

Probability of Identity by Descent in Metapopulations

Ingemar Kaj* and Martin Lascoux†

*Department of Mathematics, Uppsala University, S-751 06 Uppsala, Sweden and †Department of Genetics, Uppsala University, S-750 07 Uppsala, Sweden

Manuscript received November 29, 1998

Accepted for publication April 8, 1999

ABSTRACT

Equilibrium probabilities of identity by descent (IBD), for pairs of genes within individuals, for genes between individuals within subpopulations, and for genes between subpopulations are calculated in metapopulation models with fixed or varying colony sizes. A continuous-time analog to the Moran model was used in either case. For fixed-colony size both propagule and migrant pool models were considered. The varying population size model is based on a birth-death-immigration (BDI) process, to which migration between colonies is added. Wright's F statistics are calculated and compared to previous results. Adding between-island migration to the BDI model can have an important effect on the equilibrium probabilities of IBD and on Wright's index.

WRIGHT'S island model considers "a population subdivided into random breeding islands with populations of size N of which the proportion m consists of immigrants that may be considered a random sample of the total species" (Wright 1969). The model also assumes that (1) the populations have discrete nonoverlapping generations, (2) individual populations do not fluctuate in size, and (3) migration is uniform over all islands. Under those assumptions and at steady state, Wright showed that F_{ST} , the correlation between uniting gametes, could be written as a simple function of Nm , namely

$$F_{ST} = \frac{1}{1 + 4Nm}. \quad (1)$$

Also implicitly assumed is that the rate of mutation is weak relative to the stronger rates of migration and random genetic drift. Despite the demanding assumptions of this model, Equation 1 has been widely used to obtain indirect estimates of gene flow among populations.

Wright's F statistics, F_{IS} and F_{ST} , where the subindices refer to individuals, subpopulations, and total population, are expressed in terms of probability of identity by descent (IBD) as

$$F_{IS} = \frac{\theta_1 - \theta_2}{1 - \theta_2}$$

and

$$F_{ST} = \frac{\theta_2 - \theta_3}{1 - \theta_3},$$

where the θ_i 's are the probabilities of IBD (Malécot 1948), θ_1 for pairs of genes within individuals, θ_2 for genes between individuals within subpopulations, and θ_3 for genes between subpopulations (Cockerham and Weir 1987; Rousset 1996); precise definitions are given below. The index F_{IS} measures the extent to which genetic variation in a subpopulation is explained by variation within individuals, and F_{ST} similarly the degree to which the total genetic variation is accounted for by variation within subpopulations.

Recently, attempts have been made to relax some of the assumptions of Wright's island model (see McCauley 1993 for a review). Among these, Maruyama and Tachida (1992) derived an equation analogous to Equation 1 in the case of a metapopulation with a discrete time propagule model of extinction/recolonization (each extinct colony is replaced by a copy of a randomly selected existing colony) and selfing. This was achieved by deriving recursion equations for θ_1 , θ_2 , and θ_3 and calculating the equilibrium values. On their part, Rannala and Hartigan (1995) investigated the effect on IBD probabilities of fluctuating population sizes in a continuous time setting. They assumed a collection of disjunct patches inhabited by haploid individuals whose number changes over time due to mortality, birth, and immigration from a large external population and studied the corresponding demographic model based on a linear birth-death-immigration process. Because there is no migration between subpopulations in the model their study refers basically to a single colony population. It was found that given the size N of the system at some point t in time the conditional probability $\theta_2 = \theta_2(t, N)$ only depends on the birth and immigration rates but not on the death rate and not on N or t . This led to some conclusions on the applicability of their results also to transient situations where the popu-

Corresponding author: Martin Lascoux, Department of Genetics, Box 7003, S-750 07 Uppsala, Sweden.
E-mail: martin.lascoux@genetik.uu.se

lation size has either not yet reached a stationary distribution or does not settle into a steady state.

In the present article we extend both models. In the first part a continuous-time version of the Maruyama and Tachida model is given and extended to a migrant pool extinction model. Detailed arguments are provided for deriving the systems of ordinary differential equations satisfied by the IBD probabilities, starting out from the basic mutation/reproduction balance and successively adding migration and extinction mechanisms. In the second part of the article, migration between islands and selfing are added to a diploid version of Rannala and Hartigan's model. An approximation for Wright's index at population genetic equilibrium is found. It is shown in particular that when there is migration among islands as well as immigration from a large source with a fixed probability of IBD (the situation modeled by Rannala and Hartigan), the death rate may have an important effect on probabilities of IBD and F_{ST} .

GENERALITIES AND SCOPE

Our purpose is to model and study genetic diversity at individual, colony, and total population levels. The distribution of genetic diversity in the population is affected by five agencies: (1) mutation, (2) reproduction, (3) extinction and subsequent replacement of colonies, (4) migration between colonies and/or from an external source, and (5) population size.

We assume that mutation follows an infinite-allele model (Kimura and Crow 1964), which assumes that every mutation creates an allele that is not currently existing in the population. Under this model, two alleles that are identical in state are also identical by descent. Two types of reproduction systems are considered: asexual reproduction and sexual reproduction. Sexual reproduction includes both random and mixed mating (part of the population mating at random, the other part selfing). An extinct colony can be replaced by a given number of individuals originating from one of the existing colonies (propagule model; Slatkin 1977) or by a given number of individuals randomly sampled over the colonies (migrant pool model; Slatkin 1977). In either case, the initial size of the population is restored instantly. Alternatively, colonies can be gradually recolonized, allowing varying population sizes. Finally, colonies can exchange individuals (seeds) or single gametes (pollen).

We consider two models of basically different character, a fixed-size model where the colonies are equally large and of given size, and a variable-size model in which the sizes of the subpopulations are given by birth-and-death processes. We introduce next the various parameters and the precise dynamics in both models.

Model parameters: In both models there are n colonies. For the fixed-size model we let N denote the num-

ber of diploid individuals in each colony and k be the number of colonists. The distribution of genetic material evolves according to the nonnegative Poisson intensities

- u , mutation of an allele into a new type per gene,
- $1/2$, reproduction/compensatory death per individual,
- λ , extinction/recolonization per colony,
- m_s , seed migration between colonies per individual.

In addition, two probabilities enter as parameters, namely,

- s , selfing probability,
- m_p , pollen migration probability between colonies during mating.

The corresponding set of parameters for the variable-size model are the Poisson intensities

- u , mutation of an allele into a new type per gene,
- $1/2$, reproduction per individual,
- δ , death per individual,
- β , immigration of a new individual from the mainland per colony,
- m_s , seed migration between colonies per individual,

and the selfing probability s , $0 \leq s \leq 1$. The reproduction intensity of $1/2$ was retained to conform with the usual choice of normalization. In that case, with probability $h/2 + o(h)$ one individual, and hence with probability $h + o(h)$ one of two alleles in different individuals considered at time $t + h$, is the result of reproduction during $(t, t + h]$.

Mutation acts in the same way in both models. It is the basic cause of reduction in identity by state of the genes in the population. The other mechanisms require more precise descriptions.

Reproduction: Obviously the main difference between fixed and variable size is that in the first case the Moran model mechanism of automatic compensating deaths following the formation of every new individual applies (Moran 1958), whereas in the latter a birth event more realistically increases the size of the population by one unit. In the variable-size case the assumption $\delta > 1/2$ will guarantee that the population does not grow without bound and the assumption $\beta > 0$ will prevent permanent extinction.

To define reproduction in the fixed-size model we distinguish first of all whether reproduction occurs by random mating or asexually, and treat these cases separately. In both cases we assume the Moran model mechanism of automatic compensating deaths.

