, \qquad (1)$$

which is a beta distribution with parameters θp and $\theta(1-p)$, where $\theta = 4Nm$. In the case of k alleles among immigrants the distribution of allele frequencies for large N and $m \ll 1$ is (from Wright, 1969)

$$\phi(\alpha) = \Gamma(\theta) \prod_{i=1}^{k} \frac{\alpha_i^{\theta p_i - 1}}{\Gamma(\theta \alpha_i)}, \tag{2}$$

where $\alpha = \alpha_1, \alpha_2, ..., \alpha_k$ is the array of allele frequencies on an island, and $\mathbf{p} = p_1, p_2, ..., p_k$ is the array of allele frequencies among immigrants. Equation (2) is the density function of a Dirichlet distribution with scale parameter θ (see, e.g., Johnson and Kotz, 1972).

Wright's model may be extended to consider a collection of islands, each independently receiving immigrants from a source with constant allele frequencies. One way this situation may be generated is by assuming the immigrants entering any particular island are drawn, at random, from a very large collection of islands (Wright, 1931). This is the so-called "infinite-island" model. A mathematically equivalent situation may be obtained by assuming that each island receives immigrants exclusively (and independently) from a mainland population with constant allele frequencies. This second situation has been referred to as an island–mainland metapopulation (Hanski, 1994; Rannala and Hartigan, 1995).

The expected large-population probability of IBD for a random pair of alleles at equilibrium for Wright's (1931) discrete-generation (diploid) island model may be obtained from the moment-generating function of the compound multinomial Dirichlet (see below) and is

$$f_0 = \frac{1 + \theta f_1}{1 + \theta}, \tag{3}$$

where $f_1 = \sum_{i=1}^{k} p_i^2$. The expected genetic differentiation is approximately

$$F_{ST} = \frac{1}{1 + \theta}, \tag{4}$$

where $\theta = 4Nm$ is the number of migrant alleles exchanged per generation on each island.

Rannala and Hartigan (1995) considered a collection of haploid island populations, with continuous generations and fluctuating sizes regulated by a stochastic BDI process, with each island receiving immigrants according to an independent Poisson process with expectation ϕ (in unit time) from a source with constant allele frequencies. It was shown that the expected

probability of IBD for a pair of random alleles from a population (or sample) of size N under this model is

$$\mathsf{E}[f_0^N] = \frac{\lambda + \phi f_1}{\lambda + \phi}, \tag{5}$$

and the expected genetic differentiation among islands (for a collection of roughly 10 or more islands) is

$$F_{ST} = \frac{\lambda}{\lambda + \phi}, \tag{6}$$

where ϕ is the expected number of immigrant arrivals, and λ is the individual birthrate, in unit time. If $\theta = \phi/\lambda$ these equations are identical to those for Wright's island model at equilibrium. The reason for this similarity resides in the fact that both models have a limiting distribution of allele frequencies that is the Dirichlet with scale parameter θ.

There is an important difference between Wright's island model and the BDI model. The results for the IBD and F_{ST} under the Wright model assume a large population size on each island (i.e., use a diffusion approximation) and only hold at genetic equilibrium (i.e., the steady-state genetic distribution). For the BDI model the results for IBD and F_{ST} are exact for a population (or random sample) of size $N \geqslant 2$ (see below) and the steady-state sampling distribution is immediately achieved so that no assumption of "genetic equilibrium" is needed (i.e., the distribution of alleles does not depend on time).

2. BDI Model of Island Populations

Consider a haploid island population in which individuals reproduce by binary fission at an individual rate λ and die at an individual rate μ, and immigrants arrive at a rate ϕ in continuous time. During an infinitesimal time interval of length δt, the probability of a single birth in a population of size N is $\lambda N \delta t$, the probability of a single death is $\mu N \delta t$, and the probability of a single immigrant arrival is $\phi \delta t$. The probability that more than a single event occurs during an infinitesimal time interval of length δt is of order $o(\delta t)$ and may be neglected. The effects of mutation are assumed negligible.

