

Gene Flow and Genetic Drift in a Species Subject to Frequent Local Extinctions

MONTGOMERY SLATKIN¹

*Department of Biophysics and Theoretical Biology,
The University of Chicago, Chicago, Illinois 60637*

Received August 13, 1976

Two models of the effect of extinction and recolonization on the genetic differentiation of local populations are analyzed. One model is Wright's "island model" in which there is gene flow from a source of fixed gene frequency. The other is an island model with a continuous production of new alleles and gene flow among all the populations. Individual and group selection are not considered. It is shown that the extent of population differentiation and the direction of the effect of the colonization and extinction process depend on the manner in which the propagules that establish new colonies are formed. Two extreme cases are considered. In the "propagule pool" model all the individuals in a single propagule are derived from one population while in the "migrant pool" model, the individuals in a propagule are derived from a random sample of the entire collection of populations.

INTRODUCTION

Many species exist in discrete local populations that become extinct and are recolonized by propagules derived from other local populations. However, most models of subdivided populations assume that each local population persists indefinitely (e.g., Maruyama, 1970; Latter, 1973). While these models are appropriate for species in which the average time until the extinction of local populations is large compared with the other time scales in the system, there are many species for which that is not likely to be the case. Studies of island populations by Simberloff and Wilson (1969), Crowell (1973), and others have found relatively high extinction rates for many species including a variety of insects. The effect of gene flow between local populations of a species with significant extinction and recolonization rates has an important bearing on the possibility of group selection in that species because the effectiveness of group selection depends on the presence of genetic differences between local populations (Levins, 1970; Levin and Kilmer, 1974; Cohen and Eshel, 1976; Slatkin and Wade, 1977). I will introduce and analyze here two simple analytic models of the

¹ Present Address: Department of Zoology, University of Washington, Seattle, Washington 98195.

effect of extinction and recolonization of local populations on the genetic structure of those populations.

The two models analyzed are both variations on Wright's (1931) "island" model. There is no geographic structure and all local populations are assumed to be equally likely to receive migrants. In the first model (Model I), all the migrants come from an external source and new mutations are ignored. In the second model (Model II), all migrants come from within the collection of populations or "metapopulation," and there is a continuous production of new unique mutations (Kimura and Crow's (1964) "infinite alleles" model). We will not consider either group or individual selection.

In each of these models we will see that the results depend on the details of the assumptions about how the propagules that found new populations are formed. We will consider two possibilities, one in which the individuals in a propagule are all from the same existing population (the "propagule pool" model), and the other in which the individuals in a propagule are chosen at random from the entire metapopulation (the "migrant pool" model). It is clear that there is a continuous range of other possibilities but that these two represent the extremes.

MODEL I

Consider a diploid species with nonoverlapping generations and a single locus with two alleles, A and a , and assume that A is maintained in a source of migrants at a constant frequency x_1 . Assume that there is a large number of local populations each of effective size N and each of which has a fraction m of its individuals replaced by migrants from the source. That is, in each generation before random mating takes place a fraction m of the gametes are replaced by gametes from the source and the frequency of A in those gametes is x_1 . There is assumed to be no variability associated with the migration process, either through the number of migrants or the allele frequency of the migrant group. These effects can be easily incorporated into the present model without altering the conclusions. Finally, assume that a fraction e_0 of the local populations goes extinct and the vacant sites are all recolonized in a manner to be specified. For simplicity there is also no variation in the number of local populations, and we will be assuming that there is effectively an infinite number of populations because we will not allow for any deviations from the expected distribution.

Let x be the frequency of A in a single population and $p(x, t)$ the frequency distribution of local populations of frequency x in generation t . Because there is no individual or group selection, it is sufficient to compute only the mean and variance of $p(x, t)$. Furthermore, we will consider only the equilibrium mean and variance, although the solution to the transient problem can be found in the same manner. In the absence of selection, the mean of $p(x, t)$ at equilibrium must be x_1 , the frequency of A in the source of migrants. To compute the change in the

variance, σ_p^2 , due to drift and gene flow from the source, we can use the method of Crow and Kimura (1970, Sect. 9.2). For a population of frequency x before migration, the frequency after migration is