The asexual case is simpler. Here, in the event that the Poisson intensity signals that an individual is going to reproduce, a genetic copy of this diploid individual is produced followed by the death of one of the N original individuals. It is worth noting that the killed

individual may thus be the one which itself triggered the production of the new.

The more complicated sexual reproduction case also involves the selfing parameter s as follows. Consider one of the N individuals and suppose its Poisson clock goes off. This we understand as an instruction to first make a copy of one of its own genes that is to form the first gene in the pair of genes representing the new individual. To find the second gene in the new pair there are two possibilities. Namely, with probability s the reproducing individual selects the second gene again to be a copy of one of its own, and with probability $1 - s$ it selects the second gene to be a copy of any one of the $2N$ existing genes sampled uniformly in the subpopulation. The actual selfing probability is therefore $s + (1 - s)/N$ rather than s . Again the formation in this manner of the new individual is instantly followed by a compensatory death.

In the variable-size model, for simplicity we restrict the analysis to sexual reproduction only.

Extinction/recolonization: As already stated we cover the two schemes introduced by Slatkin (1977) within the fixed-size model. The first mechanism is the so-called propagule recolonization, which means that upon extinction of a colony, k individuals are picked randomly in one of the n existing colonies to replace it. The second is the migrant pool. The difference is that now the replacement colony is put together by sampling k individuals uniformly over the whole population. Cases intermediate between those two extreme models are also taken care of by considering the probability of common origin of two colonists (Whitlock and McCauley 1990). In all cases, the k individuals then reproduce to restore the size of the population.

For the variable-size model one can say that deaths and immigration replace extinction/recolonization. Even if colonies may go extinct from time to time, influx from an external source is guaranteed. The effect of β is that external individuals enter into each colony independently, and an additional assumption will then be made regarding the genetic relationship between existing and immigrating gametes.

Migration: To each individual is associated the seed migration intensity m_s , which, when its Poisson clock rings, forces this individual to select randomly and uniformly one of the existing individuals in the population and then exchange position with the particular individual that was chosen. For example, in the fixed-size model this yields a probability $1/nN$ that nothing at all happens. Pollen migration, on the other hand, is a refinement of the random mating mechanism. We let m_p denote the probability that one of the genes composing a newly created individual due to pollen migration should be considered as being selected randomly in the full existing population and not only within its own colony.

Now we introduce more formally the IBD probabili-

ties that we wish to study in this work. Pick two different alleles, A and B , randomly in the population. We consider the probabilities θ_k , $k = 1, 2, 3$, that A and B are identical by descent given their hierarchical "distance" k ,

- θ_1 , if A and B belong to the same individual,
- θ_2 , if A and B belong to different individuals in the same colony,
- θ_3 , if A and B are taken from different colonies in the population.

FIXED-POPULATION SIZE MODEL

Mutation and reproduction: We consider the IBD probabilities $\theta_k(t)$, $t \geq 0$, $k = 1, 2, 3$, as functions of time, and investigate to what extent they change during a small interval of time $(t, t + h]$, $h > 0$. Pick two alleles at hierarchical distance k from each other and observe them at time $t + h$.

First suppose that only mutation is in effect. Because the probability of a mutation in $(t, t + h]$ of one of the selected genes is proportional to $2uh$,

$$\theta_k(t + h) = 2uh \cdot 0 + (1 - 2uh)\theta_k(t) + o(h);$$

hence,

$$\frac{\theta_k(t + h) - \theta_k(t)}{h} = -2u\theta_k(t) + \frac{o(h)}{h},$$

and letting h tend to zero,

$$\theta'_k(t) = -2u\theta_k(t).$$

The solution

$$\theta_k(t) = \theta_k(0)e^{-2ut}, \quad t \geq 0 \tag{2}$$

ultimately tends to zero as a result of the assumed infinite-allele model.

Now we include reproduction, which stabilizes the genetic structure in the population and counteracts the diversifying effect of mutation through genetic drift toward fixation. The balance of mutation and genetic drift then leads to stationary probabilities θ_1 and θ_2 . However, there is no genetic drift over colony borders, and hence for $k = 3$ the only solution is still $\theta_3 = 0$.

Asexual reproduction: Assume that all reproduction is asexual. Then the genetic relationship between two genes in a single individual does not change in case that individual was just formed as a copy of a previously existing individual. Mutation, on the other hand, will affect their interrelationship, forcing the IBD probability to vanish asymptotically as in (2). Hence the only stationary solution for $\theta_1(t)$ is $\theta_1 = 0$.

For $\theta_2(t)$ note that with probability $h + o(h)$ one of the chosen genes belongs to an individual that during $(t, t + h]$ was produced by asexual reproduction. Then one of the two individuals involved, call it A , existed at time t and the other one, call it B , has been replaced by a copy of a randomly picked individual that existed

at time t . With probability $1/N$, B happens to be a copy of A in which case the two chosen genes must be considered belonging to the same individual. In this last case we have selected one gene in A and one gene in the copy of A called B . But the probability that two different genes in the same individual are IBD at t is $\theta_1(t)$, and thus our two genes are IBD with probability $1/2 + \theta_1(t)/2$. With probability $1 - 1/N$ their IBD probability remains at $\theta_2(t)$. Hence,

$$\theta_2(t+h) = \left[h \left(\frac{1}{N} \frac{1}{2} + \frac{1}{2} \theta_1(t) \right) + \left(1 - \frac{1}{N} \right) \theta_2(t) \right] + (1-h)\theta_2(t) \cdot (1-2uh) + 2uh \cdot 0 + o(h),$$

so that

$$\theta_2(t+h) - \theta_2(t) = -2uh\theta_2(t) + h \frac{1}{N} \left(\frac{1}{2} + \frac{1}{2} \theta_1(t) - \theta_2(t) \right) + o(h) \tag{3}$$

and thus

$$\theta_2'(t) = -2u\theta_2(t) + \frac{1}{N} \left(\frac{1}{2} + \frac{1}{2} \theta_1(t) - \theta_2(t) \right).$$

The stationary solution for $\theta_2(t)$ is obtained by taking $\theta_2'(t) = \theta_1'(t) = 0$ above; thus

$$\theta_2 = \frac{1}{2(1+2uN)},$$

which is half the value of the stationary solution of $\theta_2(t)$ in a haploid model.

Sexual reproduction: We start by deriving an equation for $\theta_2(t)$. Consider at time $t+h$ two alleles, a and b , from two different individuals in the same colony and suppose one of them, a say, belongs to an individual that was created sexually in the same time interval elapsed since time t . With probability $1/N$, allele a is copied from the individual to which b belongs, in which case they are identical by descent with probability $1/2 + \theta_1(t)/2$. With probability $1 - 1/N$, the gene a was copied elsewhere; hence their IBD probability remains $\theta_2(t)$. Some reflection shows that these same relations hold whether or not the individual decided to self. Hence,

$$\theta_2(t+h) = \left[h \left(\frac{1}{N} \frac{1}{2} + \frac{1}{2} \theta_1(t) \right) + \left(1 - \frac{1}{N} \right) \theta_2(t) \right] + (1-h)\theta_2(t) \cdot (1-2uh) + 2uh \cdot 0 + o(h).$$

This is the same equation as the one obtained for the asexual case, thus again leading to

$$\theta_2'(t) = -2u\theta_2(t) + \frac{1}{N} \left(\frac{1}{2} + \frac{1}{2} \theta_1(t) - \theta_2(t) \right). \tag{4}$$