The distribution of the population size for a stochastic BDI process has been studied by Kendall (1948, 1949), and his results are reviewed by Rannala and Hartigan (1995). The BDI process may be used to model a variety of demographic structures for islands occurring in nature. These include stochastic exponential growth on islands and a "source-sink"

island–mainland population structure (e.g., Pulliam, 1988). The genetic results presented here also apply under a stochastic "logistic" model of population demography in which population size is regulated by a density-dependent death process (Rannala and Hartigan, 1995).

The genetic structure that arises under a stochastic BDI process has been previously studied by Mode (1962), Kendall (1975), and Tavaré (1987, 1989). Recent studies have been primarily concerned with the limiting (large-population) distribution under a BDI process for the infinitely many neutral alleles model with linear mutation pressure (Tavaré, 1987, 1989). Rannala and Hartigan (1995) used the linear BDI process as a model of island populations with continuous generations and stochastic fluctuations in population size. In this paper, I present a new result for the distribution of allele types under a BDI process, used as a model of island populations, with k allele types in constant frequencies among immigrants.

2.1. *Limiting Distribution of Allele Frequencies*

The BDI process may be represented as a discrete Markov chain embedded in continuous time. By ignoring deaths and the waiting times between events, the underlying birth–immigration process can be described exactly using a Polya urn scheme (see Blackwell and MacQueen, 1973) in which each new addition to an island population, when the population is of size N, is an immigrant with probability

$$\frac{\theta}{N+\theta}, \tag{7}$$

whose allele is of type j with probability p_j, or inherits its allele from any one of the existing N individuals with probability

$$\frac{1}{N+\theta}. \tag{8}$$

In the limit, the allele frequencies α follow the Dirichlet density (see Blackwell and MacQueen, 1973; Hoppe, 1987) given by Eq. (2) above. If random deaths are introduced into the model, the survivors are a multinomial sample from the limiting Dirichlet density; the limiting density of allele frequencies among the survivors is then also Dirichlet by a classical result (see, e.g., Hoppe, 1987).

2.2. *Exact Sampling Distribution of Allele Types*

If ξ individuals arrive on an island according to an immigration–birth process and $N \leqslant \xi$ survivors are selected at random from the arrivals, these

are a multinomial sample from the limiting Dirichlet density with frequencies α,

$$\mathsf{P}(n_1, ..., n_k) = N! \prod_{i=1}^{k} \frac{\alpha_i^{n_i}}{n_i!}. \tag{9}$$

The size distribution of families of unique allele types in a random sample of size N (or equivalently an island population of size N), averaging over all elements of $\mathbf{p}$, is then the compound multinomial Dirichlet (CMD) (Mosimann, 1962),

$$\mathsf{P}(n_1, ..., n_k) = \binom{N+\theta-1}{N}^{-1} \prod_{i=1}^{k} \binom{n_i+\theta p_i-1}{n_i}. \tag{10}$$

A proof of this result is given in the Appendix. This distribution is an exact result for a population (or sample) of size N and of any age.

The moment generating function (m.g.f.) of the CMD distribution is well-known and may be applied to study various properties of the distribution of alleles in a sample of individuals from an island, such as the mean and variance of α_i, the observed frequency of the ith allele, and the expectation and variance of f_0 the probability of IBD (Malécot, 1948). The results for f_0 are given below.

2.3. *Relation to Wright's Island Model*

The CMD distribution may also be obtained by taking a sample of N individuals (with replacement) from a large island population, at equilibrium, under Wright's discrete-generation island model. As noted previously, an important difference between Wright's model and the BDI model is that the CMD distribution arises under Wright's island model only for large populations at a genetic steady-state (equilibrium) which is generally not immediately achieved. An important question for future research is the relative importance of a constant population size assumption for Wright's island model.

