On the Theory of Partially Inbreeding Finite Populations. I. Partial Selfing[1]

Edward Pollak[2]

Department of Statistics, Iowa State University, Ames, Iowa 50011

Manuscript received January 5, 1987
Revised copy accepted June 24, 1987

ABSTRACT

Some stochastic theory is developed for monoecious populations of size N in which there are probabilities β and $1 - \beta$ of reproduction by selfing and by random mating. It is assumed that $\beta \gg N^{-1}$. Expressions are derived for the inbreeding coefficient of one random individual and the coefficient of kinship of two random separate individuals at time t. The mean and between-lines variance of the fraction of copies of a locus that are identical in two random separate individuals in an equilibrium population are obtained under the assumption that there is an infinite number of possible alleles. It is found that the theory for random mating populations holds if the effective population number is $N_e = N'/(1 + F_{IS})$, where F_{IS} is the inbreeding coefficient at equilibrium when N is infinite and N' is the reciprocal of the probability that two gametes contributing to random separate adults come from the same parent. When there is a binomial distribution of successful gametes emanating from each adult, $N' = N$. An approximation to the probability that an allele A survives if it is originally present in one $A\bar{A}$ heterozygote is found to be $2(N'/N)(F_{IS}s_1 + (1 - F_{IS})s_2)$, where s_1 and s_2 are the selective advantages of AA and $A\bar{A}$ in comparison with $\bar{A}\bar{A}$. In the last section it is shown that if there is partial full sib mating and binomial offspring distributions $N_e = N/(1 + 3F_{IS})$.

ALMOST all the theory that now exists for finite populations is based upon the assumption that reproduction is wholly by random mating. There is, nevertheless, a substantial amount of inbreeding in many populations, such as those consisting of plants that reproduce normally by self-fertilization. This fact has been incorporated in models under which it is assumed that there is an infinite population and probabilities β and $1 - \beta$ of mating of relatives and mating randomly. Thus, if β is the probability of selfing, the equilibrium value of the inbreeding coefficient is $\beta/(2 - \beta)$, which is a result that was first derived by HALDANE (1924). If there is reproduction partially by full sib mating, it was shown by LI (1976) that the inbreeding coefficient in a population at equilibrium is $\beta/(4 - 3\beta)$.

The earliest reference that I know of in which there are models that incorporate a finite population size and partial inbreeding is a paper of WRIGHT (1951). He derived equations that hold when there is partial selfing or partial mating of the full sibs that are offspring of permanent couples, but only solved them in some special cases.

My main objective in this paper is to further develop the theory for partially selfing finite populations, while noting resemblances and differences between it and the usual theory for random mating populations. A secondary objective is to discuss which of the results for partial selfing are likely to apply to other partially inbreeding populations and point out one result which does not.

Thus, for example, if there is no selection or mutation and F_t is the inbreeding coefficient in generation t, it turns out that for moderately large t $1 - F_t$ is proportional to $1 - (2N_e^{-1})^t$, where N_e is a suitably defined effective population size. To this extent there is a formal resemblance to the usual theory. If, however there is an infinite number of possible alleles, we must calculate the probability that two copies of a gene in random separate individuals are identical, in place of the probability of homozygosity.

I will also derive the probability of the ultimate survival of an allele A that is initially present in one heterozygote in a large population. It turns out that it is a linear combination of the selective advantages of AA and $A\bar{A}$ in comparison with $\bar{A}\bar{A}$, rather than a multiple of the selective advantage of $A\bar{A}$, as in the usual theory. The calculations are based upon an approximation of the model by a branching process with two types of individuals, which are the homozygotes and heterozygotes that have A in their genotypes. If there is weak selection we may then use an approximate expression for the probability of survival of a line descended from a single heterozygous individual (of type $A\bar{A}$), which was derived rigorously by ESHEL (1984). An essentially equivalent formula was

---

[1] Journal Paper No. J-12528 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa, Project 2588.

[2] On leave at the Department of Animal Sciences, University of Illinois, Urbana, Illinois 61801.

Genetics 117: 353–360 (October, 1987)

derived heuristically, but checked on numerical examples, by POLLAK (1972).

I shall conclude the derivations by demonstrating that, under WRIGHT's model of partial full sib mating, $N_e \doteq N/(1 + 3F_{IS})$, where N is the population size and F_{IS} is the equilibrium value of the inbreeding coefficient when N is infinite. This contrasts with $N_e \doteq N/(1 + F_{IS})$, the expression for his model of partial selfing.