The corresponding equation for $\theta_1(t)$, however, is different. In fact, by splitting selfing from nonselfing,

weighted by probabilities s and $1 - s$, we find

$$\begin{aligned} \theta_1(t+h) = & \left[\frac{h(1-s)}{2} \left(\frac{1}{N} \frac{1}{2} + \frac{1}{2} \theta_1(t) \right) + \left(1 - \frac{1}{N} \right) \theta_2(t) \right] \\ & + \frac{hs}{2} \left(\frac{1}{2} + \frac{1}{2} \theta_1(t) \right) + \left(1 - \frac{h}{2} \right) \theta_1(t) \cdot (1-2uh) \\ & + 2uh \cdot 0 + o(h), \end{aligned}$$

leading to

$$\begin{aligned} \theta_1'(t) = & -\frac{1}{2}(1+4u)\theta_1(t) + \frac{1}{2}\theta_2(t) \\ & + \frac{1}{2} \left(\frac{1-s}{N} + s \right) \left(\frac{1}{2} + \frac{1}{2} \theta_1(t) - \theta_2(t) \right). \end{aligned} \tag{5}$$

Relation (4) shows that at equilibrium

$$\theta_2 = \frac{1 + \theta_1}{2(1 + 2uN)}. \tag{6}$$

By inserting this into (5) taking $\theta_1'(t) = 0$ we obtain the stationary solutions for mutation/genetic drift balance. They can be written

$$\begin{aligned} \theta_1 = & \left(1 + 2uN \right. \\ & \left. \cdot \frac{s(1-1/N) - 1/N}{1+4u} \right) / \left(1 + 2uN \left(2 - \frac{s(1-1/N) - 1/N}{1+4u} \right) \right) \end{aligned} \tag{7}$$

$$\theta_2 = 1 / \left(1 + 2uN \left(2 - \frac{s(1-1/N) - 1/N}{1+4u} \right) \right). \tag{8}$$

Note that the case

$$s = \frac{1/N}{1 - 1/N} = \frac{1}{N-1},$$

which corresponds to an actual selfing rate $s + (1 - s)/N = 2/N$, leads to the traditional expressions

$$\theta_1 = \theta_2 = \frac{1}{1 + 4uN}.$$

Turning to Wright's indices, we obtain for the asexual case described previously,

$$F_{is} = -\frac{1}{1 + 4uN},$$

and for the sexual system (7, 8),

$$F_{is} = \frac{s(1-1/N) - 1/N}{2(1+4u) - s(1-1/N) + 1/N}. \tag{9}$$

Negative values of F_{is} , which occur for $s < 1/(N-1)$ in the sexual case, indicate that two genes in different individuals are more likely to be IBD than two genes in the same individual.

Adding migration and extinction: The next step toward enriching the applicability of the model and taking

into account the colony structure that until now did not enter the derivations, is to add migration and extinction.

Seed migration: To include seed migration we must consider the equations for $\theta_2(t)$ and $\theta_3(t)$, noting that migration of individuals does not affect $\theta_1(t)$.

Consider first a pair of genes selected in the same colony. Then the intensity of a migration event in $(t, t + h]$, which corresponds to selecting the two genes, in fact, from different colonies is $2m_s(1 - 1/n)$. Hence

$$\theta_2(t + h) = 2m_s \left(1 - \frac{1}{n}\right) h \theta_3(t) + \left(1 - 2uh - 2m_s \left(1 - \frac{1}{n}\right) h\right) \cdot \left[h \frac{1}{N} \left(\frac{1}{2} + \frac{1}{2} \theta_1(t) - \theta_2(t)\right) + \theta_2(t) \right] + o(h). \quad (10)$$

Similarly, if two genes are selected in different colonies at time $t + h$ the intensity is $2m_s h/n$ such that they are considered to be stemming from the same colony at time t due to migration during $(t, t + h]$. Hence

$$\theta_3(t + h) = \theta_3(t) - 2(nu + m_s) \frac{1}{n} h \theta_3(t) + 2m_s \frac{1}{n} h \theta_2(t) + o(h).$$

Together with the previously derived Equation 5 for $\theta_1(t)$, these relations yield in the limit $h \rightarrow 0$ the coupled system of equations

$$\theta_1'(t) = -\frac{1}{2}(1 + 4u)\theta_1(t) + \frac{1}{2}\theta_2(t) + \frac{1}{2} \left(\frac{1-s}{N} + s\right) \left(\frac{1}{2} + \frac{1}{2}\theta_1(t) - \theta_2(t)\right) \quad (11)$$

$$\theta_2'(t) = -2 \left(u + m_s \left(1 - \frac{1}{n}\right)\right) \theta_2(t) + 2m_s \left(1 - \frac{1}{n}\right) \theta_3(t) + \frac{1}{N} \left(\frac{1}{2} + \frac{1}{2}\theta_1(t) - \theta_2(t)\right) \quad (12)$$

$$\theta_3'(t) = -2(nu + m_s) \frac{1}{n} \theta_3(t) + 2m_s \frac{1}{n} \theta_2(t). \quad (13)$$

By solving the system of Equations 11, 12, and 13 with vanishing left-hand terms for its stationary solution $(\theta_1, \theta_2, \theta_3)$ in steady state, we find the equilibrium relation

$$\theta_2 = \frac{1 + \theta_1}{2(1 + 2u\alpha N)}, \quad (14)$$

which replaces (6), and then

$$\begin{aligned} \theta_1 &= \frac{1 + 2u\alpha(s(N-1) + 1)}{(1 + 4u)(1 + 4u\alpha N) + 4u - 2u\alpha(s(N-1) + 1)} \\ \theta_2 &= \frac{1 + 4u}{(1 + 4u)(1 + 4u\alpha N) + 4u - 2u\alpha(s(N-1) + 1)} \\ \theta_3 &= \frac{m_s}{m_s + nu} \theta_2, \end{aligned} \quad (15)$$

where

$$\alpha = \frac{nm_s + nu}{m_s + nu}.$$

Pollen migration: Introduce the total migration parameter

$$m = m_s + \frac{1}{2}m_p.$$

To include migration of pollen, we start again with $\theta_1(t)$. During the mating phase the probability is m_p that one gene ends up in the new individual from a migrating pollen; hence the probability is $m_p(1 - 1/n)$ that it originated in another colony. Therefore the term $m_p(1 - 1/n)\theta_3(t)$ enters the differential equation for $\theta_1(t)$, and the corresponding reduction occurs for the other, random mating terms, more exactly

$$\begin{aligned} \theta_1'(t) &= -\frac{1}{2}(1 + 4u)\theta_1(t) + \frac{1}{2} \left(1 - m_p \left(1 - \frac{1}{n}\right)\right) \theta_2(t) \\ &+ \frac{1}{2} \left(\frac{1-s}{N} + s\right) \left(1 - m_p \left(1 - \frac{1}{n}\right)\right) \left(\frac{1}{2} + \frac{1}{2}\theta_1(t) - \theta_2(t)\right) \\ &+ \frac{1}{2} m_p \left(1 - \frac{1}{n}\right) \theta_3(t). \end{aligned} \quad (16)$$

The differential equation for $\theta_2(t)$ modifies similarly. Exploiting the notation m introduced above we see that Equation 12 changes into

$$\begin{aligned} \theta_2'(t) &= -2 \left(u + m \left(1 - \frac{1}{n}\right)\right) \theta_2(t) + 2m \left(1 - \frac{1}{n}\right) \theta_3(t) \\ &+ \frac{1}{N} \left(1 - m_p \left(1 - \frac{1}{n}\right)\right) \left(\frac{1}{2} + \frac{1}{2}\theta_1(t) - \theta_2(t)\right). \end{aligned} \quad (17)$$

