Heredity **56** (1986) 409–415

Received 27 September 1985

A Quasi-equilibrium theory of the distribution of rare alleles in a subdivided population

N. H. Barton* and M. Slatkin†

* Department of Genetics and Biometry, University College, 4 Stephenson Way, London, NW1 2HE, U.K.
† Department of Zoology, University of California, Berkeley, CA 94720, U.S.A.

The conditional average frequency of rare alleles has been shown in simulations to provide a simple and robust estimator of the number of individuals exchanged between local populations in an island model (Nm). This statistic is defined as the average frequency of an allele in those samples in which the allele is present. Here, we show that the conditional average frequency can be calculated from the distribution of allele frequencies. It is a measure of the spread of this distribution, and so is analogous to the standardised variance, F_{ST}. Analytic predictions for the island model of migration agree well with the corresponding simulation results. These predictions are based on the assumption that the rare alleles found in samples have reached a "quasi-equilibrium" distribution. As well as relating the conditional average frequency to the underlying allele frequency distribution, our results provide a more accurate method of estimating Nm from the conditional average frequency of private alleles in samples of different sizes.

INTRODUCTION

There is a variety of ways of analysing allele frequencies taken from natural populations. One goal of such analysis is to reveal the underlying population structure: that is, to estimate such parameters as population density, dispersal rate, and degree of subdivision. Wright (1935) introduced the standardised variance of allele frequency, F_{ST}, as a measure of interpopulation differentiation; this statistic has been widely used. Wright (1938; 1943; 1978) argues that F_{ST} is inversely related to the product of density and dispersal, and so can be used to estimate this product, which he termed "neighbourhood size" (*e.g.*, Dobzhansky and Wright, 1941). Slatkin (1981; 1985*a*) introduced another method, concentrating on the distributions of alleles which are so rare that they are found in only a small proportion of samples. He used computer simulations to show that the average frequency of such alleles, taken over the samples in which they are present, is inversely related to Nm, where N is the local population size and m is the proportion of migrants. In this paper we introduce a simple analytic model which predicts many features of Slatkin's simulation studies. This brings out the relation between F_{ST} and Slatkin's measures. We also present additional simulation results that extend the range of applicability of Slatkin's method.

A QUASI-EQUILIBRIUM MODEL

Our analysis is based on Wright's (1931) island model. The novel feature of our method is the use of Wright's model when the underlying assumptions are clearly not satisfied. Our approximation is similar to that of Maruyama (1972).

Consider a collection of d identical demes, which make up a diploid, monoecious species. We concentrate on one particular allele and suppose that its frequency in some deme is a random variable x. For the models we consider, the frequencies of all other alleles can be considered equivalent, with total frequency $(1-x)$. Under any given model, we can regard x as being sampled from some distribution, $\phi(x)$. If $\phi(x)$ is known, it is easy to find the distribution of the number of copies of the allele in samples of a given size from a specified number of demes. We will first show how the statistics used to describe Slatkin's simulation results can be derived from $\phi(x)$, and then discuss the assumptions needed to derive $\phi(x)$ from Wright's island model.

Slatkin's (1981; 1985*a*) simulations were of an infinite number of alleles at a single locus. He defined the "conditional average frequency" of an allele, $\bar{p}$, to be its average frequency over those samples in which it is present. This is a function of the number of samples (i) in which the allele is present; i is between 1 and the number of demes from which samples are taken, d_{sam}. Thus, each allele is characterised by two statistics, $\bar{p}$ and i. Similarly, in the analytic model, we can calculate the expected values of $\bar{p}$ and i for any particular allele, provided that we know the distribution, $\phi(x)$. As we take larger and larger samples of demes ($d_{sam} \to$ infinity) the actual statistics will converge to these expected values. Although we will only analyse cases in which $\phi(x)$ is the same for each deme, and in which the allele frequencies in different demes are independent of each other, these assumptions are not essential: the expected values of $\bar{p}$ and i depend only on the marginal distributions of allele frequencies in each deme.