PROBABILITIES OF IDENTITY OF PAIRS OF GENES

We consider a monoecious population having a constant population size N, in which there is self-fertilization with probability β and the union of gametes from random separate individuals with probability $1 - \beta$. It is also assumed that there is an infinite number of possible alleles and that u is the probability that a particular allele mutates in one generation. Let F_t and θ_t respectively denote the probabilities that two gametes uniting to produce one individual and two gametes contributing to random separate individuals of generation t carry the same allele at a locus. Then

$$\begin{aligned} F_{t+1} &= (1-u)^2\left[\frac{\beta}{2}(1+F_t) + (1-\beta)\theta_t\right], \\ \theta_{t+1} &= (1-u)^2\left[\frac{1}{N}\frac{1}{2}(1+F_t) + \left(1-\frac{1}{N}\right)\theta_t\right]. \end{aligned} \tag{1}$$

When $u = 0$, equations (1) reduce to those found by WRIGHT (1951).

We henceforth assume that N is large, u is of the same order of magnitude as N^{-1} and $\beta \gg N^{-1}$. Then at equilibrium $F_{t+1} = F_t = F$ and $\theta_{t+1} = \theta_t = \theta$, where

$$\begin{aligned} F &= \frac{1 + 2N\beta u}{1 + 4N(1-\beta/2)u}, \\ \theta &= \frac{1}{1 + 4N(1-\beta/2)u}, \end{aligned} \tag{2}$$

approximately. At a finite time t, however, $F_t = F + d_t$ and $\theta_t = \theta + e_t$, where, by (1),

$$\begin{aligned} d_{t+1} &= (1-u)^2\left[\frac{\beta}{2}d_t + (1-\beta)e_t\right], \\ e_{t+1} &= (1-u)^2\left[\frac{1}{2N}d_t + \left(1-\frac{1}{N}\right)e_t\right]. \end{aligned} \tag{3}$$

The characteristic equation corresponding to the matrix of coefficients in (3) has the roots

$$\lambda_1 = 1 - \frac{4N(1-\beta/2)u + 1}{2N(1-\beta/2)} + O(N^{-2})$$

and

$$\lambda_2 = \frac{\beta}{2} + O(N^{-1}).$$

If, therefore, we define an effective population number

$$N_e = N\left(1 - \frac{\beta}{2}\right) = \frac{N}{1 + \beta/(2-\beta)}$$

and set $M = 4N_e u$ it is evident that the general solution of (1) is approximately

$$\begin{aligned} F_t &= \frac{1 + 2N\beta u}{1 + M} + a\left(1 - \frac{M+1}{2N_e}\right)^t + b\left(\frac{\beta}{2}\right)^t, \\ \theta_t &= \frac{1}{1+M} + c\left(1 - \frac{M+1}{2N_e}\right)^t + d\left(\frac{\beta}{2}\right)^t, \end{aligned} \tag{4}$$

where a, b, c and d are constants.

In the particular case in which $u = 0$ and $F_0 = \theta_0 = 0$, equations (4) reduce to

$$\begin{aligned} F_t &\doteq 1 - \left(1 - \frac{\beta}{2-\beta}\right)\left(1 - \frac{1}{2N_e}\right)^t - \frac{\beta}{2-\beta}\left(\frac{\beta}{2}\right)^t \\ \theta_t &\doteq 1 - \left(1 - \frac{1}{2N_e}\right)^t, \end{aligned} \tag{5}$$

where F_t and θ_t are now probabilities of identity by descent rather than identity in state. Because $F_0 = r_0 = 0$, the left sides of (5) refer to identity by descent that is induced during the t generations in which partial selfing has been applied to a finite population. Equations (2), as well as equations (1) in the special case in which $u = 0$, have also been given by STROBECK (1983).

THE F-STATISTICS FOR THIS SITUATION

The form taken by the right-hand sides of (5) is no accident, but has a connection with the F-statistics introduced by WRIGHT (1943, 1951). We note first that as t increases, these equations may be written as

$$1 - F_t \doteq \left(1 - \frac{\beta}{2-\beta}\right)(1 - \theta_t).$$

This is the same in form as

$$1 - F_{IT} = (1 - F_{IS})(1 - F_{ST}).$$

Thus, at least in a formal sense, we can equate F_t to F_{IT}, θ_t to F_{ST} and $\beta/(2-\beta)$ to F_{IS}.