Then we need to take into account that, because of pollen migration, reproductions in $(t, t + h]$ now will affect θ_3 . With probability m_p/n the selected genes belong to the same colony. This leads to

$$\begin{aligned} \theta_3'(t) &= -2(nu + m) \frac{1}{n} \theta_3(t) + 2m \frac{1}{n} \theta_2(t) \\ &+ \frac{m_p}{n} \frac{1}{N} \left(\frac{1}{2} + \frac{1}{2}\theta_1(t) - \theta_2(t)\right). \end{aligned} \quad (18)$$

In steady state the relation between θ_2 and θ_3 is found from (17) and (18),

$$\theta_3 = \frac{m + m_p u}{m + nu(1 - m_p(1 - 1/n))} \theta_2, \quad (19)$$

and the analog of system (15) is obtained similarly. In particular, it is seen that Equation 14 still holds in this case with α replaced by

$$\alpha' = \frac{nm + nu}{m + nu(1 - m_p(1 - 1/n))}.$$

To complete the analysis of the migration model, it remains to do the algebraic manipulations leading to the following expressions for Wright's indices:

$$F_{IS} = [2(u + m)g(s) + [(4m_s + m_p)(1 - m_p) \cdot (1 - 1/n)/(1 + 4u)](1/N)] / [2(u + m)(2 - g(s)) - [(4m_s + m_p)(1 - m_p)(1 - 1/n)/(1 + 4u)](1/N)]$$

and

$$F_{ST} = [1 - m_p] / [(1 - m_p)(1 - (4m_s + m_p)(1 - 1/n)/(1 + 4u)) + 2(u + m)N(2 - g(s))], \tag{20}$$

where

$$g(s) = \frac{1 - m_p(1 - 1/n)}{1 + 4u} \left(s \left(1 - \frac{1}{N} \right) - \frac{1}{N} \right).$$

In particular, for $m_p = 0$,

$$F_{ST} = 1 / \left[1 - \frac{4m_s(1 - 1/n)}{(1 + 4u)} + 2(u + m_s)N / \left(2 - \frac{[s(1 - 1/N) - 1/N]}{(1 + 4u)} \right) \right].$$

Extinction/recolonization: The recolonization event consists of two steps that are supposed to occur instantly in case of extinction. First k , $1 \leq k \leq N$, founder individuals, or colonists, are selected from a single colony (propagule model) or randomly over the n colonies (migrant pool model). The $N - k$ additional individuals are produced by sexual reproduction from the k colonists. To avoid unnecessarily complicated expressions we restrict to the case $m_p = 0$, ignoring pollen migration.

The simplest case is the propagule model with $k = N$ colonizers, in which an existing colony is copied and reinstalled in full to account for recolonization in the event of extinction. It is clear that extinction and the consecutive reinstallation of a colony have no effect upon θ_1 or θ_2 . Hence we derive the stationary solution for $\theta_3(t)$, taking into consideration the possibility that one of the two colonies representing the pair of genes selected for the determination of $\theta_3(t + h)$ goes extinct during $(t, t + h]$, the probability of which is $2\lambda h$. The probability is $1/n$ that the particular replacement colony is chosen to which the other gene of the given pair belongs, and the probability is $1/nN$ that even the same individual is chosen to which the other gene of the given pair belongs. Thus

$$\begin{aligned} \theta_3(t + h) &= \theta_3(t) - 2(nu + \lambda + m_s) \frac{1}{n} h \theta_3(t) \\ &\quad + 2(\lambda + m_s) \frac{1}{n} h \theta_2(t) \\ &\quad + 2\lambda \frac{1}{n} h \left(\frac{1}{2} + \frac{1}{2} \theta_1(t) - \theta_2(t) \right) + o(h) \end{aligned}$$

and so, replacing Equation 13,

$$\theta_3'(t) = -2(nu + \lambda + m_s) \frac{1}{n} \theta_3(t) + 2(\lambda + m_s) \frac{1}{n} \theta_2(t)$$

$$+ \frac{2\lambda}{nN} \left(\frac{1}{2} + \frac{1}{2} \theta_1(t) - \theta_2(t) \right). \tag{21}$$

The modified equation for $\theta_3(t)$ results in the equilibrium probabilities

$$\begin{aligned} \theta_1 &= \frac{1 + 2u\alpha(s(N - 1) + 1)}{(1 + 4u)(1 + 4u\alpha N) + 4u - 2u\alpha(s(N - 1) + 1)} \\ \theta_2 &= \frac{1 + 4u}{(1 + 4u)(1 + 4u\alpha N) + 4u - 2u\alpha(s(N - 1) + 1)} \\ \theta_3 &= \frac{\lambda + m_s + 2\lambda u + 2\lambda m_s(1 - 1/n)}{\lambda + m_s + nu + 2\lambda m_s(1 - 1/n)} \theta_2, \end{aligned} \tag{22}$$

where

$$\alpha = \frac{\lambda + nm_s + nu}{\lambda + m_s + nu + 2\lambda m_s(1 - 1/n)}.$$

Moreover,

$$\begin{aligned} F_{IS} &= \left[\left(\frac{\lambda}{n} + u + m_s \right) \left(s \left(1 - \frac{1}{N} \right) - \frac{1}{N} \right) + 2m_s \left(1 - \frac{1}{n} \right) \right. \\ &\quad \cdot \left(1 - \frac{2\lambda}{n} \frac{1}{N} \right) / \left[\left(\frac{\lambda}{n} + u + m_s \right) \left(2(1 + 4u) \right) \right. \\ &\quad \left. \left. - s \left(1 - \frac{1}{N} \right) + \frac{1}{N} \right) - 2m_s \left(1 - \frac{1}{n} \right) \left(1 - \frac{2\lambda}{n} \frac{1}{N} \right) \right] \end{aligned}$$

and

$$\begin{aligned} F_{ST} &= \left(1 - \frac{2\lambda}{n} \right) / \left[\left(1 - \frac{2\lambda}{n} \right) \left(1 - \frac{4m_s(1 - 1/n)}{1 + 4u} \right) \right. \\ &\quad \left. + 2 \left(\frac{\lambda}{n} + u + m_s \right) \right. \\ &\quad \left. \cdot N \left(2 - \frac{s(1 - 1/N) - 1/N}{1 + 4u} \right) \right]. \end{aligned} \tag{23}$$

We observe that the case $\lambda > n/2$ is a nonbiological regime where the effect of recolonization is so strong that it is more likely for two genes to be identical by descent if they are picked from separate colonies rather than from the same. Restricting to $\lambda \leq n/2$, we note that F_{ST} is decreasing as a function of λ , u , and m_s . The effect of increasing the number of colonies n is to reduce the effect of extinctions and enhance the effect of migration on F_{ST} .

Next we turn to the general propagule and migrant pool schemata with k colonizers. The Equations 11 for $\theta_1(t)$ and 21 for $\theta_3(t)$ with $k = N$ remain unchanged, whereas Equation 12 for $\theta_2(t)$ changes. To find the modified equation for $\theta_2(t)$, suppose we pick at time $t + h$ two alleles from two different individuals in a given colony. With probability $\lambda h + o(h)$, the colony suffers extinction and is recolonized by k founders, all this within $(t, t + h]$. Let

$$a_k = P \text{ (both alleles originate from the$$

same colony-founder),

noting $a_1 = 1, a_N = 0$. If the two alleles represent two different founders of their colony we make a further distinction by introducing

$$\psi = P \text{ (two different colony founders originate from the same colony),}$$

so that in the propagule model we have $\psi = 1$ and in the migrant pool model $\psi = 1/n$.