Setting $\lambda = 1$ and $\phi = Nm$, shows that $\theta = 2\phi/\lambda$ for Wright's island model. This factor of two difference between the discrete and overlapping-generation models was first noted by Moran (1958) who studied a overlapping-generation genetic model with constant population size. The result is due to a difference in the variance of number of offspring per individual under the two models (reviewed by Tavaré, 1984).

3. Identity by Descent

In this Section I derive the expectation and variance, over replicate island populations, of the statistic $f_0^N = \sum_{i=1}^{k} n_i(n_i-1)/N(N-1)$, which is

the probability, conditional on the observed n_i, that two individuals randomly sampled from a population of size N are identical by descent. The derivation will make use of the random variables n_i which denote the number of individuals of the ith allele type obtained in a sample from an island population, where there are k allele types in total among immigrants. Define p_i to be the frequency of the ith allele type among immigrants and θ to be the scale parameter of the Dirichlet density function, which is equal to $\theta = \phi/\lambda$ for the haploid BDI model.

3.1. *Statistical Moments of n_i and α_i*

The statistical moments of the random variables n_i may be obtained using the following general formula (see Johnson and Kotz, 1969) for the factorial moments of the Dirichlet compound multinomial distribution:

$$\mu_{(r_1, r_2, \ldots, r_k)} = \mathsf{E}\left[\prod_{i=1}^{k} n_i^{(r_i)}\right] = N^{(\sum_{i=1}^{k} r_i)} \frac{\prod_{i=1}^{k} \theta p_i^{[r_i]}}{\theta^{[\sum_{i=1}^{k} r_i]}}; \tag{11}$$

where $n_i^{(r_i)} = n_i(n_i-1)\cdots(n_i-r_i+1)$; $N^{(\sum_{i=1}^{k} r_i)} = N(N-1)\cdots(N-\sum_{i=1}^{k} r_i+1)$; $\theta p_i^{[r_i]} = (\theta p_i + r_i - 1)\cdots\theta p_i$; and $\theta^{[\sum_{i=1}^{k} r_i]} = (\theta + \sum_{i=1}^{k} r_i - 1)\cdots\theta$. The relative frequency of the ith allele type on an island, denoted as α_i, for asymptotically large N follows a Dirichlet distribution with mixed moments given by the general formula (from Johnson and Kotz, 1972)

$$\mu_{(r_1, r_2, \ldots, r_k)} = \mathsf{E}\left[\prod_{j=1}^{k} \alpha_j^{r_j}\right] = \frac{\prod_{i=1}^{k} \theta p_i^{[r_i]}}{\theta^{[\sum_{i=1}^{k} r_i]}}. \tag{12}$$

Expectation of f_0. Taking expectations over the n_j gives

$$\mathsf{E}[f_0^N] = \frac{\sum_{j=1}^{k} \mathsf{E}[n_j^2] - N}{N(N-1)}. \tag{13}$$

Applying Eq. (11) above gives the following expression for the expectation of n_j^2:

$$\mathsf{E}[n_j^2] = \frac{(N-1)\, Np_j(\theta p_j + 1)}{\theta + 1} + Np_j. \tag{14}$$

Substituting Eq. (14) into Eq. (13) and simplifying gives

$$\mathsf{E}[f_0^N] = \frac{\theta[\sum_{i=1}^k p_i^2] + 1}{\theta + 1},$$

$$= \frac{\theta f_1 + 1}{\theta + 1}, \tag{15}$$

where $f_1 = \sum_{i=1}^k p_i^2$ is the probability of IBD for a random pair of immigrants.