$$x' = (1 - m)x + mx_I \quad (1)$$

with the above assumption. After random mating, the variance in x' is $x'(1 - x')/2N$. Since the expected values of both x and x' are x_I , the variance in x' between populations is simply the variance before mating plus the expectation of $x'(1 - x')/2N$. Thus, if we let $\tilde{\sigma}_p$ denote the variance in x after migration and random mating, then

$$\tilde{\sigma}_p = E[(x' - x_I)^2] + (1/2N) E[x'(1 - x')] \quad (2)$$

which by direct substitution is found to be

$$\tilde{\sigma}_p^2 = (1 - m)^2(1 - 1/2N)\sigma_p^2 + x_I(1 - x_I)/2N, \quad (3)$$

which is the result in Crow and Kimura (1970, p. 440). After extinction and recolonization the variance must be σ_p^2 , the equilibrium variance. Thus,

$$\sigma_p^2 = (1 - e_0)\sigma_p^2 + e_0\sigma_n^2, \quad (4)$$

where σ_n^2 is the variance among the newly founded populations and must be calculated for each model of the colonization process. Equation (4) is correct as long as there is no group selection that would cause the mean of the newly founded populations to be different from x_I .

The variance, σ_n^2 , depends on the manner in which the propagules that found the new populations are formed. In the propagule pool model, we assume that each established population produces a large number of propagules of size k from excess individuals each generation. With that assumption, the probability that a propagule derived from a population with frequency x has i A alleles is a binomial, $B(i; 2k, x)$. In the propagule pool model there is no mixing of individuals from different propagules so each new population is derived from individuals arising in a single population. For each new site in which a population has gone extinct, a propagule is chosen at random from the propagule pool. Thus after the colonization stage, a fraction $1 - e_0$ of the populations are of a size N and a fraction e_0 are of size k . For simplicity we assume that the newly founded populations grow to size N in the first generation after they are founded, and this assumption may not be unreasonable for species that experience frequent extinctions and recolonizations. Also we assume that the propagules are formed after mating and gene flow.

The variance in the newly founded populations is found by averaging $B(i, 2k, x)$ over all x and then taking the variance to find

$$\sigma_n^2 = (1 - (1/2k)) \bar{\sigma}_p^2 + x_i(1 - x_i)/2k, \quad (5)$$

where $\bar{\sigma}_p^2$ must be used because of the stage at which the propagules are assumed to be formed. Therefore,

$$\sigma_p^2 \Big|_{\text{prop}} = \frac{x_i(1 - x_i) \left[\left(1 - \frac{e_0}{2k}\right) \frac{1}{2N} + \frac{e_0}{2k} \right]}{1 - \left(1 - \frac{e_0}{2k}\right) (1 - m)^2 \left(1 - \frac{1}{2N}\right)}, \quad (6)$$

which we find by solving (4) using (3) and (5).

In the migrant pool model, which is the model that has been used in most previous analyses of group selection (Levins, 1970; Levin and Kilmer, 1974; Cohen and Eshel, 1976), each population is assumed to contribute to a large pool from which individuals are chosen to form propagules. Thus, in contrast to the propagule pool model, newly founded populations are derived from several existing populations. The frequency of A in the migrant pool is x_i so the variance in the frequency of A in a large number of samples of $2k$ gametes from the migrant pool is $x_i(1 - x_i)/2k$ which is then σ_n^2 . We assume gametes rather than individuals are sampled to simplify the calculations and to permit the maximum mixing of individuals forming the propagules. We note that σ_n^2 for the migrant pool model is smaller than for the propagule pool model (Eq. (5)) because of the mixing of the individuals from different populations before forming the propagules. Substituting in (2), we find

$$\sigma_p^2 \Big|_{\text{migr.}} = \frac{x_i(1 - x_i)[(1 - e_0)/2N + e_0/2k]}{1 - (1 - e_0)(1 - m)^2 (1 - 1/2N)}. \quad (7)$$

In comparing (6) and (7), there are two points worth noting. First, (7) is always greater than (6) if $e_0 > 0$ and $k < N$ because the mixing of individuals during the formation of propagules in the migrant pool model reduces the between-population variance. In fact, (7) is less than the variance in the absence of extinctions (found by setting e_0 to zero in (6) or (7)) when

$$k > N[1 - (1 - m)^2(1 - 1/2N)]. \quad (8)$$

Therefore, the degree of mixing between the individuals derived from different populations can strongly affect the between-population variance. Second, the propagule size, k , is also important and, if $N \gg k$ in the propagule pool model, then the extinction and recolonization process far outweighs the effect of genetic drift in determining the variance.