Thus, temporarily switching off mutation, reproduction, and migration and considering only extinction/recolonization,

$$\begin{aligned} \theta_2(t+h) &= \lambda h a_k \left(\frac{1}{2} + \frac{1}{2} \theta_1(t) \right) \\ &+ \lambda h (1 - a_k) \psi \left(\frac{1}{N} \left(\frac{1}{2} + \frac{1}{2} \theta_1(t) \right) + \left(1 - \frac{1}{N} \right) \theta_2(t) \right) \\ &+ \lambda h (1 - a_k) (1 - \psi) \theta_3(t) + (1 - \lambda h) \theta_2(t) + o(h), \end{aligned}$$

and therefore

$$\begin{aligned} \theta'_2(t) &= -\lambda(1 - a_k)(1 - \psi)\theta_2(t) + \lambda(1 - a_k)(1 - \psi)\theta_3(t) \\ &+ \lambda \left[a_k + (1 - a_k) \psi \frac{1}{N} \right] \left(\frac{1}{2} + \frac{1}{2} \theta_1(t) - \theta_2(t) \right). \end{aligned}$$

In combination with (12) we find the new equation

$$\begin{aligned} \theta'_2(t) &= - \left[2u + \left(2m_k \left(1 - \frac{1}{n} \right) + \lambda(1 - a_k)(1 - \psi) \right) \right] \theta_2(t) \\ &+ \left(2m_k \left(1 - \frac{1}{n} \right) + \lambda(1 - a_k)(1 - \psi) \right) \theta_3(t) \\ &+ \left(\lambda \left[a_k + (1 - a_k) \psi \frac{1}{N} \right] + \frac{1}{N} \right) \left(\frac{1}{2} + \frac{1}{2} \theta_1(t) - \theta_2(t) \right). \end{aligned} \tag{24}$$

The corresponding steady-state solutions (22) change into

$$\begin{aligned} \theta_1 &= \frac{1 + 2u\alpha_k(s(N-1) + 1)}{(1 + 4u)(1 + 4u\alpha_k N) + 4u - 2u\alpha_k(s(N-1) + 1)}, \\ \theta_2 &= \frac{1 + 4u}{(1 + 4u)(1 + 4u\alpha_k N) + 4u - 2u\alpha_k(s(N-1) + 1)}, \tag{25} \\ \theta_3 &= \frac{(\lambda + m_k)B + 2u + 2m_k(1 - 1/n) + \lambda(1 - a_k)(1 - \psi)}{(\lambda + m_k + nu)B + 2m_k(1 - 1/n) + \lambda(1 - a_k)(1 - \psi)} \theta_2, \end{aligned}$$

where

$$B = \frac{1}{\lambda} + a_k N + (1 - a_k) \psi$$

and

$$\alpha_k = \frac{(\lambda + nm_k + nu)/\lambda + n(1 - a_k)(1 - \psi)}{(\lambda + m_k + nu)B + 2m_k(1 - 1/n) + \lambda(1 - a_k)(1 - \psi)}.$$

It remains to calculate the probabilities α_k , a task that can be rephrased as follows. Suppose we start with k individuals marked $1, \dots, k$, and select from them by random sampling with replacement $2(N - k)$ allele

copies, forming $N - k$ new individuals. This scheme yields a colony of $2N$ genes classified in k categories that are represented by at least one gene each. Now select randomly with equal probabilities two individuals and in each of them a gene. We want to find the probability a_k that both genes belong to the same category (originate from the same founder). However, as long as the two chosen individuals do not both belong to the original group of size k , the probability is $1/k$ that they are marked with the same number. Hence,

$$a_k + \left(1 - \frac{\binom{k}{2}}{\binom{N}{2}} \right) \frac{1}{k} = \frac{1}{k} - \frac{k - 1}{N(N - 1)}.$$

A simpler scheme is to assume that a new colony consisting of N individuals is instantly sampled from the k founding individuals. Then

$$a_k = 1/k. \tag{26}$$

The exact expressions for Wright's indices are somewhat unwieldy in this case, so we defer the further discussion to the final section where we apply some approximations and state a result.

IBD PROBABILITY IN VARYING POPULATION SIZE

An obvious drawback with the previous model is the restriction to fixed colony sizes N throughout the population. In population genetics modeling it has been natural to apply simple birth-death-immigration (BDI) models to widen the scope toward varying population size. In particular, there is a rich probabilistic structure in haploid populations based on the demographic decomposition into the number of families of allele types with a given number of representatives in the population (Tavaré 1989). Rannala and Hartigan (1995) explore some of these ideas for computing IBD probabilities in a metapopulation run by birth, death, and immigration. Rannala (1996) considers the corresponding sampling distribution of alleles under the same model and derives an expression for F_{ST} , which then coincides with θ_2 in our notation.

We take a different approach and extend the analysis of Rannala and Hartigan to include, under equilibrium conditions, selfing and migration between colonies. Our study may also be seen as a further alternative to the extinction/recolonization scheme in the fixed-size model.

A birth-death-immigration model: We start afresh from the situation that led to Equation 15, keeping seed migration but replacing the size-preserving Moran mechanism by death and immigration, as explained in generalities and scope.

Within this setting, δ introduces the potential for local population extinctions and β that for local recoloniza-

tions. Out of several possible schemes regarding the probability of IBD between a newly immigrated gene and an existent one, we assume that all immigrants are taken from a pool of individuals such that

$$P(\text{a pair with one immigrant gene and one existing gene is IBD}) = f \geq 0.$$

We assume that only seed migration is in effect. Hence individuals are exchanged between colonies without changing the population size. Let

$$N_t^j = \text{the number of individuals at time } t \text{ in colony nr } j, \quad 1 \leq j \leq n,$$

and

$$N_t^{\text{pop}} = \sum_{j=1}^n N_t^j = \text{the total population size process.}$$

It is clear that N_t^{pop} is a continuous-time Markov birth-and-death process and that the components in the vector of colony sizes (N_t^1, \dots, N_t^n) are dependent random processes.

The conditional jump probabilities at a time t , such that $N_t^{\text{pop}} = r$, are given by

$$P(N_{t+h}^{\text{pop}} - N_t^{\text{pop}} = 1 | N_t^{\text{pop}} = r) = (n\beta + r/2)h + o(h),$$

$$P(N_{t+h}^{\text{pop}} - N_t^{\text{pop}} = -1 | N_t^{\text{pop}} = r) = \delta rh + o(h), \quad h \rightarrow 0.$$

This is a standard birth-and-death process, known as a BDI process with birth intensity $\frac{1}{2}$, death intensity δ , and immigration intensity $n\beta$. For each $\delta > \frac{1}{2}$, it has a negative binomial distribution as its stationary steady-state distribution. We let N^{pop} denote a random variable having this distribution. From properties of the BDI process, for large t

$$P(N_t^{\text{pop}} = r) \approx P(N^{\text{pop}} = r) = \binom{2n\beta + r - 1}{r} \left(1 - \frac{1}{2\delta}\right)^{2n\beta} \left(\frac{1}{2\delta}\right)^r. \tag{27}$$

Moreover, expectation and variance are given by

$$E(N^{\text{pop}}) = 2n\beta / (2\delta - 1), \quad V(N^{\text{pop}}) = 4n\beta\delta / (2\delta - 1)^2.$$

Next we study the sizes of the single colonies. We refer to their stationary distributions by means of random variables (N^1, \dots, N^n) . By symmetry we immediately have

$$E(N^j) = 2\beta / (2\delta - 1), \quad 1 \leq j \leq n.$$

Furthermore, in the special case $m_s = 0$ the n components N^j are obviously independent and thus each N^j possesses a negative binomial distribution with parameters 2β and 2δ . In fact, their sum is then also negative binomial with parameters $2n\beta$ and 2δ , in agreement with (27). However, we do want to consider the case $m_s > 0$ where the colony sizes are no more independent.