To begin with, assume that we sample every gene in each deme examined. We define the probability that an allele is absent from a sample as α_0. Then, the expected number of demes in which it will be found is:

$$\langle i \rangle = d(1-\alpha_0). \tag{1}$$

The value of α_0 can be computed from $\phi(x)$ using the approximate formula $\alpha_0 = \int_0^{1/2N} \phi(x)\,dx$. Ewens (1979) discusses the limitations of this approximation. In the application here, its validity was checked by computing the exact results for the Markov chain which describes the Wright–Fisher model. The expected frequency of the allele in a deme, given that it is present, is:

$$\langle \bar{p} \rangle = \frac{\int_0^1 x\phi(x)\,dx}{(1-\alpha_0)}. \tag{2}$$

To derive $\phi(x)$, we use Wright's island model and assume that each deme, which is of effective size 2N, receives immigrants at a rate m from a source in which the frequency of the allele is y. (In the simulations, this source is the rest of the population). Initially, we will only consider neutral alleles, for which the equilibrium distribution in each deme is, to a good approximation, a beta distribution:

$$\phi_y(x) = \frac{\Gamma(4\mathrm{Nm})}{\Gamma(4\mathrm{Nm}y)\Gamma(4\mathrm{Nm}(1-y))} \times x^{4\mathrm{Nm}y-1}(1-x)^{4\mathrm{Nm}(1-y)-1} \tag{3}$$

(where $\Gamma(\cdot)$ is the gamma function; Abramowitz and Stegun, 1965). In equation 3, the subscript y is added to emphasise that ϕ depends on y. If there are mutations away from the allele at a rate μ, then the parameters of the beta distribution are changed from 4Nmy and 4Nm$(1-y)$ to 4Nmy and 4N$[\mathrm{m}(1-y)+\mu]$(assuming $\mu \ll 1$). We will discuss selection later.

To relate this analytic model to Slatkin's simulations, we consider a set of alleles which differ in y, their frequency in the population. Since all these alleles experience the same population structure, their conditional average frequency and the proportion of demes in which each is found, will depend primarily on y. By calculating both statistics as functions of y, we can derive the relation $\bar{p}(i)$ which was used to describe the simulation results.

We are assuming, in effect, that each deme in the simulated population has reached an equilibrium between genetic drift (and possibly mutation and selection) and immigration. This assumption will be approximately correct if the timescale of change in each deme is much shorter than the timescale of change in the population as a whole. We call this a "quasi-equilibrium" approximation because a true equilibrium cannot be attained under the assumptions of the simulation model: because each mutation gives a unique allele, every new allele which enters the population must eventually be lost.

SAMPLES OF DEMES

In fig. 1, we compare some simulation results from Slatkin's (1981; 1985*a*) model with predictions of our quasi-equilibrium theory for comparable parameter values; here, every deme is sampled ($d_{sam} = d$). We can see that there is very good agreement between the analytic and simulation results for larger values of i, but some difference for small values. This is reasonable, since the assumptions of the analytic model are most likely to be met for alleles that have spread to several demes. That suggests that there might be better agreement for small values of i if only a few demes from a large population are sampled; alleles found in only a few demes of the sample are then likely to be present in many demes in the whole population, and their overall frequency is therefore likely to evolve slowly relative to the rapid fluctuations within a deme.