But there is more to it than this. Let us suppose that in generation 0 there is an infinite population. Randomly chosen sets of N individuals are chosen to be the founding members of subpopulations, or lines. These lines are then each subjected to the same regular system of inbreeding as the others, generation after generation, and develop in complete isolation from each other. Then F_t is a multiple of the expected reduction of the proportion of heterozygotes in the population as a whole, brought about by the combined effects of partial selfing and random drift. Also, if x is the frequency of a particular allele A in a random subpopulation at time t and μ_x is the average of such frequencies in the population as a whole,

$$E(x^2) = \sigma_x^2 + \mu_x^2$$

is the probability that two gametes contributing to random separate individuals in one of the subpopulations carry allele A. Since, however, two such copies of A originate from separate ancestral genes with probability $(1 - \theta_t)\mu_x^2$ and from a common ancestor with probability $\theta_t\mu_x$ it follows that

$$E(x)^2 = \mu_x[\theta_t + (1 - \theta_t)\mu_x] = \mu_x^2 + \theta_t\mu_x(1 - \mu_x).$$

Thus

$$\theta_t = \frac{\sigma_x^2}{\mu_x(1 - \mu_x)} = F_{ST},$$

so that θ_t is measuring the reduction in heterozygosity from random sampling in finite subpopulations. Therefore, finally, $F_{IS} = \beta/(2 - \beta)$ is approximately equal to a multiple of the expected amount of reduction in heterozygosity within subpopulations that is caused by partial selfing. It is also, not surprisingly, the limiting value, as t increases, of the inbreeding coefficient at time t in an infinite population in which there is selfing with probability β and random mating with probability $1 - \beta$ and an initial Hardy-Weinberg structure. We note also that

$$N_e = \frac{N}{1 + F_{IS}}, \tag{6}$$

which is a form that was also derived by LI (1976).

SOME HIGHER ORDER PROBABILITIES OF IDENTITY

The expression θ_t is equal to the expected fraction of gene pairs from random separate individuals of generation t that are identical in state. It is the analog in the partially inbred population of the expected fraction of homozygotes when $\beta = 0$. Thus, consistent with the terminology of KIMURA (1983), we may call $1 - \theta_t$ the expected value of the virtual heterozygosity at time t.

In this section, I shall derive an expression for the equilibrium value of the between-lines variance of the virtual heterozygosity from fourth order probabilities of identity. Since, however, nth order probabilities of identity at time $t + 1$ depend upon probabilities of orders 2 through $n - 1$ at time t when n is 3 or 4, it will be necessary to first calculate equilibrium values of third order probabilities. The reasoning will be based on that employed by COCKERHAM (1971).

We consider separate individuals I, II, III and IV and denote random copies of a gene from these individuals by 1, 2, 3, and 4, respectively. Two separate copies of a gene carried by an individual will be denoted by an arabic numeral and the same arabic numeral with a prime. In this notation, F_t and θ_t are written as $f_t(11')$ and $f_t(12)$.

If there are three copies of a gene, the probabilities of identity to be calculated are $f_t(123)$ and $f_t(11'2)$. These respectively refer to the identity of copies in three separate individuals and to the identity of three copies when two are in individual I and a third is in individual II. Now two among three random gametes that contribute to adults I, II and III in generation $t + 1$ originate in one parent with probability $3N^{-1} \cdot (1 - N^{-1})$. In this case the conditional probabilities that one gene is copied twice and two separate genes are copied once are both equal to ½. Hence

$$f_{t+1}(123) = (1 - u)^3\left[\frac{3}{N}\left(1 - \frac{1}{N}\right)\frac{1}{2}\,(f_t(12) + f_t(11'2)) + \left(1 - \frac{3}{N}\left(1 - \frac{1}{N}\right)\right)f_t(123)\right]. \tag{7}$$

Since individual I can arise from selfing or random mating with probabilities β and $1 - \beta$, we have as a second recurrence equation

$$f_{t+1}(11'2) = (1 - u)^3\left[\frac{\beta}{2}\,(f_t(12) + f(11'2)) + (1 - \beta)f_t(123)\right]. \tag{8}$$

At equilibrium $f_{t+1}(123) = f_t(123) = f(123)$ and $f_{t+1}(11'2) = f_t(11'2) = f(11'2)$. It can be shown that if $u = O(N^{-1})$ and $\beta \gg N^{-1}$, then

$$f(11'2) \doteq \frac{\beta}{2 - \beta}\frac{1}{1 + M} + \frac{2(1 - \beta)}{2 - \beta}f(123)$$

and

$$f(123) \doteq \frac{2}{(1 + M)(2 + M)}. \tag{9}$$

I shall now obtain recurrence equations for the fourth order probabilities of identity $f_t(12,34)$, $f_t(11',23)$ and $f_t(12,1'3)$. In this notation the identity of a pair of representatives of a locus is symbolized by preceding or following the corresponding set of numbers by a comma. Note, however, that the comma neither implies nor excludes the possibility that all four genes are identical.