To this end, we note that in the present framework migration corresponds to the death of an individual in one colony followed instantly by immigration into a

randomly chosen colony. Hence, focusing on a specific colony size, N_t^j say, the total external immigration intensity at time t is given by

$$\beta + m_s \frac{1}{n} \sum_{i \neq j} N_t^i \approx \beta + m_s \left(1 - \frac{1}{n}\right) E(N^j),$$

because a migrating gene from any other colony will end up in colony nr j with probability $1/n$. The approximation should be valid at least for large values of n in view of the law of large numbers. Hence we introduce for each component N_t^j a local immigration intensity.

$$\tilde{\beta} = \beta + m_s \left(1 - \frac{1}{n}\right) E(N^j) = \beta + \frac{2\beta m_s (1 - 1/n)}{2\delta - 1}, \tag{28}$$

and, similarly, a local death intensity by

$$\tilde{\delta} := \delta + m_s (1 - 1/n) > 1/2.$$

Indeed, a migrating gene from colony nr j is going to enter a different colony with probability $1 - 1/n$.

As an approximation for the colony size we have thus found a slightly different BDI process. Its conditional jump probabilities given that $N_t^j = r$ are

$$P(N_{t+h}^j - N_t^j = 1 | N_t^j = r) = (\tilde{\beta} + r/2)h + o(h),$$

$$P(N_{t+h}^j - N_t^j = -1 | N_t^j = r) = \tilde{\delta} rh + o(h), \quad h \rightarrow 0.$$

Consistent with our previous results the corresponding average colony size at stationarity is again

$$E(N^j) = \frac{2\tilde{\beta}}{2\tilde{\delta} - 1} = \frac{2\beta}{2\delta - 1}.$$

It is interesting to compare the variance for the components N^j with the variance of their sum N^{pop} in the dependent case $m > 0$. We obtain

$$V(N^j) = 4n\tilde{\beta}\tilde{\delta} / (2\tilde{\delta} - 1)^2 = \frac{2\beta}{2\delta - 1} \frac{2\delta + 2m_s(1 - 1/n)}{2\delta - 1 + 2m_s(1 - 1/n)}.$$

It is a simple observation that $V(N^j)$ is maximal for $m_s = 0$ [where of course $V(N^{\text{pop}}) = nV(N^1)$] and decreases toward its minimum value, coinciding with the mean $E(N^j)$, as the migration rate m is allowed to increase. In this sense migration is seen to stabilize the system and counteract fluctuations in the colony sizes.

Probabilities of IBD in the varying-population-size model: Because in the present model with positive probability some colonies or even the whole population may be extinct from time to time, the definition of the IBD probabilities θ_k is not as straightforward as in the previous constant-size model. To define θ_2 there must be at least one colony with two individuals, and to define θ_3 at least two nonempty colonies. It is reasonable to argue that otherwise corresponding weights should be assigned to let $\theta_3 = 0$ and $\theta_2 = 0$, respectively. However, we have, for example,

$$P(N^{\text{pop}} = 0) = \left(1 - \frac{1}{2\delta}\right)^{2n\beta} \approx e^{-n\beta/\delta}, \tag{29}$$

which suggests that we simply ignore this effect. Hence to derive a balanced differential equation for the IBD probability within a colony we assume that at some time t a colony j has size $N_j = r$, where $r \geq 3$, and we let $\theta_2(t)$ refer to the probability that two genes from different individuals in this colony are identical by descent.

Now we go over the various events that may occur in $(t, t + h]$. First of all, with intensity $\tilde{\delta}r$ there was a death leading to $r - 1 \geq 2$ individuals in the selected colony at time $t + h$. However, the removal of $r - 1$ such potential pairings does not change the probability of IBD for the remaining pairs (*cf.* Appendix in Rannala and Hartigan 1995).

Next, with intensity $\tilde{\beta}$ an individual immigrated between t and $t + h$, increasing the colony size from r to $r + 1$. With probability

$$\frac{\binom{r}{1}}{\binom{r+1}{2}} = \frac{2}{r+1}$$

a gene from this newly immigrated individual is selected to the pairing needed to determine θ_2 . If the colony immigrant is an external immigrant to the population this results in the probability f for the pair to be identical by descent. If the colony immigrant migrated from another colony the gene pair is related according to θ_3 . Otherwise, if the immigrant was not chosen for the pair, the probability for IBD remains the same as before immigration.

Finally, with intensity $r/2$ a birth occurs in $(t, t + h]$ such that again the size changes from r to $r + 1$. As a pair of genes A and B are selected at time $t + h$, the probability is $2/(r + 1)$ that one of them belongs to the newly reproduced individual. Let us suppose that A does. Then we have two possibilities. With probability $1/r$ gene A was produced by random sampling from the individual that gene B belongs to, in which case the probability of IBD is $1/2 + \theta_1(t)/2$. With probability $1 - 1/r$ the relationship between A and B is given simply by $\theta_2(t)$.

Summing up,

$$\begin{aligned} \theta_2(t+h) &= h \frac{2}{r+1} \left(\beta f + \frac{2m_s\beta}{2\delta-1} \left(1 - \frac{1}{n}\right) \theta_3(t) \right) \\ &+ \tilde{\beta} h \left(1 - \frac{2}{r+1}\right) \theta_2(t) \\ &+ \frac{1}{2} rh \left\{ \frac{2}{r+1} \left[\frac{1}{r} \left(\frac{1}{2} + \frac{1}{2} \theta_1(t) \right) + \left(1 - \frac{1}{r}\right) \theta_2(t) \right] \right. \\ &\quad \left. + \left(1 - \frac{2}{r+1}\right) \theta_2(t) \right\} \\ &+ (1 - 2uh - \tilde{\beta}h - rh/2) \theta_2(t) + o(h). \end{aligned}$$

Hence,

$$\begin{aligned} \theta_2'(t) &= \frac{2\beta}{r+1} f + \frac{4\beta m_s(1-1/n)}{(r+1)(2\delta-1)} \theta_3(t) \\ &+ \frac{1}{r+1} \left(\frac{1}{2} + \frac{1}{2} \theta_1(t) - \theta_2(t) \right) - 2 \left(u + \frac{\tilde{\beta}}{r+1} \right) \theta_2(t). \end{aligned}$$

By a similar argument

$$\begin{aligned} \theta_1'(t) &= \frac{1}{2} \frac{r}{r+1} \left[\left(s + \frac{1-s}{r} \right) \left(\frac{1}{2} + \frac{1}{2} \theta_1(t) - \theta_2(t) \right) + \theta_2(t) \right] \\ &- \frac{1}{2} \left(\frac{r}{r+1} + 4u \right) \theta_1(t). \end{aligned}$$

Turning to θ_3 suppose two colonies i and j have been chosen. We must determine the effect of immigration into any one of them during $(t, t + h]$. Suppose colony i of size r is subject to external immigration (or colony j of size q). Then with probability $1/(r + 1)$ the immigrant belongs to the select pair resulting in IBD probability f . Similarly, if the immigrant is a migrant from j , then again with probability $1/(r + 1)$ the IBD probability of the select pair changes, this time into θ_2 . The intensities are for the first event β and for the second approximately $mE(N)/n$. Hence

$$\begin{aligned} \theta_3'(t) &= \left(\frac{1}{r+1} + \frac{1}{q+1} \right) \left(f\beta + \frac{1}{n} mE(N) \theta_2(t) \right) \\ &- 2u\theta_3(t) - \left(\frac{1}{r+1} + \frac{1}{q+1} \right) \left(\frac{1}{n} mE(N) + \beta \right) \theta_3(t). \end{aligned}$$