Slatkin (1985*a*) showed that $\bar{p}(1)$, the average frequency of "private" alleles found in only a

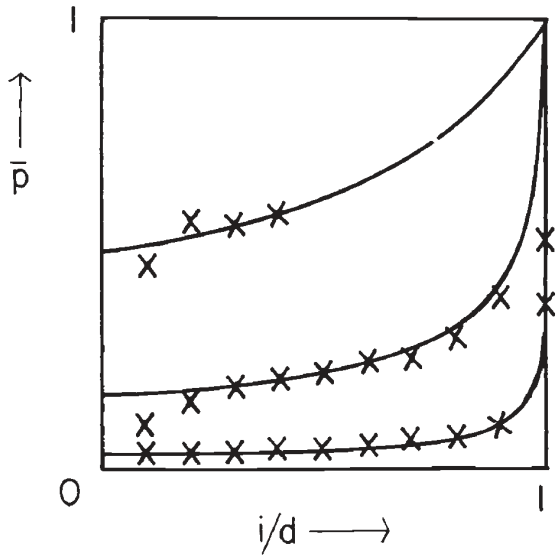

Figure 1 A comparison of the conditional average frequency, $\bar{p}(i)$, computed from the simulations, with the predictions of the analytic theory for the island model. The solid curve is a graph of $\langle\bar{p}\rangle$ against $\langle i/d\rangle$ as the parameter y varies between 0 and 1 (equations 1–3). Values from the simulations are indicated by crosses. For each curve, N = 32, and $\mu = 0{\cdot}001$. The three curves are (reading from bottom to top) for Nm = 3·2, 0·32, and 0·032; the conditional average frequency decreases as Nm increases. In the simulations, an island model with $d = 10$ demes was used; all the individuals in each deme were sampled (n = 32). In these and other simulations described in this paper, the population was initially fixed for a single allele. The simulations were then run for at least 1000 generations, with samples being taken every 200 generations after that time until at least 10,000 generations were run. When Nm is low, very few alleles reach high occupancy numbers; there are therefore some random deviations between the analytic theory and simulations for large i, and we have not been able to plot reliable values for Nm = 0·032 and $i > 4$.

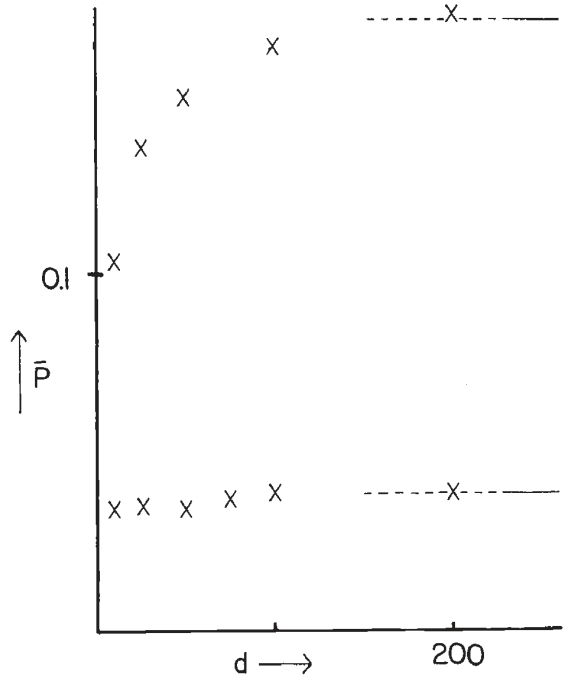

Figure 2 The dependence of the average frequency of private alleles, $\bar{p}(1)$, in a fixed number of sampled demes (d_{sam}), on the number of demes in the population (d). The dashed lines are the values predicted by the analytic theory and do not depend on d. The simulation results are indicated by crosses. The lower series of crosses gives the values for m = 0·1 (Nm = 3·2), whilst the upper series gives values for m = 0·01 (Nm = 0·32). In both cases, N = 32, $\mu = 0{\cdot}001$, and $d_{sam} = 10$.

single deme, is of particular interest; it is therefore especially important to understand the apparent discrepancy between theory and simulation for small values of i. Although we could compute $\langle\bar{p}(1)\rangle$ from the analytic theory, it is easier to calculate $\langle\bar{p}(0)\rangle$, the limit of $\bar{p}$ for extremely rare alleles. Since the theoretical graph of $\bar{p}$ against i is almost flat for small values of i (fig. 1), and since we will usually be considering relatively large samples of demes ($1/d_{sam} \ll 1$), this will cause little error.