MODEL II

The second model is the same as one analyzed by Maruyama (1970) and Latter (1973), except that the extinction and recolonization stage is added. For a diploid species, consider a single locus at which there is a mutation rate u per generation for new, selectively neutral alleles. Assume that there are n local populations each of effective size N and each of which exchanges a fraction m of its alleles with a random sample of alleles from the entire metapopulation. Following Maruyama (1970), we use the variables f_0 and f_1 as the probabilities that two alleles chosen at random from the same population and from different populations are identical.

The recursion formulas for f_0 and f_1 are

$$f_0' = (1 - u)^2 \left\{ (1 - e_0) \left[a \left(\frac{1}{2N} + \left(1 - \frac{1}{2N} \right) f_0 \right) + (1 - a) f_1 \right] + e_0 A_1 \right\} \tag{9}$$

and

$$f_1' = (1 - u)^2 \left\{ (1 - e_0) \left(1 - \frac{e_0 n}{n - 1} \right) \left[b \left(\frac{1}{2N} + \left(1 - \frac{1}{2N} \right) f_0 \right) + (1 - b) f_1 \right] + \frac{2ne_0(1 - e_0)}{n - 1} B_1 + \frac{e_0(ne_0 - 1)}{n - 1} B_2 \right\}, \tag{10}$$

where

$$a = (1 - m)^2 + \frac{m^2}{(1 - e_0)n} + \frac{2m(1 - m)}{(1 - e_0)n} \tag{11}$$

and

$$b = \frac{m(2 - m)}{(1 - e_0)n}, \tag{12}$$

where we define A_1 as the probability of identity of two alleles chosen from a newly founded population, B_1 as the probability of identity given that one of the alleles is chosen from a newly founded population and the other from an existing population, and B_2 as the probability of identity given that both alleles are from newly founded populations. The prime indicates the value in the succeeding generation. The fraction e_0 is assumed to be the same each generation. Equation (11) differs from the corresponding term in Maruyama's paper (1970) because it takes into account the possibility that one of the alleles selected at random is not from a migrant and the other is from a migrant that returned to the same population. Latter (1973) is the first to point out that this additional term is necessary because migrant gametes or individuals can return to the population from which they are derived.

We will assume that the propagules are derived only from populations that do

not go extinct, but that can be changed with no qualitative change in the results. As in Model I, the terms associated with the recolonization stage depend on how the propagules are assumed to be formed.

In order to use f_0 and f_1 as the variables in this model, we must make a slight change in the way in which new populations are assumed to be founded. We assume after the propagule of size k founds a new population, it immediately undergoes random mating to produce a population of size N . This was not necessary in Model I, but it is in this model to guarantee that all populations are the same size at the time during each generation at which f_0 and f_1 are defined. Then, f_0 and f_1 are sufficient to specify the properties of the population. As a result, there are two sampling processes during the formation of a new population, the first when k individuals are chosen to make up the propagule and the second when those k individuals mate at random to form a population of size N . Both sampling processes must be kept in mind when calculating A_1 , B_1 , and B_2 . The individuals making up each propagule are assumed to be excess individuals in each population.

In the propagule pool model, each propagule is formed from k individuals derived from a single population. The probability of identity in the propagule is

$$1/2N + (1 - 1/2N)f_0 \quad (13)$$

because there are N possible parents of the individuals in the propagule. After random mating, the probability of identity in the newly founded population is

$$A_1 = \frac{1}{2N} + \frac{1}{2k} \left(1 - \frac{1}{2N}\right) + \left(1 - \frac{1}{2k}\right) \left(1 - \frac{1}{2N}\right) f_0. \quad (14)$$

As mentioned above, there is an extra sampling stage introduced to simplify the analysis. To compute B_1 , we must account for the case in which the newly founded population is derived from the existing population and that in which it is not. Thus,

$$B_1 = \frac{1}{(1 - e_0)n} f_0 + \left(1 - \frac{1}{(1 - e_0)n}\right) f_1. \quad (15)$$