At this point we observe that the steady-state probabilities corresponding to the balanced equations we have derived will depend on r and q unless $u = 0$. In the present model the external immigration parameter β has a similar effect and purpose as mutation, namely, to guarantee fresh import of genetic material and enhance genetic diversity. The difference is that mutation acts only on the individual level while immigration occurs in colonies regardless of their sizes. We therefore make the approximation in the sequel to ignore mutation and simply restrict the case to $u = 0$. Moreover, because $\theta_1'(t) = 0$ implies

$$\left(s + \frac{1-s}{r} \right) \left(\frac{1}{2} + \frac{1}{2} \theta_1 - \theta_2 \right) + \theta_2 - \theta_1 = 0,$$

and thus

$$\begin{aligned} \theta_1 &= \frac{1 - (1-s)(1-1/r) + 2(1-s)(1-1/r)\theta_2}{1 + (1-s)(1-1/r)} \\ &\approx \frac{s + 2(1-s)\theta_2}{2-s} \end{aligned}$$

for large r , which is again the consistency relation (34), we are led to study the balanced equations

$$\tilde{\theta}_1 = \frac{s + 2(1-s)\tilde{\theta}_2}{2-s},$$

$$2\beta f + 2m_s \left(1 - \frac{1}{n}\right) E(N) \tilde{\theta}_3 + \frac{1}{2} + \frac{1}{2} \tilde{\theta}_1 - \tilde{\theta}_2 = 2\tilde{\beta} \tilde{\theta}_2,$$

$$2\beta f + \frac{2m_s E(N)}{n} \tilde{\theta}_2 = 2(m_s E(N)/n + \beta) \tilde{\theta}_3.$$

The solution is given by

$$\begin{aligned} \tilde{\theta}_1 &= \frac{1 + 2\alpha\beta + 4(2 - s)\alpha\beta f}{1 + 2(2 - s)\alpha\beta}, \\ \tilde{\theta}_2 &= \frac{1 + 2(2 - s)\alpha\beta f}{1 + 2(2 - s)\alpha\beta}, \\ \tilde{\theta}_3 &= \frac{(2\delta - 1)f + 2m_2 \tilde{\theta}_2/n}{2m_s/n + 2\delta - 1}, \end{aligned} \tag{30}$$

where now

$$\alpha = \frac{\beta + m_s E(N)}{\beta + m_s E(N)/n} = \frac{2m_s + 2\delta - 1}{2m_s/n + 2\delta - 1}.$$

In conclusion we have found, under the simplifying assumptions of mutation being negligible and the typical size of a colony being large, that the conditional IBD probabilities at equilibrium for given colony sizes are expressed by the relations (30). Again relying on, *e.g.*, (29), we let expression (30) represent the unconditional IBD probabilities for the variable-size model. Thus Wright’s genetic indices for the BDI model turn out to be

$$F_{IS} = \frac{s}{2 - s}$$

and

$$F_{ST} = 1 / \left(1 + 2\beta \left(\frac{2m_s + 2\delta - 1}{2\delta - 1}\right) (2 - s)\right). \tag{31}$$

Interestingly, the last quantity depends on neither f nor n .

DISCUSSION

Relation of fixed-size model to previous work: We compare the expressions derived so far with those previously obtained by Maruyama and Tachida (1992) and Whitlock and McCauley (1990). To recover the results of Maruyama and Tachida (1992) from (20) and (23), or the combination of them, we observe that if u , λ , and m are small compared to N , then

$$\begin{aligned} \theta_1 \approx \tilde{\theta}_1 &= \frac{1 + 2u\tilde{\alpha}sN}{1 + 2u\tilde{\alpha}N(2 - s)} \\ \theta_2 \approx \tilde{\theta}_2 &= \frac{1}{1 + 2u\tilde{\alpha}N(2 - s)} \\ \theta_3 \approx \tilde{\theta}_3 &= \frac{\lambda + m}{\lambda + m + nu} \tilde{\theta}_2 \end{aligned} \tag{32}$$

with

$$\tilde{\alpha} = \frac{\lambda + nm + nu}{\lambda + m + nu},$$

hence

$$F_{IS} \approx \frac{s}{2 - s}, \quad F_{ST} \approx \frac{1}{1 + 2(\lambda/n + u + m)N(2 - s)}. \tag{33}$$

These are the same approximate results as those found by Maruyama and Tachida (1992). However, it should be kept in mind that in the quoted article the parameters λ , u , and m_s are probabilities, whereas in our study they denote Poisson rates.

We take these findings as a basis for claiming that the continuous time approach has some advantages over the discrete analog for investigating the detailed influence on Wright’s indices of the various random mechanisms that change population genetic diversity.

We continue with some additional remarks on the approximate analysis discussed above. It is interesting that the equilibrium relation (14) is still valid for the approximative solutions, that is,

$$\tilde{\theta}_2 = \frac{1 + \tilde{\theta}_1}{2(1 + 2u\tilde{\alpha}N)}.$$

Further insight into the background of this approximation is gained from the following simple argument. Start from the reasonable assumption that θ_1 and θ_2 can only differ if individuals are subject to selfing. Suppose that a fraction s of the population is selfing, so that if an individual is picked randomly in the population it reproduces by selfing with probability s . Then we are led to the consistency relation

$$\theta_1 = s \left(\frac{1}{2} + \frac{1}{2}\theta_1\right) + (1 - s)\theta_2.$$

This gives an alternative simple relationship between θ_1 and θ_2 , and it is easy to check that $\tilde{\theta}_1$ and $\tilde{\theta}_2$ are indeed related by

$$\tilde{\theta}_1 = \frac{s + 2(1 - s)\tilde{\theta}_2}{2 - s}. \tag{34}$$

At this point we emphasize one of the refinements in our model relative to the approximation discussed in this subsection, that θ_1 and θ_2 may actually differ from reasons other than selfing. In fact, when we form a pair of genes the total mutation rate is $2u$ irrespective of whether they belong to the same individual or not, but the genetic drift differs. If we select the pair, within an individual the infinitesimal rate is only $1/2$ for reproduction to possibly affect the IBD probability, whereas if two individuals are involved they both contribute, yielding a total rate for θ_2 twice as high as that of θ_1 .

For the case of extinction/recolonization with k

founding individuals, we apply the approximation (34) and obtain from (25)

$$F_{ST} = \left(1 - \frac{2\lambda}{n} + \lambda A_k\right) / \left(1 - \frac{2\lambda}{n} + \lambda A_k + 2\left[u + m_s + \frac{\lambda}{n} + \lambda(1 - a_k)(1 - \psi)\right](2 - s)N\right),$$

$$A_k = a_k N + (1 - a_k)\psi. \tag{35}$$

One can check that F_{ST} is increasing in λ if and only if

$$a_k \geq \frac{1 - \psi + 1/n - (m_s + u)\psi}{1 - \psi - (m_s + u)\psi + (m_s + u)N}, \tag{36}$$

which is trivially true if $m_s + u \geq (1 - \psi + 1/n)/\psi$, i.e.,

$$(1 + m_s + u)\psi \geq 1 + \frac{1}{n}.$$

In particular, for the propagule case, $\psi = 1$, and with $k < N$ and large n , extinction always favors genetic fixation in the sense that F_{ST} increases with λ . This behavior has been observed previously (McCauley 1993) and should be compared to the pure propagule model with $k = N$ for which F_{ST} is a decreasing function of λ .