Fig. 1 shows that the difference between $\bar{p}(1)$, from the simulations, and $\langle\bar{p}(0)\rangle$, from the theory, can be substantial if all demes are sampled. Fig. 2 shows that as the total number of demes increases, relative to a fixed sample, the simulations converge towards the theory. There is very little dependence on d for larger neighbourhood sizes (Nm = 3·2), but stronger dependence for smaller neighbourhood sizes (Nm = 0·32). Even then, however, there is good agreement once the sample consists of less than 10 per cent of the species. In practice, of course, a smaller fraction of the total population is usually taken.

AVERAGE FREQUENCIES OF PRIVATE ALLELES

Slatkin (1985*a*) found a relatively simple relationship between the frequencies of alleles found in only one population, and Nm. As shown by the solid curve in fig. 3, $\log_{10}[\bar{p}(1)]$ is an approximately linear function of $\log_{10}$ (Nm) for intermediate values of Nm. (In fig. 3, d was chosen to be large enough to make $\bar{p}(1)$ essentially independent of d.) The simple form of this relationship makes the average frequency of private alleles, $\bar{p}(1)$, useful for estimating Nm. Fig. 3 also shows values of $\langle\bar{p}(0)\rangle$, as computed from the analytic theory, over a range of values of Nm. The curves are in good agreement, being roughly linear on a log–log scale,

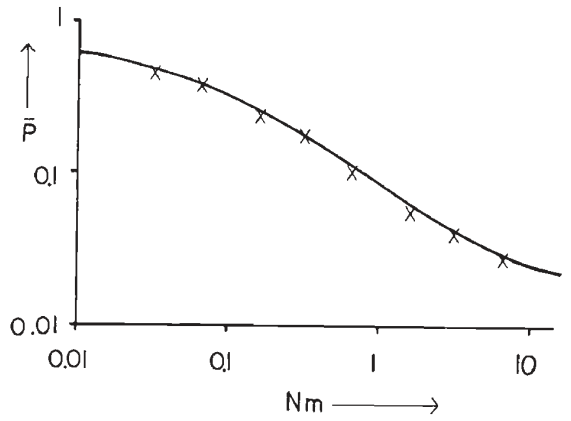

Figure 3 The dependence of the average frequency of private alleles, $\bar{p}(1)$, on Nm. The values from the simulations are indicated by crosses. The predictions of the analytic theory are given by the solid curve. $N = 32$, $\mu = 0{\cdot}001$, $d = 100$, $d_{sam} = 10$. An island model of the population was assumed.

and levelling off for small and large values of Nm. The slope of the regression line fitted to the simulation results is $-1{\cdot}14$, and to the analytic results is $-1{\cdot}09$.

The graph of the analytic values was generated by numerical computations of $\phi(x)$ (equation 3). These values were checked against the exact equilibrium distribution for the island model, calculated by inverting the transition matrix of the associated Wright-Fisher Markov process (Ewens, 1979); the results were indistinguishable. If the deme size is large ($2N \gg 1$), the frequency of private alleles is approximated by

$$\langle \bar{p}(0) \rangle = 1/4\mathrm{Nm}[\log(2N) - \psi(4\mathrm{Nm})] \qquad (4)$$

(where $\psi(x) = d(\ln \Gamma(x))/dx$ is the digamma function (Abramowitz and Stegun, 1965); $\psi(n) = \sum_{k=1}^{n-1}(1/k) - 0{\cdot}5772\ldots$). This formula becomes exact only in the limit of very large deme size; in the figures, the analytic results were calculated from the exact island model (equations 1-3), rather than from this approximation. (The same comment applies to the similar approximation derived in equation (6) below.)