The probability that two randomly chosen successful gametes among four originate from a single parent, while the other two come from two other parents, is equal to $6N^{-1}(1 - N^{-1})(1 - 2N^{-1})$. Also, there are two mutually exclusive ways in which successively chosen copies of a gene from I can be identical to copies from II and III. Therefore

$$f_t(12,34) \doteq (1-u)^4\left[\left(1-\frac{6}{N}\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\right)\cdot f_t(12,34) + \frac{1}{N}\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)[f_t(12) + f_t(11',23) + 2f_t(123) + 2f_t(12,1'3)]\right]. \quad (10)$$

Because an individual results from self-fertilization with probability β, the other two recurrence equations are

$$f_{t+1}(11',23) = (1-u)^4\left[\frac{\beta}{2}(f_t(12) + f_t(11',23)) + (1-\beta)f_t(12,34)\right] \quad (11)$$

and

$$f_{t+1}(12,1'3) = (1-u)^4\left[\frac{\beta}{2}(f_t(123) + f_t(12,1'3)) + (1-\beta)f_t(12,34)\right].$$

At equilibrium we have the approximate expressions

$$f(12,34) \doteq \frac{1}{4Nu+6}[\theta + 2f(123) + 2f(12,1'3) + f(11',23)], \quad (13)$$

$$f(11',23) \doteq \frac{1}{2-\beta}[\beta\theta + 2(1-\beta)f(12,34)] \quad (14)$$

and

$$f(12,1'3) \doteq \frac{1}{2-\beta}[\beta f(123) + 2(1-\beta)f(12,34)]. \quad (15)$$

Thus, finally, if we substitute the right sides of (2), (9), (14) and (15) into (13) we obtain

$$f(12,34) \doteq \frac{6+M}{(1+M)(2+M)(3+M)}. \quad (16)$$

Now if x_i is the frequency of A_i in the equilibrium population and N is large

$$\theta \doteq E\left(\sum_i x_i^2\right) = 1 - E(H),$$

where H is the virtual heterozygosity. Similarly

$$E\left[\left(\sum_i x_i^2\right)^2\right] = E\left(\sum_i x_i^2 \sum_j x_j^2\right) \doteq f(12,34) \doteq \frac{6+M}{(1+M)(2+M)(3+M)},$$

so that the between-lines variance of the virtual heterozygosity is

$$\mathrm{Var}(H) \doteq E\left[\left(\sum_i x_i\right)^2\right] - \theta^2 \doteq \frac{2M}{(M+1)^2\,(M+2)(M+3)}. \quad (17)$$

When $\beta = 0$, (17) reduces to the between-lines variance in the heterozygosity, which is a result derived in other ways by LI and NEI (1975) and STEWART (1976). In $\beta > 0$, it is the analogue of this expression that applies to a partially inbred population.

A MORE GENERAL EFFECTIVE POPULATION NUMBER

In what follows, I define a successful gamete to be one that contributes to a surviving adult. As when there is a completely random mating population, it is possible to replace N in the foregoing calculation by a more general expression N' that depends upon the distribution of the number of successful gametes emanating from a single individual.

Let G and I, respectively, denote the number of successful gametes produced by a particular adult and the number of offspring this adult produces by self-fertilization. Then, given $G = g$ and $I = i$, the number of ordered pairs of successful gametes from one adult that contribute to random separate adults is $g(g-1) - 2i$. Therefore, if we average over the distribution of I, we obtain $g(g-1-\beta)$ because the conditional expectation of $2I$ is $g\beta$.

Let us now assume that there is a fixed number βN of adults that arise from self-fertilization. There are then $2N$ ways to choose the first of a pair of successful gametes and $2N - 1 - N\beta$ ways to choose a second that is not a participant in a self-fertilization. The reason that βN is subtracted is that this is the number of gametes that are of the opposite sex to the one first chosen and are used in the formation of selfed progeny. Therefore the number of ordered pairs of gametes that contribute to separate adults is $2N(2N-1-\beta N)$. It then follows that if the number of progeny from selfing is a fixed number, the probability that two gametes forming random separate adults come from the same parent is equal to

$$\frac{1}{N'} = \frac{N}{2N(2N-1-N\beta)}\sum_g P(G=g)g(g-1-\beta) = \frac{1}{4N(1-\beta/2)-2}E[G(G-1-\beta)].$$

Since $E(G) = 2$ because of our assumption that all alleles are neutral, we may also write

$$N' = \frac{4N(1-\beta/2)-2}{\mathrm{Var}(G) + 2(1-\beta)}. \quad (18)$$

If N' is large we may use it in place of N in (7) and (10) because it is then practically certain that a third or fourth random gamete came from different parents than two gametes that come from the same parent. Since the second of equations (1) may be written as

$$\theta_{t+1} = (1 - u)^2 \left[\frac{1}{2N'}(1 + F_t) + \left(1 - \frac{1}{N'}\right)\theta_t\right],$$

even when N is small, all the results in the foregoing section still apply if N is replaced everywhere by N'.