The remaining case

$$(1 + m_s + u)\psi < 1 + \frac{1}{n}$$

includes migrant pool and mixed versions of propagule and migrant pool recolonization. Specializing to the choice $a_k = 1/k$ in (26), condition (36) takes the form

$$k \leq 1 + \frac{(m_s + u)N}{1 - \psi(1 + m_s + u)} \approx 1 + \frac{m_s N}{1 - \psi(1 + m_s)},$$

where the final approximation is for large N and negligible mutation. This is similar to the condition obtained by Whitlock and McCauley (1990). For alleles moving as gametes (their Equation 10a), they obtained

$$k \leq \frac{1}{2} + \frac{2mN}{1 - \phi},$$

where m denotes the proportion of migrating gametes in each generation, so that $m_s = 2m$, and ϕ is the probability of common origin of gametes. According to Whitlock and McCauley (1990, p. 1720), “when alleles move only as diploid individuals, with equal representation of parental populations,”

$$\phi = [(2k - 2)\psi + 1]/(2k - 1)$$

(Whitlock and McCauley 1990, Equation 7). The last two relations together imply that

$$k \leq 1 + \frac{2mN}{1 - \psi},$$

which is indeed similar to the condition at which we arrived.

Finally, it is instructive to look at the expressions for F_{ST} as $N \rightarrow \infty$ and k is fixed. In the migrant pool case $\phi = 1/n$,

$$F_{ST} \rightarrow \lambda a_k / \left(\lambda a_k + 2\left[u + m_s + \frac{\lambda}{n} + \lambda(1 - a_k)\left(1 - \frac{1}{n}\right)\right](2 - s)\right)$$

and in the propagule case $\phi = 1$

$$F_{ST} \rightarrow \lambda a_k / \left(\lambda a_k + 2\left[u + m_s + \frac{\lambda}{n}\right](2 - s)\right).$$

Contrary to the situation without extinction/recolonization or to the propagule model with $k = N$, for which $F_{ST} \rightarrow 0$ as $N \rightarrow \infty$, the limit value of F_{ST} is now strictly positive. This is, of course, due to the fact that the founder effect that follows recolonization restores a substantial amount of genetic drift.

Discussion of the variable size model: Writing N^{col} for the size of a typical colony we have equivalently to (31)

$$F_{ST} = \frac{1}{1 + 2(m_s + \delta - 1/2)E(N^{col})(2 - s)}, \tag{37}$$

which should be compared to the fixed-size model results, e.g., (33). This expression stresses the fact that migration and population size cannot be easily separated in the variable population model because the “migration term” now comprises the quantity $\delta - 1/2$, which together with the immigration intensity determines the average population size.

In the special case $m_s = 0$, for which $\tilde{\theta}_3 = f$ (37) simplifies further into

$$F_{ST} = \frac{1}{1 + 2\beta(2 - s)};$$

hence notably the death intensity parameter δ disappears entirely from Wright’s index. This is basically the result derived by Rannala and Hartigan (1995) in the case $s = 0$, taking into account that they assume $\theta_3 = 0$. Taking in their notation birth intensity $\lambda = 1/2$ and immigration intensity $\phi = 2\beta$, they found

$$F_{ST} \approx \tilde{\theta}_2 = \frac{\lambda + \phi f}{\lambda + \phi} = \frac{1 + 4\beta f}{1 + 4\beta} \tag{38}$$

in agreement with our result. The approach in Rannala and Hartigan (1995) is based on properties of the conditional allele distribution given a colony size N and hence does not preclude the transient situation $\delta < \lambda$ for which there is no steady-state distribution of population size or colony size. Although their conclusion is justified for the simpler BDI model, with $m_s = 0$, we have found that the probability of IBD also depends on δ and m_s . Moreover, for small values of $\delta - 1/2$ the resulting equation (31) can be highly sensitive to small changes in the parameter m_s (see also Gaggiotti 1996). In our extension of the BDI model, $m_s > 0$, the assumption $\delta > \lambda = 1/2$ is crucial for the mathematical derivations.

It is natural therefore to ask whether it is possible to include the case $m_s > 0$ in the derivation that leads to (38). To indicate heuristically what may happen in this situation, suppose that the dynamics of a large number of colonies are given by the BDI model with parameters β and δ , and suppose that migrants are exchanged at rate m_s between colonies. Focus on a single colony. At a fixed time t since the populations were initiated ($N_0 = 0$), we may argue as in (28) that, due to migration between colonies, β should be replaced by the effective immigration parameter

$$\beta + m_s E(N_t) = \beta + m_s \frac{2\beta}{1 - 2\delta} (e^{(1-2\delta)t/2} - 1).$$

If we apply this approximation and follow the approach of Rannala and Hartigan, then Equation 38 (with $f = 0$) yields

$$F_{ST} \approx 1 / \left(1 + 4\beta \left[1 + \frac{2m}{1 - 2\delta} (e^{(1-2\delta)t/2} - 1) \right] \right).$$

In the subcritical case $\delta > 1/2$ we recover Equation 31 ($s = 0$) as $t \rightarrow \infty$. For fixed t , however, the expression for F_{ST} is now continuous in $\delta > 0$ and seems to exhibit less dramatic change when we compare $m_s = 0$ and $m_s > 0$.

Our model as well as that of Rannala and Hartigan assume (i) the presence of a continuous flow of immigrants from a mainland in which allele frequencies remain constant and (ii) the absence of spatial structure. Clearly, in many biological situations these are strong assumptions. Therefore a natural next step is to study probabilities of IBD in a metapopulation where colony sizes are regulated by a birth-death process, without external immigration, and in which the migration rate among colonies depends on their spatial location. The demography of similar systems has been studied (*e.g.*, Pollett 1998) but, to our knowledge, their genetic properties remain unexplored.

We thank Dr. Steve Krone and two anonymous referees for useful comments on the manuscript.

LITERATURE CITED

- Cockerham, C. C., and B. S. Weir, 1987 Correlations, descent measures: drift with migration and mutation. *Proc. Natl. Acad. Sci. USA* **84**: 8512-8514.
- Gaggiotti, O. E., 1996 Population genetic models of source-sink metapopulations. *Theor. Popul. Biol.* **50**: 178-208.
- Kimura, M., and J. F. Crow, 1964 The number of alleles that can be maintained in a finite population. *Genetics* **49**: 725-738.
- Malécot, G., 1948 *Les Mathématiques de l'Hérédité*. Masson et Cie, Paris.
- Maruyama, K., and H. Tachida, 1992 Genetic variability and geographical structure in partially selfing populations. *Jpn. J. Genet.* **67**: 39-51.
- McCauley, D. E., 1993 Genetic consequences of extinction and recolonization in fragmented habitats, pp. 217-233 in *Biotic Interactions and Global Change*, edited by P. M. Kareiva, J. G. Kingsolver and R. B. Huey. Sinauer Associates, Sunderland, MA.
- Moran, P. A. P., 1958 Random processes in genetics. *Proc. Camb. Philos. Soc.* **54**: 60-71.
- Pollett, P. K., 1998 Modelling quasi-stationary behaviour in metapopulations. Preprint.
- Rannala, B., 1996 The sampling theory of neutral alleles in an island population of fluctuating size. *Theor. Popul. Biol.* **50**: 91-104.
- Rannala, B., and J. A. Hartigan, 1995 Identity by descent in island-mainland populations. *Genetics* **139**: 429-437.
- Rousset, F., 1996 Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics* **142**: 1357-1362.
- Slatkin, M., 1977 Gene flow and genetic drift in a species subject to frequent local extinctions. *Theor. Popul. Biol.* **12**: 253-262.
- Tavare, S., 1989 The genealogy of the birth, death, and immigration process, pp. 41-56 in *Mathematical Evolutionary Theory*, edited by M. W. Feldman. Princeton University Press, Princeton, NJ.
- Whitlock, M. C., and D. E. McCauley, 1990 Some population genetic consequences of colony formation and extinction: genetic correlation within founding groups. *Evolution* **44**: 1717-1724.
- Wright, S., 1969 *Evolution and the Genetics of Populations II, The Theory of Gene Frequencies*. University of Chicago Press, Chicago.

Communicating editor: W. Stephan