Similarly, to compute B_2 we must account for the cases in which both new populations are derived from the same and different populations. Therefore,

$$B_2 = \frac{1}{(1 - e_0)n} f_0 + \left(1 - \frac{1}{(1 - e_0)n}\right) f_1 = B_1. \quad (16)$$

We can find the equilibrium solution to (9) and (10) and the time-dependent solution for arbitrary values of the parameters by using the usual matrix methods. However, the results are complicated and not too enlightening. I will present only the equilibrium results for two ranges of parameters that are of some

biological interest. A convenient way to illustrate the effect of the colonization and extinction processes is to use n_e , the "effective number" of alleles in the metapopulation, defined by

$$1/n_c = \bar{f} = (1/n)f_0 + (1 - 1/n)f_1. \quad (17)$$

We can reasonably assume $u \ll 1$ and would most likely be using this sort of model when $n \gg 1$. Maruyama (1970) shows that, if m is small, then

$$n_e = 1 + 4nNu + nu/m, \quad (18)$$

where terms of the order of magnitude of un or smaller are ignored. Equation (18) is still a valid approximation even though Maruyama's equation must be altered to include the extra term in (11). Note in (18) that m is assumed to be small, but not necessarily the same order of magnitude as u , so that terms of order un are small relative to un/m . If we assume that both m and e_0 are small and of the same order of magnitude, then to the same degree of approximation as (18) we find

$$n_e = 1 + 4nNu \left(\frac{m}{m + e_0} \right) + \frac{un}{m + e_0}. \quad (19)$$

From (19), we can see that when gene flow and local extinctions are relatively rare events, then the extinction recolonization process can considerably reduce the between-population variance as measured by n_e . This seems to be the opposite from the result found in Model I that the between-population variance is increased by extinctions and recolonizations. This difference is due to the difference between the models. In Model I, the extinction and recolonization stage essentially introduces an additional drift term because of the sampling during the formation of propagules. This sampling does not affect the process that acts to reduce the between-population variance, the gene flow from the source. In Model II, the gene flow between the populations reduces the between-population variance and extinctions and recolonizations augment the gene flow. While the magnitude of that effect will depend on the way in which the propagules are formed, its direction is necessarily the same.

If e_0 is much larger than m , then the equilibrium solutions to (9) and (10) are simpler because, to $O(m/e_0)$, we can set $m = 0$. Again solving for the effective number of alleles, we find

$$n_e = 1 + \frac{2(1 - e_0) un}{1 - (1 - e_0)^2}, \quad (20)$$

where terms of the order of magnitude of m/e_0 , $1/n$, and u are dropped. We see

that the extinction and recolonization process can prevent the divergence of local populations even in the absence of gene flow.

For the migrant pool model, the equations are similar to those for the propagule pool model. The only difference is in A_1 , the probability of identity in a newly founded population. The probability of identity in the migrant pool from which propagules are formed is

$$\bar{f} = \frac{f_0}{(1 - e_0)n} + \left(1 - \frac{1}{(1 - e_0)n}\right) f_1, \quad (21)$$

which is the average taken over the populations that do not go extinct, and the number of parents of the migrant pool is $Nn(1 - e_0)$. Therefore, after k individuals are chosen at random to make up a propagule and they undergo random mating to form the new population, the probability of identity is

$$A_1 = \frac{1}{2nN(1 - e_0)} + \frac{1}{2k} - \frac{1}{4Nnk(1 - e_0)} + \left(1 - \frac{1}{2k}\right) \left(1 - \frac{1}{2Nn(1 - e_0)}\right) \bar{f}. \quad (22)$$

As in the results for the propagule pool model, we assume $u \ll 1$ and $n \gg 1$, and first find the effective number of alleles when e_0 and m are small and of the same order of magnitude. We find

$$n_e = 1 + 2Nnu \left(\frac{e_0 + 2m}{e_0 + m} \right) + \frac{nu}{m + e_0}, \quad (23)$$

which is larger than (19), the comparable result for the propagule pool model.

If $e_0 \gg m$, then we can find

$$n_e = 1 + \frac{2unNe_0 + nu(1 - e_0)}{E(1 - e_0 - e_0N/k)}, \quad (24)$$

where terms of order m , $1/n$, and u are ignored and where

$$E = \frac{1 - (1 - e_0)^2}{2(1 - e_0)}.$$

Equation (24) is always larger than (20), the result from the propagule pool model.