If there is a random number of progeny S from selfing, $E(S) = \beta N$. Thus (18) still applies if it is considered to refer, when N is large, to the asymptotic equality of its right side to the reciprocal of the probability that two gametes contributing to random separate adults came from the same parent. The second of equations (1) then becomes an approximate equation when N is large. Therefore, in summary, it has been shown that

$$N_e = \frac{N'}{1 + F_{IS}} = \frac{N'}{1 + \beta/(2 - \beta)}, \tag{19}$$

where N' is given by (18).

We conclude this section by noting that if $N' = N$, $\sigma_G^2 = 2(1 - N^{-1}) = 2N \times N^{-1}(1 - N^{-1})$. One way to get this variance is for there to be a binomial offspring distribution with $2N$ trials and a probability of success equal to N^{-1} at each trial.

THE PROBABILITY OF SURVIVAL OF A SLIGHTLY ADVANTAGEOUS MUTANT

The following formulation of the problem is due to ESHEL (1984). He presented it as a special case of a general model for the growth of a line, descended from a single individual, which has a slight tendency to increase in size in the long run. There is one error in his calculations, which will be pointed out. In addition, I shall generalize the expression for the approximate probability of survival, so that it holds even if $N' \neq N$.

We assume, to begin with, that there is a large population in which a mutant allele A is originally present in one individual of genotype $A\bar{A}$, where $\bar{A}$ represents any allele that differs from A. If there is not strong selection in favor of the mutant allele, its fate is almost certain to be determined in early generations, while it is rare. Thus any individual that has at least one A in its genotype is almost certain to mate with an individual of type $\bar{A}\bar{A}$ if its does not reproduce by selfing, and it is extremely unlikely that descendants of two separate individuals carrying A in generation t will mate. Hence we may say that, at least approximately, each individual carrying A at time t gives rise to a line that develops independently of lines descended from other individuals of the same generation. Since selfing is assumed to occur with a nonnegligible probability, each of such lines can contain individuals of types AA and $A\bar{A}$. Therefore we may model the process of change in the numbers of AA and $A\bar{A}$ individuals by a two dimensional branching process.

We assume that if there were no selection each individual would produce, on the average, two gametes that succeed in forming zygotes that survive to adulthood, regardless of how it reproduces. The elements of the first moment matrix of the branching process are then derived as follows. We note first that an AA individual has AA offspring if it reproduces by selfing and $A\bar{A}$ offspring if it reproduces by random mating. Second, an individual of type $A\bar{A}$ respectively produces arrays of offspring 1/4 AA + 1/2 $A\bar{A}$ + 1/4 $\bar{A}\bar{A}$ and 1/2 $A\bar{A}$ + 1/2 $\bar{A}\bar{A}$, depending upon whether it undergoes selfing or mates randomly.

Now let AA and $A\bar{A}$ be defined to be of types 1 and 2, respectively, and Y_{ij} be the number of offspring of type j produced by a parent of type i. Then if the probabilities of survival of AA, $A\bar{A}$ and $\bar{A}\bar{A}$ zygotes to adulthood are in the proportions $1 + s_1$: $1 + s_2$: 1, the means m_{ij} of the random variables Y_{ij} are

$$\begin{aligned} m_{11} &= \beta(1 + s_1) \\ m_{12} &= 2(1 - \beta)(1 + s_2) \\ m_{21} &= \tfrac{1}{4}\beta(1 + s_1) \\ m_{22} &= \tfrac{1}{2}\beta(1 + s_2) + 2(1 - \beta)\tfrac{1}{2}(1 + s_2) \\ &= (1 - \tfrac{1}{2}\beta)(1 + s_2). \end{aligned}$$

Thus the first moment matrix is

$$\mathbf{M} = \begin{bmatrix} \beta(1 + s_1) & 2(1 - \beta)(1 + s_2) \\ \tfrac{1}{4}\beta(1 + s_1) & (1 - \tfrac{1}{2}\beta)(1 + s_2) \end{bmatrix}. \tag{20}$$

A matrix like $\mathbf{M}$, which has only positive elements, has a single dominant positive characteristic root ρ. Corresponding to ρ are characteristic vectors $\mathbf{p}' = (p_1, p_2)$ and $\mathbf{v} = (v_1, v_2)'$ such that $\mathbf{p}'\mathbf{M} = \rho\mathbf{p}'$, $\mathbf{M}\mathbf{v} = \rho\mathbf{v}$, $\sum_i p_i = \mathbf{p}'\mathbf{1} = 1$ and $\sum_i p_i v_i = \mathbf{p}'\mathbf{v} = 1$.