Variance of f_0. In principle, one can derive a general formula for $\mathrm{var}[f_0^N]$ using the moments of the CMD distribution given above (Eq. (11)). The general formula is very complex and I will instead consider only the asymptotic variance of f_0^N as $N \to \infty$. First note that as N becomes large the statistic f_0^N converges to

$$\lim_{N \to \infty} f_0^N = \sum_{i=1}^k \hat{\alpha}_i^2, \tag{16}$$

where $\hat{\alpha}_i = n_i/N$ is the observed frequency of the ith allele type on the island. It has been shown that the relative frequencies of distinct types in a sample of size N from the CMD converge in distribution to the Dirichlet with parameters θ and $\mathbf{p}$ as N becomes large (Mosimann, 1962). The asymptotic variance of f_0^N is then

$$\lim_{N \to \infty} \mathrm{var}[f_0^N] = \mathsf{E}\left[\left(\sum_{i=1}^k \alpha_i^2 \right)^2 - \frac{(1 + \theta \sum_{i=1}^k \alpha_i^2)^2}{(1+\theta)^2} \right].$$

$$= \sum_{i=1}^k \mathsf{E}[\alpha_i^4] + 2 \sum_{j=2}^k \sum_{l=1}^j \mathsf{E}[\alpha_j^2 \alpha_l^2]$$

$$- \frac{(1 + \theta \sum_{i=1}^k p_i^2)^2}{(1+\theta)^2}. \tag{17}$$

Applying Eq. (12) above produces the following moments:

$$\mathsf{E}[\alpha_j^4] = \frac{p_j(1+\theta p_j)(2+\theta p_j)(3+\theta p_j)}{(1+\theta)(2+\theta)(3+\theta)}. \tag{18}$$

$$\mathsf{E}[\alpha_j^2 \alpha_l^2] = \frac{\theta p_j p_l (1+\theta p_j)(1+\theta p_l)}{(1+\theta)(2+\theta)(3+\theta)}. \tag{19}$$

Substituting these moments into Eq. (17) produces the following formula for the asymptotic variance of f_0^N,

$$\lim_{N\to\infty} \mathrm{var}[f_0^N] = \frac{\sum_{i=1}^{k} [p_i(1+\theta p_i)(2+\theta p_i)(3+\theta p_i)(1+\theta)]}{(1+\theta)^2(2+\theta)(3+\theta)} + \frac{2\sum_{j=2}^{k}\sum_{l=1}^{j-1} [\theta p_j p_l(1+\theta p_j)(1+\theta p_l)(1+\theta)]}{(1+\theta)^2(2+\theta)(3+\theta)} - \frac{(\theta\sum_{i=1}^{k} p_i^2+1)^2(2+\theta)(3+\theta)}{(1+\theta)^2(2+\theta)(3+\theta)} \tag{20}$$

For the k-allele model considered in this paper, with negligible mutation, identity by descent of individuals is equivalent to identity in state of alleles and $1-f_0$ is then a measure of heterozygosity. It follows that the expected heterozygosity (H) is

$$\mathsf{E}[H] = \frac{\theta(1-f_1)}{1+\theta} = \frac{\theta H_1}{1+\theta}, \tag{21}$$

where $H_1 = 1-f_1$ is the heterozygosity among immigrants. The variance of heterozygosity is given by Eq. (20). The expectation and variance of heterozygosity among islands are often of practical concern in problems of conservation genetics.

4. Discussion

Two assumptions of the existing mathematical models of genetic structure in subdivided populations have greatly restricted their usefulness in studying natural populations: (1) the steady-state or "equilibrium" assumption for the allele frequency distribution; (2) the constant population size assumption. The BDI process considered as a model of island populations in this paper does not require either of these assumptions and instead requires only that the allele frequency among immigrants is constant (also an assumption of Wright's discrete-generation island model). For the BDI model, the distribution of allele types in a sample from an island does not depend on the local population size or time, only the sample size, the array of allele frequencies among immigrants, and the parameter $\theta = \phi/\lambda$ (the ratio of the immigration rate to the individual birth rate).