SAMPLES OF INDIVIDUALS

The results described so far are based on the assumption that every individual in a deme is sampled. In studies of natural populations, the actual sizes of local populations are unknown. It is important, then, to know how these results depend on the proportion of individuals sampled. Slatkin (1985*a*) presented some results indicating that $\bar{p}(1)$ depends primarily on Nm and sample size, but not on N separately; here, we will show that this independence of the unknown deme size is expected under the theory.

It is simplest to assume that n individuals are sampled with replacement. The probability that an allele which is at frequency x will be found in such a sample is then $1-(1-x)^{2n}$, and equation 1 is replaced by:

$$\langle i \rangle = d\int_0^1 (1-(1-x)^{2n})\phi(x)\,dx \qquad (5)$$

(Slatkin (1981; 1985*a*) used the notation N_{sam} for the number of individuals sampled; we use n here to make the formulae easier to read). We have also made calculations assuming sampling without replacement; the results are nearly the same when $n \ll N$, as we would expect for natural populations.

Fig. 4 shows the analytic and simulation results for the dependence on sample size. The results agree qualitatively, though the simulations give rather smaller values of $\bar{p}(1)$ than would be expected theoretically. Again, we can derive an explicit formula for the frequency of private alleles, by taking the limit $y \to 0$:

$$\langle \bar{p}(0) \rangle = 1/4\mathrm{Nm}[\psi(4\mathrm{Nm}+2n) - \psi(4\mathrm{Nm})] \qquad (6a)$$

$$= 1/4\mathrm{Nm}\ln(1+2n/4\mathrm{Nm})$$

$$\mathrm{Nm} > 0{\cdot}5 \qquad (6b)$$

$$= 1/\{1+4\mathrm{Nm}[\ln(2n)+0{\cdot}5772]\}$$

$$\mathrm{Nm} < 0{\cdot}1. \qquad (6c)$$

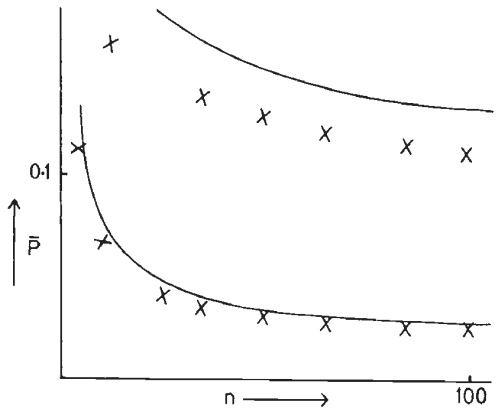

Figure 4 The dependence of the average frequency of private alleles, $\bar{p}(1)$, on the number of individuals sampled (n). In the simulations, $N = 128$, $\mu = 0{\cdot}00025$, $d = 100$, and $d_{sam} = 10$. An island model was assumed. The solid line indicates the analytic results. The lower series gives values for $m = 0{\cdot}025$ ($Nm = 3{\cdot}2$), whilst the upper series gives values for $m = 0{\cdot}0025$ ($Nm = 0{\cdot}32$).

ESTIMATING Nm

Slatkin (1985*a*) presented a method for estimating Nm from the value of $\bar{p}(1)$ if the number of individuals sampled per deme is the same as was used in the simulations, n = 25. He suggested a rough way to adjust the estimate of Nm if the numbers of individuals sampled differed from 25. The method is to find $\bar{p}(1)$ from the allele frequencies and then estimate Nm either directly from a graph of $\bar{p}(1)$ against Nm or approximately from the line fitted to the simulated values for n = 25. Then the estimate of Nm is corrected by multiplying the estimate by the ratio of the average number of individuals sampled per deme to 25. For example, if the average sample size is 50 and the uncorrected estimate of Nm is 0.5, then the corrected estimate is 0.25.