Then if q_i is the probability of ultimate extinction of a line descended from an individual of type i, the result proved by ESHEL (1984) is that

$$1 - q_i = \frac{2(\rho - 1)v_i}{\sum_i p_i \operatorname{Var}(\sum_j v_j Y_{ij})} + o(\rho - 1). \tag{21}$$

In the special case that we are considering,

$$\begin{aligned} \rho - 1 &= \frac{s_1\beta + 2s_2(1 - \beta)}{2 - \beta} + O(s_1, s_2), \\ \mathbf{p}' &= \left(\frac{\beta}{4 - 3\beta}, \frac{4(1 - \beta)}{4 - 3\beta}\right) + O(s_1, s_2) \\ \mathbf{v} &= \left(\frac{4 - 3\beta}{2 - \beta}, \frac{4 - 3\beta}{4 - 2\beta}\right)' + O(s_1, s_2) \end{aligned}$$

in my notation. The expression in the denominator of (21) can be shown to be equal to

$$\frac{v_2}{2(2-\beta)}\,[\beta\ \mathrm{Var}(2Y_{11}+Y_{12}) + 4(1-\beta)\ \mathrm{Var}(2Y_{21}+Y_{22})] + O(s_1, s_2).$$

Now if G is equal to the number of successful gametes produced by any particular individual when there is no selection, $\mathrm{Var}(2Y_{11}+Y_{12}) = \mathrm{Var}(G)$. Furthermore, regardless of whether an $A\bar{A}$ individual reproduces by selfing or random mating, half its successful gametes are expected to be of type A if there were no selection. Thus, given that $G = g$, $2Y_{21} + Y_{22}$ would, with no selection, be a binomially distributed random variable associated with g trials and a probability of success equal to ½. Therefore

$$\mathrm{Var}(2Y_{21}+Y_{22}) = E_G[\mathrm{Var}(2Y_{21}+Y_{22})\,|\,G)] + \mathrm{Var}_G[E(2Y_{21}+Y_{22})\,|\,G)] \doteq E_G(\tfrac{1}{4}\,G) + \mathrm{Var}_G(\tfrac{1}{2}\,G) = \tfrac{1}{4}\,[2+\mathrm{Var}(G)].$$

This implies that the expression in the denominator of (21) is approximately equal to

$$\frac{v_2}{2(2-\beta)}\,[\mathrm{Var}(G) + 2(1-\beta)].$$

Hence

$$1 - q_2 = \frac{4[\beta s_1 + 2(1-\beta)s_2]}{\mathrm{Var}(G) + 2(1-\beta)} + o(s_1, s_2). \tag{22}$$

Since, however, we are assuming that N is large, a comparison of expressions (18) and (22) reveals that

$$\begin{aligned} 1 - q_2 &= \frac{2}{2-\beta}\,\frac{N'}{N}\,[\beta s_1 + 2(1-\beta)s_2] \\ &\quad + o(s_1, s_2, N^{-1}) \\ &= 2\,\frac{N'}{N}\,[F_{IS}s_1 + (1-F_{IS})s_2] \\ &\quad + o(s_1, s_2, N^{-1}) \end{aligned} \tag{23}$$

is the probability that A survives if it was initially present in one individual of type $A\bar{A}$.

The form of the right-hand side of (23) is not surprising if we recall two facts. First, a partially selfed population that has the initial genotypic array $\sum_i\sum_j p_ip_jA_iA_j$ rapidly approaches one with the genotypic array $F_{IS}\sum_i p_iA_iA_i + (1-F_{IS})\sum_i\sum_j p_ip_jA_iA_j$. We have seen, in the discussion concerning the F-statistics, that this is true even if the population is finite. Thus, if a rare allele A has a frequency p, the ratio of the frequencies of AA and $A\bar{A}$ individuals is

$$\frac{F_{IS}p + (1-F_{IS})p^2}{2(1-F_{IS})p(1-p)} \doteq \frac{F_{IS}}{2(1-F_{IS})}.$$

The second fact is that the ratio of the survival probability of a line descended from an individual of type AA to that of a line descended from an individual of type $A\bar{A}$ is approximately $v_1/v_2 = 2$. It is thus natural that $1 - q_2$ should be proportional to an average of s_1 and s_2, with weights F_{IS} and $1 - F_{IS}$.

In the special case in which there is a Poisson distribution of offspring, $\mathrm{Var}(G) + 2(1-\beta) = 4 - 2\beta$, $N' = N$ and, by (22),

$$\begin{aligned} 1 - q_2 &= \frac{4[\beta s_1 + 2(1-\beta)s_2]}{4 - 2\beta} + o(s_1s_2, N^{-1}) \\ &= 2(\rho - 1) + o(s_1, s_2, N^{-1}). \end{aligned}$$

ESHEL (1984) mistakenly assumed in this case that the denominator of the first term in (21) is equal to 1, so that he wrote $2(\rho - 1)v_2$ in place of $2(\rho - 1)$.