For populations whose demographic structure may be described using a model based on a BDI process, a theory of genetic structure is possible that

does not require an assumption of "equilibrium" conditions since the steady-state sampling distribution is immediately generated by the process. One of the consequences is that more robust estimates of the gene flow parameter θ are possible using the CMD distribution described in this paper (Rannala and Hartigan, 1996). It is clear that the CMD distribution of allele types that arises under a k-allele BDI process is closely related to the celebrated Ewens sampling formula (Ewens, 1972). In particular, the Ewens formula may be generated by sampling from the limiting distribution of a BDI process in which all immigrants possess a unique allele type (see Hoppe, 1987).

It is well known that the steady-state distribution of alleles in a sample, or a population, generated by a Moran model with constant population size, linear mutation pressure, and an infinitely many neutral alleles model of mutation, is exactly described by the Ewens sampling formula (see Kelly, 1977). In this paper, it is shown that under linear migration pressure (with no mutation) the constant population size assumption may be relaxed without affecting the sampling distribution of neutral alleles.

The BDI process incorporates a basic form of "demographic stochasticity" in place of the constant population size assumption incorporated into the Fisher–Wright theory. This increases its applicability as a demographic–genetic model for many natural populations with overlapping generations. The mathematical properties of the BDI model also appear simpler than those of an equivalent Fisher–Wright model.

Appendix

Consider a haploid island population generated by a BDI process. During a small interval of time δt, the probability that a birth occurs in an island population of size N is $\lambda N \delta t$, the probability that a death occurs is $\mu N \delta t$, and the probability that an immigrant arrives is $\phi \delta t$. The probability of two or more events occuring is of order $o(\delta t)$ and may be neglected. The array of allele frequencies among immigrants is $\mathbf{p} = p_1, ..., p_k$, where there are k allele types among immigrants. Let $n_1, ..., n_k$ be the array of allele types in the island population, where n_i is the number of alleles of the ith type. It will be convenient to define $\theta = \phi/\lambda$.

Theorem. *For an island population of size N generated by a BDI process the distribution of allele types is*

$$P(n_1, ..., n_k \mid N, \theta, \mathbf{p}) = \binom{N+\theta-1}{N}^{-1} \prod_{i=1}^{k} \binom{n_i+\theta p_i-1}{n_i}. \qquad \text{(A-1)}$$

Proof. Assume that an island is initially empty and two arrivals occur according to an immigration–birth process. The probability that both individuals are of allele type i is

$$\frac{\theta p_i^2}{1+\theta}+\frac{p_i}{1+\theta}=\frac{p_i(\theta p_i+1)}{1+\theta}, \tag{A-2}$$

where the first term is the probability that two immigrants of type i arrive and the second term is the probability that one immigrant of type i arrives followed by a birth. The probability that one individual is of allele type i and the other is of allele type j is

$$\frac{2\theta p_i p_j}{1+\theta}, \tag{A-3}$$

where this is the probability that two immigrants arrive, one of allele type i and the other of allele type j, in either order. Applying these equations over k allele types generates a probability distribution identical to that obtained from Eq. (A-1) with $N=2$. Now consider an island population of size $N\geqslant 2$ generated by an immigration–birth process. Assume the distribution of allele types is given by Eq. (A-1). The population increases to size $N+1$ by a single arrival. The distribution of allele types in the population following the arrival is

$$\begin{aligned}\mathsf{P}(n_1, ..., n_k \mid N+1, \theta, \mathbf{p}) = &\sum_{i=1}^{k} \mathsf{P}(n_1, ..., n_i-1, ..., n_k \mid N, \theta, \mathbf{p}) \\ &\times \mathsf{P}(i_A \mid N, \theta, \mathbf{p}, n_i-1),\end{aligned} \tag{A-4}$$

where $\mathsf{P}(i_A \mid N, \theta, \mathbf{p}, n_i-1)$ is the transition probability from n_i-1 to n_i conditional on the population size N. The transition probability is