This method for correcting for differences in sample size would be accurate if the graphs of $\log[\bar{p}(1)]$ against log (Nm) for different values of n were parallel. Fig. 5 shows the actual results for three different samples sizes. The simulation results for different values of n are approximately linear but have slightly different slopes, a = −0·489, −0·576, and −0·612 for n = 10, 25 and 50. The values of the intercepts are b = −0·951, −1·11, and −1·21.

Although the simulation results show that Slatkin's method of correcting for differences in sample size does not work exactly, the slopes of the lines are similar enough that his method provides a reasonable approximation. For example, assume

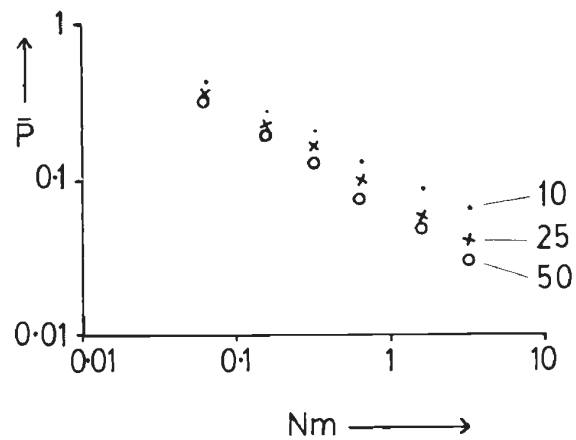

Figure 5 The dependence of the average frequency of private alleles in samples on Nm and *n*, the number of individuals sampled from each deme. In all cases, N = 128, $\mu = 0{\cdot}001$, $d = 100$, and $d_{sam} = 10$. An island model of population structure was assumed. The dots are values for n = 10, the crosses are values for n = 25, and the open circles are values for n = 50.

that in a sample of n = 10 individuals from every deme, $\bar{p}(1)$ is found to be 0·1. Using n = 25, the estimated value of Nm is 0·64, which is found by solving the regression equation for Nm with a = −0·576 and b = −1·11. Using Slatkin's method, the estimate of Nm corrected for sample size is 1·6, which is obtained by multiplying the uncorrected estimate by 25/10. The more accurate estimate of Nm is 1·3, which is obtained by using a = −0.489 and b = −0·951. The approximate estimate is too high by about 25 per cent. If instead, $\bar{p}(1) = 0.1$ and n = 50, Slatkin's method gives an estimate of 0·32 and the more accurate estimate is 0·45. These examples show that there is some error in using Slatkin's method for correcting for sample size but that the estimates obtained are of the correct order of magnitude.

SELECTION

The quasi-equilibrium model is remarkably successful in predicting the distribution of rare and neutral alleles. We can use the same approach to incorporate the effects of selection. In this paper, we will restrict our analyses to symmetric over- and under-dominance, but other types of selection could be studied in the same way.

The assumptions of the analytic model are the same as used previously, except that all heterozygotes have fitnesses $(1 - s)$ relative to homozygotes; s may be positive or negative. Using Wright's (1931) general formula for the distribution of allele frequencies under immigration, drift, and selection, we obtain:

$$\phi_y(x) = Cx^{4Nmy-1}(1-x)^{4Nm(1-y)-1}\, e^{-4Nsx(1-x)} \tag{7}$$

(where C is a normalisation constant). This replaces the beta distribution used in the previous calculations.

We could not find a simple analytic approximation for $\langle\bar{p}(0)\rangle$ using (7). However, if we assume that x is small for the alleles of interest and replace $4Nsx(1-x)$ by $4Nsx$ in (7) (which is equivalent to assuming gametic selection of strength s), then using the same methods as led to equation (6), we find

$$\langle\bar{p}(0)\rangle = \frac{1}{4N(m+s)[\psi(4N(m+s)+2n) - \psi(4N(m+s))]}. \tag{8}$$

We therefore expect the relation between the frequency of private alleles and Nm to be shifted horizontally by Ns; rare underdominant alleles will have lower conditional average frequency. This behaviour is seen in fig. 6; the values for the analytic theory were obtained using (7) and they