Now let us suppose that $\mathrm{Var}(G)$ is not necessarily equal to 2. It is then easy to show from (22) that $1 - q_2$ is an increasing function of β is $s_1/2s_2 > \mathrm{Var}(G)/[\mathrm{Var}(G) + 2]$ and a decreasing function if this inequality is reversed. So, for example, if $s_1 > s_2 = 0$, so that A is a recessive favorable allele, $1 - q_2$ increases with β. This is also true if $s_1 = 2s_2$, in which case $A\bar{A}$ has half as much of a selective advantage as AA in comparison with $\bar{A}\bar{A}$. In contrast, $1 - q_2$ is a decreasing function of β if A is a competely dominant favorable allele and $\mathrm{Var}(G) > 2$.

Now let us suppose that $s_1 > 0$ and $s_2 = -k_2 < 0$, so that $A\bar{A}$ heterozygotes are inferior in viability to individuals of genotypes AA or $\bar{A}\bar{A}$. Expression (23) then implies that the survival probability $1 - q_2$ is positive if $s_1\beta > 2(1-\beta)k_2$ and increases with N'/N if this inequality holds. Thus the conditions most favorable to the establishment of an allele A with disadvantageous heterozygotes are that β be large and $\mathrm{Var}(G)$ small.

THE EFFECTIVE POPULATION NUMBER WHEN THERE IS PARTIAL FULL SIB MATING

It is tempting to assume that an expression of the same form as (19) always remains valid if the partial inbreeding occurs in some other way than by selfing. This is false, as can be shown by making use of an expression derived by COCKERHAM (1969) for the variance of the frequency of an allele among samples of N individuals drawn from populations with positive coefficients of inbreeding and kinship.

Let us assume that the distribution of the number of successful gametes produced by an individual is binomial, so that, in particular, $N' = N$ if there is partial selfing. Then, if p is the frequency of an allele A among parents and N is the number of offspring, the variance of the mean frequency $\hat{p}$ of A among the offspring is $p(1-p)/2N$ if there is no inbreeding and no relationship between separate individuals. More generally, it follows from a formula derived by COCKERHAM (1969) that

$$\mathrm{Var}(\hat{p}) = p(1-p)\left[\frac{1+F}{2N} + \left(1 - \frac{1}{N}\right)\theta\right],$$

where F is the mean of the inbreeding coefficients and θ is the mean of the coefficients of kinship among the $N(N-1)/2$ pairs of separate offspring.

If there is partial selfing, we can replace F and θ by F_{IS} and 0, the values they would attain at equilibrium if N were infinite. This leads to the variance effective population number

$$N_{eV} = \frac{N}{1 + F_{IS}},$$

which is consistent with (6). More generally, it should still be legitimate to replace F by the equilibrium value F_{IS} of the inbreeding coefficient in an infinite population because inbreeding occurs more rapidly than the loss of variability from random genetic drift. There is no reason, however, to expect θ to be replaceable by zero in all cases because a substantial fraction of separate parents may be related under some systems of partial inbreeding. Let β be the probability that relatives mate and θ_{1t} be the average coefficient of kinship in generation t of the mating relatives. We assume that $\beta \gg N^{-1}$. Infinite population theory then leads us to suspect that $\beta\theta_{1t} = F_{t+1} = F_{IS}$, whereas θ_{2t}, the coefficient of kinship among mates that are not relatives, is negligible in comparison with θ_{1t}.

If, for example, there are $N_p = N/2$ permanent couples in every generation, βN_p pairs of parents in each generation are expected to be full sibs. Therefore

$$\frac{2(N-1)}{2N}\theta \doteq \frac{2(N-1)}{2N}\,\frac{\beta N\theta_{1t}}{N(N-1)} = \frac{2\beta}{2N}\theta_{1t} = \frac{2F_{IS}}{2N}$$

if N is large and $\beta \gg N^{-1}$, so that

$$\operatorname{Var}(\hat{p}) \doteq \frac{p(1-p)}{2N}(1 + 3F_{IS}).$$

Hence the variance effective population number is

$$N_{eV} \doteq \frac{N}{1 + 3F_{IS}},$$

which is not consistent with (6).

This expression for N_e can also be derived from inbreeding theory. The recurrence equation for $P_t = 1 - F_t$ that was derived from WRIGHT (1951) is, in the notation of this paper,

$$P_t = \left(1 + \frac{\beta}{2} - \frac{1}{N_p}\right)P_{t+1} + \left(\frac{1}{2N_p} - \frac{\beta}{4}\right)P_{t-2} + \left(\frac{1}{4N_p} - \frac{\beta}{4}\right)P_{t-3}.$$