$$\mathsf{P}(i_A \mid N, \theta, \mathbf{p}, n_i-1)=\frac{\theta p_i+n_i-1}{N+\theta}, \tag{A-5}$$

where $\theta p_i/(N+\theta)$ is the probability an immigrant of the ith allele type arrives and $(n_i-1)/(N+\theta)$ is the probability an individual is born of the ith allele type. Substituting Eq. (A-5) into Eq. (A-4) gives

$$\begin{aligned}\mathsf{P}(n_1, ..., n_k \mid N+1, \theta, \mathbf{p}) &= \binom{N+\theta-1}{N}^{-1} \sum_{i=1}^{k} \binom{n_i+\theta p_i-2}{n_i-1} \frac{\theta p_i+n_i-1}{N+\theta} \\ &\quad \times \prod_{j\neq i}^{k} \binom{n_j+\theta p_j-1}{n_j}. \\ &= \binom{[N+1]+\theta-1}{[N+1]}^{-1} \prod_{i=1}^{k} \binom{n_i+\theta p_i-1}{n_i}.\end{aligned} \tag{A-6}$$

To consider the effect of random deaths, assume a population of size $N+1$ has been generated by an immigration–birth process and has the distribution of allele types given by Eq. (A-1). Following a single death, the distribution of allele types in the population of N survivors is

$$\mathsf{P}(n_1, ..., n_k \mid N, \theta, \mathbf{p}) = \sum_{i=1}^{k} \mathsf{P}(n_1, ..., n_i+1, ..., n_k \mid N+1, \theta, \mathbf{p}) \times \mathsf{P}(i_D \mid N+1, \theta, \mathbf{p}, n_i+1), \quad \text{(A-7)}$$

where $\mathsf{P}(i_D \mid N+1, \theta, \mathbf{p}, n_i+1)$ is the transition probability from n_i+1 to n_i conditional on the population size $N+1$. The transition probability is

$$\mathsf{P}(i_D \mid N+1, \theta, \mathbf{p}, n_i+1) = \frac{n_i+1}{N+1}. \quad \text{(A-8)}$$

Substituting Eq. (A-8) into Eq. (A-7) gives

$$\begin{aligned}\mathsf{P}(n_1, ..., n_k \mid N, \theta, \mathbf{p}) &= \binom{N+\theta}{N+1}^{-1} \sum_{i=1}^{k} \binom{n_i+\theta p_i}{n_i+1} \frac{n_i+1}{N+1} \times \prod_{j \neq i}^{k} \binom{n_j+\theta p_j-1}{n_j} \\ &= \binom{N+\theta-1}{N}^{-1} \prod_{i=1}^{k} \binom{n_i+\theta p_i-1}{n_i}. \quad \text{(A-9)}\end{aligned}$$

It follows that a population of size N generated by an arbitrary sequence of arrivals and deaths of a BDI process has a distribution of allele types given by Eq. (A-1). ▮

COROLLARY. *If a random sample of N individuals is collected from a population of size $\xi \geqslant N$, generated by a BDI process, the distribution of allele types in the sample is given by Eq.* (A-1).

Proof. A random sample of N individuals is exactly equivalent to $\xi - N$ deaths occurring in sequence in a population of size ξ generated by a BDI process. The result follows from the properties of the death process considered in the proof of Theorem 1.

ACKNOWLEDGMENTS

I thank J. A. Hartigan for his many insightful comments and suggestions. Financial support for this work was provided by the National Science Foundation (DEB-9224949), the John F. Enders Research Grant of Yale University, and Sigma Xi. Personal support was provided by

a Natural Sciences and Engineering Research Council (NSERC) of Canada postgraduate scholarship and a fellowship from Yale University.

References

Blackwell, D., and MacQueen, J. B. 1973. Ferguson distributions via Polya urn schemes, *Ann. Statist.* **1**, 353–355.

Ewens, W. J. 1972. The sampling theory of selectively neutral alleles, *Theor. Popul. Biol.* **3**, 87–112.