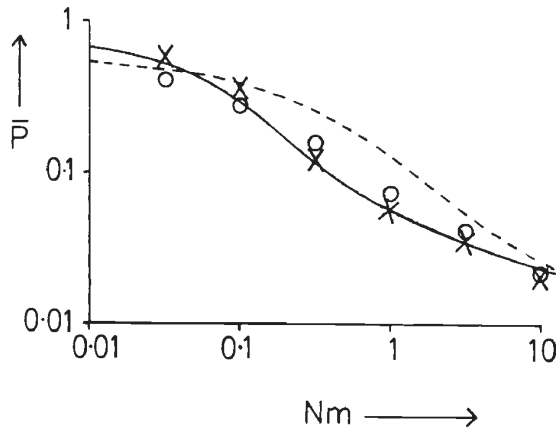

Figure 6 The dependence of the conditional average frequency, $\bar{p}(1)$, on Nm for overdominant and underdominant alleles. The continuous lines indicate the analytic results, whilst the simulations are represented by crosses or circles. The crosses, and the solid line, are for cases with uniform underdominance, in which every heterozygote has a fitness of 0·95 relative to every homozygote ($s = +0{\cdot}05$). The circles, and the dashed line, are for uniform overdominance in which every heterozygote has a fitness of 1·05 relative to every homozygote ($s = -0{\cdot}05$). In all cases, N = 32, n = 32, $\mu = 0{\cdot}001$, $d = 100$, and $d_{sam} = 10$. An island model was assumed.

agree fairly well with the simulations for underdominant alleles, even when Nm is small. However, there is a large discrepancy for overdominant alleles. This may be because a new overdominant allele will initially increase its overall frequency over a time scale $1/s$, of the same order as the time scale of changes within a deme when $m \approx s$. In contrast, the overall frequency of a neutral allele changes over a time scale $\approx 4Nd$, and so changes very slowly when the species is made up of many demes ($d \gg 1$).

ACCURACY OF THE QUASI-EQUILIBRIUM ASSUMPTION

By using the assumption that the allele frequencies in each deme have the distribution expected under Wright's island model, we can predict remarkably well several features of the simulation results. We can also use the simulation program to test the accuracy of the quasi-equilibrium assumption.

For neutral alleles, $\phi(x)$, the frequency distribution of the allele under consideration, should be given by a beta distribution (equation 4). If the quasi-equilibrium assumption is valid, then, in the simulations, an allele that has a frequency y in the population should have the frequency distribution across different demes given by equation 4, with the parameters of the beta distribution being $4Nmy$ and $4N[m(1-y)+\mu]$, where μ is the mutation rate of A to a, N is the population size, and y is the frequency of A in the population. In comparing the simulation results with the analytic theory, a represents all other alleles. In the simulations, every mutation is new, so there are no back mutations.

We used two tests of the agreement of the actual distribution of alleles with a beta distribution. One was a comparison of the numbers of demes missing the allele with the number expected under the beta distribution, and the other was a comparison of the actual and expected variances in allele frequency. The two tests depend on different features of the frequency distribution.

In the simulations, the population was sampled at a time large enough that the stationarity state had been achieved. For each allele in the population, its frequency, y, was computed. Then α_0, the expected fraction of demes missing the allele, was computed from $\phi_y(x)$ using the method recommended by Ewens (1979, pp. 157–158), and the variance in frequency among the demes, V_x, was computed using the standard formula for the variance of a beta distribution.

For each allele, the actual number of demes missing the allele was compared with a binomial distribution with parameters d (the total number of demes) and α_0 (the probability that each deme is missing the allele). The binomial distribution had to be modified to take account of the fact that each allele had to be found in at least one deme. The result of this test applied to each allele was a probability that the observed number or fewer of demes missing the allele would be obtained if the quasi-equilibrium assumption were satisfied. To test whether there were significant deviations in the variance, we used the methods described by Kendall and Stuart (1977, Ch. 10) for finding the standard error of the variance in a sample to the fourth moment of the distribution. From $\phi_y(x)$, we computed the variance and the standard error of the variance, and then compared them to the actual variance for the allele.