In each of these results the effective number of alleles, n_e , is used as a convenient measure of divergence between populations. A reduction in n_e indicates a

reduction in between population variance. Other possible measures are f_1 or f_1/f_0 , both of which show the same pattern as n_e in response to changes in e_0 . In fact for large n , $n_e \approx 1/f_1$.

DISCUSSION

The above results show that the effect of extinctions and recolonizations on the genetic structure of a metapopulation consisting of a large number of local populations is complex and depends on the kind of model being analyzed. There are two consequences of the extinction and colonization process. The first is an additional sampling process similar to genetic drift and results from the sampling of individuals to form the propagules. This can be regarded as the "founder effect." The second is an additional component to the gene flow between the local populations because the propagules are made up of individuals from one or more of the local populations. The direction and magnitude of the effect of the extinction and recolonization process depends on the role of genetic drift and gene flow in the model being considered.

We have considered two extreme models of the formation of propagules. In the propagule pool model, all individuals making up a propagule are from the same population and in the migrant pool model, all individuals making up a propagule are drawn from a random sample from the metapopulation. The propagule pool model results in a much greater effect of the sampling of individuals to form the propagule. In the migrant pool model, the effect of the sampling is much less because of the mixing of individuals from the different populations. In Model I, the additional genetic drift due to this sampling outweighs the effect of the additional gene flow between local populations. Therefore, the propagule pool model results in greater differentiation of the local populations. The gene flow from the source is the mechanism acting to prevent the populations from differentiating, and that is not affected by the gene flow resulting from the colonization and extinction process.

In Model II, the interaction between the mechanisms is different. The appearance of new mutations in each population increases the differentiation of local populations and the gene flow between the local populations acts to reduce local differentiation. Therefore, the additional gene flow resulting from the colonizations further reduces the extent of the local differentiation. The genetic drift within each population and the additional drift due to the formation of the propagules play a secondary role. Therefore, there is always less differentiation (as measured by the effective number of alleles) than in the absence of local extinctions. The propagule pool model is more effective than the migrant pool model in augmenting the effective gene flow between populations because the probability of identity by descent in a propagule derived from a single population is higher than in one derived from a larger number of populations.

ACKNOWLEDGMENTS

I thank M. J. Wade for helpful discussions of this subject and T. Nagylaki and the referees for important corrections and comments on an earlier draft of this paper. Financial support of this research was provided by NIH Grants No. R01 GM22523 and No. K04 GM00118.

REFERENCES

- COHEN, D., AND ESHEL, I. 1976. On the founder effect and the evolution of altruistic traits, *Theor. Pop. Biol.* **10**, 276-302.
- CROW, J. F., AND KIMURA, M. 1970. "Introduction to Population Genetics Theory," Harper and Row, New York.
- CROWELL, K. L. 1973. Experimental Zoogeography: introductions of mice onto small islands, *Amer. Nat.* **107**, 535-558.
- KIMURA, M., AND CROW, J. F. 1964. The number of alleles that can be maintained in a finite population, *Genetics* **49**, 725-738.
- LATTER, B. D. H. 1973. The island model of population differentiation: a general solution, *Genetics* **73**, 147-157.
- LEVIN, B. A., AND KILMER, W. L. 1974. Interdemic selection and the evolution of altruism: a computer simulation study, *Evolution* **28**, 527-545.
- LEVINS, R. 1970. Extinction, in "Some Mathematical Problems in Biology" (M. Gerstenhaber, Ed.), Vol. 2, pp. 75-108, Amer. Math. Soc., Providence, R. I.
- MARUYAMA, T. 1970. Effective number of alleles in a subdivided population, *Theor. Pop. Biol.* **1**, 273-306.
- SIMBERLOFF, D. S., AND WILSON, E. O. 1969. Experimental zoogeography of islands: the colonization of empty islands, *Ecology* **50**, 278-296.
- SLATKIN, M., AND WADE, M. J. 1977. Group selection on a quantitative character, in preparation.
- WRIGHT, S. 1931. Evolution in Mendelian populations, *Genetics* **16**, 97-159.