Hanski, I. 1994. Patch-occupancy dynamics in fragmented landscapes, *Trends Ecol. Evol.* **9**, 131–135.

Hoppe, F. M. 1987. The sampling theory of neutral alleles and an urn model in population genetics, *J. Math. Biol.* **25**, 123–159.

Johnson, N. L., and Kotz, J. S. 1969. "Discrete Distributions," Houghton Mifflin, Boston.

Johnson, N. L., and Kotz, S. 1972. "Distributions in Statistics: Continuous Multivariate Distributions," Wiley, New York.

Kelly, F. P. 1977. Exact results for the Moran neutral allele model, *Adv. Appl. Probab.* **9**, 197–201.

Kendall, D. G. 1948. On some modes of population growth leading to R. A. Fisher's logarithmic series distribution, *Biometrika* **35**, 6–15.

Kendall, D. G. 1949. Stochastic processes and population growth, *J. Roy. Statist. Soc. B* **11**, 230–264.

Kendall, D. G. 1975. Some problems in mathematical genealogy, *in* "Perspectives in Probability and Statistics: Papers in Honour of M. S. Bartlett on the Occasion of His Sixty-Fifth Birthday" (J. Gani, Ed.) pp. 325–348, Applied Probability Trust, New York.

Kimura, M. 1953. "Stepping-stone" model of population, *Ann. Report Nat. Inst. Genet.* **3**, 62–63.

Kimura, M., and Weiss, G. H. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance, *Genetics* **49**, 561–576.

Lande, R. 1992. Neutral theory of quantitative genetic variance in an island model with local extinction and colonization, *Evolution* **46**, 381–389.

Malécot, G. 1948. "Les mathématiques de l'hérédité," Freeman, San Francisco, CA; 1969 "The Mathematics of Heredity," D. M. Yermanos, trans., 1969.

Maruyama, T. 1970. Effective number of alleles in a subdivided population, *Theor. Popul. Biol.* **1**, 273–306.

Mode, C. J. 1962. Some multi-dimensional birth and death processes and their applications in population genetics, *Biometrics* **18**, 543–567.

Moran, P. A. P. 1958. The general theory of the distribution of gene frequencies. I. overlapping generations. II. non-overlapping generations, *Proc. Roy. Soc. London B* **149**, 102–116.

Mosimann, J. E. 1962. On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions, *Biometrika* **49**, 65–82.

Pulliam, H. R. 1988. Sources, sinks and population regulation, *Am. Nat.* **132**, 652–661.

Rannala, B., and Hartigan, J. A. 1995. Identity by descent in island–mainland populations, *Genetics* **139**, 429–437.

Rannala, B., and Hartigan, J. A. 1996. Estimating gene flow in island populations, *Genet. Res.*, to appear.

Tavaré, S. 1984. Line-of-descent and genealogical processes, and their applications in population genetics models, *Theor. Popul. Biol.* **26**, 119–164.

Tavaré, S. 1987. The birth process with immigration, and the genealogical structure of large populations, *J. Math. Biol.* **25**, 161–168.

TAVARÉ, S. 1989. The genealogy of the birth, death, and immigration process, *in* "Mathematical Evolutionary Theory" (M. Feldman, Ed.), pp. 41–56, Princeton Univ. Press, Princeton, NJ.

WHITLOCK, M. C. 1992. Temporal fluctuations in demographic parameters and the genetic variance among populations, *Evolution* **46**, 608–615.

WRIGHT, S. 1931. Evolution in Mendelian populations, *Genetics* **16**, 97–159.

WRIGHT, S. 1940. Breeding structure of populations in relation to speciation, *Amer. Nat.* **74**, 232–248.

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

WRIGHT, S. 1969. "Evolution and Genetics of Populations. The Theory of Gene Frequencies," Vol. 2, Univ. of Chicago Press, Chicago.

Printed in Belgium