The corresponding characteristic equation is

$$g(\lambda) = \lambda^3 - \left(1 + \frac{\beta}{2} - \frac{1}{N_p}\right)\lambda^2 - \left(\frac{1}{2N_p} - \frac{\beta}{4}\right)\lambda - \left(\frac{1}{4N_p} - \frac{\beta}{4}\right) = 0$$

and its roots are $\lambda_1 = 1$, $\lambda_2 = (\beta + \sqrt{\beta^2 + 4\beta})/4$ and $\lambda_3 = (\beta - \sqrt{\beta^2 + 4\beta})/4$ when N_p is infinite. If N_p is finite but large and $\beta \gg N^{-1}$, the largest characteristic root is

$$\lambda_1 \doteq 1 - \frac{g(1)}{g'(1)} = 1 - \frac{1}{4N_p(1 - 3\beta/4)}.$$

Hence the inbreeding effective population number is

$$N_{eI} \doteq 2N_p\left(1 - \frac{3\beta}{4}\right) = \frac{N}{1 + 3\beta/(4 - 3\beta)} = \frac{N}{1 + 3F_{IS}}.$$

DISCUSSION

The results in this paper suggest that the usual theory for neutral alleles essentially carries over to partially selfing populations if (19) is used in place of $N_e = (4N - 2)/(\sigma_G{}^2 + 2)$. It seems that the only changes that need to be made are that inbreeding coefficients and homozygosity must respectively be replaced by coefficients of kinship of random separate individuals and identity in state of two copies of a gene if one copy is in each of two such individuals. If there is selection, however, (23) differs in two ways from the approximate expression for the survival probability that applies when there is completely random mating. First, as has been mentioned in the introduction, there is a linear combination of s_1 and s_2 rather than a multiple of s_2. Second, there is a multiplier $2(N'/N)$ in place of $2(N_e/N)$ and $N' \neq N_e$ if $\beta > 0$.

The results in the section on the F-statistics suggest how a complete theory for finite partially self-fertilizing populations could be developed. What they reflect is the fact that if $\beta \gg N^{-1}$, the loss of genetic variability from inbreeding is occurring much more rapidly than is the corresponding loss associated with random genetic drift. Consequently, the inbreeding coefficient is essentially the same in a short time as F_{IS}, the equilibrium value of the inbreeding coefficient in an infinite population. Thus, we have a stochastic process with two time scales, and it is reasonable to suppose that the diffusion approximation theory developed for such processes by ETHIER and NAGYLAKI (1980) can be shown to be applicable. The results for neutral alleles give support to this conjecture because, if there were only random mating, expressions (4), (5) and (17) still hold and they can also be derived from the diffusion approximation. I suspect that diffusion theory can also be developed when there is selection if the genotypic array among the generation of parents before selection is taken to be approximately of the form $F_{IS}\sum_i p_iA_iA_i + (1 - F_{IS})\sum_i\sum_j p_ip_jA_iA_j$. This approximation should be good if $\beta \gg 0$ and selection is weak.

I see no reason to doubt that we would also have two times scales with other systems of partial inbreeding if there is a substantial probability that relatives mate. If, however, there is some other system of inbreeding than partial selfing, it has been shown that it is no longer necessarily true that $N_e \doteq N/(1 + F_{IS})$ when there are binomial offspring distributions. Consequences of partial sib mating will be discussed in more detail elsewhere.

LITERATURE CITED

COCKERHAM, C. C., 1969 Variance of gene frequencies. Evolution **23:** 72–84.

COCKERHAM, C. C., 1971 Higher order probabilities of identity of alleles by descent. Genetics **69:** 235–246.

ESHEL, I., 1984 On the survival probability of a slightly advantageous mutant in a multitype population: a multidimensional branching process model. J. Math. Biol. **19:** 201–209.

ETHIER, S. N. and T. NAGYLAKI, 1980 Diffusion approximations of Markov chains with two time scales and applications to Population Genetics. Adv. Appl. Probab. **12:** 14–49.

HALDANE, J. B. S., 1924 A mathematical theory of natural and artificial selection. Part I. Trans. Camb. Phil. Soc. **23:** 19–41.

KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution.* Cambridge University Press, New York.

LI, C. C., 1976 *First Course in Population Genetics.* The Boxwood Press, Pacific Grove, Calif.

LI, W.-H. and M. NEI, 1975 Drift variances of heterozygosity and genetic distance in transient states. Genet. Res. **25:** 229–248.

POLLAK, E., 1972 Some effects of two types of migration on the survival of a gene. Biometrics **28:** 385–400.

STEWART, F. M., 1976 Variability in the amount of heterozygosity maintained by neutral mutations. Theor. Popul. Biol. **9:** 188–201.

STROBECK, C., 1983 The use of coefficients of identity to study random drift. pp. 1–13. In: *Population Biology, Vol. 52 of Lecture Notes in Biomathematics.* Springer Verlag, New York.

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

WRIGHT, S., 1951 The genetical structure of populations. Ann. Eugen. **15:** 323–354.

Communicating editor: B. S. WEIR