We applied these tests to simulations, over the range of parameter values used in the figures. For neutral alleles in an island model with 100 demes,

the distribution of allele frequencies did appear to be drawn from a beta distribution with the correct parameters; the distribution of alleles among the demes is not distinguishable by our tests from the beta distribution expected under the assumptions of the quasi-equilibrium model. These results suggest that the ability of the quasi-equilibrium model to predict the conditional average frequencies is not fortuitous.

DISCUSSION

We have shown that many of the properties of the distribution of rare alleles in a subdivided population can be predicted from a relatively simple analytic theory that assumes an approximately equilibrium distribution of allele frequencies among the demes comprising a population. This approach contrasts with that used by Slatkin and Takahata (1985), who calculated the conditional average frequency of alleles which never leave the deme in which they originate. Slatkin and Takahata's results for this transient model give poor agreement with the simulations; our analysis suggests that this disagreement arises because most of those alleles which are found in only a single sample have in fact been present for long enough to reach a "quasi-equilibrium" under migration and drift.

This analytic theory provides a way to relate the conditional average frequencies to the overall distribution of allele frequencies in the population. There are still many unsolved problems associated with the use of allele frequency distributions to draw inferences about population structure; the most robust and efficient estimators of Nm, neighborhood size, and other properties of a subdivided population are unknown. Both F_{ST} and the conditional average frequencies can be and have been used to estimate Nm; the resulting estimates are generally consistent (Slatkin, 1985*b*). We are currently investigating how these statistics may best be used.

Acknowledgements This research has been supported in part by grants from the Royal Society of London, S.E.R.C., and the National Science Foundation. We thank J. Felsenstein for helpful discussions of this problem.

REFERENCES

ABRAMOWITZ, M. AND STEGUN, I. 1965. *Handbook of Mathematical Functions.* Dover, New York.

DOBZHANSKY, T. AND WRIGHT, S. 1941. Relations between mutation rate and accumulation of lethals in populations of *Drosophila pseudoobscura. Genetics, 26,* 23-51.

EWENS, W. J. 1979. *Mathematical Population Genetics.* Springer Verlag, Berlin.

KENDALL, M. AND STUART, A. 1977. *The Advanced Theory of Statistics.* 4th ed. MacMillan, New York.

MALÉCOT, G. 1948. *Le Mathématique de l'Hérédité.* Masson et Cie, Paris.

MARUYAMA, T. 1972. Distribution of gene frequencies in a geographically structured finite population. I. Distribution of neutral genes and of genes with small effect. *Ann. Hum. Genet., 35,* 411-423.

SLATKIN, M. 1981. Estimating levels of gene flow in natural populations. *Genetics, 99,* 323-335.

SLATKIN, M. 1985*a*. Rare alleles as indicators of gene flow. *Evolution, 39,* 53-65.

SLATKIN, M. 1985*b*. Gene flow in natural populations. *Ann. Rev. Ecol. Syst., 16,* 393-430.

SLATKIN, M. AND TAKAHATA, N. 1985. The average frequency of private alleles in a partially isolated population. *Theor. Pop. Biol., 28,* 314-331.

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

WRIGHT, S. 1935. Evolution in populations in approximate equilibrium. *J. Genetics, 30,* 257-266.

WRIGHT, S. 1938. The distribution of gene frequencies under irreversible mutation. *Proc. Natl. Acad. Sci. (USA), 245,* 253-259.

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

WRIGHT, S. 1978. *Evolution and the Genetics of Poulations. Vol. 4. Variability Within cnd Among Natural Populations.* Univ. of Chicago Press, Chicago.